Optimal algorithms for large scale quadratic programming problems

Zdeněk DOSTÁL

Department of Applied Mathematics
FEI VŠB-Technical University of Ostrava

TU Graz,
May 2006

with

David Horák, Vít Vondrák, Dalibor Lukáš, Marta Domorádová

Department of Applied Mathematics

FEI VŠB-TU Ostrava

P. Avery, C. Farhat

Stanford

M. Lesoinne

UC Boulder

Outline

1. Motivation, optimal algorithms
2. SMALE (semimonotonic augmented Lagrangians) for equality constrained quadratic programming
3. MPRGP-optimal algorithm for bound constrained quadratic programming
4. SMALBE (semimonotonic augmented Lagrangians) for bound and equality constrained quadratic programming
5. Numerical experiments

Motivation: scalable algorithms for PDE

Elliptic problems

$$f(\mathbf{u}) = \tfrac{1}{2} a(\mathbf{u},\mathbf{u}) - b(\mathbf{u}), \quad \mathbf{u} \in H_0^1(\Omega)$$

$$a(\mathbf{u},\mathbf{u}) > C\|\mathbf{u}\|^2 \text{ for } \mathbf{u} \neq \mathbf{o}, \quad a(\mathbf{u},\mathbf{v}) = a(\mathbf{v},\mathbf{u})$$

(QP) Find: $\min f(\mathbf{u})$ for $\mathbf{u} \in H_0^1(\Omega)$

Discretization and multigrid or FETI (Fedorenko 60's, … , Farhat 90's, …)

(QP_h) Find: $\min f_h(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T \mathbf{A}_h \mathbf{x} - \mathbf{b}_h \mathbf{x}$

$$C_2\|\mathbf{x}\|^2 \geq \mathbf{x}^T \mathbf{A}_h \mathbf{x} \geq C_1\|\mathbf{x}\|^2$$

$\Rightarrow$ **Solvable in** $O(1)$ **iterations**

Our goal: develop tools for extending the results to constrained problems

Challenges:

- Identify the active constraints for free

- Get rate of convergence independent of conditioning of constraints

- Use only preconditioners that preserve bound constraints (e.g. lecture M. Domorádová, Thursday), not considered here

Equality constrained problems

For $i \in \mathcal{T}$ let

$$f_i(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T \mathbf{A}_i \mathbf{x} - \mathbf{b}_i^T \mathbf{x}$$

$$\Omega_i = \{\mathbf{x} : \mathbf{B}_i \mathbf{x} = \mathbf{o}\}, \quad \|\mathbf{B}_i\| \le C_0$$

$$\mathbf{A}_i = \mathbf{A}_i^T, \ \mathbf{B}_i \text{ possibly not full rank}$$

$$C_1 \|\mathbf{x}\|^2 \le \mathbf{x}^T \mathbf{A}_i \mathbf{x} \le C_2 \|\mathbf{x}\|^2$$

(QPE$_i$) Find: $\min_{\Omega_i} f_i(\mathbf{x})$

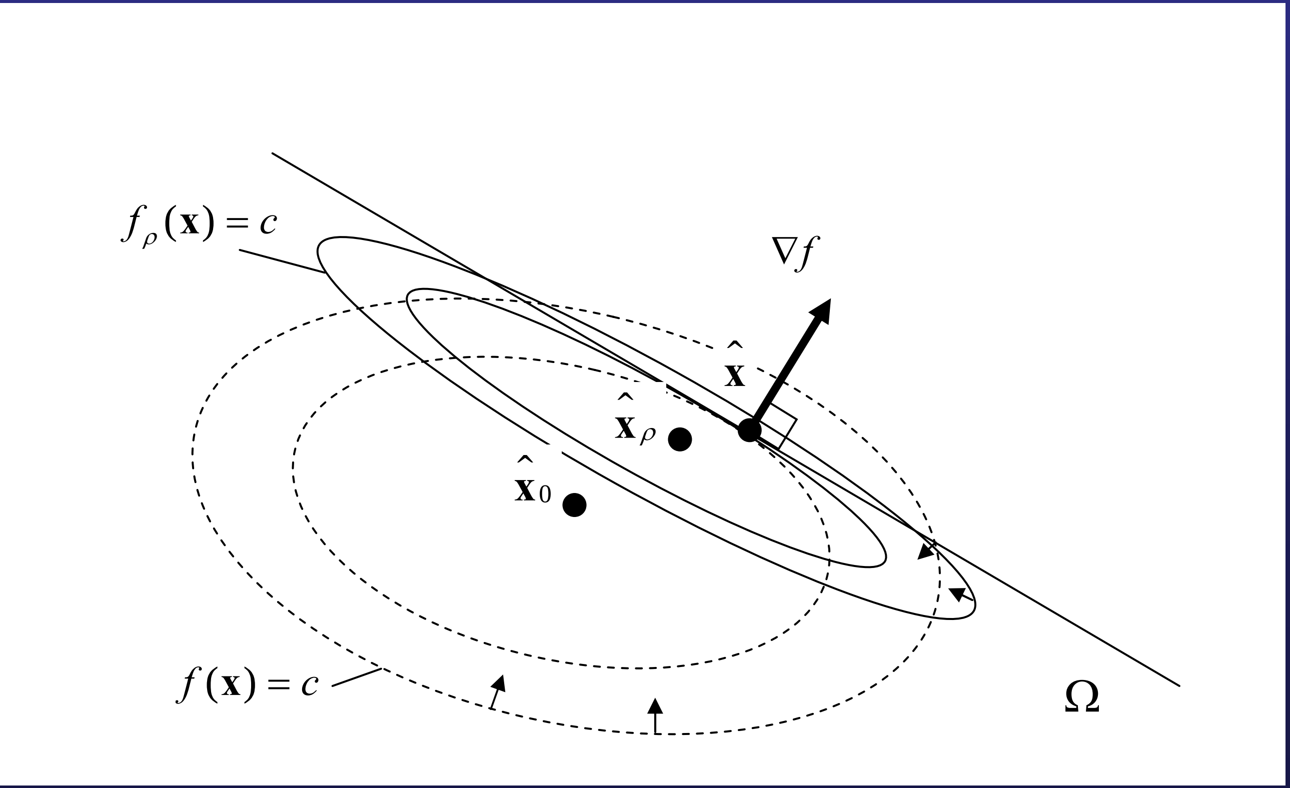
Goal: find approximate solution at *O(1)* iterations !!!

Note: we do not assume full row rank of B

Prolog: penalty method

$$f_\rho(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{2}\rho\left\|\mathbf{Bx}-\mathbf{c}\right\|^2$$

$$f_\rho(\mathbf{x}) = f(\mathbf{x}) \quad \text{on } \Omega$$



Penalty approximation of the Lagrange multipliers

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x} + \frac{1}{2}\rho\left\|\mathbf{B}\mathbf{x} - \mathbf{c}\right\|^2$$

$$\nabla f_\rho(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b} + \mathbf{B}^T\underbrace{\left(\rho(\mathbf{B}\mathbf{x} - \mathbf{c})\right)}_{\lambda}$$

Optimal estimate

$$Th.: \quad \varepsilon > 0, \quad \rho > 0, \quad \left\|\nabla f_\rho(\mathbf{x})\right\| \le \varepsilon \|\mathbf{b}\|$$

$$\Rightarrow \quad \|\mathbf{B}x - \mathbf{c}\| \le \frac{1+\varepsilon}{\sqrt{\lambda_{\min}\rho}} \|\mathbf{b}\|$$

Non optimal but linear in ρ estimate

$$Th.: \quad \varepsilon > 0, \quad \rho > 0, \quad \left\|\nabla f_\rho(\mathbf{x})\right\| \le \varepsilon \left\|\mathbf{b}\right\|$$

