# Math S-21b – Lecture #14 Notes

In the previous lecture, we explored how to interpret the presence of a complex conjugate pair of eigenvalues for an $n \times n$ matrix $\mathbf{A}$ and discovered that in this case there exists a 2-dimensional invariant subspace within which the linear transformation acts as a rotation-dilation relative to an appropriately chosen basis. We'll continue this discussion this week with another example or two and also interpret what the eigenvalues have to say about stability if the linear transformation is applied repeatedly in a discrete dynamical system. We'll also take a look at how to handle repeated eigenvalues in the case where the geometric multiplicity is strictly less that the algebraic multiplicity. We will then also answer the question of under what conditions a matrix will be not only diagonalizable, but also with an orthonormal basis of eigenvectors. This is the subject of the Spectral Theorem.

**$3 \times 3$ example with complex eigenvalues**: Consider the matrix $\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. This matrix permutes the standard basis vectors (and hence the coordinate axes) taking the *x*-axis to the *y*-axis, the *y*-axis to the *z*-axis, and the *z*-axis to the *x*-axis. We write $\lambda\mathbf{I} - \mathbf{A} = \begin{bmatrix} \lambda & 0 & -1 \\ -1 & \lambda & 0 \\ 0 & -1 & \lambda \end{bmatrix}$ and find the characteristic polynomial to be $\lambda^3 - 1 = (\lambda - 1)(\lambda^2 + \lambda + 1) = 0$. This gives the three eigenvalues $\mu = 1$, $\lambda = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$, and $\bar{\lambda} = -\frac{1}{2} - i\frac{\sqrt{3}}{2}$. These eigenvalues all have modulus 1, one real eigenvalue and a complex conjugate pair with arguments $\pm 120°$. They are equally spaced on the unit circle in the complex plane.

The eigenvalue $\mu = 1$ gives the eigenvector $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, which is, in fact, the axis of rotation for this linear transformation.

The eigenvalue $\lambda = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$ gives $\left[\begin{array}{ccc|c} -\frac{1}{2} + i\frac{\sqrt{3}}{2} & 0 & -1 & 0 \\ -1 & -\frac{1}{2} + i\frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & -1 & -\frac{1}{2} + i\frac{\sqrt{3}}{2} & 0 \end{array}\right]$. Rather than follow the usual row reduction procedure, it's usually easier to simply write down what conditions this imposes on any eigenvector $\mathbf{w} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}$. The first row gives that

$\left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)\alpha - \gamma = 0 \Rightarrow \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)\alpha = \gamma \Rightarrow \left(-1 + i\sqrt{3}\right)\alpha = 2\gamma$. Since we have one degree of freedom here, let's simply choose $\alpha = 2$ which then gives $\gamma = -1 + i\sqrt{3}$. The third row gives $-\beta + \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)\gamma = 0 \Rightarrow \beta = \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)\gamma = \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)\left(-1 + i\sqrt{3}\right) = -1 - i\sqrt{3}$.

So $\mathbf{w} = \begin{bmatrix} 2 \\ -1 - i\sqrt{3} \\ -1 + i\sqrt{3} \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} + i\begin{bmatrix} 0 \\ -\sqrt{3} \\ \sqrt{3} \end{bmatrix} = \mathbf{u} + i\mathbf{v}$, and we choose the basis $\{\mathbf{v}, \mathbf{u}\}$ as basis for the 2-dimensional subspace within which this transformation acts as rotation through $\arg(\lambda) = 120°$ with no scaling ($|\lambda| = 1$).

If we put all the eigenvectors together as $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}, \mathbf{u}\}$ and let $\mathbf{S} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & -\sqrt{3} & -1 \\ 1 & \sqrt{3} & -1 \end{bmatrix}$, we'll have:

$$[\mathbf{A}]_{\mathcal{B}} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \left[\begin{array}{c:cc} 1 & 0 & 0 \\ \hdashline 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{array}\right] = \left[\begin{array}{c:cc} 1 & 0 & 0 \\ \hdashline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$R_{120°}$}} \\ 0 & & \end{array}\right]$$

In this form, we see exactly the 120° rotation associated with this matrix. Furthermore, the vectors $\{\mathbf{v},\mathbf{u}\}$ are a basis for the plane perpendicular to the axis of rotation, a plane that remains invariant under this transformation. This is much like the subspace spanned by a real eigenvector, which is a fixed direction. For a typical $3\times3$ matrix with one real eigenvalue and a pair of complex conjugate eigenvalues, the invariant direction corresponding to the real eigenvalue need not be perpendicular to the rotational plane associated with a complex conjugate pair of eigenvalues.

