MULTIGRID METHODS FOR OBSTACLE PROBLEMS

CARSTEN GRÄSER AND RALF KORNHUBER

ABSTRACT. In this review, we intend to clarify the underlying ideas and the relations between various multigrid methods ranging from subset decomposition, to projected subspace decomposition and truncated multigrid. In addition, we present a novel globally convergent inexact active set method which is closely related to truncated multigrid. The numerical properties of algorithms are carefully assessed by means of a degenerate problem and a problem with a complicated coincidence set.

1. INTRODUCTION

Since the pioneering papers of Fichera [35] and Stampaccia [77] almost fifty years ago, variational inequalities have proved extremely useful for the mathematical description of a wide range of phenomena in material science, continuum mechanics, electrodynamics, hydrology and many others. We refer to the monographs of Baiocchi and Capelo [3], Cottle et al. [28], Duvaut and Lions [32], Glowinski [37] or Kinderlehrer and Stampaccia [50] for an introduction. Even the special case of obstacle problems covers a large and still growing number of applications ranging from contact problems in continuum mechanics to option pricing in computational finance or phase transitions in metallurgy (cf., e.g., Rodrigues [72]). In addition, the fast algebraic solution of discretized versions of highly nonlinear partial differential equations or related variational inequalities can be often traced back to a sequence of obstacle problems playing the same role as linear problems in classical Newton linearization [52, 54, 55, 58]. Finally, apart from their practical relevance, obstacle problems are fascinating mathematical objects of their own value which inherit some, but far from all essential properties from their unconstrained counterparts.

On this background, many approaches for the iterative solution of obstacle problems have been suggested and pursued. Penalty methods based on straightforward regularization are still popular in the engineering community. A mathematically well-founded approach is to incorporate the constraints by Lagrange multipliers [37]. It is an advantage of this approach that very general constraints can be treated in a systematic way. On the other hand it doubles the number of unknowns and leads to indefinite problems. Active set strategies consist of an activation/inactivation step that produces an actual guess for the coincidence set and a subsequent solution step for the resulting reduced linear problem. This concept has been very popular since the benchmarking work by Hackbusch and Mittelmann [40] and Hoppe [45, 46]. Recent new interest was stimulated by a reinterpretation of the active set approach

The authors gratefully acknowledge the support of Xue-Cheng Tai by stimulating discussions and valuable suggestions. This work has been funded in part by the Deutsche Forschungsgemeinschaft under contract Ko 1806/3-2.

in terms of nonsmooth Newton methods [43, 82]. As the existing convergence theory typically requires the exact solution of the linear subproblems the combination with inexact (multigrid) solvers is often performed on a heuristic level [48, 49, 61].

In this review we concentrate on extensions of classical multigrid methods to self-adjoint elliptic obstacle problems with box-constraints. Our aim is to bridge the gap between the underlying simple ideas motivated by linear subspace decomposition and detailed descriptions of the final implementation as multigrid V-cycles. We also intend to clarify the relations between different concepts ranging from subset decomposition [78], projected subspace decomposition [1, 2, 63] to monotone multigrid [51] and even active set strategies both with regard to convergence analysis and numerical properties. In particular, we propose a novel truncated nonsmooth Newton multigrid method which can be as well regarded as an inexact active set algorithm or a slight modification of truncated monotone multigrid. Activation/inactivation is performed by a projected Gauß-Seidel step, linear solution is replaced by just one truncated multigrid step (cf. Kornhuber and Yserentant [60]) and global convergence is achieved by damping.

Roughly speaking, it turns out that increasing flexibility goes with decreasing theoretical coverage ranging from multigrid convergence rates for multilevel subset decomposition or projected multilevel relaxation to strong mesh-dependence of truncated monotone multigrid or truncated nonsmooth Newton multigrid for badly chosen initial iterates. On the other hand, increasing flexibility seems to increase the convergence speed considerably in the case of reasonable initial iterates: Combined with, e.g., nested iteration, truncated monotone multigrid or truncated nonsmooth Newton multigrid methods converge even for complicated coincidence sets with similar convergence speed as classical linear multigrid methods for unconstrained problems. The lack of robustness of truncated monotone multigrid or active set strategies is that local inactivation by projected Gauß-Seidel or related strategies [43] might deteriorate the convergence speed, because slow next-neighbor interaction might dominate for overestimated coincidence sets. As a natural remedy, we also propose hybrid methods where local activation/inactivation is replaced by a global standard monotone multigrid step. In our numerical experiments, hybrid version prove extremely efficient for degenerate problems.

2. Continuous problem and discretization

2.1. **Constrained minimization, variational inequalities, and finite elements.** Let Ω be a bounded, polyhedral domain in the Euclidean space $\mathbb{R}^d$, $d = 1, 2, 3$ and let $H \subset H^1(\Omega)$ be a closed subspace. We consider the minimization problem

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

with the closed, convex, and non-empty set $\mathcal{K}$,

$$\mathcal{K} = \{v \in H \mid v \geq \varphi \ \text{ a.e. in } \Omega\} \subset H,$$

as generated by a suitable obstacle function $\varphi \in H^1(\Omega) \cap C(\overline{\Omega})$. We emphasize that all algorithms and convergence results to be presented can be generalized to sets $\mathcal{K}$ where also an upper obstacle is present. The energy functional $\mathcal{J}$,

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

is induced by a symmetric, H-elliptic, bilinear form $a(\cdot,\cdot)$ and a functional $\ell \in H'$. For simplicity, we restrict our considerations to

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

with $f \in L^2(\Omega)$ and $H = H_0^1(\Omega)$. Obviously, $a(\cdot,\cdot)$ defines a scalar product on H, and, by the Poincarè–Friedrichs inequality, the corresponding *energy norm*

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

is equivalent to the canonical norm in $H^1(\Omega)$, i.e.,

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

holds with a positive constant $\alpha \in \mathbb{R}$. The minimization problem (2.1) has a unique solution (cf., e.g., Stampacchia [77] or Glowinski [37, Section I.3]) and can be equivalently rewritten as the variational inequality

$$u \in \mathcal{K}: \qquad a(u, v-u) \ge \ell(v-u) \qquad \forall v \in \mathcal{K}. \tag{2.5}$$

Introducing the *coincidence set* $\Omega^\bullet$,

$$\Omega^\bullet = \{x \in \Omega \mid u(x) = \varphi(x)\}, \quad \Omega^\circ = \Omega \setminus \Omega^\bullet,$$

it turns out that u is the weak solution of the reduced linear elliptic problem

$$-\Delta u(x) = f(x) \qquad \forall x \in \Omega^\circ \tag{2.6}$$

with boundary values $u(x) = 0$ for $x \in \partial\Omega^\circ \cap \partial\Omega$ and $u(x) = \varphi(x)$ elsewhere. We emphasize that the coincidence set $\Omega^\bullet$ or, equivalently, the reduced computational domain is not known a priori.

For sufficiently regular data, e.g., for $f \in L^2(\Omega)$, $\varphi \in H^2(\Omega)$, and convex domains Ω, the solution satisfies $u \in H^2(\Omega)$ [72, Corallary 5:2.3]. In general, second order derivatives of u jump across the free boundary $\Gamma = \overline{\Omega^\bullet} \cap \overline{\Omega^\circ}$. Therefore, in contrast to linear elliptic problems, the regularity of u is limited to $u \in H^s(\Omega)$ with $s < 2.5$ even for arbitrarily smooth data. See Brézis [24] for a more general result. Regularity or stability of the free boundary Γ is considered in the monograph by Rodrigues [72].

Let us now consider a multilevel 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 a nested sequence

$$\mathcal{T}_0 \subset \mathcal{T}_1 \subset \cdots \subset \mathcal{T}_j$$

of triangulations resulting from successive refinement of a given triangulation $\mathcal{T}_0$. We assume that $\mathcal{T}_0$ is shape regular in the sense that it consists of a finite, intentionally small, number of non-degenerate simplices. Though, if not explicitly stated otherwise, all algorithms and theoretical results to be presented can be extended to adaptive refinement, we assume, for simplicity, 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 $\mathcal{T}_k$. In this way, lower and upper bounds of the interior angles are preserved in course of refinement. Such a stable decomposition of each tetrahedron into eight sub-tetrahedra is more complicated (cf., e.g., Bey [7] or Bornemann et al. [11]). Introducing the step sizes h_k,

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

we then get

$$h_j = \mathcal{O}(2^{-j}), \qquad ch_k \le \tfrac{1}{2}h_{k-1} \le Ch_k, \quad k = 1, \dots, j, \tag{2.7}$$

with positive constants c, C independent of k and j. Note that $c = C = 1$ for $d = 2$. On each level k, we choose *piecewise linear finite elements*

$$\mathcal{S}_k = \{v \in H \mid v|_t \text{ is linear } \forall t \in \mathcal{T}_k\},$$

in order to obtain a *nested sequence of finite dimensional subspaces*

$$\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_j \subset H. \tag{2.8}$$

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

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

$\mathcal{N}_k$ is denoting the set of the n_k vertices of $\mathcal{T}_k$ that are contained in Ω. The finite element approximation $u_j \in \mathcal{S}_j$ of u is obtained by replacing the set $\mathcal{K}$ by its discrete analogue $\mathcal{K}_j$,

$$\mathcal{K}_j = \{v \in \mathcal{S}_j \mid v \ge \varphi_j \text{ in } \Omega\}, \tag{2.9}$$

generated by the nodal interpolation $\varphi_j \in \mathcal{S}_j$ of φ. The resulting discrete minimization problem

$$u_j \in \mathcal{K}_j : \qquad \mathcal{J}(u_j) \le \mathcal{J}(v) \qquad \forall v \in \mathcal{K}_j \tag{2.10}$$

is equivalent to the variational inequality

$$u_j \in \mathcal{K}_j : \qquad a(u_j, v - u_j) \ge \ell(v - u_j) \qquad \forall v \in \mathcal{K}_j. \tag{2.11}$$

Introducing the discrete coincidence set $\mathcal{N}_j^\bullet$,

$$\mathcal{N}_j^\bullet = \{p \in \mathcal{N}_j \mid u_j(p) = \varphi(p)\}, \quad \mathcal{N}_j^\circ = \mathcal{N}_j \setminus \mathcal{N}_j^\bullet,$$

it is easily checked that u_j is the solution of the reduced linear problem

$$a(u_j, v) = \ell(v) \qquad \forall v \in \mathcal{S}_j^\circ = \{v \in \mathcal{S}_j \mid v(p) = 0 \; \forall p \in \mathcal{N}_j^\bullet\} \tag{2.12}$$

which is a discrete analogue of (2.6). The discrete problem (2.11) is called *non–degenerate*, if the condition

$$\ell(\lambda_p^{(j)}) - a(u_j, \lambda_p^{(j)}) < 0 \qquad \forall p \in \mathcal{N}_j^\bullet \tag{2.13}$$

is satisfied. As $\ell(\lambda_p^{(j)}) - a(u_j, \lambda_p^{(j)}) \le 0$, $p \in \mathcal{N}_j$, follows immediately from (2.11), condition (2.13) states that u_j must *not* fulfill the discretized Poisson equation in active nodes $p \in N_j^\bullet$. For non–degenerate problems, sufficiently small perturbations of the right hand side ℓ preserve the coincidence set $\mathcal{N}_j^\bullet$. Similar conditions for the continuous problem (2.5) provide the stability of the continuous free boundary Γ. We refer to Rodrigues [72, Section 6:5] for details.

The optimal error estimate $\|u - u_j\| = \mathcal{O}(h_j)$ holds for $u \in H \cap H^2(\Omega)$, $f \in L^2(\Omega)$, and $\varphi \in H^2(\Omega)$ (cf. Falk [34] or Ciarlet [26, Section 5.1]). First steps towards optimal L^2–error estimates have been made by Natterer [64]. Limited regularity of u is reflected by limited order of the discretization error. More precisely, even for arbitrarily smooth data the discretization error of piecewise quadratic finite elements only behaves like $\mathcal{O}(h_j^s)$ with $s < 1.5$ (cf. Brezzi et al. [25]).

For many practical problems, in particular in three space dimensions, it is absolutely necessary to use locally refined grids in order to reduce the number of unknowns and therefore the numerical complexity. *A posteriori estimates of the* *discretization error* providing appropriate local refinement indicators have been investigated by Veeser [83], Bartels and Carstensen [5], Kornhuber [53], Braess [13] and others. Nochetto et al. [66] derived a posteriori estimates of the coincidence set $\Omega^\bullet$ by so–called barrier sets. Convergence proofs for adaptive finite element methods have been considered by Perez et al. [68], Siebert and Veeser [75] and Braess et al. [14].

3. Linear Subspace Decomposition Methods

3.1. **Spectral properties of elliptic bilinear forms.** In this section, we consider the extreme case of an empty coincidence set $\Omega^\bullet = \emptyset$. Obviously, the reduced Poisson problem (2.6) then simplifies to the variational equality

$$u \in H : \qquad a(u,v) = \ell(v) \qquad \forall v \in H \tag{3.1}$$

which is equivalent to the unconstrained minimization problem

$$u \in H : \qquad \mathcal{J}(u) \le \mathcal{J}(v) \qquad \forall v \in H. \tag{3.2}$$

Let us state a fundamental property of elliptic bilinear forms.

Proposition 3.1. *A symmetric, H–elliptic bilinear form $a(\cdot,\cdot)$ has a countable number of positive, real eigenvalues μ_k,*

$$0 < \mu_1 \le \mu_2 \le \mu_3 \le \cdots ,$$

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

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

form an a–orthogonal basis of H.

For a proof we refer, e.g., to Raviart and Thomas [70, pp. 135] or Renardy and Rogers [71, pp. 299].

Example 3.2. *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 3.1 implies that there is an a–orthogonal splitting

$$H = V_1 + V_2 + V_3 \cdots , \qquad V_k = \operatorname{span}\{e_k\}, \tag{3.3}$$

into subspaces V_k representing a scale of increasing frequencies μ_k. Now let $u^0 \in H$ be some guess of u. Due to the a–orthogonality of the splitting (3.3), the corrections $v_k \in V_k$ as obtained by minimizing $\mathcal{J}$ *separately on each subspace* V_k provide the exact solution:

$$u = u^0 + \sum_{k=1}^{\infty} v_k, \qquad v_k = \frac{\ell(e_k) - a(u^0, e_k)}{a(e_k, e_k)} e_k. \tag{3.4}$$

Unfortunately, the eigenfunctions e_k are usually not known in practice.

3.2. **Successive subspace correction and multigrid.** We now concentrate on fast solvers for the *Ritz–Galerkin* approximation

$$u_j \in \mathcal{S}_j : \qquad a(u_j, v) = \ell(v) \qquad \forall v \in \mathcal{S}_j \tag{3.5}$$

or, equivalently,

$$u_j \in \mathcal{S}_j : \qquad \mathcal{J}(u_j) \le \mathcal{J}(v) \qquad \forall v \in \mathcal{S}_j \tag{3.6}$$

of the continuous problem (3.1) or (3.2), respectively. Similar to (3.4), we can solve (3.6) *separately on a–orthogonal subspaces* $V_k = \operatorname{span}\{e_k^{(j)}\}$, $k = 1, \dots, n_j$, spanned by the eigenfunctions $e_k^{(j)}$ of $a(\cdot,\cdot)|_{\mathcal{S}_j \times \mathcal{S}_j}$. This observation gave rise to well–known fast Fourier methods (cf., e.g., [69, Section 19.4]). Unfortunately, the explicit construction of $e_k^{(j)}$ is restricted to differential operators with constant coefficients and rectangular or cuboid domains Ω. However, in the light of Proposition 3.1, eigenfunctions must represent a scale of frequencies.

Therefore, if certain subspaces V_l represent a scale of frequencies, then the corresponding splitting $\mathcal{S}_j = V_1 + V_2 + \cdots + V_m$ might be "almost" a–orthogonal in some sense. Hence, successive minimization of energy $\mathcal{J}$ on V_l should provide a fast solver. This idea is our starting point for the construction of multigrid methods.

Let

$$\mathcal{S}_j = V_1 + V_2 + \cdots + V_m \tag{3.7}$$

be some splitting of $\mathcal{S}_j$. Then, successive minimization of energy $\mathcal{J}$ on V_l 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 3.3. *(Successive minimization)*
given: $w_0 = u_j^\nu \in \mathcal{S}_j$
for $l = 1, \dots, m$ *do:*
{
solve:
$v_l \in V_l : \quad \mathcal{J}(w_{l-1} + v_l) \le \mathcal{J}(w_{l-1} + v) \quad \forall v \in V_l$ *(local minimization)*

$w_l = w_{l-1} + v_l$ *(intermediate iterates)*
}

new iterate: $u_j^{\nu+1} = w_m = u_j^\nu + \sum_{l=1}^{m} v_l$

We now try to choose the splitting (3.7) in such a way that Algorithm 3.3 generates an iterative scheme with mesh independent convergence rates.

The subproblems for the corrections v_l can be easily solved in case of one–dimensional subspaces V_l. Straightforward nodal splitting

$$\mathcal{S}_j = \sum_{l=1}^{n_j} V_l, \qquad V_l = \operatorname{span}\{\lambda_{p_l}^{(j)}\}, \qquad l = 1, \dots, n_j, \tag{3.8}$$

produces the well–known *Gauß–Seidel relaxation*. Obviously, subspaces V_l as used in (3.8) 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 [39, pp. 49]). In order to incorporate low frequencies, we now make use of the hierarchy (2.8).

