Numer. Math. 69: 1–15 (1994)

Numerische
Mathematik
© Springer-Verlag 1994

On nonlinear generalized conjugate gradient methods

O. Axelsson[1], **A.T. Chronopoulos**[2,⋆]

[1] Faculty of Mathematics and Informatics, University of Nijimegen, Nijimegen, The Netherlands
[2] Department of Computer Science, University of Minnesota, Minneapolis, USA

Received April 29, 1992 / Revised version received November 18, 1993

Summary. The Generalized Conjugate Gradient method (see [1]) is an iterative method for nonsymmetric linear systems. We obtain generalizations of this method for nonlinear systems with nonsymmetric Jacobians. We prove global convergence results.

Mathematics Subject Classification (1991): 65H10; 65F10

1. Introduction

Nonlinear systems of equations often arise when solving initial or boundary value problems in ordinary or partial differential equations, see [14] and [15], for instance. We consider the nonlinear system of equations

$$F(\xi) = 0$$

where $F(\xi)$ is a nonlinear operator from a real Euclidean space of dimension n or Hilbert space into itself. The Euclidean norm and corresponding inner product will be denoted by $\|\cdot\|_1$ and $(\cdot,\cdot)_1$ respectively. A general different inner product with a weight function and the corresponding norm will be denoted by $(\cdot,\cdot)_0$ and $\|\cdot\|$ respectively. In the first part of this article (Sects. 2 and 3) we assume that the Jacobian of $F(\xi)$ has symmetric parts uniformly positive definite. In the final part (Sect. 4) a method is presented where this assumption is not required.

The Newton method coupled with direct linear system solvers is an efficient way to solve such nonlinear systems when the dimension of the Jacobian is small. When the Jacobian is large and sparse some kind of iterative method may be used. This can be a nonlinear iteration (for example functional iteration for contractive operators), or an inexact Newton method. In an inexact Newton the solution of the resulting linear systems is approximated by a linear iterative method. The following are typical steps in an inexact Newton method for solving this nonlinear system.

⋆ This work was partially supported by travel grants from the Office of International Education and from the Army High Performance Computing Research Center at the University of Minnesota

```
Algorithm 1.1 Inexact Newton
Choose $\xi^0$
For $k = 0, 1, \ldots$ until convergence do
    1. Solve iteratively: $F'(\xi^k)\Delta_k = -F(\xi^k)$
    2. $\xi^{k+1} = \xi^k + \Delta_k$
EndFor
```

For a given tolerance $\epsilon > 0$, convergence can be decided for example if $\|F(\xi^k)\|_1 < \epsilon$.

If the linear iterative method is a Krylov subspace method then the Jacobian is only required for performing Jacobian times vector operations. Efficient methods to compute directly sparse Jacobians have been proposed [13]. Alternatively, (given a small scalar ϵ) the Jacobian times vector operation can be approximated using the following divided difference

$$F'(\xi^0)v \approx \frac{F(\xi^0 + \epsilon v) - F(\xi^0)}{\epsilon}.$$

A very important question is how to terminate the inner and outer iterations in an inexact Newton algorithm and retain a satisfactory convergence rate. Axelsson in [3] and Dembo et al. in [9] study the convergence rates of the inexact Newton method in relation to the accuracy to which the linear systems are solved (see also [10]). This is expressed in terms of the ratio of the residuals of the inner and outer iterations:

$$\frac{\|F(\xi^k) + F'(\xi^k)\Delta_k\|_1}{\|F(\xi^k)\|_1} \quad .$$

An alternative to inexact Newton approach is to derive nonlinear iterative methods which coincide with known iterative methods for linear systems. Some nonlinear iterative methods have been derived, studied and used in various applications for steepest descent methods, SOR type and conjugate gradient type methods for nonsymmetric Jacobians (see [4], [5], [6], [7], [8], [12], [14], [15], [16], [17]).

The Generalized Conjugate Gradients (GCG) (see [1]) is an iterative method applicable to nonsymmetric linear systems. The main goal of this article is to derive nonlinear versions of GCG and to establish global convergence results. In Sect. 2, we derive the Nonlinear GCG method (NGCG) and we prove that under global conditions it converges. In Sect. 3, we discuss convergence results for NGCG with nonlinear preconditioning. In Sect. 4, we discuss a combined Approximate Newton and NGCG method and show its global convergence.

2. Nonlinear GCG

In this section, our goal is to develop an iterative method for nonlinear systems of equations with nonsymmetric Jacobians. This method will be a nonlinear extension of GCG (see [1]) in the sense that it will be identical to GCG for linear systems of equations. Global convergence will be shown for Jacobians whose symmetric parts are uniformly positive definite.

Throughout Sects. 2 and 3, the following standard assumptions will be made on the vector function $F(\xi)$. $F(\xi)$ is a nonlinear function from a ball in $\mathbb{R}^n$ around the solution, containing the approximate solutions, into $\mathbb{R}^n$. Without loss of generality we will state and prove our results assuming that $F(\xi)$ is defined on the whole space $\mathbb{R}^n$. We also assume that the Jacobian $F'(\xi)$ and the Hessian $F''(\xi)$ exist and that there exist positive constants δ_1, δ_2, δ_3 such that $\delta_1 \le \delta_2$ and for all vectors v in $\mathbb{R}^n$:

$$\begin{aligned} \delta_1 \|v\|_1^2 &\le (F'(\xi)v, v)_1 \qquad (a),\\ \|F'(\xi)v\|_1 &\le \delta_2 \|v\|_1 \qquad (b),\\ \|F''(\xi)\|_1 &\le \delta_3 \qquad (c). \end{aligned} \tag{1}$$

Assumption 1(a) states that the symmetric parts of the Jacobians are uniformly positive definite and it implies that the Jacobians are nonsingular. The following lemma follows from assumption 1(a).

