

 Open access • Journal Article • DOI:10.1088/0266-5611/4/2/007

On the uniqueness of local minima for general abstract nonlinear least-squares problems — [Source link](#)

G Chavent

Published on: 01 May 1988 - Inverse Problems (IOP Publishing)

Topics: Uniqueness, Convexity, Maxima and minima, Infimum and supremum and Monotonic function

Related papers:

- [Least-Squares Estimation, Linear Programming, and Momentum: A Geometric Parametrization of Local Minima](#)
- [Identification of the coefficient in elliptic equations](#)
- [Uniform Stability of Local Extrema of an Integral Curve of an ODE of Second Order](#)
- [Generalized Local Maximum Principles for Finite-Difference Operators](#)
- [An inverse eigenvalue problem with rotational symmetry](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/on-the-uniqueness-of-local-minima-for-general-abstract-1qyjsvy0xm>

On the uniqueness of local minima for general abstract nonlinear least-squares problems

This content has been downloaded from IOPscience. Please scroll down to see the full text.

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 194.214.160.92

This content was downloaded on 01/10/2014 at 14:56

Please note that [terms and conditions apply](#).

On the uniqueness of local minima for general abstract nonlinear least-squares problems

Guy Chavent

CEREMADE, Université Paris-Dauphine, 75775 Paris Cedex 16, France
and

Institute National de Recherche en Informatique et en Automatique, Domaine de
Voluceau, Rocquencourt, BP 105, 78153 Le Chesney Cedex, France

Received 12 June 1987

Abstract. The effectiveness of the inversion of a mapping ϕ defined on a set C by nonlinear least-squares techniques relies on, among other things, the uniqueness of local minima of the least-squares criterion, which ensures that numerical optimisation algorithm will (if they do) converge toward the global minimum of the least-squares functional. We define a number γ depending only on C and ϕ which, if the size of $\phi(C)$ is not too large with respect to its curvature, will be strictly positive, thus yielding the uniqueness of all local minima having a value smaller than γ . The condition $\gamma > 0$ will require neither convexity of C nor any monotonic property of ϕ , but involves the computation of an infimum over $\partial C \times \partial C$ of first and second derivatives of ϕ . Numerical application to the estimation of two parameters in a parabolic equation will be given.

1. Introduction

Consider

E = normed vector space (norm $\|\cdot\|_E$)

F = pre-Hilbert space (scalar product $\langle \cdot, \cdot \rangle_F$)

C = closed, C^2 -path-connected subset of E

ϕ = C^2 -mapping of C into F

$z \in F$ a given point

and the optimisation problem

$$\text{find } \hat{x} \in C \text{ such that } J(\hat{x}) \leq J(x) \forall x \in C \text{ where} \quad (1.1)$$

$$\forall x \in C, J(x) = \|\phi(x) - z\|_F^2. \quad (1.2)$$

Problem (1.1) is the general least-squares setting of the problem

$$\text{find } \hat{x} \in C \text{ such that } \phi(\hat{x}) = z. \quad (1.3)$$

when the right-hand side z does not necessarily belong to the image set $\phi(C)$.

Our goal is to find conditions on C and ϕ such that

$$\begin{aligned} &\text{problem (1.1) cannot admit two distinct local minima} \\ &\text{(and hence has at most one solution), provided that the distance} \\ &\text{from } z \text{ to } \phi(C) \text{ is taken smaller than a certain number } \gamma > 0. \end{aligned} \quad (1.4)$$

In this paper we will be able to ensure the uniqueness only of the local minima of J having a value smaller than γ (propositions 3 and 4). We also seek conditions ensuring uniqueness of all local minima.

Let us now explain our motivations.

The first question: what kind of applications have motivated the author to undertake this study? The answer is: parameter estimation problems. In this application, x is the parameter, C is the set of admissible parameters, z is the observed data and ϕ is the parameter \rightarrow output mapping resulting from the resolution of the model state equations and the observation operator. Our concern is primarily with overspecified inverse problems, where $\dim F \geq \dim E$, so that we can expect that the derivative $\phi'(x)$ to be more or less injective from E into F . In order to be more specific, we can give an example.

Example 1. We consider the 1D parabolic equation

$$\begin{aligned} \frac{\partial u}{\partial t} - \frac{\partial}{\partial y} \left(a \frac{\partial u}{\partial y} \right) + bu &= f & \forall y \in]0, 1[, \quad \forall t \in]0, T[\\ u(0, t) = u(1, t) &= 0 & \forall t > 0 \\ u(y, 0) &= u_0(y) & \forall y \in [0, 1] \end{aligned} \tag{1.5}$$

when the parameters $a \in \mathbb{R}^{+*}$ and $b \in \mathbb{R}^+$ have to be estimated from the measurement $z \in L^2(0, T)$ of the solution u at point $y = \frac{1}{2}$ against time. Here we have $x = (a, b) \in \mathbb{R}^2 = E$, C is a given closed subset of $\mathbb{R}^{+*} \times \mathbb{R}^+$, which represents the *a priori* knowledge of the experimenter about the parameter $x = (a, b)$, and ϕ is the mapping which makes the $t \rightarrow u(\frac{1}{2}, t)$ function of $L^2(0, T) = F$ correspond to a given $x = (a, b) \in C$. In this example the evaluation of $\phi(x)$ involves the resolution of the parabolic equation (1.5); the problem is obviously overspecified as $\dim F = +\infty > \dim E = 2!$

The second question: why do we address the problem of uniqueness of local minima? The only way of actually solving the parameter estimation problems described above is to undertake the minimisation of J over C on a computer. However, optimisation algorithms are only able to find local minima over a closed set. Hence the least-squares problem (1.1) will be practically solvable by an optimisation algorithm as soon as C is closed and J has at most one local minimum over C .

