Homework 9

This homework is due on Friday, October 28, 2022, at 11:59PM. Self-grades and HW Resubmissions are due on the following Friday, November 4, 2022, at 11:59PM.

1. **(OPTIONAL) Mid-Semester Survey**

   Please fill out this mid-semester survey to let us know how the class has been going so far! This survey is optional and anonymous, but you can submit a screenshot of the final page of the survey to Gradescope to receive 2 global extra credit points! We will be accepting submissions on Gradescope until Sunday, October 30 at 11:59pm.

2. Change of Basis

(a) For any given vector, we have to choose a basis to write this vector in. Typically, we choose the standard basis $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ where $\vec{e}_i$ is a vector with a 1 in the ith position and zeros everywhere else. **Given a vector $\vec{x} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$, write $\vec{x}$ as a linear combination of standard basis vectors.**

Solution: We have that $\alpha_i \vec{e}_i$ will put the term α_i in the ith component of the vector and zeros everywhere else. Hence, we can write

$$\vec{x} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \cdots + \alpha_n \vec{e}_n \tag{1}$$

(b) We can also represent the same vector $\vec{x}$ in a different basis. Let us write this new basis as $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. **Find a way to write $\vec{x}$ from the previous subpart as a linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$.** Simplify your answer as an equation with matrix-vector multiplication, and assume that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are linearly independent.

(HINT: One representation of $\vec{x}$ is the one you determined in the previous subpart. Another representation of $\vec{x}$ is $\widetilde{\alpha}_1 \vec{v}_1 + \widetilde{\alpha}_2 \vec{v}_2 + \cdots + \widetilde{\alpha}_n \vec{v}_n$. We need these two representations to be algebraically equal to indicate that they both represent the same vector. For your convenience, you may define $\vec{\widetilde{\alpha}} = \begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_n \end{bmatrix}$.)

Solution: We can first write $\widetilde{\alpha}_1 \vec{v}_1 + \widetilde{\alpha}_2 \vec{v}_2 + \cdots + \widetilde{\alpha}_n \vec{v}_n$ as

$$\widetilde{\alpha}_1 \vec{v}_1 + \widetilde{\alpha}_2 \vec{v}_2 + \cdots + \widetilde{\alpha}_n \vec{v}_n = \underbrace{\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}}_{V} \begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_n \end{bmatrix} \tag{2}$$

We want this to be equal to $\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$, so by setting these two terms equal, we have

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = V \begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_n \end{bmatrix} \tag{3}$$

Hence, we have that

$$\begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_n \end{bmatrix} = V^{-1} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \tag{4}$$

where V is invertible because $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are linearly independent.

(c) Suppose that we truncated our basis so that we now only have $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ where $m < n$ linearly independent vectors, but we could still represent $\vec{x}$ as a linear combination of these vectors. **How do you modify your method from the previous part?** You may not assume that you know $\vec{v}_{m+1}, \dots, \vec{v}_n$.

(HINT: Think about using projections. Specifically, consider projecting onto the column space of a matrix that you define.)

Solution: Following the hint, we can define a matrix $V_m = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{bmatrix}$. We want to find a vector $\begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_m \end{bmatrix}$ such that

$$V_m \begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_m \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \tag{5}$$

We can solve this as a least squares problem and obtain

$$\begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_m \end{bmatrix} = \left(V_m^\top V_m\right)^{-1} V_m^\top \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \tag{6}$$

(d) Suppose that all the vectors $\vec{v}_i$ from the previous part were orthonormal. **Simplify your answer from the previous subpart under this assumption.**

(HINT: Let $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \end{bmatrix} \in \mathbb{R}^{n \times m}$ where $n > m$. If $S = U^\top U$, then $S_{ij} = \vec{u}_i^\top \vec{u}_j$.)

