Afonso Bandeira
Bernd Gärtner

Solution for Assignment 13

1. **a)** Let $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$ be an eigenvector of $A+B$ corresponding to eigenvalue $\lambda_{\text{min}}^{(A+B)}$. By using our knowledge about Rayleigh quotients (Proposition 6.3.10), we get

$$\lambda_{\text{min}}^{(A+B)} = \frac{\mathbf{x}^\top (A+B)\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} + \frac{\mathbf{x}^\top B\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \overset{6.3.10}{\geq} \lambda_{\text{min}}^{(A)} + \lambda_{\text{min}}^{(B)}.$$

b) Since both A and B are positive semidefinite, we have $\lambda_{\text{min}}^{(A)} \geq 0$ and $\lambda_{\text{min}}^{(B)} \geq 0$. Using our result from the previous subtask, we conclude that $\lambda_{\text{min}}^{(A+B)} \geq 0$. Hence, $A+B$ is positive semidefinite.

c) This is analogous to the proof in the previous subtask: since both A and B are positive definite, we have $\lambda_{\text{min}}^{(A)} > 0$ and $\lambda_{\text{min}}^{(B)} > 0$. Using our result from the subtask a), we conclude that $\lambda_{\text{min}}^{(A+B)} > 0$. Hence, $A+B$ is positive definite.

***Remark**: Note that we actually only need one of A and B to be positive definite, as long as the other one is still positive semidefinite.*

2. Consider first the $r \times n$ matrix $B = \Sigma_r V_r^\top$ with rank r. In particular, B has full row rank and hence

$$B^\dagger = B^\top (BB^\top)^{-1} = V_r \Sigma_r (\Sigma_r V_r^\top V_r \Sigma_r)^{-1} = V_r \Sigma_r (\Sigma_r^2)^{-1} = V_r \Sigma_r^{-1}$$

where we have used Definition 4.5.3, the fact that Σ_r is a diagonal matrix, and the fact that $V_r^\top V_r = I$.

Similarly, the $m \times r$ matrix U_r has full column rank r and hence we get

$$U_r^\dagger = (U_r^\top U_r)^{-1} U_r^\top = I U_r^\top = U_r^\top$$

by Definition 4.5.1 and the fact that $U_r^\top U_r = I$.

Finally, we conclude that

$$A^\dagger = B^\dagger U_r^\dagger = V_r \Sigma_r^{-1} U_r^\top$$

by Proposition 4.5.9.

3. **a)** The main idea is to plug in the SVD of A. A crucial observation that we will need is that by orthogonality of U, we have $\| U^\top \mathbf{v} \|_2^2 = (U^\top \mathbf{v})^\top (U^\top \mathbf{v}) = \mathbf{v}^\top U U^\top \mathbf{v} = \mathbf{v}^\top \mathbf{v} = \| \mathbf{v} \|_2^2$ for all $\mathbf{v} \in \mathbb{R}^m$. Equipped with this observation, we calculate

$$\begin{aligned}
\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_2^2 &= \min_{\mathbf{x} \in \mathbb{R}^n} \|U\Sigma V^\top \mathbf{x} - \mathbf{b}\|_2^2 \\
&= \min_{\mathbf{x} \in \mathbb{R}^n} \|U^\top U\Sigma V^\top \mathbf{x} - U^\top \mathbf{b}\|_2^2 \\
&= \min_{\mathbf{x} \in \mathbb{R}^n} \|\Sigma V^\top \mathbf{x} - U^\top \mathbf{b}\|_2^2 \\
&= \min_{\mathbf{y} \in \mathbb{R}^n} \|\Sigma \mathbf{y} - \mathbf{c}\|_2^2
\end{aligned}$$

where we have substituted $\mathbf{y} = V^\top \mathbf{x}$ in the end (which works because $V^\top$ is invertible).

b) Consider the expression $\|\Sigma \mathbf{y} - \mathbf{c}\|_2^2$ and observe that we can write it as

$$\|\Sigma \mathbf{y} - \mathbf{c}\|_2^2 = \sum_{i=1}^{n} (\Sigma_{ii} y_i - c_i)^2 = \sum_{i=1}^{r} (\sigma_i y_i - c_i)^2 + \sum_{i=r+1}^{n} c_i^2.$$

We are looking to choose $\mathbf{y}$ such that this expression is minimized. Clearly, there is nothing that we can do about the term $\sum_{i=r+1}^{n} c_i^2$. But by choosing $y_i = c_i/\sigma_i$ for all $i \in [r]$, we get $\sum_{i=1}^{r} (\sigma_i y_i - c_i)^2 = 0$. Hence, this choice of $\mathbf{y}$ must be optimal. Concretely, we conclude that the optimal solution is

$$\mathbf{y}^* = \begin{bmatrix} c_1/\sigma_1 \\ \vdots \\ c_r/\sigma_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \arg\min_{\mathbf{y} \in \mathbb{R}^n} \|\Sigma \mathbf{y} - \mathbf{c}\|_2^2.$$

c) In subtask a), we substituted $\mathbf{y} = V^\top \mathbf{x}$. Hence, it would make sense to guess that $\mathbf{x}^* = V \mathbf{y}^*$. Indeed, we can verify that with this choice of $\mathbf{x}^*$ we get

