Questions

1. Towards upper-triangulation by an orthonormal basis

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 dynamical 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” 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 using induction. 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.

In order for you to better understand the steps, you can consider a concrete case

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

and figure out the general case by abstracting variables. This particular matrix has an additional special property of symmetry, but we won’t be invoking that here.

(a) Characteristic polynomial warm-ups: **Show that the characteristic polynomial of square matrix \( A \) is the same as that of the square matrix \( T^{-1}AT \) for any invertible \( T \).**

**Answer:**
Evaluating the determinant, \( \det(A - \lambda I) \) yields the characteristic polynomial of \( A \). So we must show that the determinants of \( A \) and \( TAT^{-1} \) evaluate to the same expression.

\[
\det(A - \lambda I) = \det(TT^{-1}(A - \lambda I)) = \det(T(A - \lambda I)T^{-1})
\]

This is true since determinants are oriented volumes and hence the determinant of a product is the product of determinants. So, one can reorder terms inside the determinant even if the matrices themselves don’t commute.

\[
\implies \det(A - \lambda I) = \det(TAT^{-1} - \lambda I)
\]

(b) 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 \( \vec{u}_1, \ldots, \vec{u}_{n-1} \), such that \( \text{span}(\vec{u}_0, \vec{u}_1, \ldots, \vec{u}_{n-1}) = \mathbb{R}^n \). To begin with, consider

\[
\begin{bmatrix}
1 \\
-1 \\
0
\end{bmatrix}
\]

**Can you get an orthonormal basis from what you just constructed?**
Answer: One possibility: Find a non-zero entry of $\vec{u}_0$, say $\vec{u}_0(k)$, then extend $\vec{u}_0$ with the coordinate basis $\vec{e}_0, \ldots, \vec{e}_{n-1}$ excluding $\vec{e}_k$. You can also ask students to show that this construction gives $n$ linearly independent vectors, which spans $\mathbb{R}^n$. For $[1, -1, 0]^T$, we can do

$$
\begin{bmatrix}
1 & 0 & 0 \\
-1 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
$$

You can also be completely systematic about it — add the standard basis and just Gram-Schmidt the whole thing. If you get a zero along the way, discard and move on. You are guaranteed to span the whole space by the end because the standard basis spans the whole space.

Using the Gram-Schmidt process for the basis obtained above.

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

(c) Now consider a real eigenvalue $\lambda_0$, and the corresponding eigenvector $\vec{g}_0 \in \mathbb{R}^n$ of a square matrix $M \in \mathbb{R}^{n \times n}$. From the previous part, we can extend $\vec{g}_0$ to an orthonormal basis of $\mathbb{R}^n$, denoted by $V = [\vec{v}_0, \vec{v}_1, \ldots, \vec{v}_{n-1}]$ where $\vec{v}_0 = \frac{\vec{g}_0}{\|\vec{g}_0\|}$.

Our goal is to look at what the matrix $M$ looks like in the coordinate system defined by the basis $V$. Compute $V^T M V$ by writing $V = [\vec{v}_0, R]$, where $R \triangleq [\vec{v}_1, \ldots, \vec{v}_{n-1}]$. If you prefer, you can do this and the next question with the concrete $S_{3 \times 3}$ first.

Answer:

$$V^T M V = \begin{bmatrix} \vec{v}_0^T \\ R^T \end{bmatrix} M [\vec{v}_0, R] = \begin{bmatrix} \vec{v}_0^T \\ R^T \end{bmatrix} \begin{bmatrix} \lambda_0 \vec{v}_0, MR \end{bmatrix}$$

Concrete case: Obviously $S_{3 \times 3}$ has zero as eigenvalue, let the corresponding eigenvector be just $[1, -1, 0]^T$, then we have

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

Simple calculation yields

$$V^T S_{3 \times 3} V = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{5}{6} & \frac{\sqrt{2}}{3} \\ 0 & \frac{\sqrt{2}}{6} & \frac{2}{3} \end{bmatrix}$$

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

(d) Define $Q = R^T M R$. Look at the first column and the first row of $V^T M V$ and show that

$$M = V \begin{bmatrix} \lambda_0 & \vec{d}^T \\ 0 & Q \end{bmatrix} V^T$$
Here the $\vec{a}$ is just something arbitrary.

**Answer:** We observe that $\vec{v}_0^T \vec{v}_0 = 1$, $R^T \vec{v}_0 = \vec{0}$ and $\vec{v}_0^T M R = (M \vec{v}_0)^T R = \lambda_0 \vec{v}_0^T R = \vec{0}^T$, due to the orthonormal construction. Hence we have,

$$V^T M V = \begin{bmatrix} \lambda_0 & \vec{a}^T \\ \vec{0} & Q \end{bmatrix}$$

and we get desired form by using $V V^T = I$. Note that in the homework, this will be proved via a different method. In the numerical example, we have $Q = R^T S_{3 \times 3} R = \begin{bmatrix} 5 & \sqrt{2} & 6 \\ \sqrt{2} & 6 & 2 \\ 6 & 2 & 3 \end{bmatrix}$.