Solution: Following the hint, we can define $S = V_m^\top V_m$. We have that $S_{ij} = \vec{v}_i^\top \vec{v}_j$. If $i = j$, then $S_{ij} = S_{ii} = \vec{v}_i^\top \vec{v}_i = 1$. Otherwise, if $i \neq j$, $S_{ij} = \vec{v}_i^\top \vec{v}_j = 0$. Hence, we have ones along the diagonal and zeros everywhere else, so $S = I$. Thus, our result from the previous part simplifies to

$$\begin{bmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_m \end{bmatrix} = \Big(\underbrace{V_m^\top V_m}_{I}\Big)^{-1} V_m^\top \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = V_m^\top \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \tag{7}$$

3. Cayley-Hamilton and Controllability Matrix

(a) We can define the *characteristic polynomial* of a matrix $A \in \mathbb{R}^{n\times n}$ as

$$p_A(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0\lambda^0 \tag{8}$$

where each $c_i \in \mathbb{R}$ is a constant. The characteristic polynomial has roots that are the eigenvalues of A. That is, we can equivalently define

$$p_A(\lambda) = \det\{\lambda I - A\} \tag{9}$$

We say that any of the eigenvalues of A "satisfy" the characteristic polynomial in that

$$p_A(\lambda_i) = 0 \tag{10}$$

where λ_i is the ith eigenvalue of A. Now, let A be a diagonalizable matrix, where we may write $A = V\Lambda V^{-1}$. **Prove that A satisfies its own characteristic polynomial.** In other words, prove that $p_A(A) = 0_{n\times n}$, where $0_{n\times n}$ is a $n \times n$ matrix of zeros.

(HINT: It is not correct to simply plug in $\lambda = A$ into $\det\{\lambda I - A\}$*.)*

Solution: Recall that $A^i = V\Lambda^i V^{-1}$. Hence,

$$\begin{aligned} p_A(A) &= A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0A^0 && (11)\\ &= V\Lambda^n V^{-1} + c_{n-1}V\Lambda^{n-1}V^{-1} + \cdots + c_1V\Lambda V^{-1} + c_0I && (12)\\ &= V\Big(\Lambda^n + c_{n-1}\Lambda^{n-1} + \cdots + c_1\Lambda + c_0I\Big)V^{-1} && (13)\end{aligned}$$

Notice that the ith diagonal entry of $\Lambda^n + c_{n-1}\Lambda^{n-1} + \cdots + c_1\Lambda + c_0I$ is $\lambda_i^n + c_{n-1}\lambda_i^{n-1} + \cdots + c_1\lambda_i + c_0 = p_A(\lambda_i)$. Thus, we have that

$$p_A(A) = V\begin{bmatrix} p_A(\lambda_1) & & & \\ & p_A(\lambda_2) & & \\ & & \ddots & \\ & & & p_A(\lambda_n)\end{bmatrix}V^{-1} \tag{14}$$

Note that $p_A(\lambda_i) = 0$, so

$$\begin{aligned} p_A(A) &= V\begin{bmatrix} p_A(\lambda_1) & & & \\ & p_A(\lambda_2) & & \\ & & \ddots & \\ & & & p_A(\lambda_n)\end{bmatrix}V^{-1} && (15)\\ &= V\begin{bmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{bmatrix}V^{-1} && (16)\\ &= 0_{n\times n} && (17)\end{aligned}$$

where $0_{n\times n}$ is an $n \times n$ matrix of zeros.

(b) Now, consider some vector $\vec{b} \in \mathbb{R}^n$. **Using the result from the previous part, show that $A^n\vec{b}$ is linearly dependent on $A^{n-1}\vec{b}, A^{n-2}\vec{b}, \ldots, A\vec{b}, \vec{b}$.**

Solution: From the previous part, we know that

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = 0_{n\times n} \tag{18}$$

Right multiplying both sides by $\vec{b}$, we get

$$A^n\vec{b} + c_{n-1}A^{n-1}\vec{b} + \cdots + c_1 A\vec{b} + c_0\vec{b} = \vec{0} \tag{19}$$

which we can rearrange to get

$$A^n\vec{b} = -\left(c_{n-1}A^{n-1}\vec{b} + \cdots + c_1 A\vec{b} + c_0\vec{b}\right) \tag{20}$$

Thus, $A^n\vec{b}$ is linearly dependent on $A^{n-1}\vec{b}, A^{n-2}\vec{b}, \ldots, A\vec{b}, \vec{b}$.

