AN ALGORITHM FOR SOLVING NONLINEAR LEAST-SQUARES PROBLEMS WITH A NEW CURVILINEAR SEARCH

José Mario Martínez

and

Rita Filomena Santos

RELATÓRIO TÉCNICO Nº 20/89

ABSTRACT: We propose a modification of an algorithm introduced by Martínez (1987) for solving nonlinear least-squares problems. Like in the previous algorithm, after the calculation of an approximated Gauss-Newton direction d, we obtain the next iterate on a two-dimensional subspace which includes d. However, we simplify the process of searching the new point, and we define the plane using a scaled gradient direction, instead of the original gradient. We prove that the new algorithm has global convergence properties. We present some numerical experiments.

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Ciência da Computação
Caixa Postal 6065
13.081 - Campinas - SP
BRASIL

Junho - 1989

AN ALGORITHM FOR SOLVING NONLINEAR LEAST-SQUARES PROBLEMS WITH A NEW CURVILINEAR SEARCH

by

José Mario Martínez [1]

and

Rita Filomena Santos [1]

ABSTRACT: We propose a modification of an algorithm introduced by Martínez (1987) for solving nonlinear least-squares problems. Like in the previous algorithm, after the calculation of an approximated Gauss-Newton direction d, we obtain the next iterate on a two-dimensional subspace which includes d. However, we simplify the process of searching the new point, and we define the plane using a scaled gradient direction, instead of the original gradient. We prove that the new algorithm has global convergence properties. We present some numerical experiments.

Key Words: Nonlinear least squares, Gauss-Newton Method, Curvilinear search.

[1] Applied Mathematics Laboratory, IMECC-UNICAMP, CP 6065, 13081 - Campinas, SP., Brazil.

1. INTRODUCTION

We consider the following problem:

$$\begin{array}{l} Minimize \quad \frac{1}{2}||F(x)||^2 \\ x \in \mathbb{R}^n \end{array} \tag{1}$$

where $F = (f_1, \ldots, f_m)^T$ is a C^1-function and $||\cdot||$ represents the euclidean norm in $\mathbb{R}^n$.

Martínez [2] introduced a new method for solving (1), which is specially suited for problems where m, n are large and the Jacobian matrix $J(x)$ is sparse.

The main features of Martínez's method are the following:

a) The Gauss-Newton equation is "partially" solved at each iteration using a preconditioned Conjugate Gradient algorithm.

b) The new point is obtained using a two-dimensional trust-region scheme.

The method we introduce in this paper differs from Martínez's method only in b).

In fact, we found from many numerical experiments that the two-dimensional trust-region strategy was not essential for the good behavior of the algorithm and that it could be replaced by a simpler and less prone to rounding errors curvilinear strategy.

On the other hand, instead of defining the search plane as the subspace spanned by the gradient of $||F(x)||^2$ and the truncated Gauss-Newton direction, we replace the first by a scaled gradient, which takes into account the relative size of the variables.

The new algorithm is described in Section 2, where we also prove a global convergence result, using the theory of [2]. Some numerical experiments are presented in Section 3. Finally, in Section 4, we state some conclusions, and we suggest the lines for future research.

2. THE NEW ALGORITHM

Let $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, $m \geq n$, $F \in C^1(\Omega)$, Ω an open set. Let $x^0 \in \Omega$ an arbitrary initial point, $\eta \in [0, 1)$, $\theta_1, \theta_2 \in [0, 1)$, $\theta_3 \in (0\ \frac{1}{2})$, $\overline{M} > 0$, $\underline{M} \in [\theta_1 \overline{M}, \overline{M}]$

Algorithm 2.1

Let x^k be the k-th approximation to the solution. We denote

$$F_k = F(x^k),\ J_k = J(x^k),\ g_k = J_k^T F_k = \nabla(\frac{1}{2}\ ||F(x)||^2)|_{x^k}\ ,$$

$D_k = Diag(d_1^k, \ldots, d_n^k)$, where

$$d_i^k = \begin{cases} (x_i^k)^2 & \text{if} \quad (x_i^k)^2 \in [\underline{M}, \overline{M}]\ , \\ \underline{M} & \text{if} \quad (x_i^k)^2 < \underline{M}\ , \\ \overline{M} & \text{if} \quad (x_i^k)^2 > \overline{M}. \end{cases}$$

Assume $g_k \neq 0$. To obtain x^{k+1} we perform the following steps:

Step 1: Compute J_k and g_k.