Consider the mapping $F: \ \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$, where Ω is convex and assume that F is differentiable on Ω. The following equivalence relations hold:

Lemma 2.1. *Let δ_1 be a nonnegative constant. Then the following properties are equivalent:*
(a) $(F'(\xi)v, v)_1 \ge \delta_1 \|v\|_1^2$ for all $\xi \in \Omega$, and all $v \in \mathbb{R}^n$
(b) $(F(x) - F(y),\ x - y)_1 \ge \delta_1 (x - y,\ x - y)_1$ for all $x, y \in \Omega$
(c) For any two solutions of the ordinary differential equation $x'(t) = -F(x(t))$, $t > 0$ it holds that

$$\|x(t) - y(t)\| \le \mathrm{e}^{-\delta_1 t} \|x(0) - y(0)\|. \tag{2}$$

Proof. The relation

$$(F(x) - F(y),\ x - y)_1 = \left(\int_0^1 F'(\xi(t))(x-y)dt,\ x - y \right)_1 ,$$

where $\xi(t) = y + t(x - y)$, holds. Hence, given (a), the mean value theorem for a scalar function shows that (b) holds (for some ξ on the line segment from x to y). Further,

$$\begin{aligned} \frac{d(\|x(t) - y(t)\|^2)}{dt} &= 2(x' - y',\ x - y)\\ &= -2(F(x(t)) - F(y(t)),\ x(t) - y(t))\\ &\le -2\delta_1 \|x(t) - y(t)\|^2 , \end{aligned}$$

which by integration, implies (c). Conversely, (2) implies

$$\frac{d(\|x(t) - y(t)\|^2)}{dt} \le -2\delta_1 \mathrm{e}^{-2\delta_1 t} \|x(0) - y(0)\|^2$$

or

$$-2(F(x(t)) - F(y(t)),\ x(t) - y(t)) \le -2\delta_1 \mathrm{e}^{-2\delta_1 t} \|x(0) - y(0)\|^2$$

that is

$$(F(x(t)) - F(y(t)),\ x(t) - y(t)) \ge \delta_1 \mathrm{e}^{-2\delta_1 t} \|x(0) - y(0)\|^2 .$$

Letting $t \longrightarrow 0$ yields (b). Finally, letting $x = \xi$, $x - y = tv$ in part (b) and letting $t \longrightarrow 0+$ we obtain (a). □

Remark. Any function F satisfying part (b) with $\delta_1 > 0$ is said to be strongly monotone in Ω. Using well known techniques it can be seen that Lemma 2.1 implies the existence of a unique solution of $F(\xi) = 0$ if F is strongly monotone.

Notation . We denote the Jacobian $F'(\xi^k)$ by F_k'. We denote by $(\cdot,\cdot)_0$ the inner product with respect to the weight matrix $(F_k')^{\mathrm{T}} F_k'$.

We next derive a nonlinear extension of the GCG method (see [1]). At first we describe the method and then we outline the algorithm.

Method Description. For $\mu = 0$ or 1 the inner product $(.,.)_\mu$ and its corresponding norm will be used in the method. At each iteration k, a set of vectors and scalars are computed. Also, some index parameters must be selected in advance.

(i) Index parameters. Given t and s fixed positive integers we select $s_k = \min(k, s)$ and $t_k = \min(k, t)$.
Then at each iteration (k) a (search) direction vector is computed by orthogonalizing (with respect to $(.,.)_\mu$) the (nonlinear) residual vector against s_k preceding direction vectors. The solution is approximated by solving a **nonlinear least squares** (n.l.s.) problem which minimizes the norm of the nonlinear residual over an affine subspace based at ξ^{k-1} and the search directions $\{d^{k-j}\}$, for $j = 1, \ldots, t_k$.

(ii) Vectors. Solution updates $\{\xi^k\}$, residuals $\{p^k\}$, and search directions $\{d^k\}$.

(iii) Scalars. The steplengths (used in updating the solution) $\{\alpha_{(k-j)}^{(k-1)}\}$, for $j = 1, \ldots, t_k$ and the Gram-Schmidt orthogonalization parameters (used in updating the direction vectors) $\{\beta_{k-j}^{(k-1)}\}$, for $j = 1, \ldots, s_k$. We use $\alpha_{(k-j)}^{(k-1)}$ to denote both the independent variables and the solution of the (n.l.s.) problem.

We next outline the nonlinear GCG algorithm (NGCG).

```
**Algorithm 2.1** NGCG
Initial approximation $\xi^0$
$d^0 = -p^0 = -F(\xi^0)$
**For** $k = 1$ **until** Convergence **do**
    1. $\xi^k = \xi^{k-1} + \sum_{j=1}^{t_k} \alpha_{k-j}^{(k-1)} d^{k-j}$
       where $\{\alpha_{k-j}^{(k-1)}\}$ solve the (n.l.s.) problem
       min $\|F(\ \xi^{k-1} + \sum_{j=1}^{t_k} \alpha_{k-j}^{(k-1)} d^{k-j}\ )\|_1^2$

    2. $p^k = F(\xi^k)$
    3. $\beta_{k-j}^{(k-1)} = (p^k, d^{k-j})_\mu / \|d^{k-j}\|_\mu^2,\ j = 1, \ldots, s_k$
    4. $d^k = -p^k + \sum_{j=1}^{s_k} \beta_{k-j}^{(k-1)} d^{k-j}$
**EndFor**
```

If $k \leq s$, then Algorithm 2.1 generates search direction which are fully orthogonal. In practice truncated versions are considered. The index s is chosen under computer storage constraints and Gramm-Schmidt stability considerations. We usually choose $t = s+1$. Thus, after the iteration $k = s$: $s_k = s_{k-1}$ and $t_k = t_{k-1}$. In this case, at each subsequent iteration one new search direction is added to the set of direction vectors and the oldest direction is deleted from the set.

Remark. 1. The direction vectors are orthogonal by the definition of $\beta^{(k)}$ in Algorithm 2.1 :

$$(d^k, d^{k-j})_\mu = 0, \quad j = 1, \ldots, s_k \ . \tag{3}$$

2. The steplengths $\left\{\alpha_{k-j}^{(k-1)}\right\}$ solve a nonlinear least squares problem (n.l.s.) in Algorithm 2.1 (1.) . This is equivalent to the orthogonality relations:

$$(p^k, F'_k d^{k-j})_1 = 0, j = 1, \ldots, t_k \ , \tag{4}$$

as it can be seen by partial differentiation of the error functional to be minimized.