This will ensure that the optimisation algorithm, once converged, will give the sought global minimum of J . One can also remark that the uniqueness of local minima implies (but is not equivalent to) the uniqueness of the solution \hat{x} of problem (1.1), or in terms of parameter estimation problems, the identifiability of \hat{x} from the knowledge of z and C . Of course, one other extremely important practical problem is that of the stability of the solution \hat{x} of (1.1) with respect to perturbations of the data z ; this problem will not be addressed as such in this paper, but one can remark that, when C is compact, the above uniqueness property will ensure the existence of a unique \hat{x} depending continuously on z as long as the distance of z to $\phi(C)$ is taken small enough.

The third (and last) question: what kind of conditions on C and ϕ are we looking for? The first idea is that we want data-independant conditions: for a given set C and mapping ϕ , we want to be able to decide whether property (1.4) holds or not. If it holds we will get as a by-product the upper limit $\gamma > 0$ to the distance of z from $\phi(C)$

for which the uniqueness property of local minima holds. If it does not hold the experimenter will then have to acquire more data (i.e. change the mapping ϕ) and/or augment the *a priori* available information (i.e. diminish the size of C) before checking again for property (1.4). The second idea is that such conditions will in no way be cheap! As in view of the applications, no hypothesis will be made on the shape of C and ϕ (no convexity, no monotonicity), the conditions will necessarily involve exploration all over C —which of course will require a lot of computer time as soon as the dimension of C , i.e. the number of unknown parameters, increases.

Nevertheless, we believe that such a condition will be practically useful for problems with few unknown parameters, and that it will at least help to understand what happens in nonlinear least-squares problems. As a test for the forthcoming sufficient condition for (1.4) to hold, we will add to example 1 an extremely simple example.

Example 2. Determine a real number x from the measurement (z_1, z_2) of its cosine and sine. Then we have

$$E = \mathbb{R} \quad F = \mathbb{R}^2 \quad \phi(x) = (\cos x, \sin x). \tag{1.6}$$

Of course, one has to restrict *a priori* the search for x to an interval of length smaller than 2π if we want the problem to have a chance of being well posed! So suppose we take for example

$$C = [0, X] \quad \text{with } X \text{ given, } X < 2\pi. \tag{1.7}$$

Then obviously problem (1.1) has a unique global minimum as soon as $d(z, \phi(C)) < \gamma = \sin(X/2)$ as one can see in figure 1(a) for different data z .

However, one sees also in figure 1 that there may exist, beside the global minimum, a distinct local minimum (with value larger than γ !) so that the solution of (1.1) by an optimisation algorithm may fail because condition (1.4) is not satisfied!

In order to satisfy conditions (1.4), it is sufficient to replace condition (1.7) by the stronger condition

$$C = [0, X] \quad \text{with } 0 < X < \pi. \tag{1.8}$$

Then, as seen in figure 1(b), condition (1.4) holds when

$$d(z, \phi(C)) < \sin X. \tag{1.9}$$

Conditions (1.8) plus (1.9) are clearly equivalent to (1.4), and will be used as a benchmark to indicate the precision of the condition that we will derive.

To conclude this introduction, we will recall a previous result of Spiess (1969), who considered exactly the *same problem*, namely the uniqueness of local minima of problem (1.1), but set *on an open and convex set* C . In fact he gave data-dependent sufficient conditions, i.e. conditions which, for a given datum z and a given local minimum \hat{x} , imply that \hat{x} is a global minimum. These conditions, when translated into data-independent conditions, read as follows:

Spiess conditions. If C is an open convex subset of E , ϕ is injective and C^2 over C

$$\gamma = \inf_{x,y \in C} \inf_{z \in [xy]} \frac{\|\phi'(z)(y-x)\|^2}{\|\phi''(z)(y-x, y-x)\|^2} \times \frac{\|\phi(x) - \phi(y)\|}{\|\phi(x) - \phi(z)\| + \|\phi(z) - \phi(y)\|} > 0 \tag{1.10}$$

then J has at most one local minimum over the open set C as soon as $d(z, \phi(C)) < \gamma$.

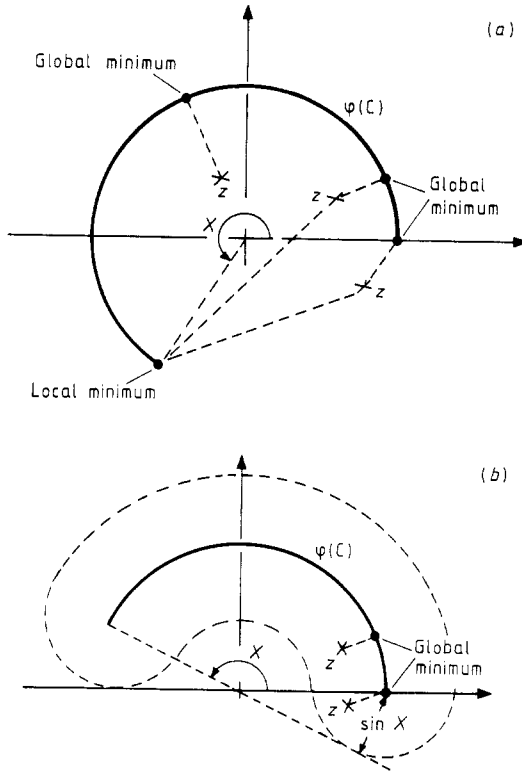


Figure 1. Determination of $x \in [0, X]$ from the measurement z of $(\cos x, \sin x)$. (a) $X < 2\pi$; (b) $X < \pi$.

