CHAPTER 5

Diagonalization

In this Chapter, we will learn how to diagonalize a matrix, when we can do it, and what else we can do if we fail to do it.

5.1 Diagonalization

Definition 5.1. A square $n \times n$ matrix $\mathbf{A}$ is **diagonalizable** if $\mathbf{A}$ is similar to a diagonal matrix, i.e.

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

for a diagonal matrix $\mathbf{D}$ and an invertible matrix $\mathbf{P}$.

Diagonalization let us simplify many matrix calculations and prove algebraic theorems. The most important application is the following. If $\mathbf{A}$ is diagonalizable, then it is easy to compute its powers:

Properties 5.2. If $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, then $\mathbf{A}^n = \mathbf{P}\mathbf{D}^n\mathbf{P}^{-1}$.

Example 5.1. Let $\mathbf{A} = \begin{pmatrix} 4 & -3 \\ 2 & -1 \end{pmatrix}$. Then $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ where

$$\mathbf{P} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{P}^{-1} = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix}$$

Then for example

$$\mathbf{D}^8 = \begin{pmatrix} 2^8 & 0 \\ 0 & 1^8 \end{pmatrix} = \begin{pmatrix} 256 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{A}^8 &= \mathbf{P}\mathbf{D}^8\mathbf{P}^{-1} \\ &= \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 256 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 766 & -765 \\ 510 & -509 \end{pmatrix} \end{aligned}$$

The Main Theorem of the Chapter is the following

Theorem 5.3 (The Diagonalization Theorem)**.** An $n \times n$ matrix $\mathbf{A}$ is diagonalizable

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

if and only if $\mathbf{A}$ has n linearly independent eigenvectors.
(Equivalently, $\mathbb{R}^n$ has a basis formed by eigenvectors of $\mathbf{A}$)

- The columns of $\mathbf{P}$ consists of eigenvectors of $\mathbf{A}$
- $\mathbf{D}$ is a diagonal matrix consists of the corresponding eigenvalues.

Proof. Since the columns of $\mathbf{P}$ is linearly independent, $\mathbf{P}$ is invertible. We have

$$\mathbf{AP} = \mathbf{A}\begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = \mathbf{PD}$$

□

Example 5.2. Let us diagonalize $\mathbf{A} = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$

Step 1: Find Eigenvalues. Characteristic equation is

$$\mathbf{p}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{Id}) = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2) = 0$$

Hence the eigenvalues are $\lambda = 7$ and $\lambda = -2$.

Step 2: Find Eigenvectors. We find by usual procedure the linearly independent eigenvectors:

$$\lambda = 7 : \mathbf{v}_1 = \begin{pmatrix}1\\0\\1\end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix}-1\\2\\0\end{pmatrix}, \qquad \lambda = -2 : \mathbf{v}_3 = \begin{pmatrix}-2\\-1\\2\end{pmatrix}$$

Step 3: P constructed from eigenvectors. Putting them in columns,

$$\mathbf{P} = \begin{pmatrix}1 & -1 & -2\\0 & 2 & -1\\1 & 0 & 2\end{pmatrix}$$

Step 4: D consists of the eigenvalues. Putting the eigenvalues according to $\mathbf{v}_i$:

$$\mathbf{D} = \begin{pmatrix}7 & 0 & 0\\0 & 7 & 0\\0 & 0 & -2\end{pmatrix}$$

and we have

$$\mathbf{A} = \mathbf{PDP}^{-1}$$

We have seen in last chapter (Theorem 4.6) that if the eigenvectors have different eigenvalues, then they are linearly independent. Therefore by the Diagonalization Theorem

Corollary 5.4. If $\mathbf{A}$ is an $n \times n$ matrix with n different eigenvalues, then it is diagonalizable.

Example 5.3. The matrix $\mathbf{A} = \begin{pmatrix}3 & 4 & 5\\0 & 0 & 7\\0 & 0 & 6\end{pmatrix}$ is triangular, hence the eigenvalues are the diagonal entries $\lambda = 3, \lambda = 0$ and $\lambda = 6$. Since they are all different, $\mathbf{A}$ is diagonalizable.

Non-Example 5.4. We have seen from Example 4.4 that the matrix $\mathbf{A} = \begin{pmatrix}3 & 0 & 0\\0 & 1 & 1\\0 & 0 & 1\end{pmatrix}$ has 2 eigenvalues $\lambda = 1, 3$ only, so we cannot apply the Corollary. In fact, each of the eigenvalue has only 1-dimensional eigenvectors. Hence $\mathbb{R}^3$ does not have a basis formed by eigenvectors and so it is not diagonalizable by the Diagonalization Theorem.

