Gram-Schmidt, Orthogonal Matrices and Schur's Lemma

Recall the vector space $\mathbb{R}^n$. The scalars are real numbers, where here they'll typically be denoted by lowercase Greek letters such as $\alpha$ or $\beta$, and the vectors are real column matrices, typically denoted here by boldface lowercase Roman letters such as $\mathbf{x}$ or $\mathbf{y}$. The *dot product* is defined by

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \Rightarrow \quad \mathbf{x} \cdot \mathbf{y} \equiv x_1 y_1 + \cdots + x_n y_n.$$

Note that using matrix multiplication we have $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$. The dot product is *symmetric* and *bilinear*. That is

$$\begin{aligned} \text{symmetric} \;&\Rightarrow\; \mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x} \\ \text{bilinear} \;&\Rightarrow\; \mathbf{z} \cdot (\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha(\mathbf{z} \cdot \mathbf{x}) + \beta(\mathbf{z} \cdot \mathbf{y}), \\ &\phantom{\Rightarrow}\; (\alpha\mathbf{x} + \beta\mathbf{y}) \cdot \mathbf{z} = \alpha(\mathbf{x} \cdot \mathbf{z}) + \beta(\mathbf{y} \cdot \mathbf{z}). \end{aligned}$$

The *euclidean length* of a vector $\mathbf{x}$ is denoted by $||\mathbf{x}||$ and is a nonnegative real number given by

$$||\mathbf{x}|| \equiv \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

The euclidean length is *homogeneous* with respect to scalar multiplication. That is

$$||\alpha\mathbf{x}|| = |\alpha| \, ||\mathbf{x}||,$$

where $|\alpha|$ denotes the absolute value of the real scalar $\alpha$. It's also important to observe that $||\mathbf{x}|| = 0 \iff \mathbf{x} = \mathbf{0}$. The dot product and euclidean length can be exemplified in e.g. $\mathbb{R}^3$ as follows. Let

$$\begin{aligned} &\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \\ \Rightarrow \quad &\mathbf{x} \cdot \mathbf{y} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32 \\ \text{and} \quad &||\mathbf{x}|| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}, \quad ||\mathbf{y}|| = \sqrt{4^2 + 5^2 + 6^2} = \sqrt{77}. \end{aligned}$$

Both the euclidean length and the dot product are geometrically significant. For example, if $\mathbf{x}'$ denotes a solid body rotation of $\mathbf{x}$ then $||\mathbf{x}'|| = ||\mathbf{x}||$. Moreover, if $\mathbf{x}$ and $\mathbf{y}$ are simultaneously rotated to $\mathbf{x}'$ and $\mathbf{y}'$ then $\mathbf{x}' \cdot \mathbf{y}' = \mathbf{x} \cdot \mathbf{y}$. That is, both the euclidean length and the dot product are invariant with respect to solid body rotation. From this the following important fact easily follows. For any two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$

$$\frac{\mathbf{x} \cdot \mathbf{y}}{||\mathbf{x}|| \, ||\mathbf{y}||} = \cos(\theta_{\mathbf{x},\mathbf{y}}),$$

where $\theta_{\mathbf{x},\mathbf{y}}$ is the planar angle between vectors $\mathbf{x}$ and $\mathbf{y}$. I'll be glad to derive this fact in class if you ask.

Two nonzero vectors, say $\mathbf{e}_1$ and $\mathbf{e}_2$, are *orthogonal* (or perpendicular) to eachother when the angle between them is $\theta_{\mathbf{e}_1,\mathbf{e}_2} = 90° \iff \mathbf{e}_1 \cdot \mathbf{e}_2 = 0$. For example, in $\mathbb{R}^4$

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix} \text{ is orthogonal to } \mathbf{e}_2 = \begin{pmatrix} 2 \\ -2 \\ 3 \\ -1 \end{pmatrix} \text{ because } \mathbf{e}_1 \cdot \mathbf{e}_2 = 0.$$

Suppose I have a set containing $m$ nonzero and mutually orthogonal vectors from $\mathbb{R}^n$. That is, a set

$$O_m = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\}$$

whose vectors satisfy $\mathbf{e}_i \neq \mathbf{0}$ for each $i$ (nonzero) and $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ for each $i \neq j$ (mutually orthogonal). I will call such a set an *orthogonal set.* An orthogonal set is <u>always</u> linearly independent. To see this is true, suppose

$$\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \cdots + \alpha_m \mathbf{e}_m = \mathbf{0},$$

and dot both sides above with the vector $\mathbf{e}_1$ to find

$$\alpha_1(\mathbf{e}_1 \cdot \mathbf{e}_1) = (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \cdots + \alpha_m \mathbf{e}_m) \cdot \mathbf{e}_1 = \mathbf{0} \cdot \mathbf{e}_1 = 0 \quad \Rightarrow \quad \alpha_1 = 0.$$

