Math 3108: Linear Algebra

Instructor: Jason Murphy

Department of Mathematics and Statistics
Missouri University of Science and Technology

Contents.

Chapter 1. Linear Equations in Linear Algebra

Our first application of linear algebra is the use of matrices to efficiently solve linear systems of equations.

A linear system of m equations with n unknowns can be represented by a matrix with m rows and $n+1$ columns:

The system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \vdots \qquad\qquad \vdots \quad &\quad\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

corresponds to the matrix

$$[A|b] = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_n \end{array}\right].$$

Similarly, every such matrix corresponds to a linear system.

- A is the $m \times n$ **coefficient matrix**:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

  a_{ij} is the element in row i and column j.

- $[A|b]$ is called the **augmented matrix**.

Example. The $2x2$ system

$$3x + 4y = 5$$
$$6x + 7y = 8$$

corresponds to the augmented matrix

$$\left[\begin{array}{cc|c} 3 & 4 & 5 \\ 6 & 7 & 8 \end{array}\right]$$

To solve linear systems, we manipulate and combine the individual equations (in such a way that the **solution set** of the system is preserved) until we arrive at a simple enough form that we can determine the solution set.

Example. Let us solve

$$\begin{aligned} 3x + 4y &= 5 \\ 6x + 7y &= 8. \end{aligned}$$

Multiply the first equation by -2 and add it to the second:

$$\begin{aligned} 3x + 4y &= 5 \\ 0x - y &= -2. \end{aligned}$$

Multiply the second equation by 4 and add it to the first:

$$\begin{aligned} 3x + 0y &= -3 \\ 0x - y &= -2. \end{aligned}$$

Multiply the first equation by $\frac{1}{3}$ and the second by -1:

$$\begin{aligned} x + 0y &= -1 \\ 0x + y &= 2. \end{aligned}$$

Example. (continued) We have transformed the linear system

$$\begin{array}{rcl} 3x + 4y & = & 5 \\ 6x + 7y & = & 8 \end{array} \quad \text{into} \quad \begin{array}{rcr} x + 0y & = & -1 \\ 0x + y & = & 2 \end{array}$$

in such a way that the solution set is preserved.

The second system clearly has solution set $\{(-1, 2)\}$.

Remark. For linear systems, the solution set S satisfies one of the following:

- S contains a single point (**consistent** system)
- S contains infinitely many points (**consistent** system),
- S is empty (**inconsistent** system).

The manipulations used to solve the linear system above correspond to **elementary row operations** on the augmented matrix for the system.

Elementary row operations.

- Replacement: replace a row by the sum of itself and a multiple of another row.
- Interchange: interchange two rows.
- Scaling: multiply all entries in a row by a nonzero constant.

Row operations **do not change the solution set** for the associated linear system.

Example. (revisited)

$$\left[\begin{array}{cc|c} 3 & 4 & 5 \\ 6 & 7 & 8 \end{array}\right] \xrightarrow{R_2 \mapsto -2R_1 + R_2} \left[\begin{array}{cc|c} 3 & 4 & 5 \\ 0 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{R_1 \mapsto 4R_2 + R_1} \left[\begin{array}{cc|c} 3 & 0 & -3 \\ 0 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{R_1 \mapsto \frac{1}{3} R_1} \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{R_2 \mapsto -R_2} \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 2 \end{array}\right].$$

(i) it is simple to determine the solution set for the last matrix
(ii) row operations preserve the solution set.

It is always possible to apply a series of row reductions to put an augmented matrix into **echelon form** or **reduced echelon form**, from which it is simple to discern the solution set.

Echelon form:

- Nonzero rows are above any row of zeros.
- The **leading entry** (first nonzero element) of each row is in a column to the right of the leading entry of the row above it.
- All entries in a column below a leading entry are zeros.

Reduced echelon form: (two additional conditions)

- The leading entry of each nonzero row equals 1.
- Each leading 1 is the only nonzero entry in its column.

Examples.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \quad \text{not in echelon from}$$

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \quad \text{echelon form, not reduced}$$

$$\begin{bmatrix} 1 & 0 & 2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \quad \text{reduced echelon form}$$

Remark. Every matrix can be put into reduced echelon form in a unique manner.

Definition.

A **pivot position** in a matrix is a location that corresponds to a leading 1 in its reduced echelon form.

A **pivot column** is a column that contains a pivot position.

Remark. Pivot positions lie in columns corresponding to dependent variables for the associated systems.

Row Reduction Algorithm.

1. Begin with the leftmost column; if necessary, interchange rows to put a nonzero entry in the first row.
2. Use row replacement to create zeros below the pivot.
3. Repeat steps 1. and 2. with the sub-matrix obtained by removing the first column and first row. Repeat the process until there are no more nonzero rows.

This puts the matrix into echelon form.

4. Beginning with the rightmost pivot, create zeros above each pivot. Rescale each pivot to 1. Work upward and to the left.

This puts the matrix into reduced echelon form.

Example.

$$
\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}
$$

$$
\xrightarrow{R_2 \mapsto -R_1 + R_2} \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ \mathbf{0} & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}
$$

$$
\xrightarrow{R_3 \mapsto -\frac{3}{2} R_2 + R_3} \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & \mathbf{0} & 0 & 0 & 1 & 4 \end{bmatrix}
$$

The matrix is now in echelon form.

Example. (continued)

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \to \begin{bmatrix} 3 & -9 & 12 & -9 & \mathbf{0} & -9 \\ 0 & 2 & -4 & 4 & \mathbf{0} & -14 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

$$\to \begin{bmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & \mathbf{1} & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

$$\to \begin{bmatrix} 3 & \mathbf{0} & -6 & 9 & 0 & -72 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

$$\to \begin{bmatrix} \mathbf{1} & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}.$$

The matrix is now in reduced echelon form.

Solving systems.

- Find the augmented matrix $[A|b]$ for the given linear system.
- Put the augmented matrix into reduced echelon form $[A'|b']$
- Find solutions to the system associated to $[A'|b']$. Express dependent variables in terms of **free variables** if necessary.

Example 1. The system

$$
\begin{aligned}
2x - 4y + 4z &= 6 \\
x - 2y + 2z &= 3 \\
x - y + 0z &= 2
\end{aligned}
$$

$$
\to \left[\begin{array}{ccc|c} 2 & -4 & 4 & 6 \\ 1 & -2 & 2 & 3 \\ 1 & -1 & 0 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -2 & 1 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right]
$$

$$
\to \quad \begin{aligned} x - 2z &= 1 \\ y - 2z &= -1. \end{aligned}
$$

The solution set is

$$
S = \{(1 + 2z, -1 + 2z, z) : z \in \mathbb{R}\}.
$$

Example 2. The system

$$\begin{aligned} 2x - 4y + 4z &= 6 \\ x - 2y + 2z &= 4 \\ x - y + 0z &= 2 \end{aligned}$$

$$\to \left[\begin{array}{ccc|c} 2 & -4 & 4 & 6 \\ 1 & -2 & 2 & 4 \\ 1 & -1 & 0 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -2 & 1 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

$$\to \quad \begin{aligned} x - 2z &= 1 \\ y - 2z &= -1, \\ 0 &= 1. \end{aligned}$$

- The solution set is empty—the system is inconsistent.
- This is always the case when a pivot position lies in the last column.

Row equivalent matrices.

Two matrices are **row equivalent** if they are connected by a sequence of elementary row operations.

Two matrices are row equivalent if and only if they have the same reduced echelon form.

We write $A \sim B$ to denote that A and B are row equivalent.

Chapter 1. Linear Equations in Linear Algebra

1.3 Vector Equations
1.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$.

A matrix with one column or one row is called a **vector**, for example

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}.$$

By using vector arithmetic, for example

$$\alpha \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \beta \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} \alpha + 4\beta \\ 2\alpha + 5\beta \\ 3\alpha + 6\beta \end{bmatrix},$$

we can write linear systems as vector equations.

The linear system

$$\begin{aligned} x + 2y + 3z &= 4 \\ 5x + 6y + 7z &= 8 \\ 9x + 10y + 11z &= 12 \end{aligned}$$

is equivalent to the vector equation

$$x \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix} + y \begin{bmatrix} 2 \\ 6 \\ 10 \end{bmatrix} + z \begin{bmatrix} 3 \\ 7 \\ 11 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 12 \end{bmatrix},$$

in that they have the same solution sets, namely,

$$S = \{(-2 + z, 3 - 2z, z) : z \in \mathbb{R}\}.$$

Geometric interpretation.

The solution set S may be interpreted in different ways:

- S consists of the points of intersection of the three planes

$$\begin{aligned} x + 2y + 3z &= 4 \\ 5x + 6y + 7z &= 8 \\ 9x + 10y + 11z &= 12. \end{aligned}$$

- S consists of the coefficients of the linear combinations of the vectors

$$\begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 6 \\ 10 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 3 \\ 7 \\ 11 \end{bmatrix}$$

  that yield the vector

$$\begin{bmatrix} 4 \\ 8 \\ 12 \end{bmatrix}.$$

Linear combinations and the span.

The set of linear combinations of the vectors $\mathbf{v_1}, \dots \mathbf{v_n}$ is called the **span** of these vectors:

$$\mathrm{span}\{\mathbf{v_1}, \dots, \mathbf{v_n}\} = \{\alpha_1 \mathbf{v_1} + \cdots + \alpha_n \mathbf{v_n} : \alpha_1, \dots, \alpha_n \in \mathbb{R}\}.$$

A vector equation

$$x\,\mathbf{v_1} + y\mathbf{v_2} = \mathbf{v_3}$$

is consistent (that is, has solutions) if and only if

$$\mathbf{v_3} \in \mathrm{span}\{\mathbf{v_1}, \mathbf{v_2}\}.$$

Example. Determine whether or not

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \text{span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} \right\}.$$

This is equivalent to the existence of a solution to:

$$x \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + y \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} + z \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The associated system is

$$\begin{aligned} x + y + z &= 1 \\ 2x + 3y + 4z &= 1 \\ 3x + 4y + 5z &= 1. \end{aligned}$$

The augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 5 & 1 \end{array}\right].$$

The reduced echelon form is

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

The system is inconsistent. Thus

$$\left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right] \notin \operatorname{span}\left\{\left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\right], \left[\begin{array}{c} 1 \\ 3 \\ 4 \end{array}\right], \left[\begin{array}{c} 1 \\ 4 \\ 5 \end{array}\right]\right\}.$$

Geometric description of span.

Let

$$S = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}, \quad T = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Then

- S is the line through the points $(0,0,0)$ and $(0,1,1)$.
- T is the plane through the points $(0,0,0)$, $(0,1,1)$, and $(1,0,1)$.

Geometric description of span. (continued)

Write

$$\mathbf{v_1} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v_2} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

The following are equivalent:

- Is $\mathbf{v_3}$ spanned by $\mathbf{v_1}$ and $\mathbf{v_2}$?
- Can $\mathbf{v_3}$ be written as a linear combination of $\mathbf{v_1}$ and $\mathbf{v_2}$?
- Is $\mathbf{v_3}$ in the plane containing the vectors $\mathbf{v_1}$ and $\mathbf{v_2}$?

Cartesian equation for span.

Recall the definition of the plane T above.

A point (x, y, z) belongs to T when the following vector equation is consistent:

$$\alpha \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

The augmented matrix and its reduced echelon form are as follows:

$$\left[\begin{array}{cc|c} 0 & 1 & x \\ 1 & 0 & y \\ 1 & 1 & z \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & y \\ 0 & 1 & x \\ 0 & 0 & z - x - y \end{array}\right].$$

Thus the Cartesian equation for the plane is

$$0 = z - x - y.$$

Matrix equations. Consider a matrix of the form

$$A = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} \end{bmatrix},$$

where the $\mathbf{a_j}$ are column vectors. The product of A with a column vector is defined by

$$A \begin{bmatrix} x \\ y \\ z \end{bmatrix} := x\mathbf{a_1} + y\mathbf{a_2} + z\mathbf{a_3}.$$

Thus all linear systems can be represented by matrix equations of the form $A\mathbf{X} = \mathbf{b}$.

Example. (Revisited) The system

$$\begin{aligned} x + y + z &= 1 \\ 2x + 3y + 4z &= 1 \\ 3x + 4y + 5z &= 1 \end{aligned}$$

is equivalent to the matrix equation $A\mathbf{X} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Remark. $A\mathbf{X} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Question. When does the vector equation $A\mathbf{X} = \mathbf{b}$ have a solution for every $\mathbf{b} \in \mathbb{R}^m$?

Answer. When the columns of A span $\mathbb{R}^m$.

An equivalent condition is the following: the reduced echelon form of A has a pivot position in every row.

To illustrate this, we study a non-example:

Non-example. Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} \xrightarrow{\text{reduced echelon form}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

This means that for any $b_1, b_2, b_3 \in \mathbb{R}$, we will have

$$A = \left[\begin{array}{ccc|c} 1 & 1 & 1 & b_1 \\ 2 & 3 & 4 & b_2 \\ 3 & 4 & 5 & b_3 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & f_1(b_1, b_2, b_3) \\ 0 & 1 & 2 & f_2(b_1, b_2, b_3) \\ 0 & 0 & 0 & f_3(b_1, b_2, b_3) \end{array}\right]$$

for some linear functions f_1, f_2, f_3. However, the formula

$$f_3(b_1, b_2, b_3) = 0$$

imposes a constraint on the choices of b_1, b_2, b_3.

That is, we cannot solve $A\mathbf{X} = \mathbf{b}$ for arbitrary choices of $\mathbf{b}$.

If instead the reduced echelon form of A had a pivot in any row, then we could use the reduced echelon form for the augmented system to find a solution to $A\mathbf{X} = \mathbf{b}$.

Chapter 1. Linear Equations in Linear Algebra

1.5 Solution Sets of Linear Systems

The system of equations $A\mathbf{X} = \mathbf{b}$ is

- **homogeneous** if $\mathbf{b} = 0$,
- **inhomogeneous** if $\mathbf{b} \neq 0$.

For homogeneous systems:

- The augmented matrix for a homogeneous system has a column of zeros.
- Elementary row operations will not change this column.

Thus, for homogeneous systems it is sufficient to work with the coefficient matrix alone.

Example.

$$\begin{aligned} 2x - 4y + 4z &= 6 \\ x - 2y + 2z &= 3 \\ x - y &= 2 \end{aligned} \;\rightarrow \left[\begin{array}{ccc|c} 2 & -4 & 4 & 6 \\ 1 & -2 & 2 & 3 \\ 1 & -1 & 0 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -2 & 1 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The solution set is

$$x = 1 + 2z, \quad y = -1 + 2z, \quad z \in \mathbb{R}.$$

On the other hand,

$$\begin{aligned} 2x - 4y + 4z &= 0 \\ x - 2y + 2z &= 0 \\ x - y &= 0 \end{aligned} \;\rightarrow \quad x = 2z, \quad y = 2z, \quad z \in \mathbb{R}.$$

The solution set for the previous inhomogeneous system $A\mathbf{X} = \mathbf{b}$ can be represented in **parametric vector form:**

$$\begin{aligned} 2x - 4y + 4z &= 6 \\ x - 2y + 2z &= 3 \\ x - y &= 2 \end{aligned} \;\rightarrow\; \begin{aligned} x &= 1 + 2z \\ y &= -1 + 2z \end{aligned}$$

$$\rightarrow \mathbf{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \;\; z \in \mathbb{R}.$$

The parametric form for the homogeneous system $A\mathbf{X} = 0$ is given by

$$\begin{aligned} x &= 2z \\ y &= 2z \end{aligned} \;\rightarrow \mathbf{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = z \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad z \in \mathbb{R}.$$

Both solution sets are parametrized by the free variable $z \in \mathbb{R}$.

Example. Express the solution set for $A\mathbf{X} = \mathbf{b}$ in parametric vector form, where

$$[A|b] = \left[\begin{array}{rrrrr|r} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 0 & 2 & 0 & 2 \\ 0 & 0 & 2 & -2 & -2 & 2 \end{array}\right]$$

Row reduction leads to

$$\left[\begin{array}{rrrrr|r} 1 & 1 & 0 & -2 & 0 & -2 \\ 0 & 0 & 1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

which means the solution set is

$$x_1 = -2 - x_2 + 2x_4, \quad x_3 = 1 - x_4 + x_5, \quad x_2, x_4, x_5 \in \mathbb{R}.$$

Example. (continued) In parametric form, the solution set is given by

$$\mathbf{X} = \begin{bmatrix} -2 - x_2 + 2x_4 \\ x_2 \\ 1 - x_4 + x_5 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix},$$

where $x_2, x_4, x_5 \in \mathbb{R}$.

To solve the corresponding homogeneous system, simply erase the first vector.

General form of solution sets. If $\mathbf{X}_p$ is any particular solution to $A\mathbf{X} = \mathbf{b}$, then any other solutions to $A\mathbf{X} = \mathbf{b}$ may be written in the form

$$\mathbf{X} = \mathbf{X}_p + \mathbf{X}_h,$$

where $\mathbf{X}_h$ is some solution to $A\mathbf{X} = 0$.

Indeed, given any solution $\mathbf{X}$,

$$A(\mathbf{X} - \mathbf{X}_p) = A\mathbf{X} - A\mathbf{X}_p = \mathbf{b} - \mathbf{b} = 0,$$

which means that $\mathbf{X} - \mathbf{X}_p$ solves the homogeneous system.

Thus, to find the general solution to the inhomogeneous problem, it suffices to

1. Find the general solution to the homogeneous problem,
2. Find any particular solution to the inhomogeneous problem.

Remark. Something similar happens in linear ODE.

Example. (Line example) Suppose the solution set of $A\mathbf{X} = \mathbf{b}$ is a line passing through the points

$$p = (1, -1, 2), \quad q = (0, 3, 1).$$

Find the parametric form of the solution set.

First note that $\mathbf{v} = \mathbf{p} - \mathbf{q}$ is parallel to this line.

As $\mathbf{q}$ belongs to the solution set, the solution set is therefore

$$\mathbf{X} = \mathbf{q} + t\mathbf{v}, \quad t \in \mathbb{R}.$$

Note that we may also write this as

$$\mathbf{X} = (1 - t)\mathbf{q} + t\mathbf{p}, \quad t \in \mathbb{R}.$$

Note also that the solution set to $A\mathbf{X} = 0$ is simply $t\mathbf{v}$, $t \in \mathbb{R}$.

Example. (Plane example) Suppose the solution set of $A\mathbf{X} = \mathbf{b}$ is a plane passing through

$$p = (1, -1, 2), \quad q = (0, 3, 1), \quad r = (2, 1, 0).$$

This time we form the vectors

$$\mathbf{v}_1 = \mathbf{p} - \mathbf{q}, \quad \mathbf{v}_2 = \mathbf{p} - \mathbf{r}.$$

(Note that $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, i.e. one is not a multiple of the other.)

Then the plane is given by

$$\mathbf{X} = \mathbf{p} + t_1\mathbf{v_1} + t_2\mathbf{v_2}, \quad t_1, t_2 \in \mathbb{R}.$$

(The solution set to $A\mathbf{X} = 0$ is then the span of $\mathbf{v}_1$ and $\mathbf{v}_2$.)

Chapter 1. Linear Equations in Linear Algebra

1.7 Linear Independence

Definition. A set of vectors

$$S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$$

is (linearly) **independent** if

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = 0 \implies x_1 = \cdots = x_n = 0$$

for any $x_1, \ldots, x_n \in \mathbb{R}$.

Equivalently, S is independent if the only solution to $A\mathbf{X} = 0$ is $\mathbf{X} = 0$, where $A = [\mathbf{v}_1 \cdots \mathbf{v}_n]$.

Otherwise, we call S (linearly) **dependent**.

Example. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}, \quad A := \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}.$$

Then

$$A \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

In particular, the equation $A\mathbf{X} = 0$ has a nontrivial solution set, namely

$$\mathbf{X} = z \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad z \in \mathbb{R}.$$

Thus the vectors are dependent.

Dependence has another useful characterization:

The vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are dependent if and only if (at least) one of the vectors can be written as a linear combination of the others.

Continuing from the previous example, we found that

$$A\mathbf{X} = 0, \quad \text{where} \quad \mathbf{X} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

(for example). This means

$$\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v_3} = 0, \quad \text{i.e.} \quad \mathbf{v_1} = 2\mathbf{v}_2 - \mathbf{v_3}.$$

Some special cases.

- If $S = \{\mathbf{v}_1, \mathbf{v}_2\}$, then S is dependent if and only if $\mathbf{v}_1$ is a scalar multiple of $\mathbf{v}_2$ (if and only if $\mathbf{v}_1$ and $\mathbf{v}_2$ are co-linear).
- If $0 \in S$, then S is always dependent. Indeed, if

$$S = \{0, \mathbf{v}_1, \cdots, \mathbf{v}_n\},$$

then a nontrivial solution to $A\mathbf{X} = 0$ is

$$0 = 1 \cdot 0 + 0\mathbf{v_1} + \cdots + 0\mathbf{v}_n.$$

Pivot columns. Consider

$$A = [\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3\mathbf{v}_4] = \begin{bmatrix} 1 & 2 & 3 & 4 \\ -2 & -4 & -5 & -6 \\ 3 & 6 & 7 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

In particular, the vector equation $A\mathbf{X} = 0$ has solution set

$$\mathbf{X} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \quad x_2, x_4 \in \mathbb{R}.$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ are dependent.

The pivot columns of A are relevant: By considering

$$(x_2, x_4) \in \{(1,0), (0,1)\},$$

we find that $\mathbf{v}_1$ and $\mathbf{v}_3$ can be combined to produce $\mathbf{v}_2$ or $\mathbf{v}_4$:

- $v_2 = 2v_1$
- $v_4 = -2v_1 + 2v_3$.

Let A be an $m \times n$ matrix. We write $A \in \mathbb{R}^{m\times n}$.

- The number of pivots is bounded above by $\min\{m, n\}$.
- If $m < n$ ('short' matrix), the columns of A are necessarily dependent.
- If $m > n$ ('tall' matrix), the rows of A are necessarily dependent.

Example.

$$A = [\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3\mathbf{v}_4] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 5 \\ 2 & 3 & 4 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The columns of A are necessarily dependent; indeed, setting the free variable $x_3 = 1$ yields the nontrivial combination

$$\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = 0.$$

Example. (continued)

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 5 \\ 2 & 3 & 4 & 5 \end{bmatrix}$$

Are the rows of A dependent or independent?

The rows of A are the columns of the **transpose** of A, denoted

$$A' = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 5 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Now note that:

- Each column of A' is a pivot column.
- $\implies$ the solution set of $A'\mathbf{X} = 0$ is $\mathbf{X} = 0$.
- $\implies$ the columns of A' are independent.
- $\implies$ the rows of A are independent.

Chapter 1. Linear Equations in Linear Algebra

1.8 Introduction to Linear Transformations
1.9 The Matrix of a Linear Transformation

Definition. A **linear transformation** from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,
- $T(\alpha \mathbf{v}) = \alpha T(\mathbf{v})$ for all $v \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

Note that for any linear transformation, we necessarily have

$$T(0) = T(0+0) = T(0) + T(0) \implies T(0) = 0.$$

Example. Let $A \in \mathbb{R}^{m \times n}$. Define $T(\mathbf{X}) = A\mathbf{X}$ for $\mathbf{X} \in \mathbb{R}^n$.