β the smallest nonzero eigenvalue of $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$

$$\Rightarrow \quad \left\|\mathbf{B}x - \mathbf{c}\right\| \le \frac{1+\varepsilon}{1+\beta\rho} \left\|\mathbf{b}\right\| \left\|\mathbf{B}\mathbf{A}^{-1}\right\| \left\|\mathbf{b}\right\| + \rho^{-1} \left\|\mathbf{c}\right\|$$

Optimality of dual penalty for FETI1

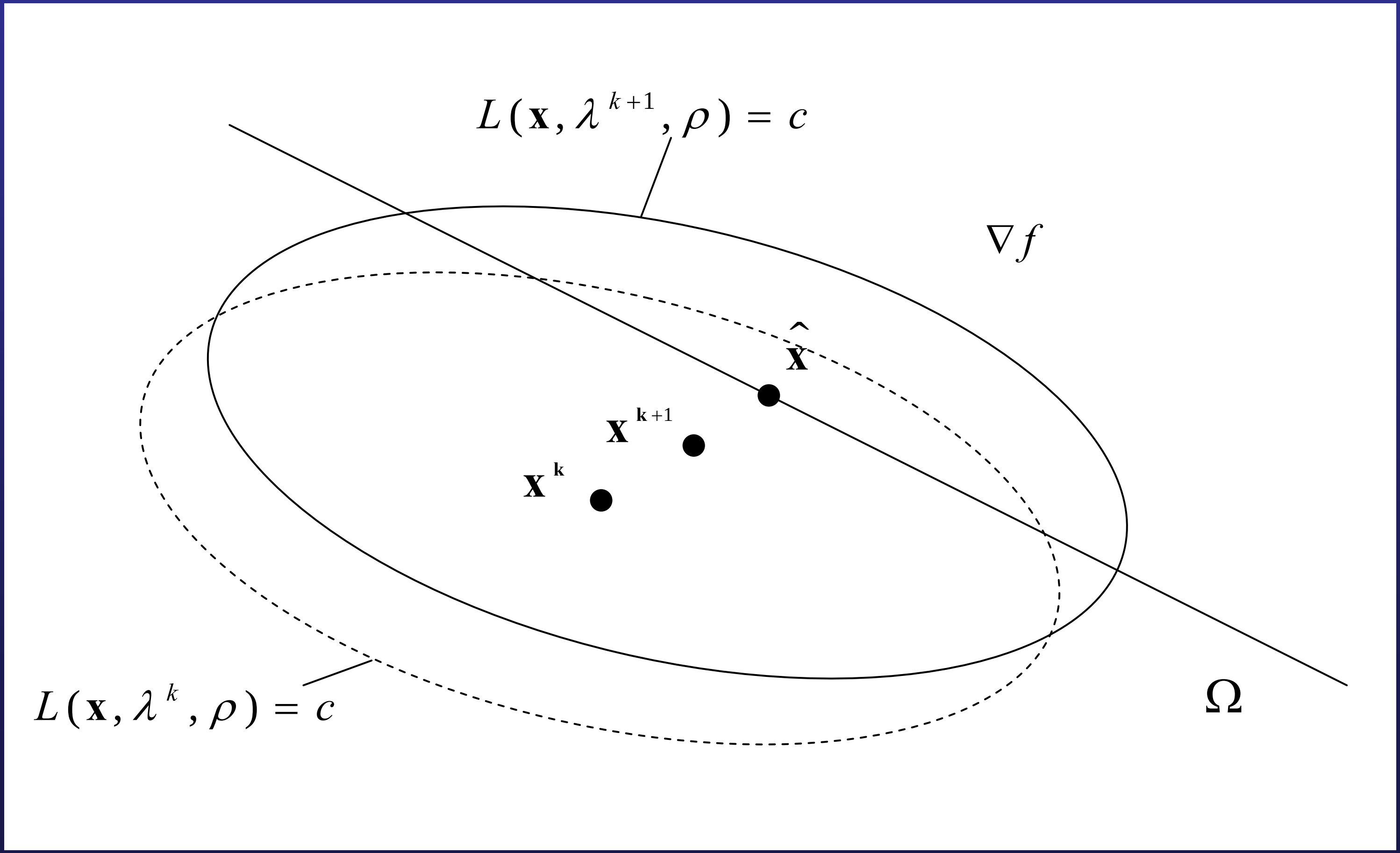
$\|\mathbf{Bx}\| / \|\mathbf{b}\|$ for varying ρ and fixed H/h			
$\rho \setminus n$	1152	139392	2130048
1	1.32e-1	1.20e-1	1.12e-1
1000	1.40e-3	1.28e-3	1.19e-3
100 000	1.40e-5	1.28e-5	1.19e-5

Augmented Lagrangian and gradient

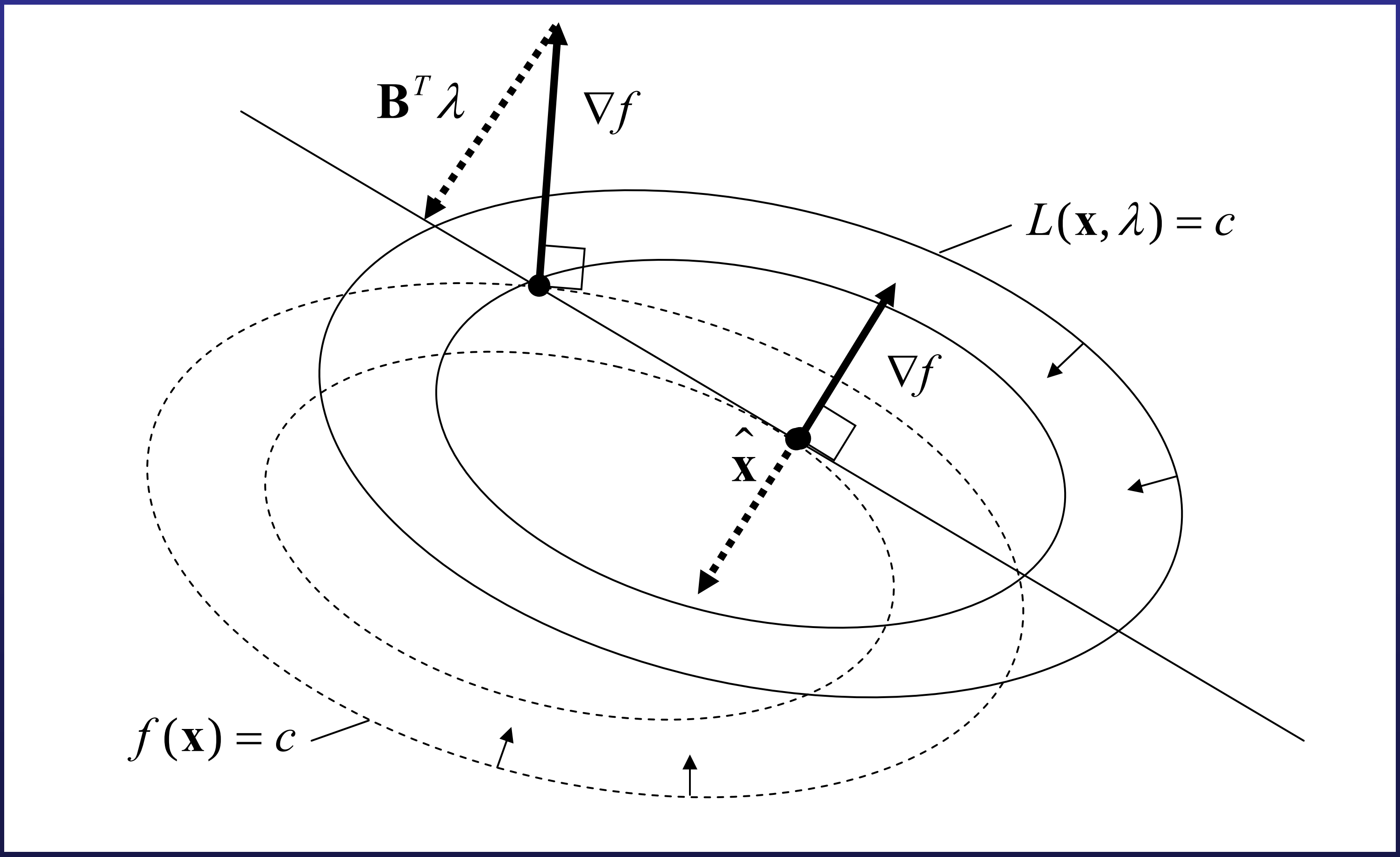
$$L(\mathbf{x}, \mu, \rho) = f(\mathbf{x}) + \mu^T(\mathbf{Bx} - \mathbf{c}) + \frac{1}{2}\rho\left\|\mathbf{Bx} - \mathbf{c}\right\|^2$$

$$\mathbf{g}(\mathbf{x}, \mu, \rho) = \nabla_{\mathbf{x}} L(\mathbf{x}, \mu, \rho) = \mathbf{Ax} - \mathbf{b} + \mathbf{B}^T \underbrace{\left(\mu + \rho(\mathbf{Bx} - \mathbf{c})\right)}_{\tilde{\mu}}$$

