Lecture 9.

Smooth minimization
of non-smooth functions

- Motivation
- Smooth approximations of non-differentiable functions
- Optimal scheme for smooth optimization
- Application examples
- Preliminary computational results
Subgradient methods
for non-smooth minimization

Advantages:

• Very simple iteration scheme.
• Low memory requirements.
• Optimal rate of convergence (uniformly in the dimension).
• Interpretation of the process.
• Extensions (saddle points, variational inequalities, stochastic optimization, etc.)

Objections:

• Low rate of convergence. (Confirmed by theory!)
• No acceleration.
• Very high sensitivity to the step-size strategy.
• Absence of dual information.
• No reliable stopping criterion.
Non-smooth unconstrained minimization

Problem:
\[
\min \{ f(x) : x \in \mathbb{R}^n \} \Rightarrow x^*, f^* = f(x^*),
\]
where \( f(x) \) is a non-smooth convex function.

Subgradients:
\[
g \in \partial f(x) \iff f(y) \geq f(x) + \langle g, y - x \rangle \quad \forall y \in \mathbb{R}^n.
\]

Main difficulties:
- \( g \in \partial f(x) \) is not a descent direction at \( x \).
- \( g \in \partial f(x^*) \) does not imply \( g = 0 \).

Example:
\[
f(x) = \max_{1 \leq j \leq m} \{ \langle a_j, x \rangle + b^{(j)} \}
\]
\[
\partial f(x) = \text{Conv} \{ a_j : \langle a_j, x \rangle + b^{(j)} = f(x) \}.
\]
Lower complexity bounds
(Nemirovsky, Yudin 1976)

Model of the problem:

\[ f(x) \text{ is given by a } \text{black-box} \text{ local oracle.} \]

- We can see only \( \mathcal{O}(x) = (f(x), g(x) \in \partial f(x)) \).
- Change of \( f \) at \( y \) does not imply a change at \( x \).

Note: “We” \( \equiv \) “method”.

Theorem:

In a black-box local setting it is impossible to converge faster than \( O\left(\frac{1}{\sqrt{k}}\right) \) uniformly in the dimension of space of variables.

(Proof: look at \( \max_{1 \leq i \leq n} x^{(i)} \).

Here:

- \( k \) is the number of calls of oracle.
- Any auxiliary computation is allowed.
- No memory limitations.

Note: Convergence is very slow.
**Question:**

We want to find an approximate solution of the problem

\[
\max_{1 \leq j \leq m} \left\{ \langle a_j, x \rangle + b^{(j)} \right\} \rightarrow \min_{x} : x \in \mathbb{R}^n, \quad (1)
\]

by a gradient scheme (\(n\) and \(m\) are quite big).

**What rate of convergence we can guarantee?**

**Answer (Complexity Theory):** In order to find an \(\epsilon\)-solution we need at least

\[
O \left( \frac{1}{\epsilon^2} \right)
\]

calls of oracle (\(\equiv\) iterations).

**Goal of the talk:** a gradient scheme for solving (1) with efficiency estimate

\[
O \left( \frac{1}{\epsilon} \right)
\]

iterations.

**Reason of speed up:** (1) is not a black box.
Complexity of smooth minimization

Problem:
\[ f(x) \rightarrow \min_{x} : x \in R^{n}, \quad (2) \]
where \( f \) is a convex function with Lipschitz-continuous gradient:
\[ \| \nabla f(x) - \nabla f(y) \|^* \leq L(f) \| x - y \| \]
for all \( x, y \in R^{n} \). (Notation: \( f \in C^{1,1}(R^{n}) \).)

Rate of convergence:

1. Gradient method: \( O \left( \frac{L(f)}{k} \right) \)
2. Optimal method: \( O \left( \frac{L(f)}{k^2} \right) \)
   [Nesterov 1983]

Complexity: \( O \left( \sqrt{\frac{L(f)}{\varepsilon}} \right) \).

Note: The difference with \( O \left( \frac{1}{\varepsilon^2} \right) \) is enormous.
Main questions:

1. Given by a non-smooth convex $f(x)$, can we find for it a smooth $\epsilon$-approximation $f_\epsilon(x)$ with

$$L(f_\epsilon) = O\left(\frac{1}{\epsilon}\right)$$

If yes, we need only $O\left(\frac{1}{\epsilon}\right)$ iterations.

2. Can we do that in a systematic way?

Conclusion:

We need a convenient model of our problem.

