A maximum curvature step and geodesic displacement for nonlinear least squares descent algorithms

Guy Chavent

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A maximum curvature step and geodesic displacement for nonlinear least squares descent algorithms

Guy Chavent

N° 4383

Février 2002

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4383--FR+ENG

A maximum curvature step and geodesic displacement for nonlinear least squares descent algorithms

Guy Chavent*

Thème 4 — Simulation et optimisation
de systèmes complexes
Projets Estime

Rapport de recherche n° 4383 — Février 2002 — 21 pages

Abstract: We adress in this paper the choice of both the step and the curve of the parameter space to be used in the line search search part of descent algorithms for the minimization of least squares objective functions. Our analysis is based on the curvature of the path of the data space followed during the line search. We define first a new and easy to compute *maximum curvature step*, which gives a guaranteed value to the residual at the next iterate, and satisfies a linear decrease condition with $\omega = \frac{1}{2}$. Then we optimize (i.e. minimize !) the guaranteed residual by performing the line search along a curve such that the corresponding path in the data space is a *geodesic* of the output set. An inexpensive implementation using a second order approximation to the geodesic is proposed. Preliminary numerical comparisons of the proposed algorithm with two versions of the Gauss-Newton algorithm show that it works properly over a wide range of nonlinearity, and tends to outperform its competitors in strongly nonlinear situations.

Key-words: optimization, least squares, line search, curvature, geodesic.

* Ceremade, Université Paris-Dauphine and Inria-Rocquencourt, BP105, 78153 Le Chesnay Cedex, France, `Guy.Chavent@inria.fr`

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Pas de courbure maximum et déplacement géodésique pour des algorithmes de descente pour les problèmes de moindres carrés non-linéaires

Résumé : Nous considérons dans ce travail le choix du pas et de la courbe sur laquelle on se deplace dans l' espace des paramètres lors de la recherche linéaire au sein d'un algorithme de descente pour la résolution d'un problème de moindre carrés. Notre analyse utilise la courbure du chemin suivi, dans l'espace des données, au cours de cette recherche linéaire. Nous définissons d'abord un nouveau *pas de courbure maximum* facile à calculer, qui donne une valeur garantie du résidu au prochain itéré, et satisfait la condition de décroissance linéaire avec $\omega = \frac{1}{2}$. Nous montrons ensuite que l'on optimise (c'est a dire minimise !) la valeur garantie du nouveau résidu en suivant dans l'espace des paramètres une courbe dont l'image est une géodésique de l'ensmble atteignable de l'espace des données. Une mise en œuvre utilisant une approximation d'ordre deux de la géodésique est développée. Une comparaison numérique préliminaire de l'algorithme avec deux versions de l'algorithme de Gauss-Newton montre qu'il se comporte correctement sur une grande gamme de non-linéarités, et tend à être plus efficace pour les fortes non-linéarités.

Mots-clés : optimisation, moindres carrés, recherche linéaire, courbure, geodesique.

1 Introduction

We consider here the resolution of a non linear least squares problem

$$(\mathcal{LS}) \quad \hat{x} \text{ minimizes } f(x) \stackrel{\text{def}}{=} \frac{1}{2}\|F(x)\|^2 \text{ over } \mathbb{R}^n \tag{1}$$

where x is the parameter of $\mathbb{R}^n$ to be estimated, and $F(x)$ is the residual between the data and the output of the model.

Descent algorithms for the resolution of $(\mathcal{LS})$ perform a line search in the parameter space by moving away from the current estimate x^k in a direction y^k specific to the algorithm, until ideally the first stationary point of $\|F(x)\|$ is attained. If we denote by p be the path of the data space followed during this step, this amounts to search for the first stationary value $\bar{r}$ of the residual $r = \|F\|$ along p.

We propose in this paper a new approach to line search algorithms, based on the curvature of the path p. Of course, one does not know the shape of p ! But one can compute easily its radius of curvature R_0 at arc length $\nu = 0$; then $R = \mu R_0$ is an a lower bound to the radius of cuvature of p in a neighborhood of 0, where $\mu \leq 1$ is a security factor, which accounts for the possible diminution or the radius of curvature along the path. This lower bound R will be the key to our study.

So we analyse first in section 2 the properties of paths p which leave p_0 in the direction v_0 with a curvature bounded by $1/R$. We show that, among all these paths, there is one (and generally only one) "worst path" p_M for which the residual $\bar{r}_M$ at the first stationary point is maximum, and simultaneously the arclength $\bar{\nu}_M$ at the first stationary point is minimum. Not surprisingly, this worst path consists, after leaving p_0 in the direction v_0, in "turning steadily away from 0" with the maximum authorized curvature $1/R$;

Then we use in section 3 these results to define a "maximum curvature step" α_M^k for the computation of x^{k+1} from x^k and y^k, which corresponds to moving forward on p up to the arclength $\bar{\nu}_M$ of the first stationary point on the worst path p_M.

This step is conservative : under the hypothesis that the radius of curvature of p has stayed above R over the $[0, \bar{\nu}_M]$ interval, one can be sure that one has not passed the first stationary point, and that the residual has decreased at least below $\bar{r}_M$. It is also optimal in the sense that it is the largest step which ensures these two properties.

Section 4 is devoted to the choice of the curve g of the parameter space along which to move from x^k to x^{k+1}. Based on the observation that the guaranteed residual $\bar{r}_M$ after a curvature step is a decreasing function of R, we are led to replace the line search by a search along a curve $\alpha \to g(\alpha)$ such that the corresponding path p has the smallest curvature. As this path is constrained to stay on the "output set"

$$D = \{F(x) \in \mathbb{R}^q \,|\, \text{for all } x \in \mathbb{R}^n\} \; .$$

this amounts to chose $g = g_G$ such that p is a geodesic of D. We show that moving along g_G optimizes the worst possible case; once the maximum curvature step α_M^k has been computed, this can be implemented at no additional computational cost by following the second order approximation g_G^{app} to g.

Finally, section 5 presents a few preliminary numerical tests on examples of increasing nonlinearity, which include comparisons with the classical Gauss-Newton algorithm with an Armijo or Quadratic line search.

