Lecture 9.

Methods with Complete Data.

– Model of Nonsmooth Function.

– Kelly Method.

– Idea of Level Method.

– Unconstrained Minimization.

– Efficiency Estimates.

– Problems with functional constraints.

Model of Nonsmooth Function.

Problem:

$$\min_{x \in Q} f(x), \tag{9.1}$$

where

- f is a Lipshitz continuous convex function;
- Q is a *simple* bounded closed convex set.

Note:

The optimal method for (9.1) is
the *subgradient method*.

However:

- The scheme of the subgradient method is very strict.
- It cannot converge faster than in theory.

What happens if our function is not so bad?

Model:

Let $X = \{x_k\}_{k=0}^{\infty}$ be a sequence in Q.

Denote

$$\hat{f}_k(X; x) = \max_{0 \le i \le k} [f(x_i) + \langle g(x_i), x - x_i \rangle],$$

where $g(x_i)$ are some subgradients of f at x_i.

The function $\hat{f}_k(X; x)$ is called the *model* of the function f.

Note:

- $f_k(X; x)$ is a piece-wise linear function.
- In view of inequality (7.10), we always have

$$f(x) \ge \hat{f}_k(X; x), \quad \forall x \in R^n.$$

- $f(x_i) = \hat{f}_k(X; x_i)$ and $g(x_i) \in \partial \hat{f}_k(X; x_i)$ for all i, $0 \le i \le k$.
- The models increase:

$$\hat{f}_{k+1}(X; x) \ge \hat{f}_k(X; x), \quad \forall x \in R^n.$$

- The model $\hat{f}_k(X; x)$ represents our *complete* information about function f, accumulated after k calls of oracle.

How we can use this knowledge?

Kelly method

The most natural scheme is as follows:

0). Choose $x_0 \in Q$.

1). $x_{k+1} \in \text{Arg} \min_{x \in Q} \hat{f}_k(X; x), \quad k \geq 0.$

This scheme is called the *Kelly method*.

Advantages:

- Very natural motivation.
- The auxiliary problem can be solved by LP-methods in finite time.

Drawbacks:

- Unstable rule for x_{k+1} (the solution of auxiliary problem may be not unique).
- Unstable practical behavior (sometimes is good, but sometimes does not work).
- Very bad *lower* complexity bound.

Example:

Consider the problem (9.1) with

$$f(y,x) \;=\; \max\{\mid y \mid, \parallel x \parallel^2\}, \quad y \in R^1,\ x \in R^n,$$

$$Q \;=\; \{z = (y,x) :\ y^2 + \parallel x \parallel^2 \le 1\}.$$

Thus,

$$z^* = (y^*, x^*) = (0,0), \quad f^* = 0.$$

Denote $Z_k^* = \mathrm{Arg}\min\limits_{z\in Q} \hat{f}_k(Z;z),\ \hat{f}_k^* = \hat{f}_k(Z_k^*)$.

Let us choose $z_0 = (1,0)$. Then

$$\hat{f}_0(Z;z) = y, \quad \Rightarrow \quad z_1 = (-1,0).$$

Further,

$$\hat{f}_1(Z;z) = \max\{y, -y\} = \mid y \mid \quad \Rightarrow \quad \hat{f}_1^* = 0.$$

Note that $\hat{f}_{k+1}^* \ge \hat{f}_k^*$, but

$$\hat{f}_k^* \le f(z^*) = 0.$$

Thus, for all $k \ge 1$ we have $\hat{f}_k^* = 0$ and $Z_k^* = (0, X_k^*)$, where

$$X_k^* = \{x \in B_2(0,1) \mid \ \parallel x_i \parallel^2 + \langle 2x_i, x - x_i \rangle \le 0, \quad i = 0,\ldots,k\}.$$

Efficiency of the cuts.

