On the Convergence of Inexact Newton Methods

Reijer Idema, Domenico Lahaye, and Cornelis Vuik

Abstract A solid understanding of convergence behaviour is essential to the design and analysis of iterative methods. In this paper we explore the convergence of inexact iterative methods in general, and inexact Newton methods in particular. A direct relationship between the convergence of inexact Newton methods and the forcing terms is presented in both theory and numerical experiments.

1 Introduction

Inexact Newton methods [1] are Newton-Raphson methods in which the Jacobian system $-J(\mathbf{x}_i)\, s_i = \mathbf{F}(\mathbf{x}_i)$ is not solved to full accuracy. Instead, in each Newton iteration the Jacobian system is solved such that

$$\frac{\|\mathbf{r}_i\|}{\|\mathbf{F}(\mathbf{x}_i)\|} \leq \eta_i, \tag{1}$$

where $\mathbf{r}_i$ is the residual vector:

$$\mathbf{r}_i = \mathbf{F}(\mathbf{x}_i) + J(\mathbf{x}_i)\, s_i. \tag{2}$$

The values η_i are called the forcing terms. Over the years a great deal of research has gone into finding good values for η_i, such that convergence is reached with the least amount of computational work. One of the most frequently used methods to calculate η_i is that of Eisenstat and Walker [3].

In this paper, we further study the relationship between the convergence of inexact Newton methods and the choice of forcing terms. We show, both in theory and numerical experiments, that if the iterate $\mathbf{x}_i$ is close enough to the solution, in iteration i the Newton method converges in some norm with a factor $(1 + \alpha)\, \eta_i$, for arbitrarily small $\alpha > 0$.

R. Idema (✉) • D. Lahaye • C. Vuik
Delft Institute of Applied Mathematics, Delft University of Technology, Delft, Netherlands
e-mail: mail@reijeridema.nl; d.j.p.lahaye@tudelft.nl; c.vuik@tudelft.nl

A. Abdulle et al. (eds.), *Numerical Mathematics and Advanced Applications - ENUMATH 2013*, Lecture Notes in Computational Science and Engineering 103, DOI 10.1007/978-3-319-10705-9_35

2 Convergence of Inexact Iterative Methods

Assume an iterative method that, given current iterate $\mathbf{x}_i$, has some way to determine a unique new iterate $\hat{\mathbf{x}}_{i+1}$. If instead an approximation $\mathbf{x}_{i+1}$ of the exact iterate $\hat{\mathbf{x}}_{i+1}$ is used to continue the process, we speak of an inexact iterative method. Inexact Newton methods are examples of inexact iterative methods. Figure 1 illustrates a single step of an inexact iterative method.

Assume that the solution $\mathbf{x}^*$, and the distances ε^c, ε^n, and $\hat{\varepsilon}$ to the solution are unknown, but that the ratio $\frac{\delta^n}{\delta^c}$ can be controlled. In inexact Newton methods this ratio is controlled using the forcing terms. The aim is then to have an improvement of the controllable error impose a similar improvement on the distance to the solution, i.e., that for some reasonably small $\alpha > 0$

$$\frac{\varepsilon^n}{\varepsilon^c} \leq (1+\alpha)\,\frac{\delta^n}{\delta^c}. \tag{3}$$

Define $\gamma = \frac{\hat{\varepsilon}}{\delta^c} > 0$, then we can write

$$\max\frac{\varepsilon^n}{\varepsilon^c} = \frac{\delta^n + \hat{\varepsilon}}{|\delta^c - \hat{\varepsilon}|} = \frac{\delta^n + \gamma\delta^c}{|1-\gamma|\,\delta^c} = \frac{1}{|1-\gamma|}\frac{\delta^n}{\delta^c} + \frac{\gamma}{|1-\gamma|}. \tag{4}$$

Therefore, to guarantee that $\mathbf{x}_{i+1}$ is closer to the solution than $\mathbf{x}_i$, it is required that

$$\frac{1}{|1-\gamma|}\frac{\delta^n}{\delta^c} + \frac{\gamma}{|1-\gamma|} < 1 \Leftrightarrow \frac{\delta^n}{\delta^c} + \gamma < |1-\gamma| \Leftrightarrow \frac{\delta^n}{\delta^c} < |1-\gamma| - \gamma. \tag{5}$$

