NONLINEAR MULTIGRID TECHNIQUES

RALF KORNHUBER

ABSTRACT. In these introductory notes, we focus on smooth and piecewise smooth semilinear elliptic partial differential equations in the framework of convex minimization problems. We present and analyze old and new nonlinear multigrid techniques. Related procedures are often used without proof in practical applications. After a brief motivation of linear multigrid, we principally concentrate on Newton multigrid for smooth problems and on monotone multigrid for the piecewise smooth case.

1. INTRODUCTION

The world is nonlinear. Hence, nonlinear phenomena play a crucial role in almost all branches of numerical analysis and scientific computing. Nonlinearity typically interacts with other specific properties of the problem under consideration so that straightforward application of standard algorithms like Newton's method usually leads to unsatisfactory results. Suitable modifications are often based on physical insight in the underlying process or on insight in the structure of the mathematical model.

In these introductory notes, we use convex minimization as a framework for smooth and piecewise smooth semilinear elliptic partial differential equations. Existence and uniqueness are addressed as well as finite element discretization and convergence. Priority is given to fast solution of discrete problems by multigrid methods. In order to reduce technicalities we only consider model problems satisfying rather restrictive conditions. We present and to analyze old and new nonlinear multigrid techniques. Our aim is to clarify the theoretical background of related procedures which are often used without proof in practical applications.

It is well–known that multigrid methods work excellent for linear self–adjoint problems: Under reasonable assumptions convergence takes place with mesh independent convergence rates. As a consequence, approximate solutions with optimal accuracy can be computed at minimal computational cost by using nested iteration. One of our principal concerns is to find out whether this property can be extended to nonlinear problems and, if yes, how this can be done.

To warm up, we first consider a linear elliptic self–adjoint partial differential equation. It is well-known that weak formulations of such problems can be rewritten as minimization problems for a quadratic energy functional. Motivated by a heuristic notion of frequencies a classical multigrid algorithm with Gauß–Seidel smoother is introduced as successive subspace correction method via successive minimization of energy.

In the next section we concentrate on smooth semilinear elliptic problems leading to a smooth energy functional. We consider existence and uniqueness and derive a

The author wishes to thank M. Weiser for his helpful comments on the manuscript.

priori error estimates for a finite element discretization involving numerical integration of nonlinear terms. Most part of the section is devoted to Newton multigrid or more generally to inexact Newton methods. Abstract formulations in standard textbooks usually focus on basic aspects and deliberately hide the difficulties that typically arise in verifying all required prerequisites for a given nonlinear partial differential equation. Our presentation is intended to fill this gap. We prove local linear convergence and global convergence of a damped version using local energy norms. Mesh independent convergence of linear multigrid provides local mesh independent convergence of Newton multigrid. This key property is leading to optimality of damped Newton multigrid with nested iteration. Appropriate choice of norms, stopping criteria for Newton multigrid iteration, the number of linear multigrid steps in each Newton step or of other parameters is briefly discussed. At the end of this section, we shortly motivate nonlinear multigrid and refer to well–known convergence results.

In the final section, we extend our considerations to the piecewise smooth case. We present so–called monotone multigrid methods consisting of one substep of a globally convergent descent method and subsequent constrained Newton linearization. The first substep is intended to deal with the non–smoothness of the problem while the second part is intended to increase local convergence speed. The role of linear problems in classical Newton linearization is taken by obstacle problems in the constrained Newton step. After a number of iterations overall convergence of monotone multigrid is dominated by convergence speed of the inner obstacle solver. Mesh independent convergence of related monotone multigrid methods for obstacle problems is observed in numerical computations but theoretical justification is subject of current research.

We only considered a very restrictive class of problems. Nevertheless, there is a huge number of articles in the literature treating the subject from many different points of view. We fear that our list of references is far from exhaustive and have to apologize for each important contribution that is missing.

2. Self–adjoint linear problems

2.1. **Continuous problem and discretization.** Let Ω be a bounded, polyhedral domain in the Euclidean space $\mathbb{R}^d$, $d = 2, 3$. We consider the minimization problem

$$u \in H : \qquad \mathcal{J}(u) \le \mathcal{J}(v) \qquad \forall v \in H \tag{2.1}$$

on the closed subspace $H = H_0^1(\Omega) \subset H^1(\Omega)$. The energy functional $\mathcal{J}$,

$$\mathcal{J}(v) = \tfrac{1}{2} a(v, v) - \ell(v), \tag{2.2}$$

is induced by the bilinear form $a(\cdot, \cdot)$,

$$a(v, w) = \int_\Omega \alpha(x) \nabla v(x) \cdot \nabla w(x)\, dx,$$

and a linear functional $\ell \in H$. We assume that the coefficient $\alpha \in C^1(\overline{\Omega})$ is positive on $\overline{\Omega}$. Hence, $a(\cdot, \cdot)$ defines a scalar product on H and the corresponding *energy norm*

$$\|v\| = a(v, v)^{1/2}, \qquad v \in H, \tag{2.3}$$

is equivalent to the canonical norm $\|\cdot\|_{H^1(\Omega)}$ of $H^1(\Omega)$, i.e.

$$\alpha_0 \|v\|_{H^1(\Omega)} \le \|v\| \le \alpha_1 \|v\|_{H^1(\Omega)} \qquad \forall v \in H \tag{2.4}$$

holds with positive constants α_0, $\alpha_1 \in \mathbb{R}$. The minimization problem (2.1) has a unique solution u and can be equivalently rewritten as variational equality

$$u \in H: \qquad a(u,v) = \ell(v) \qquad \forall v \in H. \tag{2.5}$$

For example, let $\ell(v) = (f,v)_{L^2(\Omega)}$ with $f \in L^2(\Omega)$. Then (2.5) is the weak formulation of the boundary value problem

$$-\nabla \cdot (\alpha(x)\nabla u(x)) = f(x) \qquad \forall x \in \Omega \tag{2.6}$$

and Dirichlet boundary conditions $u(x) = 0 \; \forall x \in \partial\Omega$.

Proposition 2.1. *There is a sequence of real eigenvalues*

$$0 < \mu_0 \le \mu_1 \le \mu_2 \le \cdots$$

with no finite accumulation point and $\lim_{k\to\infty} \mu_k = \infty$. *The corresponding eigenfunctions* $e_k \in H$,

$$a(e_k, v) = \mu_k (e_k, v)_{L^2(\Omega)} \qquad \forall v \in H,$$

are an a–orthogonal basis of H.

For a proof we refer e.g. to Raviart and Thomas [78], pp. 135 or Renardy and Rogers [79], pp. 299.

Example 2.2. *Choosing* $\Omega = (0,1)$ *and* $a(v,w) = (v', w')_{L^2(\Omega)}$, *we have*

$$\mu_k = (k\pi)^2, \qquad e_k(x) = \sin(k\pi x), \qquad k \in \mathbb{N}.$$

Proposition 2.1 implies that there is an a–orthogonal splitting

$$H = \mathcal{V}_0 + \mathcal{V}_1 + \mathcal{V}_2 + \cdots, \qquad \mathcal{V}_k = \operatorname{span}\{e_k\}, \tag{2.7}$$

into subspaces $\mathcal{V}_k$ representing a scale of increasing frequencies μ_k. Let $u^0 \in H$ be some guess of u. Due to a–orthogonality of the splitting (2.7), corrections $v_k \in \mathcal{V}_k$ as obtained by minimizing $\mathcal{J}$ *separately on each subspace* $\mathcal{V}_k$ provide the exact solution:

$$u = u^0 + \sum_{k=0}^{\infty} v_k, \qquad v_k = \frac{\ell(e_k) - a(u^0, e_k)}{a(e_k, e_k)} e_k. \tag{2.8}$$

Unfortunately, eigenfunctions e_k are usually not known in practice.

Let us now describe a standard finite element discretization of (2.1). A *triangulation* $\mathcal{T}$ of $\Omega \subset \mathbb{R}^d$ is a set of d–simplices such that $\bigcup_{t\in\mathcal{T}} t = \overline{\Omega}$ and such that the intersection of $t, t' \in \mathcal{T}$ is either a k–simplex with $k < d$ or empty. We consider the nested sequence

$$\mathcal{T}_0 \subset \mathcal{T}_1 \subset \cdots \subset \mathcal{T}_j, \qquad j \in \mathbb{N},$$

of triangulations resulting from successive refinement of a given, intentionally coarse triangulation $\mathcal{T}_0$. For simplicity, we assume that the triangulations are uniformly refined. In two space dimensions this means that each triangle $t \in \mathcal{T}_{k-1}$ is divided into four congruent subtriangles to obtain $\mathcal{T}_k$. For $d = 3$ the situation is more complicated (see e.g. Bey [12] or Bornemann et al. [14]). For most practical problems it is absolutely necessary to use locally refined grids in order to reduce numerical complexity. All algorithms and most of the theoretical results to be presented can be extended to this case. For basic information on adaptive multilevel methods in two and three space dimensions, we refer e.g. to Bank et al. [7], Kornhuber [65], pp. 66, Yserentant [92] and Bornemann et al. [14], respectively.

Denoting

$$h_k = \max_{t \in \mathcal{T}_k} \operatorname{diam}(t) \qquad k = 0, \ldots, j,$$

we have

$$h_j = \mathcal{O}(2^{-j}), \qquad h_{k-1} = 2h_k, \quad k = 1, \ldots, j\ . \tag{2.9}$$

Choosing *piecewise linear finite elements*

$$\mathcal{S}_k = \{v \in H_0^1(\Omega) \mid v|_t \text{ is linear } \forall t \in \mathcal{T}_k\},$$

we obtain a *nested sequence of finite dimensional subspaces*

$$\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_j \subset H. \tag{2.10}$$

Each space $\mathcal{S}_k$ is spanned by the *nodal basis*

$$\Lambda_k = \{\lambda_p^{(k)} \mid p \in \mathcal{N}_k\}, \qquad \lambda_p^{(k)}(q) = \delta_{pq} \ \ \forall p, q \in \mathcal{N}_k \ \ \text{(Kronecker–}\delta).$$

$\mathcal{N}_k$ is denoting the set of the n_k vertices of $\mathcal{T}_k$ that are contained in Ω. The *Ritz–Galerkin* approximation $u_j \in \mathcal{S}_j$ of u is obtained by replacing H by the ansatz space $\mathcal{S}_j$, i.e.

$$u_j \in \mathcal{S}_j : \qquad \mathcal{J}(u_j) \le \mathcal{J}(v) \qquad \forall v \in \mathcal{S}_j \tag{2.11}$$

or, equivalently,

$$u_j \in \mathcal{S}_j : \qquad a(u_j, v) = \ell(v) \qquad \forall v \in \mathcal{S}_j. \tag{2.12}$$

Representing u_j by the nodal basis, the discrete problem (2.12) can be reformulated as a linear system for the unknown nodal values $u_j(p)$, $p \in N_j$.

It is well–known that $\|u - u_j\| = \mathcal{O}(h_j)$, if $u \in H \cap H^2(\Omega)$ which holds for sufficiently smooth data, e.g. for $f \in L^2(\Omega)$ and convex domains Ω. We refer to monographs of Braess [16], Brenner and Scott [27] or Ciarlet [29].

In the remainder of this section, we shall concentrate on the fast solution of (2.11). Similar to (2.8), we could solve (2.11) *separately on a–orthogonal subspaces* $\mathcal{V}_k = \operatorname{span}\{e_k^{(j)}\}$, $k = 1, \ldots, n_j$, if eigenfunctions $e_k^{(j)}$ of $a(\cdot,\cdot)|_{\mathcal{S}_j \times \mathcal{S}_j}$ were known. Again, this is usually not the case, but again we always know that eigenfunctions must represent some scale of frequencies.

Conversely, if certain subspaces $\mathcal{V}_k$ represent a scale of frequencies, then the corresponding splitting $\mathcal{S}_j = \mathcal{V}_0 + \mathcal{V}_1 + \cdots + \mathcal{V}_m$ might be "almost" a–orthogonal in some sense. Then successive minimization of energy $\mathcal{J}$ on $\mathcal{V}_k$ should provide a fast solver. This idea is our starting point for the construction of multigrid methods.

2.2. **Successive subspace correction and multigrid.** Let

$$\mathcal{S}_j = \mathcal{V}_0 + \mathcal{V}_1 + \cdots + \mathcal{V}_m \tag{2.13}$$

be a given splitting of $\mathcal{S}_j$. Then, successive minimization of energy $\mathcal{J}$ on $\mathcal{V}_k$ leads to the following algorithm for computing a new iterate $u_j^{\nu+1}$ from some given $u_j^\nu \in \mathcal{S}_j$.

Algorithm 2.3. *(Successive minimization)*

given: $w_{-1} = u_j^\nu \in \mathcal{S}_j$.

for $k = 0, \ldots, m$ *solve:*

$$v_k \in \mathcal{V}_k : \quad \mathcal{J}(w_{k-1} + v_k) \le \mathcal{J}(w_{k-1} + v) \ \ \forall v \in \mathcal{V}_k, \qquad w_k = w_{k-1} + v_k, \tag{2.14}$$

new iterate: $u_j^{\nu+1} = w_m$.

We now try to choose the splitting (2.13) in such a way that Algorithm 2.3 generates an iterative scheme with mesh independent convergence rates that can be implemented with $\mathcal{O}(n_j)$ complexity.

Subproblems (2.14) can be easily solved in case of one–dimensional subspaces V_l. Straightforward nodal splitting

$$\mathcal{S}_j = \sum_{l=1}^{n_j} V_l, \qquad V_l = \operatorname{span}\{\lambda_{p_l}^{(j)}\}, \qquad l = 1, \dots, n_j, \tag{2.15}$$

produces the well–known *Gauß–Seidel relaxation.* Obviously, subspaces V_l as used in (2.15) do *not* represent a scale of frequencies. Only high–frequency functions $\lambda_{p_l}^{(j)}$ are involved. Hence, it is not astonishing that Gauß–Seidel iteration rapidly reduces high–frequency contributions of the error but scarcely affects low frequencies, cf. e.g. Hackbusch [52], pp. 49. In order to incorporate low frequencies, we make use of the hierarchy (2.10).

Collecting all basis functions $\lambda_p^{(k)}$ from all finite element spaces $\mathcal{S}_k$, we define the so–called *multilevel nodal basis* Λ,

$$\Lambda = \bigcup_{k=0}^{j} \Lambda_k = \{\lambda_l \mid l = 1, \dots, m_{\mathcal{S}}\}\ , \qquad m_{\mathcal{S}} = n_0 + \cdots + n_j .$$

Enumeration $\lambda_l = \lambda_{p_l}^{(k_l)}$ is taken from fine to coarse, i.e. $l > l'$ implies $k_l \le k_{l'}$. From an heuristic point of view, the splitting

$$\mathcal{S}_j = \sum_{l=1}^{n_j} V_l, \qquad V_l = \operatorname{span}\{\lambda_l\}, \qquad l = 1, \dots, m_{\mathcal{S}}, \tag{2.16}$$

represents a scale of frequencies ranging from low–frequency functions $\lambda_l \in \Lambda_0$ to high–frequency functions $\lambda_l \in \Lambda_j$. From the extended splitting (2.16) Algorithm 2.3 generates the *extended relaxation*

$$u_j^{\nu+1} = u_j^{\nu} + \sum_{l=1}^{m_{\mathcal{S}}} v_l, \qquad v_l = \frac{\ell(\lambda_l) - a(w_{l-1}, \lambda_l)}{a(\lambda_l, \lambda_l)} \lambda_l, \tag{2.17}$$

where $w_l = w_{l-1} + v_l$ and $w_0 = u_j^{\nu}$. It is convenient for analysis and implementation of (2.17) to arrange the computation of corrections v_l according to the refinement levels $k = 0, \dots, j$. For this reason, we consider the splitting of $\mathcal{S}_j$ into subspaces

$$\mathcal{V}_k = \mathcal{S}_k, \quad k = 0, \dots, j \tag{2.18}$$

which directly reflect the hierarchy (2.10). Gauß–Seidel iteration on $\mathcal{V}_k$ gives rise to the bilinear form $b_k(\cdot,\cdot)$,

$$b_k(v, w) = \sum_{\substack{i,l=1 \\ i \le l}}^{n_k} v(p_i) a\big(\lambda_{p_i}^{(k)}, \lambda_{p_l}^{(k)}\big) w(p_l)\ , \qquad v, w \in \mathcal{V}_k\ . \tag{2.19}$$

The nodes $p_i \in \mathcal{N}_k$ are ordered in the same way as the corresponding subspaces $V_l = \operatorname{span} \lambda_{p_l}^{(k_l)}$ on level $k_l = k$. As Gauß–Seidel iteration rapidly reduces high frequency components, $b_k(\cdot,\cdot)$ is called a *smoother* on $\mathcal{V}_k$. Now extended relaxation (2.17) can be rewritten as follows.

Algorithm 2.4. *(Successive subspace correction)*

given: $w_{j+1} = u_j^\nu \in \mathcal{S}_j$.

for $k = j, \dots, 0$ *solve:*

$$v_k \in \mathcal{V}_k : \qquad b_k(v_k, v) = \ell(v) - a(w_{k+1}, v) \qquad \forall v \in \mathcal{V}_k, \tag{2.20}$$
$$w_k = w_{k+1} + v_k.$$

new iterate: $u_j^{\nu+1} = w_0$.

Algorithm 2.4 is the prototype of a *successive subspace correction method* as introduced by Xu [87]. Selecting appropriate subspaces $\mathcal{V}_k$ and smoothers $b_k(\cdot,\cdot)$ a large number of multilevel and domain decomposition methods can be reformulated in this way. Note that Algorithm 2.3 would be recovered by choosing $b_k(\cdot,\cdot) = a(\cdot,\cdot)$.

In case of multilevel splitting (2.18) Algorithm 2.4 can be implemented as a *multigrid V–cycle with 1 pre–smoothing step.*

Algorithm 2.5. *(Multigrid V–cycle)*

given: u_j^ν

initialize: $r_j = \ell - a(u_j^\nu, \cdot), \quad a_j(\cdot,\cdot) = a(\cdot,\cdot)$

for $k = j, \dots, 1$ *do:*

{

solve:

$v_k \in \mathcal{V}_k : \quad b_k(v_k, v) = r_k(v) \quad \forall v \in \mathcal{V}_k$ *(pre–smoothing)*

$r_k = r_k - a_k(v_k, \cdot)$ *(update of residual)*

$r_{k-1} = r_k|_{\mathcal{S}_{k-1}}$

$a_{k-1}(\cdot,\cdot) = a_k(\cdot,\cdot)|_{\mathcal{S}_{k-1}\times\mathcal{S}_{k-1}}$ *(canonical restriction)*

}

solve:

$v_0 \in \mathcal{V}_0 : \quad a(v_0, v) = r_0(v) \quad \forall v \in \mathcal{V}_0$ *(approx. coarse grid solution)*

for $k = 1, \dots, j$ *do:*

{

$v_k = v_k + v_{k-1}$ *(canonical interpolation)*

}

new iterate: $u_j^{\nu+1} = u_j^\nu + v_j$

Canonical restrictions r_{k-1} and $a_{k-1}(\cdot,\cdot)$ of residual r_k and bilinear form $a_k(\cdot,\cdot)$ are defined by $r_{k-1}(v) = r_k(v)$ and $a_{k-1}(w, v) = a_k(v, w)$ for all $v, w \in \mathcal{S}_{k-1} \subset \mathcal{S}_k$, respectively. Note that $r_k = \ell - a(w_{k+1}, \cdot)$.