(c) Instead of setting $\vec{b}$ to be a vector, let it be a matrix $B \in \mathbb{R}^{n\times m}$. **Now, show that the columns of A^nB are linearly dependent on the columns of $A^{n-1}B, A^{n-2}B, \ldots, AB, B$.**
(HINT: If we were to write $B = \begin{bmatrix}\vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_m\end{bmatrix}$ where each column is n-dimensional, we can write $A^iB = \begin{bmatrix}A^i\vec{b}_1 & A^i\vec{b}_2 & \cdots & A^i\vec{b}_m\end{bmatrix}$. Make sure you convince yourself of this.)

Solution: We have that $A^nB = \begin{bmatrix}A^n\vec{b}_1 & A^n\vec{b}_2 & \cdots & A^n\vec{b}_m\end{bmatrix}$. From the previous part, we have that $A^n\vec{b}_i$ is linearly dependent on $A^{n-1}\vec{b}_i, A^{n-2}\vec{b}_i, \ldots, A\vec{b}_i, \vec{b}_i$. Since i is arbitrary here, we have that the columns of A^nB are linearly dependent on the columns of $A^{n-1}B, A^{n-2}B, \ldots, AB, B$.

(d) Consider a discrete time system of the form

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{21}$$

where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. The controllability matrix for this discrete time system is given by

$$\mathcal{C} = \begin{bmatrix}A^{n-1}B & A^{n-2}B & \cdots & AB & B\end{bmatrix} \tag{22}$$

Conclude that the rank of your controllability matrix will not change if, instead, you made your controllability matrix $\begin{bmatrix}A^nB & A^{n-1}B & \cdots & AB & B\end{bmatrix}$ (i.e., you prepended A^nB to your original controllability matrix).

Solution: From the previous part, we have that the columns of A^nB are linearly dependent on the columns of $A^{n-1}B, A^{n-2}B, \ldots, AB, B$. Hence, if we were to prepend A^nB to our original controllability matrix, the rank would not change since each column of A^nB is linearly dependent on columns already in the controllability matrix.

4. CCF Transformation and Controllability

(a) Consider the following discrete time system

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{23}$$

Suppose we define a change of basis operation given by $M\vec{z}[i] = \vec{x}[i] \iff \vec{z}[i] = M^{-1}\vec{x}[i]$. This yields a new discrete time system of the form

$$\vec{z}[i+1] = \widetilde{A}\vec{z}[i] + \widetilde{B}\vec{u}[i] \tag{24}$$

for some $\widetilde{A}$ and $\widetilde{B}$ defined in terms of M, A, and B. **What is the controllability matrix for the system in eq. (24), in terms of M, A, and B?**

Solution: We have that

$$\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] \tag{25}$$

$$M\vec{z}[i+1] = AM\vec{z}[i] + B\vec{u}[i] \tag{26}$$

$$\vec{z}[i+1] = \underbrace{M^{-1}AM}_{\widetilde{A}}\vec{z}[i] + \underbrace{M^{-1}B}_{\widetilde{B}}\vec{u}[i] \tag{27}$$

