EECS 16B Designing Information Devices and Systems II
Spring 2021 Discussion Worksheet

Discussion 10A

1. Towards Upper-Triangulation By An Orthonormal Basis

Answer: In lecture, we have been motivated by the goal of getting to a coordinate system in which the eigenvalues of a matrix representing a linear operation are on the diagonal. When this is done to the A matrix representing a time-evolving system (whether in continuous-time as a system of differential equations, or in discrete-time as a relationship between the next state and the previous one), we can view the system as a cascade of scalar systems — with each one potentially being an input to the ones that come "after" or "above" it. We saw this in lecture, but it is good to spend more time to really understand this argument.

Note that in the next homework, you will be asked to derive this in a more formal way. Here we will just provide some key steps along the way to a recursive understanding. Here, as in lecture, we will restrict attention to matrices that have all real eigenvalues.

Previously in this course, we have seen the value of changing our coordinates to be eigenbasis-aligned, because we can then view the system as a cascade of scalar systems. If we have a diagonalization, then these scalar equations are fully uncoupled, and can therefore be treated completely separately. But even when we cannot diagonalize, we can upper-triangularize in a way that allows us to solve the equations one at a time, from the "bottom up".

In this problem, to better understand the steps involved, we will use the following concrete example:

$$S_{[3\times 3]} = \begin{bmatrix} \frac{5}{12} & \frac{5}{12} & \frac{1}{6} \\ \frac{5}{12} & \frac{5}{12} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \tag{1}$$

and figure out the general case by abstracting variables. Note that there is a datahub link to a jupyter notebook on the website, which will allow you to perform the numerical calculations quickly to connect the symbolic analysis to an example, but without being time-consuming in the process.[1]

(a) Consider a non-zero vector $\vec{u}_0 \in \mathbb{R}^n$. Can you think of a way to extend it to a set of basis vectors for $\mathbb{R}^n$? In other words, find $\vec{u}_1, \cdots, \vec{u}_{n-1}$, such that $\text{span}(\vec{u}_0, \vec{u}_1, \cdots, \vec{u}_{n-1}) = \mathbb{R}^n$. To begin with, consider $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$. Can you get an orthonormal basis from what you just constructed?

Hint: what was the last discussion all about? Also, the given vector isn't normalized yet!

Answer: Starting with the provided vector, we can add all the vectors from the standard basis (here, since we're in $\mathbb{R}^3$, we will add the $\mathbb{R}^3$ basis vectors). By doing this, we guarantee that the matrix spans $\mathbb{R}^n$ (since the 3 vectors alone that we just added span $\mathbb{R}^3$, and the initial vector can be treated as "extra" for now.) Of course, for a basis, we ultimately need a *minimal* set of spanning vectors, so we will only have 3 vectors in the end.

[1] This particular matrix has an additional special property of symmetry, but we won't be invoking that here.

For $\begin{bmatrix}1 & -1 & 0\end{bmatrix}^\top$, we can form:

$$\begin{bmatrix} | & | & | & | \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Then using this matrix (considering additional vectors one by one), we can run Gram-Schmidt (which was covered in great detail in the previous discussion) to convert this matrix to an orthonormal basis. Note that since we have 4 vectors but only need 3, we will end up having to throw one out. But, Gram-Schmidt will tell us when to do this! If we ever see a zero along the way, then we discard that vector and move on. Recall that if we see a zero, this indicates that the previous vectors already in our "final" set of basis vectors are linearly dependent with the current vector (so adding the current vector doesn't add any dimensions to our span). Succinctly, the projection of the current vector on the space spanned by the existing vectors is 0.

The key is that we are guaranteed to span the whole space by the end because the standard basis spans the whole space. We can use the Gram-Schmidt process for the basis obtained above, starting with $\vec{v}_1$ (using the same notation as dis09B):

$$\vec{q}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}$$

$$\implies \vec{z}_2 = \vec{v}_2 - \left(\vec{v}_2^\top \vec{q}_1\right)\vec{q}_1$$

$$= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \left(\begin{bmatrix}1 & 0 & 0\end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}\right)\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{\sqrt{2}}{2}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}$$

$$\implies \vec{q}_2 = \frac{\vec{z}_2}{\|\vec{z}_2\|} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}$$

$$\vec{z}_3 = \vec{v}_3 - \left(\vec{v}_3^\top \vec{q}_1\right)\vec{q}_1 - \left(\vec{v}_3^\top \vec{q}_2\right)\vec{q}_2 = 0 \qquad \text{(unused in final basis)}$$

$$\vec{z}_4 = \vec{v}_4 - \left(\vec{v}_4^\top \vec{q}_1\right)\vec{q}_1 - \left(\vec{v}_4^\top \vec{q}_2\right)\vec{q}_2$$

$$\implies \vec{q}_3 = \frac{\vec{z}_4}{\|\vec{z}_4\|}$$

Once we carry out this procedure, we find that our representative orthonormal matrix is:

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(b) Now consider a real eigenvalue λ_1, and the corresponding (normalized) eigenvector $\vec{v}_1 \in \mathbb{R}^n$ of a square matrix $M \in \mathbb{R}^{n\times n}$. From the previous part, we know that we can extend $\vec{v}_1$ to an orthonormal basis of $\mathbb{R}^n$. We will denote the basis by

$$U = \begin{bmatrix} | & | & \cdots & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \\ | & | & \cdots & | \end{bmatrix}$$

where $\vec{u}_1 = \vec{v}_1$ (note that this eigenvector is already normalized).

Our goal is to look at what the matrix M looks like in the coordinate system defined by the basis U. Compute $U^\top MU$ by writing $U = \begin{bmatrix} \vec{v}_1 & R \end{bmatrix}$, where $R \triangleq \begin{bmatrix} | & | & \cdots & | \\ \vec{r}_1 & \vec{r}_2 & \cdots & \vec{r}_{n-1} \\ | & | & \cdots & | \end{bmatrix}$.

Answer: Symbolic analysis:

$$U^\top MU = \begin{bmatrix} \vec{v}_1^\top \\ R^\top \end{bmatrix} M \begin{bmatrix} \vec{v}_1 & R \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} \vec{v}_1^\top \\ R^\top \end{bmatrix} \begin{bmatrix} M\vec{v}_1 & MR \end{bmatrix} \tag{3}$$

$$= \begin{bmatrix} \vec{v}_1^\top \\ R^\top \end{bmatrix} \begin{bmatrix} \lambda_1\vec{v}_1 & MR \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} \lambda_1\vec{v}_1^\top\vec{v}_1 & \vec{v}_1^\top MR \\ \lambda_1 R^\top\vec{v}_1 & R^\top MR \end{bmatrix}. \tag{5}$$

Concrete case: $S_{[3\times 3]}$ has zero as eigenvalue since it contains a repeated column vector. So, we let the corresponding eigenvector (which can be anything) be just $\begin{bmatrix} 1 & -1 & 0 \end{bmatrix}^\top$, the starting vector from the previous subppart. Then, we have:

$$U = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & 0 \\ 0 & 1 \end{bmatrix}$$

Performing the matrix multiplication yields:

$$U^\top S_{[3\times 3]} U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{5}{6} & \frac{\sqrt{2}}{6} \\ 0 & \frac{\sqrt{2}}{6} & \frac{2}{3} \end{bmatrix}$$

From here, we can form a connection to the result of a couple subparts later, seeing that:

$$Q = R^\top S_{[3\times 3]} R = \begin{bmatrix} \frac{5}{6} & \frac{\sqrt{2}}{6} \\ \frac{\sqrt{2}}{6} & \frac{2}{3} \end{bmatrix}$$

(c) Show that $U^{-1} = U^\top$.

Answer: One way to reason through this proof is with definitions and properties. U is an orthonormal basis by our construction. $U^\top U$ performs a dot product between each of the basis vectors. Since these basis vectors are orthogonal to each other (we performed Gram-Schmidt to make them this way!), all the non-diagonal elements have to be 0. Since the basis vectors are normalized, the dot product with itself is 1. As a result, $U^\top U = I$, and $U^{-1} = U^\top$.

Also, this result was shown rigorously in lec9A. We outline the same approach below in the 3×3 case. Suppose we have an orthonormal matrix P:

$$P = \begin{bmatrix} | & | & | \\ \vec{p}_1 & \vec{p}_2 & \vec{p}_3 \\ | & | & | \end{bmatrix}$$

We can compute $P^\top P$. We use the fact that for a set of mutually orthonormal vectors, the dot product of a vector with any *other* vector in the set is 0, but the dot product of a vector with itself is 1:

$$\begin{aligned} P^\top P &= \begin{bmatrix} - & \vec{p}_1^\top & - \\ - & \vec{p}_2^\top & - \\ - & \vec{p}_3^\top & - \end{bmatrix} \begin{bmatrix} | & | & | \\ \vec{p}_1 & \vec{p}_2 & \vec{p}_3 \\ | & | & | \end{bmatrix} \\ &= \begin{bmatrix} \vec{p}_1^\top \vec{p}_1 & \vec{p}_1^\top \vec{p}_2 & \vec{p}_1^\top \vec{p}_3 \\ \vec{p}_2^\top \vec{p}_1 & \vec{p}_2^\top \vec{p}_2 & \vec{p}_2^\top \vec{p}_3 \\ \vec{p}_3^\top \vec{p}_1 & \vec{p}_3^\top \vec{p}_2 & \vec{p}_3^\top \vec{p}_3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= I \end{aligned}$$

This shows that $P^\top = P^{-1}$.