We need to estimate efficiency of the cuts for the set

$$X_k^* = \{x \in B_2(0,1) \mid \; \| x_i \|^2 + \langle 2x_i, x - x_i \rangle \leq 0, \\ i = 0, \ldots, k\}.$$

Note:

- Since x_{k+1} is an *arbitrary* point from X_k^*, at the first stage of the method we can choose

$$x_i : \quad \| x_i \| = 1.$$

  Then our the set X_k^* is as follows:

$$X_k^* = \{x \in B_2(0,1) \mid \langle x_i, x \rangle \leq \tfrac{1}{2}, i = 0, \ldots, k\}.$$

- We can do that up to the moment when the sphere

$$S_2(0,1) = \{x \in R^n \mid \| x \| = 1\}$$

  is not cut.
- Up to this moment we always have

$$f(z_i) \equiv f(0, x_i) = 1.$$

Geometric fact:

Let d be a direction in R^n, $\| d \| = 1$.
Consider the surface:

$$S(\alpha) = \{x \in R^n \mid \| x \| = 1, \ \langle d, x \rangle \geq \alpha\},$$

where $\alpha \in [0, 1]$. Denote

$$v(\alpha) = \text{vol}_{n-1}(S(\alpha)).$$

Then

$$\frac{v(\alpha)}{v(0)} \leq \left[1 - \alpha^2\right]^{\frac{n-1}{2}}.$$

Corollary:

In our case the cut is

$$\langle x_i, x \rangle \geq \tfrac{1}{2}.$$

Therefore, we cannot cut the sphere $S_2(0, 1)$ less then for

$$\left[\frac{2}{\sqrt{3}}\right]^{n-1}$$

cuts.
Note that during these iterations we still have

$$f(z_i) = 1.$$

Note:
At the first stage of the process the cuts are

$$\langle x_i, x \rangle \geq \tfrac{1}{2}.$$

Therefore, for k, $0 \leq k \leq N \equiv \left[\frac{2}{\sqrt{3}}\right]^{n-1}$, we have:

$$B_2(0, \tfrac{1}{2}) \subset X_k^*.$$

This means that after N iterations we can repeat our reasoning with $B_2(0, \frac{1}{2})$, etc.

Note that $f(0, x) = \frac{1}{4}$ for $x \in B_2(0, \frac{1}{2})$.

Thus, we have proved the following *lower* estimate for the Kelly method:

$$f(x_k) - f^* \geq \left(\frac{1}{4}\right)^{k\left[\frac{\sqrt{3}}{2}\right]^{n-1}}.$$

Complexity:

$$\frac{\ln \frac{1}{\epsilon}}{2 \ln 2} \left[\frac{2}{\sqrt{3}}\right]^{n-1}.$$

Compare:

- Ellipsoid method: $n^2 \ln \frac{1}{\epsilon}$.
- Optimal methods: $n \ln \frac{1}{\epsilon}$.
- Subgradient method: $\frac{1}{\epsilon^2}$.

Reasons of the troubles

- The test point

$$x_{k+1} \in \operatorname{Arg}\min_{x \in Q} \; \hat{f}_k(X; x)$$

is not uniquely defined.

- The set

$$\operatorname{Arg}\min_{x \in Q} \; \max_{0 \le i \le k} \; [f(x_i) + \langle g(x_i), x - x_i \rangle]$$

is not *stable*.

An *arbitrary* small variation of the data

$$\{f(x_i), g(x_i)\}_{i=0}^{k}$$

can change this set *significantly*.

Can we treat the models in a stable way?

Level Method

Denote:

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

$$\hat{f}_k^* = \min_{x \in Q} \hat{f}_k(X; x).$$

Clearly $\hat{f}_k^* \le f^* \le f_k^*$.

Let us choose some $\alpha \in (0, 1)$. Denote

$$l_k(\alpha) = (1 - \alpha)\hat{f}_k^* + \alpha f_k^*.$$

Consider the sublevel set:

$$\mathcal{L}_k(\alpha) = \{x \in Q \mid \; f_k(x) \le l_k(\alpha)\}.$$

Clearly, $\mathcal{L}_k(\alpha)$ is a closed convex set.