(Compare with the theory of self-concordant functions.)
Adjoint problem

Primal problem:

\[ f^* = \min_x \{ f(x) : x \in Q_1 \}, \]

where \( Q_1 \subset E_1 \) is convex closed and bounded.

Model of objective function:

\[ f(x) = \hat{f}(x) + \max_u \{ \langle Ax, u \rangle_2 - \hat{\phi}(u) : u \in Q_2 \}, \]

where

- \( \hat{f}(x) \) is differentiable and convex on \( Q_1 \).
- \( Q_2 \subset E_2 \) is a closed convex and bounded.
- \( \hat{\phi}(u) \) is continuous convex function on \( Q_2 \).
- linear operator \( A : E_1 \to E_2^* \).

Adjoint problem:

\[ \max_u \{ \phi(u) : u \in Q_2 \}, \]

\[ \phi(u) = -\hat{\phi}(u) + \min_x \{ \langle Ax, u \rangle_2 + \hat{f}(x) : x \in Q_1 \}. \]

Note: Adjoint problem is not unique!
Example: Consider

\[ f(x) = \max_{1 \leq j \leq m} |\langle a_j, x \rangle_1 - b^{(j)}|. \]

1. \( Q_2 \equiv E_2 = E_1^*, A = I, \)

\[ \hat{\phi}(u) \equiv f_*(u) = \max_x \{ \langle u, x \rangle_1 - f(x) : x \in E_1 \} \]

\[ = \min_{s \in \mathbb{R}^m} \left\{ \sum_{j=1}^m s^{(j)} b^{(j)} : u = \sum_{j=1}^m s^{(j)} a_j, \sum_{j=1}^m |s^{(j)}| \leq 1 \right\}. \]

2. \( E_2 = \mathbb{R}^m, \hat{\phi}(u) = \langle b, u \rangle_2, \)

\[ f(x) = \max_{1 \leq j \leq m} |\langle a_j, x \rangle_1 - b^{(j)}| \]

\[ = \max_{u \in \mathbb{R}^m} \left\{ \sum_{j=1}^m u^{(j)}[\langle a_j, x \rangle_1 - b^{(j)}] : \sum_{j=1}^m |u^{(j)}| \leq 1 \right\}. \]

3. \( E_2 = \mathbb{R}^{2m}, \hat{\phi}(u) \) is a linear, \( Q_2 \) is a simplex:

\[ f(x) = \max_{u \in \mathbb{R}^{2m}} \{ \sum_{j=1}^m (u_1^{(j)} - u_2^{(j)}) \cdot [\langle a_j, x \rangle_1 - b^{(j)}] : \sum_{j=1}^m (u_1^{(j)} + u_2^{(j)}) = 1, u \geq 0 \}. \]

Rule: Increase in dim \( E_2 \Rightarrow \) decrease in complexity of representation.
Smooth approximations

**Prox-function:** $d_2(u)$ is continuous and strongly convex on $Q_2$:

$$d_2(v) \geq d_2(u) + \langle \nabla d_2(u), v - u \rangle_2 + \frac{1}{2} \sigma_2 \| v - u \|_2^2.$$ 

Assume: $d_2(u_0) = 0$ and $d_2(u) \geq 0 \ \forall u \in Q_2$.

Fix $\mu > 0$, the smoothness parameter, and define

$$f_\mu(x) = \max_u \{ \langle Ax, u \rangle_2 - \hat{\phi}(u) - \mu d_2(u) : u \in Q_2 \}.$$ 

Denote by $u(x)$ the solution of this problem.

**Theorem:** $f_\mu(x)$ is convex and differentiable for $x \in E_1$.

Its gradient $\nabla f_\mu(x) = A^* u(x)$ is Lipschitz continuous with the constant

$$L_\mu = \frac{1}{\mu \sigma_2} \| A \|_{1,2}^2,$$

where $\| A \|_{1,2} = \max_{x,u} \{ \langle Ax, u \rangle_2 : \| x \|_1 = 1, \| u \|_2 = 1 \}$.

**Note:** 1. for any $\mu \geq 0$ and $x \in E_1$ we have

$$f_0(x) \geq f_\mu(x) \geq f_0(x) - \mu D_2,$$

where $D_2 = \max_u \{ d_2(u) : u \in Q_2 \}$.

2. All norms are very important.
Optimal method

Problem: \( \min_x \{ f(x) : x \in Q_1 \} \) with \( f \in C^{1,1}(Q_1) \).