Next dot both sides with the vector $\mathbf{e}_2$ to find

$$\alpha_2(\mathbf{e}_2 \cdot \mathbf{e}_2) = (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \cdots + \alpha_m \mathbf{e}_m) \cdot \mathbf{e}_2 = \mathbf{0} \cdot \mathbf{e}_2 = 0 \quad \Rightarrow \quad \alpha_2 = 0.$$

Continue this process to conclude we must have $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ which proves the set is indeed independent.

The dot product can also be used to build a larger orthogonal set from a smaller one. As an example, consider the two orthogonal vectors from $\mathbb{R}^4$ we considered earlier

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix} \text{ and } \mathbf{e}_2 = \begin{pmatrix} 2 \\ -2 \\ 3 \\ -1 \end{pmatrix}, \text{ and check that } \mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \notin \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\}.$$

Now, take $\mathbf{e}_3$ of the form

$$\mathbf{e}_3 = \mathbf{x} - (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2),$$

and determine scalars $\alpha_1$ and $\alpha_2$ to make $\mathbf{e}_3$ orthogonal to both $\mathbf{e}_1$ and $\mathbf{e}_2$. To this end, dot both sides with $\mathbf{e}_1$ and then with $\mathbf{e}_2$ to find

$$\begin{aligned}
& 0 = \mathbf{e}_3 \cdot \mathbf{e}_1 = (\mathbf{x} - (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2)) \cdot \mathbf{e}_1 = (\mathbf{x} \cdot \mathbf{e_1}) - \alpha_1(\mathbf{e}_1 \cdot \mathbf{e}_1) \\
\Rightarrow \quad & \alpha_1 = (\mathbf{x} \cdot \mathbf{e_1})/(\mathbf{e}_1 \cdot \mathbf{e}_1), \\
& 0 = \mathbf{e}_3 \cdot \mathbf{e}_2 = (\mathbf{x} - (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2)) \cdot \mathbf{e}_2 = (\mathbf{x} \cdot \mathbf{e_2}) - \alpha_2(\mathbf{e}_2 \cdot \mathbf{e}_2) \\
\Rightarrow \quad & \alpha_2 = (\mathbf{x} \cdot \mathbf{e_2})/(\mathbf{e}_2 \cdot \mathbf{e}_2).
\end{aligned}$$

Plug in values and compute that $\alpha_1 = 1/7$ and $\alpha_2 = 3/18$. Therefore

$$\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{7}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{18}\begin{pmatrix} 2 \\ -2 \\ 3 \\ -1 \end{pmatrix} = \frac{1}{42}\begin{pmatrix} -20 \\ 2 \\ 15 \\ 1 \end{pmatrix}.$$

Rescaling the length of $\mathbf{e}_3$ above, we now have three orthogonal vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} 2 \\ -2 \\ 3 \\ -1 \end{pmatrix}, \ \mathbf{e}_3 = \begin{pmatrix} -20 \\ 2 \\ 15 \\ 1 \end{pmatrix}.$$

Make sure you understand why in the construction done above I took $\mathbf{x} \notin \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\}$. This will guarantee that $\mathbf{e}_3 \neq \mathbf{0}$. The process just done can be continued to determine a fourth orthogonal vector, $\mathbf{e}_4$, as follows. Check that this time

$$\mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \notin \operatorname{span}\{\mathbf{e}_1, \mathbf{e_2}, \mathbf{e}_3\},$$

and take $\mathbf{e}_4$ of the form

$$\mathbf{e}_4 = \mathbf{x} - (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3).$$

Since now we want $\mathbf{e}_4$ to be orthogonal to $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$, again use the dot product to determine

$$\alpha_1 = (\mathbf{x} \cdot \mathbf{e_1})/(\mathbf{e}_1 \cdot \mathbf{e}_1), \quad \alpha_2 = (\mathbf{x} \cdot \mathbf{e_2})/(\mathbf{e}_2 \cdot \mathbf{e}_2), \quad \alpha_3 = (\mathbf{x} \cdot \mathbf{e_3})/(\mathbf{e}_3 \cdot \mathbf{e}_3).$$

I'll leave it as an exercise for you to explicitly calculate values for the scalars $\alpha_1$, $\alpha_2$ and $\alpha_3$ and finally to evaluate $\mathbf{e}_4$. It looks like it might get pretty messy though.

