SIAM J. NUMER. ANAL.
Vol. 42, No. 5, pp. 1830–1845

ASYMPTOTIC MESH INDEPENDENCE OF NEWTON'S METHOD REVISITED*

MARTIN WEISER†, ANTON SCHIELA‡, AND PETER DEUFLHARD§

Abstract. The paper presents a new affine invariant theory on asymptotic mesh independence of Newton's method for discretized nonlinear operator equations. Compared to earlier attempts, the new approach is both much simpler and more intuitive from the algorithmic point of view. The theory is exemplified at finite element methods for elliptic PDE problems.

Key words. asymptotic mesh independence, Newton's method, affine invariance

AMS subject classifications. 65J15, 65L10, 65N30

DOI. 10.1137/S0036142903434047

Introduction. The term "mesh independence" characterizes the observation that finite dimensional Newton methods, when applied to a nonlinear PDE on successively finer discretizations with comparable initial guesses, show roughly the same convergence behavior on all sufficiently fine discretizations. The "mesh independence principle" has been stated and even exploited for mesh design in papers by Allgower and Böhmer [1] and McCormick [19]. Further theoretical investigations of the phenomenon have been given in [2] by Allgower, Böhmer, Potra, and Rheinboldt. Those papers, however, lacked certain important features in the theoretical characterization that made their application to discretized PDEs difficult. This drawback has been avoided in the affine invariant theoretical study by Deuflhard and Potra in [8]; from that analysis, the modified term "asymptotic mesh independence" naturally emerged. The present paper suggests a different approach, which is also affine invariant but much simpler and more natural from the algorithmic point of view.

In a number of papers subsequent to [2], mesh independence principles for different problem settings or different algorithms were established; we mention generalized equations [11, 3], SQP methods [20, 21], shape design [18], constrained Gauss–Newton methods [15], Newton-like methods [16], and gradient projection [17].

The paper is organized as follows. In section 1 we first revisit the theoretical approaches given up to now to treat mesh independence for operator equations. In section 2 we compare discrete versus continuous Newton methods, again in affine invariant terms; in contrast to the earlier treatment in [8], we use only terminology that naturally arises from the algorithmic point of view, such as Newton sequences and approximation errors. The new theory is then exemplified at finite element methods (FEM) for elliptic PDEs (section 3).

1. Preliminary considerations.

Let a nonlinear operator equation be denoted by

$$F(x) = 0,$$

*Received by the editors August 29, 2003; accepted for publication (in revised form) April 30, 2004; published electronically January 20, 2005. This research was supported by the DFG Research Center "Mathematics for key technologies" in Berlin. A preprint of this paper appeared as ZIB-Report 03-13, Zuse Institute Berlin, Berlin, Germany, 2003.
http://www.siam.org/journals/sinum/42-5/43404.html
†Corresponding author. Zuse Institute Berlin (ZIB), 14195 Berlin, Germany (weiser@zib.de).
‡Zuse Institute Berlin (ZIB), 14195 Berlin, Germany (schiela@zib.de).
§Free University of Berlin (FU), 14195 Berlin, Germany (deuflhard@zib.de).

where $F : D \to Y$ is defined on a convex domain $D \subset X$ of a Banach space X with values in a Banach space Y. Throughout the paper we assume the existence of a unique solution x^* of this operator equation. The corresponding ordinary Newton method in Banach space may be written as

$$F'(x^k)\Delta x^k = -F(x^k), \quad x^{k+1} = x^k + \Delta x^k, \quad k = 0, 1, \ldots, \tag{1.1}$$

assuming, of course, that the derivatives are invertible. In each Newton step, the linearized operator equation must be solved, which is why this approach is often also called *quasilinearization*. For F, we assume that Theorem 1 from [7] holds, an affine invariant version of the classical Newton–Mysovskikh theorem, whose essence we recall here for the purpose of later reference.

THEOREM 1.1. *Let $F : D \to Y$ be a continuously differentiable mapping with $D \subset X$ convex. Let $\|\cdot\|$ denote the norm in the domain space X. Suppose that $F'(x)$ is invertible for each $x \in D$. Assume that, for collinear x, y, $z \in D$, the following affine invariant Lipschitz condition holds:*

$$\big\|F'(z)^{-1}\big(F'(y) - F'(x)\big)v\big\| \le \omega\|y - x\|\,\|v\|. \tag{1.2}$$

For the initial guess $x^0 \in D$ assume that

$$h_0 = \omega\|\Delta x^0\| < 2$$

and that $\bar{S}(x^0, \rho) \subset D$ for $\rho = \frac{\|\Delta x^0\|}{1-h_0/2}$.

Then the sequence $\{x^k\}$ of ordinary Newton iterates remains in $S(x^0, \rho)$ and converges to a unique solution $x^* \in \bar{S}(x^0, \rho)$. Its convergence speed can be estimated as

$$\|x^{k+1} - x^k\| \le \tfrac{1}{2}\omega\|x^k - x^{k-1}\|^2.$$

In actual computation, we can solve only discretized nonlinear equations of finite dimension, at best on a sequence of successively finer mesh levels, say,

$$F_j(x_j) = 0, \quad j = 0, 1, \ldots\,, \tag{1.3}$$

where $F_j : D_j \to Y_j$ denotes a nonlinear mapping defined on a convex domain $D_j \subset X_j$ of a finite dimensional subspace $X_j \subset X$ with values in a finite dimensional space Y_j. We assume F_j results from a Petrov–Galerkin discretization, such that $F_j(x_j) = r_j F(x_j)$ with some linear restriction $r_j : Y \to Y_j$. The corresponding finite dimensional ordinary Newton method reads

$$F_j'(x_j^k)\Delta x_j^k = -F_j(x_j^k), \quad x_j^{k+1} = x_j^k + \Delta x_j^k, \quad k = 0, 1, \ldots.$$

In each Newton step, a system of linear equations must be solved. Since $(F_j)' = r_j F'$, this system can equally well be interpreted either as a discretization of the linearized operator equation (1.1) or as a linearization of the discrete nonlinear system (1.3). Again we assume that Theorem 1.1 holds, this time for the finite dimensional mapping F_j. Let ω_j denote the corresponding affine invariant Lipschitz constant. Then the quadratic convergence of this Newton method is governed by the relation

$$\|x_j^{k+1} - x_j^k\| \le \tfrac{1}{2}\omega_j\|x_j^k - x_j^{k-1}\|^2.$$

Under the assumptions of Theorem 1.1 there exist unique discrete solutions x_j^* on each level j. Of course, we want to choose appropriate discretization schemes such that

$$\lim_{j\to\infty} x_j^* = x^*. \tag{1.4}$$

From the synopsis of the discrete and the continuous Newton method, we immediately see that any comparison of the convergence behavior on different discretization levels j will direct us toward a comparison of the affine covariant Lipschitz constants ω_j. Of particular interest is the connection with the Lipschitz constant ω of the underlying operator equation.

