Chapter 6

Diagonalization

In this chapter, all vectors are written as column vectors.

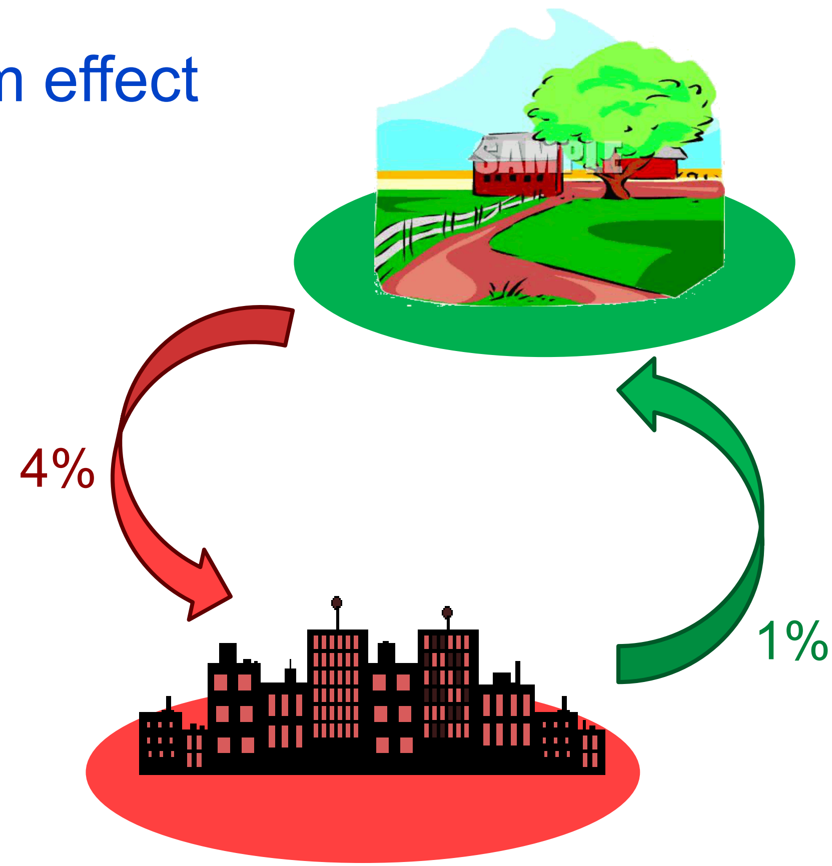
Section 6.1

Eigenvalues and Eigenvectors

An example (Example 6.1.1)

Suppose for each year, 4% of the rural population moves to the urban district while 1% of the urban population moves to the rural district.

We want to study the long term effect if things keep going like this.



Let

a_n = the rural population after n years,

b_n = the urban population after n years.

Examples (Example 6.1.1)

Observe that for $n = 1, 2, 3, \ldots,$

$$\begin{cases} a_n = 0.96\, a_{n-1} + 0.01\, b_{n-1} \\ b_n = 0.04\, a_{n-1} + 0.99\, b_{n-1} \end{cases}$$

$$\Leftrightarrow \quad \begin{bmatrix} a_n \\ b_n \end{bmatrix} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ b_{n-1} \end{bmatrix}.$$

Let $\boldsymbol{x_n} = \begin{bmatrix} a_n \\ b_n \end{bmatrix}$ and $\boldsymbol{A} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix}$.

Then $\boldsymbol{x_n} = \boldsymbol{A}\boldsymbol{x_{n-1}} = \boldsymbol{A}(\boldsymbol{A}\boldsymbol{x_{n-2}}) = \boldsymbol{A}^2\boldsymbol{x_{n-2}}$

$$= \boldsymbol{A}^3\boldsymbol{x_{n-3}} = \cdots = \boldsymbol{A}^n\boldsymbol{x_0}.$$

Examples (Example 6.1.1)

To study the long term effect, we need to compute $\boldsymbol{A}^n$ for large n.

If possible, we want to find $\lim_{n \to \infty} \boldsymbol{A}^n \boldsymbol{x_0}$.

It happen that we can write

$$\boldsymbol{A} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.95 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix}^{-1}.$$

Let $\boldsymbol{P} = \begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix}$ and $\boldsymbol{D} = \begin{bmatrix} 1 & 0 \\ 0 & 0.95 \end{bmatrix}$.

We have $\boldsymbol{A} = \boldsymbol{PDP}^{-1}$.

Examples (Example 6.1.1)

and $$\begin{aligned} \boldsymbol{A}^n &= (\boldsymbol{PDP}^{-1})^n \\ &= \underbrace{(\boldsymbol{PDP}^{-1})(\boldsymbol{PDP}^{-1})\cdots(\boldsymbol{PDP}^{-1})}_{n \text{ times}} \\ &= \boldsymbol{PD}(\boldsymbol{P}^{-1}\boldsymbol{P})\boldsymbol{D}(\boldsymbol{P}^{-1}\boldsymbol{P})\cdots(\boldsymbol{P}^{-1}\boldsymbol{P})\boldsymbol{DP}^{-1} \\ &= \boldsymbol{PDIDI}\cdots\boldsymbol{IDP}^{-1} \\ &= \boldsymbol{PDD}\cdots\boldsymbol{DP}^{-1} \\ &= \boldsymbol{PD}^n\boldsymbol{P}^{-1}. \end{aligned}$$

Note that $$\boldsymbol{D}^n = \begin{bmatrix} 1 & 0 \\ 0 & 0.95 \end{bmatrix}^n = \begin{bmatrix} 1^n & 0 \\ 0 & 0.95^n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0.95^n \end{bmatrix}.$$

See Question 2.15
and Discussion 6.2.10

Examples (Example 6.1.1)

As $\lim_{n \to \infty} \boldsymbol{D}^n = \lim_{n \to \infty} \begin{bmatrix} 1 & 0 \\ 0 & 0.95^n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$,

$$\lim_{n \to \infty} \boldsymbol{A}^n = \lim_{n \to \infty} \boldsymbol{P}\boldsymbol{D}^n\boldsymbol{P}^{-1}$$

$$= \begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 0.2 & 0.2 \\ 0.8 & 0.8 \end{bmatrix}.$$

and $\lim_{n \to \infty} \begin{bmatrix} a_n \\ b_n \end{bmatrix} = \lim_{n \to \infty} \boldsymbol{x_n} = \lim_{n \to \infty} \boldsymbol{A}^n \boldsymbol{x_0} = \begin{bmatrix} 0.2(a_0 + b_0) \\ 0.8(a_0 + b_0) \end{bmatrix}$.

Thus $\lim_{n \to \infty} a_n = 0.2(a_0 + b_0)$ and $\lim_{n \to \infty} b_n = 0.8(a_0 + b_0)$.

An example (Example 6.1.1 & Remark 6.1.2)

In the long run, 20% of the total population will stay in the rural district and 80% of the population will stay in the urban district.

In this example, the crucial step is to express $\boldsymbol{A}$ in the form $\boldsymbol{PDP}^{-1}$ where $\boldsymbol{D}$ is a diagonal matrix.

Eigenvalues and eigenvectors (Definition 6.1.3)

Let $\boldsymbol{A}$ be a square matrix of order n.

A nonzero column vector $\boldsymbol{u} \in \mathbb{R}^n$ is called an eigenvector of $\boldsymbol{A}$ if

$$\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u} \quad \text{for some scalar } \lambda.$$

The scalar λ is called an eigenvalue of $\boldsymbol{A}$

and $\boldsymbol{u}$ is said to be an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue λ.

Examples (Example 6.1.4.1)

Let $\boldsymbol{A} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Then

$$\boldsymbol{A}\boldsymbol{u} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \boldsymbol{u},$$

$$\boldsymbol{A}\boldsymbol{v} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0.95 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0.95\,\boldsymbol{v}.$$

So 1 and 0.95 are eigenvalues of $\boldsymbol{A}$,

$\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ associated with 1,

and $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A}$ associated with 0.95.

Examples (Example 6.1.4.2 modified)

Let $\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

Then

$$\boldsymbol{Bu} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 6 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 6\boldsymbol{u},$$

$$\boldsymbol{Bv} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0\boldsymbol{v},$$

The matrix $\boldsymbol{B}$ used here is different from that of Example 6.1.4.2 of the textbook.

Examples (Example 6.1.4.2 modified)

$$\boldsymbol{Bw} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0\boldsymbol{w}.$$

So 6 and 0 are eigenvalues of $\boldsymbol{B}$,

$\boldsymbol{u}$ is an eigenvector of $\boldsymbol{B}$ associated with 6,

and $\boldsymbol{v}$, $\boldsymbol{w}$ are eigenvectors of $\boldsymbol{B}$ associated with 0.