From this Non-Example, we can also deduce that

Theorem 5.5. A square matrix $\mathbf{A}$ is diagonalizable if and only if for each eigenvalue λ, the algebraic multiplicity equals the geometric multiplicity.

5.2 Symmetric Matrices

A wide class of diagonalizable matrices are given by symmetric matrices, and the diagonalization has very nice properties.

Definition 5.6. A linear operator $T \in L(V, V)$ on an inner product space is called **symmetric** if

$$T\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot T\mathbf{v}$$

If T is represented by an $n \times n$ square matrix $\mathbf{A}$ on $V = \mathbb{R}^n$, then a matrix is called symmetric if

$$\mathbf{A}^T = \mathbf{A}$$

The first important property of symmetric matrix is the orthogonality between eigenspaces.

Theorem 5.7. If $\mathbf{A}$ is symmetric, then two eigenvectors from different eigenspaces are orthogonal.

Proof. If $\mathbf{v}_1 \in V_{\lambda_1}, \mathbf{v}_2 \in V_{\lambda_2}$ are eigenvectors with eigenvalues λ_1, λ_2 such that $\lambda_1 \neq \lambda_2$, then

$$\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{A}\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{A}\mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$$

and so we must have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

Therefore, if we normalize the eigenvectors, then the matrix $\mathbf{P}$ formed from the eigenvectors will consist of orthonormal columns, i.e. $\mathbf{P}$ is an orthogonal matrix.

Definition 5.8. A matrix $\mathbf{A}$ is **orthogonally diagonalizable** if

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

for some orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$.

Theorem 5.9. An $n \times n$ matrix $\mathbf{A}$ is symmetric if and only if it is orthogonally diagonalizable.

In particular, $\mathbf{A}$ is diagonalizable means that each eigenvalue λ has the same algebraic and geometric multiplicity. That is, dimension of the eigenspace V_λ = the number of linearly independent eigenvectors with eigenvalue λ = multiplicity of the root λ of $p(\lambda) = 0$.

Example 5.5 (Exercise 5.2 cont'd)**.** We have diagonalize the matrix $\mathbf{A} = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$ before. But the matrix $\mathbf{P}$ we found is not an orthogonal matrix.

We have found before (**Step 1, Step 2.**)

$$\lambda = 7 : \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \qquad \lambda = -2 : \mathbf{v}_3 = \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}$$

Since $\mathbf{A}$ is symmetric, different eigenspaces are orthogonal to each other. So for example we see that

$$\mathbf{v}_1 \cdot \mathbf{v}_3 = \mathbf{v}_2 \cdot \mathbf{v}_3 = 0$$

So we just need to find an orthogonal basis for the eigenspace V_7.

Step 2a: Use the Gram-Schmidt process on V_7:

$$\mathbf{b}_1 = \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{b}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{b}_1 \cdot \mathbf{v}_2}{\mathbf{b}_1 \cdot \mathbf{b}_1}\right)\mathbf{b}_1 = \frac{1}{2}\begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}$$

Therefore $\{\mathbf{b}_1, \mathbf{b}_2\}$ is an orthogonal basis for V_7, and $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{v}_3\}$ is an orthogonal eigenvector basis for $\mathbb{R}^3$.

Step 2b: Normalize.

$$\mathbf{b}_1' = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{b}_2' = \frac{1}{\sqrt{18}}\begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_3' = \frac{1}{3}\begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}$$

Step 3, Step 4: Construct P and D

Putting together the eigenvectors, we have

$$\mathbf{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \end{pmatrix}$$

and $\mathbf{D} = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{pmatrix}$, consisting of the eigenvalues, is the same as before.

Theorem 5.10. If $\mathbf{A}$ is a symmetric $n \times n$ matrix, then it has n real eigenvalues (counted with multiplicity) i.e. the characteristic polynomial $p(\lambda)$ has n real roots (counted with repeated roots).

The collection of Theorems 5.7, 5.9, and 5.10 in this Section are known as the **Spectral Theorem for Symmetric Matrices**.

5.3 Minimal Polynomials

By the Cayley-Hamilton Theorem, if $p(\lambda)$ is the characteristic polynomial of a square matrix $\mathbf{A}$, then

$$p(\mathbf{A}) = \mathbf{O}$$

Although this polynomial tells us about the eigenvalues (and their multiplicities), it is sometimes too "big" to tell us information about the structure of the matrix.

