Bernd Gärtner
Robert Weismantel

Solution for Assignment 13

1. The matrix $-S^2$ is symmetric since

$$(-S^2)^\top = -(S^2)^\top = -(S^\top)^2 = -(-S)^2 = -S^2$$

where we used the assumption $S^\top = -S$.

From the lecture, we know that a symmetric matrix such as $-S^2$ is positive semidefinite if $\mathbf{x}^\top(-S^2)\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. To verify that this holds here, let $\mathbf{x} \in \mathbb{R}^n$ be arbitrary and observe that

$$\mathbf{x}^\top(-S^2)\mathbf{x} = \mathbf{x}^\top(-S)S\mathbf{x} = \mathbf{x}^\top S^\top S\mathbf{x} = \|S\mathbf{x}\|^2 \geq 0.$$

We conclude that $-S^2$ is positive semidefinite.

2. Let $\mathbf{v} \in \mathbb{R}^n$ be an arbitrary non-zero vector. We calculate

$$\mathbf{v}^\top A\mathbf{v} = \sum_{i=1}^n \sum_{j=1}^n v_i v_j A_{ij} = n\sum_{i=1}^n v_i^2 + \sum_{i<j} 2v_i v_j \geq n\sum_{i=1}^n v_i^2 + \sum_{i<j}(-v_i^2 - v_j^2) = \sum_{i=1}^n v_i^2 > 0,$$

where we have used that $0 \leq (v_i + v_j)^2 = v_i^2 + 2v_iv_j + v_j^2$ for all $i, j \in [n]$. We conclude that A is indeed positive definite.

3. We know A is symmetric, so we have to show that $\mathbf{x}^\top A\mathbf{x} \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$. We have

$$\begin{aligned}
\mathbf{x}^\top A\mathbf{x} = \sum_{ij} A_{ij}x_ix_j &= \sum_{i=1}^n A_{ii}x_i^2 + \sum_{i=1}^n \sum_{j\neq i} A_{ij}x_ix_j \\
&\geq \sum_{i=1}^n \sum_{j\neq i} |A_{ij}|\, x_i^2 - \sum_{i=1}^n \sum_{j\neq i} |A_{ij}|\,|x_i|\,|x_j| \\
&\geq \sum_{i=1}^n \sum_{j>i} |A_{ij}|\,(x_i^2 + x_j^2) - 2\sum_{i=1}^n \sum_{j>i} |A_{ij}|\,|x_i|\,|x_j| \\
&= \sum_{i=1}^n \sum_{j>i} |A_{ij}| \left(|x_i|^2 - 2\,|x_i|\,|x_j| + |x_j|^2\right) \\
&= \sum_{i=1}^n \sum_{j>i} |A_{ij}|\,(|x_i| - |x_j|)^2 \geq 0.
\end{aligned}$$

4. **a)** Let $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$ be an eigenvector of $A+B$ corresponding to eigenvalue $\lambda_{\min}^{(A+B)}$. By using our knowledge about Rayleigh quotients (Proposition 9.2.1), we get

$$\lambda_{\min}^{(A+B)} = \frac{\mathbf{x}^\top(A+B)\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} = \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} + \frac{\mathbf{x}^\top B\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} \overset{9.2.1}{\geq} \lambda_{\min}^{(A)} + \lambda_{\min}^{(B)}.$$

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

5. 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 6.4.3, 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 6.4.1 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 6.4.9.

6. **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 $\left\|U^\top \mathbf{v}\right\|_2^2 = (U^\top \mathbf{v})^\top (U^\top \mathbf{v}) = \mathbf{v}^\top U U^\top \mathbf{v} = \mathbf{v}^\top \mathbf{v} = \|\mathbf{v}\|_2^2$ 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

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

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

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

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.

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