It is easily shown that the number of point operations required for one step of Algorithm 2.5 is bounded by $\mathcal{O}(n_j)$. Reverse enumeration of subspaces $\mathcal{V}_k$ in (2.18), or, equivalently, of subspaces V_l in (2.16) corresponds to 1 post–smoothing step. Combination of pre– and post–smoothing steps or W–cycles can be formulated in a similar way.

Our heuristic reasoning is justified by the following convergence result. For a proof we refer to Neuss [73].

Theorem 2.6. *There is $\rho < 1$ depending only on the shape regularity of $\mathcal{T}_0$, on the maximal gradient of α and on the constants α_0, α_1 in (2.4) such that*

$$\|u_j - u_j^{\nu+1}\| \le \rho \|u_j - u_j^{\nu}\| \qquad \forall \nu \ge 0 \tag{2.21}$$

and all $u_j^0 \in \mathcal{S}_j$.

In practical calculations we usually want to compute an approximate solution up to a certain accuracy. In this case, the required number of multigrid steps depends on the accuracy of the initial iterate as well. Multigrid steps on level j are more costly than on coarser grids. Hence, it seems reasonable to compute initial iterates by multigrid iterations on coarser levels. This process is called *nested iteration* (cf. Hackbusch [52]) or *full multigrid* (cf. Brandt [25]).

Algorithm 2.7. *(Multigrid with nested iteration)*

given: $\tilde{u}_0 \in \mathcal{S}_j$

for $k = 1, \dots, j$ *do:*
{
initial iterate: $u_k^0 = \tilde{u}_{k-1}$
ν_j^* *steps of multigrid:* $\tilde{u}_k = u_k^{\nu_j^*}$
}

final approximation: $\tilde{u}_j$

We should keep in mind that the exact finite element solution u_j only approximates the desired continuous solution up to a certain accuracy. Hence, nested iteration should be controlled in such a way that discretization accuracy is preserved by the resulting final iterate $\tilde{u}_j$.

Theorem 2.8. *Assume that the discretization error satisfies*

$$\|u - u_k\| \le c_1 h_k \ , \qquad k = 0, 1, \dots \ , \tag{2.22}$$

that

$$\tilde{u}_0 = u_0 \tag{2.23}$$

and that the stopping criterion

$$\|u_k - \tilde{u}_k\| \le \tfrac{\sigma}{2} \|u_k - u_k^0\| \ , \qquad k = 1, 2, \dots \ , \tag{2.24}$$

holds with constants c_1 and $\sigma < 1$ independent of k. Then there is C independent of j such that

$$\|u - \tilde{u}_j\| \le C h_j \ . \tag{2.25}$$

Proof. Exploiting the assumptions (2.24), (2.22), (2.23) and (2.9), we compute

$$\begin{aligned}
\|u - \tilde{u}_j\| &\le \|u - u_j\| + \|u_j - \tilde{u}_j\| \le \|u - u_j\| + \tfrac{1}{2}\sigma \|u_j - \tilde{u}_{j-1}\| \\
&\le (1 + \tfrac{1}{2}\sigma) \|u - u_j\| + \tfrac{1}{2}\sigma \|u - \tilde{u}_{j-1}\| \\
&\le (\tfrac{\sigma}{2})^j \|u - u_0\| + (1 + \tfrac{1}{2}\sigma) \textstyle\sum_{i=0}^{j-1} (\tfrac{\sigma}{2})^i \|u - u_{j-i}\| \\
&\le \tfrac{5}{2} c_1 (1 - \sigma)^{-1} h_j
\end{aligned}$$

so that (2.25) holds with $C = \frac{5}{2} c_1 (1 - \sigma)^{-1}$. $\square$

Of course, c_1 and therefore C depends on u. Recall that $u \in H \cap H^2(\Omega)$ is a sufficient condition for discretization accuracy (2.22). The exact finite element solution u_0 on the (hopefully) coarse grid $\mathcal{T}_0$ can be computed by a direct solver. In order to control the stopping criterion (2.24) we need *a posteriori estimates of the algebraic error* $\|u_k - u_k^\nu\|$. From the triangle inequality we get for ρ from Theorem 2.6

$$(1+\rho)^{-1}\|u_k^{\nu+1} - u_k^\nu\| \le \|u_k - u_k^\nu\| \le (1-\rho)^{-1}\|u_k^{\nu+1} - u_k^\nu\| \ . \tag{2.26}$$

As a consequence, multigrid corrections provide uniform lower and upper bounds of the algebraic error.

Corollary 2.9. *Computation of $\tilde{u}_j$ with $\mathcal{O}(h_j)$ accuracy by multigrid with nested iteration (cf. Theorem 2.8) requires $\mathcal{O}(n_j)$ point operations.*

Proof. The number of point operations required for a single multigrid step on level k is bounded by cn_k with c independent of k. As the convergence rate ρ is independent of j, the number ν^* that is required to satisfy the stopping criterion (2.24) or, equivalently, $\rho^{\nu^*} < \frac{1}{2}\sigma$ is also independent of j. Uniform refinement implies $n_{k+1} \ge qn_k$ with some fixed $q > 1$. Thus overall complexity of multigrid with nested iteration is bounded by

$$c\nu^* \sum_{k=1}^{j} n_k \le c\nu^*(1-q^{-1})^{-1} n_j = \mathcal{O}(n_j) \ .$$

This concludes the proof. □

Iterative schemes with properties as stated in Corollary 2.9 are sometimes called *optimal*. More sophisticated stopping criteria even provide optimality of nested Gauß-Seidel iteration. For additional information on such *cascadic multigrid methods*, we refer to Bornemann and Deuflhard [13].

Recall that approximation of the continuous solution u up to a certain accuracy is our ultimate goal in most practical applications. In this case the final level j should not be fixed a priori. Nested iteration combined with successive refinement should be continued until the desired accuracy is reached. This procedure requires *a posteriori estimates of the discretization error*. We refer e.g. to the monographs of Verfürth [86] and Ainsworth and Oden [2] for further information.

2.3. **Concluding remarks.** At first sight, our considerations seem to be more complicated than classical motivation of multigrid (cf. e.g. Hackbusch [51], pp. 17). However, the approach via successive subspace correction can be directly extended to more complicated minimization problems which will play a crucial role in section 4.2.

However, we used a very intuitive notion of frequencies. Analytically, definition (2.18) of subspaces $\mathcal{V}_k$ is motivated by the property

$$Q_0\mathcal{S}_j = \mathcal{V}_0, \qquad (Q_k - Q_{k-1})\mathcal{S}_j \subset \mathcal{V}_k, \quad k = 1, \dots, j, \tag{2.27}$$

where Q_k denotes the L^2–projection on $\mathcal{S}_k$. In order to guarantee (2.27) for adaptively refined grids, it is sufficient to choose the subspace

$$\mathcal{V}_k = \operatorname{span}\{\Lambda_k \setminus \Lambda_{k-1}\} \subset \mathcal{S}_k$$

spanned only by *new* nodal basis functions. If subspaces $\mathcal{V}_k = \mathcal{S}_k$ were selected in case of strong local refinement, then optimal complexity of multigrid V–cycle would get lost.

Property (2.27) can be reinterpreted in terms of frequencies. It is well–known (cf. e.g. Bramble and Xu [23]) that

$$\|v - Q_k v\|_{L^2(\Omega)} \le C h_k \|v\| \qquad \forall v \in \mathcal{S}_j \tag{2.28}$$

holds with C independent of k and j. As a consequence of (2.28) and an inverse inequality, all functions $v \in (Q_k - Q_{k-1})^2 \mathcal{S}_j \subset (Q_k - Q_{k-1})\mathcal{S}_j \subset \mathcal{V}_k$ have the property

$$C^{-2} h_k^{-2} \le \frac{a(v,v)}{(v,v)} \le c h_k^{-2} \; ,$$

where $(\cdot,\cdot)$ denotes the scalar product in L^2. Hence, the scale of subspaces $\mathcal{V}_k$ still reflects a scale of frequencies. Relation of high frequencies and locality is discussed to some extend in a recent survey by Xu [90].

Convergence properties of general successive subspace correction methods (cf. Algorithm 2.4) can be analyzed in an abstract framework as developed by Bramble, Pasciak, Wang and Xu [22, 21], Bramble and Pasciak [20], Dryja and Widlund [39], Xu [87] and others. The underlying arguments were partly anticipated by Yserentant [91] in the special case of hierarchical splittings

$$\mathcal{V}_0 = I_0 \mathcal{S}_j , \qquad \mathcal{V}_k = (I_k - I_{k-1})\mathcal{S}_j , \quad k = 1, \ldots, j. \tag{2.29}$$

Here $I_k : \mathcal{S}_j \to \mathcal{S}_k$ denotes nodal interpolation.

Using this abstract theory, the most crucial point in the proof of Theorem 2.6 is to show that the splitting (2.18) is stable. This means that each $v \in \mathcal{S}_j$ can be decomposed into a sum $v = v_0 + v_1 + \cdots + v_j$ of $v_k \in \mathcal{V}_k$ in such a way that

$$\sum_{k=0}^{j} 4^k \|v_k\|^2_{L^2(\Omega)} \le K_1 \|v\|^2 \tag{2.30}$$

holds with some K_1 independent of j. Note that approximation property (2.28) together with H^1–stability of Q_0 provide (2.30) with K_1 growing linearly in j. Utilizing the equivalence of norms in certain Besov and Sobolev spaces, (2.30) was shown by Oswald [75] and Dahmen and Kunoth [30]. See Bornemann and Yserentant [15] especially for the adaptive case and Bramble and Pasciak [20] or Xu [87] for related results. In contrast to the classical multigrid convergence proofs of Hackbusch [51] and Braess and Hackbusch [17] *no additional regularity of u is required* for (2.30) and resulting mesh–independent convergence (2.21). On the other hand, we get no information how multiple smoothing would improve the convergence rate.

Note that stability of hierarchical splitting (2.29) deteriorates quadratically in two space dimensions and exponentially in the three–dimensional case. As a consequence, the convergence rates of hierarchical basis multigrid deteriorate quadratically and exponentially for two and three space dimensions, respectively (cf. Yserentant [91], Bank, Dupont and Yserentant [7], Deuflhard, Leinen and Yserentant [34]). On the other hand, hierarchical decompositions have some advantages concerning complexity and robustness which make them competitive for certain two–dimensional problems.

For additional information on successive subspace correction and multigrid, we recommend the monograph of Bramble [19] and the surveys of Xu [87, 89, 90] and Yserentant [92]. Relation of subspace correction and domain decomposition is treated by Smith, Bjørstad and Gropp [83].

3. Smooth nonlinear problems

3.1. **Continuous problem and discretization.** We consider the minimization problem

$$u \in H: \qquad \mathcal{E}(u) \le \mathcal{E}(v) \qquad \forall v \in H. \tag{3.1}$$

Recall that $H = H_0^1(\Omega)$. The energy functional $\mathcal{E}$ is given by

$$\mathcal{E}(v) = \mathcal{J}(v) + \phi(v) \tag{3.2}$$

with quadratic functional $\mathcal{J}$ as defined in (2.2) and

$$\phi(v) = \int_\Omega \Phi(v(x))\, dx. \tag{3.3}$$

Throughout this section we assume that $\Phi: \mathbb{R} \to \mathbb{R}$ is *convex*, i.e.

$$\Phi(\omega\xi + (1-\omega)\eta) \le \omega\Phi(\xi) + (1-\omega)\Phi(\eta) \qquad \forall \omega \in [0,1], \quad \xi, \eta \in \mathbb{R}, \tag{3.4}$$

and *smooth*, i.e.

$$\Phi \in C^3(\mathbb{R}), \qquad |\Phi'''(\xi)| \le \text{ const.} \quad \forall \xi \in \mathbb{R}. \tag{3.5}$$

The following Sobolev embedding will be very useful. For a proof, we refer to Adams [1], Theorem 6.2, p. 144.

Lemma 3.1. *The embedding*

$$H \to L^p(\Omega) \tag{3.6}$$

is compact for $1 \le p < \infty$ *if* $d = 2$ *and for* $1 \le p < 6$ *if* $d = 3$.

It follows from (3.4) that ϕ is convex. The growth condition in (3.5) and the continuous embedding (3.6) imply $|\phi(v)| \le C(1 + \|v\|^3)$ $\forall v \in H$ with suitable $C \in \mathbb{R}$ (cf. Reusken [80]). Hence, ϕ is *proper* (i.e. $\phi(v) > -\infty$ and $\phi \not\equiv +\infty$). Moreover, ϕ is locally Lipschitz continuous (cf. Ekeland and Temam [41], p. 12) and therefore *lower semicontinuous* (i.e. $\liminf_{\|v - v_i\| \to 0} \phi(v_i) \ge \phi(v)$ $\forall v \in H$). Finally, it is well–known (cf. [41], p. 14) that there are constants c, $C \in \mathbb{R}$ such that $\phi(v) \ge c\|v\| + C$ $\forall v \in H$.

Theorem 3.2. *The minimization problem (3.1) admits a unique solution.*

Proof. As a consequence of the above considerations the energy functional $\mathcal{E}$ is strictly convex, lower semicontinuous, proper and coercive. Hence, the assertion follows from Proposition 1.2 in [41], p. 35. □

Smoothness (3.5) of Φ implies that ϕ is twice Fréchèt differentiable on H. The derivative $\phi'(w^*)$ at $w^* \in H$ is given by

$$\phi'(w^*)(v) = \int_\Omega \Phi'(w^*(x))v(x)\, dx, \qquad v \in H, \tag{3.7}$$

and the second derivative $\phi''(w^*)$ takes the form

$$\phi''(w^*)(v, w) = \int_\Omega \Phi''(w^*(x))v(x)w(x)\, dx, \qquad v, w \in H.$$

ϕ'' is uniformly Lipschitz continuous, i.e.

$$|\phi''(w_1)(v, w) - \phi''(w_2)(v, w)| \le L\|w_1 - w_2\|\|v\|\|w\| \tag{3.8}$$

holds for all $w_1, w_2, v, w \in H$. Note that the Lipschitz constant L is invariant with respect to transformations $\tilde{v} = Bv$ of (3.1) by isomorphisms $B : H \to H$. As a consequence of convexity and the growth condition in (3.5) there is $C \in \mathbb{R}$ with

$$0 \le \phi''(w)(v,v) \le C(1 + \|w\|)\|v\|^2 \qquad \forall v, w \in H \; . \tag{3.9}$$

Proposition 3.3. *The variational problem*

$$u \in H : \qquad a(u,v) + \phi'(u)(v) = \ell(v) \qquad \forall v \in H \tag{3.10}$$

is equivalent to (3.1).

Proof. The assertion follows from smoothness and convexity of $\mathcal{E}$. □

Note that (3.10) can be regarded as weak formulation of the semilinear differential equation

$$-\nabla \cdot (\alpha(x)\nabla u(x)) + \Phi'(u(x)) = f(x) \qquad \forall x \in \Omega$$

with homogeneous boundary conditions.

In order to discretize (3.10), we define the approximate energy

$$\mathcal{E}_j(v) = \mathcal{J}(v) + \phi_j(v), \qquad v \in \mathcal{S}_j,$$

where

$$\phi_j(v) = \sum_{p \in \mathcal{N}_j} \Phi(v(p))\; h_p, \qquad h_p = \int_\Omega \lambda_p^{(j)}\; dx, \tag{3.11}$$

is resulting from $\mathcal{S}_j$–interpolation of $\Phi(v)$. The discrete analogue ϕ_j inherits the properties of of ϕ. In particular ϕ_j is convex, lower semicontinuous, proper and smooth. Derivatives at some $w^* \in \mathcal{S}_j$ are given by

$$\phi_j'(w^*)(v) = \sum_{p \in \mathcal{N}_j} \Phi_j'(w^*(p))v(p)\; h_p, \qquad v \in \mathcal{S}_j, \tag{3.12}$$

and

$$\phi_j''(w^*)(v,w) = \sum_{p \in \mathcal{N}_j} \Phi''(w^*(p))v(p)w(p)\; h_p, \qquad v, w \in \mathcal{S}_j.$$

ϕ_j'' is uniformly Lipschitz continuous, i.e. there is $L \in \mathbb{R}$, not depending on j, such that

$$|\phi_j''(w_1)(v,w) - \phi_j''(w_2)(v,w)| \le L\|w_1 - w_2\|\|v\|\|w\| \tag{3.13}$$

holds for all $w_1, w_2, v, w \in \mathcal{S}_j$. The discrete analogue of (3.9) is

$$0 \le \phi_j''(w)(v,v) \le C(1 + \|w\|)\|v\|^2 \qquad \forall v, w \in \mathcal{S}_j \; . \tag{3.14}$$

The finite element discretization

$$u_j \in \mathcal{S}_j : \qquad \mathcal{E}_j(v) \le \mathcal{E}_j(v) \qquad \forall v \in \mathcal{S}_j \tag{3.15}$$

of (3.10) is finally obtained by minimization of discrete energy on the finite element space $\mathcal{S}_j$.

Theorem 3.4. *The discrete problem (3.15) has a unique solution and can be equivalently rewritten as the variational problem*

$$u_j \in \mathcal{S}_j : \qquad a(u_j, v) + \phi_j'(u_j)(v) = \ell(v) \qquad \forall v \in \mathcal{S}_j. \tag{3.16}$$

The approximate solutions u_j converge to u,

$$\|u - u_j\| \to 0, \qquad j \to \infty \; ,$$

and the discretization error can be estimated according to

$$\|u - u_j\| \le \inf_{v \in \mathcal{S}_j} \left(\left(2 + \|\phi''(u)\| + \frac{L}{2}\|u - v\|\right)\|u - v\| + \|\phi'(v) - \phi_j'(v)\| \right) . \tag{3.17}$$

Proof. Existence and uniqueness can be shown in the same way as Theorem 3.2. Convergence follows from a more general result stated in Theorem 6.2 in [46], p. 13. We now show (3.17). Let $v \in \mathcal{S}_j$. We clearly have

$$\|u - u_j\| \le \|u - v\| + \|u_j - v\|. \tag{3.18}$$

Exploiting (3.10) and (3.16) together with

$$(\phi_j'(u_j) - \phi_j'(v))(u_j - v) = \sum_{p \in \mathcal{N}_j} \Phi''(v(p) + s_p(u_j(p) - v(p)))(u_j(p) - v(p))^2 \; h_p \ge 0$$

which holds for suitable $s_p \in (0,1)$, we obtain

$$\begin{aligned} \|u_j - v\|^2 &\le a(u - v, u_j - v) + (\phi'(u) - \phi'(v))(u_j - v) \\ &\qquad\qquad + (\phi'(v) - \phi_j'(v))(u_j - v) \\ &\le \Big(\|u - v\| + \|\phi''(u)\|\|u - v\| + \frac{L}{2}\|u - v\|^2 \\ &\qquad\qquad + \|\phi'(v) - \phi_j'(v)\| \Big) \|u_j - v\| . \end{aligned} \tag{3.19}$$

In the second estimate, we have used

$$\begin{aligned} &|(\phi'(u) - \phi'(v))(w)| \\ &\le \left| \int_0^1 (\phi''(v + s(u - v)) - \phi''(u))(u - v, w) \; ds \right| + |\phi''(u)(u - v, w)| \\ &\le \frac{L}{2}\|u - v\|^2\|w\| + \|\phi''(u)\|\|u - v\|\|w\| . \end{aligned}$$

Combining (3.18) and (3.19) we get the assertion. □

Note that

$$\|\phi''(u)\| = \sup_{v,w \in H} \frac{|\phi''(u)(v,w)|}{\|v\|\|w\|} \le C(1 + \|u\|)$$

holds with constant C occurring in (3.9). Straightforward estimation of quadrature error gives

$$\|\phi'(I_j u) - \phi_j'(I_j u))\| \le c h_j^2 |u|^2_{W^{1,\infty}(\Omega)}$$

with c depending on the constant C from (3.14) and on the shape regularity of $\mathcal{T}_0$. Recall that $I_j u$ denotes the $\mathcal{S}_j$–interpolant of u. All together we get the usual error estimate $\|u - u_j\| = \mathcal{O}(h_j)$, if the solution u is smooth enough. For regularity results we refer e.g. to Brézis, Crandall and Pazy [28].

