ISEN 629: Engineering Optimization

Lecture 6

Sergiy Butenko

Industrial and Systems Engineering
Texas A& M University

Fall 2007

Classes of differentiable functions: Lipschitz condition

Let $Q \subseteq \mathbb{R}^n$. Denote by $C_L^{k,p}(Q)$ the class of functions with the following properties:

- any $f \in C_L^{k,p}(Q)$ is k times continuously differentiable on Q.
- Its p-th derivative is Lipschitz continuous on Q with the constant L:

$$\|f^{(p)}(x) - f^{(p)}(y)\| \leq L\|x - y\|$$

for all $x, y \in Q$.

Properties:

- We always have $p \leq k$.
- If $q \geq k$ then $C_L^{q,p}(Q) \subseteq C_L^{k,p}(Q)$.
- If $f_1 \in C_{L_1}^{k,p}(Q)$, $f_2 \in C_{L_2}^{k,p}(Q)$ and $\alpha, \beta \in \mathbb{R}$, then for $L_3 = |\alpha|L_1 + |\beta|L_2$ we have $\alpha f_1 + \beta f_2 \in C_{L_3}^{k,p}(Q)$.

Lipschitz condition

Lemma (Lemma 1.2.2 in the text)

Function $f(x)$ belongs to $C_L^{2,1}(\mathbb{R}^n) \subset C_L^{1,1}(\mathbb{R}^n)$ if and only if

$$\|f''(x)\| \leq L, \quad \forall x \in \mathbb{R}^n.$$

Example:

1. Linear function $f(x) = a^T x + b \in C_0^{1,1}(\mathbb{R}^n)$ since
$$f'(x) = a, \; f''(x) = 0.$$
2. For the quadratic function $f(x) = \frac{1}{2}x^T Qx + c^T x$ we have
$$f'(x) = Qx + c, \; f''(x) = Q.$$
Thus, $f(x) \in C_L^{1,1}(\mathbb{R}^n)$ with $L = \|Q\|$.
3. For the function $f(x) = \sqrt{1 + x^2}, x \in \mathbb{R}$, we have
$$f'(x) = \frac{x}{\sqrt{1 + x^2}}, \; f''(x) = \frac{1}{(1 + x^2)^{3/2}} \leq 1.$$
So, $f(x) \in C_1^{1,1}(\mathbb{R})$.

Lipschitz condition

Lemma (1.2.3)

Let $f \in C_L^{1,1}(\mathbb{R}^n)$. Then for any $x, y \in \mathbb{R}^n$ we have

$$|f(y) - f(x) - f'(x)^T(y - x)| \leq \frac{L}{2}\|y - x\|^2.$$

Geometrically, this means that for a point $x_0 \in \mathbb{R}^n$ the graph of function f is located between the graphs of the following two quadratic functions:

$$\phi_1(x) = f(x_0) + f'(x_0)^T(x - x_0) - \frac{L}{2}\|x - x_0\|^2;$$

$$\phi_2(x) = f(x_0) + f'(x_0)^T(x - x_0) + \frac{L}{2}\|x - x_0\|^2.$$

Lipschitz condition

Consider a function $f \in C_M^{2,2}(\mathbb{R}^n)$. Then we have

$$\|f''(x) - f''(y)\| \leq M\|x - y\|, \ \forall x, y \in \mathbb{R}^n.$$

Lemma (1.2.4)

Let $f \in C_M^{2,2}(\mathbb{R}^n)$. Then for any $x, y \in \mathbb{R}^n$ we have

$$\|f'(y) - f'(x) - f''(x)(y - x)\| \leq \frac{M}{2}\|y - x\|^2,$$

$$\|f(y) - f(x) - f'(x)^T(y - x) - \frac{1}{2}(y - x)^T f''(x)(y - x)\| \leq \frac{M}{6}\|y - x\|^3.$$

Corollary

Let $f \in C_M^{2,2}(\mathbb{R}^n)$ and $\|y - x\| = r$. Then

$$f''(x) - MrI_n \preceq f''(y) \preceq f''(x) + MrI_n,$$

where I_n is the identity matrix in $\mathbb{R}^n$.

Smooth Convex Optimization

(Chapter 2 of the text)

Which properties make a problem tractable?

