Advanced Graph Algorithms and Optimization Spring 2023

Course Introduction & Convex Optimization

R. Kyng & M. Probst *Problem Set 1 — Monday, February 20th*

These exercises will not count toward your grade, but you are encouraged to solve them all. This exercise sheet contains exercises relating to lectures in Week 1.

To get feedback, you must hand in your solutions by 23.59 pm on March 2nd. Both hand-written and LaTeX solutions are acceptable, but we will only attempt to read legible text.

Exercise 1

Let us define the *α-sub-level set* of a function $f : \mathbb{R}^n \to \mathbb{R}$ to be the set $S_\alpha \stackrel{\text{def}}{=} \{\boldsymbol{x} : f(\boldsymbol{x}) \leq \alpha\}$.

(i) Prove that if a function f is convex, then all its sub-level sets are convex sets.

(ii) Is it true that a function whose sub-levels sets are all convex is necessarily convex?

Solution.

(i) If S_α is empty, then the statement is trivial. Let $\boldsymbol{x}, \boldsymbol{z} \in S_\alpha$ for some α and assume that $\theta \in [0, 1]$. We have $f(\boldsymbol{x}) \leq \alpha$ and $f(\boldsymbol{z}) \leq \alpha$ by definition. Furthermore, since f is convex, we have $f(\theta\boldsymbol{x} + (1 - \theta)\boldsymbol{z}) \leq \theta f(\boldsymbol{x}) + (1 - \theta)f(\boldsymbol{z})$. Combining the above equations yields

$$f(\theta\boldsymbol{x} + (1 - \theta)\boldsymbol{z}) \leq \theta\alpha + (1 - \theta)\alpha = \alpha.$$

This implies that $\theta\boldsymbol{x} + (1 - \theta)\boldsymbol{z}$ is in S_α.

(ii) This is not true. As a counterexample, consider the function $f(x) = x^3$. It is easy to check that all sub-level sets of f are convex. However, function f is not convex. A function whose all sub-level sets are convex is called *quasiconvex*.

Exercise 2

Recall the definition of Laplacian $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top$.

(i) We can also define Laplacian as $\boldsymbol{L} \stackrel{\text{def}}{=} \boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{A}$ is the weighted adjacency matrix, i.e. $\boldsymbol{A}(u, v) = 1/\boldsymbol{r}(u, v)$, and $\boldsymbol{D} \stackrel{\text{def}}{=} \text{diag}_{v \in V}\, \boldsymbol{w}(v)$ for $\boldsymbol{w}(v) := \sum_{(u,v) \in E} 1/\boldsymbol{r}(u, v)$. Prove that these two definitions are equivalent.

(ii) Given a function on the vertices, $\boldsymbol{x} \in \mathbb{R}^V$, the Laplacian quadratic form is

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \sum_{(u,v) \in E} \frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))^2}{\boldsymbol{r}(u, v)}.$$

Prove the above equality and building on that, show that $\boldsymbol{L}$ is positive semi-definite.

(iii) What is the kernel of $\boldsymbol{L}$, which is denoted by $\mathrm{Ker}(\boldsymbol{L})$?

Solution.

(i) For any $v \in V$, we have

$$(\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top)(v,v) = \sum_{e \in E} \boldsymbol{B}(v,e)(\boldsymbol{R}^{-1}\boldsymbol{B}^\top)(e,v).$$

Since $\boldsymbol{R}^{-1} = \mathrm{diag}_{e\in E}\, 1/\boldsymbol{r}(e)$, then $(\boldsymbol{R}^{-1}\boldsymbol{B}^\top)(e,v) = \boldsymbol{B}(v,e)/\boldsymbol{r}(e)$. Therefore, we have

$$(\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top)(v,v) = \sum_{e \in E} \boldsymbol{B}(v,e)\frac{1}{\boldsymbol{r}(e)}\boldsymbol{B}(v,e) = \sum_{(u,v)\in E} \frac{1}{\boldsymbol{r}(u,v)} = \boldsymbol{w}(v).$$

Thus, the diagonal of $\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top$ and $\boldsymbol{D} - \boldsymbol{A}$ are equivalent. For $v \neq u \in V$, we have

$$(\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top)(v,u) = \sum_{e \in E} \boldsymbol{B}(v,e)(\boldsymbol{R}^{-1}\boldsymbol{B}^\top)(e,u) = \sum_{e \in E} \boldsymbol{B}(v,e)\frac{1}{\boldsymbol{r}(e)}\boldsymbol{B}(u,e).$$

We observe that this is equal to 0 if there is no edge between v and u and $-1/\boldsymbol{r}(v,u)$ if $(v,u) \in E$. Therefore, we get $(\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top)(v,u) = (\boldsymbol{D} - \boldsymbol{A})(v,u)$.