- $T : \mathbb{R}^n \to \mathbb{R}^m$
- $T(\mathbf{X} + \mathbf{Y}) = A(\mathbf{X} + \mathbf{Y}) = A\mathbf{X} + A\mathbf{Y} = T(\mathbf{X}) + T(\mathbf{Y})$
- $T(\alpha \mathbf{X}) = A(\alpha \mathbf{X}) = \alpha A\mathbf{X} = \alpha T(\mathbf{X})$

We call T a **matrix transformation**.

Definition. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **range** of T is the set

$$R(T) := \{T(\mathbf{X}) : \mathbf{X} \in \mathbb{R}^n\}.$$

Note that $R(T) \subset \mathbb{R}^m$ and $0 \in R(T)$.

We call T **onto** (or **surjective**) if $R(T) = \mathbb{R}^m$.

Example. Determine if $\mathbf{b}$ is in the range of $T(\mathbf{X}) = A\mathbf{X}$, where

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 4 \\ 5 & 6 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 7 \\ 11 \end{bmatrix}.$$

This is equivalent to asking if $A\mathbf{X} = \mathbf{b}$ is consistent. By row reduction:

$$[A|b] = \left[\begin{array}{ccc|c} 0 & 1 & 2 & 3 \\ 3 & 0 & 4 & 7 \\ 5 & 6 & 0 & 11 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right].$$

Thus $\mathbf{b} \in R(T)$, indeed

$$T(\mathbf{X}) = \mathbf{b}, \quad \text{where} \quad \mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Example. Determine if $T(\mathbf{X}) = A\mathbf{X}$ is onto, where

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

Equivalently, determine if $A\mathbf{X} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^m$.

Equivalently, determine if the reduced form of A has a pivot in every row:

$$A \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus T is **not** onto.

Example. (Continued) In fact, by performing row reduction on $[A|\mathbf{b}]$ we can describe $R(T)$ explicitly:

$$\left[\begin{array}{ccc|c} 0 & 1 & 2 & b_1 \\ 2 & 3 & 4 & b_2 \\ 3 & 2 & 1 & b_3 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & -\frac{3}{2}b_1 + \frac{1}{2}b_2 \\ 0 & 1 & 2 & b_1 \\ 0 & 0 & 0 & \frac{5}{2}b_1 - \frac{3}{2}b_2 + b_3 \end{array}\right]$$

Thus

$$R(T) = \{\mathbf{b} \in \mathbb{R}^3 : \tfrac{5}{2}b_1 - \tfrac{3}{2}b_2 + b_3 = 0\}.$$

Definition. A linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is **one-to-one** (or **injective**) if

$$T(\mathbf{X}) = 0 \implies \mathbf{X} = 0.$$

More generally, a function f is one-to-one if

$$f(x) = f(y) \implies x = y.$$

For linear transformations, the two definitions are equivalent. In particular, T is one-to-one if:

for each $\mathbf{b}$, the solution set for $T(\mathbf{X}) = \mathbf{b}$ has at most one element.

For matrix transformations $T(\mathbf{X}) = A\mathbf{X}$, injectivity is equivalent to:

- the columns of A are independent
- the reduced form of A has a pivot in every column

Example. Let $T(\mathbf{X}) = A\mathbf{X}$, where

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 1 & 3 & 2 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

- T is **not** one-to-one, as every column does not have a pivot.
- T **is** onto, as every row has a pivot.

Summary.

For a matrix transformation $T(\mathbf{X}) = A\mathbf{X}$.

- Let B denote the reduced echelon form of A.
- T is onto if and only if B has a pivot in every row.
- T is one-to-one if and only if B has a pivot in every column.

Matrix representations.

- Not all linear transformations are matrix transformations.
- However, each linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ has a **matrix representation.**

Let $T : \mathbb{R}^n \to \mathbb{R}^m$. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ denote the standard basis vectors in $\mathbb{R}^n$, e.g.

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n.$$

Define the matrix $[T] \in \mathbb{R}^{m \times n}$ by

$$[T] = [T(\mathbf{e}_1) \cdots T(\mathbf{e}_n)].$$

- We call $[T]$ the **matrix representation** of T.
- Knowing $[T]$ is equivalent to knowing T (see below).

Matrix representations. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear,

$$[T] = [T(\mathbf{e}_1) \cdots T(\mathbf{e}_n)], \quad \mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

By linearity,

$$T(\mathbf{X}) = x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n) = [T]\mathbf{X}.$$

Example. Suppose $T : \mathbb{R}^3 \to \mathbb{R}^4$, with

$$T(\mathbf{e}_1) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad T(\mathbf{e}_2) = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad T(\mathbf{e}_3) = \begin{bmatrix} 3 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

Then

$$[T] = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \\ 4 & 4 & 4 \end{bmatrix}.$$

Matrix representations.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then $[T] \in \mathbb{R}^{m\times n}$, and so:

- $[T] \in \mathbb{R}^{m\times n}$ has m rows and n columns.
- T **onto** $\iff$ $[T]$ has pivot in every row.
- T **one-to-one** $\iff$ $[T]$ has pivot in every column.
- If $m > n$, then T cannot be onto.
- If $m < n$, then T cannot be one-to-one.

Linear transformations of the plane $\mathbb{R}^2$.

Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is linear. Then

$$T(\mathbf{X}) = [T]\mathbf{X} = [T(\mathbf{e}_1)\ T(\mathbf{e}_2)]\mathbf{X} = xT(\mathbf{e}_1) + yT(\mathbf{e}_2).$$

We consider several types of linear transformations with clear geometric meanings, including:

- shears,
- reflections,
- rotations,
- compositions of the above.

Example. (Shear) Let $\lambda \in \mathbb{R}$ and consider

$$[T] = \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix}, \quad T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + \lambda y \\ y \end{bmatrix}.$$

Then

$$\begin{aligned} [T]\begin{bmatrix} 1 \\ 0 \end{bmatrix} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \\ [T]\begin{bmatrix} 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} \lambda \\ 1 \end{bmatrix}, \\ [T]\begin{bmatrix} 0 \\ -1 \end{bmatrix} &= \begin{bmatrix} -\lambda \\ -1 \end{bmatrix}. \end{aligned}$$

Example. (Reflection across the line $y = x$)

Let
$$T(\mathbf{X}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

Note
$$[T]\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \;\; [T]\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$
$$[T]\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \;\; [T]\begin{bmatrix} -2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}.$$

Example. (Rotation by angle θ) Let

$$T(\mathbf{X}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Then

$$[T]\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix}\right],$$
$$[T]\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} -\sin\theta \\ \cos\theta \end{smallmatrix}\right].$$

Example. (Composition) Let us now construct T that
(i) reflects about the y-axis ($x = 0$) and then
(ii) reflects about $y = x$.

$$\text{(i)} \quad [T_1] = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{(ii)} \quad [T_2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We should then take

$$T(\mathbf{X}) = T_2 \circ T_1(\mathbf{X}) = T_2(T_1(\mathbf{X})),$$

that is,

$$T(\mathbf{X}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \left(\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix}$$

Note that $T = T_2 \circ T_1$ is a linear transformation, with

$$[T] = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Chapter 2. Matrix Algebra

2.1 Matrix Operations

Addition and scalar multiplication of matrices.

Let $A, B \in \mathbb{R}^{m\times n}$ with entries A_{ij}, B_{ij} and let $\alpha \in \mathbb{R}$.

We define $A \pm B$ and αA by specifying the ij^{th} entry:

$$(A \pm B)_{ij} := A_{ij} \pm B_{ij}, \quad (\alpha A)_{ij} = \alpha A_{ij}.$$

Example.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + 5 \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 31 & 37 \\ 43 & 49 \end{bmatrix}$$

Matrix addition and scalar multiplication obey the usual rules of arithmetic.

Matrix multiplication. Let $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{r \times n}$ have entries a_{ij}, b_{ij}.

The **matrix product** $AB \in \mathbb{R}^{m \times n}$ is defined via its ij^{th} entry:

$$(AB)_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj}.$$

If $\mathbf{a} \in \mathbb{R}^r$ is a (row) vector and $\mathbf{b} \in \mathbb{R}^r$ is a (column) vector, then we write

$$\mathbf{ab} = \mathbf{a} \cdot \mathbf{b} = \sum_{k=1}^{r} a_k b_k \quad (\textbf{dot product}).$$

Matrix multiplication. (Continued) If we view

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \in \mathbb{R}^{m \times r}, \quad B = [\mathbf{b}_1 \cdots \mathbf{b}_n] \in \mathbb{R}^{r \times n},$$

then

$$(AB)_{ij} = \mathbf{a}_i \cdot \mathbf{b}_j.$$

We may also write

$$AB = A[\mathbf{b}_1 \cdots \mathbf{b}_n] = [A\mathbf{b}_1 \cdots A\mathbf{b}_n],$$

where the product of a matrix and column vector is as before.

Example. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \in \mathbb{R}^{3\times 2}.$$

Then

$$AB = \begin{bmatrix} 22 & 28 \\ 49 & 64 \end{bmatrix}, \quad BA = \begin{bmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \\ 29 & 40 & 51 \end{bmatrix}.$$

Remark. You should not expect $AB = BA$ in general.

Can you think of any examples for which $AB = BA$ does hold?

Definition. The **identity matrix** $I_n \in \mathbb{R}^{n \times n}$ is given by

$$(I_n)_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Properties of matrix multiplication. Let $A \in \mathbb{R}^{m \times n}$ and $\alpha \in \mathbb{R}$. For B and C of appropriate dimensions:

- $A(BC) = (AB)C$
- $A(B + C) = AB + AC$
- $(A + B)C = AC + BC$
- $\alpha(AB) = (\alpha A)B = A(\alpha B)$
- $I_m A = AI_n = A$.

Definition. If $A \in \mathbb{R}^{m \times n}$ has ij^{th} entry a_{ij}, then the **matrix transpose** (or **transposition**) of A is the matrix $A^T \in \mathbb{R}^{n \times m}$ with ij^{th} entry a_{ji}.

One also writes $A^T = A'$.

Example.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \implies A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Thus the columns and rows are interchanged.

Properties.

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(\alpha A)^T = \alpha A^T$ for $\alpha \in \mathbb{R}$
- $(AB)^T = B^T A^T$

Proof of the last property.

$$(AB)^T_{ij} = (AB)_{ji} = \sum_k a_{jk} b_{ki}.$$

$$\begin{aligned}(B^T A^T)_{ij} &= \sum_k (B^T)_{ik} (A^T)_{kj} \\ &= \sum_k b_{ki} a_{jk}.\end{aligned}$$

Thus $(AB)^T = B^T A^T$. $\square$

Example. The transpose of a row vector is a column vector.

Let

$$\mathbf{a} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}.$$

Then $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{2\times 1}$ (column vectors), $\mathbf{a}^T, \mathbf{b}^T \in \mathbb{R}^{1\times 2}$ (row vectors):

$$\mathbf{a}^T\mathbf{b} = 11, \quad \mathbf{a}\mathbf{b}^T = \begin{bmatrix} 3 & -4 \\ -6 & 8 \end{bmatrix}.$$

Key fact. If $T_1 : \mathbb{R}^m \to \mathbb{R}^n$ and $T_2 : \mathbb{R}^n \to \mathbb{R}^k$ are linear transformations, then the matrix representation of the composition is given by

$$[T_2 \circ T_1] = [T_2][T_1].$$

Remark. The dimensions are correct:

- $T_2 \circ T_1 : \mathbb{R}^m \to \mathbb{R}^k$.
- $[T_2 \circ T_1] \in \mathbb{R}^{k \times m}$
- $[T_1] \in \mathbb{R}^{n \times m}$
- $[T_2] \in \mathbb{R}^{k \times n}$
- $[T_2][T_1] \in \mathbb{R}^{k \times m}$.

For matrix transformations, this is clear: if $T_1(\mathbf{x}) = A\mathbf{x}$ and $T_2(\mathbf{x}) = B\mathbf{x}$, then

$$T_2 \circ T_1(\mathbf{x}) = T_2(T_1(\mathbf{x})) = T_2(A\mathbf{x}) = BA\mathbf{x}.$$

Example. Recall that rotation by θ in $\mathbb{R}^2$ is given by

$$[T_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Thus rotation by 2θ is

$$[T_{2\theta}] = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}.$$

One can check that

$$[T_{2\theta}] = [T_\theta]^2 = [T_\theta][T_\theta].$$

Proof of the key fact. Recall that for $T : \mathbb{R}^n \to \mathbb{R}^m$:

$$[T] = [T(\mathbf{e}_1) \cdots T(\mathbf{e}_n)], \quad T(\mathbf{x}) = [T]\mathbf{x}.$$

So

$$\begin{aligned} [T_2 \circ T_1] &= [T_2(T_1(\mathbf{e}_1)) \cdots T_2(T_1(\mathbf{e}_n))] \\ &= [[T_2]T_1(\mathbf{e}_1) \cdots [T_2]T_1(\mathbf{e}_n)] \\ &= [T_2][T_1(\mathbf{e}_1) \cdots T_1(\mathbf{e}_n)] \quad (*) \\ &= [T_2][T_1]. \end{aligned}$$

In (*), we have used the column-wise definition of matrix multiplication. $\square$

Chapter 2. Matrix Algebra

Definition. Let $A \in \mathbb{R}^{n \times n}$ (square matrix). We call $B \in \mathbb{R}^{n \times n}$ an **inverse** of A if

$$AB = BA = I_n.$$

Remark. If A has an inverse, then it is unique. *Proof.* Suppose

$$AB = BA = I_n \quad \text{and} \quad AC = CA = I_n.$$

Then

$$B = BI_n = BAC = I_nC = C. \quad \square$$

If A has an inverse, then we denote it by A^{-1}. Note $(A^{-1})^{-1} = A$.

Remark. If $A, B \in \mathbb{R}^{n \times n}$ are invertible, then AB is invertible. Indeed,

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Example. If

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \text{then} \quad A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}.$$

Note that to solve $A\mathbf{X} = \mathbf{b}$, we may set $\mathbf{X} = A^{-1}\mathbf{b}$.

For example, the solution to

$$A\mathbf{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{is} \quad \mathbf{X} = A^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ \frac{3}{2} \end{bmatrix}.$$

Questions.

1. When does $A \in \mathbb{R}^{n \times n}$ have an inverse?
2. If A has an inverse, how do we compute it?

Note that if A is invertible (has an inverse), then:

- $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ (namely, $\mathbf{x} = A^{-1}\mathbf{b}$).
  Equivalently, A has a pivot in every row.
- If $A\mathbf{x} = 0$, then $\mathbf{x} = A^{-1}0 = 0$. Thus the columns of A are independent.
  Equivalently, A has a pivot in every column.

Conversely, we will show that if A has a pivot in every column or row, then A is invertible.

Thus all of the above conditions are equivalent.

Goal. If A has a pivot in every column, then A is invertible.

Since A is square, this is equivalent to saying that if the reduced echelon form of A is I_n, then A is invertible.

Key observation. Elementary row operations correspond to multiplication by an invertible matrix. (See below.)

With this observation, our hypothesis means that

$$E_k \cdots E_1 A = I_n$$

for some invertible matrices E_j. Thus

$$A = (E_k \cdots E_1)^{-1} = E_1^{-1} \cdots E_k^{-1}.$$

In particular, A is invertible.

Furthermore, this computes the inverse of A. Indeed,

$$A^{-1} = E_k \cdots E_1.$$

It remains to show that elementary row operations correspond to multiplication by a invertible matrix (known as **elementary matrices**).

In fact, to write down the corresponding elementary matrix, one simply applies the row operation to I_n.

Remark. A does not need to be square; the following works for any $A \in \mathbb{R}^{n \times m}$.

For concreteness, consider the 3x3 case.

- "Multiply row one by non-zero $\alpha \in \mathbb{R}$" corresponds to multiplication by

$$E = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Indeed,

$$\begin{bmatrix} \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} \alpha & 2\alpha & 3\alpha \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Note that E is invertible:

$$E^{-1} = \begin{bmatrix} \frac{1}{\alpha} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

- "Interchange rows one and two" corresponds to multiplication by

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Indeed,

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}.$$

Note that E is invertible. In fact, $E = E^{-1}$.

- "Multiply row three by α and add it to row two" corresponds to multiplication by

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{bmatrix}.$$

Indeed,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4+7\alpha & 5+8\alpha & 6+9\alpha \\ 7 & 8 & 9 \end{bmatrix}$$

Note that E is invertible:

$$E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\alpha \\ 0 & 0 & 1 \end{bmatrix}.$$

Summary. A is invertible if and only if there exist a sequence of elementary matrices E_j so that

$$E_k \cdots E_1 A = I_n.$$

Note that

$$\begin{aligned} E_k \cdots E_1[A|I_n] &= [E_k \cdots E_1 A | E_k \cdots E_1 I_n] \\ &= [I_n | E_k \cdots E_1] \\ &= [I_n | A^{-1}]. \end{aligned}$$

Thus A is invertible if and only if

$$[A|I_n] \sim [I_n|A^{-1}].$$

Example 1.

$$[A|I_2] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{cc|rr} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right] = [I_2|A^{-1}]$$

Thus A is invertible, with A^{-1} as above.

Example 2.

$$[A|I_3] = \left[\begin{array}{rrr|rrr} 3 & 3 & 3 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 3 & 5 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{rrr|rrr} 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 & 2 & -1 \end{array}\right] =: [U|B]$$

Thus A is not invertible. Note $BA = U$.

Some additional properties.

- If A is invertible, then $(A^T)^{-1} = (A^{-1})^T$.

  Indeed,
  $$A^T(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n,$$
  and similarly $(A^{-1})^T A^T = I_n$.

- Suppose AB is invertible. Then

  $$A[B(AB)^{-1}] = (AB)(AB)^{-1} = I_n.$$

  Thus
  $$A[B(AB)^{-1}\mathbf{b}] = \mathbf{b} \quad \text{for any} \quad \mathbf{b} \in \mathbb{R}^n,$$
  so that $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$. Thus A has a pivot in every row, so that A is invertible. Similarly, B is invertible.

  Conclusion. AB is invertible if and only if A, B are invertible.

Some review. Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x}, \mathbf{b} \in \mathbb{R}^m$.

Row pivots. The following are equivalent:

- A has a pivot in every row
- $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^m$
- the columns of A span $\mathbb{R}^m$
- the transformation $T(\mathbf{x}) = A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^m$
- the rows of A are independent (see below)

Column pivots. The following are equivalent:

- A has a pivot in every column
- $A\mathbf{x} = 0 \implies \mathbf{x} = 0$
- the columns of A are independent
- the transformation $T(\mathbf{x}) = A\mathbf{x}$ is one-to-one

Claim. If A has m pivots, then the rows of $A \in \mathbb{R}^{m \times n}$ are independent. (The converse is also true — why?)

Proof. By hypothesis,

$$BA = U, \quad \text{or equivalently} \quad A = B^{-1}U$$

where $B \in \mathbb{R}^{m \times m}$ is a product of elementary matrices and U has a pivot in each row.

Suppose $A^T\mathbf{x} = 0$. Then (since U^T has a pivot in each column),

$$U^T[(B^{-1})^T\mathbf{x}] = 0 \implies (B^{-1})^T\mathbf{x} = 0 \implies \mathbf{x} = 0$$

by invertibility. Thus the columns of A^T (i.e. rows of A) are independent. □

When A is square, all of the above equivalences hold, in addition to the following:

- There exists $C \in \mathbb{R}^{n \times n}$ so that $CA = I_n$.
  (This gives $A\mathbf{x} = 0 \implies \mathbf{x} = 0$.)
- There exists $D \in \mathbb{R}^{n \times n}$ so that $AD = I_n$.
  (This gives $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$.)
- A is invertible.
- A^T is invertible.

Definition. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. We say T is **invertible** if there exists $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T \circ S(\mathbf{x}) = S \circ T(\mathbf{x}) = \mathbf{x} \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^n.$$

If $T(\mathbf{x}) = A\mathbf{x}$, this is equivalent to A being invertible, with $S(\mathbf{x}) = A^{-1}\mathbf{x}$.

If T has an inverse, it is unique and denoted T^{-1}.

The following are equivalent for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$:

- T is invertible
- T is 'left-invertible' (there exists S so that $S \circ T(\mathbf{x}) = \mathbf{x}$)
- T is 'right-invertible' (there exists S so that $T \circ S(\mathbf{x}) = \mathbf{x}$).

Chapter 2. Matrix Algebra

2.5 Matrix Factorization

Definition. A matrix $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ is **lower triangular** if

$$a_{ij} = 0 \quad \text{for all} \quad i < j.$$

We call A **unit lower triangular** if additionally

$$a_{ii} = 1 \quad \text{for all} \quad i = 1, \dots, \min\{m, n\}.$$

Example. The elementary matrix E corresponding to the row replacement

$$R_j \mapsto \alpha R_i + R_j, \quad i < j$$

is unit lower triangular, as is its inverse. E.g. (in 3×3 case):

$$R_2 \mapsto \alpha R_1 + R_2 \implies E = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We can similarly define **upper triangular** and **unit upper triangular** matrices.

Note that the product of (unit) lower triangular matrices is (unit) lower triangular:

$$(AB)_{ij} = \sum_k a_{ik} b_{kj} = \sum_{j \le k \le i} a_{ik} b_{kj} = 0 \quad \text{for} \quad i < j.$$

The same is true for upper triangular matrices.

Definition. We call $P \in \mathbb{R}^{m \times m}$ a **permutation matrix** if it is a product of elementary row-exchange matrices.

LU Factorization. For any $A \in \mathbb{R}^{m \times n}$, there exists a permutation matrix $P \in \mathbb{R}^{m \times m}$ and an upper triangular matrix $U \in \mathbb{R}^{m \times n}$ (in echelon form) such that

$$PA \sim U.$$

Moreover, the elementary matrices used to reduce PA to U may all be taken to be lower triangular and of the type

$$R_j \mapsto \alpha R_i + R_j \quad \text{for some} \quad i < j.$$

Thus

$$E_k \cdots E_1 PA = U$$

for some unit lower triangular (elementary) matrices E_j, and so

$$PA = (E_1^{-1} \cdots E_k^{-1}) U = LU$$

for some unit lower triangular L.

The LU factorization is also used to solve systems of linear equations.

Example. Solve $A\mathbf{x} = \mathbf{b}$, where

$$A = LU = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

1. Solve $L\mathbf{y} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \implies \begin{cases} y_1 = 3, \\ y_1 + y_2 = 3 \end{cases} \implies \begin{cases} y_1 = 3 \\ y_2 = 0 \end{cases}$$

2. Solve $U\mathbf{x} = \mathbf{y}$.

$$\begin{bmatrix} -2 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \implies \begin{cases} -2x_1 + 2x_2 = 3 \\ x_2 = 0 \end{cases}$$

$$\implies \begin{cases} x_1 = -\frac{3}{2} \\ x_2 = 0 \end{cases} .$$

This process is computationally efficient when A is very large and solutions are required for systems in which A stays fixed but $\mathbf{b}$ varies.

See the ‘Numerical notes’ section in the book for more details.

We next compute some examples of LU factorization, beginning with the case that P is the identity matrix.