Note that

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

How to find eigenvalues (Remark 6.1.5)

Let $\boldsymbol{A}$ be a square matrix of order n.

λ is an eigenvalue of $\boldsymbol{A}$

$\Leftrightarrow$ $\boldsymbol{Au} = \lambda\boldsymbol{u}$ for some nonzero column vector $\boldsymbol{u} \in \mathbb{R}^n$

$\Leftrightarrow$ $\lambda\boldsymbol{u} - \boldsymbol{Au} = \boldsymbol{0}$ for some nonzero column vector $\boldsymbol{u} \in \mathbb{R}^n$

$\Leftrightarrow$ $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{u} = \boldsymbol{0}$ for some nonzero column vector $\boldsymbol{u} \in \mathbb{R}^n$

$\Leftrightarrow$ the linear system $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ has non-trivial solutions

$\Leftrightarrow$ $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$ (by Theorem 3.6.11).

If expanded, $\det(\lambda\boldsymbol{I} - \boldsymbol{A})$ is a polynomial in λ of degree n.

Characteristic polynomials (Definition 6.1.6)

Let $\boldsymbol{A}$ be a square matrix of order n.

The equation

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \mathbf{0}$$

is called the characteristic equation of $\boldsymbol{A}$;

and the polynomial

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A})$$

is called the characteristic polynomial of $\boldsymbol{A}$.

Examples (Example 6.1.7.1)

Let $\boldsymbol{A} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix}$.

The characteristic polynomial of $\boldsymbol{A}$ is

$$\begin{aligned}
\det(\lambda \boldsymbol{I} - \boldsymbol{A}) &= \det\left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix} \right) \\
&= \begin{vmatrix} \lambda - 0.96 & -0.01 \\ -0.04 & \lambda - 0.99 \end{vmatrix} \\
&= (\lambda - 0.96)(\lambda - 0.99) - (-0.01)(-0.04) \\
&= \lambda^2 - 1.95\lambda + 0.95 \\
&= (\lambda - 1)(\lambda - 0.95).
\end{aligned}$$

1 and 0.95 are eigenvalues of $\boldsymbol{A}$.

Examples (Example 6.1.7.2 modified)

Let $\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.

The characteristic polynomial of $\boldsymbol{B}$ is

$$\det(\lambda \boldsymbol{I} - \boldsymbol{B}) = \begin{vmatrix} \lambda - 1 & -2 & -1 \\ -2 & \lambda - 4 & -2 \\ -1 & -2 & \lambda - 1 \end{vmatrix}$$

$$= \lambda^3 - 6\lambda^2$$

$$= (\lambda - 6)(\lambda - 0)^2.$$

6 and 0 are eigenvalues of $\boldsymbol{B}$.

Examples (Example 6.1.7.3)

Let $\boldsymbol{C} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}$.

The characteristic polynomial of $\boldsymbol{C}$ is

$$\det(\lambda \boldsymbol{I} - \boldsymbol{C}) = \begin{vmatrix} \lambda - 0 & 1 & 0 \\ 0 & \lambda - 0 & -2 \\ 1 & -1 & \lambda - 1 \end{vmatrix}$$

$$= \lambda^3 - \lambda^2 - 2\lambda + 2$$

$$= (\lambda - 1)(\lambda - \sqrt{2})(\lambda + \sqrt{2}).$$

$1, \sqrt{2}$ and $-\sqrt{2}$ are eigenvalues of $\boldsymbol{C}$.

Invertible matrices (Theorem 6.1.8)

Let $\boldsymbol{A}$ be an $n \times n$ matrix. The following statements are equivalent:

1. $\boldsymbol{A}$ is invertible.
2. The linear system $\boldsymbol{Ax} = \boldsymbol{0}$ has only the trivial solution.
3. The reduced row-echelon form of $\boldsymbol{A}$ is an identity matrix.
4. $\boldsymbol{A}$ can be expressed as a product of elementary matrices.
5. $\det(\boldsymbol{A}) \neq 0$.
6. The rows of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$.
7. The columns of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$.
8. $\text{rank}(\boldsymbol{A}) = n$.
9. 0 is not an eigenvalue of $\boldsymbol{A}$.

Invertible matrices (Theorem 6.1.8)

We already learn that statements 1-8 are equivalent (see Theorem 3.6.11 and Remark 4.2.5.2).

To prove 9 $\Leftrightarrow$ 5:

0 is not an eigenvalue of $\boldsymbol{A}$

$\Leftrightarrow$ $\det(0\boldsymbol{I} - \boldsymbol{A}) \neq 0$ (by Remark 6.1.5)

$\Leftrightarrow$ $\det(-\boldsymbol{A}) \neq 0$

$\Leftrightarrow$ $(-1)^n \det(\boldsymbol{A}) \neq 0$ (by Theorem 2.5.22.1)

$\Leftrightarrow$ $\det(\boldsymbol{A}) \neq 0$.

Triangular matrices (Theorem 6.1.9)

If $\boldsymbol{A}$ is an $n \times n$ triangular matrix, say,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix},$$

> The arguments below are the same if $\boldsymbol{A}$ is a lower triangular matrix.

then

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ 0 & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda - a_{nn} \end{vmatrix}$$

$$= (\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn}).$$

So the eigenvalues of $\boldsymbol{A}$ are $a_{11}, \ a_{22}, \ \ldots, \ a_{nn}$.

Examples (Example 6.1.10)

The eigenvalues of $\begin{bmatrix} -1 & 3.5 & 14 \\ 0 & 5 & -26 \\ 0 & 0 & 2 \end{bmatrix}$ are -1, 5 and 2.

The eigenvalues of $\begin{bmatrix} -2 & 0 & 0 \\ 99 & 0 & 0 \\ 10 & -4.5 & 10 \end{bmatrix}$ are -2, 0 and 10.

Warning: In general, given a square matrix $\boldsymbol{A}$, you cannot find eigenvalues by reducing $\boldsymbol{A}$ to a triangular matrix by using row operations.

But you can use row operations to reduce $\lambda\boldsymbol{I} - \boldsymbol{A}$ to a triangular matrix in order to find $\det(\lambda\boldsymbol{I} - \boldsymbol{A})$.

Eigenspaces (Definition 6.1.11)

Let $\boldsymbol{A}$ be a square matrix of order n

and λ an eigenvalue of $\boldsymbol{A}$.

Then the solution space of the linear system

$$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$$

is called the eigenspace of $\boldsymbol{A}$ associated with the eigenvalue λ

and is denoted by E_λ (or $E_\lambda(\boldsymbol{A})$).

If $\boldsymbol{u}$ is a nonzero vector in E_λ, then $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ associated with λ.

Examples (Example 6.1.12.1)

Let $\boldsymbol{A} = \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix}$.

The eigenvalues of $\boldsymbol{A}$ are 1 and 0.95.

$$\underbrace{(1\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}}_{\boxed{\lambda = 1}} \quad \Leftrightarrow \quad \begin{bmatrix} 0.04 & -0.01 \\ -0.04 & 0.01 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow \quad \begin{bmatrix} x \\ y \end{bmatrix} = t \begin{bmatrix} 0.25 \\ 1 \end{bmatrix} \text{ for some } t \in \mathbb{R}.$$

So $E_1 = \text{span}\{\, (0.25,\ 1)^{\mathrm{T}} \,\}$.

Examples (Example 6.1.12.1)

$$(0.95\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \mathbf{0} \quad \Leftrightarrow \quad \begin{bmatrix} -0.01 & -0.01 \\ -0.04 & -0.04 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\lambda = 0.95$

$$\Leftrightarrow \quad \begin{bmatrix} x \\ y \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{for some } t \in \mathbb{R}.$$

So $E_{0.95} = \text{span}\{ (-1, 1)^T \}$.

Examples (Example 6.1.7.2 modified)

Let $\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$. The eigenvalues of $\boldsymbol{B}$ are 6 and 0.

$$(6\boldsymbol{I} - \boldsymbol{B})\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \begin{bmatrix} 5 & -2 & -1 \\ -2 & 2 & -2 \\ -1 & -2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\lambda = 6$

$$\Leftrightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \text{for some } t \in \mathbb{R}.$$

So $E_6 = \text{span}\{ (1, 2, 1)^T \}$.

Examples (Example 6.1.7.2 modified)

$$(0\boldsymbol{I} - \boldsymbol{B})\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \begin{bmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\boxed{\lambda = 0}$

$$\Leftrightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$ for some $s, t \in \mathbb{R}$.

