Truncated Nonsmooth Newton Multigrid Methods for Simplex-Constrained Minimization Problems

Carsten Gräser and Oliver Sander

Bericht Nr. 384 **März 2014**

Key words: multigrid, simplex constraints, phase-field models, convex minimization

AMS Subject Classifications: 65K15, 90C25, 49M20

Institut für Geometrie und Praktische Mathematik
RWTH Aachen

Templergraben 55, D–52056 Aachen (Germany)

This work was supported by the DFG Research Center Matheon. The authors would like to thank Ralf Kornhuber for advice and interesting discussions.

TRUNCATED NONSMOOTH NEWTON MULTIGRID METHODS FOR SIMPLEX-CONSTRAINED MINIMIZATION PROBLEMS

CARSTEN GRÄSER AND OLIVER SANDER

Abstract. We present a multigrid method for the minimization of strongly convex functionals defined on a finite product of simplices. Such problems result, for example, from the discretization of multi-component phase-field problems. Our algorithm is globally convergent, requires no regularization parameters, and achieves multigrid convergence rates. We present numerical results for the vector-valued Allen–Cahn equation and observe that the convergence rate is independent from the temperature parameter and the number of components.

AMS classification: 65K15, 90C25, 49M20

Keywords: multigrid, simplex constraints, phase-field models, convex minimization

1. Introduction

Define the set K to be the n-fold product of the $(N-1)$-dimensional Gibbs simplex (Figure 1)

$$G := \Big\{ x \in \mathbb{R}^N \;\Big|\; \sum_{i=1}^{N} x_i = 1,\ x_i \geq 0,\ i = 1, \dots, N \Big\}.$$

It is a closed, convex subset of the space $(\mathbb{R}^N)^n$. On K we consider the minimization problem

$$u \in K : \qquad \mathcal{J}(u) \leq \mathcal{J}(v) \qquad \forall v \in K \tag{1}$$

for a functional $\mathcal{J}$ which is assumed to have the following properties:

(1) The functional $\mathcal{J}$ is strongly convex and continuous on K, and differentiable on the relative interior of K.
(2) It has the decomposition

$$\mathcal{J}(u) = \mathcal{J}_0(u) + \sum_{i=1}^{n} \sum_{j=1}^{N} \phi_{i,j}(u_{i,j}) + \chi_K(u), \tag{2}$$

with $\mathcal{J}_0 : (\mathbb{R}^N)^n \to \mathbb{R}$ being strongly convex, coercive, and C^2; and $\phi_{i,j} : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ convex, lower semi-continuous, and C^2 on $(0, \infty)$. Finally, χ_K is the indicator functional of the set K, i.e., $\chi_K(u) = 0$ for $u \in K$, and $+\infty$ otherwise. The addition of the indicator functional χ_K allows to reformulate the minimization problem (1) on the entire space $(\mathbb{R}^N)^n$.

This work was supported by the DFG Research Center Matheon. The authors would like to thank Ralf Kornhuber for advice and interesting discussions.

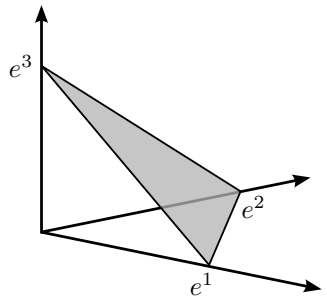

FIGURE 1. The two-dimensional Gibbs simplex

Note that we do not require that $\mathcal{J}$ be differentiable on the relative boundary ∂K of K in its affine hull. In particular, we allow that $|\nabla \mathcal{J}(x)| \to \infty$ for $x \to \partial K$. This generality is crucial to cover, e.g., the logarithmic potential used in phase-field modeling (Chapter 5). It sets us apart from competing approaches like [3], which assume $\mathcal{J}$ to be differentiable on the entire closure of K. We also note that the smoothness assumptions on $\mathcal{J}_0$ can be relaxed from C^2 to LC^1. In that case, the second partial derivatives appearing in Chapter 3 have to be replaced by suitable generalized derivatives of $\mathcal{J}_0'$, e.g., the generalized Jacobian in the sense of Clarke [7] or slanting derivatives [14].

Minimization problems like (1) arise for example from the finite element discretization of multi-component phase-field models. There, the distribution of N substances is described by N volume fractions. Since these fractions must each be contained in $[0, 1]$ and must sum up to 1 at each point, the simplex constraint structure arises. After a suitable time discretization, the phase-field behavior is governed by a strongly convex energy functional. A finite element discretization then leads to problems like (1).

The multi-component phase-field framework allows to model many physical and abstract systems. The vector-valued Allen–Cahn equations, for example, model the behavior of physical substances with phase changes (see Chapter 5). Allen–Cahn-type problems also arise as subproblems when solving the more challenging Cahn–Hilliard-type problems [11]. However, multi-component phase fields have also been used for topology optimization [2], and image segmentation [15]. All these applications benefit from efficient and robust solvers for the minimization problem (1).

Various methods have been proposed for minimization problems with this structure. In principle, many standard methods from convex optimization such as gradient projection methods [1] or interior point methods [6] can be used. However, such methods tend to be slow since they do not exploit the special structure of the problem. Of the methods that do use this structure we want to highlight the Monotone Multigrid Method proposed by Kornhuber and Krause [16]. They suggested a Gauß–Seidel type smoother where the energy is minimized locally in the directions of the simplex edges, and combined it with a special nonlinear multigrid step. This multigrid step involved local damping of the individual scalar corrections on all levels. Kornhuber and Krause proved global convergence for their scheme. However, numerical experiments showed that this solver could only be used efficiently for very low numbers of components N. Since by construction their local damping scales

with the inverse of N, the coarse grid corrections effectively vanish for increasing N. Kornhuber and Krause observed that convergence could be accelerated if the damping was omitted. However, no convergence proof was given for that case.