- Consider the unconstrained problem

  $$\min_{x \in \mathbb{R}^n} f(x),$$

  where $f(x)$ is a sufficiently smooth function.
- We want to introduce some reasonable assumptions on $f(x)$ to make our problem more tractable than it is in general.
- Next we will define a class $\mathcal{F}$ of differentiable functions that posses some desirable (with respect to minimization) properties.
- Since the methods we discussed converge to a stationary point, we want a stationary point to be a global minimizer for any $f \in \mathcal{F}$.
- We want class $\mathcal{F}$ to be sufficiently wide and include functions that are obtained from other functions in $\mathcal{F}$ by applying some basic arithmetic operations, such as multiplication by a nonnegative scalar and addition.

Basic assumptions

We formalize the desired properties of the hypothetical class $\mathcal{F}$ in the following assumptions:

1. For any $f \in \mathcal{F}$ the first-order optimality condition is sufficient for a point to be a global minimizer of our problem.
2. If $f_1, f_2 \in \mathcal{F}$ and $\alpha, \beta \geq 0$, then $\alpha f_1 + \beta f_2 \in \mathcal{F}$.
3. Any linear function $f(x) = a^T x + b$ belongs to $\mathcal{F}$.

These properties appear to be sufficient to define the class of smooth convex functions.

Smooth convex functions

For $f \in \mathcal{F}$, fix some $x_0 \in \mathbb{R}^n$ and consider the function

$$\phi(x) = f(x) - f'(x_0)^T x.$$

Then $\phi \in \mathcal{F}$ due to the assumptions 2 and 3.
We have

$$\phi'(x_0) = f'(x_0) - f'(x_0) = 0,$$

thus, according to assumption 1, x_0 is a global minimum of ϕ, so for any $x \in \mathbb{R}^n$ we have

$$\phi(x) \geq \phi(x_0) = f(x_0) - f'(x_0)^T x_0,$$

which is the same as

$$f(x) \geq f(x_0) + f'(x_0)^T (x - x_0).$$

Smooth convex functions

$$f(x) \geq f(x_0) + f'(x_0)^T (x - x_0).$$

Recall that this inequality is exactly the first-order characterization of a convex function, which can be viewed as an alternative definition of a smooth convex function:

Definition (2.1.1)

A continuously differentiable function $f(x)$ is called convex on $\mathbb{R}^n$ (denoted by $f \in \mathcal{F}^1(\mathbb{R}^n)$) if for any $x, y \in \mathbb{R}^n$ we have

$$f(y) \geq f(x) + f'(x)^T (y - x).$$

If $-f(x)$ is convex, we call $f(x)$ concave.

Smooth convex functions

- We can define classes of convex functions $\mathcal{F}_L^{k,p}(Q)$ similarly to how we defined classes $C_L^{k,p}(Q)$:

  Let $Q \subseteq \mathbb{R}^n$. Denote by $\mathcal{F}_L^{k,p}(Q)$ the class of <u>convex</u> functions with the following properties:
  - any $f \in \mathcal{F}_L^{k,p}(Q)$ is k times continuously differentiable on Q.
  - Its p-th derivative is Lipschitz continuous on Q with the constant L:

    $$\|f^{(p)}(x) - f^{(p)}(y)\| \leq L\|x - y\|$$

    for all $x, y \in Q$.

Smooth convex functions

Next we check that our three assumptions regarding $\mathcal{F}$ are satisfied by the class $\mathcal{F}^1(\mathbb{R}^n)$, and thus are the properties of this functional class.

Theorem (2.1.1)

If $f \in \mathcal{F}^1(\mathbb{R}^n)$ and $f'(x^*) = 0$ then x^* is a global minimum of $f(x)$ on $\mathbb{R}^n$.

Lemma (2.1.1)

If f_1 and f_2 belong to $\mathcal{F}^1(\mathbb{R}^n)$ and $\alpha, \beta \geq 0$ then function $f = \alpha f_1 + \beta f_2$ also belongs to $\mathcal{F}^1(\mathbb{R}^n)$.

The third assumption is also trivially satisfied. Thus, for differentiable functions our hypothetical class $\mathcal{F}$ is exactly the class of convex functions $\mathcal{F}^1(\mathbb{R}^n)$.