3.2. **Inexact Newton methods.** Choosing some basis, e.g. the nodal basis of $\mathcal{S}_j$, (3.16) can be rewritten as a nonlinear system with n_j equations for the n_j unknown coefficients of u_j. Taylor's expansion leads to well–known Newton's method (cf. Algorithm 3.5) which is expected to converge quadratically for sufficiently accurate initial iterates. A linear system has to be solved in each iteration step. Application to (3.16) (and to discretized partial differential equations in general) typically leads to large numbers of unknowns and ill conditioned coefficient matrices. This motivates *inexact Newton's method* (cf. Algorithm 3.6) involving iterative solution of linear systems up to a certain accuracy. Quadratic convergence can be preserved by increasing accuracy in course of Newton iteration. However, such strategies are still too costly in most cases so that we only strive for *local linear convergence* (cf. Corollary 3.8). Global convergence requires damping of (inexact) Newton corrections. Reduction of energy is a very powerful principle for deriving reliable *damping strategies* (cf. Theorem 3.11). Keeping in mind that we are dealing with a sequence of problems with increasing number of unknowns, any arguments relying on finite dimension of $\mathcal{S}_j$ should be avoided. This requires some care in selecting appropriate norms (and, unfortunately, some technicalities). Moreover, the number of parameters should be kept as small as possible to allow for effective implementation of theoretical results. It turns out that we need control of Lipschitz constant L from (3.13) and linear accuracy ρ from (3.27) to guarantee convergence with respect to local energy norms.

From now on, we write $F_j = \mathcal{E}_j' : \ \mathcal{S}_j \to \mathcal{S}_j'$ where $\mathcal{S}_j'$ is the dual of $\mathcal{S}_j$. Then (3.16) takes the form

$$u_j \in \mathcal{S}_j : \qquad F_j(u_j) = 0. \tag{3.20}$$

Using the duality pairing $\langle \cdot, \cdot \rangle$ between $\mathcal{S}_j$ and $\mathcal{S}_j'$, convexity of ϕ_j yields

$$\langle F_j(v) - F_j(w), v - w \rangle \geq \|v - w\|^2 \qquad \forall v, w \in \mathcal{S}_j \ . \tag{3.21}$$

so that F_j is a *strongly monotone* operator. Hence, F_j is continuously invertible (see e.g. Theorem 11.2 in Deimling [31], p. 100). The Fréchèt derivative F_j' generates the *local scalar product* $a_{w^*}(\cdot, \cdot)$,

$$a_{w^*}(v, w) = \langle F_j'(w^*)v, w \rangle = a(v, w) + \phi_j''(w^*)(v, w), \qquad v, w \in \mathcal{S}_j, \tag{3.22}$$

and the associated *local energy norm*

$$\|v\|_{w^*} = a_{w^*}(v, v)^{1/2}, \qquad v \in \mathcal{S}_j, \tag{3.23}$$

for each $w^* \in \mathcal{S}_j$. As a consequence of (3.21) and (3.14) we obtain the estimates

$$\|v\|^2 \leq \langle F_j'(w^*)v, v \rangle, \qquad |\langle F_j'(w^*)v, w \rangle| \leq C(1 + \|w^*\|)\|v\|\|w\| \tag{3.24}$$

for all $v, w, w^* \in \mathcal{S}_j$. Hence, $F_j'(w^*) : \mathcal{S}_j \to \mathcal{S}_j'$ is an isomorphism for each $w^* \in \mathcal{S}_j$ and the norms $\|\cdot\|$ and $\|\cdot\|_{w^*}$ are equivalent. Lipschitz continuity (3.13) implies

$$\begin{aligned} &|\langle (F_j'(w_1) - F_j'(w_2))v, w \rangle| \\ &\qquad\qquad \leq L\|w_1 - w_2\|\|v\|\|w\| \quad \forall w_1, w_2, v, w \in \mathcal{S}_j \ . \end{aligned} \tag{3.25}$$

In combination with (3.14) this leads to the norm equivalence

$$\|v\|_{w_1}^2 \leq (1 + L\|w_1 - w_2\|)\|v\|_{w_2}^2, \qquad \|v\|_{w_2}^2 \leq (1 + L\|w_1 - w_2\|)\|v\|_{w_1}^2 \tag{3.26}$$

for all v, w_1, $w_2 \in \mathcal{S}_j$.

We are ready to formulate well–known Newton's Method for smooth nonlinear problems.

Algorithm 3.5. *(Newton's Method)*
given: $u_j^\nu \in \mathcal{S}_j$

correction: $\Delta u_j^\nu = -F_j'(u_j^\nu)^{-1} F_j(u_j^\nu)$

new iterate: $u_j^{\nu+1} = u_j^\nu + \Delta u_j^\nu$

Newton corrections Δu_j^ν are only approximated up to a certain accuracy, if iterative linear solvers are applied. This motivates investigation of *inexact Newton methods* (cf. Bank and Rose [8], Dembo et al. [32]). Let B_j assign a linear mapping

$$B_j(w) : \mathcal{S}_j' \to \mathcal{S}_j$$

to each $w \in \mathcal{S}_j$. $B_j(w)$ is intended to approximate $F_j'(w)^{-1}$. More precisely, we assume that there is $\rho < 1$ such that

$$\|v - B_j(w)F_j'(w)v\|_w \le \rho\|v\|_w \qquad \forall v, w \in \mathcal{S}_j \ . \tag{3.27}$$

Hence, the associated linear iteration

$$\delta^{i+1} = (I - B_j(w)F_j'(w))\delta^i - B_j(w)F_j(w), \qquad i = 0, 1 \ldots, \tag{3.28}$$

converges with convergence rate ρ with respect to the norm $\|\cdot\|_w$. We consider the approximation

$$\delta u_j^\nu = -B_j(u_j^\nu)F_j(u_j^\nu) \ , \tag{3.29}$$

as resulting from one step of (3.28) with initial iterate $\delta^0 = 0$.

Algorithm 3.6. *(Inexact Newton's Method)*
given: $u_j^\nu \in \mathcal{S}_j$

approximate correction: $\delta u_j^\nu = -B_j(u_j^\nu)F_j(u_j^\nu)$

new iterate: $u_j^{\nu+1} = u_j^\nu + \delta u_j^\nu$

Let us first investigate *local convergence* of Algorithm 3.6.

Theorem 3.7. *Assume that condition (3.27) holds. Then the inexact Newton method satisfies*

$$\|u_j - u_j^{\nu+1}\|_{u_j^\nu} \le \rho\|u_j - u_j^\nu\|_{u_j^\nu} + \tfrac{1}{2}(1+\rho)L\|u_j - u_j^\nu\|_{u_j^\nu}^2 \tag{3.30}$$

for all $\nu \ge 0$. Moreover, there is a constant C_L depending only on L such that

$$\|u_j - u_j^{\nu+1}\|_{u_j} \le \left(\rho + \tfrac{1}{2}C_L(\|u_j - u_j^\nu\|_{u_j} + \|u_j - u_j^\nu\|_{u_j}^2)\right)\|u_j - u_j^\nu\|_{u_j} \tag{3.31}$$

holds for all $\nu \ge 0$.

Proof. Denote $e_{\nu+1} = u_j - u_j^{\nu+1}$ and $e_{\nu+1}^* = e_{\nu+1} - (\Delta u_j^\nu - \delta u_j^\nu) = e_\nu - \Delta u_j^\nu$. Exploiting (3.27) we get

$$\|\Delta u_j^\nu - \delta u_j^\nu\|_{u_j^\nu} \le \rho\|\Delta u_j^\nu\|_{u_j^\nu}.$$

Hence,

$$\begin{aligned}\|e_{\nu+1}\|_{u_j^\nu} &= \|e_{\nu+1}^* + (\Delta u_j^\nu - \delta u_j^\nu)\|_{u_j^\nu} \\ &\le \rho\|\Delta u_j^\nu\|_{u_j^\nu} + \|e_{\nu+1}^*\|_{u_j^\nu} \le \rho\|e_\nu\|_{u_j^\nu} + (1+\rho)\|e_{\nu+1}^*\|_{u_j^\nu}.\end{aligned}$$

We proceed by estimating $\|e^*_{\nu+1}\|_{u_j^\nu}$. Using $F_j(u_j) = 0$ and Lipschitz continuity (3.25), we compute

$$\begin{aligned}
\|e^*_{\nu+1}\|_{u_j^\nu} &= \|e_\nu + F_j'(u_j^\nu)^{-1}(F_j(u_j^\nu) - F_j(u_j))\|_{u_j^\nu} \\
&= \|\int_0^1 F_j'(u_j^\nu)^{-1}(F_j'(u_j - se_\nu) - F_j'(u_j^\nu))e_\nu \, ds\|_{u_j^\nu} \\
&= \sup_{\|v\|_{u_j^\nu}=1} \left| \int_0^1 \langle F_j'(u_j - se_\nu) - F_j'(u_j^\nu))e_\nu, v\rangle \, ds \right| \\
&\le \tfrac{1}{2} L \|e_\nu\|^2_{u_j^\nu}.
\end{aligned}$$

This proves (3.30). Norm equivalence (3.26) yields

$$\|e_{\nu+1}\|^2_{u_j} \le (1 + L\|e_\nu\|_{u_j^\nu})\|e_{\nu+1}\|^2_{u_j^\nu}, \qquad \|e_\nu\|^2_{u_j^\nu} \le (1 + L\|e_\nu\|_{u_j})\|e_\nu\|^2_{u_j}.$$

In combination with (3.30), this leads to

$$\|e_{\nu+1}\|_{u_j} \le \left(\rho + 2L(1+L)^{\frac{1}{2}}\|e_\nu\|_{u_j} + L^2(1+L)\|e_\nu\|^2_{u_j}\right) \|e_\nu\|_{u_j} \ .$$

This concludes the proof. □

As an important consequence of Theorem 3.7 we get *local linear convergence* of inexact Newton's method.

Corollary 3.8. *Let $0 < \varepsilon \le 1$ satisfy*

$$\rho + C_L \varepsilon \le \rho_0 < 1$$

and assume $\|u_j - u_j^0\|_{u_j} \le \varepsilon$. Then

$$\|u_j - u_j^{\nu+1}\|_{u_j} \le \rho_0 \|u_j - u_j^\nu\|_{u_j} \qquad \forall \nu \ge 0 \ . \tag{3.32}$$

Proof. The assertion is an immediate consequence of (3.31) □

Observe that the actual error reduction in each step tends to ρ for $\nu \to \infty$. *Local quadratic convergence* of Newton's method is recovered for $\rho = 0$. Note that local quadratic convergence holds for the inexact variant as well, if $\rho = \rho^\nu = \mathcal{O}(\|u_j - u_j^\nu\|)$.

Global convergence can be achieved by damping of the inexact Newton correction δu_j^ν. The choice of suitable damping parameters will be based on *energy reduction.*

Theorem 3.9. *Let $u_j^0 \in \mathcal{S}_j$. For $\nu \ge 0$ compute δu_j^ν from (3.29), select a damping parameter $\omega_j^\nu \in (0, 1]$ and define*

$$u_j^{\nu+1} = u_j^\nu + \omega_j^\nu \delta u_j^\nu \ . \tag{3.33}$$

Assume that damping parameters ω_j^ν are chosen such that

$$\mathcal{E}_j(u_j^\nu + \omega_j^\nu \delta u_j^\nu) \le \mathcal{E}_j(u_j^\nu) - p_0 \|\delta u_j^\nu\|^2_{u_j^\nu} \qquad \forall \nu \ge 0 \tag{3.34}$$

holds with $p_0 > 0$ independent of ν. Then $u_j^\nu \to u_j$ as $\nu \to \infty$.

Proof. As a consequence of (3.34), we have

$$\mathcal{E}_j(u_j) \le \mathcal{E}_j(u_j^\nu) \le \mathcal{E}_j(u_j^0) \qquad \forall \nu \ge 0 \ .$$

Hence, the sequence $(\mathcal{E}_j(u_j^\nu))_{\nu \ge 0}$ must converge. In combination with energy reduction (3.34) this leads to

$$\mathcal{E}_j(u_j^\nu) - \mathcal{E}_j(u_j^{\nu+1}) \ge p_0 \|\delta u_j^\nu\|^2_{u_j^\nu} \to 0 \qquad \nu \to \infty \ .$$

Utilizing (3.27) we compute

$$\|\delta u_j^\nu\|_{u_j^\nu} = \|B_j(u_j^\nu)F_j'(u_j^\nu)\Delta u_j^\nu\|_{u_j^\nu} \geq (1-\rho)\|\Delta u_j^\nu\|_{u_j^\nu} \ .$$

In the light of (3.24) we obtain $\|\Delta u_j^\nu\| \to 0$ as $\nu \to \infty$. The second inequality in (3.24) implies that $F_j'(u_j^\nu) : \mathcal{S}_j \to \mathcal{S}_j'$ is bounded, because

$$\|F_j'(u_j^\nu)\|_{\mathcal{S}_j \to \mathcal{S}_j'} = \sup_{\substack{v,w \in \mathcal{S}_j \\ v,w \neq 0}} \frac{|\langle F_j'(u_j^\nu)v, w\rangle|}{\|v\|\|w\|} \leq C(1 + \|u_j^\nu\|) \ .$$

Hence,

$$\|\Delta u_j^\nu\| = \|F_j'(u_j^\nu)^{-1}F_j(u_j^\nu)\| \geq (C(1+\|u_j^\nu\|))^{-1}\|F_j(u_j^\nu)\|_{\mathcal{S}_j'}$$

where $\|\cdot\|_{\mathcal{S}_j'}$ denotes the dual norm of $\mathcal{S}_j$ equipped with the energy norm $\|\cdot\|$. Recall that $\|u_j^\nu\|$ is bounded. Hence, $F_j(u_j^\nu) \to 0$ as $\nu \to \infty$. As F_j is continuously invertible, we conclude that $u_j^\nu \to F_j^{-1}(0) = u_j$ as $\nu \to \infty$. □

We still have to investigate whether it is possible to find damping parameters ω_j^ν with property (3.34) or not. The answer is given by the following Lemma.

Lemma 3.10. *The energy estimate*

$$\mathcal{E}_j(u_j^\nu + \omega\delta u_j^\nu) \leq \mathcal{E}_j(u_j^\nu) - p_j^\nu(\omega)\|\delta u_j^\nu\|_{u_j^\nu}^2, \qquad \omega > 0, \tag{3.35}$$

holds with

$$p_j^\nu(\omega) = (1-\delta_\nu)\omega - \tfrac{1}{2}\omega^2 - \tfrac{1}{6}\omega^3 L\|\delta u_j^\nu\|_{u_j^\nu}$$

where

$$\delta_\nu = \frac{\|\Delta u_j^\nu - \delta u_j^\nu\|_{u_j^\nu}}{\|\delta u_j^\nu\|_{u_j^\nu}} \ .$$

Proof. Elementary integration yields

$$\begin{aligned} \omega^2 \int_0^1 \int_0^1 & s\langle (F_j'(u_j^\nu + st\omega\delta u_j^\nu) - F_j'(u_j^\nu))\delta u_j^\nu, \delta u_j^\nu\rangle \ dt \ ds \\ &= \mathcal{E}_j(u_j^\nu + \omega\delta u_j^\nu) - \mathcal{E}_j(u_j^\nu) - \omega\langle F_j(u_j^\nu), \delta u_j\rangle - \tfrac{1}{2}\omega^2\|\delta u_j^\nu\|_{u_j^\nu}^2 . \end{aligned}$$

By definition of Newton's method we have

$$\langle F_j(u_j^\nu), \delta u_j\rangle = -\langle F_j'(u_j^\nu)(\Delta u_j^\nu - \delta u_j^\nu), \delta u_j^\nu\rangle - \|\delta u_j^\nu\|_{u_j^\nu}^2 .$$

After arranging terms and using Cauchy–Schwarz inequality and Lipschitz continuity (3.25) the assertion follows. □

Observe that

$$\delta_\nu \leq \delta_* = \frac{\rho}{1-\rho} \ . \tag{3.36}$$

As a consequence of Lemma 3.10, energy is decreasing, if $\delta_\nu < 1$ and $\omega \in (0, \omega_j^{\nu,\max})$ with

$$\omega_j^{\nu,\max} = \frac{4(1-\delta_\nu)}{1+\sqrt{1+\frac{8}{3}L\|\delta u_j^\nu\|_{u_j^\nu}(1-\delta_\nu)}} \leq 2 \ . \tag{3.37}$$

Maximal energy reduction is obtained for

$$\omega_j^{\nu,\mathrm{opt}} = \frac{2(1-\delta_\nu)}{1+\sqrt{1+2L\|\delta u_j^\nu\|_{u_j^\nu}(1-\delta_\nu)}} \leq 1 \ . \tag{3.38}$$

A widely used strategy to select damping factors ω_j^ν is simply to reduce some initial guess (e.g. by repeated bisection) until the *monotonicity test*

$$\mathcal{E}_j(u_j^\nu) - \mathcal{E}_j(u_j^\nu + \omega_j^\nu \delta_{u_j^\nu}) \ge q\omega_j^\nu \|\delta u_j^\nu\|_{u_j^\nu}^2 \tag{3.39}$$

is fulfilled with fixed $q \in (0, 1-\delta_*)$. Utilizing (3.35), it can be shown that (3.39) must be satisfied if $\omega_j^\nu < \omega_j^*$ holds with suitable ω_j^* independent of ν.

We now describe a more sophisticated damping strategy which guarantees global convergence together with local linear convergence (3.32).