Alternatively, Blank and coworkers presented a primal–dual active set method for the vector-valued Allen–Cahn equation with mass conservation in [3]. They considered only the simpler case of the pure obstacle potential, where $\phi_{i,j} \equiv 0$. For this method they showed local convergence.

Inspired by [16], we proposed the Polyhedral Gauß–Seidel method (PGS), which can reliably minimize strictly convex functionals that are piecewise smooth on general polyhedral decompositions of a general polyhedron contained in $(\mathbb{R}^N)^n$ [13]. For this method we showed global convergence even for inexact local minimizations. The convergence result also allows for a general acceleration step, which does not impede convergence as long as it is monotone, i.e., energy decreasing.

The product of simplices $K = \bigotimes_{i=1}^{n} G$ discussed in this article forms a special case of the general PGS framework. For this setting, we propose a coarse grid correction that leads to optimal convergence rates and is much simpler to construct than the one proposed in [16]. The resulting method is called Truncated Nonsmooth Newton Multigrid (TNNMG), because it can be interpreted as an inexact Newton method for the stationarity equation of the Gauß–Seidel smoother in the quadratic obstacle case [9]. It generalizes the corresponding method for scalar problems from [12].

When comparing with the numerical results of Kornhuber and Krause given in [16], our solver turns out to be much more efficient and robust. Indeed, the convergence rates appear to be independent of N. In Chapter 5 we show experiments for up to 18 components, with the convergence rate staying below 0.1 in all cases. Additionally, our multigrid correction is much easier to implement than the one from [16]. Our coarse grid correction involves only one global damping step. This turns out to be the right middle-ground between the too aggressive local damping of [16] (which converges only slowly), and no damping at all (which may diverge). Compared to [3], our method is much more efficient, since only one multigrid cycle is used for the coarse grid correction. Moreover, our method is *globally* convergent and is not restricted to quadratic functionals.

One final advantage of our method is its robustness with respect to certain singular limits of the functional. Several potentials used in phase-field modeling approach functions that are nonsmooth on ∂K for particular parameter choices. This does not disturb our method as it can easily handle the nonsmooth case.

We proceed as follows. In Chapter 2 we specialize the polyhedral Gauß–Seidel method to the case of simplex constraints. This leads to a globally convergent method, with an energy-decreasing acceleration step left to be specified. To obtain multigrid convergence rates we then propose a Newton-type correction in Chapter 3 and introduce a linear multigrid method for its inexact evaluation in Chapter 4. We test our algorithm with the vector-valued Allen–Cahn equation both for the logarithmic and the obstacle potential in Chapter 5 and discuss its complexity in Chapter 6. We observe optimal convergence rates independent from the temperature parameter and the number of components.

2. Accelerated Polyhedral Gauss–Seidel Methods

In [13] the polyhedral Gauß–Seidel method (PGS) was introduced for strictly convex functionals $\mathcal{J}$ that are piecewise smooth on a polyhedral partition of the admissible set. For the special case of a single polyhedron it was shown there that a Gauß–Seidel-type method that successively minimizes $\mathcal{J}$ in all edge directions converges globally. This generalized the corresponding result for simplices from [16].

In the following we describe the PGS method for the admissible set being an n-fold product K of the Gibbs simplex G. Let $e^i \in \mathbb{R}^N$ such that $(e^i)_j = \delta_{ij}$ (the Kronecker symbol). Then the edges of G are given by

$$\mathcal{E} = \big\{\eta^{ij} = e^i - e^j \in \mathbb{R}^N \mid 1 \le i < j \le N\big\},$$

and the number of such edges is $M = |\mathcal{E}| = (N-1)N/2$. To extend these local edge directions to global ones let $E^k \in \mathbb{R}^n$ with $(E^k)_l = \delta_{kl}$. Then the elements of

$$\big\{\eta \otimes E^k \in (\mathbb{R}^N)^n \mid \eta \in \mathcal{E},\ 1 \le k \le n\big\}$$

are the edges of the polyhedron $K = \bigotimes_{i=1}^n G$.

We will use these directions as search directions for a nonlinear Gauß–Seidel method. For a fixed ordering $\mathcal{E} = \{\eta^1, \dots, \eta^M\}$, denote by $V_{k,l}$ the linear subspace of $(\mathbb{R}^N)^n$ spanned by $\eta^l \otimes E^k$. The (non-direct) sum of all these spaces is the linear hull $\mathcal{V}$ of $K - w$ for any $w \in K$.

Let $u^\nu \in K$ be a given iterate. One step of the polyhedral Gauß–Seidel method takes the form

(1) Set $w^{0,M} = u^\nu$
(2) For $k = 1, \dots, n$ do
  (a) Set $w^{k,0} = w^{k-1,M}$
  (b) For $l = 1, \dots, M$ do
    Find $w^{k,l} = \operatorname*{arg\,min}_{v \in w^{k,l-1} + V_{k,l}} \mathcal{J}(v)$
(3) Set $u^{\nu+\frac{1}{2}} = w^{n,M}$
(4) Find some

$$u^{\nu+1} \in u^{\nu+\frac{1}{2}} + \mathcal{W}(u^{\nu+\frac{1}{2}}) \ : \qquad \mathcal{J}(u^{\nu+1}) \le \mathcal{J}(u^{\nu+\frac{1}{2}}),$$

where $\mathcal{W}(u^{\nu+\frac{1}{2}})$ is an additional subspace of $(\mathbb{R}^N)^n$ which may depend on $u^{\nu+\frac{1}{2}}$.

Our setting here is a special case of the more general polyhedral setting discussed in [13]. The following convergence result is therefore a corollary of the main convergence theorem from there.

