Optimal algorithms for the solution of quadratic programming problems with convex constraints

Lukáš Pospíšil

Department of Applied Mathematics, VŠB - Technical University of Ostrava, Czech Republic

Summer School, 14 July - 11 August 2013, Wisconsin - Madison
Outline

- About Czech Republic, Ostrava, University
- Motivation - about QP, contact problems
- Solving QP with inequalities
  - Active-set algorithms
  - Projected descend methods
- Solving QP with equalities
  - Augmented Lagrangians
- Conclusions (?)
About Czech Republic

Area: 78,866 km²
(30,450 sq mi)
Population: 10,513,209
We love beer! (i.e. *pivo*)

- the highest beer consumption per capita in the world (on the average 144 litres per one person during a one year)
  - 2. Germany (107)
  - 3. Austria (106)
  - 5. Poland (85)
  - 9. Spain (80)
  - 12. United States (79)
  - 41. Italy (29)

- Czech beer has been brewed since 993 AD
- now around 89 active breweries
- in pubs is the beer cheaper then coke (1 – 1.50 dollar per 0.5l)
About Ostrava

- the third largest city in the Czech Republic
- high quality black coal deposits
- steel industry and underground coal mines
- one of the most polluted in the European Union
- population: 312,000
About Ostrava

- The Lower Area of VÍTKOVICE
- today National Cultural Monument
- open as tourist district
About VŠB-TU Ostrava

- Vysoká Škola báňská - Technical University of Ostrava (High school of mining), founded in 1849
- originally founded to educate new mining engineers
- in 1970 was founded Faculty of Electrical Engineering and Computer Science
- in 1992 was founded Department of Applied mathematics by prof. Dostál
- solve large FEM problems in industry
About IT4Innovations

should be in *top 100* most powerful supercomputers in the world

funded by 85 percent from EU structural funds

Cluster for multidisciplinary computing:

- approx. 32768 cores with x86-64 architecture
- 32 cores per node, and total of 1024 nodes
- RAM 144GB per node
- estimated computing performance 866TFlops

Windows HPC cluster:

- approx. 4096 cores with x86-64 architecture
- 16 cores per node, and total of 256 nodes
- RAM 72GB per node
- estimated computing performance 68TFlops

Specialized GPU cluster:

- approx. 800 CPU + 24000 GPU cores,
- 16 CPU + 480 GPU cores per node and total of 50 nodes
- RAM 72GB per node
- estimated combined computing performance 35TFlops

will be finished in 2014
This is the end of funny part of my presentation. I am sorry about what comes next.
About QP problems

QP with convex constraints

\[
\min f(x) \quad \text{s.t. } x \in \Omega
\]

where

\- \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is (strictly) convex quadratic function

\[
f(x) := \frac{1}{2} x^T A x - b^T x
\]

with \( A \in \mathbb{R}^{n \times n} \) SPD, \( b \in \mathbb{R}^n \)

\- \( \Omega \subset \mathbb{R}^n \) is convex set described by
  \- \( x \leq l \) bound constraints
  \- \( l_1 \leq x \leq l_2 \) box constraints
  \- \( Bx \leq c \) linear inequalities
  \- \( Bx = c \) linear equalities
  \- \( x_{2i-1}^2 + x_{2i}^2 \leq r_i^2 \) separable quadratic constraints
  \- \( x_i^T H_i x_i \leq r_i^2 \) separable elliptic constraints
About QP problems

prof. RNDr. Zdeněk Dostál, DSc.

Optimal Quadratic Programming Algorithms With Applications to Variational Inequalities

Springer Optimization and Its Applications, Volume 23 (2009)
Simple example of contact problem I.

\[ -u''(x) = F(x) \]
\[ u(0) = u(1) = 0 \]

After using FEM

\[ Ku = f \]

equivalently

\[ \tilde{u} := \arg \min_{u \in \mathbb{R}^n} \frac{1}{2} u^T Ku - f^T u \]

with SPD stiffness matrix \( K \in \mathbb{R}^{n \times n} \) and vector of forces \( f \in \mathbb{R}^n \).
Simple example of contact problem II.