If we apply this condition to the simple example 2, where we take now for convenience

$$C =]\varepsilon, 2\pi - \varepsilon[\quad \varepsilon > 0 \text{ given} \tag{1.11}$$

one checks very easily that

$$\frac{\|\phi'(z)(y-x)\|^2}{\|\phi''(z)(y-x, y-x)\|} = 1 \quad \text{for any } x, y, z$$

and that

$$\frac{\|\phi(x) - \phi(y)\|}{\|\phi(x) - \phi(z)\| + \|\phi(z) - \phi(y)\|} \quad \text{is minimum for } x, y, z \text{ as in figure 2.}$$

Hence we get

$$\gamma = \frac{\sin \varepsilon}{\sqrt{2(1 + \cos \varepsilon)}} \sim \frac{1}{2} \sin \varepsilon \quad \text{for small } \varepsilon. \tag{1.12}$$

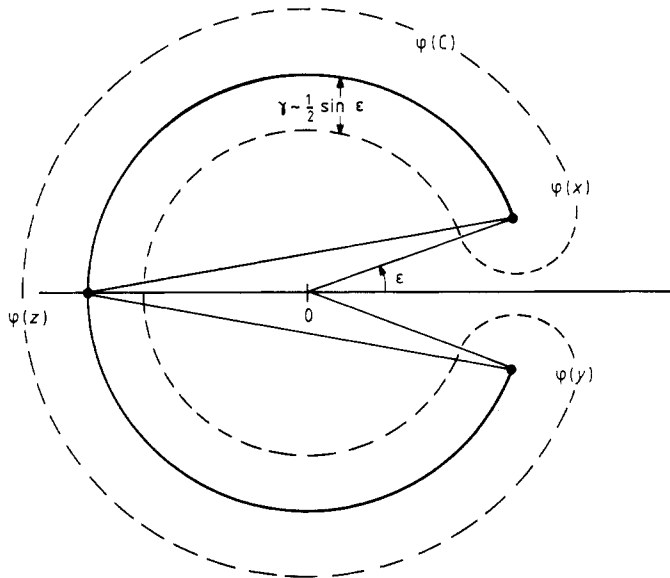


Figure 2. Application of the Spiess condition to example 2.

So γ is strictly positive, and the sufficient condition is satisfied, in this example, for all $\epsilon > 0$.

Of course, as C is taken open, the Spiess condition does not eliminate the local minima which may arise on $\phi(\partial C)$ (as in figure 1(a)), so that this condition does not answer our second question. However, it may give a reasonable idea of the kind of condition we are going to derive below, as they share the property of containing an infimum over a couple of points (x, y) of C and over a path (here the $[x, y]$ interval) connecting them.

Let us now be more technical and turn to the derivation of our sufficient condition. The hypothesis and notation given at the beginning of the introduction will hold throughout the rest of the paper and will not be repeated.

2. How to recognise the existence of two distinct local minima

Let $x, y \in C, x \neq y$, be two such local minima (see figure 3). Using the hypothesis that C is C^2 path connected, we may choose one C^2 path going from x to y , i.e. one C^2 mapping $s: \theta \rightarrow s(\theta)$ from some $[\theta_0, \theta_1]$ interval of \mathbb{R} in E , satisfying

$$\begin{aligned} s(\theta_0) &= x & s(\theta_1) &= y \\ s(\theta) &\in C & \forall \theta \in [\theta_0, \theta_1]. \end{aligned} \tag{2.1}$$

We consider now the function $f: [\theta_0, \theta_1] \rightarrow \mathbb{R}$ defined by

$$f(\theta) = \|\phi(s(\theta)) - z\|^2 \quad \forall \theta \in [\theta_0, \theta_1] \tag{2.2}$$

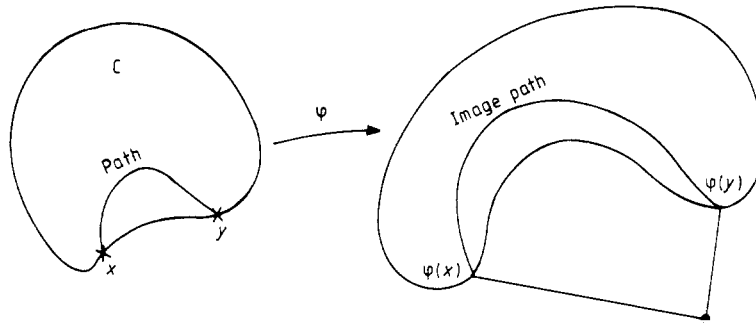


Figure 3.

which we have depicted in figure 4, in the case where $f(\theta_0) = \|\phi(x) - z\|^2 \geq f(\theta_1) = \|\phi(y) - z\|^2$.

From the properties of x and y , it is clear that one can find θ'_1 such that

$$\begin{aligned} \theta_0 < \theta'_1 \leq \theta_1 \\ f(\theta) &\geq f(\theta_0) = \|\phi(x) - z\|^2 \quad \forall \theta \in [\theta_0, \theta'_1] \\ f(\theta'_1) &= f(\theta_0) = \|\phi(x) - z\|^2. \end{aligned} \tag{2.3}$$

From here two cases may occur:

(i) $f(\theta) = f(\theta_0) \quad \forall \theta \in [\theta_0, \theta'_1]$

then of course $f''(\theta) = 0 \Rightarrow f''(\theta) \leq 0$.

(ii) $f(\theta) > f(\theta_0)$ for some subinterval of $[\theta_0, \theta'_1]$, so that there exists some $\tilde{\theta}$ in this interval such that $f''(\tilde{\theta}) \leq 0$.

So if we set

$$y' = s(\theta'_1) \tag{2.4}$$

we obtain the following proposition.

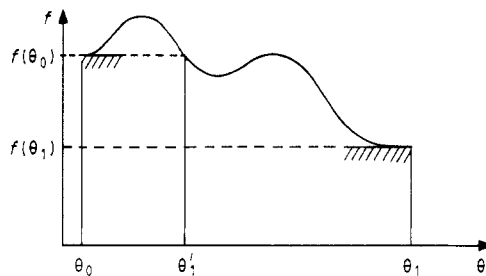


Figure 4.

Proposition 1.

If

$$z \in F$$

$$x, y = \text{local minima of } J \text{ over } C, x \neq y (\|\phi(x) - z\|^2 \geq \|\phi(y) - z\|^2)$$

$$s: [\theta_0, \theta_1] \rightarrow C = \text{path from } x \text{ to } y.$$

Then there exists

$$\theta'_1 \in]\theta_0, \theta_1] \text{ such that } f(\theta'_1) = f(\theta_0) = \|f(x) - z\|^2$$

$$\tilde{\theta} \in]\theta_0, \theta'_1[\text{ such that } f''(\tilde{\theta}) \leq 0.$$

3. Given $z \in F$, $x, y' \in C$ such that $\|\phi(x) - z\| = \|\phi(y') - z\| = 0$ and a path s from x to y' , what does $f''(\tilde{\theta}) < 0$ imply?