Example 1. Let

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

We put A into upper triangular form form via row replacements:

$$A \xrightarrow[R_3 \mapsto -\frac{1}{2}R_1 + R_3]{R_2 \mapsto -\frac{1}{2}R_1 + R_2} \begin{bmatrix} 2 & 3 & 4 & 1 \\ 0 & \frac{3}{2} & 0 & \frac{7}{2} \\ 0 & \frac{1}{2} & 1 & \frac{7}{2} \end{bmatrix}$$

$$\xrightarrow{R_3 \mapsto -\frac{1}{3}R_2 + R_3} \begin{bmatrix} 2 & 3 & 4 & 1 \\ 0 & \frac{3}{2} & 0 & \frac{7}{2} \\ 0 & 0 & 1 & \frac{7}{3} \end{bmatrix} = U.$$

We have three unit lower triangular elementary matrices E_1, E_2, E_3 so that

$$E_3 E_2 E_1 A = U, \quad \text{i.e.} \quad A = E_1^{-1} E_2^{-1} E_3^{-1} U.$$

That is, we construct L via row reduction:

$$L = E_1^{-1} E_2^{-1} E_3^{-1}, \quad A = LU.$$

Example 1. (continued) Note that

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (R_2 \mapsto -\tfrac{1}{2}R_1 + R_2)$$

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix}, \quad (R_3 \mapsto -\tfrac{1}{2}R_1 + R_3)$$

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix}, \quad (R_3 \mapsto -\tfrac{1}{3}R_2 + R_3),$$

$$E_3E_2E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{3} & 1 \end{bmatrix}, \;\; L = (E_3E_2E_1)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix},$$

In particular, $E_3E_2E_1L = I_3$.

Example 1. (continued)

Altogether, we have the LU factorization:

$$\begin{bmatrix} 2 & 3 & 4 & 1 \\ 1 & 3 & 2 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 & 1 \\ 0 & \frac{3}{2} & 0 & \frac{7}{2} \\ 0 & 0 & 1 & \frac{7}{3} \end{bmatrix}. \quad \square$$

Let us next consider an example where we will not have P equal to the identity.

Example 2. Let

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

We try to put A into echelon by using the row replacements

$$R_2 \mapsto R_1 + R_2, \quad R_3 \mapsto -2R_1 + R_3.$$

This corresponds to

$$EA := \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} A = \begin{bmatrix} 2 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

However, we now need a row interchange ($R_2 \leftrightarrow R_3$). This corresponds to multiplication by

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$ **Not** lower triangular!

Example 2. (continued)

So far, we have written $PEA = U$ with E unit lower triangular and U in echelon form. Thus (since $P = P^{-1}$),

$$A = E^{-1}PU.$$

However, $E^{-1}P$ is not lower triangular:

$$E^{-1}P = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

But if we multiply by P again, we get the desired factorization:

$$PA = LU, \quad L = PE^{-1}P = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

Chapter 2. Matrix Algebra

2.7 Applications to Computer Graphics

Definition. We call $(\mathbf{y}, h) \in \mathbb{R}^{n+1}$ (with $h \neq 0$) **homogeneous coordinates** for $\mathbf{x} \in \mathbb{R}^n$ if

$$\mathbf{x} = \frac{1}{h}\mathbf{y}, \quad \text{that is,} \quad x_j = \frac{1}{h}y_j \quad \text{for} \quad 1 \leq j \leq n.$$

In particular, $(\mathbf{x}, 1)$ are homogeneous coordinates for $\mathbf{x}$.

Homogeneous coordinates can be used to describe more general transformations of $\mathbb{R}^n$ than merely linear transformations.

Example. Let $\mathbf{x}_0 \in \mathbb{R}^n$ and define

$$T((\mathbf{x}, 1)) = \left[\begin{array}{c|c} I_n & \mathbf{x}_0 \\ \hline \mathbf{0} & 1 \end{array}\right] \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{x} + \mathbf{x}_0 \\ 1 \end{bmatrix}.$$

This is a linear transformation on homogeneous coordinates that corresponds to the translation

$$\mathbf{x} \mapsto \mathbf{x} + \mathbf{x}_0.$$

Note that translation in $\mathbb{R}^n$ is **not** a linear transformation if $\mathbf{x}_0 \neq 0$. (Why not?)

To represent a linear transformation on $\mathbb{R}^n$, say $T(\mathbf{x}) = A\mathbf{x}$, in homogeneous coordinates, we use

$$T((\mathbf{x}, 1)) = \left[\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{0} & 1 \end{array}\right] \left[\begin{array}{c} \mathbf{x} \\ 1 \end{array}\right] = \left[\begin{array}{c} A\mathbf{x} \\ 1 \end{array}\right].$$

We can then compose translations and linear transformations to produce either

$$\mathbf{x} \mapsto A\mathbf{x} + \mathbf{x}_0$$

or

$$\mathbf{x} \mapsto A(\mathbf{x} + \mathbf{x}_0).$$

Graphics in three dimensions. Applying successive linear transformations and translation to the homogeneous coordinates of the points that define an outline of an object in $\mathbb{R}^3$ will produce the homogeneous coordinates of the translated/deformed outline of the object.

See the Practice Problem in the textbook.

This also works in the plane.

Example 1. Find the transformation that translates by $(0, 8)$ in the plane and then reflects across the line $y = -x$.
Solution:

$$\left[\begin{array}{cc|c} 0 & -1 & 0 \\ -1 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 8 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} x \\ y \\ 1 \end{array}\right] = \left[\begin{array}{cc|c} 0 & -1 & -8 \\ -1 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} x \\ y \\ 1 \end{array}\right]$$

Example 2. Find the transformation that rotates points an angle θ about the point $(3, 1)$:
Solution:

$$\left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 1 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{cc|c} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & -1 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} x \\ y \\ 1 \end{array}\right]$$

Note the order of operations in each example.

Example 1. (Continued) What is the effect of the transformation in Example 1 on the following vertices:

$$(0,0), \quad (3,0), \quad (3,4).$$

Solution:

$$\left[\begin{array}{cc|c} 0 & -1 & -8 \\ -1 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{ccc} 0 & 3 & 3 \\ 0 & 0 & 4 \\ \hline 1 & 1 & 1 \end{array}\right] = \left[\begin{array}{ccc} -8 & -8 & -12 \\ 0 & -3 & -3 \\ \hline 1 & 1 & 1 \end{array}\right]$$

Thus

$$(0,0) \mapsto (-8,0), \quad (3,0) \mapsto (-8,-3), \quad (4,5) \mapsto (-12,-3).$$

Perspective projection. Consider a light source at the point $(0, 0, d) \in \mathbb{R}^3$, where $d > 0$.

A ray of light passing through a point $(x, y, z) \in \mathbb{R}^3$ with $0 \leq z < d$ will intersect the xy-plane at a point $(x^*, y^*, 0)$.

Understanding the map $(x, y, z) \mapsto (x^*, y^*)$ allows us to represent 'shadows'. (One could also imagine projection onto other $2d$ surfaces.)

By some basic geometry (similar triangles, for example), one can deduce

$$x^* = \frac{x}{1 - \frac{z}{d}}, \quad y^* = \frac{y}{1 - \frac{z}{d}}.$$

In particular, we find that

$$(x, y, 0, 1 - \tfrac{z}{d}) \quad \text{are homogeneous coordinates for} \quad (x^*, y^*, 0).$$

Perspective projection. (Continued) Note that the mapping

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{d} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \\ 1 - \frac{z}{d} \end{bmatrix}$$

takes homogeneous coordinates of (x, y, z) to the homogeneous coordinates of $(x^*, y^*, 0)$.

Using this, one can understand how the shadows of objects in $\mathbb{R}^3$ would move under translations/deformations.

Chapter 3. Determinants

Definition. The **determinant** of a matrix $A \in \mathbb{R}^{n\times n}$, denoted $\det A$, is defined inductively. Writing $A = (a_{ij})$, we have the following:

- If $n = 1$, then $\det A = a_{11}$.
- If $n \geq 2$, then

$$\det A = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det A_{1j},$$

where A_{ij} is the $(n-1)\times(n-1)$ **submatrix** obtained by removing the i^{th} row and j^{th} column from A.

Example. Consider the 2×2 case:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The 1×1 submatrices A_{11} and A_{12} are given by

$$A_{11} = d, \quad A_{12} = c.$$

Thus

$$\begin{aligned} \det A &= \sum_{j=1}^{2} (-1)^{1+j} a_{1j} \det A_{1j} \\ &= a_{11} \det A_{11} - a_{12} A_{12} \\ &= ad - bc. \end{aligned}$$

Conclusion. $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$

Examples.

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = 1 \cdot 4 - 2 \cdot 3 = -2.$$

$$\det \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} = \lambda\mu$$

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = 1 \cdot 6 - 2 \cdot 3 = 0.$$

Example. Consider

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 3 & 6 \end{bmatrix}$$

Then

$$A_{11} = \begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix},\ A_{12} = \begin{bmatrix} 1 & 4 \\ 1 & 6 \end{bmatrix},\ A_{13} = \begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix},$$

and

$$\begin{aligned} \det A &= \sum_{j=1}^{3} (-1)^{1+j} a_{1j} \det A_{1j} \\ &= 1 \cdot \det A_{11} - 2 \cdot \det A_{12} + 3 \cdot A_{13} \\ &= 6 - 4 + 0 = 2. \end{aligned}$$

Note. Note the alternating ± 1 pattern.

Definition. Given an $n \times n$ matrix A, the terms

$$C_{ij} := (-1)^{i+j} \det A_{ij}$$

are called the **cofactors** of A. Recall A_{ij} is the the $(n-1) \times (n-1)$ matrix obtained by removing the i^{th} row and j^{th} column from A.

Note

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} A_{1j} = \sum_{j=1}^{n} a_{1j} C_{1j}.$$

We call this the **cofactor expansion** of $\det A$ using row 1.

Claim. The determinant can be computed using the cofactor expansion of $\det A$ with any row or any column. That is,

$$\begin{aligned}\det A &= \sum_{j=1}^{n} a_{ij} C_{ij} \quad \text{for any} \quad i \\ &= \sum_{i=1}^{n} a_{ij} C_{ij} \quad \text{for any} \quad j.\end{aligned}$$

The first expression is the cofactor expansion using row i.

The second expression is the cofactor expansion using column j.

Example. Consider again

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 3 & 6 \end{bmatrix}, \quad \det A = 2.$$

Let's compute the determinant using column 1: Using

$$A_{11} = \begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}, \quad A_{31} = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix},$$

$$\begin{aligned} \det A &= \sum_{i=1}^{3} a_{i1} C_{i1} \\ &= a_{11} \det A_{11} - a_{21} \det A_{21} + a_{31} \det A_{31} \\ &= 1 \cdot 6 - 1 \cdot 3 + 1 \cdot (-1) = 2. \end{aligned}$$

Remark. Don't forget the factor $(-1)^{i+j}$ in C_{ij}.

• Use the flexibility afforded by cofactor expansion to simplify your computations: use the row or columns with the most zeros.

Example. Consider

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

Using row 3,

$$\det A = -6 \det A_{32} = -6 \cdot -7 = 42.$$

Properties of the determinant.

- $\det I_n = 1$
- $\det AB = \det A \det B$
- $\det A^T = \det A$
- Note that if $A \in \mathbb{R}^{n\times n}$ is invertible, then

$$1 = \det I_n = \det AA^{-1} = \det A \det(A^{-1}).$$

In particular, $\det A \neq 0$ and $\det(A^{-1}) = [\det A]^{-1}$.

Determinants of elementary matrices.

- If E corresponds to $R_i \mapsto \alpha R_i$ (for $\alpha \neq 0$), then
$$\det E = \alpha.$$
- If E corresponds to a row interchange, then
$$\det E = -1.$$
- If E corresponds to $R_j \mapsto R_i + \alpha R_j$, then
$$\det E = 1.$$

  In particular, in each case $\det E \neq 0$.

  Let's check these in the simple 2×2 case.

Determinants of elementary matrices. Recall that the elementary matrix corresponding to a row operation is obtained by applying this operation to the identity matrix.

- Scaling:

$$E = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix} \implies \det E = \alpha.$$

- Interchange:

$$E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \implies \det E = -1.$$

- Replacement $(R_2 \mapsto \alpha R_1 + R_2)$

$$E = \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \implies \det E = 1.$$

Row reduction and determinants. Suppose $A \sim U$, with

$$U = E_k \cdots E_1 A.$$

Then

$$\det A = \frac{1}{\det E_1 \cdots \det E_k} \det U.$$

Suppose that U is in upper echelon form. Then

$$\det U = u_{11} \cdots u_{nn}.$$

Indeed, this is true for any upper trianglular matrix (use the right cofactor expansion).

Thus, row reduction provides another means of computing determinants!

Example 1.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 10 \\ 3 & 8 & 9 \end{bmatrix} \xrightarrow[R_3 \mapsto -3R_1 + R_3]{R_2 \mapsto -2R_1 + R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 2 & 0 \end{bmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} = U.$$

Then

$$\det A = 1 \cdot (-1) \cdot \det U = -1 \cdot 2 \cdot 4 = -8.$$

Example 2.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 6 & 12 & 18 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus $\det A = 0$.

Invertibility.

We saw above that if $U = (u_{ij})$ is an upper echelon form for A, then

$$\det A = c \cdot u_{11} \cdots u_{nn} \quad \text{for some} \quad c \neq 0.$$

We also saw that if A is invertible, then $\det A \neq 0$. Equivalently,

$$\det A = 0 \implies A \text{ is not invertible.}$$

On the other hand, if A is not invertible then it has fewer than n pivot columns, and hence some $u_{ii} = 0$. Thus

$$A \text{ not invertible} \implies \det A = c u_{11} \cdots u_{nn} = 0.$$

So in fact the two conditions are equivalent.

Invertibility Theorem. The following are equivalent:

- A is invertible.
- The reduced row echelon form of A is I_n.
- A has n pivot columns (and n pivot rows).
- $\det A \neq 0$.

Examples.

- Recall
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 0 & 5 \\ 0 & 6 & 0 \end{bmatrix} \implies \det A = 42.$$
Thus A is invertible and $\det A^{-1} = \frac{1}{42}$.
- If $\det AB = \det A \det B \neq 0$, then A, B, AB are invertible.
- Consider the matrix
$$M(\lambda) = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}, \quad \lambda \in \mathbb{R}.$$
For which λ is $M(\lambda)$ **not** invertible?
Answer: Compute
$$\det M(\lambda) = (2-\lambda)^2 - 1 = (\lambda-1)(\lambda-3) \implies \lambda = 1, 3.$$

Chapter 4. Vector Spaces

4.1 Vector Spaces and Subspaces

Definition. A **vector space** V over a field of **scalars** F is a non-empty set together with two operations, namely **addition** and **scalar multiplication**, which obey the following rules: for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\alpha, \beta \in F$:

- $\mathbf{u} + \mathbf{v} \in V$
- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- there exists $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$
- there exists $-\mathbf{u} \in V$ such that $-\mathbf{u} + \mathbf{u} = \mathbf{0}$

- $\alpha \mathbf{v} \in V$
- $\alpha(\mathbf{u} + \mathbf{v}) = \alpha \mathbf{u} + \alpha \mathbf{v}$
- $(\alpha + \beta)\mathbf{u} = \alpha \mathbf{u} + \beta \mathbf{u}$
- $\alpha(\beta \mathbf{u}) = (\alpha\beta)\mathbf{u}$
- $1\mathbf{u} = \mathbf{u}$

Remark 1. A **field** is another mathematical object with its own long list of defining axioms, but in this class we will always just take $F = \mathbb{R}$ or $F = \mathbb{C}$.

Remark 2. One typically just refers to the vector space V without explicit reference to the underlying field.

Remark 3. The following are consequences of the axioms:

$$0\mathbf{u} = \mathbf{0}, \quad \alpha\mathbf{0} = \mathbf{0}, \quad -\mathbf{u} = (-1)\mathbf{u}.$$

Examples.

- $V = \mathbb{R}^n$ and $F = \mathbb{R}$
- $V = \mathbb{C}^n$ and $F = \mathbb{R}$ or $\mathbb{C}$
- $V = \mathbb{P}_n$ (polynomials of degree n or less), and $F = \mathbb{R}$
- $V = \mathbb{S}$, the set of all doubly-infinite sequences $(\ldots, x_{-2}, x_{-1}, x_0, x_1, \ldots)$ and $F = \mathbb{R}$
- $V = \mathcal{F}(\mathbb{D})$, the set of all functions defined on a domain $\mathbb{D}$ and $F = \mathbb{R}$.

Definition. Let V be a vector space and W a subset of V. If W is also a vector space under vector addition and scalar multipliation, then W is a **subspace** of V. Equivalently, $W \subset V$ is a subspace if

$$\mathbf{u} + \mathbf{v} \in W \quad \text{and} \quad \alpha \mathbf{v} \in W$$

for any $\mathbf{u}, \mathbf{v} \in W$ and any scalar α.

Example 1. If $\mathbf{0} \notin W$, then W is **not** a subspace of V. Thus

$$W = \{\mathbf{x} \in \mathbb{R}^2 : x_1 + x_2 = 2\} \quad \text{is not a subspace of } \mathbb{R}^2.$$

Example 2. The set

$$W = \{\mathbf{x} \in \mathbb{R}^2 : x_1 x_2 \geq 0\} \quad \text{is not a subspace of } \mathbb{R}^2.$$

Indeed, $[1, 0]^T + [0, -1]^T \notin W$.

Further examples and non-examples.

- $W = \mathbb{R}^n$ is a subspace of $V = \mathbb{C}^n$ with $F = \mathbb{R}$
- $W = \mathbb{R}^n$ is **not** a subspace of $V = \mathbb{C}^n$ with $F = \mathbb{C}$
- $W = \mathbb{P}_n$ is a subspace of $V = \mathcal{F}(\mathbb{D})$
- $W = \mathbb{S}_+$, the set of doubly-infinite sequences such that $x_{-k} = 0$ for $k > 0$ is a subspace of $\mathbb{S}$
- $W = \{(x, y) \in \mathbb{R}^2 : x, y \in \mathbb{Z}\}$ is **not** a subspace of $\mathbb{R}^2$

Span as subspace. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be a collection of vectors in $\mathbb{R}^n$. Then

$$W := \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k : c_1, \ldots, c_k \in F\}$$

is a subspace of $\mathbb{R}^n$.

Indeed, if $\mathbf{u}, \mathbf{v} \in W$ and $\alpha \in F$ then $\mathbf{u} + \mathbf{v} \in W$ and $\alpha\mathbf{u} \in W$. (Why?)

Subspaces associated with $A \in \mathbb{R}^{m \times n}$.

- The **column space** of A, denoted $\text{col}(A)$ is the span of the columns of A.
- The **row space** of A, denoted $\text{row}(A)$ is the span of the rows of A.
- The **null space** of A, denoted $\text{nul}(A)$, is

$$\text{nul}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\} \subset \mathbb{R}^n.$$

Note that $\text{nul}(A)$ is a subspace of $\mathbb{R}^n$; however,

$$\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}$$

is **not** a subspace if $\mathbf{b} \neq 0$. (Why not?)

Example. Let

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

The solution set to $A\mathbf{x} = \mathbf{0}$ is written in parametric vector form as

$$x_3 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

That is,

$$\text{nul}(A) = \text{span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Example. Let

$$W = \left\{ \begin{bmatrix} s+3t \\ 8t \\ s-t \end{bmatrix} : s, t \in \mathbb{R} \right\}.$$

This is a subspace of $\mathbb{R}^3$, since

$$\begin{bmatrix} s+3t \\ 8t \\ s-t \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} 3 \\ 8 \\ -1 \end{bmatrix},$$

and hence

$$W = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \\ -1 \end{bmatrix} \right\}.$$

Chapter 4. Vector Spaces

4.2 Null Spaces, Column Spaces, and Linear Transformations

Null space.

Recall that the **null space** of $A \in \mathbb{R}^{m\times n}$ is the solution set to $A\mathbf{x} = \mathbf{0}$, denoted $\mathrm{nul}(A)$.

Note $\mathrm{nul}(A)$ is a subspace of $\mathbb{R}^n$: for $\mathbf{x}, \mathbf{y} \in \mathrm{nul}(A)$ and $\alpha \in \mathbb{R}$,

$$A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0} \quad \text{(closed under addition)}$$
$$A(\alpha\mathbf{x}) = \alpha A\mathbf{x} = \alpha\mathbf{0} = \mathbf{0} \quad \text{(closed under scalar ultiplication).}$$

In fact, by writing the solution set to $A\mathbf{x} = \mathbf{0}$ in parametric vector form, we can identify $\mathrm{nul}(A)$ as the span of a set of vectors.

(We saw such an example last time.)

Column space.

Recall that the **column space** of $A \in \mathbb{R}^{m \times n}$ is the span of the columns of A, denoted $\mathrm{col}(A)$.

Recall that $\mathrm{col}(A)$ is a subspace of $\mathbb{R}^m$.

Note that $\mathbf{b} \in \mathrm{col}(A)$ precisely when $A\mathbf{x} = \mathbf{b}$ is consistent.

Note that $\mathrm{col}(A) = \mathbb{R}^m$ when A has a pivot in every row.

Using row reduction, we can describe $\mathrm{col}(A)$ as the span of a set of vectors.

Example.

$$[A|\mathbf{b}] = \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & b_1 \\ -1 & -1 & 0 & 0 & b_2 \\ 1 & 1 & 3 & 3 & b_3 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & b_1 \\ 0 & 0 & 1 & 1 & b_1 + b_2 \\ 0 & 0 & 0 & 0 & b_3 - 2b_2 - 3b_1 \end{array}\right]$$

Thus $\mathbf{b} \in \mathrm{col}(A)$ if and only if

$$b_3 - 2b_2 - 3b_1 = 0, \quad \text{i.e.} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ 3b_1 + 2b_2 \end{bmatrix} = b_1 \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

In particular,

$$\mathrm{col}(A) = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\}.$$

Definition. Let V and W be vector spaces. A **linear transformation** $T : V \to W$ is a function such that for all $\mathbf{u}, \mathbf{v} \in V$ and $\alpha \in F$,

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{and} \quad T(\alpha\mathbf{u}) = \alpha T(\mathbf{u}).$$

Recall that linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are represented by matrices:

$$T(\mathbf{x}) = A\mathbf{x}, \quad A = [T] = [T(\mathbf{e}_1) \cdots T(\mathbf{e}_n)] \in \mathbb{R}^{m \times n}.$$

In this case, $\operatorname{col}(A) = R(T)$ (the range of T).

For linear transformations, one defines the **kernel** of T by

$$N(T) = \{u \in V : T(\mathbf{u}) = \mathbf{0}\}.$$

For matrix transformations, $N(T) = \operatorname{nul}(A)$.

Review. We can add some new items to our list of equivalent conditions:

Row pivots. A matrix $A \in \mathbb{R}^{m \times n}$ has a pivot in every row if and only if

$$\operatorname{col}(A) = \mathbb{R}^m.$$

Column pivots. A matrix $A \in \mathbb{R}^{m \times n}$ has a pivot in every column if and only if

$$\operatorname{nul}(A) = \{\mathbf{0}\}.$$

Furthermore, if A is a square matrix, these two conditions are equivalent.

Chapter 4. Vector Spaces

4.3 Linearly Independent Sets; Bases

Definition. A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in a vector space V is **linearly independent** if

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0} \implies c_1 = \cdots = c_n = 0.$$

Otherwise, the set is **linearly dependent**.

Example 1. The set $\{\cos t, \sin t\}$ is linearly independent in $\mathcal{F}(\mathbb{R})$; indeed, if

$$c_1 \cos t + c_2 \sin t \equiv 0,$$

then $c_1 = 0$ (set $t = 0$) and $c_2 = 0$ (set $t = \frac{\pi}{2}$).

Example 2. The set $\{1, \cos^2 t, \sin^2 t\}$ is linearly dependent in $\mathcal{F}(\mathbb{R})$; indeed,

$$\cos^2 t + \sin^2 t - 1 \equiv 0.$$

For a linearly dependent set of two or more vectors, at least one of the vectors can be written as a linear combination of the others.

