Lecture 5.

Nonsmooth Minimization Methods.

– General Lower Complexity Bounds.

– Main Lemma.

– Localization Sets.

– Subgradient Method.

– Constrained Minimization Scheme.

– Optimization in finite dimension.

– Lower Complexity Bounds.

– Cutting Plane Scheme.

– Center of Gravity Method.

– Ellipsoid Method.

– Other Methods.

General Lower Complexity Bounds

Problem formulation:

$$\min_{x \in R^n} \; f(x).$$

Problem class: f is convex on R^n and Lipshitz continuous on some bounded set.

Oracle:
First-order black box: at each point $\hat{x}$ we can compute

$$f(\hat{x}), \quad g(\hat{x}) \in \partial f(\hat{x}),$$

where $g(\hat{x})$ is an *arbitrary* subgradient.

Approximate solution:
Find $\bar{x} \in R^n$ such that

$$f(\bar{x}) - f^* \leq \epsilon.$$

Methods:
Generate a sequence $\{x_k\}$:

$$x_k \in x_0 + \mathrm{Lin}\,\{g(x_0), \ldots, g(x_{k-1})\}.$$

Let us fix some constants $\mu > 0$, $\gamma > 0$.

Consider the family of functions

$$f_k(x) = \gamma \max_{1 \le i \le k} \; x^{(i)} + \frac{\mu}{2} \parallel x \parallel^2, \quad k = 1 \dots n.$$

1. The subdifferential of f_k at x is:

$$\partial f_k(x) = \mu x + \gamma \mathrm{Conv}\,\{e_i \mid i \in I(x)\},$$

$$I(x) = \{j \mid \; 1 \le j \le k, \; x^{(j)} = \max_{1 \le i \le k} \; x^{(i)}\}.$$

Therefore $\forall x, y \in B_2(0, \rho)$, $g(y) \in \partial f_k(y)$ we have

$$\begin{aligned} f_k(y) - f_k(x) \; &\le \langle g(y), y - x \rangle \le \parallel g(y) \parallel \cdot \parallel y - x \parallel \\ &\le (\mu\rho + \gamma) \parallel y - x \parallel . \end{aligned}$$

Thus, f_k is Lipshitz continuous on $B_2(0, \rho)$ with the constant

$$M = \mu\rho + \gamma.$$

2. The minimizer of f_k is

$$x_k^* : \quad (x_k^*)^{(i)} = \begin{cases} -\frac{\gamma}{\mu k}, \; 1 \le i \le k, \\ \quad 0, \; k + 1 \le i \le n. \end{cases}$$

Note that

$$R_k \; \equiv \parallel x_k^* \parallel = \tfrac{\gamma}{\mu\sqrt{k}},$$

$$f_k^* \; = -\tfrac{\gamma^2}{\mu k} + \tfrac{\mu}{2} R_k^2 = -\tfrac{\gamma^2}{2\mu k}.$$

3. **Resisting Oracle.** Denote

$$R^{p,n} = \{x \in R^n \mid \; x^{(i)} = 0, \; p+1 \le i \le n\}$$

Starting point: $x_0 = 0 \Rightarrow f_k(x_0) = 0, \; g(x_0) = e_1$.
Then $x_1 \in R^{1,n}$.

Let $x_i \in R^{p,n}$, $1 \le p \le k$.

If $\max\limits_{1 \le j \le p} x_i^{(j)} \ge 0$ **then:**

(A). Return $\quad f_k(x_i), \quad g(x_i) = \mu x_i + \gamma g_i,$

where g_i is an *arbitrary* vector from $\partial(\max\limits_{1 \le j \le p} x_i^{(j)})$.

else:

(B). Return $\quad f_k(x_i), \quad g(x_i) = \mu x_i + \gamma e_{p+1}.$

Note: Only Step (B) allows $x_{i+1} \in R^{p+1,n}$.

Therefore:

- $x_i \in R^{i,n}$, $1 \le i \le k$.
- For i: $1 \le i \le k-1$, we have:

$$f_k(x_i) \ge \gamma \max_{1 \le j \le k} x_i^{(j)} = 0.$$

Problem Class $\mathcal{P}(x_0, R, M)$:

- f is convex on R^n.
- $\| x_0 - x^* \| \leq R, \quad R > 0.$
- f is Lipshitz continuous on $B(x_0, R)$ with some constant $M > 0$.

