



---

Konrad-Zuse-Zentrum für Informationstechnik Berlin



Peter Deuffhard

Global Inexact Newton Methods  
for Very Large Scale Nonlinear Problems

---

## Contents

<b>0. Introduction</b>	<b>1</b>
<b>1. Theoretical Background</b>	<b>2</b>
1.1 Local inexact Newton method . . . . .	2
1.2 Global inexact Newton method . . . . .	5
<b>2. Derivation of the Algorithm</b>	<b>8</b>
<b>References</b>	<b>14</b>

### Abstract

Newton methods for nonlinear problems are known to require the solution of a sequence of linear problems of the same type. For very large scale problems, as understood herein, the arising linear systems can only be solved by iterative methods. Then Newton's iteration appears as outer iteration. The question of interest will be to control the accuracy of the inner iteration such that the convergence speed of Newton's method is preserved. The purpose of the paper is to combine the concept of inexact Newton methods with the concept of the affine invariant exact Newton methods — which is important for problems with ill-conditioned Jacobian matrices (such as typical 2-D or 3-D discretized partial differential equations).

---

The author wishes to thank Mrs. Erlinda Cadano-Körnig for her excellent, quick and patient T<sub>E</sub>X-typing of this manuscript.

## 0. Introduction

Newton methods for nonlinear systems are known to require the solution of linear systems per each iterative step. For problems up to moderate size standard elimination methods will be sufficient. Beyond that, there is a class of large scale problems that can be efficiently solved by *sparse elimination techniques* — this class typically includes discretized partial differential equations (PDE's) in one space dimension (1-D). For even more complex problems, which are called *very large scale* problems herein, *iterative linear solvers* may appear to be the only methods of choice. This problem class includes, for example, discretized PDE's in 2-D and 3-D on sufficiently fine meshes or, in general, engineering systems which consist of a large number of individual modules. As a typical common feature of such problems, the system module structure permits an efficient way of evaluating the associated Jacobian matrix times a given vector — which is the essential prerequisite for the applicability of iterative linear solvers.

Assume now that an iterative method for a very large scale linear problem is at hand. Then so-called *inexact* Newton methods can, in principle, be applied to tackle the associated *very large scale nonlinear* problem. In this case, the Newton iteration appears as *outer* iteration, whereas the iteration to solve the arising linear system per each Newton step appears as inner iteration. In view of an efficient implementation, the question of interest will be to determine the accuracy required in the inner iteration such that the convergence speed of Newton's method is preserved. In other words, a theoretically satisfactory matching between the termination criteria of the inner and outer iteration is needed. This question has been studied in some detail by DEMKO/EISENSTAT/STEHHAUG [2] for *local* Newton methods. Their matching criterion is based on a comparison of the norms of the residuals of the inner and outer iteration. However, there is enough theoretical and empirical evidence (from moderate size up to large scale nonlinear problems) showing that any convergence criteria based on residual norms will drastically slow down the iterative process whenever the arising *Jacobian matrices* are *ill-conditioned* — a fact which led to the development of so-called *affine invariant* Newton methods (compare DEUFLHARD [3], or the more recent monograph [4]). For exact Newton methods, which are connected with direct elimination methods, affine invariant local convergence theorems have been given by DEUFLHARD/HEINDL [6]. For *inexact* Newton methods, YPMA [9] presented local convergence results in an affine invariant setting. The latter approach, however, has two serious defaults:

- a) inexact Newton methods (with non-vanishing inner residuals) are formally no longer affine invariant — hence, the motivation of [6] for an affine

invariant theoretical characterization is no longer given,

- b) unless the not affine invariant matching criterion of [2], the affine invariant matching criterion of [9] cannot be cheaply implemented.

It is the purpose of the present paper to derive a cheaply implementable extension of affine invariant exact Newton methods to the case of inexact Newton methods. The extension is not affine invariant, unless all inner residuals vanish. In Section 1, both local and global inexact Newton methods are analyzed in theoretical terms. Recall that local Newton methods require “sufficiently good” guesses of the solution points as starting points, whereas global Newton methods, by virtue of damping strategies, are applicable to “bad” guesses as well. In Section 2, the theoretical results are used to work out details of an algorithmic implementation.

## 1. Theoretical Background

Let  $F(x) = 0$  with  $F : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $F \in C^1(D)$ , be a nonlinear system to be solved. Then inexact Newton methods are characterized by the fact that the linear systems arising in each step of the Newton iteration (*outer iteration*) are themselves solved by an iterative method (*inner iteration*).