In the earlier papers [1, 2] on mesh independence two assumptions of the kind

$$\|F_j'(x_j)^{-1}\| \le \beta_j, \quad \|F_j'(x_j+v_j) - F_j'(x_j)\| \le \gamma_j \|v_j\|$$

have been made in combination with the *uniformity* requirements

$$\beta_j \le \beta, \quad \gamma_j \le \gamma. \tag{1.5}$$

Obviously, these assumptions lack affine invariance. More important, however, and as a consequence of the noninvariance, these conditions are phrased in terms of *operator norms*, which, in turn, depend on the relation of norms in the domain and the image space of the mappings F_j and F, respectively. For typical PDEs and typical choices of norms we would obtain

$$\lim_{j\to\infty} \beta_j \to \infty,$$

which clearly contradicts the uniformity assumption (1.5). Consequently, an analysis in terms of β_j and γ_j would not be applicable to this important case.

The situation is different with the affine invariant Lipschitz constants ω_j: They depend only on the choice of *norms in the domain space*. It is easy to verify that

$$\omega_j \le \beta_j \gamma_j.$$

In section 2 below we will show that the ω_j remain bounded in the limit $j \to \infty$, as long as ω is bounded—even if either β_j or γ_j blow up. Moreover, even when the product $\beta_j\gamma_j$ remains bounded, the Lipschitz constant ω_j may be considerably lower, i.e.,

$$\omega_j \ll \beta_j \gamma_j.$$

A prerequisite for the asymptotic property (1.4) to hold is that the elements of the infinite dimensional space X can be well approximated by elements of the finite dimensional subspaces X_j. In general, however, the solution x^* has "better smoothness properties" than the generic elements of the space X. For this reason, the earlier papers [2, 8] had restricted their analysis to some smoother subset $W^* \subset X$ and explicitly assumed that

$$x^*, x^k, \Delta x^k, x^k - x^* \in W^*, \quad k = 0, 1, \ldots.$$

However, such an assumption is hard to confirm in the concrete case. That is why we will drop it for our analysis to be presented.

Next, we revisit the paper [8] in some necessary detail. In that paper a family of linear projections

$$\pi_j : X \to X_j, \quad j = 0, 1, \ldots,$$

was introduced, assumed to satisfy the *stability condition*

$$q_j = \sup_{x \in W^*, x \neq 0} \frac{\|\pi_j x\|}{\|x\|} \leq \overline{q} < \infty, \quad j = 0, 1, \ldots. \tag{1.6}$$

The projection property $\pi_j^2 = \pi_j$ immediately gives rise to the lower bound

$$q_j \geq 1. \tag{1.7}$$

As a measure of the *approximation quality* that paper defined

$$\delta_j = \sup_{x \in W^*, x \neq 0} \frac{\|x - \pi_j x\|}{\|x\|}, \quad j = 0, 1, \ldots. \tag{1.8}$$

The rather natural idea that a refinement of the discretization improves the approximation quality was expressed by the asymptotic assumption

$$\lim_{j \to \infty} \delta_j = 0. \tag{1.9}$$

The triangle inequality and (1.6) supplied the upper bound

$$q_j \leq 1 + \delta_j. \tag{1.10}$$

By combination of (1.7), (1.9), and (1.10), asymptotic stability arose as

$$\lim_{j \to \infty} q_j = 1. \tag{1.11}$$

However, as has been pointed out by Braess [6], the above theory has some weak points. In fact, from (1.6) we conclude that $x = 0$ implies $\pi_j x = 0$. The reverse, however, will not be true in general. Hence, one must be aware of pathological elements $x \neq 0$ with corresponding approximations $\pi_j x = 0$. On a uniform one-dimensional grid, such a pathological element might look just like $x(t)$ represented graphically in Figure 1.1. Insertion of such elements into (1.8) would yield

$$\delta_j \geq 1$$

on each level j, on which such pathological elements exist. If one were to accept such an occurrence on *all* levels, then this would be in clear contradiction to the desired asymptotic property (1.9) and its consequence (1.11).

In order to close this gap of that theory, one would have to relate the subset W^* and the projections π_j such that the occurrence of pathological elements would

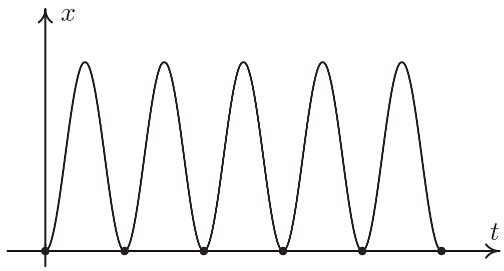

FIG. 1.1. *Pathological element $x \neq 0$ with $\pi_j x = 0$ ($\bullet$: mesh nodes).*

be asymptotically excluded. As an example, assume we have *nested* subspaces X_j, e.g., constructed by uniform mesh refinement. Suppose we begin with a "sufficiently good" initial projection π_0 on a "sufficiently" fine mesh, which already captures the main qualitative behavior of the solution x^* correctly. Then "pathological" elements would no longer be expected to occur on finer meshes in actual computation. Thus, upon carefully choosing appropriate subsets of W^*, the theory from [8] could, in principle, be repaired. However, the technicalities of such a theory tend to obscure the underlying simple idea.