Theorem 5.1 *For any class $\mathcal{P}(x_0, R, M)$ and any k, $0 \leq k \leq n-1$, there exists a function*

$$f \in \mathcal{P}(x_0, R, M)$$

such that

$$f(x_k) - f^* \geq \frac{MR}{2(1+\sqrt{k+1})}$$

for any method, generating a sequence $\{x_k\}$:

$$x_k \in x_0 + \mathrm{Lin}\,\{g(x_0), \ldots, g(x_{k-1})\}.$$

Proof. Let us choose $f(x) = f_{k+1}(x)$ with

$$\gamma = \tfrac{\sqrt{k+1}M}{1+\sqrt{k+1}}, \quad \mu = \tfrac{M}{(1+\sqrt{k+1})R}.$$

Then

$$f^* = f^*_{k+1} = -\tfrac{\gamma^2}{2\mu(k+1)} = -\tfrac{MR}{2(1+\sqrt{k+1})},$$

$$\| x_0 - x^* \| = R_{k+1} = \tfrac{\gamma}{\mu\sqrt{k+1}} = R,$$

and $f(x)$ is Lipshitz continuous on $B_2(x_0, R)$ with the constant

$$\mu R + \gamma = M.$$

Note that $x_k \in R^{k,n}$. Hence, $f(x_k) - f^* \geq -f^*$. □

Problem formulation

Consider the problem

$$\min\{f(x) \mid \; x \in Q\}, \tag{5.1}$$

where

- Q is a closed convex set,
- f is a function convex on R^n.

We are going to solve (5.1) using the subgradients $g(x_k)$.

Negative information:

- In general, the direction $-g(x)$ does not decrease the function f.
- We cannot expect $g(x) \to 0$ as x approaches x^*.

Positive information:

At any $x \in Q$ we have (see Corollary 7.4):

$$\langle g(x), x - x^* \rangle \geq 0. \tag{5.2}$$

Therefore:

- The direction $-g(x)$ decreases the distance between x and x^*.
- Inequality (5.2) cuts R^n on two half-spaces. Only one of them contains x^*.

Main Lemma

Let us fix some $\bar{x} \in R^n$.

For $x \in R^n$ with $g(x) \neq 0$ define

$$v_f(\bar{x}, x) = \frac{1}{\| g(x) \|} \langle g(x), x - \bar{x} \rangle, \quad (\leq \| x - \bar{x} \|).$$

If $g(x) = 0$, then define $v_f(\bar{x}; x) = 0$.

Interpretation: Let $\langle g(x), x - \bar{x} \rangle \geq 0$, $g(x) \neq 0$. Consider $y = \bar{x} + v_f(x) g(x) / \| g(x) \|$. Then

$$\begin{aligned} \langle g(x), x - y \rangle &= \langle g(x), x - \bar{x} \rangle \\ &\quad - v_f(\bar{x}, x) \| g(x) \| = 0, \\ \| y - \bar{x} \| &= v_f(\bar{x}, x). \end{aligned}$$

For $t \geq 0$ define

$$\mu_f(\bar{x}; t) = \max\{ f(x) - f(\bar{x}) \mid \| x - \bar{x} \| \leq t \}.$$

If $t < 0$, set $\mu_f(\bar{x}; t) = 0$.

Some properties:

- $\mu_f(\bar{x}; 0) = 0, \ t \leq 0$.
- $\mu_f(\bar{x}; t)$ is a non-decreasing function of t, $t \in R^1$.
- $f(x) - f(\bar{x}) \leq \mu_f(\bar{x}; \| x - x^* \|)$.

Lemma 5.1 *1. For any $x \in R^n$ we have:*

$$f(x) - f(\bar{x}) \le \mu_f(\bar{x}, v_f(\bar{x}; x)). \tag{5.3}$$

2. If $f(x)$ is Lipshitz continuous on $B_2(\bar{x}, R)$ with some constant M then

$$f(x) - f(\bar{x}) \le M(v_f(\bar{x}; x))_+. \tag{5.4}$$

for all $x \in R^n : v_f(\bar{x}; x) \le R$.

Proof. 1. If $\langle g(x), x - \bar{x} \rangle \le 0$, Then

$$f(\bar{x}) \ge f(x) + \langle g(x), \bar{x} - x \rangle \ge f(x)$$

and $v_f(\bar{x}; x) \le 0$.

Therefore, $\mu_f(v_f(\bar{x}; x)) = 0$ and (5.3) holds.