### 1.1 Local inexact Newton method

The *ordinary* inexact Newton method may be written as ( $k = 0, 1, \dots$ )

$$\begin{aligned} \text{a)} \quad F'(x^k)s^k &= -F(x^k) + r^k, \\ \text{b)} \quad x^{k+1} &= x^k + s^k. \end{aligned} \tag{1.1}$$

in terms of the Jacobian matrix  $F'(x^k)$ , the outer residual  $F(x^k)$  and the inner residual  $r^k$ . The termination criterion of the inner iteration controls the size of the inner residual. The question of interest is to produce an efficient matching between the termination criteria of the inner and outer iteration. The main theoretical results of DEMKO/EISENSTAT/STEIHAUG [2] are that *local convergence* can be obtained under the assumption

$$\frac{\|r^k\|}{\|F(x^k)\|} \leq \eta_k \leq \eta < 1 \tag{1.2}$$

and that *superlinear* convergence can be assured if

$$\lim_{k \rightarrow \infty} \eta_k = 0. \quad (1.3)$$

The characterization (1.2) is not affine invariant. An affine invariant local convergence analysis has been given by YPMA [9]: there (1.2) is replaced by the condition

$$\frac{\|F'(x^k)^{-1}r^k\|}{\|F'(x^k)^{-1}F(x^k)\|} \leq \nu_k \leq \nu < 1. \quad (1.4)$$

Accordingly, superlinear convergence shows up under the condition

$$\lim_{k \rightarrow \infty} \nu_k = 0. \quad (1.5)$$

Unlike (1.2), the condition (1.4) cannot be cheaply monitored in the computational context described above.

The approach to be followed herein is to treat inexact Newton methods within the frame of *Newton-like* methods (compare the basic paper of RHEINBOLDT [8] and the more recent paper of DEUFLHARD/HEINDL [6]). This can be done by formally defining some nonsingular Jacobian approximation  $M(x^k)$  such that (1.1) reads

$$M(x^k)s^k = -F(x^k). \quad (1.6)$$

Herein the *nonsingularity* of  $M(x^k)$  is a natural assumption, as long as the inner iteration is realized as an *explicit* recursion. This formulation takes into account that the Jacobian  $F'(x)$  is formally available, but its inverse  $F'(x)^{-1}$  is unavailable and must be approximated — by  $M(x)^{-1}$ . Hence, the condition

$$\frac{\|M(x^k)^{-1}r^k\|}{\|M(x^k)^{-1}F(x^k)\|} \leq \delta_k \leq \delta < 1 \quad (1.7)$$

can also be monitored in the computational process. One may easily observe that

$$\frac{\|M(x^k)^{-1}r^k\|}{\|M(x^k)^{-1}F(x^k)\|} = \frac{\|M(x^k)^{-1}(M(x^k) - F'(x^k))s^k\|}{\|s^k\|} \quad (1.8)$$

The reformulation as a Newton-like iteration leads one to the following asymptotic convergence results.

**Theorem 1** *Notation as just introduced. Let  $M(x), x \in D$ , be nonsingular, defined by (1.1) and (1.6). In addition to (1.7) assume that*

$$\|M(x^k)^{-1}(F'(x^k + t \cdot s^k) - F'(x^k))s^k\| \leq \omega_k t \|s^k\|^2, \quad t \in [0, 1]. \quad (1.9)$$

Then the asymptotic convergence rate of the inexact Newton iteration (1.1) can be characterized by

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^k\|}{\|x^k - x^{k-1}\|} \leq \lim_{k \rightarrow \infty} \delta_k \quad (1.10)$$

In particular, superlinear convergence arises, when

$$\lim_{k \rightarrow \infty} \delta_k = 0 . \quad (1.11)$$

**Proof.** Let  $\|\cdot\|$  denote a vector norm in  $\mathbb{R}^n$ . Then

$$\begin{aligned} \|x^{k+1} - x^k\| &= \|s^k\| = \left\| M(x^k)^{-1} \left[ F(x^k) - \left( F(x^{k-1}) + M(x^{k-1})s^{k-1} \right) \right] \right\| \\ &= \left\| M(x^k)^{-1} \int_{t=0}^1 \left( F'(x^{k-1} + t \cdot s^{k-1}) - M(x^{k-1}) \right) s^{k-1} dt \right\| \\ &\leq \left\| M(x^k)^{-1} \int_{t=0}^1 \left( F'(x^{k-1} + t \cdot s^{k-1}) - F'(x^{k-1}) \right) s^{k-1} dt \right\| \\ &\quad + \left\| M(x^k)^{-1} \left( F'(x^{k-1}) - M(x^{k-1}) \right) s^{k-1} \right\| \end{aligned}$$