For this reason, here we abandon that approach and turn to a different one, which seems to us both simpler and more intuitive from the algorithmic point of view: We will avoid the (anyway computationally unavailable) projections π_j and define the approximation quality δ_j differently, just exploiting usual approximation results for discretization schemes.

2. Discrete versus continuous Newton sequences. In this section, we study the comparative behavior of discrete versus continuous Newton sequences. If not explicitly stated otherwise, the notation is taken from the previous section.

We will consider the phenomenon of mesh independence of Newton's method in two steps. First, we will show that the discrete Newton sequence tracks the continuous Newton sequence closely, with a maximal distance bounded in terms of the mesh size; both of the Newton sequences behave nearly identically until, eventually, a small neighborhood of the solution is reached. Second, we prove the existence of affine invariant Lipschitz constants ω_j for the discretized problems, which approach the Lipschitz constant ω of the continuous problem in the limit $j \to \infty$; again, the distance can be bounded in terms of the mesh size. Upon combining these two lines, we finally establish the existence of locally unique discrete solutions x_j^* in a vicinity of the continuous solution x^*.

To begin with, we prove the following nonlinear perturbation lemma.

LEMMA 2.1. *Consider two Newton sequences $\{x^k\}$, $\{y^k\}$ starting at initial guesses x^0, y^0 and continuing as*

$$x^{k+1} = x^k + \Delta x^k, \quad y^{k+1} = y^k + \Delta y^k,$$

where $\Delta x^k, \Delta y^k$ are the corresponding ordinary Newton corrections. Assume the affine invariant Lipschitz condition (1.2) *is satisfied. Then the following contraction result holds:*

$$\|x^{k+1} - y^{k+1}\| \leq \omega \left(\tfrac{1}{2}\|x^k - y^k\| + \|\Delta x^k\|\right) \|x^k - y^k\|. \tag{2.1}$$

Proof. Dropping the iteration index k, we start with

$$\begin{aligned}
x + \Delta x - y - \Delta y \\
&= x - F'(x)^{-1}F(x) - y + F'(y)^{-1}F(y) \\
&= x - F'(x)^{-1}F(x) + F'(x)^{-1}F(y) - F'(x)^{-1}F(y) - y + F'(y)^{-1}F(y) \\
&= x - y - F'(x)^{-1}\big(F(x) - F(y)\big) + F'(x)^{-1}\big(F'(y) - F'(x)\big)F'(y)^{-1}F(y) \\
&= F'(x)^{-1}\left(F'(x)(x-y) - \int_{t=0}^{1} F'(y + t(x-y))(x-y)\,dt\right) \\
&\quad + F'(x)^{-1}(F'(y) - F'(x))\Delta y.
\end{aligned}$$

Upon using assumption (1.2), we conclude that

$$\begin{aligned}
\|x^{k+1} - y^{k+1}\| &\le \int_{t=0}^{1} \|F'(x^k)^{-1}\big(F'(x^k) - F'(y^k + t(x^k - y^k))\big)(x^k - y^k)\|\,dt \\
&\quad + \|F'(x^k)^{-1}(F'(y^k) - F'(x^k))\Delta y^k\| \\
&\le \frac{\omega}{2}\|x^k - y^k\|^2 + \omega\|x^k - y^k\|\,\|\Delta y^k\|,
\end{aligned}$$

which confirms (2.1). ∎

With the above auxiliary result, we are now ready to study the relative behavior of discrete versus continuous Newton sequences.

THEOREM 2.2. *In addition to the notation as already introduced, let* $x^0 = x_j^0 \in X_j$ *denote a given starting value such that the assumptions of Theorem* 1.1 *hold for the continuous Newton iteration, including*

$$h_0 = \omega\|\Delta x^0\| < 2.$$

For the discrete mapping F_j *and all arguments* $x_j \in D_j = D \cap X_j$ *define*

$$F_j'(x_j)\Delta x_j = -F_j(x_j), \quad F'(x_j)\Delta x = -F(x_j). \tag{2.2}$$

Assume that the discretization is fine enough such that

$$\|\Delta x_j - \Delta x\| \le \delta_j \le \frac{\min\{1, 2 - h_0\}}{2\omega} \tag{2.3}$$

uniformly for $x_j \in D_j$*. Assume furthermore* $\bar{S}\left(x^0, \rho_j\right) \cap X_j \subset D_j$ *for*

$$\rho_j := \frac{\|\Delta x_0\|}{1 - h_0/2} + \frac{2\delta_j}{\min\{1, 2 - h_0\}}.$$

Then the sequence of the discrete Newton iterates x_j^k *remains in* $B(x_0, \rho_j) \cap X_j$ *and the following error estimates hold:*

$$\|x_j^k - x^k\| \le \frac{2\delta_j}{\min\{1, 2 - h_0\}} \le \frac{1}{\omega} \quad \textit{for all } k \in \mathbb{N}, \tag{2.4}$$

$$\limsup_{k \to \infty} \|x_j^k - x^k\| \le 2\delta_j. \tag{2.5}$$

Proof. In [14, pp. 99, 160], Hairer, Nørsett, and Wanner introduced "Lady Windermere's fan" as a tool to prove discretization error results for evolution problems

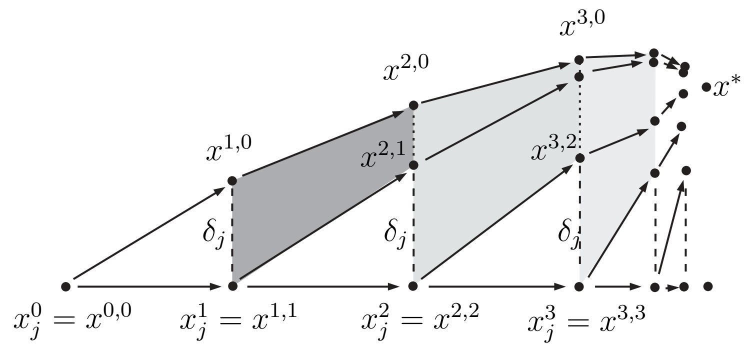

FIG. 2.1. *"Lady Windermere's fan" for the discrete and the continuous Newton method.*

based on some linear perturbation lemma. We may copy this idea and exploit our nonlinear perturbation Lemma 2.1 in the present case. The situation is represented graphically in Figure 2.1.

