Truncated Nonsmooth Newton Multigrid Methods for Convex Minimization Problems

Carsten Gräser[12] *, Uli Sack[12], and Oliver Sander[12]

[1] Department of Mathematics, Freie Universität Berlin
[2] DFG Research Center MATHEON
{graeser|usack|sander}@math.fu-berlin.de

Summary. We present a new inexact nonsmooth Newton method for the solution of convex minimization problems with piecewise smooth, pointwise nonlinearities. The algorithm consists of a nonlinear smoothing step on the fine level and a linear coarse correction. Suitable postprocessing guarantees global convergence even in the case of a single multigrid step for each linear subproblem. Numerical examples show that the overall efficiency is comparable to multigrid for similar linear problems.

1 Introduction

We consider the minimization problem

$$u \in \mathbb{R}^n : \qquad J(u) \leq J(v) \qquad \forall v \in \mathbb{R}^n \tag{1}$$

where $J : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is given by

$$J(v) = \frac{1}{2} \langle Av, v \rangle - \langle b, v \rangle + \varphi(v), \qquad \varphi(v) = \sum\nolimits_{i=1}^{n} \varphi_i(v_i) \tag{2}$$

for a symmetric positive definite matrix $A \in \mathbb{R}^{n \times n}$ and convex, lower semicontinuous and proper functions $\varphi_i : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$. We will assume that each φ_i is C^2 on a finite number of disjoint intervals $I_i^k \subset \mathbb{R}$ having the property

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

Under the above assumptions J is strictly convex, lower semicontinuous, proper, and coercive. Thus (1) has a unique solution [5].

For quadratic obstacle problems the ideas of active-set methods and monotone multigrid have been combined recently to the Truncated Nonsmooth Newton Multigrid (TNNMG) method [6]. Inspired by [7], we generalize this method to nonquadratic nonsmooth energies (2) resulting in a novel globally

* This work has been funded in part by the DFG under contract Ko 1806/3-2

convergent multigrid method. While our approach is more flexible and significantly easier to implement than the algorithm in [7] the numerical examples indicate that it is comparable to linear multigrid for resonable initial iterates which can be obtained, e.g., by nested iteration.

2 A Nonsmooth Newton Method

Problem (1) can be equivalently formulated as the following inclusion

$$(A + \partial\varphi)(u) \ni b, \tag{3}$$

where the subdifferential $\partial\varphi$ of φ is the set-valued diagonal operator given by $(\partial\varphi(v))_i = \partial\varphi_i(v_i)$. Similar to the linear case $\varphi = 0$ the nonlinear Gauß-Seidel method $u^{k+1} = u^k + \mathcal{F}u^k$ defined by successive minimization of J in the coordinate directions can be represented by the operator

$$\mathcal{F}(v) = (D + L + \partial\varphi)^{-1}(b - Rv) - v,$$

where we have used the splitting $A = D + L + R$ in the diagonal, left, and right parts. Using a monotonicity argument it can be shown that the nonlinear Gauß-Seidel method converges globally to the solution of (1) [7]. Unfortunately, as in the linear case, the convergence rates deteriorate rapidly if A is a differential operator discretized on finer and finer grids.

It follows from the global convergence of the Gauß-Seidel method that the original problem (3) is equivalent to the fixed-point equation

$$\mathcal{F}(u) = 0 \tag{4}$$

for the operator $\mathcal{F}$ which is single-valued and Lipschitz continuous. This suggests to use a nonsmooth Newton approach for (4) which leads to methods

$$u^{k+1} = u^k - H(u^k)^{-1}\mathcal{F}(u^k) \tag{5}$$

where $H(u^k)$ is a generalized linearization of $\mathcal{F}$. In order to construct such $H(u^k)$ we first derive a linearization of $f_i = (A_{ii} + \partial\varphi_i)^{-1} : \mathbb{R} \to \mathbb{R}$. Since f_i is strictly monotone and Lipschitz continuous it is differentiable almost everywhere by Rademacher's theorem [9]. An element of the generalized Jacobian in the sense of Clarke [3] is given by

$$\partial f_i(x) = \begin{cases} 0 & \text{if } \partial\varphi_i(f_i(x)) \text{ is set-valued,} \\ (a_{ii} + \varphi_i''(f_i(x)))^{-1} & \text{else.} \end{cases} \tag{6}$$

For φ_i'' we use either the derivative from the left or from the right and $(a_{ii} + \varphi_i''(f_i(x)))^{-1}$ is set to zero if both one-sided derivatives tend to infinity.