Theorem 3.11. *Assume that condition (3.27) holds with $\rho < \frac{1}{3}$ so that*

$$2(1-\delta_*) = 2\frac{1-2\rho}{1-\rho} > 1\ . \tag{3.40}$$

Choose damping parameters ω_j^ν according to

$$\omega_j^\nu = \min\{1, \sigma\omega_j^{\nu,\max}\} \qquad \forall \nu \ge 0 \tag{3.41}$$

where $\sigma \in (0,1)$ satisfies $2\sigma(1-\delta_*) > 1$. Then damped inexact Newton iteration (3.33) converges to u_j for each initial iterate $u_j^0 \in \mathcal{S}_j$. Moreover, there is $\nu_0 > 0$ such that $\omega_j^\nu = 1$ and linear convergence (3.31) holds for all $\nu \ge \nu_0$.

Proof. As a consequence of (3.36) with $\rho < \frac{1}{3}$ we have $\delta_\nu < \frac{1}{2}$. Hence, (3.41) yields

$$\mathcal{E}_j(u_j) \le \mathcal{E}_j(u_j^\nu) \le \mathcal{E}_j(u_j^0) \qquad \forall \nu \ge 0$$

so that the sequence $(\mathcal{E}_j(u_j^\nu))_{\nu\ge 0}$ is bounded. Coercivity of $\mathcal{E}_j$ implies that u_j^ν is also bounded. In the light of (3.27) and (3.24), this leads to

$$\begin{aligned}\|\delta u_j^\nu\|_{u_j^\nu} &= \|B_j(u_j^\nu)F_j'(u_j^\nu)F_j'(u_j^\nu)^{-1}F_j(u_j^\nu)\|_{u_j^\nu} \\ &\le C(\rho+1)(1+\|u_j^\nu\|)^{\frac{1}{2}}\|F_j'(u_j^\nu)^{-1}F_j(u_j^\nu)\| \\ &\le C(\rho+1)(1+\|u_j^\nu\|)^{\frac{1}{2}}\|F_j(u_j^\nu)\|_{\mathcal{S}_j'} \le \text{const.}\end{aligned}$$

Hence, $(\delta_\nu, \|\delta u_j^\nu\|_{u_j^\nu})_{\nu\ge 0} \subset Q$ holds with a compact subset $Q \subset \mathbb{R}_+^2$. The function $g(\delta_\nu, \|\delta u_j^\nu\|_{u_j^\nu}) = p_j^\nu(\omega_j^\nu)$ is continuous and nonzero on Q. Thus

$$g(\delta_\nu, \|\delta u_j^\nu\|_{u_j^\nu}) = p_j^\nu(\omega_j^\nu) \ge p_0 > 0 \qquad \forall \nu \ge 0 \tag{3.42}$$

holds with suitable $p_0 > 0$ independent of ν. Now condition (3.34) follows from (3.35) and Theorem 3.9 provides convergence of u_j^ν.

As a consequence of $\delta u_j^\nu \to 0$ we have $\sigma\omega_j^{\nu,\max} \ge 2\sigma(1-\delta_*) > 1$ for $\nu \to \infty$. Hence, $\omega_j^\nu = 1$ for sufficiently large ν. Now local linear convergence follows from Corollary 3.8. □

If $B_j(u_j^\nu)$ stands for $\kappa > 0$ steps of a certain iterative scheme with convergence rate ρ_0, then κ can be selected in such a way that $\rho = \rho_0^\kappa < \frac{1}{3}$. κ can be controlled numerically by a posteriori estimation of the linear algebraic error. Numerical approximation of $\omega^{\nu,\max}$ requires estimates of the Lipschitz constant L (see e.g. Deuflhard [33] for a detailled discussion). A related damping strategy based on successive approximation of $L\|\delta u_j^\nu\|_{u_j}$ was presented by Deuflhard and Weiser [37].

3.3. **Newton Multigrid with Nested Iteration.** An inexact Newton method involving a certain number of multigrid steps as inexact linear solver $B_j(w)$ is called *Newton multigrid method.* Theorem 3.7 provides mesh independent local linear convergence of Newton multigrid. This property turns out to be crucial for optimal numerical complexity of Newton multigrid with nested iteration (cf. Algorithm 3.12 and Theorem 3.13).

For a formal definition of inexact solver $B_j(w)$ observe that multigrid Algorithm 2.5 as applied to linear problems of the form

$$u_j^* \in \mathcal{S}_j : \qquad a_w(u_j^*, v) = \ell_w(v) \qquad \forall v \in \mathcal{S}_j \tag{3.43}$$

assigns a correction $MG_j(w)r$ to a given residual $r = \ell_w - a_w(u_j^\nu, \cdot) \in \mathcal{S}_j'$. In this way a given multigrid method for (3.43) generates a linear mapping

$$MG_j(w) : \mathcal{S}_j' \to \mathcal{S}_j \ .$$

The corresponding mapping associated with $\kappa \geq 1$ multigrid steps is denoted by

$$(MG_j(w))^\kappa : \mathcal{S}_j' \to \mathcal{S}_j \ . \tag{3.44}$$

Observe that (3.24) and (2.4) provide the norm equivalence

$$\alpha_0 \|v\|_{H^1(\Omega)} \leq \|v\|_w \leq \alpha_1 C(1 + \|w\|)^{\frac{1}{2}} \|v\|_{H^1(\Omega)} \qquad \forall v \in \mathcal{S}_j \ .$$

In analogy to Theorem 2.6 we assume that

$$\|v - MG_j(w)F_j'(w)v\|_w \leq \rho_w \|v\|_w \qquad \forall v \in \mathcal{S}_j$$

is valid with $\rho_w < 1$ for each fixed $w \in \mathcal{S}_j$ and that

$$\rho_w \leq \rho_c < 1 \qquad \forall w \in \mathcal{S}_j, \ \|w\| \leq c \tag{3.45}$$

holds with ρ_c independent of j. Choose fixed $\rho \in (0,1)$, and let κ_w denote the smallest integer with $\rho_w^{\kappa_w} \leq \rho$. We set

$$B_j(w) = (MG_j(w))^{\kappa_w} \qquad w \in \mathcal{S}_j \ .$$

Hence, $B_j(w)$ satisfies condition (3.27) and the number κ_w of inner multigrid iterations is uniformly bounded on bounded subsets of $\mathcal{S}_j$. In practical computations κ_w can be selected automatically utilizing a posteriori error estimates for linear multigrid iteration (cf. e.g. (2.26)).

As a consequence of Corollary 3.8 the resulting Newton multigrid method converges linearly with mesh independent convergence rate, if the initial iterate u_j^0 is sufficiently accurate. Exploiting the hierarchy $\mathcal{S}_0 \subset \cdots \subset \mathcal{S}_j$ the problem to find such initial iterates can be traced back to coarser spaces. Starting with appropriate $\tilde{u}_0 \in \mathcal{S}_0$, initial iterates on subsequent levels are obtained by interpolating the final approximation from the preceding level (cf. Algorithm 2.7).

Algorithm 3.12. *(Newton multigrid with nested iteration)*

given: $\tilde{u}_0 \in \mathcal{S}_j$

for $k = 1, \ldots, j$ *do:*
{
initial iterate: $u_k^0 = \tilde{u}_{k-1}$
ν_j^* *steps of damped Newton multigrid:* $\tilde{u}_k = u_k^{\nu_j^*}$
}

final approximation: $\tilde{u}_j$

Convergence of Algorithm 3.12 will be analyzed in terms of the continuous local energy norm

$$\|v\|_u^2 = a(v,v) + \phi''(u)(v,v) \qquad v \in H \ .$$

From now on, we assume that

$$\left| \|v\|_u^2 - \|v\|_{u_j}^2 \right| \le c h_j^\gamma \|v\|^2 \qquad \forall v \in \mathcal{S}_j \tag{3.46}$$

holds with c, γ independent of j. For example (3.46) is valid with $\gamma = 1$, if $\Phi \in C^4(\mathbb{R})$ with $|\Phi^{(4)}(x)| <$ const. or if no lumping is applied, i.e. $\phi_j = \phi$.

Theorem 3.13. *Assume that the discretization error satisfies*

$$\|u - u_k\|_u \le c_1 h_k \ , \qquad k = 0, 1, \dots \ , \tag{3.47}$$

that the initial guess $\tilde{u}_0 \in \mathcal{S}_0$ *has the property*

$$\|u_0 - \tilde{u}_0\|_u \le c_0 h_0 \ , \tag{3.48}$$

that the damped Newton multigrid iteration is convergent on each level and that the stopping criterion

$$\|u_k - \tilde{u}_k\|_u \le \tfrac{\sigma}{2} \|u_k - u_k^0\|_u \ , \qquad k = 1, 2, \dots \ , \tag{3.49}$$

is fulfilled where all constants c_0, c_1 *and* $\sigma < 1$ *are assumed to be independent of* k. *Then there is* C *independent of* j *such that*

$$\|u - \tilde{u}_j\|_u \le C h_j \ . \tag{3.50}$$

The number $\kappa_{u_k^\nu}$ *of inner multigrid steps is uniformly bounded, i.e.*

$$\kappa_{u_k^\nu} \le \kappa \qquad \forall \nu \ge 0, \ \ k = 1, 2, \dots \ , \tag{3.51}$$

holds with κ *independent of* ν *and* j. *Moreover, if* $\omega_j^\nu = 1$ *is valid for sufficiently large* j, *then there is* j_0 *such that*

$$\nu_j^* = \nu^* \qquad \forall j \ge j_0 \ . \tag{3.52}$$

holds with ν^* *independent of* j.

Proof. Quasioptimal accuracy (3.50) with

$$C = \tfrac{5}{2} c_1 (1 - \sigma)^{-1} + c_0$$

follows almost literally in the same way as in the linear case (cf. Theorem 2.8).

As u_k converge to u and damped Newton multigrid iteration is convergent on each level k, the iterates u_k^ν are uniformly bounded for all $\nu \ge 0$ and $k \ge 1$. Hence, (3.51) follows from (3.45).

We still have to show (3.52). Let us first state that for any $\varepsilon > 0$ there is $j_0 = j_0(\varepsilon)$ such that

$$\begin{aligned} \|u_j - u_j^0\|_{u_j} &\le (1 + c h_j^\gamma)^{\frac{1}{2}} \|u_j - u_j^0\|_u \\ &\le (1 + c h_j^\gamma)^{\frac{1}{2}} (\|u_j - u\|_u + \|u - \tilde{u}_{j-1}\|_u) \\ &\le (c_1 + 2C)(1 + c h_j^\gamma)^{\frac{1}{2}} h_j \le \varepsilon \qquad \forall j \ge j_0 \ . \end{aligned} \tag{3.53}$$

This is a consequence of (3.46), (3.47) and (3.50). Now choose ε and j_0 in such a way that $\omega_j^\nu = 1$ and (3.53) hold together with

$$(1 + c h_j^\gamma)(\rho + C_L \varepsilon) \le \rho_0 < 1 \qquad \forall j \ge j_0 \ . \tag{3.54}$$

Recall that C_L appears in Theorem 3.7. Utilizing (3.54) and local linear convergence as stated in Corollary 3.8, we get

$$\|u_j - u_j^{\nu+1}\|_u \le \rho_0 \|u_j - u_j^\nu\|_u \qquad \forall \nu \ge 0 \quad \forall j \ge j_0 \ . \tag{3.55}$$

Now choose ν^* such that

$$\rho_0^{\nu^*} \le \tfrac{\sigma}{2} \ .$$

Then $\tilde{u}_j = u_j^{\nu^*}$ satisfies the stopping criterion (3.49). This proves the assertion. □

Utilizing (3.54) and (3.55) ρ can be chosen in such a way that even $\nu_j^* = 1$ holds for sufficiently large j. However, in practical computations this would require knowledge of L. Similarly, one could prove convergence in other equivalent norms such as the energy norm $\|\cdot\|$. However, in practical computations appropriate choice of ρ would require knowledge of the corresponding constants. The norm $\|\cdot\|_u$ can be approximated by $\|\cdot\|_{u_k^\nu}$.

In the light of Theorem 3.4 discretization accuracy (3.47) is a consequence of regularity $u \in H^2(\Omega) \cap W^{1,\infty}(\Omega)$. A posteriori estimates of the discretization error have been considered e.g. by Verfürth [86]. Any initial iterate $\tilde{u}_0 \in \mathcal{S}_0$ satisfies (3.48) with $c_0 = \|u_0 - \tilde{u}_0\|_u h_0^{-1}$. Implementation of the stopping criterion (3.49) requires a posteriori estimates for the algebraic error $\|u_k - u_k^\nu\|_u$. The straightforward approximation $\|u_k^{\nu+1} - u_k^\nu\|_{u_k^{\nu+1}} \approx \|u_k - u_k^\nu\|_u$ works well if h_k is sufficiently small but might be misleading on coarser grids. In order to guarantee (3.52), we finally need a damping strategy providing $\omega_j^\nu = 1$ whenever $\|u_j - u_j^\nu\|_u$ is sufficiently small ($\omega_j^\nu \ge \omega_0 > 0$ would be sufficient). Two examples were given in the preceding section.

Property (3.52) leads to optimal numerical complexity of Newton multigrid with nested iteration.

Corollary 3.14. *Computation of $\tilde{u}_j$ with $\mathcal{O}(h_j)$ accuracy by Newton multigrid with nested iteration (cf. Theorem 3.13) requires $\mathcal{O}(n_j)$ point operations.*

Proof. The number of point operations required for a single multigrid step on level k is bounded by cn_k with c independent of k. In the light of (3.51), complexity of Newton multigrid iteration on level k is bounded by $c\nu_k^* \kappa n_k$. Uniform refinement implies $n_{k+1} \ge q n_k$ with some fixed $q > 1$. As a consequence of (3.52), we have

$$\nu_j^* \le \nu^{\max} = \max_{k=1,\dots,j_0} \nu_k^* \qquad \forall j \ge 1 \ .$$

Thus overall complexity is bounded by

$$c\nu^{\max}\kappa \sum_{k=1}^{j} n_k \le c\nu^{\max}\kappa (1-q^{-1})^{-1} n_j = \mathcal{O}(n_j) \ .$$

□

There are several multigrid methods $MG_j(w)$ that satisfy convergence condition (3.45). For example, cg–iteration preconditioned by a multigrid V–cycle with symmetric Gauß–Seidel smoothing would have this property. Without additional assumptions it can be shown that multigrid V–cycle with Gauß–Seidel smoothing (cf. Algorithm 2.5) as applied to linear problems of the form (3.43) satisfies condition (3.45) with suboptimal $\rho_c = 1 - \mathcal{O}((1+j)^{-2})$. In this case, Corollary 3.14 holds with suboptimal numerical complexity behaving like $\mathcal{O}((1+j)^2 n_j)$.

3.4. **Nonlinear multigrid and FAS.** There are various ways to avoid explicit differentiation of F_j, e.g. by using finite differences (cf. e.g. Kelley [62], p. 79). *Nonlinear multigrid* (cf. Hackbusch [52], pp. 183) avoids explicit evaluation of derivatives F_j' by multigrid techniques. Consider the *modified Newton correction*

$$-F_j'(\tilde{u})^{-1}F_j(u_j^\nu) = F_j'(\tilde{u})^{-1}(r_j) \tag{3.56}$$

with suitable $\tilde{u} \in \mathcal{S}_j$ and residual

$$r_j = F_j(u_j) - F_j(u_j^\nu) \ .$$

Recall the identity

$$F_j'(\tilde{u})^{-1} = (F_j^{-1})'(F_j(\tilde{u})) \ .$$

Then, by definition of the directional derivative, we have

$$\|(F_j^{-1})'(F_j(\tilde{u}))(r_j) - \tfrac{1}{s}(F_j^{-1}(F_j(\tilde{u}) + sr_j) - \tilde{u})\| = o(1), \qquad s \to 0 \ .$$

Altogether we get the secant approximation

$$-F_j'(\tilde{u})^{-1}F_j(u_j^\nu) \approx \frac{1}{s}(F_j^{-1}(F_j(\tilde{u}) + sr_j) - \tilde{u}) \tag{3.57}$$

with equality for $s \to 0$. The resulting iteration reads

$$u_j^{\nu+1} = u_j^\nu + \tfrac{1}{s}(w_j - \tilde{u}), \qquad F_j(w_j) = F_j(\tilde{u}) + sr_j.$$

Unfortunately, we still have to solve a nonlinear problem in each step. The basic idea of nonlinear multigrid is to compute an approximation $w_{j-1} \in \mathcal{S}_{j-1}$ from the coarse grid problem

$$w_{j-1} \in \mathcal{S}_{j-1}: \qquad F_{j-1}(w_{j-1}) = F_{j-1}(\tilde{u}_{j-1}) + sr_{j-1}.$$

The resulting coarse grid correction is

$$v_{j-1} = \tfrac{1}{s}(w_{j-1} - \tilde{u}_{j-1}) \in \mathcal{S}_{j-1}.$$

We assumed $\tilde{u} = \tilde{u}_{j-1} \in \mathcal{S}_{j-1}$ and $F_{j-1} = F_j|_{\mathcal{S}_{j-1}}$, $r_{j-1} = r_j|_{\mathcal{S}_j}$ are obtained by canonical restriction. It is clear that such a nonlinear two–grid method has to be completed by suitable smoothing on the fine grid. Moreover, global convergence requires suitable damping.

Recursive application provides a *nonlinear multigrid method.*

Algorithm 3.15. *(Nonlinear multigrid method)*

given: u_j^ν
initialize: $\ell_j = 0, \quad \tilde{u}_j = u_j^\nu$

for $k = j, \dots, 0$ *do:*
  if $k > 0$
  {
    $\bar{w}_k = \mathcal{M}_k(\tilde{u}_k, \ell_k)$ *(pre–smoothing)*
    $r_{k-1} = \ell_k - F_k(\bar{w}_k)$ *(evaluation and restriction of residual)*
    $\ell_{k-1} = F_{k-1}(\tilde{u}_{k-1}) + s_{k-1}r_{k-1}$ *(setup of coarse grid problem)*
  }
  else
  {
    $w_0 = \mathcal{M}_0(\tilde{u}_0, \ell_0)$ *(approx. coarse grid solution)*
  }

for $k = 1, \ldots, j$ *do:*

$$\begin{aligned} &\{ \\ &\quad v_{k-1} = \tfrac{1}{s_k}(w_{k-1} - \tilde{u}_{k-1}) \\ &\quad w_k = \bar{w}_k + \omega_k v_{k-1} && \textit{(coarse grid correction)} \\ &\quad w_k = \mathcal{M}_k(w_k, \ell_k) && \textit{(post–smoothing)} \\ &\} \end{aligned}$$

new iterate: $u_j^{\nu+1} = w_j$

Nonlinear counterparts of linear iterative schemes, such as Gauß–Seidel or Jacobi iteration can be used as smoothers $\mathcal{M}_k : \mathcal{S}_k \times \mathcal{S}_k' \to \mathcal{S}_k$ (cf. e.g. Hackbusch [52], pp. 184, or Reusken [80]). Note that the selection of suitable $\tilde{u}_k$, s_k and damping parameters ω_k is required in each step. For linear F_j nonlinear multigrid always reduces to usual linear multigrid.

