Relative accuracy in Quadratic Optimization with applications to Shape Design

Sergey Shpirko, PreMoLab (MFTI), Russia

June 19, 2013, Senezh (Premolab Workshop)

Joint work with Yu.Nesterov

Outline

Problem formulation

Problem: $f^* \stackrel{\text{def}}{=} \min_x \{ f(x) : \; Cx = b \}$,

where

- $f(x)$ is a homogeneous convex function:

$$f(x) = \max_s \{ \langle s, x \rangle : \; s \in Q \},$$

  with closed convex bounded Q such that $0 \in \operatorname{int} Q$,

- $C \in R^{p \times n}$ and $b \in R^p$, $b \neq 0$. Typically, p is small.

Thus, $f^* > 0$.

In order to solve this problem with *relative scale* we need to find an ellipsoid W and constants $0 < \gamma_1 \leq \gamma_2$ such that

$$\gamma_1 W \subseteq Q \subseteq \gamma_2 W$$

(see **N07**).

Finding the ellipsoid W

- *Simple sets:* can be done "by hand".
- *Q has a ν-self-concordant barrier $F(s)$.* Then we can take

$$W = \{s : \langle F''(0)s, s\rangle \le 1\} \subseteq Q.$$

In this case $\gamma_1 = 1$, $\gamma_2 = \nu + 2\sqrt{\nu}$.

- *Convex hull of many simple sets.* For example, for

$$f(x) = \max_{1 \le i \le m} |\langle a_i, x\rangle|$$

we have $Q = \mathrm{Conv}\{\pm a_i, i = 1, \ldots, m\}$. In this case, a good W can be found by special *Ellipsoid Algorithm* in $O(mn^2)$ operations (see **N08**).

This paper:

Consider the situation, typical for large-scale problems:

$$\begin{aligned} f(x) &= \max_{1 \le i \le m} f_i(A_i x), \\ f_i(u_i) &= \min_{\tau \ge 0} \{\tau : \ u_i \in \tau Q_i\}, \end{aligned}$$

where $A_i : R^n \to R^{m_i}$ are sparse, and Q_i are *simple*:

$$W_i \equiv \{u_i : \langle B_i u_i, u_i \rangle \le 1\} \subseteq Q_i \subseteq \bar{\gamma} W_i$$

for certain $B_i \succ 0$, $i = 1, \dots, m$.

Note: We need an *algorithm* for finding a good W for the set $Q = \bigcap\limits_{i=1}^{m} \{x : \ A_i x \in Q_i\}$. Equivalently, we need to find $B \succ 0$ such that

$$\frac{1}{\gamma_2} \langle Bx, x \rangle^{1/2} \le f(x) \le \frac{1}{\gamma_1} \langle Bx, x \rangle^{1/2}.$$

(This is a consequence of inclusion $\gamma_1 W \subseteq Q \subseteq \gamma_2 W$.)

This paper:

Proof.

Denote $f_A \stackrel{\text{def}}{=} \min_{\tau \geq 0}\{\tau : \ u \in \tau A\}$.

- $A \subseteq B \Rightarrow f_A \geq f_B$.
  Note: $u \in \tau A \Rightarrow u \in \tau B$.

- $$\gamma_1 W \subseteq Q \subseteq \gamma_2 W \Rightarrow$$
  $$\Rightarrow f_{\gamma_2 W} \leq f_Q \leq f_{\gamma_1 W}.$$

- $$f_{\gamma W} = \min_{\tau \geq 0}\{\tau : \ u \in \tau \gamma W\} =$$
  $$= \min_{\tau \geq 0}\{\tau : \ \langle Bu, u\rangle \leq \gamma^2 \tau^2\} \stackrel{\text{def}}{=} \tau^* = \frac{\langle Bu, u\rangle^{1/2}}{\gamma}.$$

Finding the good ellipsoid for Q

Main idea: consider

$$B(\xi) = \sum_{i=1}^{m} \xi_i A_i^T B_i A_i, \quad \xi \in R_+^m, \quad \sum_{i=1}^{m} \xi_i = 1.$$

Define $W(\xi) = \{x \in R^n : \langle B(\xi)x, x\rangle \leq 1\}$. Then for $x \in Q$ we have

$$\langle B(\xi)x, x\rangle = \sum_{i=1}^{m} \xi_i \langle B_i A_i x, A_i x\rangle \leq \bar{\gamma}^2 \sum_{i=1}^{m} \xi_i = \bar{\gamma}^2.$$

Thus, $\bar{\gamma} W(\xi) \supseteq Q$ for any $\xi \in \Delta_m$.

In order to ensure $Q \supseteq \gamma_1 W(\xi)$ for certain $\xi \in \Delta_m$ and $\gamma_1 > 0$, we need to apply a special adjustment process.