So $E_0 = \text{span}\{ (-2, 1, 0)^T, \ (-1, 0, 1)^T \}$.

Examples (Example 6.1.7.3)

Let $\boldsymbol{C} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}$. The eigenvalues of $\boldsymbol{C}$ are $1, \sqrt{2}$ and $-\sqrt{2}$.

$$\underset{\boxed{\lambda = 1}}{(1\boldsymbol{I} - \boldsymbol{C})\boldsymbol{x} = \boldsymbol{0}} \Leftrightarrow \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -2 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \text{ for some } t \in \mathbb{R}.$$

So $E_1 = \text{span}\{ (-2, 2, 1)^T \}$.

Examples (Example 6.1.7.3)

Similarly,

$$E_{\sqrt{2}} = \text{span}\{\, (-1, \sqrt{2}, 1)^T \,\}$$

and

$$E_{-\sqrt{2}} = \text{span}\{\, (-1, -\sqrt{2}, 1)^T \,\}.$$

Examples (Example 6.1.7.4)

Let $\boldsymbol{M} = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$.

Since $\boldsymbol{M}$ is a lower triangular matrix, (by Theorem 6.1.9) the eigenvalues of $\boldsymbol{M}$ are the diagonal entries of $\boldsymbol{M}$.

So 2 is the only eigenvalue of $\boldsymbol{M}$.

$$(2\boldsymbol{I} - \boldsymbol{M})\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} = t\begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{for some } t \in \mathbb{R}.$$

So $E_2 = \text{span}\{ (0, 1)^T \}$.

Section 6.2

Diagonalization

Diagonalization (Definition 6.2.1 & Example 6.2.2.1 & Example 6.2.2.2 modified)

A square matrix $\boldsymbol{A}$ is called diagonalizable if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

The matrix $\boldsymbol{P}$ is said to diagonalize $\boldsymbol{A}$.

For example,

The matrix is diagonalizable.

$$\begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 4 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0.95 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The matrix is diagonalizable.

Diagonalization (Example 6.2.2.3)

Not all square matrices are diagonalizable.

Let $\boldsymbol{M} = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$. The matrix $\boldsymbol{M}$ is not diagonalizable.

Proof: Assume that the matrix $\boldsymbol{M}$ is diagonalizable,

i.e. there exists an invertible matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$$

for some constant λ and μ.

Diagonalization (Example 6.2.2.3)

Then $$\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$$

$\Rightarrow$ $2a = a\lambda$, $2b = b\mu$, $a + 2c = c\lambda$ and $b + 2d = d\mu$.

$\Rightarrow$ $a = 0$ and $b = 0$.

However, the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix}$ is not invertible which contradicts our assumption.

So $\boldsymbol{M}$ is not diagonalizable.

In Algorithm 6.2.4, we shall learnt a systematic method to check whether a square matrix is diagonalizable.

Diagonalization and eigenvectors (Theorem 6.2.3)

Let $\boldsymbol{A}$ be a square matrix of order n.

Then $\boldsymbol{A}$ is diagonalizable if and only if $\boldsymbol{A}$ has n linearly independent eigenvectors.

Proof:

($\Rightarrow$) Suppose $\boldsymbol{A}$ is diagonalizable.

Let $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ be an invertible matrix such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix, say,

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Diagonalization and eigenvectors (Theorem 6.2.3)

$$\Rightarrow \quad \boldsymbol{A}\begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$\Rightarrow \quad \begin{bmatrix} \boldsymbol{Au_1} & \boldsymbol{Au_2} & \cdots & \boldsymbol{Au_n} \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{u_1} & \lambda_2\boldsymbol{u_2} & \cdots & \lambda_n\boldsymbol{u_n} \end{bmatrix}$$

$$\Rightarrow \quad \boldsymbol{Au_1} = \lambda_1\boldsymbol{u_1}, \quad \boldsymbol{Au_2} = \lambda_2\boldsymbol{u_2}, \quad \ldots, \quad \boldsymbol{Au_n} = \lambda_n\boldsymbol{u_n}.$$

Since $\boldsymbol{P}$ is invertible, (by Theorem 3.6.11 or Theorem 6.1.8) the columns of $\boldsymbol{P}$, $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, ..., $\boldsymbol{u_n}$, form a basis for $\mathbb{R}^n$.

So we have n linearly independent eigenvectors.

Diagonalization and eigenvectors (Theorem 6.2.3)

($\Leftarrow$) Suppose $\boldsymbol{A}$ has n linearly independent eigenvectors: $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, ..., $\boldsymbol{u}_n$,

say, $\boldsymbol{A}\boldsymbol{u}_1 = \lambda_1\boldsymbol{u}_1$, $\boldsymbol{A}\boldsymbol{u}_2 = \lambda_2\boldsymbol{u}_2$, ..., $\boldsymbol{A}\boldsymbol{u}_n = \lambda_n\boldsymbol{u}_n$ where λ_1, λ_2, ..., λ_n are eigenvalues.

Let $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ which is an $n \times n$ matrix.

Since $\{ \boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_n \}$ is linear independent and $\dim(\mathbb{R}^n) = n$, (by Theorem 3.6.7) $\{ \boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_n \}$ is a basis for $\mathbb{R}^n$.

Then (by Theorem 3.6.11 or Theorem 6.1.8) $\boldsymbol{P}$ is invertible.

Diagonalization and eigenvectors (Theorem 6.2.3)

$$\boldsymbol{AP} = \boldsymbol{A}\begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{Au}_1 & \boldsymbol{Au}_2 & \cdots & \boldsymbol{Au}_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1\boldsymbol{u}_1 & \lambda_2\boldsymbol{u}_2 & \cdots & \lambda_n\boldsymbol{u}_n \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \boldsymbol{P}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$\Rightarrow \quad \boldsymbol{P}^{-1}\boldsymbol{AP} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

So $\boldsymbol{A}$ is diagonalizable.

Diagonalize a matrix (Algorithm 6.2.4)

Let $\boldsymbol{A}$ be a square matrix of order n.

Step 1: Find all distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ (say, by solving the characteristic equation $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$).

Step 2: For each eigenvalue λ_i, find a basis S_{λ_i} for the eigenspace E_{λ_i}.

Step 3: Let $S = S_{\lambda_1} \cup S_{\lambda_2} \cup \cdots \cup S_{\lambda_k}$.

(a) If $|S| < n$, then $\boldsymbol{A}$ is not diagonalizable.

(b) If $|S| = n$, say, $S = \{ \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n \}$,

then $\boldsymbol{A}$ is diagonalizable

and $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ is an invertible matrix

that diagonalizes $\boldsymbol{A}$.

Diagonalize a matrix (Remark 6.2.5.1)

In Step 1, sometimes, the matrix $\boldsymbol{A}$ may have eigenvalues that are not real numbers but complex numbers.

For example, let $\boldsymbol{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Then $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda \end{vmatrix} = (\lambda - i)(\lambda + i)$ where $i = \sqrt{-1}$.

Hence the eigenvalues of $\boldsymbol{A}$ are i and $-i$.

We can still use the algorithm to diagonalize the matrix (but then some entries of $\boldsymbol{P}$ are complex numbers).

Diagonalize a matrix (Remark 6.2.5.2-3)

Suppose the characteristic polynomial of the matrix $\boldsymbol{A}$ can be factorized as

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k}$$

where λ_1, λ_2, ..., λ_k are distinct eigenvalues of $\boldsymbol{A}$.

Then for each λ_i, $\dim(E_{\lambda_i}) \leq r_i$.

Thus $\boldsymbol{A}$ is diagonalizable if and only if in Step 2, for each eigenvalue λ_i, $\dim(E_{\lambda_i}) = r_i$, i.e. $|S_{\lambda_i}| = r_i$.

In Step 3, the set S is always linearly independent.

Examples (Example 6.2.6.1 modified)

Let $\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.

Step 1: The eigenvalues of $\boldsymbol{B}$ are 6 and 0.

(See Example 6.1.4.2 modified.)

Step 2: $S_6 = \{ (1, 2, 1)^T \}$ is a basis for E_6

and $S_0 = \{ (-2, 1, 0)^T, (-1, 0, 1)^T \}$ is a basis for E_0.

(See Example 6.1.12.2 modified.)

Step 3: $S = \{ (1, 2, 1)^T, (-2, 1, 0)^T, (-1, 0, 1)^T \}$.