Example. Show that

$$\mathbf{p}_1(t) = 2t + 1, \quad \mathbf{p}_2(t) = t, \quad \mathbf{p}_3(t) = 4t + 3$$

are **dependent** vectors in $\mathbb{P}_1$.

We need to find a non-trivial solution to

$$x_1\mathbf{p}_1(t) + x_2\mathbf{p}_2(t) + x_3\mathbf{p}_3(t) = \mathbf{0}.$$

Expanding the left-hand side, this is equivalent to

$$t(2x_1 + x_2 + 4x_3) + (x_1 + 3x_3) = 0 \quad \text{for all} \quad t.$$

This can only happen if x_1, x_2, x_3 satisfy

$$\begin{aligned} 2x_1 + x_2 + 4x_3 &= 0 \\ x_1 + 3x_3 &= 0. \end{aligned}$$

Example. (Continued) To solve this linear system, we use the augmented matrix:

$$\left[\begin{array}{ccc} 2 & 1 & 4 \\ 1 & 0 & 3 \end{array}\right] \sim \left[\begin{array}{ccc} 1 & 0 & 3 \\ 0 & 1 & -2 \end{array}\right].$$

The solution set is therefore

$$(x_1, x_2, x_3) = (-3z, 2z, z) \quad \text{for any} \quad z \in \mathbb{R}.$$

In particular, $(-3, 2, 1)$ is a solution, and hence

$$-3\mathbf{p}_1(t) + 2\mathbf{p}_2(t) + \mathbf{p}_3(t) = 0,$$

showing that $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ are linearly dependent.

Definition. Let W be a subspace of V. A set of vectors $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a **basis** for W if

(i) B is linearly independent, and

(ii) $W = \text{span}(B)$.

The plural of basis is **bases**.

Examples.

- $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the **standard basis** for $\mathbb{R}^n$.
- $B = \{1, t, \ldots, t^n\}$ is the **standard basis** for $\mathbb{P}_n$.
- $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbb{R}^n$ is a basis for $\mathbb{R}^n$ if and only if

$$A = [\mathbf{v}_1 \cdots \mathbf{v}_n] \sim I_n.$$

Pivot in every column $\iff$ columns of A are independent,
Pivot in every row $\iff$ $\text{col}(A) = \mathbb{R}^n$.

Bases for the null space. Recall that for $A \in \mathbb{R}^{m\times n}$ we have the subspace

$$\text{nul}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\} \subset \mathbb{R}^n.$$

Suppose

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus nul(A) consists of vectors of the form

$$x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} = x_3\mathbf{u} + x_4\mathbf{v}, \quad x_3, x_4 \in \mathbb{R}.$$

In particular, $\mathbf{u}$ and $\mathbf{v}$ are independent and nul$(A) = \text{span}\{\mathbf{u}, \mathbf{v}\}$.

Thus $B = \{\mathbf{u}, \mathbf{v}\}$ is a basis for nul(A).

Bases for the column space. Consider

$$A = [\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_3\,\mathbf{a}_4] = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

By definition, $\text{col}(A) = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$.

However, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ is **not** a basis for $\text{col}(A)$. (Why not?)

We see that $\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_4}$ are independent, while $\mathbf{a}_3 = -2\mathbf{a}_1 + 2\mathbf{a}_2$:

$$[\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_4] \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad [\mathbf{a}_1\,\mathbf{a}_2 \,|\, \mathbf{a}_3] \sim \left[\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Thus $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$ are independent and

$$\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\} = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\} = \text{col}(A),$$

i.e. $B = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$ is a basis for col(A).

Bases for the row space. Recall that row(A) is the span of the rows of A.

A basis for row(A) is obtained by taking the non-zero rows in the reduced echelon form of A.

This is based on the fact that $A \sim B \implies \text{row}(A) = \text{row}(B)$.

Example. Consider

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ \mathbf{b}_4 \end{bmatrix}.$$

In particular, $B = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is a basis for row(A).

Two methods. We now have two methods for finding a basis for a subspace spanned by a set of vectors.

1. Let $W = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\} = \text{row}(A)$, where

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ \mathbf{b}_4 \end{bmatrix}.$$

Then $B = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is a basis for W.

2. Let $W = \text{span}\{\mathbf{a}_1^T, \mathbf{a}_2^T, \mathbf{a}_3^T, \mathbf{a}_4^T\} = \text{col}(A^T)$, where

$$A^T = [\mathbf{a}_1^T\ \mathbf{a}_2^T\ \mathbf{a}_3^T\ \mathbf{a}_4^T] = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 1 & 2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Then $B = \{\mathbf{a}_1^T, \mathbf{a}_2^T, \mathbf{a}_3^T\}$ is a basis for W.

Chapter 4. Vector Spaces

4.4 Coordinate Systems

Unique representations. Suppose $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for some subspace $W \subset V$. Then every $\mathbf{v} \in W$ can be written as a **unique** linear combination of the elements in B.

Indeed, B spans W by definition. For uniqueness suppose

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = d_1\mathbf{v}_1 + \cdots d_n\mathbf{v}_n.$$

Then

$$(c_1 - d_1)\mathbf{v}_1 + \cdots + (c_n - d_n)\mathbf{v}_n = \mathbf{0},$$

and hence linear independence of B (also by definition) implies

$$c_1 - d_1 = \cdots = c_n - d_n = 0.$$

Definition. Given a basis $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for a subspace W, there is a unique pairing of vectors $\mathbf{v} \in W$ and vectors in $\mathbb{R}^n$, i.e.

$$\mathbf{v} \in W \mapsto (c_1, \dots, c_n) \in \mathbb{R}^n$$

where $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. We call $(c_1, \cdots, c_n)$ the **coordinates** of $\mathbf{v}$ relative to B, or the B-**coordinates** of $\mathbf{v}$. We write

$$[\mathbf{v}]_B = (c_1, \cdots, c_n).$$

Example. If $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ then $[v]_B = \mathbf{v}$.

Example. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\mathbf{v}$ be given as follows:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ 8 \end{bmatrix}.$$

To find $[v]_B$, we need to solve $A[v]_B = \mathbf{v}$ where $A = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3]$.

For this, we set up the augmented matrix:

$$[A|\mathbf{v}] = \left[\begin{array}{ccc|c} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 8 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 8 \end{array}\right].$$

Thus $[v]_B = (2, 1, 8)$, i.e. $\mathbf{v} = 2\mathbf{v}_1 + \mathbf{v}_2 + 8\mathbf{v}_3$.

Example. Let $B' = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\} \subset \mathbb{P}_2$ and $\mathbf{p} \in \mathbb{P}_2$ be given by

$$\mathbf{p}_1(t) = 1, \quad \mathbf{p}_2(t) = 1 + t, \quad \mathbf{p}_3(t) = t^2, \quad \mathbf{p}(t) = 3 + t + 8t^2.$$

To find $[\mathbf{p}]_{B'}$, we need to write

$$3 + t + 8t^2 = x_1 + x_2(1 + t) + x_3 t^2 = (x_1 + x_2) + x_2 t + x_3 t^2.$$

This leads to the system

$$\begin{aligned} x_1 + x_2 + 0x_3 &= 3 \\ 0x_1 + x_2 + 0x_3 &= 1 \\ 0x_1 + 0x_2 + x_3 &= 8 \end{aligned}$$

This is the same system as in the last example — the solution is $[\mathbf{p}]_{B'} = (2, 1, 8)$.

Isomorphism property. Suppose $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for W. Define the function $T : W \to \mathbb{R}^n$ by

$$T(\mathbf{v}) = [\mathbf{v}]_B.$$

Using the *unique* representation property, one can check:

$$\begin{aligned} &T(\mathbf{v} + \mathbf{w}) = [\mathbf{v} + \mathbf{w}]_B = [\mathbf{v}]_B + [\mathbf{w}]_B = T(\mathbf{v}) + T(\mathbf{w}), \\ &T(\alpha\mathbf{v}) = [\alpha\mathbf{v}]_B = \alpha[\mathbf{v}]_B = \alpha T(\mathbf{v}). \end{aligned}$$

Thus T is a linear transformation. Furthermore,

$$T(\mathbf{v}) = [\mathbf{v}]_B = 0 \implies \mathbf{v} = 0, \quad \text{i.e.} \quad T \quad \text{is one-to-one} \quad .$$

We call T an **isomorphism** of W onto $\mathbb{R}^n$.

Example. (again) Let $E = \{1, t, t^2\}$. This is a basis for $\mathbb{P}_2$, and

$$[\mathbf{p}_1]_E = \mathbf{v}_1, \quad [\mathbf{p}_2]_E = \mathbf{v}_2, \quad [\mathbf{p}_3]_E = \mathbf{v}_3, \quad [\mathbf{p}]_E = \mathbf{v},$$

using the notation from the previous two examples.

In particular, finding $[\mathbf{p}]_{B'}$ is equivalent to finding $[\mathbf{v}]_B$.

Indeed, recalling the isomorphism property of $T(\mathbf{p}) = [\mathbf{p}]_E$,

$$\begin{aligned}
&\mathbf{p} = x_1\mathbf{p}_1 + x_2\mathbf{p}_2 + x_3\mathbf{p}_3 \\
&\qquad \iff T(\mathbf{p}) = x_1 T(\mathbf{p}_1) + x_2 T(\mathbf{p}_2) + x_3 T(\mathbf{p}_3) \\
&\qquad \iff [\mathbf{p}]_E = x_1[\mathbf{p}_1]_E + x_2[\mathbf{p}_2]_E + x_3[\mathbf{p}_3]_E \\
&\qquad \iff \mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3.
\end{aligned}$$

That is, $[\mathbf{p}]_{B'} = [\mathbf{v}]_B$

Chapter 4. Vector Spaces

4.5 The Dimension of a Vector Space

Question. Given a vector space V, does there exist a **finite** spanning set?

Note that every vector space V has a spanning set, namely V itself.

Also note that every vector space (except for the space containing only $\mathbf{0}$) has infinitely many vectors.

Suppose $W = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.

- If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are independent, then it is a basis for W.
- Otherwise, at least one of the vectors (say $\mathbf{v}_k$) is a linear combination of the others. Then $W = \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_{k-1}\}$.

Continuing in this way, one can obtain a finite, independent spanning set for W (i.e. a basis).

Claim. If V has a basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then every basis for V has n elements.

To see this, consider the isomorphism $T : V \to \mathbb{R}^n$ given by $T(\mathbf{v}) = [\mathbf{v}]_B$.

First, we find that a set $S \subset V$ is independent if and only if $T(S) = \{T(\mathbf{u}) : \mathbf{u} \in S\} \subset \mathbb{R}^n$ is independent. This implies that any basis in V can have *at most n* elements. (*Why?*)

Similarly, $S \subset V$ spans V if and only if $T(S)$ spans $\mathbb{R}^n$. This implies that any basis V must have *at least n* elements. (*Why?*)

In fact, using this we can deduce that isomorphic vector spaces must have the same number of vectors in a basis.

Definition.
If V has a finite spanning set, then we call V **finite dimensional**.

The **dimension** of V, denoted $\dim(V)$, is the number of vectors in a basis for V.

The dimension of $\{\mathbf{0}\}$ is zero by definition.

If V is not finite dimensional, it is **infinite dimensional.**

Examples.

- $\dim(\mathbb{R}^n) = n$
- $\dim \mathbb{P}_n = n + 1$
- If $\mathbb{P}$ is the vector space of all polynomials, $\mathbb{P}$ is infinite-dimensional.
- $\mathcal{F}(\mathbb{R})$ is infinite-dimensional

Bases and subspaces. Suppose $\dim(V) = n$. and $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset V$.

If B is independent, then B is a basis for V.

(If not, there is an independent set $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}\} \subset V$. However, this yields an independent set in $\mathbb{R}^n$ with $n + 1$ elements, a contradiction).

Similarly, if $\mathrm{span}(B) = V$, then B is a basis for V.

(If not, then there is a smaller spanning set that is independent and hence a basis. This contradicts that all bases have n elements.)

The following also hold:

- Any independent set with less than n elements may be extended to a basis for V.
- If $W \subset V$ is a subspace, then $\dim(W) \leq \dim(V)$.

Note that for $V = \mathbb{R}^n$, we have the following:

- Subspaces can have any dimension $0, 1, \ldots, n$.
- For $\mathbb{R}^3$, subspaces of dimension 1 and 2 are either lines or planes through the origin.

Example 1. Find a basis for and the dimension of the subspace W spanned by

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ 8 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 5 \\ 0 \\ 13 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 4 \\ 3 \\ 11 \end{bmatrix},$$

Then

$$A = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3\, \mathbf{v}_4] \sim \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

It follows that $\dim W = 2$ and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis.

In particular, W is a plane through the origin. We also see

$$\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2, \quad \mathbf{v}_4 = \mathbf{v}_1 + \mathbf{v}_2.$$

Example 2. Find a basis for and the dimension of the subspace

$$W = \left\{ \begin{bmatrix} a+3c \\ 2b-4c \\ -a-3c \\ a+b+c \end{bmatrix} : a, b, c \in \mathbb{R} \right\}.$$

Writing

$$\begin{bmatrix} a+3c \\ 2b-4c \\ -a-3c \\ a+b+c \end{bmatrix} = a\mathbf{u} + b\mathbf{v} + c\mathbf{w} = a \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} + c \begin{bmatrix} 3 \\ -4 \\ -3 \\ 1 \end{bmatrix}$$

shows $W = \operatorname{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. However,

$$[\mathbf{u}\,\mathbf{v}\,\mathbf{w}] \sim \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

implies $\dim(W) = 2$, with $\{\mathbf{u}, \mathbf{v}\}$ a basis.

Example. (Null space, column space, row space) Let

$$A = [\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_3\,\mathbf{a}_4\,\mathbf{a}_5] = \begin{bmatrix} 1 & -2 & -1 & -2 & -1 \\ -1 & 2 & 2 & 5 & 2 \\ 0 & 0 & 2 & 6 & 2 \end{bmatrix}$$

$$\sim \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then $A = \{\mathbf{a}_1, \mathbf{a}_3\}$ and $B = \{\mathbf{b}_1, \mathbf{b}_2\}$ are bases of col(A) and row(A), respectively. Now nul(A) is given in parametric form by

$$x_2\mathbf{u}+x_4\mathbf{v}+x_5\mathbf{w} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -3 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \quad x_2, x_4, x_5 \in \mathbb{R}.$$

Thus a basis of nul(A) is given by $C = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$.

Example. (continued) For the previous example:

- $\dim(\mathrm{nul}(A)) = 3$. This is the number of free variables in the solution set of $A\mathbf{x} = \mathbf{0}$.
- $\dim(\mathrm{col}(A)) = 2$. This is the number of pivot columns.
- $\dim(\mathrm{row}(A)) = 2$. This is the number of pivot rows.
- The total number of columns equals the number of pivot columns plus the number of free variables.

Chapter 4. Vector Spaces

4.6 Rank

Last time, we finished with the example

$$A = [\mathbf{a}_1\, \mathbf{a}_2\, \mathbf{a}_3\, \mathbf{a}_4\, \mathbf{a}_5] = \begin{bmatrix} 1 & -2 & -1 & -2 & -1 \\ -1 & 2 & 2 & 5 & 2 \\ 0 & 0 & 2 & 6 & 2 \end{bmatrix}$$

and found

$$\dim(\mathrm{nul}(A)) = 3, \quad \dim(\mathrm{col}(A)) = \dim(\mathrm{row}(A)) = 2.$$

Note that

$$A^T = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3] = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 2 & 0 \\ -1 & 2 & 2 \\ -2 & 5 & 6 \\ -1 & 2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\mathrm{col}(A^T)$, and hence $\{\mathbf{v}_1^T, \mathbf{v}_2^T\}$ is a basis for $\mathrm{row}(A)$.

Thus we have seen that $\dim(\mathrm{col}(A)) = \dim(\mathrm{row}(A))$, and that this number is equal to the number of (column or row) pivots of A. Furthermore, these are all equal to the corresponding quantities for A^T.

This is true in general.

Definition. The **rank** of $A \in \mathbb{R}^{m \times n}$ is the number of pivots of A. We denote it $\mathrm{rank}(A)$.

Rank. Fix $A \in \mathbb{R}^{m\times n}$. Note that $\mathrm{rank}(A) = \mathrm{rank}(A^T)$ and

$$n = \mathrm{rank}(A) + \dim(\mathrm{nul}(A)), \quad m = \mathrm{rank}(A^T) + \dim(\mathrm{nul}(A^T)).$$

In particular

$$n - m = \dim(\mathrm{nul}(A)) - \dim(\mathrm{nul}(A^T))$$

and if $m = n$ then $\dim(\mathrm{nul}(A)) = \dim(\mathrm{nul}(A^T))$.

Row equivalence and rank. Let $A, B \in \mathbb{R}^{m\times n}$. Note that

$$A \sim B \implies \text{rank}(A) = \text{rank}(B);$$

indeed they have the same reduced echelon form. Furthermore,

$$A \sim B \iff A = PB \quad \text{for some invertible} \quad P \in \mathbb{R}^{m\times m}.$$

The $\implies$ direction is clear; for the reverse, note $P \sim I_m$.

Example. Suppose $A = PBQ$ where $P \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{n\times n}$ are invertible. Then

$$\begin{aligned}\text{rank}A = \text{rank}PBQ &= \text{rank}BQ \\ &= \text{rank}(BQ)^T = \text{rank}Q^T B^T = \text{rank}B^T = \text{rank}B.\end{aligned}$$

As a special case, if $A = PDP^{-1}$ for some diagonal matrix D, then rank(A) is equal to the number of non-zero diagonal elements of D.

Examples.

- Suppose $A \in \mathbb{R}^{3\times 8}$ and $\text{rank}A = 3$. Then:

$$\dim(\text{nul}(A)) = 5, \quad \text{rank}(A^T) = 3.$$

- Suppose $A \in \mathbb{R}^{5\times 6}$ has $\dim(\text{nul}(A)) = 4$. Then

$$\dim(\text{col}(A)) = 2.$$

- If $A \in \mathbb{R}^{4\times 6}$, what is the smallest possible dimension of the null space? *Answer:* 2
- Suppose $A \in \mathbb{R}^{10\times 12}$ and the solution set of $A\mathbf{x} = \mathbf{b}$ has 3 free variables. If we change $\mathbf{b}$, are we guaranteed to get a consistent system?

  No. We find that $\dim(\text{nul}(A)) = 3$, so that $\text{rank}(A) = 9$. Thus A does not have a pivot in every row.

Note that if $\mathbf{u} \in \mathbb{R}^{m\times 1}$ and $\mathbf{v} \in \mathbb{R}^{1\times n}$ are nonzero, then $\mathbf{uv} \in \mathbb{R}^{m\times n}$ and $\operatorname{rank}(\mathbf{uv}) = 1$; indeed, if $v = [\beta_1 \cdots \beta_n]$ then

$$\mathbf{uv} = [\beta_1\mathbf{u} \cdots \beta_n\mathbf{u}]$$

If $A \in \mathbb{R}^{m\times n}$ is written $A = LU$ where $L \in \mathbb{R}^{m\times m}$ is lower triangular and $U \in \mathbb{R}^{m\times n}$ is upper triangular, then we can write

$$A = [\mathbf{u}_1 \cdots \mathbf{u}_m] \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{bmatrix} = \mathbf{u}_1\mathbf{v}_1 + \cdots + \mathbf{u}_m\mathbf{v}_m.$$

If $\operatorname{rank}(A) = k \geq 1$ then $\mathbf{v}_1, \ldots, \mathbf{v}_k \neq 0$ and $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_m = 0$. Thus we have written

$$A = \mathbf{u}_1\mathbf{v}_1 + \cdots + \mathbf{u}_k\mathbf{v}_k$$

as the sum of k rank one matrices.

Chapter 4. Vector Spaces

4.7 Change of Basis

Let $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be bases for a vector space V. Let us describe the 'coordinate change' transformation

$$T : \mathbb{R}^n \to \mathbb{R}^n, \quad T([\mathbf{v}]_A) = [\mathbf{v}]_B.$$

Recall $T(\mathbf{x}) = [T]\mathbf{x}$, where $[T] = [T(\mathbf{e}_1) \cdots T(\mathbf{e}_n)]$. Now,

$$\mathbf{e}_k = [\mathbf{a}_k]_A, \quad \text{so that} \quad [T] = \big[[\mathbf{a}_1]_B \cdots [\mathbf{a}_n]_B\big].$$

In conclusion,

$$[\mathbf{v}]_B = P_{A \mapsto B}[\mathbf{v}]_A, \quad \text{where} \quad P_{A \mapsto B} = \big[[\mathbf{a}_1]_B \cdots [\mathbf{a}_n]_B\big].$$

We call $P_{A \mapsto B}$ the **change of coordinate matrix** from A to B.

The columns of P are independent, so that P is invertible. In fact:

$$P_{A \mapsto B}^{-1} = P_{B \mapsto A}.$$

Let the columns of A, B be bases for $\mathbb{R}^n$ and denote $E = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Then in fact

$$A = P_{A \mapsto E} \quad \text{and} \quad B = P_{B \mapsto E}.$$

Note also that

$$P_{A \mapsto B} = P_{E \mapsto B} P_{A \mapsto E}$$

(just check $[\mathbf{v}]_B = P_{E \mapsto B} P_{A \mapsto E} [\mathbf{v}]_A$). But this means

$$P_{A \mapsto B} = P_{B \mapsto E}^{-1} P_{A \mapsto E} = B^{-1} A.$$

Thus we can use row reduction to calculate $P_{A \mapsto B}$, since

$$[B|A] \sim [I_n | B^{-1} A] = [I_n | P_{A \mapsto B}].$$

Example. Let $A = \{\mathbf{a}_1, \mathbf{a}_2\}$ and $B = \{\mathbf{b}_1, \mathbf{b}_2\}$, where

$$\mathbf{a}_1 = \begin{bmatrix} 3 \\ 8 \end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad \mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{b}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Then

$$[B|A] \sim \left[\begin{array}{cc|cc} 1 & 0 & 5 & 5 \\ 0 & 1 & -2 & -1 \end{array}\right] \sim [I_2|P_{A\mapsto B}].$$

Suppose $[\mathbf{v}]_A = [2, -3]^T$. Then

$$\mathbf{v} = [\mathbf{v}]_E = P_{A\mapsto E}[\mathbf{v}]_A = \begin{bmatrix} 3 & 4 \\ 8 & 9 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -6 \\ -11 \end{bmatrix},$$

and

$$[\mathbf{v}]_B = P_{A\mapsto B}[\mathbf{v}]_A = \begin{bmatrix} 5 & 5 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -5 \\ -1 \end{bmatrix}$$

Chapter 5. Eigenvalues and Eigenvectors

5.1 Eigenvectors and Eigenvalues

Definition. Let $A \in \mathbb{C}^{n\times n}$. Suppose $\mathbf{v} \in \mathbb{R}^n$ and $\lambda \in \mathbb{C}$ satisfy

$$A\mathbf{v} = \lambda\mathbf{v} \quad \text{and} \quad \mathbf{v} \neq 0.$$

Then $\mathbf{v}$ is an **eigenvector** of A corresponding to **eigenvalue** λ.

Equivalently, if $\operatorname{nul}(A - \lambda I_n)$ is non-trivial (i.e. does not equal $\{\mathbf{0}\}$), then the non-zero vectors in this space are eigenvectors.

