NONSMOOTH SCHUR–NEWTON METHODS FOR NONSMOOTH SADDLE POINT PROBLEMS

CARSTEN GRÄSER

Abstract. We introduce and analyze nonsmooth Schur-Newton methods for a class of nonsmooth saddle point problems. The method is able to solve problems where the primal energy decomposes into a convex smooth part and a convex separable but nonsmooth part. The method is based on nonsmooth Newton techniques for an equivalent unconstrained dual problem. Using this we show that it is globally convergent even for inexact evaluation of the linear subproblems.

1. Introduction

We consider the iterative solution of large scale nonlinear saddle point problems

$$u^* \in \mathbb{R}^n,\ w^* \in \mathbb{R}^m : \qquad \begin{pmatrix} F & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u^* \\ w^* \end{pmatrix} \ni \begin{pmatrix} f \\ g \end{pmatrix}, \tag{1}$$

where B, C are suitable matrices, and the set-valued operator $F = \partial J$ is the subdifferential of a convex functional $J = J_0 + \varphi$ that decomposes in to a smooth part J_0 and a separable nonsmooth part φ. Such problems arise e.g. from discretizations of Cahn–Hilliard equations [3], optimal control problems for linear pdes, and plasticity problems.

For discretized Cahn–Hilliard equations B is mass matrix, C a stiffness matrix, and J is the discrete analogue of a functional

$$\mathcal{J}(v) = \int_\Omega \gamma(\nabla v)^2 dx + \Big(\int_\Omega v\,dx\Big)^2 + \int_\Omega \Phi(v) dx$$

for some convex $\Phi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ and a positive 1-homogeneous convex function $\gamma : \mathbb{R}^d \to \mathbb{R}$. For an isotropic surface energy γ is a scaled euclidean norm. A classical choice for Φ is the so called logarithmic potential

$$\Phi_\theta(v) = \frac{\theta}{2}\Big((1+v)\ln(1+v) + (1-v)\ln(1-v)\Big)$$

depending on the temperature $\theta \geq 0$. For $\theta \to 0$ this degenerates to the obstacle potential

$$\Phi_0(v) = \chi_{[-1,1]}(v) = \begin{cases} 0 & \text{if } v \in [-1,1], \\ \infty & \text{else.} \end{cases}$$

For problems where J_0 is quadratic and φ is the indicator functional of a hypercube, like it is the case for the isotropic Cahn–Hilliard equation with obstacle potential, the nonsmooth Schur–Newton method was introduced in [14] and analyzed in [16]. This method is essentially a nonsmooth Newton type method for a

nonlinear Schur complement of (1). Global convergence can be shown by interpreting it as a descent method for an unconstrained dual problem. In numerical examples the method exhibits mesh independent convergence rates.

In the present paper we generalize this approach such that the case of Cahn–Hilliard equations with logarithmic potential [5, 8] and smooth anisotropy functions can also be solved with the same efficiency.

The paper is organized as follows: In Section 2 we will introduce the full problem and necessary assumptions and note some important properties. Sections 3 and 4 are devoted to the introduction of the dual problem and gradient related descent directions. Finally we derive linearizations for the nonlinear Schur complement and introduce the resulting Newton-type methods in Section 5 and 6.

2. Problem Formulation

Throughout the paper we will make the following assumptions.

(A1) $J_0 : \mathbb{R}^n \to \mathbb{R}$ is strongly convex and continuously differentiable with Lipschitz continuous derivative. I.e., there are s.p.d. matrices $\underline{H}_{J_0}, \overline{H}_{J_0} \in \mathbb{R}^{n,n}$ such that

$$\begin{aligned} \|\nabla J_0(x) - \nabla J_0(y)\| &\leq \|x-y\|_{\overline{H}_{J_0}} && \forall x, y \in \mathbb{R}^n, \\ \langle \nabla J_0(x) - \nabla J_0(x), x-y\rangle &\geq \|x-y\|^2_{\underline{H}_{J_0}} && \forall x, y \in \mathbb{R}^n. \end{aligned}$$

(A2) $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ takes the form

$$\varphi(v) = \sum_{i=1}^{n} \varphi_i(v_i).$$

Each $\varphi_i : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is convex, lower semicontinuous on $\mathbb{R}$, continuous on its domain $\operatorname{dom} \varphi_i$, and twice continuously differentiable on a finite number of disjoint nonempty open intervals (a_i^k, a_i^{k+1}), $a_i^k \in \mathbb{R} \cup \{-\infty, +\infty\}$ having the property

$$\overline{\operatorname{dom} \varphi_i} = \overline{\{x : \varphi_i(x) < \infty\}} = \bigcup\nolimits_{k=1}^{m_i} \overline{(a_i^{k-1}, a_i^k)} = \overline{(a_i^0, a_i^{m_i})}.$$

The intervals are maximal in the sense that φ_i is not twice continuously differentiable on (a_i^k, a_i^{k+2}). Furthermore, the limits

$$\lim_{\xi \nearrow a_i^{k+1}} \varphi_i''(\xi), \qquad \lim_{\xi \searrow a_i^{k}} \varphi_i''(\xi)$$

exist in $\mathbb{R} \cup \{\infty\}$ for $k = 0, \dots, (m_i - 1)$.

(A3) $B \in \mathbb{R}^{m,n}$, $f \in \mathbb{R}^n$, and $g \in \mathbb{R}^m$. $C \in \mathbb{R}^{m,m}$ is symmetric and positive semidefinite.

Under these assumptions (1) is equivalent to finding a saddle point of the associated Lagrange functional

$$\mathcal{L}(u, w) = J(u) - \langle f, u\rangle + \langle Bu - g, w\rangle - \frac{1}{2}\langle Cw, w\rangle .$$

(A4) The saddle point problem (1) has a unique solution.

Notice that the last assumption is necessary since there is no general existence and uniqueness result for problem (1). In contrast to this, the existence, uniqueness, and stability of solutions for minimization problems associated with J follows from standard arguments:

Proposition 2.1. *Assume that (A1) and (A2) hold. Then J is strongly convex, proper, lower semicontinuous, and coercive. The subdifferential $\partial J : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ has a single-valued, monotone inverse $(\partial J)^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ characterized by the variational inequality*

$$x \in \mathbb{R}^n : \qquad \langle \nabla J_0(x), v - x\rangle + \varphi(v) - \varphi(x) \geq \langle y, v - x\rangle \qquad \forall v \in \mathbb{R}^n \tag{2}$$

for $x = (\partial J)^{-1}(y)$. The operator $(\partial J)^{-1}$ is Lipschitz continuous with

$$\|(\partial J)^{-1}(y^1) - (\partial J)^{-1}(y^2)\|_{\underline{H}_{J_0}} \leq \|y^1 - y^2\|_{\underline{H}_{J_0}^{-1}}.$$

Proof. The properties of J follow directly from (A1) and (A2). Single-valuedness of $(\partial J)^{-1}$ and equivalence to the variational inequality follows from standard arguments, see, e.g., [7, Chapter II]. Lipschitz continuity can be shown adding the variational inequalities for y^1 and y^2 and using the strong monotonicity of ∇J_0. □

For the rest of the paper we assume the saddle point problem (1) satisfies (A1)–(A4).

3. Dual problem

Before we discuss the iterative solution of this problem class we derive an equivalent dual minimization problem.

Proposition 3.1. *The saddle point problem* (1) *is equivalent to*

$$w^* \in \mathbb{R}^m : \qquad H(w^*) = 0 \tag{3}$$

with the Lipschitz continuous, monotone operator $H : \mathbb{R}^m \to \mathbb{R}^m$ given by

$$H(w) = -BF^{-1}(f - B^T w) + Cw + g\,, \qquad w \in \mathbb{R}^m\,. \tag{4}$$

Proof. Due to the properties of J and F straightforward block elimination in (1) provides the equivalence.

Since H consists of a sum and a composition of F^{-1} with affine functions the Lipschitz continuity follows directly from the Lipschitz continuity of F^{-1}. By the convexity of J the operator F and thus F^{-1} is monotone. In combination with the non-negativity of C this implies monotonicity of H. □

The operator H can be regarded as a nonlinear Schur complement. For a linear saddle point problem (where F is a symmetric positive definite matrix) it reduces to the classical linear Schur complement. In contrast to the linear case, the right hand side f cannot be separated from the part depending on w in general. Note that although the saddle point problem is set-valued, the operator H is single-valued, because $F^{-1} = (\partial\varphi)^{-1}$ is single-valued or, equivalently, the minimization of $J(x) - \langle y, x\rangle$ on $\mathbb{R}^n$ admits a unique solution.

Theorem 3.1. *There is a Fréchet-differentiable, convex functional $h : \mathbb{R}^m \to \mathbb{R}$ with the property $\nabla h = H$ and the representation*

$$h(w) = -\mathcal{L}(F^{-1}(f - B^T w), w)\,, \qquad w \in \mathbb{R}^m. \tag{5}$$

Proof. By [7, Corollary 5.2, p. 22] the polar (or conjugate) functional

$$J^*(y) = \sup_{x\in\mathbb{R}^n} (\langle y, x\rangle - J(x)) = - \inf_{x\in\mathbb{R}^n} (J(x) - \langle y, x\rangle)$$

of J is convex and has the property $\partial J^* = (\partial J)^{-1} = F^{-1}$. Since $F^{-1}(y)$ is single-valued for all $y \in \mathbb{R}^n$ its polar J^* can take only finite values and the domain of the polar is $\mathbb{R}^n$. Thus J^* is continuous on the whole space $\mathbb{R}^n$ by [7, Corollary 2.3, p. 12].

By [7, Proposition 5.3, p. 23] finiteness and continuity of J^*, and single-valuedness of ∂J^* imply Gâteaux-differentiability of J^*. The continuity of $\partial J^* = F^{-1}$ implies that J^* is even Fréchet-differentiable with $\nabla J^* = F^{-1}$. Setting

$$h(w) = J^*(f - B^T w) + \frac{1}{2}\langle Cw, w\rangle + \langle g, w\rangle \tag{6}$$

we immediately get $\nabla h = H$ using the chain rule. Convexity of h directly follows from convexity of J^*, and symmetry and positivity of C. Finally, inserting

$$J^*(y) = - \left(J(F^{-1}(y)) - \left\langle y, F^{-1}(y)\right\rangle\right)$$

with $y = f - B^T w$ into (6) gives (5). □

As immediate consequence of Proposition 3.1 and Theorem 3.1 we get the equivalence of (1) to an unconstrained dual problem.

Corollary 3.1. *The set-valued saddle point problem* (1) *is equivalent to the dual unconstrained convex minimization problem*

$$w^* \in \mathbb{R}^m : \qquad h(w^*) \le h(w) \qquad \forall w \in \mathbb{R}^m. \tag{7}$$

The equivalence in Corollary 3.1 does especially imply, that the minimization problem (7) has a unique solution due to assumption (A4). Note that if C is even positive definite, then h is strongly convex which would already guarantee (A4). However, the latter need not be the case, and h is in general not even strictly convex so that we have to require uniqueness separately.