If $\gamma \geq 1$ this would mean that $\frac{\delta^n}{\delta^c} < -1$, which is impossible. Therefore, to guarantee a reduction of the distance to the solution, we need

$$\frac{\delta^n}{\delta^c} < 1 - 2\gamma \Leftrightarrow 2\gamma < 1 - \frac{\delta^n}{\delta^c} \Leftrightarrow \gamma < \frac{1}{2} - \frac{1}{2}\frac{\delta^n}{\delta^c}. \tag{6}$$

Equation (4) implies that as γ goes to 0, $\max\frac{\varepsilon^n}{\varepsilon^c}$ more and more resembles $\frac{\delta^n}{\delta^c}$. Figure 2 clearly shows that making $\frac{\delta^n}{\delta^c}$ too small leads to oversolving, as there is hardly any return of investment any more. Note that if the iterative method converges to the solution superlinearly, then γ goes to 0 with the same rate of convergence. Thus, for such a method $\frac{\delta^n}{\delta^c}$ can be made smaller and smaller in later iterations

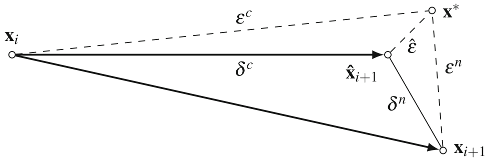

Fig. 1 Inexact iterative step

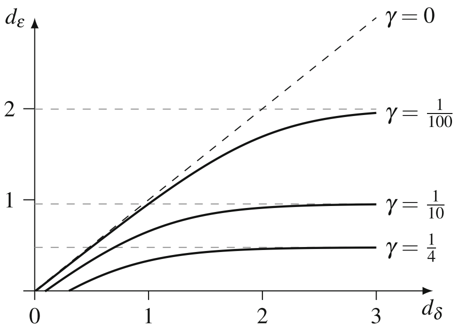

Fig. 2 Plots of equation (4) on a logarithmic scale, for several values of γ. The horizontal axis shows the number of digits improvement in the distance to the exact iterate, and the vertical axis depicts the resulting minimum digits improvement in the distance to the solution, i.e., $d_\delta = -\log \frac{\delta^n}{\delta^c}$ and $d_\varepsilon = -\log \left(\max \frac{\varepsilon^n}{\varepsilon^c}\right)$

without significant oversolving, This is in particular the case for inexact Newton methods, as convergence is quadratic once the iterate is close enough to the solution.

When using an inexact Newton method $\frac{\delta^n}{\delta^c} = \frac{\|\mathbf{x}_{i+1}-\hat{\mathbf{x}}_{i+1}\|}{\|\mathbf{x}_i-\hat{\mathbf{x}}_{i+1}\|}$ is not known, but the relative residual error $\frac{\|\mathbf{r}_i\|}{\|\mathbf{F}(\mathbf{x}_i)\|} = \frac{\|J(\mathbf{x}_i)\left(\mathbf{x}_{i+1}-\hat{\mathbf{x}}_{i+1}\right)\|}{\|J(\mathbf{x}_i)\left(\mathbf{x}_i-\hat{\mathbf{x}}_{i+1}\right)\|}$, which is controlled by the forcing terms η_i, can be used as a measure for it. In the next section, this idea is formalized in a theorem that is a variation on Eq. (3).

3 Convergence of Inexact Newton Methods

Consider the nonlinear system of equations $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, where:

- There is a solution $\mathbf{x}^*$ such that $\mathbf{F}(\mathbf{x}^*) = \mathbf{0}$,
- The Jacobian matrix J of $\mathbf{F}$ exists in a neighbourhood of $\mathbf{x}^*$,
- $J(\mathbf{x}^*)$ is continuous and nonsingular.

