A GLOBALLY CONVERGENT NONLINEAR ABS ALGORITHM [1]

Ladislav Lukšan [2] and Emilio Spedicato [3]

1 Introduction

Consider the system of nonlinear equations

$$f(x) = 0$$

where $f : R^n \to R^n$ is continuously differentiable mapping. Let $A(x)$ denote the Jacobian matrix of mapping f with

$$(A(x))_{ij} = \frac{\partial f_i(x)}{\partial x_j}, 1 \le i \le n, 1 \le j \le n.$$

In [2], a local convergence theorem was proved for the following class of scaled nonlinear ABS algorithms.

Algorithm 1

Assume that the initial estimate x_1 is selected; then the i-th cycle which defines x_{i+1} from the current estimate x_i is defined below

1. Let $y_1 = x_i$ and $H_1 = I$. Set $k = 1$.

2. Define vector $u_k \in R^n$ as a convex combination of vectors $y_1, \ldots, y_k$ (usually $u_k = y_k$ or $u_k = y_1$) and compute the Jacobian matrix $A_k = A(u_k)$.

3. Define the search vector p_k by

$$p_k = H_k^T z_k,$$

where $z_k \in R^n$ is arbitrary, save that $p_k^T A_k^T v_k \neq 0$ and $v_k \in R^n$ is arbitrary, save that $v_1, \ldots, v_k$ are linearly independent.

4. Define the stepsize α_k by

$$\alpha_k = v_k^T f(y_k) / p_k^T A_k^T v_k.$$

[1]Work partly supported by grant GA ČR 201/00/0080 and by MURST 1999 Programma di Cofinanziamento, Numerical Methods for Evolutionary Problems

[2]Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic (luksan@uivt.cas.cz)

[3]Department of Mathematics, University of Bergamo, Bergamo 24129, Italy (emilio@unibg.it)

5. Update the minor iterate y_k by

$$y_{k+1} = y_k - \alpha_k p_k.$$

If $k = n$, set $x_{i+1} = y_{n+1}$, the cycle is completed.

6. Update H_k by

$$H_{k+1} = H_k - \frac{H_k A_k^T v_k w_k^T H_k}{w_k^T H_k A_k^T v_k},$$

where $w_k \in R^n$ is arbitrary, save that $w_k^T H_k A_k^T v_k \neq 0$.

7. Increment the index k by one and go to 2.

In [1], several basic (unscaled) nonlinear ABS algorithms were tested. The tests have shown that these algorithms are very efficient when they converge. At the same time these algorithms frequently fail, since their global convergence is not guaranteed. Usually 40 (out of 64) problems with 100 variables were successfully solved by nonlinear ABS algorithms, while globally convergent methods were able to solve 60 problems (however they was usually slower). Therefore, globalization strategies for nonlinear ABS methods should be investigated.

In this contribution, we will modify one member of the scaled nonlinear ABS class to be globally convergent in the sense that for any $x_1 \in R^n$ the sequence x_i, $i \in N$, has the property that $f_i \to 0$.

Let $x_1 \in \mathcal{R}^n$, $\overline{F} \geq \|f(x_1)\|$ and $\Delta \geq 0$. Denote

$$\mathcal{L} = \{x \in \mathcal{R}^n : \|f(x)\| \leq \overline{F}\}$$

and

$$\mathcal{D}(\Delta) = \{x \in \mathcal{R}^n : \|x - y\| \leq \Delta \text{ for some } y \in \mathcal{L}\}$$

Throughout this contribution we will use the Euclidean vector norm and the spectral matrix norm respectively and suppose the following assumptions hold for an arbitrary $\Delta \geq 0$:

A1: The Jacobian matrix $A(x)$ is defined and bounded on $\mathcal{D}(\Delta)$, i.e.

$$\|A(x)\| \leq \overline{A}, \quad \forall x \in \mathcal{D}(\Delta)$$

A2: The Jacobian matrix $A(x)$ is Lipschitz continuous on $\mathcal{D}(\Delta)$, i.e.

$$\|A(y) - A(x)\| \leq \overline{L}\|y - x\| \quad \forall x, y \in \mathcal{D}(\Delta)$$

A3: The inverse Jacobian matrix $A^{-1}(x)$ is defined and bounded on $\mathcal{L}$, i.e.

$$\|A^{-1}(x)\| \leq 1/\underline{A}, \quad \forall x \in \mathcal{L}.$$

Furthermore, we will use the following merit function

$$F(x) = \frac{1}{2} f(x)^T f(x). \tag{1}$$

Note that A1 - A2 imply

$$\begin{aligned}
\|\nabla F(y) - \nabla F(x)\| &= \|A^T(y)f(y) - A^T(x)f(x)\| \\
&\leq \|A^T(y)(f(y) - f(x))\| + \|(A^T(y) - A^T(x))f(x)\| \\
&\leq \overline{A}\|f(y) - f(x)\| + \overline{L}\|y - x\|\|f(x)\| \\
&= \overline{A}\| \int_0^1 A(x + \tau(y - x))(y - x)d\tau\| + \overline{L}\|y - x\|\|f(x)\| \\
&\leq (\overline{A}^2 + \overline{LF})\|y - x\| \qquad (2)
\end{aligned}$$

whenever $x \in \mathcal{L}$ and $y \in \mathcal{D}(\Delta)$.