2 The first stationary point along a path with bounded curvature

We denote in this section by p paths of $\mathbb{R}^q$ parameterized by their arc length ν, with $W^{2,\infty}$ regularity, and by $v(\nu) = p'(\nu)$ and $a(\nu) = a''(\nu)$ the corresponding velocity and acceleration, which satisfy:

$$\|v\| = 1 \text{ and } < v, a >= 0 \ .$$

We consider as given the origin and the initial direction of the path:

$$p_0 \in \mathbb{R}^q \ , \ v_0 \in \mathbb{R}^q \ , \tag{2}$$

where v_0 is supposed to be a descent direction for the residual $r(\nu) = \|p(\nu)\|$:

$$< p_0, v_0 >\leq 0 \ . \tag{3}$$

Then we denote by:

$$R \in]0, \infty] \ , \tag{4}$$

a lower bound to the radii of curvature along p, and define:

$$\mathcal{P} = \{ \ p \in W^{2,\infty}(\mathbb{R}+) \ | \ p(0) = p_0 \ , \ v(0) = v_0 \text{ and } \|a(\nu)\| \leq 1/R\} \ . \tag{5}$$

All paths p of $\mathcal{P}$ leave the same point p_0 in the same descent direction v_0, with a curvature smaller than $1/R$. To a path $p \in \mathcal{P}$ we associate the first stationary point $\bar{\nu}$ of the residual $r(\nu) = \|p(\nu)\|$, defined by:

$$\bar{\nu} = Inf\{\nu \geq 0 \text{ such that } \frac{d}{d\nu}(r^2) = 0\} \ . \tag{6}$$

If one thinks of p as the set of points in the data space on which one wants to minimize the residual during the line search step of an optimization algorithm, then $\bar{\nu}$ is the point where one should ideally stop! So in order to study how $\bar{\nu}$ depends on p, we single out among all paths of $\mathcal{P}$ the path p_M defined by:

$$p_M \text{ turns steadily } \textit{away from zero} \text{ with the } \textit{maximum curvature } 1/R. \tag{7}$$

and we want to show that p_M is the *worst path* in the sense that it

- hits its first stationary point $\bar{\nu}_M$ sooner,
- with a residual $\bar{r}_M = r(\bar{\nu}_M)$ larger

than any other path p of $\mathcal{P}$.

Proposition 2.1 *Define:*

$\gamma = \arccos(- < v, p/||p|| >)$ *(angle between $v(\nu)$ and $-p(\nu)$)*

$y = r \sin\gamma$ *(minimum residual along the tangent to p at $p(\nu)$)*

$t = {r_0}^2 - r^2$ *(decrease of squared residual),*

Then:

$$y(t) = y_0 + \frac{G(t)}{2R} \ , \tag{8}$$

where $G(t)$ is a function which depends on p and satisfies:

$$G(0) = 0 \ , \ |G'(t)| \leq 1 \ . \tag{9}$$

The function $G_M(t) = t \ \ \forall t$ corresponds to the path p_M defined in (7).

Proof: The formula $\sin^2\gamma + \cos^2\gamma = 1$ gives, after multiplication by r^2:

$$y^2 + < v, p >^2 = {r_0}^2 - t \ ,$$

and, after differentiation with respect to ν :

$$dy = \frac{1}{2} \frac{< a, p/||p|| >}{\sin\gamma} dt \ . \tag{10}$$

But t is a monotonously increasing function of ν on $[0, \bar{\nu}]$ by definition of $\bar{\nu}$, and we can reparameterize p -and hence y- as a function of t over the $[0, \bar{t}]$ interval, where $\bar{t} = {r_0}^2 - \bar{r}^2$ and $\bar{r} = r(\bar{\nu})$.

Let n_p be the projection of $-p/||p||$ on the normal hyperplane to p at $p(\nu)$, defined by:

$$n_p + v \cos\gamma = -p/||p|| \ .$$

Then

$$< n_p, v >= 0 \text{ and } ||n_p|| = \sin\gamma \ .$$

so that

$$- < a, p/||p|| >=< a, n_p >$$

and (recall that $||a|| \leq 1/R$):

$$|- < a, p/||p|| > | \leq ||a|| \ ||n_p|| \leq \sin\gamma/R \ .$$

Hence we can associate to the path p a function χ such that:

$$\begin{cases} < a, p/||p|| >= \frac{\chi}{R} \sin\gamma \ , \\ |\chi(t)| \leq 1 \text{ a.e. on } [0, \bar{t}] \ . \end{cases} \tag{11}$$

The differential equation for y reduces then to

$$\frac{dy}{dt} = \frac{1}{2}\frac{\chi}{R} ,$$

whose solution is formula (8), where $G(t)$ denotes the primitive of χ with respect to t (not ν!):

$$G(t) = \int_0^t \chi .$$

The case $\chi_M \equiv 1$ corresponds to the case where the acceleration a satisfies $\|a\| \equiv 1/R$ and points constantly in the direction of n_p, which is the definition of p_M given in (7).

■

Proposition 2.2 *The decrease $\bar{t} = {r_0}^2 - r(\bar{\nu})^2$ of the squared residual at the first stationary point $\bar{\nu}$ is given by:*

$$\bar{t} = Inf\{t \geq 0 \text{ such that } \mu(t) = {r_0}^2\} \quad , \tag{12}$$

where:

$$\mu(t) = t + (y_0 + \frac{G(t)}{2R})^2 \leq t + (y_0 + \frac{t}{2R})^2 \stackrel{\text{def}}{=} \mu_M(t) \ , \tag{13}$$

so that:

$$\bar{t}_M \leq \bar{t} \quad \forall p \in \mathcal{P} \ . \tag{14}$$

Hence p_M is, among all paths of $\mathcal{P}$, the worst one for the minimization of the residual.

Proof: One has obviously:

$$\frac{d}{d\nu}(r^2) = 0 \ \Leftrightarrow \ < p, v >= 0 \ \Leftrightarrow \ \sin\gamma = 1 \ \Leftrightarrow \ y = r \ ,$$

and, using proposition 2.1:

$$y^2 = r^2 \ \Leftrightarrow \ (y_0 + \frac{G(t)}{2R})^2 = {r_0}^2 - t \ \Leftrightarrow \ \mu(t) = {r_0}^2 \ ,$$

which proves that (12) is equivalent to (6). Then $y_o = r_0 \sin\gamma_0 \geq 0$ implies

$$\mu(t) = t + (y_0 + \frac{G(t)}{2R})^2 \leq t + (y_0 + \frac{|G(t)|}{2R})^2 \ ,$$

which proves (13) using (9). Then (14) follows from (13).