Note:

- There is no test points inside $\mathcal{L}_k(\alpha)$.
- This set is stable with respect to the data.

Level Method's Scheme

0. Choose a point $x_0 \in Q$ and accuracy $\epsilon > 0$.

   Choose the level coefficient $\alpha \in (0, 1)$.

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

   a). Compute $\hat{f}_k^*$ and f_k^*.

   b). Terminate if $f_k^* - \hat{f}_k^* \leq \epsilon$.

   c). Set $x_{k+1} = \pi_{\mathcal{L}_k(\alpha)}(x_k)$. $\square$

Note: The most expensive operations of this method are as follows.

- Computation of $\hat{f}_k^*$. If Q is a polytope, then that is an LP-problem:

$$\begin{aligned} &\min \quad t, \\ \text{s.t.} \;\; & f(x_i) + \langle g(x_i), x - x_i \rangle \leq t, \; i = 0 \ldots k, \\ & x \in Q. \end{aligned}$$

- Computation of $\pi_{\mathcal{L}_k(\alpha)}(x_k)$. If Q is a polytope, then that is an QP-problem:

$$\begin{aligned} &\min \quad \| x - x_k \|^2, \\ \text{s.t.} \;\; & f(x_i) + \langle g(x_i), x - x_i \rangle \leq l_k(\alpha), \; i = 0 \ldots k, \\ & x \in Q. \end{aligned}$$

Properties of the Level Method

1. The minimal values of the model are increasing:
$$\hat{f}_k^* \leq \hat{f}_{k+1}^* \leq f^*.$$
2. The record values of the functions are decreasing:
$$f_k^* \geq f_{k+1}^* \geq f^*.$$
3. Denote $\Delta_k = [\hat{f}_k^*, f_k^*]$, $\delta_k = f_k^* - \hat{f}_k^*$. Then
$$\Delta_{k+1} \subseteq \Delta_k, \quad \delta_{k+1} \leq \delta_k.$$

Lemma 9.1 *Assume that for some $p \geq k$ we have:*
$$\delta_p \geq (1-\alpha)\delta_k.$$
Then
$$l_i(\alpha) \geq \hat{f}_p^*$$
for all i, $k \leq i \leq p$.

Proof:

Note that for such i we have:
$$\delta_p \geq (1-\alpha)\delta_k \geq (1-\alpha)\delta_i.$$
Therefore
$$\begin{aligned} l_i(\alpha) &= f_i^* - (1-\alpha)\delta_i \\ &\geq f_p^* - (1-\alpha)\delta_i \\ &= \hat{f}_p^* + \delta_p - (1-\alpha)\delta_i \geq \hat{f}_p^*. \end{aligned}$$
□

Denote

$$M_f = \max\{\| g \| \mid g \in \partial f(x), \ x \in Q\}.$$

Lemma 9.2 *For the sequence $\{x_k\}$ generated by the Level Method we have:*

$$\| x_{k+1} - x_k \| \geq \frac{(1-\alpha)\delta_k}{M_f}.$$

Proof:
Indeed,

$$\begin{aligned}
f(x_k) - (1-\alpha)\delta_k &\geq f_k^* - (1-\alpha)\delta_k \\
&= l_k(\alpha) \\
&\geq \hat{f}_k(x_{k+1}) \\
&\geq f(x_k) + \langle g(x_k), x_{k+1} - x_k \rangle \\
&\geq f(x_k) - M_f \| x_{k+1} - x_k \| .
\end{aligned}$$