Now, with  $\delta_{k-1}$  from (1.7),  $\omega_{k-1}$  from (1.9) and the Kantorovitch quantity (cf. [7], [4])

$$h_{k-1} := \omega_{k-1} \|s^{k-1}\| , \quad (1.12)$$

one concludes that

$$\frac{\|x^{k+1} - x^k\|}{\|x^k - x^{k-1}\|} = \frac{\|s^k\|}{\|s^{k-1}\|} \leq \|M(x^k)^{-1} M(x^{k-1})\| \cdot \left( \frac{1}{2} h_{k-1} + \delta_{k-1} \right) \quad (1.13)$$

From this, the result (1.10) and the condition (1.11) follow directly.  $\blacksquare$

**Remark.** For the sake of clarity, note that the above theoretical characterization is *not affine invariant*, unless  $\delta_k = 0$  for all iteration indices  $k$ .

One is now ready to address the crucial question for an inexact Newton method more specifically: how should the sequence  $\{\delta_k\}$  be chosen in order to preserve the local convergence properties of the ordinary Newton method. A natural requirement will be that inner iteration errors should not dominate the effect coming from the nonlinearity of the problem. From the right-hand side of (1.13), this requirement can be formulated in terms of the local condition

$$\delta_k \leq \rho \cdot h_k , \quad \rho \leq \frac{1}{2} , \quad (1.14)$$

where  $\rho$  is some safety factor to be specified in Section 2 below. Under this condition, (1.13) guarantees *local quadratic convergence*. Of course, the purely theoretical Kantorovitch quantities  $\{h_k\}$  are computationally unavailable in realistic problems. They must be replaced by cheaply available computational estimates — compare Section 2 below.

## 1.2 Global inexact Newton method

The above local convergence results for the ordinary inexact Newton method cover the situation when a “sufficiently good” initial guess  $x^0$  is available. In most nonlinear problems of interest, however, an algorithm will have to cope with “bad” initial guesses as well — or, at least, will have to find out whether a given initial guess  $x^0$  is “sufficiently good” or “bad”. In this situation, global Newton methods, which include a damping strategy, may be applied. In the above introduced notation, the associated iteration now reads

$$\begin{aligned} \text{a)} \quad & M(x^k)s^k = -F(x^k) \\ \text{b)} \quad & x^{k+1} := x^k + \lambda_k s^k, \quad 0 < \lambda_k \leq 1 \end{aligned} \tag{1.15}$$

in terms of a *damping factor*  $\lambda_k$ . Assume now that condition (1.11) holds, hence formally  $M(x^*) = F'(x^*)$ . Upon following the theoretical lines for global exact Newton methods (cf. [4]), one may derive the subsequent theorem to model global convergence of inexact Newton methods with damping strategy.

**Theorem 2** *Let  $F : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $F \in C^1(D)$ ,  $D$  open, convex. Assume  $x^0$ ,  $x^* \in D$  and  $x^*$  unique solution of  $F$  in  $D$  with  $F'(x^*) = M(x^*)$  nonsingular. Moreover, assume that*

$$\begin{aligned} \text{a)} \quad & M(x) \text{ is nonsingular for all } x \in D, \\ \text{b)} \quad & \text{the path-connected component } D_0 \text{ of } G(x^0 | F'(x^*)^{-1}) \\ & \text{in } x^0 \text{ is compact and contained in } D; \\ \text{c)} \quad & \left\| F'(x^*)^{-1} (F'(u + s \cdot v) - F'(u)) \right\| \leq s \cdot \omega_* \|v\| \\ & s \in [0, 1], \quad u + s \cdot v \in D \\ \text{d)} \quad & h_k^* := \omega_* \|s^k\| \cdot \|F'(x^k)^{-1} F'(x^*)\| < \infty \text{ for } x^k \in D_0, \\ \text{e)} \quad & r(x^k) := F'(x^k)s^k + F(x^k), \\ & \frac{\|F'(x^*)^{-1} r(x^k)\|}{\|F'(x^*)^{-1} F(x^k)\|} \leq \delta_*(x^k) \leq \delta < 1 \text{ for } x^k \in D_0. \end{aligned} \tag{1.16}$$

Then the damped inexact Newton iteration (1.15) with

$$\lambda_k \in \left[ \varepsilon, \min \left( 1, \frac{2(1 - \delta_*(x^k))}{h_k^*} - \varepsilon \right) \right] \text{ for } 0 < \varepsilon < \frac{1}{h_k^*} \quad (1.17)$$

converges globally to  $x^*$ . An associated locally optimal damping strategy comes out to be

$$\lambda_k^* := \min \left( 1, \frac{(1 - \delta_*(x^k))}{h_k^*} \right). \quad (1.18)$$

