MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 1 Notes

References

1. Elementary Linear Algebra: Application Version, Section 1.1-1.2
2. Linear Algebra with Application, Section 1.1-1.3

1 Linear Systems

1.1 Introduction to Linear Systems

An equation is linear if the variables (unknowns) are only acted upon by multiplying by constants and adding them up. A linear equation in n variables has the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

Here $a_1, a_2, ..., a_n$ are constants (fixed real numbers), called the coefficients, b is called the constant, and $x_1, x_2, ..., x_n$ are the variables.

A system of linear equations, or a linear system consists of a finite number of linear equations. In general, a linear system with n variables and m equations has the form

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

A linear system can be expressed uniquely as an augmented matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right).$$

The linear system is homogeneous if there are no constant terms

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

Then the corresponding augemented matrix is

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & 0 \end{array}\right).$$

Given a linear system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

the homogeneous system associated to it is

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

Example. (Nonhomogeneous) Linear system:

$$\begin{array}{ccccccc} 3x & + & 2y & - & z & = & 1 \\ & & 5y & + & z & = & 3 \\ x & & & + & z & = & 2 \end{array}$$

The corresponding augmented matrix is

$$\left(\begin{array}{ccc|c} 3 & 2 & -1 & 1 \\ 0 & 5 & 2 & 3 \\ 1 & 0 & 1 & 2 \end{array}\right)$$

The associated homogeneous system is

$$\begin{array}{ccccccc} 3x & + & 2y & - & z & = & 0 \\ & & 5y & + & z & = & 0 \\ x & & & + & z & = & 0 \end{array}$$

with augmented matrix

$$\left(\begin{array}{ccc|c} 3 & 2 & -1 & 0 \\ 0 & 5 & 2 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right)$$

1.2 Solutions to a Linear System

Given a linear system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

we say that

$$x_1 = c_1,\ x_2 = c_2,\ \ldots,\ x_n = c_n$$

is a <u>solution</u> to the linear system if the equations are simultaneously satisfied after making the substitution, that is,

$$\begin{array}{ccccccccc} a_{11}c_1 & + & a_{12}c_2 & + & \cdots & + & a_{1n}c_n & = & b_1 \\ a_{21}c_1 & + & a_{22}c_2 & + & \cdots & + & a_{2n}c_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}c_1 & + & a_{m2}c_2 & + & \cdots & + & a_{mn}c_n & = & b_m \end{array}$$

Example. $x = 1 = y$ is a solution to

$$\begin{array}{rcrcr} 3x & - & 2y & = & 1 \\ x & + & y & = & 2 \end{array}$$

Note that solutions may not be unique.

Example.

$$\begin{array}{rcrcr} x & + & 2y & = & 5 \\ 2x & + & 4y & = & 10 \end{array},$$

solutions: $x = 1,\ y = 2$, or $x = 3,\ y = 1$, etc.

If the solution is not unique, we need to introduce parameters, usually denoted by r, s, t, or $s_1, s_2, ..., s_k$. This means that any choice of a real number for each of the parameter is a solution to the linear system. A general solution to a linear system captures all possible solutions to the linear system.

Example.

$$\begin{array}{rcrcr} x & + & 2y & = & 5 \\ 2x & + & 4y & = & 10 \end{array},$$

General solutions: $x = 5 - 2s,\ y = s$, or $x = s,\ y = \frac{1}{2}(5 - s)$, etc.

We are able to obtain solutions of a linear system by reducing the augmented matrix to some special form known as row-echelon form or reduced row-echelon form.

In the augmented matrix, a zero row is a row with all entries 0. A leading entry of a row is the first nonzero entry of the row counting from the left.

An augmented matrix is in row-echelon form (REF) if

1. All zero rows are at the bottom of the matrix.
2. The leading entries are further to the right as we move down the rows.

An augmented matrix in REF has the form

$$\left(\begin{array}{ccccccccc|c} * & & & & & & & & & * \\ 0 & \cdots & 0 & * & & & & & & * \\ 0 & \cdots & 0 & 0 & \cdots & 0 & * & & & * \\ 0 & & & & & & & & 0 & 0 \\ \vdots & & & & & & & & \vdots & \vdots \\ 0 & \cdots & & & & & & \cdots & 0 & 0 \end{array}\right).$$

In the REF, a pivot column is a column containing a leading entry. The augmented matrix is in reduced row-echelon form (RREF) if further

3. The leading entries are 1.
4. In each pivot column, all entries except the leading entry is 0.

A matrix in RREF has the form

$$\left(\begin{array}{ccccccccc|c} 1 & & * & 0 & & * & 0 & & & * \\ 0 & \cdots & 0 & 1 & & * & 0 & & & * \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & & & * \\ 0 & & & 0 & & & 0 & 0 & & 0 \\ \vdots & & & \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & & 0 & & \cdots & 0 & 0 & & 0 \end{array}\right).$$

Note that REF of an augmented matrix is not unique, but RREF is unique.

Once the augmented matrix is in REF or RREF, we can use back substitution to find a general solution.

Example.

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 3 \end{array}\right)$$

tells us that $x + y = 2$ and $y = 3$. So $y = 3$ and $x = 2 - y = -1$.

Once in RREF, we say that the linear system is inconsistent if the last column (after the vertical line) is a pivot column.

$$\left(\begin{array}{ccc|c} * & \cdots & * & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 0 \end{array}\right).$$

This means that the system has no solution. For if the leading entry is in the last column, we will solving for

$$0x_1 + 0x_2 + \cdots + 0x_1 = 1,$$

which is impossible. Otherwise, the linear system is consistent, that is, there will be solutions to the linear system. In this case,

$$\#\text{ of parameters } = \#\text{ of nonpivot columns on the left of the vertical line}$$

Meaning that if there are at least 1 nonpivot columns before the vertical line, the linear system will have infinitely many solutions, that is, the solution is not unique.

Example. 1. The system corresponding to

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 1 & -3 & 1 \end{array}\right)$$

is consistent since the last column is not a pivot column. The system has 1 parameter since the third column is also a nonpivot column.

2. The system corresponding to

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 5 \end{array}\right)$$

is inconsistent since the last column is a pivot column.

Hence, given a linear system, to find the solutions (if it exists) our task is to reduce the augmented matrix to REF or RREF. This will be accomplished through elementary row operations.

1.3 Elementary Row Operations

There are 3 types of elementary row operations.

1. Exchanging 2 row, $R_i \leftrightarrow R_j$,
2. Adding a multiple of a row to another, $R_i + cR_j$, $c \in \mathbb{R}$,
3. Multiplying a row by a nonzero constant, aR_j, $a \neq 0$.

Remark. 1. Note that we cannot multiply a row by 0, as it may change the linear system. For example, consider

$$\begin{array}{ccccc} x & + & y & = & 2 \\ x & - & y & = & 0 \end{array}$$

It has a unique solution $x = 1,\ y = 1$. Suppose in the augmented matrix we multiply row 2 by 0,

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & -1 & 0 \end{array}\right) \xrightarrow{0R_2} \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right)$$

then the system now has a general solution $x = 2 - s$, $y = s$.

2. Elementary row operations may not commute. For example,

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{2R_1} \xrightarrow{R_2 \leftrightarrow R_1} \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 2 & 0 & 0 \end{array}\right)$$

is not the same as

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_1} \xrightarrow{2R_1} \left(\begin{array}{cc|c} 0 & 2 & 0 \\ 1 & 0 & 0 \end{array}\right)$$

But if the elementary row operations do commute, we can stack them

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow[2R_1]{2R_2} \left(\begin{array}{cc|c} 2 & 0 & 0 \\ 0 & 2 & 0 \end{array}\right)$$

3. For the second type of elementary row operation, the row we put first is the row we are performing the operation upon,

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_1 + 2R_2} \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 1 & 0 \end{array}\right)$$

instead of

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{2R_2 + R_1} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 2 & 0 \end{array}\right)$$

In fact, the $2R_2 + R_1$ is not an elementary row operation, but a combination of 2 operations, $2R_2$ then $R_2 + R_1$. Here's another example,

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_1 + R_2} \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 0 & 0 \end{array}\right)$$

and

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_2 + R_1} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 1 & 0 \end{array}\right)$$

Two augmented matrices are <u>row equivalent</u> if one can be obtained from the other by elementary row operations.

Theorem. *Two augmented matrices are row equivalent if and only have they have the same RREF.*

We will give a proof to this theorem in lecture 3.

Observe that from the RREF we are able to uniquely obtain the solution set, and from the solution set, if we know the number of equations the linear system has, we are able to reconstruct the RREF uniquely. Hence, the previous theorem gives us the following statement.

Theorem. *Two linear systems have the same solution set if and only if their augmented matrices are row equivalent.*

Example. From the REF or RREF, we are able to read off a general solution.

1.
$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right)$$
This is in REF. We let the third variable be the parameter s, then we get $y = -s$ from the second row, and $x = 1 - s - (-s) = 1$ from the first row. So a general solution is $x = 1, y = -s, z = s$.

2.
$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{array}\right)$$
This is in RREF. General solution: $x = 1, y = s, z = s$.

Example. We will now reconstruct the RREF of the augmented matrix of a linear system given a general solution.

1. $x = 1 - 2s + t,\ y = s,\ z = t$, the linear system has 3 equations. Then substituting back $y = s$ and $z = t$ into $x = 1 - 2s + t$, we get $x + 2y - z = 1$. So the RREF of the augmented matrix is
$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

2. $x = 3 - 5s,\ y = 2 + 2s,\ z = s$, the linear system has 3 equations. Substituting back, we get,
$$\begin{array}{rcrcl} x & & + & 5z & = & 3 \\ & y & - & 2z & = & 2 \end{array}$$
and thus the RREF of the augmented matrix is
$$\left(\begin{array}{ccc|c} 1 & 0 & 5 & 3 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

3. $x = 3,\ y = 2,\ z = 1$, 3 equations. RREF:
$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

1.4 Gaussian Elimination and Gauss-Jordan Elimination

Step 1: Locate the leftmost column that does not consist entirely of zeros.

Step 2: Interchange the top row with another row, if necessay, to bring a nonzero entry to the top of the column found in Step 1.

Step 3: For each row below the top row, add a suitable multiple of the top row to it so that the entry below the leading entry of the top row becomes zero.

Step 4: Now cover the top row in the augmented matrix and begin again with Step 1 applied to the submatrix that remains. Continue this way until the entire matrix is in row-echelon form.

Once the above process is completed, we will end up with a REF. The following steps continue the process to reduce it to its RREF.

Step 5: Multiply a suitable constant to each row so that all the leading entries become 1.

Step 6: Beginning with the last nonzero row and working upward, add suitable multiples of each row to the rows above to introduce zeros above the leading entries.

Remark. The Gaussian elimination and Gauss-Jordan elimination may not be the fastest way to obtain the RREF of an augmented matrix. We do not have to follow the algorithm strictly when reducing the augmented matrix.

Example.

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ -1 & 2 & -1 & 1 \\ 2 & 0 & 3 & -2 \end{array}\right) \xrightarrow[R_3-2R_1]{R_2+R_1} \left(\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 0 & 3 & 1 & 5 \\ 0 & -2 & -1 & -10 \end{array}\right) \xrightarrow{R_3+\frac{2}{3}R_2} \left(\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 0 & 3 & 1 & 5 \\ 0 & 0 & -1/3 & -20/3 \end{array}\right)$$

The augmented matrix is now in REF. By back substitution, we have

$$z = 20,\ y = \frac{1}{3}(5-z) = -5,\ x = 4 - y - 2z = -31.$$

Alternatively, we can continue to reduce it to its RREF and read off the solution.

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 0 & 3 & 1 & 5 \\ 0 & 0 & -1/3 & -20/3 \end{array}\right) \xrightarrow{-3R_3} \left(\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 0 & 3 & 1 & 5 \\ 0 & 0 & 1 & 20 \end{array}\right) \xrightarrow[R_1-2R_3]{R_2-R_3} \left(\begin{array}{ccc|c} 1 & 1 & 0 & -36 \\ 0 & 3 & 0 & -15 \\ 0 & 0 & 1 & 20 \end{array}\right)$$

$$\xrightarrow{\frac{1}{3}R_2} \left(\begin{array}{ccc|c} 1 & 1 & 0 & -36 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 20 \end{array}\right) \xrightarrow{R_1-R_2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -31 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 20 \end{array}\right)$$

Indeed, the system is consistent, with unique solution $x = -31, y = -5, z = 20$.

1.5 Solving Linear Systems: Examples

1. Solve the following linear system

$$\begin{array}{ccccccccccccc} x_1 & + & 3x_2 & - & 2x_3 & + & & + & 2x_5 & & & = & 0 \\ 2x_1 & + & 6x_2 & - & 5x_3 & - & 2x_4 & + & 4x_5 & - & 3x_6 & = & -1 \\ & + & & + & 5x_3 & + & 10x_4 & & & + & 15x_6 & = & 5 \\ 2x_1 & + & 6x_2 & & & + & 8x_4 & + & 4x_5 & + & 18x_6 & = & 6 \end{array}$$

The augmented matrix is

$$\left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 2 & 6 & 0 & 8 & 4 & 18 & 6 \end{array}\right)$$

We will begin the Gaussian elimination.

$$\left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 2 & 6 & 0 & 8 & 4 & 18 & 6 \end{array}\right) \xrightarrow{R_2-2R_1} \xrightarrow{R_4-2R_1} \left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & -2 & 0 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 0 & 0 & 4 & 8 & 0 & 18 & 6 \end{array}\right)$$

$$\xrightarrow{R_3+5R_2} \xrightarrow{R_4+4R_2} \left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & -2 & 0 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \end{array}\right) \xrightarrow{R_3 \leftrightarrow R_4} \left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & -2 & 0 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

The augmented matrix is now in REF. We may use back substitution to obtain the solution, or continue to reduce to its RREF.

$$\xrightarrow{-R_2} \xrightarrow{(1/6)R_3} \left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2-3R_3} \left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{R_1+2R_2} \left(\begin{array}{cccccc|c} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

This corresponds to the following linear system

$$\begin{array}{ccccccccccc} x_1 & + & 3x_2 & & & + & 4x_4 & + & 2x_5 & & & = & 0 \\ & + & & & x_3 & + & 2x_4 & & & & & = & 0 \\ & & & & & & & & & + & x_6 & = & 1/3 \end{array}$$

Letting $x_2 = r, x_4 = s, x_5 = t$, a general solution is

$$x_1 = -3r - 4s - 2t,\ x_2 = r,\ x_3 = -2s,\ x_4 = s,\ x_5 = t,\ x_6 = 1/3.$$

2. Solving linear system,

$$\begin{array}{ccccccc} x_1 & + & ax_2 & + & 2x_3 & = & 0 \\ x_1 & & & + & x_3 & = & 1 \\ x_1 & & & + & ax_3 & = & 2 \end{array}$$

for some fixed real number a.

$$\left(\begin{array}{ccc|c} 1 & a & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & a & 2 \end{array}\right) \xrightarrow{R_2-R_1} \xrightarrow{R_3-R_1} \left(\begin{array}{ccc|c} 1 & a & 2 & 0 \\ 0 & -a & -1 & 1 \\ 0 & -a & a-2 & 2 \end{array}\right) \xrightarrow{R_3-R_2} \left(\begin{array}{ccc|c} 1 & a & 2 & 0 \\ 0 & -a & -1 & 1 \\ 0 & 0 & a-1 & 1 \end{array}\right)$$

The augmented matrix is almost a REF. To proceed to reduce it further, we have to consider cases. If $a = 1$,

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 0 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

the system is inconsistent. Next, consider the case where $a = 0$,

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow{R_3-R_2} \xrightarrow{R_1+2R_2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Then a general solution is

$$x_1 = 2,\ x_2 = s,\ x_3 = -1.$$

Otherwise, if $a \neq 0, 1$, then the system has a unique solution

$$x_3 = \frac{1}{a-1},\ x_2 = \frac{-1}{a}(1 + x_3) = \frac{-1}{a-1},\ x_1 = -ax_2 - 2x_3 = \frac{a-2}{a-1}.$$

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 2 Notes

References

1. Elementary Linear Algebra: Application Version, Section 1.3, 1.7
2. Linear Algebra with Application, Section 2.1-2.3

2 Matrices

2.1 Introduction to Matrices

A matrix is a rectangular array of number

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

were $a_{ij} \in \mathbb{R}$ are real numbers.