Gram-Schmidt is an algorithm for building an orthogonal basis. I've shown you how it works by example in the previous paragraph. For completeness sake, let me write it here as an inductive process. Suppose $\mathbb{R}^n$ is our underlying vector space.

Step 1: Let $\mathbf{e}_1$ be a <u>nonzero</u> vector in $\mathbb{R}^n$.

Now inductively for $k = 2, \ldots, n$ do the following.

Step $k$: Find $\mathbf{x}_k \notin \operatorname{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_{k-1}\}$.

Set $\mathbf{e}_k = \mathbf{x}_k - (\alpha_1 \mathbf{e}_1 + \cdots + \alpha_{k-1}\mathbf{e}_{k-1})$,

where $\alpha_1 = (\mathbf{x}_k \cdot \mathbf{e}_1)/(\mathbf{e}_1 \cdot \mathbf{e}_1)$, $\ldots$, $\alpha_{k-1} = (\mathbf{x}_k \cdot \mathbf{e}_{k-1})/(\mathbf{e}_{k-1} \cdot \mathbf{e}_{k-1})$.

When done, Gram-Schmidt yields an orthogonal basis for $\mathbb{R}^n$.

Here's one last Gram-Schmidt example, this time on $\mathbb{R}^3$. Take

$$\mathbf{e}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

I picked $\mathbf{x}_2$ and $\mathbf{x}_3$ above so that $\{\mathbf{e}_1, \mathbf{x}_2, \mathbf{x}_3\}$ obviously forms a basis for $\mathbb{R}^3$. Compute

$$\mathbf{e}_2 = \mathbf{x}_2 - (\alpha_1 \mathbf{e}_1) \quad \Rightarrow \quad \alpha_1 = 0/2 = 0 \quad \Rightarrow \quad \mathbf{e}_2 = (\,1 \quad 0 \quad 0\,)^T .$$

Next, compute

$$\mathbf{e}_3 = \mathbf{x}_3 - (\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2) \quad \Rightarrow \quad \alpha_1 = 1/2,\ \alpha_2 = 0/1 \quad \Rightarrow \quad \mathbf{e}_3 = (\,0 \quad -1/2 \quad 1/2\,)^T .$$

This gives us an orthogonal basis for $\mathbb{R}^3$

$$\{\,(\,0 \quad 1 \quad 1\,)^T,\ (\,1 \quad 0 \quad 0\,)^T,\ (\,0 \quad -1/2 \quad 1/2\,)^T\}$$

---

1. Use Gram-Schmidt to find an orthogonal basis for $\mathbb{R}^3$ taking

$$\mathbf{e}_1 = (\,1 \quad 1 \quad 1\,)^T .$$

as your first vector. Also, please use

$$\mathbf{x}_2 = (\,0 \quad 1 \quad 0\,)^T \text{ and } \mathbf{x}_3 = (\,0 \quad 0 \quad 1\,)^T$$

in your G-S process. My answer: After rescaling lengths I got

$$\mathbf{e}_1 = (\,1 \quad 1 \quad 1\,)^T,\ \mathbf{e}_2 = (\,-1 \quad 2 \quad -1\,)^T,\ \mathbf{e}_3 = (\,-1 \quad 0 \quad 1\,)^T .$$

2. Use Gram-Schmidt to find an orthogonal basis for $\mathbb{R}^4$ taking

$$\mathbf{e}_1 = (\,1 \quad 0 \quad 1 \quad 0\,)^T .$$

as your first vector. Also, please use

$$\mathbf{x}_2 = (\,0 \quad 1 \quad 0 \quad 0\,)^T,\ \mathbf{x}_3 = (\,0 \quad 0 \quad 1 \quad 0\,)^T \text{ and } \mathbf{x}_4 = (\,0 \quad 0 \quad 0 \quad 1\,)^T$$

in your G-S process. My answer: After rescaling lengths I got

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix},\ \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix},\ \mathbf{e}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix},\ \mathbf{e}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