Let us first introduce some quantities related to the image-path $\{\phi(s(\theta)), \theta \in [\theta_0, \theta_1]\}$ in F :

$$v(\theta) = \phi'(s(\theta)) \cdot s'(\theta) \quad = \text{velocity} \tag{3.1}$$

$$a(\theta) = [\phi'(s(\theta)) \cdot s'(\theta)]' \quad = \text{acceleration.}$$

Then the first and second derivatives of $f(\theta)$ can be expressed as

$$f'(\theta) = 2\langle \phi(s(\theta)) - z, v(\theta) \rangle \tag{3.2}$$

$$f''(\theta) = 2\|v(\theta)\|^2 + 2\langle \phi(s(\theta)) - z, a(\theta) \rangle. \tag{3.3}$$

From (3.2) we get

$$|f'(\theta)| \leq 2f(\theta)^{1/2}\|v(\theta)\| \tag{3.4}$$

which together with (3.3) and the Cauchy–Schwarz inequality yield

$$f'(\theta) \geq \frac{f'(\theta)^2}{2f(\theta)} - 2f(\theta)^{1/2}\|a(\theta)\|. \tag{3.5}$$

We now define a function $g: [\theta_0, \theta'_1] \rightarrow \mathbb{R}$ by the following 1D elliptic problem:

$$\begin{aligned} -g''(\theta) &= \|a(\theta)\| & \forall \theta \in [\theta_0, \theta'_1] \\ g(\theta_0) &= g(\theta'_1) = 0. \end{aligned} \tag{3.6}$$

We may remark that this function g is independent of z (whereas f was not!), and that it is a *positive concave function*.

Plugging (3.6) into (3.5) then yields

$$\frac{f''(\theta)}{2f(\theta)^{1/2}} - \frac{f'(\theta)^2}{4f(\theta)^{3/2}} - g''(\theta) \geq 0$$

i.e.

$$\frac{d^2}{d\theta^2} (f(\theta)^{1/2} - g(\theta)) \geq 0 \tag{3.7}$$

which proves that the function $\theta \rightarrow f(\theta)^{1/2} - g(\theta)$ is convex, and hence, as $f(\theta_0) = f(\theta_1) = d$ (where d is the common value of $\|\phi(x) - z\|$ and $\|\phi(y') - z\|$),

$$f(\theta)^{1/2} \leq g(\theta) + d \quad \forall \theta \in [\theta_0, \theta_1]. \tag{3.8}$$

But on the other hand, from $f''(\tilde{\theta}) \leq 0$ we get, using (3.3) and the Cauchy-Schwartz inequality,

$$0 \geq f''(\tilde{\theta}) \geq 2\|v(\tilde{\theta})\|^2 - 2f(\tilde{\theta})^{1/2}\|a(\tilde{\theta})\|$$

and hence

$$f(\tilde{\theta})^{1/2} \geq \frac{\|v(\tilde{\theta})\|^2}{\|a(\tilde{\theta})\|}. \tag{3.9}$$

From (3.8) and (3.9) we then get

$$d \geq \frac{\|v(\tilde{\theta})\|^2}{\|a(\tilde{\theta})\|} - g(\tilde{\theta}). \tag{3.10}$$

So we have proved the following proposition.

Proposition 2.

If

$$\begin{aligned} z &\in F \\ x, y' &\in C \text{ such that } \|\phi(x) - z\| = \|\phi(y') - z\| = d \\ s: [\theta_0, \theta_1] &\rightarrow C \text{ a path from } x \text{ to } y' \end{aligned}$$

then

$$\tilde{\theta} \in [\theta_0, \theta_1], f''(\tilde{\theta}) \leq 0 \Rightarrow d \geq \frac{\|v(\tilde{\theta})\|^2}{\|a(\tilde{\theta})\|} - g(\tilde{\theta}).$$

4. A family of sufficient conditions for the uniqueness of certain local minima of problem (1.1)

Suppose we have chosen a strategy S in order to associate to every couple (x, y) of points of C a C^2 -path s from x to y :

$$S: (x, y) \in C \times C \quad s = S(x, y): [\theta_0, \theta_1] \rightarrow CD \text{ such that } \begin{cases} s(\theta_0) = x \\ s(\theta_1) = y \\ s \text{ is } C^2. \end{cases}$$

Then from propositions 1 and 2 we get immediately the following sufficient condition.

Proposition 3. (sufficient condition associated to the strategy S). Suppose that

$$\gamma = \inf_{x, y \in C} \inf_{\substack{s = S(x, y) \\ \theta \in [\theta_0, \theta_1]}} \left(\frac{\|v(\theta)\|^2}{\|a(\theta)\|} - g(\theta) \right) > 0. \tag{4.1}$$

Then the problem (1.1) has at the most one local minimum with value smaller than γ as soon as $d(z, \phi(C)) < \gamma$.

This condition does not look very useful. But before simplifying it somewhat and indicating which strategy S to choose, let us explain its meaning with a simple example.

Example 3. Suppose that

$$\begin{aligned} & C \text{ is convex (and hence } C^2\text{-path-connected!)} \\ & \phi \text{ is such that numbers } \alpha > 0, \beta > 0 \text{ exist with} \\ & \|\phi''(x) \cdot y\| \geq \alpha \|y\| \quad \forall x \in C, \quad \forall y \in E \\ & \|\phi''(x)(y, y)\| \leq \beta \|y\|^2 \end{aligned} \tag{4.2}$$

$$S(x, y) = s \quad \text{defined by} \quad \begin{aligned} & \theta_0 = 0, \theta_1 = 1 \\ & s(\theta) = x + \theta(y - x). \end{aligned} \tag{4.3}$$

Then

$$\begin{aligned} & \|v(\theta)\| \geq \alpha \|y - x\| \\ & \|a(\theta)\| \leq \beta \|y - x\|^2 \end{aligned}$$

and hence, from the maximum principle,

$$g(\theta) \leq \frac{1}{2}\beta \|y - x\|^2 \theta(1 - \theta) \leq \frac{1}{8}\beta \|y - x\|^2$$

and

$$\frac{\|v(\theta)\|^2}{\|a(\theta)\|} - g(\theta) \geq \frac{\alpha^2}{\beta} - \frac{\beta}{8} \|y - x\|^2.$$

Then from proposition 3 we get the following (weaker) sufficient condition.

