

Modified Gauss-Newton scheme with worst-case guarantees for global performance

Yu. Nesterov
 (November 2005)

In this paper we suggest a new version of Gauss-Newton method for solving a system of nonlinear equations, which combines the idea of sharp merit function with the idea of quadratic regularization. For this scheme we prove general convergence results and, under a natural non-degeneracy assumption, local quadratic convergence. We analyze the behavior of this scheme on a natural problem class, for which we get global and local worst-case complexity bounds. The implementation of each step of the scheme can be done by standard convex optimization technique.

Keywords: systems of nonlinear equations, Gauss-Newton method, trust-region methods, complexity bounds, global rate of convergence.

AMS Subject Classification: 65J15; 65K05; 90C26

1 Introduction

Motivation. The problem of solving systems of nonlinear equations is one of the most fundamental problems of numerical analysis. The reader can find the main achievements in this field and bibliographical comments in classical monographs [1, 2, 4, 6]. The standard approach to this problem consists in replacing the initial problem

$$\text{Find } x \in R^n : f_i(x) = 0, \quad i = 1, \dots, m, \quad (1)$$

by a minimization problem

$$\min_{x \in R^n} \left[f(x) \stackrel{\text{def}}{=} \phi(f_1(x), \dots, f_m(x)) \right], \quad (2)$$

Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: nesterov@core.ucl.ac.be. The research results presented in this paper have been supported by a grant “Action de recherche concertée ARC 04/09-315” from the “Direction de la recherche scientifique - Communauté française de Belgique”. The scientific responsibility rests with its author.

where the function $\phi(u)$ is non-negative and vanishes only at the origin. The choice recommended most often for the *merit function* $\phi(u)$ is the standard squared Euclidean norm:

$$\phi(u) = \|u\|_{(2)}^2 \equiv \sum_{i=1}^m \left(u^{(i)}\right)^2, \quad (3)$$

where squaring the norm has the advantage of keeping the objective function in (2) smooth enough. Of course, the problem (2), (3) can be solved by standard second-order minimization schemes. However, it is possible to reduce the order of the derivatives used by applying the *Gauss-Newton* method, in which the search direction is defined as a solution of the following auxiliary problem:

$$\min_h \left\{ \phi \left(f_1(x) + \langle f'_1(x), h \rangle, \dots, f_m(x) + \langle f'_m(x), h \rangle \right) : x + h \in D(x) \right\},$$

where $D(x)$ is a properly chosen neighborhood of point x . Under some natural non-degeneracy assumptions, for this strategy it is possible to establish a local quadratic convergence (see, for example, the estimate (10.25) in [6]).

Despite its elegance, the above approach deserves some criticism. Indeed, the transformation of problem (1) into problem (2) is done in a quite dangerous way. For example, if our system of equations is linear, then such a transformation squares the condition number of the problem. Besides increasing numerical instability, for large problems this leads to *squaring* the number of iterations, which is necessary to get an ϵ -solution of the original problem.

In this paper we suggest another approach for solving the systems of non-linear equations. At first glance, it looks very similar to the standard one: We replace our initial problem by a minimization problem (2), but with a *non-smooth* merit function. For example, a possible choice would be $\phi(u) = \|u\|$. Another difference is that at each iteration we compute the new test point as a minimizer of an auxiliary function, which is the sum of a “linearized” merit function with a quadratic proximal term. It appears that under natural assumptions for such a strategy it is possible to guarantee a monotone decrease of the non-smooth objective function in (2). In rather a general situation we can establish global and local quadratic convergence of the scheme. Moreover, for some natural non-convex problem classes we manage to derive global complexity bounds. Note also that the majority of the papers in this field deal with a variant of problem (1) with $m \geq n$ (that corresponds to a least-squares setting). In this paper the most interesting results (see Section 4) are obtained for $m \leq n$, which is indeed a natural format for systems of non-linear equations.

Contents. In Section 2 we define the modified Gauss-Newton step and prove

its main properties. In Section 3 we present the modified Gauss-Newton method. We prove that any limit point of the process satisfies the first-order optimality conditions. If the solution of the system (1) possesses (primal) non-degeneracy, then the convergence is quadratic. In Section 4 we study the class of problems with uniform *dual* non-degeneracy (see the end of this section for exact meaning of the terminology). Note that the problem from this class can have a continuous set of solutions; hence the corresponding Jacobians can be degenerate. Nevertheless, for this class of problems we establish a global efficiency estimate and prove local quadratic convergence. In the last Section 5 we discuss the results. In Section 5.1 we compare the global efficiency of the modified Gauss-Newton method with that of a modified Newton method proposed recently in [5]. In Section 5.2 we discuss the complexity of the auxiliary problems arising in the proposed scheme. We show that these problems can be solved by a standard technique developed for modern trust-region methods (see [1], Chapter 7).

Notation. For denoting (primal) finite-dimensional linear vector spaces, we always use the letter E , which may be marked by an index. This space is endowed with a fixed norm $\|\cdot\|$, which is never indexed. Thus, in order to get the right sense of the notation $\|x\|$, we need to take into account the space containing x . This space is always well defined by a context. We denote by E^* the space of linear functions on E . The value of $s \in E^*$ on $x \in E$ is denoted by $\langle s, x \rangle$. The norms of the primal and dual spaces are related in a usual way:

$$\|s\| = \max_{x \in E} \{\langle s, x \rangle : \|x\| \leq 1\}, \quad s \in E^*.$$

Thus, $\langle s, x \rangle \leq \|s\| \cdot \|x\|$.

