ISEN 629: Engineering Optimization

Lecture 17

Sergiy Butenko

Industrial and Systems Engineering
Texas A&M University

Fall 2007

Methods with polyhedral localization sets

Several methods work with localization sets in form of a polytope:

$$E_k = \{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \dots, m_k\}.$$

- *Inscribed Ellipsoid Method*: y_k is chosen as the center of the maximal ellipsoid inscribed in E_k.
- *Analytic Center Method*: y_k is chosen as the minimum of the analytic barrier $F_k(x) = -\sum_{j=1}^{m_k} \ln(b_j - a_j^T x)$.
- *Volumetric Center Method*: y_k is chosen as the minimum of the volumetric barrier $V_k(x) = \ln(\det(F_k''(x)))$, where $F_k(x)$ is the analytic barrier for the set E_k.

All these methods are polynomial with complexity bound $n(\ln \frac{1}{\epsilon})^p$ with p =1 or 2. However, the complexity of each iteration is n^3 to n^4 arithmetic operations, with y_k being computed using interior point methods.

Model of nonsmooth function

- Subgradient and ellipsoid methods are known to perform according to their theoretical bounds even on simple problems, and we cannot expect a much better performance on specific problems.
- Some more flexible schemes are based on the notion of a model of nonsmooth function.

Definition

Let $X = \{x_k : k \geq 0\}$ be a sequence in Q. Then the function

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

where $g(x_i) \in \partial f(x_i)$, is called the model of the convex function f corresponding to the sequence X.

Model of nonsmooth function

$$\hat{f}_k(X; x) = \max_{0 \leq i \leq k} [f(x_i) + g(x_i)^T (x - x_i)].$$

- $\hat{f}_k(X; x)$ is a piece-wise linear function of x and
$$f(x) \geq \hat{f}_k(X; x) \text{ for all } x \in \mathbb{R}^n.$$
- For all test points x_i, $0 \leq i \leq k$, we have
$$f(x_i) = \hat{f}_k(X; x_i), \; g(x_i) \in \partial \hat{f}_k(X; x_i).$$
- The next model is always better than the previous one:
$$\hat{f}_{k+1}(X; x) \geq \hat{f}_k(X; x) \text{ for all } x \in \mathbb{R}^n.$$
- Model $\hat{f}_k(X; x)$ represents the *complete* information on function f accumulated after k calls of the oracle.

Kelley method

0. Choose $x_0 \in Q$.
1. k-th iteration ($k \geq 0$):
   Find $x_{k+1} \in \arg\min_{x \in Q} \hat{f}_k(X; x)$.

- The auxiliary problem $\min_{x \in Q} \hat{f}_k(X; x)$ can be easily solved using linear programming.
- However, the method is still impractical due to its instability:
  - The solution of the auxiliary problem may not be unique.
  - the set $\arg\min_{x \in Q} \hat{f}_k(X; x)$ can be unstable with respect to small variations of data $\{(f(x_i), g(x_i))\}$.
  - This feature can be used to construct problem instances where the Kelly method has a very poor lower complexity bound.

Kelley method

Example

For the problem $\min_{(y,x) \in Q} f(y, x)$ with

$$f(y, x) = \max\{|y|, \|x\|^2\}, \; y \in \mathbb{R}, x \in \mathbb{R}^n,$$

$$Q = \{z = (y, x) : y^2 + \|x\|^2 \leq 1\}$$

the following lower bound can be established for the Kelley method:

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

Thus, to get an ϵ-solution we need at least

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

calls of the oracle.

Level method

Denote by

$$\begin{aligned} \hat{f}_k^* &= \min_{x \in Q} \hat{f}_k(X; x) - \text{the minimal value of the model}, \\ f_k^* &= \min_{0 \leq i \leq k} f(x_i) - \text{the record value of the model}. \end{aligned}$$

We have

$$\hat{f}_k^* \leq f^* \leq f_k^*.$$

For some $\alpha \in (0, 1)$, denote by

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

and consider the level set

$$\mathcal{L}_k(\alpha) = \{x \in Q : \hat{f}_k(X; x) \leq l_k(\alpha)\},$$

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

Level method

- Note that inside $\mathcal{L}_k(\alpha)$ there are no test points of the current model.
  - Thus, by selecting x_{k+1} in $\mathcal{L}_k(\alpha)$, we obtain a new record value of the model and

$$\mathcal{L}_{k+1}(\alpha) \subset \mathcal{L}_k(\alpha).$$

- Unlike $\arg\min_{x \in Q} \hat{f}_k(X; x)$, the set $\mathcal{L}_k(\alpha)$ is stable with respect to small variations in the data.

Level method

0. Choose a point $x_0 \in Q$, accuracy $\epsilon > 0$ and the level coefficient $\alpha \in (0,1)$.
1. k-th iteration ($k \geq 0$):
   (a) Compute $\hat{f}_k^*$ and f_k^*.
   (b) If $f_k^* - \hat{f}_k^* \leq \epsilon$, then STOP.
   (c) Set $x_{k+1} = \pi_{\mathcal{L}_k(\alpha)}(x_k)$.

Level method

The most expensive steps in the level are computing $\hat{f}_k^*$ and $\pi_{\mathcal{L}_k(\alpha)}(x_k)$. If Q is a polytope, then they can be computed by solving the following LP and convex QP, respectively:

$$\begin{aligned} \hat{f}_k^* \;=\; \min\;\; & t \\ \text{s.t.}\;\; & f(x_i) + g(x_i)^T(x - x_i) \leq t,\; i = 0, \ldots, k, \\ & x \in Q. \end{aligned}$$

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

Both problems can be solved in polynomial time.