Augmented Lagrangians



KKT conditions



SMALE-Semimonotonic Augmented Lagrangians

{Initialization}

Step 0 $\quad 1<\beta, \rho_0>0, \eta>0, M>0, \mu^0$

{Approximate solution of bound constrained problem}

Step 1 Find $\mathbf{x}^k$ such that $\left\|\mathbf{g}\left(\mathbf{x}^k, \boldsymbol{\mu}^k, \rho_k\right)\right\| \leq \min\left\{M\left\|\mathbf{B}\mathbf{x}^k-\mathbf{c}\right\|, \eta\right\}$

{Test}

Step 2 If $\left\|\mathbf{g}\left(\mathbf{x}^k, \boldsymbol{\mu}^k, \rho_k\right)\right\|$ and $\left\|\mathbf{B}\mathbf{x}^k-\mathbf{c}\right\|$ are small then $\mathbf{x}^k$ is solution

{Update Lagrange multipliers}

Step 3 $\quad \boldsymbol{\mu}^{k+1}=\boldsymbol{\mu}^k+\rho_k\left(\mathbf{B}\mathbf{x}^k-\mathbf{c}\right)$

{Update penalty parameter}

Step 4 If $L\left(\mathbf{x}^{k+1}, \boldsymbol{\mu}^{k+1}, \rho_{k+1}\right) \leq L\left(\mathbf{x}^k, \boldsymbol{\mu}^k, \rho_k\right)+\frac{\rho_{k+1}}{2}\left\|\mathbf{B}\mathbf{x}^{k+1}-\mathbf{c}\right\|^2$

then $\rho_{k+1}=\beta\rho_{\mathrm{k}}$

else $\rho_{k+1}=\rho_k$

{Repeat loop}

Step 5 $\quad k=k+1$ and return to Step 1

Basic relations for SMALE

Theorem :

Let $\{\mathbf{x}^k\}, \{\mu^k\}$ and $\{\rho^k\}$ be generated with $\bar{\alpha} \in (0, \|\mathbf{A}\|^{-1}]$ and $\Gamma > 0$.

(i) If $\rho_k \geq M^2 / \lambda_{\min}(\mathbf{A})$ then

$$L\left(\mathbf{x}^{k+1}, \mu^{k+1}, \rho_{k+1}\right) \geq L\left(\mathbf{x}^k, \mu^k, \rho_k\right) + \frac{\rho_{k+1}}{2}\left\|\mathbf{D}\mathbf{x}^{k+1}\right\|^2$$

(ii) There is $C = C(C_1, C_2, M)$ such that

$$\sum_{k=1}^{\infty} \frac{\rho_k}{2}\left\|\mathbf{B}\,\mathbf{x}^k\right\|^2 \leq C$$

Z.D., SINUM (2006), Z.D. Computing (2006)

Optimality of SMALE

Corollary :

Let $\{\mathbf{x}_i^k\}, \{\mu\}$ and $\{\rho^k\}$ be generated with $\bar{\alpha} \in (0, \|\mathbf{A}\|^{-1}]$, $\beta>0$, M>0 and $\Gamma>0$.

(i)

$$\rho_k \le \beta M^2 / \lambda_{\min}(\mathbf{A})$$

(ii) SMALE generates $\mathbf{x}^k$ that satisfies

$$\|\mathbf{g}\ (\mathbf{x}^k)\| \le \varepsilon \|\mathrm{b}\| \quad \text{and} \quad \|\mathbf{B}\mathbf{x}^k\| \le \varepsilon \|\mathrm{b}\|$$

at $O(1)$ outer iterations

(iii) SMALE with CG in inner loop generates $\mathbf{x}^k$ that satisfies

$$\|\mathbf{g}(\mathbf{x}^k)\| \le \varepsilon \|\mathrm{b}\| \quad \text{and} \quad \|\mathbf{B}\mathbf{x}^k\| \le \varepsilon \|\mathrm{b}\|$$

at $O(1)$ matrix-vector multiplications

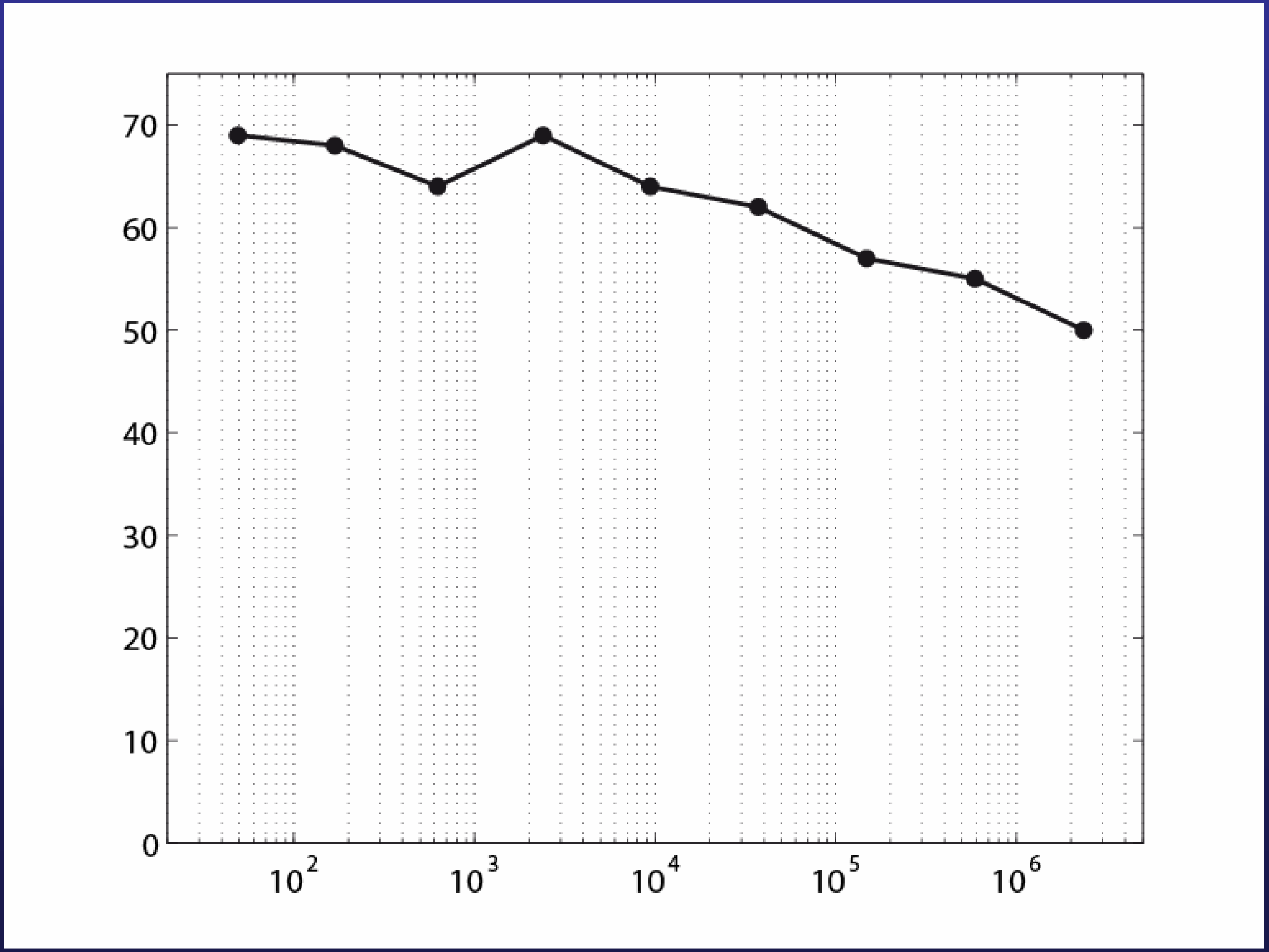
Z.D. OMS (2005), COA (2007)