The size of a matrix is given by $m \times n$, where m is the number of rows and n is the number of columns. The (i, j)-entry of the matrix is the number a_{ij} in the i-th row and j-th column, for $i = 1, ..., m$, $j = 1, ..., n$. A matrix can also be denoted as

$$\mathbf{A} = (a_{ij})_{m\times n} = (a_{ij})_{i=1}^{m}{}_{j=1}^{n}.$$

Matrices are usually denoted by upper case bolded letters.

Example. 1. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 0 & -1 \end{pmatrix}$ is a 3×2 matrix. The $(2, 1)$-entry is 3.

2. $\begin{pmatrix} 2 & 1 & 0 \end{pmatrix}$ is a 1×3 matrix. The $(1, 2)$-entry is 2.

3. $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ is a 3×1 matrix. The $(3, 1)$-entry is 3.

4. $\begin{pmatrix} 4 \end{pmatrix}$ is a 1×1 matrix.

The last example shows that all real numbers can be thought of as 1×1 matrices.

Remark. 1. To be precise, the above examples are called real-valued matrices, or matrices with real number entries. Later we will be introduced to complex-valued and even matrices with function entries.

2. The choice of using round or square brackets is a matter of taste.

Example. 1. $\mathbf{A} = (a_{ij})_{2\times 3}$, $a_{ij} = i + j$.

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}$$

2. $\mathbf{B} = (b_{ij})_{3\times 2}$, $b_{ij} = (-1)^{i+j}$.

$$\mathbf{B} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$$

3. $\mathbf{C} = (c_{ij})_{3\times 3}$, $c_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$.

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2.2 Special types of Matrices

Here are some special types of matrices.

Column and row matrices. A $1 \times n$ matrix is called a row matrix, and a $n \times 1$ matrix is called a column matrix. However, we seldom use these terms. We usually call them row vectors and column vectors, respectively, which will be introduced in lecture 5.

Square matrices. A $m \times n$ matrix is a square matrix if the number of columns is equal to the number of rows $m = n$. A square matrix of size $n \times n$ is called an order n square matrix. It is usually denoted by $\mathbf{A} = (a_{ij})_n$.

Example. Order 2: $\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$ Order 3: $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 4 \\ 5 & 6 & 6 \end{pmatrix}$

The i-th diagonal entry of a square matrix is its (i, i)-entry. The diagonal entries of a square matrix $\mathbf{A} = (a_{ij})_n$ of order n is the collection $\{a_{11}, a_{22}, ..., a_{nn}\}$.

Diagonal matrices. A square matrix with all the non diagonal entries equal 0 is called a diagonal matrix, $\mathbf{D} = (d_{ij})_n$ with $d_{ij} = 0$ for all $i \neq j$. It is usually denoted by $\mathbf{D} = \text{diag}\{d_1, d_2, ..., d_n\}$.

Example.

$$\text{diag}\{1,1\} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{diag}\{0,0,0\} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{diag}\{1,2,3,4\} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

Scalar matrices. A diagonal matrix $\mathbf{A} = \text{diag}\{a_1, a_2, ..., a_n\}$ such that all the diagonal entries are equal $a_1 = a_2 = ... = a_n$ is called a scalar matrix.

Example. $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

Identity matrices. A scalar matrix with all diagonal entries equal 1 is called an identity matrix. An identity matrix of order n is denoted as $\mathbf{I}_n$. If there is no confusion with the order of the matrix, we will write $\mathbf{I}$ instead. So a scalar matrix can be written as $c\mathbf{I}$.

Zero matrices. A matrix (of any size) with all entries equal 0 is called a zero matrix. Usually denoted as $\mathbf{0}_{m\times n}$ for the size $m \times n$ zero matrix, and $\mathbf{0}_n$ for the zero square matrix of order n. If it is clear in the context, we will just denote it as $\mathbf{0}$.

Triangular matrices. A square matrix $\mathbf{A} = (a_{ij})_n$ with all entries below (above) the diagonal equal 0, that is, $a_{ij} = 0$ for all $i > j$ ($i < j$), is called an upper (lower) triangular matrix. It is a strictly upper or lower matrix if the diagonals are equal to zero too, that is, $a_{ij} = 0$ for all $i \geq j$ ($i \leq j$).

Upper triangular: $\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & * \end{pmatrix}$ Strictly upper triangular: $\begin{pmatrix} 0 & * & \cdots & * \\ 0 & 0 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$

Lower triangular: $\begin{pmatrix} * & 0 & \cdots & 0 \\ * & * & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{pmatrix}$ Strictly lower triangular: $\begin{pmatrix} 0 & 0 & \cdots & 0 \\ * & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & 0 \end{pmatrix}$

Exercise: Is it true that every matrix is row equivalent to an upper triangular matrix? How about strictly upper matrix?

Symmetrix matrices. A square matrix $\mathbf{A} = (a_{ij})_n$ such that $a_{ij} = a_{ji}$ for all $i, j = 1, ..., n$ is called a symmetric matrix, that is, the entries are diagonally reflected along the diagonal of $\mathbf{A}$.

Example. $\begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix}$ $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 4 \\ 3 & 4 & 5 \end{pmatrix}$ $\begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 2 & 1 & 2 & 1 \end{pmatrix}$.

2.3 Matrix Operations

Two matrices are equal if they are of the same size and all the entries are equal.

Example. 1.

$$\begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix} \neq \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

for any choice of a, b, c, d, e, f.

2.

$$\begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

if and only if $a = 1$, $b = 1$, $c = 3$, $d = 2$.

Matrix Addition and Scalar Multiplication

Let $\mathbf{A} = (a_{ij})_{m\times n}$ and $\mathbf{B} = (b_{ij})_{m\times n}$ be two matrices and $c \in \mathbb{R}$ a real number. We define the following operations as such.

1. (Scalar multiplication) $c\mathbf{A} = (ca_{ij})$.
2. (Matrix addition) $\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij})_{m\times n}$

Remark. 1. Matrix addition is only defined between matrices of the same size.

2. $-\mathbf{A} = (-1)\mathbf{A}$.
3. Matrix substraction is defined to be the addition of a negative multiple of another matrix,
$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}.$$

Theorem (Properties of matrix addition and scalar multiplication)**.** *For matrices* $\mathbf{A} = (a_{ij})_{m\times n}$, $\mathbf{B} = (b_{ij})_{m\times n}$, $\mathbf{C} = (c_{ij})_{m\times n}$, *and real numbers* $a, b \in \mathbb{R}$,

(i) *(Commutative)* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$,

(ii) *(Associative)* $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$,

(iii) *(Additive identitiy)* $\mathbf{0}_{m\times n} + \mathbf{A} = \mathbf{A}$,

(iv) *(Additive inverse)* $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}_{m\times n}$,

(v) *(Distributive law)* $a(\mathbf{A} + \mathbf{B}) = a\mathbf{A} + a\mathbf{B}$,

(vi) *(Scalar addition)* $(a + b)\mathbf{A} = a\mathbf{A} + b\mathbf{A}$,

(vii) $(ab)\mathbf{A} = a(b\mathbf{A})$,

(viii) *if* $a\mathbf{A} = \mathbf{0}_{m\times n}$, *then either* $a = 0$ *or* $\mathbf{A} = \mathbf{0}$.

Proof. To show equality, we have to show that the matrices on the left and right of the equality have the same size, and that the corresponding entries are equal. It is clear that the matrices on both sides has the same size, so we will only check that the entries agree.

(i) $a_{ij} + b_{ij} = b_{ij} + a_{ij}$ follows directly from commutativity of addition of real numbers.

(ii) $a_{ij} + (b_{ij} + c_{ij}) = (a_{ij} + b_{ij}) + c_{ij}$ follows directly from associativity of addition of real numbers.

(iii) $0 + a_{ij} = a_{ij}$ follows directly from the additive identity property of real numbers.

(iv) $a_{ij} + (-a_{ij}) = 0$ follows directly from additive inverse property of real numbers.

(v) $a(a_{ij} + b_{ij}) = aa_{ij} + ab_{ij}$ follows directly from distributive property of addition of real numbers.

(vi) $(a+b)a_{ij} = aa_{ij} + ba_{ij}$ follows directly from distributive property of addition of real numbers.

(vii) $(ab)a_{ij} = a(ba_{ij})$ follows directly from associativity of multiplication of real numbers.

(viii) If $aa_{ij} = 0$, then $a = 0$ or $a_{ij} = 0$. Suppose $a \neq 0$, then $a_{ij} = 0$ for all i, j. So $\mathbf{A} = \mathbf{0}$.

$\square$

Remark. 1. Since addition is associative, we will not write the parentheses when adding multiple matrices.

2. Property (iii) and (i) imply that $\mathbf{A} + \mathbf{0} = \mathbf{A}$.

3. Property (iv) and (i) imply that $-\mathbf{A} + \mathbf{A} = \mathbf{0}$.

Matrix Multiplication

Let $\mathbf{A} = (a_{ij})_{m\times p}$ and $\mathbf{B} = (b_{ij})_{p\times n}$. The product $\mathbf{AB}$ is defined to be a $m \times n$ matrix whose (i, j)-entry is

$$\sum_{k=1}^{p} a_{ik}b_{kj} = a_{i1}b_{ij} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj}.$$

Example.

$$\underset{(2\times 3)}{\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}} \ \underset{(3\times 2)}{\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix}} = \underset{(2\times 2)}{\begin{pmatrix} 1+4-3=2 & 1+6-6=1 \\ 4+10-6=12 & 4+15-12=7 \end{pmatrix}}$$

Remark. 1. For $\mathbf{AB}$ to be defined, the number of columns of $\mathbf{A}$ must agree with the number of rows of $\mathbf{B}$. The resultant matrix has the same number of rows as $\mathbf{A}$, and the same number of columns as $\mathbf{B}$.

$$(m \times p)(p \times n) = (m \times n).$$

2. Matrix multiplication is not commutative, that is $\mathbf{AB} \neq \mathbf{BA}$ in general. For example, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \neq \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

3. If we are multiplying $\mathbf{A}$ to the left of $\mathbf{B}$, we are <u>pre-multiplying</u> $\mathbf{A}$ to $\mathbf{B}$, $\mathbf{AB}$. If we multiply $\mathbf{A}$ to the right of $\mathbf{B}$, we are <u>post-multiplying</u> $\mathbf{A}$ to $\mathbf{B}$, $\mathbf{BA}$. Pre-multiplying $\mathbf{A}$ to $\mathbf{B}$ is the same as post-multiplying $\mathbf{B}$ to $\mathbf{A}$.

Theorem (Properties of matrix multiplication)**.** *(i) (Associative) For matrices* $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times q}$, *and* $\mathbf{C} = (c_{ij})_{q\times n}$ $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

(ii) (Left distributive law) For matrices $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times n}$, *and* $\mathbf{C} = (c_{ij})_{p\times n}$, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

(iii) (Right distributive law) For matrices $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{m\times p}$, *and* $\mathbf{C} = (c_{ij})_{p\times n}$, $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.

(iv) (Commute with scalar multiplication) For any real number $c \in \mathbb{R}$, and matrices $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times n}$, $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B})$.

(v) (Multiplicative identity) For any $m \times n$ matrix $\mathbf{A}$, $\mathbf{I}_m\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$.

(vi) (Zero divisor) There exists $\mathbf{A} \neq \mathbf{0}_{m\times p}$ and $\mathbf{B} \neq \mathbf{0}_{p\times n}$ such that $\mathbf{AB} = \mathbf{0}_{m\times n}$.

(vii) (Zero matrix) For any $m \times n$ matrix $\mathbf{A}$, $\mathbf{A}\mathbf{0}_{n\times p} = \mathbf{0}_{m\times p}$ and $\mathbf{0}_{p\times m}\mathbf{A} = \mathbf{0}_{p\times n}$.

The proof is beyond the scope of this course. Interested readers may refer to the appendix.

Remark. 1. For square matrices, we define $\mathbf{A}^2 = \mathbf{AA}$, and define inductively, $\mathbf{A}^n = \mathbf{A}\mathbf{A}^{n-1}$, for $n \geq 2$. It follows that $\mathbf{A}^n\mathbf{A}^m = \mathbf{A}^{n+m}$.

2. In general $(\mathbf{AB})^n \neq \mathbf{A}^n\mathbf{B}^n$. (Why?)

Tranpose

For a $m \times n$ matrix $\mathbf{A}$, the transpose of $\mathbf{A}$, written as $\mathbf{A}^T$, is a $n \times m$ matrix whose (i, j)-entry is the (j, i)-entry of $\mathbf{A}$, that is, if $\mathbf{A}^T = (b_{ij})_{n\times m}$, then

$$b_{ij} = a_{ji}$$

for all $i = 1, ..., n$, $j = 1, ..., m$. Equivalently, the rows of $\mathbf{A}$ are the columns of $\mathbf{A}^T$ and vice versa.

Example. 1. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 5 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$ 2. $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$ 3. $\begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix}$

This gives us an alternative way to define symmetric matrices. A square matrix $\mathbf{A}$ is symmetric if and only if $\mathbf{A}^T = \mathbf{A}$.

Theorem (Properties of transpose). *(i) For any matrix $\mathbf{A}$, $(\mathbf{A}^T)^T = \mathbf{A}$.*

(ii) For any matrix $\mathbf{A}$, and real number $c \in \mathbb{R}$, $(c\mathbf{A})^T = c\mathbf{A}^T$.

(iii) For matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size, $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.

(iv) For matrices $\mathbf{A} = (a_{ij})_{m\times p}$ and $\mathbf{B} = (b_{ij})_{p\times n}$, $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

Refer to the appendix for the proof.

