Optimization and Approximation
TD 1

09 September 2020

Derivatives and descent directions

Exercise 1

Let A be an open subset of $\mathbb{R}^n$ and let

$$\begin{array}{rccc} f: & A \subseteq \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & \mathbf{x} = (x_1, \ldots, x_n)^T & \longmapsto & f(\mathbf{x}) \end{array}$$

be continuously differentiable. Consider a direction $p \in \mathbb{R}^n$ be a direction. Show that

$$\frac{\partial f}{\partial \mathbf{p}}(\mathbf{x}) = \nabla f(\mathbf{x})^T \mathbf{p}.$$

Solution:

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{p}}(\mathbf{x}) &= \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{p}) - f(\mathbf{x})}{h} &= \left[\frac{\mathrm{d}}{\mathrm{d}\alpha} f(\mathbf{x} + \alpha \mathbf{p})\right]_{\alpha=0} \\ &= \left[\nabla f(\mathbf{x} + \alpha \mathbf{p})^T \mathbf{p}\right]_{\alpha=0} &= \nabla f(\mathbf{x})^T \mathbf{p}. \end{aligned}$$

Exercise 2

Prove that for a quadratic function

$$\begin{array}{rccl} q: & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & \mathbf{x} & \longmapsto & q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A \mathbf{x} - \mathbf{b}^T \mathbf{x} = \frac{1}{2}\sum_{i,j=1}^n x_i a_{ij} x_j - \sum_{i=1}^n b_i x_i, \end{array}$$

with $A \in \mathbb{R}^{n \times n}$ symmetric and $\mathbf{b} \in \mathbb{R}^n$, it holds that

a) $\nabla q(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$.

b) $H(\mathbf{x}) = A$.

Solution: We prove the first point. $\forall \ell = 1, \ldots, n$

$$\Big(\nabla q(\mathbf{x})\Big)_\ell \quad = \quad \frac{\partial q(\mathbf{x})}{\partial x_\ell} \quad = \quad \frac{1}{2}\frac{\partial}{\partial x_\ell}\left(\sum_{i,j=1}^n x_i a_{ij} x_j\right) - \frac{\partial}{\partial x_\ell}\left(\sum_{i=1}^n b_i x_i\right).$$

Notice that

$$\frac{\partial}{\partial x_\ell}\left(\sum_{i=1}^n b_i x_i\right) \quad = \quad \frac{\partial}{\partial x_\ell}\left(\sum_{\substack{i=1\\ i\neq\ell}}^n b_i x_i + b_\ell x_\ell\right) \quad = \quad b_\ell$$

and

$$\begin{aligned}
\frac{\partial}{\partial x_\ell}\left(\sum_{i,j=1}^n x_i a_{ij} x_j\right) &= \frac{\partial}{\partial x_\ell}\left(\sum_{\substack{i,j=1\\ i,j\neq\ell}}^n x_i a_{ij} x_j + \sum_{\substack{i=1\\ i\neq\ell}}^n x_i a_{i\ell} x_\ell + \sum_{\substack{j=1\\ j\neq\ell}}^n x_\ell a_{\ell j} x_j + a_{\ell\ell} x_\ell x_\ell\right) \\
&= \frac{\partial}{\partial x_\ell}\sum_{\substack{i,j=1\\ i,j\neq\ell}}^n x_i a_{ij} x_j + \frac{\partial}{\partial x_\ell}\sum_{\substack{i=1\\ i\neq\ell}}^n x_i a_{i\ell} x_\ell + \frac{\partial}{\partial x_\ell}\sum_{\substack{j=1\\ j\neq\ell}}^n x_\ell a_{\ell j} x_j + \frac{\partial}{\partial x_\ell} a_{\ell\ell} x_\ell x_\ell \\
&= 0 + \sum_{\substack{i=1\\ i\neq\ell}}^n x_i a_{i\ell} + \sum_{\substack{j=1\\ j\neq\ell}}^n a_{\ell j} x_j + 2a_{\ell\ell} x_\ell \\
&= \sum_{\substack{i=1\\ i\neq\ell}}^n x_i a_{i\ell} + \sum_{\substack{j=1\\ j\neq\ell}}^n x_j a_{j\ell} + 2a_{\ell\ell} x_\ell \quad = \quad 2\sum_{\substack{i=1\\ i\neq\ell}}^n x_i a_{i\ell} + 2a_{\ell\ell} x_\ell \quad = \quad 2\sum_{i=1}^n a_{i\ell} x_i.
\end{aligned}$$

Then

$$\Big(\nabla q(\mathbf{x})\Big)_\ell \quad = \quad \sum_{i=1}^n x_i a_{i\ell} - b_\ell \quad = \quad \Big(A\mathbf{x} - \mathbf{b}\Big)_\ell.$$

Analogously for the second point.