Notation. At iteration k in Algorithm 2.1, we use the notation $\overline{\alpha} = \left[\alpha_{k-1}^{(k-1)}, \ldots, \alpha_{k-t_k}^{(k-1)}\right]^{\mathrm{T}}$ and $D = \left[d^{k-1}, \ldots, d^{k-t_k}\right]$ to denote the steplengths and direction vectors (respectively) used in updating the solution.

Initially (at iteration $k = 1$) the dimensions of $\overline{\alpha}$ and D are 1 and $N \times 1$ and at subsequent steps they increase to t_k and $N \times t_k$ respectively. Now, 1. of Algorithm 2.1 is expressed concisely as $\xi^k = \xi^{k-1} + D\overline{\alpha}$.

Remark. It will next be shown that the n.l.s. problems (of dimension t_k) in the NGCG algorithm has a nontrivial solution. This means that there is progress towards the solution at each iteration (i.e. $\|F(\xi^{k+1})\|_1 < \|F(\xi^k)\|_1$). The solution $\overline{\alpha}$ can be obtained by a Newton method (see [10]). In such a Newton method the Jacobian $F'(\xi + D\overline{\alpha}) * D$ must be computed. In solving the n.l.s. problems evaluation of the exact Jacobian can be avoided by using inexact line search methods (see [10]).

Lemma 2.2. *In Algorithm 2.1, assume that iteration $k-1$ $(1 \leq k)$ is well-defined. If $p^{k-1} \not\equiv 0$, then the matrix D has full rank.*

Proof. Assume that $\{d^{k-1}, \ldots, d^{k-t_k}\}$ are linearly dependent. This implies that d^{k-1} is a linear combination of $\{d^{k-2}, \ldots, d^{k-t_k}\}$. Then using NGCG (4.), we conclude that p^{k-1} can be expressed as a linear combination of $\{d^{k-2}, \ldots, d^{k-t_k}\}$. Using equality (4) we conclude that $(p^{k-1}, F_{k-1}d^{k-j})_1 = 0$ for $j = 2, \ldots, t_{k-1}$. Since (by definition) $k - t_k = k - t_{k-1}$, this yields $(p^{k-1}, F_{k-1}p^{k-1})_1 = 0$ which contradicts (1)(a) unless $p^{k-1} = 0$. □

We next prove that Algorithm 2.1 is feasible.

Lemma 2.3. *In Algorithm 2.1, assume that $s_k \leq t_k$, for all k. Given any initial vector ξ^0 in $\mathbb{R}^n$ all iterations are well-defined and $\|p^k\|_1 < \|p^{k-1}\|_1$.*

Proof. At each iteration k, we must prove that (i) the direction vectors are well-defined and $d^k \not\equiv 0$, if $p^k \not\equiv 0$ and (ii) the (n.l.s.) problem has a solution and $\|p^k\|_1 < \|p^{k-1}\|_1$. We use induction on the iteration index k.

Case k=1. (i) is obvious. We must prove (ii). For simplicity of notation we use $f_{k-1}(\overline{\alpha})$ to denote $(1/2)\|F(\xi^{k-1}+D\overline{\alpha})\|_1^2$. We must prove that there exists a nontrivial minimum of $f_{k-1}(\overline{\alpha})$. The gradient $(\nabla f_{k-1}(\overline{\alpha}))$ of $f_{k-1}(\overline{\alpha})$ equals:

$$\begin{gathered}\big[\big(\, F(\xi^{k-1} + D\overline{\alpha}), F'(\xi^{k-1} + D\overline{\alpha})d^{k-1}\ \big)_1, \ldots, \\ \big(\, F(\xi^{k-1} + D\overline{\alpha}), F'(\xi^{k-1} + D\overline{\alpha})d^{k-t_k}\ \big)_1\big]^{\mathrm{T}} \ .\end{gathered}$$

Inserting $\overline{\alpha} = 0$ we obtain

$$\begin{aligned}\nabla f_{k-1}(0) &= \left[(\, p^{k-1}, F'_{k-1}d^{k-1}\ )_1,\ \ldots, (\, p^{k-1}, F'_{k-1}d^{k-t_k}\ )_1\right]^{\mathrm{T}} \\ &= \left[-(\, p^{k-1}, F'_{k-1}p^{k-1}\ )_1, 0, \ldots, 0\right]^{\mathrm{T}} \ .\end{aligned}$$

For $k = 1$ the gradient $\nabla f_{k-1}(0)$ consists only of the first entry, which is negative because of assumption (1)(a). This implies that there exists $\overline{\alpha} \neq 0$ such that $f_{k-1}(\overline{\alpha}) < f_{k-1}(0)$.

We must now prove that the function $f_{k-1}(.)$ assumes a minimum. We insert $\xi = \xi^{k-1}$ and $y = \xi^{k-1} + D\overline{\alpha}$ in inequality (2). Since (by Lemma 2.2) D has full rank, $D\overline{\alpha} \neq 0$ for $\overline{\alpha} \neq 0$. This implies that $\|y\|_1 \to \infty$ as $\|\overline{\alpha}\|_1 \to \infty$. Hence, Lemma 2.1 implies that $F(y)$ grows unbounded as $\|\overline{\alpha}\|_1 \to \infty$. This proves that the n.l.s. problem of NGCG (1.) has a solution. We also obtain that $\|p^k\|_1 < 2f_{k-1}(0) \equiv \|p^{k-1}\|_1$.

Case m < k (induction hypothesis). We assume that (i) and (ii) hold true for iterations $m = 2, \ldots, k-1$.

Case k. We firstly prove (i). We also assume that $p^m \neq 0$, $m = 0, \ldots, k-1$, otherwise the algorithm would have terminated. The induction hypothesis states that the direction vectors are well-defined and $d^m \neq 0$, $m = 0, \ldots, k-1$. The search direction d^k is well-defined, if $\beta_{k-j}^{(k-1)}$, $j = 1, \ldots, s_k$ are well-defined. This is true because (by the induction hypothesis) $\|d^{k-j}\|_\mu \neq 0$, for $j = 1, \ldots, s_k$. If $d^k = 0$ then p^k depends linearly on d^{k-j}, $j = 1, \ldots, s_k$. Also, p^k is orthogonal to d^{k-j}, $k = 1, \ldots, t_k$ (see (4)). Since $s_k \leq t_k$ we conclude that $p^k = 0$.