Given an index set $\mathcal{J} \subset \{1, \dots, n\}$ and a matrix or vector ($n \times 1$ matrix) M we introduce the following notation for truncated versions of M

$$(M_{\mathcal{J}})_{ij} = \begin{cases} M_{ij} & \text{for } i \in \mathcal{J} \\ 0 & \text{else,} \end{cases} \qquad (M_{\mathcal{J},\mathcal{J}})_{ij} = \begin{cases} M_{ij} & \text{for } i,j \in \mathcal{J} \\ 0 & \text{else.} \end{cases}$$

Assuming a chain rule elementary computations lead to a linearization of $\mathcal{F}$ given by

$$\partial\mathcal{F}(v) = -\left(D + L + \varphi''(v+\mathcal{F}v)_{\mathcal{I}(v+\mathcal{F}v)}\right)^{-1} R_{\mathcal{I}(v+\mathcal{F}v)} - I \qquad (7)$$

with the index set of inactive components

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

Theorem 1. *If $H(u^k) = \partial\mathcal{F}(u^k)$ is used in a nonsmooth Newton step* (5) *the resulting iteration can be equivalently rewritten as the following two-step method*

$$u^{k+\frac{1}{2}} = u^k + \mathcal{F}(u^k), \qquad (8)$$

$$u^{k+1} = u^{k+\frac{1}{2}} + \mathcal{C}(u^{k+\frac{1}{2}}), \qquad (9)$$

with the linear correction

$$\mathcal{C}(v) = -\left(J''(v)_{\mathcal{I}(v),\mathcal{I}(v)}\right)^{-1} J'(v)_{\mathcal{I}(v)}. \qquad (10)$$

The proof of Theorem 1 is straightforward using the fact that $\mathcal{C}(u^{k+\frac{1}{2}})_i = 0$ for $i \notin \mathcal{I}(u^{k+\frac{1}{2}})$. For obstacle problems it can be found in [6].

Remark 1. By restriction to the $i \in \mathcal{I}(v)$ each φ_i in (9) has a classical first derivative. The second derivatives of φ_i are meant in the sense explained after (6).

Even though linearization of ∂J in (9) is restricted to locally smooth components the derivatives of φ_i might get very large leading to ill-conditioned linear systems and slow multigrid convergence. Therefore we may restrict the linearization further, e.g., to

$$\overline{\mathcal{I}}(v) = \left\{i \in \mathcal{I}(v) : |\varphi_i''(x) - \varphi_i''(y)| \le C|x-y| \;\forall x,y \in [v_i - \delta, v_i + \delta]\right\}$$

for a large constant C and a small δ.

Remark 2. Replacing $\mathcal{I}$ by some $\overline{\mathcal{I}} \subset \mathcal{I}$ leads to a truncated linearization $\overline{\partial}\mathcal{F}$ defined analogously to (7). Theorem 1 remains true for $H(u^k) = \overline{\partial}\mathcal{F}(u^k)$ if $\mathcal{I}$ is replaced by $\overline{\mathcal{I}}$.