For a linear operator $A : E_1 \rightarrow E_2$, its *operator norm* $\|A\|$ is introduced as

$$\|A\| = \max_{x \in E_1} \{\|Ax\| : \|x\| \leq 1\}.$$

Again, the spaces E_1 and E_2 are always well defined by the context. For such an operator we introduce also the minimal singular value:

$$\sigma_{\min}(A) = \min_{x \in E_1} \{\|Ax\| : \|x\| = 1\}.$$

If A is invertible, then $\sigma_{\min}(A) = 1/\|A^{-1}\|$. Note that for two linear operators A_1 and A_2 we have

$$\sigma_{\min}(A_1 A_2) \geq \sigma_{\min}(A_1) \cdot \sigma_{\min}(A_2).$$

If $\sigma_{\min}(A) > 0$, then we say that the operator A possesses *primal non-degeneracy*.

Further, for a linear operator $A : E_1 \rightarrow E_2$ we denote by A^* its *adjoint*:

$$\langle y, Ax \rangle \equiv \langle A^*y, x \rangle \quad \forall x \in E_1, y \in E_2^*.$$

Clearly, A^* maps E_2^* to E_1^* . If $\sigma_{\min}(A^*) > 0$, then we say that A possesses *dual non-degeneracy*.

Finally, for a non-linear function $F(x) : E_1 \rightarrow E_2$ we denote by $F'(x)$ its *Jacobian*, which is a linear operator from E_1 to E_2 :

$$F'(x)h = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [F(x + \alpha h) - F(x)] \in E_2, \quad h \in E_1.$$

In the special case $f(x) : E_1 \rightarrow E_2 \equiv R$, notation $f'(x)$ remains for the *gradient* of the function $f(x)$. In this case we treat $f'(x)$ as an element of E_1^* . For non-differentiable convex function $f(x)$ we denote by $\partial f(x)$ its subdifferential.

2 Modified Gauss-Newton iterate

Consider a smooth non-linear function $F(x) : E_1 \rightarrow E_2$. Our main problem of interest is to find an approximate solution to the following system of equations:

$$F(x) = 0, \quad x \in E_1. \quad (4)$$

In order to measure the quality of such a solution, we introduce a (sharp) *merit* function $\phi(u)$, $u \in E_2$, which satisfies the following conditions:

- It is convex, non-negative and vanishes only at the origin. (Hence, its level sets are bounded.)
- It is Lipschitz-continuous with unit Lipschitz constant:

$$|\phi(u) - \phi(v)| \leq \|u - v\|, \quad \forall u, v \in E_2.$$

- It has a sharp minimum at the origin:

$$\phi(u) \geq \gamma_\phi \|u\|, \quad \forall u \in E_2, \quad (5)$$

for a certain $\gamma_\phi \in (0, 1]$.

For example, we can take $\phi(u) = \|u\|$. Then $\gamma_\phi = 1$.

We can use this merit function for transforming the problem (4) into the following *unconstrained minimization* problem:

$$\min_{x \in E_1} \{ f(x) \equiv \phi(F(x)) \} \stackrel{\text{def}}{=} f^*. \quad (6)$$

Clearly, the solution x^* to the system (4) exists if and only if the optimal value f^* the problem (6) is equal to zero. The iterative scheme proposed below can be seen as a minimization method for problem (6), which employs a special structure of the objective function. The function $f(x)$ can be even non-smooth. However, we will see that it is possible to decrease its value at any x from E_1 excluding the stationary points of (6).

Let us fix some $x \in E_1$. Consider the following *local model* of our objective function:

$$\psi(x; y) = \phi(F(x) + F'(x)(y - x)), \quad y \in E_1.$$

Note that $\psi(x; y)$ is convex in y . Therefore it looks natural to choose the next approximation to the solution of (6) from the set

$$\text{Arg min}_{y \in E_1} \psi(x; y).$$

Such schemes are very well studied in the literature (see [1], [2], [4], [6]). For example, if we take

$$\phi(u) = \left[\sum_{i=1}^m (u^{(i)})^2 \right]^{1/2}, \quad u \in R^m,$$

then we get a classical *Gauss-Newton method*. However, in what follows we argue that a simple regularization of this approach allows us to get a new scheme, for which we can speak about its global efficiency.

We need to introduce the following smoothness assumption. Let \mathcal{F} be a closed convex set in E_1 with non-empty interior.

Assumption 2.1 Function $F(x)$ is differentiable on \mathcal{F} and its derivative is Lipschitz-continuous :

$$\|F'(x) - F'(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{F}, \quad (7)$$

with some $L > 0$.

In what follows we always assume that Assumption 2.1 is satisfied.

