Konrad-Zuse-Zentrum für Informationstechnik Berlin
Heilbronner Str. 10, D-10711 Berlin-Wilmersdorf

Peter Deuflhard Martin Weiser

Local Inexact Newton Multilevel FEM for Nonlinear Elliptic Problems

Preprint SC 96–29 (September 1996)

Local Inexact Newton Multilevel FEM for Nonlinear Elliptic Problems

Peter Deuflhard Martin Weiser

Contents

Abstract

The finite element setting for nonlinear elliptic PDEs directly leads to the minimization of convex functionals. Uniform ellipticity of the underlying PDE shows up as strict convexity of the arising nonlinear functional. The paper analyzes computational variants of Newton's method for convex optimization in an *affine conjugate* setting, which reflects the appropriate affine transformation behavior for this class of problems. First, an affine conjugate Newton–Mysovskikh type theorem on the local quadratic convergence of the *exact* Newton method in Hilbert spaces is given. It can be easily extended to *inexact* Newton methods, where the inner iteration is only approximately solved. For fixed finite dimension, a special implementation of a Newton–PCG algorithm is worked out. In this case, the suggested monitor for the inner iteration guarantees *quadratic* convergence of the outer iteration. In infinite dimensional problems, the PCG method may be just formally replaced by any Galerkin method such as FEM for linear elliptic problems. Instead of the algebraic inner iteration errors we now have to control the FE discretization errors, which is a standard task performed within any *adaptive* multilevel method. A careful study of the information gain per computational effort leads to the result that the quadratic convergence mode of the Newton–Galerkin algorithm is the best mode for the fixed dimensional case, whereas for an adaptive variable dimensional code a special *linear* convergence mode of the algorithm is definitely preferable. The theoretical results are then illustrated by numerical experiments with a NEWTON–KASKADE algorithm.

1 Introduction

The present paper deals with the multilevel solution of *elliptic* partial differential equations (PDEs), which come up in a variety of scientific and engineering applications. In a *finite element* setting (see e.g. the recent testbook of Braess [4]), such problems arise as *convex* minimization problems – a formulation, which directly reflects the fundamental variational principle underlying the associated physical problem. *Uniform ellipticity* of the underlying PDE shows up as *strict convexity* of the nonlinear functional in question. Among the multilevel methods, there are two basic lines: a) the nonlinear multigrid method or the full approximation scheme (FAS) (compare the classical textbooks [11, 13]), or the recent publication by Hackbusch and Reusken [12], where *nonlinear residuals* are evaluated within the multigrid cycles, and b) the Newton multigrid method (compare e.g. Bank and Rose [1]), where a linear multigrid method is applied for the computation of the Newton corrections so that *linear residuals* are evaluated within the multigrid cycles. In the second approach, the basic question of algorithmic interest is the *accuracy matching* between the outer (Newton) and the inner (linear) iteration. The present paper follows this second line, but in a way different from the known ones both in the underlying theoretical analysis and in the suggested algorithm. In order to concentrate the presentation, we only treat *local* Newton methods here and postpone the associated globalization to a forthcoming paper (cf. [9]).

In Section 2, the *ordinary* Newton method for convex optimization in Hilbert spaces is analyzed within an *affine conjugate* setting, which reflects the behavior of such problems under arbitrary affine transformation. This approach is a follow–up of two earlier affine invariant approaches: (I) the *affine covariant* approach (earlier called affine *invariant*) advocated by Deuflhard and Heindl [6] or, in the multilevel context of PDEs, by Deuflhard and Potra [8], and (II) the *affine contravariant* approach due to Hohmann [14], who had worked out an adaptive multilevel collocation method for general nonlinear ODE boundary value problems. An affine conjugate Newton–Mysovskikh theorem for the *exact* Newton method in Hilbert spaces is given, wherein the term "exact" indicates that we do not take any discretization errors into account. In Section 3.1 this local convergence theorem is extended to *inexact* Newton methods, where the inner iteration is only approximately solved. For fixed finite dimension, an affine conjugate convergence analysis and, on this basis, an implementation of a *Newton–PCG algorithm* are given so that *quadratic* convergence of the outer iteration is guaranteed. For infinite dimensional problems, any Galerkin method such as FEM for linear elliptic problems may stand for the inner iteration. A careful study of the information gain per computational effort (Section 3.2) shows that the quadratic convergence mode of the Newton–Galerkin algorithms agrees with the fixed finite dimensional case, whereas a special linear convergence mode of the algorithm nicely goes with *adaptive multilevel FEM*. Finally, in Section 4, the obtained theoretical results are illustrated by numerical experiments with a NEWTON–KASKADE code.

2 An affine conjugate Newton–Mysovskikh theorem

Consider the *minimization problem*

$$f(x) = \min \ ,$$

wherein $f : D \subset H \rightarrow \mathbb{R}$ is assumed to be a *strictly convex* C^2- functional defined in a convex neighborhood D of the minimum point x^*. Then finding the minimum point is equivalent to finding the solution point of the nonlinear elliptic problem

$$F(x) = \operatorname{grad} f(x) = f'(x)^* = 0 \ , \quad x \in D \ . \tag{2.1}$$

For such a *gradient* mapping F the Frechet–derivative $F'(x) = f''(x)$ is *selfadjoint*; for strictly convex f we even have that F' is a strictly *positive definite* operator so that $F'(x)^{1/2}$ can be defined. As for the involved Hilbert space H, let it be endowed with an inner product $\langle \cdot, \cdot \rangle$ that induces a norm $\| \cdot \|$. In addition, we will be interested in *local energy products* defined for each $x \in D$ to be symmetric bilinear forms of the kind $\langle \cdot, F'(x) \cdot \rangle$. Since $F'(x), x \in D$ is positive definite, these energy products also induce *local energy norms* of the kind $\|F'(x)^{1/2} \cdot \|$. The ordinary Newton method for the mapping F has the form

$$F'(x^k)\Delta x^k = -F(x^k) \quad x^{k+1} = x^k + \Delta x^k. \tag{2.2}$$

In *finite dimensional* problems, this is a symmetric positive definite linear system, which can be solved directly, as long as the size of the system is moderate; it will have to be solved iteratively, whenever the problem is sufficiently large scale, which then implies that inner iteration errors have to be regarded as well. In *infinite dimensional* problems, discretization errors must be additionally taken into account. The latter two cases will be treated in the subsequent Section 3, whereas here we first analyze the *exact* Newton method in the absence of any approximation errors.

