Proc. Estonian Acad. Sci. Phys. Math., 1994, 43, 2, 49 – 63
https://doi.org/10.3176/phys.math.1994.2.01

ON SOLVING ILL-CONDITIONED SYSTEMS OF NON-LINEAR EQUATIONS

Otu VAARMANN

Tallinna Tehnikaülikool (Tallinn Technical University), Akadeemia tee 1, EE-0026 Tallinn, Eesti (Estonia)

Presented by G. Vainikko

Received October 7, 1993; revised version received December 22, 1993 and February 15. 1994; accepted February 15, 1994

Abstract. Methods under consideration are intended to solve systems of nonlinear equations in n variables where the number of equations m is greater or equal to n and the Jacobian may be ill-conditioned. Two approaches for solving such problems are examined. The first one is based on the computation of the weighted pseudoinverse being, as a rule, concurrent with great computational efforts and therefore the corresponding algorithms are appropriate for solving only small dimensional problems, especially if $m>>n$. The second approach rests on solving approximately corresponding preconditioned linear equations by taking finitely many steps of an iterative procedure. Thus they are applicable to large scale nonlinear problems as well.

Key words: nonlinear equations, least squares solution, weighted pseudoinverse, preconditioner, Neumann's series, regularization.

Giving up some extra work for the numerical stability is particularly justified for problems in which the evaluation of function values and their derivatives are expensive and dominate the other costs. For instance, inverse problems are usually nonlinear and often involve very complicated subproblems with partially unknown properties. As a rule, they are ill-conditioned and numerically unstable. In this case it is reasonable to combine the algorithm with a preconditioner to reduce the condition number.

METHODS

The problem is to solve the nonlinear equation,

$$F(x) = 0, \tag{1}$$

where $F : D \subset R^n \to R^m$, $m \neq n$ and F is continuously differentiable. If the equation (1) has any solution in the classical sense, then it is sought in the

least squares sense, i.e. minimizing the functional

$$T(x) = \frac{1}{2}\| F(x) \|_2^2. \tag{2}$$

1. For solving (1) or for seeking a local minimum of (2), one might use the iterative method

$$x_{k+1} = x_k - [WF'(x_k)]^+ WF(x_k), \tag{3}$$

where $A_k := [WF'(x_k)]^+ W = [F'(x_k)]^+_{WI}$, $W \in R^{n\times m}$ and $[F'(x)]^+_{WI}$ means the WI – weighted pseudoinverse of $F'(x)$ [1].

If $m \geq n$ and $F'(x)$ and W have a full rank, then one gets

$$x_{k+1} = x_k - [WF'(x_k)]^{-1} WF(x_k), \tag{4}$$

which for $m=n$ coinsides with the standard Newton method, and taking $W = W_k = \alpha_k [F'(x_k)]^T, \alpha_k > 0$, one obtains the Gauss-Newton method for $m \geq n$. Here $[\cdot]^T$ denotes the dual mapping.

Let $P_{R(x)}$ denote the orthogonal projector of R^m on to the range of $F'(x)$, and let $\bar{P}_{R(x)}$ be defined by the relation $\bar{P}_{R(x)} = P_{R(x)} \cdot P_W$, where $P_W = W^+W$, and W is an $n \times m$ matrix approximating $[F'(x)]^+$. Let λ denote a positive scalar satisfying the inequality $\|A_k\| \leq \lambda_k = \lambda \leq \infty$ and $P_k = P_{R(F'(x_k))}$.

L e m m a 1 [2, 3]. *Let on some set* $S = \{x \in R^n : \|x - x_0\| \leq \rho\}$ *the following conditions be fulfilled*

$$\| [F'(x)]^+ \| \leq C, \quad \forall x \in S.$$

$$\| F'(x) - F'(y) \| \leq L \| x - y \|, \forall x, y \in S,$$

then

$$\| P_{R(x)} - P_{R(y)} \| \leq L_0 \| x - y \|, \quad \forall x, y \in S,$$

where

$$L_0 = CL.$$

R e m a r k 1. From Lemma 1 it follows that there exist positive scalars N' and N such that

$$\| (P_{R(y)} - P_{R(y)} P_{R(x)}) F(x) \| = \| (P_{R(y)} - P_{R(x)}) (I - P_{R(x)}) F(x) \| =$$

$$= N' \| x - y \|, \qquad \forall x, y \in S,$$

and for $x_k, x_{k+1} \in S$

$$\| (P_{k+1} - P_{k+1} P_k) F(x_k) \| \leq N \| P_k F(x_k) \|,$$

where

$$\sup_{x \in S} \| (I - P_{R(x)}) F(x) \| \le N'', \quad N' = N'' L_0 \text{ and } N = \lambda N'.$$