In addition, there exists an index  $k_0 \geq 0$  such that

$$\lambda_k^* = 1 \text{ for all } k \geq k_0. \quad (1.19)$$

The proof is omitted here — for details see [4]. In fact, the essence of Theorem 2 can be obtained from the estimation

$$\|F'(x^*)^{-1}F(x^k + \lambda s^k)\| \leq \left( 1 - \lambda(1 - \delta_*(x^k)) + \frac{\lambda^2}{2} h_k^* \right) \|F'(x^*)^{-1}F(x^k)\|. \quad (1.20)$$

At this point the main idea of affine invariant global *exact* Newton methods may be recalled. One replaces the computationally unavailable Jacobian inverse  $F'(x^*)^{-1}$  iteratively by  $F'(x^k)^{-1}$ . For global *inexact* Newton methods, one will additionally formally replace  $F'(x^k)^{-1}$  by an appropriate approximation in the framework of Newton-like methods.

In order to fix notation, let  $\Delta x^k$  denote the ordinary exact Newton correction and  $\overline{\Delta x}^{k+1}$  the simplified exact Newton correction such that

$$\begin{aligned} \text{a) } & F'(x^k)\Delta x^k = -F(x^k) \\ \text{b) } & F'(x^k)\overline{\Delta x}^{k+1} = -F(x^{k+1}). \end{aligned} \quad (1.21)$$

Then the damped exact Newton iteration

$$x^{k+1} := x^k + \lambda_k \Delta x^k, \quad 0 < \lambda_k \leq 1 \quad (1.22)$$

requires damping factors  $\lambda_k$  such that the so-called *natural monotonicity test*

$$\|\overline{\Delta x}^{k+1}\| \leq \|\Delta x^k\| \quad (1.23)$$

holds for  $k = 0, 1, \dots$ . In a damped inexact Newton method (1.15), the closest approximation of the monotonicity test (1.23) will be

$$\|\overline{s}^{k+1}\| \leq \|s^k\| \quad (1.24)$$

in terms of the *simplified* inexact Newton correction

$$\overline{M}(x^k)\overline{s}^{k+1} = -F(x^{k+1}). \quad (1.25)$$

Of course, the approximation  $\overline{M}(x^k)$  may well differ from the approximation  $M(x^k)$  defined in connection with  $s^k$ . The structure of the monotonicity test (1.24) is now analyzed in detail.

**Lemma 3** *Notation as just introduced and assumptions of Theorem 1. Then the following estimate holds*

$$\|\overline{s}^{k+1}\| \leq \|\overline{M}(x^k)^{-1}M(x^k)\| \cdot \left(1 - \lambda(1 - \delta_k) + \frac{1}{2}\lambda^2 h_k\right) \|s^k\|. \quad (1.26)$$

**Proof.** One follows standard lines of such proofs:

$$\begin{aligned} \overline{s}^{k+1} &= -\overline{M}(x^k)^{-1} \left( F(x^{k+1}) - F(x^k) - M(x^k)s^k \right) \\ &= -\overline{M}(x^k)^{-1} \left[ (1 - \lambda)M(x^k)s^k + \lambda \left( F'(x^k) - M(x^k) \right) s^k \right. \\ &\quad \left. + \int_{t=0}^{\lambda} \left( F'(x^k + t \cdot s^k) - F'(x^k) \right) s^k dt \right]. \end{aligned}$$

Upon inserting the quantities  $\delta_k$  from (1.7),  $\omega_k$  from (1.9), and  $h_k$  from (1.12), one will obtain (1.26). ■

The above result (1.26) inspires the locally optimal damping strategy

$$\overline{\lambda}_k := \min\left(1, \frac{1 - \delta_k}{h_k}\right), \quad (1.27)$$

which obviously is the local analogon of (1.18). Moreover, Lemma 3 permits one to extend condition (1.14) for  $\lambda < 1$ : if one, once more, requires that inner iteration errors should not dominate the effect coming from the nonlinearity, then (1.26) leads one to the condition

$$\delta_k \leq \rho \lambda h_k, \quad \rho \leq \frac{1}{2}. \quad (1.28)$$

Clearly, for  $\lambda h_k \ll 1$ , i.e. for an inefficient damping strategy, the above criterion turns out to be rather restrictive. If, however, the “optimal” strategy  $\lambda_k = \overline{\lambda}_k$  is inserted, then (1.28) is equivalent to

$$\delta_k \leq \frac{\rho}{1 + \rho}, \quad \text{if } h_k \geq \frac{1}{1 + \rho}. \quad (1.29)$$