Proposition 4. Suppose that (4.2) holds and that

$$\gamma = \inf_{(x,y) \in C} \left(\frac{\alpha^2}{\beta} - \frac{\beta}{8} \|y - x\|^2 \right) > 0 \tag{4.4}$$

or equivalently

$$\beta \text{ diam } C < 2\sqrt{2} \alpha \quad \text{and} \quad \gamma = \frac{\alpha^2}{\beta} - \frac{\beta}{8} (\text{diam } C)^2.$$

Then the problem (1.1) has at most one local minimum with value smaller than γ as soon as $d(z, \phi(C)) < \gamma$.

This result was already given in Chavent (1983) together with a Lipschitz continuity result of the $z \rightarrow \hat{x}$ mapping and, in the case where E is a Banach space, an existence result for \hat{x} .

However, the estimation (4.4), which involves upper and lower bounds, over all $x \in C$ and for all directions y is very rough, and may yield conditions on the size of C

that are too restrictive for practical use. A condition intermediate between (4.1) and (4.4) would be

$$\gamma = \text{Inf}_{(x,y) \in C} \left\{ \text{Inf}_{\substack{s=S(x,y) \\ \theta \in [\theta_0, \theta_1]}} \frac{\|v(\theta)\|^2}{\|a(\theta)\|} - \text{Max}_{\substack{s=S(x,y) \\ \theta \in [\theta_0, \theta_1]}} g(\theta) \right\} > 0. \tag{4.5}$$

We come back to the less constraining estimation (4.1) of proposition 3.

5. Choice of a strategy S

The problem is now to choose the strategy S, which associates to any couple $(x, y) \in C \times C$ a C^2 path s from x to y , in such a way that the number γ defined by (4.1) is the largest possible (and hence the ‘size’ condition on C the least restrictive possible).

For given $x, y \in C$, the choice of a path going from x to y can be conceptually split into two steps: (i) choose the geometry of the path; (ii) choose the time law, i.e. the parametrisation of the path. We will choose these two items separately.

5.1. For given $x, y \in C$ and a given geometry of a path from x to y , how should the time law be chosen?

We consider first a particular parametrisation $\hat{s}(\hat{\theta})$ of the path from x to y where $\hat{\theta}$ is the curvilinear abscissa on the image path $\phi \circ s(\theta)$.

Such a parametrisation satisfies, by definition

$$\|\hat{v}(\hat{\theta})\| = \|\phi'(\hat{s}(\hat{\theta})) \cdot s'(\hat{\theta})\| = 1 \tag{5.1}$$

and will exist as soon as $\phi'(x)$ is injective everywhere over C. At points where $\phi'(x)$ is not injective, $\hat{v}(\hat{\theta})$ may still exist, but $s'(\hat{\theta})$ will have to be infinite.

By deriving (5.1) we find, as usual, that the velocity $\hat{v}(\hat{\theta})$ and the acceleration $\hat{a}(\hat{\theta})$ are orthogonal on the image path:

$$\hat{a}(\hat{\theta}) = [\phi'(\hat{s}) \cdot \hat{s}']' \perp \hat{v}(\hat{\theta}) = \phi'(\hat{s}) \cdot \hat{s}' \tag{5.2}$$

We consider then any other parametrisation $s(\theta)$ of the same geometric path, which is necessarily of the form

$$s(\theta) = \hat{s}(\chi(\theta))$$

where

$$\chi: [\theta_0, \theta_1] \rightarrow [\hat{\theta}_0, \hat{\theta}_1]. \tag{5.3}$$

One checks easily that

$$\begin{aligned} v(\theta) &= \chi'(\theta) \hat{v}(\hat{\theta}) \\ a(\theta) &= \chi'(\theta)^2 \hat{a}(\hat{\theta}) + \hat{v}(\hat{\theta}) \chi''(\theta). \end{aligned} \tag{5.4}$$

But as $\hat{a}(\hat{\theta})$ and $\hat{v}(\hat{\theta})$ are orthogonal we get

$$\|a(\theta)\| = \{\chi'(\theta)^2 \|\hat{a}(\hat{\theta})\|^2 + \chi''(\theta)^2 \|\hat{v}(\hat{\theta})\|^2\}^{1/2} \tag{5.5}$$

and hence

$$\|a(\theta)\| \geq \chi'(\theta)^2 \|\hat{a}(\hat{\theta})\|. \tag{5.6}$$

In order to compare the numbers $\hat{\gamma}$ and γ associated by (4.1) to the two parametrisations $\hat{s}(\hat{\theta})$ and $s(\theta)$ of the same geometrical path, we compare the arguments of the inf in (4.1).

(i) Obviously one gets from (5.4) and (5.6)

$$\frac{\|v(\theta)\|^2}{\|a(\theta)\|^2} \leq \frac{\chi'(\theta)^2 \|\hat{v}(\hat{\theta})\|^2}{\chi'(\theta)^2 \|\hat{a}(\hat{\theta})\|^2} = \frac{\|\hat{v}(\hat{\theta})\|^2}{\|\hat{a}(\hat{\theta})\|^2}. \tag{5.7}$$