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

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

Remark 3.4. *The underlying enumeration $l = l(p,k)$ is counting all nodes p on all levels k. Conversely, a given number l characterizes the corresponding pair $(p_l, k_l) = (p,k)(l)$. For example, $\lambda_{l(p,k)} = \lambda_p^k$ or, conversely, $\lambda_l = \lambda_{p_l}^{(k_l)}$. We assume that $l = l(p,k)$ is taken from fine to coarse, i.e., $k > k'$ implies $l(p,k) < l(p,k')$.*

From an heuristic point of view, the multilevel splitting

$$\mathcal{S}_j = \sum_{l=1}^{m_{\mathcal{S}}} V_l, \qquad V_l = \operatorname{span}\{\lambda_l\}, \qquad l = 1, \dots, m_{\mathcal{S}}, \tag{3.9}$$

represents a scale of frequencies ranging from low–frequency functions $\lambda_l \in \Lambda_0$ to high–frequency functions $\lambda_l \in \Lambda_j$. We hope that such functions are "almost" a–orthogonal. In fact, it turns out that the linear independent subset $\hat{\Lambda}$,

$$\hat{\Lambda} = \Lambda_0 \cup \bigcup_{k=1}^{j} \{\lambda_p^{(k)} \mid p \in \mathcal{N}_k \setminus \mathcal{N}_{k-1}\} \subset \Lambda,$$

called *hierarchical basis* of $\mathcal{S}_j$, is in fact a–orthogonal for space dimension $d = 1$. From the multilevel splitting (3.9), Algorithm 3.3 generates the *multilevel relaxation*

$$u_j^{\nu+1} = u_j^{\nu} + \sum_{l=1}^{m_{\mathcal{S}}} v_l, \qquad v_l = \frac{\ell(\lambda_l) - a(w_{l-1}, \lambda_l)}{a(\lambda_l, \lambda_l)} \lambda_l, \tag{3.10}$$

where $w_l = w_{l-1} + v_l$ and $w_0 = u_j^{\nu}$.

The following abstract convergence result is a special case of Theorem 4.4 by Xu [85] (see also Xu and Zikatanov [89]).

Theorem 3.5. *Assume that the splitting* (3.7) *has the following two properties.*

There is a constant $C_0 > 0$ such that for all $v \in \mathcal{S}_j$ there exist $v_l \in V_l$ satisfying

$$v = \sum_{l=1}^{m} v_l, \qquad \sum_{l=1}^{m} \|v_l\|^2 \le C_0 \|v\|^2. \tag{3.11}$$

There is a constant $C_1 > 0$ such that

$$\sum_{l,k=1}^{m} |a(v_l, w_k)| \le C_1 \left(\sum_{l=1}^{m} \|v_l\|^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^{m} \|w_k\|^2 \right)^{\frac{1}{2}} \tag{3.12}$$

holds for all $v_l \in V_l$ and $w_k \in V_k$, $l, k = 1, \dots, m$.

Then the iterates (u_j^ν) produced by Algorithm 3.3 satisfy the error estimate

$$\|u_j^{\nu+1} - u_j\|^2 \le \left(1 - \frac{1}{C_0(1 + C_1)^2}\right) \|u_j^{\nu} - u_j\|^2 \qquad \forall \nu \ge 0. \tag{3.13}$$

For a–orthogonal subspaces V_l the conditions (3.11) and (3.12) clearly hold with $C_0 = C_1 = 1$. Moreover, the Cauchy–Schwartz inequality guarantees that we can always choose $C_1 \le m$. As a first example, we consider the nodal splitting (3.8).

Proposition 3.6. *The nodal splitting* (3.8) *satisfies the conditions* (3.11) *and* (3.12) *with* $C_0 = \mathcal{O}(h_j^{-2})$ *and* $C_1 = \mathcal{O}(1)$.

The iterates (u_j^ν) *produced by the Gauß–Seidel iteration satisfy*

$$\|u_j^{\nu+1} - u_j\|^2 \leq (1 - Ch_j^2)\|u_j^\nu - u_j\|^2 \qquad \forall \nu \geq 0$$

with a constant $C > 0$ *independent of* j.

Proof. Let $v \in \mathcal{S}_j$ and $v_l = v(p_l)\lambda_{p_l}^{(j)}$, $l = 1, \dots, n_j$. Exploiting an inverse inequality and the orthogonality of the basis functions $\lambda_p^{(j)}$ with respect to the lumped L^2–scalar product $\langle v, w\rangle = \sum_{p\in\mathcal{N}_j} v(p)w(p)h_p$ with the associated norm $|\cdot|_0$, we get

$$\sum_{l=1}^{n_j} \|v_l\|^2 \preceq h_j^{-2} \sum_{l=1}^{n_j} \|v_l\|^2_{L^2(\Omega)} \preceq h_j^{-2} \sum_{l=1}^{n_j} |v_l|_0^2 = h_j^{-2}|v|_0^2 \preceq h_j^{-2}\|v\|^2,$$

where $\preceq$ means that the inequality holds up to a generic constant (see [85]). Obviously, $a(\lambda_{p_l}^{(j)}, \lambda_{p_k}^{(j)}) = 0$ holds except for neighboring nodes p_l, p_k. Hence, C_1 is bounded by the maximal number of neighbors which depends on the shape regularity of $\mathcal{T}_0$. The upper bound for the convergence rate now follows directly from Theorem 3.5. □

While condition (3.12) is not an issue for the nodal splitting (3.8), it becomes problematic for the multilevel decomposition (3.9).

Remark 3.7. *The multilevel splitting* (3.9) *satisfies condition* (3.12) *with*

$$C_1 = \mathcal{O}(1) \ \textit{ for } d = 1, \qquad C_1 = \mathcal{O}(j) \ \textit{ for } d = 2, \qquad C_1 = \mathcal{O}(2^{j/2}) \ \textit{ for } d = 3.$$

The proof follows from arguments by Yserentant [90, Lemma 2.7]. Hence, even if we could now satisfy (3.11) with a generic constant C_0, Theorem 3.5 would still not exclude exponential decay of our multilevel relaxation for $d = 3$. One way out of this dilemma is to merge the one-dimensional subspaces $V_l = \operatorname{span}\{\lambda_l\}$ into larger ones. The exact solvability of the resulting larger minimization problems can be preserved by a coloring argument, as we will see later in Section 5.1. Another option is to arrange the computation of the corrections v_l according to the refinement levels in the following way.

We consider the splitting of $\mathcal{S}_j$ into subspaces

$$\mathcal{V}_k = \mathcal{S}_k, \quad k = 0, \dots, j, \tag{3.14}$$

which directly reflects the hierarchy (2.8). Gauß–Seidel relaxation on $\mathcal{V}_k$ gives rise to the bilinear form $b_k(\cdot,\cdot)$,

$$b_k(v, w) = \sum_{\substack{i,l=1\\ i\leq l}}^{n_k} v(p_i)a\big(\lambda_{p_i}^{(k)}, \lambda_{p_l}^{(k)}\big)w(p_l), \qquad v, w \in \mathcal{V}_k, \tag{3.15}$$

where 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 relaxation rapidly reduces high frequency components on $\mathcal{V}_k$, the form $b_k(\cdot,\cdot)$ is called a *smoother* on $\mathcal{V}_k$. Now the multilevel relaxation (3.10) can be rewritten as follows.

Algorithm 3.8. *(Successive subspace correction)*

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

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

$\{$

solve:

$$v_k \in \mathcal{V}_k : \quad b_k(v_k, v) = \ell(v) - a(w_{k+1}, v) \quad \forall v \in \mathcal{V}_k \qquad \textit{(pre–smoothing)}$$

$$w_k = w_{k+1} + v_k \qquad \textit{(intermediate iterates)}$$

$\}$

new iterate: $u_j^{\nu+1} = w_0 = u_j^\nu + \sum_{k=0}^{j} v_k$

Following Xu [85] such kind of algorithms are called *successive subspace correction methods*. The convergence analysis is based on a generalization of Theorem 3.5 to inexact solution on the subspaces $\mathcal{V}_k$ by so-called smoothers $b(\cdot,\cdot)$. Selecting appropriate subspaces $\mathcal{V}_k$ and smoothers $b_k(\cdot,\cdot)$, a large number of multilevel and domain decomposition methods can be reformulated and analyzed in this way.

Exploiting the linearity of the given problem, the intermediate iterates $w_k \in \mathcal{S}_j$ can be eliminated by successive updates of the residual. In this way, Algorithm 3.8 can be formulated as a classical *multigrid V–cycle*.

Algorithm 3.9. *(Multigrid V–cycle with 1 pre–smoothing step)*

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:

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

$$r_k := r_k - a_k(v_k, \cdot) \qquad \textit{(update of the residual)}$$

$$\left.\begin{array}{l} r_{k-1} = r_k|_{\mathcal{S}_{k-1}} \\ a_{k-1}(\cdot,\cdot) = a_k(\cdot,\cdot)|_{\mathcal{S}_{k-1}\times\mathcal{S}_{k-1}} \end{array}\right. \qquad \textit{(canonical restriction)}$$

$\}$

solve:

$$v_0 \in \mathcal{V}_0 : \quad b_0(v_0, v) = r_0(v) \quad \forall v \in \mathcal{V}_0 \qquad \textit{(approx. coarse grid solution)}$$

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

$\{$

$$v_k := v_k + v_{k-1} \qquad \textit{(canonical interpolation)}$$

$\}$

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

Canonical restrictions r_{k-1} and $a_{k-1}(\cdot,\cdot)$ of the residual $r_k = \ell - a(w_{k+1}, \cdot) \in \mathcal{S}_k'$ and the 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. Selecting the nodal basis Λ_k of $\mathcal{S}_k$

and using the canonical isomorphism

$$\mathcal{S}_k \ni v = \sum_{p \in \mathcal{N}_k} v(p)\lambda_p^{(k)} \longleftrightarrow (v(p))_{p \in \mathcal{N}_k} = \underline{v} \in \mathbb{R}^{n_k},$$

all substeps of Algorithm 3.9 are translated into matrix vector operations as available on the computer. It is easily checked that each step of Algorithm 3.9 requires $\mathcal{O}(n_j)$ floating point operations (flops). Reverse enumeration of the subspaces $\mathcal{V}_k$ in (3.14), or, equivalently, of the subspaces V_l in (3.9), corresponds to 1 post–smoothing step. The combination of (multiple) pre– and post–smoothing steps or W–cycles can be formulated in a similar way.

Our heuristic reasoning is confirmed by the celebrated mesh–independent convergence of multigrid methods. Extending the proof of Bramble et al. [18] for symmetric smoothers $b_k(\cdot,\cdot)$ to the actual non–symmetric case, the following convergence result was shown by Neuss [65].

Theorem 3.10. *There is a $\rho < 1$ depending only on the shape regularity of $\mathcal{T}_0$ and on the ellipticity constant α in (2.4) such that*

$$\|u_j^{\nu+1} - u_j\| \le \rho \|u_j^{\nu} - u_j\| \qquad \forall \nu \ge 0 \tag{3.16}$$

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

Multigrid steps on level j are more costly than multigrid steps on coarser grids. Hence, it seems reasonable to compute initial iterates $u_k^0 := \tilde{u}_{k-1}$ inductively for each $k = 1, \dots, j$ by a suitable number of multigrid steps on the preceding level. This procedure is called *nested iteration* (cf. Hackbusch [39, Chapter 5]) or *full multigrid* (cf. Brandt [22]). Nested iteration preserves the optimal accuracy $\|u - u_j\| = \mathcal{O}(h_j)$. More precisely, starting with $\tilde{u}_0 = u_0$ and using the stopping criterion

$$\|u_k - \tilde{u}_k\| \le \tfrac{\sigma}{2} \|u_k - u_k^0\| , \qquad k = 1, 2, \dots, j, \tag{3.17}$$

with some constant $\sigma < 1$ independent of k, we finally obtain $\|u - \tilde{u}_j\| = \mathcal{O}(h_j)$ using an overall amount of $\mathcal{O}(n_j)$ flops. Iterative schemes with this property are sometimes called *optimal.* More sophisticated stopping criteria provide optimality even of nested Gauß–Seidel relaxation. This procedure is called *cascadic multigrid* (cf. Bornemann and Deuflhard [10]), or backslash cycle. 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 check the stopping criterion (3.17) *a posteriori estimates of the algebraic error* $\|u_k - u_k^{\nu}\|$ are required. For ρ taken from Theorem 3.10, we immediately get

$$(1+\rho)^{-1}\|u_k^{\nu+1} - u_k^{\nu}\| \le \|u_k - u_k^{\nu}\| \le (1-\rho)^{-1}\|u_k^{\nu+1} - u_k^{\nu}\| \tag{3.18}$$

utilizing the triangle inequality. Obviously, multigrid corrections provide uniform lower and upper bounds of the algebraic error.

3.3. **Concluding remarks.** At first sight, our considerations seem to be more complicated than classical approaches to multigrid (cf., e.g., Hackbusch [38], pp. 17). However, the actual interpretation has the advantage that it suggests direct extensions to obstacle problems later on.

We used a very intuitive notion of frequencies. Analytically, the definition (3.14) of subspaces $\mathcal{V}_k$ is motivated by the property

$$Q_0 \mathcal{S}_j = \mathcal{V}_0, \qquad (Q_k - Q_{k-1})\mathcal{S}_j \subset \mathcal{V}_k, \quad k = 1, \dots, j, \tag{3.19}$$

where Q_k denotes the L^2–projection on $\mathcal{S}_k$. In order to guarantee (3.19) for adaptively refined grids, it is sufficient to choose the subspace

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

spanned only by the *new* nodal basis functions (see the Xu [86, 87] and the references cited therein). Straightforward selection $\mathcal{V}_k = \mathcal{S}_k$ could deteriorate the optimal complexity even up to $\mathcal{O}(n_j^2)$ in case of strongly local refinement.

We come back to the reinterpretation of (3.19) in terms of frequencies. It is well–known (cf., e.g., Bramble and Xu [20]) that

$$\|v - Q_k v\|_{L^2(\Omega)} \le C h_k \|v\| \qquad \forall v \in \mathcal{S}_j \tag{3.20}$$

holds with C independent of k and j. As a consequence of (3.20) 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. In this sense, the subspaces $\mathcal{V}_k$ represent a scale of frequencies. The relation of high frequencies and locality is discussed to some extend in a survey by Xu [88].

Convergence properties of general successive subspace correction methods (cf. Algorithm 3.8) can be analyzed in an abstract framework as developed by Bramble, Pasciak, Wang, and Xu [18, 19], Bramble and Pasciak [17], Dryja and Widlund [31], Xu [85] and others. The underlying arguments were partly anticipated by Yserentant [90] for the special case of hierarchical splittings

$$\mathcal{V}_0 = I_0 \mathcal{S}_j, \qquad \mathcal{V}_k = (I_k - I_{k-1})\mathcal{S}_j, \quad k = 1, \ldots, j, \tag{3.21}$$

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

Using this abstract theory, the most crucial point in the proof of Theorem 3.10 is to show that the splitting (3.14) 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

$$\sum_{k=0}^{j} 4^k \|v_k\|^2_{L^2(\Omega)} \le C_1 \|v\|^2 \tag{3.22}$$

holds with some C_1 independent of j. Note that the approximation property (3.20) together with H^1–stability of Q_0 provides (3.22) with C_1 growing linearly in j. Utilizing the equivalence of norms in suitable Besov and Sobolev spaces, (3.22) was first shown by Oswald [67] and Dahmen and Kunoth [29]. For extensions related to adaptively refined grids, we refer to Bornemann and Yserentant [12] and Bramble and Pasciak [17] or Xu [85]. In contrast to the classical multigrid convergence theory of Hackbusch [38] and Braess and Hackbusch [15] *no additional regularity of u is required* in order to obtain (3.22) and the resulting mesh–independent convergence (3.16). On the other hand, we get no information how multiple smoothing would improve the convergence rate.

Note that the stability of the hierarchical splitting (3.21) deteriorates quadratically in two space dimensions and exponentially for $d = 3$. As a consequence, the convergence rates of associated hierarchical basis multigrid methods deteriorate quadratically and exponentially for two and three space dimensions, respectively (cf. Yserentant [90], Bank, Dupont and Yserentant [4], Deuflhard, Leinen and Yserentant [30]). On the other hand, hierarchical splittings have some advantages concerning complexity and robustness which make them competitive for certain two–dimensional problems.

For further information on successive subspace correction and multigrid, we recommend the monograph of Bramble [16] and the surveys of Xu [85, 87, 88] and Yserentant [91]. The relation of subspace correction and domain decomposition is discussed in some detail by Smith, Bjørstad and Gropp [76].

4. Subset Decomposition Methods

We now concentrate on the obstacle problem

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

as stated in Section 2.1. As minimization over $\mathcal{S}_j$ is now replaced by minimization over $\mathcal{K}_j$, it is natural to replace the decomposition (3.7) of $\mathcal{S}_j$ into subspaces V_l by a related decomposition

$$\mathcal{K}_j = K_1 + K_2 + \cdots + K_m, \qquad K_l \subset V_l, \tag{4.1}$$

of $\mathcal{K}_j$ into closed, convex subsets K_l. We assume that there is a family of mappings $R_l : \mathcal{K}_j \to V_l$ satisfying

$$R_l v \in K_l, \qquad v = \sum_{l=1}^{m} R_l v \qquad \forall v \in \mathcal{K}_j. \tag{4.2}$$

Then successive minimization of the energy $\mathcal{J}$ on K_l leads to the following basic subset decomposition algorithm.

Algorithm 4.1. *(Subset Decomposition Method)*
given: $w_0 = u_j^\nu \in \mathcal{K}_j$.

for $l = 1, \dots, m$ *do:*
{
$D_l = -R_l u_j^\nu + K_l$ *(local defect constraints)*

solve:
$v_l \in D_l : \quad \mathcal{J}(w_{l-1} + v_l) \le \mathcal{J}(w_{l-1} + v) \quad \forall v \in D_l$ *(local minimization)*

$w_l = w_{l-1} + v_l$ *(intermediate iterates)*
}

new iterate: $u_j^{\nu+1} = w_m = u_j^\nu + \sum_{l=1}^{m} v_l$.

By construction, all intermediate iterates w_l are feasible in the sense that $w_l \in \mathcal{K}_j$ and $\mathcal{J}(w_l) \le \mathcal{J}(w_{l-1})$, $l = 1, \dots, m$.