Convergence of Lagrange multipliers

(i) Lagrange multipliers converge even for dependent constraints

(ii) The convergence is linear for sufficiently large ρ

CG iterace – string system on Winkler support, multipoint constraints, cond=5 G



Bound constrained problems

For $i \in \mathcal{T}$ let

$$f_i(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{A}_i\mathbf{x} - \mathbf{b}_i^T\mathbf{x}, \quad \Omega_i = \{\mathbf{x} : \mathbf{x} \ge \mathbf{c}_i\},$$

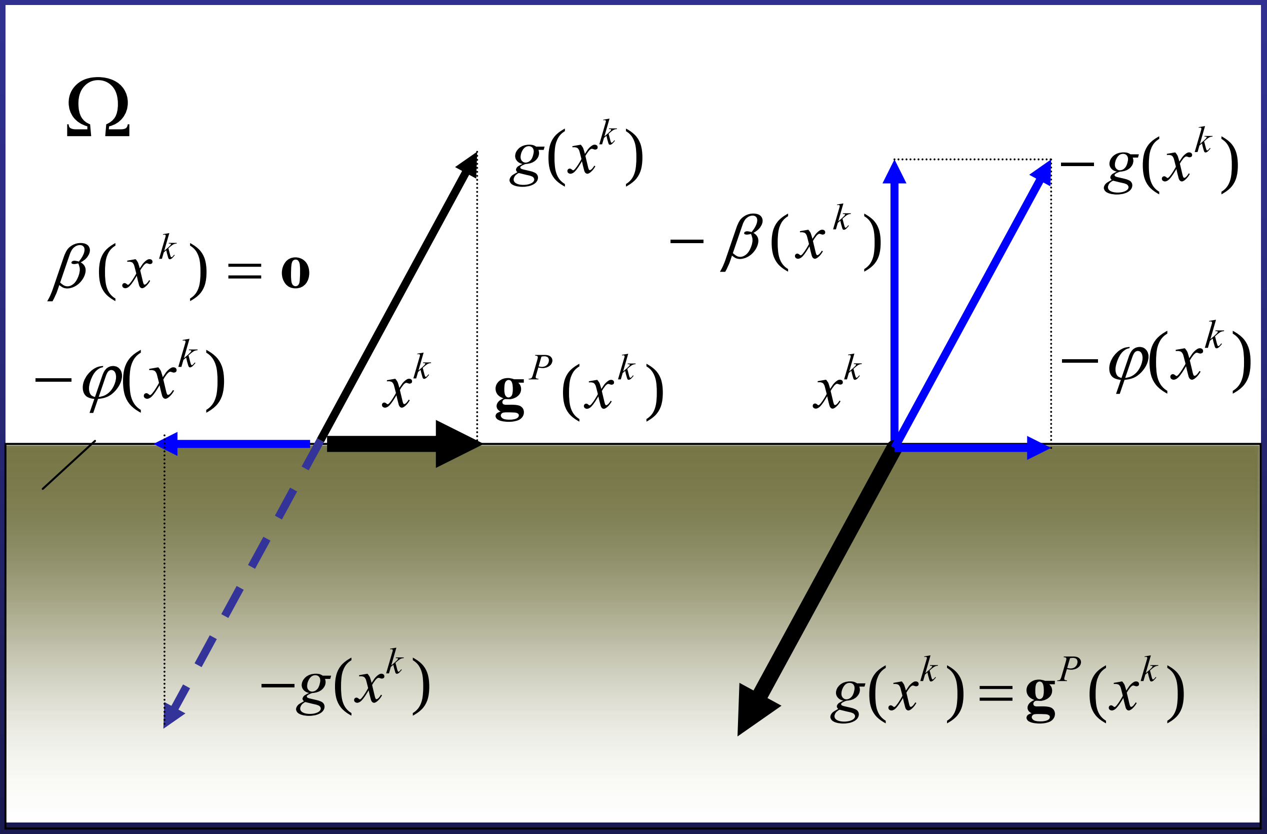
$$\mathbf{A}_i = \mathbf{A}_i^T, \quad \mathbf{x}^T\mathbf{A}_i\mathbf{x} > 0 \text{ for } \mathbf{x} \ne \mathbf{o}$$

$$C_1\|\mathbf{x}\|^2 \le \mathbf{x}^T\mathbf{A}_i\mathbf{x} \le C_2\|\mathbf{x}\|^2 \quad \text{and} \quad \|\mathbf{c}_i^+\| \le C_3$$

$$(\mathrm{QPB_i}) \qquad \text{Find: } \min_{\Omega_i} f_i(\mathbf{x})$$

Goal: find approximate solution at *O(1)* iterations !!!

