Afonso Bandeira
Bernd Gärtner

Solution for Assignment 8

1. **a)** Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ denote the columns of A. Performing the Gram-Schmidt process (Algorithm 4.4.10) yields

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$$

$$\mathbf{q}_2' = \mathbf{a}_2 - (\mathbf{a}_2^\top \mathbf{q}_1)\mathbf{q}_1 = \mathbf{a}_2 - \frac{1}{\sqrt{2}}\mathbf{q}_1 = \begin{bmatrix}0\\0\\1\end{bmatrix} - \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = \begin{bmatrix}0\\-1/2\\1/2\end{bmatrix}$$

$$\mathbf{q}_2 = \frac{\mathbf{q}_2'}{\|\mathbf{q}_2'\|} = \begin{bmatrix}0\\-1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$$

$$\mathbf{q}_3' = \mathbf{a}_3 - (\mathbf{a}_3^\top \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{a}_3^\top \mathbf{q}_2)\mathbf{q}_2 = \mathbf{a}_3 - \sqrt{2}\mathbf{q}_1 - 0\mathbf{q}_2 = \begin{bmatrix}1\\1\\1\end{bmatrix} - \begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}1\\0\\0\end{bmatrix}$$

$$\mathbf{q}_3 = \frac{\mathbf{q}_3'}{\|\mathbf{q}_3'\|} = \mathbf{q}_3'$$

where $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ is the desired set of orthonormal vectors.

b) Putting the vectors $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ into a matrix we obtain

$$Q = \begin{bmatrix}0 & 0 & 1\\ 1/\sqrt{2} & -1/\sqrt{2} & 0\\ 1/\sqrt{2} & 1/\sqrt{2} & 0\end{bmatrix}$$

and it remains to compute R. Concretely, we have

$$\begin{aligned} R = Q^\top A &= \begin{bmatrix}0 & 0 & 1\\ 1/\sqrt{2} & -1/\sqrt{2} & 0\\ 1/\sqrt{2} & 1/\sqrt{2} & 0\end{bmatrix}^\top \begin{bmatrix}0 & 0 & 1\\ 1 & 0 & 1\\ 1 & 1 & 1\end{bmatrix} \\ &= \begin{bmatrix}0 & 1/\sqrt{2} & 1/\sqrt{2}\\ 0 & -1/\sqrt{2} & 1/\sqrt{2}\\ 1 & 0 & 0\end{bmatrix} \begin{bmatrix}0 & 0 & 1\\ 1 & 0 & 1\\ 1 & 1 & 1\end{bmatrix} \\ &= \begin{bmatrix}\sqrt{2} & 1/\sqrt{2} & \sqrt{2}\\ 0 & 1/\sqrt{2} & 0\\ 0 & 0 & 1\end{bmatrix}. \end{aligned}$$

c) Let $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4$ denote the columns of B. Performing the Gram-Schmidt process (Algorithm 4.4.10) yields

$$\begin{aligned}
\mathbf{q}_1 &= \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|} = \mathbf{b}_1 \\
\mathbf{q}_2' &= \mathbf{b}_2 - (\mathbf{b}_2^\top \mathbf{q}_1)\mathbf{q}_1 = \mathbf{b}_2 - 2\mathbf{q}_1 = \begin{bmatrix} 0 & 4 & 0 & 0 \end{bmatrix}^\top \\
\mathbf{q}_2 &= \frac{\mathbf{q}_2'}{\|\mathbf{q}_2'\|} = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}^\top \\
\mathbf{q}_3' &= \mathbf{b}_3 - (\mathbf{b}_3^\top \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{b}_3^\top \mathbf{q}_2)\mathbf{q}_2 = \mathbf{b}_3 - 3\mathbf{q}_1 - 5\mathbf{q}_2 = \begin{bmatrix} 0 & 0 & 7 & 0 \end{bmatrix}^\top \\
\mathbf{q}_3 &= \frac{\mathbf{q}_3'}{\|\mathbf{q}_3'\|} = \mathbf{q}_3' = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^\top \\
\mathbf{q}_4' &= \mathbf{b}_4 - (\mathbf{b}_4^\top \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{b}_4^\top \mathbf{q}_2)\mathbf{q}_2 - (\mathbf{b}_4^\top \mathbf{q}_3)\mathbf{q}_3 = \mathbf{b}_4 - 0\mathbf{q}_1 - 6\mathbf{q}_2 - 8\mathbf{q}_3 = \begin{bmatrix} 0 & 0 & 0 & 9 \end{bmatrix}^\top \\
\mathbf{q}_4 &= \frac{\mathbf{q}_4'}{\|\mathbf{q}_4'\|} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\top
\end{aligned}$$