■

Proposition 2.3 *The arclength $\bar{\nu}$ at which a path p attains the first stationary point of the residual satisfies:*

$$\bar{\nu}_M \ \leq \ \bar{\nu} \ \ \forall p \in \mathcal{P} \ , \tag{15}$$

$$\bar{r} \leq r(\nu) \leq r_M(\nu) \quad \forall \nu \in [0, \bar{\nu}_M] \quad \forall p \in \mathcal{P} \ . \tag{16}$$

Proof: Differentiation of t using the definitions of t and γ in proposition 2.1 gives:

$$dt = -2 < p(\nu), v(\nu) > d\nu = 2r(\nu)\cos(\gamma)d\nu \ .$$

and, omitting the ν argument from now on:

$$dt = 2(r^2 - r^2\sin^2\gamma)^{1/2} d\nu \ .$$

But $r^2 = {r_0}^2 - t$ and $r\sin\gamma = y = y_0 + \frac{G(t)}{2R}$, which shows that

$$dt = 2({r_0}^2 - t - (y_0 + \frac{G(t)}{2R})^2)^{1/2} d\nu \ ,$$

or, with the notation (13):

$$dt = 2({r_0}^2 - \mu(t))^{1/2} d\nu \ .$$

Hence ν is given as a function of t by:

$$\nu = \frac{1}{2}\int_0^t \frac{dt}{({r_0}^2 - \mu(t))^{1/2}} \ . \tag{17}$$

In particular, $\bar{\nu}$ and $\bar{\nu}_M$ are given by:

$$\bar{\nu} = \frac{1}{2}\int_0^{\bar{t}} \frac{dt}{({r_0}^2 - \mu(t))^{1/2}} \quad , \quad \bar{\nu}_M = \frac{1}{2}\int_0^{\bar{t}_M} \frac{dt_M}{({r_0}^2 - \mu_M(t_M))^{1/2}} \ . \tag{18}$$

We consider now the change of variable

$$t \in [\,0, \bar{t}\,] \rightsquigarrow t_M \in [\,0, \bar{t}_M]$$

defined by:

$$\mu_M(t_M) = \mu(t) \ . \tag{19}$$

It is uniquely defined because $t \rightsquigarrow \mu_M(t)$ is monotonously increasing, and the ranges of both μ over $[\,0, \bar{t}\,]$ and μ_M over $[\,0, \bar{t}_M]$ are $[\,0, {r_0}^2\,]$. From (13) in proposition 2.2 we know that $\mu(t) \leq \mu_M(t) \ \ \forall t$, which implies that:

$$t_M \leq t \ \ \forall t \in [\,0, \bar{t}\,] \ , \tag{20}$$

and hence, using the definition of μ and μ_M:

$$(y_0 + \frac{t_M}{2R})^2 \geq (y_0 + \frac{G(t)}{2R})^2 \ . \tag{21}$$

Differentiation of (19) gives:

$$\mu'(t_M)dt_M = \mu'(t)dt,$$

where $\mu'(t_M) > 0$, but $\mu'(t)$ can be positive, negative or zero. Hence the formula for $\bar{\nu}_M$ in (18) becomes

$$\bar{\nu}_M = \frac{1}{2}\int_0^{\bar{t}} \frac{\mu'(t)}{\mu'(t_M)} \frac{dt}{({r_0}^2 - \mu(t))^{1/2}} ,$$

so that

$$\bar{\nu}_M \leq \frac{1}{2}\int_0^{\bar{t}} \frac{|\mu'(t)|}{\mu'(t_M)} \frac{dt}{({r_0}^2 - \mu(t))^{1/2}} .$$

But we see from (13) that:

$$|\mu'(t)| = |1 + \frac{G'(t)}{R}(y_0 + \frac{G(t)}{2R})| \leq 1 + \frac{|G'(t)|}{R}|y_0 + \frac{G(t)}{2R}| ,$$

i.e. using (9) and (21):

$$|\mu'(t)| \leq 1 + \frac{1}{R}(y_0 + \frac{t_M}{2R}) = \mu'(t_M) .$$

This shows that

$$\bar{\nu}_M \leq \frac{1}{2}\int_0^{\bar{t}} \frac{dt}{({r_0}^2 - \mu(t))^{1/2}} = \bar{\nu} ,$$

which is (15). In order to prove (16), let t and t_M be the decrease of squared residual along p and p_M at the same arclength $\nu = \nu_M$. In view of (17), t and t_M are linked by:

$$\int_0^t \frac{dt}{({r_0}^2 - \mu(t))^{1/2}} = \int_0^{t_M} \frac{dt}{({r_0}^2 - \mu_M(t))^{1/2}} .$$

Then (13) implies that necessarily

$$t \geq t_M \quad \forall \nu \in [0, \bar{\nu}_M]$$

which is (16).

■

Proposition 2.4 *The arc length $\bar{\nu}_M$ and residual $\bar{r}_M$ at the first stationary point along the "worst path" p_M are given by:*

$$\bar{\nu}_M = R \arctan \frac{\bar{\nu}_L}{R + \bar{r}_L} , \tag{22}$$

$$\bar{r}_M = ((R + r_L)^2 + {\nu_L}^2)^{\frac{1}{2}} - R , \tag{23}$$

where $\bar{\nu}_L$ and $\bar{r}_L$ are the arc length and residual at the first stationary point along the linearized path $p_L(\nu) = p_0 + v_0\nu$, given by:

$$\bar{\nu}_L = r_0 \cos \gamma_0 \quad , \quad \bar{r}_L = r_0 \sin \gamma_0 . \tag{24}$$

The residual $\bar{r}_M$ given by (23) decreases from r_0 (for $R = 0$, infinite curvature) to $\bar{r}_L = r_0 \sin \gamma_0$ (for $R = \infty$, zero curvature).