Definition 5.11. The **minimal polynomial** $m(\lambda)$ is the *unique* polynomial such that

$$m(\mathbf{A}) = \mathbf{O}$$

with leading coefficient 1, and has the smallest degree among such polynomials.

To see it is unique: If we have different minimal polynomials m, m', then $m(\mathbf{A}) - m'(\mathbf{A}) = \mathbf{O}$, but since m, m' have the same degree with the same leading coefficient, $m - m'$ is a polynomial with smaller degree, contradicting the fact that m has smallest degree.

Since it has the smallest degree, in particular we have

$$\deg(m) \leq \deg(p) = n$$

Example 5.6. The diagonal matrix $\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ has characteristic polynomial

$$p(\lambda) = (2 - \lambda)^3$$

but obviously $\mathbf{A} - 2\mathbf{Id} = \mathbf{O}$, hence the minimal polynomial of $\mathbf{A}$ is just

$$m(\lambda) = \lambda - 2$$

In particular,

The minimal polynomial $m(\lambda)$ of $\mathbf{A}$ has degree 1 if and only if $\mathbf{A}$ is a multiple of $\mathbf{Id}$.

Example 5.7. The diagonal matrix $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ has characteristic polynomial

$$p(\lambda) = (1-\lambda)(2-\lambda)^2$$

Since $\mathbf{A}$ is not a multiple of $\mathbf{Id}$, $m(\lambda)$ has degree at least 2. Since $(\mathbf{A}-\mathbf{Id})(\mathbf{A}-2\mathbf{Id}) = \mathbf{O}$, the polynomial

$$m(\lambda) = (\lambda-1)(\lambda-2)$$

having degree 2 is the minimal polynomial.

Example 5.8. The matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ has characteristic polynomial

$$p(\lambda) = (1-\lambda)^3$$

and it turns out that the minimal polynomial is the same also (up to a sign):

$$m(\lambda) = (\lambda-1)^3$$

From the above examples, we also observe that

Theorem 5.12. $p(\lambda) = m(\lambda)q(\lambda)$ for some polynomial $q(\lambda)$. That is $m(\lambda)$ **divides** $p(\lambda)$.

Proof. We can do a polynomial division

$$p(\lambda) = m(\lambda)q(\lambda) + r(\lambda)$$

where $r(\lambda)$ is the remainder with $\deg(r) < \deg(m)$. Since $p(\mathbf{A}) = \mathbf{O}$ and $m(\mathbf{A}) = \mathbf{O}$, we must have $r(\mathbf{A}) = \mathbf{O}$. But since $\deg(r) < \deg(m)$ and m is minimal, r must be the zero polynomial. □

Theorem 5.13. Let $\lambda_1, ..., \lambda_k$ be the eigenvalues of $\mathbf{A}$ (i.e. roots of $p(\lambda)$) Then

$$m(\lambda) = (\lambda-\lambda_1)^{s_1} \cdots (\lambda-\lambda_k)^{s_k}$$

where $1 \le s_i \le m_i$ where m_i is the algebraic multiplicity of λ_i.

Proof. To see $s_i \geq 1$, note that if $\mathbf{v}_i$ is an eigenvector for the eigenvalue λ_i, then since $m(\mathbf{A}) = \mathbf{O}$,

$$\mathbf{0} = m(\mathbf{A})\mathbf{v}_i = m(\lambda_i)\mathbf{v}_i$$

But since $\mathbf{v}_i \neq 0$, we have $m(\lambda_i) = 0$, so λ_i is a root of $m(\lambda)$. $\square$

Finally, the most useful criterion is the following result:

Theorem 5.14. An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if each $s_i = 1$. That is, $m(\lambda)$ only has linear factors.

Using this result, minimal polynomials can let us determine whether a matrix is diagonalizable or not without even calculating the eigenspaces!

Example 5.9. The matrix $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ -4 & 3 \end{pmatrix}$ has characteristic polynomial $p(\lambda) = (\lambda - 1)^2$. Since $m(\lambda) \neq \lambda - 1$ because $\mathbf{A} \neq \mathbf{Id}$, we must have $m(\lambda) = (\lambda - 1)^2$, hence $\mathbf{A}$ is not diagonalizable.

Example 5.10. The matrix $\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ -1 & 0 & 0 \end{pmatrix}$ has characteristic polynomial $p(\lambda) = \lambda(\lambda - 1)^2$, hence it has eigenvalues $\lambda = 1$ and $\lambda = 0$. The minimal polynomial can only be $\lambda(\lambda - 1)$ or $\lambda(\lambda - 1)^2$. Since

$$\mathbf{A}(\mathbf{A} - \mathbf{Id}) \neq \mathbf{O}$$

the minimal polynomial must be $m(\lambda) = \lambda(\lambda - 1)^2$, hence $\mathbf{A}$ is not diagonalizable.