Theorem 3.16. *Let j be sufficiently large. Then the parameters $\tilde{u}_k$, s_k and ω_k in Algorithm 3.15 can be chosen in such a way that damped nonlinear multigrid iteration is globally convergent. In addition, if $\|u - u_j^\nu\|_{u_j}$ is sufficiently small, then the error estimate*

$$\|u - u_j^{\nu+1}\|_{u_j} \le \rho_0 \|u - u_j^\nu\|_{u_j}$$

holds with $\rho_0 < 1$ independent of j.

For a proof we refer to Hackbusch and Reusken [55, 56]. Damping parameters are chosen according to

$$\omega_k = \min\left\{2, \frac{(F_k(\bar{w}_k) - \ell_k)(v_{k-1})}{\|v_{k-1}\|_{\bar{w}_k}^2 + 2L\|v_{k-1}\|}\right\}.$$

The approximations $\tilde{u}_k$ must be sufficiently close to u_j and s_k must be small enough. Recall that $s_k \approx 0$ provides a good secant approximation.

Theorem 3.16 states that nonlinear multigrid methods have essentially the same properties as Newton multigrid, i.e. global, locally linear convergence. Hence, optimality of nonlinear multigrid with nested iteration (cf. Corollary 3.14) can be shown in a similar way as described in the previous section.

Choosing

$$\tilde{u}_k = R_k^{k+1} \bar{w}_{k+1} \tag{3.58}$$

with suitable restriction operators $R_k^{k+1} : \mathcal{S}_{k+1} \to \mathcal{S}_k$ and $s_k = 1$ for all k we obtain the FAS (full approximation storage) algorithm of Brandt [24]. Note that equality in (3.57) for $s \to 0$ no longer helps in convergence proofs, because $s_k = 1$ is fixed.

3.5. **Concluding remarks.** In order to avoid (even more) technicalities, we imposed very restrictive assumptions on the nonlinearity. For example, we get the same results, if ϕ_j is uniformly Lipschitz on bounded subsets of $\mathcal{S}_j$, e.g. if Φ satisfies $|\Phi'''(\xi)| \le c(1 + \xi^\kappa)$ with fixed $\kappa \in \mathbb{R}$. We never used that F_j has special structure (3.22). Our assumptions providing global smoothness and uniqueness could be replaced by weaker conditions providing corresponding properties hold only on an open subset of H. In Section 3.3 we never used that $\mathcal{S}_j$ has finite dimension. Hence, convergence results remain valid, if inexact Newton iteration is applied directly to the continuous problem (3.1). In this case, inexact solution of arising linearized problems in H is performed by a linear adaptive multilevel finite element method. For further information, we refer to Deuflhard and Weiser [36].

More general results on local convergence of inexact Newton's method (cf. Theorem 3.7) can be found in the monographs of Ortega and Rheinboldt [74], Kelley [62] or Deuflhard [33]. Multigrid methods are typically used to trace back the selection of good initial iterates on a sufficiently fine grid $\mathcal{T}_j$ to much cheaper computations on a coarser grid $\mathcal{T}_0$ via nested iteration. We refer to pioneering papers by Hackbusch [49, 50] and Bank and Rose [8, 9] or Hackbusch [52], pp. 198, Reusken [80] and Heise [58, 59]. Nested iteration is closely related to the so–called *mesh independence principle* which refers to the fact that the number of Newton iterations for solving discretized nonlinear differential equations is asymptotically independent of h_j. See Allgower et al. [3, 5, 4], Deuflhard and Potra [35] or Rannacher [77] for further information. Xu [88] even showed that linearization can be based only on a very rough approximation from a much coarser mesh without loosing optimal order of accuracy.

Suitable damping strategies are indispensable for robustness of Newton multigrid with or without nested iteration. In practice, damping parameters ω_j^ν are often determined by successive bisection of some initial guess. ω_j^ν is accepted as soon as a suitable generalization of the monotonicity test (3.39) is passed. In the absence of an energy functional $\mathcal{E}_j$ an appropriate *merit function* $\tilde{\mathcal{E}}_j$ has to replace energy. Residuals $\tilde{\mathcal{E}}_j(v) = |F_j(v)|$ or corrections $\tilde{\mathcal{E}}_j(v) = |F_j(v)^{-1}F_j(v)|$, both measured in appropriate norms, are widely used. We refer to Heise [58, 59] or Deuflhard [33] for detailed discussions.

Variants of Theorem 3.16 were presented by Hackbusch [52], Reusken [80, 81], Hackbusch and Reusken [55, 56] and others. Nonlinear multigrid sometimes performs surprisingly well even for non–smooth problems but it might also happen that no convergence occurs. The algorithm of Brandt and Cryer [26] for elliptic obstacle problems is a typical example (cf. Kornhuber [63])

It is very difficult to compare Newton multigrid, nonlinear multigrid and FAS numerically. The main reason is that nonlinear problems may have very individual properties from which a given method might benefit or not. However, we refer e.g. to Hackbusch [53] for some interesting observations.

A completely different approach is based on straightforward extension of multilevel relaxation (2.17) to non-quadratic minimization problems. Related subspace correction methods have been investigated e.g. by Dryja and Hackbusch [38] or Tai and Xu [85]. We shall discuss this approach in connection with obstacle problems later on.

4. Piecewise smooth semilinear problems

4.1. **Continuous problem and discretization.** We still consider the minimization problem

$$u \in H : \qquad \mathcal{E}(u) \le \mathcal{E}(v) \qquad \forall v \in H \tag{4.1}$$

with energy functional $\mathcal{E} = \mathcal{J} + \phi$ (cf. (3.2)). But now we relax the conditions on the functional ϕ,

$$\phi(v) = \int_\Omega \Phi(v(x))\, dx \;.$$

Throughout this section we assume that

$$\Phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\} \tag{4.2}$$

is convex and that the *domain*

$$K = \{\xi \in \mathbb{R} \mid \Phi(\xi) < +\infty\}$$

of Φ is a closed interval with $0 \in K$. It follows from the convexity that Φ is locally Lipschitz in the interior of K (see e.g. Aubin [6]). We additionally require that Φ is *piecewise smooth* in the sense that

$$\Phi \in C^3(\theta_{i-1}, \theta_i), \quad |\Phi'''(\xi)| \leq \text{const.} \quad \forall \xi \in (\theta_{i-1}, \theta_i) \tag{4.3}$$

holds with a fixed partition

$$\inf K = \theta_0 < \theta_1 < \cdots < \theta_N < \theta_{N+1} = \sup K \tag{4.4}$$

of the interval K. Such problems typically arise as weak formulations of free boundary problems (see e.g. Duvaut and Lions [40], Elliott and Ockendon [43] or Friedman [45]).

Recall that a subset $\mathcal{K} \subset H$ is convex if the *indicator functional* $\chi_{\mathcal{K}}$, defined by $\chi_{\mathcal{K}}(v) = 0$, $\forall v \in \mathcal{K}$, and $\chi_{\mathcal{K}}(v) = \infty$, $\forall v \notin \mathcal{K}$, is convex. The subset dom $\phi = \{v \in H \mid \phi(v) < +\infty\} \subset H$ is the *domain* of ϕ. Finally, a sequence $(v_k)_{k\geq 0}$ *converges weakly* to $v \in H$, i.e. $v_k \rightharpoonup v$, $k \to \infty$, if $\langle w, v_k\rangle \to \langle w, v\rangle$, $k \to \infty$, holds for all $w \in H'$. We will now state some properties of the functional ϕ as resulting from the above properties of Φ.

Lemma 4.1. *The functional $\phi : H \to \mathbb{R} \cup \{+\infty\}$ is convex, lower semicontinuous and proper. The domain of ϕ is given by the non–empty, closed, and convex set*

$$\mathcal{K} = \{v \in H \mid v(x) \in K \text{ a.e. in } \Omega\} \tag{4.5}$$

and $\phi : \mathcal{K} \to \mathbb{R}$ is continuous. Moreover, we have

$$v_k \rightharpoonup v, k \to \infty \quad \Rightarrow \quad \liminf_{k\to\infty} \phi(v_k) \geq \phi(v) \tag{4.6}$$

for all $(v_k)_{k\geq 0} \subset H$ and $v \in H$.

Proof. Let us first investigate the set $\mathcal{K}$. $\mathcal{K}$ is non–empty, because $0 \in K$ provides $0 \in \mathcal{K}$. $\mathcal{K}$ is convex, because K is an interval and therefore convex. We now show that $\mathcal{K}$ is weakly closed and thus closed. Without loss of generality, let $K = (-\infty, 0]$ and consider a sequence $(v_k)_{k\geq 0} \subset \mathcal{K}$ such that $v_k \rightharpoonup v$, $k \to \infty$. Using the compact embedding of H in $L^2(\Omega)$ (cf. Lemma 3.1), we obtain $v_k \to v$, $k \to \infty$, in $L^2(\Omega)$. We assume that $v \notin \mathcal{K}$. Then we can find a subset Ω' with positive measure such that $v(x) > 0$ $\forall x \in \Omega'$, giving $\|v\|_{L^2(\Omega')} > 0$. Using $v_k \in \mathcal{K}$, we get $v(x) - v_k(x) \geq v(x)$, a.e. in Ω'. This leads to

$$\|v - v_k\|_{L^2(\Omega)} \geq \|v - v_k\|_{L^2(\Omega')} \geq \|v\|_{L^2(\Omega')} > 0 \qquad \forall k \geq 0$$

in contradiction to $v_k \to v$ in $L^2(\Omega)$.

In the next step, we show $\mathcal{K} = \text{dom } \phi$. Let $v \in \mathcal{K}$. Then $|\Phi(v(x))| \leq C(1 + |v(x)|^3)$ a.e. in Ω due to the growth condition in (4.3). Hence, $|\phi(v)| \leq C(1 + \|v\|^3_{L^3(\Omega)}) < \infty$ follows from the continuous embedding $H \subset L^3(\Omega)$. It is clear that $\phi(v) = \infty$ if $v \notin \mathcal{K}$ so that $\mathcal{K} = \text{dom } \phi$.

In order to demonstrate that ϕ is continuous on $\mathcal{K}$, we establish the stronger result

$$(v_k)_{k\geq 0} \subset \mathcal{K}, \ \ v_k \rightharpoonup v, \ k \to \infty \ \ \Rightarrow \ \ \phi(v_k) \to \phi(v), \ k \to \infty. \tag{4.7}$$

Let $(v_k)_{k\geq 0} \subset \mathcal{K}$ and $v_k \rightharpoonup v$, $k \to \infty$. We already know that $\mathcal{K}$ is weakly closed so that $v \in \mathcal{K}$. We get convergence $v_k \to v$, $k \to \infty$, in $L^p(\Omega)$, $p = 2, 4$, from Lemma 3.1. Now utilizing (4.3), we find

$$|\Phi(v(x)) - \Phi(v_k(x))| \leq c(1 + |v(x)|^2 + |v_k(x)|^2)|v(x) - v_k(x)| \qquad \text{a.e. in } \Omega$$

with c independent of v, x and k. This leads to

$$\begin{aligned} |\phi(v) - \phi(v_k)| &\leq \textstyle\int_\Omega |\Phi(v(x)) - \Phi(v_k(x))|\, dx \\ &\leq 3c \left(\textstyle\int_\Omega (1 + |v(x)|^4 + |v_k(x)|^4)\, dx\right)^{1/2} \|v - v_k\|_{L^2(\Omega)} \to 0 \quad k \to 0. \end{aligned}$$

We now prove (4.6). Let $v_k \rightharpoonup v$, $k \to \infty$. Assume that for each $k_0 \in \mathbb{N}$, there is an index $k \geq k_0$ such that $v_k \in \mathcal{K}$. Then we can find a subsequence $(v_{k_i})_{i\geq 0} \subset \mathcal{K}$ still converging weakly to v. Hence, $v \in \mathcal{K}$ and (4.7) yields $\phi(v_k) \to \phi(v)$, $k \to \infty$. In the remaining case, $v_k \notin \mathcal{K}$ $\forall k \geq k_0$ holds with some fixed $k_0 \geq 0$. Then we clearly have $\liminf_{k\to\infty} \phi(v_k) = \infty \geq \phi(v)$.

It follows from convexity of Φ that ϕ is convex. As a consequence of (4.6) ϕ is lower semicontinuous. As $\mathcal{K} \neq \emptyset$ and $\phi(v) \in \mathbb{R} \cup \{+\infty\}$ we have $\phi \not\equiv \infty$ and $\phi(v) > -\infty$ $\forall v \in H$ so that ϕ is proper. This concludes the proof. □

Theorem 4.2. *The minimization problem (4.1) admits a unique solution.*

Proof. As a consequence of Lemma 4.1 the energy functional $\mathcal{E}$ is strictly convex, lower semicontinuous, proper and coercive. Hence, the assertion follows form Proposition 1.2 in [41], p. 35. □

We emphasize that ϕ is not Fréchèt differentiable. The minimization problem (4.1) can be reformulated as a variational inequality by circumventing differentiation of ϕ. Another possibility is to extend the notion of differentiability. Recall that a bounded linear functional g on H with the property

$$\phi(v) - \phi(v_0) \geq \langle g, v - v_0 \rangle \qquad \forall v \in H$$

is called *subgradient* of ϕ at $v_0 \in H$. The *subdifferental* $\partial\phi(v_0)$ *at* v_0 is the set of all subgradients of ϕ at v_0. We refer e.g. to Deimling [31], pp. 281, for detailed information. The following results are well–known (see e.g. Ekeland and Temam [41], Proposition 2.2, p. 38 Glowinski [46], Lemma 4.1, p. 6, or Kornhuber [65], Theorem 1.6, p. 30 and Proposition 1.7, p. 34).

Proposition 4.3. *The variational inequality*

$$u \in H: \qquad a(u, v - u) + \phi(v) - \phi(u) \geq \ell(v - u) \qquad \forall v \in H \tag{4.8}$$

and the variational inclusion

$$u \in H: \qquad \ell(v) - a(u, v) \in \partial\phi(u)(v) \qquad \forall v \in H \tag{4.9}$$

are both equivalent to the minimization problem (4.1).

Note that both formulations reduce to the variational equality (3.10), if ϕ happens to be smooth.

As in the previous section, discretization of (4.1) relies on the approximate energy

$$\mathcal{E}_j(v) = \mathcal{J}(v) + \phi_j(v), \qquad v \in \mathcal{S}_j,$$

where

$$\phi_j(v) = \sum_{p \in \mathcal{N}_j} \Phi(v(p))\, h_p, \qquad h_p = \int_\Omega \lambda_p^{(j)}\, dx, \tag{4.10}$$

is resulting from $\mathcal{S}_j$–interpolation of $\Phi(v)$. Again, the discrete functional ϕ_j inherits essential properties from ϕ. In particular ϕ_j is convex, lower semicontinuous and proper. Moreover, ϕ_j is continuous on the closed set

$$\mathcal{K}_j = \text{dom } \phi_j = \{v \in \mathcal{S}_j \mid v(p) \in K \ \ \forall p \in \mathcal{N}_j\}.$$

The subdifferential of ϕ_j takes the form

$$\partial\phi_j(v_0)(v) = \sum_{p \in \mathcal{N}_j} \partial\Phi(v_0(p))v(p)h_p \qquad \forall v \in \mathcal{S}_j$$

for all $v_0 \in \text{dom } \phi_j$. Note that

$$\partial\Phi(\xi) = \left\{ \begin{array}{cl} \Phi'(\xi), & \text{if } \xi \neq \theta_i \\ [\Phi'(\theta_i)_-, \Phi'(\theta_i)_+], & \text{if } \xi = \theta_i \end{array} \right. , \qquad i = 1, \ldots, N-1,$$

holds for all $\xi \in K$, denoting $\Phi'(\theta_i)_\pm = \lim_{\xi \to \pm\theta_i} \Phi'(\xi)$.

Now the finite element discretization

$$u_j \in \mathcal{S}_j : \qquad \mathcal{E}_j(u_j) \leq \mathcal{E}_j(v) \qquad v \in \mathcal{S}_j \tag{4.11}$$

is obtained by minimization of discrete energy on the finite element space $\mathcal{S}_j$.

Theorem 4.4. *The discrete problem (4.11) has a unique solution and can be equivalently reformulated as the variational inequality*

$$u_j \in \mathcal{S}_j : \qquad a(u_j, v - u_j) + \phi_j(v) - \phi_j(u_j) \geq \ell(v - u_j) \qquad \forall v \in \mathcal{S}_j \tag{4.12}$$

or as the variational inclusion

$$u_j \in \mathcal{S}_j : \qquad \ell(v) - a(u_j, v) \in \partial\phi_j(u_j)(v) \qquad \forall v \in \mathcal{S}_j. \tag{4.13}$$

The discrete solutions u_j converge to u in the sense that

$$\|u - u_j\| \to 0 \quad \textit{and} \quad \phi_j(u_j) \to \phi(u) \qquad j \to \infty.$$

Proof. A detailed proof relying on abstract convergence results in Glowinski [46], pp. 12 has been given by Kornhuber [65], pp. 38. Note that property (4.6) plays an important role. □

Optimal a priori estimates of the discretization error are known for special cases such as obstacle problems (cf. Falk [44]) or Stefan–type problems (cf. Elliott [42]).

4.2. **Obstacle problems.** As an important special case of (4.1) let us consider the obstacle problem

$$u \in \mathcal{K} : \qquad \mathcal{J}(u) \leq \mathcal{J}(v) \qquad \forall v \in \mathcal{K} \tag{4.14}$$

with

$$\mathcal{K} = \{v \in H \mid v(x) \leq 0 \text{ a.e. in } \Omega\} \tag{4.15}$$

and its discrete analogue

$$u_j \in \mathcal{K}_j : \qquad \mathcal{J}(u_j) \leq \mathcal{J}(v) \qquad \forall v \in \mathcal{K}_j \tag{4.16}$$

with

$$\mathcal{K}_j = \{v \in \mathcal{S}_j \mid v(p) \leq 0 \; \forall p \in \mathcal{N}_j\}.$$

Choose $\Phi \equiv 0$ and $N = 1$, $\theta_0 = -\infty$, $\theta_1 = 0$ in (4.2) and (4.4) to confirm that (4.14) and (4.16) are special cases of (4.1) and (4.11), respectively. Note that problems with upper obstacles $\varphi \in H_0^1(\Omega)$ can be reduced to (4.15) by suitable transformation.

Successive minimization (cf. Algorithm 2.3) is the starting point for the construction of fast solvers for discrete obstacle problems (4.16). In analogy to the linear self–adjoint case, the splitting (2.15) generates well–known *projected Gauß–Seidel relaxation* (cf. e.g. Glowinski [46], pp. 140). In a similar way, extended splitting (2.16) decouples (4.16) into $l = 1, \dots, m_{\mathcal{S}}$ one–dimensional subproblems

$$\begin{aligned} v_l \in D_l^* : \quad & \mathcal{J}(w_{l-1} + v_l) \le \mathcal{J}(w_{l-1} + v) \quad \forall v \in D_l^* \\ & D_l^* = \{v \in V_l \mid w_{l-1} + v \in \mathcal{K}_j\} \\ & w_l = w_{l-1} + v_l \end{aligned} \tag{4.17}$$

starting with given iterate $w_0 = u_j^\nu$. The new iterate of resulting extended relaxation is $u_j^{\nu+1} = w_{m_{\mathcal{S}}}$.