Due to the leading nonlinear Gauß-Seidel step (8) global convergence can be shown if $J(u^{k+1}) \le J(u^{k+\frac{1}{2}})$ [7]. Thus the introduction of suitable damping parameters in (9) leads to global convergence. However, very small damping parameters slowing down the convergence may be necessary if

$u^{k+\frac{1}{2}} + \mathcal{C}(u^{k+\frac{1}{2}}) \notin \operatorname{dom}(J) = \{v : J(v) < \infty\}$. To overcome this problem we apply damping to a projected correction. If, additionally, we introduce inexact evaluation of $\mathcal{C}$ represented by the error ε^k the algorithm reads

$$u^{k+\frac{1}{2}} = u^k + \mathcal{F}(u^k), \tag{11}$$

$$u^{k+1} = u^{k+\frac{1}{2}} + \rho^k P^k(\mathcal{C}(u^{k+\frac{1}{2}}) + \varepsilon^k), \tag{12}$$

where P^k is the projection onto $\operatorname{dom}(J) - u^{k+\frac{1}{2}}$ and ρ^k is computed by the line search

$$\rho^k = \arg\min_{\rho\in\mathbb{R}} J(u^{k+\frac{1}{2}} + \rho P^k(\mathcal{C}(u^{k+\frac{1}{2}}) + \varepsilon^k)).$$

Since this algorithm satisfies $J(u^{k+1}) \leq J(u^{k+\frac{1}{2}})$ for arbitrary ε^k the following convergence result holds [7].

Theorem 2. *For every $u^0, \varepsilon^k \in \mathbb{R}^n$ the u^k converge to the solution u of* (3).

3 Multigrid

Now we consider the fast and inexact solution of the system (10) by multigrid methods. Since the matrix is symmetric and positive definite on the subspace

$$V^k = \{v \in \mathbb{R}^n : v_i = 0 \ \forall i \notin \overline{\mathcal{I}}(u^{k+\frac{1}{2}})\} \tag{13}$$

standard linear multigrid methods like successive or parallel subspace correction with standard transfer operators for problems in $\mathbb{R}^n$ can be applied if the following two modifications are introduced:

- If a diagonal element in a matrix is zero the subspace correction for the corresponding subspace should be zero as well. This may happen on all levels since the fine matrix has zero rows and columns.
- After the sum of all coarse corrections is prolongated to the fine space $\mathbb{R}^n$ all components $i \notin \mathcal{I}(u^{k+\frac{1}{2}})$ should be set to zero.

With these two modifications each correction in a subspace U of $\mathbb{R}^n$ is now naturally a correction in the Euclidean projection U^k of U onto V^k. Hence the subspace correction method automatically minimizes in suitable subspaces of V^k without explicit construction of these subspaces or their basis functions.

Since there is no need to solve the systems (10) to a certain accuracy applying a single multigrid step is enough to achieve global convergence. The resulting overall algorithm consists of nonlinear smoothing on the fine level and linear multigrid for a reduced linearization. As a generalization of the algorithm in [6] we call it *Truncated Nonsmooth Newton Multigrid* (TNNMG). The algorithm is open to various modifications, e.g.:

- Additional nonlinear smoothing before the linear correction
- Linear smoothing on the fine level can be omitted.
- Alternative smoothers can be applied to the linear correction.

4 Example I: Two-Body Contact in Linear Elasticity

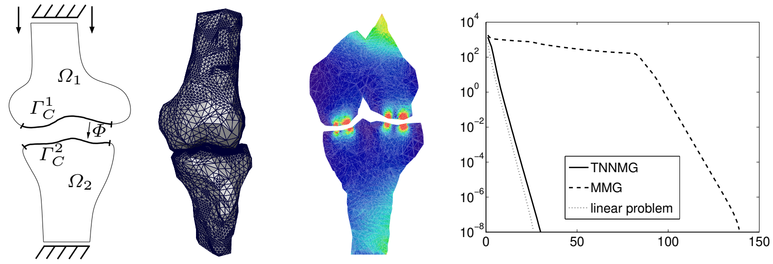

Fig. 1. Two-body contact problem. a) Schematic view; b) solution; c) vertical cut through the von-Mises stress field; d) error per number of iterations.