Before starting, however, we want to study the associated affine invariance property. For that purpose, let B denote an arbitrary *bounded linear bijective operator* that transforms H onto some K. Then we arrive at the transformed convex minimization problem

$$g(y) = f(By) = \min\ , \quad x = By\ ,$$

the transformed gradient mapping

$$G(y) = B^T F(By) = 0\ ,$$

and the transformed Fréchet–derivative

$$G'(y) = B^T F'(x) B\ , \quad x = By\ .$$

The derivative transformation is *conjugate*, which motivates the name affine conjugacy for this special affine invariance class. It is clear that all $G'(\cdot)$ are selfadjoint positive definite operators. Newton's method applied to the transformed gradient mapping reads

$$G'(y^k)\Delta y^k = -G(y^k) \quad \Longleftrightarrow \quad B^T F'(x^k) B \Delta y^k = -B^T F(x^k)\ ,$$

which shows that $\Delta x^k = B\Delta y^k$, i.e. the iterates transform exactly as the whole domain space of the mapping F, once the initial guess is also transformed accordingly. It is therefore only natural to require affine conjugacy throughout any theoretical convergence analysis of the ordinary Newton method. Such convergence theorems should then only use theoretical quantities like iterative *functional values* $f(x^k)$ or *local energy products* of quantities in domain space such as iterative *corrections* Δx^k and *errors* $x^k - x^*$. As a first step, a Newton–Mysovskikh type theorem, which meets this strict affine conjugacy requirement, is given.

Theorem 2.1 *Let $f : D \to \mathbb{R}$ be a strictly convex C^2–functional to be minimized over some open and convex domain $D \subset H$. Let $F(x) = f'(x)^*$ and $F'(x) = f''(x)$, which is then selfadjoint positive definite. In the notation just introduced in this section, assume the following affine conjugate Lipschitz condition:*

$$\left\| F'(z)^{-1/2}\big(F'(y) - F'(x)\big)(y-x) \right\| \le \omega \|F'(x)^{1/2}(y-x)\|^2 \tag{2.3}$$

for collinear $x,\ y,\ z \in D$ with some $0 \le \omega < \infty$. For well-defined iterates $x^k \in D$, let $\epsilon_k := \|F'(x^k)^{1/2}\Delta x^k\|^2$ and $h_k := \omega\|F'(x^k)^{1/2}\Delta x^k\|$. For the initial guess x^0 assume that

$$h_0 := \omega\|F'(x^0)^{1/2}\Delta x^0\| < 2 \tag{2.4}$$

and that the level set $\mathcal{L}_0 := \{x \in D \mid f(x) \le f(x^0)\}$ is compact. Then the ordinary Newton iterates remain in $\mathcal{L}_0$ and converge to the minimum point x^* at a rate estimated in terms of local energy norms by

$$\|F'(x^{k+1})^{1/2}\Delta x^{k+1}\| \le \frac{\omega}{2}\|F'(x^k)^{1/2}\Delta x^k\|^2 \Longleftrightarrow h_{k+1} \le \frac{1}{2}h_k^2 \tag{2.5}$$

or in terms of the functional by

$$-\frac{1}{6}h_k\epsilon_k \le f(x^k) - f(x^{k+1}) - \frac{1}{2}\epsilon_k \le \frac{1}{6}h_k\epsilon_k. \tag{2.6}$$

The initial distance to the minimum can be bounded as

$$f(x^0) - f(x^*) \le \frac{5}{6}\frac{\epsilon_0}{1 - h_0/2}\ . \tag{2.7}$$

Proof. For the purpose of repeated induction, let $\mathcal{L}_k$ denote the level set defined in analogy to $\mathcal{L}_0$ and let $\mathcal{L}_k \subset D$. First, in order to show that $x^{k+1} \in \mathcal{L}_k$, we start from the identity ($\lambda \in [0,1]$)

$$\begin{aligned} & f(x^k + \lambda\Delta x^k) - f(x^k) + \left(\lambda - \frac{\lambda^2}{2}\right) \|F'(x^k)^{1/2}\Delta x^k\|^2 \\ &= \int_{s=0}^{\lambda} s \int_{t=0}^{1} \left\langle \Delta x^k, \big(F'(x^k + st\Delta x^k) - F'(x^k)\big)\Delta x^k \right\rangle dtds. \end{aligned} \tag{*}$$

By means of the Cauchy–Schwarz inequality and of the Lipschitz condition (2.3) with $x = z = x^k$, $y = x^k + st\Delta x^k$, the above inner product can be bounded as

$$\begin{aligned} &\langle \Delta x^k\ ,\ \ \cdot\ \rangle \le |\langle F'(x^k)^{1/2}\Delta x^k\ , F'(x^k)^{-1/2}\cdot\ \rangle| \\ &\le \|F'(x^k)^{1/2}\Delta x^k\| \cdot \omega st \|F'(x^k)^{1/2}\Delta x^k\|^2 = st h_k \epsilon_k \end{aligned}$$

For the purpose of repeated induction, let $h_k < 2$, which then implies that

$$f(x^k + \lambda\Delta x^k) \le f(x^k) + \left(\frac{\lambda^3}{3} + \frac{\lambda^2}{2} - \lambda\right)\epsilon_k \le f(x^k) - \frac{1}{6}\epsilon_k < f(x^k) \quad \text{for } \lambda \in [0,1]\ . \tag{2.8}$$

This is the left hand side of (2.6). Therefore, the assumption $x^k + \lambda\Delta x^k \notin \mathcal{L}_k$ would lead to a contradiction for all $\lambda \in [0,1]$. Hence, $x^{k+1} \in \mathcal{L}_k \subset D$. By repeated induction and D compact, the iterates are then seen to converge to x^*. (Note that x^* is anyway unique in D under the assumptions made.) Application of the Lipschitz condition (2.3) for $z = x^{k+1}$, $y = x^k + t\Delta x^k$, $x = x^k$, then confirms the quadratic convergence result (2.5), which requires the assumption (2.4) to assure that $h_{k+1} < h_k$.