(d) Define $Q = R^\top M R$. Look at the first column and the first row of $U^\top M U$ and show that:

$$M = U \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix} U^\top$$

Here, $\vec{a}$ is a symbolic vector related to M, R, and $\vec{v}_1$ (we will show the relation!).

Answer: We found above that:

$$U^\top M U = \begin{bmatrix} \lambda_1 \vec{v}_1^\top \vec{v}_1 & \vec{v}_1^\top M R \\ \lambda_1 R^\top \vec{v}_1 & R^\top M R \end{bmatrix}$$

We can rearrange the given equation for M into the form of the above, since $U^\top = U^{-1}$:

$$M = U \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix} U^\top \implies U^\top M U = \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix}$$

Now, we need to show why:

$$\begin{bmatrix} \lambda_1 \vec{v}_1^\top \vec{v}_1 & \vec{v}_1^\top M R \\ \lambda_1 R^\top \vec{v}_1 & R^\top M R \end{bmatrix} \stackrel{?}{=} \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix}$$

We start simplifying the left side. First, we observe that $\vec{v}_1^\top \vec{v}_1 = 1 \implies \lambda_1 \vec{v}_1^\top \vec{v}_1 = \lambda_1$, because $\vec{v}_1 = \vec{u}_1$ and the set of vectors $\vec{u}$ is orthonormal by construction.

Next, $\lambda_1 R^\top \vec{v}_1 = \vec{0}$ because R consists of all of the other $\vec{u}_i$ vectors that compose our orthonormal basis; taking the dot product between any one of these and $\vec{v}_1$ yields zero (same logic as outlined in the previous part for vectors that are mutually orthogonal).

$\vec{v}_1^\top MR$ currently takes the place of $\vec{a}^\top$, suggesting that $\vec{a} = \left(\vec{v}_1^\top MR\right)^\top = R^\top M^\top \vec{v}_1$. So, finally substituting that $Q = R^\top MR$, we have:

$$U^\top MU = \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix}$$

In the numerical example with $S_{[3\times 3]}$, we have:

$$Q = R^\top S_{[3\times 3]} R = \begin{bmatrix} \frac{5}{6} & \frac{\sqrt{2}}{6} \\ \frac{\sqrt{2}}{6} & \frac{2}{3} \end{bmatrix}$$

(e) Now, we can recurse on Q to get:

$$Q = \begin{bmatrix} \vec{v}_2 & Y \end{bmatrix} \begin{bmatrix} \lambda_2 & \vec{b}^\top \\ \vec{0} & P \end{bmatrix} \begin{bmatrix} \vec{v}_2 & Y \end{bmatrix}^\top$$

where we have taken $\vec{v}_2 \in \mathbb{R}^{n-1}$, a normalized eigenvector of Q, associated with eigenvalue λ_2. Again $\vec{v}_2$ is extended into an orthonormal basis to form $\begin{bmatrix} \vec{v}_2 & Y \end{bmatrix}$.

Plug this form of Q into M above, to show that:

$$M = \begin{bmatrix} \vec{v}_1 & R\vec{v}_2 & RY \end{bmatrix} \begin{bmatrix} \lambda_1 & \breve{a}_1 & \breve{\vec{a}}_{\text{rest}}^\top \\ 0 & \lambda_2 & \vec{b}^\top \\ \vec{0} & \vec{0} & P \end{bmatrix} \begin{bmatrix} \vec{v}_1 & R\vec{v}_2 & RY \end{bmatrix}^\top$$

where we define $\breve{\vec{a}}$ to be the "adjusted" $\vec{a}$ to account for the subsitution of Q; $\breve{\vec{a}}^\top = \vec{a}^\top \begin{bmatrix} \vec{v}_2 & Y \end{bmatrix}$.