Theorem 2.1. *Under the stated assumptions, the algorithm converges to the unique minimizer of $\mathcal{J}$ on K for all initial iterates $u^0 \in K$.*

Note the additional monotone correction Step 4. It is the key to obtaining algorithms with multigrid performance, because it allows an additional correction in a space of suitable long-range interactions. Importantly, we do not require $u^{\nu+1}$ to be minimal in the space $u^{\nu+\frac{1}{2}} + \mathcal{W}(u^{\nu+\frac{1}{2}})$, but only to decrease energy.

The *Truncated Nonsmooth Newton Multigrid* (TNNMG) method is obtained as a particular choice of this correction step, which we construct in the following section. Since that correction is energy-decreasing, the global convergence of the overall method follows from Theorem 2.1.

Remark 2.1. *The convergence result in Theorem 2.1 also holds if the local minimization problems are only solved inexactly. This can lead to significant performance gains. See* [13] *for details.*

3. Truncated Newton Corrections

A naive choice of coarse correction would consist of an inexact Newton-type correction at the iterate $u^{\nu+\frac{1}{2}}$. Straightforward linearization is not possible, however, due to the nonsmoothness of the problem. Instead, to obtain an efficient method, we restrict the Newton-type correction at the iterate $u^{\nu+\frac{1}{2}} \in K$ to a subspace where $\mathcal{J}$ is locally smooth. This subspace is then precisely the space $\mathcal{W}(u^{\nu+\frac{1}{2}})$ mentioned in the general algorithm. Since $\mathcal{J}$ is non-differentiable perpendicular to K any such choice of $\mathcal{W}(u^{\nu+\frac{1}{2}})$ will be a subset of the linear hull $\mathcal{V}$ of $K - u^{\nu+\frac{1}{2}}$. We expect that this subspace varies only mildly in the neighborhood of the solution. Then the method can be expected to be very efficient if good initial iterates are available. Note, however, that the method converges even for bad initial iterates.

3.1. **Truncated Subspace Linearization.** For a given point $w \in K$ we now consider the construction of a linear subspace $\mathcal{W}(w) \subset (\mathbb{R}^N)^n$ such that $\mathcal{J}$ can be linearized in $w + \mathcal{W}(w)$. The most simple guess $\mathcal{W}(w) = \mathcal{V}$ will not work in general as $\mathcal{J}$ may not be differentiable in this space if $w \in \partial K$. To construct a more sophisticated guess we consider the active sets $a_k(w_k) := \{j \mid w_{k,j} = 0\}$ for each block $w_k \in G$. Then $\mathcal{J}$ is differentiable near w in the space

$$\prod_{k=1}^{n} \operatorname{span}\big\{\eta^{ij} \in \mathcal{E} \,\big|\, i, j \notin a_k(w_k)\big\}.$$

It can be shown that this space is the maximal subspace of $\mathcal{V}$ where $\mathcal{J}(w + \cdot)$ is differentiable near zero. It therefore makes a good first guess for the correction space $\mathcal{W}(w)$. However, there is a second problem. Using a linearization of $\mathcal{J}$ in this space to compute Newton-type corrections may lead to numerical problems due to unbounded derivatives as some $w_{i,j}$ approach 0. In order to avoid this we use the extended sets of active component given by $\mathcal{A}_k(w_k) = \{j \mid w_{k,j} = 0 \text{ or } \varphi''_{k,j}(w_{k,j}) > C\}$ for a large constant C. We can then define the truncated space $\mathcal{W}(w) \subset \mathcal{V}$ as

$$\mathcal{W}(w) = \prod_{k=1}^{n} W_k(w_k), \qquad W_k(w_k) = \operatorname{span}\big\{\eta^{ij} \in \mathcal{E} \,\big|\, i, j \notin \mathcal{A}_k(w_k)\big\}.$$

With the space $\mathcal{W}(w)$ defined we now construct an energy-decreasing step in $\mathcal{W}(w)$. Consider the linearized problem

$$v \in \mathcal{W}(w): \qquad H(w)v = -g(w), \tag{3}$$

where $g(w) = \mathcal{J}'(w)|_{\mathcal{W}(w)}$ and $H(w) = \mathcal{J}''(w)|_{\mathcal{W}(w)\times\mathcal{W}(w)}$ are the gradient and Hessian of $\mathcal{J}$ on $\mathcal{W}(w)$, respectively. We call $g(w)$ and $H(w)$ the *truncated linearization* of $\mathcal{J}$ at w.

Lemma 3.1. *There is a unique solution $v \in \mathcal{W}(w)$ of* (3).

Proof. By definition we have $\ker H(w) \perp \mathcal{W}(w)$ and $g(w) \in \mathcal{W}(w)$. Using the strong convexity of $\mathcal{J}$ we find that $H(w)$ is symmetric positive definite on $\mathcal{W}(w)$, which implies the assertion. □

While the definitions of $H(w)$ and $g(w)$ as operator and functional on $\mathcal{W}(w)$ are straightforward, in practice one has to extend them by zero to the orthogonal complement of $\mathcal{W}(w)$ in $(\mathbb{R}^N)^n$ in order to derive actual matrix and vector representations.