In order to obtain the right hand side of (2.6), we revisit the identity (*) again for $\lambda = 1$, but this time apply the Cauchy–Schwarz inequality in the other direction to obtain:

$$0 \le f(x^k) - f(x^{k+1}) \le \left(\frac{1}{2} + \frac{1}{6}h_k\right)\|F'(x^k)^{1/2}\Delta x^k\|^2 < \frac{5}{6}\epsilon_k\ .$$

Summing over all $k = 0, 1, \ldots$ we get

$$0 \le \omega^2\big(f(x^0) - f(x^*)\big) \le \sum_{k=0}^{\infty}\left(\frac{1}{2}h_k^2 + \frac{1}{6}h_k^3\right) < \frac{5}{6}\sum_{k=0}^{\infty} h_k^2\ .$$

Upon inserting the successive bounds

$$\frac{1}{2}h_{k+1} \le \left(\frac{1}{2}h_k\right)^2 \le \frac{1}{2}h_k < 1,$$

the right hand upper bound can be further treated to yield

$$\omega^2\big(f(x^0) - f(x^*)\big) < \frac{5}{6}\frac{h_0^2}{1 - h_0/2}\ ,$$

which completes the proof. ∎

We now study any possible consequences of the above local convergence theorem for *actual implementation.* Let the computable quantity Θ_k be defined as

$$\Theta_k = \frac{h_{k+1}}{h_k} = \left(\frac{\epsilon_{k+1}}{\epsilon_k}\right)^{1/2} = \left(\frac{\langle F(x^{k+1}), \Delta x^{k+1}\rangle}{\langle F(x^k), \Delta x^k\rangle}\right)^{1/2}\ .$$

As the main consequence of the above convergence theorem we have the condition

$$\Theta_k \le \frac{1}{2}h_k < 1\ ,$$

which leads to the *monotonicity criterion*

$$\Theta_k < 1 \ .$$

In particular, whenever

$$\Theta_0 \geq 1$$

then x^0 is definitely *not* within the local convergence domain as stated by the theorem. From (2.6) we may alternatively terminate the iteration as *divergent* whenever

$$f(x^{k+1}) - f(x^k) > -\frac{\epsilon_k}{6} \ .$$

Convergence may be understood to occur whenever

$$\epsilon_k \leq \text{ETOL}$$

with ETOL a user prescribed *local energy error tolerance* or whenever

$$|f(x^{k+1}) - f(x^k)| \leq \text{ETOL}/2$$

recalling that asymptotically

$$f(x^{k+1}) - f(x^k) \doteq -\frac{1}{2}\epsilon_k \ .$$

3 Inexact Newton–Galerkin methods

We keep the notation and assumptions of the preceding section, but now study *inexact* Newton methods

$$F'(x^k)\,\delta x^k = -F(x^k) + r^k \ , \quad x^{k+1} = x^k + \delta x^k \ , \tag{3.1}$$

wherein *inexact Newton corrections* δx^k instead of *exact Newton corrections* Δx^k arise, since the above linear equation is only solved up to an *inner residual* r^k. Among the inner iterations we focus our attention on those, which satisfy the *Galerkin condition*

$$\langle \delta x^k, F'(x^k)(\delta x^k - \Delta x_k)\rangle = 0 \tag{3.2}$$

or, equivalently, the condition

$$\|F'(x^k)^{1/2}(\delta x^k - \Delta x^k)\|^2 + \|F'(x^k)^{1/2}\delta x^k\|^2 = \|F'(x^k)^{1/2}\Delta x^k\|^2. \tag{3.3}$$

Examples of such inner iterations are:

- for $H = \mathbb{R}^n$ and $\langle u, v\rangle = u^T v$ the Euclidean inner product any *preconditioned conjugate gradient* (PCG) method,
- for $H = H^1$ and $\langle u, v\rangle$ the L_2-product any *finite element* method (FEM).

In any case, *affine conjugacy* will play a central role both in our theoretical characterization (as opposed to the analysis in [15]) and in the algorithmic realization to be suggested.

3.1 Theoretical convergence results

First, we want to extend the preceding Theorem 2.1 in an *affine conjugate* way from the exact to the inexact Newton iteration. Special care will be taken in designing an appropriate theoretical *accuracy matching* between inner and outer iteration, which will then, second, be discussed and worked out as an algorithm in the subsequent Section 3.2.

Theorem 3.1 *Keep the notation* $h_k := \omega\|F'(x^k)^{1/2}\delta x^k\|$, $\epsilon_k := \|F'(x^k)^{1/2}\delta x^k\|^2$ *and the assumptions of the preceding Theorem 2.1, but slightly extend* (2.3) *to the more general affine conjugate Lipschitz condition:*

$$\left\|F'(z)^{-1/2}\big(F'(y)-F'(x)\big)v\right\| \leq \omega\left\|F'(x)^{1/2}(y-x)\right\| \cdot \|F'(x)^{1/2}v\|$$

for collinear x, y, $z \in D$ *and some* $0 \leq \omega \leq \infty$. *Consider an inexact Newton–Galerkin iteration* (3.1) *satisfying* (3.2). *At any well–defined iterate* x^k, *let*

$$\frac{\|F'(x^k)^{1/2}(\delta x^k - \Delta x^k)\|}{\|F'(x^k)^{1/2}\delta x^k\|} \leq \delta_k < 1 \ . \tag{3.4}$$

For a given initial guess $x^0 \in D$ *assume that the level set* $\mathcal{L}_0 := \{x \in D \mid f(x) \leq f(x^0)\}$ *is compact. Then the following results hold:*

I. If the initial iterate x^0 *satisfies*

$$h_0 < 2 \tag{3.5}$$

and if, for some prescribed $\overline{\Theta}$ *varying in the range* $\dfrac{h_0}{2} < \overline{\Theta} < 1$, *the inner iteration is controlled such that*

$$\delta_k \leq \frac{2\overline{\Theta} - h_k}{h_k + \sqrt{4+h_k^2}}, \tag{3.6}$$

then the inexact Newton iterates x^k *remain in* $\mathcal{L}_0$ *and converge at least linearly to the minimum point* $x^* \in \mathcal{L}_0$ *such that*

$$\|F'(x^{k+1})^{1/2}\delta x^{k+1}\| \leq \overline{\Theta}\|F'(x^k)^{1/2}\delta x^k\| \Longleftrightarrow h_{k+1} \leq \overline{\Theta}\, h_k \ . \tag{3.7}$$