□

Lemma 9.3 *Let Q in the problem (9.1) be bounded:*

$$\operatorname{diam} Q \leq D.$$

If for $p \geq k$ we have $\delta_p \geq (1-\alpha)\delta_k$, then

$$p + 1 - k \leq \tfrac{M_f^2 D^2}{(1-\alpha)^2 \delta_p^2}.$$

Proof:
Denote $x_k^* \in \operatorname{Arg}\min\limits_{x \in Q} \hat{f}_k(X; x)$. In view of Lemma 9.1 we have

$$\hat{f}_i(X; x_p^*) \leq \hat{f}_p(X; x_p^*) = \hat{f}_p^* \leq l_i(\alpha)$$

for all i, $k \leq i \leq p$.
Therefore, in view of L.5.4 and L.9.2 we have:

$$\begin{aligned} \| x_{i+1} - x_p^* \|^2 &\leq \| x_i - x_p^* \|^2 - \| x_{i+1} - x_i \|^2 \\ &\leq \| x_i - x_p^* \|^2 - \tfrac{(1-\alpha)^2 \delta_i^2}{M_f^2} \\ &\leq \| x_i - x_p^* \|^2 - \tfrac{(1-\alpha)^2 \delta_p^2}{M_f^2}. \end{aligned}$$

Summarizing the inequalities in $i = k, \ldots, p$ we get:

$$(p + 1 - k) \tfrac{(1-\alpha)^2 \delta_p^2}{M_f^2} \leq \| x_k - x_p^* \|^2 \leq D^2.$$

□

Note: $p + 1 - k$ is the number of indices in the segment $[k, p]$.

Theorem 9.1 *Let* $\operatorname{diam} Q = D$. *Then the scheme of the Level Method terminates no more than after*

$$N = \frac{M_f^2 D^2}{\epsilon^2 \alpha (1-\alpha)^2 (2-\alpha)}$$

iterations. At this moment we have $f_k^* - f^* \leq \epsilon$.

Proof:
Assume that $\delta_k \geq \epsilon$, $0 \leq k \leq N$.

Let us divide the indices on the groups in the *decreasing* order:

$$\{N, \ldots, 0\} = I(0) \cup I(2) \cup \ldots \cup I(m),$$

such that

$$I(j) = [p(j), k(j)], \quad p(j) \geq k(j), \quad j = 0, \ldots, m,$$

$$p(0) = N, \quad p(j+1) = k(j) + 1, \quad k(m) = 0,$$

$$\delta_{k(j)} \leq \tfrac{1}{1-\alpha} \delta_{p(j)} < \delta_{k(j)+1} \equiv \delta_{p(j+1)}.$$

Clearly, for $j \geq 0$ we have:

$$\delta_{p(j+1)} \geq \tfrac{\delta_{p(j)}}{1-\alpha} \geq \tfrac{\delta_{p(0)}}{(1-\alpha)^{j+1}} \geq \tfrac{\epsilon}{(1-\alpha)^{j+1}}.$$

In view of L.9.2, $n(j) = p(j) + 1 - k(j)$ is bounded:

$$n(j) \leq \tfrac{M_f^2 D^2}{(1-\alpha)^2 \delta_{p(j)}^2} \leq \tfrac{M_f^2 D^2}{\epsilon^2 (1-\alpha)^2} (1-\alpha)^{2j}.$$

Therefore

$$N = \sum_{j=0}^{m} n(j) \leq \tfrac{M_f^2 D^2}{\epsilon^2 (1-\alpha)^2} \sum_{j=0}^{m} (1-\alpha)^{2j} \leq \tfrac{M_f^2 D^2}{\epsilon^2 \alpha (1-\alpha)^2 (2-\alpha)}.$$

$\square$

Remarks:

1. The optimal value of α can be found with

$$\alpha(1-\alpha)^2(2-\alpha) \quad \to \quad \max_{\alpha\in[0,1]} .$$

The solution is $\alpha^* = \frac{1}{2+\sqrt{2}}$.

Under this choice $N \leq \frac{4}{\epsilon^2} M_f^2 D^2$.

2. Comparing the complexity result with Theorem 8.1, we see that the Level Method is optimal *uniformly* in the dimension of the space.

3. The complexity of this method in *finite* dimension is not known.

4. The *gap* $\delta_k = f_k^* - \hat{f}_k^*$ provides us with the *exact* estimate of the current accuracy.

5. This method is extremely fast in practice. For the most problems we have accuracy $\epsilon = 10^{-4} - 10^{-5}$ after $3 - 4n$ iterations.