LEMMA 2.2 For any x and y from \mathcal{F} we have

$$|f(y) - \psi(x; y)| \leq \frac{1}{2}L\|y - x\|^2. \tag{8}$$

Proof Denote $d(x, y) = F(y) - F(x) - F'(x)(y - x) \in E_2$. By Proposition 3.2.12 in [4],

$$\|d(x, y)\| \leq \frac{1}{2}L\|x - y\|^2.$$

Hence, since both x and y belong to \mathcal{F} , we have

$$\begin{aligned} |f(y) - \psi(x; y)| &= |\phi(F(y)) - \phi(F(x) + F'(x)(y - x))| \\ &\leq \|d(x, y)\| \leq \frac{1}{2}L\|y - x\|^2. \end{aligned}$$

□

Inequality (8) provides us with an upper approximation of function $f(x)$:

$$f(y) \leq \psi(x; y) + \frac{1}{2}L\|y - x\|^2, \quad \forall x, y \in \mathcal{F}.$$

Let us use it for constructing a minimization scheme. Let M be a positive parameter. For the problem (6), define a *modified Gauss-Newton iterate* from a point $x \in \mathcal{F}$ as follows:

$$\boxed{V_M(x) \in \text{Arg} \min_{y \in E_1} [\psi(x; y) + \frac{1}{2}M\|y - x\|^2]}, \tag{9}$$

where "Arg" indicates that $V_M(x)$ is chosen from the set of global minima of the corresponding minimization problem.¹ Note that the auxiliary optimization problem in (9) is *convex* in y . We postpone a discussion of the complexity of finding $V_M(x)$ to Section 5.

Let us prove several auxiliary results. Denote

$$\begin{aligned} r_M(x) &= \|V_M(x) - x\|, \\ f_M(x) &= \psi(x; V_M(x)) + \frac{1}{2}Mr_M^2(x), \\ \delta_M(x) &= f(x) - f_M(x). \end{aligned}$$

¹Since we do not assume that the norm $\|x\|$, $x \in E_1$, is strongly convex, this problem may have a non-trivial set of global solutions.

Note that for a fixed x , $f_M(x)$ is *concave* in M since it can be represented as a minimum of functions, which are linear in M :

$$f_M(x) = \min_{y \in E_1} \left[\psi(x; y) + \frac{1}{2} M \|y - x\|^2 \right].$$

Consequently, the value $\frac{1}{2} r_M^2(x)$, which is equal to the derivative of $f_M(x)$ in M , is a *decreasing* function of M .

LEMMA 2.3 For any $x \in E_1$ we have

$$\delta_M(x) \geq \frac{1}{2} M r_M^2(x). \tag{10}$$

Proof Let us fix an arbitrary $x \in E_1$. Consider the function

$$\xi(t) = \min_{y \in E_1} \left[\phi(F(x) + F'(x)(y - x)) + \frac{1}{2t} \|y - x\|^2 \right]. \tag{11}$$

Note that the set $\{(y, t, \alpha) \in E_1 \times R_+^2 : \|y - x\| \leq (\alpha t)^{1/2}\}$ is convex. Consequently, the objective function of optimization problem in (11) is jointly convex in (y, t) . Therefore the function $\xi(t)$ is convex in t and

$$g(t) \equiv -\frac{1}{2t^2} r_{1/t}^2(x) \in \partial \xi(t).$$

Therefore,

$$f(x) = \xi(0) \geq \xi(t) + g(t) \cdot (-t) = \xi(t) + \frac{1}{2t} r_{1/t}^2(x).$$

Since $\xi(\frac{1}{M}) = f_M(x)$, we get (10). □

Let us compare $\delta_M(x)$ with another natural measure of local decrease of the model $\psi(x; \cdot)$. For $r > 0$ denote

$$\Delta_r(x) = f(x) - \min_{y \in E_1} \{ \psi(x; y) : \|y - x\| \leq r \}.$$

LEMMA 2.4 For any $x \in E_1$ and $r > 0$ we have

$$\delta_M(x) \geq M r^2 \cdot \kappa \left(\frac{1}{M r^2} \Delta_r(x) \right), \tag{12}$$

where

$$\kappa(t) = \begin{cases} t - \frac{1}{2}, & t \geq 1, \\ \frac{1}{2} t^2, & t \in [0, 1]. \end{cases}$$

The right-hand side of bound (12) is a decreasing function of M .

Proof Let us choose $h_r \in \text{Arg} \min_{h \in E_1} \{\psi(x; x+h) : \|h\| \leq r\}$. Then

$$\begin{aligned} f_M(x) &\leq \min_{\tau} \{\phi(F(x) + \tau F'(x)h_r) + \frac{1}{2}M\tau^2 r^2 : \tau \in [0, 1]\} \\ &= \min_{\tau} \{\phi((1-\tau)F(x) + \tau(F(x) + F'(x)h_r)) + \frac{1}{2}M\tau^2 r^2 : \tau \in [0, 1]\} \\ &\leq \min_{\tau} \{(1-\tau)\phi(F(x)) + \tau\phi(F(x) + F'(x)h_r) + \frac{1}{2}M\tau^2 r^2 : \tau \in [0, 1]\} \\ &= \min_{\tau} \{f(x) - \tau\Delta_r(x) + \frac{1}{2}M\tau^2 r^2 : \tau \in [0, 1]\}. \end{aligned}$$

Thus,

$$\delta_M(x) \geq \max_{\tau \in [0, 1]} \{\tau\Delta_r(x) - \frac{1}{2}M\tau^2 r^2\} = Mr^2 \cdot \kappa \left(\frac{1}{Mr^2} \Delta_r(x) \right).$$

Note that the right-hand side of this inequality is decreasing in M . □

Denote

$$\mathcal{L}(\tau) = \{y \in E_1 : f(y) \leq \tau\}.$$

LEMMA 2.5 *Let $\mathcal{L}(f(x)) \subseteq \text{int } \mathcal{F}$ and $M \geq L$. Then $V_M(x) \in \mathcal{L}(f(x))$.*