Remark 4.2. *The defect constraints D_l only depend on the given iterate u_j^ν and not on the intermediate corrections v_l.*

As a first example, let us consider the nodal splitting (3.8) of $\mathcal{S}_j$ into subspaces $V_l = \operatorname{span}\{\lambda_{p_l}^{(j)}\}$. In this case, the corresponding subset decomposition

$$\mathcal{K}_j = \sum_{l=1}^{n_j} K_l, \qquad K_l = \{v \in V_l \mid v(p_l) \ge \varphi_j(p_l)\} \tag{4.3}$$

and the restrictions $R_l v = v(p_l)\lambda_{p_l}^{(j)}$ are uniquely determined. Inserting these specifications into the basic Algorithm 4.1, we obtain the well–known *projected Gauß–Seidel relaxation* (cf. Glowinski [37, Chapter V]). The local corrections v_l are given by

$$v_l = \max\left\{ \frac{r_l(\lambda_{p_l}^{(j)})}{a(\lambda_{p_l}^{(j)}, \lambda_{p_l}^{(j)})}, -u_j^\nu(p_l) + \varphi(p_l) \right\} \lambda_{p_l}^{(j)} \tag{4.4}$$

denoting $r_l = \ell - a(w_{l-1}, \cdot) \in \mathcal{S}_j'$. We introduce the corresponding iteration operator $\mathcal{M}_j : \mathcal{K}_j \to \mathcal{K}_j$ defined by

$$\mathcal{M}_j(u_j^\nu) = u_j^\nu + \sum_{l=1}^{n_j} v_l \tag{4.5}$$

for further reference.

Generalizing its unconstrained counterpart, the projected Gauß–Seidel relaxation typically inherits rapidly deteriorating convergence rates for decreasing mesh size. We expect faster convergence from additional corrections representing a scale of frequencies and therefore introduce a multilevel decomposition into the subsets

$$K_l = \{v \in V_l \mid v_l \geq \phi_l\} \subset V_l = \operatorname{span}\{\lambda_l\}, \qquad l = 1, \dots, m_{\mathcal{S}}, \tag{4.6}$$

associated with the multilevel nodal basis functions $\lambda_l = \lambda_{p_l}^{(k_l)}$ (cf. Remark 3.4). Recall that the subspaces $V_l = \operatorname{span}\{\lambda_l\}$ lead to classical multigrid methods as explained in the preceding section. The local obstacles $\phi_l \in V_l$ are chosen in such a way that

$$\varphi_j = \sum_{l=1}^{m_{\mathcal{S}}} \phi_l. \tag{4.7}$$

This decomposition is not unique. However, it will turn out below that the following construction of restriction operators $R_l : \mathcal{K}_j \to K_l$ leads to an algorithm which does not depend on the actual choice of the local obstacles ϕ_l. As a starting point, we define modified interpolation operators $I_k^\ominus : \mathcal{S}_j \to \mathcal{S}_k$, $k = 0, \dots, j$, according to

$$I_k^\ominus v = \sum_{p \in \mathcal{N}_k} \underline{v}_p \lambda_p^{(k)}, \qquad \underline{v}_p = \min\{v(q) \mid q \in \mathcal{N}_j \cap \operatorname{int}\ \operatorname{supp} \lambda_p^{(k)}\}.$$

Obviously, $I_j^\ominus v = v$. Moreover, we have

$$v \geq 0 \;\Rightarrow\; I_k^\ominus v \geq 0, \qquad I_k^\ominus v \geq I_{k-1}^\ominus v \quad \forall v \in \mathcal{S}_j, \tag{4.8}$$

because $\operatorname{int}\ \operatorname{supp} \lambda_p^{(k)} \subset \operatorname{int}\ \operatorname{supp} \lambda_p^{(k-1)}$. For convenience, we set $I_{-1}^\ominus = 0$. Utilizing $l = l(p_l, k_l)$, we now define the restrictions

$$R_l v = \big(I_{k_l}^\ominus(v - \varphi_j) - I_{k_l-1}^\ominus(v - \varphi_j) + \phi_l\big)(p_l)\lambda_l, \quad l = 1, \dots, m_{\mathcal{S}}. \tag{4.9}$$

As a consequence of (4.8), the restrictions R_l satisfy the conditions (4.2).

Inserting K_l and R_l as defined in (4.6) and (4.9), respectively, into the basic Algorithm 4.1, we obtain a *multilevel subset decomposition method*. It was originally proposed and analyzed by Tai [78] under the name *constraint decomposition method*. Here we have described a slight modification by taking the minimum over the nodes $p \in \operatorname{int}\ \operatorname{supp} \lambda_p^{(k)}$ and not over $\operatorname{supp} \lambda_p^{(k)}$ in the definition of $I_k^\ominus$. This modification might lead to slightly faster convergence by slightly less restrictive coarse grid correction and clarifies the relation to monotone multigrid methods to be presented later. Similar to the projected Gauß-Seidel method the corrections provided by multilevel subset decomposition can be evaluated in closed form according to

$$v_l = \max\left\{\frac{r_l(\lambda_l)}{a(\lambda_l,\lambda_l)}, -R_l u_j^\nu(p_l) + \phi_l(p_l)\right\}\lambda_l, \tag{4.10}$$

where p_l is the supporting point of $\lambda_l = \lambda_{p_l}^{(k_l)}$.

The convergence analysis of the multilevel subset decomposition algorithm (4.10) will be based on the following abstract convergence result by Tai [78].

Theorem 4.3. *Assume that the restriction operators R_l fulfill the stability condition*

$$\sum_{l=1}^{m_{\mathcal{S}}} \|R_l v - R_l w\|^2 \le C_0 \|v-w\|^2 \qquad \forall v,w \in \mathcal{K}_j \tag{4.11}$$

with a constant $C_0 \ge 0$ and that the underlying space decomposition (3.7) *satisfies the condition* (3.12)*.*

Then, for any $u_j^0 \in \mathcal{K}_j$ the iterates $(u_j^\nu)_{\nu\ge 0}$ produced by Algorithm 4.1 satisfy the error estimates

$$\mathcal{J}(u_j^{\nu+1}) - \mathcal{J}(u_j) \le \rho\left(\mathcal{J}(u_j^\nu) - \mathcal{J}(u_j)\right) \qquad \forall \nu \ge 0 \tag{4.12}$$

and

$$\|u_j^\nu - u_j\|^2 \le 2\rho^\nu\left(\mathcal{J}(u_j^0) - \mathcal{J}(u_j)\right) \qquad \forall \nu \ge 0 \tag{4.13}$$

with

$$\rho = 1 - \frac{1}{(\sqrt{1+C^*}+\sqrt{C^*})^2}, \qquad C^* = \left(2C_1 + C_0 C_1^2\right). \tag{4.14}$$

Observe that (4.13) follows directly from (4.12) and

$$\mathcal{J}(v) - \mathcal{J}(u_j) \ge \tfrac{1}{2}\|v-u_j\|^2 \qquad \forall v \in \mathcal{K}_j.$$

Remark 4.4. *In the unconstrained case $\mathcal{K}_j = \mathcal{S}_j$, the estimate* (4.11) *is equivalent to* (3.11)*, if the restrictions R_l are linear mappings.*

As a first example, we consider the projected Gauß–Seidel relaxation (4.4).

Proposition 4.5. *The restriction operators R_l induced by the nodal splitting* (4.3) *satisfies condition* (4.11) *with the same constant $C_0 = \mathcal{O}(h_j^{-2})$ as appearing in Proposition 3.6.*

The iterates (u_j^ν) produced by the projected Gauß–Seidel relaxation (4.4) *satisfy*

$$\|u_j^\nu - u_j\|^2 \le 2(1-Ch_j^2)^\nu\left(\mathcal{J}(u_j^0) - \mathcal{J}(u_j)\right) \qquad \forall \nu \ge 0$$

with a constant $C > 0$ independent of j.

Proof. As nodal interpolation $R_l v = v(p_l)\lambda_{p_l}^{(j)}$ is linear, the conditions (4.11) and (3.11) are equivalent. □

Now we concentrate on the multilevel relaxation (4.10). In contrast to the convergence results for linear multigrid methods as presented in the preceding section, the following theorem is restricted to quasiuniform meshes and a suitable ordering of the spaces $V_l = \operatorname{span}\{\lambda_{p_l}^{(k_l)}\}$ on each level k_l. More precisely, we decompose

$$\mathcal{N}_k = \bigcup_{i=1}^{i_0} \mathcal{N}_k^i, \qquad k = 0,\dots,j, \tag{4.15}$$

by coloring the planar graph consisting of the nodes and edges of the triangulation $\mathcal{T}_k$ with i_0 colors. Such a decomposition exists with $i_0 \leq 4$ by the famous four–color theorem. By construction, we have

$$\text{int supp}\, \lambda_p^{(k)} \cap \text{int supp}\, \lambda_q^{(k)} = \emptyset \qquad \forall p, q \in \mathcal{N}_k^i,\ p \neq q. \tag{4.16}$$

We assume that the spaces $V_l = \text{span}\{\lambda_{p_l}^{(k_l)}\}$ on each level k_l are ordered according to the decomposition (4.15) in the sense that the sets $\mathcal{N}_k^i$ are enumerated one after the other. The following convergence result is due to Tai [78].

Theorem 4.6. *Assume that the space dimension is $d = 2$. Then the restriction operators R_l defined in (4.9) fulfill the condition (4.11) with*

$$C_0 = \mathcal{O}(j+1)^2. \tag{4.17}$$

For any $u_j^0 \in \mathcal{K}_j$ the iterates $(u_j^\nu)_{\nu \geq 0}$ produced by multilevel subset decomposition method (4.10) satisfy

$$\|u_j^\nu - u_j\|^2 \leq 2(1 - C(j+1)^{-2})^\nu \left(\mathcal{J}(u_j^0) - \mathcal{J}(u_j)\right) \qquad \forall \nu \geq 0$$

with a constant $C > 0$ independent of j.

Proof. Let v, $w \in \mathcal{K}_j$. Denoting

$$v^{(k)} = I_k^{\ominus}(v - \varphi_j) - I_{k-1}^{\ominus}(v - \varphi_j), \quad w^{(k)} = I_k^{\ominus}(w - \varphi_j) - I_{k-1}^{\ominus}(w - \varphi_j)$$

we can collect the summands according to the levels to obtain

$$\begin{aligned} \sum_{l=1}^{m_{\mathcal{S}}} \|R_l v - R_l w\|^2 &= \sum_{k=0}^{j} \sum_{p \in \mathcal{N}_k} |v^{(k)}(p) - w^{(k)}(p)|^2 \|\lambda_p^{(k)}\|^2 \\ &\preceq \sum_{k=0}^{j} h_k^{-2} \sum_{p \in \mathcal{N}_k} |v^{(k)}(p) - w^{(k)}(p)|^2\, h_k^2 \preceq \sum_{k=0}^{j} h_k^{-2}\, \|v^{(k)} - w^{(k)}\|_{L^2(\Omega)}^2. \end{aligned} \tag{4.18}$$

Note that the local obstacles ϕ_l appearing in (4.9) cancel out each other. Furthermore, we have used that in $d = 2$ space dimensions $\|\lambda_p^{(k)}\|$ is uniformly bounded by a constant depending only on the shape regularity of $\mathcal{T}_0$ and that the weighted sum of squared nodal values is equivalent to the L^2-norm. Now the cornerstone of the proof is the stability estimate

$$\|I_k^{\ominus}(v - \varphi_j) - I_k^{\ominus}(w - \varphi_j) - (v - w)\|_{L^2(\Omega)}^2 \preceq (1 + j - k) h_k \|v - w\|^2 \tag{4.19}$$

stated in Theorem 2 by Tai [78] which can be shown literally in the same way for the slightly modified operators $I_k^{\ominus}$. Using (4.19), we immediately get

$$\|v^{(k)} - w^{(k)}\|_{L^2(\Omega)}^2 \preceq (j - k + 1) h_k^2 \|v - w\|^2.$$

Inserting this estimate into (4.18), we obtain condition (4.11) with $C_0 \preceq (j+1)^2$.

In order to check condition (3.12) we consider the splitting

$$\mathcal{S}_j = \mathcal{V}_0 + \sum_{k=1}^{j} \sum_{i=1}^{i_0} \mathcal{V}_k^i, \qquad \mathcal{V}_0 = \mathcal{S}_0, \quad \mathcal{V}_k^i = \text{span}\{\lambda_p^{(k)} \mid p \in \mathcal{N}_k^i\}. \tag{4.20}$$

As a consequence of (4.16), the subset decomposition method induced by the decomposition

$$K_k^i = \bigcup_{p \in \mathcal{N}_k^i} K_{l(p,k)}, \quad R_k^i = \sum_{p \in \mathcal{N}_k^i} R_{l(p,k)} \qquad k = 0, \ldots, j,\ i = 1, \ldots, i_0, \tag{4.21}$$

is producing exactly the same iterates as the original multilevel method (4.10). Moreover, the decomposition (4.21) directly inherits condition (4.11) with $C_0 \preceq (j+1)^2$ from the original splitting. The reason for this reinterpretation is that for the underlying space decomposition (4.20) the condition (3.12) with $C_1 = \mathcal{O}(1)$ immediately follows from the well–known strengthened Cauchy–Schwarz inequality (cf. Yserentant [90, Lemma 2.7])

$$|a(v_k, w_l)| \preceq \left(\frac{1}{\sqrt{2}}\right)^{|k-l|} \|v_k\|\|w_l\| \qquad \forall v_k \in \mathcal{S}_k,\ w_l \in \mathcal{S}_l.$$

The final error estimate is an immediate consequence of Theorem 4.3. □

The proof of the estimate (4.17) essentially relies on the stability of nodal interpolation and therefore is restricted to $d = 2$ space dimensions. While even mesh-independent bounds are available for $d = 1$ only exponential bounds can be shown in three space dimensions.

We now concentrate on the efficient reformulation of the multilevel subset decomposition method (4.10) as a multigrid $V-$cycle. While the intermediate iterates can be removed in a similar way as in the linear case, we now have to find a way to check the constraints D_l without visiting the fine mesh. For given $u_j^\nu \in \mathcal{K}_j$ the constraints D_l can be rewritten as

$$D_l = -R_l u_j^\nu + K_l = \{z\lambda_l \mid z \geq \psi_l\}$$

where

$$\psi_l = \psi^{(k_l)}(p_l), \qquad \psi^{(k)} = -(I_k^\ominus - I_{k-1}^\ominus)(u_j^\nu - \varphi_j).$$

We emphasize that the choice of the local obstacles ϕ_l in the decomposition (4.6) has no effect on D_l and thus on the whole iteration. It is convenient to introduce the counterparts $I_k^\oplus : \mathcal{S}_j \to \mathcal{S}_k$ of $I_k^\ominus$ by

$$I_k^\oplus v = \sum_{p \in \mathcal{N}_k} \overline{v}_p \lambda_p^{(k)}, \qquad \overline{v}_p = \max\{v(q) \mid q \in \mathcal{N}_j \cap \text{int supp } \lambda_p^{(k)}\}$$

and $I_{-1}^\oplus = 0$. Note that $-I_k^\oplus v = I_k^\ominus(-v)$. The modified interpolation operators $I_j^\oplus$ satisfy the recursion formula

$$I_{k-1}^\oplus v = R_k^{k-1} I_k^\oplus v, \quad k = 1, \dots, j, \qquad I_j^\oplus v = v \qquad \forall v \in \mathcal{S}_j$$

with *monotone restriction operators* $R_k^{k-1} : \mathcal{S}_k \to \mathcal{S}_{k-1}$ defined by

$$\big(R_k^{k-1} v\big)(p) = \max\{v(q) \mid q \in \mathcal{N}_k \cap \text{int supp}\, \lambda_p^{(k-1)}\}, \qquad p \in \mathcal{N}_{k-1}, \tag{4.22}$$

for $k = j, \dots, 1$ and $R_0^{-1} = 0$. As a consequence, the interpolated defect obstacles $\chi^{(k)} = I_k^\oplus(\varphi_j - u_j^\nu)$ can be evaluated recursively according to

$$\chi^{(k-1)} = R_k^{k-1}\chi^{(k)}, \quad k = j, \dots, 0, \qquad \chi^{(j)} = \varphi_j - u_j^\nu.$$

The resulting hierarchical decomposition

$$\varphi_j - u_j^\nu = \sum_{k=0}^{j} \psi^{(k)}, \qquad \psi^{(k)} = \chi^{(k)} - \chi^{(k-1)}$$

of the defect obstacle $\varphi_j - u_j^\nu$ is illustrated in Figure 1. The left picture illustrates the monotone restriction $\chi^{(j-1)}$ (dashed) of a given defect obstacle (solid). Further monotone restriction provides $\chi^{(j-2)}$ as depicted in the right picture. A hierarchical decomposition is obtained from the increments $\psi^{(j)}$, $\psi^{(j-1)}$, and $\psi^{(j-1)}$.

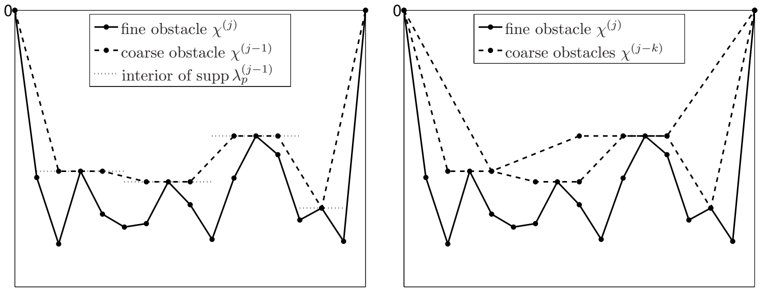

FIGURE 1. Restriction (left) and hierarchical decomposition (right) of defect obstacle

We are now ready to reformulate the multilevel subset decomposition algorithm (4.10) as a multigrid V-cycle with optimal complexity $\mathcal{O}(n_j)$.