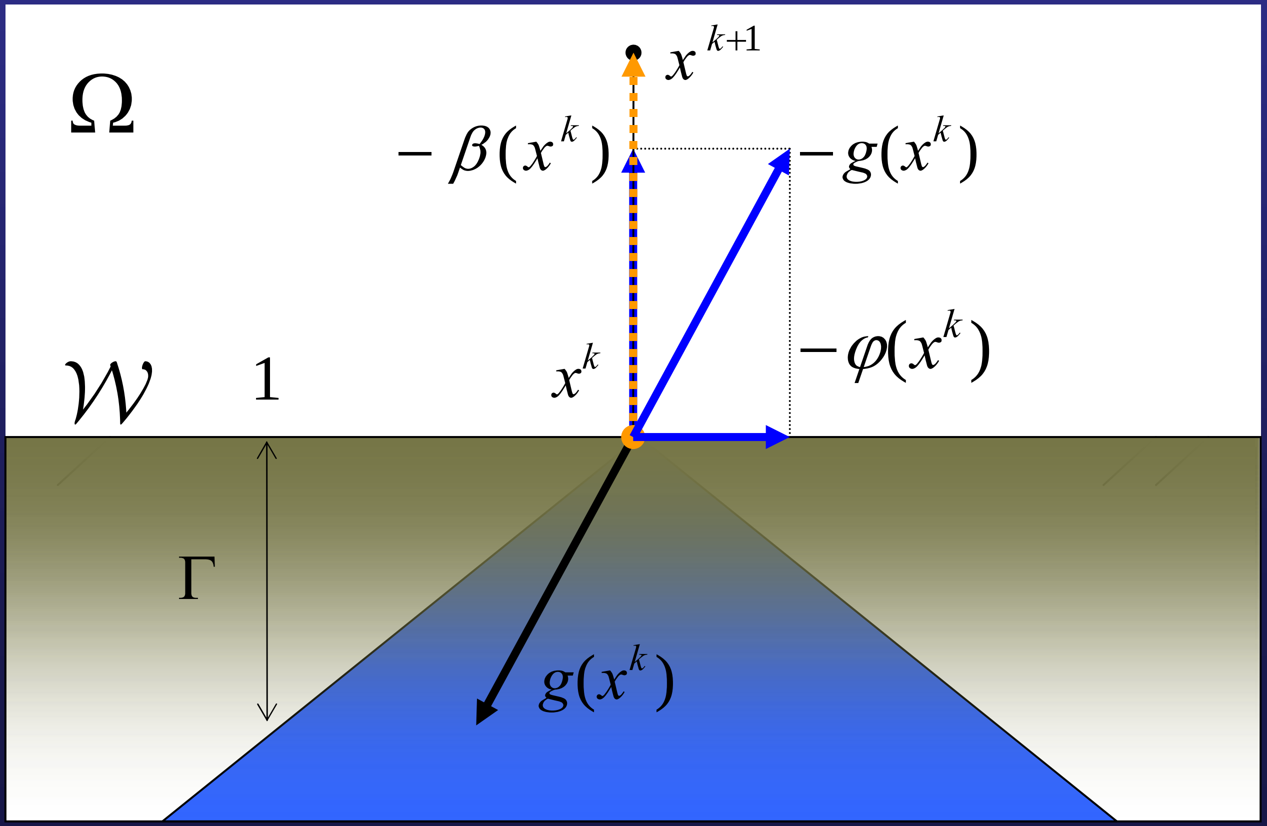
Projected gradient



Deleting indices from active set- proportioning

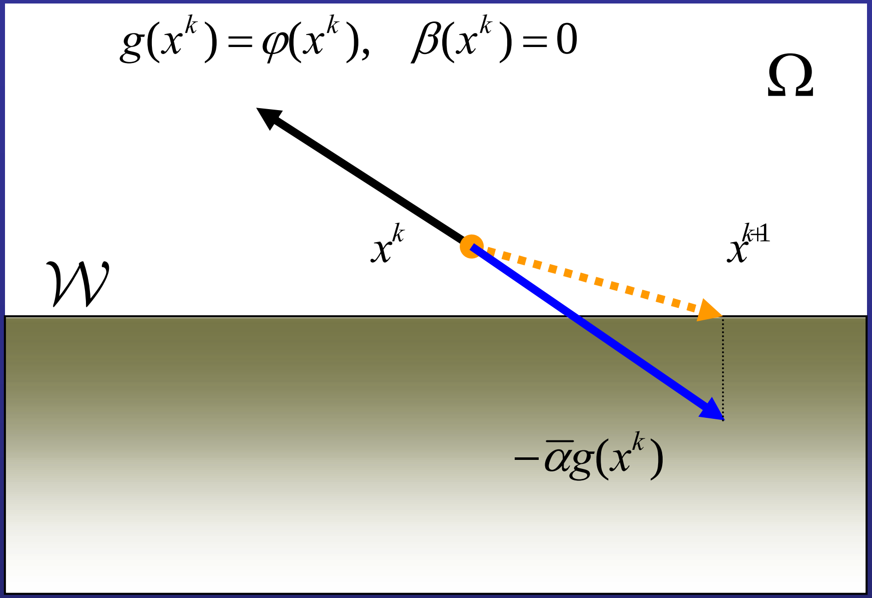
x **proportional:** $$\Gamma^2 \tilde{\varphi}^T(x)\varphi(x) \geq \|\beta(x)\|^2$$

Reduction of the active set

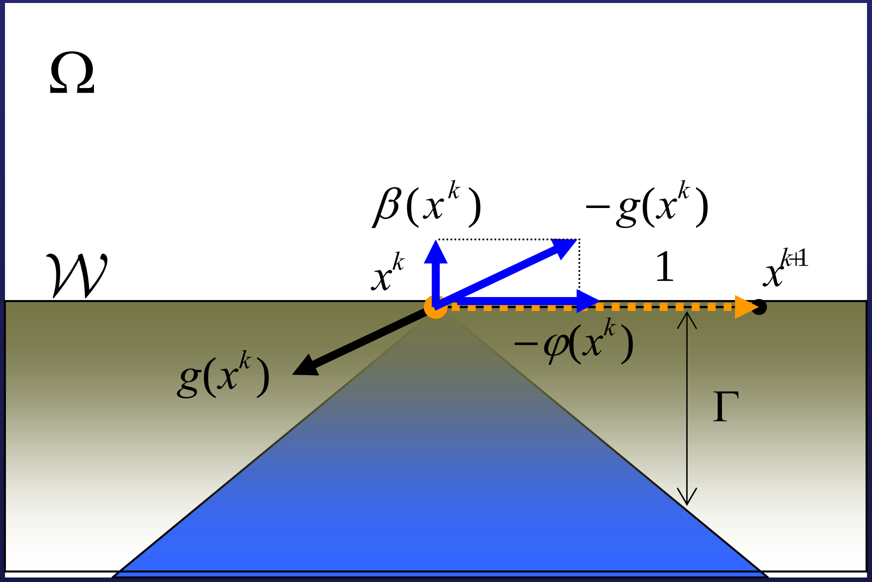


Proportional iterations

Projection step: expansion of the active set



Feasible conjugate gradient step:



MPRGP- Modified Proportioning with Reduced Gradient Projection

{Initialization}

Given $\mathbf{x}^0 \in \Omega, \quad \bar{\alpha} \in (0, \|\mathbf{A}\|^{-1}], \quad \Gamma > 0$

{Proportioning}

Step 1: if $\mathbf{x}^k$ is not proportional, then define $\mathbf{x}^{k+1}$ by proportionalization

i. e. minimalization in direction $-\beta(\mathbf{x}^k)$

{conjugate gradient}

Step 2: if $\mathbf{x}^k$ is proportional, then generate $\mathbf{x}^{k+1}$ by trial cg step

{projection}

Step 3: if $\mathbf{x}^{k+1} \in \Omega$ then use it,

else $\mathbf{x}^{k+1} = (\mathbf{x}^k - \bar{\alpha}\varphi(\mathbf{x}^k))^+$

Rate of convergence of MPRGP

Theorem :

Let $\Gamma > 0, \hat{\Gamma} = \max\{\Gamma, \Gamma^{-1}\}$, $\overline{\mathbf{x}}$ solution of (QPB), $\alpha_1 = \lambda_{\min}(\mathbf{A})$,

$\{\mathbf{x}^k\}$ generated with $\overline{\alpha} \in \left\langle 0, \|\mathbf{A}\|^{-1}\right]$. Then:

(i) The R - linear rate of convergence in the energy norm $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A}\mathbf{x}$ is given by

$$\|\mathbf{x}^k - \overline{\mathbf{x}}\|_{\mathbf{A}}^2 \le 2\eta^k \left(f(\mathbf{x}^0) - f(\overline{\mathbf{x}}) \right) \quad \text{with} \quad \eta = 1 - \frac{\overline{\alpha}\alpha_1}{2 + 2\hat{\Gamma}^2} < 1$$

(ii) The R - linear rate of convergence of the projected gradient is given by

$$\|\mathbf{g}^P(\mathbf{x}^k)\|^2 \le a\,\eta^k \left(f(\mathbf{x}^0) - f(\overline{\mathbf{x}}) \right), \quad \text{with} \quad a = \frac{36\overline{\alpha}^{-1}\alpha_1^{-1}}{\eta(1-\eta)}$$

Z.D., J. Schoeberl, Comput. Opt. Appl. (2005), Z.D. NA (2004)

Optimality of MPRGP

Theorem :

Let $\Gamma > 0, \hat{\Gamma} = \max\{\Gamma, \Gamma^{-1}\}$, $\overline{\mathbf{x}}_i$ solution of (QPB$_i$),

$\{\mathbf{x}_i^k\}$ generated with $\overline{\alpha} \in (0, C_2^{-1}]$ and $\mathbf{x}_i^0 = \max\{\mathbf{c}_i, \mathbf{o}\}$.