(ii) We have

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \boldsymbol{x}^\top(\boldsymbol{D} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}.$$

We recall that $\boldsymbol{D}$ is a diagonal matrix; thus, we have

$$\boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{x} = \sum_{v \in V} \boldsymbol{D}(v,v)\boldsymbol{x}(v)^2 = \sum_{v \in V} \boldsymbol{w}(v)\boldsymbol{x}(v)^2.$$

Furthermore,

$$\begin{aligned}
\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} &= \sum_{v \in V} \boldsymbol{x}(v)(\boldsymbol{A}\boldsymbol{x})(v) \\
&= \sum_{v \in V} \boldsymbol{x}(v) \sum_{u \in V} \boldsymbol{A}(v,u)\boldsymbol{x}(u) \\
&= \sum_{v \in V}\sum_{u \in V} \boldsymbol{A}(v,u)\boldsymbol{x}(v)\boldsymbol{x}(u) \\
&= \sum_{(u,v)\in E} \frac{2\boldsymbol{x}(v)\boldsymbol{x}(u)}{\boldsymbol{r}(u,v)}.
\end{aligned}$$

Combining the above equalities yields

$$\begin{aligned}
\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} &= \sum_{v \in V} \boldsymbol{w}(v)\boldsymbol{x}(v)^2 - \sum_{(u,v)\in E} \frac{2\boldsymbol{x}(v)\boldsymbol{x}(u)}{\boldsymbol{r}(u,v)} \\
&= \sum_{(u,v)\in E} \frac{\boldsymbol{x}(u)^2 + \boldsymbol{x}(v)^2}{\boldsymbol{r}(u,v)} - \sum_{(u,v)\in E} \frac{2\boldsymbol{x}(v)\boldsymbol{x}(u)}{\boldsymbol{r}(u,v)} \\
&= \sum_{(u,v)\in E} \frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))^2}{\boldsymbol{r}(u,v)}.
\end{aligned}$$

Finally, since the right-hand side of the above equation is non-negative, $\boldsymbol{L}$ is positive semidefinite.

(iii) Assume that the underlying graph is connected. We explain at the end how a similar argument applies to the disconnected case. Suppose that $\boldsymbol{x} \in \mathbb{R}^V$ is in the kernel of $\boldsymbol{L}$, that is, $\boldsymbol{Lx} = \boldsymbol{0}$. Let $u = \arg\min_{v \in V} \boldsymbol{x}(v)$. Then, we have

$$(\boldsymbol{Lx})(u) = \sum_{v \in V} \boldsymbol{L}(u,v)\boldsymbol{x}(v).$$

Recall that $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$. Thus, we get

$$(\boldsymbol{Lx})(u) = \sum_{(u,v)\in E} \frac{1}{\boldsymbol{r}(u,v)}\boldsymbol{x}(u) - \sum_{(u,v)\in E} \frac{1}{\boldsymbol{r}(u,v)}\boldsymbol{x}(v).$$

Let U be the set of vertices which are adjacent to u. Note that $U \neq \emptyset$ since the graph is connected. This implies that $\sum_{(u,v)\in E} \frac{1}{\boldsymbol{r}(u,v)} > 0$. Since $u = \arg\min_{v \in V} \boldsymbol{x}(v)$ and $(\boldsymbol{Lx})(u) = 0$, we can conclude that $\boldsymbol{x}(u) = \boldsymbol{x}(v)$ for any $v \in U$. By applying the same argument recursively, we have

$$\mathrm{Ker}(\boldsymbol{L}) = \{\boldsymbol{x} : \boldsymbol{x} = \alpha\mathbf{1} \text{ for } \alpha \in \mathbb{R}\}.$$

Assume that the underlying graph is not connected and has $k > 1$ connected components $\mathcal{C}_1, \cdots, \mathcal{C}_k$. Using the same argument for each of the components, we can conclude that vectors $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k$ form the basis of $\mathrm{Ker}(\boldsymbol{L})$, where $\boldsymbol{u}_i(v) = 1$ if $v \in \mathcal{C}_i$ and 0 otherwise for $i \in [k]$.

Exercise 3

(i) Prove that for a matrix $\boldsymbol{A}$ we have $\mathrm{im}(\boldsymbol{A}) = \ker(\boldsymbol{A}^\top)^\perp$, where $\mathrm{im}(\boldsymbol{A})$ denotes the image of $\boldsymbol{A}$ and $\ker(\boldsymbol{A}^\top)^\perp$ is the orthogonal complement to $\ker(\boldsymbol{A}^\top)$.

(ii) Building on part (i), prove that in a connected graph with resistances $\boldsymbol{r} \in \mathbb{R}^E_{>0}$, an electrical flow $\boldsymbol{f}$ routing demand $\boldsymbol{d}$ exists if and only if $\mathbf{1}^\top \boldsymbol{d} = 0$.

Solution.