We will now show how the algorithm can be used to efficiently solve multi-body contact problems in linear elasticity. Consider two disjoint domains Ω_1, Ω_2 in $\mathbb{R}^d, d \in \{2,3\}$, discretized by simplicial grids. The boundary $\Gamma_i = \partial\Omega_i$, $i \in \{1,2\}$, of each domain is decomposed in three disjoint parts $\Gamma_i = \Gamma_{i,D} \cup \Gamma_{i,N} \cup \Gamma_{i,C}$. Let $\mathbf{f}_i \in (L_2(\Omega_i))^d$, $i \in \{1,2\}$, be body force density fields, and $\mathbf{t}_i \in (H^{-1/2}(\Gamma_{i,N}))^d$ be fields of surface traction. The two contact boundaries $\Gamma_{i,C}$ are identified using a homeomorphism $\Phi : \Gamma_{1,C} \to \Gamma_{2,C}$ and the initial distance function $g : \Gamma_{1,C} \to \mathbb{R}$, $g(x) = \|\Phi(x) - x\|$ is defined.

Let $\mathbf{V}_h$ be the space of first-order d-valued Lagrangian finite element functions on $\Omega_1 \cup \Omega_2$ and $\{\boldsymbol{\lambda}\}$ the nodal basis in $\mathbf{V}_h$. We denote the basis functions belonging to the n_C^i nodes of the contact boundaries $\Gamma_{i,C}$, $i \in \{1,2\}$ by $\{\boldsymbol{\lambda}_C^i\}$, the corresponding coefficients in a vector v by v_C^i and the basis functions for the remaining n^I nodes by $\{\boldsymbol{\lambda}_I\}$. The two-body contact problem can then be written as a minimization problem with a quadratic part as in (1) and $n = d(n^I + n_C^1 + n_C^2)$. Here A and b are the stiffness matrix and right-hand-side vector of linear elasticity, respectively. The nonlinearity is the characteristic functional $\varphi = \chi_{\mathcal{K}}$ of the mortar discretized admissible set

$$\mathcal{K} = \{v \in \mathbb{R}^{dn} \mid N\mathsf{D}v_C^1 - N\mathsf{M}v_C^2 \le \mathsf{g}\} \tag{14}$$

with a sparse mass matrix M, a diagonal mass matrix D, a matrix N which contains the domain normals, and the weak obstacle g. Contrary to (1) $\chi_{\mathcal{K}}$ does not have pointwise structure. To overcome this we introduce the transformed basis [11]

$$\{\widetilde{\boldsymbol{\lambda}}\} = OB\{\boldsymbol{\lambda}\} = \begin{pmatrix} I & 0 & 0 \\ 0 & O_C & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & (\mathsf{D}^{-1}\mathsf{M})^T & I \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_I \\ \boldsymbol{\lambda}_C^1 \\ \boldsymbol{\lambda}_C^2 \end{pmatrix}.$$

In this basis, we get a minimization problem of the form (1) with a matrix $\widetilde{A} = OBAB^TO^T$, a right-hand side $\tilde{b} = OBb$ and

$$\tilde{\varphi}_{p,0}(v_{p,0}) = \begin{cases} 0 & v_{p,0} \leq (\mathsf{D}^{-1}\mathsf{g})_p, \ p \text{ is vertex on } \Gamma_{1,C} \\ \infty & \text{else.} \end{cases} \tag{15}$$

The matrix O_C is block-diagonal. For each vertex p on $\Gamma_{1,C}$, the $d \times d$ diagonal entry $(O_C)_{pp}$ contains the Householder reflection which maps the first canonical basis vector of $\mathbb{R}^d$ onto the domain normal at p. Due to the pointwise structure (15), a projected block Gauß-Seidel scheme converges. For the coarse grid correction (10) we compute

$$J''(v)_{\mathcal{I}(v),\mathcal{I}(v)} = \widetilde{A}_{\mathcal{I}(v),\mathcal{I}(v)} \qquad \text{and} \qquad J'(v)_{\mathcal{I}(v)} = \left[\widetilde{A}v - \tilde{b}\right]_{\mathcal{I}(v)},$$

and apply one linear multigrid step to this.