\[-u''(x) = F(x)\]
\[u(x) \geq l(x), \forall x \in (0, 1)\]
\[u(0) = u(1) = 0\]

After using FEM

\[\bar{u} := \arg\min_{u \in \Omega} \frac{1}{2} u^T Ku - f^T u\]
\[\Omega := \{ u \in \mathbb{R}^n : u_i \geq l_i, i = 1, \ldots n \}\]

with SPD stiffness matrix \(K \in \mathbb{R}^{n \times n}\), vector of forces \(f \in \mathbb{R}^n\) and convex feasible set \(\Omega \subset \mathbb{R}^n\).
FETI domain decomposition

Denote \( N \in \mathbb{N} \) as a number of domains.

\[
K := \begin{bmatrix}
K_1 \\
\vdots \\
K_N
\end{bmatrix} \in \mathbb{R}^{n_1+\cdots+n_N,n_1+\cdots+n_N}, \quad f := \begin{bmatrix} f_1 \\
\vdots \\
f_N \end{bmatrix} \in \mathbb{R}^{n_1+\cdots+n_N}
\]

\( u_i = u_j \) for border nodes, gluing conditions

QP with linear equality constraints

\[
\bar{u} := \arg \min_{u \in \Omega} \frac{1}{2} u^T Ku - f^T u
\]

\( \Omega := \{ u \in \mathbb{R}^n : u_i = u_j \} \)

with block-diagonal stiffness matrix \( K \), block vector of forces \( f \) and feasible set \( \Omega \).

Contact problem with given friction

- cantilever beam in mutual contact with rigid obstacle, prescribed given friction
- interested in displacement
- in dual formulation leads to QP with quadratic constraints.

[ Dostál Z., Kozubek T.: An optimal algorithm and superrelaxation for minimization of a quadratic function subject to separable convex constraints with applications, 2012. ]
Formulation of the QPQC problem

Primal problem

Find

\[ \bar{u} := \arg \min_{u \in \Omega_C} f(u) + j_h(u) \]

where

- \( \Omega_C := \{ u \in \Gamma_C : u_z \geq -d_C \} \) (non-penetration),
- \( f(u) := 1/2 \langle Ku, u \rangle - \langle f, u \rangle, f : \mathbb{R}^n \rightarrow \mathbb{R} \) (linear elasticity),
- \( j_h(u) := \sum_{i=1}^{m_c} \psi_i \| T_i u \|, j_h : \mathbb{R}^n \rightarrow \mathbb{R} \) (from Tresca friction law).

Dual formulation

\[ \bar{x} = \min_{x \in \Omega} \frac{1}{2} x^T A x - x^T b, \]
\[ \Omega := \{ x \in \mathbb{R}^n : \sqrt{x_{2i-1}^2 + x_{2i}^2} \leq r_i, \text{ for all } i = 1, \ldots, m \} \]

two cantilever beams in mutual contact with given friction
- different coefficients of friction in two orthogonal directions
- interested in displacement
- in dual formulation leads to QP with elliptic constraints.

QP with elliptic constraints

Problem formulation

Find

$$\bar{x} := \arg \min_{x \in \Omega} f(x)$$  \hspace{1cm} (1)

where

- \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is \textit{strictly convex quadratic function} given by

$$f(x) := \frac{1}{2} x^T A x - b^T x$$

- \( A \in \mathbb{R}^{n,n} \) is \textit{SPD}, \( b \in \mathbb{R}^n \)
- \( \Omega \subset \mathbb{R}^n \) is closed \textit{convex set} defined by separable elliptic constraints

Separable elliptic constraints

\( \Omega \subset \mathbb{R}^n, \Omega \neq \emptyset \)

\( \Omega := \Omega_1 \times \ldots \times \Omega_m \times \Omega_{\text{unconstrained}} \)

\( \Omega_j \subset \mathbb{R}^2, \Omega_j \neq \emptyset \)

\( \Omega_j := \{ x \in \mathbb{R}^2 : h_j(x) \leq 0 \}, \ j = 1, \ldots, m \)

\( h_j : \mathbb{R}^2 \rightarrow \mathbb{R} \)

\( h_j(x) := x^T H_j x - 1, \ H_j \in \mathbb{R}^{2,2}, H_j \) is \textit{SPD} \( h_j \) are \textit{separable}, \( \Omega_j \cap \Omega_i = \emptyset \)
MPGP (Modified proportioning with gradient projections) I.

