# Constrained Minimization and Multigrid

C. Gräser (FU Berlin), R. Kornhuber (FU Berlin), and O. Sander (FU Berlin)

Workshop on PDE Constrained Optimization

Hamburg, March 27 - 29, 2008

Freie Universität Berlin

Matheon

# Outline

Successive line search and multigrid (..., Yserentant 86, Xu 89, ...)

Self-adjoint elliptic obstacle problems

- projected multilevel relaxation (Mandel 84, Kh. 94, Badea 06)
- truncated monotone multigrid (Kh. 94)
- truncated nonsmooth Newton multigrid and active set strategies (Gräser & Kh. 08)

PDE constrained ($L^2$ or $H^1$) minimization with control constraints

- nonsmooth Schur-Newton (multigrid) methods (Gräser & Kh. 06)
- global convergence of primal-dual active set strategies (Hintermüller, Ito & Kunisch 03, Gräser 07)

# Outline

Successive line search and multigrid (..., Yserentant 86, Xu 89, ...)

Self-adjoint elliptic obstacle problems

- projected multilevel relaxation (Mandel 84, Kh. 94, Badea 06)
- truncated monotone multigrid (Kh. 94)
- truncated nonsmooth Newton multigrid and active set strategies (Gräser & Kh. 08)

PDE constrained ($L^2$ or $H^1$) minimization with control constraints

- nonsmooth Schur-Newton (multigrid) methods (Gräser & Kh. 06)
- global convergence of primal-dual active set strategies (Hintermüller, Ito & Kunisch 03, Gräser 07)

Message: Active Sets by Minimization

# Spectral Properties of Self-Adjoint Elliptic Bilinear Forms

quadratic minimization:

$$u \in H : \qquad \mathcal{J}(u) \leq \mathcal{J}(v) \qquad \forall v \in H$$

Poisson equation:

$$H = H_0^1(\Omega), \quad \mathcal{J}(v) = \tfrac{1}{2}a(v,v) - \ell(v), \quad a(v,w) = (\nabla v, \nabla w), \quad \ell(v) = (f, v)$$

Proposition:

eigenfunctions $e_l$, eigenvalues $\mu_l$ : $\qquad a(e_l, v) = \mu_l(e_l, v) \quad \forall v \in H, \quad l = 1, 2, \ldots$

$e_l$ $a$-orthogonal, $\qquad \mu_l \to \infty$ for $l \to \infty$

# Spectral Properties of Self-Adjoint Elliptic Bilinear Forms

quadratic minimization:

$$u \in H: \qquad \mathcal{J}(u) \leq \mathcal{J}(v) \qquad \forall v \in H$$

Poisson equation:

$$H = H_0^1(\Omega), \quad \mathcal{J}(v) = \tfrac{1}{2}a(v,v) - \ell(v), \quad a(v,w) = (\nabla v, \nabla w), \quad \ell(v) = (f,v)$$

Proposition:

eigenfunctions $e_l$, eigenvalues $\mu_l$ : $\qquad a(e_l, v) = \mu_l(e_l, v) \quad \forall v \in H, \quad l = 1, 2, \ldots$

$e_l$ $a$-orthogonal, $\qquad \mu_l \to \infty$ for $l \to \infty$

interpretation: $a$-orthogonal $e_l \quad \Longrightarrow \quad (e_l)$ scale of frequencies

# Spectral Properties of Self-Adjoint Elliptic Bilinear Forms

quadratic minimization:

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

Poisson equation:

$$H = H_0^1(\Omega), \quad \mathcal{J}(v) = \tfrac{1}{2}a(v,v) - \ell(v), \quad a(v,w) = (\nabla v, \nabla w), \quad \ell(v) = (f,v)$$

Proposition:

eigenfunctions $e_l$, eigenvalues $\mu_l$ : $\qquad a(e_l, v) = \mu_l(e_l, v) \quad \forall v \in H, \quad l = 1, 2, \dots$

$e_l$ $a$-orthogonal, $\qquad \mu_l \to \infty$ for $l \to \infty$

interpretation: $\qquad a$-orthogonal $e_l \quad \Longrightarrow \quad (e_l)$ scale of frequencies

basic idea of multigrid: "almost" $a$-orthogonal $\lambda_l \quad \Longleftarrow \quad (\lambda_l)$ scale of frequencies

# Discrete Quadratic Minimization

triangulations: $\mathcal{T}_0, \ldots, \mathcal{T}_j$ (successive refinement)

nodes: $\mathcal{N}_0 \subset \mathcal{N}_1 \cdots \subset \mathcal{N}_j$

nested finite element spaces: $\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_j$

nodal basis: $\mathcal{S}_k = \text{span}\{ \lambda_p^{(k)} \mid p \in \mathcal{N}_k \}$, dimension $|\mathcal{N}_k| = n_k$

Ritz-Galerkin discretization:

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

# Discrete Quadratic Minimization

triangulations: $\mathcal{T}_0, \ldots, \mathcal{T}_j$ (successive refinement)

nodes: $\mathcal{N}_0 \subset \mathcal{N}_1 \cdots \subset \mathcal{N}_j$

nested finite element spaces: $\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_j$

nodal basis: $\mathcal{S}_k = \text{span}\{ \lambda_p^{(k)} \mid p \in \mathcal{N}_k \}$, dimension $|\mathcal{N}_k| = n_k$

Ritz-Galerkin discretization:

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

goal: iterative solvers with $\mathcal{O}(n_j)$ complexity and mesh-independent convergence rates

# Successive Subspace Correction (. . . , Yserentant 86, Xu 92, . . . )

select search directions: $\lambda_l \subset \mathcal{S}_j, \quad l = 1, \dots, m,$ subspaces: $V_l =$ span $\{\lambda_l\}$

Algorithm (successive line search)

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

for $l = 1, \dots, m$:

{

solve: $v_l \in V_l : \quad \mathcal{J}(w_{l-1} + v_l) \leq \mathcal{J}(w_{l-1} + v) \quad \forall v \in V_l$

update: $w_l := w_{l-1} + v_l$

}

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

# Successive Subspace Correction (. . . , Yserentant 86, Xu 92, . . . )

select search directions: $\lambda_l \subset \mathcal{S}_j, \;\; l = 1, \dots, m,$ subspaces: $V_l =$ span $\{\lambda_l\}$

Algorithm (successive line search)

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

for $l = 1, \dots, m$:

{

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

update: $w_l := w_{l-1} + v_l$

}

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

$a$-orthogonal search directions $(\lambda_l)$ $\Longrightarrow$ exact solver!

# Subspace Correction and Multigrid

Gauß–Seidel relaxation: select $\lambda_{l(p)} := \lambda_p^{(j)}, \quad p \in \mathcal{N}_j$

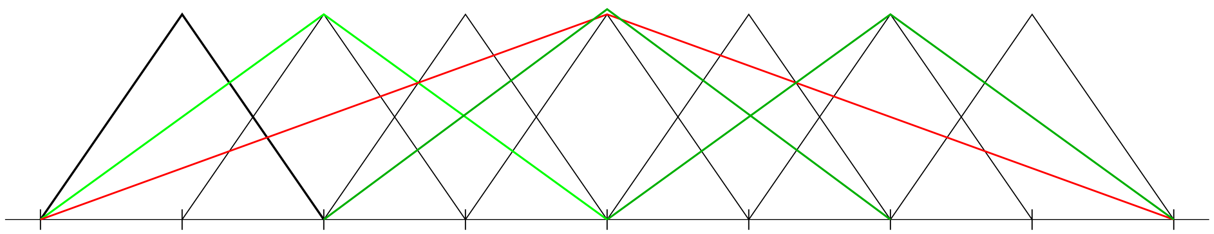
Theorem: exponential decay of convergence rates: $\rho_j \leq 1 - O(n_j^{-2})$

# Subspace Correction and Multigrid

Gauß–Seidel relaxation: select $\lambda_{l(p)} := \lambda_p^{(j)}, \quad p \in \mathcal{N}_j$

Theorem: exponential decay of convergence rates: $\rho_j \leq 1 - O(n_j^{-2})$

multilevel relaxation: select $\lambda_{l(p,k)} := \lambda_p^{(k)}, \quad p \in \mathcal{N}_k,\ k = 0, \ldots, j$



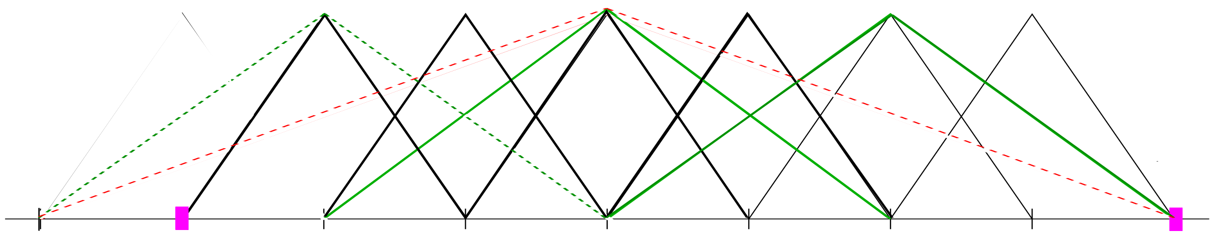
optimal complexity: $\mathcal{O}(n_j)$ (classical multigrid $V(1,0)$ cycle, Gauß–Seidel smoother)

Theorem: (..., Yserentant 86, Oswald 91, Bramble,Pasciak,Wang & Xu 91, Dahmen & Kunoth 92, Bornemann & Yserentant 93, ......)

mesh-independent convergence rates: $\rho_j \leq \rho < 1$ $\quad$ ($(\lambda_l)$ scale of frequencies)