To prove (ii) we note that the induction hypothesis and equality (4) imply that the gradient $\nabla f_{k-1}(0)$ consists of zeros except of a negative value in the first entry. The rest of the proof of (ii) is identical to case $k = 1$. □

We next define a single nonlinear steepest descent step after each iteration of NGCG has been completed. For $k = 1, 2, \ldots$, let ξ^{k-1} and p^{k-1} for $k \geq 1$ be generated in NGCG. A single nonlinear steepest descent step is defined as follows:

$$\hat{\xi}^k = \xi^{k-1} - \hat{\alpha}_{k-1}^{(k-1)} p^{k-1} \, , \tag{5}$$

where $\hat{\alpha}_{k-1}^{(k-1)}$ is the least positive real number that solves the nonlinear minimization problem

$$\min_{\alpha \in \mathbb{R}} \ \|F(\xi^{k-1} - \alpha p^{k-1})\|_1^2.$$

We next give an error bound on the decrease in the nonlinear residual that this single nonlinear steepest descent step would bring about.

Lemma 2.4. *Let our standard assumptions hold for the nonlinear function $F(\xi)$. The single steepest descent step (5) is well-defined and the following inequality holds:*

$$\|F(\, \xi^{k-1} - \hat{\alpha}_{k-1}^{(k-1)} p^{k-1} \,)\|_1 \leq c\|p^{k-1}\|_1 \, , \tag{6}$$

where $\hat{\alpha}_{k-1}^{(k-1)}$ is the steplength in (5) and $0 < c < 1$.

Proof. The proof (of (ii)) in Lemma 2.3 can be used to prove that the nonlinear steepest descent is well-defined.

For simplicity of notation let $g_{k-1}(\alpha)$ denote the scalar function $(1/2)\|F(\xi^{k-1} - \alpha p^{k-1})\|_1^2$. Then $g_{k-1}(\alpha)$ assumes a minimum at $\hat{\alpha}_{k-1}^{(k-1)}$. The first and second derivatives of $g_{k-1}(\alpha)$ are:

$$g'_{k-1}(\alpha) = (F(\xi^{k-1} - \alpha p^{k-1}), F'(\xi^{k-1} - \alpha p^{k-1})p^{k-1})_1 \quad ,$$
$$g''_{k-1}(\alpha) = ((F''(\xi^{k-1} - \alpha p^{k-1})p^{k-1}, p^{k-1} \,)_1, \; F(\xi^{k-1} - \alpha p^{k-1}))_1 + \|F'(\xi^{k-1} - \alpha p^{k-1})p^{k-1}\|_1^2.$$

Using assumption (1) we obtain the following upper bound on $|g''(\alpha)|$ for $0 < \alpha < \hat{\alpha}_{k-1}^{(k-1)}$.

$|g''_{k-1}(\alpha)| \leq (\delta_2^2 + \delta_3 \|p^{k-1}\|_1)\|p^{k-1}\|_1^2.$

A lower bound on $\hat{\alpha}_{k-1}^{(k-1)}$ can be obtained as follows:

Taylor's expansion $g'_{k-1}\left(\hat{\alpha}_{k-1}^{(k-1)}\right)$ around 0 gives:

$$g'_{k-1}\left(\hat{\alpha}_{k-1}^{(k-1)}\right) = 0 = g'_{k-1}(0) + \alpha_{k-1}^{(k-1)} g''(\overline{\alpha}),$$

where $\overline{\alpha} = t_{k-1}\hat{\alpha}_{k-1}^{(k-1)}$ for some t_{k-1} in (0, 1).

Using the upper bound on $g''(\overline{\alpha})$ derived above and assumptions (1) we can obtain now the following bound

$$\frac{\delta_1}{\left(\delta_2^2 + \delta_3\|p^{k-1}\|_1\right)} \leq \frac{(\ p^{k-1}, F'_{k-1}p^{k-1}\ )_1}{\|p^{k-1}\|_1^2\left(\delta_2^2 + \delta_3\|p^{k-1}\|_1\right)} \leq \hat{\alpha}_{k-1}^{(k-1)}.$$

We next obtain inequality (6). Using Taylor's expansion we get:

$$g_{k-1}(\alpha) \leq (1/2)\|p^{k-1}\|_1^2 - \alpha(p^{k-1}, F'_{k-1}p^{k-1})_1 + (\alpha^2/2)g''_{k-1}(\overline{\alpha}) \ \ ,$$

where $\overline{\alpha} = t\alpha$ for some t in (0, 1).

Now using the lower bound on α, the upper bound on $g''_{k-1}(\alpha)$ and assumption (1) we prove inequality (6):

$$\begin{aligned} g_{k-1}(\alpha_{k-1}^{(k-1)}) &\leq \frac{\|p^{k-1}\|_1^2}{2} - \frac{\delta_1^2\|p^{k-1}\|_1^2}{\left(\delta_2^2 + \delta_3\|p^{k-1}\|_1\right)} + \frac{\delta_1^2}{2\left(\delta_2^2 + \delta_3\|p^{k-1}\|_1\right)^2} g''_{k-1}(\overline{\alpha}) \\ &\leq \frac{1}{2}\left[1 - \frac{\delta_1^2}{\left(\delta_2^2 + \delta_3\|p^{k-1}\|_1\right)}\right]\|p^{k-1}\|_1^2\ . \ \square \end{aligned}$$

The following theorem shows the global convergence of NGCG.

Theorem 2.1. *Let ξ^0 be an initial solution. Under our standard assumptions, NGCG generates a sequence ξ^k, which converges to the unique solution ξ^* of the nonlinear operator equation $F(\xi) = 0$, and $\|\xi^k - \xi^*\|_1 \leq (1/\delta_1)\|p^k\|_1$.*

Proof. By Lemma 2.3 the iterations in Algorithm 2.1 are well-defined. At each iteration k, we use Lemma 2.4 to obtain a residual norm bound by using a single step of the nonlinear steepest descent iteration (5). This implies that in NGCG:

$$\|p^k\|_1 \leq c\|p^{k-1}\|_1, \ \ 0 < c < 1.$$

We conclude that $p^k \to 0$ as $k \to \infty$. This combined with Lemma 2.1 proves that $\{\xi^k\}$ is a Cauchy sequence and thus it converges. The proof of the uniqueness and the error bound follow from Lemma 2.1 by inserting $\xi = \xi^k$ and $y = \xi^*$ respectively. $\square$

3. Preconditioning

There exists a convergence theory for variable-step preconditioning applied to GCG (see [2]). Here we consider variable step preconditioning applied to NGCG. The nonlinear operator equation $F(\xi) = 0$ can be preprocessed using a linear or nonlinear operator.