Constrained Minimization

Problem:

$$\begin{aligned} &\min \quad f(x), \\ \text{s.t. } & f_j(x) \le 0, \; j = 1, \dots, m, \\ & x \in Q, \end{aligned} \tag{9.2}$$

where

- Q is a bounded closed convex set,
- $f(x)$ and $f_j(x)$ are Lipshitz continuous on Q.

Equivalent formulation:

Denote $\bar{f}(x) = \max_{1 \le j \le m} f_j(x)$.

Then we obtain the equivalent problem

$$\begin{aligned} &\min \quad f(x), \\ \text{s.t. } & \bar{f}(x) \le 0, \\ & x \in Q, \end{aligned} \tag{9.3}$$

Note:

- $f(x)$, $\bar{f}(x)$ are convex and Lipshitz continuous.
- We can try to solve (9.3) using the models for both of them.

Main objects

Models:

Consider a sequence $X = \{x_k\}_{k=0}^{\infty}$. Denote

$$\hat{f}_k(X;x) \;=\; \max_{0\le j\le k}\,[f(x_j) + \langle g(x_j), x - x_j\rangle] \le f(x),$$

$$\check{f}_k(X;x) \;=\; \max_{0\le j\le k}\,[\bar{f}(x_j) + \langle \bar{g}(x_j), x - x_j\rangle] \le \bar{f}(x),$$

where $g(x_j) \in \partial f(x_j)$ and $\bar{g}(x_j) \in \partial \bar{f}(x_j)$.

Parametric function:

$$f(t;x) \;=\; \max\{f(x) - t, \bar{f}(x)\},$$

$$f^*(t) \;=\; \min_{x\in Q} f(t;x).$$

Recall that (see Lecture 6):

- $f^*(t)$ is non-increasing in t.
- Denote x^* a solution to (9.3) and $t^* = f(x^*)$. Then t^* is the smallest root of $f^*(t)$.

Model of the parametric function:

$$f_k(X;t,x) \;= \max\{\hat{f}_k(X;x) - t, \check{f}_k(X;x)\} \le f(t;x)$$

$$\hat{f}_k^*(X;t) \;= \min_{x\in Q} f_k(X;t,x) \le f^*(t).$$

- $\hat{f}_k^*(X;t)$ is non-increasing in t.
- Its smallest root $t_k^*(X) \le t^*$.

Lemma 9.4

$$t_k^*(X) = \min\{\hat{f}_k(X;x) \mid \check{f}_k(X;x) \le 0, \ x \in Q\}.$$

Proof:
Denote $\hat{x}_k^*$ the solution of the above minimization problem.

And let $\hat{t}_k^* = \hat{f}_k(X;\hat{x}_k^*)$.

Then

$$\hat{f}_k^*(X;\hat{t}_k^*) \le \max\{\hat{f}_k(X;\hat{x}_k^*) - \hat{t}_k^*, \check{f}_k(X;\hat{x}_k^*)\} \le 0.$$

Thus, $\hat{t}_k^* \ge t_k^*(X)$.

Assume that $\hat{t}_k^* > t_k^*(X)$. Then there exists y:

$$\hat{f}_k(X;y) - t_k^*(X) \le 0,$$

$$\check{f}_k(X;y) \le 0.$$

However, in this case

$$\hat{t}_k^* = \hat{f}_k(X;\hat{x}_k^*) \le \hat{f}_k(X;y) \le t_k^*(X) < \hat{t}_k^*.$$

That is a contradiction. □

We will need also the function

$$f_k^*(X;t) = \min_{0 \le j \le k} f_k(X;t,x_j),$$

the *record value* of our parametric model.