Prox-function: strongly convex \( d_1(x), x \in Q_1 \):
\[
d_1(x_0) = 0, \quad d_1(x) \geq 0 \quad \forall x \in Q_1.
\]

Gradient mapping:
\[
T(x) = \arg \min_{y \in Q_1} \{ \langle \nabla f(x), y - x \rangle_1 + \frac{1}{2} L(f) \| y - x \|_2^2 \}.
\]

Method: For \( k \geq 0 \) do

1. Compute \( f(x_k), \nabla f(x_k) \).

2. Find \( y_k = T(x_k) \).

3. Find \( z_k = \arg \min_{x \in Q_1} \{ \frac{L(f)}{\sigma} d_1(x) + \sum_{i=0}^{k} \frac{i+1}{2} \langle \nabla f(x_i), x \rangle_1 \} \).

4. Set \( x_{k+1} = \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k \).

Convergence:
\[
f(y_k) - f(x^*) \leq \frac{4L(f)d_1(x^*)}{\sigma_1(k+1)^2},
\]
where \( x^* \) is the optimal solution.
Applications

Smooth problem:
\[ \bar{f}_\mu(x) = \hat{f}(x) + f_\mu(x) \rightarrow \min : x \in Q_1. \]

Lipschitz constant:
\[ L_\mu = L(\hat{f}) + \frac{1}{\mu \sigma_2} \| A \|_{1,2}^2. \]

Denote \( D_1 = \max \{ d_1(x) : x \in Q_1 \} \).

Theorem: Let us choose \( N \geq 1 \). Define
\[ \mu = \mu(N) = \frac{2 \| A \|_{1,2}}{N+1} \cdot \sqrt{\frac{D_1}{\sigma_1 \sigma_2 D_2}}. \]

After \( N \) iterations set \( \hat{x} = y_N \in Q_1 \) and
\[ \hat{u} = \sum_{i=0}^{N} \frac{2(i+1)}{(N+1)(N+2)} u(x_i) \in Q_2. \]

Then
\[ 0 \leq f(\hat{x}) - \phi(\hat{u}) \leq \frac{4 \| A \|_{1,2}}{N+1} \cdot \sqrt{\frac{D_1 D_2}{{\sigma_1 \sigma_2}}} + \frac{4L(\hat{f})D_1}{\sigma_1 \cdot (N+1)^2}. \]

Corollary. Let \( L(\hat{f}) = 0 \). To get \( \epsilon \)-solution we choose
\[ \mu = \frac{\epsilon}{2D_2}, \quad L = \frac{D_2}{2\sigma_2} \cdot \frac{\| A \|_{1,2}^2}{\epsilon}, \quad N \geq 4 \| A \|_{1,2} \sqrt{\frac{D_1 D_2}{{\sigma_1 \sigma_2}}} \cdot \frac{1}{\epsilon}. \]
Example 1: Equilibrium in matrix games

Denote $\Delta_n = \{ x \in R^n : x \geq 0, \sum_{i=1}^n x^{(i)} = 1 \}$. Consider the problem

$$\min_{x \in \Delta_n} \max_{u \in \Delta_m} \{ \langle Ax, u \rangle_2 + \langle c, x \rangle_1 + \langle b, u \rangle_2 \}.$$ 

**Minimization form:**

$$\min_{x \in \Delta_n} f(x), \quad f(x) = \langle c, x \rangle_1 + \max_{1 \leq j \leq m} [\langle a_j, x \rangle_1 + b^{(j)}],$$

$$\max_{u \in \Delta_m} \phi(u), \quad \phi(u) = \langle b, u \rangle_2 + \min_{1 \leq i \leq n} [\langle \hat{a}_i, u \rangle_2 + c^{(i)}],$$

where $a_j$ are the rows and $\hat{a}_i$ are the columns of $A$.

**1. Euclidean distance:** Let us take

$$\|x\|_1 = \left[ \sum_{i=1}^n (x^{(i)})^2 \right]^{1/2}, \quad \|u\|_2 = \left[ \sum_{j=1}^m (u^{(j)})^2 \right]^{1/2},$$

$$d_1(x) = \frac{1}{2} \| x - \frac{1}{n} e_n \|_1^2$$
and
$$d_2(u) = \frac{1}{2} \| u - \frac{1}{m} e_m \|_2^2.$$ Then

$$\|A\|_{1,2} = \lambda_{\text{max}}^{1/2}(A^T A)$$

and

$$f(\hat{x}) - \phi(\hat{u}) \leq \frac{4 \lambda_{\text{max}}^{1/2}(A^T A)}{N + 1}.$$
2. **Entropy distance.** Let us choose

\[ \|x\|_1 = \sum_{i=1}^{n} |x^{(i)}|, \quad d_1(x) = \ln n + \sum_{i=1}^{n} x^{(i)} \ln x^{(i)}, \]

\[ \|u\|_2 = \sum_{j=1}^{m} |u^{(j)}|, \quad d_2(u) = \ln m + \sum_{j=1}^{m} u^{(j)} \ln u^{(j)}. \]

