Bernd Gärtner
Robert Weismantel

Solution for Assignment 9

1. **a)** Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ denote the columns of A. Performing the Gram-Schmidt process (Algorithm 5.4.9) 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 5.4.9) yields

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

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 $\mathbf{a}_1, \dots, \mathbf{a}_n$ denote the columns of A and let $\mathbf{q}_1, \dots, \mathbf{q}_n$ denote the orthonormal vectors obtained from the Gram Schmidt process on $\mathbf{a}_1, \dots, \mathbf{a}_n$. 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. Let $\mathbf{q}_1, \dots, \mathbf{q}_m \in \mathbb{R}^m$ be the columns of Q, i.e.

$$Q = \begin{bmatrix} | & \dots & | \\ \mathbf{q}_1 & \dots & \mathbf{q}_n \\ | & \dots & | \end{bmatrix}.$$

We want to prove that $Q^\top Q = I$. Let $i, j \in [m]$ be arbitrary and consider the standard unit vectors $\mathbf{e}_i, \mathbf{e}_j \in \mathbb{R}^m$. By assumption, we have

$$\mathbf{q}_i^\top \mathbf{q}_j = (Q\mathbf{e}_i)^\top (Q\mathbf{e}_j) = \mathbf{e}_i^\top \mathbf{e}_j = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Using this, we get

$$Q^\top Q = \begin{bmatrix} \mathbf{q}_1^\top \mathbf{q}_1 & \mathbf{q}_1^\top \mathbf{q}_2 & \cdots & \mathbf{q}_1^\top \mathbf{q}_m \\ \mathbf{q}_2^\top \mathbf{q}_1 & \mathbf{q}_2^\top \mathbf{q}_2 & \ddots & \mathbf{q}_2^\top \mathbf{q}_m \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{q}_m^\top \mathbf{q}_1 & \mathbf{q}_m^\top \mathbf{q}_2 & \cdots & \mathbf{q}_m^\top \mathbf{q}_m \end{bmatrix} = I$$

and thus Q is orthogonal.

3. **a)** 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 $1 = -\sin(\theta)$.

b) 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 $|ad - bc| = 1$.

c) 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)** 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 $k \in [m]$.

5. **a)** Let us denote the four given points by $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4$, respectively. We want to find $r \in \mathbb{R}^+$ such that the sum

$$\sum_{i=1}^{4} (r - ||\mathbf{p}_i||)^2$$

is minimized. The key observation of this exercise is that this is the least squares objective of the linear system

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} r \end{bmatrix} = \begin{bmatrix} ||\mathbf{p}_1|| \\ ||\mathbf{p}_2|| \\ ||\mathbf{p}_3|| \\ ||\mathbf{p}_4|| \end{bmatrix} = \begin{bmatrix} 2 \\ \sqrt{2} \\ \sqrt{\frac{20}{9}} \\ \sqrt{\frac{10}{4}}. \end{bmatrix}$$

Using the normal equations to solve this we get

$$4r = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} r \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} ||\mathbf{p}_1|| \\ ||\mathbf{p}_2|| \\ ||\mathbf{p}_3|| \\ ||\mathbf{p}_4|| \end{bmatrix} = \sum_{i=1}^{4} ||\mathbf{p}_i||$$

and hence

$$r = \frac{1}{4} \sum_{i=1}^{4} ||\mathbf{p}_i|| = \frac{1}{4}(2 + \sqrt{2} + \sqrt{\frac{20}{9}} + \sqrt{\frac{10}{4}}).$$

b) In this more general setting, we need to solve the system

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} r \end{bmatrix} = \begin{bmatrix} ||\mathbf{p}_1|| \\ \vdots \\ ||\mathbf{p}_n|| \end{bmatrix}$$

in the least squares sense for r. Using the normal equations, this now yields

$$nr = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} r \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} ||\mathbf{p}_1|| \\ \vdots \\ ||\mathbf{p}_4|| \end{bmatrix} = \sum_{i=1}^{n} ||\mathbf{p}_i||$$

and thus

$$r = \frac{1}{n} \sum_{i=1}^{4} ||\mathbf{p}_i||.$$

6. Observe first that the concatenation $(\sigma \circ \pi) : [n] \to [n]$ (defined as $(\sigma \circ \pi)(i) := \sigma(\pi(i))$ for all $i \in [n]$) of two bijective funtions $\sigma, \pi : [n] \to [n]$ is again bijective: Indeed, if we had $\sigma(\pi(i)) = \sigma(\pi(j))$ for some distinct $i, j \in [n]$, this would also imply either $\pi(i) = \pi(j)$ or $\pi(i) \neq \pi(j)$ but $\sigma(p(i)) = \sigma(\pi(j))$, contradicting injectivity of σ or π in either case. Thus, $(\sigma \circ \pi)$ is injective. Moreover, any injective function from $[n]$ to $[n]$ is automatically surjective. We conclude that multiplying two permutation matrices $A, B \in \mathbb{R}^{n \times n}$ yields again a permutation matrix AB.

In particular, this observation implies that the matrices $P, P^2, P^3, \dots$ are all permutation matrices. Since there are only finitiely many permutation matrices of size $n \times n$, there must exist distinct indices $\ell, r \in \mathbb{N}$ such that $P^\ell = P^r$. Multiplying both sides with $(P^{-1})^\ell$ yields $I = P^{r-\ell}$. Thus, the statement holds with $k = r - \ell$.