Then $\mathbf{x}_i^k$ that satisfies

$$\|\mathbf{x}_i^k - \overline{\mathbf{x}}_i\| \le \varepsilon \|\mathbf{b}_i\| \quad \text{and} \quad \|g^P(\mathbf{x}_i^k)\| \le \varepsilon \|\mathbf{b}_i\|$$

is found at

$O(1)$ matrix-vector multiplications

Z.D., J. Schoeberl, Comput. Opt. Appl. (2005),

Finite termination

Theorem :

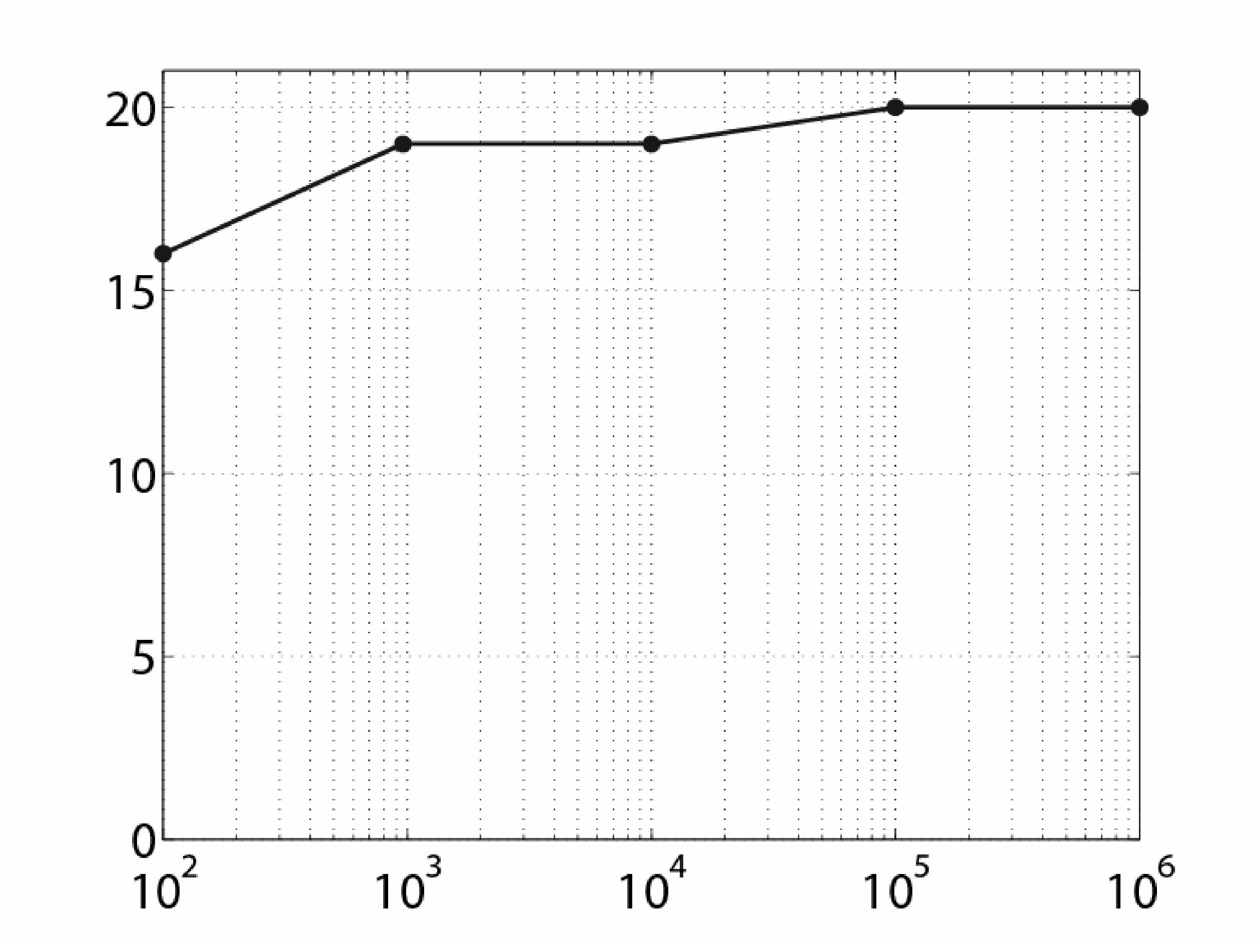
Let $\bar{\mathbf{x}}$ denote the solution of (QPB), $\{\mathbf{x}^{\mathrm{k}}\}$ generated with $\bar{\alpha} \in (0, \|\mathbf{A}\|^{-1}]$ and $\Gamma > 0$. Then

(i) If $\bar{x}_i = 0$ implies $g_i(\bar{\mathbf{x}}) = 0$ then there is $k \geq 0$ such that $\mathbf{x}^k = \bar{\mathbf{x}}$

(ii) If $\Gamma \geq 2\left(\sqrt{\kappa(\mathbf{A})} + 1\right)$ then there is $k \geq 0$ such that $\mathbf{x}^k = \bar{\mathbf{x}}$

(i) More Z.D. SIOPT (1996), (ii) Z.D., Schoeberl, COA (2005)

CG iterace – string system on Winkler support, bound constraints, cond=5



Bound and equality constrained problems

For $i \in \mathcal{T}$ let

$$f_i(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T \mathbf{A}_i \mathbf{x} - \mathbf{b}_i^T \mathbf{x}$$

$$\Omega_i = \{\mathbf{x} : \mathbf{x} \geq \mathbf{c}_i \text{ and } \mathbf{B}_i \mathbf{x} = \mathbf{o}\}, \quad \|\mathbf{B}_i\| \leq C_0$$

$$\mathbf{A}_i = \mathbf{A}_i^T,$$

$$C_1 \|\mathbf{x}\|^2 \leq \mathbf{x}^T \mathbf{A}_i \mathbf{x} \leq C_2 \|\mathbf{x}\|^2 \quad \text{and} \quad \|\mathbf{c}_i^+\| \leq C_3$$

(QPBE$_i$) Find: $\min_{\Omega_i} f_i(\mathbf{x})$

Goal: find approximate solution at *O(1)* iterations !!!

Note: we do not assume full row rank of D!!!

Augmented Lagrangian and projected gradient

$$L(\mathbf{x},\mu,\rho) = f(\mathbf{x}) + \mu^T \mathbf{B}\mathbf{x} + \frac{1}{2}\rho\left\|\mathbf{B}\mathbf{x}\right\|^2$$

$$\mathbf{g}^P(\mathbf{x},\mu,\rho) = \nabla_x L(\mathbf{x},\mu,\rho)$$

$$\mathbf{g}^P = \mathbf{g}^P(\mathbf{x},\mu,\rho) = \varphi(\mathbf{x},\mu,\rho) + \beta(\mathbf{x},\mu,\rho)$$

SMALBE-Semimonotonic augmented Lagrangians

{Initialization}

Step 0 $1 < \beta, \rho_0 > 0, \eta > 0, M > 0, \mu^0$

{Approximate solution of bound constrained problem}

Step 1 Find x^k such that $\left\| \mathbf{g}^P\left(\mathbf{x}^k, \mu^k, \rho_k\right) \right\| \le \min\left\{ M \left\| \mathbf{B}\mathbf{x}^k \right\|, \eta \right\}$

{Test}

Step 2 If $\left\| \mathbf{g}^P\left(x^k, \mu^k, \rho_k\right) \right\|$ and $\left\| \mathbf{B}\mathbf{x}^k \right\|$ are small then x^k is solution

{Update Lagrange multipliers}

Step 3 $\mu^{k+1} = \mu^k + \rho_k\left(\mathbf{B}\mathbf{x}^k\right)$

{Update penalty parameter}

Step 4 If $L\left(\mathbf{x}^{k+1}, \mu^{k+1}, \rho_{k+1}\right) \le L\left(\mathbf{x}^k, \mu^k, \rho_k\right) + \frac{\rho_{k+1}}{2}\left\| \mathbf{B}\mathbf{x}^{k+1} \right\|^2$

then $\rho_{k+1} = \beta\rho_{\mathrm{k}}$

else $\rho_{k+1} = \rho_k$

{Repeat loop}

Step 5 $k = k + 1$ and return to Step 1

Basic relations for SMALBE

Theorem :

Let $\{\mathbf{x}^k\}, \{\mu^k\}$ and $\{\rho^k\}$ be generated with $\bar{\alpha} \in (0, \|\mathbf{A}\|^{-1}]$ and $\Gamma > 0$.

