Mathematical Programming II	**Lecture 16**
ORIE 6310 Spring 2014	**March 20, 2014**
Scribe: Jialei Wang	

As promised in the last lecture, we now give the proof for Proposition 3:

Proof. Let $\overline{a} = \frac{a}{\sqrt{a^T B a}}$, so by Lemma 1,

$$-1 \leq \overline{a}^T(x-y) \leq +1 \quad \text{for } x \in E. \tag{1}$$

Now suppose $x \in E_\alpha$; then

$$(x-y)^T B^{-1}(x-y) \leq 1. \tag{2}$$

Also since

$$-1 \leq \overline{a}^T(x-y) \leq -\alpha \quad \text{by equation (1) and the definition of } E_\alpha, \tag{3}$$

we have

$$(\overline{a}^T(x-y)+\alpha)(\overline{a}^T(x-y)+1) \leq 0, \text{ or}$$
$$(x-y)^T\overline{a}\overline{a}^T(x-y) + (1+\alpha)\overline{a}^T(x-y) \leq -\alpha. \tag{4}$$

From (2) $\times (1-\sigma)$ + (4) $\times \sigma$, we get, for any $0 \leq \sigma \leq 1$,

$$(x-y)^T\left((1-\sigma)B^{-1} + \sigma\overline{a}\overline{a}^T\right)(x-y) + (1+\alpha)\sigma\overline{a}^T(x-y) \leq 1-\sigma-\sigma\alpha,$$
$$\Rightarrow (x-y+\frac{(1+\alpha)\sigma}{2}B\overline{a})^T((1-\sigma)B^{-1}+\sigma\overline{a}\overline{a}^T)(x-y+\frac{(1+\alpha)\sigma}{2}B\overline{a}) \leq 1-\sigma-\sigma\alpha+\frac{(1+\alpha)^2\sigma^2}{4}.$$

If we set $y_+ := y - \frac{(1+\alpha)\sigma}{2}B\overline{a}$ and

$$\begin{aligned}
B_+^{-1} &= \frac{1}{1-\sigma-\sigma\alpha+\frac{(1+\alpha)^2\sigma^2}{4}}((1-\sigma)B^{-1}+\sigma\overline{a}\overline{a}^T) \\
&= \frac{1-\sigma}{1-\sigma-\sigma\alpha+\frac{(1+\alpha)^2\sigma^2}{4}}\left(B^{-1}+\frac{\sigma}{1-\sigma}\overline{a}\overline{a}^T\right) \\
&= \frac{1-\sigma}{1-\sigma-\sigma\alpha+\frac{(1+\alpha)^2\sigma^2}{4}}\left(B-\frac{\frac{\sigma}{1-\sigma}B\overline{a}\overline{a}^T B}{1+\frac{\sigma}{1-\sigma}\overline{a}^T B\overline{a}}\right)^{-1} \quad \text{by the Sherman-Morrison-Woodbury formula,}
\end{aligned}$$

or

$$B_+ = \frac{1-\sigma-\sigma\alpha+\frac{(1+\alpha)^2\sigma^2}{4}}{1-\sigma}(B-\sigma B\overline{a}\overline{a}^T B),$$

this is $(x-y_+)^T B_+^{-1}(x-y_+) \leq 1$. Now plug in

$$\sigma = \frac{2(1+n\alpha)}{(1+n)(1+\alpha)} \geq 0$$

with

$$1-\sigma = \frac{n-1}{n+1}\cdot\frac{1-\alpha}{1+\alpha} \geq 0.$$

Then

$$\frac{(1+\alpha)\sigma}{2} = \frac{1+n\alpha}{1+n} = \tau,$$

and after some algebra,

$$\frac{1-\sigma-\sigma\alpha+\frac{(1+\alpha)^2\sigma^2}{4}}{1-\sigma} = \frac{(1-\alpha^2)n^2}{n^2-1} = \delta.$$

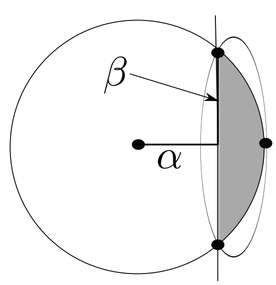
So $B_+ = \delta(B - \sigma B\overline{a}\overline{a}^T B)$ and $y_+ = y - \tau B\overline{a}$ as in the statement of the proposition. Hence $E_\alpha \subseteq E_+$. Also, its volume is

$$\begin{aligned}
\mathrm{vol}(E_+) &= \sqrt{\det B_+} \cdot \mathrm{vol}(\text{unit ball}) \\
&= \sqrt{\delta^n \cdot \det B \cdot (1 - \sigma\overline{a}^T B B^{-1} B\overline{a})} \cdot \mathrm{vol}(\text{unit ball}) \text{ (by Lemma 2)} \\
&= \mathrm{vol}(E)\left[\left(\frac{n^2}{n^2-1}\right)(1-\alpha^2)\right]^{\frac{n}{2}} \left(\frac{n-1}{n+1}\cdot\frac{1-\alpha}{1+\alpha}\right)^{\frac{1}{2}} \\
&= \mathrm{vol}(E)\left(\frac{n^2}{n^2-1}\right)^{\frac{n-1}{2}} (1-\alpha^2)^{\frac{n-1}{2}} \frac{n}{n+1}(1-\alpha).
\end{aligned}$$