Algorithm 4.7. *(Multigrid V–cycle with 1 pre–smoothing step)*

given: u_j^ν

initialize: $r_j = \ell - a(u_j^\nu, \cdot), \quad a_j(\cdot,\cdot) = a(\cdot,\cdot), \quad \chi^{(j)} = \varphi_j - u_j^\nu$

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

{

$\chi^{(k-1)} = R_k^{k-1}\chi^{(k)}$ *(monotone restriction)*

$\mathcal{D}_k = \{v \in \mathcal{S}_k \mid v \geq \psi^{(k)}\}, \quad \psi^{(k)} = \chi^{(k)} - \chi^{(k-1)}$ *(defect obstacles)*

solve:

$v^{(k)} \in \mathcal{D}_k : \quad b_k(v^{(k)}, v - v^{(k)}) \geq r_k(v - v^{(k)}), \quad \forall v \in \mathcal{D}_k$ *(pre-smoothing)*

$r_k := r_k - a_k(v^{(k)}, \cdot)$ *(update of the residual)*

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

}

solve:

$v^{(0)} \in \mathcal{D}_0 : \quad b_0(v^{(0)}, v - v^{(0)}) \geq r_0(v - v^{(0)}), \quad \forall v \in \mathcal{D}_0$ *(initial grid smoothing)*

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

{

$v^{(k)} := v^{(k)} + v^{(k-1)}$ *(canonical interpolation)*

}

new iterate: $u_j^{\nu+1} = u_j^\nu + v^{(j)}$

Algorithm 4.7 can be implemented as a $V(1,0)$-cycle multigrid method with projected Gauß-Seidel smoother and canonical restrictions and prolongations. Only the monotone restrictions R_k^{k-1} of the defect obstacles have to be added to an existing multigrid code. In order to incorporate several pre- or postsmoothing steps the original sets K_l and restrictions R_l are repeated several times after certain modifications preserving the consistency conditions (4.1) and (4.2). For example, a $V(1,1)$ cycle is obtained by $2m_{\mathcal{S}}$ restriction operators $\tilde{R}_l$ which are defined by $\tilde{R}_l = \frac{1}{2}R_l$ for $l = 1, \dots m_{\mathcal{S}}$ and repeated in reverse order for $l = m+1, \dots, 2m_{\mathcal{S}}$. Analogously, the convex sets $\tilde{K}_l$ are generated by $\tilde{\phi}_l = \frac{1}{2}\phi_l$ for $l = 1, \dots m_{\mathcal{S}}$ and repeated in reverse order for $l = m+1, \dots, 2m_{\mathcal{S}}$. We will use this setting in our numerical computations. The convergence analysis of such algorithms can be carried out along the lines explained above.

4.1. **Concluding remarks.** It seems that Algorithm 4.7 is the first multigrid method for obstacle problems that allowed for polylogarithmic bounds for the convergence rates. Moreover, to our knowledge it is still the only algorithm with proven multigrid convergence rates which can be implemented as a V-cycle (cf. Section 5).

The same theoretical framework can be used to analyze Jacobi-like versions of Algorithm 4.1. The advantage of such methods is that the corrections can be computed in parallel, because the update of the intermediate iterates is simply skipped. On the other hand, convergence has to be enforced by damping parameters $\alpha \leq n_j$ which might slow down convergence considerably in comparison with the sequential version. We refer to Tai [42, 78] for details.

Similar to linear subspace decomposition, the abstract convergence result can be also applied to overlapping domain decomposition methods. For further information, we refer to Tai [42, 78, 81] and the references cited therein.

5. Projected subspace decomposition methods

5.1. **Projected relaxation methods.** Another natural extension of linear subspace correction to obstacle problems is to perform successive *constrained* minimization on the subspaces V_l. For a given splitting (3.7) of $\mathcal{S}_j$ such kind of direct extension of Algorithm 3.3 to the obstacle problem (2.10) reads as follows.

Algorithm 5.1. *(Successive minimization)*
given: $w_0 = u_j^\nu \in \mathcal{S}_j$.
for $l = 1, \dots, m$ *do:*
{

$$D_l = \big(- w_{l-1} + \mathcal{K}_j \big) \cap V_l \qquad \textit{(local defect constraints)}$$

solve:

$$v_l \in D_l : \quad \mathcal{J}(w_{l-1} + v_l) \leq \mathcal{J}(w_{l-1} + v) \quad \forall v \in D_l \ \textit{(local minimization)}$$

$$w_l = w_{l-1} + v_l \qquad \textit{(intermediate iterates)}$$

}

new iterate: $u_j^{\nu+1} = w_m = u_j^\nu + \sum_{l=1}^{m} v_l$.

Remark 5.2. *In contrast to the subset decomposition Algorithm 4.7 the defect constraints now depend on the intermediate iterates and thus on the preceding corrections v_l.*

As a first example, we consider the nodal splitting (3.9). In this case, the basic Algorithm 5.1 is again producing the projected Gauß–Seidel relaxation $\mathcal{M}_j$ with local corrections v_l given in (4.4).

In order to generalize classical multigrid methods, we now insert the multilevel splitting (3.9) into Algorithm 5.1. The resulting *projected multilevel relaxation* was suggested by Mandel [36, 62, 63] and later investigated by many authors [1, 2, 42, 51, 81]. The corresponding corrections v_l are given explicitly by

$$v_l = \max\left\{\frac{r_l(\lambda_l)}{a(\lambda_l,\lambda_l)}, \max_{p\in\mathcal{N}_j\cap \text{int supp}\,\lambda_l} \frac{-w_{l-1}(p)+\varphi(p)}{\lambda_l(p)}\right\}\lambda_l. \tag{5.1}$$

The convergence analysis of (5.1) will be based on the following abstract convergence result by Badea, Tai, and Wang [2] for the basic Algorithm 5.1.

Theorem 5.3. *Assume that the splitting* (3.7) *has the following two properties.*

There is a constant $C_0 > 0$ such that for all v, $w \in \mathcal{K}_j$ and $s_l \in V_l$ with $w + \sum_{i=1}^{l-1} s_i \in \mathcal{K}_j$ for $l = 1,\dots,m$ there exist $z_l \in V_l$, $1,\dots,m$, satisfying

$$\begin{aligned} &v - w = \sum_{l=1}^{m} z_l, \qquad w + \sum_{i=1}^{l-1} s_i + z_l \in \mathcal{K}_j, \quad l = 1,\dots,m,\\ &\sum_{l=1}^{m} \|z_l\|^2 \le C_0\left(\|v-w\|^2 + \sum_{l=1}^{m}\|s_l\|^2\right). \end{aligned} \tag{5.2}$$

The condition (3.12) *holds with a constant $C_1 > 0$.*

Then the iterates (u_j^ν) produced by Algorithm 5.1 satisfy the error estimates

$$\mathcal{J}(u_j^{\nu+1}) - \mathcal{J}(u_j) \le \rho\left(\mathcal{J}(u_j^\nu) - \mathcal{J}(u_j)\right) \qquad \forall \nu \ge 0 \tag{5.3}$$

and

$$\|u_j^\nu - u_j\|^2 \le 2\rho^\nu\left(\mathcal{J}(u_j^0) - \mathcal{J}(u_j)\right) \qquad \forall \nu \ge 0 \tag{5.4}$$

with

$$\rho = 1 - \frac{1}{(\sqrt{1+C^*}+\sqrt{C^*})^2}, \qquad C^* = \left(2(1+\sqrt{C_0})C_1 + C_0C_1^2\right). \tag{5.5}$$

Remark 5.4. *In the unconstrained case $\mathcal{K}_j = \mathcal{S}_j$, the condition* (5.2) *is equivalent to condition* (3.11) *with v replaced by $v - w$.*

Let us first apply Theorem 5.3 to the projected Gauß–Seidel relaxation (4.4).

Proposition 5.5. *The nodal splitting* (3.8) *satisfies condition* (5.2) *with the same constant $C_0 = \mathcal{O}(h_j^{-2})$ as appearing in Proposition 3.6.*

The iterates (u_j^ν) produced by the projected Gauß–Seidel relaxation (4.4) *satisfy*

$$\|u_j^\nu - u_j\|^2 \le 2(1 - Ch_j^2)^\nu\left(\mathcal{J}(u_j^0) - \mathcal{J}(u_j)\right) \qquad \forall \nu \ge 0$$

with a constant $C > 0$ independent of j.

Proof. For given v, w and s_l the unique decomposition $z_l = (v(p_l) - w(p_l))\lambda_{p_l}^{(j)}$ clearly satisfies the estimate in (5.2) with the same constant as in (3.11). The remaining condition $w + \sum_{i=1}^{l-1} s_i + z_l \in \mathcal{K}_j$, $l = 1,\dots,n_j$, is automatically fulfilled. □

A polylogarithmic upper bound for the multilevel relaxation (5.1) has been shown just recently by Badea [1].

Theorem 5.6. *Assume that the space dimension is $d = 2$ and that dim $\mathcal{S}_0 = 1$. Assume further that the spaces $V_l = \operatorname{span}\{\lambda_{p_l}^{(k_l)}\}$ on each level k_l are ordered according to the decomposition* (4.15).

Then the multilevel splitting (3.9) *satisfies condition* (5.2) *with*

$$C_0 = \mathcal{O}(j+1)^5. \tag{5.6}$$

The iterates (u_j^ν) produced by the projected multilevel relaxation (5.1) *satisfy*

$$\|u_j^\nu - u_j\|^2 \le 2(1 - C(j+1)^{-5})^\nu \left(\mathcal{J}(u_j^0) - \mathcal{J}(u_j)\right) \qquad \forall \nu \ge 0 \tag{5.7}$$

with a constant $C > 0$ independent of j.

As in the proof of Theorem 4.6, a coloring argument yields condition (3.12) with a constant $C_1 = \mathcal{O}(1)$. The corner stone of the estimate (5.6) of the stability constant C_0 is a stability result for a modified interpolation operator $I_k^+ : \mathcal{S}_j \to \mathcal{S}_k$. Similarly to $I_k^\ominus$ appearing in Algorithm 4.7, the operator I_k^+ (we adopt the notation of Tai [78] and Badea [1], respectively) is defined by

$$I_k^+ v = \sum_{p \in \mathcal{N}_k} \underline{v}_p^+ \, \lambda_p^{(k)}, \quad \underline{v}_p^+ = \min\{v^+(q) \mid q \in \operatorname{supp} \lambda_p^{(k)}\}, \qquad v^+ = \max\{0, v\},$$

and, for $d = 2$, has the stability properties

$$\|I_k^+ v - v\|_{L^2(\Omega)} \preceq h_k^2 (1 + j - k) \|v\|^2,$$
$$\|I_k^+ v\|_{L^2(\Omega)} \preceq \|v\|_{L^2(\Omega)}, \quad \|I_k^+ v\|^2 \preceq (1 + j - k)\|v\|^2.$$

according to Lemma 4.3 in [1]. Due to the well-known Sobolev imbedding theorem even mesh-independent stability holds for $d = 1$ but only exponential bounds are available in three space dimensions. We emphasize that I_k^+ only appears in the proof and not in the algorithm itself.

In the case dim $\mathcal{S}_0 > 1$, the results of Theorem 5.6 hold for the splitting

$$\mathcal{S}_j = \mathcal{S}_0 + \sum_{l = n_0 + 1}^{m_\mathcal{S}} V_l,$$

or, equivalently, for *exact* solution on the coarsest grid.

Remark 5.7. *In contrast to the multilevel subset decomposition Algorithm 4.7 the multilevel relaxation* (5.1) cannot be implemented as a multigrid V-cycle*, because the intermediate iterates w_l enter the defect constraints D_l in an nonlinear way. More precisely, an additional interpolation to the fine grid $\mathcal{T}_j$ is necessary to evaluate the correction at each node p on each refinement level k. Therefore the numerical complexity of each iteration step is ranging from $\mathcal{O}(n_j \log n_j)$ in case of uniform refinement to even $\mathcal{O}(n_j^2)$ for highly locally refined grids.*

5.2. **Monotone multigrid methods.** The multilevel relaxation (5.1) suffers from two drawbacks

- sub-optimal complexity ranging from $\mathcal{O}(n_j \log n_j)$ up to $\mathcal{O}(n_j^2)$.
- poor asymptotic convergence speed due to poor coarse grid correction.

While the first issue has already been addressed in Remark 5.7, the second one requires further explanation. Assume, for the moment, that the coincidence set

$\mathcal{N}_j^\bullet$ is known so that we are left with the linear reduced problem (2.12). Then, all subspaces V_l with the property

$$\text{int supp}\,\lambda_l \cap \mathcal{N}_j^\bullet \neq \emptyset$$

must not contribute to the correction, because $\lambda_l \notin \mathcal{S}_j^\circ$. Such kind of poor coarse grid correction often slows down the asymptotic convergence speed of the multilevel relaxation (5.1) in comparison with the unconstrained case. Following Kornhuber [51], we therefore iteratively adapt the multilevel nodal basis functions λ_l to the reduced space $\mathcal{S}_j^\circ$. More precisely, we introduce so–called *truncated nodal basis functions* λ_l, $l > n_j$, according to

$$\tilde{\lambda}_l(p) = \begin{cases} 0 & \text{if } p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu) \\ \lambda_l(p) & \text{else} \end{cases}, \qquad p \in \mathcal{N}_j. \tag{5.8}$$

As the exact coincidence set $\mathcal{N}_j^\bullet$ is not known a priori, we truncate with respect to its actual approximation

$$\mathcal{N}_j^\bullet(\bar{u}_j^\nu) = \{p \in \mathcal{N}_j \mid \bar{u}_j^\nu(p) = \varphi(p)\}$$

provided by fine-grid smoothing

$$\bar{u}_j^\nu = \mathcal{M}_j(u_j^\nu) \tag{5.9}$$

of the given iterate u_j^ν. Note that truncation (5.8) just means nodal interpolation of λ_l to the actual approximation

$$\tilde{\mathcal{S}}_j = \tilde{\mathcal{S}}_j(\bar{u}_j^\nu) = \{v \in \mathcal{S}_j \mid v(p) = 0 \;\forall p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)\} \tag{5.10}$$

of the reduced space $\mathcal{S}_j^\circ$. The corresponding splitting

$$\mathcal{S}_j = \sum_{l=1}^{n_j} V_l + \sum_{l=n_j+1}^{m_{\mathcal{S}}} \tilde{V}_l, \qquad \tilde{V}_l = \text{span}\{\tilde{\lambda}_l\}, \tag{5.11}$$

gives rise to a *truncated multilevel relaxation.* The subspaces $V_l = \text{span}\{\lambda_l\}$, $l = 1, \dots, n_j$ reproduce the leading projected Gauß–Seidel step (5.9). The truncated subspaces $\tilde{V}_l$ improve the coarse grid correction, because now all $\tilde{\lambda}_l$ with the property int supp $\lambda_l \cap \left(\mathcal{N}_j \setminus \mathcal{N}_j^\bullet\right) \neq \emptyset$ do contribute once $\mathcal{N}_j^\bullet(\bar{u}_j^\nu) = \mathcal{N}_j^\bullet$ is known. Until then, the actual splitting (5.11) is adapted to the actual coincidence set $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ in each iteration step. The corrections $v_l \in \tilde{V}_l$ are given by

$$v_l = \max\left\{\frac{r_l(\tilde{\lambda}_l)}{a(\tilde{\lambda}_l, \tilde{\lambda}_l)}, \max_{p \in \mathcal{N}_j \cap \text{int supp}\,\tilde{\lambda}_l} \frac{-w_{l-1}(p) + \varphi(p)}{\tilde{\lambda}_l(p)}\right\} \tilde{\lambda}_l. \tag{5.12}$$

for $\tilde{\lambda}_l \neq 0$ and $v_l = 0$ otherwise.

Remark 5.8. *By construction of the coarse grid spaces $\tilde{V}_l = \text{span}\{\tilde{\lambda}_l\}$, we have*

$$\mathcal{N}_j^\bullet(\bar{u}_j^\nu) \subset \mathcal{N}_j^\bullet(u_j^{\nu+1}).$$

Hence, in contrast to the original multilevel relaxation (5.1)*, inactivation, i.e. removing nodes from $\mathcal{N}_j^\bullet$, is now exclusively performed by projected Gauß–Seidel relaxation.*

A second modification of the multilevel relaxation (5.1) is needed to allow for an implementation as a multigrid V-cycle with optimal complexity. To this end, we consider the minimization problems

$$v_l \in \tilde{D}_l : \qquad \mathcal{J}(w_{l-1} + v_l) \le \mathcal{J}(w_{l-1} + v) \qquad \forall v \in \tilde{D}_l, \tag{5.13}$$

with modified coarse–grid constraints

$$\tilde{D}_l = \{z\tilde{\lambda}_l \mid z \ge \psi_l\}, \qquad l = n_j + 1, \dots, n_{\mathcal{S}},$$

and local defect obstacles $\psi_l \in V_l$. Once ψ_l is available, the constraint $\tilde{D}_l$ can be checked without visiting the fine grid. The condition

$$0 \ge \psi_l \ge \max_{p \in \text{int supp}\, \tilde{\lambda}_l} -w_{l-1}(p) + \varphi(p) \tag{5.14}$$

on the local defect obstacles ψ_l provides

$$0 \in \tilde{D}_l \subset \big(-w_{l-1} + \mathcal{K}_j \big) \cap \tilde{V}_l$$

and thus guarantees feasibility (cf. Algorithm 5.1).

Remark 5.9. *Replacing* $-w_{l-1} + \mathcal{K}_j$ *by the coarse grid constraints* $\tilde{D}_l$ *can be regarded as an intrinsic damping of* (5.12). *More precisely, denoting the corrections provided by* (5.12) *and* (5.13) *by* v_l^* *and* v_l*, respectively, we have*

$$v_l = \omega_l v_l^* \tag{5.15}$$

with some $\omega_l \in [0, 1]$.

We now present a recursive construction of local defect obstacles ψ_l. Let

$$\psi_l = \psi^{(k_l)}(p_l)$$

with suitable defect obstacles $\psi^{(k)} \in \mathcal{S}_k$ to be defined as follows. Starting with $v^{(j)} = 0$ and $\psi^{(j)} \in \mathcal{S}_j$, defined by

$$\psi^{(j)}(p) = \begin{cases} -\infty & \text{if } p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu) \\ -\bar{u}_j^\nu(p) + \varphi(p) & \text{else} \end{cases},$$

successive update and *monotone restriction*

$$\psi^{(k-1)} = R_k^{k-1}\big(\psi^{(k)} - v^{(k)}\big)$$

inductively guarantees condition (5.14). Using the enumeration $l = l(p,k)$ (cf. Remark 3.4), the overall correction $v^{(k)}$ on level $k < j$ is given by

$$v^{(k)} = \sum_{p \in \mathcal{N}_k} v_{l(p,k)} \in \tilde{\mathcal{S}}_k = \text{span}\{\tilde{\lambda}_{l(p,k)}\} \mid p \in \mathcal{N}_k\}.$$

The restriction operator $R_k^{k-1} : \mathcal{S}_k \to \mathcal{S}_{k-1}$ defined in (4.22) has been already used in Algorithm 4.7.

