ISEN 629: Engineering Optimization

Lecture 20

Sergiy Butenko

Industrial and Systems Engineering
Texas A&M University

Fall 2007

Important inequalities

- Let $f(x)$ be a self-concordant function with the constant $M_f = 2$ (we call such self-concordant functions standard).
  - If $M_f \neq 2$, we can consider $\phi(x) = \frac{M_f^2}{4} f(x)$ with $M_\phi = 2$.
- We assume that $\mathrm{dom}\, f$ contains no straight line (so that $f''(x)$ is nondegenerate for any $x \in \mathrm{dom}\, f$).
- Denote by

$$\begin{aligned} \|u\|_x &= \left[u^T f''(x) u\right]^{1/2}, \\ \|v\|_x^* &= \left[v^T [f''(x)]^{-1} v\right]^{1/2}, \\ \lambda_f(x) = \|f'(x)\|_x^* &= \left[f'(x)[f''(x)]^{-1} f'(x)\right]^{1/2}. \end{aligned}$$

  Then
$$|v^T u| \le \|v\|_x^* \cdot \|u\|_x.$$

- $\|u\|_x$ is called the local norm of direction u with respect to x;
- $\lambda_f(x)$ is called the local norm of the gradient $f'(x)$ (or the Newton decrement of f at x).

Important inequalities

For fixed $x \in \mathrm{dom}\, f$ and $u \in \mathbb{R}^n, u \neq 0$, consider the function

$$\phi(t) = \frac{1}{[u^T f''(x + tu) u]^{1/2}}$$

with the domain $\mathrm{dom}\, \phi = \{t \in \mathbb{R} : x + tu \in \mathrm{dom}\, f\}$.

Lemma (4.1.3)

For all feasible t we have $|\phi'(t)| \le 1$.

Proof.

$$\phi'(t) = -\frac{f'''(x + tu)[u, u, u]}{2\,[u^T f''(x + tu) u]^{3/2}} \le 1$$

since f is self-concordant. □

Important inequalities

Corollary

Domain of function $\phi(t)$ contains the interval

$$(-\phi(0), \phi(0)).$$

Proof.

- From the lemma, it follows that

$$\phi(t) = \phi(0) + \phi'(\tilde{t})(t - 0) \ge \phi(0) - |t|$$

  for some $\tilde{t}$ between 0 and t. Thus, if $t \in (-\phi(0), \phi(0))$ then $\phi(t) > 0$.
- It remains to show that $\mathrm{dom}\, \phi = \{t : \phi(t) > 0\}$.
  - $f(x + tu) \to \infty$ as $x + tu$ approaches the boundary of $\mathrm{dom}\, f$.
  - Therefore, the function $u^T f''(x + tu) u$ cannot be bounded.

□

Important inequalities

Consider the ellipsoid

$$W^0(x;r) = \{y \in \mathbb{R}^n : \|y-x\|_x < r\},$$
$$W(x;r) = \mathrm{cl}(W^0(x;r)) \equiv \{y \in \mathbb{R}^n : \|y-x\|_x \le r\}.$$

This ellipsoid is called the Dikin ellipsoid of f at x.

Theorem (4.1.5)

1. *For any $x \in \mathrm{dom}\, f$ we have $W^0(x;1) \subseteq \mathrm{dom}\, f$.*
2. *For all $x, y \in \mathrm{dom}\, f$:*

$$\|y-x\|_y \ge \frac{\|y-x\|_x}{1+\|y-x\|_x}. \qquad (1)$$

3. *If $\|y-x\|_x < 1$ then*

$$\|y-x\|_y \le \frac{\|y-x\|_x}{1-\|y-x\|_x}. \qquad (2)$$

Important inequalities

Proof.

1. We have

$$(-\phi(0), \phi(0)) \subseteq \mathrm{dom}\, \phi = \{t \in \mathbb{R} : x + tu \in \mathrm{dom}\, f\}.$$

Since $\phi(0) = 1/\|u\|_x$, this implies that

$$\{y = x+tu : -1/\|u\|_x < t < 1/\|u\|_x\} = \{y = x+tu : t^2\|u\|_x^2 < 1\} \subseteq \mathrm{dom}\, f.$$

2. Choosing $u = y - x$, we obtain

$$\phi(1) = \frac{1}{\|y-x\|_y}, \ \phi(0) = \frac{1}{\|y-x\|_x},$$

and since $|\phi'(t)| \le 1$, we have $\phi(1) \le \phi(0) + 1$.

