Inverse Problems, Design and Optimization Symposium
Rio de Janeiro, Brazil, 2004

Line search global strategies for nonlinear least-squares problems based on curvature and projected curvature

Philippe Al Khoury
L.E.E.E. Université Paris X *
Inria-Rocquencourt ‡

Guy Chavent
Ceremade, Université Paris-Dauphine†
Inria-Rocquencourt

Abstract

In this paper we study line search global strategies based on one of two "curvature step-sizes": the maximum curvature step-size **MCS** ([1] and [2]), and the maximum projected curvature step-size **MPCS**. An extension of some geometrical topics, used to define the first step-size, enables defining the second one. In this extension we introduce the optimization plan, and we project the curvature on it. Using this projected curvature, an estimation is obtained for the first stationary point arc length on a search curve. Comparing arc lengths for different search curves enables defining both of curvature step-sizes. The two global strategies based on curvature step-sizes when used with different common directions, like the steepest descent, Gauss-Newton and Levenberg-Marquardt directions, gives more efficient algorithms than others with classical line search strategies based on the unit Newton step-size. Convergence of algorithms using curvature strategies is studied, and their behavior is illustrated on some numerical examples.

Key Words:
Line search global strategy, Nonlinear least-squares, Curvature step-sizes, Convergence, steepest descent, Gauss-Newton, Levenberg-Marquardt.

Nomenclature

x optimization variable
y descent direction
k subscript for the k-th iteration
α optimization step-size
$F(x) \in \mathbb{R}^q$ residual
F_i ith component of F
J Jacobian matrix of F
$f(x) \stackrel{\text{def}}{=} \frac{1}{2}\|F(x)\|^2 \in \mathbb{R}$ objective function
$\langle ., . \rangle$ the canonic scalar product.
∇f Gradient of f for the scalar product $\langle ., . \rangle$.
H Hessian of f
p path on the output set
v velocity of p
a curvature of p
a_{pr} projected curvature of p
ν arc length along p
r norm of residual on p
ρ radius of curvature of p
R lower bound for ρ on (part of) p
ρ_{pr} radius of projected curvature of p
R_{pr} lower bound for ρ_{pr} on (part of) p
MCS Maximum Curvature Step.
$MPCS$ Maximum Projected Curvature Step.

1 Introduction

Let be an application $F : \mathbb{R}^n \to \mathbb{R}^q$, both of the spaces $\mathbb{R}^n$ and $\mathbb{R}^n$ are supplied by the canonical scalar product, denoted $\langle ., . \rangle$, for witch they are Hilbert spaces. We discuss here the resolution of a nonlinear least-squares problem

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

The special structure of the objective function enables the use of some specific tools to study and solve this problem as it is now explained. First we denote by $F_i(x)$ the ith component of the m-vector $F(x)$, $J(x)$ the Jacobian matrix

*1 Chemin Desvallières 92410, Ville d'Avray France
†75775 Paris Cedex16, France
‡BP105, 78153 Le Chesnay Cedex, France

of F at x, and $H(x)$ the Hessian of f at x, so we have:

$$H(x) = J(x)^t J(x) + \sum_{i=1}^{m} F_i(x) \nabla^2 F_i(x). \quad (2)$$

Solving numerically (1) amounts to finding a stationary point x, hence satisfying

$$\nabla f(x) = 0. \quad (3)$$

A descent algorithm applied to the problem (1) defines at every iteration k:

1. A descent direction y_k, satisfying:

$$\langle \nabla f_k, y_k \rangle < 0. \quad (4)$$

2. A descent curve $g_k : \mathbb{R}^+ \to \mathbb{R}^n$ such that:

$$g_k(0) = x_k \quad , \quad g'(0) = y_k. \quad (5)$$

For this paper, it has the shape:

$$g_k(\alpha) = x_k + \alpha y_k. \quad (6)$$

3. A descent step-size α_k, which is required to decrease sufficiently the objective function f.

We say that the algorithm is convergent if it ensures:

$$\lim_{k \to \infty} \nabla f(x_k) = 0. \quad (7)$$

Applying Newton algorithm to solve (3) gives at iteration k a direction y_k solution to the following linear system:

$$H_k y_k + \nabla f_k = 0. \quad (8)$$

Then the next iterate will be $x_k + y_k$ if it is 'accepted'. The acceptance of a step-size α_k for a descent direction d is decided by a decrease condition (Armijo condition):