For solving QP on convex set with inequalities.

Active and free sets

set of constraints indeces $\mathcal{M} := \{1, \ldots, m\}$

For every $x \in \Omega$ we define a decomposition of $\mathcal{M}$ into

<table>
<thead>
<tr>
<th>Active set</th>
<th>$\mathcal{A}(x) := {j \in \mathcal{M} : h_j(x_{\mathcal{I}_j}) = 0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free set</td>
<td>$\mathcal{F}(x) := {j \in \mathcal{M} : h_j(x_{\mathcal{I}_j}) &lt; 0}$</td>
</tr>
</tbody>
</table>

Free and chopped gradient

For every $x \in \Omega$ we decompose $g(x) := \nabla f(x) = Ax - b$ into

\[
\begin{align*}
\varphi_{\mathcal{I}_j}(x) &= g_{\mathcal{I}_j} \text{ for } j \in \mathcal{F}(x), \\
\beta_{\mathcal{I}_j}(x) &= 0 \text{ for } j \in \mathcal{A}(x), \\
\beta_{\mathcal{I}_j}(x) &= g_{\mathcal{I}_j}(x) - \min\{n_{\mathcal{I}_j}^T g_{\mathcal{I}_j}(x), 0\}. n_{\mathcal{I}_j} \text{ for } j \in \mathcal{A}(x)
\end{align*}
\]

and define projected gradient by

$g^P(x) = \varphi(x) + \beta(x)$.  

MPGP (Modified proportioning with gradient projections) I.

For solving QP on convex set with inequalities.

### Active and free sets

- set of constraints indices $\mathcal{M} := \{1, \ldots, m\}$

For every $x \in \Omega$ we define a decomposition of $\mathcal{M}$ into

- **Active set** $\mathcal{A}(x) := \{j \in \mathcal{M} : h_j(x_{Ij}) = 0\}$
- **Free set** $\mathcal{F}(x) := \{j \in \mathcal{M} : h_j(x_{Ij}) < 0\}$

### Free and chopped gradient

For every $x \in \Omega$ we decompose $g(x) := \nabla f(x) = Ax - b$ into

- $\varphi_{Ij}(x) = g_{Ij}$ for $j \in \mathcal{F}(x)$,
- $\varphi_{Ij}(x) = 0$ for $j \in \mathcal{A}(x)$,
- $\beta_{Ij}(x) = 0$ for $j \in \mathcal{F}(x)$,
- $\beta_{Ij}(x) = g_{Ij}(x) - \min\{n_{Ij}^T g_{Ij}(x), 0\}.n_{Ij}$ for $j \in \mathcal{A}(x)$

and define **projected gradient** by

$$g^P(x) = \varphi(x) + \beta(x).$$

MPGP (Modified proportioning with gradient projections) II.

\[ h_j(x) < 0 \implies j \in \mathcal{F}(x) \]
\[ \varphi_{\mathcal{I}_j}(x) = g_{\mathcal{I}_j} \]
\[ \beta_{\mathcal{I}_j}(x) = 0 \]

\[ h_j(x) = 0 \implies j \in \mathcal{A}(x) \]
\[ \varphi_{\mathcal{I}_j}(x) = 0 \]
\[ \beta_{\mathcal{I}_j}(x) = g_{\mathcal{I}_j}(x) - \min\{n_{\mathcal{I}_j}^T g_{\mathcal{I}_j}(x), 0\}.n_{\mathcal{I}_j} \]
$\angle(n, -g) \in (-\pi/2, \pi/2)$

