MATH: Operations Research **2014-15 First Term**

Handout 9: Mathematical preliminaries

Instructor: Junfeng Yang *November 17, 2014*

Contents

9.1 Vector and matrix norms

Definition 9.1 (vector norm) *A mapping $\|\cdot\| : R^n \to R$ is called a vector norm if and only if*

1. $\|x\| \geq 0$ *for all* $x \in R^n$*, and* $\|x\| = 0$ *if and only if* $x = 0$*;*
2. $\|\alpha x\| = |\alpha|\|x\|$ *for all* $\alpha \in R$ *and* $x \in R^n$*.*
3. $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in R^n$*.*

Definition 9.2 (matrix norm) *A mapping $\|\cdot\| : R^{m\times n} \to R$ is a matrix norm if and only if*

1. $\|A\| \geq 0$ *for all* $A \in R^{m\times n}$*, and* $\|A\| = 0$ *if and only if* $A = 0$*;*
2. $\|\alpha A\| = |\alpha|\|A\|$ *for all* $\alpha \in R$ *and* $A \in R^{m\times n}$*.*
3. $\|A + B\| \leq \|A\| + \|B\|$ *for all* $A, B \in R^{m\times n}$*.*

Let $x = (x_1, x_2, \ldots, x_n)^T \in R^n$. Some well-known examples of vector norms:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}; \qquad (\ell_2\text{-norm})$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|; \qquad (\ell_1\text{-norm})$$

$$\|x\|_\infty = \max_{i=1,2,\ldots,n} |x_i|; \qquad (\ell_\infty\text{-norm})$$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, \quad \text{where } 1 \leq p \leq +\infty; \qquad (\ell_p\text{-norm})$$

$$\|x\|_A = \sqrt{x^T A x}, \quad \text{where } A \in S^n_{++}; \qquad (\text{ellipsoid-norm})$$

($A \in S^n_{++}$ means A is positive definite.)

Let $A \in R^{n\times n}$. Some well-known examples of matrix norms:

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p}, \quad \text{where } 1 \leq p \leq +\infty; \qquad (\text{induced } \ell_p\text{-norm})$$

$$\|A\|_1 = \max\left\{\sum_{i=1}^{n} |a_{ij}| : j = 1, 2, \ldots, n\right\}; \qquad (\text{maximum column norm})$$

$$\|A\|_\infty = \max\left\{\sum_{j=1}^{n} |a_{ij}| : i = 1, 2, \ldots, n\right\}; \qquad (\text{maximum row norm})$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}; \qquad (\text{spectral norm})$$

($A \in S^n$ implies $\|A\|_2 = \rho(A) := \max\{|\lambda_i(A)| :\ i = 1, \ldots, n\}$.)

Suppose A is nonsingular, then it holds that

$$\|A^{-1}\|_p = \frac{1}{\inf_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p}};$$

$$\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{\mathrm{tr}(A^T A)}; \qquad \text{(Frobenius norm)}$$

A vector norm $\|\cdot\|$ and a matrix norm $\|\cdot\|'$ are said to be consistent if, for every $A \in R^{n\times n}$ and $x \in R^n$,

$$\|Ax\| \leq \|A\|'\|x\|.$$

Obviously, the ℓ_p-norm has this property, i.e.,

$$\|Ax\|_p \leq \|A\|_p\|x\|_p.$$

Definition 9.3 (norm equivalence) *Let $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ be two arbitrary vector norms on R^n. The two norms are said to be equivalent if there exist $\mu_1, \mu_2 > 0$ such that*

$$\mu_1\|x\|_\alpha \leq \|x\|_\beta \leq \mu_2\|x\|_\alpha, \forall\, x \in R^n.$$

From functional analysis, any two vector norms defined on a finite dimensional space are equivalent. In particular, we have

$$\begin{aligned}
\|x\|_2 &\leq \|x\|_1 \leq \sqrt{n}\|x\|_2 \\
\|x\|_\infty &\leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty \\
\|x\|_\infty &\leq \|x\|_1 \leq n\|x\|_\infty \\
\|x\|_\infty &\leq \|x\|_2 \leq \|x\|_1 \\
\sqrt{\lambda_{\min}(A)}\|x\|_2 &\leq \|x\|_A \leq \sqrt{\lambda_{\max}(A)}\|x\|_2.
\end{aligned}$$

Let $A \in R^{m\times n}$. For matrix norms, the following inequalities hold:

$$\begin{aligned}
\|A\|_2 &\leq \|A\|_F \leq \sqrt{n}\|A\|_2 \\
\frac{1}{\sqrt{n}}\|A\|_\infty &\leq \|A\|_2 \leq \sqrt{m}\|A\|_\infty \\
\frac{1}{\sqrt{m}}\|A\|_1 &\leq \|A\|_2 \leq \sqrt{n}\|A\|_1 \\
\max_{i,j}|a_{ij}| &\leq \|A\|_2 \leq \sqrt{mn}\max_{i,j}|a_{ij}|.
\end{aligned}$$

9.2 Some important inequalities

1. Cauchy-Schwarz inequality: Let $x, y \in R^n$. It holds that

$$|x^T y| \leq \|x\|_2\|y\|_2.$$

   "$=$" holds if and only if x and y are linearly dependent.