Corollary 3.1 offers the possibility to treat the nonsmooth saddle point problem (1) as a smooth unconstrained minimization problem or as an operator equation with a Lipschitz continuous monotone operator. This simplification comes at the price of the fact that the functional h and the operator $H = \nabla h$ might be expensive to evaluate, since both involve the evaluation of $F^{-1} = (\partial J)^{-1}$ and thus the solution of an unconstrained minimization problem for the nonsmooth functional J.

4. Descent Methods for the Dual Problem

Once we have reformulated the saddle point problem (1) as the dual minimization problem (7), descent methods for unconstrained minimization of differentiable functionals can be applied.

Since the operator F^{-1} involved in h and $H = \nabla h$ is in general not directly available, it is complicated and expensive to use iterative methods based on local properties or substeps as e.g. the Gauß–Seidel or Jacobi method for the solution of (7). For this reason we consider gradient-related algorithms based on global descent directions.

Although there are numerous convergence results (see, e.g., the classic text book by Ortega and Rheinboldt [21]) for this class of methods none of them fits exactly

to the proposed problem. Thus we present modified variants for the presented problem and refer to [12] for proofs of those variants.

Throughout this section we assume that h is given by (5). However we will only need that $h : \mathbb{R}^m \to \mathbb{R}$ is a convex and continuously differentiable functions with Lipschitz continuous derivative ∇h and a unique minimizer w^*.

The results are presented in terms of the norm $\|\cdot\|_M$,

$$\|x\|_M^2 = \langle Mx, x\rangle \ , \qquad x \in \mathbb{R}^m \ ,$$

induced by a symmetric positive definite matrix $M \in \mathbb{R}^{m,m}$. Elements x' of the dual space $(\mathbb{R}^m)'$ are represented by $x \in \mathbb{R}^m$ using $x' = \langle x, \cdot\rangle$ with the Euclidean inner product $\langle\cdot,\cdot\rangle$, thus the dual space $(\mathbb{R}^m, \|\cdot\|_M)'$ is identified with $(\mathbb{R}^m, \|\cdot\|_{M^{-1}})$.

Gradient related descent methods are of the form

$$w^{\nu+1} = w^\nu + \rho_\nu d^\nu, \qquad \nu = 1, \dots \tag{8}$$

for a given initial iterative w^0. In each step, first a search direction d^ν is chosen according to the current iterate w^ν. Then, a step size ρ_ν is fixed depending on w^ν and d^ν, i.e.,

$$d^\nu = d(\nu, w^\nu), \qquad \rho_\nu = \rho(\nu, w^\nu, d^\nu) \ , \quad \nu = 0, 1, \dots \tag{9}$$

with suitable mappings d and ρ. Since it will turn out that monotonicity of the iterated is the crucial property for convergence we consider the extended algorithm

$$w^{\nu+\frac{1}{2}} = w^\nu + \rho_\nu d^\nu, \tag{10}$$

$$w^{\nu+1} = w^{\nu+\frac{1}{2}} + \mathcal{C}(w^{\nu+\frac{1}{2}}) \tag{11}$$

with an operator $\mathcal{C}$ having the property $h(w + \mathcal{C}(w)) \le h(w)$.

4.1. **Convergence Analysis.** In order to obtain a convergent method the descent directions should allow for sufficient descent of h and the step sizes must realize the descent.

Definition 4.1. *The map $d : \mathbb{N} \times \mathbb{R}^m \to \mathbb{R}^m$ is said to generate gradient-related directions (or descent directions) if for any sequence $w^\nu \subset \mathbb{R}^m$ the directions $d^\nu = d(\nu, w^\nu)$ satisfy*

$$\nabla h(w^\nu) = 0 \quad \Longleftrightarrow \quad d^\nu = 0, \qquad \forall \nu \in \mathbb{N} \tag{12}$$

and

$$-\langle \nabla h(w^\nu), d^\nu\rangle \ge c_D \, \|\nabla h(w^\nu)\|_{M^{-1}} \, \|d^\nu\|_M \, , \qquad \forall \nu \in \mathbb{N} \tag{13}$$

with a constant $c_D > 0$ (or $c_D = 0$) independent of ν.

Note that the preconditioned gradients $d(\nu, w^\nu) = -M^{-1}\nabla h(w^\nu)$ are gradient-related since (13) is satisfied with equality and $c_D = 1$. In all other cases the Cauchy–Schwarz inequality implies $c_d < 1$.

Definition 4.2. *Let $d : \mathbb{N} \times \mathbb{R}^m \to \mathbb{R}^m$ generate descent directions. Then $\rho : \mathbb{N} \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is said to generate efficient step sizes, if for any sequence $w^\nu \subset \mathbb{R}^m$ and $d^\nu = d(\nu, w^\nu)$ the step sizes $\rho_\nu = \rho(\nu, w^\nu, d^\nu)$ satisfy*

$$d^\nu \neq 0 \qquad \Rightarrow \qquad h(w^\nu + \rho_\nu d^\nu) \le h(w^\nu) - c_S \left(\frac{\langle \nabla h(w^\nu), d^\nu\rangle}{\|d^\nu\|_M} \right)^2 \qquad \forall \nu \in \mathbb{N} \tag{14}$$

with a constant $c_S > 0$ independent of ν.

The combination of gradient-related descent directions and efficient step sizes leads to a globally convergent method. Although this is a standard result that (with small modifications) can be found in many textbooks (see, e.g., [9, 21, 30]), we give a proof here since these variant do for example not include the monotone correction $\mathcal{C}$.

Theorem 4.1. *Assume that d and ρ generate gradient-related directions and efficient step sizes, respectively. Then the iterates w^ν generated by* (9), (10), *and* (11) *converge to w^* for an arbitrary initial iterate $w^0 \in \mathbb{R}^m$.*

Proof. See [12]. □

The proof is based on a finite dimensional compactness argument that allows to deduce convergence from the existence of a unique minimizer. Under the stronger assumption that h is strongly convex, i.e., if there is a $\mu > 0$ such that

$$h(\lambda x + (1-\lambda)y) \le \lambda h(x) + (1-\lambda)h(y) - \lambda(1-\lambda)\frac{\mu}{2}\|x-y\|_M^2 \qquad \forall \lambda \in [0,1],$$

we get R-linear convergence.

Theorem 4.2. *Assume that the conditions of Theorem 4.1 hold and that h is strongly convex with a constant $\mu > 0$. Then the iterates w^ν generated by* (9), (10), *and* (11) *satisfy the error estimate*

$$\|w^\nu - w^*\|_M^2 \le q^\nu \frac{2}{\mu}\left(h(w^0) - h(w^*)\right) \tag{15}$$

with $q = (1 - 2c_S c_D^2 \mu) < 1$.

Proof. See [12]. □

4.2. **Inexact Evaluation of Descent Directions.** We now consider inexact search directions $\tilde{d}^\nu$ obtained if the exact evaluation $d^\nu = d(\nu, w^\nu)$ is replaced by some approximation

$$\tilde{d}^\nu = \tilde{d}(\nu, w^\nu) \approx d(\nu, w^\nu). \tag{16}$$

Proposition 4.1. *Let d generate gradient-related directions that satisfy* (13) *with the constant $c_D > 0$, and let $\tilde{d}$ generate descent directions. Assume that there is a constant $c < c_D/2$ such that the approximations $\tilde{d}^\nu = \tilde{d}(\nu, w^\nu)$ satisfy at least one of the accuracy conditions*

$$\|d^\nu - \tilde{d}^\nu\|_M \le c\|\tilde{d}^\nu\|_M \quad \forall \nu \in \mathbb{N}, \tag{17}$$

$$\|d^\nu - \tilde{d}^\nu\|_M \le c\|d^\nu\|_M \quad \forall \nu \in \mathbb{N}, \tag{18}$$

for all sequences w^ν. Then the approximation $\tilde{d}$ does also generate gradient-related directions that satisfy (13) *with the constant $\tilde{c}_D = c_D - 2c > 0$.*

Proof. See [12]. □

Since the constant c_D needed to check the accuracy conditions in Proposition 4.1 with $c < c_D/2$ is in general not known, we replace them by the asymptotic criteria

$$\lim_{\nu\to\infty} \frac{\|d^\nu - \tilde{d}^\nu\|_M}{\|\tilde{d}^\nu\|_M} = 0 \qquad \text{and} \qquad \lim_{\nu\to\infty} \frac{\|d^\nu - \tilde{d}^\nu\|_M}{\|d^\nu\|_M} = 0, \tag{19}$$

respectively. They imply that the criteria in Proposition 4.1 with $c < c_D/2$ hold for sufficiently large ν with arbitrarily small c. To see that the whole sequence

$\tilde{d}^\nu$ is gradient-related assume that (17) or (18) is satisfied for $\nu > \nu_0$. Hence by Proposition 4.1 the estimate (13) holds for $\nu > \nu_0$ with $\tilde{c}_D$. Then it also holds for all ν with the constant

$$\tilde{\tilde{c}}_D = \min\left\{\tilde{c}_D, \min\left\{-\frac{\langle \nabla h(w^\nu), d^\nu\rangle}{\|\nabla h(w^\nu)\|_{M^{-1}}\,\|d^\nu\|_M} : \nu \le \nu_0\right\}\right\} > 0.$$

Furthermore, the constants $\tilde{c}_D, \tilde{\tilde{c}}_D$ in estimate (13) for $\tilde{d}^\nu$ tend to the constant c_D for the exact directions d^ν in this case.

4.3. **Step Size Rules.** There is a multitude of algorithms for the selection of efficient step sizes available from textbooks and surveys like [6, 20, 21, 24]. The most common choice is the step size rule by Armijo [1] (see also [6, 21]) which tracks the actual decrease of the functional h and leads to efficient step sizes. Another option are so-called approximate "exact step sizes". In contrast to the Armijo rule this approach does only depend on a single parameter. Furthermore it turned out to be more robust for the presented applications.

Proposition 4.2. *Let $d : \mathbb{N}\times\mathbb{R}^m \to \mathbb{R}^m$ generate descent directions. For a sequence $w^\nu \subset \mathbb{R}^m$ and directions $d^\nu = d(\nu, w^\nu)$ assume that a fixed parameter $\epsilon \in [0,1)$ is given. Then any step rule ρ that satisfies $\rho_\nu = \rho(\nu, w^\nu, d^\nu) \ge 0$ and*

$$\left\langle \nabla h(w^\nu + \rho_\nu d^\nu), d^\nu \right\rangle \in [\epsilon \langle \nabla h(w^\nu), d^\nu\rangle, 0]$$

generates efficient step sizes that satisfy (14) *with*

$$c_S = \frac{1-\epsilon^2}{2L}.$$

Proof. See [12]. □

Now we can obtain a sequence of efficient step sizes either by computing the first zero of $\rho \mapsto \left\langle \nabla h(w^\nu + \rho d^\nu), d^\nu \right\rangle$ exactly ($\epsilon = 0$) or by approximating it with fixed $0 \le \epsilon < 1$. The latter can be done for example using the bisection method which requires one evaluation of ∇h per bisection step.