Proof: The case where p has zero curvature is easily obtained in the previous formula by chosing $G(t) \equiv 0$. We get first from equation (8)

$$r_L(\nu)\sin\gamma_L(\nu) = r_0 \sin\gamma_0 \quad \forall \nu \ ,$$

which proves the right part of (24) as $\sin\bar{\gamma}_L = 1$. Then we see from equation (17) that

$$\bar{\nu}_L = \frac{1}{2}\int_0^{\bar{t}_L} \frac{dt}{({r_0}^2\cos^2\gamma_0 - t)^{1/2}} \ ,$$

where

$$\bar{t}_L = {r_0}^2 \cos^2\gamma_0 \ ,$$

which proves the left part of (24). Then we can rewrite (18), using $\bar{\nu}_L$ and $\bar{r}_L$ defined above, as:

$$\bar{\nu}_M = \frac{1}{2}\int_0^{\bar{t}_M} \frac{dt}{(\bar{\nu}_L^2 + (R + r_L)^2 - (R + r_L + \frac{t}{2R})^{1/2}} \ . \tag{25}$$

i.e.:

$$\bar{\nu}_M = R\int_{u_{min}}^{1} \frac{du}{(1-u^2)^{1/2}} \ , \tag{26}$$

where u and u_{min} are defined by

$$u = \frac{R + r_L + \frac{t}{2R}}{(\bar{\nu}_L^2 + (R + \bar{r}_L)^2)^{1/2}} \quad , \quad u_{min} = \frac{R + r_L}{(\bar{\nu}_L^2 + (R + \bar{r}_L)^2)^{1/2}} \ .$$

Hence:

$$\bar{\nu}_M = R\{\arcsin 1 - \arcsin \frac{R + \bar{r}_L}{\bar{\nu}_L^2 + (R + \bar{r}_L)^2}\}$$

which gives (22) after some transformations. Finally, equations (8) (12) (13) show that $\bar{r}_M$ is the positive solution of the equation:

$$r^2 + 2rR - (2R\bar{r}_L + {r_0}^2) = 0 \ ,$$

which is (23).

■

Remark In fact, formula (22) (23) could have been guessed -and in fact have- from the very beginning, by noticing that the "worst path" p_M defined by (7) is a circle of radius R in the hyperplane determined by 0, p_0 and v_0.

■

All the above results have been obtained under the hypothesis that one knew a global lower bound R of the radius of curvature along p. This will not be the case in the applications to optimization we have in mind, where such an estimate will be available only on a neighborhood of $\nu = 0$. So we replace $\mathcal{P}$ by:

$$\widetilde{\mathcal{P}} = \{\ p \in W^{2,\infty}(\mathbb{R}+) \mid p(0) = p_0 \ , v(0) = v_0\} \ , \tag{27}$$

and denote, for any $R > 0$, by:

$$\bar{\nu}_M(R) \quad \text{defined by (22)} \tag{28}$$

$$\bar{r}_M(R) \quad \text{defined by (23)} \tag{29}$$

the arclength at the stationary point of the worst path p_M with radius R, and the corresponding residual.

Proposition 2.5 *Let $p \in \widetilde{\mathcal{P}}$ and $R > 0$ be given such that:*

$$\rho(\nu) \geq R \quad \forall \nu \in [0, \bar{\nu}_M(R)] \ , \tag{30}$$

where ρ is the radius of curvature along p, given by:

$$\rho(\tau) \stackrel{\text{def}}{=} \|a(\tau)\|^{-1} = \|p''(\tau)\|^{-1} \quad . \tag{31}$$

Then:

$$\bar{\nu} \geq \bar{\nu}_M(R) \quad , \tag{32}$$

$$\bar{r} \leq r(\bar{\nu}_M(R)) \leq \bar{r}_M(R) \quad . \tag{33}$$

where

- *$\bar{\nu}$ is the arclength of first stationary point of the residual on p,*
- *$\bar{r}$ is the the value of the residual on p at its first stationary point,*

The best estimates, i.e. the largest value of $\bar{\nu}_M(R)$ and the smallest value of $\bar{r}_M(R)$, are obtained for the largest R which satisfies (30), that is for $\widetilde{R}$ solution of;

$$\tilde{\nu} = \bar{\nu}_M(\widetilde{R}) \quad , \quad \widetilde{R} = R_m(\tilde{\nu}) \ . \tag{34}$$

where:

$$R_m(\nu) = Inf\{\ \rho(\tau)\ ,\ 0 \leq \tau \leq \nu\} \quad . \tag{35}$$

denotes the smallest radius of curvature on p up to arclength ν.

Proof: Properties (32) and (33) follow immediately from (15) and (16) of proposition 2.3 applied to the path $q \in \tilde{\mathcal{P}}$ which coincides with p up to arclength $\bar{\nu}_M(R)$, and has zero curvature for $\nu \geq \bar{\nu}_M(R)$. The existence and uniqueness of $\tilde{\nu}$ and $\widetilde{R}$ satisfying (34) results from the fact that $\nu \rightsquigarrow R_m(\nu)$ defined by (35) is decreasing, and that $R \rightsquigarrow \bar{\nu}_M$ defined by (28) is increasing.

■

3 A maximum curvature step for descent algorithms

We are given in this section a mapping

$$F \quad : \mathcal{R}^n \longrightarrow \mathcal{R}^p \tag{36}$$

to be inverted in the least squares sense by minimization of:

$$f \ : \ x \in \mathcal{R}^n \rightsquigarrow f(x) = \frac{1}{2}\|F(x)\|^2 \in \mathcal{R} \ , \tag{37}$$

by a descent algorithm of the form

$$x^{k+1} = g(\alpha^k) \ . \tag{38}$$

The function $\alpha \rightsquigarrow g(\alpha)$ describes the curve of the parameter space along which one moves from x^k to x^{k+1}. It is chosen such that:

$$g(0) = x^k \quad , \quad g'(0) = y^k \ , \tag{39}$$

where:

1. x^k is the current iterate,
2. y^k is a descent direction for f at x^k, computed from $\nabla J(x^k)$ and the previous descent direction(s) (Conjugate Gradient, Quasi-Newton algorithms...), or from F^k and $J^k = F'(x^k)$ (Quasi-Newton algorithms...),
3. α^k is the step on the curve g, which is required to satisfy the so called *linear decrease condition*:
$$f(g(\alpha^k)) \leq f(x^k) + \alpha^k \omega f'(x^k).y^k \ , \tag{40}$$
for some
$$\omega \in]0, 1/2] \tag{41}$$
to be chosen by the user.