# Multigrid on Complicated Domains (Kh. & Yserentant 94)

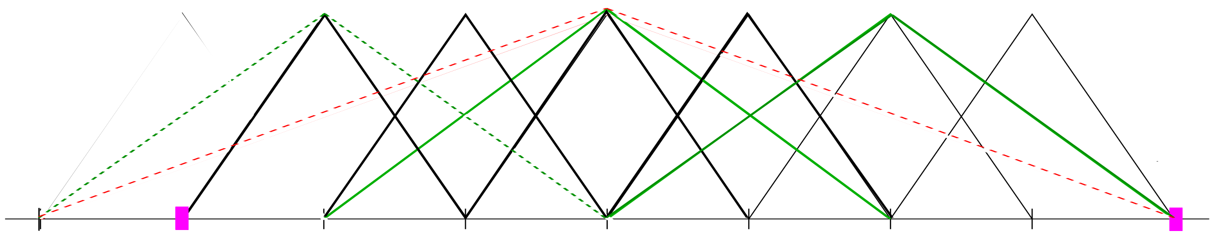
computational domain $\Omega$ not resolved by coarse grid $\mathcal{T}_k$ for $k < j$



difficulty: reduced coarse-grid correction $\Longrightarrow$ reduced convergence speed

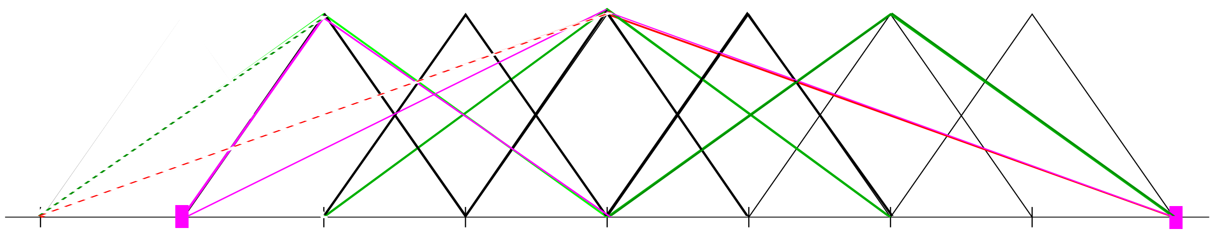
# Multigrid on Complicated Domains (Kh. & Yserentant 94)

computational domain $\Omega$ not resolved by coarse grid $\mathcal{T}_k$ for $k < j$



difficulty: reduced coarse-grid correction $\Longrightarrow$ reduced convergence speed

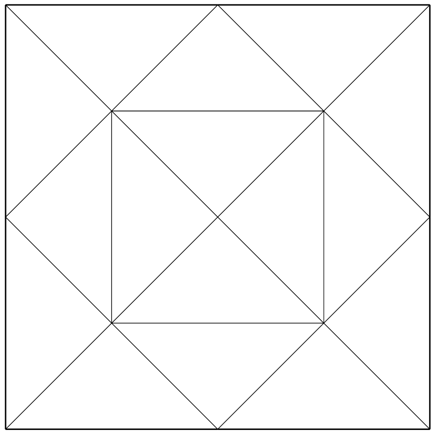
remedy: truncated search directions $\tilde{\lambda}_p^{(k)}$ (interpolation of $\lambda_p^{(k)}$ to $\Omega$)



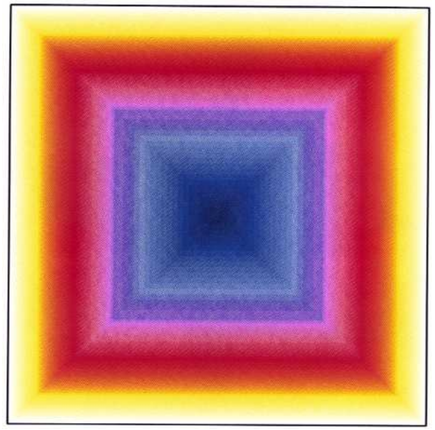
hope: preserved coarse-grid correction $\Longrightarrow$ preserved convergence speed

# A Fractal Domain

$f \equiv 1$, $V(1,0)$ cycle, symmetric Gauß-Seidel smoothing



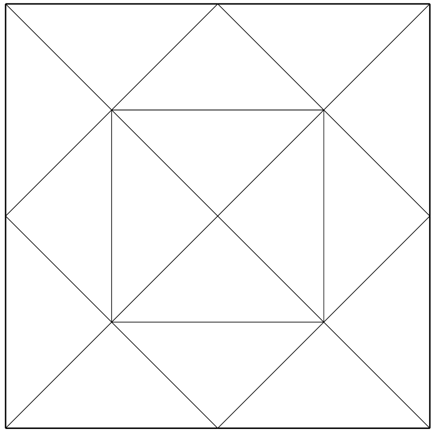
Initial triangulation $\mathcal{T}_0$



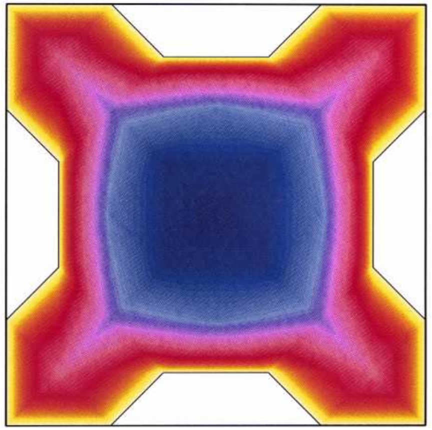
Approximate solution on $\Omega_0^\circ$

# A Fractal Domain

$f \equiv 1$, $V(1,0)$ cycle, symmetric Gauß-Seidel smoothing



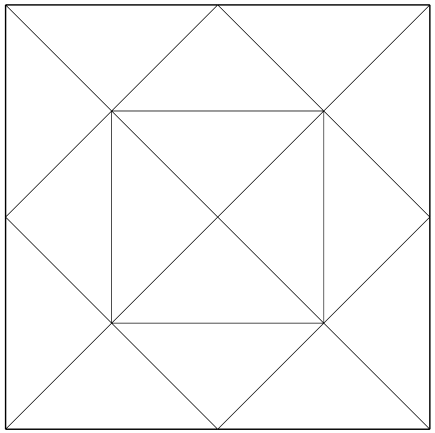
Initial triangulation $\mathcal{T}_0$



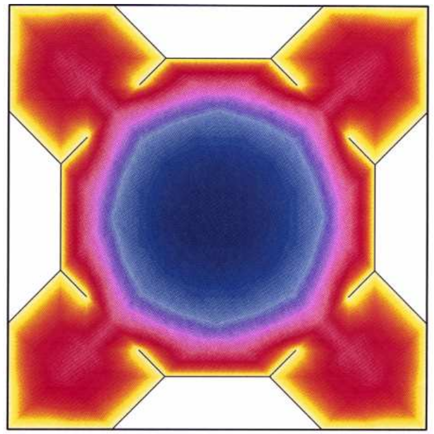
Approximate solution on $\Omega_1^\circ$

# A Fractal Domain

$f \equiv 1$, $V(1,0)$ cycle, symmetric Gauß-Seidel smoothing



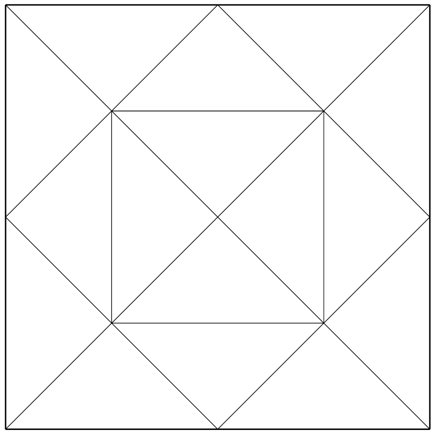
Initial triangulation $\mathcal{T}_0$



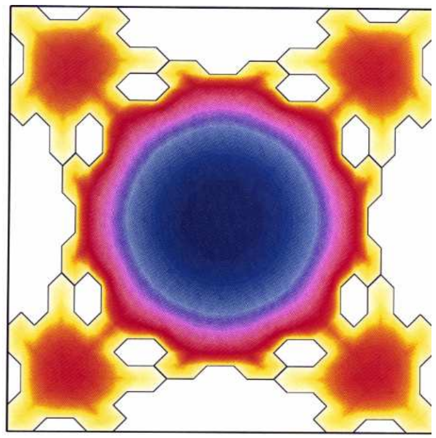
Approximate solution on $\Omega_2^\circ$

# A Fractal Domain

$f \equiv 1$, $V(1,0)$ cycle, symmetric Gauß-Seidel smoothing



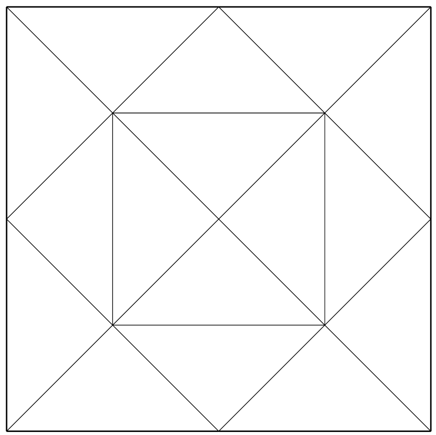
Initial triangulation $\mathcal{T}_0$



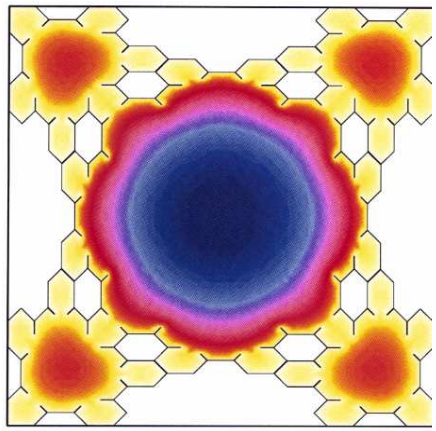
Approximate solution on $\Omega_3^\circ$

