CSM
Spring 2022

EECS 16B
Worksheet 8

1. Upper Triangularization

Recall that before we solved the system of differential equation $\frac{d}{dt}\vec{x}(t) = A\vec{x}(t) + \vec{b}u(t)$ using the change of basis V which contained the eigenvectors of A. The result of the transformation showed that $V^{-1}AV = D$ became the diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A

However, the underlying assumption for this transformation is that our matrix A was diagonalizable. In such a case that our matrix A is not diagonalizable, we can still solve the system of differential equation using upper triangularization by solving it "bottom-up" using backwards substitution.

(a) Give an example of a 2×2 matrix that is not diagonalizable. Explain why we can't use our original trick $D = V^{-1}AV$.

Solution: Any answer can suffice but the most common example of a non-diagonalizable matrix is

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

where there is only one repeated eigenvalue $\lambda = 1$ and the corresponding eigenvector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ We can no longer use $D = V^{-1}AV$ becomes we need 2 eigenvectors to construct the V basis but we only have 1 eigenvector.

(b) For this given upper triangular system of differential equations

$$\frac{d}{dt}\vec{x}(t) = \begin{bmatrix} \lambda_1 & a & \dots & A_{1,n-1} & A_{1,n} \\ 0 & \lambda_2 & \dots & A_{2,n-1} & A_{2,n} \\ \vdots & 0 & \ddots & \lambda_{n-1} & A_{n-1,n} \\ 0 & 0 & \dots & 0 & \lambda_n \end{bmatrix} \vec{x}(t)$$

Explain how we can use upper triangularization to solve the system of differential equation.
Hint: Which differential equation can we solve immediately? Can we use that solution to backward substitute in anyway?

Solution:

Notice that we can solve the differential of the the bottom most row

$$\frac{d}{dt}x_n(t) = \lambda_n x(t)$$

to be $x_n(t) = x_n(0)e^{\lambda_n t}$. Notice that in the row above it, the differential equation is at the form

$$\frac{d}{dt}x_{n-1}(t) = \lambda_{n-1}x(t) + A_{n-1,n}x_n(t).$$

Since we've solved for $x_n(t)$ previously, we can simply plug in the solution at the $n-1$ row and our differential equation becomes of the form $\frac{d}{dt}x(t) = \lambda x(t) + u(t)$ which we know how to solve using the integral solution. We can repeat this process and solve each differential equation at the each row bottom-up. This is what "backwards substitution" means.

(c) We will now perform upper triangularization on a non-diagonalizable 2×2 matrix.

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}$$

Show that $\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of A and find its corresponding eigenvalue.

Solution: Recall that $A\vec{v} = \lambda\vec{v}$ for any eigenvector-eigenvalue pair. We can simply multiply A and $\vec{u}_1$ to get $A\vec{u}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and clearly we can see that corresponding eigenvalue is $\lambda = 2$

(d) Using this eigenvector and some appropriate basis vectors of R^2, find an orthonormal basis U for the column space of A using the Gram-Schmidt algorithm. What do you notice about running the Gram-Schmidt algorithm with more vectors that needs to span a certain subspace? How does Gram-Schmidt handle this?

Solution:

We can find an orthonormal basis for any matrix using the Gram-Schmidt algorithm by filling in the basis vectors of R^3 Thus we simply run the Gram-Schmidt algorithm on the columns of the matrix below

$$U = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Running the Gram-Schmidt algorithm gets us

$\vec{u}_1 = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$ since the norm of this vector is $\sqrt{2}$

$\vec{u}_2 = \vec{s}_2 - proj_{\vec{u}_1}\vec{s}_2 = \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}$

Now we normalize $\vec{u}_2$ to be $\begin{bmatrix} \sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix}$

$\vec{u}_3 = \vec{s}_3 - proj_{\vec{u}_1}\vec{s}_3 - proj_{\vec{u}_2}\vec{s}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Since we added more vectors than we need to span R^2, Gram-Schmidt automatically will output one zero vector and we simply discard this vector. This is because we are inputting one more vector into our basis than we need. The zero vector implies that the extra vector is linearly dependent to the other vectors. Intuitively this makes sense since if we put $n+1$ vectors into Gram-Schmidt to try to span R^N, we should only need n vectors, and thus it will always output one zero vector. Thus our orthonormal basis U for A is

$$U = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix}$$

(e) The next step is to compute $Q = U^T AU$. Compute Q. What do you notice about the matrix Q? What is interesting about the diagonals of Q?