**Diagonalization over the complex numbers**

Technically, if we are willing to allow vectors and matrices to have complex entries, we can indeed diagonalize any matrix with distinct eigenvalues, including complex eigenvalues. We can also do this in the case of any repeated eigenvalues as long as the geometric multiplicities match the algebraic multiplicities. For example, the matrix $\mathbf{A}=\begin{bmatrix}2 & -1\\ 3 & 2\end{bmatrix}$ from the previous class had eigenvalues $\lambda=2+i\sqrt{3}$ and $\bar{\lambda}=2-i\sqrt{3}$. The eigenvalue $\lambda=2+i\sqrt{3}$ produced the complex eigenvector $\mathbf{w}=\begin{bmatrix}1\\ -i\sqrt{3}\end{bmatrix}$, and $\bar{\lambda}=2-i\sqrt{3}$ will have the complex conjugate of $\mathbf{w}$ as its eigenvector, $\hat{\mathbf{w}} = \begin{bmatrix} 1 \\ i\sqrt{3} \end{bmatrix}$. If we let $\mathbf{S} = \begin{bmatrix} 1 & 1 \\ -i\sqrt{3} & i\sqrt{3} \end{bmatrix}$, then

$$\mathbf{S}^{-1}\mathbf{AS} = \begin{bmatrix} 2+i\sqrt{3} & 0 \\ 0 & 2-i\sqrt{3} \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{bmatrix} = \mathbf{D}.$$

Though the rotation-dilation form provides insight into the structure and dynamics of a matrix with complex eigenvalues, it can often be simpler from a computational point of view to use this diagonal form (even though it has complex entries).

**Repeated eigenvalues**

When the algebraic multiplicity *k* of an eigenvalue $\lambda$ of **A** is greater than 1, we will usually not be able to find *k* linearly independent eigenvectors corresponding to this eigenvalue. This is the case where the geometric multiplicity is strictly less than the algebraic multiplicity of this eigenvalue. The next best thing to an eigenvector is often referred to as a "generalized eigenvector".

If, for example, a matrix **A** had $\lambda$ as an eigenvalue with algebraic multiplicity 2, but the geometric multiplicity was 1, we could certainly find an actual eigenvector $\mathbf{v}_1$ such that $\mathbf{Av}_1 = \lambda\mathbf{v}_1$, but we would not be able to product a 2nd linearly independent eigenvector. However, it can be shown (and we'll demonstrate this in an example) that we will be able to find a vector $\mathbf{v}_2$ such that $\mathbf{Av}_2 = \mathbf{v}_1 + \lambda\mathbf{v}_2$. Another way of stating this is that $(\lambda\mathbf{I} - \mathbf{A})\mathbf{v}_1 = \mathbf{0}$

and $(\lambda\mathbf{I}-\mathbf{A})\mathbf{v}_2 = -\mathbf{v}_1$, so $(\lambda\mathbf{I}-\mathbf{A})^2\mathbf{v}_2 = -(\lambda\mathbf{I}-\mathbf{A})\mathbf{v}_1 = \mathbf{0}$. If an eigenvector is a vector in $\ker(\lambda\mathbf{I}-\mathbf{A})$, then a generalized eigenvector would be in $\ker(\lambda\mathbf{I}-\mathbf{A})^2$.

$$\left.\begin{cases} \mathbf{A}\mathbf{v}_1 = \lambda\mathbf{v}_1 \\ \mathbf{A}\mathbf{v}_2 = \quad \mathbf{v}_1 + \lambda\mathbf{v}_2 \end{cases}\right\} \Rightarrow [\mathbf{A}]_{\mathcal{B}} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} = \mathbf{B}$$

In the case where the algebraic multiplicity was 3 and the geometric multiplicity was only 1, we'd also seek a vector in $\ker(\lambda\mathbf{I}-\mathbf{A})^3$, namely a vector $\mathbf{v}_3$ such that $\mathbf{A}\mathbf{v}_3 = \mathbf{v}_2 + \lambda\mathbf{v}_3$. The idea is that a generalized eigenvector is a vector such that the transformation acts on it by scaling together with a shift by the previously found vector.