Remark 3. For the transition from the finest to the second finest grid level the standard multigrid prolongation operator P is replaced by $\widetilde{P} = OB^{-1}P$. That way, discretization on the coarser levels is with respect to the nodal basis. Truncation and the transforming prolongation $\widetilde{P}$ can be combined in a single operator. This avoids having to store two fine-grid matrices.

As an example geometry we use the Visible Human data set [1]. We assume bone to be an isotropic, homogeneous, linear elastic material with $E = 17$ GPa and $\nu = 0.3$. The bottom section of the proximal tibia is clamped and a downward displacement of 6 mm is prescribed on the upper section of the femur (see Fig. 1, left). The implementation is based on the DUNE library [2].

We compare the numerical efficiencies of our solver and a monotone multigrid method (MMG), which is currently the fastest known globally convergent solver for two-body contact problems [8, 11]. It is well known that the MMG degenerates to a linear multigrid method once the active set has been found and hence shows linear multigrid convergence asymptotically.

We use nested iteration on two adaptive grids with 44777 vertices in total. Errors are computed by comparing with a precomputed reference solution $\mathbf{u}^*$. The error $e_i = \|\mathbf{u}_i - \mathbf{u}^*\|_A$ is plotted in Fig. 1. As expected, both the TNNMG and the MMG asymptotically show a linear multigrid convergence speed. However, the MMG needs more than 80 iterations to reach the asymptotic phase (see [10] for an explanation), whereas the TNNMG enters the asymptotic phase immediately. Note that iteration counts can be compared directly because both methods do a similar amount of work per iteration.

Remark 4. For two-body contact problems, the TNNMG is considerably easier to implement than the monotone multigrid method. See [10] for details.

5 Example II: The Allen-Cahn Equation

The Allen-Cahn equation is a well established diffuse interface model for phase transition phenomena as, e.g., solidification or crystallographic transformations. It can alternatively be interpreted as a regularization of the sharp interface geometric PDE for mean curvature flow (cf. [4]). The model considers an order parameter $u : \mathbb{R}^d \supset \Omega \to [-1,1]$ where the interval boundaries correspond to the pure phases, and is based on the Ginzburg-Landau energy

$$\mathcal{E}(u) = \int_\Omega \left(\frac{\varepsilon}{2} |\nabla u|^2 + \frac{1}{\varepsilon}\psi(u) \right) dx. \tag{16}$$

In the following, the potential ψ is taken to be

$$\psi(u) = \underbrace{\frac{1}{2}\theta\left((1+u)\ln\left(\frac{1+u}{2}\right) + (1-u)\ln\left(\frac{1-u}{2}\right)\right)}_{=:\,\phi_\theta(u)} + \frac{1}{2}\theta_c\left(1-u^2\right)$$

which for $\theta < \theta_c$ takes on the characteristic double-well shape. The Allen-Cahn equation

$$\varepsilon u_t = \varepsilon \Delta u - \varepsilon^{-1}\psi'(u) \tag{17}$$

results as the L^2-gradient flow of (16). For a given triangulation $\mathcal{T}$ of the domain Ω, which for simplicity we will assume to be a polygonal domain in $\mathbb{R}^2$, let $V_\mathcal{T}$ denote the space of continuous piecewise linear finite elements, $\mathcal{N}$ the set of nodes and $\{\lambda_p | \, p \in \mathcal{N}\} \subset V_\mathcal{T}$ the nodal basis. Time discretization by a semi-implicit Euler scheme and subsequent finite element discretization of (17) yields the variational problem

$$u_k \in V_\mathcal{T} : \quad a\left(u_k, v\right) - \ell_k(v) + \frac{\tau}{\varepsilon^2}(\phi'_\theta(u_k), v)_\mathcal{T} = 0 \quad \forall v \in V_\mathcal{T} \tag{18}$$

to be solved in the kth time step. Here $(\cdot,\cdot)_\mathcal{T}$ denotes the lumped L^2-product on $V_\mathcal{T}$. Furthermore