Example 5.11. The matrix $\mathbf{A} = \begin{pmatrix} 2 & -2 & 2 \\ 0 & -2 & 4 \\ 0 & -2 & 4 \end{pmatrix}$ has characteristic polynomial $p(\lambda) = \lambda(\lambda - 2)^2$, hence it has eigenvalues $\lambda = 2$ and $\lambda = 0$. The minimal polynomial can only be $\lambda(\lambda - 2)$ or $\lambda(\lambda - 2)^2$. Since

$$\mathbf{A}(\mathbf{A} - 2\mathbf{Id}) = \mathbf{O}$$

the minimal polynomial is $m(\lambda) = \lambda(\lambda - 2)$, hence $\mathbf{A}$ is diagonalizable.

5.4 Jordan Canonical Form

Finally we arrive at the most powerful tool in Linear Algebra, called the Jordan Canonical Form. This completely determines the structure of a given matrix. It is also the best approximation to diagonalization if the matrix is not diagonalizable.

The result below works **as long as $p(\lambda)$ has n roots (counted with multiplicity)**, and so it is always available if the field is $K = \mathbb{C}$, so that the characteristic polynomial $p(\lambda)$ always has n roots by the Fundamental Theorem of Algebra.

Definition 5.15. Let $\mathbf{A} = \mathbf{A}_1 \oplus \cdots \oplus \mathbf{A}_m$ denote the $n \times n$ matrix in *block form*:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \cdots & \mathbf{O} & \mathbf{A}_m \end{pmatrix}$$

such that $\mathbf{A}_i$ are square matrices of size $d_i \times d_i$, and $\mathbf{O}$ are zero matrices of the appropriate sizes. In particular $n = d_1 + d_2 + \cdots + d_m$.

For any $d \geq 1$, and $\lambda \in \mathbb{C}$, let $\mathbf{J}_\lambda^{(d)}$ be the **Jordan block** denote the $d \times d$ matrix

$$\mathbf{J}_\lambda^{(d)} = \begin{pmatrix} \lambda & 1 & & & & & \\ & \lambda & 1 & & & & \\ & & \lambda & 1 & & & \\ & & & \ddots & \ddots & & \\ & & & & \lambda & 1 & \\ & & & & & \lambda & 1 \\ & & & & & & \lambda \end{pmatrix}$$

where all the unmarked entries are 0.

Note. When $d = 1$, we have $\mathbf{J}_\lambda^{(1)} = \big(\lambda\big)$.

With these notations, we can now state the Main Big Theorem

Theorem 5.16 (Jordan Canonical Form)**.** Let $\mathbf{A} \in M_{n\times n}(\mathbb{C})$. Then $\mathbf{A}$ is similar to

$$\mathbf{J} := \mathbf{J}_{\lambda_1}^{(d_1)} \oplus \cdots \mathbf{J}_{\lambda_m}^{(d_m)}$$

where λ_i belongs to the eigenvalues of $\mathbf{A}$. (**λ_i with different index may be the same!**). This decomposition is *unique* up to permuting the order of the Jordan blocks.

Since eigenvalues, characteristic polynomials, minimal polynomials, and multiplicity etc. are all the same for similar matrices, if **we can determine the Jordan block from these data, we can determine the Jordan Canonical Form of a matrix A.**

Let us first consider a single block.

Properties 5.17. The Jordan block $\mathbf{J}_\lambda^{(d)}$ has

- only one eigenvalue λ
- characteristic polynomial $(t-\lambda)^d$
- minimal polynomial $(t-\lambda)^d$
- geometric multiplicity of λ is 1.

Now let us combine several blocks of the same eigenvalues:

Properties 5.18. The matrix $\mathbf{J}_\lambda^{(d_1)} \oplus \cdots \oplus \mathbf{J}_\lambda^{(d_k)}$ has

- only one eigenvalue λ
- characteristic polynomial $(t-\lambda)^{d_1+\cdots d_k}$
- minimal polynomial $(t-\lambda)^{\max(d_1,\ldots,d_k)}$
- geometric multiplicity of λ is k.