Since, in general, the equality $\bar{P}_{R(x)} = \bar{P}^2_{R(x)}$ does not hold, then

$$\bar{P}_{R(y)} - \bar{P}_{R(y)} \bar{P}_{R(x)} = (\bar{P}_{R(y)} - \bar{P}_{R(x)}) (I - \bar{P}_{R(x)}) + \bar{P}_{R(x)} - \bar{P}^2_{R(x)}.$$

For $W = [F'(x_0)]^+$ we have

$$\| \bar{P}_{R(y)} - \bar{P}_{R(x)} \| = \| (\bar{P}_{R(y)} - \bar{P}_{R(x)}) P_0 \| \le \| P_{R(y)} - P_{R(x)} \|,$$

$$\bar{P}_{R(x)} - \bar{P}^2_{R(x)} = P_{R(x)} \bar{P}_{R(x)} - P_{R(x)} P_0 \bar{P}_{R(x)} = (P_{R(x)} - P_{R(x)} P_0) \bar{P}_{R(x)}.$$

By Lemma 1 and Remark 1

$$\| \bar{P}_{R(x)} - \bar{P}^2_{R(x)} \| \le \tilde{N} \| x - x_0 \| \le \tilde{N} \rho.$$

$$\| (\bar{P}_{k+1} - \bar{P}_{k+1} \bar{P}_k) F(x_k) \| \le (\bar{N} + \tilde{N} \rho) \| P_k F(x_k) \|,$$

where

$$\sup_{x \in S} \| (I - \bar{P}_{R(x)}) F(x) \| \le \bar{N}'', \quad \bar{N}' = \bar{N}'' L_0, \quad \tilde{N} = L_0 \| I - P_0 \|$$

and $\bar{N} = \lambda \bar{N}'$.

First, we will examine convergence properties of (4) with $W = [F'(x_0)]^+$, and thus $\bar{P}_{R(x)} = P_{R(x)} P_0$.

Theorem 1. *Let* $m \ge n$, $x_0 \in R^n$, $S = \{x \in R^n : \| x - x_0 \| \le \rho\}$ *and the following conditions are valid in* S:

1° *operator* F *has Frechet-derivative and* $\| [F'(x)]^+ \| \le C$;

2° *derivative* F' *satisfies Lipschitz condition*

$$\| F'(x) - F'(y) \| \le L \| x - y \|;$$

3° *derivative* $F'(x)$ *has full rank.*

If $\delta = \bar{N} + \tilde{N} r + \frac{LC^2}{2} \| P_0 F(x_0) \| \le 1$ *and* $r = C \| P_0 F(x_0) \| / (1 - \delta) \le \rho$ *then the sequence generated by* (4) *with* $W = [F'(x_0)]^+$ *has a limit* x^* *which appears to be a solution of the equation* $\bar{P}_{R(x)} F(x) = 0$ *with* $\| x_k - x_0 \| \le r$ *and*

$$\| x_k - x^* \| \le r \delta^k.$$

Proof. If A and B are $m \times n$ and $n \times m$ matrices, respectively, then

$$(AB)^+ = B^+ A^+,$$

provided rank(A) = rank(B) = n [4]. On the basis of this relation one concludes that

$$A_k = \{[F'(x_0)]^+ F'(x_k)\}^+ [F'(x_0)]^+ = [F'(x_k)]^+ F'(x_0)\,[F'(x_0)]^+ =$$

$$= [F'(x_k)]^+ P_0 = [F'(x_k)]^+ P_k P_0 = [F'(x_k)]^+ \bar{P}_k.$$

According to Taylor expansion, one obtains

$$\bar{P}_{k+1} F(x_{k+1}) = (\bar{P}_{k+1} - \bar{P}_{k+1}\bar{P}_k) F(x_k) + \bar{P}_{k+1}\{(\bar{P}_k - F'(x_k) A_k) F(x_k) +$$

$$+ \int_0^1 [F'(x_k) - F'(x_k + t(x_{k+1} - x_k))] A_k F(x_k)\, dt\}.$$

Further

$$\bar{P}_k - F'(x_k) A_k = P_k \bar{P}_k - F'(x_k) A_k \bar{P}_k = (P_k - F'(x_k)\,[F'(x_k)]^+)\bar{P}_k = 0$$

since $P_k = F'(x_k)\,[F'(x_k)]^+$ and

$$\left\| \bar{P}_{k+1} F(x_{k+1}) \right\| \le \left[\bar{N} + \tilde{N} r + \frac{LC^2}{2} \left\| \bar{P}_k F(x_k) \right\| \right] \cdot \left\| \bar{P}_k F(x_k) \right\| \le$$

$$\le \delta \left\| \bar{P}_k F(x_k) \right\| \le \left\| \bar{P}_0 F(x_0) \right\| \delta^{k+1},$$

$$\left\| x_{k+1} - x_k \right\| \le C \left\| \bar{P}_k F(x_k) \right\| \le C \left\| \bar{P}_0 F(x_0) \right\| \delta^k,$$

$$\left\| x_p - x_k \right\| \le r(\delta^k - \delta^p), \qquad p \ge k,$$

$$\left\| x_p - x_0 \right\| \le r(1 - \delta^p) \le r \le \rho, \quad \left\| x^* - x_k \right\| \le r\delta^k,$$

$$x^* = \lim_{k \to \infty} x_k, \qquad \left\| P_{R(x^*)} \cdot P_o \cdot F(x^*) \right\| = \lim_{k \to \infty} \left\| \bar{P}_k F(x_k) \right\| = 0.$$