$\beta_{IJ}(x) = g_{IJ}(x) - n_{IJ}^T g_{IJ}(x) \cdot n_{IJ}$

$\angle(n, -g) \in (\pi/2, 3\pi/2)$

$\beta_{IJ}(x) = g_{IJ}(x)$
Algorithm schema

- Initialization
  - Choose $x_0 \in \Omega$
- For $k = 0, 1, 2, \ldots$ do (while $\|g^P_k\|$ is not small)
  - if $\|\varphi(x_k)\| >> \|\beta(x_k)\|$ do (proportioning condition)
    - CG step or CG halfstep.
      - make one CG step to solve problem on free set.
      - if this step means leaving $\Omega$, do only a half-step and restart CG.
  - else do
    - Gradient projection step.
      - $x_{k+1} := P_{\Omega}(x_k - \alpha g_k)$
      - restart CG
- $k := k + 1$

[ Dostál, Z., Kozubek, T.: An optimal algorithm and superrelaxation for minimization of a quadratic function subject to separable convex constraints with applications, 2012. ]
CG step
make one CG step to solve problem on free set:
\[ \alpha_{cg} = g_k^T p_k / p_k^T A p_k \]
\[ x_{k+1} = x_k - \alpha_{cg} p_k \]
\[ g_{k+1} = g_k - \alpha_{cg} A p_k \]
\[ \gamma = \varphi (x_{k+1})^T A p_k / p_k^T A p_k \]
\[ p_{k+1} = \varphi (x_{k+1}) - \gamma p_k \]
CG half-step
if step with $\alpha_{cg}$ means leaving $\Omega$, do only a half-step and restart CG:

$$\alpha_f = \max \{ \alpha : x^k - \alpha p \in \Omega \} \quad \text{(if } \alpha_f < \text{alpha}_{cg} \text{ do)}$$

$$x_{k+1} = x_k - \alpha_f p_k$$

$$g_{k+1} = g_k - \alpha_f A p_k$$

$$p_{k+1} = \varphi(x_{k+1})$$
Projection step

- $x_{k+1} = P_\Omega(x_k - \alpha g_k)$
- $g_{k+1} = Ax_{k+1} - b$

(where $\alpha \in (0, 2\|A\|^{-1})$ is constant)
Descend of function values in projections

Let $\Omega$ be a closed convex subsymmetric set, let $\bar{x}$ denote the unique solution of (1), $x \in \Omega$, $\mu := 2\|A\|^{-1}$, and $\alpha \in (0, \mu)$. Then

$$f(P_\Omega(x - \alpha g)) - f(\bar{x}) \leq \eta(\alpha) (f(x) - f(\bar{x})),$$

where

$$\eta(\alpha) := \max\{1 - \alpha \lambda_{\text{min}}(A), 1 - (\mu - \alpha) \lambda_{\text{min}}(A)\}.$$

[ Dostál Z. and Schöberl J.: Minimizing quadratic functions subject to bound constraints with the rate of convergence and finite termination, 2005. ]

[ Dostál, Z., Kozubek, T.: An optimal algorithm and superrelaxation for minimization of a quadratic function subject to separable convex constraints with applications, 2012. ]

[ Bouchala J., Dostál Z., Vodstrčil P.: Separable spherical constraints and the decrease of a quadratic function in the gradient projection, 2013 ]
MPGP (Modified proportioning with gradient projections) VIII.

**Stopping criterion**

There is used $\|g^P(x_k)\| < \epsilon \|b\|$ because

$$\forall x \in \Omega : \ |x - \overline{x}|_A^2 \leq 2(f(x) - f(\overline{x})) \leq \|g^P\|_A^{-1} \leq \lambda_{\text{min}}(A)^{-1}.\|g^P\|^2$$

**Proportioning condition**

$$\|\beta(x_k)\| \leq \Gamma \|\phi(x_k)\|, \quad \Gamma^2 = \frac{1-2\delta}{2\delta}$$

$$2\delta g^T g^P \leq \|\phi(x_k)\|^2$$

where $\delta \in (0, 1/2)$ is proportioning parameter.

[ Dostál, Z., Kozubek, T.: *An optimal algorithm and superrelaxation for minimization of a quadratic function subject to separable convex constraints with applications*, 2012. ]

MPGP (Modified proportioning with gradient projections) IX.

Number of primal/dual variables: 8262/945
Active elliptic constraints in solution: 68%

MPGP (Modified proportioning with gradient projections) X.