Now we can do the same analysis by combining different Jordan blocks. We arrive at the following structure:

Theorem 5.19. Given a matrix $\mathbf{A}$ in the Jordan canonical form:

- The eigenvalues $\lambda_1, ..., \lambda_k$ are the entries of the diagonal.
- The characteristic polynomial is

  $$p(\lambda) = (\lambda-\lambda_1)^{r_1} \cdots (\lambda-\lambda_k)^{r_k}$$

  where r_i is the number of occurrences of λ_i on the diagonal.
- The minimal polynomial is

  $$m(\lambda) = (\lambda-\lambda_1)^{s_1} \cdots (\lambda-\lambda_k)^{s_k}$$

  where s_i is the size of the largest λ_i-block in $\mathbf{A}$
- The geometric multiplicity of λ_i is the number of λ_i-blocks in $\mathbf{A}$.

Example 5.12. Assume $\mathbf{A}$ is a 6×6 matrix with characteristic polynomial

$$p(\lambda) = (\lambda - 2)^4(\lambda - 3)^2$$

and minimal polynomial

$$m(\lambda) = (\lambda - 2)^2(\lambda - 3)^2,$$

with eigenspaces $\dim V_2 = 3, \dim V_3 = 1$. Then it must have 3 blocks of $\lambda = 2$, with maximum block-size of 2 so that the $\lambda = 2$ blocks add up to 4 rows. It also has 1 block of $\lambda = 3$ with block-size 2. Hence

$$\mathbf{A} \sim \mathbf{J}_2^{(2)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_3^{(2)}$$

$$\mathbf{A} \sim \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

The uniqueness of Jordan Canonical Form says that $\mathbf{A}$ is also similar to the matrix where the Jordan blocks are in different order. For example we can have:

$$\mathbf{A} \sim \begin{pmatrix} 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

Example 5.13. Another example, let $\mathbf{A}$ be a matrix such that it has characteristic polynomial

$$p(\lambda) = \lambda^4(\lambda - 1)^3(\lambda - 2)^3$$

and minimal polynomial

$$m(\lambda) = \lambda^3(\lambda - 1)^2(\lambda - 2)$$

With this information only, we can determine

$$\mathbf{A} \sim \mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(1)} \oplus \mathbf{J}_1^{(2)} \oplus \mathbf{J}_1^{(1)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_2^{(1)}$$

$$\mathbf{A} \sim \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

It turns out that when the matrix is bigger than 6×6, sometimes we **cannot determine** the Jordan Canonical Form just by knowing $p(\lambda), m(\lambda)$ and the dimension of the eigenspaces only:

Example 5.14. Consider a 7×7 matrix $\mathbf{A}$. Let $p(\lambda) = \lambda^7$, $m(\lambda) = \lambda^3$, and $\dim V_0 = 3$. Then $\mathbf{A}$ has 3 blocks and the largest block has size 3. So it may be similar to

$$\mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(1)} \qquad \text{or} \qquad \mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(2)} \oplus \mathbf{J}_0^{(2)}$$

However, by the uniqueness of Jordan Canonical Form, we know that these two are not similar to each other, but we cannot tell which one is similar to $\mathbf{A}$ just from the given information.

To determine which one is the Jordan Canonical Form of $\mathbf{A}$, we need more techniques. In the Homework, we will discuss how one can determine exactly the size of the Jordan blocks, as well as the transformation matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{PJP}^{-1}$.

5.5 Positive definite matrix (Optional)

One application of the diagonalization of symmetric matrix allows us to analyses quadratic functions, and define “square root” and “absolute value” of a matrix, which is useful in advanced linear algebra and optimization problems.

Definition 5.20. Let $\mathbf{A}$ be a symmetric matrix. The quadratic function

$$\mathbf{Q}(\mathbf{x}) := \mathbf{x}^T \mathbf{A}\mathbf{x} = \mathbf{x} \cdot \mathbf{A}\mathbf{x}$$

is called the **quadratic form** associated to $\mathbf{A}$.

Definition 5.21. A quadratic form $\mathbf{Q}$ (or symmetric matrix $\mathbf{A}$) is called **positive definite** if

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{A}\mathbf{x} > 0, \qquad \text{for all nonzero } \mathbf{x} \in \mathbb{R}^n$$

It is called **positive semidefinite** if

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{A}\mathbf{x} \geq 0, \qquad \text{for all nonzero } \mathbf{x} \in \mathbb{R}^n$$

Example 5.15. $\mathbf{Q}(\mathbf{x}) := \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 9x^2 + y^2$ is positive definite.