Note: by definition, $\mathbf{0}$ is **not** a eigenvector. However, the scalar 0 may be an eigenvalue. Indeed, this is the case whenever $\operatorname{nul}(A)$ is non-trivial!

Examples.

- Is $\mathbf{v}$ an eigenvector of A, where

$$\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & 6 & 7 \\ 3 & 3 & 7 \\ 5 & 6 & 5 \end{bmatrix} \quad ?$$

  Check:

$$A\mathbf{v} = \begin{bmatrix} 3 & 6 & 7 \\ 3 & 3 & 7 \\ 5 & 6 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ -2 \end{bmatrix} = -2\mathbf{v},$$

  so $\mathbf{v}$ is an eigenvector with eigenvalue $\lambda = -2$.

- Is $\lambda = 2$ an eigenvalue of $A = \begin{bmatrix} 3 & 2 \\ 3 & 8 \end{bmatrix}$? We check

$$A - 2I_2 = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}.$$

  This shows $\lambda = 2$ is an eigenvalue, and non-zero multiples of $[-2, 1]^T$ are eigenvectors.

Definition. If λ is an eigenvalue of $A \in \mathbb{R}^{n \times n}$, the **eigenspace** associated with λ is

$$E_\lambda = \mathrm{nul}(A - \lambda I_n).$$

That is, E_λ contains all of the eigenvectors corresponding to eigenvalue λ, along with the zero vector.

Because it is a null space, the eigenspace is a subspace. However, you can also check the definitions directly.

Example. Let

$$A = \begin{bmatrix} 5 & 2 & -1 & -1 \\ 1 & 4 & -1 & 1 \\ 7 & 8 & -2 & 1 \\ 7 & 4 & -2 & -1 \end{bmatrix}, \quad \text{which has eigenvalue} \quad 2.$$

Note

$$A - 2I_4 = \begin{bmatrix} 3 & 2 & -1 & -1 \\ 1 & 2 & -1 & 1 \\ 7 & 8 & -4 & 1 \\ 7 & 4 & -2 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -\frac{1}{2} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus

$$E_2 = \mathrm{nul}(A - 2I_4) = \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2\},$$

where $\mathbf{v}_1 = [1, -1, 0, 1]^T$ and $\mathbf{v}_2 = [0, \frac{1}{2}, 1, 0]^T$ are two particular eigenvectors that form a basis for E_2.

Theorem. (Independence) Let S be a set of eigenvectors of a matrix A corresponding to *distinct* eigenvalues. Then S is independent.

Proof. Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_{p-1}\}$ are independent but $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ are dependent. Then there exists a *non-trivial* combination so that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = 0. \quad (*)$$

Applying A to both sides,

$$c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p v_p = 0.$$

Multiply $(*)$ by λ_p and subtract it from the above equation:

$$c_1(\lambda_1 - \lambda_p)\mathbf{v}_1 + \cdots + c_{p-1}(\lambda_{p-1} - \lambda_p)\mathbf{v}_{p-1} = 0.$$

By independence, we find $c_1 = \cdots = c_{p-1} = 0$. Combining with $(*)$ then gives $c_p = 0$. This is a contradiction.

Example. Let

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

Eigenvalue and eigenvector pairs are given by

$$\lambda_1 = 1, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

$$\lambda_2 = 2, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix},$$

$$\lambda_3 = 3, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

Row reduction confirms $[\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3] \sim I_3$, so that these vectors are independent.

Triangular matrices.

Theorem. The eigenvalues of a triangular matrix are the entries along the diagonal.

To see this, recall that

$$\dim(\mathrm{nul}(A - \lambda I_n)) = n - \mathrm{rank}(A - \lambda I_n).$$

This is positive precisely when $A - \lambda I_n$ fails to have n pivots, which occurs when λ equals one of the diagonal terms of A.

Example. Consider

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

Then

$$A - 2I_3 = \begin{bmatrix} -1 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

In particular $\mathrm{rank}(A - 2I_3) = 2$, and so $\dim(\mathrm{nul}(A - 2I_3)) = 1$. In particular, 2 is an eigenvalue.

Invertibility.

Theorem. A is invertible if and only if $\lambda = 0$ is **not** an eigenvalue of A.

Indeed, A is invertible if and only if $\text{rank}A = n$, which means

$$\dim \text{nul}(A) = \dim(\text{nul}(A - 0I_n)) = 0.$$

In particular $\lambda \in \mathbb{C}$ being an eigenvalue is equivalent to:

- $\dim \text{nul}(A - \lambda I_n) > 0$
- $\text{rank}(A - \lambda I_n) < n$
- $A - \lambda I_n$ is not invertible
- $\det(A - \lambda I_n) = (-1)^n \det(\lambda I_n - A) = 0.$

Chapter 5. Eigenvalues and Eigenvectors

5.2 The Characteristic Equation

Definition. Given $A \in \mathbb{R}^{n\times n}$, $\det(\lambda I_n - A)$ is a polynomial of degree n in λ. It is known as the **characteristic polynomial of** A. Its roots are the eigenvalues of A.

Example. Consider

$$A = \left[\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array}\right] \implies \det(\lambda I_2 - A) = (1-\lambda)^2 - 4 \implies \lambda = -1, 3.$$

One finds $E_{-1} = \text{span}\{[-1,1]^T]\}$ and $E_3 = \text{span}\{[1,1]^T\}$.

Repeated eigenvalues.

Example 1.

$$A = \begin{bmatrix} -4 & -3 & 1 \\ 4 & 3 & 0 \\ -1 & -1 & 2 \end{bmatrix} \implies \det(\lambda I_3 - A) = \lambda^3 - \lambda^2 - \lambda + 1.$$

The eigenvalues are $-1, 1, 1$, and one finds

$$E_{-1} = \mathrm{span}\{[-1,1,0]^T\}, \quad E_1 = \mathrm{span}\{[-1,2,1]^T\}.$$

Example 2.

$$B = \begin{bmatrix} -5 & -4 & 2 \\ 6 & 5 & -2 \\ 0 & 0 & 1 \end{bmatrix} \implies \det(\lambda I_3 - B) = \lambda^3 - \lambda^2 - \lambda + 1.$$

The eigenvalues are $-1, 1, 1$, and one finds

$$E_{-1} = \mathrm{span}\{[-1,1,0]^T\}, \quad E_1 = \mathrm{span}\{[-\tfrac{2}{3},1,0]^T, [\tfrac{1}{3},0,1]^T\}.$$

Complex eigenvalues.

Example. $A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \implies \det(\lambda I_2 - A) = (\lambda - 1)^2 + 4.$

The eigenvalues are $1 \pm 2i$. To find the eigenspaces, we proceed exactly as before (row reduction):

$$A - (1+2i)I_2 = \begin{bmatrix} -2i & 2 \\ -2 & -2i \end{bmatrix} \sim \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}.$$

Thus eigenvectors are of the form $x = -iy$, i.e $E_{1+2i} = \text{span}\{[-i, 1]^T\}$. Similarly, $E_{1-2i} = \text{span}\{[i, 1]^T\}$.

Remark. In this case, we need to consider $A \in \mathbb{C}^{2\times 2}$ and view the eigenspaces as subspaces of $\mathbb{C}^2$.

Similar matrices.

Definition. A matrix $B \in \mathbb{R}^{n\times n}$ is **similar** to $A \in \mathbb{R}^{n\times n}$ if there exists an invertible matrix $P \in \mathbb{R}^{n\times n}$ such that $B = P^{-1}AP$. We write $A \approx B$.

Similarity is an equivalence relation.

Note that if $B = P^{-1}AP$,

$$\begin{aligned}\det(\lambda I - B) = \det(\lambda I - P^{-1}AP) &= \det(P^{-1}[\lambda I - A]P) \\ &= \det P^{-1} \det(\lambda I - A) \det P \\ &= \det(\lambda I - A).\end{aligned}$$

As a result, **similar matrices have the same eigenvalues**.

[This also shows similar matrices have equal determinants.]

Similarity and row equivalence.

Neither implies the other.

Indeed,

$$\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

which shows that similar does not imply row requivalent.

On the other hand,

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \sim I_2,$$

which shows that row equivalent does not imply similar.

Chapter 5. Eigenvalues and Eigenvectors

5.3. Diagonalization

Definition. A matrix $A \in \mathbb{R}^{n \times n}$ is called **diagonalizable** if it is similar to a diagonal matrix.

Remark. If we can diagonalize A, then we can compute its powers easily. Indeed,

$$A = P^{-1}DP \implies A^k = P^{-1}D^kP,$$

and computing powers of a diagonal matrix is straightforward.

Characterization of diagonalizability.

Theorem. A matrix $A \in \mathbb{C}^{n\times n}$ is diagonalizable precisely when there exists a basis for $\mathbb{C}^n$ consisting of eigenvectors of A. In this case, writing $(\mathbf{v}_k, \lambda_k)$ for an eigenvector/eigenvalue pair,

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = P^{-1}AP, \quad P = [\mathbf{v}_1, \ldots, \mathbf{v}_n].$$

Indeed, if $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ then

$$D = P^{-1}AP \iff AP = PD \iff [A\mathbf{v}_1 \cdots A\mathbf{v}_n] = [\lambda_1\mathbf{v}_1 \cdots \lambda_n\mathbf{v}_n]$$

Distinct eigenvalues. If $A \in \mathbb{C}^{n\times n}$ has n distinct eigenvalues, then A has n linearly independent eigenvectors and hence A is diagonalizable.

Example. Consider

$$A = \begin{bmatrix} -1 & -2 & 1 \\ 2 & 3 & 0 \\ -2 & -2 & 4 \end{bmatrix} \implies \det(\lambda I_3 - A) = (\lambda - 1)(\lambda - 2)(\lambda - 3).$$

The eigenvalues are $\lambda = 1, 2, 3$, with eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

Thus,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = P^{-1}AP, \quad P = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3].$$

If an $n \times n$ matrix does not have n distinct eigenvalues, it may or may not be diagonalizable.

Example. (from before)

$$B = \begin{bmatrix} -5 & -4 & 2 \\ 6 & 5 & -2 \\ 0 & 0 & 1 \end{bmatrix} \implies \det(\lambda I_3 - B) = \lambda^3 - \lambda^2 - \lambda + 1.$$

The eigenvalues are $-1, 1, 1$, and one finds

$$E_{-1} = \mathrm{span}\{[-1, 1, 0]^T\}, \quad E_1 = \mathrm{span}\{[-\tfrac{2}{3}, 1, 0]^T, [\tfrac{1}{3}, 0, 1]^T\}.$$

Thus, while 1 is a repeated eigenvalue, the matrix is diagonalizable:

$$\mathrm{diag}(-1, 1, 1) = P^{-1} B P, \quad P = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3].$$

Example. (from before)

$$A = \begin{bmatrix} -4 & -3 & 1 \\ 4 & 3 & 0 \\ -1 & -1 & 2 \end{bmatrix} \implies \det(\lambda I_3 - A) = \lambda^3 - \lambda^2 - \lambda + 1.$$

The eigenvalues are $-1, 1, 1$, and one finds

$$E_{-1} = \text{span}\{[-1, 1, 0]^T\}, \quad E_1 = \text{span}\{[-1, 2, 1]^T\}.$$

In particular, one cannot form a basis of eigenvectors. The matrix is not diagonalizable.

(*Question.* Why can't some other basis diagonalize A?)

Example. We previously saw the matrix

$$A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

has eigenvalues $1 \pm 2i$, with eigenspaces spanned by

$$\mathbf{v}_1 = [-i, 1]^T, \quad \mathbf{v}_2 = [i, 1]^T.$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2\}$ are a basis for $\mathbb{C}^2$ and

$$\operatorname{diag}(1+2i, 1-2i) = P^{-1}AP, \quad P = [\mathbf{v}_1\,\mathbf{v}_2].$$

Similarity, diagonalization, linear transformations. Suppose

$$A \in \mathbb{C}^{n\times n}, \quad A' = P^{-1}AP, \quad P = [\mathbf{v}_1 \cdots \mathbf{v}_n],$$

where $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ forms a basis for $\mathbb{C}^n$. Let

$$T(\mathbf{x}) = A\mathbf{x}, \quad T'(\mathbf{x}) = A'\mathbf{x}.$$

Noting that $P = P_{B\mapsto E}$ and $P^{-1} = P_{E\mapsto B}$, we have

$$T'([\mathbf{v}]_B) = P^{-1}AP[\mathbf{v}]_B = P^{-1}A[\mathbf{v}]_E = P^{-1}T(\mathbf{v}) = [T(\mathbf{v})]_B.$$

- We call A' the (standard) matrix for T *relative to the basis* B. We may also say A' is the $B-$matrix for T. We write $[T]_B = P^{-1}AP$ and note $[T(\mathbf{v})]_B = [T]_B[\mathbf{v}]_B$.
- If B is a basis of eigenvectors, then we see that relative to the basis B the transformation simply scales along the lines containing the eigenvectors.

Chapter 5. Eigenvalues and Eigenvectors

5.4. Eigenvectors and Linear Transformations

Transformation matrix. Let $B = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for the vector space V, and $C = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ a basis for the vector space W. Given a linear transformation $T : V \to W$, we define

$$\hat{T} : \mathbb{C}^n \to \mathbb{C}^m \quad \text{by} \quad \hat{T}([\mathbf{v}]_B) = [T(\mathbf{v})]_C.$$

This is a linear transformation (*check!*).

- $\hat{T}$ expresses T using coordinates relative to the bases B, C
- We may write $\hat{T}(\mathbf{x}) = M\mathbf{x}$, where

$$M = [\hat{T}(\mathbf{e}_1) \cdots \hat{T}(\mathbf{e}_n)] = \big[[T(\mathbf{v}_1)]_C \cdots [T(\mathbf{v}_n)]_C\big].$$

- Note $[T(\mathbf{v})]_C = \hat{T}([\mathbf{v}]_B) = M[\mathbf{v}]_B$. That is, M is the matrix for T relative to the bases B and C.

Example. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be a basis for V and $C = \{\mathbf{w}_1, \mathbf{w}_2\}$ be a basis for W. Suppose $T : V \to W$, with

$$T(\mathbf{v}_1) = \mathbf{w}_1 + \mathbf{w}_2, \quad T(\mathbf{v}_2) = \mathbf{w}_1 - 3\mathbf{w}_2, \quad T(\mathbf{v}_3) = \mathbf{9w}_1.$$

The matrix for T relative to B and C is

$$M = \big[[T(\mathbf{v}_1)]_C \, [T(\mathbf{v}_2)]_C \, [T(\mathbf{v}_3)]_C\big] = \begin{bmatrix} 1 & 1 & 9 \\ 1 & -3 & 0 \end{bmatrix}$$

If $\mathbf{v} = \mathbf{v}_1 + 2\mathbf{v}_2 - 3\mathbf{v}_3$, then

$$[T(\mathbf{v})]_C = M[\mathbf{v}]_B = M \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -24 \\ -5 \end{bmatrix},$$

and so $T(\mathbf{v}) = -24\mathbf{w}_1 - 5\mathbf{w}_2$.

Example. Let $T : \mathbb{P}_2 \to \mathbb{P}_3$ be given by

$$T(p(t)) = (t+5)p(t).$$

Then T is a linear transformation. (*Check!*)

Let us find the matrix of T relative to

$$B = \{1, t, t^2\} \quad \text{and} \quad C = \{1, t, t^2, t^3\}.$$

Since

$$T(1) = 5 + t, \quad T(t) = 5t + t^2, \quad T(t^2) = 5t^2 + t^3,$$

we find

$$M = \big[[T(1)]_C [T(t)]_C [T(t^2)]_C \big] = \begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}.$$

Matrix transformations. If $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{C}^n$ and $C = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ is a basis for $\mathbb{C}^m$ and $T(\mathbf{x}) = A\mathbf{x}$ for some $A \in \mathbb{C}^{m\times n}$, then the matrix for T relative to B and C is

$$M = P_{E\mapsto C} A P_{B\mapsto E}, \quad \text{with} \quad M[\mathbf{v}]_B = [A\mathbf{v}]_C.$$

For example:

- If B and C are the elementary bases, then $M = A$ (the standard matrix for T).
- If $B = C$, then $M = P^{-1}AP = [T]_B$, where

$$P = [\mathbf{v}_1, \ldots, \mathbf{v}_n] = P_{B\mapsto E}.$$

Example. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $C = \{\mathbf{w}_1, \mathbf{w}_2\}$ be bases for $\mathbb{R}^3, \mathbb{R}^2$ given by

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}, \mathbf{v}_3 = \mathbf{e}_3, \mathbf{w}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

and

$$T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 4 & 0 \end{bmatrix}$$

Note

$$P_{B \mapsto E} = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3], \quad P_{C \mapsto E} = [P_{E \mapsto C}]^{-1} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}.$$

Then the matrix for T relative to B and C is

$$M = P_{E \mapsto C} A P_{B \mapsto E} = \begin{bmatrix} 15 & -22 & 10 \\ -8 & 13 & -6 \end{bmatrix}.$$

We use this via $[T(\mathbf{v})]_C = M[\mathbf{v}]_B$.

Example. Let

$$A = \begin{bmatrix} 2 & 3 & -3 \\ 3 & -2 & 3 \\ 3 & -3 & 4 \end{bmatrix}.$$

Find a basis B for $\mathbb{R}^3$ so that the matrix for $T(\mathbf{x}) = A\mathbf{x}$ is *diagonal*.

The eigenvalues of A are $\lambda = -2, 1, 1$, with $E_{-2} = \text{span}\{\mathbf{v}_1\}$ and $E_1 = \text{span}\{\mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $P = P_{B \mapsto E} = [\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3]$. Then the matrix for T relative to B is

$$M = P^{-1}AP = P_{E \mapsto B} A P_{B \mapsto E} = \text{diag}(-2, 1, 1).$$

Chapter 5. Eigenvalues and Eigenvectors

5.5 Complex Eigenvalues

Vectors in $\mathbb{C}^n$. Recall that for $z = \alpha + i\beta \in \mathbb{C}$ (where $\alpha, \beta \in \mathbb{R}$), we have

$$\bar{z} = \alpha - i\beta, \quad \operatorname{Re} z = \alpha = \tfrac{1}{2}(z + \bar{z}), \quad \operatorname{Im} z = \beta = \tfrac{1}{2i}(z - \bar{z}).$$

If $\mathbf{v} = (c_1, \ldots, c_n) \in \mathbb{C}^n$, then we write

$$\bar{\mathbf{v}} = (\bar{c}_1, \ldots, \bar{c}_n).$$

Or, writing $\mathbf{v} = \mathbf{x} + i\mathbf{y}$, we can write

$$\bar{\mathbf{v}} = \mathbf{x} - i\mathbf{y}, \quad \operatorname{Re} \mathbf{v} = \mathbf{x} = \tfrac{1}{2}(\mathbf{v} + \bar{\mathbf{v}}), \quad \operatorname{Im} \mathbf{v} = \tfrac{1}{2i}(\mathbf{v} - \bar{\mathbf{v}}).$$

Note that $\mathbf{v}$ and $\bar{\mathbf{v}}$ are independent if and only if $\operatorname{Re} \mathbf{v}$ and $\operatorname{Im} \mathbf{v}$ are independent. Similarly,

$$\operatorname{span}\{\mathbf{v}, \bar{\mathbf{v}}\} = \operatorname{span}\{\operatorname{Re} \mathbf{v}, \operatorname{Im} \mathbf{v}\}.$$

Conjugate pairs. Suppose $\lambda = \alpha + i\beta \in \mathbb{C}$ is an eigenvalue for $A \in \mathbb{R}^{n\times n}$ with eigenvector $\mathbf{v} = \mathbf{x} + i\mathbf{y}$. Note that

$$A\mathbf{v} = \lambda\mathbf{v} \implies A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}, \quad \text{i.e.} \quad A[\mathbf{v}\,\bar{\mathbf{v}}] = [\mathbf{v}\,\bar{\mathbf{v}}] \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}.$$

Note that

$$A(\mathbf{x}+i\mathbf{y}) = (\alpha+i\beta)(\mathbf{x}+i\mathbf{y}) \implies A\mathbf{x} = \alpha\mathbf{x} - \beta\mathbf{y}, \quad A\mathbf{y} = \beta\mathbf{x} + \alpha\mathbf{y}.$$

In particular,

$$A[\mathbf{x}\,\mathbf{y}] = [\mathbf{x}\,\mathbf{y}] \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} = r[\mathbf{x}\,\mathbf{y}]R_\theta,$$

where $r = \sqrt{\alpha^2 + \beta^2}$ and R_θ is the 2×2 rotation matrix by $\theta \in [0, 2\pi)$, defined by $\cos\theta = \frac{\alpha}{r}$ and $\sin\theta = -\frac{\beta}{r}$.

Example. The matrix

$$A = \begin{bmatrix} 1 & 5 \\ -2 & 3 \end{bmatrix}$$

has eigenvalues $\lambda = \alpha \pm i\beta$, where $(\alpha, \beta) = (2, 3)$ and eigenvectors $\mathbf{v}, \bar{\mathbf{v}} = \mathbf{x} \pm i\mathbf{y}$, where $\mathbf{x} = [1, 2]^T$ and $\mathbf{y} = [3, 0]^T$.

We can diagonalize A via

$$A = [\mathbf{v}\,\bar{\mathbf{v}}]\mathrm{diag}(\lambda, \bar{\lambda})[\mathbf{v}\,\bar{\mathbf{v}}]^{-1}.$$

Alternatively, we can write

$$A = [\mathbf{x}\,\mathbf{y}]\sqrt{13}R_\theta[\mathbf{x}\,\mathbf{y}]^{-1},$$

where $\theta \sim .9828$ radians. Writing $T(\mathbf{x}) = A\mathbf{x}$, with respect to the basis $B = [\mathbf{x}\,\mathbf{y}]$, T^n performs rotation by $n\theta$ and a dilation by $13^{\frac{n}{2}}$.

Example. Let

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 4 & 3 \\ -2 & -4 & -2 \end{bmatrix}.$$

We have the following eigenvalues and eigenvector pairs:

$$(\mathbf{u}, \lambda_1) = \left(\left[\begin{smallmatrix} 2 \\ 1 \\ -2 \end{smallmatrix}\right], 2\right), \quad (\mathbf{v}, \bar{\mathbf{v}}, \alpha \pm i\beta) = \left(\left[\begin{smallmatrix} 0 \\ -1 \\ -1 \end{smallmatrix}\right] \pm i \left[\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right], 1 \pm i\right).$$

Thus $A = PDP^{-1}$, where $P = [\mathbf{u}\,\mathbf{v}\,\bar{\mathbf{v}}]$ and $D = \mathrm{diag}(2, 1-i, 1+i)$.

Or, we can write $A = PQP^{-1}$, where $P = [\mathbf{u}\,\mathbf{x}\,\mathbf{y}]$ and

$$Q = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Example. (cont.) We can write $A = PQP^{-1}$, where $P = [\mathbf{u}\,\mathbf{x}\,\mathbf{y}]$ and

$$Q = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Thus, writing $T(\mathbf{x}) = A\mathbf{x}$ and $B = \{\mathbf{u}, \mathbf{x}, \mathbf{y}\}$, we can describe the effect of T in the B-coordinate system as follows: T scales by a factor of 2 along the **x**-axis; in the **yz**-plane, T rotates by $3\pi/4$ and scales by $\sqrt{2}$.