In particular, if W is an $(m \times m)$-unitary matrix then it can be shown that $[WF'(x)]^+ = [F'(x)]^+ W^{-1}$, and therefore x^* appears to be a solution of the nonlinear normal equation [5]. The problems of global convergence for (3) and the concept of appropriate level functions are discussed in the next section.

2. When large-scale nonlinear problems are to be solved, then the exact solution of associated linear equations at each iteration can be very expensive, not to speak of matrix inversion. In this case, the use of an inexact standard iterative method (e.g. truncated-Newton, truncated

Gauss-Newton, etc.) may be justified, which is based on solving corresponding equations approximately, frequently via the conjugate gradient method. Iterative methods solve $Kx = g$ sometimes faster than elimination methods. Moreover, finding a current approximate solution x_k with high accuracy may not be necessary when x_k is far from the solution x^*.

In particular, if the Newton equation $F'(x_k)(x_{k+1} - x_k) = -F(x_k)$ is to be solved within the tolerance $\eta_k \| F(x_k) \|$ $(\eta_k > 0)$, then the quantity η_k, accurate to a factor, coincides with the quantity γ_k, where $\| I - F'(x_k) A_k \| = \gamma_k$ and $A_k \approx [F'(x_k)]^{-1}$. The sequence $\{\eta_k\}$ is known as forcing sequence, and it can be used for controlling the computational process; specifically, the approximation rate $\gamma_k = 0 \| F(x_k) \|$ ensures that the sequence $\{x_k\}$ converges quadratically (e.g. [6, 7]). The search direction is usually assessed by the norm of the residual. However, the norm of the residual happens to be not always a good predictor of a good search direction, especially if the problem to be solved is ill-conditioned [8, 9] and hence preconditioning is needed. In particular, for symmetric and positive definite linear systems of equations $Kx = g$, a combination of iterative methods and a preconditioner based on incomplete Cholesky factorization (LU factorization) is applicable. That is, one solves iteratively the equivalent system $N^{-1}Kx = N^{-1}g$, where N is the matrix representation of the incomplete factorization. The convergence properties of the iterative process can be further improved by using a polynomical preconditioner

$$C(N^{-1}K)N^{-1}Kx = C(N^{-1}K)N^{-1}g,$$

where $C(\lambda)$ is a preconditioning polynomial (usually $C(\lambda) \approx \lambda^{-1}$) and $C(N^{-1}K)$ is the associated polynomical preconditioner [10]. If N and K are symmetric, then so is the preconditioner $C(N^{-1}K)N^{-1}$ and several variants of the conjugate gradient methods can be put to use.

The execution of one step of the iteration method (e.g. see [11])

$$x_{k+1} = x_k - \varepsilon_k \alpha_k C(\alpha_k [F'(x_k)]^T F'(x_k)) [F'(x_k)]^T F(x_k), \quad \alpha_k > 0,$$

where the parameter ε_k determines the stepsize and

$$C(\alpha_k, x_k) = C(\alpha_k [F'(x_k)]^T F'(x_k)) = I + \Psi_k + \ldots + \Psi_k^{q-1}.$$

$\Psi_k = I - \alpha_k [F'(x_k)]^T F'(x_k)$ with q $(q \geq 1)$ an integer for Ψ^{q-1} can be regarded as the execution of one step of the ordinary iteration method as applied to the preconditioned system

$$\alpha_k C(\alpha_k, x_k)\,[F'(x_k)]^{\mathrm{T}} F(x) = 0, \tag{5}$$

where $\alpha\,[F'(x)]^{\mathrm{T}}$ plays the role of W.

Let us take $A_k := \alpha_k C(\alpha_k, x_k)\,[F'(x_k)]^{\mathrm{T}}$ and rewrite it as $A_k = \alpha_k [F'(x_k)]^{\mathrm{T}}\left[P_k + E_k + \ldots + E_k^{q-1}\right]$, where $E_k = P_k - \alpha_k F'(x_k) \times [F'(x_k)]^{\mathrm{T}}$. If $0 < \alpha_k < 2/K_0^2$, where $\|F'(x)\| \le K_0$ for all $x \in S$, then there exists a positive scalar μ such that $\|E_k\| \le \mu < 1$ and $\lim\limits_{q\to\infty} \alpha_k [F'(x_k)]^{\mathrm{T}}\left[P_k + E_k + \ldots + E_k^{q-1}\right] = [F'(x_k)]^{+}$,

and for the iterative process

$$x_{k+1} = x_k - \varepsilon_k A_k F(x_k), \tag{6}$$

where the parameter ε_k determines the stepsize one obtains.