Then

\[ \sigma_1 = \sigma_2 = 1, \quad D_1 = \ln n, \quad D_2 = \ln m. \]

Moreover, since

\[ \|A\|_{1,2} = \max_{x} \{ \max_{1 \leq j \leq m} |\langle a_j, x \rangle| : \|x\|_1 = 1 \}

\[ = \max_{i, j} |A^{(i,j)}|, \]

we have

\[ f(\hat{x}) - \phi(\hat{u}) \leq \frac{4\sqrt{\ln n \ln m}}{N+1} \cdot \max_{i, j} |A^{(i,j)}|. \]

**Note:** 1. Usually \( \max_{i,j} |A^{(i,j)}| \ll \lambda_{\max}^{1/2}(A^T A). \)

2. \( \bar{f}_\mu(x) \) is easily computable:

\[ \bar{f}_\mu(x) = \langle c, x \rangle_1 + \mu \ln \left( \frac{1}{m} \sum_{j=1}^{m} e^{\langle a_j, x \rangle + b^{(j)}} / \mu \right). \]
Example 2: Continuous location problem

**Problem:** $p$ cities with population $m_j$ are located at $c_j \in \mathbb{R}^n$, $j = 1, \ldots, p$.

Construct a service center at point $x^*$, which minimizes the total distance to the center. That is

$$\text{Find } f^* = \min_x \left\{ f(x) = \sum_{j=1}^{p} m_j \|x - c_j\|_1 : \|x\|_1 \leq \bar{r} \right\}.$$

**Primal space:**

$$\|x\|_1^2 = \sum_{i=1}^{n} (x^{(i)})^2, \quad d_1(x) = \frac{1}{2} \|x\|_1^2, \quad \sigma_1 = 1, \quad D_1 = \frac{1}{2} \bar{r}^2.$$

**Adjoint space:** $E_2 = (E_1^*)^p$, $\|u\|_2^2 = \sum_{j=1}^{p} m_j (\|u_j\|_1^*)^2$, $Q_2 = \{ u = (u_1, \ldots, u_p) \in E_2 : \|u_j\|_1^* \leq 1, \ j = 1, \ldots, p \}$, $d_2(u) = \frac{1}{2} \|u\|_2^2, \quad \sigma_2 = 1, \quad D_2 = \frac{1}{2} P$.

with $P \equiv \sum_{j=1}^{p} m_j$, the total size of population.

**Operator norm:** $\|A\|_{1,2} = P^{1/2}$.

**Rate of convergence:** $f(\hat{x}) - f^* \leq \frac{2P\bar{r}}{N+1}$.

$$f_\mu(x) = \sum_{j=1}^{p} m_j \psi_\mu(\|x - c_j\|_1), \quad \psi_\mu(\tau) = \begin{cases} \frac{\tau^2}{2\mu}, & \tau \leq \mu, \\ \tau - \frac{\mu^2}{2}, & \mu \leq \tau. \end{cases}$$
Ex.3: Variational inequalities (linear operator)

Consider $B(w) = Bw + c: E \to E^*$, which is monotone:
\[ \langle Bh, h \rangle \geq 0 \quad \forall h \in E. \]

Problem:
Find $w^* \in Q : \langle B(w^*), w - w^* \rangle \geq 0 \quad \forall w \in Q, \quad (3)$
where $Q$ is a bounded convex closed set.

Merit function:
\[ \psi(w) = \max_v \{ \langle B(v), w - v \rangle : v \in Q \}. \quad (4) \]

- $\psi(w)$ is convex on $E_1$.
- $\psi(w) \geq 0$ for all $w \in Q$.
- $\psi(w) = 0$ if and only if $w$ solves (3).
- $\langle B(v), v \rangle$ is a convex function. Thus, (4) is exactly in our form.