Step 2: Obtain $w_k \in I\!R^n$ such that

$$||J_k^T J_k w_k + g_k|| \leq \eta ||g_k||. \tag{2}$$

Step 3: Obtain $v_k \in I\!R^n$ the solution of the following two-dimensional problem:

$$\text{Minimize} \quad ||J_k v + F_k||$$

$$s.t. \quad v = \lambda_1 g_k + \lambda_2 w_k\ .$$

Step 4: Set $d_1^k = -D_k g_k$. Test the following two conditions for v_k

$$v_k^T g_k \leq -\theta_1 ||v_k||\ ||g_k|| \tag{3}$$

and

$$\underline{M}||g_k|| \leq ||v_k|| \leq \overline{M}||g_k||. \tag{4}$$

If (3) and (4) are satisfied, set $d_2^k = v_k$. Otherwise $d_2^k = d_1^k$.

Step 5: Set $t = 1$. Perform steps (5.a) to (5.d)

(5.a) Set

$$d = d(t) = t^2 d_2 + \frac{g_k^T d_2^k}{g_k^T d_1^k} t(1-t) d_1 \tag{5}$$

(5.b) If

$$\frac{1}{2}||F(x^k + d)||^2 \leq \frac{1}{2}||F(x^k)||^2 + \theta_2 g_k^T d \tag{6}$$

go to (5.d).

(5.c) Let $\tilde{t}$ be such that

$$\theta_3||d(t)|| \leq ||d(\tilde{t})|| \leq (1 - \theta_3)||d(t)|| \tag{7}$$

Replace t by $\tilde{t}$. Go to (5.a)

(5.d) $d^k = d$, $x^{k+1} = x^k + d^k$.

Theorem 2.1

The algorithm 2.1 is well-defined.

Proof.

We want to prove that, if $g_k \neq 0$, we are able to arrive to Step 5.d in finite time. So, let us verify that it is possible to complete each step of the algorithm.

Step 2: Observe that the system of (normal) equations

$$J_k^T J_k w + J_k^T F_k = 0$$

always admits a solution. Therefore, it is possible to find w_k satisfying (2).

Step 3: Clearly, the solution of the two-dimensional problem is $\lambda_1 g_k + \lambda_2 w_k$, where $J_k(g_k, w_k)\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}$ is the orthogonal projection of - F_k on the column subspace of $J_k(g_k, w_k)$.

Step 4 does not present any problem. Let us analyze Step 5. Let us write

$$d = d(t) = t^2 d_2 + at(1 - t)d_1 \tag{8}$$

where $a = \dfrac{g_k^T d_2^k}{g_k^T d_1^k}$. By the definition of d_1^k and (3), we have that $a > 0$. By (7), we only need to prove that (6) is satisfied for small enough t.

In fact, by the Mean Value Theorem,

$$\frac{1}{2}||F(x^k + d(t))||^2 - \frac{1}{2}||F(x^k)||^2 = g(x^k + \xi(t)d(t))^T d(t) \tag{9}$$

where $g(x)$ denotes $\nabla(\frac{1}{2}||F(x)||^2)$ and $0 \leq \xi(t) \leq 1$.

But by (5) $d(t)$ is a positive combination of d_1 and d_2 and $g_k^T d_1 > 0 < g_k^T d_2$. Therefore $g_k^T d(t) < 0$ for $t \in [0,1]$.

Hence, by (9)

$$\frac{\frac{1}{2}||F(x^k + d(t))||^2 - \frac{1}{2}||F(x^k)||^2}{g_k^T d(t)} = \frac{g(x^k + \xi(t)d(t))^T d(t)}{g_k^T d(t)} \tag{10}$$

Taking limits on both sides of (10), we obtain

$$\frac{\frac{1}{2}||F(x^k + d(t))||^2 - \frac{1}{2}||F(x^k)||^2}{g_k^T d(t)} = 1$$

Thus, given $\theta_2 \in (0,1)$, there exists $\hat{t} > 0$ such that

$$\frac{\frac{1}{2}||F(x^k + d(t))||^2 - \frac{1}{2}||F(x^k)||^2}{g_k^T d(t)} \geq \theta_2$$

for $t \in (0, \hat{t})$.