T h e o r e m 2. *Let, in addition to the conditions* 1°-2° *of Theorem 1, the following conditions be fulfilled in S.*

$$1^{\circ}\ \delta = \delta_0 = 1 - \varepsilon_0 + \varepsilon_0 (N + \mu^q) + \frac{1}{2}\varepsilon^2 \lambda^2 L \|P_0 F(x_0)\| < 1,$$

$$\lambda \equiv \lambda_{q-1} = \min\left\{\frac{2}{K_0}(1 + \mu + \ldots + \mu^{q-1}),\ C(1+\mu^q)\right\};$$

$$2^{\circ}\ 0 < \varepsilon_0 \le \varepsilon_{k-1} \le \varepsilon_k = \min\{1, \varepsilon_{k-1}, \delta^{-1/2}\}.$$

If $r = \lambda(\|P_0 F(x_0)\|/(1-\delta) \le \rho)$, *then the sequence generated by* (6) *has a limit which appears to be a solution of the equation* $[F'(x)]^{\mathrm{T}} F(x) = 0$ *with* $\|x_k - x_0\| \le r$ *and*

$$\|x_k - x^*\| \le r\delta^k.$$

(cf. also Theorem 2 [^{2}]).

P r o o f. Since

$$\|P_k - F'(x_k) A_k\| \le \left\|P_k - \alpha_k F'(x_k)\,[F'(x_k)]^{\mathrm{T}}\right\|^q \le \mu^q$$

and $A_k = A_k P_k$, then

$$\|A_k\| = \|[F'(x_k)]^{+}(F'(x_k) A_k - P_k) + [F'(x_k)]^{+}\| \le C(1+\mu)^q.$$

On the other hand,

$$\|A_k\| \le \frac{2}{K_0}(1 + \mu + \ldots + \mu^{q-1}).$$

and in the capacity of λ_{q-1} one can take

$$\lambda_{q-1} = \min\left\{\frac{2}{K_0}(1+\mu+\ldots+\mu^{q-1}), C(1+\mu)^q\right\}.$$

Taking $U = I$ in Theorem 1 [3] and bearing in mind that α_k stands in it for the regularization parameter, the straightforward application of this theorem yields Theorem 2 under consideration here.

If $F'(x)$ has full rank, then another possibility to obtain the next iterate x_{k+1} is to solve at each iteration the linearized preconditioned equation (5), i.e. the equation $W_k F'(x_k)(x-x_k) = W_k F(x_k)$ with $W_k = A_k = \alpha_k C(\alpha_k, x_k)[F'(x_k)]^T$ or, equivalently, to compute $x_{k+1} = x_k - [W_k F'(x_k)]^{-1} W_k F(x_k)$, while $[W_k F'(x_k)]^{-1} W_k = B^{-1}(x_k) \times [F'(x_k)]^T = [F'(x_k)]^+$, $B(x) = [F'(x)]^T F'(x)$. The preconditioner W_k suffers from the disadvantage that for ill-conditioned Jacobian a rank reduction may occur when the matrix $B(x)$ is constructed. To overcome this drawback, a regularization technique is necessary.

Let H be an arbitrary symmetric positive definite matrix. One possibility to handle safely the rank-deficient (ill posed) problems is to use instead of $B(x)$ the matrix $M(x) := B(x) + \beta H$ with $\beta > 0$, not trying necessarily to find an optimal value for β but choosing it simply large enough to eliminate the singularity of the matrix $B(x)$. Doing so, one obtains the following regularized iterative process

$$x_{k+1} = x_k - \varepsilon_k D_k [F'(x_k)]^T F(x_k), \tag{7}$$

where $D_k = D_k^0\left[I+\ldots+Q_k^{q-1}\right]$, $Q_k = I - M_k D_k^0$, $M_k = M(\beta_k, x_k)$ and D_k^0 is an approximation to M_k^{-1} such that $\|Q_k\| < 1$, while $\lim\limits_{q\to\infty} D_k = M_k^{-1}$ and $\lim\limits_{\substack{q\to\infty \\ \beta_k\to 0}} D_k [F'(x_k)]^T = [F'(x_k)]^+$. In particular, if $0 < \alpha_k < 2/(K_0^2 + \alpha\|H\|)$, $\alpha_k \le \alpha$, then $D_k^0 = \alpha_k I$ is a proper choice.

On the basis of Theorem 1 [3], one can prove for (7) analogous theorem as Theorem 2 for (6).