Lemma 9.5 *Let $t_0 < t_1 \leq t^*$. Assume that*

$$\hat{f}_k^*(X; t_1) > 0.$$

Then $t_k^*(X) > t_1$ and

$$\hat{f}_k^*(X; t_0) \geq \hat{f}_k^*(X; t_1) + \tfrac{t_1 - t_0}{t_k^*(X) - t_1} \hat{f}_k^*(X; t_1). \qquad (9.4)$$

Proof. Denote $x_k^*(t) \in \operatorname{Arg\,min} f_k(X; t, x)$,

$$t_2 = t_k^*(X), \quad \alpha = \tfrac{t_1 - t_0}{t_2 - t_0} \in [0, 1].$$

Then $t_1 = (1 - \alpha)t_0 + \alpha t_2$ and (9.4) is equivalent to

$$\hat{f}_k^*(X; t_1) \leq (1 - \alpha)\hat{f}_k^*(X; t_0) + \alpha \hat{f}_k^*(X; t_2) \qquad (9.5)$$

(note that $\hat{f}_k^*(X; t_2) = 0$).

Let $x_\alpha = (1 - \alpha)x_k^*(t_0) + \alpha x_k^*(t_2)$. We have:

$$\begin{aligned}
\hat{f}_k^*(X; \ & t_1) \leq \max\{\hat{f}_k(X; x_\alpha) - t_1; \check{f}_k(X; x_\alpha)\} \\
\leq \quad \max\{ \ & (1 - \alpha)(\hat{f}_k(X; x_k^*(t_0)) - t_0) \\
& + \alpha(\hat{f}_k(X; x_k^*(t_2)) - t_2); \\
(1 - \alpha) \ & \check{f}_k(X; x_k^*(t_0)) + \alpha \check{f}_k(X; x_k^*(t_2))\} \\
\leq \ (1 - \alpha) \ & \max\{\hat{f}_k(X; x_k^*(t_0)) - t_0; \check{f}_k(X; x_k^*(t_0))\} \\
+\alpha \ & \max\{\hat{f}_k(X; x_k^*(t_2)) - t_2; \check{f}_k(X; x_k^*(t_2))\} \\
= \ (1 - \alpha) \ & \hat{f}_k^*(X; t_0) + \alpha \hat{f}_k^*(X; t_2),
\end{aligned}$$

and we get (9.5). □

We need also the following statement (compare with Lemma 6.4).

Lemma 9.6 *For any* $\Delta \geq 0$ *we have:*

$$f^*(t) - \Delta \ \leq f^*(t+\Delta),$$

$$\hat{f}_k^*(X;t) - \Delta \ \leq \hat{f}_k^*(X;t+\Delta)$$

Proof. Indeed, for $f^*(t)$ we have:

$$\begin{aligned} f^*(t+\Delta) \ &= \min_{x \in Q} \left[\max\{ f(x) - t; \bar{f}(x) + \Delta \} - \Delta \right] \\ &\geq \min_{x \in Q} \left[\max\{ f(x) - t; \bar{f}(x) \} - \Delta \right] \\ &= f^*(t) - \Delta. \end{aligned}$$

The proof of the second inequality is the same. □

Constrained Level Method

Now we are ready to present the scheme (compare with Lecture 6).

0. Choose $x_0 \in Q$ and $t_0 < t^*$.

   Choose $\kappa \in (0, \frac{1}{2})$ and accuracy $\epsilon > 0$.

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

   a). Continue the generation of $X = \{x_j\}_{j=0}^{\infty}$ by the Level Method as applied to $f(t_k; x)$. If

   $$\hat{f}_j^*(X; t_k) \geq (1 - \kappa) f_j^*(X; t_k)$$

   then stop and set $j(k) = j$.