Adjustment process

0. Set $\xi_i = \frac{1}{m}$, $i = 1, \ldots, m$. Fix the tolerance parameter $\kappa > 1$.

1. kth iteration.

a). Compute $\rho_k = \max_{1 \le i \le m} \langle B^{-1}(\xi_k), A_i^T B_i A_i \rangle$. If $\rho_k \le \kappa n$, then STOP.

b). Define $i_k : \ \rho_k = \langle B^{-1}(\xi_k), A_{i_k}^T B_{i_k} A_{i_k} \rangle$. Compute

$$\alpha_k = \arg \max_{\alpha \in [0,1)} \ln \det \left((1-\alpha) B(\xi_k) + \alpha A_{i_k}^T B_{i_k} A_{i_k} \right).$$

c). Define $\xi_{k+1} = (1-\alpha_k)\xi_k + \alpha_k e_{i_k}$. Go to a).

Comments

- Optimization problem at Step b) can be written in "small" dimension:

$$\max_{\alpha\in[0,1)} \Big[\quad (n - m_{i_k}) \ln(1-\alpha) \\ + \ln\det\left((1-\alpha)B_{i_k}^{-1} + \alpha A_{i_k} B^{-1}(\xi_k) A_{i_k}^T\right)\Big].$$

- If matrix A_i is sparse, then the update of matrices $B(\xi_k)$ and $B^{-1}(\xi_k)$ need $O(n^2)$ operations.
- It is convenient to update the small matrices $A_i B^{-1}(\xi_k) A_i^T$, $i = 1, \dots, m$. Then implementation of Step a) is easy.

Convergence result

Theorem. 1. The adjustment process is stopped at most after $O(\frac{n}{\kappa-1}\ln m)$ iterations.

2. Complexity of each step does not exceed $O(mn)$ operations.

As a result, we get a matrix $B \equiv B(\xi_k)$ such that

$$(*) \qquad \langle B^{-1}, A_i^T B_i A_i \rangle \leq \kappa n, \quad i = 1, \ldots, m.$$

Lemma. $\frac{1}{\sqrt{\kappa n}} W(\xi_k) \subseteq Q \subseteq \bar{\gamma} W(\xi_k)$.

Convergence result

Proof.

- From (*) we have $\frac{1}{\kappa n} A_i^T B_i A_i \preceq B$.
  Note
  $$\langle B^{-1}, A \rangle \leq \sigma \Rightarrow \langle I, B^{-1/2} A B^{-1/2} \rangle \leq \sigma \Rightarrow$$
  $$\Rightarrow \lambda_{\max}(B^{-1/2} A B^{-1/2}) \leq \sigma \Rightarrow B^{-1/2} A B^{-1/2} \preceq \sigma I.$$
- Hence, if $x \in \frac{1}{\sqrt{\kappa n}} W(\xi_k)$, then
  $$\frac{1}{\kappa n} \geq \langle Bx, x \rangle \geq \frac{1}{\kappa n} \langle B_i A_i x, A_i x \rangle.$$
  Hence, $A_i x \in W_i \subseteq Q_i$. □

Coming back to our problem

Problem: $\min_{Cx=b} f(x),$

where the homogeneous objective can be approximated by the computed norm:

$$\bar{\gamma}^{-1}\|x\|_B \le f(x) \le \sqrt{\kappa n}\|x\|_B,$$

where $\|x\|_B = \langle Bx, x\rangle^{1/2}$.

Denote $\mathcal{L} = \{x : \ Cx = b\}$.

Subgradient method $G_N(R)$

for $k := 0$ **to** N **do**
begin Compute $f(x_k)$ and $g(x_k)$.

$$x_{k+1} := \pi_{\mathcal{L}} \left(x_k - \frac{R}{\sqrt{N+1}} \cdot \frac{g(x_k)}{\|g(x_k)\|_*} \right).$$

end.
Output: $G_N(R) = \arg\min\{f(x) : \; x = x_0, .., x_N\}$.
Theorem. *For a fixed* δ *from* $(0,1)$, *let us choose:*

$$N = \left\lfloor \frac{\bar{\gamma}^4 \kappa^2 n^2}{\delta^2} \right\rfloor.$$

Then, $f(G_N(\hat{\rho})) \leq (1+\delta) \cdot f^*$, *where* $\hat{\rho} \stackrel{\mathrm{def}}{=} f(x_0) \cdot \bar{\gamma}$.
see **N09**

Scheme with recursive strategy

$$\text{Let} \quad \hat{N} = \left\lfloor e\bar{\gamma}^2 \kappa n \left(1 + \frac{1}{\delta}\right)^2 \right\rfloor .$$