In the following we will briefly sketch how to do this. Denote by $\mathcal{P}_S$ the orthogonal projection onto a closed convex set S, and let $\mathcal{J}'(w)$ and $\mathcal{J}''(w)$ be a generalized gradient and Hessian of $\mathcal{J}$ at $w \in (\mathbb{R}^N)^n$, respectively. Here we understand "generalized" in the sense that the entries are the partial derivatives of $\mathcal{J}$ wherever they exist, and arbitrary elsewhere. Then the global projection $\mathcal{P}_{\mathcal{W}(w)} : (\mathbb{R}^N)^n \to \mathcal{W}(w)$ is a block-diagonal matrix given by

$$\mathcal{P}_{\mathcal{W}(w)} \in (\mathbb{R}^{N\times N})^{n\times n}, \qquad \Big(\mathcal{P}_{\mathcal{W}(w)}\Big)_{ij} = \delta_{ij}\mathcal{P}_{W_i(w_i)},$$

with

$$\Big(\mathcal{P}_{W_k(w_k)}\Big)_{ij} = \begin{cases} \delta_{ij} - \frac{1}{N-|\mathcal{A}_k(w_k)|} & \text{if } i,j \notin \mathcal{A}_k(w_k), \\ 0 & \text{else.} \end{cases}$$

We can then write

$$H(w) = \mathcal{P}_{\mathcal{W}(w)}^T \mathcal{J}''(w)\mathcal{P}_{\mathcal{W}(w)}, \qquad g(w) = \mathcal{P}_{\mathcal{W}(w)}^T \mathcal{J}'(w).$$

As $\mathcal{P}_{\mathcal{W}(w)}$ is block-diagonal $H(w)$ inherits the sparsity pattern of $\mathcal{J}''(w)$, and all entries can be computed independently according to

$$H(w)_{ij} = \mathcal{P}_{W_i(w_i)}^T \Big(\mathcal{J}''(w)_{ij}\Big)\mathcal{P}_{W_j(w_j)}.$$

Using this matrix representation the solution of (3) is given by $v = -H(w)^+ g(w)$, where $(\cdot)^+$ is the Moore–Penrose pseudo-inverse. Notice that this is independent of the entries selected for non-existing partial derivatives.

Remark 3.1. *As each local subspace $W_i(w_i)$ is a subspace of $\{x \in \mathbb{R}^N \mid \sum x_i = 0\}$ the size of the local blocks can be reduced from $\mathbb{R}^{N\times N}$ to $\mathbb{R}^{(N-1)\times(N-1)}$ by representing them using a fixed basis of this space. This can lead to considerable storage and run-time savings, especially for small N. The basis change leads to an equivalent algorithm, unless the regularization described at the end of Chapter 4 is used. The properties of the regularized problem described there hold irrespectively of the choice of basis, though.*

3.2. **Post-Processing of Inexact Newton Corrections.** Using the truncated linearization we can compute Newton-like corrections at the intermediate iterate $w = u^{\nu+\frac{1}{2}}$ in the subspace $\mathcal{W}(w)$ by solving (3). However, one will normally only want to use an approximation $\tilde{v}$ of $v = -H(w)^+ g(w)$. In order to guarantee energy decrease one could use a simple damping strategy. Unfortunately this may lead to very small damping parameters if the correction is nonzero in components where w is close to ∂K. To overcome this we first project $\tilde{v}$ onto the admissible set.

Let $\mathcal{P}_{K-w}$ be the orthogonal projection onto $K-w$. Due to the product structure of K, the projection $\mathcal{P}_{K-w}\tilde{v}$ is given by

$$(\mathcal{P}_{K-w}\tilde{v})_i = \mathcal{P}_{G-w_i}\tilde{v}_i \in G - w_i, \qquad i = 1,\dots,n,$$

where $\mathcal{P}_{G-w_i}$ is the orthogonal projection onto the Gibbs simplex. This projection can be evaluated efficiently in $O(N \log N)$ steps [17]. Since $\mathcal{P}_{G-w_i}(x) = \mathcal{P}_G(x +$

$w_i) - w_i$ it is enough to consider the evaluation of $y = \mathcal{P}_G x$. The associated minimization problem is equivalent to the optimality system

$$\begin{pmatrix} I + \partial\chi_{[0,\infty)^N} & \mathbf{1} \\ \mathbf{1}^T & \end{pmatrix} \begin{pmatrix} y \\ \lambda \end{pmatrix} \ni \begin{pmatrix} x \\ 1 \end{pmatrix},$$

where $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^N$ and $\lambda \in \mathbb{R}$ is the Lagrange multiplier for the linear constraint $\sum y_i = 1$. For given λ this implies $y_i = (1 + \partial\chi_{[0,\infty)})^{-1}(x_i - \lambda) = \max(x_i - \lambda, 0)$. Inserting this in the second equation yields the nonlinear Schur complement equation

$$1 = \sum_{i=1}^{N} \max(x_i - \lambda, 0). \tag{4}$$

To solve for λ we first sort x such that $x_1 \le \cdots \le x_N$. Since $\lambda < x_N$ there is a minimal index $i_0 \le N$ such that $\lambda < x_{i_0}$. For this index (4) implies

$$1 = \sum_{i=1}^{N} \max(x_i - \lambda, 0) = \sum_{i=i_0}^{N} (x_i - \lambda),$$

and thus $\lambda = \mu_{i_0}$ where $\mu_i = (\sum_{j=i}^{N} x_j)/(N - j + 1)$. Hence we can find i_0 as the maximal index with $x_{i_0-1} \le \mu_{i_0}$. While sorting x is an $O(N \log N)$ operation finding i_0 can be done in $O(N)$ steps by successively checking for $i = N, \dots, 1$. An alternative proof for the correctness of this method has been given in [17].

The projected inexact directions $\mathcal{P}_{\left(K - u^{\nu+\frac{1}{2}}\right)} \tilde{v}$ cannot be guaranteed to always be descent directions for the energy $\mathcal{J}$. In practice, however, it is observed that they are descent directions in almost all cases. To really guarantee a correction step with non-increasing energy we perform a line search in the projected direction

$$u^{\nu+1} = u^{\nu+\frac{1}{2}} + \rho \mathcal{P}_{\left(K - u^{\nu+\frac{1}{2}}\right)} \tilde{v}, \quad \rho = \operatorname*{arg\,min}_{\sigma \in [0,\infty)} \mathcal{J}\Big(u^{\nu+\frac{1}{2}} + \sigma \mathcal{P}_{\left(K - u^{\nu+\frac{1}{2}}\right)} \tilde{v}\Big). \tag{5}$$

Then we have $u^{\nu+1} \in K$ and $\mathcal{J}(u^{\nu+1}) \le \mathcal{J}(u^{\nu+\frac{1}{2}})$ by construction. Numerical experiments show that this construction leads to very efficient methods.