## 2. Derivation of the Algorithm

The theoretical results of Section 1 are now applied to derive a global *inexact* Newton algorithm. The basic idea to be followed is to mimic an efficient global *exact* Newton algorithm — based on papers by DEUFLHARD [3] and by BOCK [1] — in terms of *leading binary digits*. The implementation of this basic idea means that the above theoretical quantities  $\delta_k$  and  $h_k$  have to be replaced by cheaply available computational estimates. The global *exact* Newton methods due to [3], [1] are already based on the replacement of the Kantorovitch quantities  $h_k$  by cheaply available computational estimates  $[h_k]$  such that

$$[h_k] \leq h_k. \quad (2.1)$$

Therefore, attention will first focus on the computational estimation of  $\delta_k$ .

Let an iterative linear solver supply the ordinary inexact Newton corrections  $s^k$  as approximations of the ordinary exact Newton corrections  $\Delta x^k$ . Then, in the notation of Section 1, the absolute error can be represented as

$$\|s^k - \Delta x^k\| = \|M(x^k)^{-1}(F'(x^k) - M(x^k))\Delta x^k\|. \quad (2.2)$$

Assume now, replacing  $s^k$  by  $\Delta x^k$  in (1.7) and (1.8), that

$$\|M(x^k)^{-1}(F'(x^k) - M(x^k))\Delta x^k\| \leq \delta_k \|\Delta x^k\|. \quad (2.3)$$

This is certainly reasonable for sufficiently small  $\delta_k < 1$ . Upon combining (2.2) and (2.3), one obtains

$$\frac{\|s^k - \Delta x^k\|}{\|\Delta x^k\|} \leq \delta_k < 1. \quad (2.4)$$

On the other hand, let the iterative linear solver yield a reasonable estimate of the relative error of the form

$$\varepsilon_k := \frac{\|s^k - \Delta x^k\|}{\|s^k\|}. \quad (2.5)$$

Then the quantities  $\varepsilon_k$ ,  $\delta_k$  are connected by

$$\frac{\varepsilon_k}{1 + \varepsilon_k} \leq \delta_k \leq \frac{\varepsilon_k}{1 - \varepsilon_k}. \quad (2.6)$$

Hence, the matching strategy to be derived involves an appropriate control of the quantity  $\varepsilon_k$ .

In order to perform the inexact natural monotonicity test (1.24) and to obtain the Kantorovitch estimates  $[h_k]$ , one needs to compute the simplified inexact

Newton corrections  $\bar{s}^{k+1}$  as approximations of the simplified exact Newton corrections  $\overline{\Delta x}^{k+1}$ . Let

$$\bar{\varepsilon}_{k+1} := \frac{\|\bar{s}^{k+1} - \overline{\Delta x}^{k+1}\|}{\|\bar{s}^{k+1}\|} \quad (2.7)$$

denote the associated relative error to be obtained from the inner iteration. In order to be able to exploit Lemma 3, assume that

$$\|\overline{M}(x^k)^{-1}(M(x^k) - \overline{M}(x^k))\| \leq \bar{\delta}_k \ll 1, \quad (2.8)$$

which implies that

$$\|\overline{M}(x^k)^{-1}M(x^k)\| \leq \frac{1}{1 - \bar{\delta}_k}. \quad (2.9)$$

On this basis, the difference between the formal Jacobian approximations  $M(x^k)$  and  $\overline{M}(x^k)$  will be subsequently ignored.

One is now ready to replace the purely theoretical locally optimal damping strategy (1.27) by its estimate

$$\text{a) } \bar{\lambda}_k := \min\left(1, \frac{1 - \hat{\varepsilon}_k}{[h_k]}\right) \quad (2.10)$$

with

$$\text{b) } \hat{\varepsilon}_k := \frac{\varepsilon_k}{1 - \varepsilon_k}.$$

Observe that (2.6) implies

$$\delta_k \leq \hat{\varepsilon}_k. \quad (2.11)$$

Moreover, an accuracy matching strategy can be derived from (1.28) and (2.11):

$$\hat{\varepsilon}_k \leq \rho \min\left(\frac{1}{1 + \rho}, [h_k]\right). \quad (2.12)$$

This strategy turns out to have desirable properties. For  $\bar{\lambda}_k = 1$  in (2.10), one obtains

$$\delta_k \leq \hat{\varepsilon}_k \leq \rho[h_k] \leq \rho h_k,$$

which is just (1.14). For  $\bar{\lambda}_k < 1$ , which implies

$$h_k \geq [h_k] \geq \frac{1}{1 + \rho},$$

one obtains

$$\delta_k \leq \hat{\varepsilon}_k \leq \frac{\rho}{1 + \rho}, \quad \varepsilon_k \leq \frac{\rho}{1 + 2\rho} \quad (2.13)$$

which is (1.29).