2 A globally convergent algorithm

We will investigate the following algorithm, which is in fact a modification of the implicit QR algorithm.

Algorithm 2

Assume that the initial estimate x_1 is selected; then the i-th cycle which defines x_{i+1} from the current estimate x_i is defined below

1. Let $y_1 = x_i$ and $H_1 = I$. Set $k = 1$.

2. Let $f_k = f(y_k)$. Compute the Jacobian matrix $A_k = A(y_k)$.

3. Define the search vector p_k by

$$p_k = H_k^T z_k,$$

   where $z_1 = A_1^T f_1$ and z_k is arbitrary for $k > 1$, save that $z_k^T H_k A_k^T f_k \neq 0$. If such a vector does not exist, set $x_{i+1} = y_k$, the cycle is completed (restart). Otherwise define the scaling vector v_k by

$$v_k = A_k p_k.$$

4. Determine the initial stepsize α_k by

$$\alpha_k = \mathrm{sign}\left(\frac{v_k^T f_k}{v_k^T v_k}\right) \min\left(|\frac{v_k^T f_k}{v_k^T v_k}|, \Delta\|p_k\|\right),$$

where $\Delta > 0$ (typically $\Delta = 10^6$), and find the smallest nonnegative integer l_k satisfying the following Armijo condition

$$F(y_k - \beta^{l_k}\alpha_k p_k) - F(y_k) \leq -\gamma\beta^{l_k}\alpha_k v_k^T f_k, \tag{3}$$

where $0 < \beta < 1$ (typically $\beta = 1/2$) and γ is a small positive number (typically $\gamma = 10^{-4}$).

5. Update the minor iterate y_k by

$$y_{k+1} = y_k - \beta^{l_k}\alpha_k p_k.$$

If $k = n$, set $x_{i+1} = y_{n+1}$, the cycle is completed.

6. Update H_k by

$$H_{k+1} = H_k - \frac{H_k A_k^T v_k z_k^T H_k}{v_k^T v_k}.$$

7. Increment the index k by one and go to 2.

Now we prove that Algorithm 2 is globally convergent in the sense that for any $x_1 \in R^n$ the sequence x_i, $i \in N$, has the property that $f_i \to 0$.

Lemma 1. *Algorithm 2 is well-defined, i.e. restart cannot occur for $k = 1$, the algorithm cannot break down because of the division by zero, and a nonpositive integer l_k satisfying the Armijo condition (3) always exists.*

Proof. Assumption A3 and $f_k \neq 0$ imply that $A_k^T f_k \neq 0$. Since $H_1 = I$, we obtain $z_1^T H_1 A_1^T f_1 = f_1^T A_1 A_1^T f_1 \neq 0$. If $z_k^T H_k A_k^T f_k \neq 0$, then $v_k \neq 0$, $v_k^T f_k \neq 0$ and α_k is well-defined. Denote $d_k = -\alpha_k p_k$. Since $v_k^T f_k \neq 0$, we obtain

$$d_k^T \nabla F(y_k) = -\alpha_k p_k^T A_k^T f_k = -\frac{(v_k^T f_k)^2}{v_k^T v_k} < 0$$

so that a nonpositive integer l_k satisfying the condition $F(y_k + \beta^l d_k) - F(y_k) \leq \gamma\beta^l d_k^T \nabla F(y_k)$ exists for arbitrary $0 < \beta < 1$ and $0 < \gamma < 1$ (see [3], Theorem 6.3.2). But the last condition is exactly (3). □

Lemma 2. *Let assumptions A1 - A3 be satisfied. Then Algorithm 2 generates a sequence x_i, $i \in N$, such that $\|f(x_{i+1})\| \leq c\|f(x_i)\|$ where $0 < c < 1$ does not depend on the cycle number.*

Proof. Since $\|f(y_{k+1})\| \leq \|f(y_k)\|$ holds by (3) in every minor iteration, it suffices to prove that $\|f(y_2)\| \leq c\|f(y_1)\|$, where $0 < c < 1$ does not depend on the cycle number. To simplify the notation, we omit index 1 and replace index 2 by +. Since $H = I$, we can write $p = A^T f = \nabla F(y)$, $v = AA^T f$ and

$$v^T f = \|A^T f\|^2 = \|p\|^2. \tag{4}$$

First we prove that $\beta^l \alpha \geq \rho$ holds where number $\rho > 0$ does not depend on the cycle number. Without any loss of generality, we will assume that $l > 0$ in (3). Using (3) and (4), we obtain

$$F(y - \beta^{l-1}\alpha p) - F(y) \geq -\gamma \beta^{l-1}\alpha \|p\|^2. \tag{5}$$