Chapter 5. Eigenvalues and Eigenvectors

5.7 Applications to Differential Equations

Scalar linear homogeneous ODE. Consider a second order ODE of the form

$$x'' + bx' + cx = 0.$$

Defining

$$x_1 = x, \quad x_2 = x', \quad \mathbf{x} = (x_1, x_2)^T,$$

we can rewrite the ODE as a 1st order 2x2 system:

$$\mathbf{x}' = A\mathbf{x}, \quad A = \left[\begin{array}{rr} 0 & 1 \\ -c & -b \end{array} \right].$$

Similarly, an n^{th} order linear homogeneous ODE can be written as a 1st order $n \times n$ system.

Matrix Exponential. How do we solve

$$\mathbf{x}' = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (*)$$

when A is a matrix and $\mathbf{x}$ is a vector? *Answer.* Same as in the scalar case! I.e. the solution to $(*)$ is

$$\mathbf{x}(t) = e^{At}\mathbf{x}_0.$$

The only question is... what does e^{At} mean? *Answer.* Same as in the scalar case!

Definition. For an $n \times n$ matrix A,

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Theorem. The solution to $(*)$ is given by $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$. We call e^{tA} the **fundamental matrix** for $\mathbf{x}' = A\mathbf{x}$.

Computing matrix exponentials. The matrix exponential is a powerful tool for solving linear systems. But how do we actually compute it?

Example 1. If $A = \text{diag}(\lambda_1, \ldots, \lambda_n)$, then

$$e^A = \text{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n}).$$

Example 2. If $A = PDP^{-1}$, then

$$e^A = Pe^D P^{-1}.$$

Combining with Example 1, we see that if A is diagonalizable, we can compute its exponential.

Example 3. $e^0 = I$.

Example 4. If A is **nilpotent** (that is, $A^{k_0} = 0$ for some k_0), then

$$e^A = \sum_{k=0}^{k_0-1} \frac{A^k}{k!} \quad \text{(a finite sum)}.$$

Example 5. If $AB = BA$, then

$$e^{A+B} = e^A e^B = e^B e^A.$$

In particular, e^A is invertible for any A, with

$$(e^A)^{-1} = e^{-A}.$$

Numerical example. Consider

$$x'' - 4x' + 3x = 0 \implies \mathbf{x}' = A\mathbf{x}, \quad A = \begin{bmatrix} 0 & 1 \\ -3 & 4 \end{bmatrix}$$

Then

$$A = PDP^{-1}, \quad P = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}, \quad D = \mathrm{diag}(1, 3)$$

Thus the fundamental matrix is

$$e^{tA} = P[\mathrm{diag}(e^t, e^{3t})]P^{-1}.$$

Using the columns of P as initial conditions, one gets the two solutions

$$P\mathrm{diag}(e^t, e^{3t})\mathbf{e}_j, \quad j = 1, 2.$$

Numerical example. (cont.) This gives two linearly independent solutions, namely

$$\mathbf{x}(t) = \begin{bmatrix} e^t \\ e^t \end{bmatrix} \quad \text{and} \quad \mathbf{x}(t) = \begin{bmatrix} e^{3t} \\ 3e^{3t} \end{bmatrix}.$$

This corresponds to the two solutions $x(t) = e^t$ and $x(t) = e^{3t}$. Any other solution is a linear combination of these (as determined by the initial conditions).

Complex eigenvalues. Note that an ODE with all real coefficients could lead to complex eigenvalues. In this case, you should diagonalize (and hence solve the ODE) via sines and cosines as in the previous section.

Chapter 6. Orthogonality and Least Squares

6.1 Inner Product, Length, and Orthogonality

Conjugate transpose. If $A \in \mathbb{C}^{m \times n}$, then we define

$$A^* = (\bar{A})^T \in \mathbb{C}^{n \times m}.$$

We call A^* the **conjugate transpose** or the **adjoint** of A. If $A \in R^{m \times n}$, then $A^* = A^T$.

Note that

- $(\alpha A + \beta B)^* = \bar{\alpha} A^* + \bar{\beta} B^*$
- $(AC)^* = C^* A^*$.

Definition. $A \in \mathbb{C}^{n \times n}$ is **hermitian** if $A^* = A$.

Note that $A \in \mathbb{R}^{n \times n}$ is hermitian if and only if it is **symmetric**, i.e. $A = A^T$.

Definition. If $\mathbf{u} = (a_1, \dots, a_n) \in \mathbb{C}^n$ and $\mathbf{v} = (b_1, \dots, b_n) \in \mathbb{C}^n$, then we define the **inner product** of $\mathbf{u}$ and $\mathbf{v}$ by

$$\mathbf{u} \cdot \mathbf{v} = \bar{a}_1 b_1 + \cdots + \bar{a}_n b_n \in \mathbb{C}.$$

Note that if we regard $\mathbf{u}, \mathbf{v}$ as $n \times 1$ matrices, then $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^* \mathbf{v}$.

Note also that for $A = [\mathbf{u}_1 \dots \mathbf{u}_k] \in \mathbb{C}^{n \times k}$ and $B = [\mathbf{v}_1 \dots \mathbf{v}_\ell] \in \mathbb{C}^{n \times \ell}$, then $A^* B \in \mathbb{C}^{k \times \ell}$,

$$A^* B = [\mathbf{u}_i \cdot \mathbf{v}_j] \quad i = 1, \dots, k, \quad j = 1, \dots, \ell.$$

Properties of the inner product. For $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$:

- $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$
- $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
- $\alpha(\mathbf{u} \cdot \mathbf{v}) = (\overline{\alpha}\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (\alpha \mathbf{v})$
- If $\mathbf{u} = (a_1, \cdots, a_n) \in \mathbb{C}^n$, then

$$\mathbf{u} \cdot \mathbf{u} = |a_1|^2 + \cdots + |a_n|^2 \geq 0,$$

  and $\mathbf{u} \cdot \mathbf{u} = 0$ only if $\mathbf{u} = 0$.

Definition. If $\mathbf{u} = (a_1, \ldots, a_n) \in \mathbb{C}^n$, then the **norm** of $\mathbf{u}$ is given by

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{|a_1|^2 + \cdots + |a_n|^2}.$$

Properties. For $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$,

- $\|\alpha \mathbf{u}\| = |\alpha| \, \|\mathbf{u}\|$
- $|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \|\mathbf{v}\|$ (Cauchy–Schwarz inequality)
- $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (triangle inequality)

The norm measures length; $\|\mathbf{u} - \mathbf{v}\|$ measures the distance between $\mathbf{u}$ and $\mathbf{v}$.

A vector $\mathbf{u} \in \mathbb{C}^n$ is a **unit vector** if $\|\mathbf{u}\| = 1$.

Example. Let $A = [\mathbf{v}_1\ \mathbf{v}_2] = \begin{bmatrix} 1 & i \\ 3+8i & 2i \end{bmatrix}$.

Then $A^* = \begin{bmatrix} 1 & 3-8i \\ -i & -2i \end{bmatrix}$. So A is not hermitian.

We have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 1 \cdot i + (3-8i) \cdot 2i = 16 + 7i$.

Note $\|\mathbf{v}_1\|^2 = 1 \cdot 1 + (3-8i)(3+8i) = 74$.

Consequently, $\frac{1}{\sqrt{74}}\mathbf{v}_1$ is a unit vector.

Definition. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$. We write $\mathbf{u} \perp \mathbf{v}$.

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{C}^n$ is an **orthogonal** set if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for each $i, j = 1, \ldots, k$ (with $i \neq j$).

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{C}^n$ is an **orthonormal** set if it is orthogonal and each $\mathbf{v}_i$ is a unit vector.

Remark. In general, we have

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

However, we have

$$\mathbf{u} \perp \mathbf{v} \implies \|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

This is the **Pythagorean theorem**.

Definition. Let $W \subset \mathbb{C}^n$. The **orthogonal complement** of W, denoted $W^\perp$, is defined by

$$W^\perp = \{\mathbf{v} \in \mathbb{C}^n : \mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for every} \quad \mathbf{w} \in W\}.$$

Subspace property. $W^\perp$ is a subspace of $\mathbb{C}^n$ satisfying $W \cap W^\perp = \{0\}$.

Indeed, $W^\perp$ is closed under addition and scalar multiplication, and $\mathbf{w} \cdot \mathbf{w} = 0 \implies \mathbf{w} = 0$.

Suppose $A = [\mathbf{v}_1 \cdots \mathbf{v}_k] \in \mathbb{C}^{n\times k}$. Then

$$[\mathrm{col}(A)]^\perp = \mathrm{nul}(A^*).$$

Indeed

$$\mathbf{0} = A^*\mathbf{x} = \begin{bmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_k^* \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{v}_1^*\mathbf{x} \\ \vdots \\ \mathbf{v}_k^*\mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{v}_k \cdot \mathbf{x} \end{bmatrix}$$

if and only if

$$\mathbf{v}_1 \cdot \mathbf{x} = \cdots = \mathbf{v}_k \cdot \mathbf{x} = 0.$$

Example 1. Let $\mathbf{v}_1 = [1, -1, 2]^T$ and $\mathbf{v}_2 = [0, 2, 1]^T$. Note that $\mathbf{v}_1 \perp \mathbf{v}_2$.

Let $W = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and $A = [\mathbf{v}_1 \mathbf{v}_2]$. Note that

$$W^\perp = [\text{col}(A)]^\perp = \text{nul}(A^*) = \text{nul}(A^T).$$

We have

$$A^T = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & \frac{5}{2} \\ 0 & 1 & \frac{1}{2} \end{bmatrix},$$

and thus $\text{nul}(A^T) = \text{span}\{\mathbf{v}_3\} = \text{span}\{[-\frac{5}{2}, -\frac{1}{2}, 1]^T\}$.

Note $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set, and $W^\perp$ is a line perpendicular to the plane W.

Example 2. Let $\mathbf{v}_1 = [1, -1, 1, -1]^T$ and $\mathbf{v}_2 = [1, 1, 1, 1]^T$. Again, $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Let $W = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and $A = [\mathbf{v}_1\, \mathbf{v}_2]$ as before. Then $W^\perp = \text{nul}(A^T)$, with

$$A^T = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

In particular, $W^\perp = \text{span}\{\mathbf{v}_3, \mathbf{v}_4\}$, with

$$\mathbf{v}_3 = [-1, 0, 1, 0]^T \quad \mathbf{v}_4 = [0, -1, 0, 1]^T.$$

Again, $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ is an orthogonal set. This time W and $W^\perp$ are planes in $\mathbb{R}^4$, with $W \cap W^\perp = \{0\}$.

Chapter 6. Orthogonality and Least Squares

6.2 Orthogonal Sets

Definition. If S is an orthogonal set that is linearly independent, then we call S an **orthogonal basis** for $\text{span}(S)$.

Similarly, a linearly independent orthonormal set S is a **orthonormal basis** for $\text{span}(S)$.

Example. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

- $S = \{\mathbf{v}_1, \mathbf{v}_2, 0\}$ is an orthogonal set, but not a basis for $\mathbb{C}^3$
- $S = \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_3\}$ is an orthogonal basis for $\mathbb{C}^3$
- $S = \{\frac{1}{\sqrt{2}}\mathbf{v}_1, \mathbf{v}_2, \frac{1}{\sqrt{2}}\mathbf{v}_3\}$ is an orthonormal basis for $\mathbb{C}^3$

Test for orthogonality. Let $A = [\mathbf{v}_1 \cdots \mathbf{v}_p] \in \mathbb{C}^{n\times p}$. Note that

$$A^*A = \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \cdots & \mathbf{v}_1 \cdot \mathbf{v}_p \\ \vdots & \ddots & \vdots \\ \mathbf{v}_p \cdot \mathbf{v}_1 & \cdots & \mathbf{v}_p \cdot \mathbf{v}_p \end{bmatrix} \in \mathbb{C}^{p\times p}.$$

Thus A^*A is diagonal precisely when $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is orthogonal.

Furthermore, $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is orthonormal precisely when $A^*A = I_p$.

Definition. A matrix $A \in \mathbb{C}^{n \times n}$ is **unitary** if $A^* A = I_n$.

The following conditions are equivalent:

- $A \in \mathbb{C}^{n \times n}$ is unitary
- $A \in \mathbb{C}^{n \times n}$ satisfies $A^{-1} = A^*$
- the columns of A are an orthonormal basis for $\mathbb{C}^n$
- $A \in \mathbb{C}^{n \times n}$ satisfies $AA^* = I_n$
- the rows of A are an orthonormal basis for $\mathbb{C}^n$

Theorem. (Independence) If $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal set of non-zero vectors, then S is independent and S is a basis for $\text{span}(S)$.

Indeed, suppose

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}.$$

Now take an inner product with $\mathbf{v}_j$:

$$\begin{aligned} 0 &= c_1\mathbf{v}_1 \cdot \mathbf{v}_j + \cdots + c_j\mathbf{v}_j \cdot \mathbf{v}_j + \cdots + c_p\mathbf{v}_p \cdot \mathbf{v}_j \\ &= 0 + \cdots + c_j\|\mathbf{v}_j\|^2 + \cdots + 0. \end{aligned}$$

Thus $c_j = 0$ for any $j = 1, \dots, p$.

Theorem. If $S = \{\mathbf{w}_1, \dots, \mathbf{w}_p\} \subset W$ are independent and $T = \{\mathbf{v}_1, \dots, \mathbf{v}_q\} \subset W^\perp$ are independent, then $S \cup T$ is independent.

Theorem. Suppose $W \subset \mathbb{C}^n$ has dimension p. Then $\dim(W^\perp) = n - p$.

Let $A = [\mathbf{w}_1 \cdots \mathbf{w}_p]$, where $\{\mathbf{w}_1, \dots, \mathbf{w}_p\}$ is a basis for W. Note

$$W^\perp = [\mathrm{col}(A)]^\perp = \mathrm{nul}(A^*) \subset \mathbb{C}^n.$$

Thus

$$\dim(W^\perp) = \dim(\mathrm{nul}(A^*)) = n - \mathrm{rank}(A^*) = n - \mathrm{rank}(A) = n - p.$$

In particular, we find

$$\dim(W) + \dim(W^\perp) = \dim(\mathbb{C}^n) = n.$$

Remark. If $B = \{\mathbf{w}_1, \dots, w_p\}$ is a basis for W and $C = \{\mathbf{v}_1, \dots, \mathbf{v}_{n-p}\}$ is a basis for $W^\perp$, then $B \cup C$ is a basis for $\mathbb{C}^n$.

Theorem. (Orthogonal decomposition) Let W be a subspace of $\mathbb{C}^n$. For every $\mathbf{x} \in \mathbb{C}^n$ there exist unique $\mathbf{y} \in W$ and $\mathbf{z} \in W^\perp$ such that $\mathbf{x} = \mathbf{y} + \mathbf{z}$.

Indeed, let $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ be a basis for W and $C = \{\mathbf{v}_1, \ldots, \mathbf{v}_{n-p}\}$ a basis for $W^\perp$. Then $B \cup C$ is a basis for $\mathbb{C}^n$, and so every $\mathbf{x}$ has a unique representation $\mathbf{x} = \mathbf{y} + \mathbf{z}$, where $\mathbf{y} \in \mathrm{span}(B)$ and $\mathbf{z} \in \mathrm{span}(C)$.

Uniqueness can also be deduced from the fact that $W \cap W^\perp = \{0\}$.

Remark. Suppose B is an **orthogonal basis** or W. Then

$$\mathbf{x} = \alpha_1 \mathbf{w}_1 + \cdots + \alpha_p \mathbf{w}_p + \mathbf{z}, \quad \mathbf{z} \in W^\perp.$$

One can compute α_j via

$$\mathbf{w}_j \cdot \mathbf{x} = \alpha_j \mathbf{w}_j \cdot \mathbf{w}_j \implies \alpha_j = \tfrac{\mathbf{w}_j \cdot \mathbf{x}}{\|\mathbf{w}_j\|^2}.$$

Projection. Let W be a subspace of $\mathbb{C}^n$. As above, for each $\mathbf{x} \in \mathbb{C}^n$ there exists a unique $\mathbf{y} \in W$ and $\mathbf{z} \in W^\perp$ so that $\mathbf{x} = \mathbf{y} + \mathbf{z}$. We define

$$\text{proj}_W : \mathbb{C}^n \to W \subset \mathbb{C}^n \quad \text{by} \quad \text{proj}_W \mathbf{x} = \mathbf{y}.$$

We call proj_W the (orthogonal) **projection** of $\mathbb{C}^n$ onto W.

Example. Suppose $W = \text{span}\{\mathbf{w}_1\}$. Then

$$\text{proj}_W \mathbf{x} = \tfrac{\mathbf{w}_1 \cdot \mathbf{x}}{\|\mathbf{w}_1\|^2} \mathbf{w}_1.$$

Note that proj_W is a linear transformation, with matrix representation given by

$$[\text{proj}_W]_E = \tfrac{1}{\|\mathbf{w}_1\|^2} \mathbf{w}_1 \mathbf{w}_1^* \in \mathbb{C}^{n \times n}.$$

Example. Let $\mathbf{w}_1 = [1,0,1]^T$ and $\mathbf{v} = [-1,2,2]^T$, with $W = \text{span}\{\mathbf{w}_1\}$. Then

$$\text{proj}_W(\mathbf{v}) = \tfrac{\mathbf{w}_1 \cdot \mathbf{v}}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 = \tfrac{1}{2}\mathbf{w}_1.$$

In fact,

$$[\text{proj}_W]_E = \tfrac{1}{\|w_1\|^2}\mathbf{w}_1\mathbf{w}_1^* = \tfrac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Thus

$$\text{proj}_W(\mathbf{x}) = \tfrac{1}{2}\begin{bmatrix} x_1 + x_3 \\ 0 \\ x_1 + x_3 \end{bmatrix}.$$

Chapter 6. Orthogonality and Least Squares

6.3 Orthogonal Projections

Orthogonal projections. Let W be a subspace of $\mathbb{C}^n$. Recall that

$$\text{proj}_W \mathbf{x} = \mathbf{y}, \quad \text{where} \quad \mathbf{x} = \mathbf{y} + \mathbf{z}, \quad \mathbf{y} \in W,\ \mathbf{z} \in W^{\perp}.$$

Let $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ be a basis for $W \subset \mathbb{C}^n$. We wish to find the B-coordinates of $\text{proj}_W \mathbf{x}$, i.e. to write

$$\mathbf{x} = \alpha_1 \mathbf{w}_1 + \cdots + \alpha_p \mathbf{w}_p + \mathbf{z}, \quad \mathbf{z} \in W^{\perp}.$$

This yields a system of p equations and p unknowns:

$$\begin{aligned} \mathbf{w}_1 \cdot \mathbf{x} &= \alpha_1 \mathbf{w}_1 \cdot \mathbf{w}_1 \cdots + \alpha_p \mathbf{w}_1 \cdot \mathbf{w}_p \\ &\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \mathbf{w}_p \cdot \mathbf{x} &= \alpha_1 \mathbf{w}_p \cdot \mathbf{w}_1 + \cdots + \alpha_p \mathbf{w}_p \cdot \mathbf{w}_p. \end{aligned}$$

Normal system. Write $A = [\mathbf{w}_1 \cdots \mathbf{w}_p] \in \mathbb{C}^{n \times p}$. The system

$$\begin{aligned} \mathbf{w}_1^* \mathbf{x} &= \alpha_1 \mathbf{w}_1^* \mathbf{w}_1 \cdots + \alpha_p \mathbf{w}_1^* \mathbf{w}_p \\ &\vdots \qquad\qquad \vdots \\ \mathbf{w}_p^* \mathbf{x} &= \alpha_1 \mathbf{w}_p^* \mathbf{w}_1 + \cdots + \alpha_p \mathbf{w}_p^* \mathbf{w}_p \end{aligned}$$

may be written as the **normal system** $A^* A \hat{x} = A^* \mathbf{x}$, where

$$\hat{x} = [\mathrm{proj}_W(\mathbf{x})]_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}.$$

One calls $A^* A \in \mathbb{C}^{p \times p}$ the **Gram matrix.**

- The normal system has at least one solution, namely $[\mathrm{proj}_W(\mathbf{x})]_B$.
- If the normal system has a unique solution, it is $[\mathrm{proj}_W(\mathbf{x})]_B$.

Theorem. (Null space and rank of A^*A) If $A \in \mathbb{C}^{n\times p}$, then $A^*A \in \mathbb{C}^{p\times p}$ satisfies

$$\mathrm{nul}(A^*A) = \mathrm{nul}(A) \quad \text{and} \quad \mathrm{rank}(A^*A) = \mathrm{rank}(A).$$

- First note $\mathrm{nul}(A) \subset \mathrm{nul}(A^*A)$.
- If instead $A\mathbf{x} \in \mathrm{nul}(A^*) = [\mathrm{col}(A)]^\perp$, then $A\mathbf{x} \in \mathrm{col}(A) \cap \mathrm{col}(A)^\perp$ and hence $A\mathbf{x} = 0$.
  Thus $\mathrm{nul}(A^*A) \subset \mathrm{nul}(A)$.
- Thus

$$\mathrm{rank}(A^*A) = p - \dim(\mathrm{nul}(A^*A)) = p - \dim(\mathrm{nul}A) = \mathrm{rank}(A).$$

Solving the normal system. If the columns of A are independent, then $A^*A \in C^{p\times p}$ and $\mathrm{rank}(A^*A) = \mathrm{rank}(A) = p$ and hence A^*A is invertible.

Solving the normal system. Suppose $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ is a basis for a subspace $W \subset \mathbb{C}^n$. Writing $A = [\mathbf{w}_1, \ldots, \mathbf{w}_p]$, we have that A^*A is invertible and the normal system

$$A^*A\hat{x} = A^*\mathbf{x}$$

has a unique solution

$$\hat{x} = [\text{proj}_W(\mathbf{x})]_B = (A^*A)^{-1}A^*\mathbf{x}.$$

We can then obtain $\text{proj}_W(\mathbf{x})$ via

$$\text{proj}_W(\mathbf{x}) = A[\text{proj}_W(\mathbf{x})]_B = A\hat{x} = A(A^*A)^{-1}A^*\mathbf{x}.$$

Example 1. If $p = 1$ (so W is a line), then

$$A^*A\hat{x} = A^*\mathbf{x} \implies [\mathbf{w}_1 \cdot \mathbf{w}_1]\hat{x} = \mathbf{w}_1 \cdot \mathbf{x},$$

leading again to

$$\text{proj}_W(\mathbf{x}) = \tfrac{\mathbf{w}_1 \cdot \mathbf{x}}{\|\mathbf{w}_1\|^2}\mathbf{w}_1.$$

Example 2. If $p > 1$ and $B = \{\mathbf{w}_1, \dots, \mathbf{w}_p\}$ is an **orthogonal basis** then

$$A^*A = \text{diag}\{\|\mathbf{w}_1\|^2, \dots, \|\mathbf{w}_p\|^2\}.$$

Recalling that $\hat{x} = (A^*A)^{-1}A^*x$, we find

$$[\text{proj}_W(\mathbf{x})]_B = \begin{bmatrix} \frac{\mathbf{w}_1 \cdot \mathbf{x}}{\|\mathbf{w}_1\|^2} \\ \vdots \\ \frac{\mathbf{w}_p \cdot \mathbf{x}}{\|\mathbf{w}_p\|^2} \end{bmatrix},$$

Thus

$$\text{proj}_W(\mathbf{x}) = \tfrac{\mathbf{w}_1 \cdot \mathbf{x}}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \cdots + \tfrac{\mathbf{w}_p \cdot \mathbf{x}}{\|\mathbf{w}_p\|^2}\mathbf{w}_p.$$

Example. Let

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 4 \\ 5 \\ -3 \\ 3 \end{bmatrix}.$$

Set $B = \{\mathbf{w}_1, \mathbf{w}_2\}$ and $A = [\mathbf{w}_1 \mathbf{w}_2]$. Then $A^*A = \operatorname{diag}\{7, 7\}$. Thus

$$\hat{x} = [\operatorname{proj}_W(\mathbf{x})]_B = (A^*A)^{-1}A^*\mathbf{x} = \begin{bmatrix} 2 \\ \frac{3}{7} \end{bmatrix},$$

$$\operatorname{proj}_W(\mathbf{x}) = 2\mathbf{w}_1 + \tfrac{3}{7}\mathbf{w}_2.$$

The projection of $\mathbf{x}$ onto $W^\perp$ is simply

$$\operatorname{proj}_{W^\perp}(\mathbf{x}) = \mathbf{x} - \operatorname{proj}_W(\mathbf{x}).$$

Example 2. If A^*A is not diagonal, then the columns of A are not an orthogonal basis for $\mathrm{col}(A)$.