Primal smoothing:
\[ \psi_\mu(w) = \max_v \{ \langle B(v), w - v \rangle - \mu d_2(v) : v \in Q \}. \]

Dual smoothing:
\[ \phi_\mu(v) = \min_w \{ \langle B(v), w - v \rangle + \mu d_1(w) : w \in Q \}. \]
(Looks better.)
Example 4: Piece-wise linear functions

1. Maximum of absolute values. Consider

$$\min_x \left\{ f(x) = \max_{1 \leq j \leq m} |\langle a_j, x \rangle_1 - b^{(j)}| : x \in Q_1 \right\}.$$ 

For simplicity choose $$\|x\|_1^2 = \sum_{i=1}^{n} (x^{(i)})^2$$, $$d_1(x) = \frac{1}{2} \|x\|^2$$.

It is convenient to choose $$E_2 = \mathbb{R}^{2m}$$,

$$\|u\|_2 = \sum_{j=1}^{2m} |u^{(j)}|, \quad d_2(u) = \ln(2m) + \sum_{j=1}^{2m} u^{(j)} \ln u^{(j)}.$$ 

Denote by $$A$$ the matrix with the rows $$a_j$$. Then

$$f(x) = \max_{u} \{ \langle \hat{A}x, u \rangle_2 - \langle \hat{b}, u \rangle_2 : u \in \Delta_{2m} \},$$ 

where $$\hat{A} = \left( \begin{array}{c} A \\ -A \end{array} \right)$$ and $$\hat{b} = \left( \begin{array}{c} b \\ -b \end{array} \right)$$. Thus, $$\sigma_1 = \sigma_2 = 1$$,

$$D_2 = \ln(2m), \quad D_1 = \frac{1}{2} \bar{r}^2, \quad \bar{r} = \max \{ \|x\|_1 : x \in Q_1 \}.$$

**Operator norm:** $$\|\hat{A}\|_{1,2} = \max_{1 \leq j \leq m} \|a_j\|^*.$$ 

**Complexity:**

$$2\sqrt{2} \bar{r} \max_{1 \leq j \leq m} \|a_j\|^* \frac{\sqrt{\ln(2m)}}{\sqrt{2} \sqrt{\ln(2m)}} \cdot \frac{1}{\epsilon}.$$ 

**Approximation:** for $$\xi(\tau) = \frac{1}{2} [e^\tau + e^{-\tau}]$$ define

$$\bar{f}_{\mu}(x) = \mu \ln \left( \frac{1}{m} \sum_{j=1}^{m} \xi \left( \frac{1}{\mu} [\langle a_j, x \rangle + b^{(j)}] \right) \right)$$
2. Sum of absolute values. Consider

\[
\min_x \left\{ f(x) = \sum_{j=1}^{m} |\langle a_j, x \rangle_1 - b^{(j)}| : x \in Q_1 \right\}.
\] (5)

Let us choose

\[
E_2 = \mathbb{R}^m, \quad Q_2 = \{ u \in \mathbb{R}^m : |u^{(j)}| \leq 1, \ j = 1, \ldots, m \},
\]

\[
d_2(u) = \frac{1}{2} \| u \|_2^2 = \frac{1}{2} \sum_{j=1}^{m} \| a_j \|_1^* \cdot (u^{(j)})^2.
\]

Then

\[
f_\mu(x) = \sum_{j=1}^{m} \| a_j \|_1^* \cdot \psi_\mu \left( \frac{\langle a_j, x \rangle_1 - b^{(j)} |}{\| a_j \|_1^*} \right),
\]

\[
\| A \|_{1,2} = P^{1/2} \equiv \left[ \sum_{j=1}^{m} \| a_j \|_1^* \right]^{1/2} \cdot
\]

On the other hand, \( D_2 = \frac{1}{2} P \) and \( \sigma_2 = 1 \). Thus, we get the following complexity bound:

\[
\frac{1}{\epsilon} \cdot \sqrt{\frac{8D_1}{\sigma_1}} \cdot \sum_{j=1}^{m} \| a_j \|_1^*.
\]

Note: The bound and the scheme allow \( m \to \infty \).
Computational experiments

Test problem:

$$\min_{x \in \Delta_n} \max_{u \in \Delta_m} \langle Ax, u \rangle_2.$$ 

Entries of $A$ are uniformly distributed in $[-1, 1]$.

**Goal:** Test of computational stability.

**Computer:** Pentium 4, 2.6GHz.

**Iteration:** $2mn$ operations.

Results for $\epsilon = 0.01$.  