Now, the solution of the local problems (5.13) on level k can be equivalently formulated as projected Gauß–Seidel smoothing

$$v^{(k)} \in \mathcal{D}_k : \qquad \tilde{b}_k(v^{(k)}, v - v^{(k)}) \ge r_k(v - v^{(k)}) \qquad \forall v \in \mathcal{D}_k$$

with the constraints

$$\mathcal{D}_k = \{v \in \tilde{\mathcal{S}}_k \mid v(p) \ge \psi^{(k)}(p)\ \forall p \in \mathcal{N}_k\} \subset \tilde{\mathcal{S}}_k,$$

the bilinear form

$$\tilde{b}_k(v,w) = \sum_{\substack{i,l=1\\ i\leq l}}^{n_k} v(p_i) a\big(\tilde{\lambda}_{p_i}^{(k)}, \tilde{\lambda}_{p_l}^{(k)}\big) w(p_l), \qquad v,w \in \tilde{\mathcal{S}}_k, \tag{5.16}$$

and the residual $r_k = \ell - a(\bar{u}_j, \cdot) - \sum_{i=k+1}^{j-1} a(v^{(i)}, \cdot) \in \mathcal{S}_j' \subset \tilde{\mathcal{S}}_k'$.

Finally, truncation of $\lambda_{l(p,k)} = \lambda_p^{(k)}$ can be performed recursively according to

$$\tilde{\lambda}_p^{(k)} = \sum_{q\in\mathcal{N}_{k+1}} \lambda_p^{(k)}(q)\ \tilde{\lambda}_q^{(k+1)}, \quad p \in \mathcal{N}_k, \tag{5.17}$$

starting with $\tilde{\lambda}_p^{(j)} = 0$, if $p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ and $\tilde{\lambda}_p^{(j)} = \lambda_p^{(j)}$, otherwise.

The resulting *truncated multigrid method* can be formulated as a multigrid $V-$cycle with optimal complexity. It is called *monotone*, because, by construction, the coarse grid correction does not increase the energy.

Algorithm 5.10. *(Truncated monotone multigrid V–cycle with 1 pre-smoothing step)*

given: u_j^ν

fine grid smoothing: $\bar{u}_j^\nu = \mathcal{M}_j(u_j^\nu), \quad v^{(j)} = 0$

initialization of residual and bilinear form: $r_j = \ell - a(\bar{u}_j^\nu, \cdot), \quad a_j(\cdot,\cdot) = a(\cdot,\cdot)$

truncation: $r_j := r_j|_{\tilde{\mathcal{S}}_j}, \quad a_j(\cdot,\cdot) := a_j(\cdot,\cdot)|_{\tilde{\mathcal{S}}_j\times\tilde{\mathcal{S}}_j}$

initialization of defect obstacle: $\psi^{(j)}(p) = -\bar{u}_j^\nu(p) + \varphi(p) \quad \forall p \in \mathcal{N}_j$

truncation: $\psi^{(j)}(p) := -\infty$*, if* $p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)$

initial restriction:

$r_{j-1} = r_j|_{\tilde{\mathcal{S}}_{j-1}}, \quad a_{j-1}(\cdot,\cdot) = a_j(\cdot,\cdot)|_{\tilde{\mathcal{S}}_{j-1}\times\tilde{\mathcal{S}}_{j-1}}, \quad \psi^{(j-1)} = R_j^{j-1}\psi^{(j)}$

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

{

solve:

$v^{(k)} \in \mathcal{D}_k : \quad \tilde{b}_k(v^{(k)}, v - v^{(k)}) \geq r_k(v - v^{(k)}) \quad \forall v \in \mathcal{D}_k$ *(pre–smoothing)*

$r_k := r_k - a_k(v^{(k)}, \cdot)$ *(update of the residual)*

$\psi^{(k)} := \psi^{(k)} - v^{(k)}$ *(update of the defect obstacle)*

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

$\psi^{(k-1)} = R_k^{k-1}\psi^{(k)}$ *(monotone restriction)*

}

solve:

$v^{(0)} \in \mathcal{D}_0 : \quad b_0(v^{(0)}, v - v^{(0)}) \geq r_0(v - v^{(0)}) \quad \forall v \in \mathcal{D}_0$ *(initial grid smoothing)*

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

{

$v^{(k)} := v^{(k)} + v^{(k-1)}$ *(canonical interpolation)*

}

new iterate: $u_j^{\nu+1} = u_j^\nu + v^{(j)}$

Implementation of the truncation step amounts to annihilating all rows and columns of the stiffness matrix and the right hand side that correspond to active nodes $p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)$. Then, according to (5.17), the algebraic formulation of the canonical restriction of the residual r_k and the bilinear form $a_k(\cdot,\cdot)$ from $\tilde{\mathcal{S}}_k$ to $\tilde{\mathcal{S}}_{k-1}$ involves the same weights as for the standard spaces $\mathcal{S}_k$ and $\mathcal{S}_{k-1}$. After prolongation, we only have to set $v^{(j)}(p) := 0$ for all $p \in \mathcal{N}_j^\bullet(u_j^\nu)$. No other changes of canonical restriction and prolongation procedures are necessary. Projected Gauß–Seidel smoothing has to be modified in order to skip all coefficients with vanishing diagonal elements. Only monotone restriction of the defect obstacles has to be newly added to an existing multigrid code.

Remark 5.11. *In addition to the initial restriction on the fine grid, truncation can be performed recursively on all levels. According to numerical computations, this variant often leads to a faster detection of the coincidence set.*

Of course, it is possible to apply the concept of modified defect constraints D_l directly to the multilevel relaxation (5.1) without any truncation of λ_l. This immediately leads to a modification of (5.1) which can be implemented as a multigrid V–cycle. As just the standard multilevel nodal basis functions are involved in this case, such algorithms are called *standard monotone multigrid method*.

Algorithm 5.12. *(Standard monotone multigrid V–cycle with 1 pre-smoothing step)*
proceed as in Algorithm 5.10 but skip the truncation of r_j, $a_j(\cdot,\cdot)$, *and* $\psi^{(j)}$

Remark 5.13. *In contrast to the truncated Algorithm 5.10, the standard version Algorithm 5.12 allows for activation and inactivation by coarse grid correction.*

Monotone multigrid methods with multiple pre– and post–smoothing or W–cycles can be derived in a similar way.

The *global convergence* of monotone multigrid methods relies on the following lemma.

Lemma 5.14. *Assume that the iterates* u_j^ν *are produced by an algorithm of the form*

$$\bar{u}_j^\nu = \mathcal{M}_j(u_j^\nu), \quad u_j^{\nu+1} = \mathcal{C}_j(\bar{u}_j^\nu) \tag{5.18}$$

with $\mathcal{M}_j : \mathcal{K}_j \to \mathcal{K}_j$ *denoting the projected Gauß–Seidel iteration* (4.4) *and some* $\mathcal{C}_j : \mathcal{K}_j \to \mathcal{K}_j$ *satisfying the monotonicity condition*

$$\mathcal{J}(\mathcal{C}_j(w)) \le \mathcal{J}(w) \qquad \forall w \in \mathcal{K}_j. \tag{5.19}$$

Then $u_j^\nu \to u_j$ *holds for any initial iterate* $u_j^0 \in \mathcal{K}_j$.

Proof. The sequence of iterates is bounded, because $\mathcal{J}(u_j^\nu) \le \mathcal{J}(u_j^0)$ holds for all ν. As $\mathcal{S}_j$ has finite dimension and $\mathcal{K}_j$ is closed, there is a subsequence $(u_j^{\nu_k})$ and $u_j^* \in \mathcal{K}_j$ such that $u_j^{\nu_k} \to u_j^*$ for $k \to \infty$. The local corrections v_l depend continuously on the intermediate iterates w_{l-1} so that, consisting of nested continuous functions, the Gauß–Seidel relaxation $\mathcal{M}_j$ is continuous on $\mathcal{K}_j$. The monotonicity (5.19) implies

$$\mathcal{J}(u_j^{\nu_{k+1}}) \le \mathcal{J}(u_j^{\nu_k+1}) \le \mathcal{J}(\mathcal{M}_j(u_j^{\nu_k})) \le \mathcal{J}(u_j^{\nu_k})$$

and passing to the limit $k \to \infty$, we obtain $\mathcal{J}(\mathcal{M}_j(u_j^*)) = \mathcal{J}(u_j^*)$. This leads to $\mathcal{M}_j(u_j^*) = u_j^*$, because the local corrections $v_l = 0$ are uniquely determined by (4.4). Denote $r = \ell - a(u_j^*, \cdot)$. Then (4.4) yields the complementarity property

$$r(\lambda_p^{(j)}) \le 0, \quad u_j^*(p) \le \varphi(p), \quad r(\lambda_p^{(j)})(-u_j^*(p) + \varphi(p)) = 0 \qquad \forall p \in \mathcal{N}_j.$$

providing $r(v - u_j^*) \le 0$ for all $v \in \mathcal{K}_j$. Hence, u_j^* solves (2.11). As $u_j = u_j^*$ is uniquely determined, each convergent subsequence of (u_j^ν) must converge to u_j. This concludes the proof. □

By construction of the algorithms, our first convergence result is an immediate consequence of Lemma 5.14.

Theorem 5.15. *The monotone multigrid algorithms 5.10 and 5.12 converge for any initial iterate* $u_j^0 \in \mathcal{K}_j$.

We now concentrate on the *asymptotic convergence speed* of the truncated monotone multigrid method 5.10.

Lemma 5.16. *Assume that the non–degeneracy condition* (2.13) *is satisfied. Then, after a sufficiently large number of steps, Algorithm 5.10 is reducing to a linear multigrid method for the reduced linear problem* (2.12) *which is generated by the splitting*

$$\mathcal{S}_j^\circ = \sum_{l=1}^{m_{\mathcal{S}}} V_l^\circ, \qquad V_l^\circ = \operatorname{span}\{\lambda_l^\circ\}, \tag{5.20}$$

with λ_l° *obtained by nodal interpolation of* λ_l *to* $\mathcal{S}_j^\circ$.

Proof. First, we show that

$$\mathcal{N}_j^\bullet(u_j^\nu) = \mathcal{N}_j^\bullet \qquad \forall \nu \ge \nu_0 \tag{5.21}$$

holds for sufficiently large ν_0. Let $p \in \mathcal{N}_j \setminus \mathcal{N}_j^\bullet$. Then the convergence $u_j^\nu \to u_j$ implies $u_j^\nu(p) > \varphi(p)$ and thus $\mathcal{N}_j^\bullet(u_j^\nu) \subset \mathcal{N}_j^\bullet$ for sufficiently large ν. Conversely, if $p \in \mathcal{N}_j^\bullet$, then the strict complementarity (2.13) and the convergence of the intermediate iterates $w_l \to u_j$ (which follows by the same arguments as used in the proof of Lemma 5.14) asymptotically provide

$$\ell(\tilde\lambda_l) - a(w_{l-1}, \tilde\lambda_l) < 0$$

for all $\tilde\lambda_l$ with $p \in \operatorname{int}\ \operatorname{supp}\tilde\lambda_l$. Here, we have set $\tilde\lambda_l = \lambda_{p_l}^{(j)}$ for $l = 1, \dots, n_j$. As a consequence, we get $p \in \mathcal{N}_j^\bullet(\bar u_{\mathcal{S}}^\nu)$ and $v_l = 0$ for all $l = n_j, \dots, m_{\mathcal{S}}$ with $p \in \operatorname{int}\ \operatorname{supp}\tilde\lambda_l$. Hence, $p \in \mathcal{N}_j^\bullet(u_j^\nu)$ and thus $\mathcal{N}_j^\bullet \subset \mathcal{N}_j^\bullet(u_j^\nu)$ for sufficiently large ν.

The equality (5.21) provides $v_l = 0$, if $p_l \in \mathcal{N}_j^\bullet(u_j^\nu)$, $l = 1, \dots, n_j$, and $\tilde\lambda_l = \lambda_l^\circ$, $l = n_j + 1, \dots, m_{\mathcal{S}}$, for $\nu \ge \nu_0$. Hence, the original splitting (5.11) asymptotically reduces to (5.20). We now show that there is some $\varepsilon \in \mathbb{R}$ such that

$$0 > -\varepsilon \ge \psi_l \qquad \forall \nu \ge \nu_1 \tag{5.22}$$

holds for all l with $\lambda_l^\circ \not\equiv 0$ and sufficiently large $\nu_1 \ge \nu_0$. First note that $\psi_l = \psi_l(w_0, \dots, w_{l-1})$ is a continuous function of the intermediate iterates. By construction, we have

$$\psi_l^* := \psi_l(u_j, \dots, u_j) = \max_{p \in \operatorname{int}\ \operatorname{supp}\lambda_l^\circ} -u_j(p) + \varphi(p) < 0$$

for $\lambda_l^\circ \not\equiv 0$. Let $\psi^* < 0$ be the maximum of all such $\psi_l^* < 0$. Then, (5.22) clearly holds for any positive $\varepsilon < -\psi^*$ and sufficiently large ν. As u_j solves (2.12) and $w_l \to u_j$, we asymptotically get $|r_l(\lambda_l^\circ)| \le \varepsilon$. In combination with (5.22), this proves the assertion. □

Utilizing Lemma 5.16, Kornhuber [51] showed asymptotic convergence rates of truncated monotone multigrid methods by adopting linear multigrid convergence theory.

Theorem 5.17. *Assume that the strict complementarity condition* (2.13) *holds and let* $d = 2$. *Then, there is a* $\nu_0 \in \mathcal{N}$ *such that the iterates* (u_j^ν) *produced by Algorithm 5.10 satisfy the error estimate*

$$\|u_j^{\nu+1} - u_j\| \le \big(1 - c(j+1)^{-4}\big)\|u_j^\nu - u_j\| \tag{5.23}$$

with a constant $c > 0$ *independent of* j.

Proof. The assertion follows directly from Theorem 2.5 by Neuss [65] using Theorem 5.3 by Kornhuber and Yserentant [60] to verify condition (V1) with $K_0 = \mathcal{O}((j+1)^2)$ and the Cauchy–Schwarz inequality to provide condition (V2) with $K_1 = \mathcal{O}(j+1)$. □

Remark 5.18. *In arbitrary space dimensions, the asymptotic convergence rate is bounded by* $1-c(j+1)^{-3}$ *under the additional condition that the coincidence set* $\mathcal{N}_j^\bullet$ *is rich enough in a certain sense. For example, coincidence sets consisting of single lines or points are excluded. We refer to Kornhuber and Yserentant [60, Section 6] for details.*

The truncated multilevel relaxation (5.12) with sub-optimal complexity shows exactly the same asymptotic behavior as Algorithm 5.10.

Remark 5.19. *The asymptotic convergence rates of the standard monotone multigrid method stated in Algorithm 5.12 are even bounded by* $1 - c(j+1)^{-2}$ *in 2D and, under the additional assumptions on* $\mathcal{N}_j^\bullet$ *mentioned above, by* $1-c(j+1)^{-1}$ *in arbitrary space dimensions. The improvements result from a strengthened Cauchy–Schwarz inequality that holds for decompositions from standard coarse grid spaces* $\mathcal{S}_k$ *(cf. Yserentant [90, Lemma 2.7]).*

The original projected multilevel relaxation (5.1) shows exactly the same asymptotic behavior as Algorithm 5.12.

Recall from the Remarks 5.8 and 5.13 that inactivation both in the truncated multilevel relaxation (5.12) and in Algorithm 5.10 is performed exclusively by projected Gauß–Seidel relaxation on the fine grid. Hence, starting with $u_j^0 = \varphi_j$, a global detaching effect of a point source f is distributed only by next–neighbor interaction. This simple example contradicts global mesh–independent convergence rates for the truncated monotone multigrid methods. Mesh-independent convergence rates for the standard version were observed in practical computations but theoretical justification still seems to be an open problem.

5.3. **Concluding remarks.** Standard and truncated multigrid methods were introduced by Mandel [62, 63] and Kornhuber [51], respectively. A related algorithm by Brandt and Cryer [23] relies on the FAS (full approximation storage) approach by Brandt [21]. In order to guarantee that the exact solution of the obstacle problem is a fixed point of the method, they modified the restriction of the residual (but not of the stiffness matrix) similar to the truncation appearing in Algorithm 5.10. The methods often exhibits similar convergence speed as truncated multigrid but sometimes fails to converge (see Bollrath [9, p. 29] or Kornhuber [51] for a comparison). Preliminary experimental results for cascadic–type iterations have been presented by Blum, Braess, and Suttmeier [8].

First results on asymptotic multigrid convergence rates are due to Kornhuber [51]. Major contributions to global bounds for the convergences rates were made by Badea, Tai and coworkers [1, 2, 42, 80]. They also used the same theoretical framework to analyze Jacobi-like versions of Algorithm 5.1, where the update of the intermediate iterates is simply skipped so that the corrections can be computed in parallel. As for related subset decomposition methods, convergence has to be enforced by severe damping which might slow down convergence considerably in comparison with the sequential version. Similar to linear subspace decomposition, the abstract convergence result can be also applied to overlapping domain decomposition methods. For further information, we refer to Badea, Tai and Wang [2] and the references cited therein.

A projected space decomposition method of domain-decomposition-type was proposed by Schöberl [74] for Signorini problems in linear elasticity. Exploiting that the number of unknowns in the bulk grows with higher order than the number of unknowns on the boundary he showed mesh-independent convergence. Block versions of monotone multigrid methods for scalar obstacle problems can be also applied directly to Signorini's problem in linear elasticity, provided that the normal directions are constant along the Signorini boundary (cf., e.g., Belsky [6]). Spatially varying normal directions can be incorporated by suitable weighting factors as suggested by Kornhuber and Krause [56]. Wohlmuth and Krause [84] extended monotone multigrid to mortar–discretized two-body contact. Their main idea is a hierarchical splitting of the ansatz space into a linear space with vanishing relative deformation and the (constrained) nodal displacements at the slave side of the (potential) contact boundary. The resulting algorithm preserves the asymptotic convergence speed of unconstrained multigrid methods even for realistic 3D geometries in biomechanical applications [59].

Projected multilevel relaxation and monotone multigrid has been extended to smooth non-quadratic energy functionals [79, 81] and also to variational inequalities of the form

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

with suitable superposition operators ϕ_j, see [52, 55]. Applications include frictional contact in elasticity [33] or phase field models [57, 58].