3.3. **The TNNMG Method.** Obviously one can replace the optimal line-search parameter ρ in (5) by approximations as long as the monotonicity criterion $\mathcal{J}(u^{\nu+1}) \le \mathcal{J}(u^{\nu+\frac{1}{2}})$ is satisfied. Notice that such approximations can always be computed using a simple bisection method. One can also check whether $\rho = 1$ does already satisfy the criterion and resort to the bisection only if this is not the case. For efficiency reasons this procedure was used in all computations to be reported below. In any case Theorem 2.1 implies global convergence for all such corrections.

Corallary 3.2. *The extended Gauß–Seidel method with the additional monotone correction* (5) *converges to the unique minimizer of $\mathcal{J}$ on K for all initial iterates $u^0 \in K$.*

In the previous chapter we have constructed corrections as solutions to the underdetermined linear system of equations (3). For efficiency reasons, inexact solutions will usually be used. It is one of the strengths of the presented method to allow many alternatives here. For example, iterations of a CG method are possible, as is the use of an actual direct solver. Since multigrid methods are known to

be amongst the most efficient solvers for symmetric positive (semi-)definite linear systems related to partial differential equations we propose to solve (3) inexactly by one linear multigrid step. We call the resulting method *Truncated Nonsmooth Newton Multigrid* (TNNMG). It can be viewed as an overall nonlinear multigrid method with nonlinear fine grid smoother and linear coarse grid corrections in suitable subspaces. A possible choice for the linear multigrid method will be discussed in the next chapter.

4. Multigrid Solution of Linear Subproblems

Assume the important special case that $(\mathbb{R}^N)^n$ is the coefficient space of a vector-valued discrete function space $\mathcal{S}_J^N$. We now briefly discuss the computation of multigrid corrections for (3) based on a hierarchy $\mathcal{S}_0 \subset \cdots \subset \mathcal{S}_J$ of subspaces of the corresponding scalar function space $\mathcal{S}_J$. To this end we identify $\mathcal{S}_k$ with $\mathbb{R}^{n_k}$, where $n_J = n$, and assume that the natural embedding of $\mathcal{S}_{k-1}$ into $\mathcal{S}_k$ is given by the prolongation matrix $\tilde{P}_k \in \mathbb{R}^{n_k \times n_{k-1}}$. This induces a hierarchy $\mathcal{S}_0^N \subset \cdots \subset \mathcal{S}_J^N$ of vector-valued function spaces and corresponding vector-valued prolongation operators $P_k \in (\mathbb{R}^{N\times N})^{n_k \times n_{k-1}}$ given by $(P_k)_{ij} = (\tilde{P}_k)_{ij} I \in \mathbb{R}^{N\times N}$.

We set $A_J = H(w)$ and construct a hierarchy of stiffness matrices by setting $A_{k-1} = P_k^T A_k P_k$. We further assume the existence of linear smoothing operators $B_k : (\mathbb{R}^N)^{n_k} \to (\mathbb{R}^N)^{n_k}$. A linear multigrid step for (3) starting from a zero initial value takes the form

(1) Set $r^J = -g(w)$
(2) For $k = J, \dots, 1$ do
  (a) Compute $v^k = B_k^{-1} r^k$
  (b) Set $r^{k-1} = P_k^T (r^k - A_k v^k)$
(3) Compute $v^0 = B_0^{-1} r^0$
(4) Set $\tilde{v} = \mathcal{P}_{\mathcal{W}(w)} \sum_{k=0}^{J} \Big(\prod_{l=k+1}^{J} P_l \Big) v^k$.

In this simple form there is a single pre-smoothing step and no post-smoothing. The extension to a V-cycle with multiple pre- and post-smoothing steps is straightforward.

We note that the only difference to a standard linear multigrid step is the projection $\mathcal{P}_{\mathcal{W}(w)}$ in the last step. This modification is needed since, while the method does in principle act in the quotient space $(\mathbb{R}^N)^n / \ker H(w) = \mathcal{W}(w)$, it represents all corrections in the larger space $(\mathbb{R}^N)^n$. Round-off errors may create spurious contributions in $\ker H(w)$. However the corrections in the quotient space are invariant under those contributions and the latter can easily be removed by the additional projection to guarantee $\tilde{v} \in \mathcal{W}(w)$.

Since the matrices A_k are in general symmetric but positive semi-definite only, we will briefly discuss the smoothers B_k. A simple choice is a block Gauß–Seidel method based on a block-triangular decomposition $A_k = D_k + L_k + L_k^T$, where $(D_k)_{ij} = \delta_{ij}(A_k)_{ii}$. However, $D_k + L_k$ may be singular because the diagonal entries $(D_k)_{ii}$ are in general only positive semi-definite. One way to avoid this is to formally replace $(D_k)_{ii}$ by $(D_k)_{ii} + \mathcal{P}_{\ker(D_k)_{ii}}$. The resulting correction is the unique solution of $(D_k + L_k)v^k = r^k$ with $v^k \perp \ker(D_k + L_k)$.