If $\alpha \geq 0$, then

$$\begin{aligned}
\frac{\mathrm{vol}(E_+)}{\mathrm{vol}(E)} &\leq \left(1+\frac{1}{n^2-1}\right)^{\frac{n-1}{2}} \left(1-\frac{1}{n+1}\right) \\
&\leq \left[\exp\left(\frac{1}{n^2-1}\right)\right]^{\frac{n-1}{2}} \cdot \exp\left(-\frac{1}{n+1}\right) \\
&= \exp\left(\frac{1}{2(n+1)}\right) \cdot \exp\left(-\frac{1}{n+1}\right) \\
&= \exp\left(-\frac{1}{2(n+1)}\right).
\end{aligned}$$

Here is a sketch of the proof that this is the minimum-volume ellipsoid, in the case $y = 0, B = I, a = -e_1$.



Suppose we consider an **arbitrary** ellipsoid $\hat{E} := \{x : ||Mx - r|| \leq 1\}$ with volume $\frac{1}{\det M} \cdot \mathrm{vol}(\text{unit ball})$. Choose $\beta = \sqrt{1-\alpha^2}$, and consider the points

$$\alpha e_1 \pm \beta e_j, j = 2, \ldots, n$$

and e_1, all in E_α. So, if the columns of M are $m_1, \ldots, m_n$, $||m_1 - r|| \leq 1$ and $||\pm\beta m_j + \alpha m_1 - r|| \leq 1$. So $||\alpha m_1 - r|| =: \gamma \leq 1$, and then we can bound $||m_1||$ and each $||m_j||$ in terms of γ. But $\det M \leq ||m_1|| \cdot ||m_2|| \ldots ||m_n||$, so we get an upper bound on $\det M$; optimize over γ to get a universal bound, which shows E_+ has the minimum volume. □

Theorem 1. *If the ellipsoid method is applied to (f, G) where $G = \emptyset$ or $\mathrm{vol}(G) \geq \delta^n$, then if $z_k = *$ after $2n(n+1)\ln\frac{2\sqrt{n}}{\delta}$ steps, $G = \emptyset$, and otherwise, we get z_k with $\epsilon(z_k, f, G) \leq \epsilon$ in $2n(n+1)\ln\frac{2\sqrt{n}}{\epsilon\delta}$ steps.*

Proof. We know each (E_k, z_k) is a localizer. Also, $E_0 = B(nI, 0) = \{x : ||x|| \leq \sqrt{n}\}$ with $\mathrm{vol}(E_0) \leq (2\sqrt{n})^n$. By Proposition 3, every $2(n+1)$ steps, the volume of E_k is cut by e.
To get from volume $(2\sqrt{n})^n$ to δ^n, then, takes

$$2n(n+1)\ln(\frac{2\sqrt{n}}{\delta}) \text{ steps.}$$

Similarly, we get the volume smaller than $(\delta\epsilon)^n$ within $2n(n+1)\ln(\frac{2\sqrt{n}}{\delta\epsilon})$ steps. □

Comments

- If $G = C = [-1, 1]^n$, then we can get an ϵ-approximation solution in $2n(n+1)\ln(\frac{1}{\epsilon})$ steps. Exercise (use the fact that E_0 is the minimum-volume ellipsoid containing C).

- The ellipsoid method is much more general: it shows that "separation ≡ optimization." We will return to this.

- Forgetting about the details of the scalars, then at each step, the algorithm moves in the direction $-B_k a_k$ (if feasible, $a_k = g(x_k)$). This looks like

  - a steepest-descent step ($B_k = I$); or more like
  - a Newton step ($B_k = [\nabla^2 f(x_k)]^{-1}$); or even more like
  - a quasi-Newton step ($B_k \approx [\nabla^2 f(x_k)]^{-1}$, update at each iteration) with a rank-one update.

  This was the viewpoint of N. Shor.

- Proposition 3 can be used to show that every convex body (compact, non-empty interior) in $\mathbb{R}^n$ can be "n-rounded". There exist B, y such that

$$E(n^{-2}B, y) \subseteq C \subseteq E(B, y).$$

Note that the left-hand side is a copy of the right-hand side, shrunk by a factor of n around its center. This ratio is best possible: let C be a simplex in $\mathbb{R}^n$.