$$f(x_k + \alpha_k y_k) \le f(x_k) + \omega \alpha_k \langle y_k, \nabla f(x_k) \rangle, \quad (9)$$

where $0 < \omega < 1$. An additional condition to prevent the step-size from being too small is added like Wolfe condition:

$$\langle y_k, \nabla f(x_k + \alpha_k y_k) \rangle \ge \omega' \langle y_k, \nabla f(x_k) \rangle, \quad (10)$$

or Goldstein condition:

$$f(x_k + \alpha_k y_k) \ge f(x_k) + \omega' \alpha_k \langle y_k, \nabla f(x_k) \rangle. \quad (11)$$

In Quasi-Newton algorithms, the descent direction y_k is defined by the equation:

$$M_k y_k + \nabla f_k = 0. \quad (12)$$

where M_k is a symmetric and positive definite matrix. The acceptance of $x_k + y_k$ as a next iterate is a consequence of the followinf theorem given by Dennis and Moré (see [4]):

Theorem 1.1 *Let f be of class $\mathcal{C}^3$. Consider the iteration $x_{k+1} = x_k + \alpha_k y_k$, where y_k has the form $-M_k^{-1}$ is the descent direction defined by (12), and α_k satisfies (9) for $\omega \le \frac{1}{2}$ and (10) for some $\omega' \in]\omega, 1[$. If (x_k) converges to a stationnary point $\hat{x}$ such that $H(\hat{x})$ is positive definite, and if:*

$$\lim_{k \to \infty} \frac{\|(M_k - H(\hat{x}))y_k\|}{\|y_k\|} = 0, \quad (13)$$

then:

- *For k sufficiently large, the unite step-size $\alpha_k = 1$ is accepted as a Wolfe step-size.*
- *If $\alpha_k = 1$, for k sufficiently large, then (x_k) converges to $\hat{x}$ superlinearly.*

In the light of this theorem we will discuss acceptance of unit step-size for non-linear least-squares descent algorithm. For the Newton direction y_N, (13) is clearly satisfied, so the unit step-size has a good chance of being accepted in a neighborhood of the solution. However computation of y_N requires the Hessian $H(x)$, which is generally expensive. Using approximations of $H(x)$ are useful to reduce the computational cost. One approximation is:

$$H(x) \approx G(x), \text{ with } G(x) = J(x)^t J(x) \quad (14)$$

Due to (2), we observe that when F is small or almost affine, $G(x)$ is considered to be a good approximation of $H(x)$. The use of $G(x)$ gives rise to the so-called Gauss-Newton direction y_{GN}, computed form the linear equation:

$$G(x) y_{GN} + \nabla f(x) = 0. \quad (15)$$

The matrix $G(x)$ is symmetric and positive semi-definite. It is positive-definite if $J(x)$ is injective. To check acceptance of the unit step-size for Gauss-Newton direction, we will check the condition (13). It means that $G(x_k)$ becomes increasingly an accurate approximation of $H(\hat{x})$ along the direction y_k. But if $G(x_k)$ is not a good approximation of H_k, this property seems difficult to be satisfied, and we are not sure of the acceptance of the unit step-size. Another famous algorithm, from the same family, is the Levenberg-Marquardt algorithm. It performs a descent direction using a matrix $L(x)$:

$$L(x) = G(x) + \lambda I, \tag{16}$$

with $\lambda \geq 0$, called the damping factor, and I is the identity matrix. The Levenberg-Marquardt direction y_{LM} is deduced from:

$$L(x)y_{LM} + \nabla f(x) = 0. \tag{17}$$

This algorithm could be seen also as an approximated Newton algorithm for a least-squares problem. The matrix L is symmetric and positive definite for all $\lambda > 0$. For the same reason as for the Gauss-Newton algorithm, the unit step-size is still not surely accepted here. Even more, in the steepest descent algorithm the unit step-size is used, if we try to see that in the light of the theorem (1.1) we find that the condition (13) is generally far from being satisfied by $M_k = I$, then we can expect that the unit step-size could be rejected as a first step-size guess. There is no theoretical result justifying the step-size guess $\alpha = 1$ for Gauss-Newton and Levenberg-Marquardt algorithms, nevertheless the fact that it is the only one mentioned in most of optimization documents, and it is widely used. When this step-size guess is chosen but not accepted, most of the algorithms reduce the step-size by backtracking and interpolating and test acceptance on every new step-size value.