In practice computing $\ker(D_k)_{ii}$ can be avoided by using a kernel-invariant local solver like, e.g., the CG method for $(D_k)_{ii}$ directly. To avoid numerical difficulties due to the infinite condition number of $(D_k)_{ii}$ we suggest the following cheap

modification for a small parameter $0 < \alpha \ll 1$:

$$\left((\tilde{D}_k)_{ii}\right)_{lm} := \begin{cases} \left((D_k)_{ii}\right)_{lm} & \text{if } l \neq m, \\ \alpha & \text{if } l = m \text{ and } \left((D_k)_{ii}\right)_{lm} = 0, \\ \left((1+\alpha)(D_k)_{ii}\right)_{lm} & \text{if } l = m \text{ and } \left((D_k)_{ii}\right)_{lm} \neq 0. \end{cases} \tag{6}$$

Then $(\tilde{D}_k)_{ii}$ is symmetric positive definite and the resulting correction can be viewed as a damped version of the one obtained for $(D_k)_{ii} + \mathcal{P}_{\ker(D_k)_{ii}}$. Numerical experiments for the example problem presented below showed that the CG method was robust for $\alpha \geq 10^{-14}$, and the overall convergence rates were hardly influenced as long as $\alpha \leq 10^{-4}$. By Theorem 2.1, the overall algorithm remains globally convergent irrespectively of the choice of α.

Remark 4.1. *We have presented the multigrid method here for a sequence of nested function spaces, but these are not actually required. A correction step based on a purely algebraic multigrid hierarchy is easily possible as well.*

5. Numerical Example: The Allen–Cahn Equation

In order to investigate the efficiency and robustness of the TNNMG method introduced above we apply it to a multicomponent Allen–Cahn equation with logarithmic and obstacle type potentials [8]. This equation is an L^2-gradient flow of the Ginzburg–Landau energy

$$E(u) = \frac{\varepsilon}{2} \int_\Omega \sum_{i=1}^N |\nabla u_i|^2 \, dx + \frac{1}{\varepsilon} \int_\Omega \Psi(u) \, dx,$$

where $\Psi : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ is a multi-well potential which can typically be split as $\Psi(u) = \Phi(u) - \frac{1}{2}\langle Cu, u\rangle$ for a convex, lower semi-continuous functional $\Phi : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ and a symmetric interaction matrix $C \in \mathbb{R}^{N\times N}$. We will concentrate on $C = I$ and the prototypic case of a logarithmic potential

$$\Phi_\theta(u) = \theta \sum_{i=1}^N u_i \ln(u_i) + \chi_G(u),$$

where $\theta > 0$ is a scalar parameter which is interpreted as a temperature. In the limiting case $\theta \to 0$, the obstacle potential $\Phi_0 = \chi_G$ is obtained. Figure 2 depicts the level sets of Ψ for $N = 3$ and different values of θ. Notice that, for $\theta \to 0$, the local minima (marked by dots) tend to the corners of G where Ψ' is singular. A weak formulation of the Allen–Cahn equation reads

$$\varepsilon\langle u_t, v - u\rangle + \varepsilon(\nabla u, \nabla(v-u)) + \frac{1}{\varepsilon}\int_\Omega (\Phi(v) - \Phi(u))dx - \frac{1}{\varepsilon}(u, v-u) \geq 0.$$

It is well-posed in $W(0,T;H^1(\Omega))$ if an initial value $u(0) = u_0$ is provided.

Time discretization using an implicit Euler scheme leads to stationary variational inequalities

$$u^k \in H^1(\Omega): \quad a(u^k, v-u^k) + \frac{1}{\varepsilon}\int_\Omega (\Phi(v) - \Phi(u^k))dx \geq l(v-u^k) \quad \forall v \in H^1(\Omega)$$

for the bilinear form $a(u,v) = (\frac{\varepsilon}{\tau} - \frac{1}{\varepsilon})(u,v) + \varepsilon(\nabla u, \nabla v)$ and right hand side $l(v) = \frac{\varepsilon}{\tau}(u^{k-1}, v)$. We note that, due to the concave term $-\epsilon^{-1}(u,v)$, the bilinear form $a(\cdot,\cdot)$ is elliptic only under the time step restriction $\tau < \varepsilon^2$. While

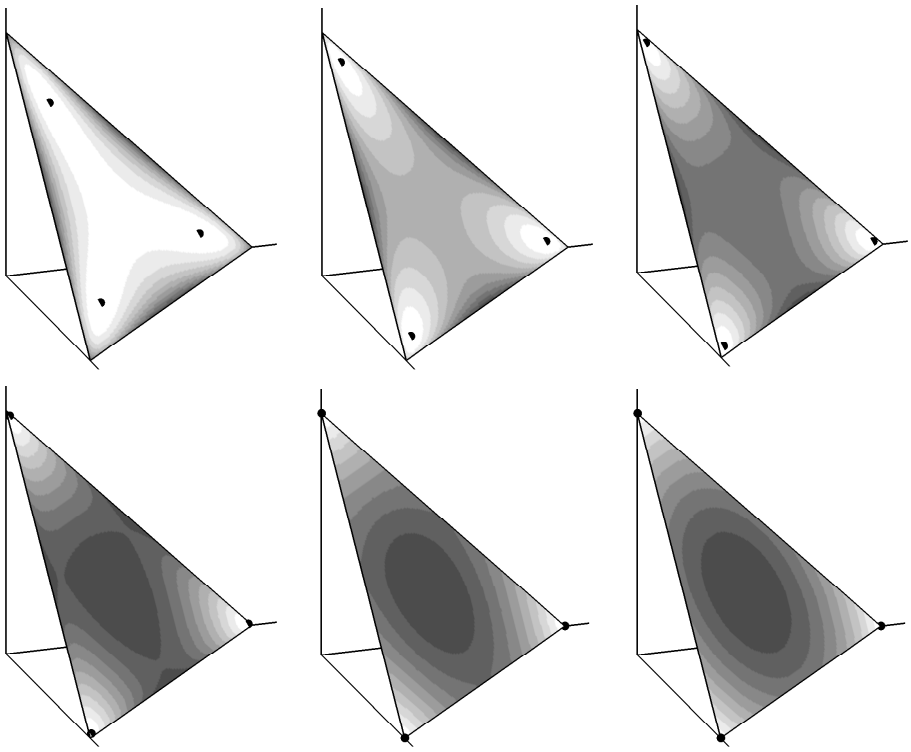

FIGURE 2. Level sets and local minima of the logarithmic potential Ψ for $N = 3$ and $\theta = 0.35$, 0.3, 0.25, 0.2, 0.1, 0. Note how the minimizers (black dots) tend to the corners of G as θ goes to zero. The derivative of Ψ at ∂G is singular for all θ.

semi-implicit discretizations have been proposed to overcome this [5], they typically lack accuracy [4, 10]. Since the structure of the stationary problems is the same for both discretizations, we concentrate on the fully implicit case here, and assume $\tau < \epsilon^2$ from now on.