$$\left.\begin{cases} \mathbf{A}\mathbf{v}_1 = \lambda\mathbf{v}_1 \\ \mathbf{A}\mathbf{v}_2 = \quad \mathbf{v}_1 + \lambda\mathbf{v}_2 \\ \mathbf{A}\mathbf{v}_3 = \qquad\qquad \mathbf{v}_2 + \lambda\mathbf{v}_3 \end{cases}\right\} \Rightarrow [\mathbf{A}]_{\mathcal{B}} = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} = \mathbf{B}$$

It can be shown that this process will <u>always</u> yield *k* linearly independent vectors corresponding to the eigenvalue λ, the first few vectors of which will be actual eigenvectors of **A**. If a matrix **A** has all real eigenvalues and if we carry out this process for all eigenvalues of **A**, we'll produce a complete basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ where we assume that all vectors corresponding to a given eigenvalue are grouped together and ordered in the way in which they were found.

For example, if **A** is a $10\times10$ matrix with nonrepeating eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$; with eigenvalue $\lambda_4$ of multiplicity 3 with only one eigenvector; and with eigenvalue $\lambda_5$ with multiplicity 4 with just two linearly independent eigenvectors; then we'll be able to produce a basis $\mathscr{B}=\{\mathbf{v}_1,\cdots,\mathbf{v}_{10}\}$ where $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ correspond, respectively, to $\lambda_1$, $\lambda_2$, and $\lambda_3$; $\mathbf{v}_4$, $\mathbf{v}_5$, and $\mathbf{v}_6$ correspond to $\lambda_4$; and where $\mathbf{v}_7$, $\mathbf{v}_8$, $\mathbf{v}_9$ and $\mathbf{v}_{10}$ correspond to $\lambda_5$; and where the matrix relative to this special basis is of the form shown at left where **S** is the change of basis matrix.

$$[\mathbf{A}]_{\mathscr{B}}=\mathbf{S}^{-1}\mathbf{A}\mathbf{S}=\left[\begin{array}{c:c:c:ccc:cccc}
\lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline
0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline
0 & 0 & \lambda_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline
0 & 0 & 0 & \lambda_4 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_4 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_4 & 0 & 0 & 0 & 0 \\ \hdashline
0 & 0 & 0 & 0 & 0 & 0 & \lambda_5 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_5 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_5 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_5
\end{array}\right]$$

If we arrange things so that, for example, the eigenvalues are listed in increasing order, the resulting matrix is called the **Jordan canonical form** of the matrix. It follows that any matrix **A** with all real eigenvalues is similar to a matrix in Jordan canonical form, with **Jordan blocks** (as indicated by the dotted lines) associated with each eigenvalue. If **A** and **B** are similar matrices, they necessarily have the same characteristic polynomials, the same eigenvalues with the same algebraic and geometric multiplicities, and hence the same Jordan canonical forms. In other words, they represent the "same" linear transformation relative to two different bases.

**Example**: Consider the matrix $\mathbf{A}=\begin{bmatrix} 0 & 4 \\ -1 & 4 \end{bmatrix}$. We write $\lambda\mathbf{I}-\mathbf{A}=\begin{bmatrix} \lambda & -4 \\ 1 & \lambda-4 \end{bmatrix}$. The characteristic polynomial is $\lambda^2-4\lambda+4=(\lambda-2)^2=0$, so there is only the one eigenvalue $\lambda=2$ with algebraic multiplicity 2.

In seeking eigenvectors, we have $\left[\begin{array}{cc|c} 2 & -4 & 0 \\ 1 & -2 & 0 \end{array}\right]\rightarrow\left[\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right]$, so any eigenvector $\mathbf{v}=\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ would have $\alpha-2\beta=0$ or $\alpha=2\beta$. We can choose the eigenvector $\mathbf{v}_1=\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

We next want a vector $\mathbf{v}_2$ such that $\mathbf{A}\mathbf{v}_2=\mathbf{v}_1+\lambda\mathbf{v}_2$. This translates into $(\lambda\mathbf{I}-\mathbf{A})\mathbf{v}_2=-\mathbf{v}_1$ or the augmented matrix $\left[\begin{array}{cc|c} 2 & -4 & -2 \\ 1 & -2 & -1 \end{array}\right]\rightarrow\left[\begin{array}{cc|c} 1 & -2 & -1 \\ 0 & 0 & 0 \end{array}\right]$. All solutions are of the form $\begin{bmatrix} -1+2t \\ t \end{bmatrix}$. Any choice of $t$ will yield a generalized vector. If we take $t=1$, we get the generalized eigenvector $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Using the basis $\mathscr{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$, the change of basis matrix will be $\mathbf{S} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{S}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$, and we have:

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 0 & 4 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 4 & 4 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} = \mathbf{B}$$

(not diagonal, but upper triangular with a 1 on the superdiagonal)

An easy calculation shows that for any matrix of the form $\mathbf{B} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, we'll have $\mathbf{B}^2 = \begin{bmatrix} \lambda^2 & 2\lambda \\ 0 & \lambda^2 \end{bmatrix}$, $\mathbf{B}^3 = \begin{bmatrix} \lambda^3 & 3\lambda^2 \\ 0 & \lambda^3 \end{bmatrix}$, and, for any positive integer $t$, $\mathbf{B}^t = \begin{bmatrix} \lambda^t & t\lambda^{t-1} \\ 0 & \lambda^t \end{bmatrix}$. This is not as simple as would be the case for a diagonal matrix, but it's still relatively simple. There are similar results for high multiplicity cases.

**Example:** Consider the matrix $\mathbf{A}=\begin{bmatrix}1 & 1 & -1\\ 1 & 1 & 0\\ 2 & -2 & 3\end{bmatrix}$. Its characteristic polynomial is

$\lambda^3-5\lambda^2+8\lambda-4$ which can be factored as $(\lambda-1)(\lambda^2-4\lambda+4)=(\lambda-1)(\lambda-2)^2=0$, so the eigenvalues are $\lambda=1$ with multiplicity 1, and $\lambda=2$ with algebraic multiplicity 2.

The eigenvalue $\lambda=1$ yields the eigenvector $\mathbf{v}_1=\begin{bmatrix}0\\1\\1\end{bmatrix}$, and the repeated eigenvalue $\lambda=2$ yields the single linearly independent eigenvector $\mathbf{v}_2=\begin{bmatrix}1\\1\\0\end{bmatrix}$.

Following the procedure outlined earlier, we can find a third basis vector $\mathbf{v}_3$ such that $\mathbf{Av}_3=\mathbf{v}_2+\lambda\mathbf{v}_3=\mathbf{v}_2+2\mathbf{v}_3$. One such vector is the vector $\mathbf{v}_3=\begin{bmatrix}1\\0\\-2\end{bmatrix}$. Using the basis $\mathcal{B}=\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$, and the matrix $\mathbf{S}=\begin{bmatrix}0 & 1 & 1\\ 1 & 1 & 0\\ 1 & 0 & -2\end{bmatrix}$, we'll get $\mathbf{S}^{-1}\mathbf{AS}=\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix}=\mathbf{B}$.

If we should need to calculate any powers of the matrix **A**, we can simply calculate $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$ and deduce from this that $\mathbf{A}^n = \mathbf{S}\mathbf{B}^n\mathbf{S}^{-1} = \mathbf{S}\left[\begin{array}{c:cc} 1 & 0 & 0 \\ \hdashline 0 & 2^n & n2^{n-1} \\ 0 & 0 & 2^n \end{array}\right]\mathbf{S}^{-1}$.

**Summary**

Any $n \times n$ matrix will yield $n$ eigenvalues, including any complex conjugate pairs and any repeated eigenvalues. We will always be able to find a preferred basis relative to which the matrix will take on a very simple form. If all the eigenvalues are real and the geometric and algebraic multiplicities match for all eigenvalues, then the matrix can be diagonalized. Even in the case where there are complex eigenvalues the matrix can still be put in diagonal form if you are willing to live with complex numbers on the diagonal. The alternative is to have $2 \times 2$ rotation-dilation blocks straddling the diagonal. In the case of repeated eigenvalues with geometric multiplicity less than the algebraic multiplicity, we can almost diagonalize except for some 1's on the superdiagonal. This is the essence of the **Jordan Canonical Form** of a matrix.

**Stability and powers of a matrix (discrete linear dynamical systems)**

We have seen that in the case of a discrete linear dynamical system where $\mathbf{x}(t+1)=\mathbf{A}\mathbf{x}(t)$, so $\mathbf{x}(t)=\mathbf{A}^t\mathbf{x}(0)$, the location of the eigenvalues in the complex plane give essentially all the qualitative information about the evolution of the system. Basically, it's just this:

(a) If all of the eigenvalues of **A** have $|\lambda_i|<1$, then all (discrete) trajectories will tend toward **0**. We would call such a system a stable system.

(b) If even one eigenvalue has $|\lambda_i|>1$, then for almost all initial states $\mathbf{x}(0)$, the trajectories will grow without bound. We would call this an unstable system.