(i) Let $\boldsymbol{a}_i \in \mathbb{R}^m$ for $i \in [n]$ denote the i-th column in $\boldsymbol{A}$. Since $(\boldsymbol{A}^\perp)^\perp = \boldsymbol{A}$, we only need to prove that $\mathrm{im}(\boldsymbol{A})^\perp = \ker(\boldsymbol{A}^\top)$.

Firstly, for $\boldsymbol{x} \in \ker(\boldsymbol{A}^\top)$, we have $\boldsymbol{A}^\top \boldsymbol{x} = \boldsymbol{0}$. This implies that $\boldsymbol{a}_i^\top \boldsymbol{x} = \boldsymbol{0}$ for $i \in [n]$. Since $\mathrm{im}(\boldsymbol{A}) = \mathrm{span}\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n\}$, we will have $\boldsymbol{x}^\top \boldsymbol{z} = \boldsymbol{0}$ for any $\boldsymbol{z} \in \mathrm{im}(\boldsymbol{A})$. This implies that $\boldsymbol{x} \in \mathrm{im}(\boldsymbol{A})^\perp$. Therefore, we have $\ker(\boldsymbol{A}^\top) \subseteq \mathrm{im}(\boldsymbol{A})^\perp$.

Secondly, for any $\boldsymbol{x} \in \mathrm{im}(\boldsymbol{A})^\perp$, we have $\boldsymbol{x}^\top \boldsymbol{a}_i = \boldsymbol{0}$ for any $i \in [n]$. This implies that $\boldsymbol{A}^\top \boldsymbol{x} = \boldsymbol{0}$. Thus, $\mathrm{im}(\boldsymbol{A})^\perp \subseteq \ker(\boldsymbol{A}^\top)$.

(ii) By definition, there exists $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{Lx} = \boldsymbol{d}$ if and only if $\boldsymbol{d} \in \mathrm{im}(\boldsymbol{L})$. From part (i), we know that $\mathrm{im}(\boldsymbol{L}) = \ker(\boldsymbol{L}^\top)^\perp$. Furthermore, we have $\boldsymbol{L} = \boldsymbol{L}^\top$. Hence, an electrical flow $\boldsymbol{f}$ routing $\boldsymbol{d}$ exists if and only if $\boldsymbol{d} \in \ker(\mathrm{L})^\perp$. This is the same as $\boldsymbol{z}^\top \boldsymbol{d} = \boldsymbol{0}$ for any $\boldsymbol{z} \in \ker(\boldsymbol{L})$. By applying Exercise 2, part (iii), this holds if and only if $\mathbf{1}^\top \boldsymbol{d} = \boldsymbol{0}$.

Exercise 4

We define[1] the gradient of a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$ as the function $\boldsymbol{\nabla} f : \mathbb{R}^n \to \mathbb{R}^n$ with $\boldsymbol{\nabla} f(\boldsymbol{x})(i) = \frac{d}{d\boldsymbol{x}(i)} f(\boldsymbol{x})$, i.e. we consider f at a point $\boldsymbol{x}$, treat the ith coordinate as a variable $\boldsymbol{x}(i)$, take a derivative w.r.t. it and then evaluate it at the point $\boldsymbol{x}$.

Now, prove that the system of linear equations $\boldsymbol{L}\boldsymbol{x} = \boldsymbol{d}$ is the same as the system obtained by setting the gradient of the function $c(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{d}$ equal to zero.

Solution.

The gradient of a function $f : S \to \mathbb{R}$ at point $\boldsymbol{x} \in S$ is denoted by $\boldsymbol{\nabla} f(\boldsymbol{x})$ is

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}(1)}, \dots, \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}(n)}\right]^\top.$$

We have

$$\boldsymbol{\nabla} c(\boldsymbol{x}) = \boldsymbol{\nabla}\left(\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{d}\right) = \boldsymbol{L}\boldsymbol{x} - \boldsymbol{d}.$$

Therefore, the system obtained by setting the gradient with respect to $\boldsymbol{x}$ of the function $c(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{d}$ equal to zero is the same as $\boldsymbol{L}\boldsymbol{x} = \boldsymbol{d}$.

Exercise 5

The goal of this exercise is to prove that

$$\max_{\boldsymbol{x} \in \mathbb{R}^V} \boldsymbol{x}^\top \boldsymbol{d} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \min_{\boldsymbol{f} \in \mathbb{R}^E} \frac{1}{2}\sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2 \quad \text{s.t. } \boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}.$$

We'll break that down into a few steps.

Let $\boldsymbol{f} \in \mathbb{R}^E$ be an arbitrary flow that satisfies $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$, i.e. it routes the demand $\boldsymbol{d}$. Let $\boldsymbol{x} \in \mathbb{R}^V$ be arbitrary voltages. *Arbitrary* means you cannot assume these are the electrical flow and voltages.