In this section, theory is presented that relates the convergence of the inexact Newton method for a problem of the above form directly to the chosen forcing terms. The following theorem is a variation on both Eq. (3), and on the inexact Newton convergence theorem presented in [1, Thm. 2.3].

Theorem 1 *Let* $\eta_i \in (0, 1)$ *and choose* $\alpha > 0$ *such that* $(1 + \alpha)\,\eta_i < 1$. *Then there exists an* $\varepsilon > 0$ *such that, if* $\|\mathbf{x}_0 - \mathbf{x}^*\| < \varepsilon$, *the sequence of inexact Newton iterates* $\mathbf{x}_i$ *converges to* $\mathbf{x}^*$, *with*

$$\|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_{i+1} - \mathbf{x}^*\right)\| < (1 + \alpha)\,\eta_i\,\|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\|. \tag{7}$$

Proof Define

$$\mu = \max[\|J\left(\mathbf{x}^*\right)\|, \|J\left(\mathbf{x}^*\right)^{-1}\|] \geq 1. \tag{8}$$

Recall that $J(\mathbf{x}^*)$ is nonsingular. Thus μ is well-defined and we can write

$$\frac{1}{\mu}\|\mathbf{y}\| \leq \|J\left(\mathbf{x}^*\right)\mathbf{y}\| \leq \mu\|\mathbf{y}\|. \tag{9}$$

Note that $\mu \geq 1$ because the induced matrix norm is submultiplicative.

Let

$$\gamma \in \left(0, \frac{\alpha\eta_i}{5\mu}\right) \tag{10}$$

and choose $\varepsilon > 0$ sufficiently small such that if $\|\mathbf{y} - \mathbf{x}^*\| \leq \mu^2\varepsilon$ then

$$\|J(\mathbf{y}) - J\left(\mathbf{x}^*\right)\| \leq \gamma, \tag{11}$$

$$\|J(\mathbf{y})^{-1} - J\left(\mathbf{x}^*\right)^{-1}\| \leq \gamma, \tag{12}$$

$$\|\mathbf{F}(\mathbf{y}) - \mathbf{F}\left(\mathbf{x}^*\right) - J\left(\mathbf{x}^*\right)\left(\mathbf{y} - \mathbf{x}^*\right)\| \leq \gamma\|\mathbf{y} - \mathbf{x}^*\|. \tag{13}$$

That such an ε exists follows from [6, Thm. 2.3.3 & 3.1.5].

First we show that if $\|\mathbf{x}_i - \mathbf{x}^*\| < \mu^2\varepsilon$, then Eq. (7) holds. Write

$$\begin{aligned} J\left(\mathbf{x}^*\right)\left(\mathbf{x}_{i+1} - \mathbf{x}^*\right) = &\left[I + J\left(\mathbf{x}^*\right)\left(J\left(\mathbf{x}_i\right)^{-1} - J\left(\mathbf{x}^*\right)^{-1}\right)\right] \cdot [\mathbf{r}_i + \\ &\left(J(\mathbf{x}_i) - J\left(\mathbf{x}^*\right)\right)\left(\mathbf{x}_i - \mathbf{x}^*\right) - \left(\mathbf{F}(\mathbf{x}_i) - \mathbf{F}\left(\mathbf{x}^*\right) - J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\right)]. \end{aligned} \tag{14}$$

Taking norms gives

$$\begin{aligned} \|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_{i+1} - \mathbf{x}^*\right)\| \leq &\left[1 + \|J\left(\mathbf{x}^*\right)\|\|J(\mathbf{x}_i)^{-1} - J\left(\mathbf{x}^*\right)^{-1}\|\right] \cdot [\|\mathbf{r}_i\| + \\ &\|J(\mathbf{x}_i) - J\left(\mathbf{x}^*\right)\|\|\mathbf{x}_i - \mathbf{x}^*\| + \|\mathbf{F}(\mathbf{x}_i) - \mathbf{F}\left(\mathbf{x}^*\right) - J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\|], \\ &\leq [1 + \mu\gamma] \cdot \left[\|\mathbf{r}_i\| + \gamma\|\mathbf{x}_i - \mathbf{x}^*\| + \gamma\|\mathbf{x}_i - \mathbf{x}^*\|\right], \\ &\leq [1 + \mu\gamma] \cdot \left[\eta_i\|\mathbf{F}(\mathbf{x}_i)\| + 2\gamma\|\mathbf{x}_i - \mathbf{x}^*\|\right]. \end{aligned} \tag{15}$$

Here the definitions of η_i and μ were used, together with Eqs. (11)–(13).