6. From truncated multigrid to inexact active set methods

6.1. **A nonsmooth Newton–like method and inexact variants.** The truncated monotone multigrid method stated in Algorithm 5.10 has the flavor of an *active set approach*: The actual coincidence set $\mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ is fixed by the leading projected Gauß–Seidel relaxation and then is essentially preserved by subsequent coarse grid correction. We will now clarify this analogy by deriving a *nonsmooth Newton–like method* which will turn out to be closely related to Algorithm 5.10.

Conversely, reduction of energy suggests a natural way to provide global convergence of *inexact active set methods* involving approximate solutions of the arising linear problems as resulting from, say, one multigrid step.

The main idea is to reformulate the obstacle problem (2.10) as the nonlinear system

$$u_j \in \mathcal{S}_j : \quad \mathcal{F}_j(u_j) = 0, \qquad \mathcal{F}_j = \mathcal{M}_j - I, \tag{6.1}$$

where $\mathcal{M}_j$ denotes projected Gauß–Seidel relaxation (4.4) and I is the identity operator on $\mathcal{S}_j$. Observe that $\mathcal{F}_j$ is not differentiable, but Lipschitz–continuous.

In order to derive a nonsmooth Newton–like method for (6.1), it is convenient to identify finite element functions with the coefficient vectors of their nodal basis representation:

$$v = \sum_{i=1}^{n_j} v_i \lambda_{p_i}^{(j)} \in \mathcal{S}_j \quad \longleftrightarrow \quad \underline{v} = (v_i)_{i=1}^{n_j} \in \mathbb{R}^{n_j}.$$

In this way, (6.1) can be reformulated as the nonsmooth, nonlinear system

$$\underline{u} \in \mathbb{R}^{n_j} : \quad F(\underline{u}) = 0, \qquad F = M - I, \quad M(\underline{v}) = \underline{\mathcal{M}_j(v)},$$

with I denoting the unit matrix in $\mathbb{R}^{n_j}$. From now on, we drop underlines, as long as no confusion is likely to occur, e.g., we simply write $\varphi = (\varphi_i)_{i=1}^{n_j}$ instead of $\underline{\varphi}_j$. We finally introduce the usual stiffness matrix A,

$$A = L + D + R, \qquad A = (a_{ik})_{i,k=1}^{n_j}, \quad a_{ij} = a(\lambda_{p_i}^{(j)}, \lambda_{p_k}^{(j)}),$$

which is split into its lower diagonal, diagonal, and upper diagonal parts L, D, and R, respectively.

Let us consider some fixed $w \in \mathbb{R}^{n_j}$, denoting $\bar{w} = M(w)$. We define the corresponding set of active indices

$$N^\bullet = N^\bullet(\bar{w}) = \{i \in N \mid \bar{w}_i = \varphi_i\}, \qquad N = \{1, 2, \ldots, n_j\},$$

and the subset

$$U = U(N^\bullet) = \{v \in \mathbb{R}^{n_j} \mid (M(v))_i = \varphi_i \Leftrightarrow i \in N^\bullet\} \in \mathbb{R}^{n_j}$$

of all vectors such that projected Gauß-Seidel relaxation provides the same active set $N^\bullet$. It will turn out that F is affine linear on U. In order to derive an explicit representation of $y = F(w)$, $w \in U$, we reformulate (4.4) by straightforward calculation to obtain

$$y_i = (\varphi - w)_i \quad \forall i \in N^\bullet, \qquad \big((L + D)y\big)_i = (b - Aw)_i \quad \forall i \in N \setminus N^\bullet.$$

Using the truncation matrix $T = T(N^\bullet) = (T_{ik}) \in \mathbb{R}^{n_j, n_j}$ defined by

$$T_{ik} = \begin{cases} 1, & \text{if } i = k \in N \setminus N^\bullet \\ 0, & \text{else} \end{cases},$$

this leads to the desired formula

$$F(w) = \big(T(L + D) + I - T\big)^{-1} \big(T(b - Aw) + (I - T)(\varphi - w)\big). \tag{6.2}$$

By construction, (6.2) holds for arbitrary $w \in U$. As the mapping $w \mapsto N^\bullet(w) \subset N$ is well-defined on $\mathbb{R}^{n_j}$, we have

$$\mathbb{R}^{n_j} = \bigcup_{N^\bullet \subset N} U(N^\bullet), \qquad N_1^\bullet \neq N_2^\bullet \Rightarrow U(N_1^\bullet) \cap U(N_2^\bullet) = \emptyset.$$

Hence, for given $w \in \mathbb{R}^{n_j}$, (6.2) suggests the linearization

$$\partial F(w) = -\big(T(L+D) + I - T\big)^{-1}\big(TA + (I-T)\big), \quad T = T(N^\bullet(\bar{w})),\ \bar{w} = M(w).$$

As the sets $U(N^\bullet) \subset \mathbb{R}^{n_j}$ might degenerate to lower–dimensional objects, ∂F might not be a generalized derivative in the sense of Clarke [27, Chapter 2]. Hence, the associate iterative scheme

$$u^{\nu+1} = u^\nu - \big(\partial F(u^\nu)\big)^{-1} F(u^\nu) \tag{6.3}$$

can be regarded as a *nonsmooth Newton–like method*.

We derive a more convenient representation of $u^{\nu+1}$. Inserting the above definition of ∂F and the representation (6.2) into (6.3), we get the linear system

$$(TA + I - T)u^{\nu+1} = Tb + (I-T)\varphi, \quad T = T(N^\bullet(\bar{u}^\nu)),\ \bar{u}^\nu = M(u^\nu) \tag{6.4}$$

for $u^{\nu+1}$. By definition of T and by application of $I - T$ to equation (6.4) we get

$$(I-T)\bar{u}^\nu = (I-T)u^{\nu+1} = (I-T)\varphi.$$

Using these identities, (6.4) can be reformulated as the linear system

$$(TAT + I - T)v^\nu = T(b - A\bar{u}^\nu), \quad T = T(N^\bullet(\bar{u}^\nu)),\ \bar{u}^\nu = M(u^\nu) \tag{6.5}$$

for the correction $v^\nu = u^{\nu+1} - \bar{u}^\nu \in \mathbb{R}^{n_j}$.

In terms of finite element functions this system takes the form

$$v_j^{\nu+1} \in \tilde{\mathcal{S}}_j : \qquad a(v_j^{\nu+1}, v) = \ell(v) - a(\bar{u}_j^\nu, v) \qquad \forall v \in \tilde{\mathcal{S}}_j, \qquad \bar{u}_j = \mathcal{M}_j(u_j^\nu), \tag{6.6}$$

with $\tilde{\mathcal{S}}_j$ defined in (5.10). We are ready to state a crucial observation of this section.

Remark 6.1. *An inexact variant of the nonsmooth Newton–like method* (6.3) *is obtained by simply ignoring the coarse–grid defect obstacles $\psi^{(k)}$ in Algorithm 5.10.*

Let $\tilde{v}_j^\nu$ be some approximation of the solution $v_j^{\nu+1}$ of (6.6). In order to enforce global convergence, we enforce decreasing energy by successive projection and damping. In a multigrid context, it is natural to use fine–grid smoothing

$$w_j^\nu = \mathcal{M}_j(\bar{u}_j^\nu + \tilde{v}_j^\nu)$$

for the projection of $\bar{u}_j^\nu + \tilde{v}_j^\nu$ to $\mathcal{K}_j$ (simple lumped L^2–projection leads to very similar results). Subsequent damping leads to the new iterate

$$u_j^{\nu+1} = \bar{u}_j^\nu + \omega^\nu(w_j^\nu - \bar{u}_j^\nu), \qquad \omega^\nu = \operatorname*{argmin}_{\omega\in[0,1]} \mathcal{J}(\bar{u}_j^\nu + \omega(w_j^\nu - \bar{u}_j^\nu)). \tag{6.7}$$

By Lemma 5.14, the resulting iterative scheme is globally convergent for *any choice* of $\tilde{v}_j^\nu \in \mathcal{S}_j$. Approximating the solution of (6.6) by one truncated multigrid sweep, we obtain the following algorithm (recall the definition (5.10) of $\tilde{\mathcal{S}}_j$).

Algorithm 6.2. *(Truncated nonsmooth Newton multigrid V–cycle with global damping)*

given: u_j^ν
fine grid smoothing: $\bar{u}_j^\nu = \mathcal{M}_j(u_j^\nu), \quad v^{(j)} = 0$

initialization of residual and bilinear form: $r_j = \ell - a(\bar{u}_j^\nu, \cdot), \quad a_j(\cdot,\cdot) = a(\cdot,\cdot)$
truncation: $r_j := r_j|_{\tilde{\mathcal{S}}_j}, \quad a_j(\cdot,\cdot) := a_j(\cdot,\cdot)|_{\tilde{\mathcal{S}}_j\times\tilde{\mathcal{S}}_j}$

initial restriction:

$r_{j-1} = r_j|_{\tilde{\mathcal{S}}_{j-1}}, \quad a_{j-1}(\cdot,\cdot) = a_j(\cdot,\cdot)|_{\tilde{\mathcal{S}}_{j-1}\times\tilde{\mathcal{S}}_{j-1}}$

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

{

solve:

$v^{(k)} \in \tilde{\mathcal{S}}_k : \quad \tilde{b}_k(v^{(k)}, v - v^{(k)}) = r_k(v - v^{(k)}) \quad \forall v \in \tilde{\mathcal{S}}_k$ *(pre–smoothing)*

$r_k := r_k - a_k(v^{(k)}, \cdot)$ *(update of the residual)*

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

}

solve:

$v^{(0)} \in \tilde{\mathcal{S}}_0 : \quad b_0(v^{(0)}, v - v^{(0)}) = r_0(v - v^{(0)}) \qquad \forall v \in \tilde{\mathcal{S}}_0$ *(approx. sol. on* $\mathcal{T}_0$*)*

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

{

$v^{(k)} := v^{(k)} + v^{(k-1)}$ *(canonical interpolation)*

}

$w_j^\nu = \mathcal{M}_j(\bar{u}_j^\nu + v^{(j)})$ *(projection)*

$\omega^\nu = \operatorname*{argmin}_{\omega\in[0,1]} \mathcal{J}(\bar{u}_j^\nu + \omega(w_j^\nu - \bar{u}_j^\nu))$ *(global damping)*

new iterate: $u_j^{\nu+1} = \bar{u}_j^\nu + \omega^\nu(w_j^\nu - \bar{u}_j^\nu)$

Implementation of the truncation step amounts to annihilating all rows and columns of the stiffness matrix and the right hand side that correspond to active nodes $p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)$. Then, essentially, any standard multigrid implementation can be used for the coarse grid correction $v^{(j)}$. The only modifications are that Gauß–Seidel smoothing has to skip all coefficients with vanishing diagonal elements and that we have to set $v^{(j)}(p) := 0$ for all $p \in \mathcal{N}_j^\bullet(\bar{u}_j^\nu)$ after prolongation.

Remark 6.3. *Algorithm 6.2 is open for various kinds of modifications, such as cg–acceleration of coarse grid correction, multiple nonlinear post–smoothing before or after global damping or line search* $\omega \in \mathbb{R}$ *instead of mere damping.*

We abandon these options in favor of a better comparison with the other algorithms presented in this paper. In any case, convergence follows from Lemma 5.14.

Theorem 6.4. *Algorithm 6.2 converges for any initial iterate* $u_j^0 \in \mathcal{S}_j$.

Remark 6.5. *Inactivation in Algorithm 6.2 is performed exclusively by projected Gauß–Seidel relaxation, as in truncated monotone multigrid.*

As a consequence of Remark 6.5, we can hardly expect global mesh–independent convergence rates of Algorithm 6.2. Indeed, if we start with $u_j^0 = \phi_j$ in case of a point source f and an exact solution $u_j > \phi_j$, then the number of next-neighbor inactivation steps cannot be independent of n_j. This was already pointed out at the end of Section 5.2. On the other hand, we expect high convergence speed for reasonable initial iterates, i.e., if the coincidence set $\mathcal{N}_j^\bullet$ is resolved sufficiently well.

6.2. **Concluding remarks.** The truncated nonsmooth Newton multigrid method as stated in Algorithm 6.2 can be regarded as an active set strategy combined with a linear multigrid method for the resulting reduced linear problems. Other active set strategies combined with multigrid are well–known for quite a while. We refer, e.g., to the poineering work of Hackbusch and Mittelmann [40] or the now often called primal dual active set strategies by Hoppe [45, 46, 47]. In contrast to our approach, primal dual active set strategies are essentially iterating on the complementarity condition. For a recent re–interpretation in terms of non–smooth analysis, we refer, e.g., to M. Ulbrich [82] or Hintermüller, Ito, and Kunisch [43]. The convergence proofs typically rely on M–matrix properties of the stiffness matrix and assume that the reduced linear subproblems are solved exactly. Both assumptions are seldom met in practical computations. For example, adaptive mesh refinement involving edge bisection already violates the M–matrix property. Nevertheless, inexact versions on adaptively refined grids are often used on a heuristic basis.

Numerical experiments for Signorini's problem by Krause [61] indicate that the convergence properties of the overall iteration strongly depend on the choice of the multigrid solver for the reduced linear subproblems. According to his computations, truncated linear multigrid, as involved in Algorithm 6.2 seems to be the best choice. Truncated linear multigrid was introduced by Hoppe and Kornhuber [48] and later analyzed by Kornhuber and Yserentant [60]. Numerical experiments for two-body contact problems by Sander [73] confirm the observations of Krause [61] that truncated nonsmooth Newton multigrid 6.2 usually converges faster than truncated monotone multigrid 5.10.

7. Numerical Assessment

7.1. **Numerical test problems.** For most of the multigrid algorithms to be described below numerical experiments with "generic" test problems like, e.g., simplified elasto–plastic torsion [37, Chapter II] have been already reported elsewhere [51]. The two following examples are particularly designed to check the robustness of multigrid algorithms with respect to complicated and unstable coincidence sets. We always consider the square $\Omega = (-1, 1)^2$, the initial triangulation $\mathcal{T}_0$ consisting of four congruent triangles and apply $j = 9$ uniform refinement steps to obtain the final triangulation $\mathcal{T}_j$ with 523 265 interior nodes.

7.1.1. *Spiral problem.* The fact that the coincidence set $\mathcal{N}_j^\bullet$ usually has no representation on coarser grids sets a particular challenge for any kind of multigrid approach to obstacle problems. In order to highlight this intrinsic difficulty, we consider the obstacle function

$$\varphi(x(r, \phi)) = \sin(2\pi/r + \pi/2 - \phi) + \frac{r(r + 1)}{r - 2} - 3r + 3.6, \quad r \neq 0,$$

and $\varphi(0) = 3.6$ with polar coordinates $x(r, \phi) = re^{i\phi}$. For the right hand side $f = 0$ this choice leads to the spiral set $\mathcal{N}_j^\bullet$ as illustrated in the left picture of Figure 2. Obviously, functions $v \in \mathcal{S}_0$ do not have much to do with the approximate solution u_j on the final level $j = 9$.

7.1.2. *Degenerate problem.* Many multigrid solvers for obstacle problems reduce to linear multigrid methods for the reduced linear problem (2.12), once the coincidence set $\mathcal{N}_j^\bullet$ has been detected. For sufficiently good initial iterates, such methods take advantage of the good convergence properties of linear multigrid. In order to check

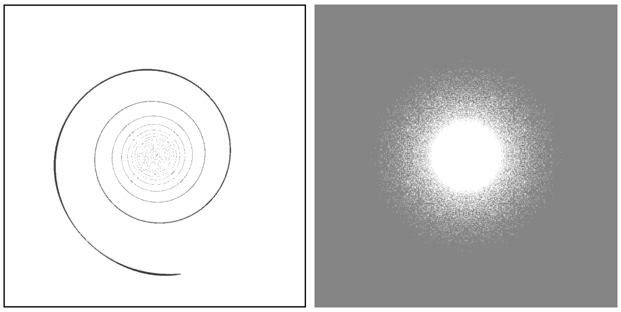

FIGURE 2. Coincidence sets of spiral problem (left) and degenerate problem (right)

their robustness, we consider a degenerate continuous problem with an approximate coincidence set $\mathcal{N}_j^\bullet$ that is extremely hard to find. Related practical problems occur, e.g., in the simulation of deep drawing processes. For the given obstacle

$$\varphi(x_1, x_2) = -(x_1^2 - 1)(x_2^2 - 1),$$

the selection of the right hand side $f = -\Delta\varphi$ causes $u = \varphi$ on Ω to be the solution of (2.1). As a consequence, the solution u_j of the discretized obstacle problem, the solution of its unconstrained analogue, and the obstacle φ_j converge to a common limit. As a consequence, the coincidence set $\mathcal{N}_j^\bullet$ becomes arbitrarily unstable with increasing refinement and thus can be hardly detected by iterative solvers. According to the right picture in Figure 2, the (black) coincidence set is not only unstable but fairly complicated as well.

7.2. **Multilevel Subset Decomposition Versus Projected Multilevel Relaxation.** In our numerical experiments, we check the robustness of the convergence behavior with respect to given "bad" initial iterates and "good" initial iterates as obtained by nested iteration. The convergence behavior is illustrated by *iteration histories* showing the algebraic error $\|u_j - u_j^\nu\|$ over the iteration steps ν. In order to evaluate possible mesh–dependence, we introduce the *asymptotic convergence rates*

$$\rho_k = \sqrt[\nu^*]{\frac{\|u_j - u_k^{\nu^*}\|}{\|u_j - u_j^1\|}}, \qquad k = 0, \dots, j, \tag{7.1}$$

where ν^* is chosen such that $\|u_j - u_k^{\nu^*}\| < 10^{-11}$ and the initial iterate u_j^0 is obtained by nested iteration.

We first compare the convergence behavior of the multilevel subset decomposition method MSD stated in Algorithm 4.7 and of the projected multilevel relaxation (4.10) called PMLR. We always consider $V(1,1)$ cycles.