The derived matching strategy will lead to the following pattern of iterative behavior: as long as the outer iterates  $x^k$  are still *far away* from the solution point  $x^*$ , *only few* inner iterations are required to realize the Newton method with damping; once the outer iterates are *close to*  $x^*$ , then the number of inner iterations *increases superlinearly* according to the *superlinear convergence criterion* (1.11). This structure is of basic importance both in terms of computing time and reliability of computations.

As for the accuracy of the computational estimate  $[h_k]$ , the following result will be helpful.

**Lemma 4** *Notation as introduced above. Assume that the inequalities*

$$[h_k] \leq h_k < 2 \max([h_k], 1 - \hat{\varepsilon}_k), \quad (2.14)$$

*hold. Then, with  $\overline{M}(x^k) = M(x^k)$  and  $\lambda_k = \overline{\lambda}_k$ , inexact natural monotonicity*

$$\|\overline{s}^{k+1}\| \leq \|s^k\| \quad (2.15)$$

*is guaranteed.*

**Proof.** One starts from the estimate (1.26) with  $\lambda = \overline{\lambda}_k$  as in (2.10). Let  $\overline{\lambda}_k < 1$  first, so (2.14) implies  $h_k < 2[h_k]$ . Hence

$$\begin{aligned} \frac{\|\overline{s}^{k+1}\|}{\|s^k\|} &\leq 1 - \frac{1 - \hat{\varepsilon}_k}{[h_k]}(1 - \delta_k) + \frac{1}{2} \frac{(1 - \hat{\varepsilon}_k)^2}{[h_k]^2} \cdot h_k \\ &\leq 1 - \frac{1 - \hat{\varepsilon}_k}{[h_k]}(\hat{\varepsilon}_k - \delta_k) = 1 - \lambda(\hat{\varepsilon}_k - \delta_k). \end{aligned}$$

For  $\overline{\lambda}_k = 1$ , (2.14) yields  $h_k < 2(1 - \hat{\varepsilon}_k)$ , which implies

$$\frac{\|\overline{s}^{k+1}\|}{\|s^k\|} \leq \delta_k + \frac{1}{2}h_k < 1 - (\hat{\varepsilon}_k - \delta_k).$$

Upon recalling (2.11), the result (2.15) is confirmed. ■

With these preparations, the computational estimates  $[h_k]$  can be studied in more detail. In [4] the a-priori estimate (for  $k > 0$ )

$$[h_k] := \frac{\|\overline{\Delta x^k} - \Delta x^k\|}{\|x^k - x^{k-1}\|} \cdot \frac{\|\Delta x^k\|}{\|\overline{\Delta x^k}\|} \leq h_k \quad (2.16)$$

has been defined as a slight modification of an earlier suggestion [3]. In [1], an a-posteriori estimate (for  $k \geq 0$ )

$$[h_k] := \frac{2\|\overline{\Delta x}^{k+1}(\lambda) - (1-\lambda)\Delta x^k\|}{\lambda^2\|\Delta x^k\|} \leq h_k, \quad (2.17)$$

has been proposed based on a new trial iterate  $x^{k+1} = x^k + \lambda\Delta x^k$ . Up to now, these estimates have only been used to supply reasonable damping factors  $\lambda_k$  — repeatedly, if natural monotonicity is violated. In the present context, these estimates are additionally needed to control the required accuracy of the inner iteration. If one replaces

$$\Delta x^k, \overline{\Delta x}^k \longrightarrow s^k, \overline{s}^k,$$

then approximation errors enter into the above estimates. Let

$$[\overline{h}_k] := \frac{\|\overline{s}^k - s^k\|}{\|x^k - x^{k-1}\|} \cdot \frac{\|s^k\|}{\|\overline{s}^k\|} \quad (2.18)$$

denote the approximate a-priori estimate and

$$[\overline{h}_k] := \frac{2\|\overline{s}^{k+1}(\lambda) - (1-\lambda)s^k\|}{\lambda^2\|s^k\|} \quad (2.19)$$

the approximate a-posteriori estimate. Upon recalling the proof of Lemma 3 and setting  $\overline{M}(x^k) \doteq M(x^k)$ , the a-posteriori estimate (2.19) can be further analyzed to yield

$$[\overline{h}_k] \leq h_k + \frac{2\delta_k}{\lambda} \quad (2.20)$$

and, by insertion of (1.28)

$$[\overline{h}_k] \leq (1+2\rho)h_k, \quad \rho \leq \frac{1}{2}. \quad (2.21)$$