Consequently an accepted step-size guess for Gauss-Newton algorithms is a vital question in order to reduce the price, paid for bad step-size guess, in supplementary evaluations of the function F.

Here we propoe new step-sizes based on the curvature or on the projected curvature, in some sens to precise, at the current point. These steps are more adaptable on variations of curves forme than the unit step-size is. The radius of curvature R or the radius of projected curvature R_{pr} is evaluated, and then used to guess a step-size α. The advantage of the method is the total separation of two problems, computing a descent direction, and supplying an accepted step-size for the computed direction. The curvature and projected curvature steps could be applied for every descent direction y. If applied the convergence of the algorithm should be revised in the light of Zoutendijk lemma ([3]). Here we prove convergence for the steepest descent, Gauss-Newton, and Levenberg-Marquardt directions.

2 Curvature steps

In this section, we introduce the (*Maximum Projected Curvature Step (MPCS)* as an extension of the Maximum Curvature Step (MCS). We omit some details here, where the most of them could be found in ([1] and [2]).

2.1 Moving on a path of bounded projected curvature

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

$$\|v\| = 1, \text{ and } \langle a, v \rangle = 0. \tag{18}$$

The curvature radius $\rho \in]0, \infty]$ is defined by:

$$\frac{1}{\rho(\nu)} = \|a(\nu)\|. \tag{19}$$

When the curvature $a(\nu)$ is not nul, the osculating plan, denoted by $\mathcal{S}$, is the plan:

$$\mathcal{S}(\nu) = Vect\{v(\nu), a(\nu)\}. \tag{20}$$

We call optimization plan, denoted by $\mathcal{O}$, the plan:

$$\mathcal{O}(\nu) = Vect\{p(\nu), v(\nu)\}. \tag{21}$$

if the $p(\nu)$ and $v(\nu)$ are not paralell, otherwise $\mathcal{O}(\nu) = \mathcal{S}(\nu)$ (when the curvature is not nul). When the curvature is nul, we consider that every plan containing p and v is an osculating and an optimization plan. The results for this case could be considered as asymtotic ones for the case of non nul curvature corresponding to

$\|a\| \to 0$. For the following we shall consider that the curvature is not nul. We define the vector:

$$n(\nu) = \begin{cases} \frac{a}{\|a\|}(\nu), \text{ if } p \parallel v \text{ else:} \\ \frac{p-\langle p,v\rangle v}{\|p-\langle p,v\rangle v\|}(\nu), \end{cases} \tag{22}$$

We introduce the *projected* curvature, denoted by $a_{pr}(\nu)$, to be the projection of $a(\nu)$ on the optimization plan $\mathcal{O}(\nu)$, it equals:

$$a_{pr}(\nu) = \langle a, n\rangle(\nu) n(\nu), \tag{23}$$

We define the projected curvature radius $\rho_{pr}(\nu)$, by:

$$\frac{1}{\rho_{pr}(\nu)} = \|a_{pr}(\nu)\|. \tag{24}$$

We denote by $\psi(\nu)$ the angle between the optimization and the osculating plans. Then clearly we have:

$$\rho(\nu) = \rho_{pr}(\nu)\cos(\psi(\nu)). \tag{25}$$

The origin p_0 and the initial velocity v_0 are given:

$$p_0 \in \mathbb{R}^q \ , \ v_0 \in \mathbb{R}^q \ , \tag{26}$$

We suppose that v_0 is a descent direction for the residual function $r(\nu) = \|p(\nu)\|$:

$$\langle p_0, v_0\rangle < 0 \ . \tag{27}$$

We introduce the set:

$$\begin{aligned} \mathcal{P} \ = \ & \{ \ p \in W^{2,\infty}(\mathbb{R}^+) \ \text{ such that:} \\ & p(0) = p_0 \ , \ v(0) = v_0 \}. \end{aligned} \tag{28}$$