Proof Assume $V_M(x) \notin \mathcal{L}(f(x))$. Consider the points

$$y(\alpha) = x + \alpha \cdot (V_M(x) - x), \quad \alpha \in [0, 1].$$

Since $y(0) = x \in \text{int } \mathcal{F}$, we can define the value $\bar{\alpha} \in (0, 1)$ such that $y(\bar{\alpha})$ lies on the boundary of the set \mathcal{F} . Note that

$$f(y(\bar{\alpha})) \geq f(x) \geq f_M(x),$$

and $r_M(x) > 0$. By our assumption, $\bar{\alpha} \in (0, 1)$. Denote

$$d = F(y(\bar{\alpha})) - F(x) - \bar{\alpha}F'(x)(V_M(x) - x) \in E_2.$$

In view of Proposition 3.2.12 in [4], $\|d\| \leq \frac{L}{2} \bar{\alpha}^2 r_M^2(x)$. Therefore,

$$\begin{aligned} f(x) &\leq f(y(\bar{\alpha})) = \phi(F(x) + \bar{\alpha}F'(x)(y(1) - x) + d) \\ &\leq \phi((F(x) + \bar{\alpha}F'(x)(V_M(x) - x)) + \|d\|) \\ &\leq (1 - \bar{\alpha})f(x) + \bar{\alpha}\phi((F(x) + F'(x)(V_M(x) - x)) + \frac{1}{2}M\bar{\alpha}^2 r_M^2(x)) \\ &\leq (1 - \bar{\alpha})f(x) + \bar{\alpha}f_M(x) - \frac{1}{2}M\bar{\alpha}(1 - \bar{\alpha})r_M^2(x). \end{aligned}$$

Thus, $f(x) \leq f_M(x) - \frac{1}{2}M(1 - \bar{\alpha})r_M^2(x)$, and that is a contradiction to (10).
□

LEMMA 2.6 *Let both x and $V_M(x)$ belong to \mathcal{F} . Then*

$$f_M(x) \leq \min_{y \in \mathcal{F}} [f(y) + \frac{1}{2}(L + M)\|y - x\|^2]. \quad (13)$$

Proof For $y \in \mathcal{F}$ denote $d(x, y) = F(y) - F(x) - F'(x)(y - x) \in E_2$. By Proposition 3.2.12 [4],

$$\|d(x, y)\| \leq \frac{1}{2}L\|x - y\|^2.$$

Hence, since both x and $V_M(x)$ belong to \mathcal{F} , we have

$$\begin{aligned} f_M(x) &= \min_{y \in \mathcal{F}} [\phi(F(x) + F'(x)(y - x)) + \frac{1}{2}M\|y - x\|^2] \\ &= \min_{y \in \mathcal{F}} [\phi(F(y) - d(x, y)) + \frac{1}{2}M\|y - x\|^2] \\ &\leq \min_{y \in \mathcal{F}} [f(y) + \frac{1}{2}(L + M)\|y - x\|^2]. \end{aligned}$$

□

COROLLARY 2.7 *Let x^* be a solution to the problem (6) and $\mathcal{L}(f(x)) \subseteq \mathcal{F}$. Then*

$$f_M(x) \leq f^* + \frac{1}{2}(L + M)\|x - x^*\|^2. \quad (14)$$

Proof It is enough to substitute $y = x^*$ in the right-hand side of (13). □

3 Modified Gauss-Newton process

Now we can analyze convergence of the following process. Let us fix some $L_0 \in (0, L]$.

Modified Gauss-Newton method	
Initialization: Choose $x_0 \in R^n$.	
<p>Iteration k, ($k \geq 0$):</p> <p>1. Find $M_k \in [L_0, 2L]$ such that</p> $f(V_{M_k}(x_k)) \leq f_{M_k}(x_k).$ <p>2. Set $x_{k+1} = V_{M_k}(x_k)$.</p>	(15)

Since $f_M(x) \leq f(x)$, this process is monotone:

$$f(x_{k+1}) \leq f(x_k). \tag{16}$$

If the constant L is known, then in Item 1 of this scheme we can use $M_k \equiv L$. In the opposite case, it is possible to apply a simple search procedure. The reader could consult [5], Section 5.2, where two efficient strategies are discussed for a similar optimization scheme. Let us present now the convergence results.

Let $x_0 \in \text{int } \mathcal{F}$ be a starting point for the above minimization process. We need to assume the following.

Assumption 3.1 The set \mathcal{F} is big enough: $\mathcal{L}(f(x_0)) \subseteq \mathcal{F}$.

In what follows we always suppose Assumption 3.1 be satisfied. In view of (16) this assumption implies that $\mathcal{L}(f(x_k)) \subseteq \mathcal{F}$ for any $k \geq 0$.