It will be convenient to denote by

$$z^k = g''(0) \tag{42}$$

the initial acceleration on the curve g. The usual situation where one moves from x^k to x^{k+1} along a straight line of the parameter space is obtained by requiring that g satisfies the differential equation:

$$g''(\alpha) = 0 \ \ \forall \alpha \geq 0 \ \ , \tag{43}$$

which, together with (39) gives:

$$g_S(\alpha) \stackrel{\text{def}}{=} x^k + \alpha y^k \ . \tag{44}$$

We consider in this section the curve g as given, for example -but not necessarily- by (43) (44), and discuss the choice of the step α^k. We recall first in sections 3.1 and 3.2 the classical *Armijo* and *Quadratic* steps (cf for example [2]), which we have implemented in our numerical tests for comparison purpose. Then we introduce in section 3.3 a new *Maximum Curvature* step by application of the results of section 2 on the first stationary point of a path.

3.1 The Armijo step

In this approach, one has first to make a first guess α of the step. If the linear decrease condition (40) is not satisfied, one replaces α by $\mu\alpha$ for some $\mu \in]0,1]$ and tries again. The Armijo step α_A^k is then the first α for which condition (40) is satisfied.

In Gauss-Newton type algorithms, where y^k is determined by requiring that $x^k + y^k$ solves the linearized problem, $\alpha = 1$ is a reasonable first guess (condition (40) would then be satisfied with $\omega = 1/2$ if F were affine and $g = g_S$ had been chosen).

3.2 The Quadratic step

Here also, one has to make a first guess α of the step. If condition (40) is satisfied, one sets $\alpha_Q = \alpha$. If not, one uses $f(g(\alpha))$, $f(x^k)$ and $f'(x^k).y^k$, which have just been evaluated to check (40), to compute a quadratic approximation of $(f \circ g)''(0)$. Then one sets:

$$\tilde{\alpha} = -\frac{f'(x^k).y^k}{(f \circ g)''(0)} \quad . \tag{45}$$

When F happens to be affine, and one moves along the straight line $g = g_S$, the function $\alpha \rightsquigarrow f(g(\alpha))$ is quadratic. Then the step $\tilde{\alpha}$ produces exactly the minimum of f along the straight line, and (40) is satisfied with $\omega = 1/2$. But f is nonlinear, and for security reasons, one does not want to test (40) for values outside the $[0, \alpha]$ interval. Hence one projects $\tilde{\alpha}$ on the interval $[\tau\alpha, (1-\tau)\alpha]$ for some $\tau \in [0,1]$. This gives the new value of α_Q. If (40) holds, one proceeds to the next iteration. If not, the algorithm stops with the proper diagnostic.

3.3 The maximum curvature step

We associate to g a path $\tilde{p}$ of the data space defined by:

$$\tilde{p}(\alpha) = F(g(\alpha)) \quad \forall \alpha \geq 0 \ , \tag{46}$$

and denote by

$$\nu(\alpha) = \int_0^\alpha \|F'(g(\tau)).g'(\tau)\| \, d\tau \tag{47}$$

the arclength function along $\tilde{p}$, and by p the reparameterization of $\tilde{p}$ by the arclength ν. The curve g and the mapping F are supposed regular enough for p to have a finite curvature which varies continuously with ν, i.e. to satisfy:

$$p \in \mathcal{C}^{2,\infty}(\mathbb{R}^+) \ . \tag{48}$$

We propose here to use the specific least-squares structure (37) of f to define a new "maximum curvature step" α_M^k which will produce a guaranteed decrease of f and satisfy (40) with $\omega \simeq 1/2$: rather than evaluating one single real number $f''(x^k).(y^k,y^k)$ as it is done in the quadratic step, one could as well evaluate at a similar cost the vector $\tilde{p}''(0)$, which gives information on the shape of the path p in the data space, and allows to use results of section 2.

With the notations:

$$\begin{cases} F^k = & \tilde{p}(0) = & F(x^k) \ , \\ V^k = & \tilde{p}'(0) = & F'(x^k).y^k \ , \\ A^k = & \tilde{p}''(0) = & F''(x^k).(y^k,y^k) + F'(x^k).z^k \ . \end{cases} \tag{49}$$

the quantities associated to p in section 2 are given by:

$$p_0 = F^k \ , \qquad v_0 = V^k/\|V^k\| \ , \tag{50}$$

$$r_0 = \|F^k\| \ , \quad f'(x^k).y^k = -\|V^k\| < p_0, v_0 > \ , \tag{51}$$

$$\bar{\nu}_L = - < F^k, v_0 > \ , \quad \bar{r}_L = (\|F^k\|^2 - < F^k, v_0 >^2)^{1/2} \ , \tag{52}$$

and the radius of cuvature R^k of p at $\nu = 0$ is given by:

$$(1/R^k)^2 = \|A^k\|^2 - < A^k, v_0 >^2 \ . \tag{53}$$

(58) In order to apply the results of proposition 2.5, one needs a lower bound R of the radius of curvature of p on a neighborhood of $\nu = 0$. But we know the radius of curvature R^k at $\nu = 0$. Hence it is natural to take R of the form:

$$R = \kappa^{k+\frac{1}{2}} R^k \quad \text{with} \quad 0 \le \kappa^{k+\frac{1}{2}} \le 1 \ , \tag{54}$$

where $\kappa^{k+\frac{1}{2}}$ is a security factor which accounts for the possible increase of the curvature along the path. We can now define the *maximum curvature step α_M^k along g* by:

$$\nu(\alpha_M^k) = \bar{\nu}_M(R) = R \arctan \frac{\bar{\nu}_L}{R + \bar{r}_L} \tag{55}$$

where ν is the arc length function defined in (47), and $\bar{\nu}_M(R)$ is the arclength of the first stationary point on the "worst path" p_M with curvature $1/R$, defined in (28).