$$\ell_k(v) = \left(1 + \frac{\theta_c \tau}{\varepsilon^2}\right)(u_{k-1}, v)_\mathcal{T}, \quad a\left(v, w\right) = \tau(\nabla v, \nabla w) + (v, w)_\mathcal{T}$$

are a linear functional and a symmetric, positive definite bilinear form, respectively. Thus (18) is equivalent to the minimization problem in $V_\mathcal{T}$ for

$$J(v) = a(v,v) - \ell_k(v) + \frac{\tau}{\varepsilon^2}\sum_{p\in\mathcal{N}} \phi_\theta(v(p)) \int_\Omega \lambda_p(x)\, dx.$$

Identifying $V_\mathcal{T}$ with $\mathbb{R}^n$ we are now in the setting described in Section 1.

For the numerical example we choose $\Omega = [-1,1]^2$, $\varepsilon = 10^{-2}$, $\theta_c = 1$, $\tau = 10^{-4}$, and an initial value as shown in Fig. 2. We use one nonlinear smoothing step and a V(3,3)-cycle for the linearized system with nested iteration and compare the averaged convergence rates versus the temperature θ. Furthermore the error $e_j = \|u_1^j - u_1^*\|_a$ in the jth TNNMG-step of the first time step is considered where u_1^* is a precomputed reference solution.

As Figs. 2c+d indicate, the TNNMG-method exhibits very fast convergence and robustness wrt. θ which is remarkable as ϕ_θ is singular for $\theta \to 0$. Note that the convergence rates here never exceed 0.1. Experiments have shown that introducing additional nonlinear smoothing steps does not significantly accelerate convergence any further in this testcase whereas using less linear smoothing results in a considerable slowdown.

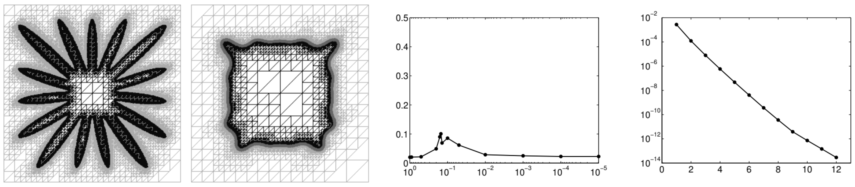

Fig. 2. a) Initial value; b) solution at time step 200; c) averaged convergence rates vs. temperature for TNNMG in the first time step; d) error vs. number of TNNMG iterations at $\theta = 0.15$ ($\sim$ 200.000 nodes).

References

[1] The Visible Human Project. `http://www.nlm.nih.gov/research/visible/visible_human.html`.

[2] P. Bastian, M. Blatt, A. Dedner, C. Engwer, R. Klöfkorn, R. Kornhuber, M. Ohlberger, and O. Sander. A generic interface for parallel and adaptive scientific computing. Part II: Implementation and tests in DUNE. *Computing*, accepted.

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

[4] K. Deckelnick, G. Dziuk, and C.M. Elliot. Computation of geometric partial differential equations and mean curvature flow. *Acta Numerica*, 14, 2005.

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

[6] C. Gräser and R. Kornhuber. Multigrid methods for obstacle problems. *J. Comp. Math.*, submitted.

[7] R. Kornhuber. On constrained Newton linearization and multigrid for variational inequalities. *Numer. Math.*, 91:699–721, 2002.

[8] R. Kornhuber, R. Krause, O. Sander, P. Deuflhard, and S. Ertel. A monotone multigrid solver for two body contact problems in biomechanics. *Comp. Vis. Sci*, 11:3–15, 2008.

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

[10] O. Sander. *Multidimensional Coupling in a Human Knee Model*. PhD thesis, Freie Universität Berlin, 2008.

[11] B. Wohlmuth and R. Krause. Monotone methods on nonmatching grids for nonlinear contact problems. *SIAM J. Sci. Comput.*, 25(1):324–347, 2003.