II. If, for some $\rho > 0$, *the initial iterate* x^0 *satisfies*

$$h_0 < \frac{2}{1+\rho} \tag{3.8}$$

and the inner iteration is controlled such that

$$\delta_k \leq \frac{\rho h_k}{h_k + \sqrt{4+h_k^2}}, \tag{3.9}$$

then the inexact Newton iterates x^k *remain in* $\mathcal{L}_0$ *and converge quadratically to the minimum point* $x^* \in \mathcal{L}_0$ *such that*

$$\|F'(x^{k+1})^{1/2}\delta x^{k+1}\| \leq (1+\rho)\frac{\omega}{2}\|F'(x^k)^{1/2}\delta x\|^2 \Longleftrightarrow h_{k+1} \leq \frac{1+\rho}{2}h_k^2 \ . \tag{3.10}$$

III. The convergence in terms of the functional can be estimated by

$$-\frac{1}{6}h_k\epsilon_k \leq f(x^k) - f(x^{k+1}) - \frac{1}{2}\epsilon_k \leq \frac{1}{6}h_k\epsilon_k \ . \tag{3.11}$$

Proof We adopt the notation of the proof of the preceding Theorem 2.1. For the purpose of repeated induction, let $x^k \in \mathcal{L}_k \subset D$. As before, we start from the identity

$$\begin{aligned} & f(x^{k+1}) - f(x^k) + \frac{1}{2}\epsilon_k \\ &= \int_{s=0}^{1} s \int_{t=0}^{1} \left\langle \delta x^k \ , \big(F'(x^k + st\delta x^k) - F'(x^k)\big)\delta x^k \right\rangle dtds + \langle \delta x^k \ , \ \ r^k \rangle \ . \end{aligned}$$

The second term vanishes due to (3.2), since

$$\langle \delta x^k, r^k \rangle = \langle \delta x^k, F'(x^k)(\delta x^k - \Delta x^k)\rangle = 0 \ .$$

Upon treating the integrand within the first term by the Cauchy–Schwarz inequality and the Lipschitz condition with $z = x = x^k$, $y - x = st\delta x^k$ in both directions, the result (3.11) is confirmed – just as in the proof for the exact Newton method. This result, however, is not yet applicable for any convergence statement, since the behavior of the Kantorovitch type quantities h_k still needs to be studied. For this purpose, we start by observing that (3.3) implies

$$\|F'(x^{k+1})^{1/2}\delta x^{k+1}\| \leq \|F'(x^{k+1})^{1/2}\Delta x^{k+1}\|.$$

We therefore estimate the local energy norms as

$$\begin{aligned}\|F'(x^{k+1})^{1/2}\delta x^{k+1}\| &\leq \left\| F'(x^{k+1})^{-1/2}\left(\int\limits_{t=0}^{1}\big(F'(x^k+tdx^k)-F'(x^k)\big)\delta x^k dt + r^k\right)\right\| \\ &\leq \frac{\omega}{2}\big\|F'(x^k)^{1/2}\delta x^k\|^2 + \|F'(x^{k+1})^{-1/2}r^k\| \ .\end{aligned}$$

For further treatment of the second term, we define

$$z := F'(x^k)^{1/2}(\delta x^k - \Delta x^k) \ , \ \ w := F'(x^{k+1})^{-1/2}F'(x^k)^{1/2}z$$

and estimate

$$\begin{aligned}\|w\|^2 &= \big\langle z \ , \ \ F'(x^k)^{1/2}F'(x^{k+1})^{-1}F'(x^k)^{1/2}z\big\rangle \\ &\leq \|z\|^2 + \big|\big\langle w \ , \ \ F'(x^{k+1})^{-1/2}\big(F'(x^{k+1})-F'(x^k)\big)F'(x^k)^{-1/2}z\big\rangle\big| \\ &\leq \|z\|^2 + h_k\|w\|\cdot\|z\| \ .\end{aligned}$$

This quadratic inequality can be solved to yield

$$\|w\| \leq \frac{1}{2}\left(h_k + \sqrt{4+h_k^2}\right)\|z\| \ .$$

Collecting all estimates then confirms the *linear* convergence result

$$\Theta_k := \frac{\|F(x^{k+1})^{1/2}\delta x^{k+1}\|}{\|F'(x^k)^{1/2}\delta x^k\|} \leq \frac{1}{2}\left(h_k + \left(h_k + \sqrt{4+h_k^2}\right)\delta_k\right) \ . \tag{3.12}$$

If we assume (3.6), then the condition (3.5) is necessary to obtain

$$\Theta_k \leq \overline{\Theta} < 1,$$

which is just the result (3.7). Consequently, by an argument elaborated in detail in the proof of Theorem 2.1, we have $x^{k+1} \in \mathcal{L}_k \subset D$. With D compact and repeated induction using the upper bound from (3.11), we arrive at the statement that the iterates x^k remain in $\mathcal{L}_0$ and converge to x^*.

As for *quadratic* convergence, we just impose in (3.12) that the second right hand term, which represents the perturbation by the inner iteration, should be somehow matched with the first term, which represents the quadratic convergence pattern of the exact Newton iteration. This is condition (3.9), which is constructed such that the convergence relations (3.10) are obtained. Finally, in order to assure that $h_{k+1} < h_k$, we need the initial assumption (3.8), which completes the proof. ■

In passing, we note that the above result (3.11) permits the same estimate of the functional distance to the minimum $f(x^*)$ as in Theorem 2.1.

Remark. Without proof we state that similar, but less favorable results can be obtained, when the inner iteration is *not* required to satisfy a Galerkin condition. In particular, equation (3.12) must then be replaced by

$$\Theta_k \leq \frac{1}{2(1-\delta_{k+1})}\left(h_k + \left(h_k + \sqrt{4+h_k^2}\right)\delta_k\right) \ ,$$

which additionally imposes a condition such as

$$\delta_{k+1} \leq \delta_k < \tfrac{1}{2}$$

to allow for $\Theta_k < 1$. Such conditions will play a role e.g. when *Gauss–Seidel* or *Gauss-Jacobi* iterations are selected as inner iterations.