Let $\langle g(x), x - \bar{x} \rangle > 0$. For

$$y = \bar{x} + v_f(\bar{x}; x) g(x) / \parallel g(x) \parallel$$

we have $\langle g(x), y - x \rangle = 0$, $\parallel y - \bar{x} \parallel = v_f(\bar{x}; x)$.

Therefore $f(y) \ge f(x) + \langle g(x), y - x \rangle = f(x)$, and

$$\begin{aligned} f(x) - f(\bar{x}) \le f(y) - f(\bar{x}) &\le \mu_f(\bar{x}, \parallel y - \bar{x} \parallel) \\ &= \mu_f(\bar{x}, v_f(\bar{x}; x)). \end{aligned}$$

2. If $0 \le v_f(\bar{x}; x) \le R$, then $y \in B_2(\bar{x}, R)$. Hence,

$$\begin{aligned} f(x) - f(\bar{x}) \le f(y) - f(\bar{x}) &\le M \parallel y - \bar{x} \parallel \\ &= M v_f(\bar{x}; x). \end{aligned}$$

□

Localization Sets:

Let $\{x_i\}_{i=0}^k$ be a sequence in Q.

Define

$$S_k = \{x \in Q \mid \ \langle g(x_i), x_i - x\rangle \geq 0, \ i = 0 \dots k\}.$$

We call this set the *localization set* of problem (5.1) (generated by the sequence $\{x_i\}_{i=0}^k$).

Main inclusion: for all k we have

$$x^* \in S_k.$$

Denote

$$v_i = v_f(x^*; x_i) \ (\geq 0), \quad v_k^* = \min_{0 \leq i \leq k} \ v_i.$$

Thus,

$$v_k^* = \max\{r \mid \ \langle g(x_i), x_i - x\rangle \geq 0 \quad \forall x \in B_2(x^*, r), i = 0 \dots k\}.$$

Lemma 5.2 *Let*

$$f_k^* = \min_{0 \leq i \leq k} \ f(x_i).$$

Then

$$f_k^* - f^* \leq \mu_f(v^*, v_k^*).$$

Proof. Using Lemma 5.1, we have:

$$\mu_f(x^*, v_k^*) = \min_{0 \leq i \leq k} \mu_f(v_i) \geq \min_{0 \leq i \leq k} [f(x_i) - f^*] = f_k^* - f^*.$$

□

Subgradient Method

Problem:

$$\min\{f(x) \mid \ x \in Q\}, \tag{5.5}$$

where

- Q is a *simple* closed convex set,
- f is a function convex on R^n.

Oracle: $\bar{x} \quad \Rightarrow \quad f(\bar{x}), \ g(\bar{x})$.

Note: $\| \ g(\bar{x}) \ \|$ is not very informative.

Therefore we can use the *direction* $g(\bar{x}) / \| \ g(\bar{x}) \ \|$.

Scheme:

0. Choose $x_0 \in Q$ and a sequence $\{h_k\}_{k=0}^{\infty}$:

$$h_k \geq 0, \quad h_k \to 0.$$

1. kth iteration ($k \geq 0$).

   a). Compute $f(x_k)$, $g(x_k)$.

   b). Set

$$x_{k+1} = \pi_Q \left(x_k - h_k \frac{g(x_k)}{\| \ g(x_k) \ \|} \right). \tag{5.6}$$

Theorem 5.2 *Let f be Lipshitz continuous on the ball $B_2(x^*, R)$ with the constant M and $\| x_0 - x^* \| \leq R$. Then*

$$f_k^* - f^* \leq M \frac{R^2 + \sum\limits_{i=0}^{k} h_i^2}{2 \sum\limits_{i=0}^{k} h_i}. \tag{5.7}$$

Proof. Denote $r_i = \| x_i - x^* \|$. Then (see L.4.5)

$$\begin{aligned} r_{i+1}^2 &= \left\| \pi_Q \left(x_i - h_i \frac{g(x_i)}{\|g(x_i)\|} \right) - x^* \right\|^2 \\ &\leq \left\| x_i - h_i \frac{g(x_i)}{\|g(x_i)\|} - x^* \right\|^2 \\ &= r_i^2 - 2h_i v_i + h_i^2. \end{aligned}$$

Summurazing these inequalities for $i = 0 \ldots k$ we get:

$$r_0^2 + \sum_{i=0}^{k} h_i^2 = 2 \sum_{i=0}^{k} h_i v_i + r_{k+1}^2 \geq 2 v_k^* \sum_{i=0}^{k} h_i.$$

Thus,

$$v_k^* \leq \frac{R^2 + \sum\limits_{i=0}^{k} h_i^2}{2 \sum\limits_{i=0}^{k} h_i}.$$

It remains to use Lemma 5.2. □

Step-Size Strategies

Denote $\Delta_k = \frac{R^2 + \sum\limits_{i=0}^{k} h_i^2}{2\sum\limits_{i=0}^{k} h_i}$.