(i) Prove that

$$\frac{1}{2}\sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2 = \boldsymbol{x}^\top \boldsymbol{d} - \left(\sum_{(u,v)\in E} (\boldsymbol{x}(u) - \boldsymbol{x}(v))(\boldsymbol{f}(u,v)) - \frac{1}{2}\boldsymbol{r}(u,v)\boldsymbol{f}(u,v)^2\right)$$

Hint: use that $\boldsymbol{x}^\top(\boldsymbol{B}\boldsymbol{f} - \boldsymbol{d}) = 0$.

(ii) Prove that

$$(\boldsymbol{x}(u) - \boldsymbol{x}(v))(\boldsymbol{f}(u,v)) - \frac{1}{2}\boldsymbol{r}(u,v)\boldsymbol{f}(u,v)^2 \le \frac{1}{2}\frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))^2}{\boldsymbol{r}(u,v)}.$$

[1]We will give a more formal definition of Frechet derivatives later, which is formally what we mean by 'gradient'.

(iii) Conclude that $\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} \geq \boldsymbol{x}^\top \boldsymbol{d} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}$.

(iv) Assume we are given $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{f}}$ such that

$$\boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{d} \text{ and } \tilde{\boldsymbol{f}} = \boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}$$

Prove that $\boldsymbol{B}\tilde{\boldsymbol{f}} = \boldsymbol{d}$ and

$$\tilde{\boldsymbol{x}}^\top \boldsymbol{d} - \frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} = \frac{1}{2}\tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}}.$$

(v) Show

$$\tilde{\boldsymbol{x}} \in \arg\max_{\boldsymbol{x}\in\mathbb{R}^V} \boldsymbol{x}^\top \boldsymbol{d} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}$$

and

$$\tilde{\boldsymbol{f}} \in \arg\min_{\boldsymbol{f}\in\mathbb{R}^E} \frac{1}{2}\sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2 \quad \text{s.t. } \boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}.$$

Solution.

(i) Since $\boldsymbol{f} \in \mathbb{R}^E$ satisfies $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$, for $\boldsymbol{x} \in \mathbb{R}^V$, we have that

$$\begin{aligned}
& \boldsymbol{x}^\top(\boldsymbol{B}\boldsymbol{f} - \boldsymbol{d}) = 0 \\
\Rightarrow\quad & \boldsymbol{x}^\top \boldsymbol{d} = \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{f} = (\boldsymbol{B}^\top \boldsymbol{x})\boldsymbol{f} = \sum_{(u,v)\in E} (\boldsymbol{x}(u) - \boldsymbol{x}(v))\boldsymbol{f}(u,v) \\
\Rightarrow\quad & 0 = \boldsymbol{x}^\top \boldsymbol{d} - \sum_{(u,v)\in E} (\boldsymbol{x}(u) - \boldsymbol{x}(v))\boldsymbol{f}(u,v) \\
\Rightarrow\quad & \frac{1}{2}\sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2 = \boldsymbol{x}^\top \boldsymbol{d} - \sum_{(u,v)\in E} (\boldsymbol{x}(u) - \boldsymbol{x}(v))\boldsymbol{f}(u,v) + \frac{1}{2}\sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2 \\
\Rightarrow\quad & \frac{1}{2}\sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2 = \boldsymbol{x}^\top \boldsymbol{d} - \left(\sum_{(u,v)\in E} (\boldsymbol{x}(u) - \boldsymbol{x}(v))(\boldsymbol{f}(u,v)) - \frac{1}{2}\boldsymbol{r}(u,v)\boldsymbol{f}(u,v)^2\right).
\end{aligned}$$

(ii) Denote $h : \mathbb{R} \to \mathbb{R}$ is a function of $\boldsymbol{f}(u,v)$:

$$h(\boldsymbol{f}(u,v)) = (\boldsymbol{x}(u) - \boldsymbol{x}(v))(\boldsymbol{f}(u,v)) - \frac{1}{2}\boldsymbol{r}(u,v)\boldsymbol{f}(u,v)^2.$$

It is noticed that h is quadratic function of $\boldsymbol{f}(u,v)$, thus its maximum is attained when its derivative h' is equal to zero:

$$h'(\boldsymbol{f}^*(u,v)) = \boldsymbol{x}(u) - \boldsymbol{x}(v) - \boldsymbol{r}(u,v)\boldsymbol{f}^*(u,v) = 0 \;\Rightarrow\; \boldsymbol{f}^*(u,v) = \frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))}{\boldsymbol{r}(u,v)}.$$

Hence

$$h(\boldsymbol{f}^*(u,v)) = \frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))^2}{\boldsymbol{r}(u,v)} - \frac{1}{2}\frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))^2}{\boldsymbol{r}(u,v)} = \frac{1}{2}\frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))^2}{\boldsymbol{r}(u,v)}.$$