---

A nonzero vector in $\mathbb{R}^n$ can be *normalized* to have <u>unit</u> length by scalar multiplying by the reciprocal of its length. That is, when $\mathbf{x} \neq \mathbf{0}$, the vector

$$\widehat{\mathbf{x}} \equiv \frac{1}{||\mathbf{x}||}\,\mathbf{x} \quad \Rightarrow \quad ||\widehat{\mathbf{x}}|| = ||\frac{1}{||\mathbf{x}||}\,\mathbf{x}|| = \frac{1}{||\mathbf{x}||}\,||\mathbf{x}|| = 1.$$

Such vectors are called *unit vectors* and are often indicated by placing the chapeau symbol $(\widehat{\ })$ on top of the vector name as done above. When each vector in an orthogonal basis is normalized to have unit length we call the scaled basis an *orthonormal basis.* I may also apply the chapeau symbol

$$\{\widehat{\mathbf{e}}_1, \widehat{\mathbf{e}}_2, \ldots, \widehat{\mathbf{e}}_n\}$$

to help clarify this fact.

Consider an $n \times n$ matrix $O$ whose columns are vectors from an orthonormal basis

$$O = \begin{pmatrix} \widehat{\mathbf{e}}_1 & \widehat{\mathbf{e}}_2 & \cdots & \widehat{\mathbf{e}}_n \\ | & | & & | \end{pmatrix}$$

and its transpose

$$O^T = \begin{pmatrix} \widehat{\mathbf{e}}_1^T & \text{——} \\ \widehat{\mathbf{e}}_2^T & \text{——} \\ \vdots & \\ \widehat{\mathbf{e}}_n^T & \text{——} \end{pmatrix}.$$

Matrix multiplication (i.e. dotting rows with columns) clearly shows

$$(O^T O)_{i,j} = \widehat{\mathbf{e}}_i \cdot \widehat{\mathbf{e}}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \Rightarrow \quad O^T O = I \quad \Rightarrow \quad O^T = O^{-1},$$

and in this case we'll call $O$ an *orthogonal matrix* or more properly an *orthonormal matrix*.

To help you see what an orthonormal matrix is, consider the orthogonal basis you derived in exercise 1. Normalize each orthogonal basis vector to get

$$\widehat{\mathbf{e}}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \widehat{\mathbf{e}}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \quad \widehat{\mathbf{e}}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore,

$$O = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

is an orthonormal matrix, and you should check that $O^T O = I$.

Here's an important fact that everybody should know. The product of two orthonormal matrices is an orthonormal matrix. This is easily seen as follows. Suppose $O_1^T = O_1^{-1}$ and $O_2^T = O_2^{-1}$ and let $O_3 = O_1 O_2$ denote the matrix product. Calculate

$$O_3^T = (O_1 O_2)^T = O_2^T O_1^T = O_2^{-1} O_1^{-1} = (O_1 O_2)^{-1} = O_3^{-1}.$$

Therefore, $O_3$ is an orthonormal matrix. We'll use this fact later.

---

3. Consider an orthogonal basis for $\mathbb{R}^2$ given by

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

Use these to build an orthonormal matrix $O_3$.

4. Consider an orthogonal basis for $\mathbb{R}^2$ given by

$$\mathbf{e}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 3 \\ -2 \end{pmatrix}.$$

Use these to build an orthonormal matrix $O_4$.

5. Compute the product $O = O_3 O_4$ where $O_3$ is from exercises 3 and $O_4$ is from exercise 4 and confirm $O$ is an orthonormal matrix.

6. Compute $M^T M$ with $M$ given by

$$M = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

See exercise 2. Explain why the diagonal matrix you got is not the identity.

---

The result of Schur's lemma has important theoretical consequences as opposed to being a powerful computational tool. Its general statement reads as follows. Every square matrix is unitarily similar to an upper triangular matrix. A *unitary matrix* is the complex generalization of the orthonormal matrix we saw above. Specifically, $U \in \mathbb{C}^{n\times n}$ is unitary if $U^*U = I$ where the notation $U^*$ (read $U$-star) stands for the complex conjugate transpose of $U$. To say a matrix $A$ is unitarily similar to an upper triangular matrix means there is a unitary matrix $U$ such that

$$U^* A U = T$$

where $T$ is upper triangular. What I'll prove here is given that $A \in \mathbb{R}^{n\times n}$ has only real eigenvalues, there is an orthonormal matrix $O$ such that

$$O^T A O = T.$$

Interestingly, the eigenvalues of $A$, counting multiplicity, are the diagonal entries of $T$.