3.2 Accuracy matching

We now want to exploit the above local convergence theorem for actual computation. For the *outer* iteration, the algorithmic techniques are essentially the same as in the *exact* Newton method – see the end of Section 2. As a slight generalization of the situation of Theorem 3.1 we set certain default parameters $\overline{\Theta}_k < 1$ and require the *inexact monotonicity criterion*

$$\Theta_k = \left(\frac{\epsilon_{k+1}}{\epsilon_k}\right)^{1/2} = \left(\frac{\langle F(x^{k+1})\ ,\ \ \delta x^{k+1}\rangle}{\langle F(x^k)\ ,\ \ \delta x^k\rangle}\right)^{1/2} \leq \overline{\Theta}_k < 1\ . \tag{3.13}$$

We will regard the outer iteration as *divergent*, whenever $\Theta_k > \overline{\Theta}_k$ holds. Note that Θ_k monotonely increases in the course of the inner iteration for any Galerkin type method such as the Newton-PCG or the Newton-FEM. It is therefore sufficient to concentrate here on the question of *how to match inner and outer iterations*.

In view of actual computation, a key property of any inexact Newton–Galerkin method shows up in the estimate (3.11), which does *not* contain any pollution effect from the inner iteration, but only depends on the unknown Kantorovitch quantities h_k. That is why this relation is perfectly suited for the construction of a cheap *computational estimate*

$$[h_k] := \frac{6}{\epsilon_k}|f(x^{k+1}) - f(x^k) + \tfrac{1}{2}\epsilon_k| \leq h_k, \tag{3.14}$$

which is a guaranteed lower bound of the associated quantity within the brackets $[\cdot]$. (Note that this nice feature strictly depends on the Galerkin property of the inner iteration – as can be seen from the proof above.) For $k = 0$ and given x^0, we cannot but choose any (sufficiently small) $\delta_0 < \overline{\Theta}_0$ ad hoc and run the inner iteration until the threshold condition (3.4) is passed. Then we may evaluate $[h_0]$ from (3.14). Whenever $[h_0] \geq 2\overline{\Theta}_0$, we terminate the iteration as *divergent* in accordance with (3.5) and (3.8). Otherwise, we continue with either the linear or the quadratic convergence mode indicated in Theorem 3.1.

Linear convergence mode. For $k \geq 0$, we now assume that $[h_k] < 2\overline{\Theta}_k < 2$. As for the termination of the inner iteration, we would like to assure the condition

$$\delta_k \leq \frac{2\overline{\Theta}_k - h_k}{h_k + \sqrt{4 + h_k^2}}$$

The main feature of this upper bound is that $\delta_k \to \overline{\Theta}_k \to 1$ is permitted when $k \to \infty$. In words: *the closer the iterates come to the solution point, the less work is necessary within the inner iteration to assure linear convergence of the outer iteration.* Since the above upper bound is unavailable, we will replace it by the computationally available estimate

$$[\delta_k] := \frac{2\overline{\Theta}_k - [h_k]}{[h_k] + \sqrt{4 + [h_k]^2}} \geq \frac{2\overline{\Theta}_k - h_k}{h_k + \sqrt{4 + h_k^2}}. \tag{3.15}$$

Obviously, this estimate may be "too large", since the above right hand side is a monotone *decrasing* function of h_k and $[h_k] \leq h_k$. Fortunately, as described above, the difference between computational estimate and theoretical quantity can be ignored asymptotically. For $k = 0$, we evaluate (3.15) with $[h_0]$ inserted from (3.14). For $k > 0$, we suggest to evaluate (3.15) with $[h_k] := \Theta_{k-1}[h_{k-1}]$ and $[h_{k-1}]$ from (3.14). In any case, we require the monotonicity (3.13) and run the inner iteration at each step k until either the actual value of δ_k obtained in the course of the inner iteration – see (3.4) – is less than the associated estimate above or divergence occurs with $\Theta_k > \overline{\Theta}_k$. Ideas for a choice of the parameters $\overline{\Theta}_k$ may come from the context of the problem to be solved. If nothing specific is known, a common value will be set throughout. On the other hand, with these parameters at our disposal, not only linear convergence is covered by this algorithmic mode – so that the clumsy name *at least linear* convergence mode would be appropriate.

Quadratic convergence mode. For $k \geq 0$, we again assume that $[h_k] < 2$. As for the termination of the inner iteration, we now want to obey the condition

$$\delta_k \leq \rho \frac{h_k}{h_k + \sqrt{4 + h_k^2}} \ , \quad \rho > 0 \ .$$

In contrast to the linear convergence mode, the main feature of the upper bound here is that $\delta_k \to 0$ is forced when $k \to \infty$. In words: *the closer the iterates come to the solution point, the more work needs to be done in the inner iteration to assure quadratic convergence of the outer iteration.* As before, we will replace the unavailable upper bound by the computationally available estimate

$$[\delta_k] := \frac{\rho\,[h_k]}{[h_k] + \sqrt{4 + [h_k]^2}}. \tag{3.16}$$

in terms of computational estimates $[h_k]$. Since the above right hand side is a monotone *increasing* function of $[h_k]$, the relation $[h_k] \leq h_k$ here implies that

$$[\delta_k] \leq \rho \frac{h_k}{h_k + \sqrt{4 + h_k^2}} \ .$$

This means that we are really able to *assure* the above theoretical condition by our computational strategy! As for the computational estimates to be inserted above, we may well use formula (3.14) based on functional evaluations. In this mode, however, we have a further simple possibility to construct cheap computational estimates. Recalling (3.10), we may also define

$$[h_k] := \frac{2\Theta_k}{1+\rho} \leq h_k. \tag{3.17}$$

If we compute both estimates, then the *maximum* of the two lower bounds will be preferable. For $k > 0$, the "more advanced" estimate $[h_k] := \Theta_{k-1}[h_{k-1}]$ will be preferably inserted into formula (3.16). For $k = 0$, the two estimates $[h_0]$ may be formally identified to determine the matching parameter ρ via

$$1 + \rho = \frac{\Theta_0 \epsilon_0}{3|f(x^1) - f(x^0) + \frac{1}{2}\epsilon_0|}$$