Observe that for this rule and the Armijo rule a sequence of evaluations of either h or ∇h is required. In view of Theorem 3.1 this will involve one solution of the minimization problem associated with F^{-1} per evaluation of h or ∇h, which can be very expensive. In order to mitigate this disadvantage we will now present a method that allows to decide a priori if we can choose $\rho_\nu = 1$ for a given w^ν and ν or if some kind of line search is needed. While (13) does only give information about the angle between $\nabla h(w^\nu)$ and $d(\nu, w^\nu)$ we will need the stronger condition that $d(\nu, w^\nu) \to 0$ implies $\nabla h(w^\nu) \to 0$.

Let $\alpha_{-1} > 0$ and $\sigma \in (0,1)$ and define for $w^\nu, d^\nu \in \mathbb{R}^m$ the sequence

$$\alpha_\nu = \begin{cases} \|d^\nu\|_M & \text{if } \|d^\nu\|_M \le \sigma\alpha_{\nu-1}, \\ \alpha_{\nu-1} & \text{else.} \end{cases} \tag{20}$$

For a step size rule ρ that generates efficient step sizes we will switch off the step rule if the norm of the direction decreases by the factor σ in the following sense:

$$\tilde{\rho}_\nu = \begin{cases} 1 & \text{if } \|d^\nu\|_M \le \sigma\alpha_{\nu-1}, \\ \rho(\nu, w^\nu, d^\nu) & \text{else.} \end{cases} \tag{21}$$

Note that the sequence $\tilde{\rho}_\nu$ can easily be computed in practice. If $\|d^\nu\|_M \leq \sigma\alpha_{\nu-1}$ is not true, the step size $\tilde{\rho}_\nu$ is computed using the step size rule ρ. If the criterion is satisfied for some ν, the step size $\tilde{\rho}_{\nu'} = 1$ is used and the new bound $\alpha_\nu = \|d^\nu\|_M$ is computed. Thus the criterion for the ν-th step is checked with the bound $\alpha_{\nu-1} = \|d^{\nu'}\|_M$ where ν' is the last iteration step that satisfied the criterion.

It is also possible to simplify the criterion for the selection of $\tilde{\rho}_\nu = 1$ to the stronger criterion

$$\|d^\nu\|_M \leq \sigma \min_{\mu<\nu} \|d^\mu\|_M,$$

and the convergence proof of the following theorem remains essentially the same.

Theorem 4.3. *Assume that d and ρ generate gradient-related directions and efficient step sizes, respectively. Furthermore, assume that $d(\nu, v^\nu) \to 0$ implies $\nabla h(v^\nu) \to 0$ for any sequence v^ν. If $\tilde{\rho}_\nu$ is computed by (20) and (21) for some $\alpha_{-1} > 0$ and $\sigma \in (0,1)$ and $d^\nu = d(\nu, w^\nu)$, then the iterates w^ν obtained by*

$$w^{\nu+\frac{1}{2}} = w^\nu + \tilde{\rho}_\nu d^\nu,$$
$$w^{\nu+1} = w^{\nu+\frac{1}{2}} + \mathcal{C}(w^{\nu+\frac{1}{2}}),$$

converge to w^* for an arbitrary initial iterate $w^0 \in \mathbb{R}^m$.

In contrast to the previous results Theorem 4.3 is a not a standard result. However we skip the technical proof for ease of presentation here.

Proof. See [12].

□

We will see that an important example for directions satisfying the extra assumption of Theorem 4.3 is given by

$$d(\nu, w^\nu) = -S_\nu^{-1} \nabla h(w^\nu)$$

with symmetric positive definite matrices S_ν that are bounded uniformly from above and below with respect to ν. If such directions are evaluated inexactly one does in general not know a priori if the inexact directions satisfy

$$\tilde{d}(\nu, v^\nu) \to 0 \qquad \Rightarrow \qquad \nabla h(w^\nu) \to 0.$$

In this case the following generalization of Theorem 4.3 can be used.

Corollary 4.1. *Let d and ρ satisfy the assumptions of Theorem 4.3 and let $\tilde{d}$ satisfy the assumptions of Proposition 4.1 with the accuracy condition (17), i.e.,*

$$\|d^\nu - \tilde{d}^\nu\|_M \leq c\|\tilde{d}^\nu\|_M \quad \forall \nu \in \mathbb{N}.$$

If $\tilde{\rho}_\nu$ is computed by (20) and (21) for some $\alpha_{-1} > 0$ and some $\sigma \in (0,1)$ with d^ν replaced by $\tilde{d}^\nu = \tilde{d}(\nu, w^\nu)$, then the iterates w^ν obtained by

$$w^{\nu+\frac{1}{2}} = w^\nu + \tilde{\rho}_\nu \tilde{d}^\nu,$$
$$w^{\nu+1} = w^{\nu+\frac{1}{2}} + \mathcal{C}(w^{\nu+\frac{1}{2}}),$$

converge to w^* for an arbitrary initial iterate $w^0 \in \mathbb{R}^m$.

Proof. See [12]. □

Notice that the above result does no longer hold if $\tilde{d}$ does only satisfy the second accuracy condition (18) of Proposition 4.1.

5. Derivatives of the Nonlinear Schur Complement

The convergence speed of gradient-related descent algorithms depends heavily on the selection of the descent directions d^ν. If h is C^2 the directions

$$d^\nu = -(\nabla^2 h(w^\nu))^{-1} \nabla h(w^\nu) \tag{22}$$

lead to a damped Newton method for the operator $H = \nabla h$. If H is not differentiable but Lipschitz continuous we want to define directions similar to (22), replacing $(\nabla^2 h(w^\nu))$ by symmetric positive definite matrices $S(w^\nu) \in \mathbb{R}^{m,m}$ that represent generalized linearizations of H at w^ν.

In the following h and $H = \nabla h$ are given as in Theorem 3.1. For the special case of a quadratic obstacle problem with additional linear constraint such linearizations $S(w^\nu)$ were introduced by Gräser and Kornhuber [16]. There the piecewise linearity of H in that case was used. Here we will generalize this approach using piecewise smoothness of H instead.

5.1. Derivatives of F^{-1}.

Since H is a composition and sum of affine functions with F^{-1} the crucial part in the derivation of linearizations of H are linearizations of F^{-1}. In order to derive such linearizations for F^{-1} we first look at the functionals φ_i.

Lemma 5.1. *The limits*

$$\varphi'_{i,-}(x) := \lim_{\xi \nearrow x} \varphi'_i(\xi), \qquad \varphi''_{i,-}(x) := \lim_{\xi \nearrow x} \varphi''_i(\xi) \qquad \forall x \in (a_i^0, a_i^{m_i}],$$
$$\varphi'_{i,+}(x) := \lim_{\xi \searrow x} \varphi'_i(\xi), \qquad \varphi''_{i,+}(x) := \lim_{\xi \searrow x} \varphi''_i(\xi) \qquad \forall x \in [a_i^0, a_i^{m_i}).$$

exist in $\mathbb{R} \cup \{-\infty, \infty\}$.

Proof. The existence of the limits for $x \in (a_i^k, a_i^{k+1})$ and of the limits of φ''_i at the a_i^k is guaranteed by (A2). We only have to show the existence of the limits of φ'_i for $x = a_i^k$. First we note that φ'_i is monotone. Furthermore, it is bounded on each interval (a_i^{k-1}, a_i^k) with $k < m_i$ since φ_i cannot be convex on (a_i^{k-1}, a_i^{k+1}) otherwise. Thus $\lim_{\xi \nearrow a_i^k} \varphi'_i(\xi)$ exists and is finite for $k < m_i$ and either finite or ∞ for $k = m_i$. Limits from above can be shown analogously. □

For simplicity we will use the notation $\varphi''_i(x)$ also if the one sided directional derivatives $\varphi''_{i,-}(x)$ and $\varphi''_{i,+}(x)$ do not coincide. In this case $\varphi''_i(x)$ denotes the maximum of both.

In principle the linearization of F^{-1} will be defined piecewise and the components i where φ_i lacks regularity need special care. To do so we introduce the inactive sets

$$\mathcal{I}(v) := \{i : \partial\varphi_i(v_i) \text{ is single-valued}\}.$$

For convenience we also define the corresponding active sets

$$\mathcal{A}(v) := \{1, \dots, n\} \setminus \mathcal{I}(v).$$

It will be convenient to introduce the following notion of truncated matrices and vectors.

Definition 5.1. *Let $\mathcal{I}, \mathcal{J} \subset \mathbb{N}$ be index sets, $x \in \mathbb{R}^n$ a vector, and $M \in \mathbb{R}^{m,n}$ a matrix. Then define the truncated matrix $M_{\mathcal{I},\mathcal{J}} \in \mathbb{R}^{m,n}$ and the truncated vector $x_{\mathcal{I}} \in \mathbb{R}^n$ by*

$$(M_{\mathcal{I}\mathcal{J}})_{ij} := \begin{cases} M_{ij} & \text{if } i \in \mathcal{I} \text{ and } j \in \mathcal{J}, \\ 0 & \text{else}, \end{cases} \qquad (x_{\mathcal{I}})_i := \begin{cases} x_i & \text{if } i \in \mathcal{I}, \\ 0 & \text{else}. \end{cases}$$

Furthermore, define the abbreviation $M_{\mathcal{I}} := M_{\mathcal{I},\mathcal{I}}$.

In order to extract a decomposition of $\mathbb{R}^n$ into nontrivial subsets where F^{-1} is smooth we have to distinguish different active configurations. Since the functions φ_i may have multiple points a_i^k where they are not smooth, an active configuration is not completely determined by the active set itself. To distinguish different configurations we also have to take the values at the active component into account. The equivalence classes

$$[c] := \{v \in \operatorname{dom}\varphi : \mathcal{A}(v) = \mathcal{A}(c), v_i = c_i \, \forall i \in \mathcal{A}(c)\}$$

defined for $c \in \operatorname{dom}\varphi$ containing all vectors with the same active configuration provide exactly this distinction. Hence we can address an active configuration by $[c]$ for one representative. By definition x and y have the same active configuration if and only if $[x] = [y]$ and hence the representative is obviously not unique. The set of all possible active configurations is given by

$$\mathbb{A} := \{[c] : c \in \operatorname{dom}\varphi\}.$$

By Assumption (A2) the set $\mathbb{A}$ is finite and $\operatorname{dom}\varphi$ can be decomposed according to

$$\operatorname{dom}\varphi = \bigcup_{[c]\in\mathbb{A}} [c].$$

Since F has a single-valued inverse we have $\operatorname{dom}\varphi \supset F^{-1}(\mathbb{R}^n)$ and thus $\mathbb{R}^n = F(\operatorname{dom}\varphi)$ can be decomposed according to

$$\mathbb{R}^n = \bigcup_{[c]\in\mathbb{A}} F([c]), \qquad F([c]) = \bigcup_{x\in[c]} F(x) = \{y : F^{-1}(y) \in [c]\}$$

using the images of $[c]$ under F. Note that in general $\operatorname{dom}\varphi \supset F^{-1}(\mathbb{R}^n)$ is a real inclusion and equality does not hold. This is due to the possibility of $\varphi_i'(x_i) \to \infty$ for $x_i \to a_i^{m_i}$. In this case $F(x)$ is even empty for all $x \in \operatorname{dom}\varphi$ with $x_i = a_i^{m_i}$.