Properties of smooth convex functions

Recall the original definition of a convex function that we gave...

Theorem (2.1.2)

Continuously differentiable f belongs to the class $\mathcal{F}^1(\mathbb{R}^n)$ if and only if for any $x, y \in \mathbb{R}^n$ and $\alpha \in [0,1]$ we have

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y).$$

Here is another equivalent first-order characterization:

Theorem (2.1.3)

Continuously differentiable f belongs to the class $\mathcal{F}^1(\mathbb{R}^n)$ if and only if for any $x, y \in \mathbb{R}^n$ we have

$$(f'(x) - f'(y))^T (x - y) \geq 0.$$

Properties of smooth convex functions

Proof: The necessity follows from adding the inequalities

$$f(x) \geq f(y) + f'(y)^T (x-y), \quad f(y) \geq f(x) + f'(x)^T (y-x).$$

To show the sufficiency, assume that the inequality $(f'(x) - f'(y))^T (x-y) \geq 0$ holds for all $x, y \in \mathbb{R}^n$. Denote by $x_\tau = x + \tau(y-x)$. Then

$$\begin{aligned} f(y) &= f(x) + \int_0^1 f'(x + \tau(y-x))^T (y-x) d\tau \\ &= f(x) + f'(x)^T (y-x) + \int_0^1 (f'(x_\tau) - f'(x))^T (y-x) d\tau \\ &= f(x) + f'(x)^T (y-x) + \int_0^1 \tfrac{1}{\tau} (f'(x_\tau) - f'(x))^T (x_\tau - x) d\tau \\ &\geq f(x) + f'(x)^T (y-x). \end{aligned}$$

$\square$

Properties of smooth convex functions

Recall the second-order characterization of a convex function...

Theorem (2.1.4)

Two times continuously differentiable function f belongs to $\mathcal{F}^2(\mathbb{R}^n)$ if and only if for any $x \in \mathbb{R}^n$ we have

$$f''(x) \succeq 0.$$

Properties of smooth convex functions

Lemma (2.1.2)

If $f \in \mathcal{F}^1(\mathbb{R}^m), b \in \mathbb{R}^m$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ then

$$\phi(x) = f(Ax + b) \in \mathcal{F}^1(\mathbb{R}^n).$$

Proof: For $x, y \in \mathbb{R}^n$, denote by $\bar{x} = Ax + b, \bar{y} = Ay + b$. We have

$$\begin{aligned} \phi(y) = f(\bar{y}) &\geq f(\bar{x}) + f'(\bar{x})^T (\bar{y} - \bar{x}) \\ &= \phi(x) + f'(\bar{x})^T A(y-x) \\ &= \phi(x) + (A^T f'(\bar{x}))^T (y-x) \\ \{\phi'(x) = A^T f'(Ax+b)\} &= \phi(x) + \phi'(x)^T (y-x). \end{aligned}$$

$\square$

Properties of smooth convex functions

Theorem (2.1.5)

Each conditions below, holding for all $x, y \in \mathbb{R}^n$ and $\alpha \in [0,1]$, are equivalent to inclusion $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$:

$$0 \le f(y) - f(x) - f'(x)^T(y-x) \le \frac{L}{2}\|x-y\|^2, \qquad (1)$$

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

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

$$(f'(x) - f'(y))^T(x-y) \le L\|x-y\|^2, \qquad (4)$$

$$\alpha f(x) + (1-\alpha) f(y) \ge f(\alpha x + (1-\alpha) y) + \frac{\alpha(1-\alpha)}{2L}\|f'(x) - f'(y)\|^2, \qquad (5)$$

$$\alpha f(x) + (1-\alpha) f(y) \le f(\alpha x + (1-\alpha) y) + \alpha(1-\alpha)\frac{L}{2}\|x-y\|^2. \qquad (6)$$

Properties of smooth convex functions

The following theorem characterizes the class $\mathcal{F}_L^{2,1}(\mathbb{R}^n)$:

Theorem (2.1.6)

Two times continuously differentiable function f belongs to $\mathcal{F}_L^{2,1}(\mathbb{R}^n)$ if and only if for any $x \in \mathbb{R}^n$ we have

$$0 \preceq f''(x) \preceq L I_n.$$