We have that the controllability matrix for the z system is $\mathcal{C}_z = \begin{bmatrix}\widetilde{A}^{n-1}\widetilde{B} & \widetilde{A}^{n-2}\widetilde{B} & \cdots & \widetilde{A}\widetilde{B} & \widetilde{B}\end{bmatrix}$ where $\widetilde{A}^i\widetilde{B} = \left(M^{-1}A^iM\right)\left(M^{-1}B\right) = M^{-1}A^iB$. Hence, we can rewrite the controllability matrix as

$$\mathcal{C}_z = \begin{bmatrix}M^{-1}A^{n-1}B & M^{-1}A^{n-2}B & \cdots & M^{-1}AB & M^{-1}B\end{bmatrix} \tag{28}$$

$$= M^{-1}\begin{bmatrix}A^{n-1}B & A^{n-2}B & \cdots & AB & B\end{bmatrix} \tag{29}$$

(b) Consider the change of basis given by $\vec{z}[i] = T^{-1}\vec{x}[i]$ where, under this change of basis transformation, we have the following discrete time system

$$\vec{z}[i+1] = A_{\text{CCF}}\vec{z}[i] + B_{\text{CCF}}\vec{u}[i] \tag{30}$$

Using the result from the previous part, determine an expression for T in terms of $\mathcal{C}$, the controllability matrix of the original system in eq. (23), and $\mathcal{C}_{\text{CCF}}$, the controllability matrix of the system in eq. (30).

Solution: From the previous part, we can set $M = T$ and $\mathcal{C}_z = \mathcal{C}_{\text{CCF}}$ to obtain

$$\mathcal{C}_{\text{CCF}} = T^{-1}\mathcal{C} \tag{31}$$

$$T\mathcal{C}_{\text{CCF}} = \mathcal{C} \tag{32}$$

$$T = \mathcal{C}\mathcal{C}_{\text{CCF}}^{-1} \tag{33}$$

(c) We know that the controllability matrix for a system in CCF will always be full rank. **Using this, prove that you can find a transformation matrix T as in the previous part if and only if your original system is controllable.** *(HINT: To prove this, you can first show that, if such a T exists, then* *your original system is controllable. Then, you can show that, if your original system is controllable, there will exist such a transformation matrix T.) (HINT: Recall that T must be invertible (equivalently, full rank) in order for it to be a valid transformation matrix. You may use without proof the fact that* $\text{rank}(AB) = \min(\text{rank}(A), \text{rank}(B))$.*)*

Solution: Following the hint, we have that, if T exists, then it must be full rank. Also, $\text{rank}(\mathcal{C}_{\text{CCF}}) = n$. From the previous part, we end up with

$$\text{rank}(\mathcal{C}_{\text{CCF}}) = \text{rank}(T^{-1}\mathcal{C}) \tag{34}$$

$$\text{rank}(\mathcal{C}_{\text{CCF}}) = \min(\text{rank}(T^{-1}), \text{rank}(\mathcal{C})) \tag{35}$$

$$n = \min(n, \text{rank}(\mathcal{C})) \tag{36}$$

Notice that $\text{rank}(\mathcal{C}) \leq n$, so $\min(n, \text{rank}(\mathcal{C})) = \text{rank}(\mathcal{C})$ and we conclude that $\text{rank}(\mathcal{C}) = n$. Next, we need to prove that if the original system is controllable (i.e., $\text{rank}(\mathcal{C}) = n$), then T exists. We already know how to compute T, so we need to show that $\text{rank}(T) = n$ (which would make it a valid basis transformation matrix). We have that

$$\text{rank}(T) = \text{rank}(\mathcal{C}\mathcal{C}_{\text{CCF}}^{-1}) \tag{37}$$

$$\text{rank}(T) = \min(\text{rank}(\mathcal{C}), \text{rank}(\mathcal{C}_{\text{CCF}}^{-1})) \tag{38}$$

$$\text{rank}(T) = \min(n, n) = n \tag{39}$$

so $\text{rank}(T) = n$ and it is thus a valid transformation matrix.