THEOREM 3.2 *For any $k \geq 0$ and $r > 0$ we have*

$$\begin{aligned} f(x_k) - f^* &\geq \frac{1}{2}L_0 \sum_{i=k}^{\infty} r_{M_i}^2(x_i) \geq \frac{1}{2}L_0 \sum_{i=k}^{\infty} r_{2L}^2(x_i), \\ f(x_k) - f^* &\geq r^2 \sum_{i=k}^{\infty} M_i \kappa \left(\frac{1}{M_i r^2} \Delta_r(x) \right) \geq 2Lr^2 \sum_{i=k}^{\infty} \kappa \left(\frac{1}{2Lr^2} \Delta_r(x) \right). \end{aligned} \quad (17)$$

Proof Indeed, in view of the rules of Step 1 in (15),

$$f_{M_i}(x_i) \geq f(x_{i+1}), \quad M_i \geq L_0, \quad r_{M_i}(x_i) \geq r_{2L}(x_i).$$

Thus, inequality (10) justifies the first inequality in (17). In order to prove the second one, we apply (12) and use the bound $M_i \leq 2L$ imposed by (15). \square

COROLLARY 3.3 *Let the sequence $\{x_k\}_{k=0}^{\infty}$ be generated by the scheme (15). Then*

$$\lim_{k \rightarrow \infty} \|x_k - x_{k+1}\| = 0, \quad \lim_{k \rightarrow \infty} \Delta_r(x_k) = 0,$$

and therefore the set of limit points X^ of this sequence is connected. For any \bar{x} from X^* we have $\Delta_r(\bar{x}) = 0$.*

Let us justify now the local convergence of the scheme (15).

THEOREM 3.4 *Let point $x^* \in \mathcal{L}(f(x_0))$ with $F(x^*) = 0$ be a non-degenerate solution to problem (4):*

$$\sigma \equiv \sigma_{\min}(F'(x^*)) > 0.$$

Let γ_ϕ be defined by (5). If $x_k \in \mathcal{L}(f(x_0))$ and

$$\|x_k - x^*\| \leq \frac{2}{L} \cdot \frac{\sigma \gamma_\phi}{3 + 5\gamma_\phi},$$

then $x_{k+1} \in \mathcal{L}(f(x_0))$ and

$$\|x_{k+1} - x^*\| \leq \frac{3(1+\gamma_\phi)L \|x_k - x^*\|^2}{2\gamma_\phi(\sigma - L \|x_k - x^*\|)} \leq \|x_k - x^*\|. \quad (18)$$

Proof Since $f(x^*) = 0$, in view of inequality (14) and Proposition 3.2.12 [4],

we have

$$\begin{aligned}
\frac{3L}{2}\|x_k - x^*\|^2 &\geq f_{M_k}(x_k) \geq \psi(x_k; x_{k+1}) \geq \gamma_\phi \|F(x_k) + F'(x_k)(x_{k+1} - x_k)\| \\
&= \gamma_\phi \|F'(x^*)(x_{k+1} - x^*) + (F(x_k) - F(x^*) - F'(x^*)(x_k - x^*)) \\
&\quad + (F'(x_k) - F'(x^*))(x_{k+1} - x_k)\| \\
&\geq \gamma_\phi \left[\|F'(x^*)(x_{k+1} - x^*)\| - \frac{L}{2}\|x_k - x^*\|^2 - L\|x_k - x^*\| \cdot \|x_{k+1} - x_k\| \right] \\
&\geq \gamma_\phi \left[(\sigma - L\|x_k - x^*\|) \cdot \|x_{k+1} - x^*\| - \frac{3L}{2}\|x_k - x^*\|^2 \right].
\end{aligned}$$

□

4 Global rate of convergence

In order to get global complexity results for method (15), we need to introduce an additional non-degeneracy assumption.

Assumption 4.1 The operator $F'(x) : E_1 \rightarrow E_2$ possesses a uniform *dual* non-degeneracy:

$$\sigma_{\min}(F'(x)^*) \geq \sigma > 0 \quad \forall x \in \mathcal{L}(f(x_0)).$$

Note that this assumption implies $\dim E_2 \leq \dim E_1$. The role of Assumption 4.1 in our analysis can be seen from the following standard result.¹

LEMMA 4.2 *Let linear operator $A : E_1 \rightarrow E_2$ possess dual non-degeneracy: $\sigma_{\min}(A^*) > 0$. Then for any $b \in E_2$ there exists a point $x(b) \in E_1$ such that*

$$Ax(b) = b, \quad \|x(b)\| \leq \frac{\|b\|}{\sigma_{\min}(A^*)}.$$

(That can be easily derived by singular-value decomposition of matrix A .)

An important consequence of Lemma 4.2 is as follows.

LEMMA 4.3 *Let the operator $F'(x)$ possess dual non-degeneracy, that is $\sigma_{\min}(F'(x)^*) > 0$. Then for any $M > 0$ we have*

$$r_M(x) \leq \frac{\|F(x)\|}{\sigma_{\min}(F'(x)^*)}. \quad (19)$$

¹Different variants of this statement are widespread in the literature. Its most general form can be found in [3].

Proof Indeed, in view of Lemma 4.2 there exists h^* such that $F(x) + F'(x)h^* = 0$ and

$$\|h^*\| \leq \frac{\|F(x)\|}{\sigma_{\min}(F'(x)^*)}.$$

Therefore

$$\begin{aligned} \frac{M}{2} r_M^2(x) &\leq \psi(x; V_M(x)) + \frac{M}{2} r_M^2(x) = \min_{h \in E_1} [\psi(x; x+h) + \frac{M}{2} \|h\|^2] \\ &\leq \frac{M}{2} \|h^*\|^2 \leq \frac{M \|F(x)\|^2}{2\sigma_{\min}^2(F'(x)^*)}. \end{aligned}$$

□

Now we can justify the global rate of convergence of scheme (15).