Example.

$$\begin{aligned}
\left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix}\right)^T &= \begin{pmatrix} 2 & 1 \\ 12 & 7 \end{pmatrix}^T = \begin{pmatrix} 2 & 12 \\ 1 & 7 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 & -1 \\ 1 & 3 & -2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix}^T \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T
\end{aligned}$$

Which of the following statements are true? Justify.

(a) If $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices of the same size, then so is $\mathbf{A}+\mathbf{B}$.

(b) If $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices (with the appropriate sizes), then so is $\mathbf{AB}$.

2.4 Revisit Linear System

Given a linear system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

We can represent it as a <u>matrix equation</u> $\mathbf{Ax}=\mathbf{b}$, or

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

or as <u>vector equation</u>

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

A homogeneous system is thus written as $\mathbf{Ax}=\mathbf{0}$. From property (vii) of matrix multiplication, we have $\mathbf{x}=\mathbf{0}$ as a solution to any homogeneous system. This is called the <u>trivial solution</u>. If there is a solution $\mathbf{x}\neq\mathbf{0}$, then we say that the homogeneous linear system admits <u>nontrivial solutions</u>.

Example.

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

admits a nontrivial solution $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Appendix for Lecture 2

Theorem (Properties of matrix multiplication). *(i) (Associative) For matrices* $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times q}$, *and* $\mathbf{C} = (c_{ij})_{q\times n}$ $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

(ii) (Left distributive law) For matrices $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times n}$, *and* $\mathbf{C} = (c_{ij})_{p\times n}$, $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$.

(iii) (Right distributive law) For matrices $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{m\times p}$, *and* $\mathbf{C} = (c_{ij})_{p\times n}$, $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$.

(iv) (Commute with scalar multiplication) For any real number $c \in \mathbb{R}$, *and matrices* $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times n}$, $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B})$.

(v) (Multiplicative identity) For any $m \times n$ *matrix* $\mathbf{A}$, $\mathbf{I}_m\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$.

(vi) (Zero divisor) There exists $\mathbf{A} \neq \mathbf{0}_{m\times p}$ *and* $\mathbf{B} \neq \mathbf{0}_{p\times n}$ *such that* $\mathbf{AB} = \mathbf{0}_{m\times n}$.

(vii) (Zero matrix) For any $m \times n$ *matrix* $\mathbf{A}$, $\mathbf{A}\mathbf{0}_{n\times p} = \mathbf{0}_{m\times p}$ *and* $\mathbf{0}_{p\times m}\mathbf{A} = \mathbf{0}_{p\times n}$.

Proof. We will check that the corresponding entries on each side agrees. The check for the size of matrices agree is trivial and is left to the reader.

(i) The (i,j)-entry of $(\mathbf{AB})\mathbf{C}$ is

$$\sum_{l=1}^{q}(\sum_{k=1}^{p} a_{ik}b_{kl})c_{lj} = \sum_{l=1}^{q}\sum_{k=1}^{p} a_{ik}b_{kl}c_{lj}.$$

The (i,j)-entry of $\mathbf{A}(\mathbf{BC})$ is

$$\sum_{k=1}^{p} a_{ik}(\sum_{l=1}^{q} b_{kl}c_{lj}) = \sum_{k=1}^{p}\sum_{l=1}^{q} a_{ik}b_{kl}c_{lj}.$$

Since both sums has finitely many terms, the sums commute and thus the (i,j)-entry of $(\mathbf{AB})\mathbf{C}$ is equal to the (i,j)-entry of $\mathbf{A}(\mathbf{BC})$.

(ii) The (i,j)-entry of $\mathbf{A}(\mathbf{B}+\mathbf{C})$ is $\sum_{k=1}^{p} a_{ik}(b_{kj}+c_{kj}) = \sum_{k=1}^{p}(a_{ik}b_{kj}+a_{ik}c_{kj}) = \sum_{k=1}^{p} a_{ik}b_{kj} + \sum_{k=1}^{p} a_{ik}c_{kj}$, which is the (i,j)-entry of $\mathbf{AB}+\mathbf{AC}$.

(iii) The proof is analogous to left distributive law.

(iv) Left to reader.

(v) Note that $I = (\delta_{ij})$, where $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$. So the (i,j)-entry of $\mathbf{I}_m\mathbf{A}$ is

$$\delta_{i1}a_{1j} + \cdots + \delta_{ii}a_{ij} + \cdots + \delta_{im}a_{mj} = 0a_{1j} + \cdots + 1a_{ij} + \cdots + 0a_{mj} = a_{ij}.$$

The proof for $\mathbf{A} = \mathbf{A}\mathbf{I}_n$ is analogous.

(vi) Consider for example $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

(vii) Left to reader, if you have read till this far, surely this proof is trivial to you.

$\square$

Theorem (Properties of transpose). *(i) For any matrix* $\mathbf{A}$*,* $(\mathbf{A}^T)^T = \mathbf{A}$*.*

(ii) For any matrix $\mathbf{A}$*, and real number* $c \in \mathbb{R}$*,* $(c\mathbf{A})^T = c\mathbf{A}^T$*.*

(iii) For matrices $\mathbf{A}$ *and* $\mathbf{B}$ *of the same size,* $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$*.*

(iv) For matrices $\mathbf{A} = (a_{ij})_{m\times p}$ *and* $\mathbf{B} = (b_{ij})_{p\times n}$*,* $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$*.*

Proof. We will only proof (iv). The rest is left to the reader. The (j,i)-entry of $\mathbf{AB}$ is

$$\sum_{k=1}^{p} a_{jk}b_{ki},$$

which is the (i,j)-entry of $(\mathbf{AB})^T$. The (i,j)-entry of $\mathbf{B}^T\mathbf{A}^T$ is

$$\sum_{k=1}^{p} b_{ki}a_{jk} = \sum_{k=1}^{p} a_{jk}b_{ki},$$

which is exactly the (i,j)-entry of $(\mathbf{AB})^T$. $\square$

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 3 Notes

References

1. Elementary Linear Algebra: Application Version, Section 1.4-1.6
2. Linear Algebra with Application, Section 2.4-2.5

2.5 Block Multiplication

Let $\mathbf{A}$ be an $m \times n$ matrix,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The rows of $\mathbf{A}$ are the $1 \times n$ submatrices of $\mathbf{A}$,

$$\mathbf{r}_1 = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}, \mathbf{r}_2 = \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}, ..., \mathbf{r}_m = \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

and the columns of $\mathbf{A}$ are the $m \times 1$ submatrices of $\mathbf{A}$

$$\mathbf{c}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{c}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, ..., \mathbf{c}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

In general, a $p \times q$ submatrix of an $m \times n$ matrix $\mathbf{A}$, $p \leq m$, $q \leq n$, is formed by taking a $p \times q$ block of the entries of the matrix $\mathbf{A}$.

Example. $\begin{pmatrix} 4 & 6 & 1 \\ 2 & 2 & 1 \end{pmatrix}$ is a 2×3 submatrix of $\mathbf{A} = \begin{pmatrix} 1 & 3 & 5 & 7 & 1 \\ 2 & 4 & 6 & 1 & 2 \\ 1 & 2 & 2 & 1 & 3 \end{pmatrix}$, taken from row 2 and 3, and columns 2 to 4 of $\mathbf{A}$.

Observe that matrix multiplication respects submatrices, in the sense that if we take $k \times p$ submatrices of $\mathbf{A}$ and multiply to $p \times l$ submatrices of $\mathbf{B}$, we obtain $k \times l$ submatrices of $\mathbf{AB}$. We call this block multiplication.

Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 2 \\ 3 & 2 & 1 \\ 4 & 4 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 3 & 5 & 7 & 1 \\ 2 & 4 & 6 & 1 & 2 \\ 1 & 2 & 2 & 1 & 3 \end{pmatrix}$. Then multiplying the submatrix of $\mathbf{A}$, $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 2 \end{pmatrix}$ consisting of the first 2 rows of $\mathbf{A}$ to the submatrix of $\mathbf{B}$,

$\begin{pmatrix} 1 & 3 \\ 2 & 4 \\ 1 & 2 \end{pmatrix}$ consisting of the first 2 columns of $\mathbf{B}$, we get $\begin{pmatrix} 2 & 5 \\ 7 & 15 \end{pmatrix}$, which is a 2×2 submatrix of $\mathbf{AB}$ consisting of the first 2 rows and first 2 columns of $\mathbf{AB}$.

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 2 \\ 3 & 2 & 1 \\ 4 & 4 & 2 \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 & 7 & 1 \\ 2 & 4 & 6 & 1 & 2 \\ 1 & 2 & 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 5 & 7 & 8 & 4 \\ 7 & 15 & 21 & 11 & 11 \\ 8 & 9 & 29 & 24 & 10 \\ 14 & 32 & 48 & 34 & 18 \end{pmatrix}.$$

In particular, we have the following cases.

1. If $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$, then

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 & \cdots & \mathbf{A}\mathbf{b}_n \end{pmatrix},$$

that is, the columns of $\mathbf{AB}$ is just $\mathbf{A}$ pre-multiplying to the columns of $\mathbf{B}$.

2. If $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$, then

$$\mathbf{AB} = \begin{pmatrix} \mathbf{a}_1\mathbf{b}_1 & \mathbf{a}_1\mathbf{b}_2 & \cdots & \mathbf{a}_1\mathbf{b}_n \\ \mathbf{a}_2\mathbf{b}_1 & \mathbf{a}_2\mathbf{b}_2 & \cdots & \mathbf{a}_2\mathbf{b}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m\mathbf{b}_1 & \mathbf{a}_m\mathbf{b}_2 & \cdots & \mathbf{a}_m\mathbf{b}_n \end{pmatrix}.$$

Indeed, the (i, j)-entry of $\mathbf{AB}$ is

$$\mathbf{a}_i\mathbf{b}_j = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{ip} \end{pmatrix} \begin{pmatrix} b_{ij} \\ b_{2j} \\ \vdots \\ b_{pj} \end{pmatrix} = \sum_{k=1}^{p} a_{ik}b_{kj}.$$

Exercise: If $\mathbf{A}_i$ is a $m_i \times p$ matrix, for $i = 1, 2$, and $\mathbf{B}_i$ is a $p \times n_i$ matrix for $i = 1, 2$, show that the following block multiplication holds,

$$\begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1\mathbf{B}_1 & \mathbf{A}_1\mathbf{B}_2 \\ \mathbf{A}_2\mathbf{B}_1 & \mathbf{A}_2\mathbf{B}_2 \end{pmatrix},$$

where $\begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix}$ is a $(m_1 + m_2) \times p$ matrix, where the first m_1 rows are from $\mathbf{A}_1$, and the $m_1 + 1$ to $m_1 + m_2$ rows are from $\mathbf{A}_2$, and $\begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix}$ is a $p \times (n_1 + n_2)$ matrix, where the first n_1 columns are from $\mathbf{B}_1$ and the $n_1 + 1$ to $n_1 + n_2$ columns are from $\mathbf{B}_2$.

Suppose now we have p linear systems with the same coefficient matrix $\mathbf{A} = (a_{ij})_{m \times n}$, for $k = 1, ..., p$,

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_{1k} \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_{2k} \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_{mk} \end{array}.$$

We can represent this as a matrix equation $\mathbf{AX} = \mathbf{B}$, where $\mathbf{X} = (x_{ij})_{n\times p}$, and $\mathbf{B} = (b_{ij})_{m\times p}$,

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} \end{pmatrix}.$$

This can be reprenseted as a p simultaneous augmented matrix

$$\left(\begin{array}{cccc|c|c|c|c} a_{11} & a_{12} & \cdots & a_{1n} & b_{11} & b_{12} & & b_{1p} \\ a_{21} & a_{22} & \cdots & a_{2n} & b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_{m1} & b_{m2} & & b_{mp} \end{array}\right).$$

Then we may proceed to perform row operations to the augmented matrix above, and solve for all p linear systems simultaneously.

Example. 1. Solve the following 2 linear systems

$$\begin{array}{rcrcrcl} x & + & 2y & - & 3z & = & 1 \\ 2x & + & 6y & - & 11z & = & 1 \\ x & - & 2y & + & 7z & = & 1 \end{array} \quad \text{and} \quad \begin{array}{rcrcrcl} x & + & 2y & - & 3z & = & 1 \\ 2x & + & 6y & - & 11z & = & 2 \\ x & - & 2y & + & 7z & = & 1 \end{array}$$

can be represented as a matrix equation

$$\begin{pmatrix} 1 & 2 & -3 \\ 2 & 6 & -11 \\ 1 & -2 & 7 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix},$$

and the augmented matrix is

$$\left(\begin{array}{ccc|c|c} 1 & 2 & -3 & 1 & 1 \\ 2 & 6 & -11 & 1 & 2 \\ 1 & -2 & 7 & 1 & 1 \end{array}\right).$$

The reduced row-echelon form is

$$\left(\begin{array}{ccc|c|c} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & -5/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right).$$

So the first linear system is inconsistent, and the second has a general solution $x = 1 - 2s, y = (5/2)s, z = s$.

2. Solve the following 3 linear systems

$$\begin{array}{rcrcrcl} 3x & + & 2y & - & z & = & a \\ 5x & - & y & + & 3z & = & b \\ 2x & + & y & - & z & = & c \end{array}, \text{ for } \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

The matrix equation is

$$\begin{pmatrix} 3 & 2 & -1 \\ 5 & -1 & 3 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & 0 \end{pmatrix},$$

and the augmented matrix is

$$\left(\begin{array}{ccc|c|c|c} 3 & 2 & -1 & 1 & 2 & 1 \\ 5 & -1 & 3 & 2 & 1 & 1 \\ 2 & 1 & -1 & 3 & 1 & 0 \end{array}\right).$$

The reduced row-echelon form is

$$\left(\begin{array}{ccc|c|c|c} 1 & 0 & 0 & 5/3 & 2/9 & -1/9 \\ 0 & 1 & 0 & -11/3 & 7/9 & 10/9 \\ 0 & 0 & 1 & -10/3 & 2/9 & 8/9 \end{array}\right).$$

Hence, all 3 linear systems has unique solutions,

$$\begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 5 \\ -11 \\ 10 \end{pmatrix}, \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 2 \\ 7 \\ 2 \end{pmatrix}, \begin{pmatrix} x_{13} \\ x_{13} \\ x_{33} \end{pmatrix} = \frac{1}{9}\begin{pmatrix} -1 \\ 10 \\ 8 \end{pmatrix}.$$

2.6 Inverse of a Matrix

Recall that for a nonzero real number $c \in \mathbb{R}$, the multiplicative inverse (or just inverse) of c, denoted as $\frac{1}{c}$, is defined to be the number such that when multiplied to c gives 1, $\frac{1}{c} \times c = 1$ (this is why 0 has no inverse, for $a \times 0 = 0$ for any real number $a \in \mathbb{R}$, and so there can be no $a \in \mathbb{R}$ such that $a \times 0 = 1$).