Examples (Example 6.2.6.1 modified)

Let $\boldsymbol{P} = \begin{bmatrix} 1 & -2 & -1 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. Then $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

In Step 3(b), if we rearrange the order of columns of $\boldsymbol{P}$, what will happen to the diagonal matrix $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}$?

For example, let $\boldsymbol{Q} = \begin{bmatrix} -2 & 1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.

Then $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Examples (Example 6.2.6.2)

Let $\boldsymbol{C} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}$.

Step 1: The eigenvalues of $\boldsymbol{C}$ are $1, \sqrt{2}$ and $-\sqrt{2}$.

(See Example 6.1.4.3.)

Step 2: $S_1 = \{ (-2, 2, 1)^T \}$ is a basis for E_1

$S_{\sqrt{2}} = \{ (-1, \sqrt{2}, 1)^T \}$ is a basis for $E_{\sqrt{2}}$

and $S_{-\sqrt{2}} = \{ (-1, -\sqrt{2}, 1)^T \}$ is a basis for $E_{-\sqrt{2}}$.

(See Example 6.1.12.3.)

Step 3: $S = \{ (-2, 2, 1)^T, (-1, \sqrt{2}, 1)^T, (-1, -\sqrt{2}, 1)^T \}$.

Examples (Example 6.2.6.1)

Let $\boldsymbol{P} = \begin{bmatrix} -2 & -1 & -1 \\ 2 & \sqrt{2} & -\sqrt{2} \\ 1 & 1 & 1 \end{bmatrix}$.

Then $\boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & -\sqrt{2} \end{bmatrix}$.

Examples (Example 6.2.6.3)

Let $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ -3 & 5 & 2 \end{bmatrix}$.

Step 1: The eigenvalues of $\boldsymbol{A}$ are 1 and 2.

(See Theorem 6.1.9.)

Step 2: For $\lambda = 1$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ 3 & -5 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} 1 \\ -1 \\ 8 \end{bmatrix} \text{ for some } t \in \mathbb{R}.$$

So $S_1 = \{ (1, -1, 8)^T \}$ is a basis for E_1.

Examples (Example 6.2.6.3)

For $\lambda = 2$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ -3 & -5 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} = t \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ for some } t \in \mathbb{R}.$$

So $S_2 = \{ (0, 0, 1)^T \}$ is a basis for E_2.

Step 3: $S = \{ (1, -1, 8)^T, (0, 0, 1)^T \}$.

Since $|S| = 2 < 3$, $\boldsymbol{A}$ is not diagonalizable.

(Since the characteristic polynomial of $\boldsymbol{A}$ is

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - 1)(\lambda - 2)^2.$$

By Remark 6.2.5.2, after computing S_2, we already know that $\boldsymbol{A}$ is not diagonalizable.)

Diagonalization and eigenvalues (Theorem 6.2.7 & Example 6.2.8)

Let $\boldsymbol{A}$ be a square matrix of order n. If $\boldsymbol{A}$ has n distinct eigenvalues, then $\boldsymbol{A}$ is diagonalizable.

(Please read our textbook for a proof of the result.)

Let $\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 4 \end{bmatrix}$.

The eigenvalues of $\boldsymbol{A}$ are 1, 2, 3 and 4.

(See Theorem 6.1.9.)

Since $\boldsymbol{A}$ is a 4×4 matrix with 4 distinct eigenvalues, $\boldsymbol{A}$ is diagonalizable.

Power of matrices (Remark 6.2.10)

Let $\boldsymbol{A}$ be a square matrix of order n and $\boldsymbol{P}$ an invertible matrix such that

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

For any positive integer m,

$$\boldsymbol{A}^m = \boldsymbol{P} \begin{bmatrix} {\lambda_1}^m & 0 & \cdots & 0 \\ 0 & {\lambda_2}^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\lambda_n}^m \end{bmatrix} \boldsymbol{P}^{-1}.$$

Power of matrices (Remark 6.2.10)

Suppose $\boldsymbol{A}$ is invertible, i.e. $\lambda_i \neq 0$ for all i. (See Theorem 6.1.8.)

Then

$$\boldsymbol{A}^{-1} = \boldsymbol{P}\begin{bmatrix} \lambda_1^{-1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^{-1} \end{bmatrix}\boldsymbol{P}^{-1}$$

and any positive integer m,

$$\boldsymbol{A}^{-m} = \boldsymbol{P}\begin{bmatrix} \lambda_1^{-m} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-m} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^{-m} \end{bmatrix}\boldsymbol{P}^{-1}.$$

An example (Example 6.2.11.1)

Let $\boldsymbol{A} = \begin{bmatrix} -4 & 0 & -6 \\ 2 & 1 & 2 \\ 3 & 0 & 5 \end{bmatrix}$.

First (following Algorithm 6.2.4) we find an invertible matrix

$$\boldsymbol{P} = \begin{bmatrix} -2 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \text{ such that } \boldsymbol{P}^{-1}\boldsymbol{AP} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Then $\boldsymbol{A}^m = \boldsymbol{P} \begin{bmatrix} (-1)^m & 0 & 0 \\ 0 & 1^m & 0 \\ 0 & 0 & 2^m \end{bmatrix} \boldsymbol{P}^{-1}$.

An example (Example 6.2.11.1)

In particular,

$$\begin{bmatrix} -4 & 0 & -6 \\ 2 & 1 & 2 \\ 3 & 0 & 5 \end{bmatrix}^{10} = \begin{bmatrix} -2 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} (-1)^{10} & 0 & 0 \\ 0 & 1^{10} & 0 \\ 0 & 0 & 2^{10} \end{bmatrix} \begin{bmatrix} -2 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} -1022 & 0 & -2046 \\ 0 & 1 & 0 \\ 1023 & 0 & 2047 \end{bmatrix}.$$

Fibonacci numbers (Example 6.2.11.2)

Let $(a_0, a_1, a_2, \ldots)$ be a sequence of numbers such that

$$a_0 = 0,$$

$$a_1 = 1$$

and $$a_n = a_{n-1} + a_{n-2} \text{ for } n \geq 2.$$

(These numbers are known as the Fibonacci numbers.)

First, we formulate the problem in terms of a matrix equation:

$$\begin{cases} a_n = \phantom{a_{n-1} +} a_n \\ a_{n+1} = a_{n-1} + a_n \end{cases} \Leftrightarrow \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_n \end{bmatrix}.$$

Fibonacci numbers (Example 6.2.11.2)

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_n \end{bmatrix}$$

Let $\boldsymbol{x_n} = \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix}$ and $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$.

Then $\boldsymbol{x_n} = \boldsymbol{A}\boldsymbol{x_{n-1}} = \boldsymbol{A}^2\boldsymbol{x_{n-2}} = \cdots = \boldsymbol{A}^n\boldsymbol{x_0}$.

In order to compute $\boldsymbol{A}^n$, (following Algorithm 6.2.4) we find an invertible matrix

$$\boldsymbol{P} = \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix} \text{ such that } \boldsymbol{P}^{-1}\boldsymbol{AP} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}.$$

Fibonacci numbers (Example 6.2.11.2)

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \boldsymbol{x}_n = \boldsymbol{A}^n \boldsymbol{x}_0$$

$$= \overbrace{\begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix}^{-1}}^{\boldsymbol{A}^n} \overbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}^{\boldsymbol{x}_0}$$

$$= \begin{bmatrix} \underbrace{\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n}_{a_n} \\ \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \end{bmatrix}$$

Section 6.3

Orthogonal Diagonalization

An example (Example 6.3.1 modified)

Let $\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.

When trying to diagonalize $\boldsymbol{B}$ (as in Example 6.2.6.1 modified) instead of using bases

$$S_6 = \{ (1, 2, 1)^T \} \text{ and } S_0 = \{ (-2, 1, 0)^T, (-1, 0, 1)^T \}$$

for E_6 and E_0, respectively,

we choose orthonormal bases

$$T_3 = \left\{ \frac{1}{\sqrt{6}}(1, 2, 1)^T \right\} \text{ and } T_0 = \left\{ \frac{1}{\sqrt{5}}(-2, 1, 0)^T, \frac{1}{\sqrt{30}}(-1, -2, 5)^T \right\}$$

for E_6 and E_0, respectively.

An example (Example 6.3.1 modified)

Let $\boldsymbol{R} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{5}} & \frac{-1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{5}{\sqrt{30}} \end{bmatrix}$.

Note that $\boldsymbol{R}$ is an orthogonal matrix,

i.e. $\boldsymbol{R}^{-1} = \boldsymbol{R}^{\mathrm{T}}$.