(i) If $\rho_k \geq M^2 / \lambda_{\min}(\mathbf{A})$ then

$$L(\mathbf{x}^{k+1}, \mu^{k+1}, \rho_{k+1}) \geq L(\mathbf{x}^k, \mu^k, \rho_k) + \frac{\rho_{k+1}}{2}\|\mathbf{B}\mathbf{x}^{k+1}\|^2$$

(ii) There is $C = C(C_1, C_2, \bar{\alpha}, \Gamma, M)$ such that

$$\sum_{k=1}^{\infty} \frac{\rho_k}{2}\|\mathbf{B}\mathbf{x}^k\|^2 \leq C$$

Z.D. SINUM (2005), Z.D.(2006)

Optimality of SMALBE

Corollary :

Let $\{\mathbf{x}_i^k\}, \{\mu\}$ and $\{\rho^k\}$ be generated with $\bar{\alpha} \in (0, \|\mathbf{A}\|^{-1}]$,

$\beta>0$, M>0 and $\Gamma>0$.

(i)

$$\rho_k \leq \beta M^2 / \lambda_{\min}(\mathbf{A})$$

(ii) SMALBE generates $\mathbf{x}^k$ that satisfies

$$\|g^P(\mathbf{x}^k)\| \leq \varepsilon \|b\| \quad \text{and} \quad \|\mathbf{B}\mathbf{x}^k\| \leq \varepsilon \|b\|$$

at $O(1)$ outer iterations

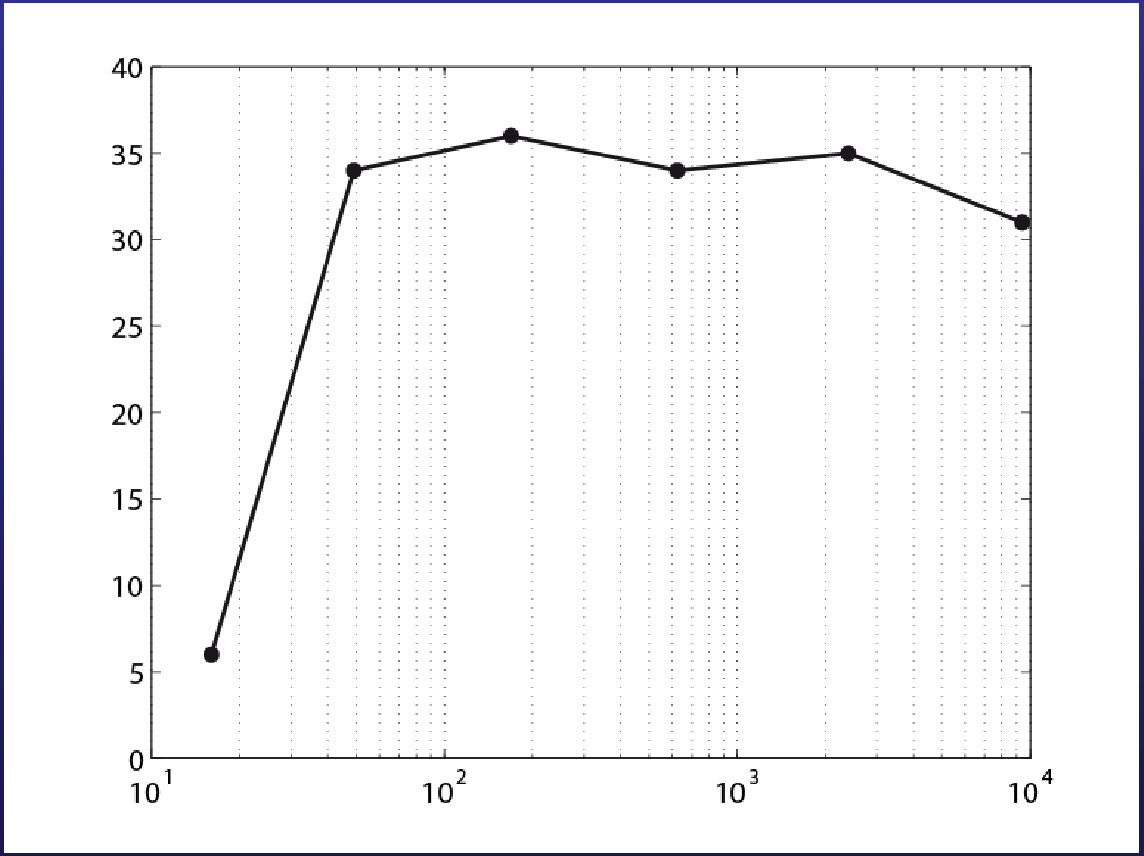
(ii) SMALBE with MPRGP in inner loop generates $\mathbf{x}^k$ that satisfies

$$\|g^P(\mathbf{x}^k)\| \leq \varepsilon \|b\| \quad \text{and} \quad \|\mathbf{B}\mathbf{x}^k\| \leq \varepsilon \|b\|$$