Observe that in general $-w_{l-1} + \mathcal{K}_j \not\subset \mathcal{S}_k$ if $k < j$. Hence, implementation of (4.17) in analogy to Algorithm 2.5 requires additional interpolation of corrections $v_k \in \mathcal{S}_k$, $k < j$, in order to check the constraints. As a consequence, computational complexity ranges from $\mathcal{O}(n_j \log n_j)$ in case of uniform refinement up to $\mathcal{O}(n_j^2)$ for certain adaptively refined meshes. In order to preserve optimal complexity, we consider approximate subproblems

$$\begin{aligned} v_l \in D_l : \qquad & \mathcal{J}(w_{l-1} + v_l) \le \mathcal{J}(w_{l-1} + v) \qquad \forall v \in D_l \\ & w_l = w_{l-1} + v_l \end{aligned} \tag{4.18}$$

as resulting from approximate constraints

$$D_l = \{v \in V_l \mid v(p) \le \psi_l \lambda_l(p) \ \forall p \in \mathcal{N}_j\} \ .$$

Suitable defect obstacles $\psi_l \in \mathbb{R}$ can be obtained by successive update and *monotone restriction* of initial defect obstacle $-u_j^\nu(p)$ (cf. e.g. Algorithm 4.5 below). Note that defect obstacles ψ_l should be chosen in such a way that

$$0 \in D_l \subset D_l^* \ . \tag{4.19}$$

Utilizing some enumeration of multilevel nodal basis $\lambda_l = \lambda_{p_l}^{(k_l)}$ from fine to coarse (cf. Section 2.2) and denoting

$$\mathcal{D}_k = \{v \in \mathcal{S}_k \mid v(p) \le \psi_p^{(k)} \ \forall p \in \mathcal{N}_k\} \subset \mathcal{S}_k$$

with $\psi_{p_l}^{(k_l)} = \psi_l$, successive approximate minimization (4.18) can be reformulated as a multigrid V–cycle. The bilinear forms $b_k(\cdot,\cdot)$ are given by (2.19).

Algorithm 4.5. *(Standard monotone multigrid V–cycle)*

given: u_j^ν

initialize: $r_j = \ell - a(u_j^\nu, \cdot), \quad a_j(\cdot,\cdot) = a(\cdot,\cdot), \quad \psi_p^{(j)} = -u_j^\nu(p) \ \forall p \in \mathcal{N}_j$

for $k = j, \dots, 1$ *do:*

{

solve:

$v_k \in \mathcal{D}_k : \quad b_k(v_k, v - v_k) \ge r_k(v - v_k) \quad \forall v \in \mathcal{D}_k$ *(pre–smoothing)*

$r_k = r_k - a_k(v_k, \cdot)$ *(update of residual)*

$\psi_p^{(k)} = \psi_p^{(k)} - v_k(p) \ \forall p \in \mathcal{N}_k$ *(update of defect obstacle)*

$r_{k-1} = r_k|_{\mathcal{S}_{k-1}}$

$a_{k-1}(\cdot,\cdot) = a_k(\cdot,\cdot)|_{\mathcal{S}_{k-1} \times \mathcal{S}_{k-1}}$ *(canonical restriction)*

$\psi_p^{(k-1)} = \min\{\psi_q^{(k)} \mid q \in \mathcal{N}_k \cap \text{int supp}\, \lambda_p^{(k-1)}\}$ *(monotone restriction)*

}

solve:

$v_0 \in \mathcal{D}_0: \quad b_0(v_0, v - v_0) \geq r_0(v - v_0) \qquad \forall v \in \mathcal{D}_0,$ *(approx. solution on $\mathcal{T}_0$)*

for $k = 1, \ldots, j$ *do:*

{

$v_k = v_k + v_{k-1}$ *(canonical interpolation)*

}

new iterate: $u_j^{\nu+1} = u_j^\nu + v_j$

Algorithm 4.5 is called *standard monotone multigrid method*, because *standard* multilevel nodal basis functions give rise to the underlying splitting (2.16) and, by construction, energy of iterates is *monotonically* decreasing. Note that projected Gauß–Seidel smoothing is performed on each level. Again, post–smoothing or W–cycles can be formulated in a similar way.

Theorem 4.6. *The standard monotone multigrid method as given by Algorithm 4.5 is globally convergent.*

Proof. A more general convergence result (e.g. Theorem 4.9) will be shown later. □

Let us investigate the behavior of approximate *coincidence sets* $\mathcal{N}_j^\bullet(u_j^\nu)$

$$\mathcal{N}_j^\bullet(u_j^\nu) = \{p \in \mathcal{N}_j \mid u_j^\nu(p) = 0\} \subset \mathcal{N}_j$$

and corresponding sets of *regular nodes* $\mathcal{N}_j^\circ(u_j^\nu) = \mathcal{N}_j \setminus \mathcal{N}_j^\bullet(u_j^\nu)$.

Lemma 4.7. *Assume that the discrete problem (4.16) is non–degenerate in the sense that*

$$\ell(\lambda_p^{(j)}) - a(u_j, \lambda_p^{(j)}) > 0 \qquad \forall p \in \mathcal{N}_j^\bullet(u_j). \tag{4.20}$$

Then there is $\nu_0 > 0$ *such that*

$$\mathcal{N}_j^\bullet(u_j^\nu) = \mathcal{N}_j^\bullet(u_j) \qquad \forall \nu \geq \nu_0.$$

Proof. Let $p \in \mathcal{N}_j^\circ(u_j)$. Then $u_j(p) < 0$. Convergence of $(u_j^\nu)_{\nu\geq 0}$ and equivalence of norms on $\mathcal{S}_j$ implies $u_j^\nu(p) < 0$ for sufficiently large ν. As Ω is bounded we conclude $\mathcal{N}_j^\circ(u_j) \subset \mathcal{N}_j^\circ(u_j^\nu)$.

It remains to show $\mathcal{N}_j^\bullet(u_j) \subset \mathcal{N}_j^\bullet(u_j^\nu)$. Linear superposition of (4.20) yields

$$\ell(\lambda_l) - a(u_j, \lambda_l) > 0$$

if $\text{int supp}\, \lambda_l \cap \mathcal{N}_j^\bullet(u_j) \neq \emptyset$. As a consequence of

$$\mathcal{J}(u_j^{\nu+1}) \leq \mathcal{J}(w_l^\nu) \leq \mathcal{J}(u_j^\nu)$$

the intermediate iterates w_l^ν also converge to u_j. Hence,

$$\ell(\lambda_l) - a(w_l^\nu, \lambda_l) > 0 \tag{4.21}$$

holds, if $\text{int supp}\, \lambda_l \cap \mathcal{N}_j^\bullet(u_j) \neq \emptyset$ and ν is large enough.

Now let $p \in \mathcal{N}_j^\bullet(u_j)$ and assume that ν is large enough to guarantee (4.21). Projected Gauß–Seidel smoothing on level j provides the intermediate iterate w_{n_j} (cf. (4.17)). Assume that $p \notin \mathcal{N}_j^\bullet(w_{n_j})$, or equivalently,

$$w_{n_j}(p) = w_{l_p}(p) = w_{l_p-1}(p) + v_{l_p}(p) < 0$$

with $\lambda_{l_p} = \lambda_p^{(j)}$. Then (4.17) yields

$$a(v_{l_p}, v) = \ell(v) - a(w_{l_p-1}, v) \qquad \forall v \in V_{l_p}$$

in contradiction to (4.21). Hence, $p \in \mathcal{N}_j^\bullet(w_{n_j})$. Consider some fixed $l = n_j + 1, \ldots, m_\mathcal{S}$ with $p \in \operatorname{int}\operatorname{supp}\lambda_l$. Then (4.21) in combination with (4.18) and (4.19) implies $v_l = 0$ in a similar way. Hence, $p \in \mathcal{N}_j^\bullet(u_j^{\nu+1})$ so that we have shown $\mathcal{N}_j^\bullet(u_j) \subset \mathcal{N}_j^\bullet(u_j^\nu)$ for sufficiently large ν. □

Note that continuous versions of nondegeneracy condition (4.20) guarantee stability of the free boundary (see e.g. Rodrigues [82], pp. 198).

If the exact coincidence set $\mathcal{N}_j^\bullet(u_j)$ is known, then (4.16) can be reformulated as reduced linear problem

$$u_j \in \mathcal{S}_j^\circ : \qquad a(u_j, v) = \ell(v) \qquad \forall v \in \mathcal{S}_j^\circ \tag{4.22}$$

on the reduced space

$$\mathcal{S}_j^\circ = \{v \in \mathcal{S}_j \mid v(p) = 0 \ \forall p \in N_j^\bullet(u_j)\} \subset \mathcal{S}_j \ .$$

Note that (4.22) approximates a linear elliptic partial differential equation on the reduced computational domain $\Omega^\circ(u_j)$,

$$\Omega^\circ(u_j) = \Omega \setminus \{t \in \mathcal{T}_j \mid t \cap \mathcal{N}_j^\circ(u_j) = \emptyset\} \ .$$

Convergence of iterates in combination with Lemma 4.7 implies that the nonlinear iteration as described in Algorithm 4.5 becomes a linear successive subspace correction method for (4.22) (cf. Algorithm 2.4) if ν is large enough. The underlying splitting of $\mathcal{S}_j^\circ$ is given by

$$\mathcal{S}_j^\circ = \sum_{k=0}^{j} \mathcal{V}_k^\circ, \qquad \mathcal{V}_k^\circ = \mathcal{S}_k \cap \mathcal{S}_j^\circ$$

and the bilinear forms $b_k(\cdot,\cdot)$ are associated with Gauß–Seidel relaxation.

Hence, linear convergence theory (cf. Bramble [19], Xu [87, 89, 90] or Yserentant [92]) can be applied to derive asymptotic upper bounds for the convergence rates. Utilizing estimates for modified hierarchical splittings (cf. Kornhuber and Yserentant [70]), the following result is a consequence of Theorem 2.5 by Neuss [73].

Theorem 4.8. *Let $d = 2$. Assume that the discrete problem (4.16) is non–degenerate in the sense of (4.20). Then there is a ν_0 depending on j such that linear convergence*

$$\|u_j - u_j^{\nu+1}\| \leq (1 - c(j+1)^{-2})\|u_j - u_j^\nu\| \tag{4.23}$$

holds with constant $c > 0$ depending only on the initial triangulation $\mathcal{T}_0$, on the maximal gradient of α and on the constants α_0 and α_1 in (2.4).

Optimal convergence rates in arbitrary space dimensions can be shown if the coincidence set is large enough in a certain sense. We refer to Kornhuber and Yserentant [70] and Oswald [76] for details.

We emphasize that Theorem 4.8 does *not* state local linear convergence. Indeed, for each $\varepsilon > 0$ there is $u_j^0 \in \mathcal{S}_j$ such that $\|u_j - u_j^0\| \leq \varepsilon$ but $\mathcal{N}_j^\bullet(u_j^0) \neq \mathcal{N}_j^\bullet(u_j)$.

Linear multigrid convergence theory providing linear convergence (4.23) cannot be applied in this case.

Let $\|u_j - u_j^0\| \le Ch_j$. Then Theorem 4.8 guarantees the existence of $\nu_0(h_j, j)$ such that linear convergence (4.23) holds for $\nu \ge \nu_0(h_j, j)$. It is not (yet) known wether $\nu_0(h_j, j)$ is uniformly bounded in j. As a consequence, we can *not* use the arguments of the preceding sections to show that nested iteration provides approximations with $\mathcal{O}(h_j)$ accuracy by $\mathcal{O}((1+j)^2 n_j)$ point operations.

Let us now describe a modification of Algorithm 4.5 based on suitable adaptation of trial spaces V_l to the coincidence set. In general, the reduced computational domain $\Omega^\circ(u_j)$ can not be resolved by subsets of the coarse grids $\mathcal{T}_k$, $k < j$. This leads to bad representation of low frequencies which in turn leads to unsatisfying convergence rates as compared to classical linear multigrid (cf. Algorithm 2.5). In order to improve convergence speed, we now consider a slight modification of Algorithm 4.5.

Observe that each step of Algorithm 4.5 can be decomposed into two substeps

$$\begin{aligned} \bar{u}_j^\nu &= \mathcal{M}_j u_j^\nu \\ u_j^{\nu+1} &= \mathcal{C}_j \bar{u}_j^\nu \end{aligned} \quad . \tag{4.24}$$

Here $\mathcal{M}_j : \mathcal{S}_j \to \mathcal{S}_j$ denotes the leading projected Gauß–Seidel smoothing on the fine grid and $\mathcal{C}_j : \mathcal{S}_j \to \mathcal{S}_j$ stands for all subsequent coarse grid corrections. The coincidence set $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ of the so called *smoothed iterate* $\bar{u}_j^\nu$ can be regarded as an approximation of $\mathcal{N}_j^\bullet(u_j)$. Recall that the proof of Lemma 4.7 yields $\mathcal{N}_j^\bullet(\bar{u}_j^\nu) = \mathcal{N}_j^\bullet(u_j)$ for sufficiently large ν. We now simply freeze $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ and ignore all entries from $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ in course of coarse grid correction.

Modifications of Algorithm 4.5 (Truncated monotone multigrid V–cycle)

modified restrictions of stiffness matrix and of residual:
set all entries from the approx. coincidence set $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ *to zero*

modified restrictions of defect obstacle:
set all entries from the approx. coincidence set $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ *to* $+\infty$

modified interpolation of corrections:
set all entries from the approx. coincidence set $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ *to zero*

The above modifications of Algorithm 4.5 can be reinterpreted as an adaptation (i.e. truncation) of search directions λ_l in successive minimization (4.18) to the approximate reduced computational domain $\Omega^\circ(\bar{u}_j^\nu)$. For a one–dimensional analogue truncation is illustrated in Figure 1. The two critical nodes are marked by dots and $\tilde{\lambda}_p^{(0)}$ is the corresponding truncation of dashed $\lambda_p^{(0)}$. Hence the resulting iterative scheme is called *truncated monotone multigrid method.* Global convergence (cf. Theorem 4.6) carries over to the truncated version. In the non–degenerate case, truncated multigrid also reduces to a linear multigrid method for (4.22). The underlying splitting of $\mathcal{S}_j^\circ$ consists of larger subspaces $\tilde{\mathcal{V}}_k^\circ \supset \mathcal{V}_k^\circ$. In general, we have $\tilde{\mathcal{V}}_k^\circ \not\subset \mathcal{S}_k$ so that this linear reduced scheme is slightly more difficult to analyze. Similar arguments as in the proof of Theorem 4.8 provide asymptotic convergence

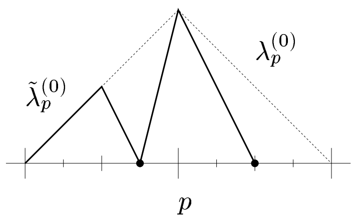

FIGURE 1. Truncated nodal basis function

rates behaving like $1-\mathcal{O}((j+1)^{-3})$. However, it has been observed in many numerical experiments that truncated monotone multigrid usually converges faster than the standard variant.

As an example, we consider a well–known model problem (cf. Glowinski [46], pp. 41, or Kornhuber [63]). The quadratic form $a(\cdot,\cdot)$, the right hand side $\ell(\cdot)$ and the constraints $\mathcal{K} \subset H_0^1(\Omega)$ are selected according to

$$a(v,w) = \int_\Omega \nabla v \ \cdot \nabla w \ dx, \qquad \ell(v) = 5 \int_\Omega \ v \ dx$$

and

$$\mathcal{K} = \{v \in H_0^1(\Omega) | \ v(x) \le \operatorname{dist}(x, \partial\Omega), \ \text{a.e. in } \Omega\},$$

respectively. The computational domain is the unit square $\Omega = (0,1) \times (0,1)$ and the initial triangulation $\mathcal{T}_0$ consists of 4 congruent triangles. Starting with initial iterates $u_j^0 = u_{j-1}^{\nu_j^*}$ for $j = 1, \ldots, 8$ and $u_0^0 = 0$ (nested iteration), we use standard monotone multigrid (cf. Algorithm 4.5) and the truncated variant (cf. Modifications of Algorithm 4.5) to compute iterates u_j^ν, $\nu = 0, \ldots, \nu_j^*$. The number ν_j^* of iteration

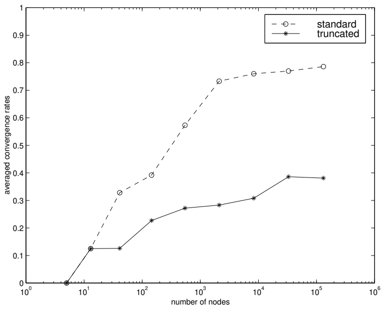

FIGURE 2. Nested iteration: Averaged convergence rates

steps is selected in such a way that $\varepsilon_j^{\nu_j^*} = \|u_j - u_j^{\nu_j^*}\| < 10^{-12}$. Figure 2 shows averaged convergence rates ρ_j given by

$$\rho_j = \sqrt[\nu_j^*]{\varepsilon_j^{\nu_j^*} / \varepsilon_j^0}, \ \ j = 0, \ldots, 8,$$

where $\varepsilon_j^0 = \|u_j - u_j^0\|$. Observe that averaged convergence rates seem to saturate for both methods indicating that nested iteration leads to optimal schemes. The truncated variant is about twice as fast as the standard method. Similar convergence rates are obtained for classical linear multigrid (cf. Algorithm 2.5) as applied to the same problem without any constraints.

We have seen that local linear convergence showed up numerically. Moreover, other numerical experiments indicate that convergence rates might be even globally bounded uniformly in j. Theoretical justification is subject of current research.

4.3. **Constrained Newton Linearization and Monotone Multigrid.** We now present an extension of Newton multigrid (cf. section 3.3) to piecewise smooth problems of the form (4.1). The basic idea is to constrain Newton linearization to a closed convex neighborhood of u_j on which ϕ_j is smooth. Such a neighborhood is not known a priori but has to be approximated in course of the iteration. This leads to a two stage method. Each step consists of a nonlinear Gauß–Seidel relaxation and approximate solution of an obstacle problem of the form (4.16) which now plays the same role as classical Newton linearization played in the smooth case. We give a more precise formulation of this approach in order to provide a general framework for *monotone multigrid methods* to be constructed later on.

Assume that $\mathcal{M}_j : \mathcal{S}_j \to \mathcal{S}_j$ satisfies

$$\left.\begin{aligned} \mathcal{E}_j(\mathcal{M}_j(w)) &< \infty \\ \mathcal{E}_j(\mathcal{M}_j(w)) &\le \mathcal{E}_j(w) \end{aligned}\right. \qquad \forall w \in \mathcal{S}_j, \tag{4.25}$$

where

$$\mathcal{E}_j(\mathcal{M}_j(w)) = \mathcal{E}_j(w) \Leftrightarrow w = u_j. \tag{4.26}$$

In addition, we require that

$$\limsup_{\nu\to\infty} \mathcal{E}_j(\mathcal{M}_j(w^\nu)) \le \mathcal{E}_j(\mathcal{M}_j(\lim_{\nu\to\infty} w^\nu)) \tag{4.27}$$

holds for each convergent sequence $(w^\nu)_{\nu\ge 0} \subset \mathcal{K}_j$.