Therefore, for $\boldsymbol{f}$ being an arbitrary flow s.t. $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$ and $\boldsymbol{x}$ being an arbitrary voltages, we have

$$(\boldsymbol{x}(u) - \boldsymbol{x}(v))(\boldsymbol{f}(u,v)) - \frac{1}{2}\boldsymbol{r}(u,v)\boldsymbol{f}(u,v)^2 \leq \frac{1}{2}\frac{(\boldsymbol{x}(u) - \boldsymbol{x}(v))^2}{\boldsymbol{r}(u,v)}.$$

(iii) In Exercise 2, part (ii), we have proved that

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \sum_{(u,v)\in E} \frac{(\boldsymbol{x}(u)-\boldsymbol{x}(v))^2}{\boldsymbol{r}(u,v)}.$$

Combining it with the previous two questions in this exercise, we have

$$\sum_{(u,v)\in E} (\boldsymbol{x}(u)-\boldsymbol{x}(v))(\boldsymbol{f}(u,v)) - \frac{1}{2}\boldsymbol{r}(u,v)\boldsymbol{f}(u,v)^2 \le \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x},$$

thus,

$$\frac{1}{2}\sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2 = \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} \ge \boldsymbol{x}^\top \boldsymbol{d} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}.$$

(iv) To prove $\boldsymbol{B}\tilde{\boldsymbol{f}} = \boldsymbol{d}$, we have

$$\boldsymbol{B}\tilde{\boldsymbol{f}} = \boldsymbol{B}(\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}) = (\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top)\tilde{\boldsymbol{x}} = \boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{d}.$$

We also have

$$\begin{aligned}
&\tilde{\boldsymbol{x}}^\top \boldsymbol{d} - \frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} \\
=&\tilde{\boldsymbol{x}}^\top (\boldsymbol{L}\tilde{\boldsymbol{x}}) - \frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} \\
=&\frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} \\
=&\frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}} && \boldsymbol{L} = \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top \\
=&\frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{R}\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}} && \boldsymbol{R}\boldsymbol{R}^{-1} = \boldsymbol{I} \\
=&(\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}})^\top \boldsymbol{R}(\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}) && \boldsymbol{R} \text{ is diagonal} \\
=&\frac{1}{2}\tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}}
\end{aligned}$$

(v) According to optimality condition, maximum is attained when

$$\boldsymbol{\nabla}(\tilde{\boldsymbol{x}}^\top \boldsymbol{d} - \frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}}) = \boldsymbol{d} - \boldsymbol{L}\tilde{\boldsymbol{x}} = 0,$$

thus,

$$\tilde{\boldsymbol{x}} \in \arg\max_{\boldsymbol{x}\in\mathbb{R}^V} \boldsymbol{x}^\top \boldsymbol{d} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x},$$

where $\boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{d}$.

To prove the optimality of $\tilde{\boldsymbol{f}}$, we know from part (iii) that for any $\boldsymbol{f}$ s.t. $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$ and any $\boldsymbol{x}$, we have

$$\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} \ge \boldsymbol{x}^\top - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x},$$

thus, for any $\boldsymbol{f}$ s.t. $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$, we must have

$$\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} \ge \tilde{\boldsymbol{x}}^\top - \frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} = \frac{1}{2}\tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}}.$$

That is

$$\tilde{\boldsymbol{f}} \in \arg\min_{\boldsymbol{f} \in \mathbb{R}^E} \frac{1}{2} \sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2$$
$$\text{s.t. } \boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}.$$

Exercise 6

Recall that the following theorem gives us a sufficient (though not necessary) condition for optimality.

Theorem (Extreme Value Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and $\mathcal{F} \subseteq \mathbb{R}^n$ be nonempty, bounded, and closed. Then, the optimization problem* $\min f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{F}$ *has an optimal solution.*

Prove the above theorem. You might use the following two theorems.

Theorem (Bolzano-Weierstrass)**.** *Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*

Theorem (Boundedness Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and $\mathcal{F} \subseteq \mathbb{R}^n$ be nonempty, bounded, and closed. Then f is bounded on $\mathcal{F}$.*

Solution.

Let α be the infimum of f over $\mathcal{F}$ (i.e. the largest value for which any point $\boldsymbol{x} \in \mathcal{F}$ respects $f(\boldsymbol{x}) \geq \alpha$); by the Boundedness Theorem, such a value exists, as f is lower-bounded, and the set of lower bounds has a greatest lower bound, α.