2. Let A be an $n \times n$ symmetric and positive definite matrix, $x, y \in R^n$, then
$$|x^T A y| \le \|x\|_A \|y\|_A.$$
"$=$" holds if and only if x and y are linearly dependent.

3. Let A be an $n \times n$ symmetric and positive definite matrix, $x, y \in R^n$, then
$$|x^T y| \le \|x\|_A \|y\|_{A^{-1}}.$$
"$=$" holds if and only if x and $A^{-1}y$ are linearly dependent.

4. Young inequality: Assume that $p, q \in R$, $p, q > 1$ satisfy $1/p + 1/q = 1$. For $a, b \in R$, it holds that
$$ab \le \frac{a^p}{p} + \frac{b^q}{q},$$
and "$=$" holds if and only if $a^p = b^q$.

5. Holder inequality: Let $x, y \in R^n$. It holds that
$$|x^T y| \le \|x\|_p \|y\|_q,$$
where $p, q \in R$, $p, q > 1$ satisfy $1/p + 1/q = 1$.

6. Minkowski inequality: Let $x, y \in R^n$. For $p \ge 1$, it holds that
$$\|x + y\|_p \le \|x\|_p + \|y\|_p.$$

9.3 Multivariate calculus

Continuity, differentiability, derivative and gradient. Let $f : R^n \to R$ be a mapping.

- f is said to be continuous on R^n if ...
- f is said to be continuously differentiable at $x \in R^n$ if $\frac{\partial f}{\partial x_i}(x)$ exists and is continuous, $i = 1, 2, \ldots, n$.
- The gradient of f at x is defined as
$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$
- The derivative of f at x is defined as
$$f'(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right) = \nabla f(x)^T.$$
- If f is continuously differentiable at each $x \in R^n$, then f is said to be continuously differentiable on R^n and is denoted by $f \in C^1(R^n)$.

Directional derivative.

- Let $x, d \in R^n$. The directional derivative of f at x in the direction d is defined as

$$f'(x; d) := \lim_{\theta \to 0+} \frac{f(x + \theta d) - f(x)}{\theta} = \nabla f(x)^T d.$$

  It represents how fast f changes at x in the direction d.

- For any $x, x + d \in R^n$, if $f \in C^1(R^n)$, then

$$f(x + d) = f(x) + \int_0^1 \nabla f(x + td)^T d\, dt.$$

  ($h(t) = f(x + td)$. Then, $h(1) = h(0) + \int_0^1 h'(t)dt$.)

$$f(x + d) = f(x) + \nabla f(x + \theta d)^T d, \text{for some } \theta \in (0, 1).$$

  ($h(1) = h(0) + h'(\theta)$ for some $\theta \in (0, 1)$.)

$$f(x + d) = f(x) + \nabla f(x)^T d + o(\|d\|).$$

  ($h(1) = h(0) + h'(0) + o(1)$.)

Hessian matrix.

- A continuously differentiable function $f : R^n \to R$ is called twice continuously differentiable at $x \in R^n$ if $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ exists and is continuous, $i, j = 1, 2, \ldots, n$.

- The Hessian matrix of f is defined as the $n \times n$ symmetric matrix with elements

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix}_{n \times n}$$

- If f is twice continuously differentiable at every point in R^n, then f is said to be twice continuously differentiable on R^n and is denoted by $f \in C^2$.

- Let $f \in C^2$. For any $x, d \in R^n$, the second directional derivative of f at x in the direction d is defined as

$$f''(x; d) = \lim_{\theta \to 0+} \frac{f'(x + \theta d; d) - f'(x; d)}{\theta},$$

  which is equal to $d^T \nabla^2 f(x) d$.

- For any $x, d \in R^n$, there exists $\theta \in (0, 1)$ such that

$$f(x + d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x + \theta d) d,$$

  or

$$f(x + d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d + o(\|d\|^2).$$

Vector-valued functions.