Example 5.16. Let $\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$. Then $\mathbf{Q}(\mathbf{x}) := \mathbf{x}^T \mathbf{A}\mathbf{x} = 5x^2 + 8xy + 5y^2$ is positive definite. We can see that it represents ellipses as follows: We can diagonalize the matrix by $\mathbf{A} = \mathbf{PDP}^T$ where

$\mathbf{D} = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$ and $\mathbf{P} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Then

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x}^T \mathbf{P}\mathbf{D}\mathbf{P}^T \mathbf{x} = (\mathbf{P}^T\mathbf{x})^T \mathbf{D}(\mathbf{P}^T\mathbf{x})$$

Therefore if we let $\widehat{\mathbf{x}} = \begin{pmatrix} \widehat{x} \\ \widehat{y} \end{pmatrix} = \mathbf{P}^T\mathbf{x}$, i.e. rotating the basis by $\mathbf{P}^{-1}$, then

$$\mathbf{Q}(\widehat{\mathbf{x}}) = 9\widehat{x}^2 + \widehat{y}^2$$

and it is represented by an ellipse.

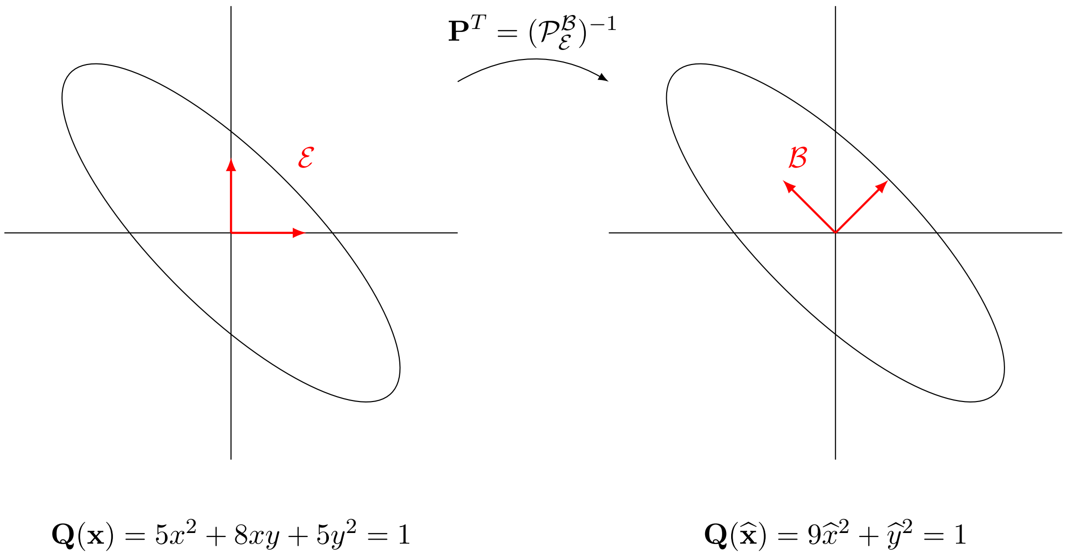

Figure 5.1: Pictorial explanation of similar matrix.

Theorem 5.22. A quadratic form $\mathbf{Q}$ associated to a symmetric matrix $\mathbf{A}$ is positive (semi)definite if and only if $\lambda_i > 0$ ($\lambda_i \geq 0$) for all the eigenvalues of $\mathbf{A}$.

Proof. Substitute $\mathbf{x} \rightsquigarrow \mathbf{P}\mathbf{x}$ where $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ is the diagonalization. $\square$

Remark. If all eigenvalues are $\lambda_i < 0$ ($\lambda_i \leq 0$), we call the quadratic form **negative (semi)definite**. Otherwise if some are positive and some are negative, it is called **indefinite**.

We can always find a "square root" of $\mathbf{A}$ if it is positive (semi)definite.

Theorem 5.23. If $\mathbf{A}$ is positive (semi)definite, then there exists exactly one positive (semi)definite matrix $\mathbf{B}$ such that

$$\mathbf{B}^2 = \mathbf{A}$$

We call $\mathbf{B}$ the **square root** of $\mathbf{A}$ and denote it by $\sqrt{\mathbf{A}}$. It is given by

$$\sqrt{\mathbf{A}} = \mathbf{P}\mathbf{D}^{\frac{1}{2}}\mathbf{P}^T$$

where $\mathbf{D}^{\frac{1}{2}}$ is the diagonal matrix where we take the square root of the entries of $\mathbf{D}$.