1. If $\sum\limits_{i=0}^{\infty} h_i = \infty$ and $h_i \to 0$ then $\Delta_k \to 0$.

2. Let us fix N, the number of steps. Then the optimal strategy is

$$h_i = \frac{R}{\sqrt{N+1}}, \quad i = 0 \ldots N. \tag{5.8}$$

In this case

$$\Delta_N = \frac{R}{\sqrt{N+1}}.$$

Using Theorem 5.1 we conclude:

The subgradient method (5.6), (5.8) is optimal for the problem (5.5) (uniformly in the dimension n).

3. If we take

$$h_i = \frac{r}{\sqrt{i+1}},$$

the Δ_k is proportional to

$$\frac{R^2 + r\ln(k+1)}{2r\sqrt{k+1}}$$

(sub-optimal rate).

Minimization with functional constraints

Problem:

$$\min\{f(x) \mid \; x \in Q,\; f_j(x) \leq 0, i = 1 \ldots m\}, \qquad (5.9)$$

where

- Q is a simple bounded closed convex set,

$$\| x - y \| \leq R, \quad x, y \in Q.$$

- f and f_j are convex on R^n.

Define $\bar{f}(x) = \left(\max\limits_{1 \leq j \leq m} \; f_j(x) \right)_+$. Then our problem is:

$$\min\{f(x) \mid \; x \in Q,\; \bar{f}(x) \leq 0\}.$$

Let us fix some x^*, a solution to (5.9).

Note that $\bar{f}(x^*) = 0$ and $v_{\bar{f}}(x^*; x) \geq 0 \; \forall x \in R^n$.

Therefore, in view of Lemma 5.1 we have:

$$\bar{f}(x) \leq \mu_{\bar{f}}(x^*; v_{\bar{f}}(x^*; x)).$$

If f_j are Lipshitz continuous on Q with constant M then $\forall x \in R^n$

$$\bar{f}(x) \leq M \cdot v_{\bar{f}}(x^*; x).$$

Constrained Minimization Scheme

Assumption: $\operatorname{diam} Q \leq R$.

0. Choose $x_0 \in Q$ and the sequence $\{h_k\}_{k=0}^{\infty}$:

$$h_k = \frac{R}{\sqrt{k+0.5}}.$$

1. kth iteration ($k \geq 0$).

   a). Compute $f(x_k)$, $g(x_k)$, $\bar{f}(x_k)$, $\bar{g}(x_k)$ and set

$$p_k = \begin{cases} g(x_k), & \text{if } \bar{f}(x_k) < \parallel \bar{g}(x_k) \parallel h_k, \ (A), \\ \bar{g}(x_k), & \text{if } \bar{f}(x_k) \geq \parallel \bar{g}(x_k) \parallel h_k, \ (B). \end{cases}$$

   b). Set

$$x_{k+1} = \pi_Q\left(x_k - h_k \frac{p_k}{\parallel p_k \parallel}\right). \qquad (5.10)$$

Theorem 5.3 *Let f be Lipshitz continuous on $B_2(x^*, R)$ with constant M_1 and*

$$M_2 = \max_{1 \leq j \leq m} \{\parallel g \parallel: \ g \in \partial f_j(x), \ x \in B_2(x^*, R)\}.$$

Then for all $k \geq 3$ we have: $\exists i', 0 \leq i' \leq k$, such that

$$f(x_{i'}) - f^* \leq \frac{\sqrt{3} M_1 R}{\sqrt{k-1.5}}, \quad \bar{f}(x_{i'}) \leq \frac{\sqrt{3} M_2 R}{\sqrt{k-1.5}}.$$

Without proof.

Complexity Bounds in Finite Dimension

Problem formulation:

$$\min_{x \in R^n} f(x). \tag{5.11}$$

Problem class: f is convex on R^n and Lipshitz continuous on some bounded set.
Dimension n is relatively *small*.