- Let $F : R^n \to R^m$ be a vector-valued function, i.e.,

$$F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

  F is continuously differentiable at $x \in R^n$ if each f_i is so.

- The derivative $F'(x) \in R^{m\times n}$ of F at x is called the Jacobian matrix of F at x and is defined by

$$\begin{aligned} F'(x) = J(x) &= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}_{m\times n} \\ &= \begin{pmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{pmatrix} = \begin{bmatrix} \nabla f_1(x), & \nabla f_2(x), & \dots, & \nabla f_m(x) \end{bmatrix}^T. \end{aligned}$$

- If $F : R^n \to R^m$ is continuously differentiable in R^n, then for any $x, x+d \in R^n$, it holds that

$$F(x+d) - F(x) = \int_0^1 J(x+td)d\, dt.$$

Let $f : R^n \to R^m$ and $g : R^m \to R$ be continuously differentiable. Let $h = g \circ f : R^n \to R$. Then the chain rule is

$$h'(x) = g'(f(x))f'(x) \quad \text{or} \quad \nabla h(x) = \nabla f(x)\nabla g(f(x)),$$

where $\nabla f = (\nabla f_1, \nabla f_2, \dots, \nabla f_m) \in R^{n\times m}$.

For general mapping.

- Let E_1, E_2 be two finite dimensional Euclidean spaces;
- Let $\mathcal{M}(E_1, E_2)$ be the set of linear operators from E_1 to E_2.

Let $f : E_1 \longrightarrow E_2$ be a mapping from E_1 to E_2. Then, for any $x \in E_1$, $f'(x)$ is a linear operator from E_1 to E_2, that is

$$f' : E_1 \longrightarrow \mathcal{M}(E_1, E_2).$$

Therefore,

$$\begin{aligned} f'' : \quad & E_1 \longrightarrow \mathcal{M}\Big(E_1, \mathcal{M}(E_1, E_2)\Big) \\ f''' : E_1 & \longrightarrow \mathcal{M}\Big(E_1, \mathcal{M}\Big(E_1, \mathcal{M}(E_1, E_2)\Big)\Big), \qquad \dots \end{aligned}$$

Assume differentiability. For any $x, d \in E_1$, it holds that

$$f(x+d) = f(x) + \int_0^1 f'(x+td)d\, dt.$$

$$f(x+d) = f(x) + f'(x)d + \frac{1}{2!}f''(x)[dd] + \dots + \frac{1}{n!}f^{(n)}(x)[d^n] + \dots$$

9.4 Convex sets

Definition 9.4 (凸集) 设$C \subset R^n$. 称C为凸集，如果对$\forall x \in C, y \in C$ 以及$\forall \alpha \in (0,1)$, 都有

$$\alpha x + (1-\alpha)y \in C.$$

R^n中凸集的性质：

1. If C is a convex set and $\beta \in R$, the set $\beta C := \{\beta x : x \in C\}$ is convex.
2. If C and D are convex sets, then the set

$$C + D = \{x + y : x \in C, y \in D\}$$

   is convex.
3. The intersection of any collection of convex sets is convex.

Definition 9.5 (凸组合) 设$x_i \subset R^n$, $\alpha_i \in [0,1]$, $i = 1, 2, \ldots, k$, 并且$\sum_{i=1}^k \alpha_i = 1$, 其中$k$为任意正整数。称$x = \sum_{i=1}^k \alpha_i x_i$ 为$x_1, x_2, \ldots, x_k$的凸组合。

Definition 9.6 (凸包) 设S 为R^n 的子集。R^n 中包含S 的最小的凸集称为S 的凸包，记为$conv(S)$.

Theorem 9.7 (凸包的等价刻画) $conv(S)$由S中点的所有凸组合组成，即

$$conv(S) = \left\{ x = \sum_{i=1}^k \alpha_i x_i \in R^n \;\middle|\; \begin{array}{l} x_i \in S; \\ \alpha_i \in [0,1], i = 1, 2, \ldots, k, \\ \sum_{i=1}^k \alpha_i = 1; \\ k \textit{ is any positive integer.} \end{array} \right\}$$

9.5 Convex functions

Definition 9.8 *Let $C \subset R^n$ be a nonempty convex set and $f : C \to R$.*

- *f is said to be convex on C if*

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y)$$

  holds for all $x, y \in C$ and $\alpha \in (0,1)$.

- *f is said to be strictly convex on C if*

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$$

  holds for all $x, y \in C$, $x \ne y$, and $\alpha \in (0,1)$.