where $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4$ is the desired set of orthonormal vectors.

d) This is not always true. The $n \times n$ matrix $-I$ is a counterexample for any $n \in \mathbb{N}^+$. It already has orthonormal columns, hence Gram-Schmidt would leave it unaltered. Moreover, its columns are not exactly the standard unit vectors: the sign is wrong. Therefore, this is indeed a counterexample.

Note that this is already a full solution. But we still provide a proof that the answer to the question would be yes if we had required the diagonal entries to be strictly positive (and not just non-zero).

Let A be an arbitrary upper triangular $n \times n$ matrix with strictly positive entries on its diagonal. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ denote the columns of A and let $\mathbf{q}_1, \ldots, \mathbf{q}_n$ denote the orthonormal vectors obtained from the Gram Schmidt process on $\mathbf{a}_1, \ldots, \mathbf{a}_n$. We claim that $\mathbf{q}_i = \mathbf{e}_i$ for all $i \in [n]$. Assume for a contradiction that this is not the case and let $i \in [n]$ be the smallest index such that $\mathbf{q}_i \neq \mathbf{e}_i$. Note that we have $\mathbf{a}_1 = c\mathbf{e}_1$ for some constant $c \in \mathbb{R}^+$ and hence $\mathbf{q}_1 = \frac{\mathbf{a}_1}{c} = \mathbf{e}_1$. Hence, we must have $i > 1$. Observe that by definition of the Gram-Schmidt process and because the last $n - i$ entries of $\mathbf{a}_i$ are zero (triangular shape of A), we also get that the last $n - i$ entries of $\mathbf{q}_i$ are zero. We claim that the first $i - 1$ entries of $\mathbf{q}_i$ are zero as well. To see this, assume for a moment that there is $j < i$ such that the j-th entry of $\mathbf{q}_i$ is non-zero. Then $\mathbf{q}_j^\top \mathbf{q}_i = \mathbf{e}_j^\top \mathbf{q}_i \neq 0$ which contradicts the orthogonality of $\mathbf{q}_j$ and $\mathbf{q}_i$. Hence, we conclude that the first $i - 1$ entries of $\mathbf{q}_i$ are zero. In particular, we established that the only non-zero entry of $\mathbf{q}_i$ is the i-th entry. Since $\mathbf{q}_i$ must be a unit vector (by the Gram-Schmidt process), we get $\mathbf{q}_i = \mathbf{e}_i$, a contradiction.

2. We prove this by showing that the columns of P are orthonormal, i.e. they all have unit length and they are pairwise orthogonal. Clearly, all columns have unit length since they are standard unit vectors. Thus, it remains to show pairwise orthogonality. Let $i \neq j$ with $i, j \in [n]$ be arbitrary. We want to show that the scalar product $\mathbf{e}_{p(i)} \cdot \mathbf{e}_{p(j)}$ is 0. We observe that we have $\mathbf{e}_{p(i)} \cdot \mathbf{e}_{p(j)} = 0$ if and only if $p(i) \neq p(j)$. Moreover, by injectivity of p and since $i \neq j$, we must have $p(i) \neq p(j)$. Hence, we conclude that the two vectors are orthogonal. This holds for any $i \neq j$ and hence we conclude that the columns of P are pairwise orthogonal.