Therefore, using $g_k^T d(t) < 0$, we obtain (6).

This completes the proof. □

Theorem 3.2

Assume that (x^k) is generated by algorithm 3.1. Then:

(a) If there exists $c > 0$ such that $||g_k|| \leq c$ for all $k = 0, 1, 2, \ldots$ and $x^* \in \Omega$ is a limit point of (x^k), then $J(x^*)^T F(x^*) = 0$.

(b) Let $\epsilon > 0$. If $\{x \in \Omega : ||F(x)||^2 \leq ||F(x^0)||^2\}$ is compact, then there exists $k \in \mathbb{N}$ such that $||J(x^k)^T F(x^k)|| \leq \epsilon$

(c) Let x^* be a strict local minimizer of f in Ω, $\epsilon > 0$. Then, there exists $\epsilon_1 > 0$ such that $||x^k - x^*|| \leq \epsilon$ for all $k \geq k_0$, provided $||x^{k_0} - x^*|| \leq \epsilon_1$.

(d) If x^* is a strict local minimizer of $||F(x)||^2$ in Ω, then there exists $\epsilon > 0$ such that $lim\ x^k = x^*$, whenever $||x^0 - x^*|| \leq \epsilon$.

Proof.

We will show that Algorithm 2.1 is a particular case of a slight extension of Algorithm 3.1 of [2]. This extension consists in replacing the inequality $||d^k|| \leq ||d_2^k||$ in (6), by $||d^k|| \leq K||d_2^k||$ for some constant K, independent of k, for all $k = 0, 1, 2, \ldots$. Revising the proof of Theorem 3.1 of [2], we easily verify that it remains valid for this extension. The same holds for Corollary 3.1, Lemma 3.1 and Theorem 3.2 of [2]. Let us now prove that our Algorithm 2.1 is in fact a particular case of the extended Algorithm 3.1 of [2].

Define $f(x) = \frac{1}{2}||F(x)||^2$, $g(x) = \nabla f(x)$.

From (3), (5) and the definition of d_1, we verify that $d^k \in C(d_1^k, d_2^k)$. Let us verify that d_1^k satisfies:

$$g_k^T d_1^k \leq -\theta_1 ||d_1^k|| \, ||g_k|| \tag{11}$$

In fact,

$$\frac{|g_k^T d_1^k|}{||g_k|| \, ||d_1^k||} = \frac{|g_k^T D_k g_k|}{||g_k|| \, ||D_k g_k||} \geq \frac{\underline{M}||g_k||^2}{\overline{M}||g_k|| \, ||g_k||} = \underline{M}/\overline{M} \geq \theta_1.$$

Hence, (11) is proved.

Now, by (3), (4) and the choice of d_2^k we also have:

$$g_k^T d_2^k \leq -\theta_1 ||d_2^k|| \, ||g_k||.$$

So, the axiom (7) of [2] is satisfied.

By the definition of d_1^k, we have

$$\underline{M}||g_k|| \leq ||d_1^k|| \leq \overline{M}||g_k||.$$

Hence, by (3), (4), the axiom (8) of [2] is also satisfied. Moreover, by (6), axiom (9) of [2] also holds. It only remains to prove the inequality

$$||d^k|| \leq K||d_2^k|| \tag{12}$$

Consider the expression (8) for $d(t)$. The derivative $d'(t)$ is:

$$d'(t) = 2td_2 + a(1 - 2t)d_1$$

Therefore, if $\gamma(t) = \|d(t)\|^2$, we have, for $t \in [0,1]$,

$$\begin{aligned}
\gamma'(t) &= 2\, d'(t)^T d(t) = 2(2td_2 - a(1-2t)d_1)^T (t^2 d_2 + at(1-t)d_1)) \\
&= 2(2t^3 d_2^T d_2 + 2at^2(1-t)d_1^T d_2 - at^2(1-2t)d_1^T d_2 - a^2 t(1-t)(1-2t)d_1^T d_1) \\
&= 4t^3\|d_2\|^2 + [4at^2(1-t) - 2at^2(1-2t)]d_1^T d_2 - 2a^2 t(1-t)(1-2t)\|d_1\|^2 \\
&\leq 4\|d_2\|^2 + 10a\|d_1\|\,\|d_2\| + 6a^2\|d_1\|^2 \\
&\leq 4\,\overline{M}^2\|g_k\|^2 + \frac{10\,\|g_k\|\,\|d_2\|\,\|d_1\|\,\|d_2\|}{\theta_1\,\|g_k\|\,\|d_1\|} + \frac{6\,\|g_k\|\,\|d_2\|\,\|d_1\|^2}{\theta_1\,\|g_k\|\,\|d_1\|} \\
&\leq 4\overline{M}^2\,\|g_k\|^2 + \frac{10\,\overline{M}^2\|g_k\|^2}{\theta_1} + \frac{6\,\overline{M}^2\|g_k\|^2}{\theta_1} \\
&\leq (4\overline{M}^2 + \frac{10\,\overline{M}^2}{\theta_1} + \frac{6\overline{M}^2}{\theta_1})\,\|g_k\|^2 = C_1\|g_k\|^2. \qquad (13)
\end{aligned}$$

Therefore, for $t \in [0,1]$,

$$\|d(t)\|^2 = \gamma(t) \leq \gamma(1) + \max_{t\in[0,1]} |\gamma'(t)| \leq$$

$$\leq \|d_2^k\|^2 + C_1\|g_k\|^2 \leq \|d_2^k\|^2 + \frac{C_1\,\|d_2^k\|^2}{\underline{M}^2}$$

$$= \left(1 + \frac{C_1}{\underline{M}^2}\right)\|d_2^k\|^2$$

Hence, (12) is satisfied with $K = \sqrt{1 + \frac{C_1}{\underline{M}^2}}$. This completes the proof. □

Remark.

Let us explain the geometrical meaning of formula (5).

We already know that $d(1) = d_2$, $d(0) = 0$, $d'(0) = ad_1$, $a > 0$ and that $d(t)$ lies in the positive cone generated by d_1 and d_2 for all $t \in [0,1]$. However, there are many simple curves which satisfy these properties, and we want to explain why we chose the expression $g_k^T d_2^k / g_k^T d_1^k$ for the parameter a.

Let h be the orthogonal projection of d_2 on the orthogonal complement of the line generated by d_1, related to the norm $\|.\|_{D_k^{-1}}$ ($\|z\|^2_{D_k^{-1}} = z^T D_k^{-1} z$ for all $z \in \mathbb{R}^n$). Therefore,

$$h = d_2 - \frac{d_2^T D_k^{-1} d_1}{d_1^T D_k^{-1} d_1} d_1$$

But $d_1 = -D_k g_k$, hence,

$$h = d_2 - \frac{g_k^T d_2}{g_k^T d_1} d_1$$

Each point z in the plane spanned by $\{d_1, h\}$ may be expressed as $z = y_1 d_1 + y_2 h$. d_2 corresponds to $y_1 = \frac{g_k^T d_2}{g_k^T d_1}$, $y_2 = 1$. We consider in this plane the parabola defined by:

$$P = \{z = y_1 d_1 + y_2 h \mid y_2 = \frac{g_k^T d_2}{g_k^T d_1} y_1^2\},$$

which is the simpler curve in the coordinates $(y_1\ y_2)$ with the desirable characteristics. After some calculations, we verify that this curve is precisely the one defined by formula (5).

3. NUMERICAL EXPERIMENTS

For the numerical implementation of Algorithm 2.1, we used the conjugate gradient strategy and preconditioning scheme given in [2]. The implementation of the new curvilinear strategy is straightforward. For obtaining $\bar{t}$ at step (5.c) of the algorithm we tried a cubic interpolation procedure (see [1]) and a bisection procedure, and we saw no meaningful differences between the results. We ran the new algorithm with the problems described in [2]. The results were very similar to the ones reported in [2] for Martínez's algorithm, both in number of iterations as in functional evaluations. However, the computer time was slightly smaller due to the simplification of the curvilinear procedure.