We shall see that the above conditions are sufficient for global convergence of the iteration $u_j^{\nu+1} = \mathcal{M}_j(u_j^\nu)$. However, the convergence speed may be unacceptable low. In analogy to (4.24), we introduce slightly more general *monotone iterations*

$$\begin{aligned} \bar{u}_j^\nu &= \mathcal{M}_j(u_j^\nu) \\ u_j^{\nu+1} &= \mathcal{C}_j(\bar{u}_j^\nu) \end{aligned} \tag{4.28}$$

where the additional substep $\mathcal{C}_j$ is intended to accelerate the convergence speed. Adopting multigrid terminology, $\mathcal{M}_j$ is called *fine grid smoother*, $\bar{u}_j^\nu$ is the *smoothed iterate* and $\mathcal{C}_j$ is called *coarse grid correction*.

Theorem 4.9. *Assume that the fine grid smoother $\mathcal{M}_j$ satisfies conditions (4.25) - (4.27) and that the coarse grid correction $\mathcal{C}_j$ has the monotonicity property*

$$\mathcal{E}_j(\mathcal{C}_j(w)) \le \mathcal{E}_j(w) \qquad \forall w \in \mathcal{S}_j. \tag{4.29}$$

Then the monotone iteration (4.28) is globally convergent.

Proof. Let us first show that the sequence of iterates $(u_j^\nu)_{\nu\ge 0}$ is bounded. Recall that there are constants c, $C \in \mathbb{R}$, such that $\phi_j(v) \ge c\|v\| + C$ $\forall v \in \mathcal{S}_j$, because

ϕ_j is convex, lower semicontinuous and proper (cf. e.g. Ekeland and Temam [41], p. 14). As a consequence, we have

$$\mathcal{E}_j(v) \geq \tfrac{1}{2}\|v\|^2 + (c - \|\ell\|)\|v\| + C \qquad \forall v \in \mathcal{S}_j, \tag{4.30}$$

so that $\|v\| \to \infty$ implies $\mathcal{E}_j(v) \to \infty$. Hence, $(u_j^\nu)_{\nu \geq 0}$ must be bounded, because

$$\mathcal{E}_j(u_j^\nu) \leq \mathcal{E}_j(\bar{u}_j^1) < \infty \qquad \forall \nu \geq 1$$

follows from (4.25) and (4.29).

Now, let $(u_j^{\nu_k})_{k \geq 0}$ be an arbitrary, convergent subsequence of $(u_j^\nu)_{\nu \geq 0}$ with limit $u^* \in \mathcal{S}_j$. Such a subsequence exists, because u_j^ν is bounded and $\mathcal{S}_j$ has finite dimension. Note that $u^* \in \mathcal{K}_j$, because $(u_j^{\nu_k})_{k \geq 1} \subset \mathcal{K}_j$ and $\mathcal{K}_j$ is closed. We now prove $u_j^* = u_j$. In the light of (4.26), it is sufficient to show

$$\mathcal{E}_j(\mathcal{M}_j(u^*)) = \mathcal{E}_j(u^*). \tag{4.31}$$

Observe that (4.25) and (4.29) imply

$$\mathcal{E}_j(u_j^{\nu_{k+1}}) \leq \mathcal{E}_j(u_j^{\nu_k + 1}) \leq \mathcal{E}_j(\mathcal{M}_j(u_j^{\nu_k})) \leq \mathcal{E}_j(u_j^{\nu_k}).$$

In virtue of the continuity of $\mathcal{E}_j$ on $\mathcal{K}_j$, this leads to

$$\lim_{k \to \infty} \mathcal{E}_j(\mathcal{M}_j(u_j^{\nu_k})) = \mathcal{E}_j(u_j^*).$$

Now the equality (4.31) is a consequence of conditions (4.25) and (4.27).

As $(u_j^{\nu_k})_{k \geq 0}$ was an arbitrary convergent subsequence, the whole sequence u_j^ν must converge to u_j. This completes the proof. □

As a by–product, we also get convergence of the smoothed iterates

$$\lim_{k \to \infty} \bar{u}_j^\nu = u_j. \tag{4.32}$$

We now concentrate on the construction of appropriate fine grid smoother $\mathcal{M}_j$ and coarse grid correction $\mathcal{C}_j$.

Successive minimization of energy (cf. Algorithm 2.3) is directly applicable to the discrete minimization problem (4.11). Choosing the splitting (2.15) in one–dimensional subspaces $V_l = \operatorname{span}\{\lambda_{p_l}^{(j)}\}$ we obtain the well–known *nonlinear Gauß–Seidel relaxation*.

Algorithm 4.10. *(Nonlinear Gauß–Seidel relaxation)*

given: $w_0 = u_j^\nu \in \mathcal{S}_j$

for $l = 1, \dots, n_j$ *do:*

{

$$\begin{aligned} v_l \in V_l: \quad & \mathcal{J}(w_{l-1} + v_l) + \Phi(u_j^\nu(p_l) + v_l(p_l))h_{p_l} \\ & \qquad \leq \mathcal{J}(w_{l-1} + v) + \Phi(u_j^\nu(p_l) + v(p_l))h_{p_l} \quad \forall v \in V_l, \\ w_l = w_{l-1} + v_l & \end{aligned} \tag{4.33}$$

}

new iterate: $u_j^{\nu+1} = \mathcal{M}_j^{\mathrm{GS}}(u_j^\nu) = w_{n_j}$.

Lemma 4.11. *Nonlinear Gauß–Seidel iteration $\mathcal{M}_j^{\mathrm{GS}}$ satisfies conditions (4.25)–(4.27).*

Proof. Condition (4.25) is clear by construction. Condition (4.26) follows from the uniqueness of solutions of local problems (4.33) and the fact that the unknowns are not coupled by the nonlinearity ϕ_j. Finally, (4.27) holds, because $\mathcal{M}_j$ is continuous. $\square$

Lemma 4.11 heavily relies on the special structure (3.11) of ϕ_j. For dangerous generalizations and a simple counterexample we refer to Glowinski [46], pp. 149.

Next, we construct a coarse grid correction $\mathcal{C}_j$. For ease of presentation, we assume $K = \mathbb{R}$ and $N = 1$ so that there is only one singularity of Φ at $\theta_1 \in \mathbb{R}$.

Let us introduce the subset of *regular nodes*

$$\mathcal{N}_j^\circ(\bar{u}_j^\nu) = \{p \in \mathcal{N}_j \mid \bar{u}_j^\nu(p) \neq \theta_1\}$$

and denote $\mathcal{N}_j^\bullet(\bar{u}_j^\nu) = \mathcal{N}_j \setminus \mathcal{N}_j^\circ(\bar{u}_j^\nu)$. We consider the functional $\phi_{\bar{u}_j^\nu}$,

$$\phi_{\bar{u}_j^\nu}(v) = \sum_{p \in \mathcal{N}_j^\circ(\bar{u}_j^\nu)} \Phi(v(p))\, h_p + \sum_{p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)} \Phi(\theta_1)\, h_p\,, \qquad v \in \mathcal{S}_j.$$

Observe that Φ is smooth on a neighborhood of $\bar{u}_j^\nu(p)$ that does not contain θ_1. This motivates the definition of the closed, convex set

$$\mathcal{K}_{\bar{u}_j^\nu} = \{w \in \mathcal{S}_j \mid \underline{\varphi}_{\bar{u}_j^\nu}(p) \le w(p) \le \overline{\varphi}_{\bar{u}_j^\nu}(p),\ p \in \mathcal{N}_j\}$$

where for $p \in \mathcal{N}_j^\circ(\bar{u}_j^\nu)$

$$\underline{\varphi}_{\bar{u}_j^\nu}(p) = \begin{cases} -\infty, & \text{if } \bar{u}_j^\nu(p) < \theta_1 \\ \theta_1, & \text{if } \bar{u}_j^\nu(p) > \theta_1 \end{cases} \qquad \overline{\varphi}_{\bar{u}_j^\nu}(p) = \begin{cases} \theta_1, & \text{if } \bar{u}_j^\nu(p) < \theta_1 \\ +\infty, & \text{if } \bar{u}_j^\nu(p) > \theta_1 \end{cases},$$

and for $p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ $\underline{\varphi}_{\bar{u}_j^\nu}(p) = \overline{\varphi}_{\bar{u}_j^\nu}(p) = \theta_1$. By construction, $\phi_{\bar{u}_j^\nu}$ is *smooth on* $\mathcal{K}_{\bar{u}_j^\nu}$, i.e. the Lipschitz condition

$$|\phi''_{\bar{u}_j^\nu}(w_1)(v,w) - \phi''_{\bar{u}_j^\nu}(w_2)(v,w)| \le L\|w_1 - w_2\|\|v\|\|w\| \qquad \forall w_1, w_2 \in \mathcal{K}_{\bar{u}_j^\nu} \tag{4.34}$$

holds for all $v, w \in \mathcal{S}_j$ with suitable $L \in \mathbb{R}$. The relation of $\phi_{\bar{u}_j^\nu}$ and $\mathcal{K}_{\bar{u}_j^\nu}$ to our original problem (4.11) is clarified by the following lemma.

Lemma 4.12. *Let (4.11) satisfy the* non–degeneracy condition

$$\ell(\lambda_p^{(j)}) - a(u_j, \lambda_p^{(j)}) \in \partial\phi_j(u_j)(\lambda_p^{(j)}) \qquad \forall p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu). \tag{4.35}$$

Consider monotone iteration (4.28) with nonlinear Gauß–Seidel smoothing $\mathcal{M}_j$ and arbitrary $\mathcal{C}_j$ with property (4.29). Then there is a ν_0 such that (4.12) can be equivalently rewritten as the constrained smooth problem

$$u_j \in \mathcal{K}_{\bar{u}_j^\nu}: \qquad \mathcal{E}_{\bar{u}_j^\nu}(u_j) \le \mathcal{E}_{\bar{u}_j^\nu}(v) \qquad \forall v \in \mathcal{K}_{\bar{u}_j^\nu} \qquad \forall \nu \ge \nu_0 \tag{4.36}$$

denoting $\mathcal{E}_{\bar{u}_j^\nu} = \mathcal{J} + \phi_{\bar{u}_j^\nu}$.

Proof. Using (4.35), it can be shown in a similar way as in Lemma 4.7 that

$$\mathcal{N}_j^\circ(u_j^\nu) = \mathcal{N}_j^\circ(\bar{u}_j^\nu) = \mathcal{N}_j^\circ(u_j) \tag{4.37}$$

holds for sufficiently large ν. Together with convergence (4.32), this implies $u_j \in \mathcal{K}_{\bar{u}_j^\nu}$ $\forall \nu \ge \nu_0$ with suitable ν_0. u_j is the global minimizer of $\mathcal{E}_j$. As

$$\mathcal{E}_j(v) = \mathcal{E}_{u_j}(v) = \mathcal{E}_{\bar{u}_j^\nu}(v) \qquad \forall v \in \mathcal{K}_{\bar{u}_j^\nu} \qquad \forall \nu \ge \nu_0 \tag{4.38}$$

u_j must also be the minimizer of $\mathcal{E}_{\bar{u}_j^\nu}$ on $\mathcal{K}_{\bar{u}_j^\nu}$, i.e. the solution of (4.36). $\square$

As a consequence of Lemma 4.12, exact solution of the *constrained smooth problem*

$$u_{\bar u_j^\nu} \in \mathcal{K}_{\bar u_j^\nu} : \qquad \mathcal{E}_{\bar u_j^\nu}(u_{\bar u_j^\nu}) \le \mathcal{E}_{\bar u_j^\nu}(v) \qquad \forall v \in \mathcal{K}_{\bar u_j^\nu} \tag{4.39}$$

and monotone iteration (4.28) with coarse grid correction $\mathcal{C}_j \bar u_j^\nu = u_{\bar u_j^\nu}$ would provide the exact solution u_j after a finite number of steps. We can not expect that the exact solution of (4.39) is available. In order to compute approximate solutions we shall now derive a constrained variant of inexact Newton methods.

Denoting $F_{\bar u_j^\nu} = \mathcal{E}'_{\bar u_j^\nu}$, a *constrained Newton correction* $\Delta \bar u_j^\nu$ can be computed from the variational inequality

$$\begin{aligned} \Delta \bar u_j^\nu \in -\bar u_j^\nu + \mathcal{K}_{\bar u_j^\nu} : \quad \langle F'_{\bar u_j^\nu}(\bar u_j^\nu)\Delta \bar u_j^\nu, v - \Delta \bar u_j^\nu\rangle \ &\ge \\ -\langle F_{\bar u_j^\nu}(\bar u_j^\nu), v - \Delta \bar u_j^\nu\rangle \quad &\forall v \in -\bar u_j^\nu + \mathcal{K}_{\bar u_j^\nu} . \end{aligned} \tag{4.40}$$

Observe that (4.40) reduces to classical Newton linearization in Algorithm 3.5, if ϕ_j is smooth, i.e. $\mathcal{K}_{\bar u_j^\nu} = \mathcal{S}_j$. For fixed constraints $-\bar u_j^\nu + \mathcal{K}_{\bar u_j^\nu}$ (4.40) associates a correction $\Delta \bar u_j^\nu \in \mathcal{S}_j$ to the residual $-F_{\bar u_j^\nu}(\bar u_j^\nu) \in \mathcal{S}'_j$. Inexact corrections

$$\delta \bar u_j^\nu = B_{\bar u_j^\nu}(\bar u_j^\nu)(-F_{\bar u_j^\nu}(\bar u_j^\nu)) \tag{4.41}$$

are obtained from some approximate solution operator $B_{\bar u_j^\nu}(\bar u_j^\nu) : \mathcal{S}'_j \to \mathcal{S}_j$ for (4.40). In analogy to (3.44), κ steps of standard monotone multigrid (cf. Algorithm 4.5) as applied to the obstacle problem (4.40) generate the nonlinear mapping

$$(MG^{\mathrm{STD}}_{\bar u_j^\nu}(\bar u_j^\nu))^\kappa : \mathcal{S}'_j \to \mathcal{S}_j \ .$$

Note that notation is consistent with section 3. Index $\bar u_j^\nu$ and $w \in \mathcal{K}_{\bar u_j^\nu}$ occurring in $B_{\bar u_j^\nu}(w)$ and $MG^{\mathrm{STD}}_{\bar u_j^\nu}(w)$ refer to actual constraints $-\bar u_j^\nu + \mathcal{K}_{\bar u_j^\nu}$ and the argument of Newton linearization $F'_{\bar u_j^\nu}(w)$, respectively.

For simplicity, we only consider the case $\kappa = 1$. A *standard monotone multigrid method* for the iterative solution of (4.11) is finally obtained by

$$\begin{aligned} \bar u_j^\nu &= \mathcal{M}_j^{\mathrm{GS}}(u_j^\nu) \\ u_j^{\nu+1} &= \mathcal{C}_j(\bar u_j^\nu) = \bar u_j^\nu + \omega_j^\nu MG^{\mathrm{STD}}_{\bar u_j^\nu}(\bar u_j^\nu)(-F_{\bar u_j^\nu}(\bar u_j^\nu)) \end{aligned} \tag{4.42}$$

with suitable damping parameter ω_j^ν providing monotonicity (4.29).

As damped corrections remain in $-\bar u_j^\nu + \mathcal{K}_{\bar u_j^\nu}$, i.e. in the smooth regime of ϕ_j, energy estimates of the form (3.35) (cf. Lemma 3.10) are still valid. Hence, the damping strategies as presented in section 3.2 immediately generalize to the constrained smooth case.

Optimality of Newton multigrid with nested iteration (cf. Theorem 3.13) heavily relies on local linear convergence of inexact Newton methods (cf. Corollary 3.8) and on mesh–independent convergence of linear multigrid. Recall that mesh–independent convergence rates for Algorithm 4.5 are not (yet?) available. However, it can be shown that the convergence speed of $MG^{\mathrm{STD}}_{\bar u_j^\nu}(\bar u_j^\nu)$ (cf. Theorem 4.8) asymptotically dominates the convergence speed of the overall iteration (4.42) which is in direct analogy to Newton multigrid methods. Errors are measured with respect to the local energy norm

$$\|v\|_{u_j} = \langle F'_{u_j}(u_j)v, v\rangle^{\frac{1}{2}} \ , \qquad v \in \mathcal{S}_j \ ,$$

which generalizes definition (3.23) to the piecewise smooth case.

Theorem 4.13. *Let $d = 2$ and assume that discrete non–degeneracy (4.35) holds. Assume that damping parameters $\omega_j^\nu \in [0,1]$ are chosen in such a way that monotonicity (4.29) holds and $\omega_j^\nu = 1$ is valid if $\|u_j - u_j^\nu\|_{u_j}$ is sufficiently small. Then the iterates u_j^ν provided by the standard monotone multigrid method (4.42) converge to u_j for each initial iterate $u_j^0 \in \mathcal{S}_j$ and there is ν_0 depending on j such that linear convergence*

$$\|u_j - u_j^{\nu+1}\|_{u_j} \le (1 - C(j+1)^{-2})\|u_j - u_j^\nu\|_{u_j} \qquad \forall \nu \ge \nu_0 \tag{4.43}$$

holds with constant $C > 0$ depending only on on the initial triangulation $\mathcal{T}_0$ on the maximal gradient of α and on the constants α_0 and α_1 in (2.4).

Proof. Global convergence follows from Theorem 4.9 and Lemma 4.11. We only sketch the proof of the asymptotic error estimate (4.43) referring to Kornhuber [67] for details. First note that by Lemma 4.12 and (4.38) the constrained smooth problem (4.39) can be rewritten as the unconstrained smooth problem

$$u_j \in \bar{\mathcal{S}}_j^\circ : \qquad \mathcal{E}_{u_j}(u_j) \le \mathcal{E}_{u_j}(v) \qquad v \in \bar{\mathcal{S}}_j^\circ \tag{4.44}$$

where $\bar{\mathcal{S}}_j^\circ = \{v \in \mathcal{S}_j \mid v(p) = \theta_1 \ \forall p \in \mathcal{N}_j^\bullet(u_j)\}$. Recall from (4.32) that $\bar{u}_j^\nu \to u_j$ and (4.37) implies $\bar{u}_j^\nu \in \bar{\mathcal{S}}_j^\circ$ for sufficiently large ν. Hence, $\Delta\bar{u}_j^\nu \to 0$ for $\nu \to \infty$ and

$$\Delta\bar{u}_j^\nu \in \mathcal{S}_j^\circ = \{v \in \mathcal{S}_j \mid v(p) = 0 \ \forall p \in \mathcal{N}_j^\bullet(u_j)\}$$

holds if ν is large enough. As a consequence, (4.40) reduces asymptotically, i.e. for large ν, to classical Newton linearization

$$\Delta\bar{u}_j^\nu \in \mathcal{S}_j^\circ : \qquad F'_{u_j}(\bar{u}_j^\nu)\Delta\bar{u}_j^\nu = -F_{u_j}(\bar{u}_j^\nu) \tag{4.45}$$

of (4.44) and $MG_{u_j}^{STD}(\bar{u}_j^\nu)$ becomes a linear subspace correction method for (4.45). It can be shown by similar arguments as used in the proof of Theorem 5.6 in [67] that

$$\|v - MG_{u_j}^{STD}(w)F'_{u_j}(w)\|_w \le (1 - c(j+1)^{-2})\|v\|_w \qquad \forall v \in \mathcal{S}_j^\circ$$

holds for uniformly bounded $w \in \bar{\mathcal{S}}_j^\circ$. The constant $c > 0$ depends only on the upper bound of w, on the initial triangulation $\mathcal{T}_0$ on the maximal gradient of α and on the constants α_0 and α_1 in (2.4).