$$\|\Sigma \mathbf{y}^* - \mathbf{c}\|_2^2 = \|\Sigma V^\top \mathbf{x}^* - \mathbf{c}\|_2^2 = \|U \Sigma V^\top \mathbf{x}^* - U U^\top \mathbf{b}\|_2^2 = \|U \Sigma V^\top \mathbf{x}^* - U U^\top \mathbf{b}\|_2^2 = \|A \mathbf{x}^* - \mathbf{b}\|_2^2$$

and by $\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_2^2 = \min_{\mathbf{y} \in \mathbb{R}^n} \|\Sigma \mathbf{y} - \mathbf{c}\|_2^2$ and optimality of $\mathbf{y}^*$ we conclude that $\mathbf{x}^*$ is optimal, i.e.

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathbb{R}^n} \|A \mathbf{x}^* - \mathbf{b}\|_2^2.$$

4. **a)** We prove this by direct calculation

$$\| \mathbf{x} \|_2^2 = \sum_{i=1}^{n} x_i^2 \le \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i||x_j| = (\sum_{i=1}^{n} |x_i|)^2 = \| \mathbf{x} \|_1^2 .$$

Observe that the inequality $\sum_{i=1}^{n} x_i^2 \le \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i||x_j|$ holds because all terms appearing on the left actually appear on the right as well (but on the right we have some additional non-negative terms).

b) Without loss of generality, assume that all entries in $\mathbf{x}$ are non-negative (if there was a negative entry, simply switch its sign and observe that both norms remain the same). Next, observe that $\| \mathbf{x} \|_1 = \sum_{i=1}^{n} |x_i| = \sum_{i=1}^{n} x_i = \mathbf{1}^\top \mathbf{x}$ where $\mathbf{1} \in \mathbb{R}^n$ is the all-ones vector. By Cauchy-Schwarz, we obtain $\mathbf{1}^\top \mathbf{x} \le \| \mathbf{1} \|_2 \| \mathbf{x} \|_2$. It remains to calculate $\| \mathbf{1} \|_2 = (\sum_{i=1}^{n} 1)^{\frac{1}{2}} = \sqrt{n}$ to conclude that

$$\| \mathbf{x} \|_1 = \mathbf{1}^\top \mathbf{x} \le \| \mathbf{1} \|_2 \| \mathbf{x} \|_2 = \sqrt{n} \| \mathbf{x} \|_2 .$$

5. **a)** Recall that the trace of a matrix is the sum of its diagonal entries. Consider the matrix $A^\top A$. The j-th diagonal entry of $A^\top A$ is exactly the norm of the j-th column of A which is given by $\sum_{i=1}^{m} A_{ij}^2$. Hence, the trace of $A^\top A$ is given by

$$\text{Tr}(A^\top A) = \sum_{j=1}^{n} \sum_{i=1}^{m} A_{ij}^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2 = \| A \|_F^2 .$$

b) By Remark 7.1.13 we know that the squared singular values of A are the eigenvalues of the matrix $A^\top A$. Moreover, by Proposition 6.1.11 we know that the trace of $A^\top A$ is equal to the sum of its eigenvalues. Hence, we conclude

$$\text{Tr}(A^\top A) = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2$$

and the result follows by combining this with the previous subtask.

c) By definition, we have

$$\| A \|_{op} = \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \| \mathbf{x} \|_2 = 1}} \| A\mathbf{x} \|_2 \,.$$

Now observe that we can rewrite the squared version of this as

$$\max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \| \mathbf{x} \|_2^2 = 1}} \| A\mathbf{x} \|_2^2 = \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{0\}} \frac{\| A\mathbf{x} \|_2^2}{\| \mathbf{x} \|_2^2} = \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{0\}} \frac{\mathbf{x}^\top A^\top A\mathbf{x}}{\mathbf{x}^\top \mathbf{x}}.$$

The matrix $A^\top A$ is symmetric and its largest eigenvalue is σ_1^2, hence we get $\max_{\mathbf{x} \in \mathbb{R}^n \setminus \{0\}} \frac{\mathbf{x}^\top A^\top A\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \sigma_1^2$ by Proposition 6.3.10. It remains to observe that

$$\arg\max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \| \mathbf{x} \|_2 = 1}} \| A\mathbf{x} \|_2 = \arg\max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \| \mathbf{x} \|_2^2 = 1}} \| A\mathbf{x} \|_2^2$$

and hence

$$\| A \|_{op} = \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \| \mathbf{x} \|_2 = 1}} \| A\mathbf{x} \|_2 = \sqrt{\max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \| \mathbf{x} \|_2^2 = 1}} \| A\mathbf{x} \|_2^2} = \sqrt{\sigma_1^2} = \sigma_1.$$

d) This follows from b) and c) as

$$\| A \|_{op} = \sigma_1 = \sqrt{\sigma_1^2} \leq \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2} = \| A \|_F \,.$$

e) Using previous subtasks, we obtain

$$\| A \|_F^2 = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2 \leq \min\{m,n\} \sigma_1^2$$

and hence

$$\| A \|_F \leq \sqrt{\min\{m,n\}} \sigma_1 = \sqrt{\min\{m,n\}} \, \| A \|_{op} \,.$$