(ii) In order to compare $g(\theta)$ and $\hat{g}(\hat{\theta})$, we set

$$\tilde{g}(\theta) = \hat{g}(\chi(\theta)) \tag{5.8}$$

and will compare $g(\theta)$ and $\tilde{g}(\theta)$. One first checks easily that $\tilde{g}(\theta)$ satisfies the following equation:

$$-\frac{1}{\chi'(\theta)} \frac{d}{d\theta} \left(\frac{1}{\chi'(\theta)} \frac{d}{d\theta} \tilde{g}(\theta) \right) = \|\hat{a}(\theta)\| \quad \forall \theta \in [\theta_0, \theta_1] \tag{5.9}$$

$$\tilde{g}(\theta_0) = \tilde{g}(\theta_1) = 0.$$

Comparing with the equation defining $g(\theta)$,

$$-g''(\theta) = \|a(\theta)\| \quad \forall \theta \in [\theta_0, \theta_1] \tag{5.10}$$

$$g(\theta_0) = g(\theta_1) = 0$$

we get, using (5.6) and the maximum principle,

$$\text{Max}_\theta g(\theta) \geq \text{Max}_\theta \tilde{g}(\theta) \quad \forall \theta \in [\theta_0, \theta_1]. \tag{5.11}$$

Summarising the results (5.7) and (5.11) we have

$$\hat{\gamma} \geq \gamma \quad \text{for } \gamma \text{ defined by (4.5)}. \tag{5.12}$$

Conclusion. For a given geometric path going from x to y , the best parametrisation, when γ is defined by (4.5), is obtained when θ is the curvilinear abscissa on the image path. In other cases, in particular when γ is defined by (4.1), the curvilinear abscissa is at least the most intrinsic parametrisation.

In the following, we will omit the hat on s , θ , etc, and θ will always denote the curvilinear abscissa on the image path.

With this parametrisation, the formula (4.1) simplifies somewhat, and moreover gains a geometrical interpretation: now the radius of curvature $\rho(\theta)$ of the image path at point $\phi \circ s(\theta)$, since $\|v(\theta)\| = 1$ and $v(\theta) \perp a(\theta)$, is given by

$$\rho(\theta) = 1/\|a(\theta)\|. \tag{5.13}$$

Hence, when the parameter θ is chosen to be the curvilinear abscissa along the image path, (4.1) reduces to

$$\gamma = \text{Inf}_{\substack{x, y \in C \\ s=S(x, y) \\ \theta \in [\theta_0, \theta_1]}} \text{Inf}_{\theta} (\rho(\theta) - g(\theta)) > 0 \tag{5.14}$$

where $g(\theta)$ is defined by

$$-g''(\theta) = 1/\rho(\theta) \quad \text{for } \theta_0 \leq \theta \leq \theta_1 \tag{5.15}$$

$$g(\theta_0) = g(\theta_1) = 0.$$

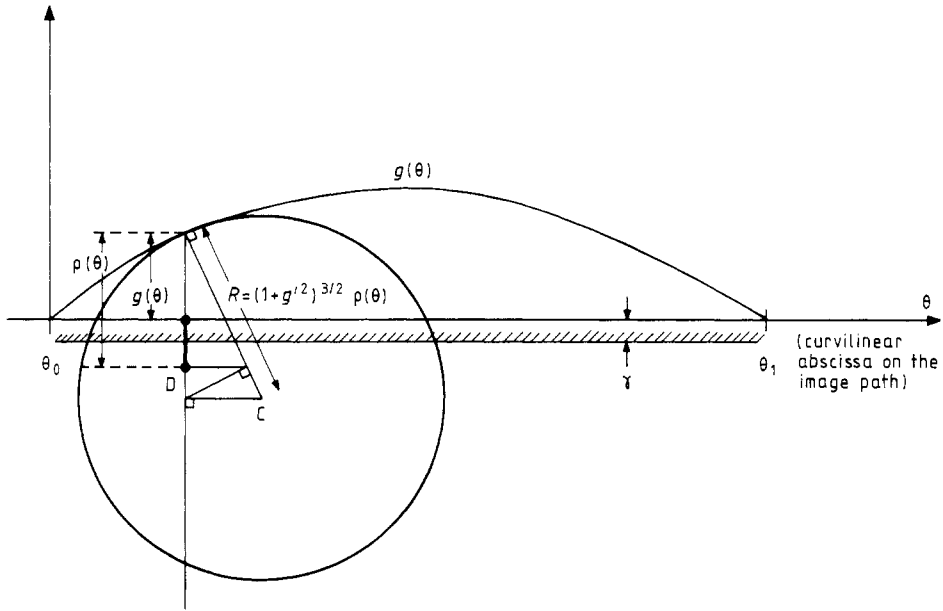


Figure 5.

We have illustrated in figure 5 a geometrical construction of γ from the data of the $\theta \rightarrow g(\theta)$ function: the point D should never get above the horizontal hatched line.

5.2. Given $x, y \in C$, given a path s from x to y , and given two points x', y' belonging to that path s , how do γ (associated to x, y) and γ' (associated to x', y') compare?

If x' and y' correspond to the parameters θ'_0 and θ'_1 of the interval $[\theta_0, \theta_1]$, it is clear from figure 6 and from the maximum principle for elliptic equations that

$$\gamma' \geq \gamma.$$

Conclusion. As soon as the strategy S chosen is stable with respect to restriction (i.e. if $S(x', y') = S(x, y)|_{[\theta'_0, \theta'_1]}$), it is sufficient, in order to calculate γ to consider only couples of the boundary ∂C of C .

Hence (4.1) or (4.14) reduces to

$$\gamma = \inf_{x, y \in \partial C} \inf_{\substack{s=S(x, y) \\ \theta \in [\theta_0, \theta_1]}} (\rho(\theta) - g(\theta)) > 0 \tag{5.16}$$

where θ is the curvilinear abscissa along the image path and where $g(\theta)$ is defined by (5.15).

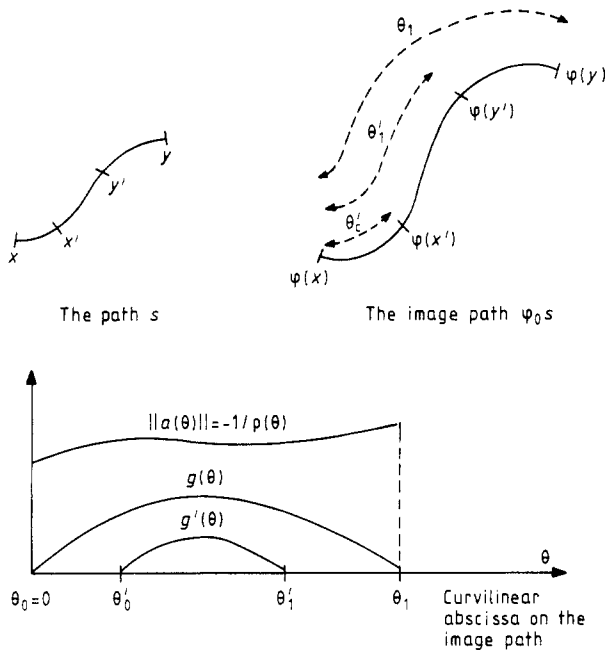


Figure 6.