Let ν be large enough to guarantee $\omega_j^\nu = 1$ and choose arbitrary $\varepsilon \in (0,1)$. Utilizing Corollary 3.8 we find ν_0 such that

$$\|u_j - u_j^{\nu+1}\|_{u_j} \le (1 - c(j+1)^{-2} + \varepsilon)\|u_j - \bar{u}_j^\nu\|_{u_j} \qquad \forall \nu \ge \nu_0 \ .$$

Moreover, it is known from [67], Lemma 5.2 that ν_0 can be chosen such that additionally

$$\|u_j - \bar{u}_j^\nu\|_{u_j} \le (1 + \varepsilon)\|u_j - u_j^\nu\|_{u_j} \qquad \forall \nu \ge \nu_0 \ .$$

Finally, select $\varepsilon \in (0,1)$ such that $\varepsilon(j+1)^2 \le \frac{1}{6}c$. Then

$$(1 - c(j+1)^{-2} + \varepsilon)(1 + \varepsilon) \le 1 - \tfrac{1}{2}c(j+1)^{-2}$$

and the assertion follows with $C = \frac{1}{2}c$. □

Similar convergence results hold for *truncated monotone multigrid methods* as resulting from the choice

$$B_{\bar{u}_j^\nu}(\bar{u}_j^\nu) = (MG_{\bar{u}_j^\nu}^{\mathrm{TRC}}(\bar{u}_j^\nu))^\kappa$$

with $\kappa \geq 1$ and truncated obstacle solver $MG^{\mathrm{TRC}}_{\bar{u}_j^\nu}$ as obtained from above modifications of Algorithm 4.5.

It has been observed in many numerical experiments that multigrid convergence (4.43) starts immediately if initial iterates are obtained by nested iteration (see e.g. Kornhuber [64, 65, 66, 67]). As already mentioned above a theoretical justification is the subject of current research.

4.4. **Concluding remarks.** For obstacle problems standard monotone multigrid (cf. Algorithm 4.5) and the truncated variant (cf. Modifications of Algorithm 4.5) were introduced by Mandel [71, 72] and Kornhuber [63], respectively. Note that the truncated method is related to algorithms of Brandt and Cryer [26], see [63] for a comparison. Very recently Tai [84] proved that a related multigrid method based on successive sub*set* decomposition provides logarithmic convergence rates for the energy functional $\mathcal{J}(u_j^\nu)$. Convergence rates are valid from the first iteration. It seems to be an open problem whether similar results hold for the iterates u_j^ν as well.

Though active nodes p, i.e. nodes where the values of the actual iterate and the obstacle coincide, can not be inactivated in course of coarse grid correction, truncated multigrid should not be mixed up with *active set methods*. In contrast to monotone multigrid such algorithms reduce the given obstacle problem to a sequence of *linear problems* of the form (4.22). In this case, linear multigrid can be used for fast iterative solution while approximation of the contact set $\mathcal{N}_j^\bullet(u_j)$ is based on some active set strategy (cf. e.g. Hackbusch and Mittelmann [54], Hoppe [60] or Hoppe and Kornhuber [61]).

Monotone multigrid for piecewise smooth problems (cf. Section 4.3) has been generalized successfully to unbounded derivatives of Φ on (θ_{i-1}, θ_i) and to contact problems with reduced friction in linear elasticity. We refer to Kornhuber [67] and Kornhuber and Krause [68, 69] for further information.

Penalty methods use Newton–type linearization after suitable regularization of ϕ_j (cf. e.g. Glowinski et al. [48]). Regularization parameters have to be selected carefully because reasonable convergence speed might have to be paid by unacceptable discretization errors and vice versa. Dual methods (cf. Glowinski [46], Bertsekas [11], Glowinski and LeTallec [47]) provide very flexible and well–developed tools for the treatment of minimization problems. This approach typically leads to a sequence of saddle point problems for primal variables and Lagrange multipliers which in turn can be solved by suitable multigrid methods (cf. e.g. Bank et al. [10], Braess and Sarazin [18], Zulehner [93]).

References

[1] R.A. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[2] M. Ainsworth and J.T. Oden. *A Posteriori Error Estimation in Finite Element Analysis.* Wiley, Chichester, 2000.

[3] A. Allgower, S. McCormick, and D. Pryor. A general mesh independence principle for Newton's method applied to second order boundary value problems. *Computing*, 23:233–246, 1979.

[4] E. Allgower and K. Böhmer. Application of mesh independence principle to mesh refinement strategies. *SIAM J. Numer. Anal.*, 24:1335–1351, 1987.

[5] E. Allgower, K. Böhmer, and F. Potra. A mesh independence principle for operator equations and their discretizations. *SIAM J. Numer. Anal.*, 23:160–169, 1986.

[6] J.P. Aubin. *Optima and Equilibria.* Springer, Berlin, 1993.

[7] R.E. Bank, T.F. Dupont, and H. Yserentant. The hierarchical basis multigrid method. *Numer. Math.*, 52:387–404, 1988.

[8] R.E. Bank and D.J. Rose. Global approximate Newton methods. *Num. Math.*, 37:279–295, 1981.

[9] R.E. Bank and D.J. Rose. Analysis of a multilevel iterative method for nonlinear finite element equations. *Math. Comp.*, 39:453–665, 1982.

[10] R.E. Bank, B.D. Welfert, and H. Yserentant. A class of iterative methods for solving saddle point problems. *Numer. Math.*, 56:645–666, 1990.

[11] D.P. Bertsekas. *Constrained optimization and Lagrange multiplier methods.* Academic Press, New York, 1984.

[12] J. Bey. *Finite-Volumen- und Mehrgitterverfahren für elliptische Randwertprobleme.* Teubner, Stuttgart, 1998.

[13] F.A. Bornemann and P. Deuflhard. The cascadic multigrid method for elliptic problems. *Numer. Math.*, 75:135–152, 1996.

[14] F.A. Bornemann, B. Erdmann, and R. Kornhuber. Adaptive multilevel methods in three space dimensions. *Int. J. Numer. Meth. Engrg.*, 36:3187–3203, 1993.

[15] F.A. Bornemann and H. Yserentant. A basic norm equivalence in the theory of multilevel methods. *Numer. Math.*, 64:455–476, 1993.

[16] D. Braess. *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics.* Cambridge University Press, Cambridge, 1997.

[17] D. Braess and W. Hackbusch. A new convergence proof for the multigrid method including the V–cycle. *SIAM J. Numer. Anal.*, 20:967–975, 1983.

[18] D. Braess, R. Sarazin. An efficient smoother for the Stokes problem. *Appl. Numer. Math.*, 23:3–19, 1997.

[19] J.H. Bramble. *Multigrid Methods.* Pitman Research Notes in Mathematics. Longman, Harlow, 1993.

[20] J.H. Bramble and J.E. Pasciak. New estimates for multilevel algorithms including the V–cycle. *Math. Comp.*, 60:447–471, 1993.

[21] J.H. Bramble, J.E. Pasciak, J. Wang, and J. Xu. Convergence estimates for multigrid algorithms without regularity assumptions. *Math. Comp.*, 57:23–45, 1991.

[22] J.H. Bramble, J.E. Pasciak, J. Wang, and J. Xu. Convergence estimates for product iterative methods with application to domain decomposition. *Math. Comp.*, 57:1–21, 1991.

[23] J.H. Bramble and J. Xu. Some estimates for a weighted L^2 projection. *Math. Comp.*, 56:463–476, 1991.

[24] A. Brandt. Multi–level adaptive solutions to boundary–value problems. *Math. Comp.*, 31:333–390, 1977.

[25] A. Brandt. Guide to multigrid development. In Hackbusch and Trottenberg [57], pages 220–312.

[26] A. Brandt and C.W. Cryer. Multigrid algorithms for the solution of linear complementary problems arising from free boundary problems. *SIAM J. Sci. Stat. Comput.*, 4:655–684, 1983.

[27] S. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods.* Springer, New York, 1994.

[28] H. Brézis, M.G. Crandall, and A. Pazy. Perturbations of maximal monotone sets in Banach space. *Commun. Pure Appl. Math.*, 23:123–144, 1970.

[29] P.G. Ciarlet. Basic error estimates for elliptic problems. In *Finite Element Methods (Part 1)*, volume II of *Handbook of Numerical Analysis*, pages 17–351, Amsterdam, 1991. North–Holland.

[30] W. Dahmen and A. Kunoth. Multilevel preconditioning. *Numer. Math.*, 63:315–344, 1992.

[31] K. Deimling. *Nonlinear Analysis.* Springer, Berlin, 1985.

[32] R.S. Dembo, S.C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19:401–408, 1982.

[33] P. Deuflhard. *Newton-Techniques for Nonlinear Problems.* Springer, Berlin, Heidelberg, in preparation.

[34] P. Deuflhard, P. Leinen, and H. Yserentant. Concepts of an adaptive hierarchical finite element code. *IMPACT Comput. Sci. Engrg.*, 1:3–35, 1989.

[35] P. Deuflhard and F.A. Potra. Asymptotic mesh independence of Newton-Galerkin methods via a refined Mysovskii theorem. *SIAM J. Numer. Anal.*, 29(5):1395–1412, 1992.

[36] P. Deuflhard and M. Weiser. Local inexact Newton multilevel FEM for nonlinear elliptic problems. In G. Etgen, W. Fitzgibbon, J.L. Lions, J. Périaux, and M. Wheeler, editors, *Computational Science for the 21st Century*, pages 129–138. Wiley, 1997.

[37] P. Deuflhard and M. Weiser. Global inexact Newton multilevel FEM for nonlinear elliptic problems. In W. Hackbusch and G. Wittum, editors, *Multigrid methods V*, Lecture Notes in Computational Science and Engineering, Berlin, Heidelberg, 1998. Springer.

[38] M. Dryja and W. Hackbusch. On the nonlinear domain decomposition method. *BIT*, 32:296–311, 1997.

[39] M. Dryja and O.B. Widlund. Multilevel additive methods for elliptic finite element problems. In W. Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations*, Braunschweig, 1991. Vieweg.

[40] G. Duvaut and J.L. Lions. *Les inéquations en mécanique et en physique.* Dunaud, Paris, 1972.

[41] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* North–Holland, Amsterdam, 1976.

[42] C.M. Elliott. On the finite element approximation of an elliptic variational inequality arising from an implicit time discretization of the Stefan problem. *IMA J. Numer. Anal.*, 1:115–125, 1981.

[43] C.M. Elliott and J.R. Ockendon. *Weak and Variational Methods for Moving Boundary Problems.* Volume 53 of *Research Notes in Mathematics.* Pitman, London, 1982.

[44] R.S. Falk. Error estimates for the approximation of a class of variational inequalities. *Math. Comp.*, 28:963–971, 1974.

[45] A. Friedman. *Variational Principles and Free Boundary Problems.* Wiley, New York, 1982.

[46] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems.* Springer, New York, 1984.

[47] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator–Splitting Methods in Nonlinear Mechanics.* SIAM, Philadelphia, 1989.

[48] R. Glowinski, J.L. Lions, and Trémolières. *Numerical Analysis of Variational Inequalities.* North–Holland, Amsterdam, 1981.

[49] W. Hackbusch. On the fast solutions of nonlinear elliptic equations. *Numer. Math.*, 32:83–95, 1979.

[50] W. Hackbusch, Error analysis of the nonlinear multi-grid method of the second kind. *Apl. Mat.*, 26:18–29, 1981.

[51] W. Hackbusch. Multi–grid convergence theory. In Hackbusch and Trottenberg [57].

[52] W. Hackbusch. *Multi-Grid Methods and Applications.* Springer, Berlin, 1985.

[53] W. Hackbusch. Comparison of different multi–grid variants for nonlinear equations. *ZAMM*, 72:148–151, 1992.

[54] W. Hackbusch and H.D. Mittelmann. On multigrid methods for variational inequalities. *Numer. Math.*, 42:65–76, 1983.

[55] W. Hackbusch and A. Reusken. On Global Multigrid Convergence for Nonlinear Problems. In Hackbusch, W. (ed.) *Robust Multigrid Methods,* pages 105–113. Vieweg, 1988.

[56] W. Hackbusch and A. Reusken. Analysis of a damped nonlinear multilevel method. *Numer. Math.*, 55:225–246, 1989.

[57] W. Hackbusch and U. Trottenberg, editors. *Multigrid Methods*, volume 960 of *Lecture Notes in Mathematics*, Berlin, 1982. Springer.

[58] B. Heise. Nonlinear field calculations with multigrid-Newton methods. *IMPACT Comput. Sci. Eng.*, 5:75–110, 1992.

[59] B. Heise. Analysis of a fully discrete finite element method for a nonlinear magnetic field problem. *SIAM J. Numer. Anal.*, 31:745–759, 1994.

[60] R.H.W. Hoppe. Multigrid algorithms for variational inequalities. *SIAM J. Numer. Anal.*, 24:1046–1065, 1987.

[61] R.H.W. Hoppe and R. Kornhuber. Adaptive multilevel–methods for obstacle problems. *SIAM J. Numer. Anal.*, 31(2):301–323, 1994.

[62] C.T. Kelley. *Iterative methods for linear and nonlinear equations.* SIAM, Philadelphia, 1996.

[63] R. Kornhuber. Monotone multigrid methods for elliptic variational inequalities I. *Numer. Math.*, 69:167 – 184, 1994.

[64] R. Kornhuber. Monotone multigrid methods for elliptic variational inequalities II. *Numer. Math.*, 72:481 – 499, 1996.

[65] R. Kornhuber. *Adaptive Monotone Multigrid Methods for Nonlinear Variational Problems.* Teubner, Stuttgart, 1997.

[66] R. Kornhuber. On robust multigrid methods for non–smooth variational problems. In W. Hackbusch and G. Wittum, editors, *Multigrid Methods V*, pages 173–188, Berlin, 1998. Springer.

[67] R. Kornhuber. On constrained Newton linearization and multigrid for variational inequalities. Preprint Nr. A–2–01, Freie Universität Berlin, 2001, submitted to Numer. Math. .

[68] R. Kornhuber and R. Krause. On monotone multigrid methods for the Signorini problem. Preprint Nr. A–9–99, Freie Universität Berlin, 1999.

[69] R. Kornhuber and R. Krause. Adaptive multigrid methods for Signorini's problem in linear elasticity. Preprint Nr. A–15–2000, Freie Universität Berlin, 2000, to appear in Comp. Vis. Sci. .

[70] R. Kornhuber and H. Yserentant. Multilevel methods for elliptic problems on domains not resolved by the coarse grid. *Contemporary Mathematics*, 180:49–60, 1994.

[71] J. Mandel. A multilevel iterative method for symmetric, positive definite linear complementarity problems. *Appl. Math. Optimization*, 11:77–95, 1984.

[72] J. Mandel. Étude algébrique d'une méthode multigrille pour quelques problèmes de frontière libre. *C.R. Acad. Sci.*, 298:469–472, 1984. Ser. I.

[73] N. Neuss. V–cycle convergence with unsymmetric smoothers and application to an anisotropic model problem. *SIAM J. Numer. Anal.*, 35:1201–1212, 1998.

[74] J.M. Ortega and W.C. Rheinboldt. *Iterative solution of nonlinear equations in several variables.* Academic Press, Orlando, 1970.

[75] P. Oswald. On function spaces related to finite element approximation theory. *Z. Anal. Anwend.*, 9:43–64, 1990.

[76] P. Oswald. Stable subspace splittings for Sobolev spaces and their applications. Forschungsergebnisse der Friedrich-Schiller-Universität Jena Math/93/7, Jena, 1993.

[77] R. Rannacher. On the convergence of Newton–Raphson method for strongly nonlinear finite element equations. In P. Wriggers and W. Wagner, editors, *Nonlinear Computational Mechanics*, Providence, 1991. Springer.

[78] P.A. Raviart and J.M. Thomas. *Introduction a l'analyse numérique des équations aux dérivées partielles.* Masson, Paris, 1992.

[79] M. Renardy and R.C. Rogers. *An Introduction to Partial Differential Equations.* Springer, New York, 1992.

[80] A. Reusken. Convergence of the multigrid full approximation scheme for a class of elliptic mildly nonlinear boundary value problems. *Numer. Math.*, 52:251–277, 1988.

[81] A. Reusken. Convergence of the multilevel full approximation scheme including the V–cycle. *Numer. Math.*, 53:663–686, 1988.

[82] J.F. Rodrigues. *Obstacle Problems in Mathematical Physics.* North–Holland, Amsterdam, 1987.

[83] B. Smith, P. Bjørstad, and W. Gropp. *Domain Decomposition.* Cambridge University Press, 1996.

[84] X.C. Tai. Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. Preprint, University of Bergen, Norway, 2000.

[85] X.C. Tai and J. Xu. Global and uniform convergence of subspace correction methods for some convex optimization methods. Preprint, Pennsylvania State University, 1999.

[86] R. Verfürth. *A Review of a Posteriori Error Estimation and Adaptive Mesh–Refinement Techniques.* Wiley-Teubner, Chichester, 1996.

[87] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34:581–613, 1992.

[88] J. Xu. Two–grid discretization techniques for linear and non–linear pde. *SIAM J. Numer. Anal.*, 33:1759–1777, 1996.

[89] J. Xu. An introduction to multilevel methods. In M. Ainsworth, J. Levesley, W.A. Light, and M. Marletta, editors, *Wavelets, Multilevel Methods and Elliptic PDEs*, pages 213–302, Oxford, 1997. Clarendon Press.

[90] J. Xu. The method of subspace corrections. *J. Comp. Appl. Math.*, to appear.

[91] H. Yserentant. On the multi-level splitting of finite element spaces. *Numer. Math.*, 49:379–412, 1986.

[92] H. Yserentant. Old and new convergence proofs for multigrid methods. *Acta Numerica*, pages 285–326, 1993.
[93] W. Zulehner. A class of smoothers for saddle point problems. Technical report, Fakultät Mathematik, Universität Linz, November 1998.

FU Berlin, Math. Institut II, Arnimallee 2-6, D-14195 Berlin, Germany
E-mail address: kornhuber@math.fu-berlin.de