Here's a $2 \times 2$ example to start things off. Consider

$$A = \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}, \text{ which has } \lambda = 1,\ \mathbf{r} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

(Check that its other eigenvalue is $\lambda = 2$.) Construct an orthogonal basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ by taking $\mathbf{e}_1 = \mathbf{r}$ and Gram-Schmidt to get $\mathbf{e}_2 = \frac{3}{10}\begin{pmatrix} -1 & 3 \end{pmatrix}^T$. From this build the orthonormal matrix

$$O = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix},$$

and compute

$$\begin{aligned} O^T A O &= \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix} \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \\ &= \frac{1}{10} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 3 & 19 \\ 1 & 13 \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ 0 & 2 \end{pmatrix}. \end{aligned}$$

Why did this similarity transformation triangularize $A$? Well, notice

$$O^T A O = \begin{pmatrix} \widehat{\mathbf{e}}_1^T & \text{——} \\ \widehat{\mathbf{e}}_2^T & \text{——} \end{pmatrix} A \begin{pmatrix} \widehat{\mathbf{e}}_1 & \widehat{\mathbf{e}}_2 \\ | & | \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{e}}_1^T A \widehat{\mathbf{e}}_1 & \widehat{\mathbf{e}}_1^T A \widehat{\mathbf{e}}_2 \\ \widehat{\mathbf{e}}_2^T A \widehat{\mathbf{e}}_1 & \widehat{\mathbf{e}}_2^T A \widehat{\mathbf{e}}_2 \end{pmatrix},$$

and since $\widehat{\mathbf{e}}_1$ is an eigenvector with eigenvalue 1 and $\{\widehat{\mathbf{e}}_1, \widehat{\mathbf{e}}_2\}$ is an orthonormal basis

$$\widehat{\mathbf{e}}_1^T A \widehat{\mathbf{e}}_1 = \widehat{\mathbf{e}}_1^T 1 \widehat{\mathbf{e}}_1 = 1 \text{ and } \widehat{\mathbf{e}}_2^T A \widehat{\mathbf{e}}_1 = \widehat{\mathbf{e}}_2^T 1 \widehat{\mathbf{e}}_1 = 0.$$

Therefore

$$O^T A O = \begin{pmatrix} 1 & \widehat{\mathbf{e}}_1^T A \widehat{\mathbf{e}}_2 \\ 0 & \widehat{\mathbf{e}}_2^T A \widehat{\mathbf{e}}_2 \end{pmatrix} = T.$$

Also notice that $T_{1,1}$ is equal to the eigenvalue $\lambda = 1$. But why is $T_{2,2} = \widehat{\mathbf{e}}_2^T A \widehat{\mathbf{e}}_2$ equal to the other eigenvalue $\lambda = 2$? Well, recall similar matrices always have the same characteristic polynomial

$$\begin{aligned} B = S^{-1} A S \quad \Rightarrow \quad & \det(B - \lambda I) = \det(S^{-1} A S - \lambda I) = \det(S^{-1}(A - \lambda I) S) \\ &= \det(S^{-1}) \det(A - \lambda I) \det(S) = \det(A - \lambda I), \end{aligned}$$

and $T$ and $A$ are similar. Since $T$ is triangular $\det(T - \lambda I) = (T_{1,1} - \lambda)(T_{2,2} - \lambda) \Rightarrow$ the diagonal entries $T_{1,1}$ and $T_{2,2}$ must be eigenvalues of $A$.

Here's a $3 \times 3$ example. Consider

$$A = \begin{pmatrix} 3 & -1 & 1 \\ -2 & 4 & 2 \\ -1 & 1 & 5 \end{pmatrix}, \text{ which has } \lambda = 2, \ \mathbf{r} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

(Check that its other eigenvalues are $\lambda = 4$ and $\lambda = 6$.) Gram-Schmidt taking $\mathbf{e}_1 = \mathbf{r}$ to obtain an orthogonal basis with vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \mathbf{e}_2 = \frac{1}{2} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \ \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \Rightarrow \quad O_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix},$$

and now compute

$$O_1^T A O_1 = \begin{pmatrix} 2 & 1 & 3\sqrt{2}/2 \\ 0 & 5 & \sqrt{2}/2 \\ 0 & \sqrt{2} & 5 \end{pmatrix} = \begin{pmatrix} 2 & 1 \ \ 3\sqrt{2}/2 \\ 0 & \\ 0 & \widetilde{A}_2 \end{pmatrix} \text{ where } \widetilde{A}_2 = \begin{pmatrix} 5 & \sqrt{2}/2 \\ \sqrt{2} & 5 \end{pmatrix}.$$