   Global Stop: Terminate if $f_j^*(X; t_k) \leq \epsilon$.

   b). Set $t_{k+1} = t_{j(k)}^*(X)$. □

Note: We are interested in the analytical complexity of this method. Therefore:

1. The complexity of computation of $t_j^*(X)$ and $\hat{f}_j^*(X; t)$ is not important for us now.

2. We need to estimate the rate of convergence of the *master process*.

3. We need to estimate the complexity of Step 1a).

Master Process

Lemma 9.7

$$f^*_{j(k)}(X;t_k) \le \tfrac{t_0-t^*}{1-\kappa}\left[\tfrac{1}{2(1-\kappa)}\right]^k.$$

Proof:
Denote

$$\sigma_k = \tfrac{f^*_{j(k)}(X;t_k)}{\sqrt{t_{k+1}-t_k}}, \quad \beta = \tfrac{1}{2(1-\kappa)} \quad (<1).$$

Since $t_{k+1} = t^*_{j(k)}(X)$, in view of Lemma 9.5 for $k \ge 1$ we have:

$$\begin{aligned}\sigma_{k-1} &= \tfrac{f^*_{j(k-1)}(X;t_{k-1})}{\sqrt{t_k-t_{k-1}}} \ge \tfrac{\hat{f}^*_{j(k)}(X;t_{k-1})}{\sqrt{t_k-t_{k-1}}} \\ &\ge \tfrac{2\hat{f}^*_{j(k)}(X;t_k)}{\sqrt{t_{k+1}-t_k}} \ge \tfrac{2(1-\kappa)f^*_{j(k)}(X;t_k)}{\sqrt{t_{k+1}-t_k}} = \tfrac{\sigma_k}{\beta}.\end{aligned}$$

Thus, $\sigma_k \le \beta\sigma_{k-1}$ and we obtain

$$\begin{aligned}&f^*_{j(k)}(X;t_k) = \sigma_k\sqrt{t_{k+1}-t_k} \\ &\le \beta^k\sigma_0\sqrt{t_{k+1}-t_k} = \beta^k f^*_{j(0)}(X;t_0)\sqrt{\tfrac{t_{k+1}-t_k}{t_1-t_0}}.\end{aligned}$$

Further, in view of L.9.6, $t_1 - t_0 \ge \hat{f}^*_{j(0)}(X;t_0)$.

Therefore

$$\begin{aligned}&f^*_{j(k)}(X;t_k) \le \beta^k f^*_{j(0)}(X;t_0)\sqrt{\tfrac{t_{k+1}-t_k}{\hat{f}^*_{j(0)}(X;t_0)}} \\ &\le \tfrac{\beta^k}{1-\kappa}\sqrt{\hat{f}^*_{j(0)}(X;t_0)(t_{k+1}-t_k)} \le \tfrac{\beta^k}{1-\kappa}\sqrt{f^*(t_0)(t_0-t^*)}.\end{aligned}$$

It remains to note that $f^*(t_0) \le t_0 - t^*$ (L.9.6). $\square$

Complexity Analysis

Let $f_j^*(X; t_k) \leq \epsilon$. Then there exist j^* such that

$$f(t_k; x_{j^*}) = f_j^*(X; t_k) \leq \epsilon.$$

Therefore we have:

$$f(t_k; x_{j^*}) = \max\{f(x_{j^*}) - t_k; \bar{f}(x_{j^*})\} \leq \epsilon.$$

Since $t_k \leq t^*$, we conclude that

$$\begin{aligned} f(x_{j^*}) &\leq t^* + \epsilon, \\ \bar{f}(x_{j^*}) &\leq \epsilon. \end{aligned} \tag{9.6}$$

In view of Lemma 9.7, we can get (9.6) at most in

$$N(\epsilon) = \frac{1}{\ln[2(1-\kappa)]} \ln \frac{t_0 - t^*}{(1-\kappa)\epsilon}$$

full iterations of the master process.

(The last iteration of the process is terminated by the Global Stop rule).

Note:

κ is an absolute constant (for example, $\kappa = \frac{1}{4}$).

