ISEN 629: Engineering Optimization

Lecture 9

Sergiy Butenko

Industrial and Systems Engineering
Texas A&M University

Fall 2007

Optimal methods

Next, we need to make sure that the condition of Lemma (2.2.1) is satisfied. Assume that we already have x_k such that

$$\phi_k^* \geq f(x_k).$$

Then by the above lemma, we have

$$\begin{aligned}\phi_{k+1}^* \quad \geq \quad & (1-\alpha_k)f(x_k) + \alpha_k f(y_k) - \tfrac{\alpha_k^2}{2\gamma_{k+1}}\|f'(y_k)\|^2 \\ & + \tfrac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}} f'(y_k)^T(v_k - y_k).\end{aligned}$$

Since $f(x_k) \geq f(y_k) + f'(y_k)^T(x_k - y_k)$, we get

$$\begin{aligned}\phi_{k+1}^* \quad \geq \quad & f(y_k) - \tfrac{\alpha_k^2}{2\gamma_{k+1}}\|f'(y_k)\|^2 \\ & +(1-\alpha_k)f'(y_k)^T(\tfrac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k).\end{aligned}$$

We want to have $\phi_{k+1}^* \geq f(x_{k+1})$...

Optimal methods

From $f(x) - f(y) \leq f'(y)^T(x-y) + \frac{L}{2}\|x-y\|^2$, we can ensure the inequality

$$f(y_k) - \frac{1}{2L}\|f'(y_k)\|^2 \geq f(x_{k+1})$$

by, e.g., taking the gradient step $x_{k+1} = y_k - h_k f'(y_k)$ with $h_k = \frac{1}{L}$ and using the inequality $f(y) - f(x) - f'(x)^T(y-x) \leq \frac{L}{2}\|x-y\|^2$ with $y = x_{k+1}$ and $x = y_k$. Let us define α_k as follows:

$$L\alpha_k^2 = \gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu.$$

Then $\frac{\alpha_k^2}{2\gamma_{k+1}} = \frac{1}{2L}$ and we have

$$\phi_{k+1}^* \geq f(x_{k+1}) + (1-\alpha_k)f'(y_k)^T\left(\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k\right).$$

Since we are free to choose y_k, we can find it from the equation

$$\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k = 0,$$

yielding $\qquad y_k = \frac{\alpha_k\gamma_k v_k + \gamma_{k+1}x_k}{\gamma_k + \alpha_k\mu}$.

General scheme of optimal method

0. Choose $x_0 \in \mathbb{R}^n$ and $\gamma_0 > 0$. Set $v_0 = x_0$.
1. k-th iteration ($k \geq 0$):
   a) Compute $\alpha_k \in (0,1)$ from equation

   $$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu.$$

   Set $\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu$.

   b) Choose

   $$y_k = \frac{\alpha_k\gamma_k v_k + \gamma_{k+1}x_k}{\gamma_k + \alpha_k\mu}$$

   and compute $f(y_k)$ and $f'(y_k)$.

   c) Find x_{k+1} such that

   $$f(x_{k+1}) \leq f(y_k) - \frac{1}{2L}\|f'(y_k)\|^2$$

   d) Set $v_{k+1} = \frac{(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k f'(y_k))}{\gamma_{k+1}}$.

Optimal methods

Theorem (2.2.1)

The above scheme generates a sequence $\{x_k : k \geq 0\}$ *such that*

$$f(x_k) - f^* \leq \lambda_k [f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2],$$

where $\lambda_0 = 1$ *and* $\lambda_k = \prod_{i=0}^{k-1}(1-\alpha_i)$.

Proof: Choose $\phi_0(x) = f(x_0) + \frac{\gamma_0}{2}\|x - v_0\|^2$. Then $f(x_0) = \phi_0^*$ and we get $f(x_k) \leq \phi_k^*$ for any k. Therefore, we can apply Lemma (2.2.1) $\square$

Optimal methods

To estimate the rate of convergence of $\{f(x_k) : k \geq 0\}$, we can use the rate of convergence of $\{\lambda_k\}$.

Lemma (2.2.4)

If $\gamma_0 \geq \mu$ *then*

$$\lambda_k \leq \min\left\{\left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L} + k\sqrt{\gamma_0})^2}\right\}.$$

Optimal methods

Theorem (2.2.2)

Choose $\gamma_0 = L$. *Then our scheme generates a sequence* $\{x_k : k \geq 0\}$ *such that*

$$f(x_k) - f^* \leq L \min\left\{\left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4}{(k+2)^2}\right\}\|x_0 - x^*\|^2.$$

This means that our scheme is optimal for unconstrained minimization of functions from $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, $\mu \geq 0$.