We will define the linearization of F^{-1} piecewise on sets where the operator is smooth. Thus we do not only need to handle the active components but also the "smoothness intervals" (a_i^{k-1}, a_i^k) the inactive components are contained in. To this end it is convenient to first define the set $\mathcal{E}$ of all multi-indices needed to identify these intervals by

$$\mathcal{E} := \{\eta \in \mathbb{N}^n : 1 \le \eta_i \le m_i\}.$$

Now we can define the sets of all vectors corresponding to an active configuration $[c] \in \mathbb{A}$ and the open and closed smoothness intervals $\eta \in \mathcal{E}$ by

$$[c]_\eta := \{x \in [c] : x_i \in (a_i^{\eta_i - 1}, a_i^{\eta_i}) \text{ for } i \in \mathcal{I}(c)\},$$
$$[[c]]_\eta := \{x \in [c] : x_i \in [a_i^{\eta_i - 1}, a_i^{\eta_i}] \text{ for } i \in \mathcal{I}(c)\}.$$

Both sets are $(n - |\mathcal{A}(c)|)$-dimensional hypercubes since all active components $i \in \mathcal{A}(c)$ are fixed, while the others can take values in a nontrivial interval. The set

$[[c]]_\eta$ is in general only a subset of $\overline{[c]}_\eta$ since $\operatorname{dom}\varphi_i$ may not contain a_i^0 and $a_i^{m_i}$. These sets provide a decomposition of $[c]$ in the sense that

$$[c] = \bigcup_{\eta\in\mathcal{E}} [[c]]_\eta, \qquad\qquad \overline{[c]} = \bigcup_{\eta\in\mathcal{E}} \overline{[c]}_\eta.$$

Note that these decompositions are not disjoint in general.

If a linearization of F^{-1} is to be defined piecewise it is important that the sets where it is defined do not degenerate to lower-dimensional objects or, equivalently, that active configurations are stable in a certain sense. This is provided by the following lemma that guarantees that each set $F([[c]]_\eta)$ is contained in the closure of an open set.

Lemma 5.2. *Let $[c] \in \mathbb{A}$ with $F(c) \neq \emptyset$ and $\eta \in \mathcal{E}$. Then*

$$F([[c]]_\eta) \subset \overline{F([c]_\eta)^\circ} \subset \overline{\operatorname{int} F([[c]]_\eta)} \neq \emptyset$$

holds for the open set

$$F([c]_\eta)^\circ := \left\{ y \in F([c]_\eta) : \left\{ \begin{array}{ll} y \in (\nabla J_0(F^{-1}(y)))_i + \operatorname{int}\partial\varphi_i(F^{-1}(y)_i) & \forall i \in \mathcal{A}(c), \\ F^{-1}(y)_i \in (a_i^{\eta_i-1}, a_i^{\eta_i}) & \forall i \in \mathcal{I}(c) \end{array} \right. \right\}.$$

Proof. Let $[c] \in \mathbb{A}$ with $F(c) \neq \emptyset$ and $\eta \in \mathcal{E}$. Since $\partial\varphi_i(c_i)$ is set-valued for $i \in \mathcal{A}(c)$, an element of $F([c]_\eta)^\circ$ can easily be constructed, which shows that $F([c]_\eta)^\circ \neq \emptyset$. Next we show that $F([c]_\eta)^\circ$ is open.

Let $y \in F([c]_\eta)^\circ$, $x = F^{-1}(y)$ be fixed and $x' = F^{-1}(y')$ for some y' with $\|y - y'\|_\infty < \epsilon$. By (A2) and continuity of F^{-1} we instantly get

$$x'_i \in (a_i^{\eta_i-1}, a_i^{\eta_i})$$

and thus $\mathcal{A}(x') \subset \mathcal{A}(x)$ if ϵ is small enough.

To show $\mathcal{A}(x') \supset \mathcal{A}(x)$ assume that $x'_{\mathcal{I}(x)}$ is known and fixed. Then $x'_{\mathcal{A}(x)}$ is the unique solution of

$$F(x'_{\mathcal{I}(x)} + x'_{\mathcal{A}(x)})_i \ni y'_i \qquad \forall i \in \mathcal{A}(x). \tag{23}$$

By continuity of F^{-1} and ∇J_0 the residual defined by $r(b, v) := b - \nabla J_0(v)$ satisfies

$$\Big\| r(y, x) - r(y', x'_{\mathcal{I}(x)} + x_{\mathcal{A}(x)}) \Big\|_\infty < \max\Big\{ \operatorname{dist}\big(\partial P_{x,i}, r(y, x)_i\big) : i \in \mathcal{A}(x) \Big\}$$

for the border $\partial P_{x,i}$ of the set $P_{x,i} = \partial\varphi_i(x_i)$ if ϵ is small enough. In this case we have $r(y', x'_{\mathcal{I}(x)} + x_{\mathcal{A}(x)}) \in \operatorname{int}\partial\varphi_i(x_i)$. Hence $x'_{\mathcal{A}(x)} = x_{\mathcal{A}(x)}$ solves (23) which yields $\mathcal{A}(x') = \mathcal{A}(x)$. Thus $x' \in [c]_\eta$ and even more $y' \in F([c]_\eta)^\circ$. Since y was arbitrary, $F([c]_\eta)^\circ$ must be open and we have $F([c]_\eta)^\circ \subset \operatorname{int} F([[c]]_\eta)$.

Now let $y \in F([[c]]_\eta) \setminus F([c]_\eta^\circ$ with $x = F^{-1}(y)$ be fixed. Then it is easy to give a sequence $x^k \in [c]_\eta$ with $x^k \to x$. For the sequence

$$y^k = \nabla J_0(x^k) + (y - \nabla J_0(x) + z^k)_{\mathcal{A}(c)} + (\partial\varphi(x^k))_{\mathcal{I}(c)}$$

with

$$z_i^k = \frac{\epsilon}{k} \begin{cases} 1 & \text{if } i \in \mathcal{A}(c) \text{ and } (y - \nabla J_0(x))_i = \min\partial\varphi_i(c_i), \\ -1 & \text{if } i \in \mathcal{A}(c) \text{ and } (y - \nabla J_0(x))_i = \max\partial\varphi_i(c_i), \\ 0 & \text{else} \end{cases}$$

and ϵ small enough we have

$$(y^k - \nabla J_0(x^k))_i = \begin{cases} (y - \nabla J_0(x) + z^k)_i \in \operatorname{int} \partial\varphi_i(c_i) & \text{if } i \in \mathcal{A}(c), \\ \partial\varphi_i(x^k) & \text{if } i \in \mathcal{I}(c) \end{cases}$$

and hence $x^k = F^{-1}(y^k)$ and $y^k \in F([c])^\circ_\eta$. Since ∇J_0 is continuous and φ_i is continuously differentiable on $(a_i^{\eta_i - 1}, a_i^{\eta_i + 1})$ for $i \in \mathcal{I}(c)$ we have

$$y^k \to \nabla J_0(x) + (y - \nabla J_0(x))_{\mathcal{A}(c)} + (\partial\varphi(x))_{\mathcal{I}(c)} = y,$$

which proves the assertion. □

Since $[c]$ decomposes into the sets $[[c]]_\eta$, Lemma 5.2 implies

$$F([c]) \subset \overline{\operatorname{int} F([c])}.$$

While this lemma shows that the sets $F([c])$ do not degenerate it does not give insight into their structure. The following remark sheds some light on the geometry of these sets.

Remark 5.1. *Define the points where φ_i is not differentiable by*

$$\{\tilde{a}_i^0, \dots, \tilde{a}_i^{\widetilde{m}_i}\} := \{a : \partial\varphi_i(a) \text{ is set-valued}\} \subset \{a_i^0, \dots, a_i^{m_i}\}.$$

Then the configuration $[c']$ without active component, i.e. $\mathcal{A}(c') = \emptyset$, is clearly given by the open set

$$[c'] = \prod_{i=1}^n \bigcup_{k=1}^{\widetilde{m}_i} (\tilde{a}_i^{k-1}, \tilde{a}_i^k) = \bigcup_{\substack{(k_1,\dots,k_n) \\ 1 \le k_i \le \widetilde{m}_i}} \prod_{i=1}^n (\tilde{a}_i^{k_i - 1}, \tilde{a}_i^{k_i})$$

and a representative is, e.g., given by $c'_i = \frac{1}{2}(\tilde{a}_i^0 + \tilde{a}_i^1)$. Note that $[c']$ is the union of n-dimensional open hypercubes $Q_{(k_1,\dots,k_n)}$. If the arguments of Lemma 5.2 are applied with the indicator functions of these hypercubes instead of φ it can be seen that $\nabla J_0(Q_{(k_1,\dots,k_n)}) \subset \overline{\operatorname{int} \nabla J_0(Q_{(k_1,\dots,k_n)})}$. Hence the images of the hypercubes under ∇J_0 do not degenerate in the sense that all points are limits of sequences in their interior.

If at least one component of c is active the set $[c]$ is the union of hypercubes Q with dimension less then n, and hence no longer open in $\mathbb{R}^n$. To be precise the length of these hypercubes in any direction e_i with $i \in \mathcal{A}(c)$ is zero. However, the set $(\partial\varphi([c]))_{\mathcal{A}(c)}$ is a hypercube that has nonzero lengths exactly in the directions e_i with $i \in \mathcal{A}(c)$.

Figure 1 and Figure 2 show an example of the decomposition of $\operatorname{dom}\varphi$ *and $\mathbb{R}^2$ into the sets $[c] \in \mathbb{A}$ and $F([c])$, respectively. For simplicity it is assumed that all a_i^k differ from 0, such that $c_i = 0$ means that the i-th component is not active. While the sets $[c]$ are 1-dimensional edges or 0-dimensional vertices if one or two components of c are active, the corresponding images $F([c])$ of all such sets have a nontrivial interior. Note that $F([c])$ has a curved boundary in general but edges parallel to the i-th axis if $c_i = \tilde{a}_i^k$ for some k. For example the set $F([(0,0)])$ of all $F(x)$ such that φ_i is smooth at x_i for all i might have all edges curved. Conversely, the set $F([(a_1^1, 0)])$ of all $F(x)$ such that the first component is fixed to the kink a_1^1 (and thus active) has two straight edges parallel to the first axis. In case of a quadratic function J_0 all $F([c])$, $[c] \in \mathbb{A}$, are (possibly unbounded) parallelepipeds.*

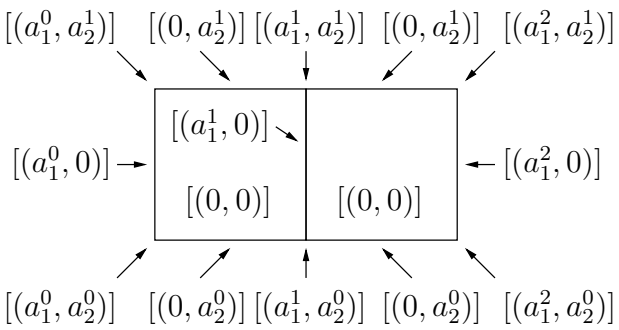

FIGURE 1. Decomposition of $\operatorname{dom}\varphi \subset \mathbb{R}^2$ into the sets $[c]$, $c \in \mathbb{A} \subset \mathbb{R}^2$.