Whenever $[h_0]$ is "too close" to 2, then the matching factor ρ will be "too small", which involves "too much" work in the inner iteration. In this context we may note that reversing formula (3.17) leads to

$$\Theta_k = \frac{1+\rho}{2}[h_k].$$

If we set the default parameters $\overline{\Theta}_k$ in the linear convergence mode such that the same relation as above holds for the next iterative step (inserting the "more advanced" estimate), then we would end up with

$$\overline{\Theta}_{k+1} = \frac{1+\rho}{2}[h_{k+1}] = \Theta_k^2.$$

With this (slightly strict) specification and exclusive use of the functional based estimates (3.14) we are able to realize the quadratic convergence mode within the framework of the "at least linear" convergence mode without considering the safety factor ρ at all. Still, the problem with "too small" ρ-values is hidden implicitly. That is why the relaxed specification

$$\overline{\Theta}_{k+1} = 2\Theta_k^2, \quad \overline{\Theta}_0 \approx \tfrac{1}{2} \tag{3.18}$$

is recommended instead. Remember, however, that the choice of $\overline{\Theta}_0$ (just as the choice of ρ) governs the size of the accepted local convergence domain – as can be seen in Theorem 3.1, inequality (3.8).

Computational complexity. After the above derivation of the inexact Newton–Galerkin method in the linear and the quadratic mode, we are left with the question of when to use which of the two modes. In a first step, we might just look at the iterative contraction factors Θ_k and the associated *information gain* at iteration step k:

$$I_k = \log \frac{1}{\Theta_k} = |\log \Theta_k|$$

Maximization of I_k will directly lead us to the quadratic convergence mode only. However, as will be worked out now, the *amount of computational work* A_k involved to to realize the above information gain must be taken into account as well. We therefore will rather have to look at the *information gain per unit work* and solve the problem

$$i_k = \frac{I_k}{A_k} = \frac{1}{A_k} \log \frac{1}{\Theta_k} = \max \ . \tag{3.19}$$

In what follows, we will give computational complexity models for two typical Newton–Galerkin methods to exmplify the procedure how to design an efficient mode of the algorithm.

Finite dimensional problems: Newton – PCG. Consider a nonlinear elliptic problem with *fixed* finite dimension such as a discretized nonlinear elliptic PDE, which has been treated by a grid generator before starting the solution process. Hence, we have the "simple" case $H = \mathbb{R}^N$, which might not be too simple, in fact, when the fixed dimension N is large. For the inner iteration we assume that some *preconditioned conjugate gradient* (PCG) method has been selected without further specification of the preconditioner. It may be worth noting that this Newton–PCG may also be interpreted as some *nonlinear* cg method (compare Glowinski [10]) with *rare Jacobian updates*. At iteration step k, the computational work involved is

- evaluation of the Jacobian matrix $F'(x^k)$, which is typically sparse in the case of a discretized PDE so that a computational amount $\sim N$ needs to be counted
- work per PCG iteration: $O(N)$,
  number of PCG iterations $\sim \sqrt{\kappa} \log \frac{1}{\delta_k}$
  with κ the condition number of the preconditioned matrix.

Summing up, the total amount of work can be roughly estimated as

$$A_k \sim \left(c_1 + c_2 \log \frac{1}{\delta_k} \right) N \sim \text{const} + \log \frac{1}{\delta_k}.$$

From this, we obtain the information gain per unit work as

$$i_k \sim \frac{\left| \log \left(\frac{1}{2} \left(h_k + \delta_k (h_k + \sqrt{4 + h_k^2}\ ) \right) \right) \right|}{\text{const} + |\log \delta_k|} \ .$$

Asymptotically, for $h_k \to 0$, we obtain the simplification

$$i_k \sim \frac{|\log(\delta_k)|}{\text{const} + |\log \delta_k|} \ .$$

Since const > 0, the right hand side is a monotone *decreasing* function of δ_k, which implies that the minimum among any reasonable δ_k will maximize i_k. Therefore we are clearly led to the *quadratic convergence mode* of the algorithm, when the iterates are sufficiently close to the solution point – with the tacit hope that the asymptotic analysis carries over to the whole local convergence domain (a hope, which depends on the above constant).

Remark. It may be worth noting that the above analysis would lead to the same decision, if PCG were replaced by some linear multigrid method.

Infinite dimensional problems: adaptive Newton–FEM. Consider a nonlinear elliptic PDE problem in *finite element* formulation, which is a strictly convex minimization problem over some infinite dimensional space H. For the inner iteration we assume an *adaptive* multilevel FEM such as KASKADE [2, 3, 7]. Let d denote the underlying spatial dimension. At iteration step k on discretization level j let N_k^j be the number of nodes and ϵ_k^j the local energy. With $l = l_k$ we mean that discretization level, at which we achieve the prescribed tolerance δ_k. The important difference to the fixed dimension case now is that within an *adaptive multilevel* method the approximating dimension of the problem depends on the required accuracy. In the linear elliptic case we have the rough relation for the relative discretization error (on energy equilibrated meshes)

$$\left(\frac{N_k^0}{N_k^l}\right)^{\dfrac{2}{d}} \sim \frac{\epsilon_k^\infty - \epsilon_k^i}{\epsilon_k^\infty} \leq \delta_k^2$$

With a suitable preconditioner such as BPX [16, 5] the number of PCG iterations *inside* KASKADE is essentially independent of the number of nodes. Therefore the amount of work involved within one linear FEM call can be estimated as

$$A_k \sim N_k^i \sim \frac{N_k^0}{\delta_k^d}.$$

From this we end up with the following estimate for the information gain per unit work

$$i_k \sim \delta_k^d \left|\log\left(\tfrac{1}{2}\left(h_k + \delta_k(h_k + \sqrt{4 + h_k^2})\right)\right)\right| \; .$$

In the asymptotic case $h_k \to 0$ we arrive at

$$i_k \sim \delta_k^d \log \frac{1}{\delta_k} \; .$$

This simple function has its maximum at

$$\delta_{\text{opt}} = e^{-1/d}.$$

Even if we have to take the above rough complexity model *cum grano salis*, the message of this analysis is nevertheless clear: in the *adaptive* multilevel FEM, the *linear convergence mode* of the associated Newton–FEM is preferable. As a rough orientation, the above model suggests to set the default parameters $\overline{\Theta}_k = \overline{\Theta}^d$ with

$$\left(\overline{\Theta}^1, \overline{\Theta}^2, \overline{\Theta}^3\right) > \left(e^{-1}, e^{-1/2}, e^{-1/3}\right) \approx (0.37, 0.61, 0.72). \tag{3.20}$$

On the other hand, the adaptive refinement process anyway leads to natural values of the contraction factors Θ_k depending on the order of the applied finite elements: for linear FE we expect $\Theta_k \approx 0.5$. Upon taking possible variations into account that may come from the discretization error estimates, we have chosen the standard value $\overline{\Theta} = 0.7$ throughout the numerical experiments presented in the next section.