Hence, in order to define, if possible, the inverse of a matrix, we first need to identity the matrix that serves the same role as 1 in real numbers. We indeed have such an object. Recall that the identity matrix has the multiplicative identity property, for any $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{I}_m\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n.$$

Remark. We may define an inverse of a $m \times n$ matrix $\mathbf{A}$ to be a matrix $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{I}$. However, since matrix multiplication is non-commutative, we might not have $\mathbf{AB} = \mathbf{I}$. In fact, if $m \neq n$, then for $\mathbf{AB}$ and $\mathbf{BA}$ to be defined, $\mathbf{B}$ must be of size $n \times m$, and so $\mathbf{AB}$ is not even of the same size as $\mathbf{BA}$. Therefore, we shall consider them separately.

Let $\mathbf{A}$ be a $m \times n$ matrix. A left inverse of $\mathbf{A}$ is a matrix $n \times m$ matrix $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{I}_n$. A right inverse of $\mathbf{A}$ is a $n \times m$ matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{I}_m$.

Example. 1. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$

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

Remark. 1. Note that if $\mathbf{B}$ is a left inverse of $\mathbf{A}$, then $\mathbf{A}$ is a right inverse of $\mathbf{B}$.

2. In the definition, we need not specify the size of the left or right inverse $\mathbf{B}$, since identity matrices are square matrices, the size of $\mathbf{B}$ is determined by $\mathbf{A}$.

3. Not every matrix has a left or right inverse. For example, show that $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ cannot have left or right inverse. More generally, if $n > m$, then any $m \times n$ matrix cannot have a left inverse, and if $m > n$, then any $m \times n$ matrix cannot have a right inverse. The proof of this statement is postponed till lecture 9.

4. Left and right inverses are not unique. For example,

$$\begin{pmatrix} 1 & 0 & s \\ 0 & 1 & t \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ a & b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

for any $a, b, s, t \in \mathbb{R}$.

5. For this reason and more, we will not delve too much into inverses for non square matrices. It turns out that if we restrict our attention to square matrices, we have (almost) all the desired properties that inverses of real numbers have.

A square matrix $\mathbf{A}$ of order n is <u>invertible</u> if there exists a square matrix $\mathbf{B}$ of order n such that $\mathbf{AB} = \mathbf{I}_n = \mathbf{BA}$. A square matrix is <u>singular</u> if it is not invertible.

Remark. 1. As before, it is not required to specify in the definition the size of $\mathbf{B}$, it is a consequence of the definition.

2. For $\mathbf{A}$ to be invertible, we need to check that there exists a $\mathbf{B}$ that is simultaneously a left and right inverse of $\mathbf{A}$, that is, we need to show both $\mathbf{BA} = \mathbf{I}_n$ and $\mathbf{AB} = \mathbf{I}_n$.

3. It turns out that to show that a square matrix $\mathbf{A}$ is invertible, suffice to show that it has a left inverse, or a right inverse. Moreoever, if $\mathbf{BA} = \mathbf{I}$, then neessarily $\mathbf{AB} = \mathbf{I}$ and vice versa. The actual theorem and proof will be given in the next lecture.

Theorem (Uniqueness of inverse)**.** *If* $\mathbf{B}$ *and* $\mathbf{C}$ *are both inverses of a square matrix* $\mathbf{A}$*, then* $\mathbf{B} = \mathbf{C}$*.*

Proof. By definition, $\mathbf{BA} = \mathbf{I}_n = \mathbf{AC}$. So

$$\mathbf{B} = \mathbf{BI} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{IC} = \mathbf{C}.$$

□

So since inverses are unique, we can denote the inverse of an invertible matrix $\mathbf{A}$ by $\mathbf{A}^{-1}$. That is, if $\mathbf{A}$ is invertible, there exists a unique matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{AA}^{-1} = \mathbf{I}_n = \mathbf{A}^{-1}\mathbf{A}.$$

Example. 1. The identity matrix $\mathbf{I}$ is invertible with $\mathbf{I}^{-1} = \mathbf{I}$.

2. $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. So $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}$.

3. $\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}$. So $\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$.

Theorem (Cancellation law for matrices)**.** *Let* $\mathbf{A}$ *be a invertible matrix of order* n.

(i) If $\mathbf{B}$ *and* $\mathbf{C}$ *are* $n \times m$ *matrices with* $\mathbf{AB} = \mathbf{AC}$*, then* $\mathbf{B} = \mathbf{C}$.

(ii) If $\mathbf{B}$ *and* $\mathbf{C}$ *are* $m \times n$ *matrices with* $\mathbf{BA} = \mathbf{CA}$*, then* $\mathbf{B} = \mathbf{C}$.

Proof. (i) Pre-multiply $\mathbf{A}^{-1}$ to both sides of $\mathbf{AB} = \mathbf{AC}$, we get $\mathbf{B} = \mathbf{IB} = \mathbf{A}^{-1}\mathbf{AB} = \mathbf{A}^{-1}\mathbf{AC} = \mathbf{IC} = \mathbf{C}$.

(ii) Post-multiply $\mathbf{A}^{-1}$ to both sides of $\mathbf{BA} = \mathbf{CA}$, we get we get $\mathbf{B} = \mathbf{BI} = \mathbf{BAA}^{-1} = \mathbf{CAA}^{-1} = \mathbf{CI} = \mathbf{C}$.

□

Exercise:

1. Let $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}$. By writing $\mathbf{B} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, using the equation $\mathbf{AB} = \mathbf{I} = \mathbf{BA}$, find the inverse of $\mathbf{A}$. What do you notice about the equations derived from $\mathbf{AB} = \mathbf{I}$ and $\mathbf{I} = \mathbf{BA}$?

2. Show that if $\mathbf{A}$ is invertible, the linear system $\mathbf{Ax} = \mathbf{b}$ is consistent.

Theorem (Properties of inverses)**.** *Let* $\mathbf{A}$ *be an invertible matrix of order* n.

(i) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

(ii) For any nonzero real number $a \in \mathbb{R}$*,* $(a\mathbf{A})$ *is invertible with inverse* $(a\mathbf{A})^{-1} = \frac{1}{a}\mathbf{A}^{-1}$.

(iii) $\mathbf{A}^T$ *is invertible with inverse* $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

(iv) If $\mathbf{B}$ *is an invertible matrix of order* n*, then* $(\mathbf{AB})$ *is invertible with inverse* $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Proof. (i) Since $\mathbf{AA}^{-1} = \mathbf{I}_n = \mathbf{A}^{-1}\mathbf{A}$, then $\mathbf{A}$ is the inverse of $\mathbf{A}^{-1}$.

(ii) We directly check that $a\mathbf{A}(\frac{1}{a}\mathbf{A}^{-1}) = \frac{a}{a}\mathbf{I}_n = \mathbf{I}_n = (\frac{1}{a}\mathbf{A}^{-1})(a\mathbf{A})$.

(iii) Since $\mathbf{I}$ is symmetric, $\mathbf{I} = \mathbf{I}^T = (\mathbf{AA}^{-1})^T = (\mathbf{A}^{-1})^T\mathbf{A}^T$, and similarly, $\mathbf{I} = \mathbf{A}^T(\mathbf{A}^{-1})^T$. Hence, $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

(iv) We directly check that $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_n$ and $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}_n$

□

Remark. 1. By induction, one can prove that the product of invertible matrices is invertible, and $(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k)^{-1} = \mathbf{A}_k^{-1}\cdots\mathbf{A}_2^{-1}\mathbf{A}_1^{-1}$ if $\mathbf{A}_i$ is an invertible matrix for $i = 1, ..., k$.

2. We define the negative power of an invertible matrix to be

$$\mathbf{A}^{-n} = (\mathbf{A}^{-1})^n$$

for any positive integer n.

Exercise: Show that $\mathbf{AB}$ is invertible if and only if both $\mathbf{A}$ and $\mathbf{B}$ are invertible. Hint: If $\mathbf{AB}$ is invertible, let $\mathbf{C}$ be the inverse. Pre and post multiply $\mathbf{AB}$ with $\mathbf{C}$.

Theorem. *An order 2 square matrix*

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is invertible if and only if $ad - bc \neq 0$, in which case the inverse is given by the formula

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The formula is obtained by using the adjoint of $\mathbf{A}$, which is beyond the scope of this module. However, readers may verify that indeed we have $\mathbf{AA}^{-1} = \mathbf{I}_2 = \mathbf{A}^{-1}\mathbf{A}$.

2.7 Elementary Matrices

A square matrix of order n $\mathbf{E}$ is called an elementary matrix if it can be obtained from the identity matrix $\mathbf{I}_n$ by performing a single elementary row operation

$$\mathbf{I}_n \xrightarrow{r} \mathbf{E},$$

where r is an elementary row operation. The elementary row operations is said to be the row operation corresponding to the elementary matrix.

Theorem (Elementary matrices and elementary row operations)**.** *Let $\mathbf{A}$ be an $n \times m$ matrix and let $\mathbf{E}$ be the elementary matrix*

(i)

$$\begin{array}{rc} \begin{matrix} \\ \\ i \\ \\ j \\ \\ \\ \end{matrix} & \begin{pmatrix} 1 & 0 & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & a & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \\ & \begin{matrix} & & i & & j & & \end{matrix} \end{array} \textit{ if } i < j, \textit{ or } \begin{array}{rc} \begin{matrix} \\ \\ j \\ \\ i \\ \\ \\ \end{matrix} & \begin{pmatrix} 1 & 0 & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & a & & 1 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \\ & \begin{matrix} & & i & & j & & \end{matrix} \end{array} \textit{ if } i > j,$$

(ii)

$$\begin{array}{rc} \begin{matrix} \\ \\ i \\ \\ j \\ \\ \\ \end{matrix} & \begin{pmatrix} 1 & 0 & & & & & 0 \\ & \ddots & & & & & \\ 0 & & & & 1 & & 0 \\ & & & \ddots & & & \\ 0 & & 1 & & & & 0 \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \\ & \begin{matrix} & & i & & j & & \end{matrix} \end{array},$$

(iii)

$$\begin{array}{cc} i & \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & c & & \\ & & & \ddots & \\ 0 & & & & 1 \end{pmatrix}, \; c \neq 0. \\ & \quad\quad\quad i \end{array}$$

Then the product $\mathbf{EA}$ *is obtained from* $\mathbf{A}$ *via performing the corresponding elementary row operations*

(i) adding a *times row* j *to row* i $(R_i + aR_j)$,

(ii) exchanging row i *with row* j $(R_i \leftrightarrow R_j)$,

(iii) multiplying row i *by* c,

respectively. In words, it means that pre-multiplying an elementary matrix is equivalent to performing the corresponding elementary row operation.

Proof. The reader can directly verify by computing the product $\mathbf{EA}$ for each of the three cases. □

Example. 1.

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix} \xrightarrow{R_2+2R_1} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 3 & 4 & 5 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix}$$

2.

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 4 & 2 \\ 1 & 2 & 3 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix}$$

3.

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix} \xrightarrow{-2R_2} \begin{pmatrix} 1 & 1 & 1 & 0 \\ -2 & -4 & -6 & 2 \\ 2 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix}$$

Corollary. *Suppose the matrix* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by performing row operations* $r_1, r_2, ..., r_k$,

$$\mathbf{A} \xrightarrow{r_1}\xrightarrow{r_2} \cdots \xrightarrow{r_k} \mathbf{B}.$$

Let $\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_k$ *be the corresponding elementary matrices. Then*

$$\mathbf{B} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}.$$

Proof. We will proof by induction. Let $\mathbf{A}_1$ be the matrix obtained by performing r_1 on $\mathbf{A}$. Then by the theorem above, $\mathbf{A}_1 = \mathbf{E}_1\mathbf{A}$. Suppose now $\mathbf{A}_l$ is the matrix obtained by performing row operations $r_1, r_2, ..., r_l$, $l < k$. By the induction hypothesis, $\mathbf{A}_l = \mathbf{E}_l \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$. Let $\mathbf{A}_{l+1}$ be obtained from $\mathbf{A}_l$ by performing row operation r_l. Then by the theorem above,

$$\mathbf{A}_{l+1} = \mathbf{E}_{l+1}\mathbf{A}_l = \mathbf{E}_{l+1}\mathbf{E}_l \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}.$$

The inductive step is shown, and hence, the statement is proved. □

Example.

$$\begin{pmatrix}1&1&1&0\\1&2&3&-1\\2&1&4&2\end{pmatrix}\xrightarrow{R_2+2R_1}\xrightarrow{R_2\leftrightarrow R_3}\begin{pmatrix}1&1&1&0\\2&1&4&2\\3&4&5&-1\end{pmatrix}=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}\begin{pmatrix}1&0&0\\2&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&1&1&0\\1&2&3&-1\\2&1&4&2\end{pmatrix}$$

Question: What if we post-multiply elementary matrices to a matrix?

Lemma. *Elementary matrix* $\mathbf{E}$ *is invertible, and the inverse* $\mathbf{E}^{-1}$ *is also an elementary matrix.*

Proof. We claim that the inverse of the elementary matrix corresponding to

(i) $R_i + aR_j$,

(ii) $R_i \leftrightarrow R_j$, and

(iii) cR_i for $c \neq 0$

is the elementary matrix corresponding to

(i) $R_i - aR_j$,

(ii) $R_i \leftrightarrow R_j$, and

(iii) $\frac{1}{c}R_i$,

respectively. The verification is left to the reader. $\square$

Example. 1. $\begin{pmatrix}1&0&0\\2&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\-2&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0\\-2&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\2&1&0\\0&0&1\end{pmatrix}$

2. $\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$

3. $\begin{pmatrix}1&0&0\\0&1&0\\0&0&5\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1/5\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1/5\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&5\end{pmatrix}$

2.8 Inverses and elementary matrices

Theorem. *A square matrix is invertible if and only if its reduce row-echelon form is the identity matrix.*

Equivalently, a square matrix is singular if and only if it has a REF with a zero row.

Corollary. *A square matrix is invertible if and only if it is a product of elementary matrices.*

Example. 1.

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_2 - R_1} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1 - R_2} \xrightarrow{R_2 - R_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{A} = \mathbf{I},$$

or, $\mathbf{A}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$

2.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{R_3 - R_1} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \xrightarrow{R_3 - R_2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence, $\mathbf{A}$ is singular.

Recall that two matrices $\mathbf{A}$ and $\mathbf{B}$ are row equivalent if one can be obtained from the other by performing elementary row operations. This is equivalent to $\mathbf{B}$ being obtained from $\mathbf{A}$ by pre-multiplying some elementary row operations.

Exercise: Show that $\mathbf{A}$ and $\mathbf{B}$ are row equivalent if and only if $\mathbf{A} = \mathbf{PB}$ for some invertible matrix $\mathbf{P}$.

Algorithm for finding inverses

The theorem and corollary above provides us with a way to check if a matrix is invertible, and also to find the inverse if it exists. Let $\mathbf{A}$ be a square matrix of order n.

Step 1: Form a new $n \times 2n$ matrix $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{I}_n \end{array} \right)$.

Step 2: Reduce the matrix $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{I} \end{array} \right) \longrightarrow \left(\begin{array}{c|c} \mathbf{R} & \mathbf{B} \end{array} \right)$ to its reduced row-echelon form.

Step 3: If $\mathbf{R} \neq \mathbf{I}$, then $\mathbf{A}$ is not invertible. If $\mathbf{R} = \mathbf{I}$, $\mathbf{A}$ is invertible with inverse $\mathbf{A}^{-1} = \mathbf{B}$.