The discrete Newton sequence starting at the given initial point $x_j^0 = x^{0,0}$ is written as $\{x^{k,k}\}$. The continuous Newton sequence, written as $\{x^{k,0}\}$, starts at the same initial point $x^0 = x^{0,0}$ and runs toward the solution point x^*. In between we define further continuous Newton sequences, written as $\{x^{i,k}\}, k = i, i+1, \ldots$, which start at the discrete Newton iterates $x_j^i = x^{i,i}$ and also run toward x^*. Note that the existence or even uniqueness of a discrete solution point x_j^* is not implied by the assumptions of the theorem.

For the purpose of repeated induction, we assume that

$$\|x_j^{k-1} - x^0\| < \rho_j,$$

which certainly holds for $k = 1$. In order to characterize the deviation between discrete and continuous Newton sequences, we introduce the two majorants

$$\omega\|\Delta x^k\| \le h_k, \quad \|x_j^k - x^k\| \le \epsilon_k.$$

Recall from Theorem 1.1 that

$$h_{k+1} = \frac{1}{2}h_k^2. \tag{2.6}$$

For the derivation of a second majorant recursion, we apply the triangle inequality in the form

$$\|x^{k+1,k+1} - x^{k+1,0}\| \le \|x^{k+1,k+1} - x^{k+1,k}\| + \|x^{k+1,k} - x^{k+1,0}\|.$$

The first term can be treated using assumption (2.3) so that

$$\begin{aligned}\|x^{k+1,k+1} - x^{k+1,k}\| = \|x_j^k + \Delta x_j^k - \left(x^{k,k} + \Delta x^{k,k}\right)\| &= \|\Delta x_j^k - \Delta x^{k,k}\| \\ &\le \delta_j.\end{aligned} \tag{2.7}$$

For the second term, we may apply our nonlinear perturbation Lemma 2.1 (see the shaded regions in Figure 2.1) to obtain

$$\|x^{k+1,k} - x^{k+1,0}\| \le \omega \left(\frac{1}{2}\|x^{k,k} - x^{k,0}\| + \|\Delta x^{k,0}\|\right) \|x^{k,k} - x^{k,0}\|.$$

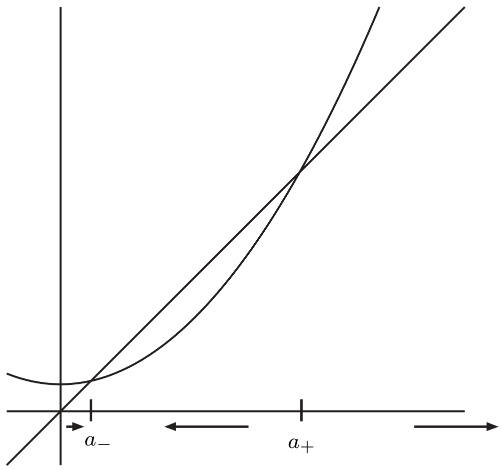

FIG. 2.2. *Fixed point recursion* a_k.

Combining these results then leads to

$$\|x^{k+1,k+1} - x^{k+1,0}\| \le \delta_j + \frac{\omega}{2}\epsilon_k^2 + h_k\epsilon_k.$$

The above right-hand side may be defined to be ϵ_{k+1}. Hence, together with (2.6), we arrive at the following set of majorant equations:

$$\begin{aligned} h_{k+1} &= \frac{1}{2}h_k^2, && h_0 = \omega\|\Delta x_0\|, \\ \epsilon_{k+1} &= \delta_j + \frac{1}{2}\omega\epsilon_k^2 + h_k\epsilon_k, && \epsilon_0 = 0. \end{aligned}$$

Now for $\beta \ge 1$ we multiply the second recursion by $\beta\omega$ and add both recursions. This yields the following recursion for $\beta\omega\epsilon_k + h_k$:

$$\beta\omega\epsilon_{k+1} + h_{k+1} = \beta\omega\delta_j + \frac{1}{2}(\beta\omega\epsilon_k + h_k)^2 - \left[\tfrac{1}{2}(\beta-1)\beta\omega^2\epsilon_k^2\right].$$

Since the term in squared brackets is positive, the sequence a_k defined by

$$a_{k+1} = \beta\omega\delta_j + \frac{1}{2}a_k^2, \qquad a_0 = h_0, \tag{2.8}$$

is a majorant to $\beta\omega\epsilon_k + h_k$. Solving (2.8) yields the equilibrium points

$$a_\pm = 1 \pm \sqrt{1 - 2\beta\omega\delta_j} \tag{2.9}$$

if $2\beta\omega\delta_j \le 1$, which is always possible to guarantee by choosing $1 \le \beta \le (2\omega\delta_j)^{-1}$ due to (2.3). The sequence converges monotonically toward the stable fixed point a_- in case $h_0 < a_+$ (see Figure 2.2). We consider the two cases $h_0 \le 1$ and $h_0 > 1$ separately. If $h_0 \le 1$, we choose

$$\beta = \frac{1}{2\omega\delta_j},$$

such that $h_0 \le a_- = 1$. Due to monotonicity the sequence a_k is bounded from above by $a_- = 1$. We then derive the upper bound

$$\epsilon_k \le \frac{a_-}{\beta\omega} \le 2\delta_j.$$

Both (2.4) and (2.5) are covered by this result. For $1 < h_0 < 2$, we choose $\sigma > 0$ sufficiently small and

$$\beta = \frac{h_0(2-h_0)}{(2+\sigma)\omega\delta_j},$$

such that both $\beta \ge 1$ and $h_0 < a_+$ are satisfied. Due to monotonicity, the sequence a_k is bounded from above by $a_0 = h_0$, and we obtain

$$\epsilon_k \le \frac{h_0}{\beta\omega} = \frac{(2+\sigma)\delta_j}{2-h_0}. \tag{2.10}$$

Since (2.10) holds for all sufficiently small $\sigma > 0$, we obtain

$$\epsilon_k \le \frac{2\delta_j}{2-h_0},$$

which proves (2.4). The asymptotic result (2.5) is now an immediate consequence of $a_k \to a_-$.