4 Numerical experiments with NEWTON–KASKADE

In this section, we want to demonstrate properties of the above derived adaptive Newton–multilevel FEM. For the inner iteration we pick the linear elliptic FEM code KASKADE [2] with linear finite elements specified throughout. Of course, any other *adaptive* linear multigrid method could equally well be used. As an illustrative example, we select

$$f(u) = \int\limits_\Omega \left(1 + |\nabla u|^2\right)^p - gu \, dx \; , \;\; p \geq \frac{1}{2} \; , \;\; x \in \Omega \subset \mathbb{R}^d \; .$$

This gives rise to the weak formulations

$$\Big\langle F(u), v \Big\rangle = \int_0^1 \big(2p(1+|\nabla u|^2)^{p-1}\langle \nabla u, \nabla v\rangle - gv\big)dx\ ,$$

$$\Big\langle w, F'(u)v \Big\rangle = \int_0^1 2p\big(2(p-1)(1+|\nabla u|^2)^{p-2}\langle \nabla w, \nabla u\rangle\langle \nabla u, \nabla v\rangle + (1+|\nabla u|^2)^{p-1}\langle \nabla w, \nabla v\rangle\big)dx$$

Obviously, with $\langle\cdot,\cdot\rangle$ the Euclidean inner product in $\mathbb{R}^d$, the term $\Big\langle v, F'(u)v \Big\rangle$ is strictly positive so that our theoretical assumptions hold. As for the computational cost, the evaluation of f is cheap, the evaluation of F is expensive and comparable to the evaluation of F'.

Linear versus quadratic convergence mode (1D). We set $\Omega = [0,1]$, $p = 2$, $g \equiv 16$ and Dirichlet boundary conditions. For a starting guess u^0 we choose the piecewise linear finite element function on three nodes with $u^0(0) = 0$, $u^0(0.5) = 0.5$ and $u^0(1) = 0$. In Table 1, computational results for the *quadratic* convergence mode are given, which nicely show the quadratic convergence performance of our computational strategy (3.18). Associated graphical information is represented in Fig. 1. In Table 2, comparative results for the *linear* convergence mode are

k	$\|F'(u^k)^{1/2}\delta u^k\|$	Θ_k	# nodes
0	1.23491	0.394465	5
1	0.487128	0.264594	33
2	0.128891	0.0791279	513
3	0.0101989	$\leq$ 0.00626	32769
4	$\leq$ 6.38e-05		$>$ 131073

Table 1: Quadratic convergence history ($d = 1$). To be compared with Table 2.

k	$\|F'(u^k)^{1/2}\delta u^k\|$	Θ_k	# nodes
0	1.23491	0.367242	5
1	0.453511	0.43875	9
2	0.198978	0.452755	17
3	0.0900881	0.491008	33
4	0.044234	0.499634	65
5	0.0221008	0.499936	129

Table 2: Linear convergence history ($d = 1$). To be compared with Table 1.

k	$\|F'(u^k)^{1/2}\delta u^k\|$	Θ_k	# nodes
0	0.440166	0.327883	205
1	0.144323	0.401937	684
2	0.0580088	0.506047	2242
3	0.0293552	0.467398	9106
4	0.0137205	$\leq$ 0.60	35716

Table 3: Linear convergence history ($d = 2$).

listed. The better efficiency compared with the quadratic mode is striking – in agreement with our computational complexity analysis (Section 3.2). Graphical information is represented in Fig. 2. The comparison with Fig. 1 shows that the iterates u^2 and u^3 respectively are quite close, whereas the corrections δu^2 and δu^3 (on a *smaller scale* than u^2, u^3, of course) differ visibly. In both modes of NEWTON–KASKADE nearly uniform grids appeared, reflecting the overall smoothness of the solution.

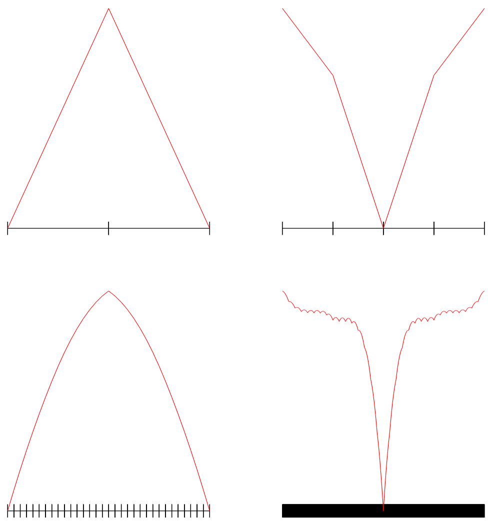

Figure 1: u^0, δu^0 and u^2, δu^2 obtained in the quadratic convergence mode ($d = 1$). To be compared with Fig. 2. (δu^2–scale reduction $\sim 1 : 25$)

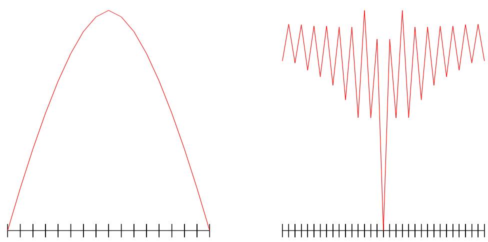

Figure 2: u^3, δu^3 obtained in the linear convergence mode ($d = 1$). To be compared with Fig. 1. (δu^3–scale reduction $\sim 1 : 100$)

Linear convergence mode (2D). In order to give some more insight into NEWTON–KASKADE, we also solved the above test problem for a not too simple $\Omega \subset \mathbb{R}^2$ with $p = 0.7$, $g \equiv 0$, and a mixture of Dirichlet and Neumann boundary conditions. In Fig. 3 (top) the coarse mesh obtained after some initialization step is documented together with u^0. Further iterates in Fig. 3 and the convergence history in Table 3 are given to illustrate the performance of NEWTON–KASKADE in the (standard) linear convergence mode.

