CSM
Spring 2022

EECS 16B
Worksheet 7

1. Controlling a System

We are given a discrete time state space system, where $\vec{x}$ is our **state vector**, A is the **state space model** of the system, B is the **input matrix**, and $\vec{u}$ is the **control input vector**.

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

We define a system to be controllable if: **given a set of inputs, we can move the system from any initial state to any final state by choosing a sequence of inputs.** This has an important physical meaning; if a physical system is controllable, that means that we can get anywhere in the state space. For example, if a robot is controllable, it is able to travel anywhere in the system it is living in (given enough time and control inputs).

We will start with the assumption that the system started at rest at the zero vector: $\vec{x}[0] = \vec{0}$, and our state space will be an n-dimensional vector in $\mathbb{R}^n$.

(a) **How can you write out $\vec{x}[1]$ using the state space equation? How about $\vec{x}[2]$?**

Solution: Using the state space model $\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i]$,

$$\vec{x}[1] = A\vec{x}[0] + B\vec{u}[0] = A\vec{0} + B\vec{u}[0] = B\vec{u}[0]$$

At time step $t = 2$,

$$\vec{x}[2] = A\vec{x}[1] + B\vec{u}[1] = A(B\vec{u}[0]) + B\vec{u}[1]$$

(b) **Given these two observations and the initial condition $\vec{x}[0] = \vec{0}$, where can $\vec{x}$ reach in our state space reach after 2 time steps (i.e. at $t = 2$)?** **Solution:** From the previous part, we know that

$$\vec{x}[2] = A(B\vec{u}[0]) + B\vec{u}[1] = AB\vec{u}[0] + B\vec{u}[1]$$

Let's say B has two columns, although we can generalize this for any number of columns.

We can then equivalently write out our equation as:

$$\vec{x}[2] = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 \end{bmatrix} \begin{bmatrix} u_1[0] \\ u_2[0] \end{bmatrix} + \begin{bmatrix} \vec{b}_1 & \vec{b}_2 \end{bmatrix} \begin{bmatrix} u_1[1] \\ u_2[1] \end{bmatrix} = u_1[0]A\vec{b}_1 + u_2[0]A\vec{b}_2 + u_1[1]\vec{b}_1 + u_2[1]\vec{b}_2.$$

Since $\vec{u}$ is a vector we have full control over, we can reach anywhere in:

$$\text{span}(A\vec{b}_1, A\vec{b}_2, \vec{b}_1, \vec{b}_2)$$

In general, we could reach anywhere in the span of the columns of B and AB.

(c) **Given n observations, where in our state space can $\vec{x}$ reach at time $t = n$?** **Solution:** We will start off in a similar manner by writing out $\vec{x}(3)$ in terms of $\vec{x}[2]$ and $\vec{u}[2]$.

$$\vec{x}(3) = A\vec{x}[2] + B\vec{u}[2] = A(AB\vec{u}[0] + B\vec{u}[1]) + B\vec{u}[2] = A^2B\vec{u}[0] + AB\vec{u}[1] + B\vec{u}[2].$$

Using a similar argument as the previous part, we will see that we can reach anywhere in the span of the columns of B, AB, A^2B. We can continue doing this, and see that after n time steps,

$$\begin{aligned}\vec{x}[i] &= A\vec{x}[i-1] + B\vec{u}[i-1] \\ &= A^{n-1}B\vec{u}[0] + A^{n-2}B\vec{u}[1] + A^{n-3}B\vec{u}[2] + \cdots + AB\vec{u}[i-2] + B\vec{u}[i-1]\end{aligned}$$

Therefore, after n timesteps, we can reach anywhere in the span of the columns of $\{B, AB, A^2B, \cdots, A^{n-1}B\}$.

(d) **Show that if the columns of A^kB are linearly dependent, then the columns of $A^{k+1}B$ are also linearly dependent.** **Solution:** We first make the observation that the columns of A^kB are:

$$A^kB = \begin{bmatrix} A^k\vec{b}_1 & A^k\vec{b}_2 & \cdots & A^k\vec{b}_m \end{bmatrix}$$

Where $\vec{b}_i$ is the i^{th} column of the matrix B.