We will explain why $\mathbf{A}^{-1} = \mathbf{B}$. Since $\left(\begin{array}{c|c} \mathbf{R} & \mathbf{B} \end{array} \right)$ is the RREF of $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{I}_n \end{array} \right)$, there an invertible matrix $\mathbf{P}$ such that

$$\left(\begin{array}{c|c} \mathbf{R} & \mathbf{B} \end{array} \right) = \mathbf{P} \left(\begin{array}{c|c} \mathbf{A} & \mathbf{I} \end{array} \right) = \left(\begin{array}{c|c} \mathbf{PA} & \mathbf{PI} \end{array} \right) = \left(\begin{array}{c|c} \mathbf{PA} & \mathbf{P} \end{array} \right).$$

Since $\mathbf{A}$ is invertible, $\mathbf{R} = \mathbf{I}$ and the equation above tells us that $\mathbf{B} = \mathbf{P}$ and $\mathbf{I} = \mathbf{PA} = \mathbf{BA}$, and thus $\mathbf{A}^{-1} = \mathbf{B}$.

Example. Find the inverse of $\begin{pmatrix} 2 & 7 & 1 \\ 1 & 4 & -1 \\ 1 & 3 & 0 \end{pmatrix}$.

$$\left(\begin{array}{ccc|ccc} 2 & 7 & 1 & 1 & 0 & 0 \\ 1 & 4 & -1 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_3} \left(\begin{array}{ccc|ccc} 1 & 3 & 0 & 0 & 0 & 1 \\ 1 & 4 & -1 & 0 & 1 & 0 \\ 2 & 7 & 1 & 1 & 0 & 0 \end{array}\right) \xrightarrow[R_2 - R_1]{R_3 - 2R_1}$$

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 & 0 & -2 \end{array}\right) \xrightarrow{R_2 + R_3} \xrightarrow{\frac{1}{2}R_2} \left(\begin{array}{ccc|ccc} 1 & 3 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1/2 & 1/2 & -3/2 \\ 0 & 1 & 1 & 1 & 0 & -2 \end{array}\right) \xrightarrow[R_1 - 3R_2]{R_3 - R_2}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -3/2 & -3/2 & 11/2 \\ 0 & 1 & 0 & 1/2 & 1/2 & -3/2 \\ 0 & 0 & 1 & 1/2 & -1/2 & -1/2 \end{array}\right).$$

So

$$\begin{pmatrix} 2 & 7 & 1 \\ 1 & 4 & -1 \\ 1 & 3 & 0 \end{pmatrix}^{-1} = \frac{1}{2} \begin{pmatrix} -3 & -3 & 11 \\ 1 & 1 & -3 \\ 1 & -1 & -1 \end{pmatrix}.$$

Appendix for Lecture 3

Theorem. *A square matrix is invertible if and only if its reduce row-echelon form is the identity matrix.*

Proof. Let $\mathbf{R}$ be the reduced row-echelon form of a square matrix $\mathbf{A}$ of order n. Then $\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{PA}$, where $\mathbf{P} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1$, for some elementary matrices $\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_k$. Equivalently,

$$\mathbf{A} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}\mathbf{R}.$$

Note that $\mathbf{R}$ is the identity matrix if and only if it has n pivot columns.

Suppose $\mathbf{R}$ has n pivot columns. Then $\mathbf{R} = \mathbf{I}$, and so

$$\mathbf{A} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1},$$

which shows that $\mathbf{A}$ is invertible since it is a product of invertible matrices.

Suppose $\mathbf{R}$ has less than n pivot columns. Then the last row of $\mathbf{R}$ must be a zero row. Write

$$\mathbf{R} = \begin{pmatrix} \mathbf{Q} \\ \mathbf{0}_{1\times n} \end{pmatrix},$$

for some $(n-1) \times n$ matrix $\mathbf{Q}$. Then for any square matrix $\mathbf{B}$ of order n,

$$\mathbf{PAB} = \mathbf{RB} = \begin{pmatrix} \mathbf{Q} \\ \mathbf{0}_{1\times n} \end{pmatrix} \mathbf{B} = \begin{pmatrix} \mathbf{QB} \\ \mathbf{0}_{1\times n}\mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{QB} \\ \mathbf{0}_{1\times n} \end{pmatrix},$$

which can never be equal to the identity since it has a zero row. So $\mathbf{PA}$ cannot be invertible, and thus $\mathbf{A}$ cannot be invertible. $\square$

Corollary. *A square matrix is invertible if and only if it is a product of elementary matrices.*

Proof. Using the notations from the previous theorem, we have $\mathbf{A} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}\mathbf{R}$, and by a lemma above, each $\mathbf{E}_i^{-1}$ is an elementary matrix, for $i = 1, ..., k$. By the theorem above, if $\mathbf{A}$ is invertible, $\mathbf{R} = \mathbf{I}$, and so $\mathbf{A}$ is a product of elementary matrices.

Conversely, if $\mathbf{A}$ is a product of elementary matrices, then since elementary matrices are invertible, $\mathbf{A}$ is invertible. $\square$

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 4 Notes

References

1. Elementary Linear Algebra: Application Version, Section 1.4-1.6, 2.1-2.2
2. Linear Algebra with Application, Section 2.4-2.5, 3.1-3.2

2.9 Inverse and Linear system

Theorem (Invertibility and homogeneous sytem)**.** *A square matrix* $\mathbf{A}$ *of order* n *is invertible if and only if the homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

Proof. Suppose $\mathbf{A}$ is invertible. Let $\mathbf{u}$ be a solution to the homogeneous linear system $\mathbf{Ax} = \mathbf{0}$. Then $\mathbf{Au} = \mathbf{0}$ and by pre-multiplying both sides of the equation by $\mathbf{A}^{-1}$, we get

$$\mathbf{u} = \mathbf{A}^{-1}\mathbf{0} = \mathbf{0},$$

which shows that the only solution is the trivial one.

Suppose $\mathbf{A}$ is not invertible. Reduce the augmented matrix $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{0} \end{array}\right) \longrightarrow \left(\begin{array}{c|c} \mathbf{R} & \mathbf{0} \end{array}\right)$ to its RREF. Since $\mathbf{A}$ is not invertible, $\mathbf{R}$ is not the identity matrix, and thus must have a non-pivot column. Thus, any general solution of the homogeneous system $\mathbf{Ax} = \mathbf{0}$ must at least 1 parameter, and so the system admits nontrivial solutions. $\square$

Example. 1. We have shown in lecture 3 that $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$ is invertible. Thus, $\mathbf{A}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0}$ has only the trivial solution. Indeed,

$$\mathbf{A} = \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) \xrightarrow{R_2 - R_1} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_3} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

$$\xrightarrow{R_1 - R_2} \xrightarrow{R_2 - R_3} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right),$$

and thus $x = 0, y = 0z = 0$ is the only solution.

2. We have shown in lecture 3 that $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ is singular. Thus, $\mathbf{A}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0}$ should admit non-trivial solutions. Indeed,

$$\mathbf{A} = \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right) \xrightarrow{R_3 - R_1} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right) \xrightarrow{R_3 - R_2} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right),$$

with general solution $x = -s$, $y = 0$, $z = s$.

Next, we will prove the claim that for a square matrix to be invertible, suffice to check that it has either a left or a right inverse.

Theorem. *Let* $\mathbf{A}$ *be a square matrix of order* n. *Suppose* $\mathbf{BA} = \mathbf{I}$ *for some square matrix* $\mathbf{B}$. *Then* $\mathbf{A}$ *is invertible and* $\mathbf{A}^{-1} = \mathbf{B}$.

Proof. Consider the homogeneous system $\mathbf{Ax} = \mathbf{0}$. If $\mathbf{u}$ is a solution to the system, then

$$\mathbf{u} = \mathbf{BAu} = \mathbf{B0} = \mathbf{0}.$$

So the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, and thus $\mathbf{A}$ is invertible. Finally,

$$\mathbf{B} = \mathbf{B}(\mathbf{AA}^{-1}) = (\mathbf{BA})\mathbf{A}^{-1} = \mathbf{A}^{-1}.$$

□

The theorem shows that a left inverse of $\mathbf{A}$ is the inverse of $\mathbf{A}$. Applying transpose to both sides, and using the fact that the inverse of a transpose is the transpose of the inverse, we get that right inverse is also the inverse. That is, if $\mathbf{AB} = \mathbf{I}$, then

$$\mathbf{I} = \mathbf{B}^T\mathbf{A}^T,$$

and so applying the theorem to $\mathbf{A}^T$, we have

$$\mathbf{B}^T = (\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T,$$

and hence $\mathbf{B} = \mathbf{A}^{-1}$

Theorem. *A square matrix* $\mathbf{A}$ *of order* n *is invertible if and only if the system* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution for all* $\mathbf{b}$.

Remark. Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{Ax} = \mathbf{b}$ represent a linear system. Consider the augmented matrix $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{b} \end{array}\right)$. Suppose the system has a unique solution. Let

$$\left(\begin{array}{c|c} \mathbf{A} & \mathbf{b} \end{array}\right) \xrightarrow{r_1}\xrightarrow{r_2} \cdots \xrightarrow{r_k} \left(\begin{array}{c|c} \mathbf{R} & \mathbf{u} \end{array}\right),$$

where $\mathbf{R}$ is the RREF of $\mathbf{A}$. Then necessarily $\mathbf{R}$ has the form

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{0}_{m-n\times n} \end{pmatrix}.$$

Let $\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_k$ be the corresponding elementary matrices. Then

$$\mathbf{u} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{b}$$

is the unique solution to $\mathbf{Ax} = \mathbf{b}$.

Example. 1. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. For any $\mathbf{b} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$,

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & a \\ 0 & 1 & 1 & b \\ 0 & 0 & 1 & c \end{array}\right) \xrightarrow{R_1-R_2}\xrightarrow{R_2-R_3} \left(\begin{array}{ccc|c} 1 & 0 & 0 & a-b \\ 0 & 1 & 0 & b-c \\ 0 & 0 & 1 & c \end{array}\right).$$

That is, $\begin{pmatrix} a-b \\ b-c \\ c \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{A}^{-1}\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is the unique solution to $\mathbf{Ax} = \mathbf{b}$.

2. The matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ is not invertible. Consider $\mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$,

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right) \xrightarrow{R_3-R_1} \xrightarrow{R_3+R_2} \left(\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right).$$

The system is inconsistent, and hence, $\mathbf{Ax} = \mathbf{b}$ has no solution.

We will summarize all the equivalent statements for invertibility of a matrix.

Theorem. *Let* $\mathbf{A}$ *be a square matrix of order* n*. The following statements are equivalent.*

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}$ *has a left inverse.*

(iii) $\mathbf{A}$ *has a right inverse.*

(iv) The reduced row-echelon form of $\mathbf{A}$ *is the identity matrix* $\mathbf{I}_n$*.*

(v) $\mathbf{A}$ *is a product of elementary matrices.*

(vi) The homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(vii) The linear system $\mathbf{Ax} = \mathbf{b}$ *has a unique solution for every* $\mathbf{b}$*.*

2.10 Determinants

We will define the determinant of $\mathbf{A}$ of order n by induction.

1. For $n = 1$, $\mathbf{A} = (a)$, $\det(\mathbf{A}) = a$.

2. For $n = 2$, $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\det(\mathbf{A}) = ad - bc$.

Suppose we have defined the determinant of all square matrices of order $\leq n - 1$. Let $\mathbf{A}$ be a square matrix of order n.

- Define $\mathbf{M}_{ij}$, called the (i, j) matrix minor of $\mathbf{A}$, to be the matrix obtained from $\mathbf{A}$ be deleting the i-th row and j-th column.

  Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ -1 & 1 & 3 \\ 3 & 2 & 1 \end{pmatrix}$. Then

  (i) $\mathbf{M}_{23} = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$,

  (ii) $\mathbf{M}_{12} = \begin{pmatrix} -1 & 3 \\ 3 & 1 \end{pmatrix}$, and

  (iii) $\mathbf{M}_{31} = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix}$.

- For a square matrix of order n, the (i,j)-cofactor of $\mathbf{A}$, denoted as A_{ij}, is the (real) number given by

$$A_{ij} = (-1)^{i+j}\det(\mathbf{M}_{ij}).$$

This definition is well-defined since $\mathbf{M}_{ij}$ is a square matrix of order $n-1$, and by induction hypothesis, the determinant is well defined. Take note of the sign of the (i,j)-entry, $(-1)^{i+j}$. Here's a visualization of the sign of the entries of the matrix

$$\begin{pmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \ddots & \end{pmatrix}.$$

Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ -1 & 1 & 3 \\ 3 & 2 & 1 \end{pmatrix}$. Then

(i) $A_{23} = (-1)^5(2-6) = 4$,

(ii) $A_{12} = (-1)^3(-1-9) = 10$, and

(iii) $A_{31} = (-1)^4(6+1) = 7$.

- The determinant of $\mathbf{A}$ is defined to be

$$\det(\mathbf{A}) = a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n} = \sum_{k=1}^{n} a_{1k}A_{1k}.$$

This is called the cofactor expansion along row 1.

The determinant of $\mathbf{A}$ is also denoted as $\det(\mathbf{A}) = |\mathbf{A}|$.

So for order 3 matrices, the determinant is defined to be

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} = aei - afh - bdi + bfg + cdh - ceg.$$

There is an easy way to compute determinant for order 3 matrices.

$$\det(\mathbf{A}) = \left|\begin{array}{ccc|cc} \mathrm{a} & \mathrm{b} & \mathrm{c} & \mathrm{a} & \mathrm{b} \\ \mathrm{d} & \mathrm{e} & \mathrm{f} & \mathrm{d} & \mathrm{e} \\ \mathrm{g} & \mathrm{h} & \mathrm{i} & \mathrm{g} & \mathrm{h} \end{array}\right|$$

Example. Compute the determinant of $\begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

$$\begin{aligned} \begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} &= \begin{vmatrix} 2 & 6 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix} - 5\begin{vmatrix} 0 & 6 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{vmatrix} - 2\begin{vmatrix} 0 & 2 & 6 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{vmatrix} \\ &= (2) - 5(0) + 0 - 2(0) = 2. \end{aligned}$$

How about if we try compute using cofactor expansion along the first column?

$$\begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1 \begin{vmatrix} 2 & 6 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix} - 0 \begin{vmatrix} 5 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix} + 0 \begin{vmatrix} 5 & 1 & 2 \\ 2 & 6 & 3 \\ 0 & 0 & 1 \end{vmatrix} - 0 \begin{vmatrix} 5 & 1 & 2 \\ 2 & 6 & 3 \\ 0 & 1 & 2 \end{vmatrix} = 2$$

The are the same!

Theorem. *The determinant of a square matrix* $\mathbf{A}$ *of order* n *can be computed by cofactor expansion along any rows or columns,*

$$\det(\mathbf{A}) = \sum_{k=1}^{n} a_{ik} A_{ik} = \sum_{l=1}^{n} a_{lj} A_{lj}.$$

The proof requires knowledge of the symmetric groups, which is beyond the scope of this module.