Answer: From part (d), we know that

$$M = U \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix} U^\top$$

and that:

$$U = \begin{bmatrix} \vec{v}_1 & R \end{bmatrix}$$

So we get that,

$$M = U \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix} U^\top \tag{6}$$

$$= \begin{bmatrix} \vec{v}_1 & R \end{bmatrix} \begin{bmatrix} \lambda_1 & \vec{a}^\top \\ \vec{0} & Q \end{bmatrix} \begin{bmatrix} \vec{v}_1^\top \\ R^\top \end{bmatrix} \tag{7}$$

In the given definition of Q, let's denote $\begin{bmatrix}\vec{v}_2 & Y\end{bmatrix}$ as U_2, since this is the orthonormal basis that upper triangularizes Q (note the middle matrix of Q, which we can call T_2, is block upper-triangular). We can then write that:

$$Q = U_2 \underbrace{\begin{bmatrix}\lambda_2 & \vec{b}^\top \\ \vec{0} & P\end{bmatrix}}_{T_2} U_2^\top \tag{8}$$

We had an expression for Q previously; $R^\top MR$. We can equate the two representations and simplify:

$$U_2 T_2 U_2^\top = R^\top MR \tag{9}$$
$$T_2 = U_2^\top R^\top MRU_2 \tag{10}$$
$$= (RU_2)^\top MRU_2 \tag{11}$$

We know that T_2 is an upper triangular matrix, so what the final equation above indicates is that the new orthonormal basis that upper triangularizes *M better that the original U basis* is the RU_2 basis. That is, instead of using $\begin{bmatrix}\vec{v}_1 & R\end{bmatrix}$, we want $\begin{bmatrix}\vec{v}_1 & RU_2\end{bmatrix} = \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix}$.

It is worth taking a moment to recap the logic; we wanted M to be upper triangular in a particular basis, so we showed that we could construct a basis U for which M was block upper-triangular. But we don't just want the first row and column of the middle matrix T to satisfy the upper-triangular form; we want the *entire* middle matrix to be upper triangular! This requires that $Q = R^\top MR$ is also upper triangular. We showed above that the R basis is insufficient to guarantee that Q is block upper-triangular, but the RU_2 basis is sufficient. Note that now we have a new problem! P might not be upper triangular, which means again, the original T would not be fully upper triangular. This is why we have to recurse into the bottom left entry of each block upper-triangular matrix.

We can substitute the "better" orthonormal basis (the one that upper-triangularizes the first 2 rows and columns of M) now in our expression for M, and solve for what the middle matrix T is. Below, we use the fact that $\begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix}$ is orthonormal and the inverse equals the transpose; we will confirm this in discussion 10B.

$$M = \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix} T \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix}^\top \tag{12}$$

$$\begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix}^\top M \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix} = \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix}^\top \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix} T \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix}^\top \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix} \tag{13}$$

$$\implies \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix}^\top M \begin{bmatrix}\vec{v}_1 & R\vec{v}_2 & RY\end{bmatrix} = T \tag{14}$$

$$\implies \begin{bmatrix}\vec{v}_1^\top M\vec{v}_1 & \vec{v}_1^\top MRU_2 \\ U_2^\top R^\top M\vec{v}_1 & U_2 R^\top MRU_2\end{bmatrix} = T \tag{15}$$

In part d), we got some practice with simplifying the entries of a matrix like the one on the left. We do the same here:

i. $\vec{v}_1^\top M\vec{v}_1 = \vec{v}_1^\top \lambda_1 \vec{v}_1 = \lambda_1 \vec{v}_1^\top \vec{v}_1 = \lambda_1$
ii. $\vec{v}_1^\top MRU_2 = \left(\vec{v}_1^\top MR\right) U_2 = \vec{a}^\top U_2 = \breve{\vec{a}}^\top$
iii. $U_2^\top R^\top M\vec{v}_1 = U_2^\top R^\top \lambda_1 \vec{v}_1 = \lambda_1 U_2^\top R^\top \vec{v}_1 = \vec{0}$ (since the rows of $R^\top$ are all orthogonal to $\vec{v}_1$.)
iv. $U_2 R^\top MRU_2 = U_2 QU_2 = T_2$ (by definition, U_2 is the basis that upper-triangularizes Q.)

So finally, we can write that:

$$M = \begin{bmatrix} \vec{v}_1 & R\vec{v}_2 & RY \end{bmatrix} \begin{bmatrix} \lambda_1 & \breve{a}_1 & \breve{\vec{a}}_{\text{rest}}^\top \\ 0 & \lambda_2 & \vec{b}^\top \\ \vec{0} & \vec{0} & P \end{bmatrix} \begin{bmatrix} \vec{v}_1 & R\vec{v}_2 & RY \end{bmatrix}^\top \tag{16}$$

We can be precise and write that $\breve{\vec{a}}_{\text{rest}}^\top = \breve{\vec{a}}_{2:n-1}^\top$.
The numerical results are:

$$Q = \begin{bmatrix} \frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \\ \frac{-\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \\ \frac{-\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix}^\top \tag{17}$$

$$M = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & \frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & \frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix}^\top \tag{18}$$

Contributors:

- Neelesh Ramachandran.
- Yuxun Zhou.
- Edward Wang.
- Anant Sahai.
- Sanjit Batra.
- Pavan Bhargava.