Now let's suppose that the columns of A^kB are linearly dependent. We can say that if we have scalars α_i such that

$$\alpha_1 A^k\vec{b}_1 + \cdots + \alpha_m A^k\vec{b}_m = \vec{0},$$

then at least one of the α_i must be nonzero. Now if we left multiply by A, we see that:

$$A(\alpha_1 A^k\vec{b}_1 + \cdots + \alpha_m A^k\vec{b}_m) = \alpha_1 A^{k+1}\vec{b}_1 + \cdots + \alpha_m A^{k+1}\vec{b}_m = \vec{0}.$$

As a result, we observe that if we have a linear combination of the columns of $A^{k+1}B$ equal to the zero vector, at least one of the scalars is nonzero. This shows that the columns of $A^{k+1}B$ must be linearly dependent.

(e) To summarize our work from the previous parts, we now define a controllability matrix

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

Based on the previous parts, when can we say our system is controllable? (i.e. we can reach anywhere in our state space using our system)

Solution: Using our knowledge from part (c), we can say that $\vec{x}$ can reach anywhere in the span of the columns of $\mathscr{C}$ after n time steps. We also saw in part (d), that if A^kB had linearly dependent columns, then $A^{k+1}B$ will also have linearly dependent columns.

This means if one of our measurements at time $t = k$, was linearly dependent from our previous measurements, then the future measurements $t = k+1$ onward, will also be linearly dependent from the previous ones. In other words, every redundant measurement taken after a redundant one will continue to be redundant.

We must take at least n measurements, since we want to span all of $\mathbb{R}^n$, and B may just be a single vector. Therefore, we conclude by saying that our system can reach anywhere in our state-space, $\mathbb{R}^n$, if $\mathscr{C}$ is a matrix of rank n.

This does not mean that $\mathscr{C}$ has to be invertible, it just means that $\mathscr{C}$ must have n linearly independent columns for the system to be controllable.

2. Gram-Schmidt Process

A set of vectors is orthonormal **if and only if**:

- Any pair of vectors are orthogonal (for any vectors $\vec{u}, \vec{v} \in S$ where $\vec{u} \neq \vec{v}$, the angle between them is $90°$).
- Each vector is of unit length (for vector $\vec{v} \in S$, $\|\vec{v}\| = 1$).

Gram-Schmidt orthonormalization is a process that allows us to take such a set S of arbitrary, linearly-independent vectors, and create a new set S' of vectors that span the exact same space, yet are now orthonormal.

(a) For linearly independent vectors $\vec{v}_1, \vec{v}_2$, **justify that Span($\vec{v}_1, \vec{v}_2$) is the same as Span($\vec{v}_1, \vec{v}_2 - \alpha\vec{v}_1$).**

Solution: The definition of span is all vectors that can be "reached" using a linear combination of a set of vectors. So,

$$span(\vec{v}_1, \vec{v}_2) = \vec{v}_i = \beta_1\vec{v}_1 + \beta_2\vec{v}_2 \text{ for arbitrary } \vec{v}_i \text{ and all real } \beta_1, \beta_2$$

Following a similar train of thought,

$$\begin{aligned} \text{span}(\vec{v}_1, \vec{v}_2 - \alpha\vec{v}_1) &= \beta_3\vec{v}_1 + \beta_4(\vec{v}_2 - \alpha\vec{v}_1) \\ &= \beta_3\vec{v}_1 + \beta_4\vec{v}_2 - \beta_4\alpha\vec{v}_1 \\ &= (\beta_3 - \beta_4\alpha)\vec{v}_1 + \beta_4\vec{v}_2 \end{aligned}$$

All α and β are arbitrary constants that we can set. If we set $\beta_4 = \beta_2$ and $\beta_3 = \beta_1 + \alpha\beta_2$, then these spans are the same!