Finally, with application of the triangle inequality

$$\|x_j^{k+1} - x^0\| \le \|x^{k+1} - x^0\| + \epsilon_{k+1} < \frac{\|\Delta x_0\|}{1-h_0/2} + \frac{2\delta_j}{\min\{1, 2-h_0\}} = \rho_j,$$

the induction and therefore the whole proof are completed. □

We are interested in the question of whether a discrete solution point x_j^* exists. The above tracking theorem, however, states only that the discrete Newton sequence stays close to the continuous Newton sequence and therefore has an accumulation point close to the continuous solution.

COROLLARY 2.3. *Under the assumptions of Theorem* 2.2*, there exists at least one accumulation point*

$$\hat{x}_j \in \bar{S}\,(x^*, 2\delta_j) \cap X_j \;\subset\; S\left(x^*, \frac{1}{\omega}\right) \cap X_j,$$

which need not be a solution point of the discrete equations $F_j(x_j) = 0$.

In order to prove more, Theorem 1.1 directs us to study whether a Lipschitz condition of the kind (1.2) additionally holds.

LEMMA 2.4. *Assume Theorem* 1.1 *holds for the mapping* $F : X \to Y$. *For collinear* $x_j, y_j, z_j \in X_j$, *define* $u_j \in X_j$ *and* $u \in X$ *according to*

$$F'(x_j)u = \left(F'(z_j) - F'(y_j)\right) v_j, \tag{2.11}$$

$$F_j'(x_j)u_j = \left(F_j'(z_j) - F_j'(y_j)\right) v_j \tag{2.12}$$

for arbitrary $v_j \in X_j$. *Assume that the discretization method satisfies*

$$\|u - u_j\| \le \sigma_j \|z_j - y_j\| \, \|v_j\|. \tag{2.13}$$

Then there exist constants

$$\omega_j \leq \omega + \sigma_j \,, \tag{2.14}$$

such that the affine invariant Lipschitz condition

$$\|u_j\| \leq \omega_j \|z_j - y_j\| \, \|v_j\|$$

holds.

Proof. The proof is a simple application of the triangle inequality:

$$\begin{aligned}\|u_j\| \leq \|u\| + \|u_j - u\| \leq \omega\|z_j - y_j\| \, \|v_j\| + \sigma_j \|z_j - y_j\| \, \|v_j\| \\ = (\omega + \sigma_j) \, \|z_j - y_j\| \, \|v_j\|. \qquad \square\end{aligned}$$

Finally, the existence of a unique discrete solution x_j^* close to the continuous solution x^* is a direct consequence.

COROLLARY 2.5. *Under the assumptions of Theorem* 2.2 *and Lemma* 2.4 *the discrete Newton sequence* $\{x_j^k\}, k = 0, 1, \dots,$ *converges q-quadratically to a unique discrete solution point*

$$x_j^* \in \bar{S}\,(x^*, 2\delta_j) \cap X_j \;\subset\; S\left(x^*, \frac{1}{\omega}\right) \cap X_j.$$

Proof. We just need to apply Theorem 1.1 to the finite dimensional mapping F_j with the starting value $x_j^0 = x^0$, and the affine invariant Lipschitz constant ω_j from (2.14). $\square$

Summarizing, we come to the following conclusion, at least in terms of the analyzed upper bounds: If the asymptotic properties

$$\lim_{j\to\infty} \delta_j = 0, \quad \lim_{j\to\infty} \sigma_j = 0,$$

can be shown to hold, then the convergence speed of the discrete ordinary Newton method is asymptotically just the same as that of the continuous ordinary Newton method. Moreover, if related initial guesses x^0 and x_j^0 and a common termination criterion are chosen, then even the number of iterations will be nearly the same.

3. Application to discretization schemes. In order to apply the abstract mesh independence principles of section 2 to discretization schemes for differential equations, we have to show two features. First,

$$\|\Delta x - \Delta x_j\| \leq \delta_j, \qquad \lim_{j\to\infty} \delta_j = 0, \tag{3.1}$$

where Δx is the exact and Δx_j is the approximate solution of the Newton equations (2.2), respectively.

Second,

$$\|u - u_j\| \leq \sigma_j \|z_j - y_j\| \, \|v_j\|, \qquad \lim_{j\to\infty} \sigma_j = 0, \tag{3.2}$$

where u and u_j are the solutions of the Lipschitz equations (2.11) and (2.12), respectively.

The structure of the argumentation will be straightforward. The first step is to apply classical error estimates for the numerical method under consideration. These estimates usually depend on the regularity of the exact solution y of the linear correction problems. The second step is then to show appropriate regularity results for y.

We concentrate on FEM for elliptic PDEs. Collocation methods for ODEs are discussed in [22].

FEM for semilinear elliptic PDEs. Assume $f : \mathbb{R} \to \mathbb{R}$ is monotonically increasing and locally Lipschitz continuously differentiable with

$$|f'(x) - f'(y)| \leq L(1 + \max(|x|, |y|))|x - y|. \tag{3.3}$$

This implies the growth condition $f = \mathcal{O}(|x|^3)$, which in turn implies that the nonlinear superposition (or Nemyckii) operator $\mathbf{f}$ generated by f maps $H_0^1(\Omega)$ continuously into $L_2(\Omega)$ on some convex polygonal domain $\Omega \subset \mathbb{R}^d$, $d \leq 3$, via the embedding $H_0^1(\Omega) \hookrightarrow L_6(\Omega)$ (cf. [4, 12]). We define the continuous problem $F(x) = 0$ as the boundary value problem

$$F(x) = -\mathrm{div}(\kappa \nabla x) + \mathbf{f}(x) = 0, \quad x \in H_0^1(\Omega), \tag{3.4}$$

with $0 < \underline{\kappa} \leq \kappa \leq \overline{\kappa}$. The discretizations F_j are provided by finite element approximations on shape-regular triangulations $\mathcal{T}_j$ with mesh size $\tau_j = \max_{T \in \mathcal{T}_j} \mathrm{diam}\, T$. We consider piecewise linear finite element spaces $X_j \subset H_0^1(\Omega)$ on the triangulations $\mathcal{T}_j$.