Oracle:
First-order black box: at each point $\hat{x}$ we can compute

$$f(\hat{x}), \quad g(\hat{x}) \in \partial f(\hat{x}),$$

where $g(\hat{x})$ is an *arbitrary* subgradient.

Approximate solution:
Find $\bar{x} \in R^n$ such that

$$f(\bar{x}) - f^* \leq \epsilon.$$

Methods:
Generate a sequence $\{x_k\}$:

$$x_k \in x_0 + \text{Lin}\,\{g(x_0), \ldots, g(x_{k-1})\}.$$

Similar problem:

$$\text{Find } x^* \in Q, \tag{5.12}$$

where Q is a convex set.

Oracle:

For any $\bar{x} \in R^n$ there are two possible answers:

- $\bar{x} \in Q$.
- Return $\bar{g}$:

$$\langle \bar{g}, \bar{x} - x \rangle \geq 0 \quad \forall x \in Q.$$

Assumption:

$$\exists x^* \in Q : \quad B_2(x^*, \epsilon) \subseteq Q \tag{5.13}$$

for some $\epsilon > 0$.

Example:

If we know f^* in the problem (5.11), then

$$\bar{Q} = \{(t, x) \in R^{n+1} \mid \ t \geq f(x), \ t \leq f^* + \bar{\epsilon}\}.$$

(Question:

What is the relation between $\bar{\epsilon}$ and ϵ in (5.13)?)

Resisting Oracle for (5.12)

It forms a sequence of boxes $\{B_k\}_{k=0}^{\infty}$:

$$B_k = \{x \in R^n \mid \; a_k \leq x \leq b_k\}$$

with the centers $c_k = \frac{1}{2}(a_k + b_k)$.

Initialize: $a_0 := -Re$, $b_0 := Re$, $m := 0$, $i := 1$.

Procedure Sep($x \in R^n$ **).**

If $x \notin B_0$ **then** Return a separator of x from B_0.

else

1. Find the maximal $k \in [0 \dots m] : \; x \in B_k$.
2. **If** $k < m$ **then** Return g_k

   else { Generate a new box }

   If $x^{(i)} \geq c_m^{(i)}$ **then**

   $b_{m+1} := b_m + (c_m^{(i)} - b_m^{(i)})e_i;$
   $a_{m+1} := a_m;\; g_m := e_i.$

   else $a_{m+1} := a_m + (c_m^{(i)} - a_m^{(i)})e_i;$

   $b_{m+1} := b_m;\; g_m := -e_i.$

   endif

   $m := m + 1;$

   $i := i + 1;$ **If** $i > n$ **then** $i := 1$.

   Return g_m;

   endif

endif;

Note:

- $B_{k+1} \subset B_k$.
- $\mathrm{vol}_n B_{k+1} = \frac{1}{2}\mathrm{vol}_n B_k$.
- For any $k \geq 0$ we have:

$$b_{k+n} - a_{k+n} = \tfrac{1}{2}(b_k - a_k).$$

Lemma 5.3 *For all $k \geq 0$ we have the inclusion:*

$$B_k \supset B_2(c_k, r_k), \quad r_k = \frac{R}{2}\left(\frac{1}{2}\right)^{-\frac{k}{n}}. \tag{5.14}$$

Proof. Indeed, for all $k \in [0 \dots n-1]$ we have

$$\begin{aligned} B_k &\supset B_n = \{x \mid \; c_n - \tfrac{1}{2}Re \leq x \leq c_n + \tfrac{1}{2}Re\} \\ &\supset B_2(c_n, \tfrac{1}{2}R). \end{aligned}$$

Therefore, $B_k \supset B_2(c_k, \frac{1}{2}R)$ and (5.14) holds.

Let $k = nl + p$, $p \in [0 \dots n-1]$. Since

$$b_k - a_k = \left(\tfrac{1}{2}\right)^{-l}(b_p - a_p),$$

we conclude that

$$B_k \supset B_2(c_k, \tfrac{1}{2}R\left(\tfrac{1}{2}\right)^{-l}).$$

It remains to note that $r_k \leq \frac{1}{2}R\left(\frac{1}{2}\right)^{-l}$. □

Theorem 5.4 *The lower complexity bound for the feasibility problem (5.12), (5.13) with*

$$Q \subseteq B_\infty(0, R)$$

is as follows:

$$n \ln \frac{R}{2\epsilon}$$

calls of the oracle.

Proof. We have seen that:

- The number of generated boxes does not exceed the number of calls of the oracle.
- After k iterations the last box contains the ball $B_2(c_{m_k}, r_k)$.