We can also construct the "absolute value" of *any* matrix $\mathbf{A}$:

Theorem 5.24. Let $\mathbf{A}$ be any $m \times n$ matrix. Then $\mathbf{A}^T\mathbf{A}$ is a positive semidefinite matrix, and

$$|\mathbf{A}| := \sqrt{\mathbf{A}^T\mathbf{A}}$$

is called the **absolute value** of $\mathbf{A}$.

Proof. $\mathbf{A}^T\mathbf{A}$ is symmetric, and $\mathbf{x} \cdot \mathbf{A}^T\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$. □

This is used in the construction of Singular Value Decomposition in the next section.

Example 5.17. Let $\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{P}^T$ where $\mathbf{P} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Then

$$\sqrt{\mathbf{A}} = \mathbf{P}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{P}^T = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Let $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ -2/5 & -11/5 \end{pmatrix}$. Then $\mathbf{B}^T\mathbf{B} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$, therefore by above $|\mathbf{B}| = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

5.6 Singular Value Decomposition (Optional)

We know that not all matrix can be diagonalized. One solution to this is to use Jordan Canonical Form, which give us an approximation. Another approach is the **Singular Value Decomposition**, and this can even be applied to rectangular matrix! This method is also extremely important in data analysis.

Recall that if $\mathbf{Av} = \lambda\mathbf{v}$ is the eigenvector, the effect is "stretching by λ" along the direction of $\mathbf{v}$. We want to consider all such directions if possible, even for rectangular matrix.

Definition 5.25. Let $\mathbf{A}$ be $m \times n$ matrix. The **singular values** of $\mathbf{A}$ is the eigenvalues σ_i of $|\mathbf{A}| = \sqrt{\mathbf{A}^T\mathbf{A}}$.

If $\mathbf{A}$ has rank r, then we have r nonzero singular values. We arrange them as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Since $\mathbf{A}^T\mathbf{A}$ is a positive definite symmetric matrix, it has an orthonormal set of eigenvectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ with positive eigenvalues $\{\lambda_1, ..., \lambda_n\}$. Then

$$\|\mathbf{Av}_i\|^2 = \mathbf{v}_i^T\mathbf{A}^T\mathbf{Av}_i = \mathbf{v}_i^T\lambda_i\mathbf{v}_i = \lambda_i$$

therefore the singular values $\sigma_i = \sqrt{\lambda_i} = \|\mathbf{Av}_i\|$ of $\mathbf{A}$ is precisely the length of the vector $\mathbf{Av}_i$.

Let us denote a "quasi-diagonal" matrix of size $m \times n$ and rank $r \leq m, n$:

$$\mathbf{\Sigma} = \begin{array}{cc} \begin{array}{cc} r & n-r \end{array} & \\ \begin{pmatrix} \mathbf{D} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} & \begin{array}{c} r \\ m-r \end{array} \end{array}$$

where $\mathbf{D}$ is a diagonal matrix. (When $r = m$ or n, we omit the rows or columns of zeros).

Theorem 5.26 (Singular value decomposition)**.** Let $\mathbf{A}$ be an $m \times n$ matrix with rank r. Then we have the factorization

$$\mathbf{A} = \mathbf{U\Sigma V}^T$$

where

- $\mathbf{\Sigma}$ is as above with $\mathbf{D}$ consists of the first r singular values of $\mathbf{A}$
- $\mathbf{U}$ is an $m \times m$ orthogonal matrix
- $\mathbf{V}$ is an $n \times n$ orthogonal matrix

$\mathbf{V} = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}$ where the columns are the orthonormal eigenvectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of $\mathbf{A}^T\mathbf{A}$.

For $\mathbf{U}$, extend the orthogonal set $\{\mathbf{Av}_1, ..., \mathbf{Av}_r\}$ to a basis of $\mathbb{R}^m$, and normalize to obtain an orthonormal basis $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$. Then $\mathbf{U} = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{pmatrix}$.

Example 5.18. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$. Then $\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ and it has eigenvalues $\lambda_1 = 4, \lambda_2 = 2, \lambda_3 = 0$ with orthonormal eigenvectors

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

Therefore

$$\mathbf{V} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

Also $\sigma_1 = \sqrt{\lambda_1} = 2, \sigma_2 = \sqrt{\lambda_2} = \sqrt{2}$. Therefore

$$\mathbf{\Sigma} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

Finally

$$\begin{aligned} \mathbf{u}_1 &= \frac{\mathbf{A}\mathbf{v}_1}{\|\mathbf{A}\mathbf{v}_1\|} = \frac{\mathbf{A}\mathbf{v}_1}{\sigma_1} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \mathbf{u}_2 &= \frac{\mathbf{A}\mathbf{v}_2}{\|\mathbf{A}\mathbf{v}_2\|} = \frac{\mathbf{A}\mathbf{v}_2}{\sigma_2} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{aligned}$$

therefore

$$\mathbf{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

and

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

is the Singular Value Decomposition of $\mathbf{A}$.