7.2.1. *Spiral problem.* In our first experiment, we consider the discrete problem stated in Section 7.1.1 with a complicated, spiral coincidence set. We first observe that the convergence behavior of MSD and PMLR is almost independent of the choice of the initial iterate x_j^0. This is illustrated in Figure 3 and the left picture in Figure 4, showing the iteration histories for initial iterates on the obstacle $u_j^0 = \varphi_j$, above the obstacle $u_j^0 = \varphi_j + 10$ and obtained by nested iteration. For bad initial iterates as considered in Figure 3, the leading fast reduction of the error is due to fast reduction of high-frequency components. The asymptotic convergence rates both of MSD and PMLR seem to saturate with increasing refinement level j as depicted in the right picture of Figure 4. This is in good accordance with the theoretical results stated in Theorem 4.6 and Theorem 5.6, respectively. However, due to the dynamic adaptation of the underlying splitting of the defect obstacles $u_j^\nu - \varphi_j$, PMLR shows a considerable faster convergence speed than MSD as soon as the exact coincidence set $\mathcal{N}_j^\bullet$ is detected. While the asymptotic convergence rates of PMLR seems to be bounded by 0.7, the convergence rates of MSD exceed 0.9 on level $j = 9$.

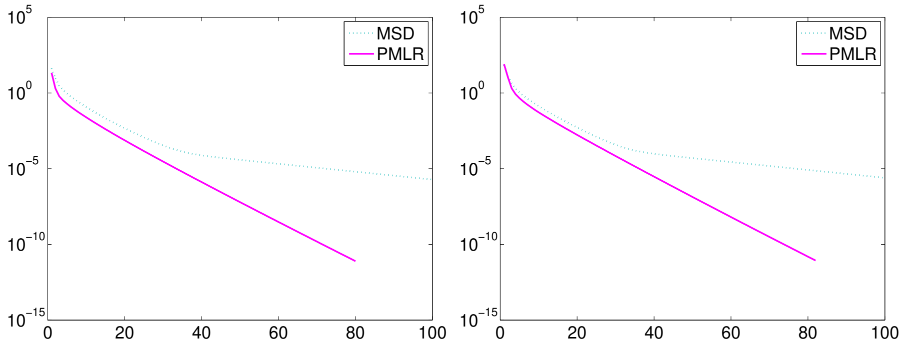

FIGURE 3. Spiral problem: Iteration histories for the initial iterates $u_j^0 = \varphi_j$ (left) and $u_j^0 = \varphi_j + 10 > \varphi_j$ (right)

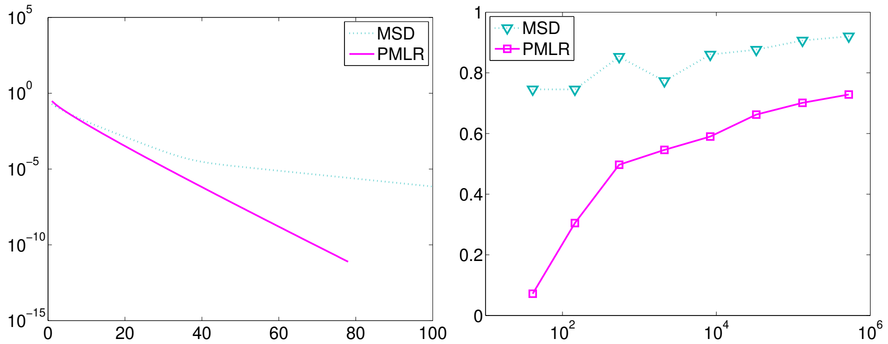

FIGURE 4. Spiral problem: Iteration history for nested iteration (left) and asymptotic convergence rates (right)

7.2.2. *Degenerate problem.* Comparing the convergence behavior of MSD and PMLR for the degenerate problem stated in Section 7.1.2 we first observe a similar robustness of both methods with respect to different initital iterates as before. This is illustrated by the iteration histories for initial iterates on the obstacle $u_j^0 = \varphi_j$, above the obstacle $u_j^0 = \varphi_j + 10$ and obtained by nested iteration as depicted in Figure 5 and the left picture in Figure 6. Observe that this time PMLR exhibits leading fast convergence even for nested iteration, illustrating that the resulting initial iterate is not too good in this (almost) degenerate case. In contrast to the spiral example the exact coincidence set is now hard to detect. As a consequence, apart from a slightly faster leading convergence of PMLR, the convergence speed of MSD and PMLR is very similar for all the three initial iterates. As before, the asymptotic convergence rates seem to saturate for both methods. According to the right picture in Figure 6 they seem to be bounded by 0.7 for both methods. We should not forget at this point that PMLR has a higher complexity than MSD (cf. Remark 5.13). More precisely, for $j = 9$ levels the cpu time for each step is about twice as large for PMLR than for MSD.

To sum up, both the convergence behavior of PMLR and MSD is almost invariant with respect to different initial iterates and/or degeneracy of the problem. Such kind of robustness nicely reflects the convergence analysis reported in Section 5.1, where non–degeneracy does not play a role as well.

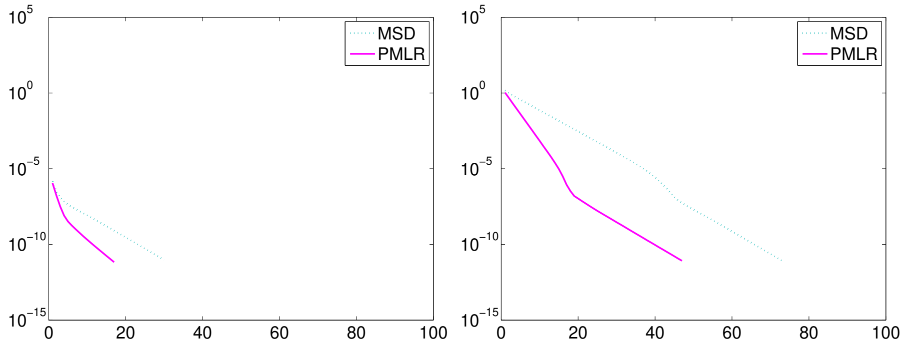

FIGURE 5. Degenerate problem: Iteration histories for the initial iterates $u_j^0 = \varphi_j$ (left) and $u_j^0 = \varphi_j + 10 > \varphi_j$ (right)

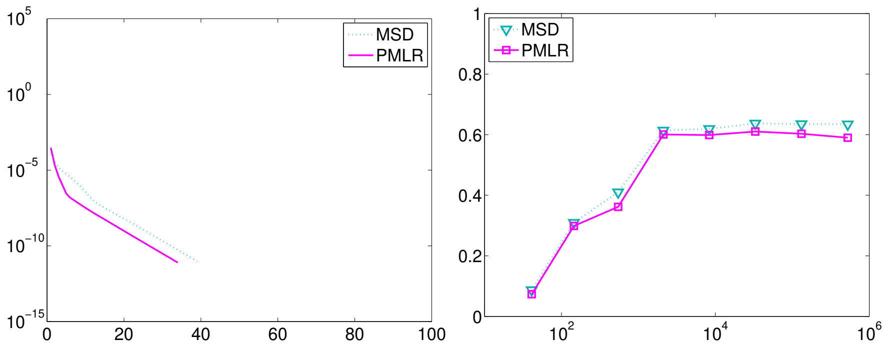

FIGURE 6. Degenerate problem: Iteration history for nested iteration (left) and asymptotic convergence rates (right)

7.3. **Projected Multilevel Relaxation Versus Monotone Multigrid.** We investigate the convergence behavior of the truncated monotone multigrid method TMMG stated in Algorithm 5.10 and of the standard version SMMG stated in Algorithm 5.12. As we are interested to see to which extend the convergence properties of PMLR are preserved by its V-cycle counterpart SMMG, we include the results of PMLR for a comparison. We always consider $V(1,1)$ cycles.

7.3.1. *Spiral problem.* We consider the discrete problem stated in Section 7.1.1. The iteration histories for the "bad" initial iterates $u_j^0 = \varphi_j$, $u_j^0 = \varphi_j + 10 > \varphi_j$ and nested iteration are shown in Figure 7 and the left picture of Figure 8, respectively.

Let us first notice that the intrinsic damping (cf. Remark 5.9) leading to optimal complexity hardly slows down the convergence of SMMG as compared to PMLR. Optimal resolution of the constraints at the cost of suboptimal complexity improves the convergence speed of PMLR only for $u_j^0 = \varphi_j + 10$. Both methods provide almost identical results otherwise. In particular, the convergence behavior of SMMG is still almost independent of the choice of initial iterates. This is different for TMMG. Starting with $u_j^0 = \varphi_j$ the iteration begins very slowly, because inactivation is performed exclusively by projected Gauß–Seidel relaxation (cf. Remark 5.8). However, once the coincidence set is approximated sufficiently well, the iteration accelerates tremendously. The asymptotic convergence rates of TMMG are essentially the same as for the corresponding linear multigrid method applied to an *unconstrained* Poisson problem on Ω. This supports our heuristic reasoning in Section 5. For $u_j^0 = \varphi_j + 10 > \varphi_j$ coarse grid correction can contribute right from the start. This leads to faster convergence throughout the iteration as illustrated in Figure 7 (right). Again, the asymptotic linear convergence speed is reached once the coincidence set is approximated sufficiently well. According to Figure 8 (left) asymptotic linear convergence starts immediately for nested iteration. In this case, the performance of TMMG can be hardly distinguished from classical multigrid for unconstrained problems. Figure 8 shows ρ_k over the corresponding number of unknowns n_k, $k = 2, \ldots, 9$. The asymptotic convergence rates seem to saturate at about 0.73 (PMLR and SMMG) and 0.41 (TMMG). In comparison with previous computations for generic problems [44, 51], the complicated coincidence set does not seem to deteriorate the convergence properties of all three methods.

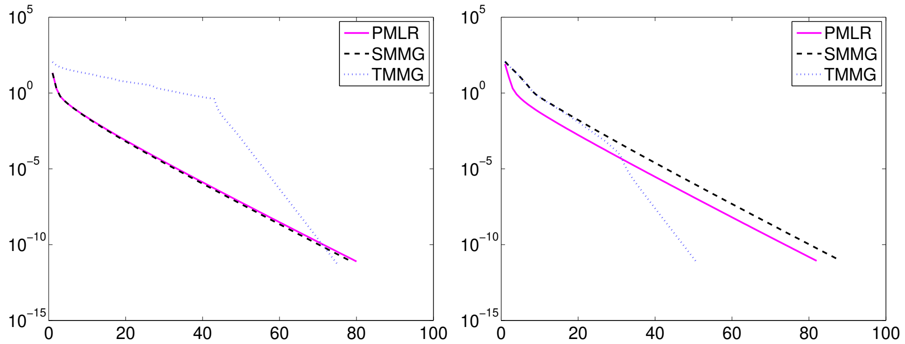

FIGURE 7. Spiral problem: Iteration histories for the initial iterates $u_j^0 = \varphi_j$ (left) and $u_j^0 = \varphi_j + 10 > \varphi_j$ (right)

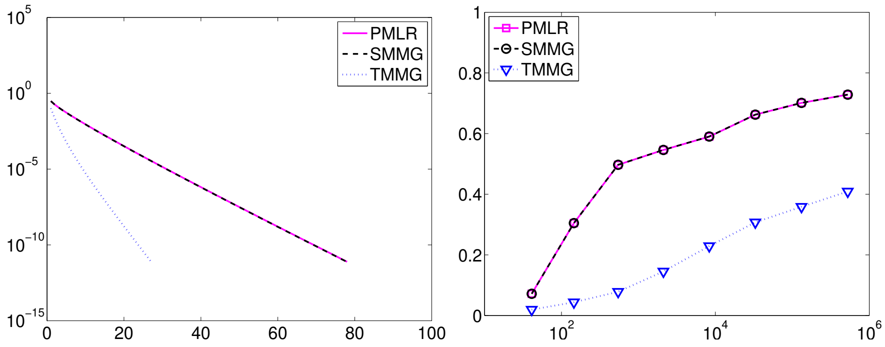

FIGURE 8. Spiral problem: Iteration history for nested iteration (left) and asymptotic convergence rates (right)

7.3.2. *Degenerate problem.* We consider the discrete problem stated in Section 7.1.2. The iteration histories of PMLR, SMMG, and TMMG for the initial iterates $u_j^0 = \varphi_j$, $u_j^0 = 0 > \varphi_j$, and nested iteration are depicted in Figure 9 and Figure 10 (left), respectively.

Let us first compare the convergence properties of PMLR and SMMG. Both methods perform quite similarly for $u_j^0 = \varphi_j$. However, the decelerating effect of intrinsic damping in SMMG is clearly visible in the remaining two cases. This effect is stronger than before, because in this example u_j and φ_j are almost identical. Starting with $u_j^0 = \varphi_j$, TMMG also behaves qualitatively as before: After a transient phase dominated by slow inactivation by projected Gauß–Seidel relaxation, the iteration accelerates considerably, reaching linear multigrid convergence speed. However, this time it takes about 80 steps until the coincidence set is approximated sufficiently well. If $u_j^0 = 0$ or nested iteration is used, the behavior of TMMG now can be hardly distinguished from SMMG. In contrast to the non–degenerate case, truncation now does not have much of a positive effect, because the exact coincidence set is detected not until one or two steps before the desired accuracy is reached. As a consequence, the asymptotic convergence rates depicted in Figure 10 (right) are now comparable for all three methods. Summing up, SMMG and PMLR do not suffer from degeneracy while, apart from possible slow inactivation, TMMG now essentially behaves like SMMG.

7.4. **Truncated Monotone Multigrid Versus Truncated Nonsmooth Newton Multigrid.** We apply the truncated nonsmooth Newton multigrid method TNMG stated in Algorithm 6.2 to the test problems introduced in Section 7.1. For a comparison we include the results of the related truncated monotone multigrid method TMMG stated in Algorithm 5.10. We always consider $V(1,1)$ cycles.

In our numerical experiments, we check the robustness of the convergence behavior with respect to given "bad" initial iterates and "good" initial iterates as obtained by nested iteration. The convergence behavior is illustrated by *iteration histories* showing the algebraic error $\|u_j - u_j^\nu\|$ over the iteration steps ν. In order to evaluate possible mesh–dependence, we consider the *asymptotic convergence rates* ρ_k, $k = 0, \dots, j$ as introduced in (7.1).

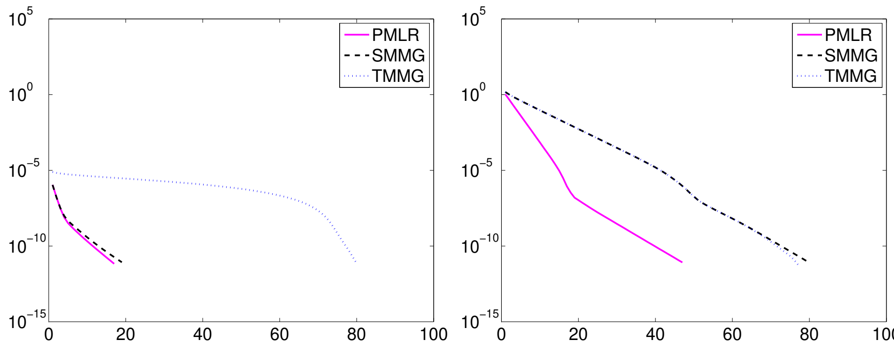

FIGURE 9. Degenerate Problem: Iteration histories for the initial iterates $u_j^0 = \varphi_j$ (left) and $u_j^0 = 0 > \varphi_j$ (right)

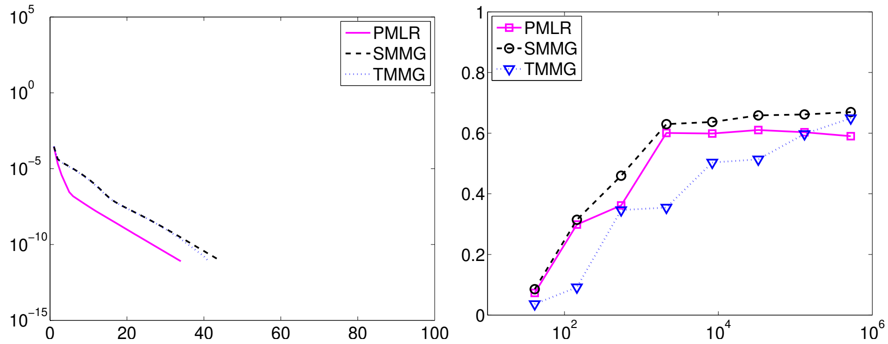

FIGURE 10. Degenerate Problem: Iteration history for nested iteration (left) and asymptotic convergence rates (right)

7.4.1. *Spiral problem.* We consider the discrete problem stated in Section 7.1.1. The iteration histories for the "bad" initial iterates $u_j^0 = \varphi_j$, $u_j^0 = \varphi_j + 10 > \varphi_j$ and for nested iteration are shown in Figure 11 and the left picture of Figure 12, respectively. As expected from Remark 6.5, inactivation by projected Gauß–Seidel iteration initially slows down the convergence speed for the initial iterate $u_j^0 = \varphi_j$. Fast convergence throughout the iteration is observed for $u_j^0 = \varphi_j + 10 > \varphi_j$. While TNMG performes slightly better than TMMG for those "bad" initial iterates, both methods provide almost identical results, if nested iteration is applied. In this case the performance of both TMMG and TMMG can be hardly distinguished from classical multigrid for unconstrained problems. The asymptotic convergence rates, apparently saturating at about 0.41, are almost the same, cf. Figure 12 (right). In view of its simplicity and its potential for further improvements (line search, cg–acceleration) these results suggest that TNMG should be preferred to TMMG.