Example. Compute the determinant of $\begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}$.

Cofactor expansion along the first column.

$$\begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1 \begin{vmatrix} 2 & 6 & 3 \\ 0 & 1 & 2 \\ 0 & 1 & 1 \end{vmatrix} - 0 \begin{vmatrix} 5 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 1 & 1 \end{vmatrix} + 0 \begin{vmatrix} 5 & 1 & 2 \\ 2 & 6 & 3 \\ 0 & 1 & 1 \end{vmatrix} - 0 \begin{vmatrix} 5 & 1 & 2 \\ 2 & 6 & 3 \\ 0 & 1 & 2 \end{vmatrix} = 2(1-2) = -2.$$

Corollary. *The determinant of* $\mathbf{A}$ *and* $\mathbf{A}^T$ *are equal,*

$$\det(\mathbf{A}) = \det(\mathbf{A}^T).$$

This statement is proved by induction on the size n of the matrix $\mathbf{A}$, and using fact that cofactor expansion along first row of $\mathbf{A}$ is equal to cofactor expansion along the first column of $\mathbf{A}^T$.

Corollary. *If a square matrix* $\mathbf{A}$ *has a zero row or column, then* $\det(\mathbf{A}) = 0$.

Proof. Cofactor expand along the zero row or column. □

Corollary. *The determinant of a triangular matrix is the multiplication of the diagonal entries. If* $\mathbf{A} = (a_{ij})_n$ *is a triangular matrix, then*

$$\det(\mathbf{A}) = a_{11}a_{22} = \cdots a_{nn} = \Pi_{k=1}^{n} a_{ii}.$$

Sketch of proof:

Upper triangular matrix , cofactor expand along first column,

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

Lower triangular matrix, cofactor expand along the first row,

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

In general, we can always reduce a square matrix $\mathbf{A}$ to a triangular matrix $\mathbf{R}$ (for example, REF). Then $\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$ for some elementary matrices. If there is a relationship between the determinant of $\mathbf{A}$ and $\mathbf{R}$, this will make computation easier, since the determinant of $\mathbf{R}$ is just the product of the diagonals. We will state the results, but omit the proof.

Theorem. *If $\mathbf{A}$ has 2 identical row or columns, then $\det(\mathbf{A}) = 0$.*

Theorem (Determinant and elementary row operations)**.** *If $\mathbf{B}$ is obtained from $\mathbf{A}$ be performing the following elementary row operations*

(i) $\mathbf{A} \xrightarrow{R_i+aR_j} \mathbf{B}$,

(ii) $\mathbf{A} \xrightarrow{R_i\leftrightarrow R_j} \mathbf{B}$,

(iii) $\mathbf{A} \xrightarrow{cR_j} \mathbf{B}$, $c \neq 0$,

then

(i) $\det(\mathbf{B}) = \det(\mathbf{A})$,

(ii) $\det(\mathbf{B}) = -\det(\mathbf{A})$,

(iii) $\det(\mathbf{B}) = c\det(\mathbf{A})$,

respectively.

Theorem (Determinant of elementary matrices)**.** *Suppose $\mathbf{E}$ is an elementary matrix corresponding to the row operation*

(i) $R_i + cR_j$, *then* $\det(\mathbf{E}) = 1$.

(ii) $R_i \leftrightarrow R_j$, *then* $\det(\mathbf{E}) = -1$.

(iii) cR_j, *then* $\det(\mathbf{E}) = c$.

Theorem (Determinant of row equivalent matrices)**.** *Let $\mathbf{A}$ and $\mathbf{R}$ be square matrices such that*

$$\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$$

for some elementary matrices $\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_k$. Then

$$\det(\mathbf{R}) = \det(\mathbf{E}_k) \cdots \det(\mathbf{E}_2)\det(\mathbf{E}_1)\det(\mathbf{A}).$$

So if $\mathbf{R}$ is a REF of $\mathbf{A}$, that is, $\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$, then

$$\begin{aligned}
\det(\mathbf{R}) &= \det(\mathbf{E}_k) \cdots \det(\mathbf{E}_2)\det(\mathbf{E}_1)\det(\mathbf{A}) \\
\Rightarrow \det(\mathbf{A}) &= \det(\mathbf{E}_1)^{-1}\det(\mathbf{E}_2)^{-1} \cdots \det(\mathbf{E}_k)^{-1}\det(\mathbf{R}) \\
&= \det(\mathbf{E}_1)^{-1}\det(\mathbf{E}_2)^{-1} \cdots \det(\mathbf{E}_k)^{-1}\Pi_{k=1}^{n} r_{ii},
\end{aligned}$$

where for $i = 1, ..., n$, r_{ii} is the i-th diagonal entry of $\mathbf{R}$. Since the determinant of elementary matrices can be obtained easily (recall that inverse of elementary matrix is an elementary matrix), the determinant of $\mathbf{A}$ is thus easily found.

Example.

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 3 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix} \xrightarrow[\mathbf{E}_2:R_3-2R_4]{\mathbf{E}_1:R_1-2R_2} \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & -3 & -6 \\ 0 & 1 & 2 & 3 \end{pmatrix} \xrightarrow[\mathbf{E}_4:-\frac{1}{3}R_3]{\mathbf{E}_3:R_4-R_1} \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 4 \end{pmatrix} \xrightarrow{\mathbf{E}_5:R_4-R_3}$$
$$\begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 \end{pmatrix} \xrightarrow{\mathbf{E}_6:R_1\leftrightarrow R_2} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

So

$$\begin{aligned}
2 &= \det(\mathbf{E}_6)\det(\mathbf{E}_5)\det(\mathbf{E}_4)\det(\mathbf{E}_3)\det(\mathbf{E}_2)\det(\mathbf{E}_1)\det(\mathbf{A}) \\
&= (-1)(1)(1)(-\frac{1}{3})(1)(1)\det(\mathbf{A})
\end{aligned}$$

$\Rightarrow \det(\mathbf{A}) = 6$.

Theorem (Determinant of product of matrices)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be square matrices of the same size. Then*

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

By induction, we get

$$\det(\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_k) = \det(\mathbf{A}_1)\det(\mathbf{A}_2) \cdots \det(\mathbf{A}_k).$$

Corollary (Determinant of inverse)**.** *If* $\mathbf{A}$ *is invertible, then*

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

Proof. Since the identity matrix $\mathbf{I}$ is a triangular matrix, $\det(\mathbf{I}) = 1$. Then

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

So $\det(\mathbf{A})^{-1} = \det(\mathbf{A}^{-1})$. $\square$

Corollary (Equivalence of invertibility and determinant)**.** *A square matrix* $\mathbf{A}$ *is invertible if and only if* $\det(\mathbf{A}) \neq 0$.

Proof. Recall that $\mathbf{A}$ is invertible if and only if its RREF $\mathbf{R}$ is the identity. If $\mathbf{R}$ is not the identity, then it will have at least a zero row and thus $\det(\mathbf{R}) = 0$. Hence

$$\begin{aligned}\det(\mathbf{A}) &= \det(\mathbf{E}_1)^{-1}\det(\mathbf{E}_2)^{-1}\cdots\det(\mathbf{E}_k)^{-1}\det(\mathbf{R})\\ &= \det(\mathbf{E}_1)^{-1}\det(\mathbf{E}_2)^{-1}\cdots\det(\mathbf{E}_k)^{-1}0 = 0.\end{aligned}$$

On the other hand, if $\mathbf{R} = \mathbf{I}$, then $\det(\mathbf{A}) = \det(\mathbf{E}_1)^{-1}\det(\mathbf{E}_2)^{-1}\cdots\det(\mathbf{E}_k)^{-1}$, and since the determinant of elementary matrices are nonzero, $\det(\mathbf{A}) \neq 0$. $\square$

We will add this equivalence of invertibility to list.

Theorem. *Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.*

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}$ *has a left inverse.*

(iii) $\mathbf{A}$ *has a right inverse.*

(iv) The reduced row-echelon form of $\mathbf{A}$ *is the identity matrix* $\mathbf{I}_n$.

(v) $\mathbf{A}$ *is a product of elementary matrices.*

(vi) The homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(vii) The linear system $\mathbf{Ax} = \mathbf{b}$ *has a unique solution for every* $\mathbf{b}$.

(viii) The determinant of $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$.

Corollary (Determinant of scalar multiplication)**.** *For any square matrix* $\mathbf{A}$ *of order n and scalar $c \in \mathbb{R}$,*

$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}).$$

We will present two proofs.

Proof 1: $c\mathbf{A} \xrightarrow{\frac{1}{c}R_1}\xrightarrow{\frac{1}{c}R_2} \cdots \xrightarrow{\frac{1}{c}R_k} \mathbf{A}$. The elementary operations are multiplying a row by $\frac{1}{c}$, and since the determinant of the corresponding elementary matrices are $\frac{1}{c}$, $\det(A) = (\frac{1}{c})^n \det c\mathbf{A}$. Hence, $c^n\det(\mathbf{A}) = \det(c\mathbf{A})$.

Proof 2: Observe that $c\mathbf{A} = c\mathbf{IA} = \operatorname{diag}\{c, ..., c\}\mathbf{A}$. So

$$\det(c\mathbf{A}) = \det(\operatorname{diag}\{c, ..., c\})\det(\mathbf{A}) = c^n\det(\mathbf{A}).$$

Example. Let $\mathbf{A} = \begin{pmatrix}1 & 5 & 1 & 2\\ 0 & 2 & 6 & 3\\ 0 & 0 & 1 & 2\\ 0 & 0 & 1 & 1\end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix}1 & 0 & 1 & 1\\ 0 & 1 & 1 & -1\\ 0 & 0 & 1 & 2\\ 0 & 0 & 0 & 3\end{pmatrix}$.

$$\mathbf{A} \xrightarrow{R_4 - R_3} \begin{pmatrix}1 & 5 & 1 & 2\\ 0 & 2 & 6 & 3\\ 0 & 0 & 1 & 2\\ 0 & 0 & 0 & -1\end{pmatrix}.$$

So $\det(\mathbf{A}) = -2$

1. $\det(3\mathbf{A}) = 3^4(-2) = -162$.
2. $\det(3\mathbf{A}\mathbf{B}^{-1}) = 3^4(-2)(-3)^{-1} = -18$.
3. $\det((3\mathbf{B})^{-1}) = (3^4 \times 3)^{-1} = 3^{-5}$.

Appendix for Lecture 4

Theorem. *A square matrix* $\mathbf{A}$ *of order* n *is invertible if and only if the system* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *has a unique solution for all* $\mathbf{b}$.

Proof. Suppose $\mathbf{A}$ is invertible. Then for any $\mathbf{b}$, $\mathbf{u} = \mathbf{A}^{-1}b$ is a solution to the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. Suppose $\mathbf{u}$ and $\mathbf{v}$ are two solution, $\mathbf{A}(\mathbf{u}-\mathbf{v}) = \mathbf{A}\mathbf{u} - \mathbf{A}\mathbf{v} = \mathbf{b} - \mathbf{b} = \mathbf{0}$. This shows that $(\mathbf{u}-\mathbf{v})$ is a solution to the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$. Since $\mathbf{A}$ is invertible, the homogeneous system has only the trivial solution. Hence, $\mathbf{u} = \mathbf{v}$.

Suppose now $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b}$. In particular, let $\mathbf{b} = \mathbf{e}_i$ for $i = 1, .., n$, where $\mathbf{e}_i$ is the $n \times 1$ matrix with 1 in the i-th row, and 0 everywhere else. Let $\mathbf{b}_i$ be the solution to $\mathbf{A}\mathbf{x} = \mathbf{e}_i$ for $i = 1, ..., n$, and let $\mathbf{B} = \begin{pmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{pmatrix}$. Then by the block multiplication,

$$\mathbf{AB} = \mathbf{A}\begin{pmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{pmatrix} = \begin{pmatrix}\mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 & \cdots & \mathbf{A}\mathbf{b}_n\end{pmatrix} = \begin{pmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n\end{pmatrix} = \mathbf{I}.$$

This shows that $\mathbf{B}$ is a right inverse, and thus the inverse of $\mathbf{A}$. Hence, $\mathbf{A}$ is invertible. □

Observe that the first part of the proof can be adapted to show that for any size $m \times n$ matrix $\mathbf{A}$, if $\mathbf{A}$ has a left inverse, then the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent, the solution is not unique if $m \neq n$.

We will now proof the two claims in lecture 1

Theorem (Uniqueness of RREF)**.** *Suppose* $\mathbf{R}$ *and* $\mathbf{S}$ *are two reduced row-echelon forms of* $\mathbf{A}$*, then* $\mathbf{R} = \mathbf{S}$.

Proof. First note that there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{PR} = \mathbf{S}. \tag{1}$$

This is because $\mathbf{A}$ is row equivalent to $\mathbf{R}$ and $\mathbf{S}$, and so there are invertible matrices $\mathbf{P}_1, \mathbf{P}_2$ such that $\mathbf{A} = \mathbf{P}_1\mathbf{R}$ and $\mathbf{A} = \mathbf{P}_2\mathbf{S}$. Let $\mathbf{P} = \mathbf{P}_2^{-1}\mathbf{P}_1$. We will prove by induction on the numbers of rows n of $\mathbf{R}$ and $\mathbf{S}$.

Suppose $n = 1$. Then $\mathbf{R}$, $\mathbf{S}$ are row matrices and $\mathbf{P}$ is a nonzero real number. Since the leading entries of $\mathbf{R}$ and $\mathbf{S}$ must be 1, by the equation (1), $\mathbf{P} = 1$. So $\mathbf{R} = \mathbf{S}$.