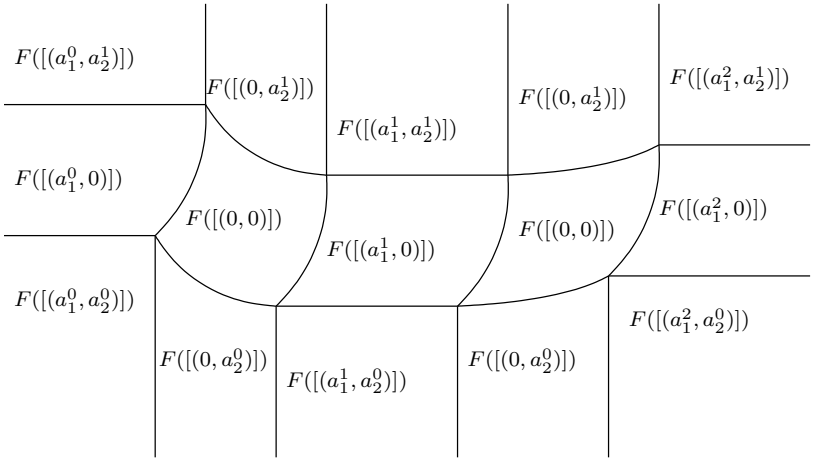

FIGURE 2. Decomposition of $\mathbb{R}^2$ into the sets $F([c])$, $c \in \mathbb{A} \subset \mathbb{R}^2$.

The essence of Lemma 5.2 is that the subsets in the following decomposition

$$\mathbb{R}^n = \bigcup_{[c]\in\mathbb{A}} F([c]) = \bigcup_{[c]\in\mathbb{A}} \bigcup_{\eta\in\mathcal{E}} F([[c]]_\eta)$$

of $\mathbb{R}^n$ are non-degenerate. Since F^{-1} is smooth on each of these subsets $F([[c]]_\eta)$ this allows us to define a piecewise linearization on these sets. We will only state the main result here. For the technical details we refer to [12].

Theorem 5.1. *Let J_0 be twice continuously differentiable. Then an element of the generalized derivative in the sense of Clarke at $y = F(x)$ is given by the Moore–Penrose pseudoinverse $(\partial^2 J(x)_{\mathcal{I}'(x)})^+$ of the reduced Hessian*

$$\partial^2 J(x)_{\mathcal{I}'x} := \Big(\nabla^2 J_0(x) + \varphi''(x)\Big)_{\mathcal{I}'(x)} \tag{24}$$

with the reduced inactive set

$$\mathcal{I}'(v) := \left\{ i \in \mathcal{I}(v) : \max\{\varphi''_{i,-}(v_i), \varphi''_{i,+}(v_i)\} < \infty \right\}.$$

I.e., we have

$$\left(\partial^2 J(x)_{\mathcal{I}'(x)}\right)^+ \in \partial_B(F^{-1})(y) \subset \partial_C(F^{-1})(y). \tag{25}$$

where ∂_B and ∂_C denote the B-subdifferential (cf. [25, 31]) and the generalized Jacobian in the sense of Clarke (cf. [4]), respectively.

Furthermore, F^{-1} is differentiable on each set $F([c]_\eta)^\circ$ with $[c] \in \mathbb{A}$ and $\eta \in \mathcal{E}$ and the derivative is given by the matrix $(\partial^2 J(x)_{\mathcal{I}'(x)})^+$.

Proof. See [12]. □

Theorem 5.2. *Let J_0 be twice continuously differentiable and let all φ_i'' be uniformly bounded from above by a constant $c_{\varphi''}$ where $\partial\varphi_i$ is single-valued. Then there is a constant $c > 0$ such that F^{-1} is strongly monotone with respect to the semi-norm introduced by $cI_{\mathcal{I}_0}$, i.e.*

$$\left\langle F^{-1}(u) - F^{-1}(v), u - v \right\rangle \geq c \left\langle u - v, u - v \right\rangle_{\mathcal{I}_0} \qquad \forall u, v \in \mathbb{R}^n, \tag{26}$$

where $\mathcal{I}_0$ is the smallest inactive set, i.e.

$$\mathcal{I}_0 := \bigcap_{y \in \mathbb{R}^n} \mathcal{I}'(F^{-1}(y)) = \mathbb{N} \setminus \{i \in \mathbb{N} : \exists \xi \in \mathbb{R} : \partial\varphi_i(\xi) \text{ is set-valued}\}.$$

Proof. See [12]. □

By Rademacher's theorem (see, e.g., [19]) ∇J_0 is in general only differentiable on a set $\mathcal{D}_{\nabla J_0}$ such that $\mathbb{R}^n \setminus \mathcal{D}_{\nabla J_0}$ has measure zero under the assumptions (A1)–(A4). If this is the case we make the following additional assumption.

(A5) For all $x \in \mathbb{R}^n$ the matrix $\partial^2 J_0(x) \in \mathbb{R}^{n,n}$ is symmetric and positive definite. It coincides with the classical Hessian $\nabla^2 J_0$ if $x \in \mathcal{D}_{\nabla J_0}$.

Normally one would chose $\partial^2 J_0$ to be some generalized linearization of ∇J_0. If ∇J_0 is not differentiable everywhere we can still define

$$\left(\partial^2 J(x)_{\mathcal{I}'(x)}\right)^+ = \left(\partial^2 J_0(x) + \varphi''(x)\right)^+_{\mathcal{I}'(x)}$$

and use this as generalized linearization of ∇J_0 at $x = F^{-1}(y)$, analogously to (24).

5.2. **Derivatives of H.** If F^{-1} is a continuously differentiable operator we can easily derive a linearization of the nonlinear Schur complement

$$H(w) = -BF^{-1}(f - B^T w) + Cw + g$$

using the chain rule. The result is

$$\nabla H(w) = B\nabla(F^{-1})(f - B^T w)B^T + C.$$

If F itself is also differentiable we have $\nabla(F^{-1})(y) = (\nabla F)(F^{-1}(y))^{-1}$ and $\nabla H(w)$ as given above is just the Schur complement of the linear saddle point problem

$$u \in \mathbb{R}^n, w \in \mathbb{R}^m : \qquad \begin{pmatrix} (\nabla F)(F^{-1}(f - B^T w_0)) & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ w \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix},$$

which is the linearization of the nonlinear saddle point problem (1) at $(u_0, w_0)^T$ with $u_0 = F^{-1}(f - B^T(w_0))$.

In the general case these derivatives do not exist. While F is not even a single-valued operator we know from Propositions 2.1 and 3.1 that F^{-1} and H are Lipschitz continuous. Thus one could in principle select elements of the generalized Jacobian in the sense of Clarke [4]

$$\partial_C H(w) = \operatorname{co} \partial_B H(w).$$

However, it will be complicated to compute elements of this set since the generalized Jacobian ∂_C does not satisfy the chain rule in general. Nevertheless we use a chain rule to obtain a generalized linearization $S(w)$ of H at w which is not necessarily an element of $\partial_C H(w)$. Based on the linearization of F^{-1} derived in the previous subsection this approach results in

$$S(w) := B\Big(\partial^2 J(u)_{\mathcal{I}'(u)}\Big)^+ B^T + C$$

as linearization of H at w with $u = F^{-1}(f - B^T w)$.

Proposition 5.1. *Let J_0 be twice continuously differentiable and let* $\operatorname{rank} B = n$. *Then*

$$S(w) \in \partial_B H(w) \subset \partial_C H(w) \qquad \forall w \in \mathbb{R}^m.$$

Proof. If $\operatorname{rank} B = n$ the mapping defined by $G(w) = f - B^T w$ is surjective. In the proof of Theorem 5.1 the generalized derivative of F^{-1} was derived as a limit of classical derivatives that are defined on disjoint open sets $F([c]_\eta)^\circ$ where F^{-1} is differentiable. Furthermore, the space $\mathbb{R}^n$ can be decomposed according to

$$\mathbb{R}^n = \bigcup_{[c]\in\mathbb{A},\eta\in\mathcal{E}} \overline{F([c]_\eta)^\circ} = \overline{\bigcup_{[c]\in\mathbb{A},\eta\in\mathcal{E}} F([c]_\eta)^\circ}.$$

Since F^{-1} is differentiable on $F([c]_\eta)^\circ$ this is also true for $F^{-1} \circ G$ and H on $G^{-1}(F([c]_\eta)^\circ)$. By the classical chain rule we have $\nabla H(w) = S(w)$ at $w \in G^{-1}(F([c]_\eta)^\circ)$. Now let

$$w \in R := \overline{\bigcup_{[c]\in\mathbb{A},\eta\in\mathcal{E}} G^{-1}(F([c]_\eta)^\circ)}.$$

Having only a finite number of sets $F([c]_\eta)^\circ$ we can, without loss of generality, assume that there is a sequence $w^k \to w$ with $w^k \in G^{-1}(F([c]_\eta)^\circ)$ for a single fixed set $F([c]_\eta)^\circ$. Then we have $S(w) = \lim_{k\to\infty} S(w^k) \in \partial_B H(w)$.

To complete the proof we assume that there is a $w \in \mathbb{R}^m \setminus R$. Then there is an open ball $B_\epsilon(w)$ such that $B_\epsilon(w) \cap G^{-1}(F([c]_\eta)^\circ) = \emptyset$ for all c, η. By the open mapping theorem (see, e.g., [33]) $G(B_\epsilon(w))$ is also open. Thus it must intersect at least one $F([c]_\eta)^\circ$ which contradicts the assumption and shows that $\mathbb{R}^m = R$. □

Remark 5.2. *While Proposition 5.1 seems to give a reasonable characterization of $S(w)$, the assumption* $\operatorname{rank} B = n$ *is quite restrictive for the following reason. If the saddle point problem arises from a minimization problem with linear constraints we have $C = 0$ in general, and a well posed problem will have $m \le n$ linear constraints only. Combined with* $\operatorname{rank} B = n$ *this results in B to be a regular square matrix and hence the solution $u = B^{-1}g$ is completely determined by the linear constraint.*

The following example shows that the assertion of Proposition 5.1 is in general not valid if $\operatorname{rank} B < n$.