# A Fractal Domain

$f \equiv 1$, $V(1,0)$ cycle, symmetric Gauß-Seidel smoothing



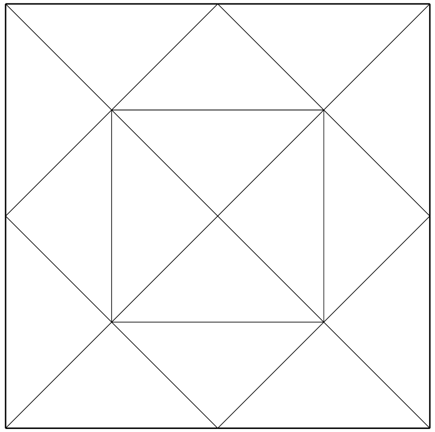
Initial triangulation $\mathcal{T}_0$



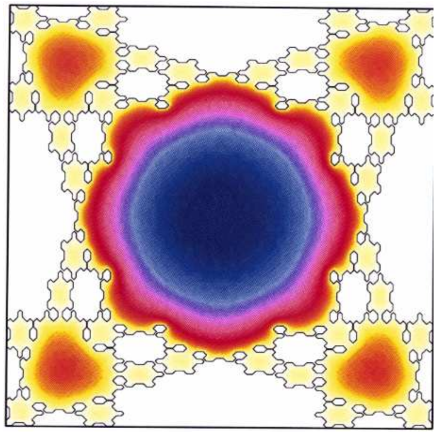
Approximate solution on $\Omega_4^\circ$

# A Fractal Domain

$f \equiv 1$, $V(1,0)$ cycle, symmetric Gauß-Seidel smoothing



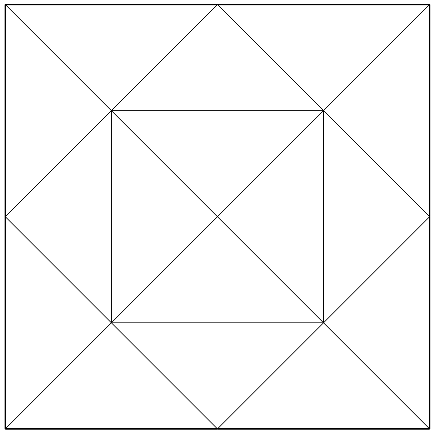
Initial triangulation $\mathcal{T}_0$



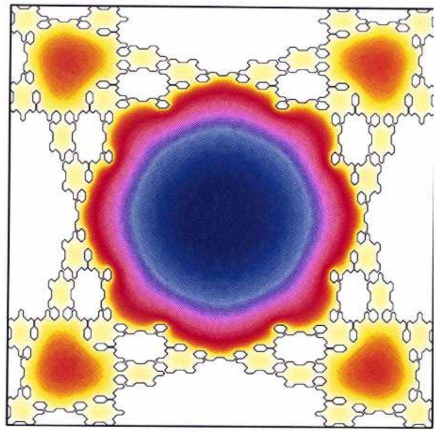
Approximate solution on $\Omega_5^\circ$

# A Fractal Domain

$f \equiv 1$, $V(1,0)$ cycle, symmetric Gauß-Seidel smoothing



Initial triangulation $\mathcal{T}_0$



Approximate solution on $\Omega_5^\circ$

convergence rates:

| | j=2 | j=3 | j=4 | j=5 | j=6 | j=7 |
|---|---|---|---|---|---|---|
| std | 0.30 | 0.61 | 0.58 | 0.68 | 0.65 | 0.71 |
| trc | 0.28 | 0.30 | 0.32 | 0.32 | 0.33 | 0.34 |
| $(0,1)^2$ | 0.28 | 0.30 | 0.31 | 0.32 | 0.33 | 0.33 |

# Self-Adjoint Elliptic Obstacle Problems

constrained minimization:

$$u_j \in \mathcal{K}_j : \quad \mathcal{J}(u_j) \leq \mathcal{J}(v) \quad \forall v \in \mathcal{K}_j = \{v \in \mathcal{S}_j \mid v(p) \in [\alpha(p), \beta(p)] \ \ \forall p \in \mathcal{N}_j\}$$

Algorithm (successive line search)

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

for $l = 1, \ldots, m$:

{

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

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

update: $\quad w_l := w_{l-1} + v_l$

}

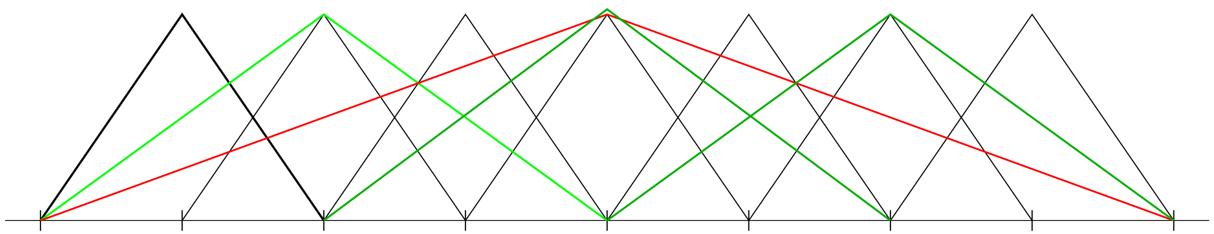
new iterate: $\quad u^{\nu+1} := w_m = u^\nu + \sum_{l=1}^{m} v_l$

# Projected Multilevel Relaxation PMLR

projected Gauß–Seidel relaxation: select $\lambda_{l(p)} := \lambda_p^{(j)}, \quad p \in \mathcal{N}_j$

Theorem: exponential decay of convergence rates: $\rho_j \leq 1 - O(n_j^{-2})$

projected multilevel relaxation: select $\lambda_{l(p,k)} := \lambda_p^{(k)}, \quad p \in \mathcal{N}_k, \; k = 0, \ldots, j$



suboptimal complexity: $\mathcal{O}(n_j \log(n_j))$ (additional prolongations to check $v_l \in D_l$)

Theorem: (..., Badea, Tai and Wang 03, Badea 06)

polylogarithmic convergence rates in 2D: $\rho_j \leq 1 - \mathcal{O}(j+1)^{-5}$

# Monotone Iterations (Kh. 94)

two-stage selection:

a) $\lambda_l = \lambda_{p_l}^{(j)}, \quad l = 1, \ldots, n_j$

b) arbitrary $\lambda_l, \quad l = n_j + 1, \ldots, m$

two-stage (energy-)monotone iteration:

a) leading projected Gauß–Seidel step: $u_j^\nu \to \overline{u}_j^\nu = GS(u_j^\nu)$

b) (energy-) monotone "coarse-grid correction"

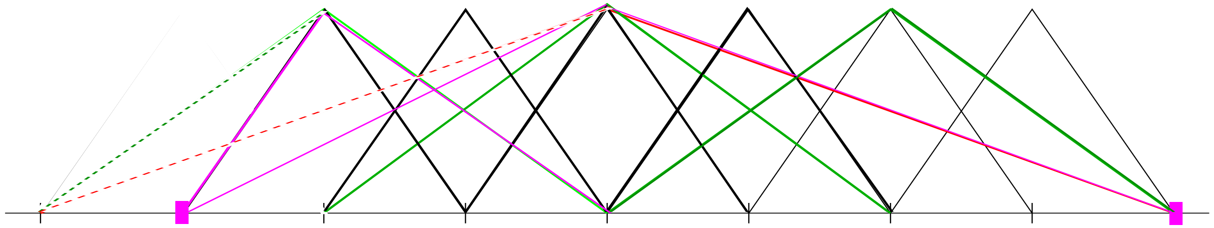
Theorem: Monotone iterations converge globally

goal: improve convergence speed of PMLR by adapting $\lambda_p^{(k)}$ to the active set $\mathcal{N}_j^\bullet$

# Truncated Monotone Multigrid TMMG (Kh. 94)

complicated reduced computational domain: active set $\mathcal{N}_j^\bullet$ not resolved by $\mathcal{T}_k$

a) use truncated search directions $\tilde{\lambda}_p^{(k)}$ (interpolation of $\lambda_p^{(k)}$ to $\mathcal{N}_j^\bullet$)



b) replace unknown active set $\mathcal{N}_j^\bullet$ by its actual approximation $\mathcal{N}_j^\bullet(\overline{u}_j)$

optimal complexity: $\mathcal{O}(n_j)$ by monotone restriction of defect constraints

Theorem:

- global convergence
- asymptotic polylogarithmic convergence rates $\rho_j \leq 1 - \mathcal{O}(1+j)^{-2}$
- no global mesh-independent convergence rates (next-neighbor inactivation by projected Gauß–Seidel $\overline{u}_j^\nu = GS(u_j^\nu)$)

