Mathematical Programming II	**Lecture 14**
ORIE 6310 Spring 2014	**March 13, 2014**
Scribe: Cory Girard	

Suppose we are given $(f, G) \in \mathcal{F}$, where f is convex on $C := [-1, 1]^n$, G is a convex subset of C with either $G = \emptyset$ or $\text{vol}(G) \geq \delta^n$, and $\max f(C) - \min f(C) \leq 1$.

Definition 1 *A pair (H, z) is a **localizer** for (f, G) if either $z = (*)$ and $G \subseteq H$ or $z \in G$ and $f(x) \leq f(z), x \in G \Longrightarrow x \in H$. So $\{x \in G : f(x) \leq f(z)\} \subseteq H$.*

For simplicity, define $f(*) := \infty$.

Proposition 1 *If (H, z) is a localizer for (f, G) with $\theta := \left(\frac{\text{vol}(H)}{\text{vol}(C)}\right)^{1/n}$, and $\theta < \frac{\delta}{2}$, then*
a) if $z = (*)$, then $G = \emptyset$;
b) if $z \in G$, then $\epsilon(z, f, G) \leq \frac{2\theta}{\delta}$.

Proof: $\text{vol}(H) = \theta^n \text{vol}(C) = (2\theta)^n < \delta^n$, where $\text{vol}(G) \geq \delta^n$ if $G \neq \emptyset$, so a) follows.

Now suppose $z \in G$. Let z_* be any minimizer of f over G, and consider $G(\epsilon) := \{(1-\epsilon)z_* + \epsilon x : x \in G\}$ for any $\epsilon > \frac{2\theta}{\delta}$. $\text{vol}(G(\epsilon)) = \epsilon^n \text{vol}(G) \geq (\epsilon\delta)^n$ while $\text{vol}(H) = (2\theta)^n < \text{vol}(G(\epsilon))$. So there is some $x \in G(\epsilon) \backslash H$. So $f(x) \geq f(z)$. Hence, for some $\hat{x} \in G$, we have

$$\begin{aligned}
f(z) &\leq f(x) \\
&= f((1-\epsilon)z_* + \epsilon\hat{x}) \\
&\leq (1-\epsilon)f(z_*) + \epsilon f(\hat{x}) \\
&= \min f(G) + \epsilon(f(\hat{x}) - f(z_*)) \\
&\leq \min f(g) + \epsilon.
\end{aligned}$$

Since $\epsilon > \frac{2\theta}{\delta}$ was arbitrary, $f(z) \leq \min f(G) + \frac{2\theta}{\delta}$. $\square$

The Method of Centers of Gravity (MCG) (D.J.Newman, A. Yu. Levin, 1965)

Start with the localizer $(H, z) = (C, (*))$. At iteration k, let x_k be the center of gravity of H_k, where (H_k, z_k) is the current localizer:

$$x_k = \frac{\int_{H_k} x d\lambda}{\int_{H_k} d\lambda}.$$

Call the oracle at x_k. If $x_k \notin G \cap \text{int}(C)$, and the oracle gives a separating hyperplane $G \subseteq \{x \in C : v_k^T x \leq v_k^T x_k\}$, then set $z_{k+1} = z_k$ and $a_k := v_k$. If $x_k \in G \cap \text{int}(C)$ and the origin gives $f(x_k)$ and $g(x_k) \in \partial f(x_k)$, then set $z_{k+1} = \text{argmin}\{f(x_k), f(z_k)\}$ and $a_k := g(x_k)$.

In either case, $H_{k+1} := \{x \in H_k : a_k^T x \leq a_k^T x_k\}$. Stop if $\left(\frac{\text{vol}(H_{k+1})}{\text{vol}(C)}\right)^{1/n} \leq \frac{\epsilon\delta}{2}$, or $\left(\frac{\text{vol}(H_{k+1})}{\text{vol}(C)}\right)^{1/n} < \frac{\delta}{2}$ and $z_{k+1} = (*)$.

