Advanced Graph Algorithms and Optimization Spring 2023

Spectral Graph Theory

R. Kyng & M. Probst Gutenberg *Problem Set 6— Wednesday, March 29nd*

The exercises for this week will not count toward your grade, but you are highly encouraged to solve them all. This exercise sheet has exercises related to week 6. We encourage you to start early so you have time to go through everything.

To get feedback, you must hand in your solutions by 23:59 on April 7. Both hand-written and LaTeX solutions are acceptable, but we will only attempt to read legible text.

Notation

Througout the following exercises, we will use the following notation:

- S^n is the set of symmetric real matrices $n \times n$ matrices.
- S^n_+ is the set of positive semi-definite $n \times n$ matrices.
- S^n_{++} is the set of positive definite $n \times n$ matrices.

Whenever we say a matrix is positive semi-definite or positive definite, we require it to be real and symmetric.

Exercise 1

In this exercise, we want you to complete the proof of Theorem 9.3.3 in Chapter 9. Refer to the lectures notes for definitions of the terms used here.

1. Prove that Equation (9.4) is satisfied, i.e. that for all edges $e \in E$ we have $\|\boldsymbol{X}_e\| \leq \frac{1}{\alpha}$.
2. Prove that Equation (9.5) is satisfied, i.e. that $\left\|\sum_e \mathbb{E}\left[\boldsymbol{X}_e^2\right]\right\| \leq \frac{1}{\alpha}$.
3. Explain how we can use a scalar Chernoff bound to prove that $\left|\tilde{E}\right| \leq O(\epsilon^{-2}\log(n/\delta)n)$ with probability at least $1-\delta/2$. You may pick any constant that suits you to establish the $O(\cdot)$ bound.

Solution.

1. We recall that

$$\boldsymbol{Y}_e = \begin{cases} \frac{\boldsymbol{w}(e)}{p_e}\boldsymbol{b}_e\boldsymbol{b}_e^\top & \text{with probability } p_e \\ \boldsymbol{0} & \text{otherwise} \end{cases}$$

for $p_e = \min(1, \alpha \left\|\Phi(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top)\right\|)$ and $\Phi(\boldsymbol{A}) = \boldsymbol{L}^{+/2}\boldsymbol{A}\boldsymbol{L}^{+/2}$. Furthermore, we have $\boldsymbol{X}_e = \Phi(\boldsymbol{Y}_e) - \mathbb{E}\left[\Phi(\boldsymbol{Y}_e)\right]$. Thus,

$$\boldsymbol{X}_e = \begin{cases} (\frac{1}{p_e} - 1)\Phi(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top) & \text{with probability } p_e \\ -\Phi(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top) & \text{otherwise.} \end{cases}$$

Note that when $p_e = 1$, we have $\boldsymbol{X}_e = \boldsymbol{0}$ always. So we only need to bound the norm when $p_e = \alpha \left\|\Phi(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top)\right\|$. Therefore, we can compute

$$\|\boldsymbol{X}_e\| \le \max\{\frac{1}{p_e} - 1, 1\} \left\|\Phi(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top)\right\| \le \frac{1}{p_e} \left\|\Phi(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top)\right\| = \frac{1}{\alpha}.$$

2. Note that $\sum_e \mathbb{E}\left[\boldsymbol{X}_e^2\right] \succeq 0$, because each term is PSD. So we only need to give an upper bound to bound the norm. Let us upper-bound $\mathbb{E}\left[\boldsymbol{X}_e^2\right]$. We can focus on $p_e = \alpha \left\|\Phi(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top)\right\|$, because in the other case $\boldsymbol{X}_e$ is identically zero. We have that

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{X}_e^2\right] &= \left(\frac{1}{p_e} - 1\right) \Phi\left(\boldsymbol{w}\left(e\right) \boldsymbol{b}_e\boldsymbol{b}_e^\top\right)^2 \\ &\preceq \left(\frac{1}{p_e} - 1\right) \left\|\Phi\left(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top\right)\right\| \Phi\left(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top\right) \\ &\preceq \frac{1}{\alpha}\Phi\left(\boldsymbol{w}(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top\right). \end{aligned}$$