Clearly the $2 \times 2$ submatrix $\widetilde{A}_2$ is not upper triangular. However, we can easily find another orthonormal transformation to make it so. It's also easy to see that $\widetilde{A}_2$ must have eigenvalues $\lambda = 4$ and $\lambda = 6$ – i.e. the other two eigenvalues of the $3 \times 3$ matrix $A$. For eigenvalue $\lambda = 4$, I calculated that $\widetilde{A}_2$ has eigenvector $\mathbf{r} = (-\sqrt{2}/2 \quad 1)^T$. Take $\mathbf{e}_1 = \mathbf{r}$ and Gram-Schmidt to find $\mathbf{e}_2 = (1 \quad \sqrt{2}/2)^T$, and normalize to get

$$\widehat{\mathbf{e}}_1 = \frac{2}{\sqrt{6}} \begin{pmatrix} -\sqrt{2}/2 \\ 1 \end{pmatrix}, \ \widehat{\mathbf{e}}_2 = \frac{2}{\sqrt{6}} \begin{pmatrix} 1 \\ \sqrt{2}/2 \end{pmatrix} \quad \Rightarrow \quad \widetilde{O}_2 = \frac{2}{\sqrt{6}} \begin{pmatrix} -\sqrt{2}/2 & 1 \\ 1 & \sqrt{2}/2 \end{pmatrix},$$

where $\widetilde{O}_2$ is a $2 \times 2$ orthonormal matrix. Again, multiply out to find

$$\widetilde{O}_2^T \widetilde{A}_2 \widetilde{O}_2 = \begin{pmatrix} 4 & \sqrt{2}/2 \\ 0 & 6 \end{pmatrix}.$$

We're essentially done now. The only problem is $\widetilde{O}_2$ is a $2\times 2$ matrix – not $3\times 3$. However, we can embed it in a $3 \times 3$ matrix as follows.

$$O_2 \equiv \begin{pmatrix} 1 & 0 \quad 0 \\ 0 & \\ 0 & \widetilde{O}_2 \end{pmatrix} \quad \Rightarrow \quad O_2^T O_2 = \begin{pmatrix} 1 & 0 \quad 0 \\ 0 & \\ 0 & \widetilde{O}_2^T \end{pmatrix} \begin{pmatrix} 1 & 0 \quad 0 \\ 0 & \\ 0 & \widetilde{O}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

which says $O_2$ is a $3\times 3$ orthonormal matrix. Therefore, since as shown earlier the product of orthonormal matrices is orthonormal, define $O = O_1 O_2$ and find

$$\begin{aligned} O^T A O = O_2^T \left(O_1^T A O_1\right) O_2 &= \begin{pmatrix} 1 & 0 \quad 0 \\ 0 & \\ 0 & \widetilde{O}_2^T \end{pmatrix} \begin{pmatrix} 2 & 1 \quad 3\sqrt{2}/2 \\ 0 & \\ 0 & \widetilde{A}_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \quad 0 \\ 0 & \\ 0 & \widetilde{O}_2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & * \quad * \\ 0 & \\ 0 & \widetilde{O}_2^T \widetilde{A}_2 \widetilde{O}_2 \end{pmatrix} = \begin{pmatrix} 2 & * & * \\ 0 & 4 & \sqrt{2}/2 \\ 0 & 0 & 6 \end{pmatrix}. \end{aligned}$$

The stars $(*)$ above indicate possibly nonzero numbers above the main diagonal. Their particular values are not significant to the current discussion. Note $A$'s eigenvalues on the main diagonal of $O^T A O$.