Example 5.1. *For* $K = \{x \in \mathbb{R}^2 : x_i \geq 0, i = 1, 2\}$ *consider the saddle point problem*

$$\begin{pmatrix} F & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ w \end{pmatrix} = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \partial\chi_K \quad \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right) \begin{pmatrix} u_1 \\ u_2 \\ w \end{pmatrix} \ni \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}.$$

Then we have $(F^{-1}(y))_i = \max\{0, y_i\}$ *and thus the nonlinear Schur complement is*

$$H(w) = \max\{0, w\} - \max\{0, -w\} + w = 2w.$$

Hence $\nabla H(w) = 2$ *and* $\partial_C H(w) = \partial_B H(w) = \{2\}$. *On the other hand we have* $S(w) = 2$ *for* $w \neq 0$ *but* $S(0) = 1$.

This problem occurs since the line $f - B^T w$ *crosses three domains where* F^{-1} *is smooth. While these domains have a nonempty interior themselves the line intersects the one leading to* $F^{-1}(y) = 0$ *only at the single point* $y = 0$ *of its border. Thus the preimage of this domain under* $f - B^T(\cdot)$ *collapses to the single point* $w = 0$.

6. Schur Nonsmooth Newton Methods

We now consider the algorithms obtained if a linearization of the nonlinear Schur complement is used as preconditioner for search directions

$$d^\nu = -S(w^\nu)^{-1} \nabla h(w^\nu) \tag{27}$$

in the general descent algorithms given by (8), or (10) and (11). Convergence will follow from the convergence results for gradient-related descent methods.

6.1. Algorithms and Convergence.

Before showing convergence of the algorithms we consider the solvability of the system

$$S(w^\nu) d^\nu = -\nabla h(w^\nu).$$

We immediately get

$$\langle S(w)x, y\rangle = \left\langle \left(\partial^2 J(u)_{\mathcal{I}'(u)}\right)^+ B^T x, B^T y \right\rangle + \langle Cx, y\rangle, \qquad x, y \in \mathbb{R}^m,$$

for $u = F^{-1}(f - B^T w)$. Hence by (A1)–(A3) and (A5) the matrix $S(w)$ is symmetric and positive semidefinite. However we have no guarantee that it is invertible.

Even if this matrix is invertible it will often not be possible to solve the above system directly, and the application of iterative schemes does in general involve multiplications by $\left[\partial^2 J_0(u) + \varphi''(u)\right]^+_{\mathcal{I}'(u)}$. While this is in principle possible the possibly large derivatives φ_i'' might prevent convergence. In order to overcome this problem we reduce the inactive set further to

$$\mathcal{I}''(v) := \{i \in \mathcal{I}'(v) : \varphi_i''(v_i) < (C_\varphi)_{i,i}\}$$

for a positive definite diagonal matrix $C_\varphi \in \mathbb{R}^{m,m}$. The induced truncated linearization of H at w with $u = F^{-1}(f - B^T w)$ is given by

$$S'(w) := B\left(\partial^2 J(u)_{\mathcal{I}''(u)}\right)^+ B^T + C.$$

Note that replacing $\mathcal{I}'(v)$ by $\mathcal{I}''(v)$ does essentially mean to set very small derivatives of $F^{-1}$to zero. This additional truncation of the matrix ensures that the diagonal elements of $\partial^2 J(u)_{\mathcal{I}''(u)}$ remain bounded independently of $\varphi''(u)$.

Independently of this truncation the matrix $S'(w)$ may not be invertible. In the most extreme case $S'(w) = 0$ if all components are active while the system results from a constraint minimization problem, i.e. $\mathcal{I}''(u) = \{1, \dots, n\}$ and $C = 0$. Although this does not happen in many application problems, it is not uncommon that $S'(w)$ has a nontrivial kernel.

Since the kernel of $\left[\partial^2 J_0(u) + \varphi''(u)\right]_{\mathcal{I}''(u)}$ and thus the kernel of $S'(w)$ with $u = F^{-1}(f - B^T w)$ depends only on $\mathcal{I}''(u)$, the same is true for the orthogonal projection $P_{\ker(S'(w))} : \mathbb{R}^m \to \ker(S'(w))$. Hence for a fixed symmetric positive definite matrix $\tilde{C}$ we can define the symmetric positive semidefinite matrix

$$\tilde{C}(\mathcal{I}''(u)) := P_{\ker(S'(w))}^T \tilde{C} P_{\ker(S'(w))} \in \mathbb{R}^{m,m},$$

and introduce the regularized linearization of H given by

$$S''(w) := S'(w) + \tilde{C}(\mathcal{I}''(u)).$$

If $v_{w,1}, \dots, v_{w,l}$ is an orthonormal basis of $\ker(S'(w))$ then it is easy to see that $P_{\ker(S'(w))}$ and $\tilde{C}(\mathcal{I}''(u))$ are given by

$$P_{\ker(S'(w))} = \sum_{i=1}^{l} v_{w,i} v_{w,i}^T, \qquad \tilde{C}(\mathcal{I}''(u)) = \sum_{i,j=1}^{l} \langle v_{w,i}, v_{w,j} \rangle_{\tilde{C}} \, v_{w,i} v_{w,j}^T.$$

Lemma 6.1. *$S''(w)$ is symmetric and positive definite for all $w \in \mathbb{R}^m$.*

Proof. Let $x_1, x_2 \in \mathbb{R}^m$ and $x_i^\bullet = P_{\ker(S'(w))} x_i$, $x_i^\circ = x_i - x_i^\bullet$. Then symmetry and definiteness follow from

$$\langle S''(w) x_1, x_2 \rangle = \langle S'(w) x_1^\circ, x_2^\circ \rangle + \left\langle \tilde{C} x_1^\bullet, x_2^\bullet \right\rangle.$$

□

Theorem 6.1. *The directions generated by $d(\nu, w) = -S''(w)^{-1} \nabla h(w)$ are gradient-related and guarantee $\nabla h(v^\nu) \to 0$ for any sequence $v^\nu \in \mathbb{R}^m$ with $d(\nu, v^\nu) \to 0$.*

Proof. The equivalence $d(\nu, w) = 0 \Leftrightarrow \nabla h(w) = 0$ in (12) follows from the fact that each $S''(w)$ is regular. To prove the estimate (13) let $w \in \mathbb{R}^m$ and define the reduced space

$$V_{\mathcal{I}} := \operatorname{span}\{e_i : i \in \mathcal{I}\} = \{v \in \mathbb{R}^n : v = v_{\mathcal{I}}\} \tag{28}$$

for any index set $\mathcal{I}$. For $u = F^{-1}(f - B^T w)$ we then have

$$\begin{aligned} \left\langle \partial^2 J(u)_{\mathcal{I}''(u)} v, v \right\rangle &\le \left\langle \overline{H}_{J_0} v, v \right\rangle + \langle C_\varphi v, v \rangle \\ &\le \lambda_{\max}(\overline{H}_{J_0} + C_\varphi) \langle v, v \rangle \qquad \forall v \in V_{\mathcal{I}''(u)}, \end{aligned}$$

and

$$\begin{aligned} \left\langle \partial^2 J(u)_{\mathcal{I}''(u)} v, v \right\rangle &\ge \left\langle \underline{H}_{J_0} v, v \right\rangle \\ &\ge \lambda_{\min}(\underline{H}_{J_0}) \langle v, v \rangle \qquad \forall v \in V_{\mathcal{I}''(u)}. \end{aligned}$$

Since the eigenvalues of $\partial^2 J(u) = \partial^2 J_0(u) + \varphi''(u)$ restricted to the indices in $\mathcal{I}''(u)$ are bounded, the same is true for the restricted inverse. Thus the following estimate

holds for all $v \in \mathbb{R}^n$

$$\begin{aligned}\lambda_{\max}(\overline{H}_{J_0} + C_\varphi)^{-1} \left\langle I_{\mathcal{I}''(u)} v, v \right\rangle &\leq \left\langle \left(\partial^2 J(u)_{\mathcal{I}''(u)}\right)^+ v, v \right\rangle \\ &\leq \lambda_{\min}(\underline{H}_{J_0})^{-1} \left\langle I_{\mathcal{I}''(u)} v, v \right\rangle \\ &\leq \lambda_{\min}(\underline{H}_{J_0})^{-1} \left\langle v, v \right\rangle .\end{aligned}$$

Using these estimates for $S''(w)$ we get for $v \in \mathbb{R}^m$

$$\begin{aligned}\min\left\{\frac{1}{\lambda_{\max}(\overline{H}_{J_0} + C_\varphi)}, 1\right\} \left\langle \left(B I_{\mathcal{I}''(u)} B^T + C + \tilde{C}(\mathcal{I}''(u))\right) v, v \right\rangle \\ \leq \langle S''(w) v, v \rangle \leq \max\left\{\frac{1}{\lambda_{\min}(\underline{H}_{J_0})}, 1\right\} \left\langle \left(B B^T + C + \tilde{C}(\mathcal{I}''(u))\right) v, v \right\rangle .\end{aligned}$$

Recalling that

$$\ker(I_{\mathcal{I}''(u)}) = \ker\left(\partial^2 J(u)_{\mathcal{I}''(u)}\right)^+$$

it is clear that the matrix on the left of the inequality is regular. Hence the matrices $S''(w)$ are bounded

$$\gamma_{\mathcal{I}''(u)} \left\langle v, v \right\rangle \leq \left\langle S''(w) v, v \right\rangle \leq \Gamma_{\mathcal{I}''(u)} \left\langle v, v \right\rangle$$

with constants $\gamma_{\mathcal{I}''(u)}, \Gamma_{\mathcal{I}''(u)} > 0$ depending only on the inactive set $\mathcal{I}''(u)$. Using this we get

$$\left\langle y, S''(w)^{-1} y \right\rangle \geq \gamma_{\mathcal{I}''(u)} \|S''(w)^{-1} y\|^2 \geq \frac{\gamma_{\mathcal{I}''(u)}}{\Gamma_{\mathcal{I}''(u)}} \|S''(w)^{-1} y\| \|v\|,$$

and thus (13) with

$$c_D = \min_{\mathcal{J} \subset \{1,\ldots,n\}} \frac{\gamma_{\mathcal{J}}}{\Gamma_{\mathcal{J}}}.$$

Finally we note that we get $\nabla h(v^\nu) \to 0$ from

$$\|\nabla h(v^\nu)\| \leq \|S''(v^\nu)\| \, \|d(\nu, v^\nu)\| \leq \max_{\mathcal{J} \subset \{1,\ldots,n\}} \Gamma_{\mathcal{J}} \, \|d(\nu, v^\nu)\|$$

for any sequence v^ν with $d(\nu, v^\nu) \to 0$. □

While this proof allows to apply the generic convergence results to the descent method obtained using the directions $d^\nu = -S''(w^\nu)^{-1} \nabla h(w^\nu)$ for the whole problem class, it is suboptimal in the following sense:

Since all estimates are derived for the Euclidean norm, the constant c_D incorporates the condition number of $\nabla^2 J_0(u)$, which may be large for discretized partial differential equations. For special cases it may be possible to derive much better estimates if a suitable norm for w is used. However, such improvements would only be visible in the convergence result of Theorem 4.2, since the more general result in Theorem 4.1 uses a compactness argument that does not give bounds.