Exercise 3

Let $f : \mathbb{R}^2 \to \mathbb{R}$ (resp. $f : \mathbb{R}^3 \to \mathbb{R}$) be a continuously differentiable function. Prove that the gradient of f is perpendicular to the level curves (resp. level surfaces) of f.

Solution:

1. Assume $n = 2$. We show that the slope of the gradient is the negative reciprocal of the slope of the level curves (this ensures that the gradient is perpendicular to the level curves.) Note that for a 2D vector $[a, b]^T$ the slope is b/a. We consider a level curve $f(x, y) = c$. Implicitly this gives a relation between x and y, which means that y can be thought of as a function of x, say $y = y(x)$.We then rewrite the equation of the level curve as $f(x, y(x)) = c$. The chain rule gives:
$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0$$

and so

$$\frac{dy}{dx} = -\frac{\partial f}{\partial x} \Big/ \frac{\partial f}{\partial y}$$

which is the slope of the level curve. Now, $\nabla f(x,y) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]^T$ and its slope is $\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}$. The slope of the gradient is the negative reciprocal of the slope of the level curve.

2. Assume now $n = 3$. Let $r(t) = (x(t), y(t), z(t))$ be a curve on the level surface. Let $g(t) = f(x(t), y(t), z(t))$. Since the curve is on the level surface we have that $g(t) = f(x(t), y(t), z(t)) = c$. Differentiating this equation with respect to t gives

$$\frac{dg}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = 0$$

that in vector form is

$$\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right]\left[\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right]^T = 0$$

that is $\nabla f(\mathbf{x})^T r'(t) = 0$. Since the dot product is 0, we have shown that the gradient is perpendicular to the tangent to any curve that lies on the level surface.

Exercise 4

a) Graphically depict some level curves of the function

$$f(x, y) = 2(x-1)^2 + y^2 + 1.$$

Choose a point on one of the level curves and depict the gradient and an example of descent direction in this point.

b) Let f be a differentiable function, $f : \mathbb{R}^n \to \mathbb{R}$. Let $\mathbf{d} \in \mathbb{R}^n$ be a non-zero vector with the following property:

$$\|\nabla f(\mathbf{x}) + \mathbf{d}\| < \|\nabla f(\mathbf{x})\|. \tag{1}$$

Show that $\mathbf{d}$ is a descent direction in $\mathbf{x}$.

Solution:

a) The level curves are ellipses with focal points $(1, 0)$ and $(0, 0)$. In each point, the gradient is perpendicular to the level curves, the descent directions form an obtuse angle with the gradient.

b) From (1) it holds:

$$\|\nabla f(\mathbf{x}) + \mathbf{d}\|^2 = \|\nabla f(\mathbf{x})\|^2 + \|\mathbf{d}\|^2 + \nabla f(\mathbf{x})^T\mathbf{d} < \|\nabla f(\mathbf{x})\|^2.$$

Then, $\|\mathbf{d}\|^2 + \nabla f(\mathbf{x})^T\mathbf{d} < 0$ and $0 < \|\mathbf{d}\|^2 < -\nabla f(\mathbf{x})^T\mathbf{d}$. Consequently it holds $\nabla f(\mathbf{x})^T\mathbf{d} < 0$.

Exercise 5

Consider the function $f(x, y) = (x + y^2)^2$. At the point $\mathbf{z}_0 = (x_0, y_0) = (1, 0)^T$, consider the direction $\mathbf{p}_0 = (-1, -1)^T$.

a) Show that $\mathbf{p}_0$ is a descent direction.

b) Assume to solve the problem of minimizing f by using a line-search strategy, which employs the direction $\mathbf{p}_0$ at the first step and the point (x_0, y_0) as a starting guess. What is the best value of α that should be used in the line-search at fist step?

c) Find all the stationary points of f and establish if they are local/global minima.

d) Is this function f strictly convex?