Optimal methods

Proof: We will use a property of $f \in \mathcal{F}^{1,1}_L(\mathbb{R}^n)$:

$$f(y) - f(x) - f'(x)^T(y - x) \leq \frac{L}{2}\|x - y\|^2, \ \forall x, y \in \mathbb{R}^n.$$

Since for $y = x^*$: $f'(x^*) = 0$, we have

$$f(x_0) - f^* \leq \frac{L}{2}\|x_0 - x^*\|^2.$$

Hence, from the previous theorem and lemma,

$$\begin{aligned} f(x_k) - f^* &\leq \lambda_k[f(x_0) - f^* + \tfrac{\gamma_0}{2}\|x_0 - x^*\|^2] \\ &\leq \min\left\{\left(1 - \sqrt{\tfrac{\mu}{L}}\right)^k, \tfrac{4L}{(2\sqrt{L}+k\sqrt{\gamma_0})^2}\right\} L\|x_0 - x^*\|^2 \\ &= L\min\left\{\left(1 - \sqrt{\tfrac{\mu}{L}}\right)^k, \tfrac{4}{(k+2)^2}\right\}\|x_0 - x^*\|^2. \end{aligned}$$

Next, we will show that our scheme is indeed optimal for $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$.

Optimal methods

Recall the lower complexity bound for $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$:

$$f(x_k) - f^* \geq \frac{\mu}{2}\left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k} R^2 \geq \frac{\mu}{2}\exp\left(-\frac{4k}{\sqrt{Q_f}-1}\right) R^2,$$

where $Q_f = L/\mu,\ R = \|x_o - x^*\|$. Hence, if $f(x_k) - f^* \leq \epsilon$, we have

$$\frac{\mu}{2}\exp\left(-\frac{4k}{\sqrt{Q_f}-1}\right) R^2 \leq \epsilon \Rightarrow k \geq \frac{\sqrt{Q_f}-1}{4}\left[\ln\frac{1}{\epsilon} + \ln\frac{\mu}{2} + 2\ln R\right].$$

For our scheme:

$$f(x_k) - f^* \leq LR^2\left(1 - \sqrt{1/Q_f}\right)^k \leq LR^2 \exp\left(-\frac{k}{\sqrt{Q_f}}\right).$$

To get an upper bound on k for our method, we have:

$$LR^2 \exp\left(-\frac{k}{\sqrt{Q_f}}\right) \geq \epsilon \Rightarrow k \leq \sqrt{Q_f}\left[\ln\frac{1}{\epsilon} + \ln\frac{\mu}{2} + 2\ln R\right].$$

Optimal methods

Thus, we have:

$$k \geq \frac{\sqrt{Q_f}-1}{4}\left[\ln\frac{1}{\epsilon} + \ln\frac{\mu}{2} + 2\ln R\right]$$

and

$$k \leq \sqrt{Q_f}\left[\ln\frac{1}{\epsilon} + \ln\frac{\mu}{2} + 2\ln R\right],$$

and the main term in the upper bound, $\sqrt{Q_f}\ln\frac{1}{\epsilon}$, is proportional to the lower bound.

So, the proposed method is indeed optimal for $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ $\square$

Optimal methods

If we use the constant-step gradient iteration to find x_{k+1} in our general scheme, we obtain the following scheme.

0. Choose $x_0 \in \mathbb{R}^n$ and $\gamma_0 > 0$. Set $v_0 = x_0$.
1. k-th iteration ($k \geq 0$):
   a) Compute $\alpha_k \in (0,1)$ from equation
   $$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu.$$
   Set $\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu$.
   b) Choose
   $$y_k = \frac{\alpha_k\gamma_k v_k + \gamma_{k+1}x_k}{\gamma_k + \alpha_k\mu}$$
   and compute $f(y_k)$ and $f'(y_k)$.
   c) $\boxed{\text{Set } x_{k+1} = y_k - \frac{1}{L}f'(y_k).}$
   d) Set $v_{k+1} = \frac{(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k f'(y_k))}{\gamma_{k+1}}$.