Hence, we decided to run a number of small-dimensional nonlinear least squares problems, in order to detect differences in the performance of the two algorithms. We used $\eta = 10^{-4}$, $\theta_1 = 10^{-7}$, $\theta_2 = 10^{-4}$, $\overline{M} = 10^3$, $\underline{M} = 10^{-3}$. The results are reported in Table 1. For each case, we report:

FUNCTION: Function number, corresponding to the test collection given in [4].

X 0: Initial point.

(NI, FE, FF): Number of iterations, function evaluations and final value of $||F(x)||^2$ respectively.

All the tests were run in a VAX11/785 at the State University of Campinas, using the Fortran 77 compiler, single precision and the VMS Operational System.

Function	m,n	x0	NI, FE, FF	
			Martínez	New
7	3,3	(-1, 0, 0)	21,34,0.	8, 10, .8E-27
8	15,3	(1, 1, 1)	5, 6, .8E-2	5, 6, .8E-2
9	15,3	(0.4, 1, 0)	2, 3, .1E-7	2, 3, .1E-7
11	6,3	(5, 2.5, 0.15)	24, 25, .9E-2	24, 25, .9E-2
12	9,3	(0, 10, 20)	7, 8, 0.	7, 8, 0.
13	4,4	(3, -1, 0, 1)	15, 16, .4E-13	15, 16, .4E-13
15	11,4	(.25,.39,.415,.39)	8, 12, .3E-3	6, 8, .3E-3
17	33,5	(.5,1.5,-1,0.01,0.02)	10, 15, .5E-4	15, 26, .5E-4
25	6,4	$(\frac{5}{6}, \frac{4}{6}, \frac{3}{6}, \frac{2}{6})$	9, 10, 0.	9, 10, 0.
26	6,6	$(\frac{1}{6}, \ldots, \frac{1}{6})$	11, 18, .3E-13	29, 37, .4E-12
31	6,6	(-1, ... , -1)	7, 8, .8E-13	7, 8, .8E-13

Table 1 – Performance of Martínez's method versus the new method.

4. CONCLUSIONS

In this paper we introduced a modification of the algorithm of Martínez [2] for solving nonlinear least-squares problems.

In the new algorithm we use a curvilinear search which, unlike the one used in [2], is not related to trust-region type ideas. However, we are able to prove a global convergence theorem. Therefore, we showed that, from the theoretical point of view, the

trust-region strategy of [2] is not essential to guarantee good convergence properties. This result was expected since, in [2], the convergence of the trust-region based algorithm is obtained as a particular case of a larger class, where the trust-region idea does not appear. The global convergence of the new algorithm is proved showing that it is a particular case of a slight generalization of that class.

Among the plane curves which join the current point x^k and the first trial point, the trust-region curve should be the best, in the sense that lower values of the objective function on this curve are expected. The numerical experiments seem to confirm this prediction, but the superiority of the trust-region strategy over the new curvilinear strategy in terms of iterations and functional evaluations is not very impressive. Moreover, in some cases, the new curvilinear strategy performed better than the trust-region strategy. On the other hand, the trust-region strategy is much more expensive than the new curvilinear search, hence, the last one represents a real practical improvement in many cases.

Moreover, as we mentioned in Section 4, we detected no differences in terms of iterations or function evaluations for large sparse problems. This fact may be explained as follows: In a small-dimensional problem, a two-dimensional trust-region strategy is more like an n-dimensional trust-region strategy than in a large-dimensional problem. As a extreme case, if $n = 2$, the two-dimensional strategy coincides with the complete trust-region strategy. Therefore, for small n, the two-dimensional strategy inherits a good amount of the excellent stability properties of trust-region algorithms (see [1]). Conversely, if n is large, the two-dimensional trust-region algorithm is very different from the n-dimensional one and so, it seems not to be important the choice of the plane curve used for finding the new point.