The first stationary point of r is defined by:

$$\bar{\nu} = Inf\{\nu \geq 0 \text{ such that } \frac{d}{d\nu}(r^2) = 0\} \ , \tag{29}$$

the corresponding residual is denoted $r(\overline{\nu}) = \bar{r}$. We will study the path $p(\nu)$ for arc lengths $\nu \in [0, \overline{\nu}]$. For a path $p \in \mathcal{P}$, we denote by:

$$\nu_L(\nu) \ = \ |\langle p, v\rangle|(\nu), \ \text{, and} \tag{30}$$

$$r_L(\nu) \ = \ \sqrt{r^2(\nu) - \nu_L^2(\nu)}, \tag{31}$$

the linearized displacement and residual for the current point $p(\nu)$ respectively. Studying the linearized residual $r_L(\nu)$ on $[0, \overline{\nu}]$ gives the following inequality for $\overline{\nu}$ (see [1] and [2]):

Theorem 2.1 *For a path $p \in \mathcal{P}$, let $R \in]0, \infty]$ be such that:*

$$\forall \nu \in [0, \overline{\nu}] \ , \ \rho_{pr}(\nu) \geq R. \tag{32}$$

Then the arclength $\overline{\nu}$ of the first stationary point of p satisfies:

$$\overline{\nu} \geq \ \bar{\nu}_M(R), \tag{33}$$

$$\bar{r} \leq r(\bar{\nu}_M(R)) \leq \bar{r}_M(R) \tag{34}$$

where:

$$\bar{\nu}_M(R) = R \arctan \frac{\nu_L(0)}{R + r_L(0)}, \ \textit{and} \tag{35}$$

$$\bar{r}_M(R) = ((R + r_L(0))^2 + {\nu_L}^2(0))^{\frac{1}{2}} - R, \tag{36}$$

denote respectively the worst displacement on p without encountering the first stationary point, and the corresponding worst residual.

We notice that, in practice, the global information (32) about p is difficult to get. But a local information is still useful as it is shown in the following proposition:

Proposition 2.2 *Let $p \in \mathcal{P}$ and $R_{pr} > 0$ satisfy:*

$$\rho_{pr}(\nu) \geq R_{pr} \quad \forall \nu \in [0, \bar{\nu}_M(R_{pr})] \ , \tag{37}$$

where ρ_{pr} is the radius of projected curvature (24). Then:

$$\bar{\nu} \geq \bar{\nu}_M(R_{pr}) \ \ , \tag{38}$$

$$\bar{r} \leq r(\bar{\nu}_M(R_{pr})) \leq \bar{r}_M(R_{pr}) \ \ . \tag{39}$$

If we replace the condition (37) by:

$$\rho(\nu) \geq R \quad \forall \nu \in [0, \bar{\nu}_M(R)] \ , \tag{40}$$

then the results (38) and (39) hold for R. In fact (25) shows that (40) implies (37) for R. The largest value $\tilde{R}_{pr}$ of R_{pr} satisfying (37) on $[0, \bar{\nu}_M(\tilde{R}_{pr})]$ realizes the smallest value of $\bar{r}_M(R_{pr})$. So it is a solution to:

$$\tilde{\nu} = \bar{\nu}_M(\widetilde{R}) \quad , \quad \widetilde{R}_{pr} = R_{m,pr}(\tilde{\nu}) \ . \tag{41}$$

where:

$$R_{m,pr}(\nu) = Inf\{\ \rho_{pr}(\tau)\ ,\ \tau \in [0,\nu]\}\ \ . \quad (42)$$

refers to the smallest radius of projected curvature on p up to ν. If we denote by:

$$R_m(\nu) = Inf\{\ \rho(\tau)\ ,\ \tau \in [0,\nu]\}\ \ . \quad (43)$$

From (25), it follows that:

$$R_{m,pr}(\nu) \geq R_m(\nu), \quad (44)$$

but $\bar{\nu}_M$ is an increasing function while $\bar{r}_M$ is a decreasing one, then using the projected curvature to define the step enables to go farther towords the objective point, and to get a better decrease of the residual than using the curvature could do. Now we can define the maximum *projected* curvature step (respectively maximum curvature step) using formula (35) applied for some R_{pr} (respectively R) satisfying (37) (respectively (40)).