7.4.2. *Degenerate problem.* We consider the discrete problem stated in Section 7.1.2. The iteration histories for the initial iterates $u_j^0 = \varphi_j$, $u_j^0 = 0 > \varphi_j$ and for nested iteration are shown in Figure 13 and the left picture of Figure 14, respectively. Note that $u_j^0 = \varphi_j$ provides the minimal initial error in this case. Nevertheless, starting from the obstacle, next–neighbor inactivation by projected Gauß–Seidel

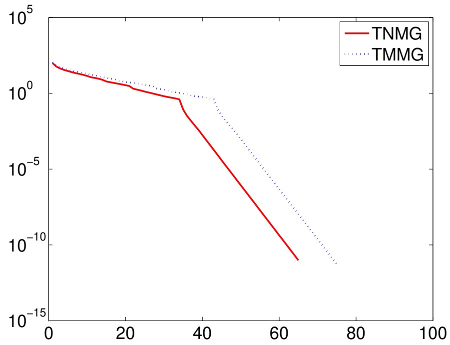

FIGURE 11. Spiral problem: Iteration histories for the initial iterates $u_j^0 = \varphi_j$ (left) and $u_j^0 = 0 > \varphi_j$ (right)

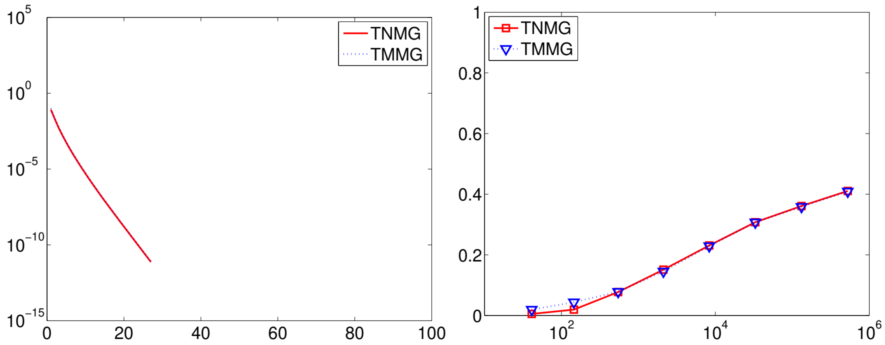

FIGURE 12. Spiral problem: Iteration history for nested iteration (left) and asymptotic convergence rates (right)

relaxation slows down the convergence speed of TNMG until the coincidence set is approximated sufficiently well. As in the preceding experiment, TNMG performs only slightly better than TMMG in this case. However, starting from $u_j^0 = 0 > \varphi_j$, TMMG is clearly outperformed by TNMG: While intrinsic local damping is reducing the effect of coarse grid correction in TMMG, global damping as used in TNMG is less pessimistic in this case.

The situation is quite different, if nested iteration is applied. After a very fast reduction of the high–frequency components of the error, the coarse grid correction of TNMG produces some kind of undershoot in the sense that the approximate coincidence set is much too large. This leads to slow convergence during the following inactivation phase. Finally, once the coincidence set is approximated sufficiently well, asymptotic linear convergence is reached. This behavior occurs on all refinement levels $k = 2, \dots, 9$. Hence, the increasing asymptotic convergence rates depicted in the right picture of Figure 14 rather reflects the increasing length of the intermediate inactivation phase than the asymptotic convergence speed itself. More robust global inactivation by standard monotone multigrid will be discussed below.

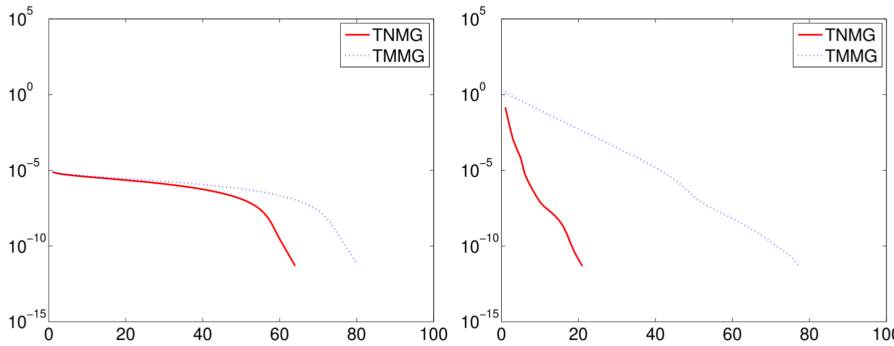

FIGURE 13. Degenerate Problem: Iteration histories for the initial iterates $u_j^0 = \varphi_j$ (left) and $u_j^0 = 0 > \varphi_j$ (right)

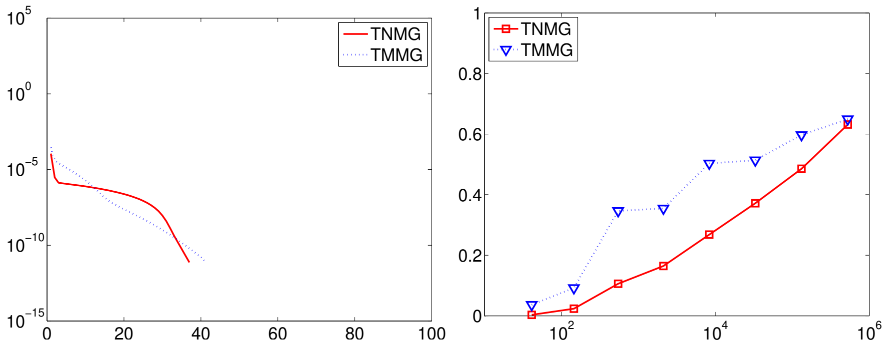

FIGURE 14. Degenerate Problem: Iteration history for nested iteration (left) and asymptotic convergence rates (right)

7.5. **Concluding remarks.** Our numerical experiments show that there is a certain tradeoff between robustness and convergence speed. In perfect agreement with theory the multilevel relaxation MSD, the projected multilevel relaxation PMLR and its V-cycle counterpart show a remarkable amount of robustness, but much slower convergence than classical multigrid methods for related unconstrained problems. Conversely, truncated monotone multigrid TMMG and the related truncated nonsmooth Newton multigrid method TNMG are very fast for generic problems and initial iterates but might run into trouble for certain bad initial iterates (cf. Remark 5.8 and 5.13) and degenerated cases.

Similar to truncated monotone multigrid TMMG or truncated nonsmooth Newton multigrid TNMG the inactivation in present primal dual active set strategies is based on local next–neighbor interaction [45, 46, 43]. This explains why mesh–independent convergence of such methods is still (and might remain) open. In addition, local inactivation typically leads to deteriorating convergence speed, if the approximate coincidence set is much too large. As a remedy, it seems natural to use *global* inactivation by one step of the standard monotone multigrid STDMG. The resulting hybrid multigrid method HMG thus combines robustness with fast asymptotic convergence. Note that, by construction, such hybrid multigrid method are globally convergent. To illustrate the possible benefit of the hybrid approach, we again consider the degenerate problem stated in section 7.1.2 with initial iterates obtained by nested iteration. The iteration history of HMG is shown in the left picture of Figure 15. For a fair comparison with SMMG and TNMG, we counted each step of HMG twice. Combining the robustness of multilevel relaxation with fast asymptotic convergence of truncated multigrid, HMG clearly outperforms SMMG and TNMG. According to the right picture in Figure 15 the convergence rates seem to saturate at about 0.3. A theoretical justification is left to future research.

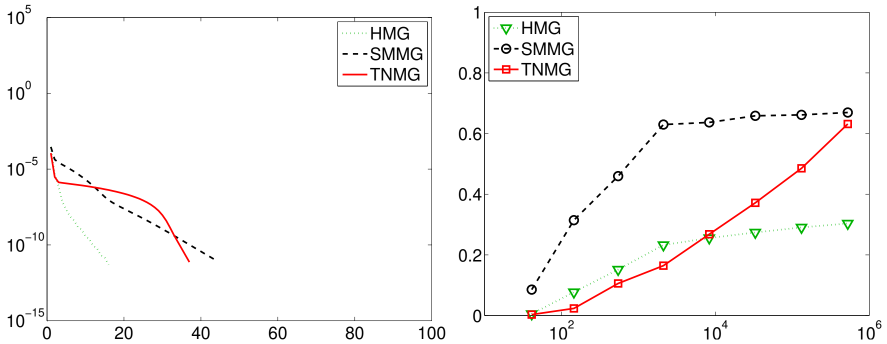

FIGURE 15. Degenerate Problem: Iteration history for nested iteration (left) and asymptotic convergence rates (right)

REFERENCES

[1] L. Badea. Convergence rate of a Schwarz multilevel method for the constrained minimization of nonquadratic functionals. *SIAM J. Numer. Anal.*, 44(2):449–477, 2006.

[2] L. Badea, X.-Ch. Tai, and J. Wang. Convergence rate analysis of a multiplicative Schwarz method for variational inequalities. *SIAM J. Numer. Anal.*, 41(3):1052–1073, 2003.

[3] C. Baiocchi and A. Capelo. *Variational and Quasivariational Inequalities*. Wiley, Chichester, 1984.

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

[5] S. Bartels and C. Carstensen. Averaging techniques yield reliable a posteriori finite element error control for obstacle problems. *Numer. Math.*, 99(2):225–249, 2004.

[6] V. Belsky. A multi–grid method for variational inequalities in contact problems. *Computing*, 51:293–311, 1993.

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

[8] H. Blum, D. Braess, and F.T. Suttmeier. A cascadic multigrid algorithm for variational inequalities. *Comput. Vis. Sci.*, 7:153–157, 2004.

[9] C. Bollrath. *Zwei Mehrgitterverfahren zur numerischen Berechnung von stationären Strömungen durch poröse Medien mit freiem Rand*. PhD thesis, Universität Bochum, 1985.

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

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

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

[13] D. Braess. A posteriori error estimators for obstacle problems – another look. *Numer. Math.*, 101:415–421, 2005.

[14] D. Braess, C. Carstensen, and R.H.W. Hoppe. Convergence analysis of a conforming adaptive finite element method for an obstacle problem. *Numer. Math.*, 107:455–471, 2007.

[15] 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.

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

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

[18] 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.

[19] 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.

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

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

[22] A. Brandt. Guide to multigrid development. In Hackbusch and Trottenberg [41], pages 220–312.

[23] 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.

[24] H. Brézis. Seuil de régularité pour certaines problèmes unilateraux. *C.R. Ac. Sci. Paris*, 273:35–37, 1971.

[25] F. Brezzi, W.W. Hager, and P.A. Raviart. Error estimates for the finite element solution of variational inequalities I. *Numer. Math.*, 28:431–443, 1977.

[26] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North–Holland, Amsterdam, 1978.

[27] F.H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley and Sons, New York, 1983.

[28] R.W. Cottle, J.S. Pang, and R.E. Stone. *The Linear Complementarity Problem*. Academic Press, Boston, 1992.

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

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

[31] 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.

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

[33] K. Fackeldey and R.H. Krause. Solving frictional contact problems with multigrid efficiency. In O. Widlund and D. Keyes, editors, *Domain Decomposition Methods in Science and Engineering*, volume 55 of *LNCSE*, pages 549–556. Springer, 2006.

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

[35] G. Fichera. Problemi elastotatici con vincoli unilaterali: il problema di Signorini con ambigue condizioni al contorno. *Atti. Acad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Nat. Sez. Ia*, 7(8):91–140, 1963-1964.

[36] E. Gelman and J. Mandel. On multilevel iterative method for optimization problems. *Math. Prog.*, 48:1–17, 1990.

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

[38] W. Hackbusch. Multi–grid convergence theory. In Hackbusch and Trottenberg [41], pages 177–219.

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

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

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

[42] B. Heimsund, X.-C. Tai, and J. Xu. Rate of convergence for parallel subspace correction methods for nonlinear variational inequalities. In N.D̃ebit et al., editor, *Domain Decomposition Methods in Science and Engineering*, pages 127–138. CIMNE, 2002.

[43] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set method as a semi-smooth Newton method. *SIAM J. Opt.*, 13:865–888, 2003.

[44] M. Holtz and A. Kunoth. B-spline-based monotone multigrid methods. *SIAM J. Numer. Anal.*, 45:1175–1199, 2007.

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

[46] R.H.W. Hoppe. Two-sided approximations for unilateral variational inequalities by multigrid methods. *Optimization*, 18:867–881, 1987.

[47] R.H.W. Hoppe. Une méthode multigrille pour la solution des problèmes d'obstacle. M^2 *AN*, 24:711–736, 1990.

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

[49] S. Hüeber, M. Mair, and B.I. Wohlmuth. A priori error estimates and an inexact primal-dual active set strategy for linear and quadratic finite elements applied to multibody contact problems. *Appl. Num. Math.*, 54:555–576, 2005.

[50] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Their Applications.* Academic Press, New York, 1980.

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

[52] R. Kornhuber. Monotone multigrid methods for elliptic variational inequalities II. *Numer. Math.*, 72:481 – 499, 1996.
[53] R. Kornhuber. A posteriori error estimates for elliptic variational inequalities. *Computers Math. Applic.*, 31:49–60, 1996.
[54] R. Kornhuber. Nonlinear multigrid techniques. In J.F. Blowey, J.P. Coleman, and A.W. Craig, editors, *Theory and Numerics of Differential Equations*, Universitext, pages 179–229. Springer, 2001.
[55] R. Kornhuber. On constrained Newton linearization and multigrid for variational inequalities. *Numer. Math.*, 91:699–721, 2002.
[56] R. Kornhuber and R. Krause. Adaptive multigrid methods for Signorini's problem in linear elasticity. *Comp. Visual. Sci.*, 4:9–20, 2001.
[57] R. Kornhuber and R. Krause. On multigrid methods for vector-valued Allen-Cahn equations. In I. Herrera et al., editor, *Domain Decomposition Methods in Science and Engineering*, pages 307 – 314, Mexico City, Mexico, 2003. UNAM.
[58] R. Kornhuber and R. Krause. Robust multigrid methods for vector-valued Allen–Cahn equations with logarithmic free energy. *Comput. Vis. Sci.*, 9:103–116, 2006.
[59] R. Kornhuber, R. Krause, O. Sander, P. Deuflhard, and S. Ertel. A monotone multigrid solver for two body contact problems in biomechanics. *Comp. Visual. Sci.*, 11:3–15, 2008.
[60] R. Kornhuber and H. Yserentant. Multilevel methods for elliptic problems on domains not resolved by the coarse grid. *Contemporary Mathematics*, 180:49–60, 1994.
[61] R. Krause. From inexact active set strategies to nonlinear multigrid methods. In P. Wriggers and U. Nackenhorst, editors, *Analysis and Simulation of Contact Problems*, volume 27 of *Lecture Notes in Applied and Computational Mechanics*. Springer, 2006.
[62] J. Mandel. Etude algébrique d'une méthode multigrille pour quelques problèmes de frontière libre. *C.R. Acad. Sci.*, 298:469–472, 1984. Ser. I.
[63] J. Mandel. A multilevel iterative method for symmetric, positive definite linear complementarity problems. *Appl. Math. Optimization*, 11:77–95, 1984.
[64] F. Natterer. Optimale L_2–Konvergenz finiter Elemente bei Variationsungleichungen. *Bonn. Math. Schr.*, 89:1–12, 1976.
[65] N. Neuss. V-cycle convergence with unsymmetric smoothers and application to an anisotropic model problem. *SIAM J.Numer.Anal.*, 35:1201–1212, 1998.
[66] R.H. Nochetto, K.G. Siebert, and A. Veeser. Fully localized a posteriori error estimators and barrier sets for contact problems. *SIAM J. Numer. Anal.*, 42:2118–2135, 2005.
[67] P. Oswald. On function spaces related to finite element approximation theory. *Z. Anal. Anwend.*, 9:43–64, 1990.
[68] F.A. Pérez, J.M. Cascón, and L. Ferragut. A numerical adaptive algorithm for the obstacle problem. In M. Bubak, editor, *Proceedings of Computational science – ICCS, Part II*, volume 3037 of *Lecture Notes in Computer Science*, pages 130–137. Springer, 2004.
[69] W.H. Press, S.A. Teukolsky, W. Vetterling, and B.P. Flannery. *Numerical Recipes in C.* Cambridge University Press, New York, 1988.
[70] P.A. Raviart and J.M. Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles.* Masson, Paris, 1992.

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

[72] J.F. Rodrigues. *Obstacle Problems in Mathematical Physics*. Number 134 in Mathematical Studies. North–Holland, Amsterdam, 1987.

[73] O. Sander. *Multidimensional Coupling in a Human Knee Model*. PhD thesis, FU Berlin, 2008.

[74] J. Schöberl. Solving the Signorini problem on the basis of domain decomposition techniques. *Computing*, 60(4):323–344, 1998.

[75] K. Siebert and A. Veeser. A unilaterally constrained quadratic minimization with adaptive finite elements. *SIAM J. Optim.*, 18:260–289, 2007.

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

[77] G. Stampacchia. Formes bilinéaires coercitives sur les ensembles convexes. *C.R. Acad. Sci. Paris*, 258:4413–4416, 1964.

[78] X.-C. Tai. Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. *Numer. Math.*, 93(4):755–786, 2003.

[79] X.-C. Tai and M. S. Espedal. Rate of convergence of some space decomposition method for linear and non-linear elliptic problems. *SIAM J. Numer. Anal.*, 35:1558–1570, 1998.

[80] X.-C. Tai and P. Tseng. Convergence rate analysis of an asynchronous space decomposition method for convex minimization. *Math. Comp.*, 71:1105–1135, 2001.

[81] X.-C. Tai and J. Xu. Global and uniform convergence of subspace correction methods for some convex optimization problems. *Math. Comp.*, 71:105–124, 2001.

[82] M. Ulbrich. Nonsmooth Newton-like methods for variational inequalities and constrained optimization problems in function spaces. Habilitationsschrift, TU München, 2002.

[83] A. Veeser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems. *SIAM J. Numer. Anal.*, 39:146–167, 2001.

[84] B. Wohlmuth and R. Krause. Monotone multigrid methods on nonmatching grids for nonlinear multibody contact problems. *SIAM J. Sci. Comput.*, 25:324–347, 2003.

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

[86] J. Xu. An introduction to multigrid convergence theory. In R.H. Chan et al., editor, *Iterative Methods in Scientific Computing*, pages 169–241. Springer, 1997.

[87] J. Xu. An introduction to multilevel methods. In M. Ainsworth et al., editor, *Wavelets, Multilevel Methods and Elliptic PDEs*, pages 213–302, Oxford, 1997. Clarendon Press.

[88] J. Xu. The method of subspace corrections. *J. Comp. Appl. Math.*, 128:335–362, 2001.

[89] J. Xu and L. Zikatanov. The method of alternating projections and the method of subspace corrections in Hilbert space. *J. Am. Math. Soc.*, 15(3):573–597, 2002.

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

[91] H. Yserentant. Old and new convergence proofs for multigrid methods. *Acta Numerica*, pages 285–326, 1993.

Carsten Gräser, Freie Universität Berlin, Institut für Mathematik, Arnimallee 6, D - 14195 Berlin, Germany
E-mail address: graeser@math.fu-berlin.de

Prof. Dr. Ralf Kornhuber, Freie Universität Berlin, Institut für Mathematik, Arnimallee 6, D - 14195 Berlin, Germany
E-mail address: kornhuber@math.fu-berlin.de