Solution:

$$Q = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix}$$

Notice that the matrix is upper triangular and the diagonals of Q corresponds to the eigenvalues of A

(f) Notice that we have computed $U^T AU = Q = \begin{bmatrix} \lambda_1 & 2 \\ 0 & \lambda_2 \end{bmatrix}$. Show that you can write $A = U \begin{bmatrix} \lambda_1 & 2 \\ 0 & \lambda_2 \end{bmatrix} U^T$.

Hint: What is special about the matrix U?

Solution: From earlier, we can write $Q = U^T AU$. Since U is an orthonormal matrix, we can show that

$$Q = U^T AU$$

Multiply both sides with U to get

$$UQ = UU^T AU = AU$$

Multiply both sides with U^T to get

$$UQU^T = AUU^T = A$$

Thus we finally have

$$A = UQU^T = U \begin{bmatrix} \lambda_1 & 2 \\ 0 & \lambda_2 \end{bmatrix} U^T$$

(g) Now let's extend this to a 3×3 matrix. Play around with the Jupyter Notebook to see the process of upper triangularizing a 3×3 matrix.

Solution: On Jupyter Notebook

(h) Notice that we got lucky that we only needed to run Gram-Schmidt once to find a basis which upper triangularizes our 2×2 matrix A. For higher dimension matrices like the 3×3 matrix in the Jupyter Notebook, however, this is not the case and multiple iterations of part (d)-(f) is required to construct such basis. For a general matrix $n \times n$, how many times do we need to perform Gram-Schmidt in order to construct a basis that upper triangularizes A?

Hint: In the 3×3 matrix example in the Jupyter Notebook, how many times was Gram-Schmidt algorithm ran? Why did it run that many times?

Solution: We need to run it $n-1$ times to construct an orthonormal basis to upper triangularize our matrix A. This is because for higher dimensional matrix, running Gram-Schmidt once doesn't guarantee an upper triangular matrix when multiplying $U^T AU$ because you'll produce a smaller submatrix which isn't upper triangular. Thus we will need to keep repeating the process on smaller submatrices until the 2×2 case since the smaller submatrix of that is a 1x1 matrix, which is a scalar.

(i) In your own words, write down the general algorithm for upper triangularizing an $n \times n$ matrix A.

Solution: Step 1: Find an eigenvalue-eigenvector pair for A.

Step 2: Using the appropriate basis vectors and the eigenvector from Step 1, run Gram-Schmidt to construct an orthonormal basis. Don't forget to throw out $\vec{0}$ since we have one extra vector everytime we do this step.

Step 3: Repeat Step 1 and 2 on the smaller submatrix until the 2×2 case.

Step 4: Using what Gram-Schmidt outputs, construct the orthonormal basis U.

2. Spectral Intuition

An amazing result in Linear Algebra is the Spectral Theorem which says that any symmetric matrix is orthogonally diagonalizable. This means that a symmetric matrix will always have n linearly independent eigenvectors that are all mutually orthogonal. We will show that some of these properties are true for the symmetric matrix A^TA to help motivate the SVD.

(a) Show that A^TA is a symmetric matrix.

Solution: Remember that matrix M is symmetric, if $M^T = M$, therefore it remains to show that $(A^TA)^T = A^TA$. We can also remember that for two matrices, $A, B, (AB)^T = B^TA^T$.
Applying the fact above, we see that $(A^TA)^T = A^T(A^T)^T = A^TA$.

(b) Show that every eigenvalue λ_i of A^TA is greater than or equal to zero.
Hint: Consider $\|A\vec{v}\|_2^2$. *where* $\vec{v}$ *is an eigenvector of* A^TA *with eigenvalue* λ.

Solution: $\|A\vec{v}\|_2^2 = (A\vec{v})^T(A\vec{v}) = \vec{v}^TA^TA\vec{v} = \vec{v}^T(\lambda\vec{v}) = \lambda\vec{v}^T\vec{v} = \lambda\|\vec{v}\|_2^2$.
Therefore we see that:

$$\lambda = \frac{\|A\vec{v}\|_2^2}{\|\vec{v}\|_2^2} \geq 0 \tag{1}$$

$\vec{v}$ is an eigenvector, so it must be nonzero, meaning its norm squared will be greater than 0.
$\|A\vec{v}\|_2^2$ will also be greater than or equal to zero, but may be zero if $A\vec{v} = 0$ or $\vec{v} \in \text{Nul}(A)$.