Using the fact that $\|\Phi(\boldsymbol{L})\| \le 1$, we can conclude

$$\left\|\sum_e \mathbb{E}\left[\boldsymbol{X}_e^2\right]\right\| \le \left\|\frac{1}{\alpha}\Phi(\boldsymbol{L})\right\| \le \frac{1}{\alpha}.$$

3. For each edge e, let Bernoulli random variable Z_e be 1 with probability p_e. Then, $\left|\tilde{E}\right| = \sum_e Z_e$ is the sum of $|E|$ independent Bernoulli random variables. Let us define $\mu' = 40\epsilon^{-2}n\log(n/\delta)$. We know that $\mathbb{E}\left[\left|\tilde{E}\right|\right] \le \mu'$. Therefore, by applying the Chernoff bound we get

$$\begin{aligned} \Pr\left[\left|\tilde{E}\right| \ge 2\mu'\right] &\le Pr\left[\left|\tilde{E}\right| \ge 2\,\mathbb{E}\left[\left|\tilde{E}\right|\right]\right] \\ &\le \exp\left(-\frac{\mathbb{E}\left[\left|\tilde{E}\right|\right]}{4}\right) \\ &\le \exp\left(-\frac{c\epsilon^{-2}n\log(n/\delta)}{4}\right) \\ &= \left(\frac{\delta}{n}\right)^{\frac{c\epsilon^{-2}n}{4}} \\ &\le \frac{\delta}{2} \end{aligned}$$

where we used that we can lower bound $\mathbb{E}\left[\left|\tilde{E}\right|\right]$ by $c\epsilon^{-2}n\log(n/\delta)$ for some constant $c > 0$.

Exercise 2

Consider $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{v}, \boldsymbol{u} \in \mathbb{R}^n$.

1. Assume that $\boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^\top$ is invertible. Determine c such that
$$\left(\boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^\top\right)^{-1} = \boldsymbol{I} - \frac{\boldsymbol{u}\boldsymbol{v}^\top}{c}.$$

2. Assume that both $\boldsymbol{A}$ and $\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^\top$ are invertible. Prove that
$$\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^\top\right)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u}}.$$
Hint: You might use that $(\boldsymbol{B}\boldsymbol{C})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}$ for two invertible matrices $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{n\times n}$.

Solution.

1. We want to find c such that
$$\boldsymbol{I} = \left(\boldsymbol{I} - \frac{\boldsymbol{u}\boldsymbol{v}^\top}{c}\right)\left(\boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^\top\right).$$
This is equivalent to
$$\boldsymbol{I} = \boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^\top - \frac{\boldsymbol{u}\boldsymbol{v}^\top}{c} - \frac{\boldsymbol{u}\boldsymbol{v}^\top}{c}\boldsymbol{u}\boldsymbol{v}^\top.$$
Notice that $\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{u}\boldsymbol{v}^\top = \boldsymbol{u}(\boldsymbol{v}^\top\boldsymbol{u})\boldsymbol{v}^\top = (\boldsymbol{v}^\top\boldsymbol{u})\boldsymbol{u}\boldsymbol{v}^\top$. Thus, the above equation holds if and only if
$$1 - \frac{1}{c} - \frac{\boldsymbol{v}^\top\boldsymbol{u}}{c} = 0.$$
Therefore, we set $c = 1 + \boldsymbol{v}^\top\boldsymbol{u}$ and conclude hat
$$\left(\boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^\top\right)^{-1} = \boldsymbol{I} - \frac{\boldsymbol{u}\boldsymbol{v}^\top}{1 + \boldsymbol{v}^\top\boldsymbol{u}}. \tag{1}$$

2. We have that
$$\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^\top\right)^{-1} = \left(\boldsymbol{A}\left(\boldsymbol{I} + \boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\right)\right)^{-1}.$$
By applying the fact that $(\boldsymbol{B}\boldsymbol{C})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}$ for two invertible matrices $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{n\times n}$ and then using Equation (1), we get
$$\begin{aligned}\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^\top\right)^{-1} &= \left(\boldsymbol{I} + \boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\right)^{-1}\boldsymbol{A}^{-1} \\ &= \left(\boldsymbol{I} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top}{1 + \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u}}\right)\boldsymbol{A}^{-1} \\ &= \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u}}.\end{aligned}$$
This equality is known as the *Sherman-Morrison* formula.