(d) Consider the following discrete-time dynamics model:

$$\vec{x}[i+1] = \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{A} \vec{x}[i] + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\vec{b}} \vec{u}[i] \tag{40}$$

Find the transformation matrix T such that the dynamics model for $\vec{z}[i] = T^{-1}\vec{x}[i]$ is in CCF. You may use a calculator/computer to perform any computations, if you wish.

(HINT: First, find the characteristic polynomial of A. Use this to determine what A_{CCF} and $\vec{b}_{\text{CCF}}$ should be, and then use this to determined $\mathcal{C}_{\text{CCF}}$.)

Solution: Firstly, we can compute $\mathcal{C}$ as follows:

$$\vec{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{41}$$

$$A\vec{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{42}$$

so $\mathcal{C} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ which is full rank. Hence, the transformation matrix T will exist. Following the hint, the characteristic polynomial of A is

$$p_A(\lambda) = \det\{A - \lambda I\} \tag{43}$$

$$= \det\left\{ \begin{bmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{bmatrix} \right\} \tag{44}$$

$$= (\lambda - 1)^2 \tag{45}$$

$$= \lambda^2 - 2\lambda + 1 \tag{46}$$

Here, we pattern match $a_2 = 2$ and $a_1 = -1$. Recall that our A matrix in CCF will be

$$A_{\text{CCF}} = \begin{bmatrix} 0 & 1 \\ a_1 & a_2 \end{bmatrix} \tag{47}$$

$$= \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix} \tag{48}$$

and

$$\vec{b}_{\text{CCF}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{49}$$

by the definition of CCF. Hence, $\mathcal{C}_{\text{CCF}}$ can be computed as follows:

$$\vec{b}_{\text{CCF}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{50}$$

$$A_{\text{CCF}}\vec{b}_{\text{CCF}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \tag{51}$$

so $\mathcal{C}_{\text{CCF}} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$. Thus,

$$T = \mathcal{C}\mathcal{C}_{\text{CCF}}^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \tag{52}$$

5. QR System ID

(a) Suppose we are given the following discrete time dynamical system:

$$x[i+1] = ax[i] + b_1u_1[i] + b_2u_2[i] + \cdots + b_{n-1}u_{n-1}[i] \tag{53}$$

We would like to estimate $a, b_1, b_2, \ldots, b_{n-1}$ using system ID. Suppose we have collected data up to $x[m]$, where $m < n$. **Set up a linear system of the form $D\vec{p} = \vec{s}$ to solve this system ID problem. Show that D has dimensions $m \times n$.**

Solution: We can set up our system ID problem as

$$\underbrace{\begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[m] \end{bmatrix}}_{\vec{s}} = \underbrace{\begin{bmatrix} x[0] & u_1[0] & u_2[0] & \cdots & u_{n-1}[0] \\ x[1] & u_1[1] & u_2[1] & \cdots & u_{n-1}[1] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[m-1] & u_1[m-1] & u_2[m-1] & \cdots & u_{n-1}[m-1] \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} a \\ b_1 \\ b_2 \\ \vdots \\ b_{n-1} \end{bmatrix}}_{\vec{p}} \tag{54}$$

Here, D has m rows and n columns, so $D \in \mathbb{R}^{m \times n}$.

(b) As we saw in the previous part, we have a wide matrix D. Assuming that D is rank m, we would technically have infinitely many solutions for $a, b_1, b_2, \ldots, b_{n-1}$. We can find the solution with the smallest norm using QR decomposition.

We can write $D^\top = \begin{bmatrix} \vec{d}_1 & \vec{d}_2 & \cdots & \vec{d}_m \end{bmatrix}$ where each $\vec{d}_i \in \mathbb{R}^n$. We can also define an orthonormal matrix $Q \in \mathbb{R}^{n \times n}$ which can be written as $Q = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_m & \vec{q}_{m+1} & \vec{q}_{m+2} & \cdots & \vec{q}_n \end{bmatrix}$, where $\text{Span}(\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_m) = \text{Span}\left(\vec{d}_1, \vec{d}_2, \ldots, \vec{d}_m\right)$. **In this case, what is $\vec{d}_j^\top \vec{q}_i$ for $j \in \{1, \ldots, m\}$ and $i \in \{m+1, \ldots, n\}$? Explain your answer.**

(HINT: If we say that $\text{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) = \text{Span}(\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n)$, then we may say that $\vec{v}_i$ can be written as a linear combination of $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ (and equivalently, $\vec{u}_i$ can be written as a linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$).)