Alternatively, for solving (1) one might at each iteration to solve the linearized preconditioned equation $W_k F'(x_k)(x-x_k) = W_k F(x_k)$ with $W_k = D_k [F(x_k)]^T$, or to compute $x_{k+1} - x_k = -[W_k F'(x_k)]_R^+ W_k F(x_k)$, where $[\cdot]^+$ denote the reflexive generalized inverse $[W_k F'(x_k)]_R^+ = B_k^+ D_k^{-1}$, and

$$\lim_{\beta_k\to 0} [W_k F'(x_k)]_R^+ W_k = [F'(x_k)]^+.$$

R e m a r k 2. For improving the accuracy of the approximation for $[F'(x)]^+$ based on the Neumann series, one can use the Chebyshev acceleration procedure (e.g. see [11]). When x_i $(i = 1,...,n)$ vary widely in magnitude, it is reasonable to transform the variable space $\bar{x} = Dx$ where D is a diagonal matrix.

MOTIVATION OF METHODS

The usual convergence analysis is based on the monotonicity test of the objective function

$$T(x_{k+1}) \leq T(x_k), \tag{8}$$

where $\{x_k\}$ is a sequence generated by a particular local method. But frequently $\|F(x_k)\|$ is not monotonically decreasing. For badly nonlinear $F(x)$ monotonicity in $T(x)$ may only ultimately occur when iterates $\{x_k\}$ are sufficiently close to the solution x^*. Sometimes the use of relaxation parameter determining the appropriate step size produces a decrease of $\|F(x)\|$ but at a cost of a lower speed of convergence.

That is why one is interested in solving a weighted least squares problem

$$\Phi(x) := \frac{1}{2}\| WF(x) \|_2^2, \tag{9}$$

where W is a weighting matrix. In some case a simple weighting strategy is not sufficient and a dynamic weighting strategy is needed, i.e. one recalculates the matrix W at each iteration once again or periodically after several iterations. Thus, in general, $W=W(x)$. The introduction of a weighting matrix means that one changes the notion of the distance, i.e. instead of $\|y\|_2^2 = y^T y$ one uses $\|y\|_W^2 = y^T W^T W y$, where $\|\cdot\|_W$ means, in general, a seminorm. In particular, if W is a symmetric positive definite matrix, it is usually called a scaling matrix or preconditioner. The basic idea of preconditioning is to introduce a preliminary scaling on the vector of independent variables and dependent variables. For example, instead of solving a linear equation $Kx=g$ explicitely one often solves the equivalent preconditioned system

$$MKx = Mg, \tag{10}$$

where M is a scaling matrix. Specifically, using a diagonal weighting matrix D as a preconditioner rescales the dependent variables, i.e. one changes their units. If M is a proper approximation to K^{-1} or K^+, it will be a lot more efficient to solve (10) than $Kx=g$, because the spectral condition number $\text{cond}_2(MK) = \kappa(MK)$ is considerably less than $\kappa(K)$ or than that

$\kappa(K^{\mathrm{T}}K) = \kappa^2(K)$ which is the case when the Gauss transformation is used.

For a given matrix A the objective function of nonlinear least squares (NLSQ) problems may be written as

$$T(x|A) := \frac{1}{2}\|AF(x)\|_2^2 . \tag{11}$$

The solution (11) defines a level function $T(x|A)$ and an associated with him level set in x_k [8]

$$G_k(A) = \{x \in S | T(x|A) \leq T(x_k|A)\}, x \in S \subseteq R^n.$$

The condition

$$T(x_{k+1}|A) \leq T(x_k|A) \tag{12}$$

is equivalent to the requirement

$$x_{k+1} \in G_k(A) \text{ for some A.}$$

Obviously,

$$0 = T(x^*|J(x^*)^+) \leq T(x|J(x^*)^+)$$

for all $x \in S$, where x^* denotes a solution of the NLSQ problem, $A = J(x^*)^+$ and $J(x) = F'(x)$. Thus the monotonicity test

$$x_{k+1} \in G_k(J(x^*)^+) \tag{13}$$

would be as natural as (12). Note that this test uses a norm in the space of independent variables R^n, which in many cases is more preferable than the usual monotonicity test (8) specifying a norm in the space R^m. For instance, when for expanding the domain of convergence of a Gauss-Newton type method one uses a suitably chosen relaxation (damping) parameter λ to guarantee the convergence of it from a poor starting point, then for steep valley it is desirable to avoid the bottom of the valley as long as the full step of the method is unacceptable [12]. The property of global convergence is a criterion of robustness of the algorithm and therefore finding an appropriate globalization procedure is highly desirable for solving ill-conditioned NLSQ-s. Usually the line search or the trust region strategy is used for globalization of a method. Another efficient way for choosing damping parameter for ill-conditioned problems based on test (12) has been proposed by Deuflhard [8].