(c) If there are any complex conjugate pairs of eigenvalues, then any trajectory within the invariant 2-dimensional subspace associated with this pair of eigenvalues will exhibit rotation. If the modulus of these eigenvalues is greater than 1 the trajectory will spiral out; and if the modulus of the eigenvalues is less than 1 it will spiral in. These will be discrete spirals, of course. If the modulus is equal to one, the trajectory will move around an ellipse in this invariant plane.

## Orthogonal diagonalizability and the Spectral Theorem

Now that we've addressed the issue of when an $n \times n$ matrix **A** can be diagonalized (there exists a basis consisting of eigenvectors of **A**), we now ask:

> **Q**: When is it possible to diagonalize an $n \times n$ matrix **A** and also have the basis of eigenvectors be an orthonormal basis?

**Definition**: A (real) $n \times n$ matrix **A** is called ***orthogonally diagonalizable*** if there exists an orthonormal basis for $\mathbf{R}^n$ consisting of eigenvectors of the matrix **A**.

We have previously established, for example, that eigenvectors corresponding to distinct eigenvalues must be linearly independent, but they certainly did not also have to be orthogonal. The necessary and sufficient condition for a matrix to be orthogonally diagonalizable is the subject of the Spectral Theorem.

**Spectral Theorem**: A (real) $n \times n$ matrix **A** is orthogonally diagonalizable if and only if **A** is symmetric.

> **Note**: The name of this theorem (as well as the name *spectrum* for the set of eigenvalues of a given matrix) derives from the analogous theorem for *linear operators* acting on a space of functions. Specifically, in Hilbert Space (an inner product space used in the study of quantum mechanics), the Schrödinger operator (denoted by *H*) acts on wave functions (denoted by $\psi$) in the same way that a

matrix acts on vectors. Instead of seeking eigenvectors such that $\mathbf{Av} = \lambda\mathbf{v}$ where $\lambda$ is an eigenvalue, in Hilbert space we would seek wave functions such that $H\psi = E\psi$ where $E$ is the energy level. Energy is proportional to frequency, so the set of all energy levels is in proportion with the set of all frequencies, hence the appropriateness of the name *spectrum*. The analogue of a matrix being symmetric is that the operator $H$ be *self-adjoint*. In this context, we are then assured that a self-adjoint operator will yield orthonormal *eigenfunctions*.

The proof of the Spectral Theorem for matrices is quite simple in one direction, but more involved (and more interesting) in the other direction. We'll address them separately.

**Prop**: If **A** is orthogonally diagonalizable, then **A** is symmetric.
**Proof**: If **A** is orthogonally diagonalizable, then it admits an orthonormal basis of eigenvectors $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ with corresponding eigenvalues $\{\lambda_1, \cdots, \lambda_n\}$. All of these eigenvalues will be real numbers, and there may be multiplicity greater than 1 for some eigenvalues. The crucial fact is that in this case the change-of-basis matrix $\mathbf{S} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}$ will be an <u>orthogonal</u> matrix because it has orthonormal columns, so $\mathbf{S}^{-1} = \mathbf{S}^{\mathrm{T}}$. Since $\mathbf{S}^{-1}\mathbf{AS} = \mathbf{S}^{\mathrm{T}}\mathbf{AS} = \mathbf{D}$, a diagonal matrix, we have that $\mathbf{A} = \mathbf{SDS}^{\mathrm{T}}$, so $\mathbf{A}^{\mathrm{T}} = (\mathbf{SDS}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{SD}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}} = \mathbf{SDS}^{\mathrm{T}} = \mathbf{A}$ (where we used the fact that $\mathbf{D}^{\mathrm{T}} = \mathbf{D}$ for any diagonal matrix). So **A** is symmetric.

**Projections and Reflections**: The Spectral Theorem gives us another way of understanding why the matrix of any orthogonal projection or reflection must necessarily be symmetric. (The converse is false, by the way.) Simply note that if *V* is any subspace of $\mathbf{R}^n$, we can use Gram-Schmidt to produce an orthonormal basis $\{\mathbf{v}_1,\cdots,\mathbf{v}_k\}$ for *V*, and an orthonormal basis $\{\mathbf{v}_{k+1},\cdots,\mathbf{v}_n\}$ for $V^\perp$. In the case of orthogonal projection, *V* will be the eigenspace associated with the eigenvalue 1, and $V^\perp$ will be the eigenspace associated with the eigenvalue 0. Combining these basis vectors, $\mathcal{B}=\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$ will then be an orthonormal basis for $\mathbf{R}^n$ consisting of eigenvectors, so orthogonal projection will therefore be orthogonally diagonalizable. The Spectral Theorem then says that its matrix must be symmetric. We can use this same basis to draw the same conclusion for reflection across the subspace *V*. The only difference will be that $V^\perp$ will be the eigenspace associated with the eigenvalue $-1$.