Set $\hat{x}_0 = x_0$.
For $t \geq 1$ **iterate**: $\hat{x}_t := G_{\hat{N}}\left(\bar{\gamma} f(\hat{x}_{t-1})\right);$

$$\textbf{if } f(\hat{x}_t) \geq \frac{1}{\sqrt{e}} f(\hat{x}_{t-1}), \textbf{ then } T := t \textbf{ and } \text{Stop} \qquad (*).$$

Theorem. *The number of points generated by the process (*) is bounded:*

$$T \leq 1 + 2\ln(\bar{\gamma}\sqrt{\kappa n}).$$

The total number of lower-level gradient steps in the process (*) does not exceed:

$$e \cdot \bar{\gamma}^2 \kappa n \cdot \left(1 + \frac{1}{\delta}\right)^2 \cdot \left(1 + 2\ln(\bar{\gamma}\sqrt{\kappa n})\right).$$

Shape Design: problem formulation

Given: type of material, total weight, and a set of external loads.

Find: the optimal shape able to resist to the load.

Finite Element Method

- The domain $\Omega \Rrightarrow$ many nonoverlapping cells $\Omega_1, .., \Omega_m$
- $T(x) = T_i$ for $\forall x \in \Omega_i, \ i = 1, .., m$
- A linear space V of virtual displacements of Ω is approximated by its finite dimensional subspace V^n.

Problem

$$\max_T \min_v \left\{ 1/2 \sum_{i=1}^m t_i \int_{\Omega_i} \langle T_i J(x), J(x) \rangle dx - \langle \bar{f}, v \rangle \right\}, \qquad (**)$$

$$T_i \succeq 0 \quad \langle T_i, I \rangle = 1 \quad \sum_{i=1}^m t_i = const.$$

Optimization problem

$$\max_{v}\left\{\frac{\langle \bar{f}, v\rangle^2}{\max_{T_i}\int_{\Omega_i}\langle T_i J(x), J(x)\rangle dx}: \quad T_i \geq 0 \ \langle T_i, I\rangle = 1\right\}.$$

$$\Downarrow$$

$$\max_{v\in Q}\langle \bar{f}, v\rangle, \quad Q \equiv \left\{v: \ \max_{T_i}\int_{\Omega_i}\langle T_i J(x), J(x)\rangle dx \leq 1\right\} \quad (***)$$

or

$$\min_{\langle \bar{f}, v\rangle = 1}\ \max_{T_i}\int_{\Omega_i}\langle T_i J(x), J(x)\rangle dx \quad (****)$$

2D-case

Let us consider an element with vertices $(0,0),(1,0),(0,1),(1,1)$ and their displacements v_3, v_4, v_1, v_2.

$$V(x) = x_1 \cdot x_2 \cdot v_2 + (1-x_1) \cdot x_2 \cdot v_1 + x_1 \cdot (1-x_2) \cdot v_4 + (1-x_2) \cdot (1-x_1) \cdot v_3$$

$$J(x) = \left(\frac{\partial V_1}{\partial x_1}, \frac{\partial V_2}{\partial x_2}, \frac{1}{\sqrt{2}} \left(\frac{\partial V_1}{\partial x_2} + \frac{\partial V_2}{\partial x_1} \right) \right)^T .$$

$$max_{T_i} \int_{\Omega_i} \langle T_i J(x), J(x) \rangle dx \equiv \max_{1 \le i \le m} \sigma_{i\,\max}(u),$$

$$u = \begin{pmatrix} \frac{-v_1^1+v_2^1-v_3^1+v_4^1}{2} & 0 & \frac{-v_1^1+v_2^1+v_3^1-v_4^1}{\sqrt{12}} \\ \frac{v_1^2+v_2^2-v_3^2-v_4^2}{2} & \frac{-v_1^2+v_2^2+v_3^2-v_4^2}{\sqrt{12}} & 0 \\ \frac{v_1^1+v_2^1-v_3^1-v_4^1-v_1^2+v_2^2-v_3^2+v_4^2}{2\sqrt{2}} & \frac{-v_1^1+v_2^1+v_3^1-v_4^1}{\sqrt{24}} & \frac{-v_1^2+v_2^2+v_3^2-v_4^2}{\sqrt{24}} \end{pmatrix}$$

2D-case

$$W_i = \{\|u\|_F^2 \leq 1\} \equiv \{\sum_{j=1}^3 \sigma_j^2(u) \leq 1\} \subseteq Q_i$$

(Same for all elements)

$$Q_i = \{u : \sigma_{max}(u) \leq 1\} \subseteq \sqrt{3}W$$

$$\bar{\gamma} = \sqrt{3}.$$