Theorem 3.1 *If $\chi^{k+\frac{1}{2}}$ is small enough for p and R to satisfy (30), the maximum curvature step α_M^k defined by (55) satisfies:*

$$\nu(\alpha_M^k) = \quad \bar{\nu}_M(R) \quad \le \bar{\nu} \tag{56}$$

$$f(g(\alpha_M^k)) \le \quad \tfrac{1}{2}\bar{r}_M(R)^2 \quad \le f(x^k) + \alpha_M^k \frac{1}{2} \frac{\nu(\alpha_M^k)}{\nu_L(\alpha_M^k)} f'(x^k).y^k \tag{57}$$

where $\bar{\nu}_M(R)$ and $\bar{r}_M(R)$ are given by (28) (29), and $\alpha \rightsquigarrow \nu_L(\alpha)$ is the arclength along the tangent to p at the origin, defined by:

$$\nu_L(\alpha) = \alpha \ \|V^k\| \ . \tag{58}$$

The linear decrease condition (40) is hence satisfied with

$$\omega = \frac{1}{2} \ \frac{\nu(\alpha_M^k)}{\nu_L(\alpha_M^k)} \ \simeq \ \frac{1}{2} \ , \tag{59}$$

Proof: The left part of (56) is the definition of α_M^k, and the right part is (32) of proposition 2.5 applied to p. Then we get from equation (33) in proposition 2.5 that:

$$f(g(\alpha_M^k)) = \frac{1}{2} r(\bar{\nu}_M(R))^2 \leq \frac{1}{2} \bar{r}_M(R)^2 \ , \tag{60}$$

which proves the left part of (57). Then we define ω_M^k by:

$$\frac{1}{2}\bar{r}_M(R)^2 = f(x^k) + \alpha_M^k \frac{\nu(\alpha_M^k)}{\nu_L(\alpha_M^k)} \omega_M^k f'(x^k).y^k \ . \tag{61}$$

We can rewrite this equation, with the notations (50) (51), as:

$$\frac{1}{2}\bar{r}_M(R)^2 = \frac{1}{2} r_0^2 - \alpha_M^k \frac{\nu(\alpha_M^k)}{\nu_L(\alpha_M^k)} \omega_M^k \|V^k\| < p_0, v_0 > \ , \tag{62}$$

or,using (58) and the definitions (55) of α_M^k and (52) of $\bar{\nu}_L$:

$$\frac{1}{2}\bar{r}_M(R)^2 = \frac{1}{2} r_0^2 - \bar{\nu}_M(R)\bar{\nu}_L \omega_M^k \ . \tag{63}$$

Solving this equation for ω_M^k and using the formula of proposition 2.4 for $\bar{\nu}_M(R)$ and $\bar{r}_M(R)$ gives:

$$\omega_M^k = \frac{1-\cos\theta}{\theta \sin\theta} \quad \text{where:} \quad \tan\theta = \frac{\bar{\nu}_L}{R + \bar{r}_L} \ . \tag{64}$$

But the function $\theta \rightsquigarrow \omega_M^k$ increases from $1/2$ for $\theta = 0$ to $2/\pi$ for $\theta = \pi/2$, which proves the right part of (57).

■

We discuss now the implementation of the maximum curvature step.

In order to determine α_M^k by (55), we need first to render equation (55) solvable with respect to α by replacing the arclength function $\alpha \rightsquigarrow \nu(\alpha)$ by a simpler function, as for example

- its linear approximation:

$$\nu_L(\alpha) = \alpha \ \|V^k\| \ \ \text{(as in (58))}, \tag{65}$$

in which case α_M^k is given by:

$$\alpha_M^k \|V^k\| = \bar{\nu}_M(R) \ , \tag{66}$$

- or its quadratic approximation:

$$\nu_Q(\alpha) = \alpha \ \|V^k\| + \frac{\alpha^2}{2} < \frac{V^k}{\|V^k\|}, A^k > \ , \tag{67}$$

in which case α_M^k is the root of smallest absolute value of the second degree equation:

$$\nu_Q(\alpha) = \bar{\nu}_M(R) \ . \tag{68}$$

Now that equation (55) is technically solvable, we need to know its right-hand side to actually compute α_M^k. Equation (52) gives immediately the values of $\bar{\nu}_L$ and $\bar{r}_L$, so we are left with the choice of the lower bound R to radius of curvature of p near $\nu = 0$.

Given $\omega \in [0, 1/2]$, we want to satisfy the linear decrease condition (40) at each iteration for this value of ω. But theorem 3.1 shows that this condition is necessarily satisfied for $\alpha = \alpha_M^k$, provided that R is chosen small enough for (30) to hold. So we shall compute α_M^k from $R = \kappa^{k+\frac{1}{2}} R^k$ with the current value of $\kappa^{k+1/2}$, and accept the step if (40) is satistied. If not, $\kappa^{k+\frac{1}{2}}$ is multiplied by some $\mu \leq 1$, and α^k is evaluated again. In the implementation of this algorithm used in section 5, the security factor κ was initialized to 1 at the beginning of each iteration, but other strategies can be considered.

When only $f(x^k)$ and $\nabla f(x^k)$ are computationally available, the cost incurred by the computation of α_M^k is that required to evaluate (exactly, or approximately by finite differences) the two directional derivatives V^k and A^k of the forward model F in the same direction y^k (two evaluations of F).

When $f(x^k)$ and the Jacobian $J^k \stackrel{\text{def}}{=} F'(x^k)$ are available, then V^k can be computed by the matrix product $J^k y^k$, so the only cost incurred by α_M^k is that of the evaluation of A^k (one evaluation of F).

4 Moving along Geodesics

We consider in this section the case where both $F(x^k)$ and its Jacobian $J^k = F'(x^k)$ are available for the computation of x^{k+1}, and discuss the choice of the curve $\alpha \rightsquigarrow g(\alpha)$ used to move from x^k to x^{k+1} with a maximum curvature step α_M^k

There is of course no hope of finding the curve g which gives the best decrease to f. But the maximum curvature step, based on the worst case analysis of section 2, has been shown in theorem 3.1 to ensure that $f(x^{k+1})$ is smaller than the "guaranteed residual" $\frac{1}{2}\bar{r}_M(R)^2$. So we will choose the curve g which gives the smallest guaranteed residual $\frac{1}{2}\bar{r}_M(R)^2$.

As we know from proposition 2.4, $\bar{r}_M$ decreases from r_0 to $\bar{r}_L = r_0 \sin\gamma_0$ when R increases from 0 (infinite curvature) to ∞ (zero curvature). This leads to choose for g the curve g_G whose image p by F has the smallest possible curvature, i.e., as p is contrained to stay on $F(\mathbb{R}^n)$, such that p is a geodesic of $F(\mathbb{R}^n)$. Such a function $g(\alpha)$ is the solution of the differential equation (see [1] for example):

$$J^T(g)J(g)\,g'' + J^T(g)\,F''(g).(g',g') = 0\ , \tag{69}$$

$$g(0) = x^k \quad , \quad g'(0) = y^k\ . \tag{70}$$

and the arclength function $\nu(\alpha)$ simplifies to:

$$\nu_G(\alpha) = \alpha\|V^k\| = \nu_L(\alpha)\ , \tag{71}$$

where $\nu_L(\alpha)$ is the arclength along the tangent to p at the origin, defined in (58).