Further write, using that by definition $\mathbf{F}(\mathbf{x}^*) = \mathbf{0}$,

$$\mathbf{F}(\mathbf{x}_i) = \left[J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\right] + \left[\mathbf{F}(\mathbf{x}_i) - \mathbf{F}\left(\mathbf{x}^*\right) - J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\right]. \tag{16}$$

Again taking norms gives

$$\begin{aligned}\|\mathbf{F}(\mathbf{x}_i)\| &\le \|J(\mathbf{x}^*)(\mathbf{x}_i-\mathbf{x}^*)\| + \|\mathbf{F}(\mathbf{x}_i)-\mathbf{F}(\mathbf{x}^*)-J(\mathbf{x}^*)(\mathbf{x}_i-\mathbf{x}^*)\| \\ &\le \|J(\mathbf{x}^*)(\mathbf{x}_i-\mathbf{x}^*)\| + \gamma\|\mathbf{x}_i-\mathbf{x}^*\|. \end{aligned} \tag{17}$$

Substituting Eq. (17) into Eq. (15) then leads to

$$\begin{aligned}&\|J(\mathbf{x}^*)(\mathbf{x}_{i+1}-\mathbf{x}^*)\| \\ &\le (1+\mu\gamma)\left[\eta_i\left(\|J(\mathbf{x}^*)(\mathbf{x}_i-\mathbf{x}^*)\| + \gamma\|\mathbf{x}_i-\mathbf{x}^*\|\right) + 2\gamma\|\mathbf{x}_i-\mathbf{x}^*\|\right] \\ &\le (1+\mu\gamma)\left[\eta_i(1+\mu\gamma)+2\mu\gamma\right]\|J(\mathbf{x}^*)(\mathbf{x}_i-\mathbf{x}^*)\|. \end{aligned} \tag{18}$$

Here Eq. (9) was used to write $\|\mathbf{x}_i-\mathbf{x}^*\| \le \mu\|J(\mathbf{x}^*)(\mathbf{x}_i-\mathbf{x}^*)\|$.

Finally, using that $\gamma \in \left(0, \frac{\alpha\eta_i}{5\mu}\right)$, and that both $\eta_i < 1$ and $\alpha\eta_i < 1$—the latter being a result from the requirement that $(1+\alpha)\eta_i < 1$—gives

$$\begin{aligned}(1+\mu\gamma)\left[\eta_i(1+\mu\gamma)+2\mu\gamma\right] &\le \left(1+\frac{\alpha\eta_i}{5}\right)\left[\eta_i\left(1+\frac{\alpha\eta_i}{5}\right)+\frac{2\alpha\eta_i}{5}\right] \\ &= \left[1+\frac{2\alpha\eta_i}{5}+\frac{\alpha^2\eta_i^2}{25}+\frac{2\alpha}{5}+\frac{2\alpha^2\eta_i}{25}\right]\eta_i \\ &< \left[1+\frac{2\alpha}{5}+\frac{\alpha}{25}+\frac{2\alpha}{5}+\frac{2\alpha}{25}\right]\eta_i \\ &< (1+\alpha)\,\eta_i. \end{aligned} \tag{19}$$

Equation (7) follows by substituting Eq. (19) into Eq. (18).

Given that Eq. (7) holds if $\|\mathbf{x}_i-\mathbf{x}^*\| < \mu^2\varepsilon$, we now proceed to prove Theorem 1 by induction.

For the base case

$$\|\mathbf{x}_0-\mathbf{x}^*\| < \varepsilon \le \mu^2\varepsilon. \tag{20}$$

Thus Eq. (7) holds for $i = 0$.