Level method

- The optimal values of the model increase, and the record values decrease:

$$\hat{f}_k^* \leq \hat{f}_{k+1}^* \leq f^* \leq f_{k+1}^* \leq f_k^*.$$

- Denote by $\Delta_k = [\hat{f}_k^*, f_k^*]$ and $\delta_k = f_k^* - \hat{f}_k^*$. Then

$$\Delta_{k+1} \subseteq \Delta_k \text{ and } \delta_{k+1} \leq \delta_k.$$

We call δ_k the gap of the model $\hat{f}_k(X; x)$.

Level method

Lemma (3.3.1)

Assume that for some $p \geq k$ we have $\delta_p \geq (1-\alpha)\delta_k$. Then for all $i, k \leq i \leq p$, we have

$$l_i(\alpha) \geq \hat{f}_p^*.$$

Proof.

Recall that $\delta_k = f_k^* - \hat{f}_k^*$ and $\delta_{k+1} \leq \delta_k$.
For any $i, k \leq i \leq p$, we have $\delta_p \geq (1-\alpha)\delta_k \geq (1-\alpha)\delta_i$.
Therefore,

$$l_i(\alpha) = f_i^* - (1-\alpha)\delta_i \geq f_p^* - (1-\alpha)\delta_i = \hat{f}_p^* + \delta_p - (1-\alpha)\delta_i \geq \hat{f}_p^*.$$

$\square$

Level method

Let

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

Lemma (3.3.2)

For the sequence $\{x_k\}$ generated by the level set method we have

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

Proof.

Since $x_{k+1} \in \mathcal{L}_k(\alpha) \Rightarrow \hat{f}_k(X; x_{k+1}) \leq l_k(\alpha)$, we have

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

$\square$

Level method

Lemma (3.3.3)

Let $\mathrm{diam}(Q) \leq D$. If for some $p \geq k$ we have $\delta_p \geq (1-\alpha)\delta_k$, then

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

Proof.

Let $x_k^* \in \arg\min_{x\in Q} \hat{f}_k(X; x)$. Then for all $i, k \leq i \leq p$, we have

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

Applying the inequality $\|x - \pi_S(x_0)\|^2 \leq \|x - x_0\|^2 - \|\pi_S(x_0) - x_0\|^2$ with $x_0 = x_i$, $x = x_p^*$ and $S = \mathcal{L}_k(\alpha)$, we obtain

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

Summing up these inequalities for $i = k, \ldots, p$, we get

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

$\square$

Level method

Theorem (3.3.1)

Let $\mathrm{diam}(Q) = D$. Then the level method needs at most $N = \left\lfloor \frac{M_f^2 D^2}{\epsilon^2\alpha(1-\alpha)^2(2-\alpha)} \right\rfloor + 1$ iterations to guarantee $f_k^* - f^* \leq \epsilon$.

Proof.

Assume that $f_k^* - f^* \geq \epsilon$. Then $\delta_k = f_k^* - \hat{f}_k^* \geq f_k^* - f^* \geq \epsilon$. Let $\{N, \ldots, 0\} = I(0) \cup I(1) \cup \cdots \cup I(m)$, where

$$\begin{aligned} I(j) &= \{p(j), \ldots, k(j)\},\ p(j) \geq k(j),\ j = 0, \ldots, m, \\ p(0) &= N,\ p(j+1) = k(j) + 1,\ k(m) = 0, \\ \delta_{k(j)} &\leq \tfrac{1}{1-\alpha}\delta_{p(j)} < \delta_{k(j)+1} \equiv \delta_{p(j+1)}. \end{aligned}$$

Then for $j \geq 0$ we have $\delta_{p(j+1)} \geq \frac{\delta_{p(j)}}{1-\alpha} \geq \frac{\delta_{p(0)}}{(1-\alpha)^{j+1}} \geq \frac{\epsilon}{(1-\alpha)^{j+1}}$. Thus,

$$n(j) \equiv p(j) + 1 - k(j) \leq \frac{M_f^2 D^2}{(1-\alpha)^2\delta_{p(j)}^2} \leq \frac{M_f^2 D^2}{\epsilon^2(1-\alpha)^2}(1-\alpha)^{2j}$$

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

$\square$

Level method

$$N = \left\lfloor \frac{M_f^2 D^2}{\epsilon^2\alpha(1-\alpha)^2(2-\alpha)} \right\rfloor + 1$$

We can obtain the optimal value of the level parameter α by solving the following problem:

$$\max_{\alpha\in[0,1]} \alpha(1-\alpha)^2(2-\alpha) = \max_{\alpha\in[0,1]} (1-\alpha)^2(1-(1-\alpha)^2) = \max_{\gamma\in[0,1]} \gamma(1-\gamma)$$

yielding $\gamma^* = (1-\alpha^*)^2 = 1/2$, i.e. $\alpha^* = 1 - \frac{1}{\sqrt{2}}$.

Hence, with $\alpha = 1 - \frac{1}{\sqrt{2}}$ we have

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

- Note that $f_k^* - \hat{f}_k^* \leq \epsilon \Rightarrow f_k^* - f^* \leq f_k^* - f^* \leq \epsilon$.
- For most real-life optimization problems, the accuracy $\epsilon \in [10^{-5}, 10^{-4}]$ is achieved after $N \in [3n, 4n]$ iterations.