(c) Show that if λ_i and λ_j are distinct eigenvalues of A^TA, then the respective eigenvectors $\vec{v}_i$ and $\vec{v}_j$ are orthogonal.
Hint: Write out the eigenvector relationships: $A^TA\vec{v}_i = \lambda_i\vec{v}_i$ *and* $A^TA\vec{v}_j = \lambda_i\vec{v}_j$ *and then try taking the transpose of the second equation.*

Solution: Let $\vec{v}_i$ and $\vec{v}_j$ be eigenvectors of A^TA with distinct eigenvalues λ_i and λ_j.
Then,

$$A^TA\vec{v}_i = \lambda_i\vec{v}_i \text{ and } A^TA\vec{v}_j = \lambda_j\vec{v}_j \tag{2}$$

We take the transpose of the second equation on the right to get:

$$\vec{v}_j^TA^TA = \lambda_j\vec{v}_j^T \tag{3}$$

If we left multiply the first equation by $\vec{v}_j^T$ and right multiply the second equation by $\vec{v}_i$ and we get:

$$\vec{v}_j^TA^TA\vec{v}_i = \lambda_i\vec{v}_j^T\vec{v}_i = \lambda_j\vec{v}_j^T\vec{v}_i \tag{4}$$

Therefore we can say that $\lambda_i\vec{v}_j^T\vec{v}_i = \lambda_j\vec{v}_j^T\vec{v}_i$ or

$$\lambda_i\vec{v}_j^T\vec{v}_i - \lambda_j\vec{v}_j^T\vec{v}_i = (\lambda_i - \lambda_j)\vec{v}_j^T\vec{v}_i = 0 \tag{5}$$

Since we assumed $\lambda_i \neq \lambda_j, \vec{v}_j^T\vec{v}_i$ must be zero, but this implies that $\vec{v}_i$ and $\vec{v}_j$ are orthogonal.

(d) Show that if A^TA has a repeated eigenvalue, λ, meaning the eigenspace of λ has dimension greater than or equal to two, we can pick an orthonormal basis for the eigenspace.

Solution: We can pick an orthonormal basis for the eigenspace of λ, by using Gram-Schmidt.

(e) It can be shown through induction that the matrix A^TA is has n linearly independent eigenvectors. Conclude by showing that we can pick n mutually orthonormal eigenvectors for the matrix A^TA. *HINT: Recall what gram-schmidt does.*

Solution: We've shown that eigenvectors with distinct eigenvalues are orthogonal.
We've also shown that for repeated eigenvalues, we can pick an orthogonal eigenbasis.
Therefore, we can pick n linearly independent vectors of A^TA that are all mutually orthogonal, and can normalize each one to have norm 1 to make them mutually orthonormal.

3. (Optional) Basic SVD Practice

This is a review of the individudal steps of finding the Singular Value Decomposition of an $m \times n$ matrix A.

The final answer will be of the form $A = U\Sigma V^T$ where U is a $m \times m$ orthonormal matrix, V is a $n \times n$ orthonormal matrix, and Σ is a $m \times n$ matrix that is a diagonal matrix with 0s padded on the right or below depending on the dimensions m and n.

We give the following procedure to compute the SVD of a $m \times n$ matrix A.

(i) **Step 1: Compute the symmetric matrix $A^T A$ or AA^T.**

$A^T A$ will be of dimension $n \times n$, and AA^T will be of dimension $m \times m$.

For a tall, skinny matrix, where $m > n$, the SVD will be easier to calculate using $A^T A$ while for a short, fat matrix, where $m < n$, the SVD will be easier to calculate using AA^T.

(ii) **Step 2: Find the eigenvalues and eigenvectors of $A^T A$ or AA^T.**

If $m > n$, find the eigenvalues $(\lambda_1, \dots, \lambda_n)$ and eigenvectors $(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$ of $A^T A$.
If $m < n$, we find the eigenvalues $(\lambda_1, \dots, \lambda_m)$ and eigenvectors $(\vec{u}_1,\ \vec{u}_2, \dots, \vec{v}_m)$ of AA^T.
By the spectral theorem for real symmetric matrices, these eigenvectors are orthonormal.

