3M1 Examples Paper 1 Notes

February 12, 2020

1 Question 5

Why does a Hermitian positive semi-definite (PSD) matrix $\boldsymbol{A}$ have non-negative eigenvalues?

Consider any arbitrary $\mathbf{x} \in \mathbb{C}^n$. Assuming that $\boldsymbol{A}$ is diagonalizable, there is a set of eigenvectors that is linearly independent – call it $\mathbf{u}_1, ..., \mathbf{u}_n$ and normalize them to have unit 2-norm. We know that they are orthogonal to each other because $\boldsymbol{A}$ is Hermitian. We can expand $\mathbf{x}$ in the *eigenbasis*:

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + ... + \alpha_n \mathbf{u}_n. \tag{1}$$

Then, by hypothesis, $\mathbf{x}^H \boldsymbol{A} \mathbf{x} \geq 0$, so we have

$$\begin{aligned}
&\left(\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\right)^H \boldsymbol{A} \left(\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\right) \\
&= \left(\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\right)^H \left(\sum_{i=1}^{n} \lambda_i \alpha_i \mathbf{u}_i\right) \\
&= \sum_{i,j} \alpha_i \alpha_j \lambda_i \mathbf{u}_i^H \mathbf{u}_j \\
&= \sum_{i=1}^{n} \alpha_i^2 \lambda_i \geq 0.
\end{aligned}$$

We can set $\alpha_i = 1$ for each $i = 1, ..., n$ to get that $\lambda_i \geq 0$ for all i.

2 Question 6a

An alternative solution is proposed. Let $\mathbf{x}_1$ be an eigenvector of $\boldsymbol{A}$ with eigenvalue λ_1, and $\mathbf{x}_2$ be an eigenvector of $\boldsymbol{A}^T$ with eigenvalue λ_2, then

$$
\begin{aligned}
\boldsymbol{A}\mathbf{x}_1 &= \lambda_1 \mathbf{x}_1 \\
\boldsymbol{A}^T\mathbf{x}_2 &= \lambda_2 \mathbf{x}_2 \\
\implies \mathbf{x}_1^T\boldsymbol{A}^T\mathbf{x}_2 &= \lambda_2 \mathbf{x}_1^T\mathbf{x}_2 \\
\text{and} \qquad (\mathbf{x}_1^T\boldsymbol{A}^T\mathbf{x}_2)^T &= \mathbf{x}_2^T\boldsymbol{A}\mathbf{x}_1 \\
&= \lambda_1 \mathbf{x}_2^T\mathbf{x}_1.
\end{aligned} \tag{2}
$$

So if $\mathbf{x}_2^T\mathbf{x}_1 \neq 0$, then we have $\lambda_1 = \lambda_2$. I think this solution should be fine[1].

3 Question 12a

Showing the second equality of the displayed equation:

$$
\begin{aligned}
\mathbf{r}^H\boldsymbol{A}\mathbf{z} &= \mathbf{b}^H(\boldsymbol{A}(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H - \boldsymbol{I})\boldsymbol{A}\mathbf{z} \\
&= \mathbf{b}^H(\boldsymbol{A}\underbrace{(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{A}}_{\boldsymbol{I}} - \boldsymbol{A})\mathbf{z} \\
&= \mathbf{b}^H(\boldsymbol{A} - \boldsymbol{A})\mathbf{z} = 0.
\end{aligned}
$$

We emphasize again that $(\boldsymbol{A}^H\boldsymbol{A})^{-1} \neq \boldsymbol{A}^{-1}\boldsymbol{A}^{-H}$, since $\boldsymbol{A}$ is not square and is hence not invertible.