Let F_r and F_l denote the right and left preconditioners respectively. We assume that F_r^{-1} and F_l^{-1} exist. The preconditioned nonlinear operator equation can take one of the following three forms:

(1) right preconditioning:

$$G(y) = F(F_r(y)) = 0, y = F_r^{-1}(\xi) \ . \tag{7}$$

(2) left preconditioning:

$$G(\xi) = F_l(F(\xi)) = 0 \ . \tag{8}$$

(3) split preconditioning:

$$G(y) = F_l(F(F_r(y))) = 0, y = F_r^{-1}(\xi) \ . \tag{9}$$

In the case of right and left preconditioning F_r and F_l must locally approximate the inverse of F. In the case of split preconditioning the operator $F_l F F_r$ must be close to the identity operator. In all cases the preconditioned operator equation is “easier” to solve if the preconditioned operator G has a Jacobian with “better” spectral properties than the Jacobian of F.

Notation. We denote the Jacobian $G'(\xi^k)$ by G'_k. We denote by $(\cdot,\cdot)_0$ the inner product with respect to the weight matrix $(G'_k)^{\mathrm{T}} G'_k$.

We next formulate the preconditioned version of Algorithm 2.1 using right preconditioning. The algorithms for the other types of preconditionings are similar.

Algorithm 3.1 (Right preconditioned NGCG)

```
Initial approximation $\xi^0$
$y^0 = F_r^{-1}(\xi^0)$
$d^0 = -p^0 = -G(y^0)$
For $k = 1$ until Convergence do
    1. $y^k = y^{k-1} + \sum_{j=1}^{t_k} \alpha_{k-j}^{(k-1)} d^{k-j}$
       where $\{\alpha_{k-j}^{(k-1)}\}$ solve the (n.l.s.) problem
       min $\|G(\ y^{k-1} + \sum_{j=1}^{t_k} \alpha_{k-j}^{(k-1)} d^{k-j}\ )\|_1^2$
    2. $p^k = G(y^k)$
    3. $\beta_{k-j}^{(k-1)} = (p^k, d^{k-j})_\mu / \|d^{k-j}\|_\mu^2, j = 1, \ldots, s_k$ and $\mu = 0$ or $1$ .
    4. $d^k = -p^k + \sum_{j=1}^{s_k} \beta_{k-j}^{(k-1)} d^{k-j}$
EndFor
$\xi^k = F_r(y^k)$
```

The following remarks can be made on Algorithm 3.1:

(i) $p^k = G(y^k) \equiv F(\xi^k)$

(ii) The Jacobian of the operator $G(y)$ equals $F'(\xi)F_r'(y)$. Thus, $G_k' = G'(y^k) = F'(\xi^k)F_r'(y^k)$. At each iteration, only multiplication by G_k' is required. This may be approximated by the directional derivative approximation:

$$G_k' v \approx \frac{G(y^k + \epsilon v) - G(y^k)}{\epsilon} \quad .$$

(iii) ξ^k is only computed after the termination of the algorithm. Intermediate computation requires the additional work $\xi^k = F_r(y^k)$ per iteration.

We next state the global convergence result for the right preconditioned NGCG method as a corollary of Theorem 2.1. A similar result can be obtained for the other two types of preconditioning.

Corollary 3.1. *Let the standard assumptions (in Sect. 1) also hold for the vector function $G(y)$ for all $y \in \mathbb{R}^n$. Moreover, we assume that $F_r'^{-1}$ exists and there is a positive constant δ_4 so that*

(10) $$\delta_4 \|v\|_1^2 \leq (F_r'^{-1}(y)v, v)_1 \ ,$$

for all vectors v in $\mathbb{R}^n$. Then the right preconditioned NGCG method generates a well-defined iteration for any initial approximation ξ^0. The sequence $\xi^k = F_r(y^k)$ converges to a unique solution ξ^* of the nonlinear operator equation $F(\xi) = 0$ and

(11) $$\|\xi^k - \xi^*\|_1 \leq (1/\delta_1\delta_4)\|F(\xi^k)\|_1 \quad .$$

Proof. Theorem 2.1 applied to the nonlinear system $G(y) = 0$ implies that y^k converges to a unique solution y^* of $G(y) = 0$. Since $\xi = F_r(y)$ and F_r' exist ξ^k converges to ξ^* and $F(\xi^*) = 0$. The error inequality is obtained from (10). Firstly, we use the mean value theorem (similarly to proving (2)) to prove the following inequality:

$$\delta_4 \|\xi^k - \xi^*\|_1 \leq \|F_r^{-1}(\xi^k) - F_r^{-1}(\xi^*)\|_1 = \|y^k - y^*\|_1 \ .$$

This inequality combined with $\|y^k - y^*\|_1 \leq (1/\delta_1)\|F(y^k)\|_1$ proves inequality (11). □

We next show that a combined Newton NGCG method converges for any initial approximation.

4. The approximate Newton direction NGCG method (NNGCG)

Let $\{\rho_k\}, 0 \leq \rho_k < 1$ be a nonincreasing sequence, and consider the iteration method: given ξ^0, for k = 0, 1, . . . compute p^{k+1} such that

(12) $$\|F(\xi^k) + F'(\xi^k)p^{k+1}\| \leq \rho_k \|F(\xi^k)\|_1.$$

In the standard Newton method we have $\rho_k = 0$. Here we only approximately solve the linear equation. Let

(13) $$d^{k+1} = -p^{k+1} + \sum_{j=0}^{s_k} \beta_{k+1-j}^{(k)} d^{k-j}, 0 \leq s_k \leq k,$$

where $\beta_j^{(k)}$ are computed to make

$$(d^{k+1}, d^{k-j})_1 = 0, 0 \leq j \leq s_k. \tag{14}$$

(12) can be viewed as a nonlinear preconditioning step. Let the integers r_k satisfy $0 \leq r_{k-1} \leq r_k \leq r_{k-1} + 1$ and let

$$\xi^{k+1} = \arg\min\{\|F(\xi)\|; \xi = \xi^k + \sum_{j=0}^{r_k} \alpha_{k+1-j} d^{k+1-j}\}. \tag{15}$$

Repeat until convergence.