(b) Let us iteratively generate the orthonormal set S' from S. Start by arbitrarily picking a vector $\vec{v}_1$ from S: **how can we create a new vector $\vec{u}_1$ for S' so that it spans the same space as the original vector, yet ensures that orthonormality is preserved for our new set?**

Solution: Since we are only taking into consideration a single vector, we do not need to worry about orthogonality. So,

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_2\|}$$

(c) Now, consider the next vector $\vec{v}_2$ in S. **How can we convert $\vec{v}_2$ to some $\vec{u}_2$ that is orthonormal with respect to the $\vec{u}_1$ that we've already added to S'?**

Solution: Now, we do need to consider orthogonality. In order for two vectors $\vec{u}, \vec{v}$ to be orthogonal, their dot product has to be 0. In other words, this means that there is nothing–no trace, no component–of $\vec{u}$ in $\vec{v}$, and vice versa. We can do this by defining the following where:

$$\vec{p}_2 = \vec{v}_2 - (\vec{v}_2{}^T\vec{u}_1)\vec{u}_1$$

By subtracting the projection of $\vec{v}_2$ onto $\vec{u}_1$, we are removing the $\vec{u}_1$ component of $\vec{v}_2$ from $\vec{v}_2$. So, $\vec{p}_2$ is now orthogonal to $\vec{u}_1$, and mutual orthogonality is preserved for all vectors in our new set S'. $\vec{p}_2$ also needs to be normalized, so:

$$\vec{w}_2 = \frac{\vec{p}_2}{|\vec{p}_2|}$$

(d) **Write out the process for creating an arbitrary vector $\vec{u}_i$ for our orthonormal set S'.**

Solution: Define the following vector:

$$\vec{p}_i = \vec{v}_i - \sum_{j=1}^{i-1} (\vec{v}_i{}^T \vec{w}_j)\vec{w}_j$$

By subtracting the component of every other vector already in S', this ensures that our new vector $\vec{w}_i$ is orthogonal. Then, we normalize this:

$$\vec{w}_i = \frac{\vec{p}_i}{|\vec{p}_i|}$$

Rinse and repeat for all vectors in S to generate a new, orthonormal set S' that we can use in place of S.

(e) Let $A = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix}$.

We will use Gram-Schmidt to find an orthonormal basis $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \end{bmatrix}$ for the $\mathrm{Col}(A)$.

i. **Find $\vec{u}_1$.**

Solution: For the first vector, we do not need to worry about orthogonality. Therefore we just have to normalize.

$$\vec{u}_1 = \frac{\vec{a}_1}{\|\vec{a}_1\|} = \frac{1}{\sqrt{4}} \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

ii. **Find $\vec{u}_2$.**

Solution: We first compute $\vec{q}_2$ as a vector orthogonal to $\vec{u}_1$ by subtracting the projection of $\vec{a}_2$ onto $\vec{u}_1$

$$\vec{q}_2 = \vec{a}_2 - \langle \vec{a}_2, \vec{u}_1 \rangle \vec{u}_1 = \vec{a}_2 - \frac{1}{2}\langle \vec{a}_2, \vec{a}_1 \rangle \frac{1}{2}\vec{a}_1 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 3 \end{bmatrix} - \frac{8}{4} \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

If we were to normalize this vector, we would get:

$$\vec{u}_2 = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

iii. **Find $\vec{u}_3$.**

Solution: We again compute $\vec{q}_3$ as a vector orthogonal to rest by subtracting its projections onto

$\vec{a}_1$ and $\vec{a}_2$.

$$\vec{q}_3 = \vec{a}_3 - \langle \vec{a}_3, \vec{u}_1 \rangle \vec{u}_1 - \langle \vec{a}_3, \vec{u}_2 \rangle \vec{u}_2 = \vec{a}_3 - \frac{1}{4}\langle \vec{a}_3, \vec{a}_1 \rangle \vec{a}_1 - \frac{1}{4}\langle \vec{a}_3, \vec{q}_2 \rangle \vec{q}_2$$
$$= \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} - \frac{4}{4}\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} - \frac{16}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \end{bmatrix}$$

Normalizing $\vec{q}_3$, we get:

$$\vec{u}_3 = \frac{1}{\sqrt{16}}\begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

3. Projection Matrices

We know from Gram-Schmidt, that if we have a basis $S = \{\vec{v}_1, \cdots \vec{v}_n\}$, for a vector space V, then we could pick an orthonormal basis $U = \{\vec{u}_1, \cdots \vec{u}_n\}$ for the same vector space, since we showed that the span of U was equivalent to the span of S, or the vector space V.