Corollary 6.1. *Let $w^0 \in \mathbb{R}^m$. Then the sequence w^ν defined by*

$$\begin{aligned} d^\nu &= -S''(w^\nu)^{-1} \nabla h(w^\nu), \\ w^{\nu+\frac{1}{2}} &= w^\nu + \rho^\nu d^\nu, \\ w^{\nu+1} &= w^{\nu+\frac{1}{2}} + \mathcal{C}(w^{\nu+\frac{1}{2}}) \end{aligned}$$

converges to the solution w^* of (7) if the step size rule $\rho^\nu = \rho(\nu, w^\nu, d^\nu)$ generates efficient step sizes.

The same is true if d^ν is replaced by descent directions $\tilde{d}^\nu$ such that $\|d^\nu - \tilde{d}^\nu\|$ satisfies the accuracy condition (17) of Proposition 4.1 and if ρ^ν is replaced by $\tilde{\rho}^\nu$ in the sense of Theorem 4.3.

Proof. From Theorem 6.1 and Proposition 4.1 it follows that we have gradient-related descent directions. Thus we can apply Theorem 4.1 if $\rho^\nu = \rho(\nu, w^\nu, d^\nu)$ is chosen. If ρ^ν is replaced by $\tilde{\rho}^\nu$ in the sense of Theorem 4.3 we only have to note that $d(\nu, v^\nu) \to 0$ implies $\nabla h(v^\nu) \to 0$, by Theorem 6.1. □

The algorithm in Corollary 6.1 is essentially an inexact damped Newton-type method for the operator $H = \nabla h$. For $\mathcal{C} = 0$ it takes the form

$$w^{\nu+1} = w^\nu - \rho^\nu S''(w^\nu)^{-1} H(w^\nu)$$

with $\rho^\nu = \rho(\nu, w^\nu, -S''(w^\nu)^{-1}\nabla h(w^\nu))$. Since $S''(w)$ plays the role of a generalized linearization of the nonsmooth nonlinear Schur complement H at w we call this a "Schur Nonsmooth Newton method".

Lemma 6.2. *Let J_0 be twice continuously differentiable and let all φ_i'' be uniformly bounded from above by a constant $c_{\varphi''}$, whenever $\partial\varphi_i$ is single-valued. Then h is strongly convex if $S(w)$ is symmetric positive definite for all $w \in \mathbb{R}^m$.*

Proof. Let $w_1, w_2 \in \mathbb{R}^n$ and $x_i = f - B^T w_i$. By Theorem 5.2 we have for some $c > 0$

$$\begin{aligned}
&\langle H(w_1) - H(w_2), w_1 - w_2\rangle \\
&\qquad = \left\langle F^{-1}(x_1) - F^{-1}(x_2), x_1 - x_2\right\rangle + \langle w_1 - w_2, w_1 - w_2\rangle_C \\
&\qquad \geq c\,\langle x_1 - x_2, x_1 - x_2\rangle_{\mathcal{I}_0} + \langle w_1 - w_2, w_1 - w_2\rangle_C \\
&\qquad = \langle w_1 - w_2, w_1 - w_2\rangle_{BcI_{\mathcal{I}_0}B^T + C}\,.
\end{aligned}$$

Now let $x \in \operatorname{dom}\varphi$ such that $\mathcal{I}(x) = \mathcal{I}_0$ and $y \in F(x)$. Then the kernels of $(\partial^2 J(u)_{\mathcal{I}_0})^+$ and $cI_{\mathcal{I}_0}$ coincide. Thus the reduced Schur complement

$$BcI_{\mathcal{I}_0}B^T + C$$

must also be positive definite because $S(w)$ is. Hence $\nabla h = H$ is strongly monotone and h is strongly convex. □

Corollary 6.2. *Let J_0, φ, $S(w)$, and ρ satisfy the assumptions of Lemma 6.2 and Corollary 6.1, and let $(C_\varphi)_i > c_{\varphi''}$. Then $S(w) = S'(w) = S''(w)$ holds true and the method in Corollary 6.1 converges R-linearly.*

The same is true if d^ν is replaced by descent directions $\tilde{d}^\nu$ such that $\|d^\nu - \tilde{d}^\nu\|$ satisfies the accuracy condition (17) of Proposition 4.1.

Proof. Combine Theorem 6.1, Lemma 6.2 and Theorem 4.2. □

In general one would expect local superlinear convergence of a Newton-type method. Unfortunately our preconditioners $S''(w)$ are in general not contained in $\partial_C H(w)$ for the following reasons:

- As shown by Example 5.1 we may have $S'(w) \notin \partial_C H(w)$ if $\operatorname{rank} B \neq n$ due to the lack of a chain rule.

- If ∇J_0 is not differentiable it may not be possible to choose $\partial^2 J_0(w) \in \partial_C(\nabla J_0(w))$. Even if this is possible the lack of a chain rule may lead to $S'(w) \notin \partial_C H(w)$.
- In case of unbounded second derivatives of φ additional truncation is introduced.
- $S'(w)$ may not be invertible and thus needs to be regularized.

In all of the above cases the classical convergence analysis of semismooth Newton methods as introduced by Kummer [18], Pang [23], Qi and Sun [26] cannot be applied. The remaining case is considered in the following proposition.

Proposition 6.1. *Let J_0, φ, and $S(w)$ satisfy the assumptions of Lemma 6.2 and let $(C_\varphi)_i > c_{\varphi''}$ and $\operatorname{rank} B = n$. Then $S(w) = S'(w) = S''(w)$ holds true and the sequence w^ν defined by*

$$w^{\nu+1} = w^\nu - S''(w^\nu)^{-1}\nabla h(w^\nu) \tag{29}$$

converges superlinearly to the solution w^* of (7) if $\|w^0 - w^*\|$ is small enough.

Proof. See [12]. □

This result is unsatisfactory not only because of the restrictive assumptions (cf. Remark 5.2). It also does not give any information on the domain of convergence.

Proposition 6.2. *Let the assumptions of Proposition 6.1 be satisfied and assume that the solution w^* of (3) satisfies the non-degeneracy condition*

$$\exists \eta^* \in \mathcal{E} : \qquad f - B^T w^* \in F([u^*, \eta^*])^\circ \tag{30}$$

with $u^* = F^{-1}(f - B^T w^*)$. Then (29) reduces to a classical Newton method for H in the open neighborhood

$$U := (f - B^T(\cdot))^{-1}(F([u^*, \eta^*])^\circ).$$

Analogously the method of Corollary 6.1 with $\mathcal{C} = 0$ reduces to a damped classical Newton method on U.

Proof. We only have to note that F^{-1} is differentiable on $F([u^*, \eta^*])^\circ$ and that $f - B^T(\cdot)$ is continuous. □

In view of Proposition 6.2 the result of Proposition 6.1 is almost useless. Provided that the non-degeneracy condition on w^* holds, one can simply apply the convergence theory for classical smooth Newton methods in a small neighborhood U' contained in U. Since Proposition 6.1 does not ensure that the domain of convergence is larger than U' it does not give any additional information. If the inactive set $\mathcal{I}(u^*)$ of w^* and the set $F([u^*, \eta^*])^\circ$ are not known, then there is no hope that the local result can be applied. Moreover the determination of $\mathcal{I}(u^*)$ and $F([u^*, \eta^*])^\circ$ is generally not a simpler task then solving the original problem.

6.2. **Computational Aspects.** As already mentioned the terms h and $\nabla h = H$ are in general not explicitly available. In order to obtain an efficient method it is crucial to have fast iterative schemes to evaluate these quantities.

Before dealing with this problem we note that for $\mathcal{C} = 0$ the Schur Nonsmooth Newton method in Corollary 6.1 can equivalently written as

$$u^\nu = F^{-1}(f - B^T w^\nu), \tag{31}$$

$$w^{\nu+1} = w^\nu + \rho^\nu \underbrace{S''(w^\nu)^{-1}(Bu^\nu - Cw^\nu - g)}_{=:d^\nu} \tag{32}$$

with $\rho^\nu = \rho(\nu, w^\nu, d^\nu)$. This is a preconditioned Uzawa method for the original saddle point problem (1). If F is a linear operator it reduces to the classical Uzawa method and $S''(w)$ reduces to the linear Schur complement. In this case the preconditioned method obviously terminates within one step. If F is associated with a quadratic obstacle problem and the preconditioner is omitted standard convergence results for Uzawa methods can be applied yielding even an a priori fixed interval of allowed step sizes [10, 11].

The first substep amounts to the evaluation of F^{-1}, which is equivalent to the solution of the minimization problem

$$u^\nu = \operatorname*{arg\,min}_{u \in \mathbb{R}^n} \left(J(u) - \left\langle f - B^T w^\nu, u \right\rangle \right).$$

However such problems can efficiently solved using the *truncated nonsmooth Newton multigrid* (TNNMG) method. The TNNMG method was introduced in [15] for quadratic obstacle problems and later generalized to energy functionals of the above type [12, 13, 17] The the assumptions for this method are a subset of tmethod has recently been

The evaluation of F^{-1} is also needed if h or ∇h have to be evaluated in order to compute ρ^ν using a step size rule. This leads to multiple evaluations of F^{-1} per iteration step in general. If this is expensive it may be advantageous to adaptively switch off the step rule using the criterion (21) of Theorem 4.3. In view of the interpretation of the method as a Newton-type method one can hope that the norms of the directions decrease for good initial iterates. In this case the step rule will not be switched on only one evaluation of F^{-1} remains. However, the adaptive criterion (21) ensures that the method does still converge globally if this is not the case.