2.2 Application on the image of a descent curve

We consider in this section the resolution of the least squares problem (1) by a descent algorithm. At a point x, let $\alpha \rightsquigarrow g(\alpha)$ be a descent curve in the parameter space, for a descent direction y. We associate with g a path $\tilde{p}$ in the data space defined by:

$$\tilde{p}(\alpha) = F(g(\alpha)) \quad \forall \alpha \geq 0\ , \quad (45)$$

and we denote by

$$\nu(\alpha) = \int_0^\alpha \|\tilde{p}'(\tau)\|\ d\tau \quad (46)$$

the arc length function along $\tilde{p}$. Now let p be reparameterization of $\tilde{p}$ by the arclength ν. The averge value of $\|\tilde{p}'\|$ on $[0,\alpha]$ is denoted by $\overline{\|\tilde{p}'\|}(\alpha)$, it equals:

$$\overline{\|\tilde{p}'\|}(\alpha) = \frac{\nu(\alpha)}{\alpha}. \quad (47)$$

The curve g and the mapping F are supposed regular enough to satisfy:

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

First we evaluate:

$$\begin{aligned} \tilde{p}(0) &= F(x_k)\ , \\ \tilde{p}'(0) &= F'(x_k).y_k\ , \\ \tilde{p}''(0) &= F''(x_k).(y_k, y_k) + F'(x_k).g''(0). \end{aligned} \quad (49)$$

then vectors determining osculating and optimization plans of p, defined in the subsection 2.1 are given by:

$$\begin{aligned} p(0) &= \tilde{p}(0)\ , \\ v(0) &= \tilde{p}'(0)/\|\tilde{p}'(0)\|\ , \\ a(0) &= (\tilde{p}''(0) - \langle \tilde{p}''(0), v(0)\rangle v(0))/\|\tilde{p}'(0)\|^2. \\ n(0) &= (p - \langle p, v\rangle v)(0)/\|p - \langle p, v\rangle v\|(0) \end{aligned} \quad (50)$$

Then we deduce from the formula (22) and (23) the projected curvature $a_{pr}(0)$ and its radius $\rho_{pr}(0)$.

Now we can define our step-sizes:

Definition 2.3 *Let R_{pr} be a lower bound for ρ_{pr} satisfying the condition (37). We define the maximum projected curvature step-size* ***MPCS*** *α_M to be the solution of the equation:*

$$\begin{aligned} \nu(\alpha_M) &= \bar{\nu}_M(R_{pr}) \\ &= R_{pr} \arctan \frac{\nu_L(0)}{R_{pr} + r_L(0)}. \end{aligned} \quad (51)$$

The step-size defined by the equation (51), where a lower bound R satisfying the condition (40) is used for, this step-size is called the maximum curvature step ***MCS****.*

The following theorem available for both **MCS** and **MPCS**.

Theorem 2.4 *We denote by R_m the real satisfying one of the two conditions (37) or (40), and we denote α_M the cooresponding step-size defined by (51). Then:*

$$\nu(\alpha_M) \leq \bar{\nu}\ , \quad (52)$$

$$f(g(\alpha_M)) \leq f(x) + \alpha_M \omega(\alpha_M) f'(x).y\ , \quad (53)$$

where $\alpha \rightsquigarrow \omega(\alpha)$ is defined by:

$$\omega(\alpha) = \frac{1}{2} \frac{\overline{\|\tilde{p}'\|}(\alpha)}{\|\tilde{p}'(0)\|}, \quad (54)$$

It satisfies:

$$\lim_{\alpha \to 0} \omega(\alpha) = \frac{1}{2}. \quad (55)$$

The inequality (53) is a decreasing condition satisfied then by $\alpha_M(R_m)$ when R_m satisfies one of two inequalities (37) and (40). The value of the decreasing coefficient ω usually used for the decreasing condition (9) is $\omega = 10^{-4}$, we deduce that $\alpha_M(R_m)$ has a good chance of being accepted. The limite (55) talls that small step-size are surely accepted.

We discuss now how to find numerically R_m and then deduce $\alpha_M(R_m)$. We choose to do this for the step-size **MPCS**, as the two step have similar definitions. We take R_m of the form:

$$R_m = \kappa_0 \tau^i \rho_{pr}(0) \quad \text{with} \quad 0 < \tau < 1, \qquad (56)$$

where $0 \leq \kappa_0$ depends on the choosen step-size, it equals to 0.9 for the **MPCS** and to 1.5 for the **MCS**. The integer i depends on ρ_{pr} variations, its value is increased if the decreasing condition (9) is not satisfied by $\alpha_M(R_m)$. So $\kappa_0 \tau^i$ is a security factor which accounts for the possible increase of the projected curvature along the path.