Here $s_k + 2$ is the dimension of the set of search direction vectors at stage k and the functional $\|F(\xi)\|$ is minimized on a subspace spanned by $r_k + 1$ vectors. The norm is defined by the inner product $(\cdot, \cdot)_0$.

We let $s_k \leq r_k - 1$. (Hence, in the simplest case where $r_k = 0$, we let $d^{k+1} = -p^{k+1}$.) To compute p^{k+1} from (12) one can use preconditioned iteration methods, for instance. Recently it has been shown that efficient such methods exist even when $F'(\xi^k)$ is indefinite (see [2]). We shall prove convergence of the above algorithm for any initial approximation under the following assumptions:

(i.a) $F'(\xi^k)$ is nonsingular and there exists a constant β such that

$$\beta = \sup_k \|F'(\xi^k)^{-1}\|. \tag{16}$$

or

(i.b) $F'(\xi^k)$ is nonsingular but $F'(\xi)$ may be singular at the limit point where $F(\xi) = 0$. In this case we assume that

$$\delta = \sup_k \|F'(\xi^k)^{-1} F(\xi^k)\| \tag{17}$$

exists.

(ii) $F'(\cdot)$ is Hölder continuous, i.e. there exists a $\gamma, 0 < \gamma \leq 1$, such that

$$K_\gamma = (1 + \gamma) \sup_{k, \xi \neq \xi^k} \frac{\| \int_0^1 \left[F'\left(\xi^k + t(\xi - \xi^k)\right) - F'(\xi^k)\right] F'(\xi^k)^{-1} dt\|}{\|\xi - \xi^k\|^\gamma}, \tag{18}$$

where $K_\gamma < \infty$.

Note that this is a relative Hölder bound, since the difference in the bracket is multiplied with $F'(\xi^k)^{-1}$. This indicates that the constant is frequently not large. Furthermore, in practice, it suffices to take the supremum in a ball about the points ξ^k.

We shall analyze the convergence of the NNGCG algorithm first for the case (i.a), (ii). The analyses for the case (i.b), (ii) will be similar.

Note first that for any $\tau_k, 0 < \tau_k \leq 1$,

$$\min_\xi \left\{ \|F(\xi)\|, \xi = \xi^k + \sum_{j=0}^{r_k} \alpha_{k+1-j} d^{k+1-j} \right\} \tag{19}$$
$$\leq \|F(\tilde{\xi})\|, \tilde{\xi} = \xi^k + \tau_k p^{k+1}.$$

For later use, note that $\tau_k p^{k+1} = \tilde{\xi} - \xi^k$.

Next note that

$$\begin{aligned} (20) \qquad F(\tilde{\xi}) &= F(\xi^k) + F'(\xi^k)(\tilde{\xi} - \xi^k) + \\ & \int_0^1 [F'(\xi^k + t(\tilde{\xi} - \xi^k)) - F'(\xi^k)](\tilde{\xi} - \xi^k) dt \\ &= (1 - \tau_k)F(\xi^k) + \tau_k(F(\xi^k) + F'(\xi^k)p^{k+1}) \\ &+ \tau_k \int_0^1 [F'(\xi^k + t(\tilde{\xi} - \xi^k)) - F'(\xi^k)]F'(\xi^k)^{-1}F'(\xi^k)p^{k+1} dt \ . \end{aligned}$$

Note that (12) shows that

$$(21) \qquad \|F'(\xi^k)p^{k+1}\| \leq (1 + \rho_k)\|F(\xi^k)\|$$

and

$$\|p^{k+1}\| \leq \|F'(\xi^k)^{-1}\| \ \|F'(\xi^k)p^{k+1}\| \ ,$$

that is

$$(22) \qquad \|p^{k+1}\| \leq \beta(1 + \rho_k)\|F(\xi^k)\| \ .$$

Also,

$$\begin{aligned} (23) \qquad \|F'(\xi^k)^{-1}F(\xi^k) + p^{k+1}\| &\leq \|F'(\xi^k)^{-1}\| \ \|F(\xi^k) + F'(\xi^k)p^{k+1}\| \\ &\leq \rho_k \|F'(\xi^k)^{-1}\| \ \|F(\xi^k)\| \leq \rho_k \kappa_k \|F'(\xi^k)^{-1}F(\xi^k)\|, \end{aligned}$$

where

$$\kappa_k = \|F'(\xi^k)\| \ \|F'(\xi^k)^{-1}\|,$$

i.e. κ_k is the condition number of the operator $F'(\xi^k)$.

Hence, an alternative estimate of p^{k+1} is

$$(24) \qquad \|p^{k+1}\| \leq (1 + \rho_k \kappa_k)\|F'(\xi^k)^{-1}F(\xi^k)\| \ .$$

(20) and (21, 22) show now that

$$\begin{aligned} \frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|} &\leq 1 - \tau_k + \tau_k \rho_k + \\ \tau_k(1 + \rho_k) & \frac{\| \int_0^1 \left[F'(\xi^k + t(\tilde{\xi} - \xi^k)) - F'(\xi^k)\right] F'(\xi^k)^{-1} dt \|}{\|\tilde{\xi} - \xi^k\|^\gamma} \|\tilde{\xi} - \xi^k\|^\gamma, \end{aligned}$$

that is, using (ii) and (22),

$$\frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|} \leq 1 - \tau_k + \tau_k \rho_k + [\tau_k(1 + \rho_k)]^{1+\gamma} \frac{K_\gamma}{1 + \gamma} [\beta \|F(\xi^k)\|]^\gamma$$

or

$$(25) \qquad \frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|} \leq 1 - \tau_k + \tau_k \rho_k + [\tau_k(1 + \rho_k)]^{1+\gamma} \frac{\tilde{K}_{\gamma,k}}{1 + \gamma},$$

where

$$\tilde{K}_{\gamma,k} = K_\gamma [\beta \|F(\xi^k)\|]^\gamma \ .$$

For the proof of the global convergence theorem we may take τ_k to be

$$\tau_k = \min\{1, \hat{\tau}_k\},$$

where $\hat{\tau}_k$ minimizes the upper bound in (25), that is,

$$\hat{\tau}_k = \left(\frac{1-\rho_k}{(1+\rho_k)^{1+\gamma}}\tilde{K}_{\gamma,k}^{-1}\right)^{1/\gamma} = \left(\frac{1-\rho_k}{(1+\rho_k)^{1+\gamma}}K_{\gamma}^{-1}\right)^{1/\gamma}\beta^{-1}\|F(\xi^k)\|^{-1}. \tag{26}$$