3. If $\|y-x\|_x < 1$, then $\phi(0) > 1$ and since $|\phi'(t)| \le 1$, we have $\phi(1) \ge \phi(0) - 1$.

□

Important inequalities

Theorem (4.1.6)

Let $x \in \mathrm{dom}\, f$. Then for any $y \in W^0(x;1)$ we have

$$(1-\|y-x\|_x)^2 f''(x) \preceq f''(y) \preceq \frac{1}{(1-\|y-x\|_x)^2} f''(x).$$

Thus,

- At any point $x \in \mathrm{dom}\, f$ there is an ellipsoid

$$W^0(x;1) = \{x \in \mathbb{R}^n : (y-x)^T f''(x)(y-x) < 1\} \subseteq \mathrm{dom}\, f.$$

- Inside the ellipsoid $W(x;r)$ with $r \in [0,1)$ function f is almost quadratic (the Hessian is almost the same when r is small):

$$(1-r)^2 f''(x) \preceq f''(y) \preceq \frac{1}{(1-r)^2} f''(x) \ \ \forall y \in W(x;r).$$

Important inequalities

Theorem (4.1.7)

For any $x, y \in \mathrm{dom}\, f$ we have

$$(f'(y) - f'(x))^T (y-x) \ge \frac{\|y-x\|_x^2}{1+\|y-x\|_x}, \qquad (3)$$

$$f(y) \ge f(x) + f'(x)^T(y-x) + \omega(\|y-x\|_x), \qquad (4)$$

where $\omega(t) = t - \ln(1+t)$.

Theorem (4.1.8)

Let $x, y \in \mathrm{dom}\, f$ and $\|y-x\|_x < 1$. Then

$$(f'(y) - f'(x))^T (y-x) \le \frac{\|y-x\|_x^2}{1-\|y-x\|_x}, \qquad (5)$$

$$f(y) \le f(x) + f'(x)^T(y-x) + \omega_*(\|y-x\|_x), \qquad (6)$$

where $\omega_*(t) = -t - \ln(1-t)$.

Important inequalities

Theorem (4.1.9)

Each of the inequalities (1)-(6) *is an equivalent characterization (necessary and sufficient characteristic) of a standard self-concordant function.*

Proof.

The proof consists of the following steps:

1. Definition of self-concordant function $\Rightarrow$ (1) $\Rightarrow$ (3)$\Rightarrow$ (4)$\Rightarrow$ definition of self-concordant function;
2. Definition of self-concordant function $\Rightarrow$ (2) $\Rightarrow$ (5)$\Rightarrow$ (6)$\Rightarrow$ definition of self-concordant function.

□

Important inequalities

Theorem (4.1.10)

For any $x, y \in \mathrm{dom}\, f$ *we have*

$$f(y) \geq f(x) + f'(x)^T(y-x) + \omega(\|f'(y) - f'(x)\|_y^*). \quad (7)$$

If $\|f'(y) - f'(x)\|_y^* < 1$ *then*

$$f(y) \leq f(x) + f'(x)^T(y-x) + \omega_*(\|f'(y) - f'(x)\|_y^*). \quad (8)$$

Important inequalities

In the above theorems we used the functions

$$\omega(t) = t - \ln(1+t) \text{ and } \omega_*(\tau) = -\tau - \ln(1-\tau).$$

Since

$$\begin{aligned} \omega'(t) &= \tfrac{t}{1+t} \geq 0, \quad \omega''(t) = \tfrac{1}{(1+t)^2} > 0, \\ \omega_*'(\tau) &= \tfrac{\tau}{1-\tau} \geq 0, \quad \omega_*''(\tau) = \tfrac{1}{(1-\tau)^2} > 0, \end{aligned}$$

$\omega(t)$ and $\omega_*(\tau)$ are convex functions.