3. a) Recall that

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

and hence

$$R_\theta^\top = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

In order to prove that R_θ is orthogonal, it suffices to show $R_\theta^\top R_\theta = I$. For this, we need the trigonometric identity $\sin^2(\theta) + \cos^2(\theta) = 1$ to calculate

$$R_\theta^\top R_\theta = \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & \cos(\theta)\sin(\theta) - \sin(\theta)\cos(\theta) \\ \sin(\theta)\cos(\theta) - \cos(\theta)\sin(\theta) & \sin^2(\theta) + \cos^2(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Hence, R_θ is indeed orthogonal.

b) Consider the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Clearly, A is an orthogonal matrix. Moreover, A is not a rotation matrix because there is no $\theta \in \mathbb{R}$ satisfying both $1 = \sin(\theta)$ and $1 = -\sin(\theta)$.

c) Assume that A is orthogonal. Recall the formula for the 2×2 inverse

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Since A is orthogonal, we must have $A^\top = A^{-1}$. From this, we deduce $a = \frac{d}{ad-bc}$, $d = \frac{a}{ad-bc}$, $c = \frac{-b}{ad-bc}$, and $b = \frac{-c}{ad-bc}$. Note that $ad - bc \neq 0$ since A is invertible.

Assume first $a \neq 0$. Then we obtain $ad - bc = \frac{d}{a} = \frac{a}{d}$ since we also must have $d \neq 0$. This implies $|a| = |d|$ and $|ad - bc| = 1$.

On the other hand, if we have $a = 0$ then we must have $b \neq 0$ and $c \neq 0$. Thus, we get $ad - bc = \frac{-b}{c} = \frac{-c}{b}$ and therefore $|b| = |c|$ and $|ad - bc| = 1$.

d) Consider the matrix A that we get by setting $a = d = \sqrt{2}$ and $b = c = 1$. Clearly, we have $|ad - bc| = 2 - 1 = 1$. But A is not orthogonal since in particular, its two columns $\begin{bmatrix} \sqrt{2} & 1 \end{bmatrix}^\top$ and $\begin{bmatrix} 1 & \sqrt{2} \end{bmatrix}^\top$ are not orthogonal (and also they are not unit vectors).

4. **a)** We obtained the new datapoints from the old one by shifting them along the t-axis. Imagine a line that optimally fits the old datapoints. By shifting this line the same amount in the direction of the t-axis, we should get a line that optimally fits the new datapoints. In other words, it should be possible to obtain the optimal line for the new datapoints from the optimal line for the old datapoints by a shift along the t-axis. This should also work the other way around, i.e. given an optimal line for the new datapoints we should be able to shift the line by $-c$ in the direction of the t-axis to get an optimal line for the old datapoints.

It remains to think about this shifting operation in terms of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$. The main insight here is that shifting a line does not change its slope. Hence we expect to have $\alpha_1 = \alpha_1'$ but not $\alpha_0 = \alpha_0'$.

b) We compute the scalar product and check that it yields zero. Concretely, we have

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \begin{bmatrix} t_1' \\ \vdots \\ t_m' \end{bmatrix} = \sum_{k=1}^{m} t_m' = \sum_{k=1}^{m} (t_m + c) = \sum_{k=1}^{m} (t_k - \frac{1}{m} \sum_{i=1}^{m} t_i) = \sum_{k=1}^{m} t_k - m\frac{1}{m} \sum_{i=1}^{m} t_i = \sum_{k=1}^{m} t_k - \sum_{i=1}^{m} t_i = 0$$

and hence the columns of A' are indeed orthogonal.

c) We prove that for all $\boldsymbol{\alpha}' \in \mathbb{R}^2$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}' + \begin{bmatrix} c\alpha_1' \\ 0 \end{bmatrix}$, we have $||A'\boldsymbol{\alpha}' - \mathbf{b}||^2 = ||A\boldsymbol{\alpha} - \mathbf{b}||^2$. From this, it follows that $\boldsymbol{\alpha}'$ is optimal if and only if $\boldsymbol{\alpha}$ is optimal.