THEOREM 3.1. *Let a bounded set $D \subset H_0^1(\Omega)$ be given. Then there exist constants $M_1, M_2 < \infty$ depending only on D and the problem setting $P = (\Omega, \kappa, f)$, such that the Newton-FEM discretizations F_j satisfy the Newton approximation condition* (3.1) *with $\delta_j = M_1 \tau_j$,*

$$\|\Delta x - \Delta x_j\|_{H^1} \leq M_1 \tau_j \quad \textit{uniformly for } x_j \in D \cap X_j, \tag{3.5}$$

and the Lipschitz approximation condition (3.2) *with $\sigma_j = M_2 \tau_j$,*

$$\|u - u_j\|_{H^1} \leq M_2 \tau_j \|z_j - y_j\|_{H^1} \|v_j\|_{H^1} \tag{3.6}$$

uniformly for all $y_j, z_j \in D \cap X_j$ and $v_j \in X_j$.

Proof. First we prove (3.5). Let Δx satisfy $F'(x_j)\Delta x = -F(x_j)$ and let Δx_j be its FEM approximation. Returning to (2.7) we notice that $x^{k+1,k}$ is more regular than $\Delta x^{k,k}$. Thus we introduce $w = x_j + \Delta x$, which satisfies

$$-\mathrm{div}(\kappa \nabla w) + \mathbf{f}'(x_j) w = -\mathbf{f}(x_j) + \mathbf{f}'(x_j) x_j. \tag{3.7}$$

The growth condition (3.3) implies $\mathbf{f}(x_j) \in L_2$ and $\mathbf{f}'(x_j) \in L_3$, such that the right-hand side of (3.7) is contained in L_2. We may estimate

$$\begin{aligned}
\|\mathbf{f}(x_j) - \mathbf{f}'(x_j)x_j\|_{L_2} &= \left\| \int_{t=0}^1 (\mathbf{f}'(tx_j) - \mathbf{f}'(x_j))x_j \, dt + \mathbf{f}(0) \right\|_{L_2} \\
&\leq \int_{t=0}^1 L(1-t)\|(1 + |x_j|)x_j^2\|_{L_2} \, dt + c \\
&\leq \frac{L}{2}(\|x_j^2\|_{L_2} + \|x_j^3\|_{L_2}) + c \\
&= c(\|x_j\|_{L_4}^2 + \|x_j\|_{L_6}^3 + 1) \\
&\leq c(\|x_j\|_{H^1}^2 + \|x_j\|_{H^1}^3 + 1) \\
&\leq c,
\end{aligned}$$

where c denotes a generic constant independent of the discretization and x_j. Since the Helmholtz term in (3.7) is positive semidefinite due to the monotonicity of f, the inverse of the differential operator can be bounded in terms of the ellipticity constant of its main part only, which is independent of x_j. Thus we obtain

$$\|w\|_{H^1} \leq c\|\mathbf{f}(x_j) - \mathbf{f}'(x_j)x_j\|_{L_2} \leq c.$$

Using Hölder's inequality and the embedding $H_0^1(\Omega) \hookrightarrow L_6(\Omega)$, we estimate

$$\|\mathbf{f}'(x_j)w\|_{L_2} \leq \|w\|_{L_6}\|\mathbf{f}'(x_j)\|_{L_3} \leq \|w\|_{H^1}\, c(1+\|x_j\|_{L_6}^2) \leq c.$$

We now rewrite (3.7) as

$$-\mathrm{div}(\kappa\nabla w) = -\mathbf{f}(x_j) + \mathbf{f}'(x_j)x_j - \mathbf{f}'(x_j)w.$$

Since the right-hand side is contained in L_2, the solution w is H^2-regular (cf. [13]) with

$$\|w\|_{H^2} \leq c\|\mathbf{f}(x_j) - \mathbf{f}'(x_j)x_j + \mathbf{f}'(x_j)w\|_{L_2} \leq c.$$

We thus obtain an approximation error

$$\|w_j - w\|_{H^1} \leq c\tau_j\|w\|_{H^2} \leq c\tau_j$$

for its FEM approximation $w_j = x_j + \Delta x_j$ (cf. [5, p. 79]), uniformly for all x_j. For the approximation error $\Delta x_j - \Delta x$ we now obtain

$$\|\Delta x_j - \Delta x\|_{H^1} = \|w_j - w\|_{H^1} \leq c\tau_j.$$

Second, we prove (3.6). u is defined by

$$F'(x_j)u = (F'(z_j) - F'(y_j))v_j = (\mathbf{f}'(z_j) - \mathbf{f}'(y_j))v_j.$$

As before, the right-hand side is contained in L_2 and the solution u is H^2-regular, such that we obtain

$$\|u_j - u\|_{H^1} \leq c\tau_j\|(\mathbf{f}'(z_j) - \mathbf{f}'(y_j))v_j\|_{L_2}.$$

Upon using Hölder's inequality twice we conclude that

$$\begin{aligned}\|(\mathbf{f}'(z_j) - \mathbf{f}'(y_j))v_j\|_{L_2} &\leq \|L^2(1+\max(|y_j|,|z_j|))^2(z_j-y_j)^2v_j^2\|_{L_1}^{1/2}\\ &\leq \left(\|L^2(1+\max(|y_j|,|z_j|))^2\|_{L_3}\|(z_j-y_j)^2\|_{L_3}\|v_j^2\|_{L_3}\right)^{1/2}\\ &= L\|1+\max(|y_j|,|z_j|)\|_{L_6}\|z_j-y_j\|_{L_6}\|v_j\|_{L_6} \leq c,\end{aligned}$$

which completes the proof. ◻

Combining Theorem 2.2 and Lemma 2.4 with Theorem 3.1 we obtain asymptotic mesh independence for FEM approximations of semilinear elliptic equations.

COROLLARY 3.2. *Assume that there exists a convex and bounded set $D \subset H^1$, such that on D the assumptions of Theorem* 1.1 *(in particular $\omega < \infty$) and Theorem* 3.1 *are satisfied for the nonlinear equation* (3.4).