The MPGP was developed for solving QP with spherical constraints. Numerical experiments show, it is not suitable for solving QP with badly scaled elliptic constraints. (the problem is in outer normal).

\[ \beta(x_k) = g_k - \min\{n^T g_k, 0\} \cdot n \]

MwPGP (Modified weakly proportioning with gradient projections) I.

Reduced projected gradient

\[ \tilde{\beta}^{\alpha}_{I_j}(x) = \begin{cases} 0 & \text{for } j \in \mathcal{F}(x), \\ g_{I_j}(x) & \text{for } j \in \mathcal{A}(x) \text{ and } n^T_{I_j}g_{I_j} > 0, \\ \frac{1}{\alpha}(x_{I_j} - \mathbf{P}_{\Omega_j}(x_{I_j} - \alpha g_{I_j})) & \text{for } j \in \mathcal{A}(x) \text{ and } n^T_{I_j}g_{I_j} \leq 0. \end{cases} \]

\[ \tilde{g}^P_{\alpha} : = \varphi(x) + \tilde{\beta}^{\alpha}(x) \]

Stopping criterion

We can use \( \|\tilde{g}^P_{\alpha}(x_k)\| < \epsilon.\|b\| \) because

\[ \forall x \in \Omega : \|\tilde{g}^P_{\alpha}(x)\|^2 \leq \|g^P(x)\|^2 \]

Proportioning condition

\[ \|\beta(x_k)\| \leq \Gamma \|\varphi(x_k)\|, \quad \Gamma^2 = \frac{1-2\delta}{2\delta} \]

\[ 2\delta\|\tilde{g}^P_{\alpha}(x_k)\|^2 \leq \|\varphi(x_k)\|^2 \]

where \( \delta \in (0, 1/2) \) is proportioning parameter.

spherical: [ Bouchala J., Dostál Z., Vodstrčil P.: Separable spherical constraints and the decrease of a quadratic function in the gradient projection, 2013]
MwPGP (Modified weakly proportioning with gradient projections) II.

Proves are based on special property of spherical (soon also elliptic) constraints.

---

**Subsymmetric sets**

A closed convex set $\Omega \subseteq \mathbb{R}^n$ is *subsymmetric* if for any $x \in \Omega$, $y \in \mathbb{R}^n$, $s := x - y$, and $\delta \in (0, 1)$

$$\|P_\Omega(y + \delta s) - y\| \geq \|P_\Omega(y - \delta s) - y\|.$$

---

It was already proved that half-intervals, spheres, halfspaces, and their products are subsymmetric, but not all convex sets are subsymmetric.

[ Bouchala J., Dostál Z., Vodstrčil P.: *Separable spherical constraints and the decrease of a quadratic function in the gradient projection*, 2013]

Elliptic constraints are also subsymmetric. See our new paper.

Algorithm schema

- **Initialization**
  - Choose \( x_0 \in \Omega \)
- For \( k = 0, 1, 2, \ldots \) do (while \( \| g_k^P \| \) is not small)
  - if \( \| \varphi(x_k) \| > > \| \beta(x_k) \| \) do (proportioning condition)
    - make one CG step to solve problem on free set.
    - if this step means leaving \( \Omega \), do only a half-step and restart CG.
  - else do
    - Proportioning step.
      - make one CG step to solve problem on active set.
      - restart CG on free set
  - \( k := k + 1 \)

[ Dostál, Z., Kozubek, T.: *An optimal algorithm and superrelaxation for minimization of a quadratic function subject to separable convex constraints with applications*, 2012. ]
SMALSE-M (semimonotonic augmented Lagrangian for separable and equality constrained QP problems) I.