4 Question 13

A few clarifications are in order:

- Why is the rank of $\boldsymbol{A}$ necessarily equal to m? Firstly, $n \geq m$ by the pigeonhole principle (convince yourself this is true...), so the rank of $\boldsymbol{A}$, $r(\boldsymbol{A}) \leq m$. However, The rows are linearly independent (LI), so the rank is equal to m.
- Why does the column space span all of $\mathbb{C}^m$? If $n = m$, we must have all m of the columns be LI, because row rank = column rank, and thus the columns span all of $\mathbb{C}^m$. If $n > m$, we are adding more vectors to the column space, but the span will not be smaller. It stays as $\mathbb{C}^m$. *Exercise: The row space is NOT always* $\mathbb{C}^n$*! Why?*
- Recall that the *left null space* is defined as the set of vectors:

$$
\{\mathbf{x} \,|\, \boldsymbol{A}^T\mathbf{x} = 0\} \tag{3}
$$

[1]The small print: My concern with this solution is that for any given λ, if the *geometric multiplicity* of that eigenvalue is 1, then if $\mathbf{x}$ is an eigenvector then only vectors of the form $c\mathbf{x}$ are also eigenvectors of this eigenvalue. Therefore, in the specific case that both eigenvalues λ_1 and λ_2 have geometric multiplicity of 1, and $\mathbf{x}_1^T\mathbf{x}_2 = 0$, we cannot find another pair of eigenvectors that will allow us to dodge this zero, and we cannot show that $\lambda_1 = \lambda_2$ for this case.

It is the orthogonal complement to the column space, which means that any vector in the left null space dotted with a vector from the column space is zero. *Exercise: Prove this fact.* When the column space is all of $\mathbb{C}^m$, it must be that the left null space is just the zero vector – It is the only vector that, when dotted with *any* vector in $\mathbb{C}^m$, gives zero.

Please go through this question again and make sure you understand every step.

5 Question 16d

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. If we substitute $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$, we can compute $\boldsymbol{A}^+$ as

$$\begin{aligned}\boldsymbol{A}^+ &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T \qquad \text{because } \boldsymbol{A} \text{ is tall}\\ &= (\boldsymbol{V\Sigma}^T\boldsymbol{U}^T\boldsymbol{U\Sigma V}^T)^{-1}\boldsymbol{V\Sigma}^T\boldsymbol{U}^T\\ &= \boldsymbol{V\Sigma}_n^{-2}\boldsymbol{\Sigma}^T\boldsymbol{U}^T,\end{aligned}$$

where $\boldsymbol{\Sigma}_n \in \mathbb{R}^{n\times n}$ is a diagonal matrix with the singular values of $\boldsymbol{A}$ along its diagonal. It is different from $\boldsymbol{\Sigma}$ which is $\in \mathbb{R}^{m\times n}$. Now, that is equal to

$$\boldsymbol{V\Sigma}^+\boldsymbol{U}^T,$$

where $\boldsymbol{\Sigma}^+ \in \mathbb{R}^{n\times m}$. Then,

$$\begin{aligned}\boldsymbol{AA}^+ &= \boldsymbol{U\Sigma V}^T\boldsymbol{V\Sigma}^+\boldsymbol{U}^T\\ &= \boldsymbol{U\Sigma\Sigma}^+\boldsymbol{U}^T\\ &= \boldsymbol{U}_n\boldsymbol{U}_n^T.\end{aligned}$$

where $\boldsymbol{U}_n$ is the first n columns of $\boldsymbol{U} \in \mathbb{R}^{m\times m}$. It is no longer an orthogonal matrix because it is not full-rank.

As we remarked in the supervision, the psuedoinverse takes two forms whenever $\boldsymbol{A}$ is tall and whenever $\boldsymbol{A}$ is wide:

- $\boldsymbol{A}$ tall:
$$\boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$$
$$\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}$$
$$\boldsymbol{AA}^+ = \boldsymbol{P} \qquad \text{(Projector to column space)}$$

- $\boldsymbol{A}$ wide:
$$\boldsymbol{A}^+ = \boldsymbol{A}^T(\boldsymbol{AA}^T)^{-1}$$
$$\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{P} \qquad \text{(Projector to } \textit{row} \text{ space)}$$
$$\boldsymbol{AA}^+ = \boldsymbol{I}$$

Food for thought: 1. The formula for tall matrices cannot be applied to wide matrices (and vice versa). Why not? 2. Show that for a non-singular square matrix, the formulae are equivalent and reduce to $\boldsymbol{A}^{-1}$

6 The invertible matrix theorem

The *invertible matrix theorem* (link) lists 23 equivalent conditions for the invertibility of a square matrix. The extraordinary motivated student may try to go ahead to prove and convince themselves these are true!