Lemma (4.1.4)

For any $t \geq 0$ *and* $\tau \in [0,1)$ *we have*

$$\omega'(\omega_*'(\tau)) = \tau, \ \omega_*'(\omega'(t)) = t,$$

$$\omega(t) = \max_{0 \leq \xi < 1} [\xi t - \omega_*(\xi)], \ \omega_*(\tau) = \max_{\xi \geq 0} [\xi \tau - \omega(\xi)],$$

$$\omega(t) + \omega_*(\tau) \geq \tau t,$$

$$\omega_*(\tau) = \tau \omega_*'(\tau) - \omega(\omega_*'(\tau)), \ \omega(t) = t\omega'(t) - \omega_*(\omega'(t)).$$

Minimizing the self-concordant function

Consider the problem

$$\min\{f(x) : x \in \mathrm{dom}\, f\}.$$

As before, we assume that f is a standard self-concordant function and $\mathrm{dom}\, f$ contains no straight line. Recall that $\lambda_f(x) = \|f'(x)\|_x^*$.

Theorem (4.1.11)

Let $\lambda_f(x) < 1$ *for some* $x \in \mathrm{dom}\, f$. *Then the solution* x_f^* *of our problem exists and is unique.*

Proof.

For any $y \in \mathrm{dom}\, f$ we have

$$\begin{aligned} f(y) &\geq f(x) + f'(x)^T(y-x) + \omega(\|y-x\|_x) \\ &\geq f(x) - \|f'(x)\|_x^* \cdot \|y-x\|_x + \omega(\|y-x\|_x) \\ &= f(x) - \lambda_f(x) \cdot \|y-x\|_x + \omega(\|y-x\|_x). \end{aligned}$$

Minimizing the self-concordant function

Therefore, for any $y \in \mathcal{L}_f(f(x)) = \{y \in \mathbb{R}^n : f(y) \le f(x)\}$ we have

$$\frac{\omega(\|y - x\|_x)}{\|y - x\|_x} \le \lambda_f(x) < 1.$$

Since the function $\frac{\omega(t)}{t} = 1 - \frac{1}{t}\ln(1+t)$ is strictly increasing in t, we have $\|y - x\|_x \le \bar{t}$, where $\bar{t}$ is a unique positive root of the equation

$$(1 - \lambda_f(x))t = \ln(1+t).$$

Thus, $\mathcal{L}_f(f(x))$ is bounded and x_f^* exists. The uniqueness follows from the inequality

$$f(y) \ge f(x_f^*) + \omega(\|y - x_f^*\|_{x_f^*})$$

for all $y \in \mathrm{dom}\, f$. □

Minimizing the self-concordant function

The following example shows that the assumption $\lambda_f(x) < 1$ cannot be relaxed.

Example

For a fixed $\epsilon > 0$, consider a univariate function

$$f_\epsilon(x) = \epsilon x - \ln(x), x > 0.$$

- This function is self-concordant.
- Since $f_\epsilon'(x) = \epsilon - \frac{1}{x}$, $f_\epsilon''(x) = \frac{1}{x^2}$, we have $\lambda_{f_\epsilon}(x) = |1 - \epsilon x|$.
- Thus, for $\epsilon = 0$ we have $\lambda_{f_0}(x) = 1$ for any $x > 0$.
- f_0 is not bounded from below.
- If $\epsilon > 0$ then $x_{f_\epsilon}^* = \frac{1}{\epsilon}$.

Damped Newton method

0. Choose $x_0 \in \mathrm{dom}\, f$.
1. $x_{k+1} = x_k - \frac{1}{1+\lambda_f(x_k)}[f''(x_k)]^{-1} f'(x_k), k \ge 0.$

Theorem (4.1.12)

For any $k \ge 0$: $f(x_{k+1}) \le f(x_k) - \omega(\lambda_f(x_k))$.

Proof.

Denote by $\lambda = \lambda_f(x_k)$. Then $\|x_{k+1} - x_k\|_x = \frac{\lambda}{1+\lambda} = \omega'(\lambda)$ and

$$\begin{aligned} f(x_{k+1}) &\le f(x_k) + f'(x_k)^T(x_{k+1} - x_k) + \omega_*(\|x_{k+1} - x_k\|_x) \\ &= f(x_k) - \frac{\lambda^2}{1+\lambda} + \omega_*(\omega'(\lambda)) \\ &= f(x_k) - \lambda\omega'(\lambda) + \omega_*(\omega'(\lambda)) = f(x_k) - \omega(\lambda). \end{aligned}$$

□