Theorem 4.1 *Let p_G (resp. p_S) be the path on $F(\mathbb{R}^n)$ associated to the curve g_G (resp. g_S) defined by (70) (71) (resp. (44)), and $\tilde{\nu}_G$, $\widetilde{R}_G$ (resp $\tilde{\nu}_S$, $\widetilde{R}_S$) the corresponding solutions of (34).*

1. *There exists $\nu_{max} > 0$ such that:*

$$\widetilde{R}_{max,S} \le \widetilde{R}_{max,G}\ , \tag{72}$$

where

$\widetilde{R}_{max,S}$ is the largest R such that $R \le \widetilde{R}_S$ and $\bar{\nu}_M(R) \le \nu_{max}$,

$\widetilde{R}_{max,G}$ is the largest R such that $R \le \widetilde{R}_G$ and $\bar{\nu}_M(R) \le \nu_{max}$.

2. *For any R_S and R_G such that:*

$$0 \le R_S \le \widetilde{R}_{max,S} \quad , \quad 0 \le R_G \le \widetilde{R}_{max,G}\ , \tag{73}$$

the maximum curvature step $\alpha^k_{M,S}$ along the half line g_S satifies (57) (with R replaced by R_S), and the maximum curvature step $\alpha^k_{M,G}$ along the geodesic g_G satisfies:

$$f(g(\alpha^k_{M,G})) \le \frac{1}{2}\bar{r}_M(R)^2 \le f(x^k) + \frac{1}{2}\alpha^k_M f'(x^k).y^k\ , \tag{74}$$

which means that the linear decrease condition (40) holds with $\omega = \frac{1}{2}$.

3. *If moreover the Gauss-Newton step y^k given by:*

$$J^{k^T} J^k y^k + J^{k^T} F^k = 0 \tag{75}$$

is small enough, then (72) is satified with a strict inequality. Hence , if R_G is chosen close enough to $\widetilde{R}_{max,G}$ in (73), the guaranteed residual along the "geodesic" g_G is necessarily smaller than the one along the "straight line" g_S, for any R_S satisfying (73).

Proof: The geodesic have at each point the smallest curvature among all paths of same direction. Hence there exists $\nu_{max} > 0$ such that;

$$\rho_G(\nu) \geq \rho_S(\nu) \quad \forall \nu \in [0, \nu_{max}] \ . \tag{76}$$

The theorem follows then easily from proposition 2.5 and theorem 3.1.

■

From a computational point of view, following the geodesic is an expensive operation, as it requires the resolution of the differential equation (70) by a numerical scheme. If an Euler scheme were used, for example, each step would have the same cost as the computation of the Gauss-Newton direction y^k (compare (69) and (75)), plus the cost of the second directional derivative $F''(g).(g', g')$.

Moreover, we see from theorem 4.1 that the geodesic is, from the point of view of achievable guaranteed residuals, better than the half line only on the neighborhood $[0, \nu_{max}]$ of $\nu = 0$.

Hence we shall replace g_G by its second order approximation g_G^{app} defined by:

$$g_G^{app}(\alpha) = x^k + \alpha y^k + \frac{\alpha^2}{2} z^k \ , \tag{77}$$

where $z^k = g''(0)$ is the solution of (compare with (75)):

$$J^{k^T} J^k z^k + J^{k^T} F''(x_k).(y_k, y_k) = 0 \ . \tag{78}$$

As we have seen in section 3, the implementation of the maximum curvature step requires the computation of $F''(x_k).(y_k, y_k)$ anyway for the evaluation of R^k (see equations (49) (53)). Hence the only cost incurred by moving along the geodesic rather than along a straight line is that of the resolution of the linear system (78), which is the same as the Gauss-Newton system (75), but with $F''(x_k).(y_k, y_k)$ in the right-hand side instead of F^k.

5 Numerical results

We have performed some preliminary tests, which we present now. We shall invert the function:

$$F = \Phi - d \ , \tag{79}$$

where $\Phi: \quad \Omega =]-1, +\infty[\times\mathbb{R} \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ is a regularized version of the Powell example, defined by:

$$\Phi(x_1, x_2) = \begin{pmatrix} x_1 \\ \frac{10\,x_1}{x_1+1} + 2x_2^2 \\ \epsilon\, x_2 \end{pmatrix} , \tag{80}$$

and where d is the data to be inverted:

$$d = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \tag{81}$$

The vector d does not belong to $\Phi(\Omega)$, so the minimum residual is strictly positive. The last component is set to zero, which correspond to the situation where no information on x_2 is available.

When the regularization parameter ϵ goes to zero, the function Φ tends to the function of the Powell example, which has a singularity along the line $x_2 = 0$. We have used the values $\epsilon = 0.1$, which corresponds to a relatively smooth problem, and $\epsilon = 0.01$, which corresponds to a quite stiff problem, where the curvature of $\Phi(\Omega)$ varies very quickly when x_2 changes sign.

The solution of the minimization problem is, $\forall \epsilon > 0$:

$$\hat{x} = \begin{pmatrix} 0.125 \\ 0 \end{pmatrix} , \tag{82}$$

which, when $\epsilon \to 0$, tends to be on the singularity of $\Phi(\Omega)$.

The minimization of $f(x) = \frac{1}{2}\|F(x)\|^2$ has been performed by three variations of the Gauss-Newton algorithm: at each iteration, the descent direction y^k is the Gauss-Newton direction computed by (75), only the way x^{k+1} is computed from x^k, y^k and J^k changes:

GN/Armijo/Straight: one moves from x^k to y^k straight in the direction y^k, with a step determined by the Armijo backtracking algorithm of section 3.1. This is a classical algorithm, which we use as reference. The parameters are the initial guess of the step at each iteration, set to $\alpha = 1$, the reduction factor, set to $\mu = 0.5$, and the coefficient ω in the linear decrease condition (40), set to $\omega = 10^{-4}$.

GN/Quadratic/Straight: same as above, but with a step determined by the quadratic algorithm of section 3.2. This algorithm is expected to be more efficient for smooth problems, so it can be a more demanding reference. The parameters are the initial guess of the step, still set to $\alpha = 1$, the security coefficient for the projection, set to $\tau = 10^{-2}$, and the coefficient ω, still set to $\omega = 10^{-4}$.