## Truncated Nonsmooth Newton Multigrid TNMG (Gräser & Kh. 08)

modifications of TMMG to obtain TNMG

b) ignore defect constraints ($v_l \in V_l$ instead of $v_l \in D_l = (-w_{l-1} + \mathcal{K}_j) \cap V_l$)

c) enforce monotonicity of coarse-grid correction by projection and damping

## Truncated Nonsmooth Newton Multigrid TNMG (Gräser & Kh. 08)

modifications of TMMG to obtain TNMG

b) ignore defect constraints ($v_l \in V_l$ instead of $v_l \in D_l = (-w_{l-1} + \mathcal{K}_j) \cap V_l$)

c) enforce monotonicity of coarse-grid correction by projection and damping

alternative derivation of TNMG:

a) nonsmooth Newton linearization of $GS(u) - u = 0$ at $u_j^\nu$

b) linear truncated multigrid step for the resulting linear system

c) projection and damping

Theorem:

- global convergence
- asymptotic polylogarithmic convergence rates $\rho_j \leq 1 - \mathcal{O}(1+j)^{-2}$
- no global mesh-independent convergence rates
  (next-neighbor inactivation by projected Gauß–Seidel $\overline{u}_j^\nu = GS(u_j^\nu)$)

# Multigrid versus Primal–Dual Active Set (Hintermüller, Ito & Kunisch 03)

| | primal–dual | TNMG | TMMG |
|---|---|---|---|
| inactivation | proj. Jacobi | proj. Gauß–Seidel | proj. Gauß–Seidel |
| linear solve | exact | linear MG step | loc. proj. MG step |
| convergence | maximum principle | monotonicity | monotonicity |

common lack of robustness:

no global mesh-independent convergence rates due to local inactivation

hybrid approach HMG:

a) global inactivation by PMLR (not by $GS$!): $\overline{u_j}^\nu = \mathrm{PMLR}(u_j^\nu)$

b) linear truncated multigrid step

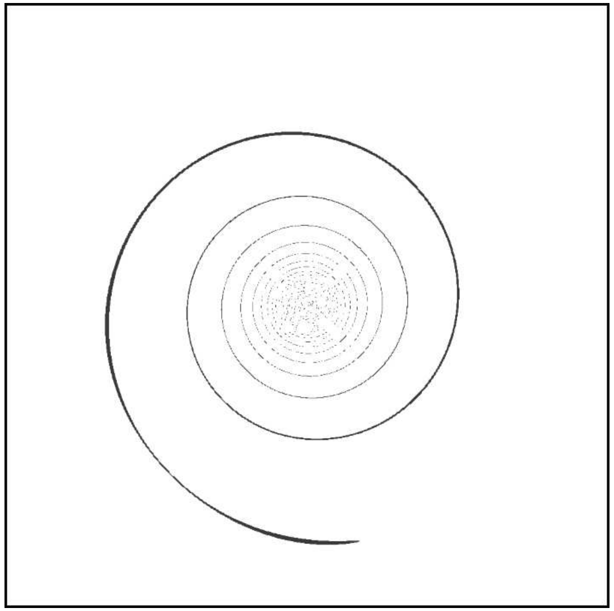
c) projection and damping

Theorem: (Gräser & Kh. 08) global convergence, asymptotic polylogarithmic convergence

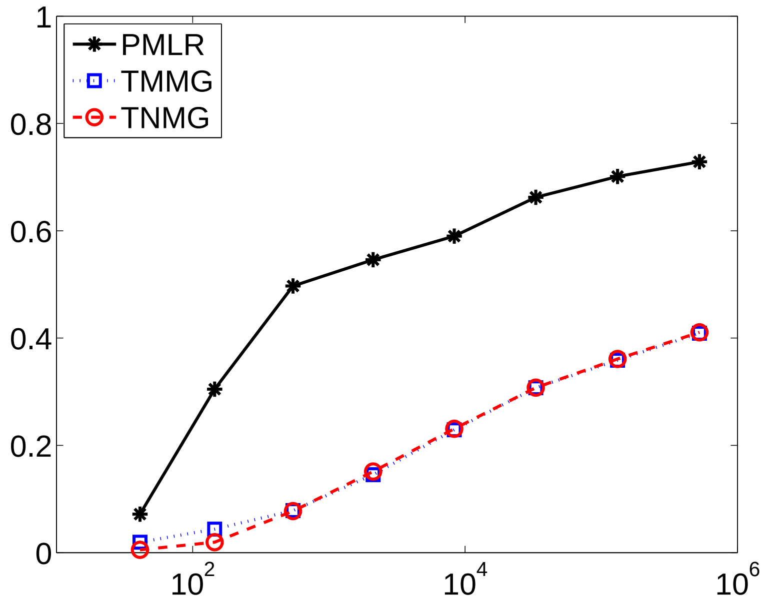
# Spiral Active Set (Gräser & Kh 08)

lower obstacle: $u_j \geq \varphi_j$

uniform refinement: $j = 9$ with $523\,265$ unknowns

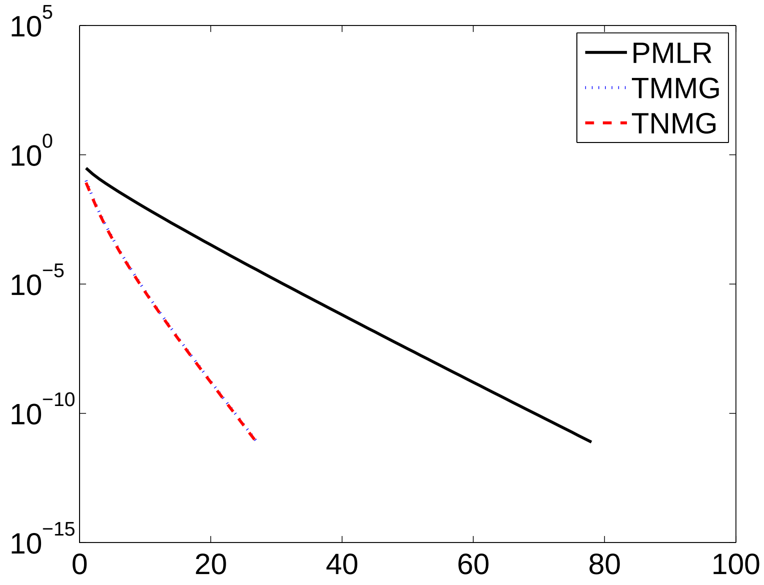


active set $\mathcal{N}_j^\bullet$

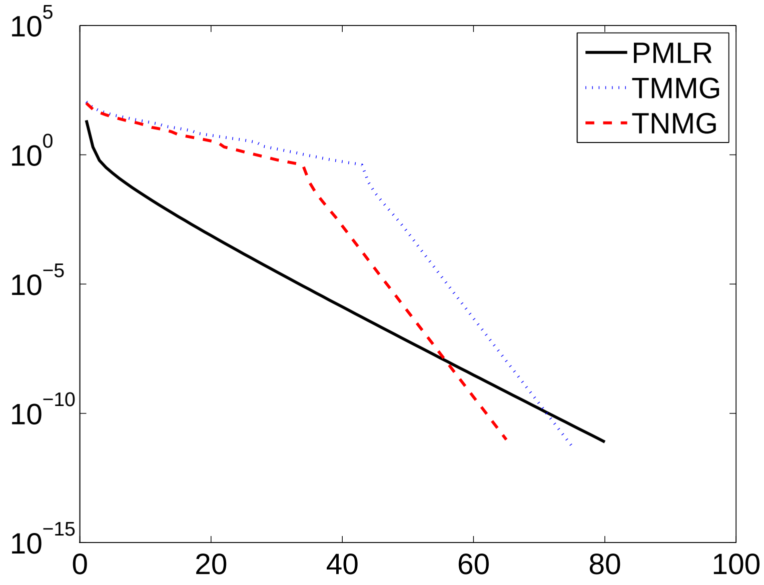



asymptotic convergence rates

# Iteration Histories




nested iteration

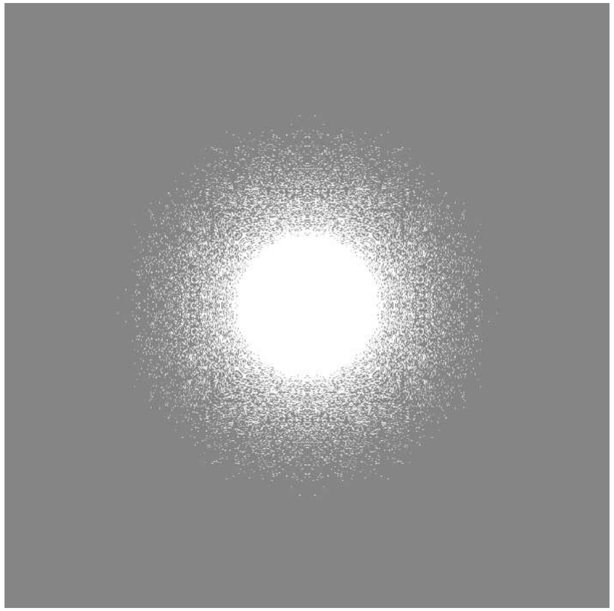



starting from the obstacle

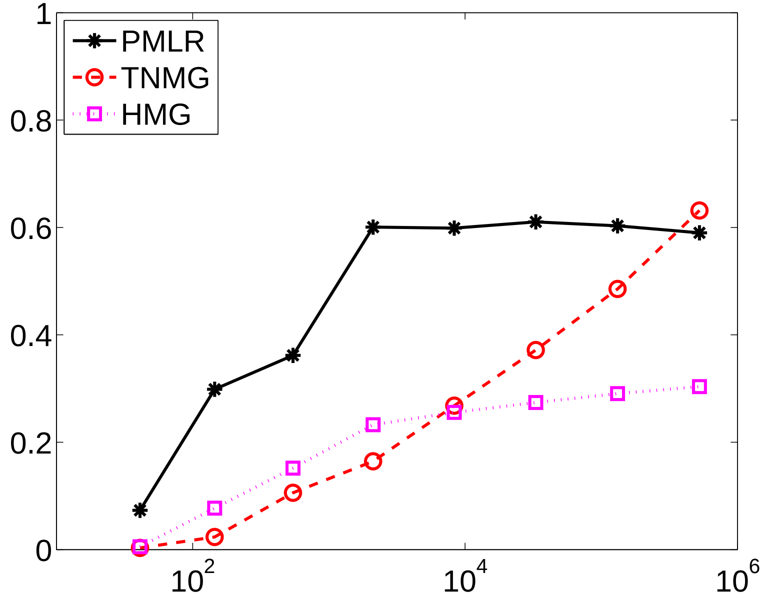
# Degenerate Problem (Gräser & Kh 08)