<u>Note</u>: We came up with this same conclusion earlier by noting that if $\mathbf{B}=\begin{bmatrix}\uparrow & & \uparrow\\ \mathbf{v}_1 & \cdots & \mathbf{v}_k\\ \downarrow & & \downarrow\end{bmatrix}$ where $\{\mathbf{v}_1,\cdots,\mathbf{v}_k\}$ is an orthonormal basis for *V*, then $\mathbf{A}=\mathbf{B}\mathbf{B}^{\mathrm{T}}$ will be the matrix for orthogonal <u>projection</u> onto the subspace *V*. We easily compute that $\mathbf{A}^{\mathrm{T}}=(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{\mathrm{T}}=\mathbf{B}\mathbf{B}^{\mathrm{T}}=\mathbf{A}$. The matrix

for <u>reflection</u> across the subspace $V$ is given by $\mathbf{R} = 2\mathbf{A} - \mathbf{I} = 2\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{I}$, and we compute that $\mathbf{R}^{\mathrm{T}} = (2\mathbf{A} - \mathbf{I})^{\mathrm{T}} = 2\mathbf{A}^{\mathrm{T}} - \mathbf{I}^{\mathrm{T}} = 2\mathbf{A} - \mathbf{I} = \mathbf{R}$.

**Example**: Given the symmetric matrix $\mathbf{A} = \begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix}$, find its eigenvalues and (an orthonormal basis of) eigenvectors.

**Solution**: $\lambda\mathbf{I} - \mathbf{A} = \begin{bmatrix} \lambda - 8 & 2 \\ 2 & \lambda - 5 \end{bmatrix}$, so $p_{\mathbf{A}}(\lambda) = \lambda^2 - 13\lambda + 36 = (\lambda - 9)(\lambda - 4)$. This yields eigenvalues $\lambda_1 = 9$ and $\lambda_2 = 4$. We can easily produce the corresponding eigenvectors $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ for these eigenvalues, but in the context of the Spectral Theorem we want to <u>normalize these to be unit vectors</u>. So we choose $\mathbf{v}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\mathbf{v}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. It's clear that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is then an orthonormal basis of eigenvectors. The change-of-basis matrix $\mathbf{S} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$ is an orthogonal matrix and, due to our choice of orientation of the vectors, a simple rotation matrix.

**Example**: Given the symmetric matrix $\mathbf{B}=\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}$, find its eigenvalues and (an orthonormal basis of) eigenvectors.

**Solution**: $\lambda\mathbf{I}-\mathbf{B}=\begin{bmatrix}\lambda-1&-1&-1\\-1&\lambda-1&-1\\-1&-1&\lambda-1\end{bmatrix}$, so

$p_{\mathbf{B}}(\lambda)=(\lambda-1)(\lambda^2-2\lambda)+1(-\lambda)-1(\lambda)=\lambda^3-3\lambda^2=\lambda^2(\lambda-3)$. The eigenvalues are $\lambda_1=3$ (multiplicity 1) and $\lambda_2=\lambda_3=0$ (multiplicity 2). For the eigenvalue $\lambda_1=3$, we have:

$$\left[\begin{array}{ccc|c}2&-1&-1&0\\-1&2&-1&0\\-1&-1&2&0\end{array}\right]\xrightarrow{RREF}\left[\begin{array}{ccc|c}1&0&-1&0\\0&1&-1&0\\0&0&0&0\end{array}\right].$$

This gives the (normalized) eigenvector $\mathbf{v}_1=\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$.

For the repeated eigenvalue $\lambda_2 = \lambda_3 = 0$, we have $\left[\begin{array}{ccc|c} -1 & -1 & -1 & 0 \\ -1 & -1 & -1 & 0 \\ -1 & -1 & -1 & 0 \end{array}\right] \xrightarrow{RREF} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$.

This yields eigenvectors of the form $\begin{bmatrix} -s-t \\ s \\ t \end{bmatrix} = s\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. Though it's tempting to choose these two spanning vectors as our basis vectors, they are not orthogonal, and in the context of the Spectral Theorem we are seeking an orthonormal basis or eigenvectors. We can start by (arbitrarily) choosing $\mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and then seek a vector $\begin{bmatrix} -s-t \\ s \\ t \end{bmatrix}$ that's orthogonal to $\mathbf{v}_2$.