GN/Max_Curv/Geodesic: one moves from x^k to y^k along the approximate geodesic g_G^{app} defined in (77) of section 4, with the maximum curvature step $\alpha_{M,G}^k$ defined by (55) of section 3.3 applied to g_G^{app}. The parameters are the initial value of the security factor for the radius of curvature at each iteration, set to $\kappa = 1$, and the reduction factor, set to $\mu = 0.5$.

Each algorithm was started at one of the two initial points:

$$x_1^0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{or} \quad x_2^0 = \begin{pmatrix} 6 \\ 5 \end{pmatrix} , \tag{83}$$

and was stopped when the norm of the gradient was a given fraction of its initial value:

$$\|{J^k}^T F^k\| \leq 10^{-4} \, \|{J^0}^T F^0\| \; . \tag{84}$$

We present in tables 1 to 4 the comparative results in four situations. Each table displays:

- the number of Gauss-Newton iterations,
- the total number of reductions of the step α (for the GN/Armijo/Straight and GN/Quadratic/Straight algorithms) or the security factor κ for the radius of curvature (for the GN/Max_Curv/Geodesic algorithm),
- the total number of function evaluations (computation of the Jacobian J^k and the directional derivative $F''(x^k)(y^k, y^k)$ have been counted for one),
- the average value of the step α^k per Gauss Newton iteration,
- the solution $[x_1, x_2]$ found by the algorithm,
- the exit type of the algorithm: **Normal** when it stops because condition (84) is satisfied, **Descent** when it stops because the linear decrease condition (40)cannot be satisfied,

The first comparison was made on a rather smooth problem ($\epsilon = 0.1$) and with the initial guess $x_1^0 = [2, 1]$. The results are as follows:

Algorithm	GN Iter	Reductions	Funct eval
GN/Armijo/straight	261	1173	1695
GN/Quadratic/Straight	56	51	163
GN/Max_curv/Geodesic	501	68	1571
Algorithm	Mean step	Solution	Exit type
GN/Armijo/straight	0.060	$[0.1250; -0.0001]$	Normal
GN/Quadratic/Straight	0.1286	$[0.1250; -0.0006]$	Normal
GN/Max_curv/Geodesic	0.018	$[0.1249; 10^{-41}]$	Normal

As expected, the GN/Quadratic/Straight algorithm is the most effective on this smooth problem. The GN/Max_curv/Geodesic algorithm makes equal game in this case with the GN/Armijo/straight algorithm.

The next table show how the algorithms performs on the same problem ($\epsilon = 0.1$), but with the worse initial guess $x_2^0 = [6, 5]$:

Algorithm	GN Iter	Reductions	Funct eval
GN/Armijo/straight	257	1165	1695
GN/Quadratic/Straight	63	57	163
GN/Max_curv/Geodesic	48	61	1571
Algorithm	**Mean step**	**Solution**	**Exit type**
GN/Armijo/straight	0.065	$[0.1268; 0.0062]$	Normal
GN/Quadratic/Straight	0.09	$[0.1248; 0.0564]$	Normal
GN/Max_curv/Geodesic	0.129	$[0.1268; -0.0021]$	Normal

This time we see that the GN/Max_curv/Geodesic algorithm takes a slight advantage over GN/Armijo/straight: it does a few less function evaluations and also less iterations. But the GN/Quadratic/Straight algorithm is still the best performer.

We test now the case of more strongly non linear problems: we divide ϵ by 10, so one has now $\epsilon = 0.01$. We begin with the closest initial guess $x_1^0 = [2, 1]$. The results are:

Algorithm	GN Iter	Reductions	Funct eval
GN/Armijo/straight	7111	78846	93068
GN/Quadratic/Straight	4		
GN/Max_curv/Geodesic	15970	739	48469
Algorithm	**Mean step**	**Solution**	**Exit type**
GN/Armijo/straight	0.002	$[0.1250; 10^{-5}]$	Normal
GN/Quadratic/Straight	0.4417	$[0.9997; 0.1708]$	Descent
GN/Max_curv/Geodesic	0.0007	$[0.1250; 10^{-6}]$	Normal

The problem being less regularized, we see that the GN/Quadratic/Straight algorithm stops far from the solution, and that it takes more iterations to reach the solution for the two other algorithms. The GN/Armijo/straight algorithm does twice as much function evaluations as the GN/Max_curv/Geodesic algorithm, although it does less iterations.

The last comparison is made with the same regularization parameter $\epsilon = 0.01$, but with the worse initial guess $x_2^0 = [6, 5]$:

Algorithm	GN Iter	Reductions	Funct eval
GN/Armijo/straightccc	6455	72294	85204
GN/Quadratic/Straight	5		
GN/Max_curv/Geodesic	2163	99	6588
Algorithm	**Mean step**	**Solution**	**Exit type**
GN/Armijo/straight	0.002	$[0.1268; -0.0013]$	Normal
GN/Quadratic/Straight	0.5776	$[0.9970; -0.1375]$	Descent
GN/Max_curv/Geodesic	0.002	$[0.1268; 0.0003]$	Normal

The GN/Max_curv/Geodesic algorithm takes here a great advantage over the others. It is the only one that finds the solution in a reasonable number of steps, and it does on average 13 times less function evaluations than the GN/Armijo/straight.

6 Conclusion

The proposed algorithm, which is based on the optimization of the worst situation, seems to behave in a very robust way over a wide range of situation and nonlinearity. Unsurprisingly, it tends to outperform the two reference algorithms in situation of strong nonlinearity. Further numerical experimentation is required to assess its practical interest.

Acknowledgment

We thank Jean Charles Gilbert and Paul Armand for fruitful discussions, and Yves Laroque for the numerical tests.

References

[1] M. P. do Carmo: *Riemannian Geometry*, Boston, Birkhauser, 1992.

[2] J- C. Gilbert: *Optimisation différentiable: Théorie et Algorithmes*, in preparation.

[3] T. Rapcsac: *Smooth nonlinear optimization in* $\mathbb{R}^n$, Kluwer Academic Publishers, 1998.

[4] C. Udriste: *Convex functions and optimization methods on Riemannian manifolds*, Kluwer Academic Publishers, 1994.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399