□

Theorem 5.5 *The lower complexity bound for the minimization problem (5.11) with*

$$Q \subseteq B_\infty(0, R)$$

and $f \in \mathcal{F}_M^{0,0}(B_\infty(0, R))$*, is as follows:*

$$n \ln \frac{MR}{8\epsilon}$$

calls of the oracle.

(The structure of the resisting oracle is similar.)

Cutting Plane Scheme

Problem:

$$\min\{f(x) \mid\ x \in Q\}, \tag{5.15}$$

where

- Q is a closed convex set such that

$$\operatorname{int} Q \neq \emptyset, \quad \operatorname{diam} Q = D < \infty.$$

- f is a function convex on R^n.

Q is not "simple" anymore!

Oracle:
For any $\bar{x} \in R^n$ it returns a vector g which is:

- a subgradient of f at $\bar{x}$, if $x \in Q$,
- a separator of $\bar{x}$ from Q, if $x \notin Q$.

Example:
If $Q = \{x \in R^n \mid\ \bar{f}(x) \leq 0\}$ then for $x \notin Q$ any

$$\bar{g} \in \partial \bar{f}(x)$$

separates x from Q.

Localization Sets in Finite Dimension

Let $X \equiv \{x_i\}_{i=0}^{\infty}$ be a sequence in Q.

Consider a sequence of bounded sets:

$$\begin{aligned} S_0(X) &= Q, \\ S_{k+1}(X) &= \{x \in S_k(X) \mid \ \langle g(x_k), x_k - x \rangle \geq 0\}. \end{aligned}$$

Main inclusion: for all k we have: $x^* \in S_k$.

Denote $\quad v_i = v_f(x^*; x_i) \ (\geq 0), \quad v_k^* = \min\limits_{0 \leq i \leq k} v_i$.

Theorem 5.6

$$v_k^* \leq D \left[\frac{\mathrm{vol}_n S_k(X)}{\mathrm{vol}_n Q} \right]^{\frac{1}{n}}.$$

Proof. Let $\alpha = v_k^*/D \ (\leq 1)$. Since $Q \subseteq B_2(x^*, D)$ and Q is convex, we conclude that

$$\begin{aligned} (1-\alpha)x^* + \alpha Q &\subseteq (1-\alpha)x^* + \alpha B_2(x^*, D) \\ &= B_2(x^*, v_k^*), \\ (1-\alpha)x^* + \alpha Q &\equiv [(1-\alpha)x^* + \alpha Q] \cap Q \\ &\subseteq B_2(x^*, v_k^*) \cap Q \subseteq S_k(X). \end{aligned}$$

Therefore

$$\mathrm{vol}_n S_k(X) \geq \mathrm{vol}_n [(1-\alpha)x^* + \alpha Q] = \alpha^n \mathrm{vol}_n Q.$$

□

Note: Usually the sets $S_k(X)$ cannot be treated explicitly.

Realistic Scheme:

0. Choose a bounded set $E_0 \supseteq Q$.

1. kth iteration ($k \geq 0$).

   a). Choose $y_k \in E_k$

   b). If $y_k \in Q$ then compute $f(y_k)$, $g(y_k)$.

   If $y_k \notin Q$ then compute $\bar{g}(y_k)$, which separates y_k from Q.

   c). Set

$$g_k = \begin{cases} g(y_k), & \text{if } y_k \in Q, \\ \bar{g}(y_k), & \text{if } y_k \notin Q. \end{cases}$$

   d). Choose

$$E_{k+1} \supseteq \{x \in E_k \mid \langle g_k, y_k - x \rangle \geq 0\}.$$

Let $Y = \{y_k\}_{k=0}^{\infty}$, $X = Y \cap Q$. Define

$$i(k) = \text{ number of } y_j, 0 \le j < k, \\ \text{such that } y_j \in Q.$$

Thus, if $i(k) > 0$ then $X \neq \emptyset$.

Lemma 5.4 *For any $k \ge 0$ we have:*

$$S_{i(k)} \subseteq E_k.$$

Proof. 1. $i(0) = 0 \Rightarrow S_0 = Q \subseteq E_0$.