In order to compute the maximum projected curvature step we use the linear approximation of $\alpha \rightsquigarrow \nu(\alpha)$:

$$\nu(\alpha) \approx \alpha \, \|\tilde{p}'(0)\|, \qquad (57)$$

in which case α_M is given by:

$$\alpha_M \|\tilde{p}'(0)\| = \bar{\nu}_M(R). \qquad (58)$$

We have now all we need to introduce our new algorithms, and to study their convergence.

3 Curvature and projected curvature algorithms, discription and convergence

These two algorithms are named corresponding to the step-size used. First we give the following general discription of both of them then we will study their global convergence.

Algorithm 3.1 *(Curvature Algorithms)*

1. *Initialization:* $k = 0$, x_0.
2. *While* $(k \leq k_{\max})$
   (a) *Comput* F_k, J_k, *and* ∇f_k.
   (b) *If* $(\nabla f_k \leq \epsilon_{df})$ *then* $x^* = x_k$
   (c) *Else evaluate for the descent curve* g_k *the quantities:* $\tilde{p}_k(0)$, $\tilde{p}'_k(0)$, $\tilde{p}''_k(0)$, $p_k(0)$, $v_k(0)$, *and* $a_k(0)$.
   (d) *Evaluate* $\rho_k(0)$ *for* ***MCS*** *or* $\rho_{pr,k}(0)$ *for* ***MPCS***,
   (e) *Initialization* $i = 0$.
   (f) *While* $(i \leq i_{max})$
      i. *Evaluate* R_m *by* (56) *and* α_M *by* (51) *and* (58).
      ii. *If the condition* (9) *is satisfied then put* $R_{m,k} = R_m$ *and* $\alpha_k = \alpha_M$ *and go to* (2g).
      iii. *Else increase* i.
   (g) *Calculate the next iterate:*

   $$x_{k+1} = g_k(\alpha_k).$$

   (h) *increase* k.

The main tool to study the convergence of a descent algorithm is the Zoudendijk condition ([3]). This convergence depends on proprieties of the sequence $(R_{m,k})_{k\geq 0}$ defined previously at $(2(f)ii)$. All the results of this section are available for both **MCS** and **MPCS**.

Proposition 3.2 *(Zoutendijk Condition)* *Let the sequence* x_k *satisfy the following two conditions:*

1. *There is a strictly positive real* C *such that for all* k*:*

$$f(x_{k+1}) \leq f(x_k) - C\|\nabla f(x_k)\|^2 \cos^2 \theta_k. \qquad (59)$$

2. *The sequence* $(f(x_k))_{k\geq 0}$ *is bounded from below.*

Then the series $\sum_{k\geq 0} \|\nabla f(x_k)\|^2 \cos^2 \theta_k$ *converges. If furthermore the sequence* $\cos^2 \theta_k$ *bounded away from zero then:*

$$\lim_{k\to\infty} \nabla f(x_k) = 0. \qquad (60)$$

Proposition 3.3 *Let* $(x_k)_{k\geq 0}$ *be a sequence generated by a descent algorithm of maximum projected curvature step to solve problem* (1). *If the following facts hold :*

1. *The sequence of linear applications* $(F'(x_k))_{k\geq 0}$ *is* M *bounded.*
2. *The sequence* $(R_{m,k})_{k\geq 0}$ *is bounded away from zero by* R_{min}.

Then we have for every iteration k:

$$f(x_{k+1}) \leq f(x_k) - C\|\nabla f(x_k)\|^2 \cos^2\theta_k.$$

Where θ_k *describes the angle between the gradient* $\nabla f(x_k)$ *and the descent direction* y_k. *And the constant* C *could have the value:*

$$C = \frac{\pi}{8M^2}\frac{R_{min}}{R_{min} + \|F(x_0)\|} \tag{61}$$

To get convergence, we need to study the descent angle θ_k for the chosen descent direction.