As pointed out in [8], the choice $A = W(x) = J(x)^+$ turns out to be a most natural choice, since

$$x_{k+1} = x_k - \text{grad}\,(T(x|J(x_k)^+)\Big|_{x=x_k},$$

i.e. the Gauss-Newton direction for the problem (2) is the steepest descent direction of $T(x|J(x_k)^+)$. Indeed, since

$J(x)^* J(x)^{+^*} = (J(x^+) J(x))^* = J(x^+) J(x)$ and $J(x^+) J(x) J(x^+) = J(x^+)$, one obtains

$$\text{grad}\,(T(x|J(x_k)^+))\big|_{x=x_k} = J(x)^* J(x)^{+^*} J(x_k)^+ F(x)\big|_{x=x_k} =$$

$$= J(x_k)^+ F(x_k).$$

Therefore $T(x|J(x_k)^+)$ is called the natural level function. But, in general,

$$x^* \notin G_k(J(x_k)^+)$$

except for a special case $x_k = x^*$ (unfortunately, the element x^* is unknown) and therefore a global convergence proof cannot be based on the natural monotonocity test (13) and other convergence criteria are necessary. P. Deuflhard introduced the concept of appropriate level functions [13] and showed that the natural level function $T(x|J(x_k)^+)$ is locally appropriate in x_k, and there exists some neighbourhood $U(x_k)$ with $J(y)^+ \in A_0(x_k)$ for all $y \in U(x_k)$, where $A_0(x_k)$ denotes the set of matrices for each of them, the level function $T(x|A)$ is locally appropriate in x_k [8, 13].

Let $r(x) = (I - J(x) J(x)^+) F(x)$, $\Delta x = -J(x)^+ F(x)$, and for fixed x_k let $A(x)$ be some matrix set such that

$$(\Delta x_k)^{\mathrm{T}} \text{grad}\, T(x_k|A) < 0, \text{ if } \Delta x_k \neq 0, \ \forall x_k \in S, \text{ for all } A \in A(x_k).$$

D e f i n i t i o n [8]. *A level function* $T(x|A)$ *is said to be locally appropriate in* x_k *if and only if* $A \in A_0(x_k)$, *where a special sub-set* $A_0(x) \subset A(x)$ *is defined by*

$$\|Ar(x_k)\| < \|AF(x_k)\|, \text{ if } \Delta x_k \neq 0, \text{ for all } A \in A_0(x_k).$$

In general, the definition yields a not easily controllable condition $\sigma(x) := \|WW^+ - WJ(x) J^+(x) W^+\| < 1$ for $T(x|W)$ to be a locally appropriate level function. In particular, for $W = [F'(x_0)]^+$ and for all $x \in S$, one obtains

$$\sup_{x \in S} \| [F'(x_0)]^+ \ (P_0 - P_0 P_{R(x)}) F'(x_0) \| < 1.$$

i.e. $\sigma(x) \leq \kappa(J(x_0)) L_0 \rho < 1$.

ASPECTS OF NUMERICAL REALIZATION

A common complaint about normal equation methods is that they are less accurate than orthogonal methods, since the latter ones preserve the condition number κ of the original problem, while during the solution of the normal system the condition number is squared. This argument is partially true since it ignores the presence of the term κ^2 in the bound of the approximate pseudosolution error for the perturbed problem. Let us consider the system

$$Kx = g, \tag{14}$$

and a perturbed system

$$(K + \delta K)\, x = g + \delta g. \tag{15}$$

A sufficient condition of rank conservation of the problem (12) is the inequality $\|K^+\| \|\delta K\| < 1$. On the basis of the results by Wedin [14] and Stewart [15], one obtains

$$\|\delta \bar{x}\| / \|\bar{x}\| \le \bar{\kappa} \{ (1 + \kappa\rho)\, \alpha + \gamma\beta \}, \tag{16}$$

$$\kappa = \kappa(K) = \|K\| \|K^+\|, \; \bar{\kappa} = \frac{\kappa}{1 - \kappa\alpha} = \frac{\|K\| \|K^+\|}{1 - \|\delta K\| \|K^+\|}, \; \alpha = \frac{\|\delta K\|}{\|K\|},$$

$$\beta = \frac{\|\delta g\|}{\|g\|}, \; \gamma = \frac{\|g\|}{\|K\| \|\bar{x}\|}, \; \rho = \frac{\|\bar{r}\|}{\|K\| \|\bar{x}\|},$$

where $\bar{x}$ is the pseudosolution of the problem (14), $\bar{r}$ is residual corresponding to $\bar{x}$, and $\hat{x} = \bar{x} + \delta\bar{x}$ is the pseudosolution of the perturbed problem (15). Hence, for problems with large residual $\bar{r}$ the right-hand side of (16) includes the term with κ^2 whose adverse influence is unavoidable regardless of the successful choice of the numerical method as applied to the problem. Thus, the use of a preconditioner might be fruitful.