- to find the minimizer of a strictly convex quadratic function subject to separable convex inequality and linear equality constraints

\[
\text{find } \bar{x} : = \arg \min_{x \in \Omega_{SE}} f(x), \quad \Omega_{SE} : = \{x \in \Omega : Cx = 0\}
\]

- use the augmented Lagrangian function

\[
L(x, \lambda, \varrho) : = \frac{1}{2} x^T (A + \varrho C^T C)x - x^T b + \lambda^T Cx
\]

- gradient of \( L(x, \lambda, \varrho) \) is given by

\[
\nabla_x L(x, \lambda, \varrho) = (A + \varrho C^T C)x - b + C^T \lambda
\]

- SMALSE-M is Uzawa-type algorithm:
  - find \( x_k \) corresponding to \( \lambda_k \)
    - (if \( \Omega = \mathbb{R}^n \))
      \[
      \text{solve } \nabla_x L(x, \lambda, \varrho) = 0
      \]
    - (if \( \Omega \) consists of linear inequalities)
      \[
      \text{solve } \arg \min_{x \in \Omega} \| \nabla_x L(x, \lambda, \varrho) \|
      \]
  - update Lagrange multipliers using \( x \) from previous step
    \[
    \lambda_{k+1} = \lambda_k + \rho Cx_k
    \]
Algorithm schema

- Initialization
  - Choose $\eta > 0$, $\beta > 1$, $M_0 > 0$, $\varrho > 0$, $\lambda^0 \in \mathbb{R}^m$, $\alpha \in (0, 2\|A\|^{-1})$
  - For $k = 0, 1, 2, \ldots$ do (while $\|\tilde{g}_\alpha^P(x^k)\|$ and $\|Cx^k\|$ are not small)
    - Step 1. {Inner iteration with adaptive precision control.}
      
      Find $x^k \in \Omega$ such that $\|\tilde{g}_\alpha^P(x^k, \lambda^k, \varrho)\| \leq \min\{M_k\|Cx^k\|, \eta\}$
    
    - Step 2. {Updating the Lagrange multipliers.}
      
      $\lambda^{k+1} = \lambda^k + \varrho C x^k$
    
    - Step 3. {Update $M$ provided the increase of the Lagrangian is not sufficient.}
      - if $k > 0$ and
      
      $L(x^k, \lambda^k, \varrho) < L(x^{k-1}, \lambda^{k-1}, \varrho) + \frac{\varrho}{2}\|Cx^k\|^2$

      set $M_{k+1} = M_k / \beta$
    - else set $M_{k+1} = M_k$

Numerical example I.

minimize \( q(x) = \frac{1}{2} \|x - a\|^2 \) subject to \( \|x\|_\infty \leq 1 \) and \( Bx = c \).

Fig.: Projection to the intersection of a cube with a hyperplane.

Fig.: Projection to the intersection of a cube with \( m \) hyperplanes.

Numerical example II.

<table>
<thead>
<tr>
<th>primal</th>
<th>dual</th>
<th>SPG</th>
<th>MPGP</th>
<th>MPGP-BB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>it</td>
<td>cg</td>
</tr>
<tr>
<td>375</td>
<td>75</td>
<td>36</td>
<td>9</td>
<td>44</td>
</tr>
<tr>
<td>1029</td>
<td>147</td>
<td>45</td>
<td>20</td>
<td>64</td>
</tr>
<tr>
<td>2187</td>
<td>243</td>
<td>27</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>3993</td>
<td>363</td>
<td>33</td>
<td>15</td>
<td>47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>primal</th>
<th>dual</th>
<th>SPG</th>
<th>MPGP</th>
<th>MPGP-BB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>it</td>
<td>cg</td>
</tr>
<tr>
<td>375</td>
<td>75</td>
<td>1566</td>
<td>977</td>
<td>2542</td>
</tr>
<tr>
<td>1029</td>
<td>147</td>
<td>923</td>
<td>553</td>
<td>1475</td>
</tr>
<tr>
<td>2187</td>
<td>243</td>
<td>588</td>
<td>366</td>
<td>953</td>
</tr>
<tr>
<td>3993</td>
<td>363</td>
<td>1020</td>
<td>547</td>
<td>1566</td>
</tr>
</tbody>
</table>

Future work

- finish new paper about weakly proportioning and MwPGP algorithm
- write the code fully parallel in C programming language:

**FLLOP (FETI Light Layer On PETSc)**
[ Hapla V., Horák D., Merta M.: *Use of direct solvers in TFETI massively parallel implementation*, LNCS, 2013 ]
[ Hapla V., Horák D.: *TFETI coarse problem massively parallel implementation*, ECCOMAS, 2012 ]

study adaptive step-size rules in projection steps
Thank you for your attention!