Thus, consider an arbitrary $\boldsymbol{\alpha}' \in \mathbb{R}^2$ and let $\boldsymbol{\alpha} = \boldsymbol{\alpha}' + \begin{bmatrix} c\alpha_1' \\ 0 \end{bmatrix}$. We compute

$$\begin{aligned}
||A'\boldsymbol{\alpha}' - \mathbf{b}||^2 &= \sum_{k=1}^{m}(b_k - (\alpha_0' + \alpha_1' t_k'))^2 \\
&= \sum_{k=1}^{m}(b_k - (\alpha_0' + \alpha_1'(t_k + c)))^2 && (t_k' = t_k + c) \\
&= \sum_{k=1}^{m}(b_k - ((\alpha_0 - c\alpha_1') + \alpha_1'(t_k + c)))^2 && (\alpha_0' = \alpha_0 - c\alpha_1') \\
&= \sum_{k=1}^{m}(b_k - ((\alpha_0 - c\alpha_1) + \alpha_1(t_k + c)))^2 && (\alpha_1' = \alpha_1) \\
&= \sum_{k=1}^{m}(b_k - (\alpha_0 - c\alpha_1 + \alpha_1 t_k + \alpha_1 c))^2 \\
&= \sum_{k=1}^{m}(b_k - (\alpha_0 + \alpha_1 t_k))^2 \\
&= ||A\boldsymbol{\alpha} - \mathbf{b}||^2
\end{aligned}$$

and therefore conclude the proof.

d) By the closed form solution and subtask b), we get that the optimal $\boldsymbol{\alpha}'$ satisfies

$$\begin{bmatrix} \alpha_0' \\ \alpha_1' \end{bmatrix} = \begin{bmatrix} \frac{1}{m}\sum_{k=1}^{m} b_k \\ (\sum_{k=1}^{m} t_k' b_k)/(\sum_{k=1}^{m} {t_k'}^2) \end{bmatrix}.$$

By subtask c), the optimal $\boldsymbol{\alpha}$ is hence given by

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{m}\sum_{k=1}^{m} b_k + c(\sum_{k=1}^{m} t_k' b_k)/(\sum_{k=1}^{m} {t_k'}^2) \\ (\sum_{k=1}^{m} t_k' b_k)/(\sum_{k=1}^{m} {t_k'}^2) \end{bmatrix}$$

where $c = -\frac{1}{m}\sum_{k=1}^{m} t_k$ and $t_k' = t_k + c$. This is a closed form solution because the right-hand side can be directly computed from the datapoints.

5. **a)** Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ be the columns of A. We first compute all the scalar products between columns of A. In particular, we get

$$\mathbf{a}_1^\top \mathbf{a}_1 = m, \quad \mathbf{a}_1^\top \mathbf{a}_2 = \sum_{k=1}^{m} t_k, \quad \mathbf{a}_1^\top \mathbf{a}_3 = \sum_{k=1}^{m} t_k^2, \quad \mathbf{a}_2^\top \mathbf{a}_2 = \sum_{k=1}^{m} t_k^2, \quad \mathbf{a}_2^\top \mathbf{a}_3 = \sum_{k=1}^{m} t_k^3, \quad \mathbf{a}_3^\top \mathbf{a}_3 = \sum_{k=1}^{m} t_k^4$$

and therefore

$$A^\top A = \begin{bmatrix} m & \sum_{k=1}^{m} t_k & \sum_{k=1}^{m} t_k^2 \\ \sum_{k=1}^{m} t_k & \sum_{k=1}^{m} t_k^2 & \sum_{k=1}^{m} t_k^3 \\ \sum_{k=1}^{m} t_k^2 & \sum_{k=1}^{m} t_k^3 & \sum_{k=1}^{m} t_k^4 \end{bmatrix}.$$

b) For $A^\top A$ to be diagonal, we need to have $\sum_{k=1}^{m} t_k = 0$, $\sum_{k=1}^{m} t_k^2 = 0$, and $\sum_{k=1}^{m} t_k^3 = 0$. The first and last condition are not so interesting, but note that the condition $\sum_{k=1}^{m} t_k^2 = 0$ implies $t_k = 0$ for all $k \in [m]$ because we clearly have $t_k^2 \geq 0$ for all $k \in [m]$.