THEOREM 4.4 *Let Assumptions 2.1, 3.1 and 4.1 be satisfied.*

1). *Suppose that the sequence $\{x_k\}_{k=0}^{\infty}$ be generated by method (15). If $f(x_k) \geq \frac{\sigma^2}{2L} \gamma_\phi^2$, then*

$$f(x_{k+1}) \leq f(x_k) - \frac{\sigma^2}{4L} \gamma_\phi^2. \quad (20)$$

Otherwise,

$$f(x_{k+1}) \leq \frac{L}{\sigma^2 \gamma_\phi^2} f^2(x_k) \leq \frac{1}{2} f(x_k). \quad (21)$$

2). *Suppose that the sequence $\{x_k\}_{k=0}^{\infty}$ be generated by method (15) with $M_k \equiv L$. If $f(x_k) \geq \frac{\sigma^2}{L} \gamma_\phi^2$, then*

$$f(x_{k+1}) \leq f(x_k) - \frac{\sigma^2}{2L} \gamma_\phi^2. \quad (22)$$

Otherwise,

$$f(x_{k+1}) \leq \frac{L}{2\sigma^2 \gamma_\phi^2} f^2(x_k) \leq \frac{1}{2} f(x_k). \quad (23)$$

Proof Let us prove the first part of the theorem. Since the operator $F'(x_k)$ is non-degenerate, in view of Lemma 4.2 there exists a solution h_k^* to the system of linear equations $F(x_k) + F'(x_k)h = 0$ with a bounded norm:

$$\|h_k^*\| \leq \frac{1}{\sigma} \|F(x_k)\| \leq \frac{1}{\sigma \gamma_\phi} f(x_k).$$

Therefore, in view of the step-size rules in the scheme (15) and the upper bound on the values M_k , we have

$$\begin{aligned} f(x_{k+1}) &\leq \min_{h \in E_1} \left[\phi(F(x_k) + F'(x_k)h) + \frac{1}{2}M_k \|h\|^2 \right] \\ &\leq \min_{t \in [0,1]} \left[\phi(F(x_k) + tF'(x_k)h_k^*) + L \|th_k^*\|^2 \right] \\ &\leq \min_{t \in [0,1]} \left[\phi((1-t)F(x_k)) + \frac{L}{\sigma^2 \gamma_\phi^2} t^2 f^2(x_k) \right] \\ &\leq \min_{t \in [0,1]} \left[(1-t)f(x_k) + \frac{L}{\sigma^2 \gamma_\phi^2} t^2 f^2(x_k) \right] \end{aligned}$$

Thus, if $f(x_k) \leq \frac{\sigma^2}{2L} \gamma_\phi^2$, then the minimum in the latter one-dimensional problem is attained at $t = 1$ and we get inequalities (21). In the opposite case, the minimum is attained at $t = \frac{\sigma^2 \gamma_\phi^2}{2L f(x_k)}$ and we get estimate (20).

The second part of the theorem can be proved in a similar way. \square

Using Theorem 4.4, we can establish some properties of the problem (6).

THEOREM 4.5 *Let Assumptions 2.1, 3.1 and 4.1 be satisfied. Then there exists a solution x^* to the problem (6) such that $f(x^*) = 0$ and*

$$\|x^* - x_0\| \leq \frac{2}{\sigma} \|F(x_0)\|. \quad (24)$$

Proof Let us choose $\phi(u) = \|u\|$. Then $\gamma_\phi = 1$. Let us apply now method (15) with $M_k \equiv L$ to corresponding problem (6) with $f(x) = \|F(x)\|$.

Assume first, that $f(x_0) > \frac{\sigma^2}{L}$. In accordance to the second statement of Theorem 4.4, as far as $f(x_k) \geq \frac{\sigma^2}{L}$ we have

$$f(x_k) - f(x_{k+1}) \geq \frac{\sigma^2}{2L}. \quad (25)$$

Denote by N the length of the first stage of the process:

$$f(x_N) \geq \frac{\sigma^2}{L} \geq f(x_{N+1}).$$

Summing up inequalities (25) for $k = 0, \dots, N$, we get

$$N + 1 \leq \frac{2L}{\sigma^2} (f(x_0) - f(x_{N+1})). \quad (26)$$

On the other hand, in view of inequality (10) we have

$$f(x_k) - f(x_{k+1}) \geq \frac{L}{2} \|x_k - x_{k+1}\|^2. \quad (27)$$

Summing up these inequalities for $k = 0, \dots, N$, we get

$$\begin{aligned} f(x_0) - f(x_{N+1}) &\geq \frac{L}{2} \sum_{k=0}^N \|x_k - x_{k+1}\|^2 \geq \frac{L}{2(N+1)} \left(\sum_{k=0}^N \|x_k - x_{k+1}\| \right)^2 \\ &\geq \frac{L}{2(N+1)} \|x_0 - x_{N+1}\|^2. \end{aligned}$$

Now, using the estimate (26), we obtain

$$\|x_0 - x_{N+1}\| \leq \left[\frac{2(N+1)}{L} (f(x_0) - f(x_{N+1})) \right]^{1/2} \leq \frac{2}{\sigma} (f(x_0) - f(x_{N+1})). \quad (28)$$

Further, in view of Theorem 4.4, at the second stage of the process we can guarantee that

$$f(x_{k+1}) \leq \frac{L}{2\sigma^2} f^2(x_k) \leq \frac{1}{2} f(x_k), \quad k \geq N+1. \quad (29)$$

Thus, $f(x_{N+k+1}) \leq (\frac{1}{2})^k f(x_{N+1})$ for $k \geq 0$. Hence, in view of inequality (19) we have

$$\|x_{N+k+2} - x_{N+k+1}\| \leq \frac{1}{\sigma} (\frac{1}{2})^k f(x_{N+1}), \quad k \geq 0.$$

Thus, the sequence $\{x_k\}_{k=0}^{\infty}$ converges to a point x^* with $F(x^*) = 0$ and

$$\|x^* - x_{N+1}\| \leq \frac{2}{\sigma} f(x_{N+1}).$$

Taking into account this inequality and (28), we get (24).