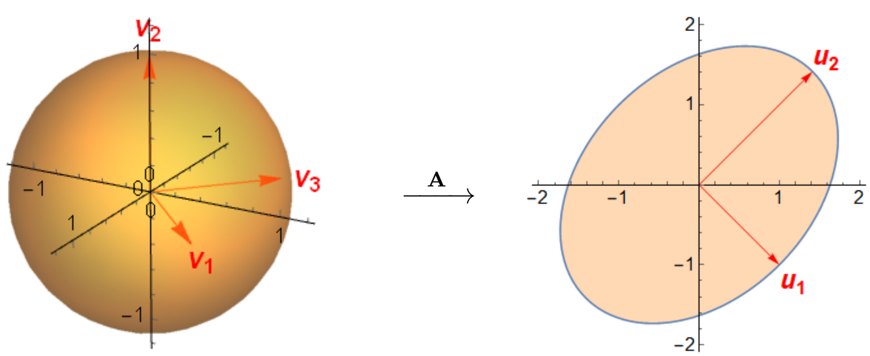

Figure 5.2: Multiplication by $\mathbf{A}$. It squashed the $\mathbf{v}_3$ direction to zero.

One useful application of SVD is to **find the bases of the fundamental subspaces**.

Theorem 5.27. Let $\mathbf{A}$ be $m \times n$ matrix with rank r, and $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the SVD.

Assume $\mathbf{U} = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{pmatrix}$ and $\mathbf{V} = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}$. Then

- $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is an orthonormal basis of Col$\mathbf{A}$.
- $\{\mathbf{u}_{r+1}, ..., \mathbf{u}_m\}$ is an orthonormal basis of Nul$(\mathbf{A}^T)$.
- $\{\mathbf{v}_1, ..., \mathbf{v}_r\}$ is an orthonormal basis of Row$\mathbf{A}$ = Col$\mathbf{A}^T$.
- $\{\mathbf{v}_{r+1}, ..., \mathbf{v}_n\}$ is an orthonormal basis of Nul$\mathbf{A}$.

Another application is the **least-square solution** which works like the example from QR decomposition in Chapter 3.

Definition 5.28. Let $\mathbf{U}_r = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r \\ | & & | \end{pmatrix}, \mathbf{V}_r = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_r \\ | & & | \end{pmatrix}$ be the submatrix consists of the first r columns. Then

$$\mathbf{A} = \begin{pmatrix} \mathbf{U}_r & * \end{pmatrix} \begin{pmatrix} \mathbf{D} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{V}_r^T \\ * \end{pmatrix} = \mathbf{U}_r \mathbf{D} \mathbf{V}_r^T.$$

The **pseudoinverse** of $\mathbf{A}$ is defined to be

$$\mathbf{A}^+ = \mathbf{V}_r \mathbf{D}^{-1} \mathbf{U}_r^T$$

The pseudoinverse satisfies for example

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$$

and

$$\mathbf{A}\mathbf{A}^+ = \mathrm{Proj}_{\mathrm{Col}\mathbf{A}}$$

because

$$\mathbf{A}\mathbf{A}^+ = (\mathbf{U}_r \mathbf{D} \mathbf{V}_r^T)(\mathbf{V}_r \mathbf{D}^{-1} \mathbf{U}_r^T) = \mathbf{U}_r \mathbf{D}\mathbf{D}^{-1} \mathbf{U}_r^T = \mathbf{U}_r \mathbf{U}_r^T = \mathrm{Proj}_{\mathrm{Col}\mathbf{A}}$$

Theorem 5.29. Given the equation $\mathbf{x} = \mathbf{A}\mathbf{b}$, the least-square solution is given by

$$\widehat{\mathbf{x}} = \mathbf{A}^{+}\mathbf{b} = \mathbf{V}_r\mathbf{D}^{-1}\mathbf{U}_r^T\mathbf{b}$$

Proof. Since $\mathbf{A}\widehat{\mathbf{x}} = \mathbf{A}\mathbf{A}^{+}\mathbf{b} = \mathrm{Proj}_{\mathrm{Col}\mathbf{A}}\mathbf{b}$, $\mathbf{A}\widehat{\mathbf{x}}$ is the closest point to $\mathbf{b}$ in Col$\mathbf{A}$. □