Thus $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{R} = \boldsymbol{R}^{-1}\boldsymbol{B}\boldsymbol{R} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Orthogonal diagonalization (Definition 6.2.2 & Remark 6.3)

A square matrix $\boldsymbol{A}$ is called orthogonally diagonalizable if there exists an $\boxed{\text{orthogonal matrix } \boldsymbol{P}}$ such that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

i.e. $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}$

The matrix $\boldsymbol{P}$ is said to orthogonally diagonalize $\boldsymbol{A}$.

(The orthogonal diagonalization gives us a perfect tool to study a widely used mathematical object called "quadratic forms" which will be discussed in Section 6.4.)

Orthogonal diagonalization (Theorem 6.3.4)

A square matrix $\boldsymbol{A}$ is orthogonally diagonalizable if and only if $\boldsymbol{A}$ is symmetric, i.e. $\boldsymbol{A}^{\mathsf{T}} = \boldsymbol{A}$.

Proof: (We only prove the "only if" part.)

($\Rightarrow$) Suppose $\boldsymbol{A}$ is orthogonally diagonalizable, i.e. there exists an orthogonal matrix $\boldsymbol{P}$ such that

$\boldsymbol{P}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$ where $\boldsymbol{D}$ is a diagonal matrix.

(Diagonal matrices are symmetric, i.e. $\boldsymbol{D}^{\mathsf{T}} = \boldsymbol{D}$.)

Since $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathsf{T}}$,

$$\boldsymbol{A} = (\boldsymbol{P}^{\mathsf{T}})^{-1}\boldsymbol{D}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}.$$

So $\boldsymbol{A}^{\mathsf{T}} = (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}})^{\mathsf{T}} = (\boldsymbol{P}^{\mathsf{T}})^{\mathsf{T}}\boldsymbol{D}^{\mathsf{T}}\boldsymbol{P}^{\mathsf{T}} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}} = \boldsymbol{A}$.

Orthogonally diagonalize a matrix (Algorithm 6.3.5)

Let $\boldsymbol{A}$ be a symmetric matrix of order n.

Step 1: Find all distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ (by solving the characteristic equation $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \boldsymbol{0}$).

Step 2: Find each eigenvalue λ_i,

(a) find a basis S_{λ_i} for the eigenspace E_{λ_i}, and then

(b) use the Gram-Schmidt Process (Theorem 5.2.19) to transform S_{λ_i} to an orthonormal basis T_{λ_i}.

Step 3: Let $T = T_{\lambda_1} \cup T_{\lambda_2} \cup \cdots \cup T_{\lambda_k}$,

say, $T = \{ \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \}$.

Then $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ is an orthogonal matrix that orthogonally diagonalizes $\boldsymbol{A}$.

Orthogonally diagonalize a matrix (Remark 6.3.6)

In Step 1, the eigenvalues of a symmetric matrix are always real numbers.

Suppose the characteristic polynomial of the symmetric matrix $\boldsymbol{A}$ can be factorized as

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k}$$

where λ_1, λ_2, ..., λ_k are distinct eigenvalues of $\boldsymbol{A}$.

In Step 2, for each λ_i, $\dim(E_{\lambda_i}) = r_i$, i.e. $|S_{\lambda_i}| = |T_{\lambda_i}| = r_i$. (See Remark 6.2.5.2.)

In Step 3, the set T is always orthonormal.

Since T is always orthonormal, (by Theorem 5.4.6) the matrix $\boldsymbol{P}$ in Step 3 is an orthogonal matrix.

Orthogonally diagonalize a matrix (Remark 6.3.6.3)

If $\boldsymbol{A}$ is symmetric, eigenvectors from different eigenspaces of $\boldsymbol{A}$ are always orthogonal to each other.

Proof. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be eigenvectors from eigenspaces E_λ and E_μ, respectively, with $\lambda \neq \mu$. Then

$$(\boldsymbol{Au}) \cdot \boldsymbol{v} = (\boldsymbol{Au})^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} \underset{\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}}{=} \boldsymbol{u}^{\mathrm{T}}\boldsymbol{Av} = \boldsymbol{u}^{\mathrm{T}}(\boldsymbol{Av}) = \boldsymbol{u} \cdot (\boldsymbol{Av}).$$

So

$$\lambda(\boldsymbol{u} \cdot \boldsymbol{v}) = (\lambda\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (\mu\boldsymbol{v}) = \mu(\boldsymbol{u} \cdot \boldsymbol{v}).$$

$\boldsymbol{Au} = \lambda\boldsymbol{u}$ and $\boldsymbol{Av} = \mu\boldsymbol{v}$

Since $\lambda \neq \mu$, we have $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ and hence $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal to each other.

Examples (Example 6.3.7.1)

Let $\boldsymbol{A} = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$.

Step 1: The characteristic polynomial of $\boldsymbol{A}$ is

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \lambda^2 - 2\lambda + \frac{3}{4} = (\lambda - \tfrac{1}{2})(\lambda - \tfrac{3}{2}).$$

The eigenvalues of $\boldsymbol{A}$ are $\frac{1}{2}$ and $\frac{3}{2}$.

Step 2: For $\lambda = \frac{1}{2}$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$ for some $t \in \mathbb{R}$.

Examples (Example 6.3.7.1)

So $S_{\frac{1}{2}} = \{ (1, 1)^T \}$ is a basis for $E_{\frac{1}{2}}$.

Then $T_{\frac{1}{2}} = \left\{ \dfrac{1}{||(1, 1)^T||}(1, 1)^T \right\} = \left\{ \dfrac{1}{\sqrt{2}}(1, 1)^T \right\}$

is an orthonormal basis for $E_{\frac{1}{2}}$.

For $\lambda = \dfrac{3}{2}$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{for some } t \in \mathbb{R}.$$

So $S_{\frac{3}{2}} = \{ (-1, 1)^T \}$ is a basis for $E_{\frac{3}{2}}$.

Examples (Example 6.3.7.1)

Then $T_{\frac{3}{2}} = \left\{ \frac{1}{\|(-1,\ 1)^{\mathrm{T}}\|}(-1,\ 1)^{\mathrm{T}} \right\} = \left\{ \frac{1}{\sqrt{2}}(-1,\ 1)^{\mathrm{T}} \right\}$

is an orthonormal basis for $E_{\frac{3}{2}}$.

Step 3:

Let $\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$.

Then $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}$.

Examples (Example 6.3.7.2)

Let $\boldsymbol{B} = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{bmatrix}$.

Step 1: The characteristic polynomial of $\boldsymbol{A}$ is

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \lambda^4 - 8\lambda^3 + 16\lambda^2 = \lambda^2(\lambda - 4)^2.$$

The eigenvalues of $\boldsymbol{A}$ are 0 and 4.

Examples (Example 6.3.7.2)

Step 2: For $\lambda = 0$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -3 & -1 \\ 1 & -1 & -1 & -3 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = s \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

for some $s, t \in \mathbb{R}$.

$S_0 = \{ (1, 1, 0, 0)^T, \ (2, 0, -1, 1)^T \}$ is a basis for E_0.

By the Gram-Schmidt Process, we obtain an orthonormal basis for E_0:

$$T_0 = \left\{ \frac{1}{\sqrt{2}}(1, 1, 0, 0)^T, \ \frac{1}{2}(1, -1, -1, 1)^T \right\}.$$

Examples (Example 6.3.7.2)

For $\lambda = 4$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} 3 & 1 & -1 & 1 \\ 1 & 3 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = s \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

for some $s, t \in \mathbb{R}$.

$S_4 = \{ (1, -1, 2, 0)^{\mathrm{T}}, (-1, 1, 0, 2)^{\mathrm{T}} \}$ is a basis for E_4.

By the Gram-Schmidt Process, we obtain an orthonormal basis for E_0:

$$T_4 = \left\{ \frac{1}{\sqrt{6}}(1, -1, 2, 0)^{\mathrm{T}}, \frac{1}{\sqrt{12}}(-1, 1, 1, 3)^{\mathrm{T}} \right\}.$$

Examples (Example 6.3.7.2)

Step 3:

Let $\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{12}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{2} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ 0 & \frac{-1}{2} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ 0 & \frac{1}{2} & 0 & \frac{3}{\sqrt{12}} \end{bmatrix}$.

Then $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$.