In this question, we will look into a special class of matrices called **projection matrices.** A projection matrix is a matrix P such that $P^2 = P$. Intuitively, if we projected a vector onto some subspace, then taking any vector in that subspace and reprojecting it will give the same vector.

(a) Let visit the least squares problem from 16A. Suppose we have some equation $A\vec{x} = \vec{b}$ that has no solution. Why does it have no solution? Because A has more rows than columns; it is a "tall" matrix where we have more equations than unknowns. How can we estimate the solution $\hat{x}$ using Least Squares?

Solution: Remember that we can set up the following least squares problem:

$$\min \|\vec{r}\| = \min \left\|\vec{b} - A\vec{x}\right\|$$

We use a geometric argument to show that the residual $\vec{r}$ is minimized, when it is orthogonal to $\text{Col}(A)$. This means that the residual $\vec{r}$ is in $\text{Nul}(A^T)$ from the Fundamental Theorem of Linear Algebra. Therefore, we multiply both sides by A^T to get:

$$A^T A\vec{x} = A^T\vec{b} + A^T\vec{r} = A^T\vec{b}$$

$A^T A$ is a square matrix that is invertible, when A is of full rank. Therefore, we multiply both sides by the inverse to get

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

This was the formula for least-squares from 16A.

(b) We can actually view the least squares problem as projection of the vector $\vec{b}$ onto the $\text{Col}(A)$. What would the projection matrix P that projects $\vec{b}$ onto $\text{Col}(A)$ be? Verify that it indeed is a projection matrix.

Solution: We are projecting a vector $\vec{b}$ and the result of the projection is $A\hat{x} = A(A^T A)^{-1} A^T \vec{b}$. Therefore, we should suspect that $P = A(A^T A)^{-1} A^T$ is our projection matrix. We can indeed verify that is a projection matrix by computing its square:

$$P^2 = A(A^T A)^{-1} A^T A (A^T A)^{-1} A^T = A(A^T A)^{-1} A^T = P$$

(c) From the result above, we see that we have to invert $A^T A$ every time. This takes a lot of computational power and is rather slow. If A were a matrix with orthonormal columns, how could this help us?

Solution: Let A be orthonormal. If this were the case, then $A^T A = I$. Recall that $I^{-1} = I$. This means we don't have to do any work for inverting this matrix, and the least-squares formula reduces to $\hat{x} = A^T \vec{b}$.

(d) We would like A to be orthonormal. We cannot, however, arbitrarily replace A with an orthonormal matrix Q. What must be true about Q for the least-squares equation to hold?

Solution: Both A and Q must span the same space. Geometrically, least-squares projects $\vec{b}$ onto the column space of A, so if we want the projection to hold when using Q, it must project to the same spot on the same column space.

(e) We will create this orthonormal matrix Q by considering the columns of A as a basis for $\text{Col}(A)$ and performing Gram-Schmdit on these basis vectors. We can then show that the projection of $\vec{b}$ onto $\text{Col}(A)$ is the same as projecting it onto $\text{Col}(Q)$ through the Orthogonal Decomposition Theorem. That is, $A(A^TA)^{-1}A^T\vec{b} = Q(Q^TQ)^{-1}Q^T\vec{b}$. What is the least square solution $\hat{x}$ in terms of Q?

Solution:

$$\hat{x} = (Q^TQ)^{-1}Q^T\vec{b} = IQ^T\vec{b} = Q^T\vec{b}$$