Then there exists a constant M_1 and a mesh size $\tau_0 > 0$, such that for all discretizations X_j with corresponding mesh size $\tau_j < \tau_0$ and starting values $x^0 = x_j^0 \in X_j$ with

$$h_0 = \omega\|\Delta x^0\|_{H^1} < 2 \tag{3.8}$$

and

$$\bar{S}\left(x^0, \frac{\|\Delta x_0\| + 2M_1\tau_j}{1-h_0/2}\right) \subset D, \tag{3.9}$$

the discrete Newton sequence remains in D, and its distance to the continuous Newton sequence is bounded by

$$\|x_j^k - x^k\|_{H^1} \leq \frac{2M_1\tau_j}{1-h_0/2}. \tag{3.10}$$

Moreover, both sequences converge q-quadratically to solutions x_j^* and x^*, respectively, with

$$\|x_j^* - x^*\|_{H^1} \leq 2M_1\tau_j. \tag{3.11}$$

Proof. Application of Theorem 3.1 on D yields constants $M_1, M_2 < \infty$ such that $\|\Delta x - \Delta x_j\|_{H^1} \leq M_1\tau_j$ and $\|u-u_j\|_{H^1} \leq M_2\tau_j\|z_j-y_j\|_{H^1}\|v_j\|_{H^1}$ hold for all $x_j \in D$ in terms of (3.1) and (3.2). We will verify Corollary 3.2 for

$$\tau_j < \tau_0 := \min\left\{\frac{1-h_0/2}{2\omega M_1}, \frac{1}{2M_2}\left(\sqrt{\omega^2 + \frac{M_2(1-h_0/2)}{M_1}} - \omega\right)\right\}.$$

Note that the continuous Newton sequence satisfying (3.8) and (3.9) remains in $S(x^0, \frac{\|\Delta x_0\|_{H^1}}{1-h_0/2}) \subset D$ due to Theorem 1.1. Because of

$$\tau_j < \tau_0 \leq \frac{1-h_0/2}{2\omega M_1} \leq \frac{\min\{1, 2-h_0\}}{2\omega M_1},$$

condition (2.3) is clearly satisfied, such that we can apply Theorem 2.2 and obtain (3.10), (3.11), and $x_j^k \in D$.

Now we turn to q-quadratic convergence of the discrete Newton sequence. A direct consequence of (3.10) is the estimate

$$\begin{aligned}\|\Delta x_j^k\|_{H^1} &\leq \|x_j^{k+1} - x^{k+1}\|_{H^1} + \|\Delta x^k\|_{H^1} + \|x^k - x_j^k\|_{H^1}\\ &\leq \frac{4M_1\tau_j}{1-h_0/2} + \|\Delta x^k\|_{H^1}.\end{aligned}$$

As $\lim_{k\to\infty}\|\Delta x^k\|_{H^1} = 0$ by Theorem 1.1 we can find an index k_0 such that

$$\|\Delta x_j^k\|_{H^1} \leq \frac{8M_1\tau_j}{1-h_0/2} \qquad \text{for all } k \geq k_0.$$

Application of Lemma 2.4 yields $\omega_j \leq \omega + \tau_j M_2$ and therefore

$$h_j^k := \omega_j\|\Delta x_j^k\|_{H^1} \leq (\omega + \tau_j M_2)\frac{8M_1\tau_j}{1-h_0/2} \qquad \text{for all } k \geq k_0.$$

Now

$$\tau_j < \tau_0 \leq \frac{1}{2M_2}\left(\sqrt{\omega^2 + \frac{M_2(1-h_0/2)}{M_1}} - \omega\right)$$

implies $h_j^{k_0} < 2$, such that Theorem 1.1 yields q-quadratic convergence of the discrete Newton iteration starting at $x_j^{k_0}$. ∎

FEM for strongly nonlinear elliptic PDEs. For strongly nonlinear PDEs with a second order differential operator depending on the solution, the analytic treatment of the approximation conditions (3.1) and (3.2) is considerably more difficult. The global regularity of the right-hand side is, in general, only H^{-1}, which results in sharp edges in the Newton correction. These bucklings, however, coincide geometrically with the edges of the triangulation, such that the finite element approximation quality does not deteriorate. This effect is indeed observed in actual computation.

The regularity theory necessary for addressing such problems is beyond the scope of the present paper. As a substitute, we give a numerical example from [10], where the phenomenon of asymptotic mesh independence may be studied.

Example: Parametric minimal surface. Consider the variational problem

$$\min \int_\Omega \sqrt{1+|\nabla u|^2}\, dx$$

subject to the boundary conditions

$$\begin{aligned} u &= \cos(x)\cos(y) && \text{on } \Gamma_D = \partial\Omega\backslash\Gamma_N, \\ \frac{\partial u}{\partial n} &= 0 && \text{on } \Gamma_N \end{aligned}$$

on $\Omega = [-\pi/2, 0]^2$. The functional gives rise to the first and second order expressions

$$\begin{aligned} \langle F(u), v\rangle &= \int_\Omega \bigl(1+|\nabla u|^2\bigr)^{-1/2} \nabla u^T \nabla v\, dx, \\ \langle F'(u)v, w\rangle &= \int_\Omega \Bigl(-\bigl(1+|\nabla u|^2\bigr)^{-3/2} \nabla w^T (\nabla u \nabla u^T) \nabla v \\ &\qquad + \bigl(1+|\nabla u|^2\bigr)^{-1/2} \nabla w^T \nabla v\Bigr)\, dx. \end{aligned}$$

We define two different problem settings by choosing

(a) $\Gamma_N = [-\pi/2, 0] \times \{0\}$,

(b) $\Gamma_N = [-\pi/2, 0] \times \{0\} \cup \{0\} \times [-\pi/4, 0]$.

Note that by symmetry, problem (a) represents a Dirichlet problem on a convex domain, whereas the deliberate choice of boundary conditions (b) leads to a Dirichlet problem on a highly nonconvex slit domain, on which no physically meaningful solution exists.