(e) **What can you say about the characteristic polynomial** $\det(\lambda I - Q)$ **of** $Q **in relationship to the characteristic polynomial of the original** $M$? **Recall that** $Q$ **is an** $(n - 1) \times (n - 1)$ **matrix.**

**Answer:** The idea here is to see that the characteristic polynomial of $Q$ must be the characteristic polynomial of the original $M$ divided by the factor $(\lambda - \lambda_0)$. This is because of the structure of the block matrix. The determinant must be the first corner times the determinant of the bottom corner block. The oriented volume interpretation gives that to you nearly instantly, especially via the connection between determinants and the operations of Gaussian Elimination.

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

$$Q = [\vec{u}_0, \vec{Y}] \begin{bmatrix} \lambda_1 & \vec{b}^T \\ \vec{0} & P \end{bmatrix} [\vec{u}_0, \vec{Y}]^T$$

where we have taken $\vec{u}_0 \in \mathbb{R}^{n-1}$, an eigenvector of $Q$, associated with eigenvalue $\lambda_1$. Again $\vec{u}_0$ is extended into an orthonormal basis $[\vec{u}_0, \vec{u}_1, \ldots, \vec{u}_{n-2}]$ of $\mathbb{R}^{n-1}$. We denote $\vec{Y} \overset{\Delta}{=} [\vec{u}_1, \ldots, \vec{u}_{n-2}]$.

Plug this into $M$ to show that:

$$M = [\vec{v}_0, R\vec{u}_0, R\vec{Y}] \begin{bmatrix} \lambda_0 & \lambda_1 & \vec{a}^T \\ \vec{0} & 0 & \vec{b}^T \\ \vec{0} & \vec{0} & P \end{bmatrix} [\vec{v}_0, R\vec{u}_0, R\vec{Y}]^T$$

Again, using the concrete case may help you first.

**Answer:**

From part (d), we know that

$$M = V \begin{bmatrix} \lambda_0 & \vec{a}^T \\ \vec{0} & Q \end{bmatrix} V^T$$

and that

$$V = [\vec{v}_0, R]$$

So we get that,

$$M = V \begin{bmatrix} \lambda_0 & \vec{a}^T \\ \vec{0} & Q \end{bmatrix} V^T = [\vec{v}_0, R] \begin{bmatrix} \lambda_0 & \vec{a}^T \\ \vec{0} & Q \end{bmatrix} [\vec{v}_0^T, R^T]$$

Since,

$$Q = [\vec{u}_0, \vec{Y}] \begin{bmatrix} \lambda_1 & \vec{b}^T \\ \vec{0} & P \end{bmatrix} [\vec{u}_0, \vec{Y}]^T$$
\[
\begin{bmatrix}
\lambda_0 & a_1 \\
0 & \lambda_1
\end{bmatrix}
\begin{bmatrix}
\bar{R}^T \\
\bar{P}
\end{bmatrix}
\begin{bmatrix}
\bar{v}_0, R\bar{u}_0, RY \\
\bar{v}_0, R\bar{u}_0, RY
\end{bmatrix}
\]

The numerical one has
\[
Q = \begin{bmatrix}
\frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} & 0 \\
\frac{\sqrt{2}}{3} & \frac{\sqrt{3}}{3} & 1
\end{bmatrix}
\begin{bmatrix}
\frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \\
\frac{\sqrt{2}}{3} & \frac{\sqrt{3}}{3}
\end{bmatrix}
\]

(g) **Show that the matrix** \([\bar{v}_0, R\bar{u}_0, RY]\) **is still orthonormal.**

**Answer:** By construction, we have \(\bar{v}_0^T R\bar{u}_0 = 0\), \(\bar{v}_0^T RY = 0\) because \(\bar{v}_0\) is orthogonal to columns of \(R\), and \((R\bar{u}_0)^T RY = \bar{u}_0^T R^T RY = \bar{u}_0^T Y = 0\) because we constructed \([\bar{u}_0, Y]\) as an orthonormal basis of \(\mathbb{R}^{n-1}\).

Check for the numerical one, that
\[
RU = \begin{bmatrix}
\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{3}}{2} \\
0 & \frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3}
\end{bmatrix}
\begin{bmatrix}
\frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} & 0 \\
\frac{\sqrt{2}}{3} & \frac{\sqrt{3}}{3} & 1
\end{bmatrix}
\]

which is orthogonal to \(\bar{v}_0 = \begin{bmatrix}
\frac{\sqrt{2}}{2} \\
\frac{\sqrt{3}}{2}
\end{bmatrix}\) and we finally get
\[
S = \begin{bmatrix}
\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{2} \\
\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{2}
\end{bmatrix}
\begin{bmatrix}
0 & 0 & 0 \\
\frac{1}{2} & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\frac{\sqrt{3}}{3} & \frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \\
\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3}
\end{bmatrix}
\]

(h) **Perform the above process recursively - what will you get in the end?**

**Answer:** The final matrix would be a scalar, i.e., having dimensions \([1 \times 1]\) \((n = 1)\) having an eigenvalue and eigenvector of 1. Although the the above recursive process is intuitive and essentially rigorous, it is still not a very formal proof. Consult the homework for how this can be cast as a formal induction proof.

---

**Contributors:**

- Yuxun Zhou.
- Edward Wang.
- Anant Sahai.
- Sanjot Batra.