Section 6.4

Quadratic Forms and Conic Sections

Quadratic forms (Definition 6.4.1)

The expression

$$
\begin{aligned}
Q(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=i}^{n} q_{ij} x_i x_j
&= q_{11}x_1^2 + q_{12}x_1x_2 + q_{13}x_1x_3 + \cdots + q_{1n} x_1x_n \\
&\quad + q_{22}x_2^2 + q_{23}x_2x_3 + \cdots + q_{2n} x_2x_n \\
&\quad + q_{33}x_3^2 + \cdots + q_{3n} x_3x_n \\
&\quad + \cdots + \\
&\quad + q_{nn} x_n^2,
\end{aligned}
$$

where q_{ij}'s are real numbers, is called a quadratic form in n variables x_1, x_2, ..., x_n.

Quadratic forms (Definition 6.4.1)

Define an $n \times n$ symmetric matrix $\boldsymbol{A} = (a_{ij})$ such that

$$a_{ij} = \begin{cases} q_{ii} & \text{if } i = j \\ \frac{1}{2}q_{ij} & \text{if } i < j \\ \frac{1}{2}q_{ji} & \text{if } i > j. \end{cases}$$

Then

$$Q(x_1, x_2, \ldots, x_n) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \overbrace{\begin{bmatrix} q_{11} & \frac{1}{2}q_{12} & \cdots & \frac{1}{2}q_{1n} \\ \frac{1}{2}q_{12} & q_{22} & \cdots & \frac{1}{2}q_{2n} \\ \vdots & \vdots & & \vdots \\ \frac{1}{2}q_{1n} & \frac{1}{2}q_{2n} & \cdots & q_{nn} \end{bmatrix}}^{\boldsymbol{A}} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$.

Quadratic forms (Definition 6.4.1)

Proof: With $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ and $\boldsymbol{A} = (a_{ij})$,

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{i1} x_i & \sum_{i=1}^{n} a_{i2} x_i & \cdots & \sum_{i=1}^{n} a_{in} x_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} a_{i1} x_i x_1 + \sum_{i=1}^{n} a_{i2} x_i x_2 + \cdots + \sum_{i=1}^{n} a_{in} x_i x_n$$

Quadratic forms (Definition 6.4.1)

$$
\begin{aligned}
= \quad & a_{11}x_1x_1 && + a_{12}x_1x_2 && + \cdots + a_{1n}x_1x_n \\
+ \; & a_{21}x_2x_1 && + a_{22}x_2x_2 && + \cdots + a_{2n}x_2x_n \\
& \quad \vdots && \quad \vdots && \qquad \vdots \\
+ \; & a_{n1}x_nx_1 && + a_{n2}x_nx_2 && + \cdots + a_{nn}x_nx_n
\end{aligned}
$$

$$
\begin{aligned}
= \quad & q_{11}x_1^2 && + q_{12}x_1x_2 && + \cdots + q_{1n}x_1x_n \\
& && + q_{22}x_2^2 && + \cdots + q_{2n}x_2x_n \\
& && && \ddots \qquad \vdots \\
& && && \qquad + q_{nn}x_n^2
\end{aligned}
$$

$\boxed{q_{ii} = a_{ii}}$

$$= Q(x_1, x_2, \ldots, x_n).$$

$\boxed{q_{ij} = a_{ji} + a_{ij} \text{ for } i < j}$

Quadratic forms (Definition 6.4.1 & Example 6.4.2)

The quadratic form can also be regarded as a mapping $Q: \mathbb{R}^n \to \mathbb{R}$ defined by $Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$.

For examples, $Q_1(x, y) = x^2 - xy + y^2$ is a quadratic form in x, y which can be written as:

$$Q_1(x, y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

and $Q_2(x, y, z) = x^2 + 2y^2 + z^2 + 2xz$ is a quadratic form in x, y, z which can be written as:

$$Q_2(x, y, z) = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

An application of quadratic forms (Remarks 6.4.3)

Quadratic forms appear quite often in various areas.

For example, in probability and statistics, the density function of a multivariate normal distribution of n random variables is given by

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{A})}} e^{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{A}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}$$

for $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ and $\boldsymbol{A}$ is an $n \times n$ symmetric matrix.

Simplifying quadratic forms (Remarks 6.4.3)

Let $Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ be a quadratic form in n variables x_1, x_2, ..., x_n where $\boldsymbol{A}$ is an $n \times n$ symmetric matrix and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$.

Since $\boldsymbol{A}$ is symmetric, (by Theorem 6.3.4) it is orthogonally diagonalizable,

i.e. (by Algorithm 6.3.5) we can find an orthogonal matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Simplifying quadratic forms (Remarks 6.4.3)

Define new variables $y_1, \; y_2, \; \ldots, \; y_n$ such that

$$\boldsymbol{y} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{P}^{-1}\boldsymbol{x}$$

where $\boldsymbol{y} = (y_1, y_2, \; \ldots, y_n)^{\mathrm{T}}$.

Then

$$\begin{aligned} Q(\boldsymbol{x}) &= Q(\boldsymbol{Py}) = (\boldsymbol{Py})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{Py}) \\ &= \boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{APy} \\ &= \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \lambda_1 {y_1}^2 + \lambda_2 {y_2}^2 + \cdots + \lambda_n {y_n}^2. \end{aligned}$$

Examples (Example 6.4.5.1)

Consider $Q_1(x, y) = x^2 - xy + y^2$.

We find an orthogonal matrix (by Algorithm 6.4.5)

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

such that

$$\boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}.$$

Examples (Example 6.4.5.1)

Define new variables x', y' such that

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(x+y) \\ \frac{1}{\sqrt{2}}(-x+y) \end{bmatrix}.$$

Then

$$Q_1(x, y) = \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

$$= \frac{1}{2}x'^2 + \frac{3}{2}y'^2 = \frac{1}{4}(x+y)^2 + \frac{3}{4}(-x+y)^2.$$

Examples (Example 6.4.5.2)

Consider $Q_2(x, y, z) = x^2 + 2y^2 + z^2 + 2xz$.

We find an orthogonal matrix (by Algorithm 6.4.5)

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

such that

$$\boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \boldsymbol{P} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Examples (Example 6.4.5.2)

Define new variables x', y', z' such that

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(x+z) \\ y \\ \frac{1}{\sqrt{2}}(-x+z) \end{bmatrix}.$$

Then

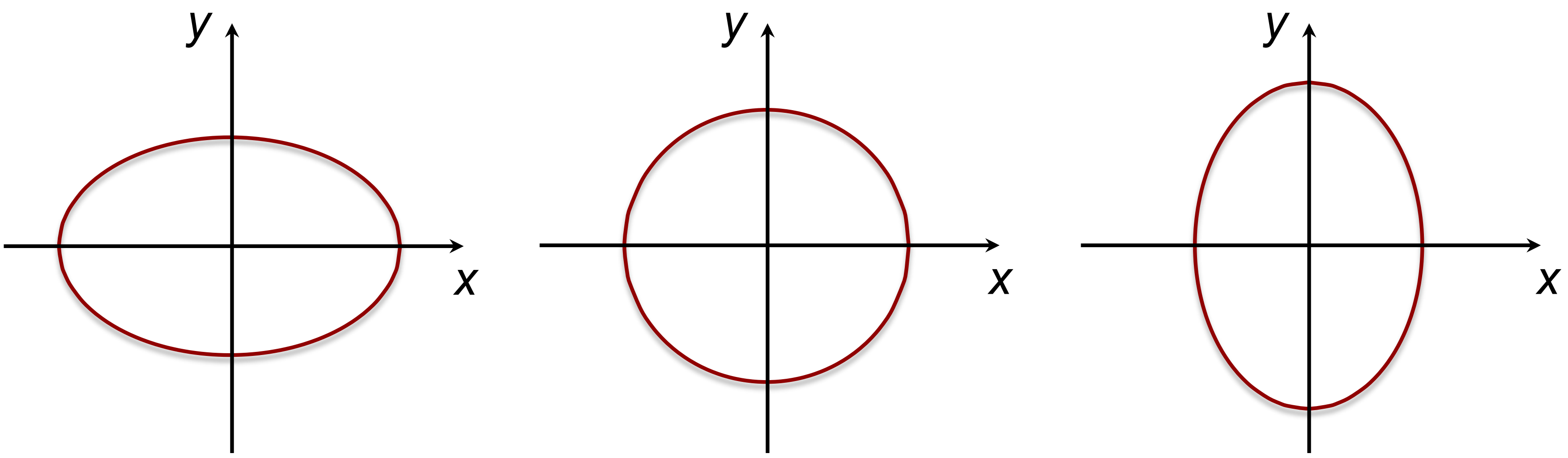
$$Q_2(x, y, z) = \begin{bmatrix} x' & y' & z' \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$$

$$= 2x'^2 + 2y'^2 = (x+z)^2 + 2y^2.$$

Circles and ellipses (Definition 6.4.6.1)