5.3. How to choose the geometrical path from x to y

For a given $x, y \in \partial C$, we are now looking for a path s from x to y , which will be parametrised by the curvilinear abscissa θ along the image path $\phi \circ s$, such that the quantity

$$\text{Inf}_{\theta \in [\theta_0, \theta_1]} (\rho(\theta) - g(\theta)) \tag{5.17}$$

appearing in (4.1) or (5.14) is maximum.

The first remark is that the quantity (5.17) depends only on the geometrical properties of the image path $\phi \circ s$ going from $\phi(x)$ to $\phi(y)$: θ is the curvilinear abscissa along this path, $\rho(\theta)$ is the radius of curvature of this path and $g(\theta)$ is defined from θ_0, θ_1 and $\rho(\theta)$.

So we can replace (at least conceptually!) the task of choosing a path s from x to y in C by that of choosing a path S from $\phi(x)$ to $\phi(y)$ in $\phi(C)$ in such a way that the quantity (5.17) is maximised.

In this new setting the mapping ϕ is used only, together with the set C , for the definition of the set $\phi(C)$ in which the sought path S has to stay.

The second remark is that, whenever the segment $[\phi(x), \phi(y)]$ is fully included in $\phi(C)$ then one can choose $S = [\phi(x), \phi(y)]$, which yields $\rho(\theta) = +\infty$ and $g(\theta) = 0$, hence $\gamma = +\infty$ so that S is obviously the sought optimal solution!

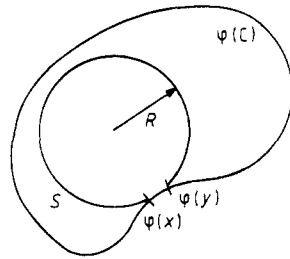


Figure 7.

The third remark is that, if one chooses a path S from $\phi(x)$ to $\phi(y)$ with both large radii of curvature and a large length $\theta_1 - \theta_0$, like the one depicted in figure 7, for which we have

$$\begin{aligned} \rho(\theta) &= R > 0 & \forall \theta \in [\theta_0, \theta_1] \\ \theta_0 &= 0, & \theta_1 \approx 2\pi R. \end{aligned}$$

Then the function g is of the form

$$g(\theta) = \frac{1}{2R} (\theta - \theta_0)(\theta_1 - \theta) \quad \forall \theta \in [\theta_0, \theta_1]$$

and is maximum at the point $\theta = \frac{1}{2}(\theta_0 + \theta_1) = \pi R$,

$$g(\pi R) \approx \frac{1}{2}\pi^2 R$$

so that

$$\rho(\pi R) - g(\pi R) \approx (1 - \frac{1}{2}\pi^2)R \leq 0.$$

So we see that paths S with both large ρ and large length are not optimal.

From the two last remarks, we may conjecture that the optimal S is the minimum-length path going from $\phi(x)$ to $\phi(y)$ in $\phi(C)$. But this still remains to be proved.

We can propose two strategies for the choice of the path s from x to y .

Strategy 1. Determine a path s from x to y in such a way that $S = \phi \circ s$ is the minimum-length path in $\phi(C)$ going from $\phi(x)$ to $\phi(y)$. This procedure may (if our conjecture is true!) yield the optimal number γ , and hence the less constraining condition on the size of C . However, from the practical point of view, such a strategy seems very difficult to implement, as one would have to solve, for each couple $x, y \in C$, a complicated optimisation problem in a high dimensional space.

Strategy 2. Choose s as the minimum length path in C going from x to y . This procedure is surely non-optimal, but will guarantee that the corresponding image path $S = \phi \circ s$ will not be too long, as soon as upper bounds on $\|\phi'(x)\|$ are available. Moreover, as the set C is defined by explicit constraints, and is usually of non-void interior, the minimum-length path in C from x to y can be determined relatively easily (in many cases it will be the $[x, y]$ interval).

To conclude this section, let us see, using the very simple example 2 of the §1 how close our final condition (5.14)–(5.15) with strategy 1 or 2 comes to the solution of this example.

We have seen in §1 (see figure 1 and formula (1.7)) that the least-squares functional for the search of a real number x from measurements of its cosine and sine had a unique local minimum with value smaller than $\gamma = \sin^2 X$ (which of course is the global minimum) as soon as we search for x in the interval $[0, X]$ with $X < 2\pi$.

If we now apply conditions (5.15) and (5.16) to this problem, we have to compute the argument of the infimum in (5.16) only for a path going from 0 to X . Obviously, the path

$$s(\theta) = \theta \quad 0 \leq \theta \leq X \tag{5.18}$$

has all the desired properties: θ is the curvilinear abscissa on the arc of the circle which is the image of the interval $[0, X]$ by the ϕ function defined in (1.6); s yields the minimal length path as well in the image set as in the parameter set so strategies 1 and 2 are equivalent here.

Along this path one has

$$\rho(\theta) = 1 \quad (\text{radius of curvature of image circle})$$

and hence

$$g(\theta) = \frac{1}{2}\theta(X - \theta)$$

which is maximum at $\theta = \frac{1}{2}X$

$$g(\theta) \leq g(\frac{1}{2}X) = \frac{1}{8}X^2.$$

Hence we get from (5.16) the condition

$$\gamma = 1 - \frac{1}{8}X^2 > 0$$

or

$$X < 2\sqrt{2} \approx 2.828$$

which is to be compared with the best possible condition $\chi < 2\pi$ exhibited for this example in the introduction. We see that the result is not too bad, but as $2\sqrt{2} < 2\pi$ we cannot conclude whether the condition (5.15)–(5.16) with strategy 1 is optimal or not. But one may remark that $2\sqrt{2} < \pi$, which proves that, for this example, our condition $\gamma > 0$ yields in fact the uniqueness of *all* local minima.