In view of Lemma 4, the first binary digit in  $[\overline{h}_k]$  needs to be correct, which means to choose  $\rho \ll 1$  sufficiently small. Observe that (2.13) for the boundary value  $\rho = \frac{1}{2}$  implies

$$\delta_k \leq \widehat{\varepsilon}_k = \frac{1}{3}, \quad \varepsilon_k = \frac{1}{4}, \quad \text{if } \overline{\lambda}_k < 1. \quad (2.22)$$

In words, more than *two leading binary digits* of accuracy are required in the ordinary inexact Newton correction  $s^k$ . Upon asking for at least three digits, one obtains

$$\rho \leq \frac{1}{6} \Rightarrow \delta_k \leq \widehat{\varepsilon}_k \leq \frac{1}{7}, \quad \varepsilon_k \leq \frac{1}{8}, \quad \text{if } \overline{\lambda}_k < 1. \quad (2.23)$$

An associated detailed analysis of the approximate a-priori estimate (2.18) would not lead to useful terms. There is, however, a very natural handling of the two estimates  $[\bar{h}_k]$  in combination with the *inner* iteration towards  $s^k$  and  $\bar{s}^{k+1}$ .

In principle, global Newton methods as *outer* iteration can be combined with any iterative solver as *inner* iteration. In order to construct an efficient combination of inner/outer iteration, the inner iterative solver is required to supply the (final) inexact Newton corrections together with a reliable estimate of their *relative accuracy*. Note that the final residual is not needed. Obtaining the relative errors, however, may involve some slight modifications of known methods. A recent presentation of some old and new iterative solvers is given in [5]. The structure of the global inexact Newton method inspires natural *starting values for the inner iteration*: for the iteration towards  $\bar{s}^{k+1}$  one may take  $(1 - \lambda_k)s^k$ , for the iteration towards  $s^{k+1}$  one may take  $\bar{s}^{k+1}$ . In both cases, only the contribution from the nonlinearity needs to be collected. Thus one may directly compute the terms

$$\begin{aligned} \text{a) } \delta s^k &:= s^k - \bar{s}^k \\ \text{b) } \delta \bar{s}^{k+1} &:= \bar{s}^{k+1} - (1 - \lambda_k)s^k \end{aligned} \quad (2.24)$$

without cancellation of leading digits. At the same time, one may observe that  $\|\delta s^k\|$  is required in the a-priori estimate (2.18), while  $\|\delta \bar{s}^{k+1}\|$  is required in the a-posteriori estimate (2.19). This fact can be easily exploited.

First, assume that  $s := s^k$ ,  $\lambda := \lambda_k$  is given and  $\bar{s} := \bar{s}^{k+1}$ ,  $\delta \bar{s} := \delta \bar{s}^{k+1}$  must be computed. Let the inner iteration for  $\bar{s}$  be written as

$$\begin{aligned} \text{a) } \bar{s}_0 &:= (1 - \lambda)s \\ \text{b) } \bar{s}_{i+1} &:= \bar{s}_i + \bar{\Delta}_i, \quad i = 0, 1, \dots, \end{aligned} \quad (2.25)$$

which is equivalent to

$$\begin{aligned} \text{a) } \delta \bar{s}_0 &= 0 \\ \text{b) } \delta \bar{s}_{i+1} &:= \delta \bar{s}_i + \bar{\Delta}_i, \quad i = 0, 1, \dots \\ \text{c) } \bar{s}_{i+1} &:= \bar{s}_0 + \delta \bar{s}_{i+1}, \quad i = 0, 1, \dots \end{aligned} \quad (2.26)$$

Let  $\alpha(\bar{\Delta}_i)$  denote an estimate of the *absolute* error of both  $\bar{s}$  and  $\delta \bar{s}$ . Then, upon recalling the requirements of Lemma 4 and (2.23), a reasonable termination criterion for the inner iteration towards  $\bar{s}^{k+1}$  appears to be

$$\alpha(\bar{\Delta}_i) \leq \frac{1}{4} \max \left( \|\delta \bar{s}_{i+1}\|, \frac{1}{2} \|\bar{s}_{i+1}\| \right). \quad (2.27)$$

This criterion guarantees sufficient accuracy of the a-posteriori estimate (2.19) and, at the same time, assures a satisfactory performance of the inexact monotonicity test (2.15).