Now suppose $n > 1$. Write $\mathbf{R} = \begin{pmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n\end{pmatrix}$ and $\mathbf{S} = \begin{pmatrix}\mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_n\end{pmatrix}$. By equation (1), we have

$$\mathbf{P}\mathbf{r}_j = \mathbf{s}_j, \tag{2}$$

for $j = 1, ..., n$. Since $\mathbf{P}$ is invertible, by the theorem above, $\mathbf{R}$ and $\mathbf{S}$ must have the same zero columns. By deleting the zero columns and forming a new matrix, we may assume that $\mathbf{R}$ and $\mathbf{S}$ has no zero columns. With this assumption, and the fact that $\mathbf{R}$ and $\mathbf{S}$ are in RREF, necessarily the first column of both $\mathbf{R}$ and $\mathbf{S}$ must have 1 in the first entry and 0 everywhere else. By the equation (1), the first column of $\mathbf{P}$ also have 1 in the first entry and zero everywhere else. So we write $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{P}$ in is submatrices,

$$\mathbf{P} = \begin{pmatrix} 1 & \mathbf{p}' \\ 0 & \\ \vdots & \mathbf{P}' \\ 0 & \end{pmatrix}, \mathbf{R} = \begin{pmatrix} 1 & \mathbf{r}' \\ 0 & \\ \vdots & \mathbf{R}' \\ 0 & \end{pmatrix}, \text{ and } \mathbf{S} = \begin{pmatrix} 1 & \mathbf{s}' \\ 0 & \\ \vdots & \mathbf{S}' \\ 0 & \end{pmatrix},$$

where $\mathbf{p}', \mathbf{r}', \mathbf{s}'$ are row matrices. By the equation (1) and block multiplication, we have $\mathbf{P}'\mathbf{R}' = \mathbf{S}'$. Note that $\mathbf{P}'$ is invertible. Since $\mathbf{R}$ and $\mathbf{S}$ are in RREF, $\mathbf{R}'$ and $\mathbf{S}'$ are in RREF too. Hence, by the induction hypothesis, $\mathbf{R}' = \mathbf{S}'$. We are left to show that $\mathbf{r}' = \mathbf{s}'$. Since $\mathbf{R}' = \mathbf{S}'$, and both $\mathbf{R}$ and $\mathbf{S}$ are in RREF, $\mathbf{R}$ and $\mathbf{S}$ must have the same pivot columns, say columns $i_1, i_2, ..., i_r$. In these columns, the entries of $\mathbf{r}'$ and $\mathbf{s}'$ must be zero. For the nonzero entries, by equation (2), and the fact that the entries of the columns agree from second row onward, the entries in the first row of each column agrees too, that is $\mathbf{r}' = \mathbf{s}'$ too. Thus the inductive step in complete, and the statement is proven. □

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

Proof. Suppose $\mathbf{A}$ and $\mathbf{B}$ has the same RREF $\mathbf{R}$. Then there are invertible matrices $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{PA} = \mathbf{R}$ and $\mathbf{QB} = \mathbf{R}$. Then

$$\mathbf{Q}^{-1}\mathbf{PA} = \mathbf{Q}^{-1}\mathbf{R} = \mathbf{B}.$$

Since $\mathbf{Q}^{-1}\mathbf{P}$ is invertible, it can be written as a product of elementary matrices, and so $\mathbf{A}$ is row equivalent to $\mathbf{B}$.

Suppose now $\mathbf{A}$ is row equivalent to $\mathbf{B}$. Let $\mathbf{P}$ be an invertible matrix such that $\mathbf{PA} = \mathbf{B}$. Let $\mathbf{R}$ be the RREF of $\mathbf{A}$ and $\mathbf{S}$ be the RREF of $\mathbf{B}$. Then $\mathbf{R} = \mathbf{UA}$ and $\mathbf{S} = \mathbf{VB}$ for some invertible matrices $\mathbf{U}$ and $\mathbf{V}$. Then

$$\mathbf{VPU}^{-1}\mathbf{R} = \mathbf{VPA} = \mathbf{VB} = \mathbf{S},$$

which shows that $\mathbf{R}$ is row equivalent to $\mathbf{S}$. By the uniqueness of RREF, $\mathbf{R} = \mathbf{S}$. □

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 5 Notes

References

1. Elementary Linear Algebra: Application Version, Section 3.1-3.2, 3.4, 6.1-6.2
2. Linear Algebra with Application, Section 4.1-4.2

3 Vectors in Euclidean Spaces

3.1 Introduction to Euclidean Spaces

A (real) n-vector (or vector) is a collection of n ordered real numbers,

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \text{ where } v_i \in \mathbb{R} \text{ for } i = 1, ..., n.$$

The collection of all n-vectors is called a vector space, which in this course, is synonymous to the Euclidean n-space. It is denoted as $\mathbb{R}^n$,

$$\mathbb{R}^n = \left\{ \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \middle| \, v_i \in \mathbb{R} \text{ for } i = 1, ..., n. \right\}$$

Remark.

1. We can similarly define complex n-vectors, denoted as $\mathbb{C}^n$. Later, we will also define vector-valued functions, which are vectors with functions entries.
2. Strictly speaking, the above is called a column vector. A column vector is a $n \times 1$ matrix (column matrix). A row vector is a $1 \times n$ matrix (row matrix), $\mathbf{v} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$.
3. The space of all column vectors is isomorphic to the space of row vectors (meaning they have all the same properties we care about, except for only how they look). So in most subjects, they are both denoted as $\mathbb{R}^n$ (in fact any real n-dimensional space is isomorphic to $\mathbb{R}^n$, but we digressed).
4. In this course, if not explicitly mentioned, a vector would mean a column vector.
5. Geometrically, we can think of $\mathbb{R}^2$ as a plane, and $\mathbb{R}^3$ as our physical universe (without considering time).
6. A vector $\mathbf{v}$ can be interpreted as an arrow, with the tail placed at the origin $\mathbf{0}$, and the head of the arrow at $\mathbf{v}$, or it could represent a point in the Euclidean n-space.

3.2 Vectors addition and Scalar Multiplication

Since vectors are (column or row) matrices, the properties of matrix addition and scalar multiplication holds for vectors. For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and scalars $a, b \in \mathbb{R}$,

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$,

(ii) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$,

(iii) $\mathbf{0} + \mathbf{v} = \mathbf{v}$,

(iv) $\mathbf{v} - \mathbf{v} = \mathbf{0}$,

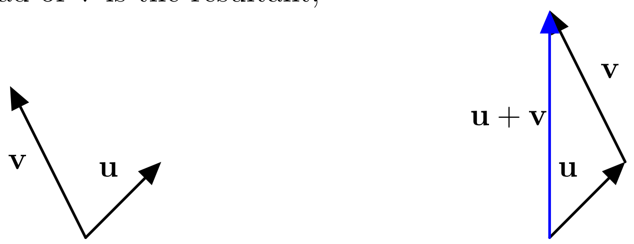
(v) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$,

(vi) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$,

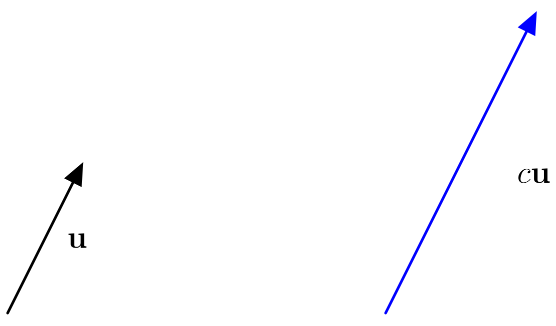
(vii) $(ab)\mathbf{u} = a(b\mathbf{u})$,

(viii) if $a\mathbf{u} = \mathbf{0}$, then $a = 0$ or $\mathbf{u} = 0$.

Geometrically, adding $\mathbf{u}$ to $\mathbf{v}$ can be visualized by joining the tail of $\mathbf{v}$ at the head of $\mathbf{u}$, and the head of $\mathbf{v}$ is the resultant,



and a scalar multiple of a vector is scaling (stretching or compressing) the vector,



3.3 Solutions to linear systems (revisit)

There are 2 ways to express subsets of $\mathbb{R}^n$, implicit and explicit form.

- Implicit form: $\left\{ \mathbf{v} \in \mathbb{R}^n \;\middle|\; \mathbf{v} \text{ fulfills some conditions.} \right\}$,
- Explicit form: $\left\{ \mathbf{u} + s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots s_k\mathbf{v}_k \;\middle|\; s_1, s_2, ..., s_k \in \mathbb{R} \right\}$.

Example. 1. Implicit form: $\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \;\middle|\; x = -y, z = 1 \right\}$ = Explicit form: $\left\{ \begin{pmatrix} s \\ -s \\ 1 \end{pmatrix} \;\middle|\; s \in \mathbb{R} \right\}$

2. Implicit form: $\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \;\middle|\; x = y = z \right\}$ = Explicit form: $\left\{ s \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \;\middle|\; s \in \mathbb{R} \right\}$

Consider the linear system

$$\begin{array}{rcrcrcl} 3x & + & 2y & - & z & = & 1 \\ & & y & - & z & = & 0 \end{array}$$

The implicit form is

$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \;\middle|\; 3x + 2y - z = 1, y - z = 0 \right\}$$

A general solution is $x = \frac{1}{3}(1 - s)$, $y = s$, $z = s$, $s \in \mathbb{R}$. We may present this in vector form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{3}(1-s) \\ s \\ s \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + s\begin{pmatrix} -1/3 \\ 1 \\ 1 \end{pmatrix}, s \in \mathbb{R}.$$

The set containing all possible s is then called the <u>solution set</u> of the linear system, and is denoted as

$$\left\{ \tfrac{1}{3}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + s\begin{pmatrix} -1/3 \\ 1 \\ 1 \end{pmatrix} \;\middle|\; s \in \mathbb{R} \right\} \subseteq \mathbb{R}^3.$$

In order words, from the linear system, we get the implicit form of the solution set, and the explicit form is derived from a general solution.

Remark. Note that even though general solutions are not unique, the solution set is, different expressions of general solutions gives the same solution set for a linear system.

3.4 Dot Product

How do we multiply vectors? Given two (column) vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we cannot multiply them since their size don't match. However, if we transpose one of the vectors, we can multiply,

1. $\mathbf{u}\mathbf{v}^T = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1v_1 & u_1v_2 & \cdots & u_1v_n \\ u_2v_1 & u_2v_2 & \cdots & u_2v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nv_1 & u_nv_2 & \cdots & u_nv_n \end{pmatrix} = (u_iv_j)_n$, or

2. $\mathbf{u}^T\mathbf{v} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \sum_{i=1}^{n} u_iv_i.$

The first multiplication is known as outer product, denoted as $\mathbf{u} \otimes \mathbf{v}$, and the second is known as inner product, or dot product, denoted as $\mathbf{u} \cdot \mathbf{v}$. In this course, we will only be discussing inner product.

Example. 1. $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = 2 + 4 - 2 = 4.$

2. $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1 + 0 - 1 = 0.$

3. $\begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \end{pmatrix} = 2 - 6 = -4.$

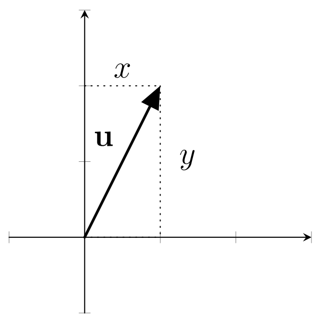
The norm of a vector $\mathbf{u} \in \mathbb{R}^n$ is defined to be the square root of the inner product of $\mathbf{u}$ with itself, and is denoted as $\|\mathbf{u}\|$,

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}.$$

Geometric meaning of norm. The distance between the point $\begin{pmatrix} x \\ y \end{pmatrix}$ and the origin in $\mathbb{R}^2$ is given by

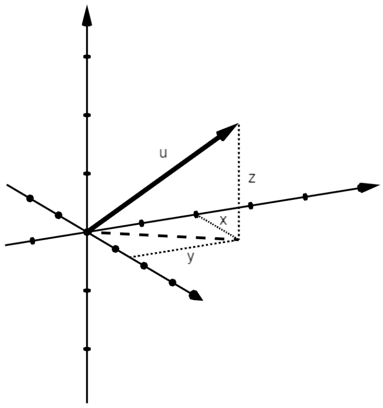
$$\text{distance} = \sqrt{x^2 + y^2} = \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|.$$

That is, in $\mathbb{R}^2$, the norm of a vector can be interpreted as its distance from the origin.



Similarly, in $\mathbb{R}^3$, the distance of a vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ to the origin is

$$\text{distance} = \sqrt{x^2 + y^2 + z^2} = \left\| \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\|.$$



We may thus generalize and define the distance between a vector $\mathbf{v}$ and the origin in $\mathbb{R}^n$ is its norm, $\|\mathbf{v}\|$.

Observe that the distance between two vector $\mathbf{v} = (v_i)$ and $\mathbf{u} = (u_i)$ is

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2} = \|\mathbf{u} - \mathbf{v}\|.$$

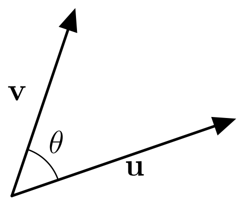
Example. 1. $\left\| \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \right\| = \sqrt{1^2 + 2^2 + (-1)^2} = \sqrt{6}.$

2. $d\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \end{pmatrix} \right) = \left\| \begin{pmatrix} 1-0 \\ 3-5 \end{pmatrix} \right\| = \sqrt{1^2 + (-2)^2} = \sqrt{5}.$

The angle between two nonzero vectors, $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ is the number θ with $0 \leq \theta \leq \pi$ such that

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

This is a natural definition because once again, in $\mathbb{R}^2$, this is indeed the definition of the trigonometric function cosine.



Theorem (Properties of dot product and norm)**.** *Let* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *be vectors and* $a, b, c \in \mathbb{R}$ *be scalars.*

(i) (Symmetric) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

(ii) (Scalar multiplication) $c\mathbf{u} \cdot \mathbf{v} = (c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v})$.

(iii) (Distribution) $\mathbf{u} \cdot (a\mathbf{v} + b\mathbf{w}) = a\mathbf{u} \cdot \mathbf{v} + b\mathbf{u} \cdot \mathbf{w}$.

(iv) (Positive definite) $\mathbf{u} \cdot \mathbf{u} \geq 0$ *with equality if and only if* $\mathbf{u} = \mathbf{0}$.

(v) $\|c\mathbf{u}\| = |c|\|\mathbf{u}\|$.

Proof. Let $\mathbf{u} = (u_i)$, $\mathbf{v} = (v_i)$, and $\mathbf{w} = (w_i)$.

(i) $\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i = \sum_{i=1}^n v_i u_i = \mathbf{v} \cdot \mathbf{u}$. Alternatively, since $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, it is symmetric. So $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v} = (\mathbf{u}^T\mathbf{v})^T = \mathbf{v}^T\mathbf{u} = \mathbf{v} \cdot \mathbf{u}$.

(ii) $c\sum_{i=1}^n u_i v_i = \sum_{i=1}^n (cu_i)v_i = \sum_{i=1}^n u_i(cv_i)$.

(iii) $\sum_{i=1}^n u_i(av_i + bw_i) = \sum_{i=1}^n (au_iv_i + bu_iw_i) = a\sum_{i=1}^n u_iv_i + b\sum_{i=1}^n u_iw_i$.

(iv) $\mathbf{u} \cdot \mathbf{u} = \sum_{i=1}^n u_i^2 \geq 0$ since $u_i \in \mathbb{R}$ are real numbers. It is clear that a sum of square of real numbers is equal to 0 if and only if all the numbers are 0.

(v) $\|c\mathbf{u}\| = \sqrt{\sum_{i=1}^n (cu_i^2)} = \sqrt{c^2}\sqrt{\sum_{i=1}^n u_i^2} = |c|\|\mathbf{u}\|$.

$\square$

A vector $\mathbf{u}$ is a unit vector if $\|\mathbf{u}\| = 1$. We can normalize every nonzero vector by multiplying it by the reciprocal of its norm,

$$\mathbf{u} \longrightarrow \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Then

$$\left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\right) \cdot \left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\right) = \frac{\mathbf{u} \cdot \mathbf{u}}{\|\mathbf{u}\|^2} = 1.$$

Example. Let $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$. We have computed that $\|\mathbf{u}\| = \sqrt{6}$ and so

$$\frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

is a unit vector.

We say that vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal if

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

Suppose $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal.