This gives that $2s + t = 0$, so if we were to pick $s = 1$, we'll get $t = -2$, so the (normalized) eigenvector will be $\mathbf{v}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$. You can then check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis.

The change-of-basis matrix in this case will be $\mathbf{S} = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{bmatrix}$. It may not be pretty, but it is orthogonal, and $\mathbf{S}^{-1} = \mathbf{S}^{\mathrm{T}}$. It's a rotation with a reflection, hence reverses orientation. That is, although $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is right-handed, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is left-handed. Also

$$\mathbf{S}^{-1}\mathbf{B}\mathbf{S} = \mathbf{S}^{\mathrm{T}}\mathbf{B}\mathbf{S} = \mathbf{D} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Now, back to the other half of the proof of the Spectral Theorem.

**Prop**: If **A** is symmetric, then **A** is orthogonally diagonalizable.

We prove this as a series of three separate facts. By default, we treat vectors as column vectors.

**(a)** If **A** is symmetric, then all the eigenvalues of **A** must be real.

**Proof of (a)**: Suppose one of the eigenvalues of **A** was $\lambda = a + ib$ with eigenvector $\mathbf{v} = \mathbf{x} + i\mathbf{y}$. We have previously shown that the eigenvalue $\overline{\lambda} = a - ib$ will then have eigenvector $\hat{\mathbf{v}} = \mathbf{x} - i\mathbf{y}$. Consider the dot product $(\mathbf{A}\mathbf{v}) \cdot \hat{\mathbf{v}} = (\lambda\mathbf{v}) \cdot \hat{\mathbf{v}} = \lambda(\mathbf{x} + i\mathbf{y}) \cdot (\mathbf{x} - i\mathbf{y}) = \lambda(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)$.

But $(\mathbf{Av})\cdot\hat{\mathbf{v}} = \hat{\mathbf{v}}^{\mathrm{T}}\mathbf{Av} = (\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{Av})^{\mathrm{T}} = \mathbf{v}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{v}} = \mathbf{v}^{\mathrm{T}}\mathbf{A}\hat{\mathbf{v}} = (\mathbf{A}\hat{\mathbf{v}})\cdot\mathbf{v} = (\bar{\lambda}\hat{\mathbf{v}})\cdot\mathbf{v} = \bar{\lambda}(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)$, where we used the fact that $\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{Av}$ is a $1\times1$ matrix and is therefore symmetric. Equating these expressions, we see that $\lambda(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) = \bar{\lambda}(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)$, so $(\lambda - \bar{\lambda})(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) = 0$. Since $\mathbf{v} = \mathbf{x} + i\mathbf{y}$ is a nonzero eigenvector, $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \neq 0$. So it must be the case that $\lambda - \bar{\lambda} = 0$. Therefore $\bar{\lambda} = \lambda$, and $\lambda$ must be real.

**(b)** If $\mathbf{A}$ is symmetric and if $\lambda_1 \neq \lambda_2$ are two distinct (real) eigenvalues with (nonzero) eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, then $\mathbf{v}_1$ and $\mathbf{v}_2$ must be orthogonal.

**Proof of (b)**: The proof is very similar to the previous one, except we begin by considering $(\mathbf{Av}_1)\cdot\mathbf{v}_2 = \boxed{\mathbf{v}_2{}^{\mathrm{T}}\mathbf{Av}_1} = (\lambda_1\mathbf{v}_1)\cdot\mathbf{v}_2 = \lambda_1(\mathbf{v}_1\cdot\mathbf{v}_2)$. The boxed expression is a $1\times1$ matrix and is therefore symmetric, so

$$\boxed{\mathbf{v}_2{}^{\mathrm{T}}\mathbf{Av}_1} = (\mathbf{v}_2{}^{\mathrm{T}}\mathbf{Av}_1)^{\mathrm{T}} = \mathbf{v}_1{}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{v}_2 = \mathbf{v}_1{}^{\mathrm{T}}\mathbf{Av}_2 = \mathbf{Av}_2\cdot\mathbf{v}_1 = \lambda_2\mathbf{v}_2\cdot\mathbf{v}_1 = \lambda_2(\mathbf{v}_1\cdot\mathbf{v}_2).$$

Therefore $\lambda_1(\mathbf{v}_1\cdot\mathbf{v}_2) = \lambda_2(\mathbf{v}_1\cdot\mathbf{v}_2)$, so $(\lambda_1 - \lambda_2)(\mathbf{v}_1\cdot\mathbf{v}_2) = 0$. Since $\lambda_1 \neq \lambda_2$ it follows that $\mathbf{v}_1\cdot\mathbf{v}_2 = 0$, i.e. that the eigenvectors are orthogonal. We can then normalize them to get orthonormal eigenvectors.