Next, assume that  $\bar{s} := \bar{s}^k$  is given and  $s := s^k$ ,  $\delta s := \delta s^k$  must be computed. In this case, the inner iteration may read

$$\begin{aligned} \text{a) } s_0 &:= \bar{s} \\ \text{b) } s_{i+1} &:= s_i + \Delta_i, \quad i = 0, 1, \dots, \end{aligned} \quad (2.28)$$

which is equivalent to

$$\begin{aligned} \text{a) } \delta s_0 &:= 0 \\ \text{b) } \delta s_{i+1} &:= \delta s_i + \Delta_i, \quad i = 0, 1, \dots \\ \text{c) } s_{i+1} &:= s_0 + \delta s_{i+1}; \quad i = 0, 1, \dots \end{aligned} \quad (2.29)$$

With  $\alpha(\Delta_i)$  an estimate of the *absolute* error of both  $s$  and  $\delta s$ , one ends up with the comparable criterion

$$\alpha(\Delta_i) \leq \frac{1}{4} \max \left( \|\delta s_{i+1}\|, \frac{1}{2} \|s_{i+1}\| \right). \quad (2.30)$$

This criterion guarantees sufficient accuracy of the a-priori estimate (2.18) — thus defining a first guess of the damping factor  $\bar{\lambda}_k$ . If  $\bar{\lambda}_k < 1$  and if the criterion (2.30) has been activated such that

$$\alpha(\Delta_i) \leq \frac{1}{8} \|s_{i+1}\|, \quad (2.31)$$

then the ordinary inexact Newton correction  $s^k := s_{i+1}$  is already sufficiently accurate. Otherwise, the inner iteration towards  $s^k$  is continued until the requirement (2.12) is met — now, with a reasonable estimate  $[\bar{h}_k]$  at hand.

For  $\bar{\lambda}_k = 1$ , an alternative device can be used. In this case, the exact Kantorovitch quantities are known to satisfy a majorant equation (cf. [7])

$$h_k = \frac{1}{2} h_{k-1}^2. \quad (2.32)$$

One may therefore assume a similar behavior for the estimates

$$[\bar{h}_k] \doteq \frac{1}{2} [\bar{h}_{k-1}]^2, \quad (2.33)$$

which allows one to take the previous a-posteriori estimate  $[\bar{h}_{k-1}]$ .

Summarizing, the damping strategy (2.10), where  $[h_k]$  is replaced by  $[\bar{h}_k]$ , can now be performed — just as in the exact Newton iteration. Numerical experiments with the thus derived algorithm will be presented elsewhere.

## References

- [1] Bock, H.G.: *Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics*. In: Ebert/Deuffhard/Jäger (ed.): *Modelling of Chemical Reaction Systems*. Berlin-Heidelberg-New York: Springer Ser. Chem. Phys., vol. **18** (1981), p. 102-125 (1981).
- [2] Demko, R. S., Eisenstat, S. C., Steihaug, T.: *Inexact Newton Methods*. SIAM J. Numer. Anal. **18**, p. 400-408 (1982).
- [3] Deuffhard, P.: *A Relaxation Strategy for the Modified Newton Method*. In: Bulirsch/Oettli/Stoer (ed.): *Optimization and Optimal Control*. Springer Lecture Notes **477**, p. 59-73 (1975).
- [4] Deuffhard, P.: *Newton Techniques for Highly Nonlinear Problems — Theory, Algorithms, Codes*. Academic Press, Inc. (To be published).
- [5] Deuffhard, P., Freund, R., Walter, A.: *Fast Secant Methods for the Iterative Solution of Large Nonsymmetric Linear Systems*. Konrad-Zuse-Zentrum für Informationstechnik Berlin: Preprint, to appear (1990).
- [6] Deuffhard, P., Heindl, G.: *Affine Invariant Convergence Theorems for Newton's Method and Extensions to Related Methods*. SIAM J. Numer. Anal. **16**, p. 1-10 (1979).
- [7] Kantorovitch, L., Akhilov, G.: *Functional analysis in normed spaces*. Moscow: Fizmatgiz (1959).
- [8] Rheinboldt, W. C.: *A unified convergence theory for a class of iterative processes*. SIAM J. Numer. Anal. **5**, p. 42-63 (1968).
- [9] Ypma, T. J.: *Local Convergence of Inexact Newton Methods*. SIAM J. Numer. Anal. **21**, p. 583-590 (1984).

---