lower obstacle: $u_j \geq \varphi_j = I_{\mathcal{S}_j}\varphi, \ -\Delta\varphi = f$

uniform refinement: $j = 9$ with $523\,265$ unknowns

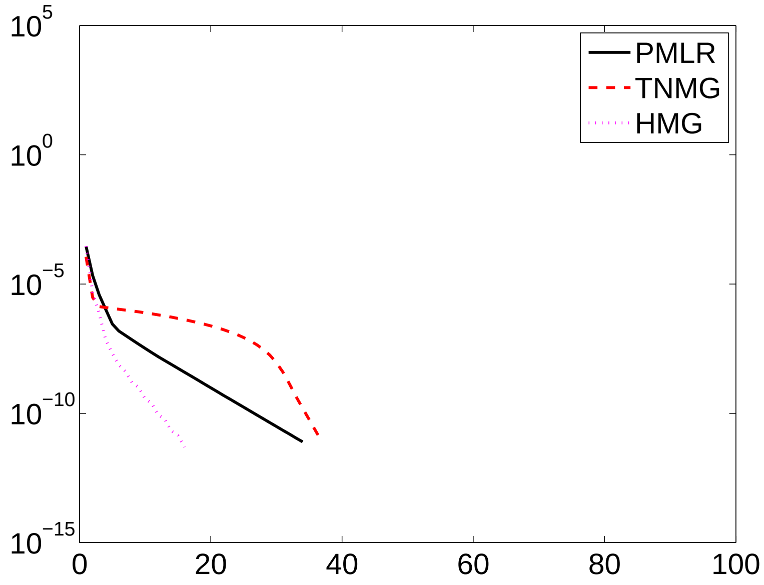


active set $\mathcal{N}_j^\bullet$

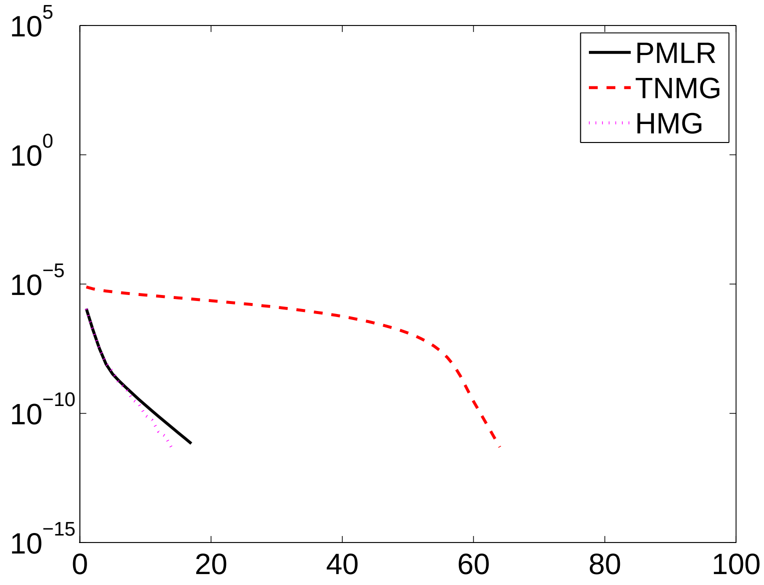



asymptotic convergence rates

# Iteration Histories

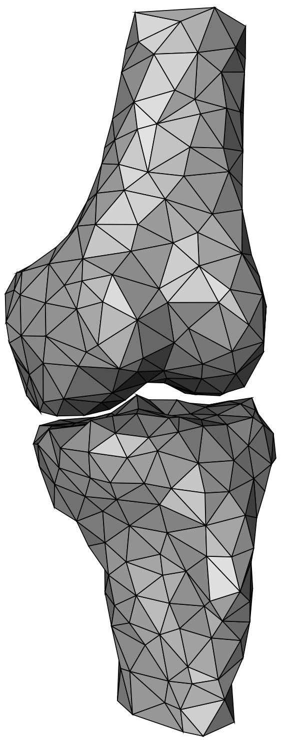



nested iteration

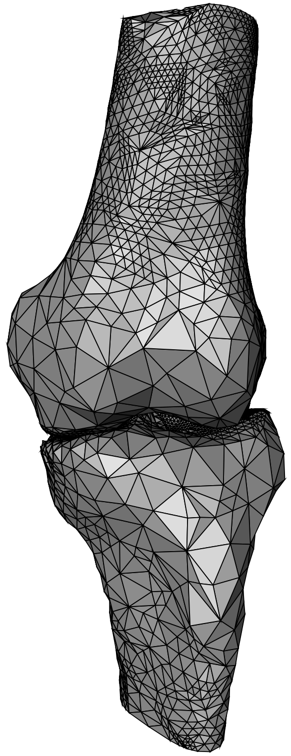



starting from the obstacle

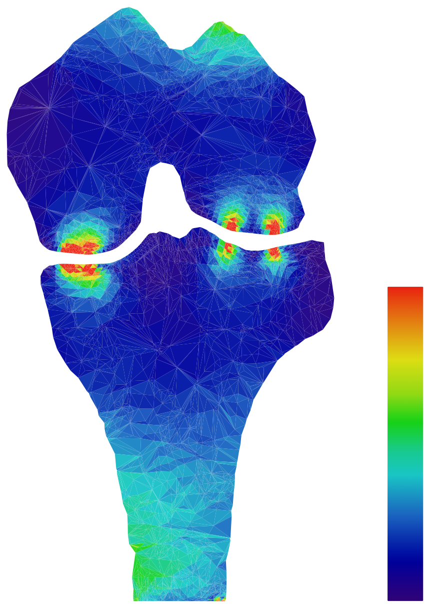
# Two-Body Contact in Linear Elasticity Kh. & Krause 01, Wohlmuth & Krause 03, Sander 08



coarse mesh (2 372 dof)



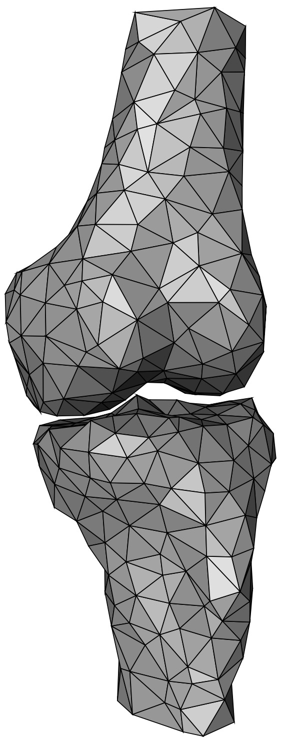
refined mesh (82 980 dof)



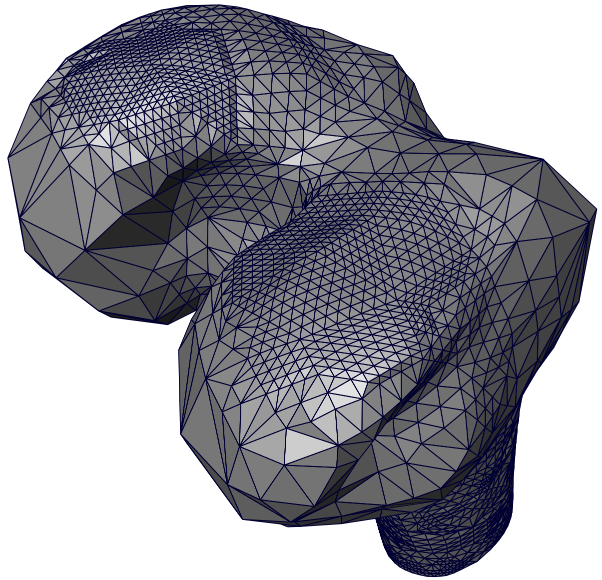
stress (planar cut)

implemented in DUNE

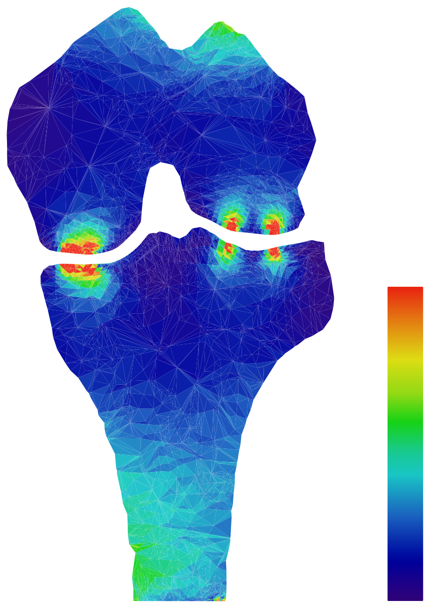
# Two-Body Contact in Linear Elasticity Kh. & Krause 01, Wohlmuth & Krause 03, Sander 08



coarse mesh (2 372 dof)



refined mesh (82 980 dof)