If all the eigenvalues are distinct, then we can use this to produce an orthonormal basis of eigenvectors. The only remaining potential problem is the case of repeated eigenvalues. We saw in the previous example that we were still able to construct an orthonormal basis in this case, but how would you prove this in general? Well, it's complicated. A proof by induction follows.

**(c)** Generally, if **A** is symmetric, then **A** is orthogonally diagonalizable – even in the case of eigenvalues with multiplicity greater than 1.

**Proof of (c)**: The $1\times 1$ case is obvious. $\mathbf{A}=\left[a\right]$ is symmetric and is already diagonal.

Assume the conclusion is true for any symmetric $(n-1)\times(n-1)$ matrix **B**. We want to prove that it's also true for any $n\times n$ symmetric matrix **A**. To this end, let $\lambda$ be an eigenvalue of **A** with unit eigenvector $\mathbf{v}_1$. Extend this to an orthonormal basis $\mathcal{B}=\left\{\mathbf{v}_1,\cdots,\mathbf{v}_n\right\}$ for $\mathbf{R}^n$ (not necessarily consisting of eigenvectors other than $\mathbf{v}_1$). Then $\mathbf{P}=\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow\end{bmatrix}$ will be an orthogonal matrix. Consider the matrix $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}=\mathbf{P}^{\mathrm{T}}\mathbf{A}\mathbf{P}=[\mathbf{A}]_{\mathcal{B}}$.

By construction, the 1st column of $\mathbf{P}^{\mathrm{T}}\mathbf{AP}$ is $\lambda\mathbf{e}_1 = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. But $(\mathbf{P}^{\mathrm{T}}\mathbf{AP})^{\mathrm{T}} = \mathbf{P}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{P} = \mathbf{P}^{\mathrm{T}}\mathbf{AP}$, so

$\mathbf{P}^{\mathrm{T}}\mathbf{AP}$ is symmetric. It must therefore be of the form $\mathbf{P}^{\mathrm{T}}\mathbf{AP} = \left[\begin{array}{c:ccc} \lambda & 0 & \cdots & 0 \\ \hdashline 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{array}\right]$ where **B** is a

symmetric $(n-1)\times(n-1)$ matrix. By the induction hypothesis, **B** is orthogonally diagonalizable, so there exists an orthogonal $(n-1)\times(n-1)$ matrix **Q** such that $\mathbf{Q}^{\mathrm{T}}\mathbf{BQ} = \mathbf{D}_{n-1}$, a diagonal $(n-1)\times(n-1)$ matrix.

If we let $\mathbf{R} = \left[\begin{array}{c:ccc} 1 & 0 & \cdots & 0 \\ \hdashline 0 & & & \\ \vdots & & \mathbf{Q} & \\ 0 & & & \end{array}\right]$, then **R** will be an orthogonal matrix.

If we then let $\mathbf{S} = \mathbf{PR}$, we have:

$$\mathbf{S}^{\mathrm{T}}\mathbf{A}\mathbf{S} = (\mathbf{P}\mathbf{R})^{\mathrm{T}}\mathbf{A}(\mathbf{P}\mathbf{R}) = \mathbf{R}^{\mathrm{T}}(\mathbf{P}^{\mathrm{T}}\mathbf{A}\mathbf{P})\mathbf{R} = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{Q}^{\mathrm{T}} & \\ 0 & & & \end{array}\right]\left[\begin{array}{c|ccc} \lambda & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{array}\right]\left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{Q} & \\ 0 & & & \end{array}\right]$$

$$= \left[\begin{array}{c|ccc} \lambda & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{Q}^{\mathrm{T}}\mathbf{B}\mathbf{Q} & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c|ccc} \lambda & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{D}_{n-1} & \\ 0 & & & \end{array}\right] = \mathbf{D}$$

So **A** is orthogonally diagonalizable. This completes the proof by induction.

In the next class we'll apply the Spectral Theorem to prove the **Principal Axes Theorem** which can be used to understand **quadratic forms**.

**Notes by Robert Winters**