Our work above and induction on dimension will verify the result of Schur's lemma in our restricted case when $A \in \mathbb{R}^{n\times n}$ and has only real eigenvalues. Suppose $1 \le k < n$ and suppose there is an orthonormal matrix $O_k$ which reduces $A$ as follows.

$$O_k^T A O_k = \begin{pmatrix} T_k & * \\ 0 & \widetilde{A}_{k+1} \end{pmatrix},$$

where $T_k \in \mathbb{R}^{k\times k}$ is upper triangular with eigenvalues $\lambda_1, \ldots, \lambda_k$ along its diagonal, $\widetilde{A}_{k+1} \in \mathbb{R}^{n-k\times n-k}$, the $*$ represents an $k \times n-k$ block of possibly nonzero real numbers and 0 represents a $n-k\times k$ block of zeros. Because $A$ is similar to $O_k^T A O_k$, they have the same characteristic polynomial. The fundamental theorem of algebra allows us to factor $A$'s characteristic polynomial

$$\det(A - \lambda I) = (\lambda_1 - \lambda)\cdots(\lambda_k - \lambda)\Big[(\lambda_{k+1} - \lambda)\cdots(\lambda_n - \lambda)\Big],$$

and cofactors allows us to write

$$\det(O_k^T A O_k - \lambda I) = (\lambda_1 - \lambda)\cdots(\lambda_k - \lambda)\,\Big[\det(\widetilde{A}_{k+1} - \lambda I)\Big].$$

Equating the square bracketed terms in each, i.e. $\big[\cdots\big]$, find that

$$\det(\widetilde{A}_{k+1} - \lambda I) = (\lambda_{k+1} - \lambda)\cdots(\lambda_n - \lambda).$$

Therefore, $\widetilde{A}_{k+1}$ has eigenvalue $\lambda = \lambda_{k+1}$ and an associated eigenvector $\mathbf{r} \in \mathbb{R}^{n-k\times n-k}$. Take $\mathbf{e}_1 = \mathbf{r}$ and use Gram-Schmidt to build an orthogonal basis

$$\mathbb{R}^{n-k} = \operatorname{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_{n-k}\} \quad \text{giving} \quad \widetilde{O}_{k+1} = \begin{pmatrix} \widehat{\mathbf{e}}_1 & \widehat{\mathbf{e}}_2 & \cdots & \widehat{\mathbf{e}}_{n-k} \\ | & | & & | \end{pmatrix},$$

where the columns of $\widetilde{O}_{k+1}$ are the normalized orthogonal basis vectors $\widehat{\mathbf{e}}_1, \ldots, \widehat{\mathbf{e}}_{n-k}$. Clearly, $\widetilde{O}_{k+1}$ is an orthonormal matrix. Moreover

$$\begin{aligned} \widetilde{O}_{k+1}^T \widetilde{A}_{k+1} \widetilde{O}_{k+1} &= \begin{pmatrix} \widehat{\mathbf{e}}_1^T & \text{---} \\ \widehat{\mathbf{e}}_2^T & \text{---} \\ \vdots & \\ \widehat{\mathbf{e}}_{n-k}^T & \text{---} \end{pmatrix} \begin{pmatrix} \widetilde{A}_{k+1}\widehat{\mathbf{e}}_1 & \widetilde{A}_{k+1}\widehat{\mathbf{e}}_2 & \cdots & \widetilde{A}_{k+1}\widehat{\mathbf{e}}_{n-k} \\ | & | & & | \end{pmatrix} \\ &= \begin{pmatrix} \lambda_{k+1} & * \\ 0 & \widetilde{A}_{k+2} \end{pmatrix}, \end{aligned}$$

where $*$ is a possibly nonzero $n-k-1$ row vector and $0$ is a $n-k-1$ column vector containing only zeros. This last identity follows from the fact that $\widehat{\mathbf{e}}_1$ is a normalized eigenvector of $\widetilde{A}_{k+1}$ and the normalized basis vectors are orthogonal, that is

$$\begin{aligned} \widehat{\mathbf{e}}_1^T \widetilde{A}_{k+1} \widehat{\mathbf{e}}_1 &= \widehat{\mathbf{e}}_1^T \lambda_{k+1} \widehat{\mathbf{e}}_1 = \lambda_{k+1} \\ \widehat{\mathbf{e}}_i^T \widetilde{A}_{k+1} \widehat{\mathbf{e}}_1 &= \widehat{\mathbf{e}}_i^T \lambda_{k+1} \widehat{\mathbf{e}}_1 = 0 \quad \text{for } i = 2, \ldots, n-k. \end{aligned}$$

Finally, compute the product

$$O_{k+1} = O_k \begin{pmatrix} I & 0^T \\ 0 & \widetilde{O}_{k+1} \end{pmatrix},$$

where $I$ is the $k \times k$ identity, $0$ is a $n-k\times k$ block of zeros and $\widetilde{O}_{k+1}$ is the $n-k\times n-k$ orthonormal matrix constructed above. Clearly, $O_{k+1}$ is an orthonormal matrix and it brings us to the next induction step

$$O_{k+1}^T A O_{k+1} = \begin{pmatrix} T_{k+1} & * \\ 0 & \widetilde{A}_{k+2} \end{pmatrix}.$$

This completes the general $n \times n$ induction argument.