<table>
<thead>
<tr>
<th>$m \backslash n$</th>
<th>100</th>
<th>300</th>
<th>1000</th>
<th>3000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>808</td>
<td>1011</td>
<td>1112</td>
<td>1314</td>
<td>1415</td>
</tr>
<tr>
<td></td>
<td>0&quot;</td>
<td>0&quot;</td>
<td>3&quot;</td>
<td>12&quot;</td>
<td>44&quot;</td>
</tr>
<tr>
<td>300</td>
<td>910</td>
<td>1112</td>
<td>1415</td>
<td>1617</td>
<td>1819</td>
</tr>
<tr>
<td></td>
<td>0&quot;</td>
<td>2&quot;</td>
<td>10&quot;</td>
<td>35&quot;</td>
<td>135&quot;</td>
</tr>
<tr>
<td>1000</td>
<td>1112</td>
<td>1213</td>
<td>1415</td>
<td>1718</td>
<td>2020</td>
</tr>
<tr>
<td></td>
<td>2&quot;</td>
<td>8&quot;</td>
<td>32&quot;</td>
<td>115&quot;</td>
<td>451&quot;</td>
</tr>
</tbody>
</table>

Number of iterations: $40 - 50\%$ of predicted values.
### Table 2

<table>
<thead>
<tr>
<th>( m \setminus n )</th>
<th>100</th>
<th>300</th>
<th>1000</th>
<th>3000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>6970 ( 2'' )</td>
<td>8586 ( 8' )</td>
<td>9394 ( 29'' )</td>
<td>10000 ( 91'' )</td>
<td>10908 ( 349'' )</td>
</tr>
<tr>
<td>300</td>
<td>7778 ( 8'' )</td>
<td>10101 ( 27'' )</td>
<td>12424 ( 97'' )</td>
<td>14242 ( 313'' )</td>
<td>15656 ( 1162'' )</td>
</tr>
<tr>
<td>1000</td>
<td>8788 ( 30'' )</td>
<td>11010 ( 105'' )</td>
<td>13030 ( 339'' )</td>
<td>15757 ( 1083'' )</td>
<td>18282 ( 4085'' )</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>( m \setminus n )</th>
<th>100</th>
<th>300</th>
<th>1000</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>67068 ( 25'' )</td>
<td>72073 ( 80'' )</td>
<td>74075 ( 287'' )</td>
<td>80081 ( 945'' )</td>
</tr>
<tr>
<td>300</td>
<td>85086 ( 89'', 42% )</td>
<td>92093 ( 243'' )</td>
<td>101102 ( 914'' )</td>
<td>112113 ( 3302'' )</td>
</tr>
<tr>
<td>1000</td>
<td>97098 ( 331'' )</td>
<td>100101 ( 760'' )</td>
<td>116117 ( 2936'' )</td>
<td>139140 ( 11028'' )</td>
</tr>
</tbody>
</table>
Comparing the bounds

**Gradient:** $2 \cdot 4 \cdot \frac{m n}{\epsilon} \sqrt{\ln n \ln m}.$

**Short-step path-following method ($n \geq m$):**

$\left(7.2 \sqrt{n \ln \frac{1}{\epsilon}}\right) \cdot \frac{m(m+1)}{2} n.$

<table>
<thead>
<tr>
<th>$m$</th>
<th>$n$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
<td>$b$</td>
</tr>
<tr>
<td>100</td>
<td>300</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
<td>$b$</td>
</tr>
<tr>
<td>100</td>
<td>1000</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
<td>$b$</td>
</tr>
<tr>
<td>100</td>
<td>3000</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
<td>$b$</td>
</tr>
<tr>
<td>100</td>
<td>10000</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
<td>$b$</td>
</tr>
<tr>
<td>300</td>
<td>300</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
<td>$b$</td>
</tr>
<tr>
<td>300</td>
<td>1000</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
<td>$b$</td>
</tr>
<tr>
<td>300</td>
<td>3000</td>
<td>$g$</td>
<td>$g$</td>
<td>$=$</td>
<td>$b$</td>
</tr>
<tr>
<td>300</td>
<td>10000</td>
<td>$g$</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
</tr>
<tr>
<td>1000</td>
<td>1000</td>
<td>$g$</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
</tr>
<tr>
<td>1000</td>
<td>3000</td>
<td>$g$</td>
<td>$g$</td>
<td>$g$</td>
<td>$b$</td>
</tr>
<tr>
<td>1000</td>
<td>10000</td>
<td>$g$</td>
<td>$g$</td>
<td>$=$</td>
<td>$=$</td>
</tr>
</tbody>
</table>

$g$ - gradient method, $b$ - barrier method