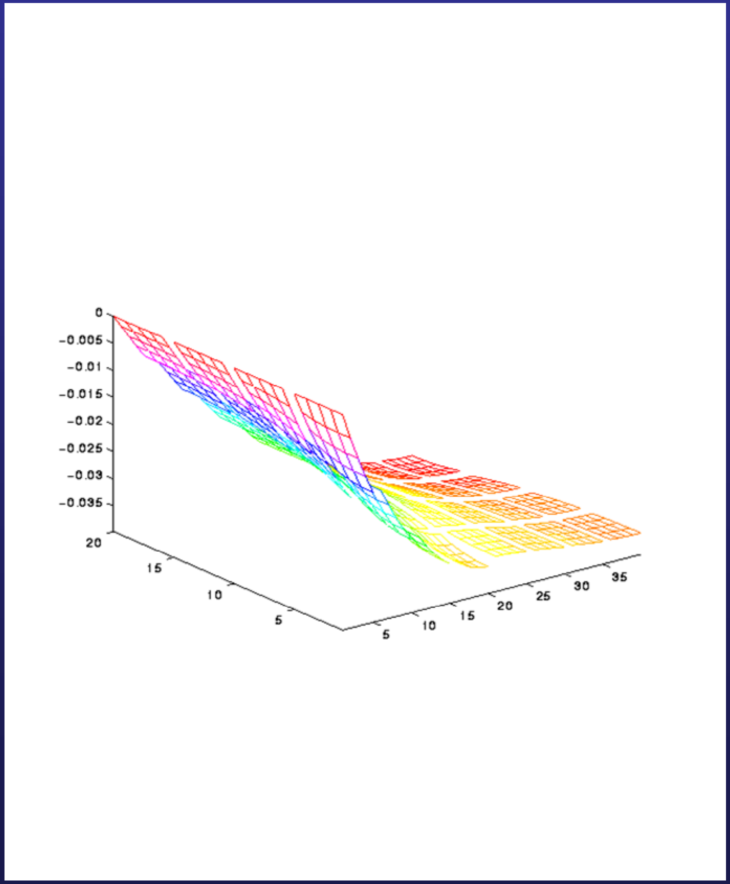
at $O(1)$ matrix-vector multiplications

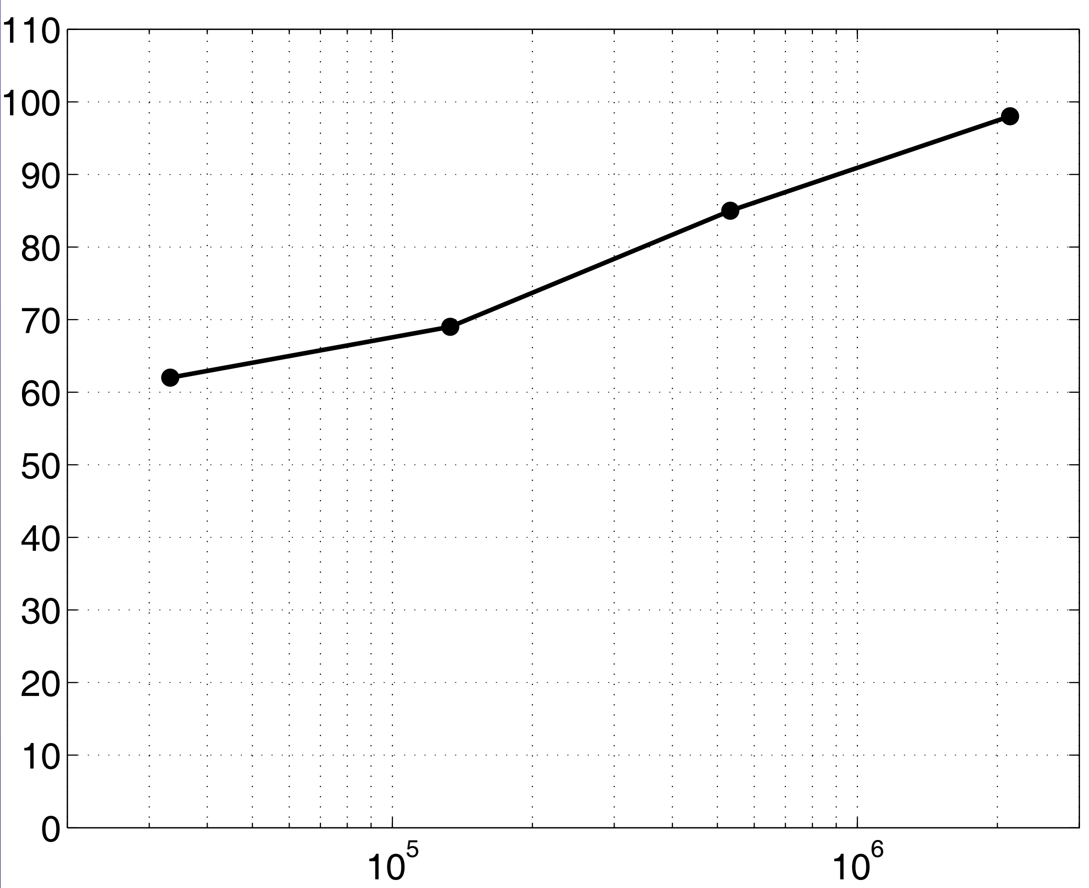
Z.D. SINUM (2006), Z.D. Computing (2007)

CG iterations – string system on Winkler support, bound and multipoint constraints, cond=5

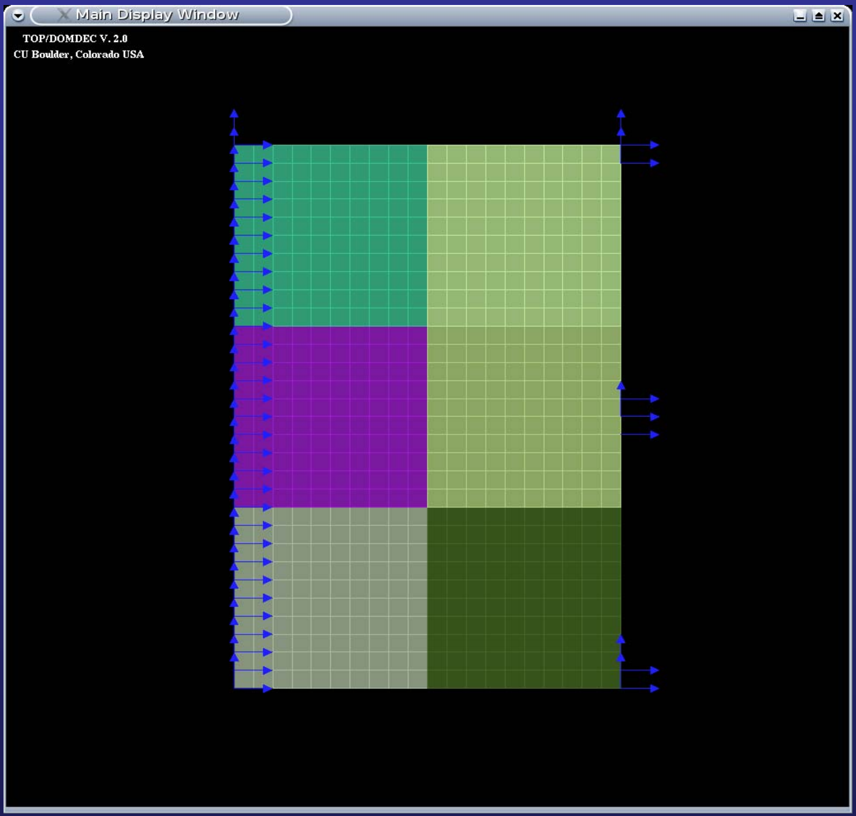


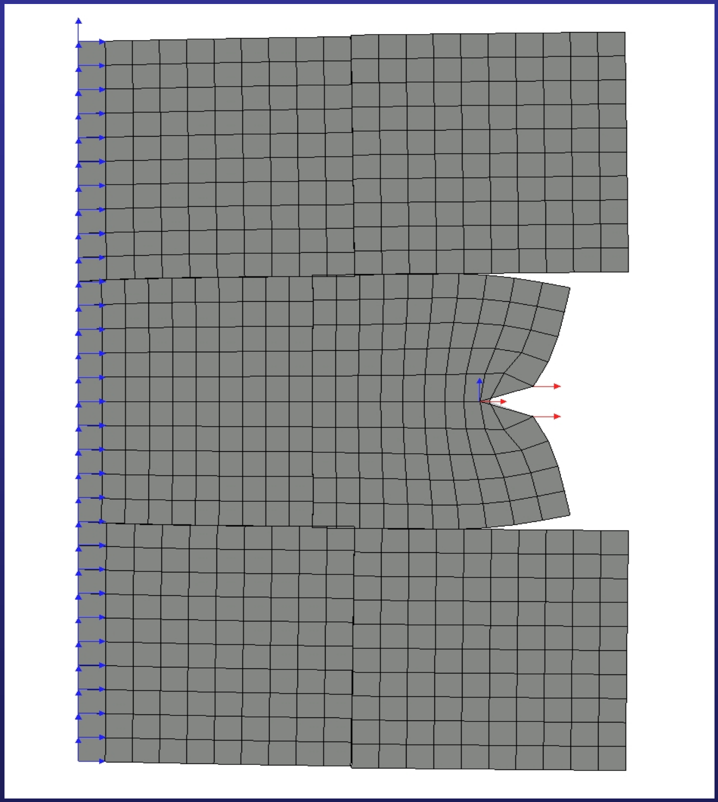
Solution and numerical scalability of TFETI for *n* ranging from 50 to 2 130 048 (C/PETSc)





Solution and numerical scalability of FETI
2D semicoercive benchmark, 6 bodies





Subdomains	dof	Contact conditions	It FETI-1	It FETI-DP
96	118098	565	103	82
384	466578	1125	129	90

Related work

1. Projectors introduced by Calamai, More, Toraldo
2. Efficiency of inexact working set strategy with preconditioning in face considered by O'Leary
3. Adaptive precision control introduced by Friedlander and Martinez
4. Basic algorithm for bound and equality constraints was introduced by Conn, Gould and Toint and used in LANCELOT
5. Precision control that we use introduced Hager, used by Z.D., Friedlander, Santos and Gomes

Conlusions

1. New algorithms for bound and equalityconstrained problems were introduced
2. Qualitatively new results were proved
3. Theoretical results demonstrated by numerical experiments
4. The results were applied to develop scalable algorithms for elliptic boundary variational inequalities
5. Current reserach: preconditioning with improved rate of convergence (Thursday – Domorádová)