If $f(x_0) \leq \frac{\sigma^2}{L}$, then we can apply the latter reasoning from the very beginning:

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{\sigma} \sum_{k=0}^{\infty} f(x_k) \leq \frac{1}{\sigma} f(x_0) \sum_{k=0}^{\infty} (\frac{1}{2})^k = \frac{2}{\sigma} f(x_0).$$

□

Applying exactly the same arguments as in the proof of Theorem 4.5, it is possible to justify the following statement.

THEOREM 4.6 *Let Assumptions 2.1, 3.1 and 4.1 be satisfied. Suppose the sequence $\{x_k\}_{k=0}^{\infty}$ be generated by method (15) as applied to the problem (6). Then this sequence converges to a single point x^* with $F(x^*) = 0$.*

Let us conclude this section with the following remark. We have seen that Assumptions 2.1, 3.1 and 4.1 guarantee existence of a solution to the problem (4). Denote

$$D = \min_x \{\|x - x_0\| : x \in \mathcal{L}(f(x_0)), F(x) = 0\}.$$

In view of Corollary 2.7 and the bounds on M_k in method (15), we can always guarantee that

$$f(x_1) \leq \frac{3}{2}LD^2. \quad (30)$$

Thus, in view of Theorem 4.4, the number of iterations N of method (15), which is necessary for reaching the region of quadratic convergence can be bounded as follows:

$$N \leq 1 + \frac{4L}{\sigma^2\gamma_\phi^2} f(x_1) \leq 1 + 6 \left(\frac{LD}{\sigma\gamma_\phi} \right)^2. \quad (31)$$

We will refer to this bound as to an upper complexity estimate of the class of problems described by Assumptions 2.1, 3.1 and 4.1. This bound is justified by the modified Gauss-Newton method (15).

5 Discussion

5.1 Comparative analysis for scheme (15)

For other methods proposed so far for solving systems of non-linear equations, we did not manage to find in the literature any global worst-case efficiency estimates. Therefore we have to compare the efficiency of method (15) with the only general-purpose scheme, for which such estimates are known. That is a modified Newton scheme for unconstrained minimization proposed recently in [5]. Note that the fields of applications of both methods intersect. Indeed, any problem of solving a system of non-linear equations can be transformed into a problem of unconstrained minimization using a kind of merit function. On the other hand, any unconstrained minimization problem can be reduced to a system of non-linear equations, which correspond to the first-order optimality conditions.

Consider the following unconstrained minimization problem:

$$\min_{x \in E_1} \varphi(x), \quad (32)$$

where $\varphi(x)$ is a twice differentiable strongly convex function, whose Hessian is Lipschitz continuous. Thus, we assume that there exist positive σ and L such that the conditions

$$\begin{aligned} \langle \varphi''(x)h, h \rangle &\geq \sigma \|h\|^2, \\ \|\varphi''(x+h) - \varphi''(x)\| &\leq L \|h\|, \end{aligned} \quad (33)$$

are satisfied for any x and h from E_1 . Denote $D = \|x_0 - x^*\|$. Then in [5], Section 6, it is shown that the complexity of the problem (32) for the modified Newton method (3.3) [5] depends on the characteristic

$$\zeta = \frac{LD}{\sigma}$$

(we use notation of our paper). If $\zeta < 1$, then the problem (32) is easy. In the opposite case, the number of iterations of the modified Newton scheme, which is necessary for reaching the region of quadratic convergence, is bounded by

$$N_1 = 6.25\sqrt{\zeta}, \quad (34)$$

(see (6.1) in [5]).

Note that the problem (32) can be posed in the form (4):

$$\text{Find } x : F(x) \stackrel{\text{def}}{=} \varphi'(x) = 0. \quad (35)$$

Note that $F'(x) = \varphi''(x)$. Therefore, in view of conditions (33), our problem (35) satisfies Assumptions 2.1, 3.1 and 4.1. Let us choose $f(x) = \|F(x)\|$. Then, in view of (31), the number of iterations of the modified Gauss-Newton scheme (15), which is necessary in order to come to the region of quadratic convergence, is bounded by

$$N_2 = 1 + 6\zeta^2. \quad (36)$$

Clearly, the estimate (34) is much better than (36). However, this observation just confirms a standard rule that a specialized procedure must be more efficient than a general purpose scheme. The question is: How much? Needless to say that at this moment we know nothing about lower complexity bounds

of the problem class described by Assumptions 2.1, 3.1 and 4.1. So, there are chances that the complexity (36) can be improved by other methods.