(iii) **Step 3: Compute the singular values $\sigma_i = \sqrt{\lambda_i}$ where λ_i are the sorted in descending order eigenvalues of $A^T A$ or AA^T.**

We know these are all non-negative because $(A\vec{v}_i)^T(A\vec{v}_i) = \|A\vec{v}_i\|^2$ and $(A\vec{v}_i)^T(A\vec{v}_i) = \vec{v}_i^T(A^T A)\vec{v}_i = \lambda_i \vec{v}_i^T \vec{v}_i = \lambda_i$. The corresponding normalized eigenvectors $\vec{v}_i$ form the V matrix.

(iv) **Step 4: Find the corresponding vectors of the U or V matrix**

If $m > n$, we use the nonzero values of σ_i, and $\vec{v}_i$, to find corresponding vectors of the U matrix, $\vec{u}_i$ by computing $\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$.
If $m < n$, we use the nonzero values of σ_i and $\vec{u}_i$, to find corresponding vectors of the V matrix, $\vec{v}_i$ by computing $\vec{v}_i = \frac{A^T \vec{u}_i}{\sigma_i}$
These are normalized since $\sigma_i = \|A\vec{v}_i\| = \|A^T \vec{u}_i\|$ by the argument above, and orthogonal since $(A\vec{v}_i)^T(A\vec{v}_j) = \vec{v}_i^T(A^T A)\vec{v}_j = \lambda_j \vec{v}_i^T \vec{v}_j = 0$ if $i \neq j$, since V is an orthonormal matrix.

(v) **Step 5 (for finding the full SVD): Use Gram-Schmidt to complete the U or V matrix**

If $m > n$ we can complete the U matrix by finding $\vec{u}_{n+1}, \dots, \vec{u}_m$ through Gram-Schmidt.
If $m < n$ we will complete the V matrix by finding $\vec{v}_{m+1}, \dots, \vec{v}_n$ through Gram-Schmidt.

Alternatively we can solve for $\vec{v}_i$ by computing the null-space of A or $\vec{u}_i$ by computing the null-space of A^T and then performing Gram-Schmidt on the basis for the respective null-space.

Recall the following SVD forms:

Compact SVD	$A = U_r \Sigma_r V_r{}^T$
Outer Product SVD	$\sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T$
Full SVD	$A = U\Sigma V^T$

(a) Given the matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Compute the following:

i. compact SVD
ii. outer product SVD
iii. full SVD

Solution:

(i) **Step 1:** The matrix A is a 2×3 matrix. Therefore, we will compute AA^T.

$$AA^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

(ii) **Step 2:** We then compute the eigenvalues and eigenvectors of AA^T.

$$\begin{aligned} \det(AA^T - \lambda I) &= \det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \right) \\ &= (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1) = 0 \end{aligned}$$

The eigenvalues will then be $\lambda = 3, 1$. To find the eigenvectors of AA^T, we look at the null-space of $AA^T - \lambda I$.

$$AA^T - 3I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

A basis for the null-space of $AA^T - 3I$ is:

$$\vec{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$AA^T - I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

A basis for the null-space of $AA^T - I$ is:

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

Remember that these vectors form the U matrix.

(iii) **Step 3:** The singular values are computed by taking the square root of the eigenvalues of AA^T.

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3},\ \sigma_2 = \sqrt{\lambda_2} = 1$$

(iv) **Step 4:** We now solve for the vectors in the V matrix through the equation $\vec{v}_i = \frac{A^T \vec{u}_i}{\sigma_i}$

$$\vec{v}_1 = \frac{A^T \vec{u}_1}{\sqrt{3}} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$$

$$\vec{v}_2 = \frac{A^T \vec{u}_2}{1} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

(v) **Step 5:** We still have one last vector $\vec{v}_3$ to solve for. To do this, we compute the null-space of A A basis for the null-space of A can be computed through Gaussian-Elimination as:

$$\vec{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

We however, must normalize $\vec{v}_3$ so that V has orthonormal columns. As a result, we see that

$$\vec{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}.$$

The final result of our full SVD will be:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{\sqrt{6}}{3} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \tag{6}$$

The compact SVD is:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{\sqrt{6}}{3} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{7}$$

The outer product SVD is:

$$A = \sqrt{3} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{\sqrt{6}}{3} & \frac{1}{\sqrt{6}} \end{bmatrix} + 1 \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{8}$$

Contributors:

- Nareauphol Liu.
- Taejin Hwang.