Solution:

a) $\mathbf{p}_0$ is a descent direction because $\nabla f(\mathbf{z}_0)^T \mathbf{p}_0 = (2, 0)(-1, -1)^T = -2 < 0$.

b) To find α we have to solve $\min_{\alpha>0} f(\mathbf{x}_0 + \alpha \mathbf{p}_0)$. Let us define $\phi(\alpha) = f(\mathbf{x}_0 + \alpha \mathbf{p}_0) = (1 - \alpha + \alpha^2)^2$. It holds $\phi'(\alpha) = 2(1 - \alpha + \alpha^2)(-1 + 2\alpha) = 0$. The optimal value is then $\alpha = 1/2$.

c) The stationary points are $(0, 0)$ and all the points of the form $(-y^2, y)$ for $y \in \mathbb{R}$ and these are all global minima because the value of f in these points is zero, which is the lower bound of f ($f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$).

d) The function cannot be strictly convex because it has more than one global minimum.

Exercise 6

Give an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that both second order partial derivatives $\frac{\partial f}{\partial x \partial y}$ and $\frac{\partial f}{\partial y \partial x}$ exist, but $\frac{\partial f}{\partial x \partial y} \neq \frac{\partial f}{\partial y \partial x}$.

Solution: The Peano function:

$$f(x, y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

The partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ exist and are everywhere continuous. However, the second order partial derivatives are not continuous at $(0, 0)$, and the symmetry fails.

Convex sets

Exercise 7

Show that the *hyperbolic set* $\{\mathbf{x} \in \mathbb{R}^2_+ \mid x_1 x_2 \geq 1\}$ is convex. As a generalization, show that $\{\mathbf{x} \in \mathbb{R}^n_+ \mid \prod_{i=1}^n x_i \geq 1\}$ is convex.

Solution: We have:

$$\prod_{i=1}^{n} x_i \geq 1 \Leftrightarrow \sum_{i=1}^{n} \log x_i \geq 0$$

Since $f(x) = \log x$ is a concave function, we have:

$$\log(tx + (1-t)y) \geq t \log x + (1-t) \log y$$

for all $x, y \in \mathbb{R}$ and $0 \leq t \leq 1$.

Given two point $\mathbf{x}, \mathbf{y}$ in the hyperbolic set, we have:

$$\begin{aligned}\sum_{i=1}^{n} \log(tx_i + (1-t)y_i) &\geq \sum_{i=1}^{n} t \log x_i + (1-t) \log y_i \\ &= t\left(\sum_{i=1}^{n} \log x_i\right) + (1-t)\left(\sum_{i=1}^{n} \log y_i\right) \geq 0\end{aligned}$$

Exercise 8

Which of the following sets are convex?

a) A *slab*, i.e., a set of the form $\{\mathbf{x} \in \mathbb{R}^n \mid \ \alpha \leq \mathbf{a}^T\mathbf{x} \leq \beta\}$, for $\alpha, \beta \in \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^n$.

b) A *rectangle*, i.e, a set of the form $\{\mathbf{x} \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \ldots, n\}$, for $\alpha_i, \beta_i \in \mathbb{R}$. A rectangle is sometimes called a *hyperrectangle* when $n > 2$.

c) A *wedge*, i.e, $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}_1^T\mathbf{x} \leq b_1, \mathbf{a}_2^T\mathbf{x} \leq b_2\}$, for $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^n$ and $b_1, b_2 \in \mathbb{R}$.

d) The set of points closer to a given point than a given set, i.e., given $S \subset \mathbb{R}^n$,

$$\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}_0\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2, \forall \mathbf{y} \in S\}.$$

e) The set of points closer to one set than another, i.e., given $S, T \subseteq \mathbb{R}^n$

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{dist}(\mathbf{x}, S) \leq \mathbf{dist}(\mathbf{x}, T)\}$$

where

$$\mathbf{dist}(\mathbf{x}, S) = \inf_{\mathbf{z} \in S} \|\mathbf{x} - \mathbf{z}\|_2$$

f) The set $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} + S_2 \subseteq S_1\}$, where $S_1, S_2 \subseteq \mathbb{R}^n$ with S_1 convex.

g) The set of points whose distance to $\mathbf{a} \in \mathbb{R}^n$ does not exceed a fixed fraction θ of the distance to $\mathbf{b} \in \mathbb{R}^n$, i.e., the set $\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{a}\|_2 \leq \theta\|\mathbf{x} - \mathbf{b}\|_2\}$. You can assume that $\mathbf{a} \neq \mathbf{b}$ and $0 \leq \theta \leq 1$.