Let

$$\mathcal{F}_k := \{\boldsymbol{x} \in \mathcal{F} : \alpha \leq f(\boldsymbol{x}) \leq \alpha + 2^{-k}\}.$$

$\mathcal{F}_k$ cannot be empty, since if it were, then $\alpha + 2^{-k}$ would be a strictly greater lower bound on f than α. For each k, let $\boldsymbol{x}_k$ be some $\boldsymbol{x} \in \mathcal{F}_k$. $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ is a bounded sequence as $\mathcal{F}_k \subseteq \mathcal{F}$, so the Bolzano-Weierstrass theorem we know that there is a convergent subsequence, $\{\boldsymbol{y}_k\}_{k=1}^{\infty}$, with limit $\bar{\boldsymbol{y}}$. Because the set is closed, $\bar{\boldsymbol{y}} \in \mathcal{F}$. By continuity $f(\bar{\boldsymbol{y}}) = \lim_{k\to\infty} f(\boldsymbol{y}_k)$, while by construction, $\lim_{k\to\infty} f(\boldsymbol{y}_k) = \alpha$.

Thus, the optimal solution is $\bar{\boldsymbol{y}}$.

Exercise 7

Prove or sketch a proof of Taylor's Theorem.

Theorem (Taylor's Theorem, multivariate first-order remainder form)**.** *If $f : S \to \mathbb{R}$ is continuously differentiable over $[\boldsymbol{x}, \boldsymbol{y}]$, then for some $\boldsymbol{z} \in [\boldsymbol{x}, \boldsymbol{y}]$, we have $f(\boldsymbol{y}) = f(\boldsymbol{x}) + \boldsymbol{\nabla} f(\boldsymbol{z})^\top (\boldsymbol{y} - \boldsymbol{x})$.*

Solution.

First, we prove the univariate case, which can be represented as

$$f(y) = f(x) + f'(z)(y - x).$$

We construct a new function

$$g(w) = \frac{f(y) - f(x)}{y - x}(w - x) + f(x) - f(w),$$

which can be proved to be continuously differentiable over $[x, y]$, and $g(x) = g(y)$. Then according to Rolle's theorem, $\exists z \in (x, y)$ such that

$$g'(z) = \frac{f(y) - f(x)}{y - x} - f'(z) = 0.$$

Rearranging the equality and we can get

$$f(y) = f(x) + f'(z)(y - x).$$

Now, we prove multivariate case, which is expresses as

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{y} - \boldsymbol{x}).$$

We play a trick to transform multivariate case to univariable case. Set $\phi(t) = f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$. It is noticed that $\phi(0) = f(\boldsymbol{x})$, $\phi(1) = f(\boldsymbol{y})$. According to the Chain Rule, we have

$$\phi'(t) = \nabla f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))^T(\boldsymbol{y} - \boldsymbol{x}).$$

In particular, ϕ is differentiable. According to the conclusion made in the univariate case, $\exists \tilde{t} \in (0, 1)$ such that

$$\phi(1) = \phi(0) + \phi'(\tilde{t})(1 - 0),$$

or equivalently,

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{y} - \boldsymbol{x}),$$

where $\boldsymbol{z} = \boldsymbol{x} + \tilde{t}(\boldsymbol{y} - \boldsymbol{x})$.

Exercise 8

Let $f_1(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x})$ be a collection of convex functions all with the same domain and define $f(\boldsymbol{x}) \overset{\text{def}}{=} \max_{1 \le i \le k} f_i(\boldsymbol{x})$. Prove that $f(\boldsymbol{x})$ is convex.

Solution.

Notice that the domain of f is trivially convex. Consider arbitrary $\boldsymbol{x}, \boldsymbol{y}$ in the domain of f and let θ be in $[0, 1]$. Then by applying the fact that $f_i(\boldsymbol{x})$ is convex for any $i \in [k]$, we get

$$\begin{aligned} f(\theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y}) &= \max_{1 \le i \le k} f_i(\theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y}) \\ &\le \max_{1 \le i \le k} \theta f_i(\boldsymbol{x}) + (1 - \theta) f_i(\boldsymbol{y}) \\ &\le \theta \max_{1 \le i \le k} f_i(\boldsymbol{x}) + (1 - \theta) \max_{1 \le i \le k} f_i(\boldsymbol{y}) \\ &= \theta f(\boldsymbol{x}) + (1 - \theta) f(\boldsymbol{y}). \end{aligned}$$

This implies that $f(\boldsymbol{x})$ is convex.

Exercise 9

Assume that $f(x, y)$ is a convex function and S is a convex non-empty set. Prove that

$$g(x) = \inf_{y \in S} f(x, y)$$

is convex, provided $g(x) > -\infty$ for all x.

Solution.

Note that the domain of g is the set

$$\mathrm{dom}(g) = \{x : (x, y) \in \mathrm{dom}(f) \text{ for some } y \in S\}.$$

Let us first prove that $\mathrm{dom}(g)$ is a convex set. Consider $x, x' \in \mathrm{dom}(g)$ and an arbitrary $\theta \in [0, 1]$. Then, there exist $y, y' \in S$ such that (x, y) and (x', y') are in the domain of f. Since $\mathrm{dom}(f)$ is convex, $\theta(x, y) + (1 - \theta)(x', y')$ is in $\mathrm{dom}(f)$. Furthermore, we have that $\theta y + (1 - \theta)y' \in S$ because S is a convex set. Therefore, $\theta x + (1 - \theta)x'$ is in $\mathrm{dom}(g)$.