The second substep (32) involves the evaluation of $S''(w^\nu)^{-1}$. It can be written as the linear saddle point problem

$$\overline{u}^\nu \in \mathbb{R}^n,\ d^\nu \in \mathbb{R}^m : \qquad \begin{pmatrix} A^\nu & (B^\nu)^T \\ B^\nu & -C^\nu \end{pmatrix} \begin{pmatrix} \overline{u}^\nu \\ d^\nu \end{pmatrix} = \begin{pmatrix} 0 \\ g^\nu \end{pmatrix} \tag{33}$$

with

$$\begin{aligned} A^\nu &= \left(\partial^2 J_0(u) + \varphi''(u) \right)_{\mathcal{I}''(u^\nu)}, \\ B^\nu &= B_{\mathbb{N}, \mathcal{I}''(u^\nu)}, \\ C^\nu &= C + \tilde{C}(\mathcal{I}''(u^\nu)), \\ g^\nu &= \nabla h(w^\nu) = g + Cw^\nu - Bu^\nu, \end{aligned}$$

for an auxiliary variable $\overline{u}^\nu$. Since A^ν represents a linearization of $F = \partial J$ on the reduced space

$$V_{\mathcal{I}''(u^\nu)} = \operatorname{span}\{e_i : i \in \mathcal{I}''(u^\nu)\} = \mathbb{R}^n / \ker A^\nu, \tag{34}$$

this system can be regarded as a regularized linearization of the saddle point problem (1) on the reduced space $V_{\mathcal{I}''(u^\nu)} \times \mathbb{R}^m$. By construction the linear Schur complement of (33) is given by

$$S''(w^\nu) = B(A^\nu)^+ B^T + C + \tilde{C}(\mathcal{I}''(u^\nu)) = B^\nu (A^\nu)^+ (B^\nu)^T + C + \tilde{C}(\mathcal{I}''(u^\nu)).$$

Proposition 6.3. *The linear saddle point problem* (33) *has a unique solution* $(\overline{u}^\nu, d^\nu) \in V_{\mathcal{I}''(u^\nu)} \times \mathbb{R}^m$ *given by* $d^\nu = -S''(w^\nu)^{-1} g^\nu$ *and* $\overline{u}^\nu = -(A^\nu)^+ (B^\nu)^T d^\nu$. *The solutions of* (33) *in* $\mathbb{R}^n \times \mathbb{R}^m$ *are given by* $(\overline{u}^\nu + v^\nu, d^\nu) \in \mathbb{R}^n \times \mathbb{R}^m$ *with* $v^\nu \in V_{\mathcal{I}''(u^\nu)}^\perp = \{v \in \mathbb{R}^n : v = v_{\mathcal{I}''(u^\nu)}\} = \ker A^\nu$.

Proof. Replace A^ν by $A^\nu + I - I_{\mathcal{I}''(u^\nu)}$ and the right hand side of the first equation by v with $v \in V_{\mathcal{I}''(u^\nu)}^\perp$. Then a simple block elimination yields that $(\overline{u}^\nu + v, d^\nu) \in \mathbb{R}^n \times \mathbb{R}^m$ with $(\overline{u}^\nu, d^\nu)$ as given above is the unique solution of this modified system. Now the identity

$$(M_\mathcal{I})^+ = \left((M_\mathcal{I} + I - I_\mathcal{I})^{-1}\right)_\mathcal{I} = \left((M_\mathcal{I})^+\right)_\mathcal{I}$$

for the pseudoinverse together with the invariance of the original system under modifications $\overline{u}^\nu + v$ with $v \in \ker A^\nu$ provide the assertion. □

In view of this result the solution of (33) can either be obtained by considering the system on the subspace $V_{\mathcal{I}''(u^\nu)} \times \mathbb{R}^m = (\mathbb{R}^n / \ker A^\nu) \times \mathbb{R}^m$ only or by adding the orthogonal projection onto the kernel given by $P_{\ker A^\nu} = I - I_{\mathcal{I}''(u^\nu)} = I_{\mathbb{N} \setminus \mathcal{I}''(u^\nu)}$ to A^ν in order to make the part of $\overline{u}^\nu$ in $V_{\mathcal{I}''(u^\nu)}^\perp$ unique.

While there are general methods to solve the nonlinear convex minimization problems associated with F^{-1} the situation looks different for the linear saddle point problem. Since the problem is linear and symmetric it is possible to use a direct solver or Krylov methods like GMRES [27] or MINRES [22]. Due to the indefinite matrix there is no general multigrid method. However, there are multigrid methods that work well in special cases. Some of those methods require the saddle point problem to be related to a quadratic minimization problem with linear constraints, i.e. $C^\nu = 0$. Since this does not hold in general for the subproblems (33) we note that they can also be reformulated in the following way.

Proposition 6.4. *The linear saddle point problem* (33) *is equivalent to the saddle point problem*

$$(35) \quad \overline{u}^\nu \in \mathbb{R}^n, d_0^\nu \in \mathbb{R}^m, d^\nu \in \mathbb{R}^m : \quad \begin{pmatrix} A^\nu & 0 & (B^\nu)^T \\ 0 & C^\nu & -C^\nu \\ B^\nu & -C^\nu & 0 \end{pmatrix} \begin{pmatrix} \overline{u}^\nu \\ d_0^\nu \\ d^\nu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ g^\nu \end{pmatrix}$$

in the sense that $(\overline{u}^\nu, d_0^\nu, d^\nu)$ *is a solution of* (35) *iff* $(\overline{u}^\nu, d^\nu)$ *is a solution of* (33) *and* $Cd_0^\nu = Cd^\nu$. *The solutions of* (35) *are unique in* $V_{\mathcal{I}''(u^\nu)} \times (\mathbb{R}^m / \ker C) \times \mathbb{R}^m$ *and the Schur complement is given by* $S''(w^\nu)$.

Again we can construct a system that is uniquely solvable in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ by adding $P_{\ker A^\nu}$ to A^ν and $P_{\ker C^\nu}$ to the appearance of C^ν on the diagonal of (35) without changing the part of the solution in $V_{\mathcal{I}''(u^\nu)} \times (\mathbb{R}^m / \ker C) \times \mathbb{R}^m$.

One class of multigrid methods for systems of the form (35) uses the smoother by Braess and Sarazin [2]. Each application of this smoother incorporates the solution of a linear problem for $(B^\nu - C^\nu)((B^\nu)^T - C^\nu)^T$. While this reduces to a discretized second order elliptic problem for the Stokes problem it is not appropriate if B^ν or C^ν themselves result from a second order differential operator.

Another approach is to construct a smoother by successively solutving small local saddle point problems that couple only a few primal and dual unknowns in a so-called patch. Such smoothers were introduced by Vanka [32] for the Navier–Stokes equations. For the case of a parallel solution of the local problems, i.e. block Jacobi patch smoothers, convergence results were established by Zulehner [34, 35], Schöberl and Zulehner [28], and Simon and Zulehner [29].

References

[1] L. Armijo. Minimization of functions having Lipschitz–continuous first partial derivatives. *Pazific J. Math.*, 204:126–136, 1966.

[2] D. Braess and R. Sarazin. An efficient smoother for the Stokes problem. *Appl. Numer. Math.*, 23(1):3–19, 1997.

[3] J. W. Cahn and J. E. Hilliard. Free energy of a non-uniform system i. interfacial free energy. *Jnl. of Chemical Physic*, 28:258–267, 1958.

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

[5] M. I. M Copetti and C. M. Elliott. Numerical analysis of the Cahn-Hilliard equation with a logarithmic free energy. *Numerische Mathematik*, 63:39–65, 1992.

[6] P. Deuflhard. *Newton Methods for Nonlinear Problems*. Number 35 in Springer Series in Computational Mathematics. Springer, Berlin Heidelberg New York, 1. edition, 2004.

[7] I. Ekeland and R. Temam. *Convex Analysis*. North-Holland, 1976.

[8] C. M. Elliott. The Cahn-Hilliard model for the kinetics of phase separation. In J. F. Rodrigues, editor, *Mathematical Models for Phase Change Problems*, volume 88 of *International Series of Numerical Mathematics*. Birkhäuser, Basel, 1989.

[9] C. Geiger and C. Kanzow. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer, Berlin Heidelberg New York, 1999.

[10] R. Glowinski, J. L. Lions, and R. Trémolières. *Numerical Analysis of Variational Inequalities*. Number 8 in Studies in Mathematics and its Applications. North-Holland Publishing Company, Amsterdam New York Oxford, 1981.

[11] C. Gräser. Analysis und Approximation der Cahn-Hilliard Gleichung mit Hindernispotential. Diplomarbeit, Freie Universität Berlin, 2004.

[12] C. Gräser. *Convex Minimization and Phase Field Models*. PhD thesis, Freie Universität Berlin, 2011.

[13] C. Gräser. Truncated nonsmooth Newton multigrid methods for anisotropic convex minimization problems. in preparation, Matheon Berlin, 2013.

[14] C. Gräser and R. Kornhuber. On preconditioned Uzawa-type iterations for a saddle point problem with inequality constraints. In O. B. Widlund and D. E. Keyes, editors, *Domain Decomposition Methods in Science and Engineering XVI*, pages 91–102, Heidelberg, 2006. Springer.

[15] C. Gräser and R. Kornhuber. Multigrid methods for obstacle problems. *J. Comp. Math.*, 27(1):1–44, 2009.

[16] C. Gräser and R. Kornhuber. Nonsmooth Newton methods for set-valued saddle point problems. *SIAM J. Numer. Anal.*, 47(2):1251–1273, 2009.

[17] C. Gräser, U. Sack, and O. Sander. Truncated nonsmooth Newton multigrid methods for convex minimization problems. In M. Bercovier, M. Gander,

R. Kornhuber, and O. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, LNCSE, pages 129–136. Springer, 2009.

[18] B. Kummer. Newton's method based on generalized derivatives for nonsmooth functions: Convergence analysis. In W. Oettli and D. Pallaschke, editors, *Advances in optimization (Lambrecht, 1991)*, pages 171–194, Berlin, 1992. Springer.

[19] A. Nekvinda and L. Zajíček. A simple proof of the Rademacher theorem. *Časopis Pěst. Mat*, 113(4):337–341, 1988.

[20] J. Nocedal. Theory of algorithms for unconstrained optimization. *Acta Numerica*, 1:199–242, 1992.

[21] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[22] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numerical Analysis*, 12:617–629, 1975.

[23] J. S. Pang. Newton's method for b-differentiable equations. *Mathematics of Operations Research*, 15(2):311–341, 1990.

[24] M. J. D. Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287–336, 1998.

[25] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18(1):227–244, 1993.

[26] L. Qi and J. Sun. A nonsmooth version of Newtons's method. *Mathematical Programming*, 58:353–367, 1993.

[27] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 7:856–869, 1986.

[28] J. Schöberl and W. Zulehner. On Schwarz-type smoothers for saddle point problems. *Numer. Math.*, 95(2):377–399, 2003.

[29] R. Simon and W. Zulehner. On Schwarz-type smoothers for saddle point problems with applications to PDE-constrained optimization problems. *Numer. Math.*, 111:445–468, 2009.

[30] P. Spellucci. *Numerische Verfahren der nichtlinearen Optimierung*. Birkhäuser, Basel Berlin, 1993.

[31] M. Ulbrich. *Nonsmooth Newon-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. Habilitationsschrift, Technische Universität München, 2002.

[32] S. P. Vanka. Block-implicit multigrid solution of Navier-Stokes equations in primitive variables. *J. Comput. Phys.*, 65:138–158, 1986.

[33] D. Werner. *Funktionalanalysis*. Springer, Berlin Heidelberg New York, 3. edition, 2000.

[34] W. Zulehner. A class of smoothers for saddle point problems. *Computing*, 65 (3):227–246, 2000.

[35] W. Zulehner. Analysis of iterative methods for saddle point problems: A unified approach. *Math. Comput.*, 71(238):479–505, 2002.

Carsten Gräser, Freie Universität Berlin, Institut für Mathematik, Arnimallee 6, D - 14195 Berlin, Germany

E-mail address: graeser@mi.fu-berlin.de