Solution:

a) If $\alpha \leq \mathbf{a}^T\mathbf{x}_1 \leq \beta$ and $\alpha \leq \mathbf{a}^T\mathbf{x}_2 \leq \beta$, then $\forall t, 0 \leq t \leq 1$:

$$\alpha \leq t(\mathbf{a}^T\mathbf{x}_1) + (1-t)(\mathbf{a}^T\mathbf{x}_2) = \mathbf{a}^T(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \leq \beta$$

Thus, this is a convex set.

b) Convex set. Similar to question a).

c) Convex set. Similar to question a)

d) Consider the set $T(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{z} - \mathbf{x}_0\|_2 \leq \|\mathbf{z} - \mathbf{y}\|_2\}$, then.

$\|\mathbf{z} - \mathbf{x}_0\|_2 \leq \|\mathbf{z} - \mathbf{y}\|_2 \Leftrightarrow \quad -2\mathbf{z}^T\mathbf{x}_0 + \|\mathbf{x}_0\|^2 \leq -2\mathbf{z}^T + \|\mathbf{y}\|^2 \Leftrightarrow \quad 2\mathbf{z}^T(\mathbf{y} - \mathbf{x}_0) \leq \|\mathbf{y}_0\|^2 - \|\mathbf{x}_0\|^2$

Therefore $T(\mathbf{y})$ is a convex set (a half space). Moreover, we have:

$$\{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2, \forall \mathbf{y} \in S\} = \cap_{\mathbf{y} \in S} T(\mathbf{y})$$

Since intersection of convex sets is a convex set, it concludes the proof.

e) It is not a convex set. Taking $S = \{1, -1\}$, $T = \{0\}$.

f) Convex set. Suppose that $\mathbf{x}_1 + S_2 \subseteq S_1$ and $\mathbf{x}_2 + S_2 \subseteq S_1$. We need to prove that $(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) + S_2 \subseteq S_1$. Indeed, $\forall \mathbf{y} \in S_2$, we have:

$$\mathbf{x}_1 + \mathbf{y} \in S_1$$
$$\mathbf{x}_2 + \mathbf{y} \in S_1$$

Since S_1 is convex, $t(\mathbf{x}_1 + \mathbf{y}) + (1-t)(\mathbf{x}_2 + \mathbf{y}) \in S_1$. Therefore, $(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) + \mathbf{y} \in S_1$, which is what we need to prove.

g) Convex set, in fact a ball:

$$\begin{aligned} \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{a}\|_2 \leq \theta\|\mathbf{x} - \mathbf{b}\|_2\} &= \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{a}\|_2^2 \leq \theta^2\|\mathbf{x} - \mathbf{b}\|^2\} \\ &= \{\mathbf{x} \mid (1-\theta^2)\mathbf{x}^T\mathbf{x} - 2(\mathbf{a} - \theta^2\mathbf{b})^T\mathbf{x} + (\mathbf{a}^T\mathbf{a} - \theta^2\mathbf{b}^T\mathbf{b}) \leq 0\} \end{aligned}$$

If $\theta = 1$, it is a half space. If $\theta < 1$, this is a quadratic inequality describing a ball.

Convex functions

Exercise 9

Prove that the following functions are convex:

a) Every norm in $\mathbb{R}^n$ (that is if $f : \mathbb{R}^n \to \mathbb{R}$ is a norm, then f is a convex function).

b) $f(\mathbf{x}) = \max_i x_i$ where x_i, $i = 1, \ldots, n$, is the i-th component of $\mathbf{x} \in \mathbb{R}^n$.

c) $f(x, y) = x^2/y$, with $y > 0$.

Solution:

a) Norm is convex:

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq f(t\mathbf{x}) + f((1-t)\mathbf{y}) = tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$

The inequality follows from the triangle inequality, and the equality follows from homogeneity of a norm.

b) Max function. For $a, b \in \mathbb{R}$, it holds $\max(a+b) \leq \max a + \max b$. Then

$$f(t\mathbf{x} + (1-t)\mathbf{y}) = \max_i(tx_i + (1-t)y_i) \leq \max_i tx_i + \max(1-t)y_i = tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$

c) The Hessian matrix is positive semidefinite:

$$\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succcurlyeq 0$$

Exercise 10

Suppose that $S \subseteq \mathbb{R}^n$ is a convex set and $f : \mathbb{R}^n \to \mathbb{R}$ is an affine function. Prove that $f(S)$ is a convex set.