One can still compute the projection via

$$\mathrm{proj}_W(\mathbf{x}) = A(A^*A)^{-1}A^*\mathbf{x}.$$

Distance minimization. Orthogonal projection is related to minimizing a distance. To see this, supose $\mathbf{w} \in W$ and $\mathbf{x} \in \mathbb{C}^n$. By the Pythagorean theorem,

$$\|\mathbf{x} - \mathbf{w}\|^2 = \|\text{proj}_{W^\perp}(\mathbf{x})\|^2 + \|\text{proj}_W(\mathbf{x}) - \mathbf{w}\|^2,$$

and thus

$$\min_{\mathbf{w} \in W} \|\mathbf{x} - \mathbf{w}\| = \|\mathbf{x} - \text{proj}_W(\mathbf{x})\| = \|\text{proj}_{W^\perp}(\mathbf{x})\|.$$

Example. Let $W = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\} \subset \mathbb{R}^3$. Then $\|\text{proj}_{W^\perp}(\mathbf{x})\|$ is the distance from $\mathbf{x}$ to the plane spanned by $\mathbf{w}_1$ and $\mathbf{w}_2$.

Conclusion. Let $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ be a basis for $W \subset \mathbb{C}^n$, $A = [\mathbf{w}_1 \cdots \mathbf{w}_p]$, and $\mathbf{x} \in \mathbb{C}^n$.

- $\mathrm{nul}(A^*A) = \mathrm{nul}(A)$, $\mathrm{rank}(A^*A) = \mathrm{rank}(A) = p$, and so A^*A is invertible
- The solution to $A^*A\hat{x} = A^*\mathbf{x}$ is $\hat{x} = (A^*A)^{-1}A^*\mathbf{x}$
- $\hat{x} = (A^*A)^{-1}A^*\mathbf{x} = [\mathrm{proj}_W(\mathbf{x})]_B$
- $\mathrm{proj}_W : \mathbb{C}^n \to W$ is given by $\mathrm{proj}_W(\mathbf{x}) = A(A^*A)^{-1}A^*\mathbf{x}$
- $\mathrm{proj}_{W^\perp}(\mathbf{x}) = \mathbf{x} - \mathrm{proj}_W(\mathbf{x})$
- $\mathbf{x} = \mathrm{proj}_W(\mathbf{x}) + \mathrm{proj}_{W^\perp}(\mathbf{x})$
- if B is orthogonal, $\mathrm{proj}_W(\mathbf{x}) = \frac{\mathbf{w}_1 \cdot \mathbf{x}}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \cdots + \frac{\mathbf{w}_p \cdot \mathbf{x}}{\|\mathbf{w}_p\|^2}\mathbf{w}_p$
- $\min_{\mathbf{w} \in W} \|\mathbf{x} - \mathbf{w}\| = \|\mathbf{x} - \mathrm{proj}_W(\mathbf{x})\|$

Chapter 6. Orthogonality and Least Squares

6.4 The Gram–Schmidt Process

Orthogonal projections. Recall that if $B = \{\mathbf{w}_1, \cdots, \mathbf{w}_p\}$ is an independent set and $A = [\mathbf{w}_1 \cdots \mathbf{w}_p]$, then

$$\text{proj}_W(\mathbf{x}) = A(A^*A)^{-1}A^*\mathbf{x}, \quad W = \text{span}(B)$$

If B is **orthogonal**, then

$$\text{proj}_W(\mathbf{x}) = \tfrac{\mathbf{w}_1 \cdot x}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \cdots + \tfrac{\mathbf{w}_p \cdot x}{\|\mathbf{w}_p\|^2}\mathbf{w}_p.$$

This may be written

$$\text{proj}_W(\mathbf{x}) = \text{proj}_{W_1}(\mathbf{x}) + \cdots + \text{proj}_{W_p}(\mathbf{x}), \quad W_j = \text{span}\{\mathbf{w}_j\}.$$

If B is not orthogonal, then we may apply an algorithm to B to obtain an orthogonal basis for W.

Gram-Schmidt algorithm. Let $A = \{\mathbf{w}_1, \cdots, \mathbf{w}_p\}$.

Let $\mathbf{v}_1 := \mathbf{w}_1$ and $\Omega_1 := \text{span}\{\mathbf{v}_1\}$.

Let $\mathbf{v}_2 = \text{proj}_{\Omega_1^\perp}(\mathbf{w}_2) = \mathbf{w}_2 - \text{proj}_{\Omega_1}(\mathbf{w}_2), \quad \Omega_2 := \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$

$\dots$

Let $\mathbf{v}_{j+1} = \text{proj}_{\Omega_j^\perp}(\mathbf{w}_{j+1}), \quad \Omega_{j+1} := \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_{j+1}\}$

Here $j = 1, \dots, p-1$. This generates a pairwise orthogonal set $B = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ with $\text{span}(B) = \text{span}(A)$. Note that

$$\mathbf{v}_{j+1} = \mathbf{0} \iff \mathbf{w}_{j+1} \in \Omega_j.$$

Matrix representation. Write $V_i = \text{span}\{\mathbf{v}_i\}$. Since $\{\mathbf{v}_i\}$ are orthogonal, we can write

$$\text{proj}_{\Omega_j}(\mathbf{w}_{j+1}) = \sum_{k=1}^{j} \text{proj}_{V_k}(\mathbf{w}_{j+1}) = \sum_{k=1}^{j} r_{k,j+1}\mathbf{v}_k,$$

where $r_{k,j+1} = \frac{\mathbf{v}_k \cdot \mathbf{w}_{j+1}}{\|\mathbf{v}_k\|^2}$ if $\mathbf{v}_k \neq 0$ and
$r_{k,j+1}$ can be anything if $\mathbf{v}_k = 0$.

Thus, using $\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - \sum_{k=1}^{j} r_{k,j+1}\mathbf{v}_k$, we find

$$\mathbf{w}_{j+1} = [\mathbf{v}_1 \cdots \mathbf{v}_{j+1}] \begin{bmatrix} r_{1,j+1} \\ \vdots \\ r_{j,j+1} \\ 1 \end{bmatrix}, \quad j = 1, \ldots, p-1.$$

Matrix representation (continued). The Gram-Schmidt algorithm therefore has the matrix representation

$$[\mathbf{w}_1 \cdots \mathbf{w}_p] = [\mathbf{v}_1 \cdots \mathbf{v}_p] R,$$

where

$$R = \begin{bmatrix} 1 & r_{1,2} & \cdots & r_{1,p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & r_{p-1,p} \\ & & & 1 \end{bmatrix}$$

- This shows that any matrix $A = [\mathbf{w}_1 \cdots \mathbf{w}_p] \in \mathbb{C}^{n \times p}$ may be factored as $A = QR$, where the columns of Q are orthogonal and $R \in \mathbb{C}^{p \times p}$ is an invertible upper triangular matrix.
- The non-zero vectors in $\{\mathbf{v}_1, \ldots, \mathbf{v}_j\}$ form an orthogonal basis for Ω_j.
- R is unique when each $\mathbf{v}_j$ is non-zero.

Example 1. Let

$$\mathbf{w}_1 = [1,0,1,0]^T, \quad \mathbf{w}_2 = [1,1,1,1]^T,$$
$$\mathbf{w}_3 = [1,-1,1,-1]^T, \quad \mathbf{w}_4 = [0,0,1,1]^T.$$

We apply Gram–Schmidt:

- $\mathbf{v}_1 = \mathbf{w}_1, \quad \Omega_1 = \mathrm{span}\{\mathbf{v}_1\}$
- $\mathbf{v}_2 = \mathbf{w}_2 - \frac{\mathbf{v}_1\cdot\mathbf{w}_2}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 = [0,1,0,1]^T, \quad \Omega_2 = \mathrm{span}\{\mathbf{v}_1,\mathbf{v}_2\}$
- $\mathbf{v}_3 = \mathbf{w}_3 - \frac{\mathbf{v}_1\cdot\mathbf{w}_3}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{v}_2\cdot\mathbf{w}_3}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 = 0, \quad \Omega_3 = \mathrm{span}\{\mathbf{v}_1,\mathbf{v}_2\}$
- $\mathbf{v}_4 = \mathbf{w}_4 - \frac{\mathbf{v}_1\cdot\mathbf{w}_4}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{v}_2\cdot\mathbf{w}_4}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 = \frac{1}{2}[-1,-1,1,1]^T$
  $\Omega_4 = \mathrm{span}\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_4\}$.

In particular, $\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_4\}$ is an orthogonal basis for $\mathrm{span}\{\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3,\mathbf{w}_4\}$.

Example 1. (cont.) Let $A = [\mathbf{w}_1 \mathbf{w}_2 \mathbf{w}_3 \mathbf{w}_4]$ and $Q = [\mathbf{v}_1 \mathbf{v}_2 \, \mathbf{0} \, \mathbf{v}_4]$. Then we can write $A = QR$, where

$$R = \begin{bmatrix} 1 & 1 & 1 & \frac{1}{2} \\ 0 & 1 & -1 & \frac{1}{2} \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

for any c. These coefficients are determined by evaluating the inner products above, cf.

$$r_{k,j+1} = \tfrac{\mathbf{v}_k \cdot \mathbf{w}_{j+1}}{\|\mathbf{v}_k\|^2} \quad \text{if} \quad \mathbf{v}_k \neq 0.$$

Example 2. Let

$$\mathbf{w}_1 = [1, 1, 1, 0]^T, \quad \mathbf{w}_2 = [0, 1, -1, 1]^T.$$

Note $\mathbf{w}_1 \perp \mathbf{w}_2$. Extend $\{\mathbf{w}_1, \mathbf{w}_2\}$ to an orthogonal basis for $\mathbb{C}^4$.

Write $A = [\mathbf{w}_1 \mathbf{w}_2]$ and $W = \mathrm{col}(A)$.

First want a basis for $W^\perp = \mathrm{nul}(A^*)$. Since

$$A^* = A^T = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{bmatrix},$$

we get $\mathrm{nul}(A^*) = \mathrm{span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where

$$\mathbf{x}_1 = [-2, 1, 1, 0]^T, \quad \mathbf{x}_2 = [1, -1, 0, 1]^T.$$

Now $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{x}_1, \mathbf{x}_2\}$ is a basis for $\mathbb{C}^4$, but not orthogonal.

Example 2. (Cont.) We now apply Gram–Schmidt to $\{\mathbf{x}_1, \mathbf{x}_2\}$.

- $\mathbf{v}_1 = \mathbf{x}_1 = [-2, 1, 1, 0]^T$
- $\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{x}_2}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 = [0, -\frac{1}{2}, \frac{1}{2}, 1]^T$

Thus $B = \{\mathbf{w}_1, \mathbf{w}_2\}$ is an orthogonal basis for $W = \text{col}(A)$, $C = \{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for $W^\perp$, and $B \cup C$ is an orthogonal basis for $\mathbb{C}^4$.

Chapter 6. Orthogonality and Least Squares

6.5 Least-Squares Problems

The normal system. For $A \in \mathbb{C}^{n \times p}$ and $\mathbf{b} \in \mathbb{C}^n$, the equation

$$A^* A\mathbf{x} = A^* \mathbf{b}$$

is called the **normal system** for $A\mathbf{x} = \mathbf{b}$.

- The normal system arose when computing the orthogonal projection onto a subspace, where the columns of A were assumed to be a basis $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ for $W = \mathrm{col}(A)$.
- The system was $A^* A\hat{x} = A^* \mathbf{x}$, with solution

$$\hat{x} = [\mathrm{proj}_W(\mathbf{x})]_B = (A^* A)^{-1} A^* \mathbf{x}, \quad A\hat{x} = \mathrm{proj}_W(\mathbf{x}).$$

- Invertibility of $A^* A \in \mathbb{C}^{p \times p}$ followed from $\mathrm{rank}(A^* A) = \mathrm{rank}(A) = p$.

In general, we need not assume that $\mathrm{rank}(A) = p$...

Claim. $A^*A\mathbf{x} = A^*\mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{C}^n$.

To see this we first show $\text{col}(A^*A) = \text{col}(A^*)$.

- Indeed, if $\mathbf{y} \in \text{col}(A^*A)$, then we may write $\mathbf{y} = A^*[A\mathbf{x}]$, so that $\mathbf{y} \in \text{col}(A^*)$.
- On the other hand, we have previously shown that $\text{rank}(A^*A) = \text{rank}(A^*)$. Thus $\text{col}(A^*A) = \text{col}(A^*)$.

Since $A^*\mathbf{b} \in \text{col}(A^*)$, the claim follows.

If $\hat{x}$ is a solution to the normal system, then

$$A^*(\mathbf{b} - A\hat{x}) = 0 \implies \mathbf{b} - A\hat{x} \in [\text{col}(A)]^\perp.$$

On the other hand, $A\hat{x} \in \text{col}(A)$, which shows

$$A\hat{x} = \text{proj}_W(\mathbf{b}), \quad \text{where} \quad W := \text{col}(A).$$

I.e. solutions to the normal system give combinations of the columns of A equal to $\text{proj}_W(\mathbf{b})$.

Least squares solutions of $A\mathbf{x} = \mathbf{b}$. We have just seen that $A^*A\mathbf{x} = A^*\mathbf{b}$ is *always* consistent, even if $A\mathbf{x} = \mathbf{b}$ is not!

We saw that the solution set of $A^*A\mathbf{x} = A^*\mathbf{b}$ is equivalent to the solution set of

$$A\mathbf{x} = \mathrm{proj}_W(\mathbf{b}), \quad \text{where} \quad W = \mathrm{col}(A). \quad (*)$$

Indeed, we just saw that any solution to the normal system satisfies $(*)$, while applying A^* to $(*)$ gives

$$A^*A\mathbf{x} = A^*\mathrm{proj}_W(\mathbf{b}) = A^*\mathbf{b} \quad (\text{cf.} \quad \mathrm{col}(A)^\perp = \mathrm{nul}A^*)$$

Note that $A\mathbf{x} = \mathbf{b}$ is consistent precisely when $\mathbf{b} \in \mathrm{col}(A)$, i.e. when $\mathbf{b} = \mathrm{proj}_W(\mathbf{b})$.

Thus the normal system is equivalent to the original system precisely when the original system is consistent.

Least squares solutions of $A\mathbf{x} = \mathbf{b}$ There is a clear geometric interpretation of the solution set to the normal system: let $\hat{x}$ be a solution to the normal system $A^*A\mathbf{x} = A^*\mathbf{b}$. Then, with $W = \text{col}(A)$,

$$\|\mathbf{b} - A\hat{x}\| = \|\mathbf{b} - \text{proj}_W(\mathbf{b})\| = \min_{w \in W} \|\mathbf{b} - \mathbf{w}\| = \min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{b} - A\mathbf{x}\|.$$

Thus $\hat{x}$ minimizes $\|\mathbf{b} - A\mathbf{x}\|$ over all $\mathbf{x} \in \mathbb{C}^n$.

The solution to the normal system for $A\mathbf{x} = \mathbf{b}$ is called the **least squares solution** of $A\mathbf{x} = \mathbf{b}$.

Example. Let

$$A = [\mathbf{w}_1 \mathbf{w}_2 \mathbf{w}_3] = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

The system $A\mathbf{x} = \mathbf{b}$ is inconsistent, since

$$[A|\mathbf{b}] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

The normal system $A^*A\mathbf{x} = A^*b$ is consistent, since

$$[A^*A|A^*\mathbf{b}] \sim \left[\begin{array}{ccc|r} 1 & 0 & 1 & 1/3 \\ 0 & 1 & 1 & -1/3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The least squares solutions to $A\mathbf{x} = \mathbf{b}$ are therefore

$$\hat{x} = \begin{bmatrix} 1/3 \\ -1/3 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \quad z \in \mathbb{C}.$$

Example. (cont.) We can also compute that

$$\text{proj}_W \mathbf{b} = A\hat{x} = \tfrac{1}{3}\mathbf{w}_1 - \tfrac{1}{3}\mathbf{w}_2 + \mathbf{0} = \tfrac{1}{3}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix},$$

where $W = \text{col}(A)$.

The **least squares error** for $A\mathbf{x} = \mathbf{b}$ is defined by

$$\min_{x \in \mathbb{C}^n} \|\mathbf{b} - A\mathbf{x}\| = \|\mathbf{b} - A\hat{x}\|.$$

In this case, one can check that $\|\mathbf{b} - A\hat{x}\| = \frac{2}{3}\sqrt{3}$.

- This is a measurement of the smallest error possible when approximating $\mathbf{b}$ by a vector in $\text{col}(A)$.

Chapter 6. Orthogonality and Least Squares

6.6 Applications to Linear Models

Linear models. Suppose you have a collection of data from an experiment, given by

$$\{(x_j, y_j) : j = 1, \ldots, n\}.$$

You believe there is an underlying relationship describing this data of the form

$$\beta_1 f(x) + \beta_2 g(x) + \beta_3 = h(y),$$

where f, g, h are known but $\beta = (\beta_1, \beta_2, \beta_3)^T$ is not.

Assuming a relation of this form and accounting for experimental error, we have

$$\beta_1 f(x_j) + \beta_2 g(x_j) + \beta_3 = h(y_j) + \varepsilon_j$$

for $j = 1, \ldots, n$ and some small $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$.

Linear models. (Cont.) In matrix form, we have

$$X\beta - \mathbf{y} = \varepsilon, \quad X = \begin{bmatrix} f(x_1) & g(x_1) & 1 \\ \vdots & \vdots & \vdots \\ f(x_n) & g(x_n) & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} h(y_1) \\ \vdots \\ h(y_n) \end{bmatrix}.$$

Terminology:

- X is the design matrix,
- β is the parameter vector,
- $\mathbf{y}$ is the observation vector,
- ε is the residual vector.

The goal is to find β to minimize $\|X\beta - \mathbf{y}\|^2$.

To this end, we solve the normal system $X^*X\beta = X^*\mathbf{y}$. This solution gives the **least squares best fit**.

Example 1. (Fitting to a quadratic polynomial). Find a least squares best fit to the data

$$(-1,0), \quad (0,1), \quad (1,2), \quad (2,4)$$

for the model given by $y = \beta_1 x^2 + \beta_2 x + \beta_3$. The associated linear model is

$$X\beta = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \beta = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix} + \varepsilon = \mathbf{y} + \varepsilon.$$

Example 1. (cont.) The normal system $X^*X\beta = X^*\mathbf{y}$ has solution $\hat{\beta} = [.25, 1.05, .85]^T$, which implies the least squares best fit to the data is

$$y = .25x^2 + 1.05x + .85.$$

The least squares error is $\|X\hat{\beta} - \mathbf{y}\| = .0224$.

Example 2. Kepler's first law asserts that the orbit of a comet (parametrized by (r, θ)) is described by $r = \beta + e(r\cos\theta)$, where β, e are to be determined.

The orbit is elliptical when $0 < e < 1$, parabolic when $e = 1$, and hyperbolic when $e > 1$.
Given observational data

$$(\theta, r) = \{(.88, 3), (1.1, 2.3), (1.42, 1.65), (1.77, 1.25), (2.14, 1.01)\},$$

what is the nature of the orbit?

Example 2. (cont.) The associated linear model is

$$\begin{bmatrix} 1 & r_1 \cos\theta_1 \\ \vdots & \vdots \\ 1 & r_5 \cos\theta_5 \end{bmatrix} \begin{bmatrix} \beta \\ e \end{bmatrix} = \begin{bmatrix} r_1 \\ \vdots \\ r_5 \end{bmatrix} + \varepsilon.$$

We can rewrite this as $X\beta = \mathbf{y} + \varepsilon$. The solution to the normal system $X^*X\beta = X^*\mathbf{y}$ is given by $[\hat{\beta}, \hat{e}] = [1.45, .81]$.

We conclude that the orbit is most likely elliptical.

Chapter 7. Symmetric Matrices and Quadratic Forms

7.1 Diagonalization of Symmetric Matrices

Schur Triangular Form.

Definition. A matrix $P \in \mathbb{C}^{n\times n}$ is **unitary** if $P^*P = I_n$.

Schur Factorization. Any $A \in \mathbb{C}^{n\times n}$ can be written in the form $A = PUP^*$ where $P \in \mathbb{C}^{n\times n}$ is unitary and $U \in \mathbb{C}^{n\times n}$ is upper triangular.

This can be proven by induction. The case $n = 1$ is clear.

Now suppose the result holds for $(n-1)\times(n-1)$ matrices and let $A \in \mathbb{C}^{n\times n}$.

Let $\{\lambda_1, \mathbf{v}_1\}$ be an eigenvalue/eigenvector pair for A with $\|\mathbf{v}_1\| = 1$.

Extend $\mathbf{v}_1$ to an orthonormal basis $\{\mathbf{v}_1 \ldots, \mathbf{v}_n\}$ for $\mathbb{C}^n$ and set $P_1 = [\mathbf{v}_1, \cdots \mathbf{v}_n]$.

Schur Factorization. (cont.) Note $P_1^* = P_1^{-1}$. We may write

$$AP_1 = P_1 \begin{bmatrix} \lambda_1 & \mathbf{w} \\ \mathbf{0} & M \end{bmatrix}, \quad M \in \mathbb{C}^{(n-1)\times(n-1)}, \quad \mathbf{w} \in \mathbb{C}^{1\times(n-1)}.$$

By assumption, we can write $M = QU_0Q^*$, Q unitary and U is upper triangular.

Now set

$$P_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix}, \quad P = P_1P_2.$$

Then P is unitary (check!) and

$$P^*AP = P_2^* \begin{bmatrix} \lambda_1 & \mathbf{w} \\ \mathbf{0} & M \end{bmatrix} P_2 = \begin{bmatrix} \lambda & \mathbf{w}Q \\ \mathbf{0} & U_0 \end{bmatrix},$$

which completes the proof.

Schur Triangular Form.

This result shows that every $A \in \mathbb{C}^{n \times n}$ is similar to an upper triangular matrix $U \in \mathbb{C}^{n \times n}$ via a change of coordinate matrix $P \in \mathbb{C}^{n \times n}$ that is **unitary**.

That is: every matrix A is unitarily similar to an upper triangular matrix.

Definition. (Normal matrices) A matrix $A \in \mathbb{C}^{n \times n}$ is **normal** if

$$A^* A = AA^*.$$

Examples of normal matrices.

- If $A^* = A$ (i.e. A is **hermitian**), then A is normal.
- If $A \in \mathbb{R}^{n \times n}$ is symmetric ($A = A^T$), then A is normal.
- If $A^* = -A$ (**skew-adjoint**), then A is normal.
- If A is unitary ($A^* A = I_n$), then A is normal.