The induction hypothesis that Eq. (7) holds for $i = 0, \ldots, k-1$ then gives

$$\begin{aligned}\|\mathbf{x}_k-\mathbf{x}^*\| &\le \mu\|J(\mathbf{x}^*)(\mathbf{x}_k-\mathbf{x}^*)\| \\ &< \mu(1+\alpha)^k\,\eta_{k-1}\cdots\eta_0\|J(\mathbf{x}^*)(\mathbf{x}_0-\mathbf{x}^*)\| \\ &< \mu\|J(\mathbf{x}^*)(\mathbf{x}_0-\mathbf{x}^*)\| \\ &\le \mu^2\|\mathbf{x}_0-\mathbf{x}^*\| \\ &< \mu^2\varepsilon. \end{aligned} \tag{21}$$

Thus Eq. (7) also holds for $i = k$, completing the proof. □

In words, Theorem 1 states that for arbitrarily small $\alpha > 0$, and any choice of forcing terms $\eta_i \in (0, 1)$, Eq. (7) holds if the current iterate is close enough to the solution. This does not mean that for a certain iterate $\mathbf{x}_i$, one can choose α and η_i arbitrarily small and expect Eq. (7) to hold, as ε depends on the choice of α and η_i.

If we define oversolving as using forcing terms η_i that are too small for the iterate, in the context of Theorem 1, then the theorem can be characterised by saying that a convergence factor $(1 + \alpha)\,\eta_i$ is attained if η_i is chosen such that there is no oversolving. Using Eq. (10), $\eta_i > \frac{5\mu\gamma}{\alpha}$ can then be seen as a theoretical bound on the forcing terms that guards against oversolving.

A note on preconditioning is in order. Right preconditioning does not change the residual, and thus it does not change the interpretation of the forcing term η_i in Theorem 1. However, left preconditioning changes the residual such that η_i is closer to the ratio $\frac{\delta^n}{\delta^c}$. As a result, a theoretical relation closer to Eq. (3) is expected. Indeed, following the proof of Theorem 1 for a left-preconditioned problem, we get

$$\|M^{-1} J\left(\mathbf{x}^*\right)\left(\mathbf{x}_{i+1} - \mathbf{x}^*\right)\| < (1 + \alpha)\,\eta_i \|M^{-1} J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\|, \tag{22}$$

where norms of the form $\|M^{-1} J\left(\mathbf{x}^*\right)(\mathbf{y} - \mathbf{x}^*)\|$ are close to $\|\mathbf{y} - \mathbf{x}^*\|$ for a good preconditioner M.

A relation between the nonlinear residual norm $\|\mathbf{F}(\mathbf{x}_i)\|$ and the error norm $\|J\left(\mathbf{x}^*\right)(\mathbf{x}_i - \mathbf{x}^*)\|$ can also be derived within the neighbourhood of the solution where Theorem 1 holds. This shows that the nonlinear residual norm is indeed a good measure of convergence of the Newton method.

Theorem 2 *Let* $\eta_i \in (0, 1)$ *and choose* $\alpha > 0$ *such that* $(1 + \alpha)\,\eta_i < 1$. *Then there exists an* $\varepsilon > 0$ *such that, if* $\|\mathbf{x}_0 - \mathbf{x}^*\| < \varepsilon$, *then*

$$\left(1 - \frac{\alpha\eta_i}{5}\right)\|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\| < \|\mathbf{F}(\mathbf{x}_i)\| < \left(1 + \frac{\alpha\eta_i}{5}\right)\|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\|. \tag{23}$$

Proof Using that $\mathbf{F}\left(\mathbf{x}^*\right) = \mathbf{0}$ by definition, again write

$$\mathbf{F}(\mathbf{x}_i) = \left[J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\right] + \left[\mathbf{F}(\mathbf{x}_i) - \mathbf{F}\left(\mathbf{x}^*\right) - J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\right]. \tag{24}$$