Now, consider $x, x' \in \mathrm{dom}(g)$ and $\theta \in [0, 1]$. Assume that

$$y = \arg\inf_{y \in S} f(x, y) \quad \text{and} \quad y' = \arg\inf_{y \in S} f(x', y).$$

We have that

$$\begin{aligned}
g(\theta x + (1 - \theta)x') &= \inf_{y \in S} f(\theta x + (1 - \theta)x', y) \\
&\leq f(\theta x + (1 - \theta)x', \theta y + (1 - \theta)y') \\
&= f(\theta(x, y) + (1 - \theta)(x', y')) \\
&\leq \theta f(x, y) + (1 - \theta)f(x', y') \\
&= \theta g(x) + (1 - \theta)g(x').
\end{aligned}$$

This finishes the proof.

Exercise 10

For each function below, determine whether it is convex or not.

1. $f(x) = |x|^6$ on $x \in \mathbb{R}$
2. $f(x) = \exp(x)$ on $x \in (0, \infty)$
3. $f(x, y) = \sqrt{x + y}$ on $(x, y) \in (0, 1) \times (0, 1)$
4. $f(x, y) = xy$ on $(x, y) \in (-1, 1) \times (-1, 1)$

Solution.

1. We have $f(x) = |x|^6 = x^6$ on $x \in \mathbb{R}$. It suffices to show that for any $x, y \in \mathbb{R}$
$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$
which is equivalent to
$$y^6 \geq x^6 + 6x^5(y - x) \Leftrightarrow y^6 - 6x^5 y + 5x^6 \geq 0.$$
This is trivially true for $x = 0$. Thus, assume that $x \neq 0$ and let $z = y/x$. Then, it suffices to show that
$$h(z) = z^6 - 6z + 5 \geq 0.$$
We observe that $\nabla h(z) = 6z^5 - 6$. Thus, $h(z)$ reaches its minimum at $z = 1$, which is equal to 0. Therefore $h(z) \geq 0$ for any z.

2. We have $f(x) = \exp(x)$ on $x \in (0, \infty)$. It suffices to show that for any $x, y \in (0, \infty)$
$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$
which is equivalent to
$$\exp(y) \geq \exp(x) + \exp(x)(y - x) \Leftrightarrow \exp(y - x) \geq 1 + (y - x).$$
This inequality holds because we have
$$\exp(z) \geq 1 + z.$$
When proving this, we'll take for granted that
   (a) $\exp(z) = \sum_{i=0}^{\infty} \frac{z^i}{i!}$
   (b) $\exp(z) < 1$ for $z < 0$, $\exp(z) = 1$ for $z = 0$, and $\exp(z) > 1$ for $z > 0$.

   Let $f(z) = \exp(z) - (1 + z) = \sum_{i=2}^{\infty} \frac{z^i}{i!}$. Note $f(0) = 0$, and $f'(z) = \sum_{i=2}^{\infty} \frac{z^{i-1}}{(i-1)!} = \exp(z) - 1$, so $f'(z) > 0$ for $z > 0$ and $f'(z) < 0$ for $z < 0$. Thus $z = 0$ is the minimizer of f, and so for all z we have $f(z) \geq 0$, i.e. $\exp(z) \geq 1 + z$.

3. Consider the function $f(x, y) = \sqrt{x + y}$ on $(x, y) \in (0, 1) \times (0, 1)$. We have
$$\nabla f(x, y) = \left[\frac{1}{2\sqrt{x + y}}, \frac{1}{2\sqrt{x + y}} \right].$$
For $f(x, y)$ to be convex it is necessary that for any $\boldsymbol{a}, \boldsymbol{b} \in (0, 1) \times (0, 1)$
$$f(\boldsymbol{b}) \geq f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})^\top (\boldsymbol{b} - \boldsymbol{a}).$$
Set $\boldsymbol{a} = (1/4, 0)$ and $\boldsymbol{b} = (1/2, 0)$. Then, $f(\boldsymbol{b}) = 1/\sqrt{2}$ and $f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})^\top (\boldsymbol{b} - \boldsymbol{a}) = 3/4$. Since $1/\sqrt{2} < 3/4$, we can conclude that $f(x, y)$ is not convex.