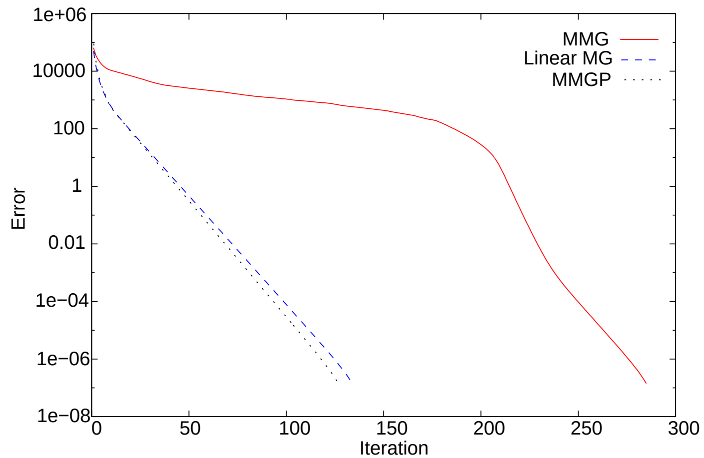
stress (planar cut)

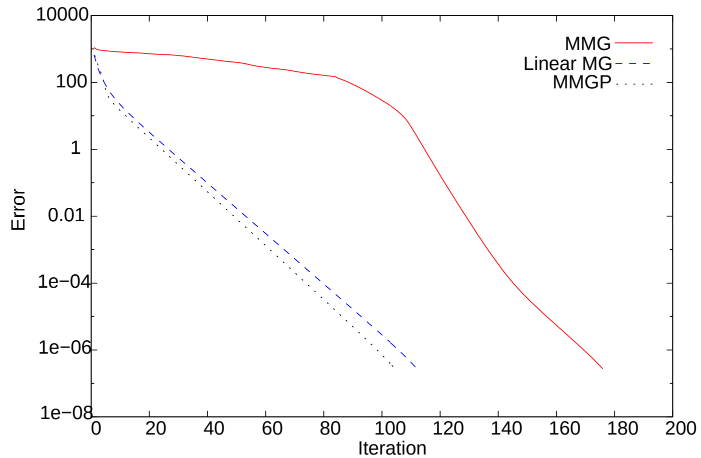
implemented in DUNE

# Comparison with Linear Multigrid

coarse-grid solver: interior point method IPOPT Waechter & Biegler 04

adaptive refinement: hierarchical error estimate, $j = 4$




initial iterate: $u_j^0 = 0$ nested iteration

linear multigrid convergence rates ($\rho = 0.56$)

# PDE-Constrained Minimization with Control Constraints

minimize $\quad \mathcal{J}(y,u) = \frac{1}{2}a_1(y,y) + \frac{1}{2}a_2(u,u) - \ell(y,u)$

pde constraints $\quad (\nabla y, \nabla v) = \langle u, v\rangle \quad \forall v \in \mathcal{S}_j$

box constraints: $\quad u \in \mathcal{K}_j = \{u \in \mathcal{S}_j \mid |u| \leq 1\}$

$L^2$ control problem:

$$a_1(y,v) = \langle y, v\rangle, \quad a_2(u,v) = \varepsilon\langle u, v\rangle$$ (lumped $L^2$-scalar product: diagonal)

spatial Cahn-Hilliard problem (deep-quench limit):

$$a_1(y,z) = \tau(\nabla y, \nabla z) + \langle y, 1\rangle\langle z, 1\rangle, \quad a_2(u,v) = \gamma(\nabla u, \nabla v) + \langle u, 1\rangle\langle v, 1\rangle$$

# Set-Valued Saddle-Point Problems

$$\begin{pmatrix} F & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} w \\ \lambda \end{pmatrix} \ni \begin{pmatrix} f \\ g \end{pmatrix}$$

$w = (\underline{y}, \underline{u})^T$, $\quad F = A + \partial \chi_{\mathbb{R}^{n_1} \times K_j}$

$A \in \mathbb{R}^{n,n}$ s.p.d., $\quad B \in \mathbb{R}^{m,n}$, $\quad C \in \mathbb{R}^{m,m}$ symmetric, positive semi-definite

generalization:
$F = \partial\varphi$, $\quad \varphi : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$
strictly convex, l.s.c., proper, coercive: $F^{-1}$ single valued, Lipschitz

# Non-linear Schur Complement

$$H(\lambda) = 0\,, \qquad H(\lambda) = -BF^{-1}(f - B^T\lambda) + C\lambda + g$$

Proposition

Let $\varphi^* : \mathbb{R}^n \mapsto \mathbb{R}$ denote the polar functional of $\varphi$ with $F = \partial\varphi$

- $\mathcal{J}(\lambda) = \varphi^*(f - B^T\lambda) + \frac{1}{2}(C\lambda, \lambda) + (g, \lambda)$ is Fréchet–differentiable
- $H = \nabla\mathcal{J}$
- $H(\lambda) = 0$ is equivalent to unconstrained (!) convex minimization

$$\lambda \in \mathbb{R}^m : \qquad \mathcal{J}(\lambda) \leq \mathcal{J}(v) \qquad \forall v \in \mathbb{R}^m$$

## Gradient-Related Descent Methods (/. . . , Ortega & Rheinboldt 70, . . . )

$$\lambda^{\nu+1} = \lambda^{\nu} + \rho_\nu d_\nu \;, \qquad d_\nu = -H_\nu^{-1} \nabla \mathcal{J}(\lambda^\nu) \;, \quad H_\nu \text{ s.p.d.}$$

Assumption on $d_\nu$: $\gamma \|v\|^2 \le (H_\nu v, v) \le \Gamma \|v\|^2 \qquad \forall \nu \in \mathbb{N}$

Assumption on $\rho_\nu$: $\mathcal{J}(\lambda^\nu + \rho_\nu d_\nu) \le \mathcal{J}(\lambda^\nu) - c(\nabla \mathcal{J}(\lambda^\nu), d_\nu)^2 / \|d_\nu\|^2$

Theorem: The iteration is globally convergent.

# Damping Strategies

Armijo's strategy: select $\sigma$, $\beta \in (0,1)$

accept $\rho$, if $\mathcal{J}(\lambda^\nu + \rho d_\nu) \leq \mathcal{J}(\lambda^\nu) + \sigma\rho(\nabla\mathcal{J}((\lambda^\nu, d_\nu))$, else set $\rho := \beta\rho$ and try again

drawback: evaluation of $\mathcal{J}$ is expensive

inexact damping:

approximate the solution of $(H(\lambda^\nu + \rho d_\nu), d_\nu) = 0$ by bisection

computational cost: evaluation of $F^{-1}$ in each bisection step

numerical experiments: usually: 1 step, exceptions: up to 8 steps

both strategies satisfy the assumption on $\rho$

# Inexact Evaluation of $H_\nu^{-1}$

Proposition:

Let $\tilde{d}_\nu \approx d_\nu = -H_\nu^{-1} \nabla \mathcal{J}(\lambda^\nu)$.

Then the accuracy conditions

$$\|d_\nu - \tilde{d}_\nu\| \leq \frac{1}{\nu} \|d_\nu\| \; , \qquad (H(\lambda^\nu), \tilde{d}_\nu) < 0 \qquad \forall \nu \in \mathbb{N}$$

preserve convergence.

# Selection of $H_\nu$: Nonsmooth Schur-Newton Methods

Newton-like methods: $\lambda^{\nu+1} = \lambda^\nu - \rho_\nu H_\nu^{-1} H(\lambda^\nu)$

Newton methods: $H_\nu = H'(\lambda^\nu)$

non-smooth Newton methods: $H_\nu \in \partial_B H(\lambda^\nu) \subset \partial H(\lambda^\nu)$

Proposition:
Let rank $B = n$ and $w^\nu = F^{-1}(f - B^T)\lambda^\nu$. Then

$$H_\nu := S(w^\nu) = B\hat{A}^{-1}B^T + C \in \partial H(\lambda^\nu)$$

where $\hat{A} = TAT + I - T$, with $T_{ii} = 0$, if $i \in N^\bullet(w^\nu)$ else $T_{ii} = 1$

# Selection of $H_\nu$: Nonsmooth Schur-Newton Methods

Newton-like methods: $\lambda^{\nu+1} = \lambda^\nu - \rho_\nu H_\nu^{-1} H(\lambda^\nu)$

Newton methods: $H_\nu = H'(\lambda^\nu)$

non-smooth Newton methods: $H_\nu \in \partial_B H(\lambda^\nu) \subset \partial H(\lambda^\nu)$

Proposition:
Let rank $B = n$ and $w^\nu = F^{-1}(f - B^T)\lambda^\nu$. Then

$$H_\nu := S(w^\nu) = B\hat{A}^{-1}B^T + C \in \partial H(\lambda^\nu)$$

where $\hat{A} = TAT + I - T$, with $T_{ii} = 0$, if $i \in N^\bullet(w^\nu)$ else $T_{ii} = 1$

$L^2$-control: Newton-like (rank $B = m < n$) Cahn-Hilliard: Newton (rank $B = n$)

# Convergence Results for Nonsmooth Schur-Newton Methods

Theorem:

- global convergence for appropriate damping
- finite termination for non-degenerate problems (locally superlinear, quadratic)
- inexact versions converge globally, if
  - obstacle problem: computation of $\mathcal{N}_j^{\bullet}(w^\nu)$
  - linear saddle point problem: approximation up to sufficient accuracy
- linear convergence, if $S(w)$ is s.p.d. for all $w \in \mathbb{R}^n$

# Convergence Results for Nonsmooth Schur-Newton Methods

Theorem:

- global convergence for appropriate damping
- finite termination for non-degenerate problems (locally superlinear, quadratic)
- inexact versions converge globally, if
  - obstacle problem: computation of $\mathcal{N}_j^{\bullet}(w^\nu)$
  - linear saddle point problem: approximation up to sufficient accuracy
- linear convergence, if $S(w)$ is s.p.d. for all $w \in \mathbb{R}^n$

$L^2$-control: global linear convergence

Cahn-Hilliard: global convergence

# Interpretations

preconditioned Uzawa iteration:

- evaluation of $F^{-1}$: obstacle problem (up to $\mathcal{N}_j^\bullet(w^\nu)$)
  exact ($L^2$-control), multigrid or projected Gauß-Seidel (Cahn-Hilliard)

- (inexact) evaluation of $S(w^\nu)^{-1}$: linear saddle point problem
  exact, multigrid (Vanka 86, Zulehner & Schöberl 03)

# Interpretations

preconditioned Uzawa iteration:

- evaluation of $F^{-1}$: obstacle problem (up to $\mathcal{N}_j^\bullet(w^\nu)$)
  exact ($L^2$-control), multigrid or projected Gauß-Seidel (Cahn-Hilliard)

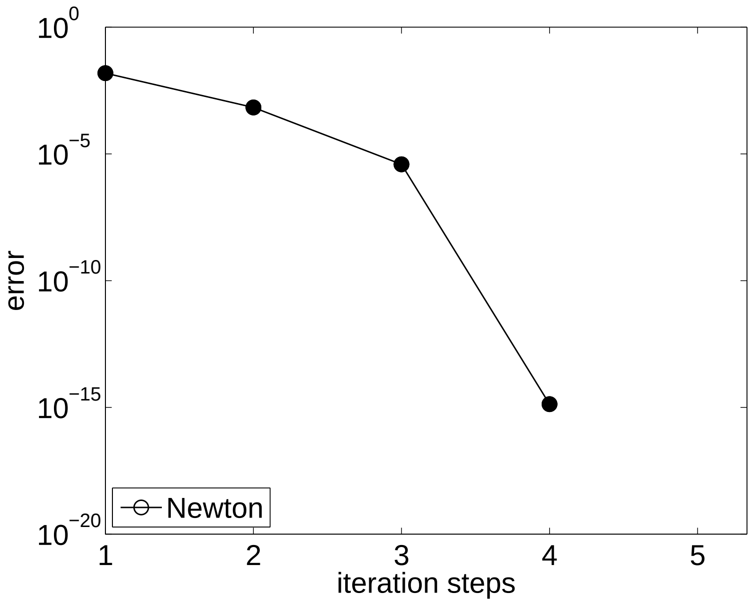
- (inexact) evaluation of $S(w^\nu)^{-1}$: linear saddle point problem
  exact, multigrid (Vanka 86, Zulehner & Schöberl 03)

generalization and globalization of primal-dual active set strategies: (Gräser 07)

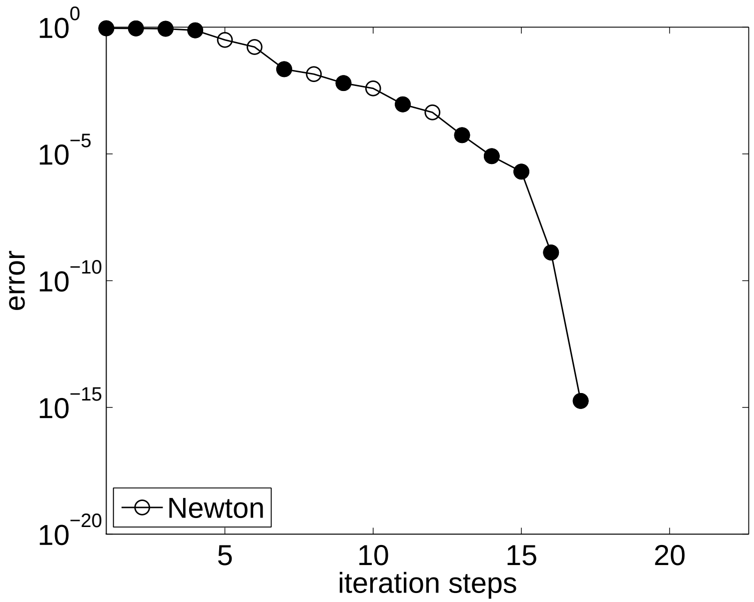
$L^2$-control: nonsmooth Schur-Newton $\Longleftrightarrow$ primal-dual active set (Hintermüller, Ito, Kunisch 03)

# $L^2$-Control: Iteration Histories for Varying $\varepsilon$

uniform refinement: $j = 7$ with $32\,513$ unknowns



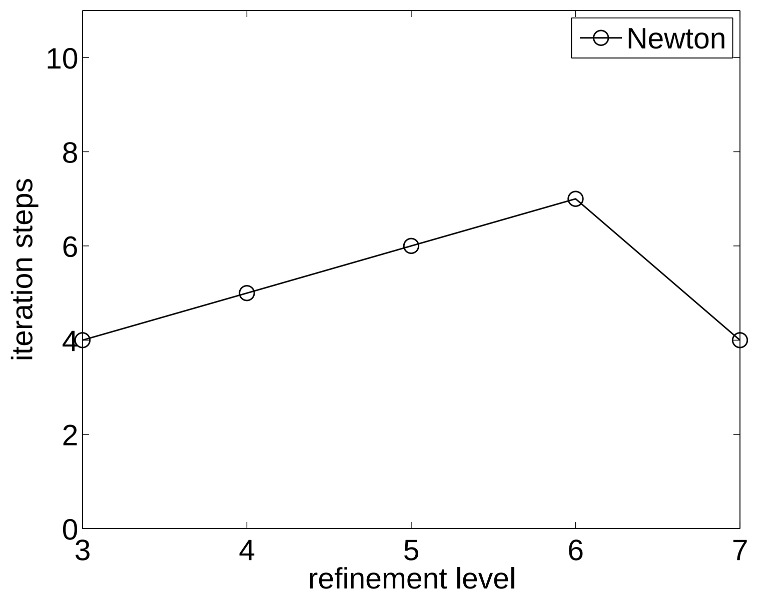

$\varepsilon = 10^{-4}$



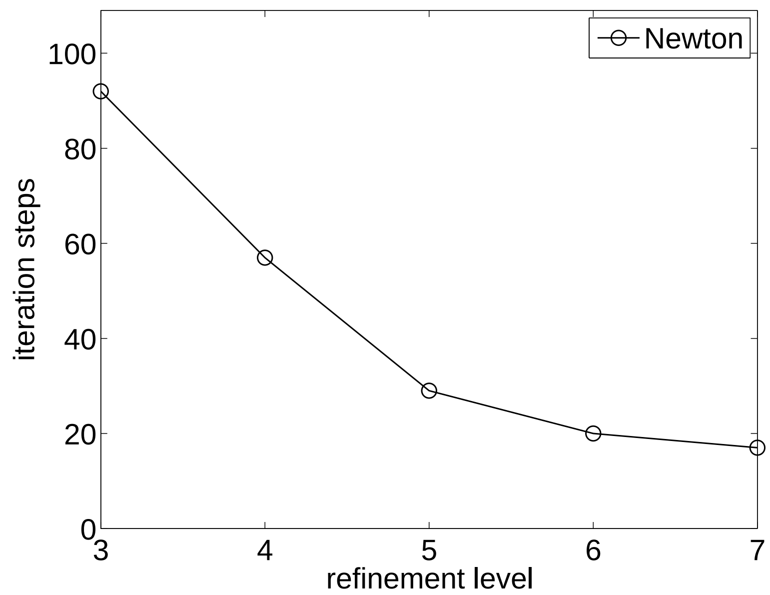

$\varepsilon = 10^{-8}$

# $L^2$-Control: Mesh Dependence for Varying $\varepsilon$

number of iteration steps to roundoff error




$\varepsilon = 10^{-4}$

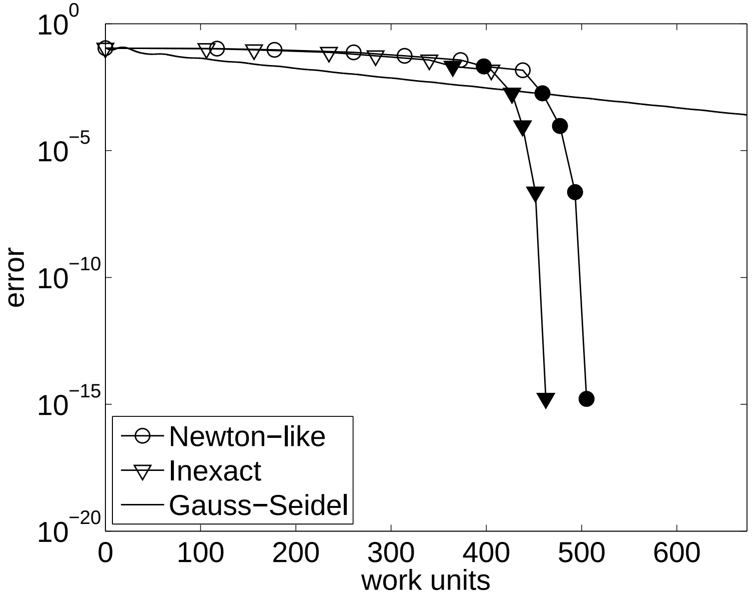



$\varepsilon = 10^{-8}$

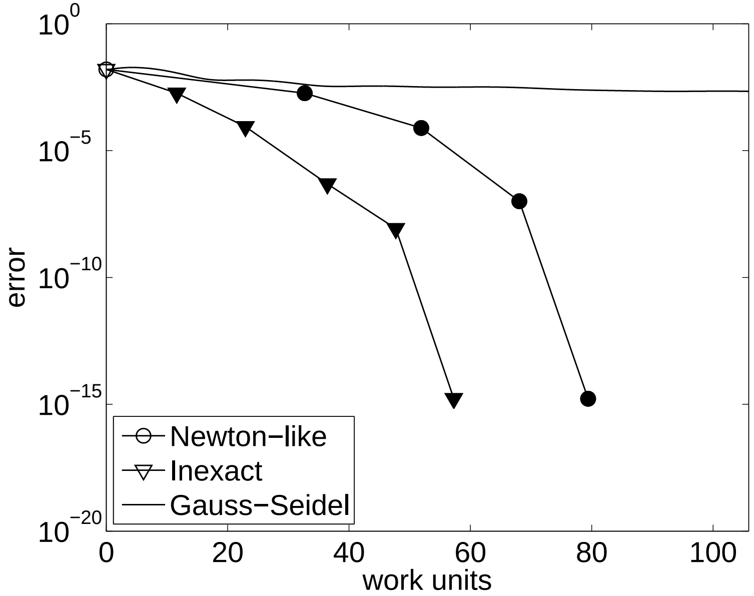
# Cahn-Hilliard: Iteration Histories for Varying $u^0$