Exercise 3

Consider a matrix function $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$. For $\boldsymbol{X}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times n}$, we define

$$Df\left(\boldsymbol{X}\right)\left[\boldsymbol{Y}\right] = \frac{\partial}{\partial t}\Big|_{t=0} f\left(\boldsymbol{X} + t\,\boldsymbol{Y}\right).$$

Remark. *Note that if we think of $\boldsymbol{X}$ and $\boldsymbol{Y}$ each as a vector of numbers, then this is the (matrix-valued) directional derivative of f at $\boldsymbol{X}$ in the direction of $\boldsymbol{Y}$.*

Consider $f(\boldsymbol{X}) = \boldsymbol{X}^{-1}$ for an invertible matrix $\boldsymbol{X} \in \mathbb{R}^{n\times n}$. Prove that

$$Df(\boldsymbol{X})[\boldsymbol{Y}] = -\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{X}^{-1}.$$

Hint: You might need to use Exercise 3.

Solution.

We can write $\boldsymbol{Y} = \sum_{i,j} \boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top$, where $\boldsymbol{e}_i$ denotes the vector with a 1 in the i-th coordinate and 0's elsewhere. Thus, we have that

$$Df\left(\boldsymbol{X}\right)\left[\boldsymbol{Y}\right] = \sum_{i,j} Df\left(\boldsymbol{X}\right)\left[\boldsymbol{Y}\left(i,j\right)\boldsymbol{e}_i\boldsymbol{e}_j^\top\right].$$

Therefore, if we prove that $Df(\boldsymbol{X})[\boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top] = -\boldsymbol{X}^{-1}\boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top\boldsymbol{X}^{-1}$ then we can conclude that $Df(\boldsymbol{X})[\boldsymbol{Y}] = -\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{X}^{-1}$.

We know that

$$Df\left(\boldsymbol{X}\right)\left[\boldsymbol{Y}\left(i,j\right)\boldsymbol{e}_i\boldsymbol{e}_j^\top\right] = \lim_{t\to 0}\frac{\left(\boldsymbol{X} + t\,\boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top\right)^{-1} - \boldsymbol{X}^{-1}}{t}. \tag{2}$$

Furthermore, by applying Exercise 3 we get

$$\left(\boldsymbol{X} + t\,\boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top\right)^{-1} = \boldsymbol{X}^{-1} - \frac{\boldsymbol{X}^{-1}t\,\boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top\boldsymbol{X}^{-1}}{1 + t\,\boldsymbol{Y}(i,j)\boldsymbol{e}_j^\top\boldsymbol{X}^{-1}\boldsymbol{e}_i}. \tag{3}$$

Combining Equations (2) and (3) implies that

$$Df\left(\boldsymbol{X}\right)\left[\boldsymbol{Y}\left(i,j\right)\boldsymbol{e}_i\boldsymbol{e}_j^\top\right] = \lim_{t\to 0}\frac{-\boldsymbol{X}^{-1}\boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top\boldsymbol{X}^{-1}}{1 + t\,\boldsymbol{Y}(i,j)\boldsymbol{e}_j^\top\boldsymbol{X}^{-1}\boldsymbol{e}_i} = -\boldsymbol{X}^{-1}\boldsymbol{Y}(i,j)\boldsymbol{e}_i\boldsymbol{e}_j^\top\boldsymbol{X}^{-1}.$$

This finishes the proof.