6. Numerical application

For historical reasons, the numerical application we are going to present was not made using (5.14) with the curvilinear abscissa in the data-space parametrisation, but using a weaker version of (4.1) with a time law that has a constant velocity in the parameter space. The geometry of the path going from x to y was given by strategy S_2 of §4.3 (minimum length in the parameter space) and C was taken to be convex.

Hence, for any $x, y \in C$ the path $s(\theta)$ was

$$s(\theta) = x + \theta(y - x) \quad \theta \in [0, 1]. \tag{6.1}$$

The sufficient condition for (4.1) was obtained in the following way: using the fact that, for any $v \in H_0(0, 1) = \{v \in L^2(0,1) | v' \in L^2(0,1), v(0) = v(1) = 0\}$, one has $|v(\theta)| \leq \frac{1}{2} \|v'\|_{L^2(0,1)}$, and that, from its definition (3.6) as the solution of an elliptic boundary value

problem, the function g satisfies $\|g'\|_{L^2(0,1)}^2 \leq \|g\|_{L^2(0,1)} \times \|a\|_{L^2(0,1)}$, we get the following upper bound for $g(\theta)$:

$$|g(\theta)| \leq (1/2\pi)\|a\|_{L^2(0,1)}. \tag{6.2}$$

Thus a sufficient condition for (4.1) to hold is:

$$\gamma = \text{Inf}_{x,y \in \partial C} \left(\text{Inf}_{\theta \in [0,1]} \frac{\|v(\theta)\|^2}{\|a(\theta)\|} - \frac{1}{2\pi} \|a\|_{L^2(0,1)} \right) > 0. \tag{6.3}$$

We applied the condition (6.3) to example 1 of §1. However, rather than checking, for *a priori* given admissible parameter sets C , whether condition (6.3) holds, we used an alternative approach: supposing we have been given by an engineer some nominal value $\bar{x} = (\bar{a}, \bar{b}) \in \mathbb{R}^{+*} \times \mathbb{R}^+$ of the unknown parameters, we tried to answer the question ‘how large can the parameter set C be chosen around (\bar{a}, \bar{b}) while still maintaining the uniqueness of the local minima of problem (1.1) over C ?’. This amounts to finding ‘around’ a given \bar{x} the ‘largest’ set \bar{C} for which $\gamma = 0$ so that any set C strictly included in \bar{C} will yield a strictly positive γ . This was done by computing the values of the argument of the infimum in (6.3) for segments of increasing length centred at \bar{x} and lying on a finite number of straight lines going through \bar{x} , until one reaches the zero value in each direction. At this stage, all couples $[x, 2\bar{x}-x] \in \partial C$ that are symmetrical with respect to \bar{x} were tested. Then the couples (x, y) with $y \neq x$ were tested, eventually diminishing the length of the $[x, 2\bar{x}-x]$ interval if the argument of (6.3) happens to be negative for the $[x, y]$ segment. Of course, this procedure will produce domains dependent on the order in which the (x, y) segments are tested in the second part of the algorithm.

The numerical results, taken from Charles (1985), are shown in figure 8. The interesting point to be noted is that the size of the ‘maximal’ sets given by condition (6.3) is already plausible from a practical point of view. Using condition (5.16) would yield still larger sets, with no basic increase in computational time. On the other hand,

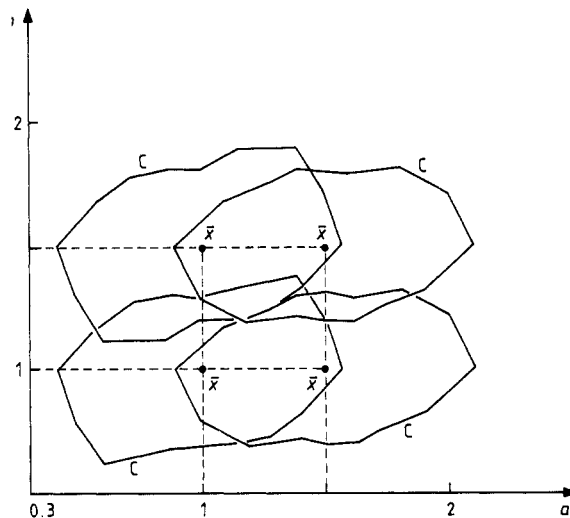


Figure 8. ‘Maximal’ sets C obtained for example 1 around different nominal values $\bar{x} = (\bar{a}, \bar{b})$ of the parameters.

the use of the much more restrictive condition (4.4) would lead, in this example, to a maximal set of the size of a point in figure 8, and it is thus inadequate for practical use.

7. Conclusion

We have studied the uniqueness of the local minima of general nonlinear least-squares problems, under the main hypothesis that the mapping to be inverted is regular C^2 and has an injective derivative. For this case we have derived a sufficient condition that involves a minimisation, over all 'geodesic' curves of the image set, of a quantity that involves the radius of curvature of the 'geodesic' curve and a function related to the radius of curvature through the resolution of an elliptic problem (see (5.16)). This condition has been optimised among a class of possible sufficient conditions, but it is not known whether it is the best possible condition. However, numerical examples have shown that the proposed condition makes it possible to obtain practically interesting results for a two-parameter estimation problem.

References

- Charles J-L 1985 Détermination d'un domaine de stabilité pour un problème d'estimation de paramètres
Rapport de DEA, Université Paris-IX Dauphine
- Chavent G 1983 Local stability of the output least square parameter estimation technique *Mathematica Applicada e Computational* **2** 3–22
- Spiess J 1969 Eindeutigkeitsätze bei der Nichtlinearen Approximation in Strikt-Konvexen Räumen
Dissertation zur Erlangung des Doktorgrades Fachbereich Mathematik, Universität Hamburg (in German)