Optimal methods

Note that

$$\begin{aligned}
y_k &= \tfrac{1}{\gamma_k+\alpha_k\mu}(\alpha_k\gamma_k v_k + \gamma_{k+1}x_k) \Rightarrow v_k = \tfrac{(\gamma_k+\alpha_k\mu)y_k - \gamma_{k+1}x_k}{\alpha_k\gamma_k} \\
x_{k+1} &= y_k - \tfrac{1}{L}f'(y_k), \\
v_{k+1} &= \tfrac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k f'(y_k))] \\
&= \tfrac{1}{\gamma_{k+1}}[\tfrac{(1-\alpha_k)}{\alpha_k}((\gamma_k+\alpha_k\mu)y_k - \gamma_{k+1}x_k) + \alpha_k\mu y_k - \alpha_k f'(y_k)] \\
&= \tfrac{1}{\gamma_{k+1}}\left[\tfrac{(1-\alpha_k)\gamma_k}{\alpha_k}y_k + \mu y_k\right] - \tfrac{1-\alpha_k}{\alpha_k}x_k - \tfrac{\alpha_k}{\gamma_{k+1}}f'(y_k) \\
&= x_k + \tfrac{1}{\alpha_k}(y_k - x_k) - \tfrac{1}{\alpha_k L}f'(y_k)\ [\text{since } \gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu = L\alpha_k^2] \\
&= x_k + \tfrac{1}{\alpha_k}(x_{k+1} - x_k). \text{ Hence,} \\
y_{k+1} &= \tfrac{1}{\gamma_{k+1}+\alpha_{k+1}\mu}(\alpha_{k+1}\gamma_{k+1}v_{k+1} + \gamma_{k+2}x_{k+1}) \\
&= x_{k+1} + \beta_k(x_{k+1} - x_k),
\end{aligned}$$

where $\beta_k = \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}\mu)}$.

Optimal methods

Since

$$\alpha_k^2 L = (1-\alpha_k)\gamma_k + \mu\alpha_k = \gamma_{k+1},$$

we have

$$\begin{aligned}\beta_k &= \tfrac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}\mu)} = \tfrac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}^2 L-(1-\alpha_{k+1})\gamma_{k+1})}\\ &= \tfrac{\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}L)} = \tfrac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}.\end{aligned}$$

Note that

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + \frac{\mu}{L}\alpha_{k+1},$$

so we can completely eliminate the sequence $\{\gamma_k : k \geq 0\}$.

Optimal methods

We obtain the following scheme.

0. Choose $x_0 \in \mathbb{R}^n$ and $\alpha_0 \in (0,1)$.
   Set $y_0 = x_0$ and $q = \mu/L$.
1. k-th iteration ($k \geq 0$):
   a) Compute $f(y_k)$ and $f'(y_k)$. Set
   $$x_{k+1} = y_k - \frac{1}{L}f'(y_k).$$
   b) Compute $\alpha_{k+1} \in (0,1)$ from equation
   $$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q\alpha_{k+1}.$$
   Set $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$,
   $$y_{k+1} = x_{k+1} + \beta_k(x_{k+1}-x_k).$$

Optimal methods

Theorem (2.2.3)

If in the above scheme

$$\alpha_0 \geq \sqrt{\frac{\mu}{L}},$$

then

$$\begin{aligned}f(x_k) - f^* \leq \min\ &\left\{\left(1-\sqrt{\tfrac{\mu}{L}}\right)^k, \tfrac{4L}{(2\sqrt{L}+k\sqrt{\gamma_0})^2}\right\}\\ &\times[f(x_0) - f^* + \tfrac{\gamma_0}{2}\|x_0 - x^*\|^2],\end{aligned}$$

where $\gamma_0 = \frac{\alpha_0(\alpha_0 L-\mu)}{1-\alpha_0}$.

Proof: The condition $\alpha_0 \geq \sqrt{\frac{\mu}{L}}$ is equivalent to $\gamma_0 \geq \mu$. Therefore the statement of this theorem follows from Theorem (2.2.1) and Lemma (2.2.4). □

Optimal methods

If we choose $\alpha_0 = \sqrt{\mu/L}$ then $\gamma_0 = \mu$, $\alpha_k = \sqrt{\mu/L}$, $\beta_k = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$ and we obtain the following scheme.

0. Choose $y_0 = x_0 \in \mathbb{R}^n$.
1. k-th iteration ($k \geq 0$):

$$\begin{aligned}x_{k+1} &= y_k - \tfrac{1}{L}f'(y_k),\\ y_{k+1} &= x_{k+1} + \tfrac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(x_{k+1}-x_k).\end{aligned}$$