Taking norms, and using Eqs. (13) and (9), gives

$$\begin{aligned} \|\mathbf{F}(\mathbf{x}_i)\| &\leq \|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\| + \|\mathbf{F}(\mathbf{x}_i) - \mathbf{F}\left(\mathbf{x}^*\right) - J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\| \\ &\leq \|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\| + \gamma\|\mathbf{x}_i - \mathbf{x}^*\| \\ &\leq \|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\| + \mu\gamma\|J\left(\mathbf{x}^*\right)\mathbf{x}_i - \mathbf{x}^*\| \\ &= (1 + \mu\gamma)\,\|J\left(\mathbf{x}^*\right)\left(\mathbf{x}_i - \mathbf{x}^*\right)\|. \end{aligned} \tag{25}$$

Similarly, it holds that

$$\begin{aligned}
\|\mathbf{F}(\mathbf{x}_i)\| &\geq \|J(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}^*)\| - \|\mathbf{F}(\mathbf{x}_i) - \mathbf{F}(\mathbf{x}^*) - J(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}^*)\| \\
&\geq \|J(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}^*)\| - \gamma\|\mathbf{x}_i - \mathbf{x}^*\| \\
&\geq \|J(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}^*)\| - \mu\gamma\|J(\mathbf{x}^*)\mathbf{x}_i - \mathbf{x}^*\| \\
&= (1 - \mu\gamma)\|J(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}^*)\|. \qquad (26)
\end{aligned}$$

The theorem now follows from (10). □

4 Numerical Experiments

Both classical Newton-Raphson convergence theory [2, 6], and the inexact Newton convergence theory by Dembo et al. [1], require the current iterate to be close enough to the solution. What exactly is "close enough" depends on the problem, and is in practice generally too difficult to calculate. Decades of practice have shown that the theoretical convergence is reached within a few Newton steps for most problems. Thus the theory is not just of theoretical, but also of practical importance.

In this section, experiments are presented to illustrate the practical merit of Theorem 1. For simplicity, we test an idealised version of relation (7):

$$\|\mathbf{x}_{i+1} - \mathbf{x}^*\| < \eta_i \|\mathbf{x}_i - \mathbf{x}^*\|. \qquad (27)$$

The experiments in this section are performed on a power flow problem [4, 5] that results in a nonlinear system of approximately 256k equations, with a Jacobian matrix that has around 2M nonzeros. The linear Jacobian systems are solved using GMRES [7], preconditioned with a high quality ILU factorisation of the Jacobian.

In Figs. 3–5, the results are shown for different amounts of GMRES iterations per Newton step. In all cases two Newton steps with just a single GMRES iteration were performed at the start but omitted from the figure.

Figure 3 has a distribution of GMRES iterations that leads to a fast solution of the problem. Practical convergence nicely follows theory. This suggests that $\mathbf{x}_2$ is close enough to the solution to use the chosen forcing terms without oversolving.

Figure 4 shows the convergence for a more exotic distribution of GMRES iterations, illustrating that practice can also follow theory for such a scenario.