I know the induction argument given above might be somewhat confusing to many of you. That's why I gave the $2\times 2$ and $3 \times 3$ examples first. Pay more attention to the examples and don't get mired in the mud of uninteresting details in the general case. Schur's idea is really quite simple though. Just knock off one column at a time.

---

7. Find an orthonormal matrix $O$ so that $O^T AO$ is upper triangular when $A = \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix}$.

My upper triangular: $O^T AO = \begin{pmatrix} 1 & -5 \\ 0 & 2 \end{pmatrix}$.

8. Find an orthonormal matrix $O$ so that $O^TAO$ is upper triangular when $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

My upper triangular: $O^TAO = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$.

9. Let $A = \begin{pmatrix} -1 & 6 & 0 \\ -1 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Find orthonormal $O$ so the $O^TAO$ is upper triangular.

My upper triangular: $O^TAO = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 7 \\ 0 & 0 & 2 \end{pmatrix}$.

---

I want to make a couple of simple but important closing comments. Notice the Schur triangular matrix found in exercise 8 was better than just upper triangular, in fact it was diagonal, $O^TAO = \Lambda$ where $\Lambda = \text{diag}(1, 3)$. In this case, Schur's lemma gives us both $A$'s eigenvalues as well as its eigenvectors. Is there something special here about $A$? The answer is yes. In exercise 8 the matrix $A$ is symmetric. Let me show you this wasn't just a lucky problem.

First, when $A$ is real and symmetric, that is when $A \in \mathbb{R}^{n\times n}$ and $A^T = A$, its eigenvalues must be real. To see this, suppose it has a complex eigenvalue $\lambda$ with associated complex eigenvector $\mathbf{r}$. Notice

$$A\mathbf{r} = \lambda\mathbf{r} \quad \Rightarrow \quad (A\mathbf{r})^* = (\lambda\mathbf{r})^* \quad \Rightarrow \quad \mathbf{r}^*A^* = \bar{\lambda}\mathbf{r}^*,$$

where the star superscript $(*)$ denotes the complex conjugate transpose of a complex matrix and the overbar ( $\bar{\ }$ ) denotes the complex conjugate of a complex scalar. Multiply the left most expression above on the left by $\mathbf{r}^*$, resp. multiply the right most expression on the right by $\mathbf{r}$ to get

$$\mathbf{r}^*A\mathbf{r} = \lambda\,\mathbf{r}^*\mathbf{r}, \quad \text{resp.} \quad \mathbf{r}^*A^*\mathbf{r} = \bar{\lambda}\,\mathbf{r}^*\mathbf{r}.$$

Subtract the first from the second and use the fact that $A^* = A$

$$0 = \mathbf{r}^*A^*\mathbf{r} - \mathbf{r}^*A\mathbf{r} = (\bar{\lambda} - \lambda)\,\mathbf{r}^*\mathbf{r}.$$

An eigenvector $\mathbf{r}$ is never $\mathbf{0}$, and this tells us $\mathbf{r}^*\mathbf{r}$ is a positive real number. Therefore

$$\bar{\lambda} - \lambda = 0 \quad \Rightarrow \quad \bar{\lambda} = \lambda \quad \Rightarrow \quad \lambda \in \mathbb{R}.$$

Finally, because $A$ and $\lambda$ are real, we can assume without loss of generality that the associate eigenvector $\mathbf{r}$ is real.

Our version of Schur's lemma (we assumed real eigenvalues) applies to a real symmetric matrix $A$, and so there is an orthonormal matrix $O$ and an upper triangular matrix $T$

such that

$$O^T AO = T \quad \Rightarrow \quad (O^T AO)^T = T^T \quad \Rightarrow \quad O^T A^T O = T^T \quad \Rightarrow \quad O^T AO = T^T.$$

However, $T = T^T$ implies the upper triangular matrix $T$ is in fact diagonal. Therefore, a symmetric matrix is always diagonalizable. Moreover, it can be assumed that each eigenvector of $A$ is orthogonal to every other of its eigenvectors, i.e. its eigenvectors can be assumed to form an orthogonal basis.