Noting that $0 < \tau_k \leq 1$, the upper bound becomes then

$$\frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|} \leq \begin{cases} \rho_k + \frac{(1+\rho_k)^{1+\gamma}}{1+\gamma}K_\gamma\beta^\gamma\|F(\xi^k)\|^\gamma, & \text{if } \hat{\tau}_k \geq 1 \\ 1 - \hat{\tau}_k\frac{\gamma}{1+\gamma}(1-\rho_k), & \text{if } \hat{\tau}_k < 1. \end{cases} \tag{27}$$

Note now that the upper bound function in (25) as a function of $\tau = \tau_k > 0$ is strictly less than one, initially for sufficiently small values of τ. Hence this and (19) show that

$$\frac{\|F(\xi^{k+1})\|}{\|F(\xi^k)\|} < 1, k = 0, 1, \ldots$$

As this holds in particular for $k = 0$, we have

$$\frac{\|F(\xi^1)\|}{\|F(\xi^0)\|} \leq 1 - \varepsilon, \text{ for some } \varepsilon\ , 0 < \varepsilon < 1,$$

where in fact we can let ε be defined by

$$\varepsilon = \begin{cases} 1 - \rho_0 - \frac{(1+\rho_0)}{1+\gamma}^{1+\gamma}K_\gamma\beta^\gamma\|F(\xi^0)\|^\gamma, & \text{if } \hat{\tau}_0 \geq 1 \\ \hat{\tau}_0\frac{\gamma}{1+\gamma}(1-\rho_0), & \text{if } \hat{\tau}_0 < 1. \end{cases} \tag{28}$$

Note now that because of the minimization property of the algorithm, $\|F(\xi^k)\|$ does not increase with k. Hence (26) shows that $\hat{\tau}_k$ does not decrease, so (27) shows that

$$\frac{\|F(\xi^{k+1})\|}{\|F(\xi^k)\|} \leq 1 - \varepsilon,\ \ k \geq 1,$$

and by induction,

$$\frac{\|F(\xi^{k+1})\|}{\|F(\xi^0)\|} \leq (1-\varepsilon)^{k+1},$$

which shows the global convergence. We state this result:

Theorem 4.1. *Let $F(\cdot)$ be a nonlinear differentiable mapping on $\mathbb{R}^n$ and assume that (i.a) and (ii) hold, where ξ^k is defined in algorithm NNGCG. Then the algorithm converges for any initial approximation ξ^0 and*

$$\|F(\xi^k)\| \leq (1-\varepsilon)^k\|F(\xi^0)\|,\ k \geq 1,$$

where ε is defined in (28). □

Remark. (26) shows that as k increases eventually $1 \leq \hat{\tau}_k$, because $\|F(\xi^k)\| \longrightarrow 0$. Hence (35) shows that if we choose $\rho_k = O(\|F(\xi^k)\|)$ then the algorithm eventually converges with a superlinear rate, namely $\|F(\xi^{k+1})\| = O(\|F(\xi^k)\|^{1+\gamma})$.

Consider now the case where $F'(\cdot)$ satisfies (i.b) and (ii), i.e. $F'(\xi)$ may be singular at the limit point. In this case (20), (21) and (23, 24) show that

$$\begin{aligned}\frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|} &\leq 1-\tau_k+\tau_k\rho_k+\tau_k(1+\rho_k)\frac{K_\gamma}{1+\gamma}\|\tilde{\xi}-\xi^k\|^\gamma\\ &\leq 1-\tau_k+\tau_k\rho_k+\tau_k^{1+\gamma}(1+\rho_k)\frac{K_\gamma}{1+\gamma}(1+\rho_k\kappa_k)^\gamma\cdot\\ &\quad \|F'(\xi^k)^{-1}F(\xi^k)\|^\gamma,\end{aligned}$$

or

$$\frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|}\leq 1-\tau_k+\tau_k\rho_k+\frac{\tau_k^{1+\gamma}}{1+\gamma}\cdot\tilde{K}_{\gamma,k}, \tag{29}$$

where

$$\tilde{K}_{\gamma,k}=(1+\rho_k)K_\gamma(1+\rho_k\kappa_k)^\gamma\|F'(\xi^k)^{-1}F(\xi^k)\|^\gamma$$

Now we choose τ_k such that

$$\tau_k=\min\{1,\hat{\tau}_k\},$$

where $\hat{\tau}_k$ minimizes the upper bound in (29), that is,

$$\hat{\tau}_k=\left(\frac{1-\rho_k}{K_{\gamma,k}}\right)^{1/\gamma}=\left(\frac{1-\rho_k}{1+\rho_k}K_\gamma^{-1}\right)^{1/\gamma}\frac{1}{1+\rho_k\kappa_k}\|F'(\xi^k)^{-1}F(\xi^k)\|^{-1}. \tag{30}$$

Note that for any k, (29) and (17) show that

$$\frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|}\leq 1-\tau_k+\tau_k\rho_k+\frac{\tau_k^{1+\gamma}}{1+\gamma}C, \tag{31}$$

where

$$C=2K_\gamma(1+c)^\gamma\cdot\delta^\gamma\ .$$

Here δ is defined in (23) and $c=\sup_k\rho_k\kappa_k$. We assume here that the numbers ρ_k are chosen such that $c<\infty$. In particular, this means that (12) is solved particularly accurately at the final steps of the algorithm NNGCG, when $F'(\xi^k)$ becomes more singular and κ_k increases.