The adaptive Newton-multilevel code Newton-KASKADE [9, 10] has been run on both problems, providing affine invariant computational estimates $[\omega_j] \le \omega_j$ on each mesh refinement level j. On each level, a few Newton steps have been computed using the approximation from the level before, and the maximum estimate encountered in these steps has been selected as $[\omega_j]$. As can be seen from Table 3.1, the Lipschitz constants for the well-defined problem (a) remain bounded and rather independent of the refinement level, apart from some fluctuation due to the finite sampling of ω_j. In contrast to that, the estimates for the Lipschitz constant of problem (b) are dramatically increasing by five orders of magnitude. This indicates that the problem has finite dimensional solutions on each of the successive meshes, each unique within the corresponding finite dimensional Kantorovich ball with radius $\rho_j \sim 1/\omega_j$; however, these balls shrink from radius $\rho_0 \approx 1$ to $\rho_{12} \approx 10^{-5}$. Frank extrapolation of this effect insinuates the conjecture that there exists no continuous unique solution of the underlying minimization problem.

Table 3.1
Estimated Lipschitz constants $[\omega_j]$ on different refinement levels j.

	Problem (a)		Problem (b)	
j	♯nodes	$[\omega_j]$	♯nodes	$[\omega_j]$
0	4	1.32	5	7.5
1	7	1.17	10	4.2
2	18	4.55	17	7.3
3	50	6.11	26	9.6
4	123	5.25	51	22.5
5	158	20.19	87	50.3
6	278	19.97	105	1486.2
7	356	9.69	139	2715.6
8	487	8.47	196	5178.6
9	632	11.73	241	6837.2
10	787	44.21	421	12040.2
11	981	49.24	523	167636.0
12	1239	20.10	635	1405910.0
13	1610	32.93		
14	2054	37.22		

Acknowledgment. One of the authors (P.D.) wishes to thank Dietrich Braess, Bochum, for detailed discussions about the earlier paper [8]; these discussions eventually led to our decision to analyze asymptotic mesh independence in a simpler and more intuitive framework. Moreover, the authors gratefully acknowledge his constructive comments on an earlier draft of this paper.

REFERENCES

[1] E. L. Allgower and K. Böhmer, *Application of the mesh independence principle to mesh refinement strategies*, SIAM J. Numer. Anal., 24 (1987), pp. 1335–1351.

[2] E. L. Allgower, K. Böhmer, F. A. Potra, and W. C. Rheinboldt, *A mesh-independence principle for operator equations and their discretizations*, SIAM J. Numer. Anal., 23 (1986), pp. 160–169.

[3] W. Alt, *Mesh-independence of the Lagrange-Newton method for nonlinear optimal control problems and their discretizations*, Ann. Oper. Res., 101 (2001), pp. 101–117.

[4] J. Appell and P. P. Zabrejko, *Nonlinear Superposition Operators*, Cambridge University Press, Cambridge, UK, 1990.

[5] D. Braess, *Finite Elements*, 2nd ed., Cambridge University Press, Cambridge, UK, 2001.

[6] D. Braess, *private communication*, 2002.

[7] P. Deuflhard and G. Heindl, *Affine invariant convergence theorems for Newton's method and extensions to related methods*, SIAM J. Numer. Anal., 16 (1979), pp. 1–10.

[8] P. Deuflhard and F. Potra, *Asymptotic mesh independence of Newton–Galerkin methods via a refined Mysovskii theorem*, SIAM J. Numer. Anal., 29 (1992), pp. 1395–1412.

[9] P. Deuflhard and M. Weiser, *Local inexact Newton multilevel FEM for nonlinear elliptic problems*, in Computational Science for the 21st Century, M.-O. Bristeau, G. Etgen, W. Fitzigibbon, J.-L. Lions, J. Periaux, and M. Wheeler, eds., Wiley-Interscience, New York, 1997, pp. 129–138.

[10] P. Deuflhard and M. Weiser, *Global inexact Newton multilevel FEM for nonlinear elliptic problems*, in Multigrid Methods V, Lect. Notes Comput. Sci. Eng. 3, W. Hackbusch and G. Wittum, eds., Springer-Verlag, Berlin, 1998, pp. 71–89.

[11] A. Dontchev, W. Hager, and V. Veliov, *Uniform convergence and mesh independence of Newton's method for discretized variational problems*, SIAM J. Control Optim., 39 (2000), pp. 961–980.

[12] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Grundlehren Math. Wiss. 224, Springer-Verlag, Berlin, New York, 1977.

[13] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Monographs and Studies in Mathematics 24, Pitman, Boston, 1985.

[14] E. Hairer, S. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations.* I*: Nonstiff Problems*, Springer Ser. Comput. Math. 8, 2nd ed., Springer-Verlag, Berlin, 1993.

[15] M. Heinkenschloss, *Mesh independence for nonlinear least squares problems with norm constraints*, SIAM J. Optim., 3 (1993), pp. 81–117.

[16] C. Kelley and E. Sachs, *Mesh independence of Newton-like methods for infinite dimensional problems*, J. Integral Equations Appl., 3 (1991), pp. 549–573.

[17] C. T. Kelley and E. W. Sachs, *Mesh independence of the gradient projection method for optimal control problems*, SIAM J. Control Optim., 30 (1992), pp. 477–493.

[18] M. Laumen, *Newton's mesh independence principle for a class of optimal shape design problems*, SIAM J. Control Optim., 37 (1999), pp. 1070–1088.

[19] S. McCormick, *A revised mesh refinement strategy for Newton's method applied to nonlinear two-point boundary value problems*, in Numerical Treatment of Differential Equations Applications, Lecture Notes in Math. 679, Springer-Verlag, Berlin, 1978, pp. 15–23.

[20] S. Volkwein, *Mesh-independence for an augmented Lagrangian-SQP method in Hilbert spaces*, SIAM J. Control Optim., 38 (2000), pp. 767–785.

[21] S. Volkwein, *Mesh-independence of Lagrange-SQP methods with Lipschitz-continuous Lagrange multiplier updates*, Optim. Methods Softw., 17 (2002), pp. 77–111.

[22] M. Weiser, A. Schiela, and P. Deuflhard, *Asymptotic Mesh Independence of Newton's Method Revisited*, preprint, ZIB Report 03-13, Zuse Institute Berlin, Berlin, Germany, 2003.