2. Let $S_{i(k)} \subseteq E_k$ for some $k \ge 0$. There are two possibilities:

a). $i(k+1) = i(k) \Leftrightarrow y_k \notin Q$. Then

$$\begin{aligned} E_{k+1} &\supseteq \{x \in E_k \mid \langle \bar{g}(y_k), y_k - x\rangle \ge 0\} \\ &\supseteq \{x \in S_{i(k+1)} \mid \langle \bar{g}(y_k), y_k - x\rangle \ge 0\} \\ &= S_{i(k+1)} \end{aligned}$$

since $S_{i(k+1)} \subseteq Q$ and $\bar{g}(y_k)$ separates y_k from Q.

b). $i(k+1) = i(k) + 1 \Leftrightarrow y_k \in Q$. Then

$$\begin{aligned} E_{k+1} &\supseteq \{x \in E_k \mid \langle g(y_k), y_k - x\rangle \ge 0\} \\ &\supseteq \{x \in S_{i(k)} \mid \langle g(y_k), y_k - x\rangle \ge 0\} \\ &= S_{i(k)+1} \end{aligned}$$

since $y_k = x_{i(k)}$. □

Corollary 5.1 *1. For any $k : i(k) > 0$ we have:*

$$v^*_{i(k)}(X) \le D \left[\frac{\mathrm{vol}_n S_{i(k)}(X)}{\mathrm{vol}_n Q} \right]^{\frac{1}{n}} \le D \left[\frac{\mathrm{vol}_n E_k}{\mathrm{vol}_n Q} \right]^{\frac{1}{n}}.$$

2. If $\mathrm{vol}_n E_k < \mathrm{vol}_n Q$ *then* $i(k) > 0$.

Proof. 1) is already proved and 2) follows from the inclusion:

$$Q = S_0 = S_{i(k)} \subseteq E_k$$

for all $k : i(k) = 0$. □

Conclusion:

If we manage to ensure $\mathrm{vol}_n E_k \to 0$, then we obtain a convergent method.

We need to decrease $\mathrm{vol}_n E_k$ as fast as possible.

Center of Gravity Method

Let S be a bounded convex set in R^n, $\mathrm{int}\, S \neq \emptyset$.

Define

$$cg(S) = \frac{1}{\mathrm{vol}_n S} \int_S x dx.$$

This point is called *the center of gravity* of the set S.

Lemma 5.5 *Let g be a direction in R^n. Define*

$$S_+ = \{x \in S \mid \langle g, cg(S) - x \rangle \geq 0\}.$$

Then

$$\frac{\mathrm{vol}_n S_+}{\mathrm{vol}_n S} \leq 1 - \frac{1}{e}.$$

(Accept without proof.)

Scheme:

0. Set $S_0 = Q$.

1. kth iteration ($k \geq 0$).

   a). Choose $x_k = cg(S_k)$.

   b). Compute $f(x_k)$, $g(x_k)$.

   c). Choose

$$S_{k+1} = \{x \in S_k \mid \langle g(x_k), x_k - x \rangle \geq 0\}.$$

Denote $f_k^* = \min_{0 \le j \le k} f(x_j)$.

Theorem 5.7 *If f is Lipshitz continuous on the ball $B_2(x^*, D)$ with some constant M, then for any $k \ge 0$ we have:*

$$f_k^* - f^* \le MD\left(1 - \frac{1}{e}\right)^{-\frac{k}{n}}.$$

Proof: Follows from L.5.2, T.5.6 and L.5.5. □

Conclusion:

1. This method is optimal in finite dimension (see T.5.5).

2. Its rate of convergence depends on n only.

3. It is absolutely impractical since the computation of $cg(S)$ is more difficult than our initial problem.

Ellipsoid Method

Let H be a positive definite symmetric $n \times n$ matrix. Denote

$$E(H, \bar{x}) = \{x \in R^n \mid \langle H^{-1}(x - \bar{x}), x - \bar{x}\rangle \le 1\}.$$

Let g be a direction in R^n. Consider

$$E_+ = \{x \in E(H, \bar{x}) \mid \langle g, \bar{x} - x\rangle \ge 0\}.$$

Lemma 5.6 *Let*

$$\bar{x}_+ = \bar{x} - \frac{1}{n+1} \cdot \frac{Hg}{\langle Hg, g\rangle^{1/2}},$$

$$H_+ = \frac{n^2}{n^2 - 1}\left(H - \frac{2}{n+1} \cdot \frac{Hgg^T H}{\langle Hg, g\rangle}\right).$$

Then $E_+ \subset E(H_+, \bar{x}_+)$ *and*

$$\operatorname{vol}_n E(H_+, \bar{x}_+) \le \left(1 - \frac{1}{(n+1)^2}\right)^{\frac{n}{2}} \operatorname{vol}_n E(H, \bar{x}).$$

(Accept without proof.)