Proposition 2 *In MCG, each (H_k, z_k) is a localizer.*

Proof: By induction on k; trivial for $k = 0$.

Assume true for k. If $x_k \notin G \cap \text{int}(C)$, then $z_{k+1} = z_k$ and $a_k = v_k$ with $G \subseteq \{x \in C : v_k^T x \leq v_k^T x_k\}$. So $G \backslash H_{k+1} = G \backslash H_k$. So if $x \in G \backslash H_{k+1}, x \in G \backslash H_k$, so $f(x) \geq f(z_k) = f(z_{k+1})$. If $x_k \in G \cap \text{int}(C)$, then we get $f(x_k)$ and $g(x_k) = a_k$. Then take any $x \in G \backslash H_{k+1}$; either $x \in G \backslash H_k$ so $f(x) \geq f(z_k) \geq f(z_{k+1})$, or $x \in H_k$ and $g(x_k)^T x \geq g(x_k)^T x_k$, so $f(x) \geq f(x_k) + g(x_k)^T (x - x_k) \geq f(x_k) \geq f(z_{k+1})$. □

Proposition 3 *(Grünbaum, Mityagin) If $D \subseteq \mathbf{R^n}$ is a convex compact set with center of gravity x, then for* ***any*** $0 \neq a \in \mathbf{R^n}$,

$$\text{vol}(\{y \in D : a^T y \leq a^T x\}) \leq \left(1 - \frac{n}{n+1}\right)^n \text{vol}(D) \leq \frac{e-1}{e} \text{vol}(D).$$

Theorem 1 *(Yudin and Nemirovski) If the method of centers of gravity performs $2.2n \ln \frac{2}{\delta}$ iterations and still has $z_k = (*), G = \emptyset$. If it produces $z_k \in G$, then within $2.2n \left(\ln \frac{2}{\delta} + \ln \frac{1}{\epsilon}\right)$ steps it produces z_k with $\epsilon(z_k, f, G) \leq \epsilon$.*

Proof: After k steps, we have localizer (H_k, z_k) with $\left(\frac{\text{vol}(H_k)}{\text{vol}(C)}\right)^{1/n} \leq \left(\frac{e-1}{e}\right)^{k/n}$. Note $\frac{1}{\ln \frac{e-1}{e}} < 2.2$. Then $\left(\frac{\text{vol}(H_k)}{\text{vol}(C)}\right)^{1/n} < \frac{2}{\delta}$ within $\frac{n \ln \frac{2}{\delta}}{\frac{e-1}{e}} < 2.2n \ln 2/\delta$ steps. Similarly, within $2.2n \left(\ln \frac{2}{\delta} + \ln \frac{1}{\epsilon}\right))$ steps, $\left(\frac{\text{vol}(H_k)}{\text{vol}(C)}\right)^{1/n} < \delta\epsilon$, so we have an ϵ-optimal z_k. □

In general, computing the center of gravity is"hard," so how can we circumvent it?

a) Use only nice sets H_k for which the center of gravity is easy to compute (e.g., the ellipsoid method).

b) Use a different notion of center, which is "easy to compute."

(a) leads to the *ellipsoid method* of Yudin and Nemirovski (1976) and Shor (1977). This is a simple modification of MCG: at every iteration, we have a localizer (E_k, z_k), with E_k an ellipsoid. We call the oracle at x_k, the center of the ellipsoid, and update z_k as above, but then set E_{k+1} to be the minimum volume ellipsoid containing

$$E_k^{1/2} := \{x \in E_k : a_k^T x \leq a_k^T x_k\}.$$

Questions: 1) How do we represent E_k?

2) How fast do the volumes of E_k's shrink?

1) $E_k = E(B_k, x_k) := \{x \in \mathbf{R^n} : (x - x_k)^T B_k^{-1} (x - x_k) \leq 1\}$ with some symmetric, positive definite B_k.

2) $\frac{\text{vol}(E_{k+1})}{\text{vol}(E_k)} < \exp\{-\frac{1}{2(n+1)}\}$.