Moreover, if $J(x)$ is highly ill-conditioned, even a very small residual cannot guarantee an accurate solution. In other words, the rounded solution of an equation may have a large residual. The situation is more complicated when an accurate solution has a large residual, specifically when $S(x) := [F''(x)]^T F(x)$ is larger in some sense than $J(x)^T J(x)$ (obviously, if $\|F(x)\|$ is large, then it may cause $\|S(x)\|$ to be also large). It is known that in the latter case the sequence $\{x_k\}$ generated by the

Gauss-Newton method is always divergent even if $J(x)^T J(x)$ is invertible [16]. Then Newton-type methods (e.g. augmented Gauss-Newton methods) as applied to (8) or (11) might be useful since they exploit more completely information on $\nabla^2 T(x)$ or methods with the convergence order $p \geq 3$, as they are based on a quadratic model. Besides, sometimes, the latter permit to reduce the run times in seconds on the computer. Yet one reason for using methods with the convergence order higher than that of the Newton method is that variants of the Gauss-Newton method may experience great difficulty or fail completely if $\nabla T(x)$ is singular or ill-conditioned at the solution point x^*, since they are based on a linear model. On the other hand, using even very rough approximations, the second-order derivatives in methods based on a quadratic model may provide their numerical stability [17, 18].

NUMERICAL EXPERIMENTS

A computer experimentation which is aimed at comparing the relative performance of methods was carried out for:

1. The Fridman method

$$x_{k+1} = x_k - \alpha_k g(x_k), \; g(x) = J(x)^T F(x), \; \alpha_k = \|F(x_k)\|^2 / \|g(x_k)\|^2.$$

2. The two-step gradient method of Măruşter [19]

$$x_{k+1} = x_k - \alpha_k g(x_k) + \beta_k J(x_k)^T J(x_k)\, g(x_k),$$

where the parameters α_k and β_k are determined by the system

$$\beta_k \|J(x_k)\, g(x_k)\|^2 + \alpha_k \|g(x_k)\|^2 + \|F(x_k)\| = 0,$$

$$\beta_k \|J(x_k)^T J(x_k)\, g(x_k)\|^2 + \alpha_k \|J(x_k)\, g(x_k)\|^2 + \|g(x_k)\|^2 = 0.$$

3. The modified Newton method.
4. The Newton method a) with derivatives b) with finite-differences.
5. The method defined by (4) with $W_k = A_{k-1}$.
6. The Gauss-Newton method.
7. The cubically convergent method [20]

$$x_{k+1} = y_k - [F(2y_k - x_k; x_k)]^{-1} F(y_k),$$

$$y_k = x_k - [F(2y_{k-1} - x_{k-1}; x_{k-1})]^{-1} F(x_k),$$

where $f(.;.)$ denotes a finite-difference approximation to $J(x)$.

8. The combination of the latter with the Newton method with finite-differences.

9. The Kogan method [21]

$$x_{k+1} = x_k - \left[F'\left(x_k - \frac{1}{2}\Gamma_k(x_k)\right)\right]^{-1} F(x_k),$$

where $\Gamma_k = [F'(x_k)]^{-1}$.

This paper is not intended to give a numerical comparative study of methods under consideration with the other related methods. For the sake of illustration only some results on their numerical behavior are presented in the table.

Method		1	2	3	4		5	6	7	8	9
Test	Dimension				*a*	*b*					
1	2	*k*>300	4	3	3	3	3	2	2	2	2
2	4	*k*>300	*k*>300	*k*>300	32	31	32	*k*>300	-	38	21
3	2	*k*>300	26	-	13	19	13	14	-	12	9
4	4	*k*>300	*k*>300	*k*>300	15	47	15	15	-	17	28
5	3	*k*>300	134	-	11	-	11	11	-	10	6
6	6	*k*>300	*k*>300	*k*>300	13	45	13	13	-	14	9
7	5	25	6	-	6	6	6	6	-	5	4
8	10	*k*>300	*k*>300	-	91	80	-	-	-	92	*k*>300
9	10	*k*>300	*k*>300	6	4	3	4	4	3	3	3
10	10	16	7	6	4	4	4	4	3	3	3
11	10	*k*>300	*k*>300	*k*>300	8	6	8	8	-	6	7
12	10	15	27	*k*>300	15	20	15	15	-	16	10
13	10	114	33	22	5	5	5	6	-	5	4
14	10	35	15	60	7	6	7	7	-	6	5

The set of test problems containing 14 problems for systems of nonlinear equations of Argonne National Laboratory plus Freudenstein and Roth function and Box three-dimensional function for nonlinear least

squares was taken from [22], and the same starting points were used as in [22].

All the problems were run on a EC 1060 computer under a FORTRAN-IV compiler. The calculus was performed in double precision and stopped $\|x_{k+1} - x_k\| \leq 10^{-9}$. Termination of the routine also occurred when the number of iterations exceeded the given maximum value 300. The table gives iteration number k at which the presigned accuracy was achieved and "-" denotes the failure (nonconvergence).