Spatial discretization by first-order finite elements using a lumped approximation of $\int_\Omega \Phi(u)dx$ then leads to discrete minimization problems as introduced in Chapter 1 for $\mathcal{J}$ taking the form (2) with

$$\mathcal{J}_0(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle, \qquad \phi_{i,j}(z) = \frac{\theta}{\varepsilon} z\ln(z)\omega_i.$$

Here A is a positive definite matrix and $\omega_i > 0$ is the integral over the i-th scalar nodal basis function.

All numerical tests are carried out for the parameters $\varepsilon = 0.05$, $\Omega = [0,1]^2$, and $\tau = 0.002 < \varepsilon^2$ on a mesh hierarchy obtained by uniform refinements of an initial mesh with two triangles. The evolution of the phase field is depicted in Figure 3.

The TNNMG-method was applied with 3 polyhedral Gauß–Seidel pre- and post-smoothing steps on the fine mesh. For the coarse correction we used one linear multigrid V-cycle with 3 block-Gauß–Seidel pre- and post-smoothing steps. The diagonal blocks were regularized according to (6) and the obtained $(N-1)\times(N-1)$

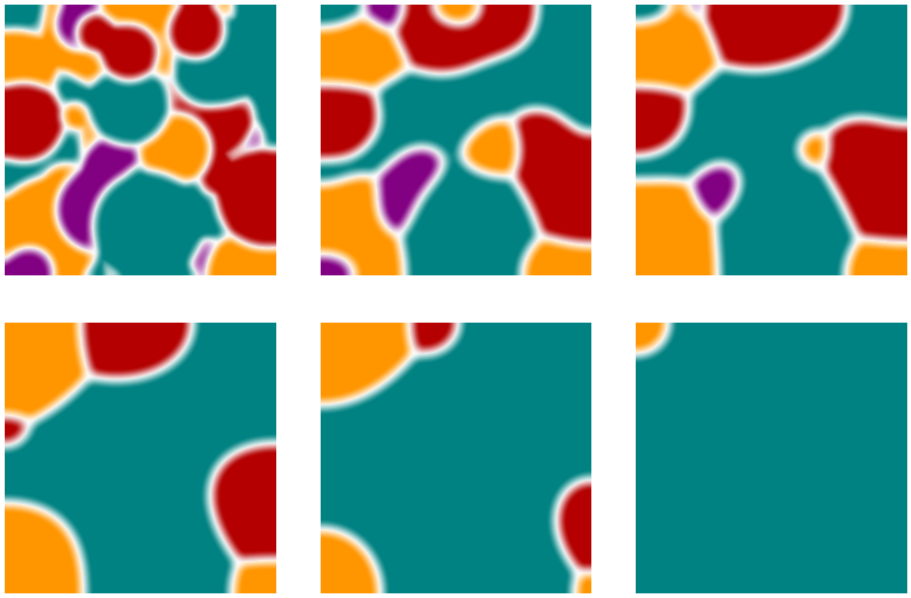

FIGURE 3. Solution for $\theta = 10^{-5}$ and $N = 4$ at time steps 1, 10, 20, 50, 100, and 200

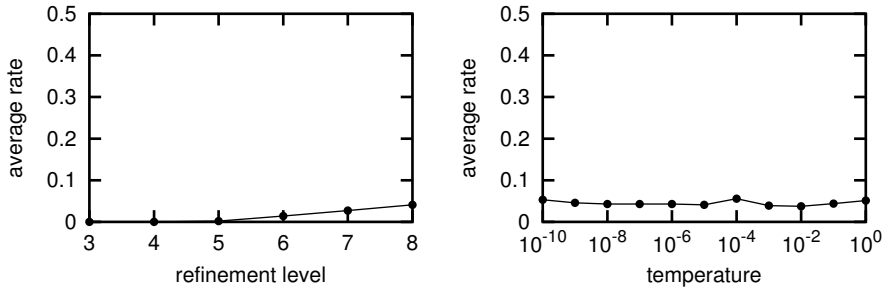

FIGURE 4. Average convergence rate over refinement level J (left) and over temperature θ (right)

local linear systems were solved up to machine accuracy using a CG method. A reasonable initial value was computed by nested iteration. All presented rates are averaged convergence rates computed for a reference solution obtained by solving to a higher accuracy.

We present the averaged convergence rates for the solution of the first time step. If not stated otherwise the experiments are for $\theta = 10^{-5}$, $N = 4$, and the 8th refinement level. We solve the problem until the energy norm of the correction drops below 10^{-11}. First, we investigate the mesh dependence of the solver. Figure 4 (left) shows that the convergence rates is hardy influenced by the mesh size and goes up to only ≈ 0.04 on the finest mesh. Next we investigate the dependence

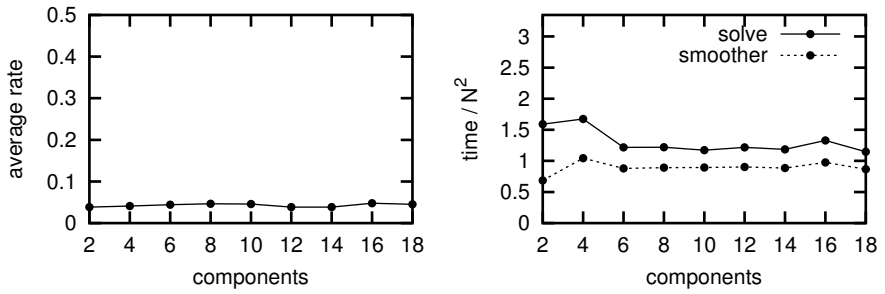

FIGURE 5. Average convergence rate (left) and relative computational time/N^2 over number of components N (right)

on $\theta \geq 0$. Figure 4 (right) shows that the rate is bounded by ≈ 0.06 for θ in the range between 10^{-10} and 1. In the limiting case $\theta = 0$ (not depicted in the figure) the rate was ≈ 0.04. Finally we investigate the dependence of the convergence rate on the number of components. Figure 5 (left) shows that the rate is bounded by ≈ 0.05 for $N = 2, \ldots, 18$ components.