Solution: Let $x, y \in f(S)$. Then there exist $\mathbf{a}, \mathbf{b} \in S$ such that $x = f(\mathbf{a}), y = f(\mathbf{b})$. Then for $t \in (0,1)$ $tx + (1-t)y = tf(\mathbf{a}) + (1-t)f(\mathbf{b}) = f(t\mathbf{a} + (1-t)\mathbf{b}) \in f(S)$.

Exercise 11

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable and convex function.

a) Show that the set of global minimizers of f is a convex set.

b) Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $f(\mathbf{y}) < f(\mathbf{x})$. Show that $\mathbf{y} - \mathbf{x}$ is a descent direction in $\mathbf{x}$.

Solution:

a) Let G be the set of global minimizers of f. Let $\mathbf{x}, \mathbf{y} \in G$ and assume $t_in(0,1)$. We want to prove that $t\mathbf{x} + (1-t)\mathbf{y} \in G$. It holds

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}) = tf(\mathbf{x}) + (1-t)f(\mathbf{x}) = f(\mathbf{x})$$

then also $t\mathbf{x} + (1-t)\mathbf{y}$ is a global minimizer.

b) We show that in the direction $\mathbf{y} - \mathbf{x}$ the function decreases. The points in such direction are, given $t > 0$ $\mathbf{x} + t(\mathbf{y} - \mathbf{x}) = (1-t)\mathbf{x} + t\mathbf{y}$. From the convexity of f, we have $f((1-t)\mathbf{x} + t\mathbf{y}) < (1-t)f(\mathbf{x}) + tf(\mathbf{y}) < f(\mathbf{x})$, where the last inequality follows from $f(\mathbf{y}) < f(\mathbf{x})$.

Exercise 12

Let A be a convex subset of $\mathbb{R}^n$. Suppose the function $f : A \longrightarrow \mathbb{R}$ is twice differentiable over A. Prove that the followings are equivalent:

(i) f is convex.

(ii) $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$, for all $\mathbf{x}, \mathbf{y} \in A$.

(iii) $\nabla^2 f(\mathbf{x}) \succeq 0$, for all $\mathbf{x} \in A$.

Hint: For proving $(i) \Rightarrow (ii)$, notice that $f(t\mathbf{y}+(1-t)\mathbf{x}) = f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))$. For $(ii) \Rightarrow (i)$ define $z = tx+(1-t)y$ and use the inequality once for on $(\mathbf{x},\mathbf{z})$ and once for $(\mathbf{y},\mathbf{z})$. For $(ii) \iff (iii)$, first prove it for $n = 1$ and then generalize it to $n > 1$. For the generalization you can use the following theorem:

Theorem 1 (Convexity along lines)**.** *A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \longrightarrow \mathbb{R}$ given by $g(t) = f(\mathbf{x}+t\mathbf{y})$ is convex (as a univariate function) for all $\mathbf{x}$ in the domain of f and all $\mathbf{y} \in \mathbb{R}^n$.*

Solution:

- $(i) \Rightarrow (ii)$ From definition of convexity

$$f(t\mathbf{y} + (1-t)\mathbf{x}) \le tf(\mathbf{y}) + (1-t)f(\mathbf{x}) \text{ that is}$$
$$f(\mathbf{x} + t(\mathbf{y}-\mathbf{x})) \le f(\mathbf{x}) + t(f(\mathbf{y}) - f(\mathbf{x}))$$

Then
$$f(\mathbf{y}) - f(\mathbf{x}) \ge \frac{f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - f(\mathbf{x})}{t} \text{ for all } t \in (0,1].$$
Passing to the limit for $t \to 0$ we get the thesis.