- *f is said to be strongly (or uniformly) convex on C if there exists $\mu > 0$ such that*

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) - \frac{1}{2}\mu\alpha(1-\alpha)\|x - y\|^2$$

  holds for all $x, y \in C$ and $\alpha \in (0,1)$.

Theorem 9.9 *Let $C \subset R^n$ be a nonempty open convex set and let $f : C \to R$ be a differentiable function. Then*

- *f is convex if and only if*
$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \ \forall\, x, y \in C.$$
- *f is strictly convex if and only if*
$$f(y) > f(x) + \nabla f(x)^T (y - x), \ \forall\, x, y \in C, x \neq y.$$
- *f is strongly (or uniformly) convex if and only if*
$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}\mu \|x - y\|^2, \ \forall\, x, y \in C,$$
where $\mu > 0$ is a constant.

Proof:

- Convex case. Necessity. For $x, y \in C$, $\alpha \in (0, 1)$, it holds that
$$f(\alpha y + (1 - \alpha)x) \leq \alpha f(y) + (1 - \alpha) f(x),$$
or, equivalently,
$$f(x + \alpha(y - x)) - f(x) \leq \alpha(f(y) - f(x)).$$
Dividing α and letting $\alpha \to 0+$ yield the result.
Sufficiency. Take $x, y \in C$, $\alpha \in (0, 1)$ and let $z = \alpha x + (1 - \alpha)y$.
$$f(x) \geq f(z) + \nabla f(z)^T (x - z)$$
$$f(y) \geq f(z) + \nabla f(z)^T (y - z).$$
α times first plus $1 - \alpha$ times second implies that f is convex.
- Strictly convex case. Necessity. Let $x, y \in C$ $(x \neq y)$ and $\alpha \in (0, 1)$. Strict convexity of f implies that
$$f(x + \alpha(y - x)) - f(x) < \alpha(f(y) - f(x)).$$
The results follows from
$$f(x + \alpha(y - x)) - f(x) \geq \alpha \nabla f(x)^T (y - x).$$
Sufficiency. Take $x, y \in C$ $(x \neq y)$, $\alpha \in (0, 1)$ and let $z = \alpha x + (1 - \alpha)y$. Clearly $z \neq x$ and $z \neq y$.
$$f(x) > f(z) + \nabla f(z)^T (x - z)$$
$$f(y) > f(z) + \nabla f(z)^T (y - z).$$
α times first plus $1 - \alpha$ times second implies that f is strictly convex.
- Strongly (uniformly) convex case. First, it is easy to show that f is strongly convex with constant $\mu > 0$ if and only if $g = f - \frac{\mu}{2}\| \cdot \|^2$ is convex, i.e.,
$$g(y) \geq g(x) + \nabla g(x)^T (y - x), \ \forall\, x, y \in C,$$
or, equivalently,
$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}\mu \|x - y\|^2, \ \forall\, x, y \in C.$$

■

Theorem 9.10 (Convexity meets differentiability) *Let $C \subset R^n$ be a nonempty open convex set and let $f : C \to R$ be a differentiable function. Then*

- *f is convex if and only if*
$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0, \ \forall\, x, y \in C.$$
- *f is strictly convex if and only if*
$$\langle \nabla f(x) - \nabla f(y), x - y \rangle > 0, \ \forall\, x, y \in C, x \neq y.$$
- *f is strongly (or uniformly) convex if and only if*
$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2, \ \forall\, x, y \in C,$$
where $\mu > 0$ is a constant.

Proof:

- Necessity for all three cases. Suppose f is strongly convex with constant $\mu > 0$. Then, for any $x, y \in C$, it holds that
$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}\mu \|x - y\|^2,$$
$$f(x) \geq f(y) + \nabla f(y)^T (x - y) + \frac{1}{2}\mu \|x - y\|^2.$$
Addition of the two yields
$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2.$$
If f is only convex, let $\mu = 0$ in the above. If f is strictly convex, let $\mu = 0$ and $x \neq y$, then "$\geq$" can be replaced by "$>$" in the above three.

- Sufficiency for convex case. For any fixed $x, y \in C$, it holds that
$$f(y) - f(x) = \nabla f(\xi)^T (y - x),$$
where $\xi = x + t(y - x)$ for some $t \in (0, 1)$. Thus
$$\langle \nabla f(\xi) - \nabla f(x), y - x \rangle = \frac{1}{t} \langle \nabla f(\xi) - \nabla f(x), \xi - x \rangle \geq 0,$$
and
$$f(y) - f(x) = \nabla f(\xi)^T (y - x) \geq \nabla f(x)^T (y - x),$$
i.e., f is convex.