Exercise 4

1. Consider $\boldsymbol{A} \in S_{++}^n$ and matrix $\boldsymbol{\Delta} \in S_+^n$. Prove that $(\boldsymbol{A} + \boldsymbol{\Delta})^{-1} \preceq \boldsymbol{A}^{-1}$.

2. Let T be a convex set. We say that a function $f : T \to \mathbb{R}^{n\times n}$, is operator convex if for any two matrices $\boldsymbol{A}, \boldsymbol{B} \in T$ and any $\theta \in [0, 1]$

$$f\left(\theta \boldsymbol{X} + (1-\theta)\ \boldsymbol{Y}\right) \preceq \theta f\left(\boldsymbol{X}\right) + (1-\theta)\, f\left(\boldsymbol{Y}\right).$$

Prove that $f(\boldsymbol{X}) = \boldsymbol{X}^{-1}$ is operator convex over the set $T = S^n_{++}$.

Hint: You could first show that operator convexity is implied by the second directional derivative $D^2 f(\boldsymbol{X})[\boldsymbol{Y}, \boldsymbol{Y}]$ *being positive semi-definite for all* $\boldsymbol{Y} \in S^n$ *and* $\boldsymbol{X} \in S^n_{++}$.

Solution.

1. From Exercise 10 in problem set 2, we know that if $\boldsymbol{A} \preceq \boldsymbol{B}$ for two matrices $\boldsymbol{A}, \boldsymbol{B} \in S^n_{++}$, then $\boldsymbol{B}^{-1} \preceq \boldsymbol{A}^{-1}$. By setting $\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{\Delta}$, we can conclude that $(\boldsymbol{A}+\boldsymbol{\Delta})^{-1} \preceq \boldsymbol{A}^{-1}$.

   Note that by setting $\boldsymbol{\Delta} = \boldsymbol{B} - \boldsymbol{A}$, we can prove the statement of Exercise 10 from Problem Set 2. Thus, it would be interesting to provide an alternative proof which does not use this exercise. We provide such an alternative proof.

   We know that
$$(\boldsymbol{A}+\boldsymbol{\Delta})^{-1} = \boldsymbol{A}^{-1} + \int_{t=0}^{1} \frac{d}{dt}\left(\boldsymbol{A}+t\boldsymbol{\Delta}\right)^{-1} dt.$$
   By applying Exercise 4, we get
$$(\boldsymbol{A}+\boldsymbol{\Delta})^{-1} = \boldsymbol{A}^{-1} + \int_{t=0}^{1} -\left(\boldsymbol{A}+t\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}\left(\boldsymbol{A}+t\boldsymbol{\Delta}\right)^{-1} dt.$$
   Consider an arbitrary vector $\boldsymbol{x}$, then we have that
$$\boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{\Delta})^{-1}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{A}^{-1}\boldsymbol{x} - \int_{t=0}^{1} \boldsymbol{x}^\top\left(\boldsymbol{A}+t\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}\left(\boldsymbol{A}+t\boldsymbol{\Delta}\right)^{-1}\boldsymbol{x}\ dt.$$
   We observe that $\boldsymbol{x}^\top\left(\boldsymbol{A}+t\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}\left(\boldsymbol{A}+t\boldsymbol{\Delta}\right)^{-1}\boldsymbol{x} \geq 0$. Therefore, we can conclude that
$$\boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{\Delta})^{-1}\boldsymbol{x} \leq \boldsymbol{x}^\top\boldsymbol{A}^{-1}\boldsymbol{x}$$
   which implies that $(\boldsymbol{A}+\boldsymbol{\Delta})^{-1} \preceq \boldsymbol{A}^{-1}$.

