Similarity, diagonalization, and eigenvalues

Remark. From now on all matrices are square, i.e., $n \times n$ for some $n$ and our field of scalars is the complex numbers, $\mathbb{C}$.

**Definition 1** $A$ is similar to $B$ (written $A \sim B$) iff there exists an invertible matrix $P$ such that $A = PBP^{-1}$.

**Theorem 2**  
1. $A \sim A$  
2. $A \sim B$ implies $B \sim A$, and  
3. $A \sim B$ and $B \sim C$ implies $A \sim C$.

**proof:**
For (1) $A = IAI^{-1}$.
For (2) if $A = PBP^{-1}$, then $B = QAQ^{-1}$ where $Q = P^{-1}$.
For (3) suppose $A = PBP^{-1}$ and $B = QCQ^{-1}$. Then

$$A = (PQ)C(PQ)^{-1}.$$ 

QED

**Definition 3** $A$ is diagonalizable iff $A$ is similar to a diagonal matrix.

**Definition 4** $\lambda \in \mathbb{C}$ is an eigenvalue of $A$ iff for some $X \neq Z$ we have $AX = \lambda X$. Such an $X$ is called an eigenvector of $A$.

**Theorem 5** An $n \times n$ matrix $A$ is diagonalizable iff $\mathbb{C}^n$ has a basis consisting of eigenvectors of $A$.

**proof:**
Suppose first that $\mathbb{C}^n$ has a basis $X_1, \ldots, X_n$ consisting of eigenvectors of $A$, so $AX_i = \lambda_i X_i$. Let $P$ be the $n \times n$ matrix with $\text{col}_j(P) = X_j$ for
each \( j = 1, \ldots, n \) and let \( D \) be the diagonal matrix with \( \lambda_1, \lambda_2, \ldots, \lambda_n \) on its diagonal, i.e.,

\[
D = \begin{bmatrix}
\lambda_1 & 0 & 0 & \cdots & 0 \\
0 & \lambda_2 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & 0 \\
0 & 0 & 0 & \cdots & \lambda_n
\end{bmatrix}.
\]

Since the \( X_j \)'s are a basis \( P \) is invertible.

**Claim** \( AP = PD \).

**proof:**

For any \( j \)

\[
col_j(AP) = A \cdot \text{col}_j(P) \\
= \lambda_j \cdot \text{col}_j(P)
\]

since \( \text{col}_j(P) = X_j \) an eigenvector associated to \( \lambda_j \). Also

\[
col_j(PD) = P \cdot \text{col}_j(D) \\
= P \cdot \text{col}_j(\lambda_jI) \\
= \lambda_j \cdot P \cdot \text{col}_j(I) \\
= \lambda_j \cdot \text{col}_j(PI) \\
= \lambda_j \cdot \text{col}_j(P)
\]

Hence \( AP \) and \( PD \) have the same columns, and so they are equal. This proves the Claim and so \( A = PDP^{-1} \).

Now we prove the converse. Suppose that \( A = PDP^{-1} \). Where \( D \) is a diagonal matrix with \( \lambda_1, \ldots, \lambda_n \) on its diagonal. We have that

\[
AP = PD
\]

and by the same argument as the Claim, we have for each \( j = 1, \ldots, n \) that

\[
A \cdot \text{col}_j(P) = \lambda_j \cdot \text{col}_j(P)
\]

Then since \( P \) is invertible the columns \( P \) are a basis of \( \mathbb{C}^n \) and the formula implies that each \( \text{col}_j(P) \) is an eigenvector of \( A \).

QED
**Theorem 6** If $\lambda_1, \lambda_2, \ldots, \lambda_m$ are distinct eigenvalues of $A$ and $AX_i = \lambda_i X_i$ for $i = 1, \ldots, m$ are (nontrivial) eigenvectors of $A$, then $X_1, X_2, \ldots, X_m$ are linearly independent.

**proof:**

Nontrivial means that $X_i \neq Z$ all $i$.

Suppose they are linearly dependent. Then (by an exercise) either $X_1 = Z$ or there exists $k$ such that $X_k \in \text{span}(\{X_1, \ldots, X_{k-1}\})$. If we choose $k$ minimal such that $X_k \in \text{span}(\{X_1, \ldots, X_{k-1}\})$, then it follows that $X_1, \ldots, X_{k-1}$ are linearly independent.

Now let $a_1, \ldots, a_{k-1}$ be such that

$$X_k = a_1 X_1 + \cdots + a_{k-1} X_{k-1}.$$  

Multiplying by $A$ gives us

$$AX_k = a_1 AX_1 + \cdots + a_{k-1} AX_{k-1}$$

and using $AX_i = \lambda_i X_i$ gives us

$$\lambda_k X_k = a_1 \lambda_1 X_1 + \cdots + a_{k-1} \lambda_{k-1} X_{k-1}.$$  

Case 1. $\lambda_k = 0$.

In this case we have

$$Z = a_1 \lambda_1 X_1 + \cdots + a_{k-1} \lambda_{k-1} X_{k-1}.$$  

and since $X_1, \ldots, X_{k-1}$ are linearly independent we have that $\lambda_i a_i = 0$ for all $i = 1, \ldots, k - 1$. But the $\lambda$’s are all distinct, so $\lambda_i \neq \lambda_k = 0$ for all $i = 1, \ldots, k - 1$. Hence $a_i = 0$ and so $X_k = a_1 X_1 + \cdots + a_{k-1} X_{k-1} = Z$, which contradicts the fact that $X_k$ is nontrivial.

Case 2. $\lambda_k \neq 0$.

In this case we can divide by $\lambda_k$ and get

$$X_k = a_1 \frac{\lambda_1}{\lambda_k} X_1 + \cdots + a_{k-1} \frac{\lambda_{k-1}}{\lambda_k} X_{k-1}.$$  

subtracting this from

$$X_k = a_1 X_1 + \cdots + a_{k-1} X_{k-1}$$
gives

\[ Z = (a_1 - a_1 \frac{\lambda_1}{\lambda_k})X_1 + \cdots + (a_{k-1} - a_{k-1} \frac{\lambda_{k-1}}{\lambda_k})X_{k-1}. \]

Since \( X_1, \ldots, X_{k-1} \) are linearly independent we have that \( a_i - a_i \frac{\lambda_1}{\lambda_k} = 0 \) for all \( i = 1, \ldots, k-1 \). Since \( X_k \) is nontrivial at least one \( i \leq k-1 \) exists such that \( a_i \neq 0 \), but then

\[ a_i - a_i \frac{\lambda_i}{\lambda_k} = 0 \]

implies that \( \lambda_i = \lambda_k \) contradicting the fact that the \( \lambda \)'s are distinct.

In either case we get a contradiction and so our “suppose not” is impossible and therefore \( X_1, \ldots, X_m \) are linearly independent.

QED

**Corollary 7** If an \( n \times n \) matrix \( A \) has \( n \) distinct eigenvalues, then it is diagonalizable.

**proof:**

Since any set of \( n \) independent vectors in \( \mathbb{C}^n \) must be a basis, it must be that \( \mathbb{C}^n \) has a basis consisting of eigenvectors of \( A \) and so \( A \) is diagonalizable.

QED