- Sufficiency for strictly convex case. Let $x \neq y$ and replace "$\geq$" by "$>$" in the above.

- Sufficiency for strongly convex case. Let
$$\phi(t) = f(x + t(y - x)) = f(u),$$
where $u = x + t(y - x)$, $t \in (0, 1)$. Note that
$$\phi'(t) = \langle \nabla f(u), y - x \rangle, \quad \phi'(0) = \langle \nabla f(x), y - x \rangle,$$
$$\phi'(t) - \phi'(0) = \frac{1}{t} \langle \nabla f(u) - \nabla f(x), u - x \rangle \geq \frac{\mu}{t} \|u - x\|^2 = t\mu \|y - x\|^2.$$
$$\phi(1) - \phi(0) - \phi'(0) = \int_0^1 (\phi'(t) - \phi'(0)) dt \geq \frac{\mu}{2} \|y - x\|^2,$$
which, by the definition of ϕ, implies that f is strongly convex.

■

Notation: $A \succeq B$ or $A - B \succeq 0$ means that $A - B$ is positive semi-definite; $A \succ B$ or $A - B \succ 0$ means that $A - B$ is positive definite.

Theorem 9.11 (Convexity meets second order differentiability) *Let $C \subset R^n$ be a nonempty open convex set, and let $f : C \to R$ be twice continuously differentiable. Then*

- *f is convex if and only if $\nabla^2 f(x) \succeq 0$ for any $x \in C$.*
- *f is strictly convex if $\nabla^2 f(x) \succ 0$ for any $x \in C$.*
- *f is uniformly convex with constant $\mu > 0$ if and only if*

$$d^T \nabla^2 f(x) d \geq \mu \|d\|^2, \ \forall \ x \in C, d \in R^n,$$

i.e., $\nabla^2 f$ is uniformly positive definite in C (the minimum eigenvalue of $\nabla^2 f(x)$ is greater or equal to μ for all $x \in C$).

Proof: Hints: Consider either

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(\xi)(y - x),$$

where $\xi \in (x, y)$, or

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) + o(\|y - x\|^2).$$

■

9.6 Convex optimization

Let $f : R^n \to R$ and $X \subset R^n$. An optimization problem in general form is as follows

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \ & x \in X. \end{aligned}$$

Definition 9.12 (convex optimization) *The optimization problem $\min\{f(x) : s.t. \ x \in X\}$ is called a convex optimization problem if $X \subset R^n$ is a convex set and $f : X \to R$ is a convex function.*

Definition 9.13 (local minimizer) *1. A point $x^* \in X$ is called a local minimizer if there exists $\delta > 0$ such that*

$$f(x^*) \leq f(x), \forall x \in X \cap \{x \in R^n : \|x - x^*\| < \delta\}.$$

2. A point $x^* \in X$ is called a strict local minimizer if there exists $\delta > 0$ such that

$$f(x^*) < f(x), \forall x \in X \cap \{x \in R^n : \|x - x^*\| < \delta\}, \ x \neq x^*.$$

Definition 9.14 (global minimizer) *1. $x^* \in X$ is a global minimizer if $f(x^*) \leq f(x)$ for all $x \in X$.*

2. *$x^* \in X$ is a strict global minimizer if $f(x^*) < f(x)$ for all $x \in X$, $x \neq x^*$.*

For general optimization problem, it is usually very difficult to find a global optimal solution since any iterative algorithms tend to be trapped around local optimal solutions. The most important feature of convex optimization is that any local optimal solution is also globally optimal.

Theorem 9.15 *Any local optimal solution of a convex optimization problem is also globally optimal.*

9.7 Homework

1. Let $f : R^n \to R$. The epigraph of f is defined by

$$epi(f) := \{(z, x) : \; f(x) \leq z, z \in R, x \in R^n\}.$$

   Prove that f is convex if and only if $epi(f)$ is a convex subset of $R \times R^n$.

2. Suppose $f : R^n \to R$ is a convex function.

   (a) For $z \in R$, the level set of f is defined by

$$L(z) := \{x \in R^n : \; f(x) \leq z\}.$$

   Prove that, if nonempty, $L(z)$ must be convex.

   (b) Define $g : R_{++} \times R^n \to R$ by

$$g(t, x) := t\, f(x/t)$$

   Prove that $g(t, x)$ is convex in (t, x).

3. Let $A \in R^{m \times n}$ and $g : R^m \to R$ be continuously differentiable. Let $h(x) = g(Ax)$ for $x \in R^n$. Prove that

$$h'(x) = g'(Ax)A \quad \text{or} \quad \nabla h(x) = A^T \nabla g(Ax).$$