2. First, we will show that $D^2 f(\boldsymbol{X})[\boldsymbol{Y}, \boldsymbol{Y}]$ being positive semi-definite for all $\boldsymbol{Y} \in S^n$ and $\boldsymbol{X} \in S^n_{++}$ implies operator-convexity. Define for $t \in [0, 1]$ a function $h(t) = \boldsymbol{z}^\top f(\boldsymbol{X} + t(\boldsymbol{Y} - \boldsymbol{X}))\boldsymbol{z}$ for fixed $\boldsymbol{z} \in \mathbb{R}^n$ and $\boldsymbol{X}, \boldsymbol{Y} \in S^n_{++}$. This is a valid definition since $\boldsymbol{X} + t(\boldsymbol{Y} - \boldsymbol{X})$ is in S^n_{++} for $t \in [0, 1]$. Observe that:
$$\frac{\partial}{\partial t}h(t) = \boldsymbol{z}^\top Df(\boldsymbol{X}+t(\boldsymbol{Y}-\boldsymbol{X}))[\boldsymbol{Y}-\boldsymbol{X}]\boldsymbol{z}$$
$$\frac{\partial^2}{\partial t^2}h(t) = \boldsymbol{z}^\top D^2 f(\boldsymbol{X}+t(\boldsymbol{Y}-\boldsymbol{X}))[\boldsymbol{Y}-\boldsymbol{X}, \boldsymbol{Y}-\boldsymbol{X}]\boldsymbol{z} \geq 0$$
   The last inequality follows from $D^2 f(\boldsymbol{X}+t(\boldsymbol{Y}-\boldsymbol{X}))[\boldsymbol{Y}-\boldsymbol{X}, \boldsymbol{Y}-\boldsymbol{X}]$ being positive semi-definite by assumption. Therefore, we know that h is convex:
$$\boldsymbol{z}^\top\left[\theta f(\boldsymbol{X}) + (1-\theta)f(\boldsymbol{Y})\right]\boldsymbol{z} = \theta h(0) + (1-\theta)h(1) \geq h(1-\theta) = \boldsymbol{z}^\top f(\theta\boldsymbol{X} + (1-\theta)\boldsymbol{Y})\boldsymbol{z}$$

This inequality holds for any $\boldsymbol{z} \in \mathbb{R}^n$ which proves that f is operator convex.

What is left to show is that $D^2 f(\boldsymbol{X})[\boldsymbol{Y}, \boldsymbol{Y}]$ is indeed positive semi-definite. Using Exercise 4:

$$\begin{aligned}
D^2 f(\boldsymbol{X})[\boldsymbol{Y}, \boldsymbol{Y}] &= \frac{\partial^2}{\partial t^2}\Big|_{t=0} f(\boldsymbol{X} + t\,\boldsymbol{Y}) \\
&= \frac{\partial}{\partial t}\Big|_{t=0} - (\boldsymbol{X} - t\,\boldsymbol{Y})^{-1}\,\boldsymbol{Y}\,(\boldsymbol{X} - t\,\boldsymbol{Y})^{-1} \\
&= -\frac{\partial(\boldsymbol{X} - t\,\boldsymbol{Y})^{-1}}{\partial t}\Big|_{t=0} \boldsymbol{Y}\,\boldsymbol{X}^{-1} - \boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\frac{\partial(\boldsymbol{X} - t\,\boldsymbol{Y})^{-1}}{\partial t}\Big|_{t=0} \\
&= \boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1} + \boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1} \\
&= 2\boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1}
\end{aligned}$$

Remember that $\boldsymbol{X}$ being positive definite implies $\boldsymbol{X}^{-1}$ being positive definite. Thus, we have for any $\boldsymbol{x} \in \mathbb{R}^n$:

$$\boldsymbol{x}^\top D^2 f(\boldsymbol{X})[\boldsymbol{Y}, \boldsymbol{Y}]x = 2\boldsymbol{x}^\top \boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1}\,\boldsymbol{Y}\,\boldsymbol{X}^{-1}\boldsymbol{x} = 2\boldsymbol{z}^\top \boldsymbol{X}^{-1}\boldsymbol{z} \geq 0$$

where $\boldsymbol{z} = \boldsymbol{Y}\,\boldsymbol{X}^{-1}\boldsymbol{x}$. Hence, $D^2 f(\boldsymbol{X})[\boldsymbol{Y}, \boldsymbol{Y}]$ is positive semi-definite which completes the proof.