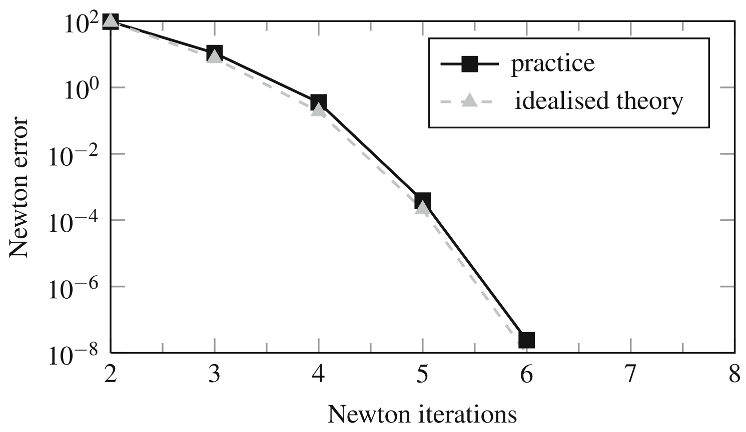

Fig. 3 GMRES iteration distribution 1, 1, 4, 6, 10, 14

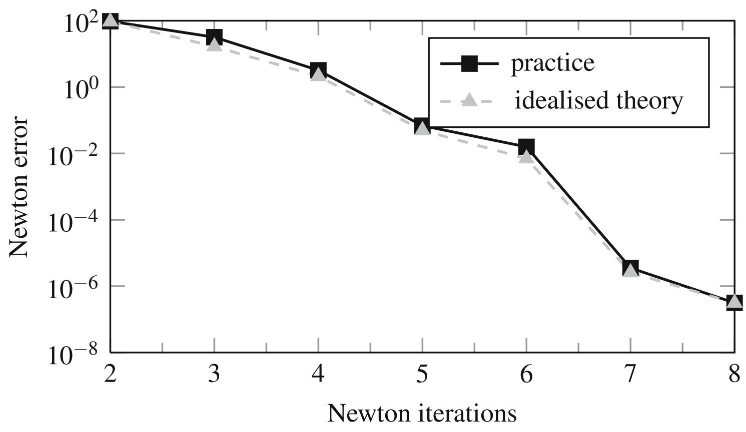

Fig. 4 GMRES iteration distribution 1, 1, 3, 4, 6, 3, 11, 3

Figure 5 illustrates the impact of oversolving. Practical convergence is nowhere near the idealised theory because extra GMRES iterations are performed that do not further improve the Newton error. In terms of Theorem 1 this means that the iterates $\mathbf{x}_i$ are not close enough to the solution to be able to take the forcing terms η_i as small as they were in this example.

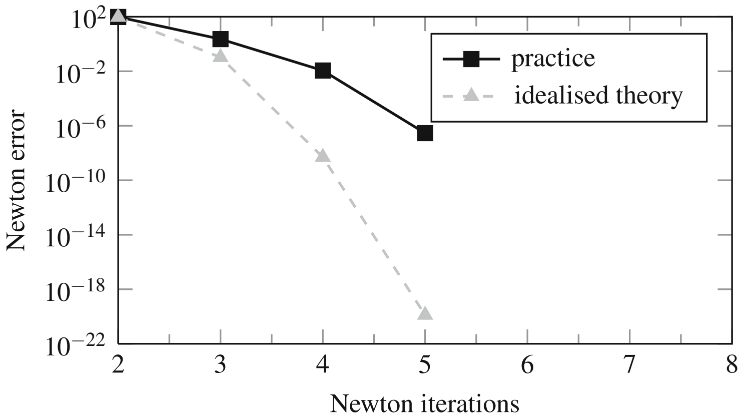

Fig. 5 GMRES iteration distribution 1, 1, 9, 19, 30

Conclusions

A proper choice of tolerances in inexact iterative methods is very important to minimize computational work. In the case of inexact Newton methods these tolerances are called the forcing terms.

In this paper we explored the relation between the choice of tolerances and the convergence of inexact iterative methods, and in particular the relation between the forcing terms and the convergence of inexact Newton methods. We proved that, under certain conditions, in each iteration an inexact Newton method converges with a factor near equal to the forcing term of that iteration, and numerical experiments were used to illustrate the results.

References

1. R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods. SIAM J. Numer. Anal. **19**(2), 400–408 (1982)
2. J.E. Dennis Jr., R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (SIAM, Philadelphia, 1996)
3. S.C. Eisenstat, H.F. Walker, Choosing the forcing terms in an inexact Newton method. SIAM J. Sci. Comput. **17**(1), 16–32 (1996)
4. R. Idema, D.J.P. Lahaye, C. Vuik, L. van der Sluis, Scalable Newton-Krylov solver for very large power flow problems. IEEE Trans. Power Syst. **27**(1), 390–396 (2012)
5. R. Idema, G. Papaefthymiou, D.J.P. Lahaye, C. Vuik, L. van der Sluis, Towards faster solution of large power flow problems. IEEE Trans. Power Syst. **28**(4), 4918–4925 (2013)
6. J.M. Ortega, W.C. Rheinboldt, Iterative solution of nonlinear equations in several variables (SIAM, Philadelphia, 2000)
7. Y. Saad, M.H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput. **7**, 856–869 (1986)