Notice that, in contrast to, e.g., the heat equation, the good convergence rates cannot be explained by the small time step size because the resulting discrete flow is still far from the identity due to the concave term in the Ψ.

6. Complexity

It remains to discuss the complexity of the proposed method. Assume that $\mathcal{J}_0$ is quadratic with a sparse matrix A and each A_{ii} being diagonal as given in the example problem. Then one nonlinear PGS step has $O(nN^2)$ complexity given that each scalar minimization problem can be solved in constant time. For the coarse grid correction we assume that a multigrid step for the related hierarchy of scalar valued function spaces has optimal complexity $O(n)$. In view of the dense $(N-1)\times(N-1)$ blocks in the linearization the complexity of the coarse correction then ranges from $O(nN^2)$ to $O(nN^3)$ depending on the local linear solver.

To also assess the implementation and cache effects we measured the computational time needed to solve the example problem for $N = 2, \ldots, 18$ components using a CG method as local linear solver. In order to visualize the dependence on N, Figure 5 (right) depicts the time needed to solve the minimization problem, scaled by N^{-2}. The plot reveals that the overall effort (solid line) as well as the share of the nonlinear fine grid smoother PGS (dotted line) scale by N^2, and that the latter dominates the overall effort.

7. Conclusions

We introduced the TNNMG method for simplex-constrained minimization problems by augmenting the polyhedral Gauß–Seidel (PGS) method with a coarse grid correction based on a linearization in a suitable subspace. Global convergence follows from a convergence result for PGS in [13]. Numerical examples illustrate that the TNNMG method exhibits mesh independent convergence rates comparable to

those of linear multigrid for elliptic problems. Furthermore the rates are independent on the local simplex dimension N and robust for degenerating singular nonlinearities.

In the light of these properties the presented method is a significant improvement over the truncated monotone multigrid (MMG) method proposed in [16] where mesh independence and robustness with respect to the nonlinearity were only achieved for a variant without convergence proof. While robustness with respect to N was not discussed systematically for MMG, it seems that the rate deteriorates when going from $N = 3$ to $N = 5$ which is in accordance with the overly pessimistic damping parameter approaching zero as $N \to \infty$. Regarding the implementation we note that TNNMG has $O(nN^2)$ complexity whereas this is a lower bound for MMG. However we are not aware of any possible MMG implementation better than $O(nN^3)$. Furthermore the coarse grid correction in TNNMG is much easier to implement than MMG as only a linear multigrid step is required.

References

[1] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, Mass., 2. edition, 2003.

[2] L. Blank, H. Garcke, L. Sarbu, T. Srisupattarawanit, V. Styles, and A. Voigt. Phase-field approaches to structural topology optimization. In *Constrained Optimization and Optimal Control for Partial Differential Equations*, pages 245–256. Springer, 2012.

[3] L. Blank, H. Garcke, L. Sarbu, and V. Styles. Nonlocal Allen–Cahn systems: analysis and a primal–dual active set method. *IMA J. Numer. Anal.*, 2013.

[4] L. Blank, H. Garcke, L. Sarbu, and V. Styles. Primal–dual active set methods for Allen–Cahn variational inequalities with non-local constraints. *Numer. Methods Partial Differential Equations*, 29(3):999–1030, 2013.

[5] J. F. Blowey and C. M. Elliott. The Cahn–Hilliard gradient theory for phase separation with non-smooth free energy part II: Numerical analysis. *European J. Appl. Math.*, 3:147–179, 1992.

[6] S. Bopyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2006.

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

[8] H. Garcke, B. Nestler, and B. Stoth. On anisotropic order parameter models for multi-phase systems and their sharp interface limits. *Physica D: Nonlinear Phenomena*, 115(1):87–108, 1998.

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

[10] C. Gräser, R. Kornhuber, and U. Sack. Time discretizations of anisotropic Allen–Cahn equations. *IMA J. Numer. Anal.*, 33(4):1226–1244, 2013.

[11] C. Gräser, R. Kornhuber, and U. Sack. Nonsmooth Schur–Newton methods for vector-valued Cahn–Hilliard equations. Preprint, Serie A 01/2013, Freie Universität Berlin, submitted.

[12] 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.

[13] C. Gräser and O. Sander. Polyhedral Gauß–Seidel converges. *J. Numer. Math.*, accepted.

[14] M. Hintermüller, K. Ito, and K. Kunisch. The primal–dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13(3):865–888, 2003.

[15] D. A. Kay and A. Tomasi. Color image segmentation by the vector-valued Allen–Cahn phase-field model: a multigrid solution. *IEEE Trans. Image Process.*, 18(10):2330–2339, 2009.

[16] 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.

[17] W. Wang and M. Á. Carreira-Perpiñán. Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. Technical Report arXiv:1309.1541, 2013.

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

Oliver Sander, RWTH Aachen, Institut für Geometrie und Praktische Mathematik, Templergraben 55, 52056 Aachen, Germany
E-mail address: sander@igpm.rwth-aachen.de