In fact, as compared with the modified Newton method [5], the scheme (15) has one important advantage. The auxiliary problem of computation of the new test point at each iteration of the modified Newton method [5] is solvable in polynomial time only if this method is based on a Euclidean norm. On the contrary, in the modified Gauss-Newton scheme we are absolutely free in the choice of the norms in the spaces E_1 and E_2 . As we will see in Section 5.2, any such a choice results in a convex auxiliary problem. Therefore the norms can be chosen in a reasonable way, which makes the ratio $\frac{L}{\sigma}$ as small as possible.

5.2 Implementation issues

Let us study the complexity of the auxiliary problem (9). For simplicity, let us assume that we choose $f(x) = \|F(x)\|$. So, our problem becomes as follows:

$$\text{Find } f_M(x) = \min_{h \in E_1} [\|F(x) + F'(x)h\| + \frac{1}{2}M\|h\|^2]. \quad (37)$$

Note that sometimes this problem looks easier in its dual form:

$$\begin{aligned} & \min_{h \in E_1} [\|F(x) + F'(x)h\| + \frac{1}{2}M\|h\|^2] \\ &= \min_{h \in E_1} \max_{\substack{s \in E_2^* \\ \|s\| \leq 1}} [\langle s, F(x) + F'(x)h \rangle + \frac{1}{2}M\|h\|^2] \\ &= \max_{\substack{s \in E_2^* \\ \|s\| \leq 1}} \min_{h \in E_1} [\langle s, F(x) + F'(x)h \rangle + \frac{1}{2}M\|h\|^2] \\ &= \max_{s \in E_2^*} [\langle s, F(x) \rangle - \frac{1}{2M}\|F'(x)^*s\|^2 : \|s\| \leq 1]. \end{aligned}$$

Thus, the problem dual to (37) is just a quadratic maximization problem with a simple constraint. Since it is convex, we can apply the standard highly efficient optimization schemes.

Let us show that in the case of Euclidean norms, the problem (37) can be solved by a standard linear algebra technique.

LEMMA 5.1 *Let us introduce in E_1 and E_2 some Euclidean norms:*

$$\|x\| = \langle Q_1x, x \rangle^{1/2}, \quad x \in E_1, \quad \|u\| = \langle Q_2u, u \rangle^{1/2}, \quad u \in E_2.$$

Then the problem (37) can be represented in a dual form as follows:

$$f_M(x) = \min_{\lambda \in R} \left[\frac{1}{2}\lambda + \frac{1}{2} \langle (\lambda Q_2 + \frac{1}{M} F'(x) Q_1^{-1} F'(x)^*)^{-1} F(x), F(x) \rangle : \lambda \geq 0 \right]. \quad (38)$$

If λ^* is an optimal solution to (38), then the solution to (37) is given by

$$h^* = -\frac{1}{M} Q_1^{-1} F'(x)^* (\lambda^* Q_2 + \frac{1}{M} F'(x) Q_1^{-1} F'(x)^*)^{-1} F(x). \quad (39)$$

Proof Indeed

$$\begin{aligned} f_M(x) &= \min_{h \in E_1} \max_{s \in E_2^*} \left[\langle s, F(x) + F'(x)h \rangle + \frac{M}{2} \langle Q_1 h, h \rangle : \langle s, Q_2 s \rangle \leq 1 \right] \\ &= \max_{s \in E_2^*} \left[\langle s, F(x) \rangle - \frac{1}{2M} \langle Q_1^{-1} F'(x)^* s, F'(x)^* s \rangle : \langle s, Q_2 s \rangle \leq 1 \right] \\ &= \max_{s \in E_2^*} \min_{\lambda \geq 0 \in R} \left[\langle s, F(x) \rangle - \frac{1}{2M} \langle Q_1^{-1} F'(x)^* s, F'(x)^* s \rangle + \frac{1}{2} \lambda (1 - \langle s, Q_2 s \rangle) \right] \\ &= \min_{\lambda \geq 0 \in R} \left[\frac{1}{2} \lambda + \frac{1}{2} \langle (\lambda Q_2 + \frac{1}{M} F'(x) Q_1^{-1} F'(x)^*)^{-1} F(x), F(x) \rangle \right]. \end{aligned}$$

□

Note that the one-dimensional optimization problem in (38) can be solved efficiently by a standard technique developed for modern trust-region methods (see [1], Chapter 7).

Finally, let us mention that in Euclidean case our step strategy (39) can be seen as a variant of Levenberg-Marquart method with a special rule for the choice of the proximal parameter λ^* . In non-Euclidean case, the corresponding strategy could be also interpreted as a variant of the trust-region approach. However, the main advantage of our approach is that it is fully automatic. Moreover, it has unambiguous interpretation, which is crucial for justifying the global and local properties of the process (15).

References

- [1] Conn, A.B., Gould, N.I.M. and Toint Ph.L., 2000, *Trust Region Methods* (Philadelphia: SIAM).
- [2] Dennis, J.E. and Schnabel, R.B., 1996, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, (Philadelphia: SIAM) 2nd edition.
- [3] Kantorovich, L.V. and Akilov, G.P., 1964, *Functional analysis in normed spaces* (New York: Pergamon Press).
- [4] Ortega, J.M. and Rheinboldt, W.C., 1970, *Iterative Solution of Nonlinear Equations in Several Variables* (New York: Academic Press).
- [5] Nesterov, Yu. and Polyak, B., Cubic regularization of a Newton scheme and its global performance. Accepted by *Mathematical Programming*.
- [6] Nocedal, J. and Wright, S.J., 1999, *Numerical Optimization* (New York: Springer-Verlag).