In conclusion, we discussed the methods mainly on the theoretical basis. One reason for doing so is that mathematical properties exhibit indisputable features of the algorithms; in contrast to the computer experiments their numerical results presented in papers and books often show contradictory aspects. Also, mathematical aspects fix limits of what can be expected from the use of an algorithm.

ACKNOWLEDGEMENT

The author wishes to express his special thanks to Prof. G. Vainikko for very helpful discussions.

REFERENCES

1. Elden, L. A note on weighted pseudoinverses with application to the regularization of Fredholm integral equations of the first kind: Report Lith-MAT-R-1975-11, Department of Mathematics, Linköping University (Sweden), 1975.
2. Ваарманн О., Ломп М. Изв. АН ЭССР . Физ. Матем., 1982, **31**, 4, 410-417.
3. Vaarmann, O. Eesti NSV TA Toim. Füüs.-Matem., 1989, **38**, 2, 146-153.
4. Albert, A. Regression and the Moor-Penrose pseudoinverse. Academic Press, New York-London, 1972.
5. Deuflhard, P. & Heindl, G. SIAM J. Numer. Anal., 1979, **16**, 1-10.
6. Vaarmann, 0. Proc. Estonian Acad. Sci. Phys. Math., 1991, **40**, 2, 99-104.
7. Vaarmann, O. In: Tikhonov, A. N. (ed.). Ill-Posed Problems in Natural Sciences. Utrecht: VSP/ Moscow: TVP Sci. Publ. 1992, 191-201.
8. Deuflhard, P. & Apostolescu, V. Preprint N° 51. Universität Heidelberg, Institute für Angewandte Mathematik. January, 1980.
9. Nash, S & Sofer, A. Oper. Res. Lett., 1990, **4**, 4, 219-221.
10. Ashby, S. F., Manteuffel, T. A. & Otto, J. S. SIAM J. Sci. Comput., 1992, **13**, 1, 1-29.
11. Ваарманн О. Изв. АН ЭССР . Физ. Матем., 1971, **20**, 4, 386-394.
12. Lindqvist, S.-g. Programs for nonlinear least squares problems a user's guide. Computing centre, the Swedish University of Åbo (Finland), Report N° 9, Jan. 1981, 19.
13. Deuflhard, P. Numer. Math., 1974, **22**, 289-315.
14. Wedin, P.-A. Bit, 1973, **13**, 217-232.
15. Stewart, G. W. SIAM Rev., 1977, **19**, 4, 634-662.
16. Пугачев Б. П. ЖВМ и МФ, 1978, **18**, 6, 1593-1595.
17. Ehle, G. & Schwetlick, H. SIAM J. Numer. Anal., 1976, **13**, 3, 433-443.
18. Schnabel, R. B. & Frank, P. D. SIAM J. Numer. Anal., 1984, **21**, 5, 815-843.

19. Măruşter, St. On the two-step gradient method for nonlinear equations. Seminarul De Informatic Si Analiz Numerica. Timişoara, 1985, 20.
20. Ваарманн О., Полль В. Изв. АН ЭССР . Физ. Матем., 1977, **26**, 2, 123-127.
21. Коган Т. И. Сиб. мат. журн., 1967, **8**, 4, 958-960.
22. Moré, J. J., Garbow, B. S. & Hillstrom, K. E. ACM Trans. Math. Soft., 1987, **7**, 1, 17-41.

EBASTABIILSETE MITTELINEAARSETE VÕRRANDISÜSTEEMIDE LAHENDAMISEST

Otu VAARMANN

On vaadeldud m võrrandi ja n otsitavaga $(m \geq n)$ mittelineaarse võrrandisüsteemi lahendamist juhul, kui selle Jacobi maatriksi konditsiooniarv võib olla väga suur. On uuritud sellise ülesande lahendusmeetodeid, mis põhinevad: 1) Jacobi maatriksi kaalutud pseudopöördmaatriksi arvutamisel; 2) vastavate lineaarsete võrrandite ligikaudsel lahendamisel mõne iteratsioonimeetodi abil. On esitatud saadud teoreetilisi uurimistulemusi illustreerivad arvutustulemused.

О РЕШЕНИИ ПЛОХО ОБУСЛОВЛЕННЫХ НЕЛИНЕЙНЫХ СИСТЕМ УРАВНЕНИЙ

Оту ВААРМАНН

Для решения плохо обусловленных систем нелинейных уравнений с m уравнениями и n неизвестными $(m \geq n)$ применяются итерационные методы, основанные на вычислении взвешенной псевдоинверсии для якобиана и на приближенном решении соответствущих линейных уравнений с помощью некоторого итерационного метода. Доказаны теоремы о сходимости рассматриваемых методов. Для иллюстрации полученных теоретических выводов приводятся результаты численных экспериментов.