- $(ii) \Rightarrow (i)$ Define $\mathbf{z} = t\mathbf{x} + (1-t)\mathbf{y}$. We use (ii) on $(\mathbf{x},\mathbf{z})$ and on $(\mathbf{y},\mathbf{z})$:

$$f(\mathbf{x}) \ge f(\mathbf{z}) + \nabla f(\mathbf{z})^T(\mathbf{x}-\mathbf{z}), \tag{2}$$
$$f(\mathbf{y}) \ge f(\mathbf{z}) + \nabla f(\mathbf{z})^T(\mathbf{y}-\mathbf{z}). \tag{3}$$

Multiplying (2) by t, (3) by $(1-t)$ and adding, we get

$$\begin{aligned} tf(\mathbf{x}) + (1-t)f(\mathbf{y}) &\ge f(\mathbf{z}) + \nabla f(\mathbf{z})^T(t\mathbf{x} + (1-t)\mathbf{y} - z) = f(\mathbf{z}) \\ &= f(t\mathbf{x} + (1-t)\mathbf{y}) \end{aligned}$$

- $(ii) \Rightarrow (iii)$ Assume $n = 1$. Assume $y > x$. It holds

$$f(y) \ge f(x) + \nabla f(x)^T(y-x),$$
$$f(x) \ge f(y) + \nabla f(y)^T(x-y).$$

Then
$$f'(x)(y-x) \le f(y) - f(x) \le f'(y)(y-x).$$
Dividing the first and the last term by $(y-x)^2$ we obtain

$$\frac{f'(y) - f'(x)}{(y-x)} \ge 0$$

for all x, y, $x \neq y$. As we let $y \to x$, we get $f''(x) \ge 0$, $\forall x \in A$.

- $(iii) \Rightarrow (ii)$ Assume $n = 1$. By the second order Taylor's formula we have:

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2}f''(z)(y-x)^2, \text{ for all } z \in [x,y].$$

From the assumption we get the thesis.

Generalization to $n > 1$: to establish $(ii) \iff (iii)$ we use Theorem 1. We define $g(\alpha) = f(\mathbf{x}_0 + \alpha\mathbf{v})$ for $\mathbf{x}_0 \in A$ and $\mathbf{v} \in \mathbb{R}^n$. f is convex if g is convex $\forall \mathbf{x}_0 \in A$ and $\forall \mathbf{v} \in \mathbb{R}^n$. We have just proved that g (that is defined in dimension 1) is convex if and only if $g''(\alpha) \geq 0$. Then

$$g''(\alpha) = \mathbf{v}^T \nabla^2 f(\mathbf{x}_0 + \alpha\mathbf{v})\mathbf{v} \geq 0,$$

$\forall \mathbf{x}_0 \in A$ and $\forall \mathbf{v} \in \mathbb{R}^n$. Hence, f is convex if and only if $\nabla^2 f(\mathbf{x}) \succeq 0$, for all $\mathbf{x} \in A$.

Exercise 13

An optimality criterion for constrained convex problems:
Let $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ be a differentiable convex function over a convex set A. Consider the following constrained problem:

$$\min_{\mathbf{x} \in X} f(\mathbf{x})$$

with $X = \{\mathbf{x} \in \mathbb{R}^n \mid f_i(\mathbf{x}) \leq 0, i = 1, \ldots, m\}$, and f_i, $i = 1, \ldots, m$ convex functions. Show that $\mathbf{x} \in X$ is a local minimum if and only if for all $\mathbf{y} \in X$,

$$\nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \geq 0. \tag{4}$$

Hint: notice that X is a convex set.

Solution: First suppose that $\mathbf{x}$ satisfies (4). Then for $\mathbf{y} \in X$ we have, by (4) that $f(\mathbf{y}) \geq f(\mathbf{x})$. This shows $\mathbf{x}$ is an optimal point. Conversely, suppose $\mathbf{x}$ is optimal, but the condition 4 does not hold, i.e., for some $\mathbf{y} \in X$ we have $\nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) < 0$. Consider the point $\mathbf{z}(t) = t\mathbf{y} + (1 - t)\mathbf{x}$, where $t \in [0, 1]$ is a parameter. Since $\mathbf{z}(t)$ is on the line segment between $\mathbf{x}$ and $\mathbf{y}$, and the feasible set X is convex, $\mathbf{z}(t)$ is feasible. We claim that for small positive t we have $f(\mathbf{z}(t)) < f(\mathbf{x})$, which will prove that $\mathbf{x}$ is not optimal. To show this, note that

$$\frac{d}{dt} f(\mathbf{z}(t))\bigg|_{t=0} = \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) < 0,$$

so for small positive t, we have $f(\mathbf{z}(t)) < f(\mathbf{x})$.