Circle or ellipse (in standard form):

$$\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1, \text{ i.e. } \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1.$$



$|\alpha| > |\beta|$ $|\alpha| = |\beta|$ $|\alpha| < |\beta|$

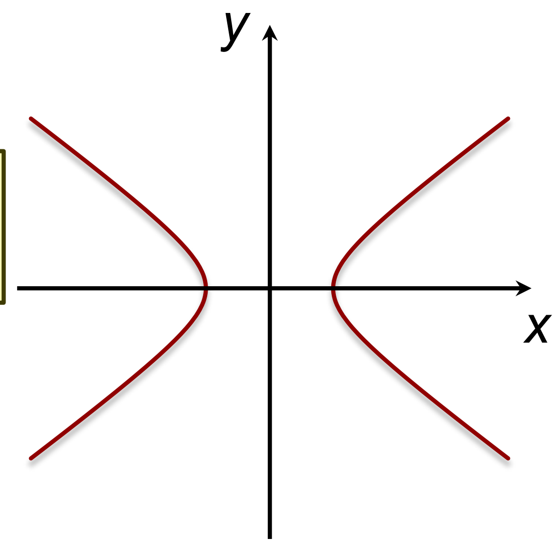
Hyperbolas (Definition 6.4.6.2)

Hyperbola (in standard form):

$$\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1,$$

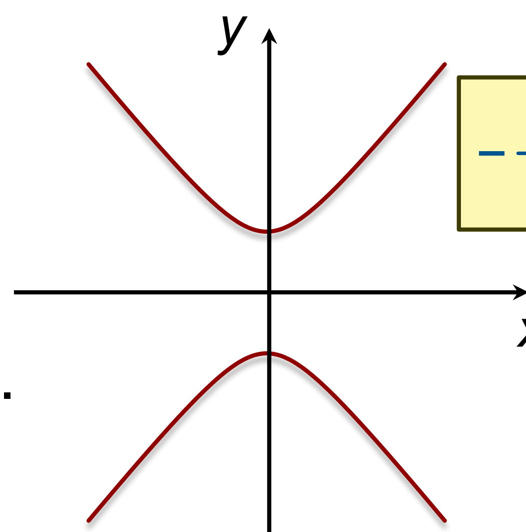
i.e. $$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & -\frac{1}{\beta^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1,$$

$$\boxed{\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1}$$



or $$-\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1,$$

i.e. $$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} -\frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1.$$



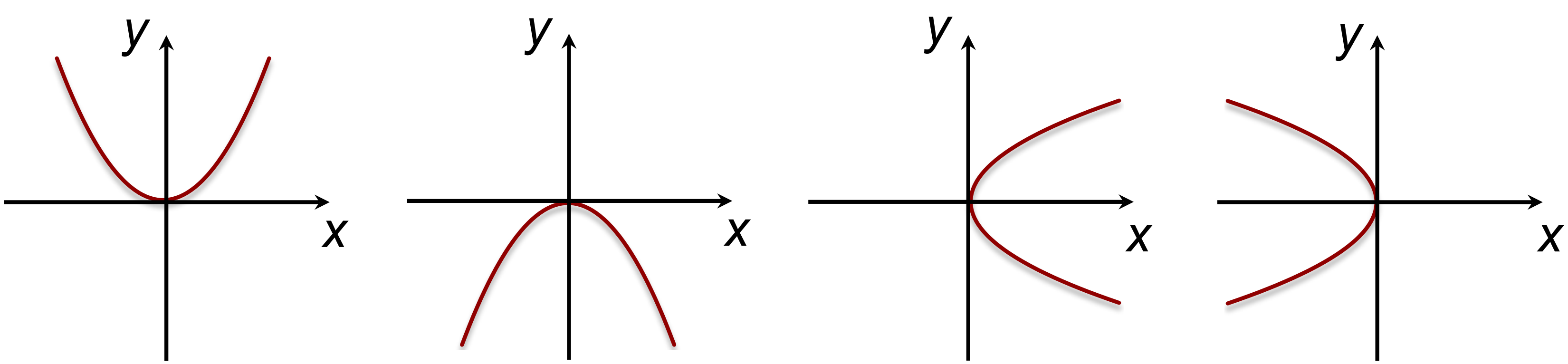
$$\boxed{-\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1}$$

Parabolas (Definition 6.4.6.3)

Parabolas (in standard form):

$$x^2 = \alpha y, \quad \text{i.e.} \quad \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 & -\alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0,$$

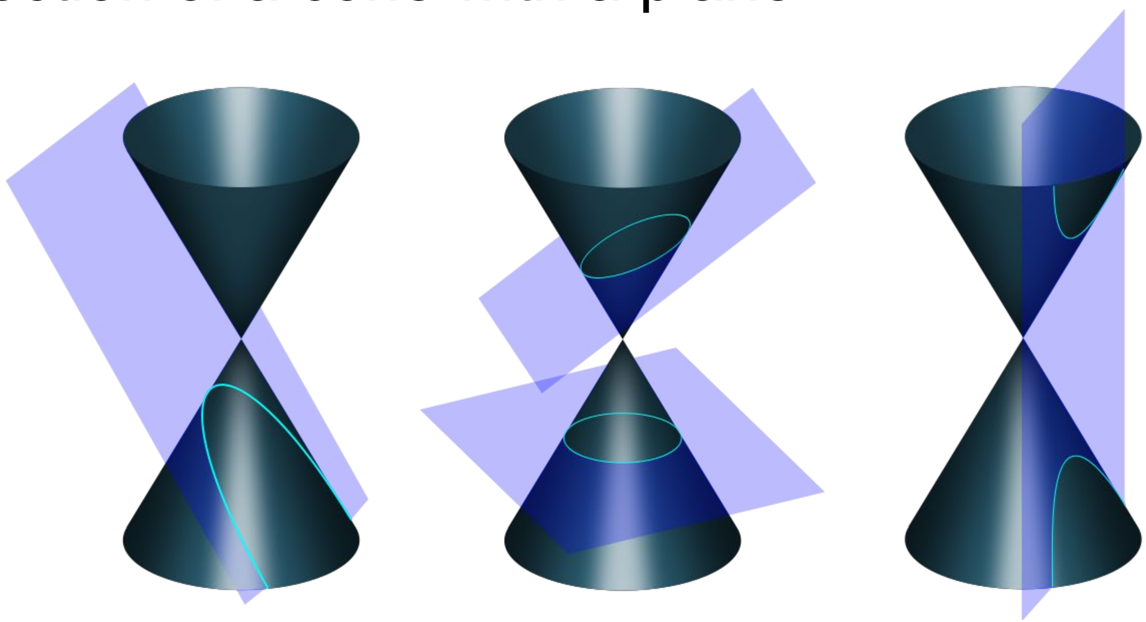
$$\text{or} \quad y^2 = \alpha x, \quad \text{i.e.} \quad \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -\alpha & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0.$$



$x^2 = \alpha y$, $\alpha > 0$ | $x^2 = \alpha y$, $\alpha < 0$ | $y^2 = \alpha x$, $\alpha > 0$ | $y^2 = \alpha x$, $\alpha < 0$

Conic sections (Definition 6.4.6)

Geometrically, a conic section is a curve obtained as the intersection of a cone with a plane.



parabola

circle
&
ellipse

hyperbola

Conic sections (Definition 6.4.6)

Algebraically, such a curve can be defined by a quadratic equation

$$ax^2 + bxy + cy^2 + dx + ey = f \qquad (*)$$

where a, b, c, d, e, f are real numbers and a, b, c are not all zero.

We can rewrite the formula as

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = f.$$

or $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} = f$

where $\boldsymbol{x} = (x, y)^{\mathrm{T}}$, $\boldsymbol{b} = (d, e)^{\mathrm{T}}$ and $\boldsymbol{A} = \begin{bmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{bmatrix}$.

Degenerated conic sections (Definition 6.4.6)

A conic section is called degenerated if it is the empty set, a point, a line or a pair of lines.

For example,

(a) $x^2 + y^2 = -1$: The solution set is an empty set.

(b) $x^2 + y^2 = 0$: The solution set is $\{ (0, 0) \}$ which is a point in $\mathbb{R}^2$.

(c) $x^2 - 4y^2 = 0$: The solution set gives us a pair of lines $x + 2y = 0$ and $x - 2y = 0$ in $\mathbb{R}^2$.

(d) $x^2 + 2xy + y^2 = 0$: The solution set gives us the line $x + y = 0$ in $\mathbb{R}^2$.