Theorem 3.4 *The quantity* $\cos(\theta_k)$ *is lower bounded below by a strictly positive number* B *in any of the following cases*

1. *The descent direction is the steepest descent, then we have* $B = 1$.
2. *If the sequence of Jacobians* $(F'(x_k))_{k\geq 0}$ *is bounded by a constant* M, *and uniformly* β *injective, then for the Gauss-Newton direction we have* $B = (\beta/M)^2$.
3. *If the sequence of Hessians* $(f''(x_k))_{k\geq 0}$ *is bounded by* M *and et uniformly* β *coercive, then for the Newton direction we have* $B = (\beta/M)$.
4. *If the sequence of Matrix* $(M_k)_{k\geq 0}$ *is bounded by a constant* M *and et uniformly* β *coercive, then for the Quasi-Newton direction we have* $B = (\beta/M)$.
5. *For Levenberg-Marquardt direction, if the damping factor* λ_k *equals to :*

$$\lambda_k = \frac{B}{1-B}\|F_k^{'t}F_k'\|, \tag{62}$$

then $\cos(\theta_k)$ *is bounded below by* B.

We can see that the Levenberg-Marquardt direction gives a choice of the lower bound B, and it does not need the Jacobian to be injective. We shall see in the following section results for some of the previous directions.

4 Numerical Experiments

We compare the behaviour of five algorithms on a testing example. These algorithms are :

- Armijo Algorithm that uses a unit step-size as an initial guess, if it is rejected quadrature interpolation and backtracking are applied to find an adequate step-size.
- Wolfe algorithm based on testing both conditions (9) and (10) to accept a step-size, but if it is not accepted the algorithm apply backtracking and cubic interpolation to find a new one.
- Goldstein algorithm, it replaces Wolfe condition (10) by Goldstein condition (11), and it applies backtracking and quadratic interpolation to look for a new step-size if the guessed value, initiated to one, is not accepted.
- The Maximum curvature step-size (MCS), see ([1]).
- The Maximum projected curvature step-size (MPCS) introduced in section 2.

One of the five following raisons implies to stop the iterative process:

- In the main loop, where the descent direction y is copmuted, the maximum number of iteration $k_{\max}$ is reached ($T = 0$).
- In the inner loop, where a step-size is performed, the maximum number of step-size evaluation $i_{\max}$ is reached ($T = 1$).
- The norm of the gradient ∇f_k is less than a given precision $\epsilon_{\nabla f}$ ($T = 2$).
- The variation in the objective function $f_k - f_{k+1}$ for an accepted step-size is less than a given precision ϵ_f ($T = 3$).
- The norm of $\alpha_k y_k$ is less than a given precision ϵ_x ($T = 4$).

For all tests we took:

$$(k_{\max}, i_{\max}) = (4000, 20),$$

and:

$$(\epsilon_{\nabla f}, \epsilon_f, \epsilon_x) = (10^{-6}, 10^{-24}, 10^{-24})$$

We mentionne that stoping for the last three reasons means that we have the best result for a given choice of precisions. We consider that there is no convergence if one of the first two cases holds. In the tables 1, 2, and 3, we have used the following notations:

- N_1 refers to the number of descent direction calculus iterations.
- N_2 indicates the number of model evaluations.
- N_3 is the number of evaluated and non accepted step-sizes.

4.1 Acceptance of Curvature Steps and the unit step-size for different descent directions

We choose to invert a regularized version of the Powell example F ([5]), because it has a modulable difficulty. It is defined by:

$$F = \Phi - d \ , \tag{63}$$

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

$$\Phi(x_1, x_2) = \begin{pmatrix} x_1 \\ \frac{10\,x_1}{x_1+1} + 2x_2^2 \\ \epsilon\, x_2 \end{pmatrix} \ , \tag{64}$$

The vector d represents the data to be inverted:

$$d = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \tag{65}$$

Since d does not belong to $\Phi(\Omega)$, we look for its projection, for which the minimum residual will be strictly positive. For the two first tests we put $\epsilon = 0.01$, but for the theird ϵ is equal to zero. Results of five descent algorithms with the Gauss-Newton direction are given in table 1. Here Goldstein, MCS, and MPCS convege, and MPCS algorithm realizes the least cost.