Internal Process

Denote

$$M_f = \max\{\| g \| \mid g \in \partial f(x) \cup \partial \bar{f}(x),\ x \in Q\}.$$

1. *Full step.*

This step is terminated by the rule

$$\hat{f}^*_{j(k)}(X; t_k) \geq (1-\kappa) f^*_{j(k)}(X; t_k)$$

The corresponding inequality for the *gap* is:

$$f^*_{j(k)}(X; t_k) - \hat{f}^*_{j(k)}(X; t_k) \leq \kappa f^*_{j(k)}(X; t_k).$$

In view of Theorem 9.1, this happens at most after

$$\frac{M_f^2 D^2}{\kappa^2 (f^*_{j(k)}(X; t_k))^2 \alpha (1-\alpha)^2 (2-\alpha)}$$

iterations of the internal process.

Since at the full step $f^*_{j(k)}(X; t_k)) \geq \epsilon$, we conclude that

$$j(k) - j(k-1) \leq \frac{M_f^2 D^2}{\kappa^2 \epsilon^2 \alpha (1-\alpha)^2 (2-\alpha)}$$

for any full iteration of the master process.

2. *Last step.*

This step was terminated by Global Stop rule:

$$f_j^*(X; t_k) \leq \epsilon.$$

Since the normal stopping criterion did not work, we conclude that

$$f_{j-1}^*(X; t_k) - \hat{f}_{j-1}^*(X; t_k) \geq \kappa f_{j-1}^*(X; t_k) \geq \kappa\epsilon.$$

Therefore, in view of Theorem 9.1, the number of iterations at the last step does not exceed

$$\frac{M_f^2 D^2}{\kappa^2 \epsilon^2 \alpha (1-\alpha)^2 (2-\alpha)}$$

Total estimate:

$$(N(\epsilon)+1)\frac{M_f^2 D^2}{\kappa^2\epsilon^2\alpha(1-\alpha)^2(2-\alpha)}$$

$$= \frac{M_f^2 D^2}{\kappa^2\epsilon^2\alpha(1-\alpha)^2(2-\alpha)}\left[1 + \frac{1}{\ln[2(1-\kappa)]}\ln\frac{t_0 - t^*}{(1-\kappa)\epsilon}\right]$$

$$= \frac{M_f^2 D^2 \ln\frac{2(t_0-t^*)}{\epsilon}}{\epsilon^2\alpha(1-\alpha)^2(2-\alpha)\kappa^2\ln[2(1-\kappa)]}.$$

Remarks:

1. The reasonable choice is

$$\alpha = \kappa = \tfrac{1}{2+\sqrt{2}}.$$

2. The principal term in the complexity estimate is in order of

$$\tfrac{1}{\epsilon^2} \ln \tfrac{2(t_0 - t^*)}{\epsilon}.$$

Thus, the Constrained Level Scheme is suboptimal (see Lecture 7).

3. At each iteration of the master process we need to solve the problem of finding $t^*_{j(k)}(X)$.

In view of Lemma 9.4, that is

$$\min\{\hat{f}_k(X;x) \mid \check{f}_k(X;x) \le 0,\ x \in Q\}.$$

This is equivalent to the following:

$$\begin{aligned} \min \quad & t, \\ \text{s.t.} \quad & f(x_j) + \langle g(x_j), x - x_j\rangle \le t,\ \ j = 0,\dots,k, \\ & \bar{f}(x_j) + \langle \bar{g}(x_j), x - x_j\rangle \le 0,\ \ j = 0,\dots,k, \\ & x \in Q. \end{aligned}$$

If Q is a polytope, that can be solved by finite LP-methods (simplex method).

If Q is more complicated, we need to use Interior-Point Methods.

4. The practical convergence of the scheme is very fast.

5. We can use a better model for the constraints. Since

$$\bar{f}(x) = \max_{1 \le i \le m} f_i(x),$$

it is possible to deal with

$$\check{f}_k(X; x) = \max_{0 \le j \le k} \max_{1 \le i \le m} [f_i(x_j) + \langle g_i(x_j), x - x_j \rangle],$$

where $g_i(x_j) \in \partial f_i(x_j)$.

This increases the convergence, but complicates the iteration.

6. In practical schemes we have to use some technique for dropping the old elements of the model.