Note now that there exists a $\tau=\tau_k$ for which (31) is strictly smaller than one, so

$$\frac{\|F(\tilde{\xi})\|}{\|F(\xi^k)\|}\leq 1-\varepsilon, \tag{32}$$

for some, ε, $0<\varepsilon<1$.

As in the proof of the previous theorem, this shows that there exists an $\varepsilon, 0<\varepsilon<1$, such that

$$\frac{\|F(\xi^{k+1})\|}{\|F(\xi^k)\|}\leq 1-\varepsilon,\ k\geq 0$$

and hence global convergence. We have shown:

Theorem 4.2. *Let $F(\cdot)$ be a nonlinear differentiable mapping on $\mathbb{R}^n$ and assume that (i.b) and (ii) hold, where ξ^k is defined in algorithm NNGCG. If (12) is solved sufficiently accurately so that $\rho_k\kappa_k\leq c$ holds for any k, for some constant c, where κ_k is the condition number of $F'(\xi^k)$, then the algorithm converges for any initial approximation ξ^0 and*

$$\|F(\xi^k)\|\leq(1-\varepsilon)^k\|F(\xi^0)\|,\ k\geq 1,$$

for some $\varepsilon, 0 < \varepsilon < 1$. □

The above shows that as long as $F'(\xi^k)$ is nonsingular and $\|F'(\xi^k)^{-1}F(\xi^k)\|$ is uniformly bounded, the algorithm NNGCG can be made to converge by properly handling the relative accuracy parameter sequence $\{\rho_k\}$. In particular, we can solve $F(\xi) = 0$ even if $F(\cdot)$ has a multiple root.

Note that the theorems hold for any version of the NNGCG method, including the truncated versions, where $r_k < k + 1$. In particular they hold for $r_k = 0$, when ξ^{k+1} is computed using just a linesearch.

We conclude with a remark concerning automatic differentiation. The algorithm NNGCG requires updating the Jacobian $F'(\cdot)$ at every iteration step. However, using iterative solution methods to compute the search direction p^{k+1}, the Jacobian is never required in explicit form, but only used implicitly to compute matrix-vector products. Recent improvements of automatic differentiation methods show that such products can be computed with a complexity of the same order as a function evaluation (like $(F(\xi^k), F(\xi^k))_0$). For a survey of such results, see [13]. Hence, it is not computationally efficient in general to avoid updating the Jacobian for a differentiable mapping at every iteration step.

5. Conclusions

We have presented and analyzed a nonlinear generalization of GCG for solving nonlinear algebraic systems of equations. It has been shown that under the assumption that the Jacobian is positive definite and the Hessian is bounded the methods are guaranteed to converge globally to a unique solution. We also proved convergence results for this nonlinear iterative method used in conjunction with nonlinear preconditioning. Under the weaker assumption of a nonsingular and Hölder continuous Jacobian matrix it has also been shown that the combined Newton and NGCG method converges globally. This result includes functions for which the Jacobian is singular in the limit point. The damped (and inexact) Newton methods in [3] and [9] require special choice of steplengths. On the other hand in the NNGCG method the corresponding coefficients α_j are computed automatically by the algorithm.

Acknowledgement. Some comments made by two anonymous referees, which helped improve the presentation of some parts of the paper are gratefully acknowledged.

References

1. Axelsson, O. (1987): A generalized conjugate gradient, least squares method. Numer. Math. **51**, 209–227
2. Axelsson O., Vassilevski, P.S. (1992): Construction of variable-step preconditioners for inner-outer iteration methods. In: Beauwens, R., de Groen, P., eds., Iterative Methods in Linear Algebra, pp. 1–14. North-Holland
3. Axelsson, O. (1982): On global convergence of iterative methods. In: Ansorge, R., Meis, Th., Törnig, W., eds., Iterative Solution of Nonlinear Systems of Equations, LNM ♯953, pp. 1–19. Springer
4. Brown, P.N., Saad, Y. (1990): Hybrid Krylov methods for nonlinear systems of equations. SIAM J. Sci. Stat. Comp. **11**(3), 450–481
5. Chronopoulos, A.T. (1992): A Non-linear CG-like iterative method. J. Comp. Appl. Math. **40**
6. Chronopoulos, A.T., Zlatev, Z. (1992): Iterative methods for nonlinear operator equations. Applied Math. Computation **51**(2,3), 167–180

7. Daniel, J.W. (1967): The conjugate gradient method for linear and nonlinear operator equations. SIAM J. Numer. Anal. **4**, 10–26
8. Daniel, J.W. (1967): Convergence of the conjugate gradient method with computationally convenient modifications. Numer. Math. **10**, 125–131
9. Dembo, R.S., Eisenstat, S.C., Steihaug, T. (1982): Inexact Newton methods. SIAM J. Numer. Anal. **19**, 400–408
10. Dennis, J.E., Schnabel, R.B. (1983): Numerical methods for unconstrained optimization and nonlinear equations. Prentice-Hall, Englewood Cliffs, NJ
11. Eisenstat, S.C., Walker, H.F. (1991) Globally convergent inexact Newton methods. SIAM J. Optimization (to appear)
12. Golub, G.H., Kannan, R. (1986): Convergence of a two stage Richardson process for nonlinear equations. BIT, 209–216
13. Griewank, A. (1990): Direct calculation Newton steps without accumulating Jacobians. In: Coleman, T.F., Yuying, Li, eds., Large-scale numerical optimization, pp. 115–137. SIAM
14. Gummel, H.K. (1964) A self-consistent iterative scheme for one-dimensional steady state transistor calculations. IEEE Trans. Electron Devices **ED-11**, 455–465
15. Kerkhoven, T. (1986): A proof of convergence of Gummel's algorithm for realistic boundary conditions. SIAM J. Numer. Anal. **23(6)**, 1121–1137
16. O'Leary, D. (1982): A discrete Newton algorithm for minimizing a function of many variables. Math. Programming **23**, 20–33
17. Saaty, T.L. (1981): Modern nonlinear equations. Dover Publications Inc., NY

This article was processed by the author using the LaTeX style file *pljour1* from Springer-Verlag.