Table 1: Gauss-Newton Algorithms

Algorithm	N_1	N_2	N_3	T
Armijo	4000	53794	45794	0
Wolfe	4000	31958	23958	0
Goldstein	1318	4157	1522	2
MCS	1548	6054	1412	2
MPCS	262	985	199	3

If we check acceptance of the step-size for the last iterations on figures 1, 2 and 3, we see that

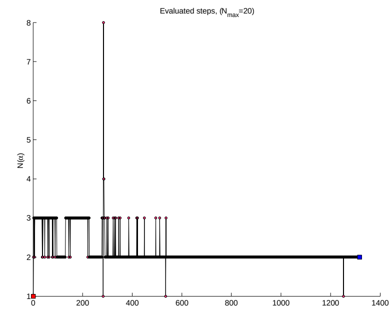

Figure 1: Goldstein algorithm; evaluated steps

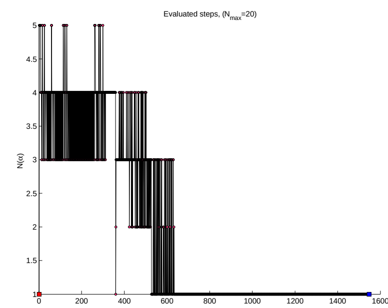

Figure 2: MCS algorithm; evaluated steps

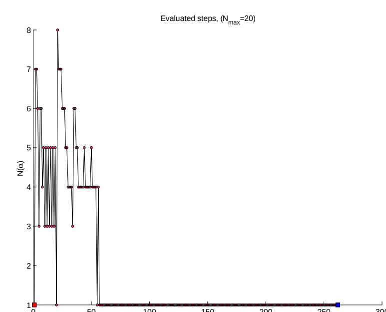

Figure 3: MPCS algorithm; evaluated steps

first evaluation of curvature step-sizes is accepted in the neighborhood of the solution, and we dont have to reduce the step-size, it is not the case for Goldstein step-size, it still needs two evaluations to get an acceptable step-size. This fact could be seen clearly in the next example, where the steepest descent direction is used.

Results in table 2 show that we have convergence of four algorithms, where the best algorithm is the MPCS algorithm, it realizes the least cost. Both MCS and MPCS are often accepted:

Table 2: Steepest Descent Algorithms

Algorithm	N_1	N_2	N_3	T
Armijo	703	4882	3477	2
Wolfe	565	2023	894	2
Goldstein	4	31	23	1
MCS	91	272	1	2
MPCS	75	223	0	2

In the last example, the third component of Φ is set to zero $\epsilon = 0$, so no information on x_2 is available. The descent direction is set to be Levenberg-Marquardt direction, where the damping coefficient is deduced from (62) for $B = 0.1$. The results are given in table 3, where a reference trust region algorithm for least squares problems, denoted by RC, was called in Matlab environment. Results show the relative efficeincy of MCS and MPCS, they realize a low iteration number N_1.

Table 3: Levenberg-Marquardt Algorithms

Algorithm	N_1	N_2	N_3	T
Armijo	188	375	0	2
Wolfe	188	375	0	2
Goldstein	188	375	0	2
MCS	35	103	0	2
MPCS	79	235	0	2
RC	61	62	-	4

Conclusion

Both maximum curvature step-size and maximum projected curvature step-size show a good behaviour with a frequent advantage for the second. The convergence and the adaptativity of both of them with several descent directions is shown theoriticaly and numerically. They have more chance than the unit step-size of being accepted when other descent directions than the Newton direction is used in a descent algorithm. A descent algorithm of Levenberg-Marquardt direction with any of curvature steps is a rubustic algorithm for solving strong non linear least squares, it gives good results even if while computing the Jacobians become singular.

Acknowledgment

The authors thank Jean Charles Gilbert for fruitful discussions.

References

[1] G. Chavent: *Curvature steps and geodesic moves for nonlinear least squares descent algorithms*, Inverse Problems in Engineering.

[2] G. Chavent: *Curvature steps and geodesic moves for nonlinear least squares descent algorithms*, Inria Report, 2002.

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

[4] J. E. Dennis, and Jorge J. Moré:*Quasi-Newton Methods, Motivation and Theory*, SIAM Review, Vol. 19, No. 1, January 1977.

[5] M. J. D. Powell:*A hybrid method for non-linear equations*, in *"Numerical Methods for Non-linear Algebric Equations"*, *Ed.* P. Rabinowitz, Gordon and Breach, London, 1970.