Theorem. If $A \in \mathbb{C}^{n \times n}$ is normal and $(\lambda, \mathbf{v})$ is an eigenvalue/eigenvector pair, then $\{\bar{\lambda}, \mathbf{v}\}$ is an eigenvalue/eigenvector pair for A^*.

Indeed,

$$\begin{aligned}
\|(A - \lambda I)\mathbf{v}\|^2 &= [(A - \lambda I)\mathbf{v}]^*(A - \lambda I)\mathbf{v} \\
&= \mathbf{v}^*(A^* - \bar{\lambda} I)(A - \lambda I)\mathbf{v} \\
&= \mathbf{v}^*(A - \lambda I)(A^* - \bar{\lambda} I)\mathbf{v} \\
&= \|(A^* - \bar{\lambda} I)\mathbf{v}\|^2.
\end{aligned}$$

Theorem. (Spectral theorem for normal matrices)

• A matrix $A \in \mathbb{C}^{n \times n}$ is normal if and only if it is unitarily similar to a diagonal matrix. That is, A is normal if and only if

$$A = PDP^*$$

for some diagonal $D \in \mathbb{C}^{n \times n}$ and unitary $P \in \mathbb{C}^{n \times n}$. •

One direction is easy: if $A = PDP^*$ for P unitary and D diagonal, then

$$A^*A = AA^*. \quad \text{(Check!)}$$

Therefore we focus on the reverse direction.

Spectral Theorem. (cont.)

Now suppose $A \in \mathbb{C}^{n\times n}$ is normal, i.e. $AA^* = A^*A$. We begin by writing the Schur factorization of A, i.e.

$$A = PUP^*, \quad P = [\mathbf{v}_1 \cdots \mathbf{v}_n],$$

where P is unitary and $U = [c_{ij}]$ is upper triangular.

First note that $AP = PU$ implies $A\mathbf{v}_1 = c_{11}\mathbf{v}_1$, and hence (since A is normal) $A^*\mathbf{v}_1 = \bar{c}_{11}\mathbf{v}_1$.

However, $A^*P = PU^*$, so that

$$\bar{c}_{11}\mathbf{v}_1 = A^*\mathbf{v}_1 = \bar{c}_{11}\mathbf{v}_1 + \cdots + \bar{c}_{1n}\mathbf{v}_n$$

By independence of $\mathbf{v}_2, \ldots, \mathbf{v}_n$, we deduce $c_{1j} = 0$ for $j = 2, \ldots, n$.

Spectral Theorem. (cont.)

We have shown

$$U = \begin{bmatrix} c_{11} & \mathbf{0} \\ \mathbf{0} & \tilde{U} \end{bmatrix},$$

where $\tilde{U} \in \mathbb{C}^{(n-1)\times(n-1)}$ is upper triangular.

But now $AP = PU$ gives $A\mathbf{v}_2 = c_{22}\mathbf{v}_2$, and arguing as above we deduce $c_{2j} = 0$ for $j = 3, \ldots, n$.

Continuing in this way, we deduce that U is diagonal. $\square$

Spectral Theorem. (cont.)

To summarize, $A \in \mathbb{C}^{n \times n}$ is normal ($AA^* = A^*A$) if and only if it can be written as $A = PDP^*$ where $P = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ is unitary and $D = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$. Note

- P unitary means $P^{-1} = P^*$
- A is unitarily similar to a diagonal matrix
- $\{\lambda_j, \mathbf{v}_j\}$ are eigenvalue-eigenvector pairs for A

Theorem. (Spectral Theorem for Self-Adjoint Matrices)

• A matrix $A \in \mathbb{C}^{n\times n}$ is self-adjoint ($A = A^*$) if and only if it is unitarily similar to a real diagonal matrix, i.e. $A = PDP^*$ for some unitary $P \in \mathbb{C}^{n\times n}$ and some diagonal $D \in \mathbb{R}^{n\times n}$. •

Indeed, this follows from the spectral theorem for normal matrices. In particular,

$$PDP^* = A = A^* = PD^*P \implies D = D^*,$$

which implies that $D \in \mathbb{R}^{n\times n}$.

Note this implies that self-adjoint matrices have *real* eigenvalues.

Eigenvectors and eigenvalues for normal matrices. Suppose A is a normal matrix.

- Eigenvectors associated to different eigenvalues are orthogonal:

$$\mathbf{v}_1 \cdot A\mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2,$$
$$\mathbf{v}_1 \cdot A\mathbf{v}_2 = A^*\mathbf{v}_1 \cdot \mathbf{v}_2 = \lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2.$$

- If the eigenvalues are all real, then A is self-adjoint. (This follows from the spectral theorem.)

Spectral decomposition. If $A \in \mathbb{C}^{n\times n}$ is a normal matrix, then we may write $A = PDP^*$ as above. In particular,

$$A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^* + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^*$$

Recall that

$$\tfrac{1}{\|\mathbf{v}_k\|^2} \mathbf{v}_k \mathbf{v}_k^* = \mathbf{v}_k \mathbf{v}_k^*$$

is the projection matrix for the subspace $V_k = \text{span}\{\mathbf{v}_k\}$.

Thus, a normal matrix can be written as the sum of scalar multiples of projections on to the eigenspaces.

Chapter 7. Symmetric Matrices and Quadratic Forms

7.2 Quadratic Forms

Definition. Let $A \in \mathbb{C}^{n\times n}$ be a self-adjoint matrix. The function

$$Q(\mathbf{x}) = \mathbf{x}^* A\mathbf{x}, \quad \mathbf{x} \in \mathbb{C}^n$$

is called a **quadratic form**. Using self-adjointness of A, one finds

$$Q : \mathbb{C}^n \to \mathbb{R}.$$

- If $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$, we call Q **positive definite.**
- If $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq 0$, we call Q **positive semidefinite.**
- We define negative definite, negative semidefinite similarly.
- We call Q **indefinite** if it attains both positive and negative values.

Characteristic forms. Expanding the inner product, we find that

$$\mathbf{x}^* A\mathbf{x} = \sum_{j=1}^{n} a_{jj}|x_j|^2 + 2\sum_{i<j} \mathrm{Re}(a_{ij}x_i x_j).$$

For $A \in \mathbb{R}^{n\times n}$ and $\mathbf{x} \in \mathbb{R}^n$, this reduces to

$$\mathbf{x}^T A\mathbf{x} = \sum_{j=1}^{n} a_{jj}x_j^2 + 2\sum_{i<j} a_{ij}x_i x_j.$$

Example.

$$\mathbf{x}^T \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & -5 \\ 3 & -5 & -6 \end{bmatrix} \mathbf{x} = x_1^2 + 4x_2^2 - 6x_3^2 - 4x_1x_2 + 6x_1x_3 - 10x_2x_3.$$

Characterization of definiteness. Let $A \in \mathbb{C}^{n\times n}$, $Q(\mathbf{x}) = \mathbf{x}^* A\mathbf{x}$.

- There exists an orthonormal basis $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ s.t. $A = PDP^*$, where $P = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ and $D = \operatorname{diag}(\lambda_1, \cdots, \lambda_n) \in \mathbb{R}^{n\times n}$.

Then, with $\mathbf{y} = P^{-1}\mathbf{x}$

$$Q(\mathbf{x}) = \mathbf{x}^* PDP^*\mathbf{x} = (P^{-1}\mathbf{x})^* DP^{-1}\mathbf{x} = y^* Dy$$
$$= \lambda_1|y_1|^2 + \cdots + \lambda_n|y_n|^2.$$

We conclude:

Theorem. If $A \in \mathbb{C}^{n\times n}$ is self-adjoint, then $Q(\mathbf{x}) = \mathbf{x}^* A\mathbf{x}$ is positive definite if and only if the eigenvalues of A are all positive.

(Similarly for negative definite, or semidefinite...)

Quadratic forms and conic sections. The equation

$$ax_1^2 + 2bx_1x_2 + cx_2^2 + dx_1 + ex_2 = f$$

can be written as

$$\mathbf{x}^T A\mathbf{x} + [d\ e]\mathbf{x} = f, \quad A = A^T = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

By the spectral theorem, there is a basis of eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ that diagonalizes A. That is,

$$A = PDP^T, \quad P = [\mathbf{v}_1\ \mathbf{v}_2], \quad D = \mathrm{diag}(\lambda_1, \lambda_2).$$

Writing $\mathbf{y} = P^T\mathbf{x}$, the equation becomes

$$\mathbf{y}^T D\mathbf{y} + [d'\ e']\mathbf{y} = f, \quad [d'\ e'] = [d\ e]P,$$

i.e. $\lambda_1 y_1^2 + \lambda_2 y_2^2 + d'y_1 + e'y_2 = f$.

Principle axis theorem. The change of variables $\mathbf{y} = P^T\mathbf{x}$ gives

$$\mathbf{x}^T A\mathbf{x} + [d\ e]\mathbf{x} = f \iff \mathbf{y}^T D\mathbf{y} + [d'\ e']\mathbf{y} = f.$$

The nature of the conic section can be understood through the quadratic form $\mathbf{y}^T D\mathbf{y}$.

Note that this transforms $\mathbf{x}^* A\mathbf{x}$ into a quadratic form $\mathbf{y}^* D\mathbf{y}$ with no cross-product term.

Example. Consider $x_1^2 - 6x_1x_2 + 9x_2^2$. This corresponds to

$$A = \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}.$$

The eigenvalues are $\lambda = 10, 0$ (the quadratic form is positive semidefinite), with eigenspaces

$$E_0 = \operatorname{span}([3, 1]^T), \quad E_{10} = \operatorname{span}([1, -3]^T).$$

Consequently $A = PDP^T$, with

$$P = \tfrac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}.$$

Writing $\mathbf{y} = P^T\mathbf{x}$ leads to the quadratic form

$$10y_1^2 + 0y_2^2 = 10y_1^2.$$

Example. (cont.) Consider the conic section described by

$$x_1^2 - 6x_1x_2 + 9x_2^2 + 3x_1 + x_2 = 1.$$

This can be written $\mathbf{x}^T A\mathbf{x} + [3\ 1]\mathbf{x} = 1$. Continuing from above, this is equivalent to

$$10y_1^2 + [3\,1]P\mathbf{y} = 10y_1^2 + \sqrt{10}y_2 = 1,$$

i.e. $y_2 = \frac{\sqrt{10}}{10} - \sqrt{10}y_1^2$.

In the y_1y_2 plane, the conic section is a parabola. To go from $\mathbf{x}$ coordinates to $\mathbf{y}$ coordinates, we apply P, which is a rotation.

Chapter 7. Symmetric Matrices and Quadratic Forms

7.3 Constrained Optimization

Recall: A self-adjoint matrix $A \in \mathbb{C}^{n \times n}$ is unitarily similar to a real diagonal matrix. Consequently, we can write

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^* + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^*,$$

where $\lambda_n \leq \cdots \leq \lambda_1 \in \mathbb{R}$ and $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal basis.

Quadratic forms and boundedness. Let A be self-adjoint. Continuing from above,

$$\begin{aligned}\mathbf{x}^*A\mathbf{x} &= \lambda_1\mathbf{x}^*\mathbf{u}_1(\mathbf{u}_1^*\mathbf{x}) + \cdots + \lambda_n\mathbf{x}^*\mathbf{u}_n(\mathbf{u}_n^*\mathbf{x}) \\ &= \lambda_1|\mathbf{u}_1^*\mathbf{x}|^2 + \cdots + \lambda_n|\mathbf{u}_n^*\mathbf{x}|^2.\end{aligned}$$

Since $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal basis,

$$\mathbf{x} = (\mathbf{u}_1^*\mathbf{x})\mathbf{u}_1 + \cdots + (\mathbf{u}_n^*\mathbf{x})\mathbf{u}_n \implies \|\mathbf{x}\|^2 = |\mathbf{u}_1^*\mathbf{x}|^2 + \cdots + |\mathbf{u}_n^*\mathbf{x}|^2.$$

We deduce

$$\lambda_n\|x\|^2 \le \mathbf{x}^*A\mathbf{x} \le \lambda_1\|\mathbf{x}\|^2.$$

Rayleigh principle. We continue with A as above and set

$$\Omega_0 = \{\mathbf{0}\}, \quad \Omega_k := \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}.$$

Then for $\mathbf{x} \in \Omega_{k-1}^{\perp}$ we have

$$\begin{aligned} \|x\|^2 &= |\mathbf{u}_k^* \mathbf{x}|^2 + \cdots + |\mathbf{u}_n^* \mathbf{x}|^2, \\ \mathbf{x}^* A \mathbf{x} &= \lambda_k |\mathbf{u}_k^* \mathbf{x}|^2 + \cdots + \lambda_n |\mathbf{u}_n^* \mathbf{x}|^2. \end{aligned}$$

Thus (using $\lambda_n \leq \cdots \leq \lambda_1$) $\quad \lambda_n \|x\|^2 \leq \mathbf{x}^* A \mathbf{x} \leq \lambda_k \|x\|^2$.

$$\implies \lambda_n \leq \mathbf{x}^* A \mathbf{x} \leq \lambda_k \quad \text{for all} \quad \mathbf{x} \in \Omega_{k-1}^{\perp} \quad \text{with} \quad \|x\| = 1.$$

But since $\mathbf{u}_n^* A \mathbf{u}_n = \lambda_n$ and $\mathbf{u}_k^* A \mathbf{u}_k = \lambda_k$, we deduce the **Rayleigh principle:** for $k = 1, \ldots, n$,

$$\begin{aligned} \min_{\|\mathbf{x}\|=1} \mathbf{x}^* A \mathbf{x} = \min_{\|\mathbf{x}\|=1,\ \mathbf{x} \in \Omega_{k-1}^{\perp}} \mathbf{x}^* A \mathbf{x} &= \lambda_n, \\ \max_{\|\mathbf{x}\|=1,\ \mathbf{x} \in \Omega_{k-1}^{\perp}} \mathbf{x}^* A \mathbf{x} &= \lambda_k. \end{aligned}$$

Example. Let $Q(x_1, x_2) = 3x_1^2 + 9x_2^2 + 8x_1x_2$, which corresponds to

$$A = \begin{bmatrix} 3 & 4 \\ 4 & 9 \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = 11$ and $\lambda_2 = 2$, with

$$\Omega_1 = \mathrm{nul}(A - 11I_2) = \mathrm{span}\{[1, 2]^T\},$$
$$\Omega_1^\perp = \mathrm{nul}(A - I_2) = \mathrm{span}\{[-2, 1]^T\}.$$

Note

$$\min_{\|\mathbf{x}\|=1} \mathbf{x}^* A\mathbf{x} = \lambda_2 = 1, \quad \max_{\|\mathbf{x}\|=1} \mathbf{x}^* A\mathbf{x} = \lambda_1 = 11,$$

By the Rayleigh principle, the minimum is obtained on $\Omega_1^\perp$, while the maximum restricted to this set is also equal to $\lambda_2 = 1$.

Example. (cont.)

The contour curves $Q(x_1, x_2) = \text{const}$ are ellipses in the x_1x_2 plane.

Using the change of variables $\mathbf{y} = P^*\mathbf{x}$, where

$$P = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$$

is a rotation by $\theta \sim 63.44^\circ$, one finds $Q(\mathbf{x}) = 11y_1^2 + y_2^2$.

Thus the contour curves $Q(x_1, x_2) = \text{const}$ are obtained by rotating the contour curves of $11x_1^2 + x_2^2 = \text{const}$ by θ.

Chapter 7. Symmetric Matrices and Quadratic Forms

7.4 The Singular Value Decomposition

Singular values. For a matrix $A \in \mathbb{C}^{n \times p}$, the matrix $A^*A \in \mathbb{C}^{p \times p}$ is self-adjoint. By the spectral theorem, there exists an orthonormal basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ for $\mathbb{C}^p$ consisting of eigenvectors for A^*A with real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$.

Noting that $\mathbf{x}^*(A^*A)\mathbf{x} = (A\mathbf{x})^*A\mathbf{x} = \|A\mathbf{x}\|^2 \geq 0$ for all $\mathbf{x}$, we deduce

$$\lambda_j = \lambda_j \|\mathbf{v}_j\|^2 = \mathbf{v}_j^*(A^*A)\mathbf{v}_j \geq 0 \quad \text{for all} \quad j.$$

Definition. With the notation above, we call $\sigma_j := \sqrt{\lambda_j}$ the **singular values** of A.

- If $\operatorname{rank} A = r$, then $\sigma_{r+1} = \cdots = \sigma_p = 0$.
- In this case $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for $\operatorname{col}(A^*)$, while $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_p\}$ is an orthonormal basis for $\operatorname{nul}(A)$.

Singular Value Decomposition. Let $A \in \mathbb{C}^{n\times p}$ with rank$A = r$ as above. The vectors

$$\mathbf{u}_j = \frac{1}{\sigma_j} A\mathbf{v}_j, \quad j = 1, \dots, r$$

form an orthonormal basis for col(A). Indeed,

$$\mathbf{u}_i \cdot \mathbf{u}_j = \frac{\mathbf{v}_i^*(A^*A\mathbf{v}_j)}{\sigma_i\sigma_j} = \frac{\lambda_j}{\sigma_i\sigma_j}\mathbf{v}_i^*\mathbf{v}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases}$$

Next let $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_n\}$ be an orthonormal basis for col$(A)^\perp$. Defining the unitary matrices $V = [\mathbf{v}_1 \cdots \mathbf{v}_p]$ and $U = [\mathbf{u}_1 \cdots \mathbf{u}_n]$,

$$AV = U\Sigma, \quad \Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{n\times p}, \quad D = \mathrm{diag}\{\sigma_1, \cdots, \sigma_r\}.$$

We call $A = U\Sigma V^*$ the **singular value decomposition** of $A \in \mathbb{C}^{n\times p}$.

SVD and linear transformations. Let $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation $T : \mathbb{C}^p \to \mathbb{C}^n$.

Writing $A = U\Sigma V^*$ as above, we have $B = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ and $C = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ are orthonormal bases for $\mathbb{C}^p$ and $\mathbb{C}^n$. Then

$$U^*(A\mathbf{x}) = \Sigma(V^*\mathbf{x}) \implies [T(\mathbf{x})]_C = \Sigma[\mathbf{x}]_B,$$

i.e. there are orthonormal bases for $\mathbb{C}^p$ and $\mathbb{C}^n$ s.t. T can be represented in terms of the matrix Σ.

Transformations of $\mathbb{R}^2$. If $T : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $T(\mathbf{x}) = A\mathbf{x}$, then there exist unitary matrices U, V so that $A = UDV^T$ for $D = \mathrm{diag}(\sigma_1, \sigma_2)$.

Unitary matrices in $\mathbb{R}^{2\times 2}$ represent rotations/reflections of the plane.

Every linear transformation of the plane is the composition of three transformations: a rotation/reflection, a scaling transformation, and a rotation/refection.

Moore–Penrose inverse of $A \in \mathbb{C}^{n\times p}$. Write $V_r = [\mathbf{v}_1 \cdots \mathbf{v}_r] \in \mathbb{C}^{p\times r}$ and $U_r = [\mathbf{u}_1 \cdots \mathbf{u}_r] \in \mathbb{C}^{n\times r}$. Then

$$A = U\Sigma V^* = U_r D V_r^*$$

represents a reduced SVD for A.

Definition. The **Moore–Penrose pseudo inverse** of $A \in \mathbb{C}^{n\times p}$ is defined by

$$A^+ = V_r D^{-1} U_r^* \in \mathbb{C}^{p\times n}.$$

- $AA^+ = U_r U_r^* = \mathrm{proj}_{\mathrm{col}(A)} \in \mathbb{C}^{n\times n}$
- $A^+A = V_r V_r^* = \mathrm{proj}_{\mathrm{col}(A^*)} \in \mathbb{C}^{p\times p}$
- $AA^+A = A, \quad A^+AA^+ = A^+,$
- $A^+ = A^{-1}$ whenever $r = p = n$.

Least squares solutions for $A \in \mathbb{C}^{n \times p}$. Recall that the least squares solutions of $A\mathbf{x} = \mathbf{b}$ are the solutions to the normal system $A^*A\mathbf{x} = A^*\mathbf{b}$. Equivalently, they are solutions to $A\mathbf{x} = \mathrm{proj}_{\mathrm{col}A}\mathbf{b}$.

When $\mathrm{rank}A^*A = r < p$, there are infinitely many least squares solutions.

Note that since $AA^+ = \mathrm{proj}_{\mathrm{col}A}$, we have

$$AA^+\mathbf{b} = \mathrm{proj}_{\mathrm{col}A}(\mathbf{b}) \implies A^+\mathbf{b} \quad \text{is a least squares solution.}$$

On the other hand, using $A^+\mathbf{b} \in \mathrm{col}(A^*)$, we have for any other least squares solution $\hat{x}$,

$$A\hat{x} - AA^+\mathbf{b} = 0 \implies \hat{x} - A^+\mathbf{b} \in \mathrm{nul}(A) = \mathrm{col}(A^*)^\perp,$$

so $A^+\mathbf{b} \perp \hat{x} - A^+\mathbf{b}$. Consequently,

$$\|\hat{x}\|^2 = \|A^+\mathbf{b}\|^2 + \|\hat{x} - A^+\mathbf{b}\|^2.$$

Thus $A^+\mathbf{b}$ is the least squares solution of smallest length.

Four fundamental subspaces. Let $A \in \mathbb{C}^{n \times p}$. Consider

- $\text{col} A, \quad \text{col} A^{\perp} = \text{nul} A^*$
- $\text{col} A^* = \text{row}(\bar{A}), \quad \text{col}(A^*)^{\perp} = \text{nul} A$

Recall the SVD of $A \in \mathbb{C}^{n \times p}$ with $\text{rank}(A) = r$ yields an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ consisting of eigenvectors of A^*A, and an orthonormal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ obtained by completing

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}, \quad \text{where} \quad \mathbf{u}_j = \frac{1}{\sigma_j} A\mathbf{v}_j.$$

Since $A^*A\mathbf{v}_j = \lambda_j \mathbf{v}_j$, $A\mathbf{v}_j = \sigma_j \mathbf{u}_j$:

- $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for $\text{col}(A^*A) = \text{col}(A^*) = \text{row}(\bar{A})$
- $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_p\}$ is an orthonormal basis for $\text{col}(A^*)^{\perp} = \text{nul}(A)$
- $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is an orthonormal basis for $\text{col}(A)$
- $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for $\text{col}(A)^{\perp} = \text{nul}(A^*)$

Review: Matrix Factorizations Let $A \in \mathbb{C}^{n \times p}$.

- Permuted LU factorization: $PA = LU$, where $P \in \mathbb{C}^{n \times n}$ is an invertible permutation matrix, $L \in \mathbb{C}^{n \times n}$ is invertible and lower triangular, and $U \in C^{n \times p}$ is upper triangular.
- QR factorization: $A = QR$, where the columns of $Q \in \mathbb{C}^{n \times p}$ are generated from the columns of A by Gram-Schmidt and $R \in \mathbb{C}^{p \times p}$ is upper triangular.
- SVD: $A = U\Sigma V^*$, where $U \in \mathbb{C}^{n \times n}$, $V \in \mathbb{C}^{p \times p}$ are unitary,

$$D = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{n \times p}, \quad D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r).$$

For $A \in \mathbb{C}^{n \times n}$:

- Schur factorization: $A = PUP^*$ where P is unitary and U is upper triangular.
- Spectral theorems: $A = PDP^*$, where P is unitary and D is diagonal. This holds if and only if A is normal. The matrix D is real if and only if A is self-adjoint.