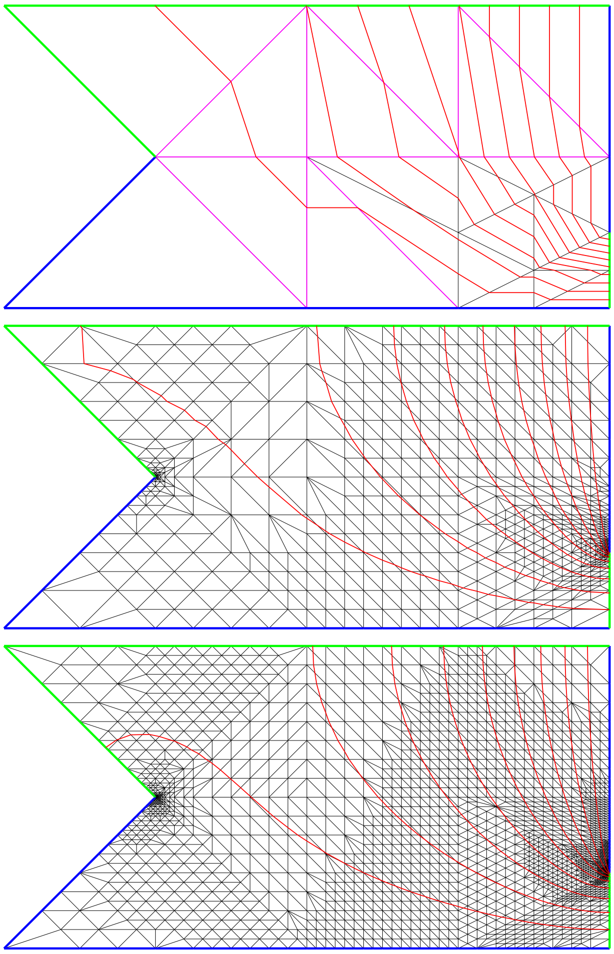

Figure 3: u^0, u^2, u^3 obtained in the linear convergence mode ($d = 2$). Dirichlet boundary conditions: black lines, Neumann boundary conditions: grey lines. Refinement levels $j = 1$, 10, 14.

Conclusion

In this paper, local inexact Newton–Galerkin algorithms have been designed on an affine conjugate theoretical basis. The algorithms permit a firm grip on both the inner and outer iterations. Within an adaptive multilevel setting, the computational costs are dominated by the last iterate, which requires the finest grid — a feature of this algorithm shared with adaptive nested nonlinear multigrid methods.

Acknowledgement. The authors gratefully acknowledge the helpful support by R. Beck in the extension from KASKADE to a first working version of NEWTON–KASKADE.

References

[1] R.E. Bank, D.J. Rose: *Analysis of a multilevel iterative method for nonlinear finite element equations*. Math. Comput. **39**, pp. 453–465 (1982).

[2] R. Beck, B. Erdmann, R. Roitzsch: *KASKADE 3.0 An Object–Oriented Adaptive Finite Element Code*. Technical Report TR 95–4, Konrad–Zuse–Zentrum für Informationstechnik Berlin (1995).

[3] F.A. Bornemann, B. Erdmann, R. Kornhuber: *Adaptive Multilevel–Methods in Three Space Dimensions*. Int. J. Numer. Methods Eng. **36** pp. 3187–3203 (1993).

[4] D. Braess: *Finite Elemente*. Springer–Verlag: Berlin Heidelberg, 302 pages (1992).

[5] J.H. Bramble, J.E. Pasciak, J. Xu: *Parallel Multilevel Preconditioners*. Math. Comp. **55**, pp. 1–22 (1990).

[6] P. Deuflhard, G. Heindl: *Affine Invariant Convergence Theorems for Newton's Method and Extensions to Related Methods*. SIAM J. Numer. Anal. **16**, pp. 1–10 (1979).

[7] P. Deuflhard, P. Leinen, H. Yserentant: *Concepts of an Adaptive Hierarchical Finite Element Code*. IMPACT **1**, pp. 3–35 (1989).

[8] P. Deuflhard, F. A. Potra: *Asymptotic Mesh Independence of Newton–Galerkin Methods Via a Refined Mysovskii Theorem*. SIAM J. Numer. Anal. **29**, pp. 1395–1412 (1992).

[9] P. Deuflhard, M. Weiser: *Global Inexact Newton Multilevel FEM for Nonlinear Elliptic Problems*. In Preparation (1996).

[10] R. Glowinski: *Lectures on Numerical Methods for Nonlinear Variational Problems*. Tata Institute of Fundamental Research, Bombay, Springer (1980).

[11] W. Hackbusch: *Multigrid methods and applications*. Springer–Verlag: Berlin, Heidelberg, New York, Tokyo (1995).

[12] W. Hackbusch, A. Reusken: *Analysis of a Damped Nonlinear Multilevel Method*. Numer. Math. **55**, pp. 225–246 (1989).

[13] W. Hackbusch, U. Trottenberg (eds.): *Multigrid methods*. Proceedings, Köln–Porz, 1981. Lect. Notes Math. Vol. **960**. Springer–Verlag: Berlin, Heidelberg, New York, Tokyo (1982).

[14] A. Hohmann: *Inexact Gauss Newton Methods for Parameter Dependent Nonlinear Problems*. Freie Universität Berlin, Institut für Mathematik: Dissertation (May 1994).

[15] R. Rannacher: *On the Convergence of the Newton–Raphson Method for Strongly Nonlinear Finite Element Equations*. In: P. Wriggers, W. Wagner (eds.), Nonlinear Computational Mechanics — State of the Art, Springer (1991).

[16] J. Xu: *Theory of Multilevel Methods*. Report No. AM 48, Department of Mathematics, Pennsylvania State University (1989).