uniform refinement: $j = 9$ with $2 \times 523\,265$ unknowns

work unit: linear $V(3, 3)$ cycle for the saddle point problem




bad initial iterate: $u^0 = 0$

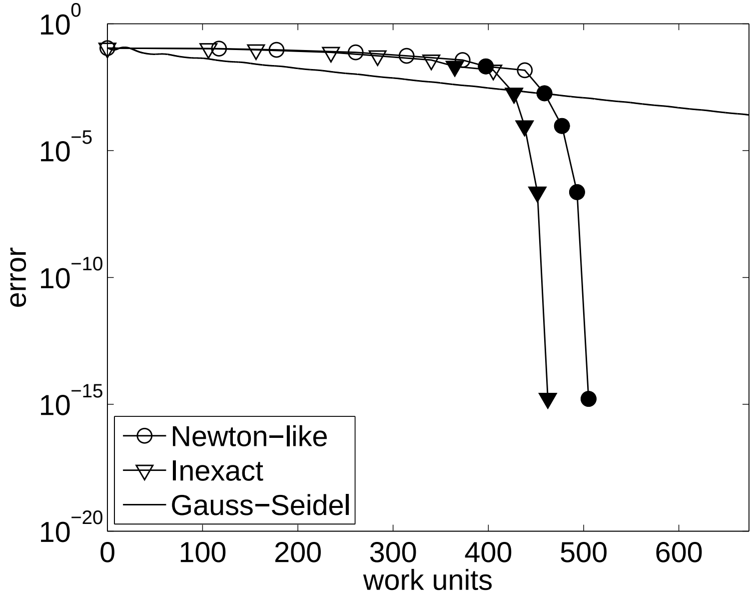



good initial iterate: previous time step

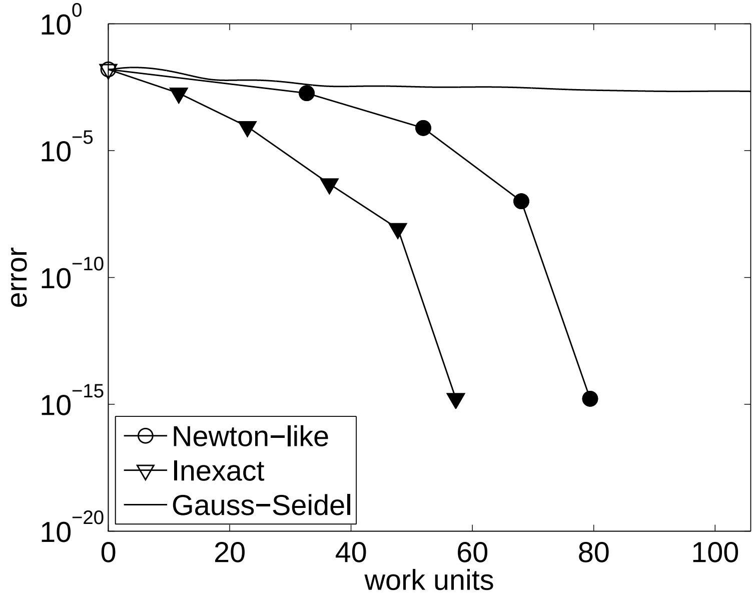
# Cahn-Hilliard: Iteration Histories for Varying $u^0$

uniform refinement: $j = 9$ with $2 \times 523\,265$ unknowns

work unit: linear $V(3,3)$ cycle for the saddle point problem




bad initial iterate: $u^0 = 0$

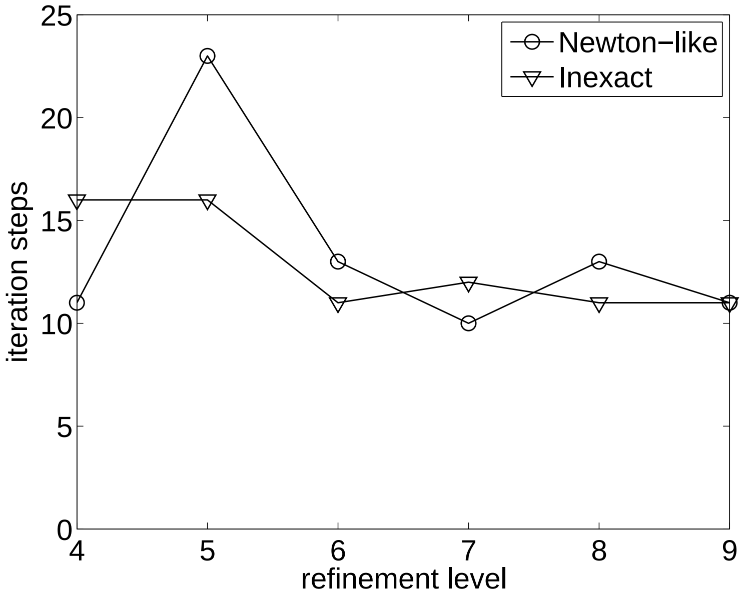



good initial iterate: previous time step

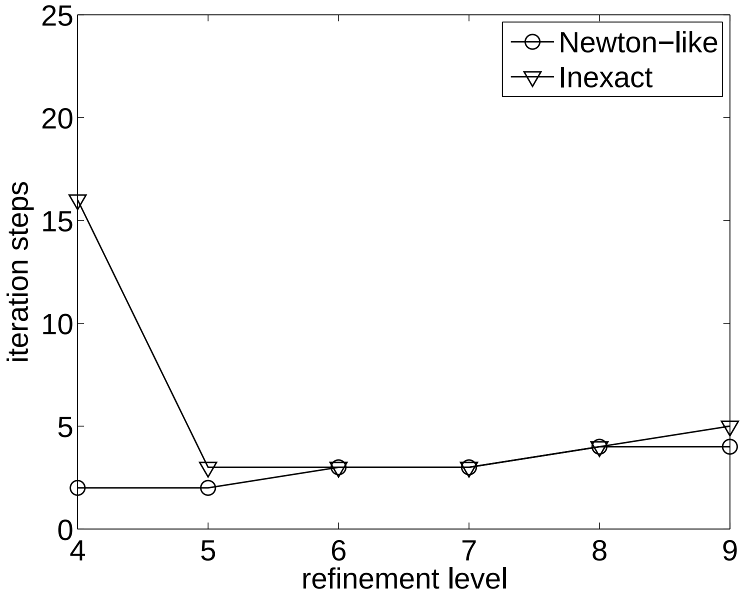
averaged convergence rate $\rho = 0.58$

# Cahn-Hilliard: Mesh-Dependence for Varying $u^0$

number of iteration steps to roundoff error




bad initial iterate: $u^0 = 0$




good initial iterate: previous time step

# Where Does the CPU Time Go?

"bad" initial iterate (inexact version):

| INEXACT | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # tests | 7 | 3 | 5 | 3 | 3 | 1 | 3 | 1 | 0 | 0 | 0 |
| % Armijo | 88.7 | 85.9 | 88.1 | 76.1 | 74.2 | 49.2 | 69.3 | 44.4 | 0.1 | 0.1 | 0.1 |
| % obstacle | 7.2 | 0.0 | -0.0 | -0.0 | 0.0 | 0.0 | 0.0 | -0.0 | 0.0 | 27.2 | 24.0 |
| % linear | 4.1 | 14.0 | 11.9 | 23.8 | 25.7 | 50.7 | 30.7 | 55.5 | 99.7 | 72.6 | 75.7 |
| work units | 106.1 | 50.1 | 78.5 | 49.0 | 56.4 | 24.5 | 40.5 | 21.8 | 11.0 | 13.4 | 10.9 |

"good" initial iterate by nested iteration: table starts at row 9 (no damping)

numerical bottleneck: • damping • linear saddle point solver

# Conclusion

## Obstacle and PDE Constrained Optimal Control Problems

- globally convergent inexact active set strategies by minimization techniques
- numerical experiments:
  mesh-independence and linear multigrid convergence speed (nested iteration)

# Outlook

- proof of multigrid convergence rates for the hybrid active set methods
- proof of mesh-independent convergence rates for Schur-Newton methods
- nonsmooth Schur-Newton methods in function space (Gräser & Schiela 08)
- applications (biomechanics, hydrology, phase-field models, . . . )