Solution: Each $\vec{d}_j$ can be written as a linear combination of $\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_m$ since their spans are equal. More concretely,

$$\vec{d}_j = \alpha_1\vec{q}_1 + \alpha_2\vec{q}_2 + \cdots + \alpha_m\vec{q}_m \tag{55}$$

Hence, we have that

$$\vec{d}_j^\top \vec{q}_i = (\alpha_1\vec{q}_1 + \alpha_2\vec{q}_2 + \cdots + \alpha_m\vec{q}_m)^\top \vec{q}_i \tag{56}$$

$$= \alpha_1\vec{q}_1^\top \vec{q}_i + \alpha_2\vec{q}_2^\top \vec{q}_i + \cdots + \alpha_m\vec{q}_m^\top \vec{q}_i \tag{57}$$

Since $i \in \{m+1, \ldots, n\}$, we have that $\vec{q}_1^\top \vec{q}_i = 0, \vec{q}_2^\top \vec{q}_i = 0, \ldots, \vec{q}_m^\top \vec{q}_i = 0$, so $\vec{d}_j^\top \vec{q}_i = 0$ for $j \in \{1, \ldots, m\}$ and $i \in \{m+1, \ldots, n\}$.

(c) Suppose that $D^\top$ can be written as

$$D^\top = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_m \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ 0 & r_{22} & \cdots & r_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{mm} \end{bmatrix} \tag{58}$$

Using this result and the result from the previous part, show that the QR decomposition of $D^\top$ can be written as

$$D^\top = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0_{(n-m)\times m} \end{bmatrix} \tag{59}$$

Using eq. (59), write an expression for $Q_1^\top \vec{p}$ where $D\vec{p} = \vec{s}$, and show that the value of $Q_2^\top \vec{p}$ does not matter. Here, $R_1 \in \mathbb{R}^{m\times m}$ is a square, upper triangular matrix, $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \in \mathbb{R}^{n\times n}$ is an orthonormal matrix, and $0_{(n-m)\times m} \in \mathbb{R}^{(n-m)\times m}$ is a matrix of all zeros. Q_1 is $n \times m$ and Q_2 is $n \times (n-m)$. Note that R_1 is invertible.

(HINT: Equation (59) uses block matrix form. When multiplying block matrices, they obey the same rules as regular matrix-vector multiplication. That is, $\begin{bmatrix} M & N \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = MA + NB$. When transposing block matrices, we may write $\begin{bmatrix} A \\ B \end{bmatrix}^\top = \begin{bmatrix} A^\top & B^\top \end{bmatrix}$.) (HINT: First, simplify eq. (59) using the previous hint. Then, use the previous problem to find a potential candidate for Q_2. Use the previous part again to confirm that this candidate would work by computing R_{ij} using the formula provided in lecture (for $j \in \{1, \dots, m\}$ and $i \in \{m+1, \dots, n\}$).)

Solution: Simplifying eq. (59), we have

$$D^\top = Q_1 R_1 + Q_2 0_{(n-m)\times m} = Q_1 R_1 \tag{60}$$

From eq. (58), we can set

$$Q_1 = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_m \end{bmatrix} \tag{61}$$

$$R_1 = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ 0 & r_{22} & \cdots & r_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{mm} \end{bmatrix} \tag{62}$$

For Q_2, we can set it as

$$Q_2 = \begin{bmatrix} \vec{q}_{m+1} & \vec{q}_{m+2} & \cdots & \vec{q}_n \end{bmatrix} \tag{63}$$

which would mean that $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix} = Q$ is an orthonormal matrix. However, we need to check that $R = \begin{bmatrix} R_1 \\ 0_{(n-m)\times m} \end{bmatrix}$ is a valid upper triangular matrix for the QR decomposition.