Note: $E(H_+, \bar{x}_+)$ is the ellipsoid of the *minimal* volume containing E_+.

Algorithmic scheme:

0. Choose $y_0 \in R^n$ and $R > 0$:
$$B_2(y_0, R) \supseteq Q.$$
Set $H_0 = R^2 I_n$.

1. kth iteration ($k \geq 0$).

   a). If $y_k \in Q$ then compute $f(y_k)$, $g(y_k)$.
   If $y_k \notin Q$ then compute $\bar{g}(y_k)$, which separates y_k from Q.

   b). Set
$$g_k = \begin{cases} g(y_k), & \text{if } y_k \in Q, \\ \bar{g}(y_k), & \text{if } y_k \notin Q. \end{cases}$$

   c). Set
$$\begin{aligned} y_{k+1} &= y_k - \tfrac{1}{n+1} \cdot \tfrac{H_k g_k}{\langle H_k g_k, g_k \rangle^{1/2}}, \\ H_{k+1} &= \tfrac{n^2}{n^2-1} \left(H_k - \tfrac{2}{n+1} \cdot \tfrac{H_k g_k g_k^T H_k}{\langle H_k g_k, g_k \rangle} \right). \end{aligned}$$

For this method
$$E_k = \{ x \in R^n \mid \ \langle H_k^{-1}(x - x_k), x - x_k \rangle \leq 1 \}.$$

Denote $X = Y \cap Q$, $f_k^* = \min_{0 \le j \le k} f(x_j)$.

Theorem 5.8 *Let f be Lipshitz continuous on the ball $B_2(x^*, R)$ with some constant M.*

Then for $i(k) > 0$ we have:

$$f_{i(k)}^* - f^* \le MR\left(1 - \frac{1}{(n+1)^2}\right)^{\frac{k}{2}} \cdot \left[\frac{\mathrm{vol}_n B_0(x_0, R)}{\mathrm{vol}_n Q}\right]^{\frac{1}{n}}.$$

Proof: Follows from L.5.2, C.5.1 and L.5.6. □

If $Q \supseteq B_2(\bar{x}, \rho)$ then

$$\left[\frac{\mathrm{vol}_n B_2(x_0, R)}{\mathrm{vol}_n Q}\right]^{\frac{1}{n}} \le \frac{R}{\rho}$$

and we obtain:

$$f_{i(k)}^* - f^* \le \frac{MR^2}{\rho}\left(1 - \frac{1}{(n+1)^2}\right)^{\frac{k}{2}} \le \frac{1}{\rho} MR^2 \cdot e^{-\frac{k}{2(n+1)^2}}.$$

Note that $i(k) > 0$ for

$$k > 2(n+1)^2 \ln \frac{R}{\rho}.$$

Conclusion:

1. Each iteration of this method is very cheap: only $O(n^2)$ a.o.

2. Its efficiency estimate is

$$2(n+1)^2 \ln \frac{MR^2}{\rho\epsilon}.$$

It is not optimal (see T.5.5), but it has

- linear dependence on $\ln \frac{1}{\epsilon}$,
- polynomial dependence on n and the logarithms of the class parameters (M, R, ρ).

For problem classes, in which the oracle has polynomial complexity, such algorithms are called *polynomial.*

Other Methods

There are several methods, which work with the localization sets in the form

$$E_k = \{x \in R^n \mid \langle a_j, x\rangle \le b_j,\ j = 1 \ldots m_k\}.$$

1. Inscribed Ellipsoid Method:

$$y_k = \text{Center of the maximal ellipsoid } W_k :$$

$$W_k \subset E_k.$$

2. Analytic Center Method:

$$y_k = \arg\min\{F_k(x) \mid \ x \in \operatorname{int} E_k\},$$

$$F_k(x) = -\sum_{j=1}^{m_k} \ln(b_j - \langle a_j, x\rangle).$$

3. Volumetric Center Method:

$$y_k = \arg\min\{V_k(x) \mid \ x \in \operatorname{int} E_k\},$$

$$V_k(x) = \ln\det F_k''(x).$$

These methods are polynomial with the complexity

$$n\left(\ln\frac{1}{\epsilon}\right)^p, \quad p = 1 \text{ or } 2.$$

The complexity of their steps is larger: $n^3 - n^4$ a.o.