On the other hand, using the mean value theorem with $0 < \mu < 1$, we can write (see (2))

$$\begin{aligned} F(y - \beta^{l-1}\alpha p) - F(y) &= -\beta^{l-1}\alpha p^T \nabla F(y - \mu\beta^{l-1}\alpha p) \\ &\leq -\beta^{l-1}\alpha(p^T \nabla F(y) - \|p\| \|\nabla F(y - \mu\beta^{l-1}\alpha p) - \nabla F(y)\| \\ &\leq -\beta^{l-1}\alpha(\|p\|^2 - \|p\|^2 \beta^{l-1}\alpha(\overline{A}^2 + \overline{LF})). \end{aligned} \tag{6}$$

Coupling both these inequalities, we obtain

$$\gamma \geq 1 - \beta^{l-1}\alpha(\overline{A}^2 + \overline{LF}),$$

which gives

$$\beta^l \alpha \geq \beta(1-\gamma)\frac{1}{(\overline{A}^2 + \overline{LF})} \triangleq \rho.$$

Using this result together with (3), (4) and A3, we can deduce that

$$\begin{aligned} \|f\|(\|f_+\| - \|f\|) &\leq \frac{1}{2}(\|f_+\| + \|f\|)(\|f_+\| - \|f\|) = F_+ - F \\ &\leq -\gamma\rho\|A^T f\|^2 \leq -\gamma\rho\underline{A}^2\|f\|^2. \end{aligned}$$

Thus

$$\|f_+\| \leq (1 - \gamma\rho\underline{A}^2)\|f\| = \left(1 - \beta\gamma(1-\gamma)\frac{\underline{A}^2}{(\overline{A}^2 + \overline{LF})}\right)\|f\| \triangleq c\|f\|.$$

□

Theorem 1. *Let assumptions A1 - A3 be satisfied. Then Algorithm 2 generates a sequence x_i, $i \in N$, such that $f(x_i) \to 0$.*

Proof. Using Lemma 2, we obtain

$$\sum_{i=1}^{\infty} \|f_i\| = \sum_{i=1}^{\infty} c^{i-1}\|f_1\| = \frac{1}{1-c}\|f_1\| < \infty,$$

which implies $f_i \to 0$. □

Algorithm 2 is highly impractical, since the Jacobian matrix must be evaluated in every minor iteration. This is a consequence of scaling. We need this scaling for obtaining global convergence. On the other hand we can replace the computation of the Jacobian matrix by a suitable restart procedure (see [4]). In this case, the correct Jacobian matrix is not required in every minor iteration. But if the Armijo condition (3) is not satisfied for $l_k \leq 3$, say, then the correct Jacobian matrix should be recomputed.

In [2], the local and superlinear convergence of Algorithm 1 is proved. It is shown that the stepsize α_k is acceptable in the neighbourhood of the solution. This implies that also Algorithm 2 is superlinearly convergent.

Notice that for $k > 1$, the Armijo condition (3) can be replaced by a simpler condition $F(y_k - \beta^{l_k}\alpha_k p_k) \leq F(y_k)$.

3 Computational experiments

In this section, we present computational experiments with two versions of Algorithm 2: `nl_iqr1` (Algorithm 2 with the computation of Jacobian matrices in all minor iterations), `nl_iqr2` (Algorithm 2 with the computation of Jacobian matrices only in major iterations). Algorithm 2 is also compared with three basic ABS algorithms proposed in [1]: `nl_huanh1` (Huang algorithm in the basic form), `nl_huanh2` (Huang algorithm in the product form), `nl_ilu` (implicit LU algorithm). Tests consist in seeking solutions to 64 nonlinear systems with 20 equations and unknowns. Computations were performed using a PC-Pentium 3 (800 MHz) computer in double precision arithmetic. The results are arranged in the following table, where NI is the number of iterations, NF is the number of function evaluations, NJ is the number of Jacobian evaluations and NS is the number of succesfully solved problems.

Algorithm	NI	NF	NJ	NS
`nl_huang1`	954	6567	954	45
`nl_huang2`	1177	8882	1177	44
`nl_ilu`	920	6421	920	41
`nl_iqr1`	1422	32358	27089	63
`nl_iqr2`	935	54321	935	60

Table 1 – results for 20 variables

The results contained in Table 1 show that the globally convergent Algorithm 2 is very robust compared with basic nonlinear ABS algorithms (its version `nl_iqr1` solves 63 problems successfully, it only fails for one problem from our collection). On the other hand, Algorithm 2 is slower than the basic nonlinear ABS algorithms, since the vector of nonlinear functions (and for `nl_iqr1` also the Jacobian matrix) has to be evaluated in every minor iteration.

References

[1] Bodon, E., Del Popolo A., Lukšan L., Spedicato, E., *Numerical performance of ABS codes for systems of nonlinear equations,* Report QDMSIA 1/01, University of Bergamo, 2001.

[2] Deng, N., Spedicato, E., Zhu, M., *A local convergence theorem for nonlinear ABS algorithms,* Comp. Appl. Math., Vol. 13. No. 1, pp. 49-59, 1994.

[3] Dennis, J.E., and Schnabel, R.B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[4] Lukšan L., Vlček J.: Computational experience with globally convergent descent methods for large sparse systems of nonlinear equations. Optimization Methods and Software, Vol.8, 1998, pp.201-223.