4. Consider function $f(x, y) = xy$ on $(x, y) \in (-1, 1) \times (-1, 1)$. Set $\boldsymbol{a} = (-1/4, 1/4)$, $\boldsymbol{b} = (1/2, -1/2)$, and $\theta = 1/2$. Then, we have
$$f(\theta \boldsymbol{a} + (1 - \theta)\boldsymbol{b}) = -\frac{1}{64} > -\frac{10}{64} = -\frac{1}{32} - \frac{1}{8} = \theta f(\boldsymbol{a}) + (1 - \theta) f(\boldsymbol{b}).$$
This implies that $f(x, y)$ is not convex.

Remark. *It is often much easier to show convexity by directly showing the Hessian of a twice differentiable function is positive semi-definite. In following lectures, we will learn about this approach.*

Bonus Exercise 11: Jensen's Inequality

This exercise will teach you about Jensen's inequality, one of the most important inequalities that we use when studying convex functions.

1. Assume that $S \subseteq \mathbb{R}^n$ is a convex set and that the function $f : S \to \mathbb{R}$ is convex. Suppose that $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in S$ and $\theta_1, \cdots, \theta_n \geq 0$ with $\theta_1 + \cdots + \theta_n = 1$. Prove that

$$f(\theta_1 \boldsymbol{x}_1 + \cdots + \theta_n \boldsymbol{x}_n) \leq \theta_1 f(\boldsymbol{x}_1) + \cdots + \theta_n f(\boldsymbol{x}_n).$$

   Remark. *This is typically known as Jensen's inequality and can be extended to infinite sums. If $\mathcal{D}$ is a probability distribution on S, and $\boldsymbol{X} \sim \mathcal{D}$, then $f(\mathbb{E}[\boldsymbol{X}]) \leq \mathbb{E}[f(\boldsymbol{X})]$ whenever both integrals are finite.*

2. Prove that $\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^n x_i$.

3. Prove that $\frac{1}{\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}} \leq \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}$.

Solution.

1. We prove this statement by applying induction on n. As the base case, the statement is true for $n = 2$ because f is convex. As the induction hypothesis, assume that for any $n \geq 2$

$$f\left(\theta_1 \boldsymbol{x}_1 + \cdots + \theta_n \boldsymbol{x}_n\right) \leq \theta_1 f\left(\boldsymbol{x}_1\right) + \cdots + \theta_n f\left(\boldsymbol{x}_n\right).$$

   for any choice of $\theta_1, \cdots, \theta_n \geq 0$ such that $\theta_1 + \cdots + \theta_n = 1$.

   Now, we show that this also holds for $n+1$. We can assume without loss of generality that $\theta_{n+1} \in (0, 1)$. We have

$$\sum_{i=1}^{n+1} \theta_i f\left(\boldsymbol{x}_i\right) = (1 - \theta_{n+1}) \sum_{i=1}^{n} \frac{\theta_i}{(1 - \theta_{n+1})} f\left(\boldsymbol{x}_i\right) + \theta_{n+1} f\left(\boldsymbol{x}_{n+1}\right).$$

   Note that $\sum_{i=1}^n \frac{\theta_i}{(1-\theta_{n+1})}$ is equal to 1. Thus, by applying the induction hypothesis, we get

$$\sum_{i=1}^{n+1} \theta_i f\left(\boldsymbol{x}_i\right) \geq (1 - \theta_{n+1}) f\left(\sum_{i=1}^{n} \frac{\theta_i}{(1 - \theta_{n+1})} \boldsymbol{x}_i\right) + \theta_{n+1} f\left(\boldsymbol{x}_{n+1}\right).$$

   Observe that $\sum_{i=1}^n \frac{\theta_i}{(1-\theta_{n+1})} \boldsymbol{x}_i \in S$ since it is a convex combination of points in S. Again, by using the convexity of f, we have

$$\sum_{i=1}^{n+1} \theta_i f\left(\boldsymbol{x}_i\right) \geq f\left((1 - \theta_{n+1}) \sum_{i=1}^{n} \frac{\theta_i}{(1 - \theta_{n+1})} \boldsymbol{x}_i + \theta_{n+1} \boldsymbol{x}_{n+1}\right) = f\left(\sum_{i=1}^{n+1} \theta_i \boldsymbol{x}_i\right).$$

2. We observe that

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^{n} x_i.$$

is equivalent to

$$\log\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \leq \log\frac{1}{n}\sum_{i=1}^{n} x_i.$$

This is the same as

$$\sum_{i=1}^{n} -\frac{1}{n}\log x_i \geq -\log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right).$$

which is true by using the previous part and the fact that function $f(x) = -\log x$ is convex for $x \in (0, \infty)$.

3. By the previous part, we know that

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^{n} x_i$$

which, by applying the inquality to values $1/x_i$ instead of x_i yields

$$\left(\prod_{i=1}^{n} \frac{1}{x_i}\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}.$$

Dividing both sides by $\left(\prod_{i=1}^{n} \frac{1}{x_i}\right)^{\frac{1}{n}} \cdot \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}$ gives

$$\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \leq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}.$$