Conic sections which are not the empty set, a point, a line or a pair of lines are called non-degenerated.

Conic sections (Definition 6.4.6)

In general, a non-degenerated conic section

$$ax^2 + bxy + cy^2 + dx + ey = f \quad \text{or} \quad \boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathsf{T}}\boldsymbol{x} = f$$

represents a circle, an ellipse, a hyperbola or a parabola which is moved out from the standard position by a translation and/or a rotation (see Section 7.3).

To identify the type of the conic section the equation represented, we need to change the coordinates.

First, we can find an orthogonal matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Conic sections (Definition 6.4.6)

Then by defining new variables x', y' such that

$$\boldsymbol{y} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{P}^{-1}\boldsymbol{x}, \quad \text{where} \quad \boldsymbol{y} = (x', y')^{\mathrm{T}},$$

i.e. $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y}$. (See Discussion 6.4.4.)

The equation of the curve can be simplified as

$$\lambda_1 x'^2 + \lambda_2 y'^2 + gx' + hy' = f \quad \text{where} \quad \begin{bmatrix} g & h \end{bmatrix} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{P}.$$

It is known that

(a) if either $\lambda_1 = 0$ or $\lambda_2 = 0$, the curve is a parabola;

(b) if $\lambda_1 = \lambda_2$, the curve is a circle;

(c) if $\lambda_1\lambda_2 > 0$, the curve is an ellipse; and

(d) if $\lambda_1\lambda_2 < 0$, the curve is a hyperbola.

Examples (Example 6.4.7.1)

Consider the quadratic equation

$$x^2 - xy + y^2 - x - y = 1,$$

i.e. $\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1.$

Let $\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$. Then $\boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}$.

(See Example 6.4.5.1.)

Examples (Example 6.4.7.1)

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Using the new variables x', y'

such that $\begin{bmatrix} x' \\ y' \end{bmatrix} = \boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} x \\ y \end{bmatrix}$,

we have

$$x^2 - xy + y^2 = \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{1}{2}x'^2 + \frac{3}{2}y'^2.$$

We can regard (x', y') as the coordinates of the point (x, y) using a new coordinate system

with x'-axis in the direction of $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

and y'-axis in the direction of $\left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

Examples (Example 6.4.7.1)

Substituting x' and y' into the quadratic equation,

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -1 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 1$$

$$\Leftrightarrow \quad \begin{bmatrix} x' & y' \end{bmatrix}\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} -1 & -1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} = 1$$

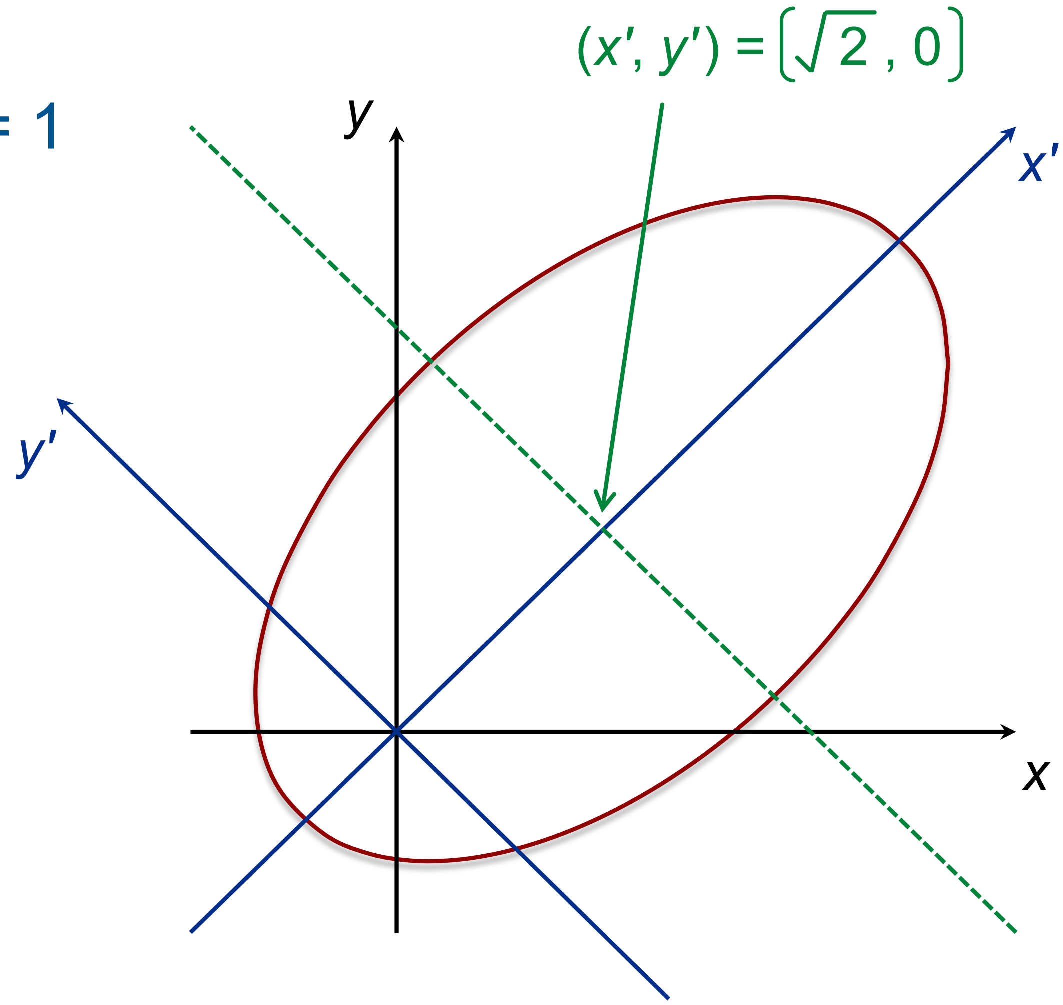
$$\Leftrightarrow \quad \frac{1}{2}x'^2 + \frac{3}{2}y'^2 - \sqrt{2}\,x' = 1$$

$$\Leftrightarrow \quad \frac{1}{2}\left(x' - \sqrt{2}\right)^2 + \frac{3}{2}y'^2 = 2$$

Examples (Example 6.4.7.1)

$$\Leftrightarrow \quad \frac{\left(x' - \sqrt{2}\right)^2}{4} + \frac{y'^2}{4/3} = 1$$

which resembles the standard form of an ellipse.



Examples (Example 6.4.7.2)

Consider the quadratic equation

$$2x^2 + 24xy + 9y^2 + 20x - 6y = 5,$$

i.e. $\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 12 \\ 12 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 20 & -6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5.$

We obtain (by Algorithm 6.3.5) an orthogonal matrix

$$\boldsymbol{P} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \text{ such that } \boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} 2 & 12 \\ 12 & 9 \end{bmatrix} \boldsymbol{P} = \begin{bmatrix} 18 & 0 \\ 0 & -7 \end{bmatrix}.$$

Examples (Example 6.4.7.2)

$$\boldsymbol{P} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}$$

Using the new variables x', y'

such that $\begin{bmatrix} x' \\ y' \end{bmatrix} = \boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} x \\ y \end{bmatrix}$,

we have

$$2x^2 + 24xy + 9y^2 = \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} 18 & 0 \\ 0 & -7 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = 18x'^2 - 7y'^2.$$

We can regard (x', y') as the coordinates of the point (x, y) using a new coordinate system

with x'-axis in the direction of $\left(\frac{3}{5}, \frac{4}{5}\right)$

and y'-axis in the direction of $\left(-\frac{4}{5}, \frac{3}{5}\right)$.

Examples (Example 6.4.7.2)

Substituting x' and y' into the quadratic equation,

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 2 & 12 \\ 12 & 9 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 20 & -6 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 5.$$

$$\Leftrightarrow \quad \begin{bmatrix} x' & y' \end{bmatrix}\begin{bmatrix} 18 & 0 \\ 0 & -7 \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} 20 & -6 \end{bmatrix}\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} = 5$$

$$\Leftrightarrow \quad 18x'^2 - 7y'^2 + \frac{36}{5}x' - \frac{98}{5}y' = 5$$

$$\Leftrightarrow \quad 18\left(x' + \frac{1}{5}\right)^2 - 7\left(y' + \frac{7}{5}\right)^2 = -8$$

Examples (Example 6.4.7.2)

$$\Leftrightarrow \quad -\frac{\left(x' + \frac{1}{5}\right)^2}{4/9} + \frac{\left(y' + \frac{7}{5}\right)^2}{8/7} = 1$$

which resembles the standard form of an hyperbola.