From lecture, we know that $R_{ij} = \vec{d}_j^\top \vec{q}_i$. For $j \in \{1, \dots, m\}$ and $i \in \{1, \dots, m\}$, we have that $R_{ij} = r_{ij} = \vec{d}_j^\top \vec{q}_i$ from eq. (58). For $j \in \{1, \dots, m\}$ and $i \in \{m+1, \dots, n\}$, we know that $R_{ij} = 0$ from the previous part. Hence, our R matrix is valid.

To find an expression for $Q_1^\top \vec{p}$, we can use the fact that $D^\top = QR$ to write $D = R^\top Q^\top$. Hence,

$$R^\top Q^\top \vec{p} = \vec{s} \tag{64}$$

$$\begin{bmatrix} R_1^\top & 0_{(n-m)\times m}^\top \end{bmatrix} \begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix} \vec{p} = \vec{s} \tag{65}$$

$$R_1^\top Q_1^\top \vec{p} + 0_{(n-m)\times m}^\top Q_2^\top \vec{p} = \vec{s} \tag{66}$$

$$R_1^\top Q_1^\top \vec{p} = \vec{s} \tag{67}$$

$$Q_1^\top \vec{p} = \left(R_1^\top\right)^{-1} \vec{s} \tag{68}$$

Note that in eq. (66) we have $0^\top_{(n-m)\times m}$ multiplying $Q_2^\top \vec{p}$, so it does not matter what $Q_2^\top \vec{p}$ is (it can be a vector with any numbers, but since it multiplies with a 0 matrix the result will always be $\vec{0}$).

(d) From the previous part, we determined that the value of $Q_2^\top \vec{p}$ did not matter. Hence, we can set $Q_2^\top \vec{p} = \vec{0}$ for the purposes of minimizing $\|\vec{p}\|$ (the reason why we do this will be covered a little bit later, but take this as a given for now). **Solve for $\vec{p}$ using the QR decomposition of $D^\top$, assuming $Q_2^\top \vec{p} = \vec{0}$.** *(HINT: The following identity holds true: $\begin{bmatrix} A\vec{x} \\ B\vec{x} \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} \vec{x}$.) (HINT: Stack the two expressions for $Q_1^\top \vec{p}$ and $Q_2^\top \vec{p}$ to obtain an expression for $\begin{bmatrix} Q_1^\top \vec{p} \\ Q_2^\top \vec{p} \end{bmatrix}$. Use the previous hint to determine your final expression for $\vec{p}$.)*

Solution: Following the hint, we have $Q_1^\top \vec{p} = \left(R_1^\top\right)^{-1} \vec{s}$ from the previous part, and $Q_2^\top \vec{p} = \vec{0}$. Thus,

$$\begin{bmatrix} Q_1^\top \vec{p} \\ Q_2^\top \vec{p} \end{bmatrix} = \begin{bmatrix} \left(R_1^\top\right)^{-1} \vec{s} \\ \vec{0} \end{bmatrix} \tag{69}$$

$$\begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix} \vec{p} = \begin{bmatrix} \left(R_1^\top\right)^{-1} \vec{s} \\ \vec{0} \end{bmatrix} \tag{70}$$

$$Q^\top \vec{p} = \begin{bmatrix} \left(R_1^\top\right)^{-1} \vec{s} \\ \vec{0} \end{bmatrix} \tag{71}$$

$$\vec{p} = Q \begin{bmatrix} \left(R_1^\top\right)^{-1} \vec{s} \\ \vec{0} \end{bmatrix} \tag{72}$$

Contributors:

- Anish Muthali.