The first trial point at each iteration of our algorithm is chosen applying a conjugate gradient algorithm to the Gauss-Newton linear equation. Alternative choices should be investigated. For nonlinear systems of equations, Martínez [3] introduced recently a family of superlinearly convergent quasi-Newton methods based on direct secant updates of matrix factorizations. The aim of many methods of Martínez's family is to save computer time of linear algebra calculations, in the cases where evaluation of derivatives is not a problem. The methods of this family may be modified using curvilinear strategies like the one introduced in this paper, in order to obtain global convergence results. This is going to be the subject of future research.

ACKNOWLEDGEMENTS

We acknowledge FAPESP, CNPq, FINEP and CAPES for financial support.

REFERENCES

[1] J.E. Dennis and R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice Hall Series in Comput. Math., Prentice Hall, N.J., 1983.

[2] J.M. Martínez, "An Algorithm for solving Sparse Nonlinear Least-Squares Problems", Computing 39 (1987) pp. 307-325.

[3] J.M. Martínez, "A Family of Quasi-Newton Methods for Nonlinear Equations with Direct Secant Updates of Matrix Factorizations", SIAM J. Numer. Anal. (1989), to appear.

[4] J.J. Moré, B.S. Garbow and K.E. Hillstrom, "Testing Unconstrained Optimization Software", ACM TOMS 7 (1981) pp. 136-140.

RELATÓRIOS TÉCNICOS — 1989

01/89 — Uniform Approximation of Continuous Functions With Values in [0, 1] — *João B. Prolla.*

02/89 — On Some Nonlinear Iterative Relaxation Methods in Remote Sensing — *A. R. De Pierro.*

03/89 — A Parallel Iterative Method for Convex Programming with Quadratic Objective — *Alfredo N. Iusem* and *Alvaro R. De Pierro.*

04/89 — Fifth Force, Sixth Force, and all that: a Theoretical (Classical) Comment — *Erasmo Recami* and *Vilson Tonin-Zanchin.*

05/89 — An Application of Singer's Theorem to Homogeneous Polynomials — *Raymundo Alencar.*

06/89 — Summhammer's Experimental Test of the Non-Ergodic Interpretation of Quantum Mechanics — *Vincent Buonomano.*

07/89 — Privileged Reference Frames in General Relativity — *Waldyr A. Rodrigues Jr.* and *Mirian E. F. Scanavini.*

08/89 — On the Numerical Solution of Bound Constrained Optimization Problems — *Ana Friedlander* and *José Mario Martínez.*

09/89 — Dual Extremun Principles for the Heat Equation Solved by Finite Element Methods I — *Vera Lucia da Rocha Lopes* and *José Vitório Zago.*

10/89 — Local Convergence Theory of Inexact Newton Methods Based on Structured Least Chance Updates — *José Mario Martínez*

11/89 — Real Spin-Clifford Bundle and the Spinor Structure of Space-Time — *Waldyr A. Rodrigues Jr.* and *Vera L. Figueiredo.*

12/89 — A Multiplier Theorem on Weighted Orlicz Spaces — *B. Bordin* and *J. B. Garcia.*

13/89 — Dual Extremum Principles For The Heat Equation Solved By Finite Element Methods II — *Vera Lucia da Rocha Lopes* and *José Vitório Zago.*

14/89 — Dirac and Maxwell Equations in the Clifford and Spin-Clifford Bundels — *W. A. Rodrigues Jr.* and *E. Capelas de Oliveira.*

15/89 — Formal Structures, The Concepts of Covariance Invariance, Equivalent Reference Frames, and the Principle of Relativity — *W. A. Rodrigues Jr., M. E. F. Scanavini* and *L. P. de Alcantara.*

16/89 — Local Minimizers of a Quadratic Function With a Spherical Constraint — *José Mario Martínez.*

17/89 — On Pseudo-Convex Polycircular Domains In Banach Spaces — *Mário C. Matos.*

18/89 — On Circular and Special Units of an Abelian Number Field — *Trajano Nóbrega.*

19/89 — Implementing Algorithms for Solving Sparse Nonlinear Systems of Equations — *Márcia A. Gomes-Ruggiero, José Mario Martínez* and *Antonio Carlos Moretti.*