- Case 1: Either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$.
- Case 2: Otherwise,

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = 0$$

tells us that $\theta = \frac{\pi}{2}$, that is, $\mathbf{u}$ and $\mathbf{v}$ are perpendicular.

That is, $\mathbf{u}, \mathbf{v}$ are orthogonal if and only if either one of them is the zero vector or they are perpendicular to each other.

Example.

$$\begin{pmatrix}1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\0\\-1\end{pmatrix}=0,\quad \begin{pmatrix}1\\0\\0\end{pmatrix}\cdot\begin{pmatrix}0\\1\\0\end{pmatrix}=0,\quad \begin{pmatrix}1\\1\\0\end{pmatrix}\cdot\begin{pmatrix}0\\0\\1\end{pmatrix}=0.$$

Exercise: Suppose $\mathbf{u},\mathbf{v}$ are orthogonal. Show that for any $s,t\in\mathbb{R}$ scalars, $s\mathbf{u},t\mathbf{v}$ are also orthogonal.

A set $S=\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_k\}\subseteq\mathbb{R}^n$ of vectors is orthogonal if $\mathbf{v}_i\cdot\mathbf{v}_j=0$ for every $i\neq j$, that is, vectors in S are pairwise orthogonal. A set $S=\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_k\}\subseteq\mathbb{R}^n$ of vectors is orthonormal if

$$\mathbf{v}_i\cdot\mathbf{v}_j=\begin{cases}0 & \text{if } i\neq j,\\ 1 & \text{if } i=j.\end{cases}$$

That is, S is orthogonal, and all the vectors are unit vectors.

Remark. Every orthogonal set of nonzero vectors can be normalized to an orthonormal set.

Example.

1. $S=\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$ is an orthonormal set.
2. $S=\left\{\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}0\\1\\-1\end{pmatrix},\begin{pmatrix}-1\\0\\1\end{pmatrix}\right\}$ is not an orthogonal set since $\begin{pmatrix}0\\1\\-1\end{pmatrix}\cdot\begin{pmatrix}-1\\0\\1\end{pmatrix}=-1\neq 0$.
3. $S=\left\{\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}1\\0\\-1\end{pmatrix},\begin{pmatrix}1\\-2\\1\end{pmatrix}\right\}$ is an orthogonal but not orthonormal set. It can be normalized to an orthonormal set

$$\left\{\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix},\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix},\frac{1}{\sqrt{6}}\begin{pmatrix}1\\-2\\1\end{pmatrix}\right\}.$$

4. $S=\left\{\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\\0\end{pmatrix},\begin{pmatrix}-1/\sqrt{2}\\1/\sqrt{2}\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$ is an orthonormal set.
5. $S=\left\{\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}1\\-1\\0\end{pmatrix},\begin{pmatrix}0\\0\\0\end{pmatrix}\right\}$ is an orthogonal set but it cannot be normalized to an orthonormal set since it contains the zero vector.

Appendix to Lecture 5

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ be a subset. Form a $n \times k$ matrix $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$, where the columns of $\mathbf{A}$ are the vectors in S. Consider the product

$$\begin{aligned}
\mathbf{A}^T\mathbf{A} &= \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \cdots \\ \mathbf{u}_k^T \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \cdots & \mathbf{u}_1^T\mathbf{u}_k \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \cdots & \mathbf{u}_2^T\mathbf{u}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_k^T\mathbf{u}_1 & \mathbf{u}_k^T\mathbf{u}_2 & \cdots & \mathbf{u}_k^T\mathbf{u}_k \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \mathbf{u}_1 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_k \\ \mathbf{u}_2 \cdot \mathbf{u}_1 & \mathbf{u}_2 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_2 \cdot \mathbf{u}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_k \cdot \mathbf{u}_1 & \mathbf{u}_k \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_k \cdot \mathbf{u}_k \end{pmatrix}.
\end{aligned}$$

Then S is orthogonal if and only if the product $\mathbf{A}^T\mathbf{A}$ is a diagonal matrix, and S is orthonormal if and only if $\mathbf{A}^T\mathbf{A}$ is the identity matrix of order k.

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 6 Notes

References

1. Elementary Linear Algebra: Application Version, Section 4.2
2. Linear Algebra with Application, Section 5.1, 6.2

3.5 Linear Combination and Linear Span

A linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k \in \mathbb{R}^n$ is

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k, \text{ for some } c_1, c_2, ..., c_k \in \mathbb{R}.$$

Remark. One can think of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ as the possible directions, and $c_1, c_2, ..., c_k$ as the amount of units to walk in the respective directions.

The collection of all possible linear combinations of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k \in \mathbb{R}^n$ is call the span,

$$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = \left\{ \ c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \ \middle| \ c_1, c_2, ..., c_k \in \mathbb{R} \ \right\}.$$

It is straight forward to compute a linear combination of a vectors. We may ask the reverse question. Is a given vector $\mathbf{v}$ a linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$? Equivalently, whether $\mathbf{v} \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$. This is asking for whether we are able to find real numbers $c_1, c_2, ..., c_k$ such that

$$\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k.$$

Example. Is $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ a linear combination of $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$?

We are looking for $a, b, c \in \mathbb{R}$, if possible, such that

$$a\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + b\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Comparing the entries, we end up with a linear system

$$\left\{ \begin{array}{ccccccc} a & + & b & + & c & = & 1 \\ a & - & b & + & 2c & = & 2 \\ a & & & + & c & = & 3 \end{array} \right.$$

Then

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & 2 \\ 1 & 0 & 1 & 3 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -3 \end{array} \right).$$

Observe that the constant is the vector $\mathbf{v}$, and the columns of the coefficient matrix (the left hand side) are the vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$.

So the system is consistent, and we have

$$6\begin{pmatrix}1\\1\\1\end{pmatrix}-2\begin{pmatrix}1\\-1\\0\end{pmatrix}-3\begin{pmatrix}1\\2\\1\end{pmatrix}=\begin{pmatrix}1\\2\\3\end{pmatrix}.$$

Theorem. *Let $\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ be vectors in $\mathbb{R}^n$. Then $\mathbf{v} \in span\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ if and only if the linear system associated to the augmented matrix*

$$\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v} \end{array}\right)$$

is consistent.

So this shows that a vector $\mathbf{v} \in \mathbb{R}^n$ is in $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ if and only if we can find a vector $\mathbf{u} = \begin{pmatrix}c_1\\c_2\\\vdots\\c_k\end{pmatrix} \in \mathbb{R}^k$ such that

$$\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}\begin{pmatrix}c_1\\c_2\\\vdots\\c_k\end{pmatrix} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{v}.$$

Summarising, we have the following statement.

Corollary. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ be a subset and $\mathbf{A} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}$ be a $n \times k$ matrix whose columns are the vectors in S. Then $\mathbf{v} \in \mathbb{R}^n$ if and only if there is a $\mathbf{u} \in \mathbb{R}^k$ such that $\mathbf{A}\mathbf{u} = \mathbf{v}$, that is, the system $\mathbf{A}\mathbf{x} = \mathbf{v}$ is consistent.*

Example.

1. Let $\mathbf{u} = \begin{pmatrix}4\\-3\\0\end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix}1\\-1\\0\end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix}1\\2\\1\end{pmatrix}$.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 4\\ 1 & -1 & 2 & -3\\ 1 & 0 & 1 & 0 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & -1\\ 0 & 1 & 0 & 4\\ 0 & 0 & 1 & 1 \end{array}\right).$$

   This means that $\begin{pmatrix}1 & 1 & 1\\1 & -1 & 2\\1 & 0 & 1\end{pmatrix}\begin{pmatrix}-1\\4\\1\end{pmatrix} = \begin{pmatrix}4\\-3\\0\end{pmatrix} = -\begin{pmatrix}1\\1\\1\end{pmatrix} + 4\begin{pmatrix}1\\-1\\0\end{pmatrix} + \begin{pmatrix}1\\2\\1\end{pmatrix} = \begin{pmatrix}4\\-3\\0\end{pmatrix}$, and hence $\mathbf{u} \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

2. Let $\mathbf{v} = \begin{pmatrix}1\\2\\3\end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix}1\\0\\1\end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix}0\\1\\-1\end{pmatrix}$.

$$\left(\begin{array}{cc|c} 1 & 0 & 1\\ 0 & 1 & 2\\ 1 & -1 & 3 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & 1\\ 0 & 1 & 2\\ 0 & 0 & 4 \end{array}\right).$$

   Hence, $\mathbf{v} \notin \text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

3. Let $\mathbf{e}_i$ be the i-th column of the order n identity matrix $\mathbf{I}_n$ for $i = 1, ..., n$. Then for any $\mathbf{w} = (w_i)$,

$$\mathbf{w} = w_1\mathbf{e}_1 + w_2\mathbf{e}_2 + \cdots + w_n\mathbf{e}_n.$$

This shows that $\text{span}\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\} = \mathbb{R}^n$. This set is called the <u>standard basis</u> of $\mathbb{R}^n$.

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ be a subset. Now instead of asking if a specific vector $\mathbf{v} \in \mathbb{R}^n$ is in the span, we may ask if all the vectors in $\mathbb{R}^n$ is in the span, that is, whether $\text{span}(S) = \mathbb{R}^n$. This is equivalent to asking if for every $\mathbf{v}$ in $\mathbb{R}^n$, the system $\mathbf{Ax} = \mathbf{v}$ is consistent, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ is the $n \times k$ matrix whose columns are the vectors in S.

Example. 1. $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} \right\}$.

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 1 & 2 & 3 & y \\ 1 & 1 & 2 & z \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 2x - y \\ 0 & 1 & 1 & -x + y \\ 0 & 0 & 0 & -x + z \end{array} \right)$$

So $\text{span}(S) \neq \mathbb{R}^3$, since for example, $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \notin \text{span}(S)$ or $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \notin \text{span}(S)$.

2. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\}$.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & x \\ 1 & -1 & 2 & y \\ 1 & 0 & 1 & z \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & -x - y + 3z \\ 0 & 1 & 0 & x - z \\ 0 & 0 & 1 & x + y - 2z \end{array} \right)$$

So for any $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$,

$$(-x - y + 3z) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (x - z) \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + (x + y - 2z) \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

From the examples above, observe that we can always pick a $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ such that the last entry in the right hand side of the RREF of the augmented matrix is nonzero. This seems to indicate that S spans $\mathbb{R}^n$ if and only if the RREF of $\mathbf{A}$ does not have zero rows. This is indeed the case. For the general proof, readers may refer to the appendix.

Corollary. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ be a subset and $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ be a $n \times k$ matrix whose columns are the vectors in S. Then span$(S) = \mathbb{R}^n$ if and only if the reduced row-echelon form $\mathbf{R}$ of $\mathbf{A}$ has no zero rows.*

Observe that if $k < n$, then $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ cannot span $\mathbb{R}^n$, since the reduced row-echelon form $\mathbf{R}$ of $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ can have at most k pivot columns, and hence, it can at most have k nonzero rows. So $\mathbf{R}$ must have $n-k$ number of zero rows,

$$\mathbf{R} = \begin{pmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & * \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

We will state this as a result.

Theorem. *A subset $S \subseteq \mathbb{R}^n$ containing less than n vectors cannot span $\mathbb{R}^n$.*

So n is the lower bound on the number of vectors needed to span $\mathbb{R}^n$. Is this lower bound achieved? Yes, for example, the standard basis $E = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ spans $\mathbb{R}^n$. So n is the "most efficient" number of vectors needed to span $\mathbb{R}^n$.

Remark. We will learn later that the above corollary is equivalent to

$$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = \mathbb{R}^n \Leftrightarrow \text{rank}(\mathbf{A}) = n,$$

where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.

We will see eventually that spanning sets are not only ubiquitous, they are fundamental as they produce geometrical objects call subspaces. Here we present their fundamental properties that make them important.

Theorem. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$. Then*

(i) (Contains the origin) $\mathbf{0} \in span(S)$, and

(ii) (Closed under linear combination) for any $\mathbf{u}, \mathbf{v} \in span(S)$ and $\alpha, \beta \in \mathbb{R}$,

$$\alpha\mathbf{u} + \beta\mathbf{v} \in span(S).$$

Proof. (i) Take the trivial combination,

$$0\mathbf{u}_1 + 0\mathbf{u}_2 + \cdots + 0\mathbf{u}_k = \mathbf{0}.$$

(ii) Suppose $\mathbf{u}, \mathbf{v} \in \text{span}(S)$. Then we can find real numbers $c_1, c_2, ..., c_k, d_1, d_2, ..., d_k \in \mathbb{R}$ such that

$$\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \text{ and } \mathbf{v} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k.$$

Then

$$\begin{aligned} \alpha\mathbf{u} + \beta\mathbf{v} &= \alpha(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) + \beta(d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k) \\ &= (\alpha c_1 + \beta d_1)\mathbf{u}_1 + (\alpha c_2 + \beta d_2)\mathbf{u}_2 + \cdots + (\alpha c_k + \beta d_k)\mathbf{u}_k \end{aligned}$$

tells us that $\alpha\mathbf{u} + \beta\mathbf{v}$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$, and thus it is in the span.

□

By induction using property (ii), we have property (ii'), that if $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m \in \text{span}(S)$, then any linear combination of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ is also in $\text{span}(S)$. This is an important result and we will state it as a theorem.

Theorem. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$. If $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m \in span(S)$, then for any $c_1, c_2, ..., c_m \in \mathbb{R}$,*

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m \in span(S).$$

That is, $span\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\} \subseteq span(S)$.

Observe that the vectors used to span a set is not unique, for example, both $S = \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$ and $T = \left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$ span the whole x, y plane in $\mathbb{R}^3$. So given two spanning sets S and T, the theorem above gives us a way to tell if they span the same set, or even if one of them is contained in the other.

For suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, and $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$, both subsets of $\mathbb{R}^n$. Then if $\mathbf{v}_i \in \text{span}(S)$ for every $i = 1, ..., m$, the theorem above tells us that $\text{span}(T) \subseteq \text{span}(S)$. If further $\mathbf{u}_i \in \text{span}(T)$ for every $i = 1, ..., k$, then we too have $\text{span}(S) \subseteq \text{span}(T)$. Thus, equality $\text{span}(S) = \text{span}(T)$ holds. So rephrasing the theorem, we get the following statement.

Theorem. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$, both subsets of $\mathbb{R}^n$. Then $span(T) \subseteq span(S)$ if and only if $\mathbf{v}_i \in span(S)$ for every $i = 1, ..., m$.*

The algorithm to check if $\text{span}(T) \subseteq \text{span}(S)$ is thus as follows. We want to check if the system associated to the augmented matrix

$$\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_i \end{array}\right)$$

is consistent for all $i = 1, ..., m$. Recall from lecture 2 that we can do this simultaneously, that is, check if the systems associated to the augmented matrix

$$\left(\begin{array}{cccc|c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{array}\right)$$

is consistent.

Example. 1. $S = \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}, T = \left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$.

To check if $\text{span}(T) \subseteq \text{span}(S)$,

$$\left(\begin{array}{cc|c|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

is consistent.

To check if $\text{span}(S) \subseteq \text{span}(T)$,

$$\left(\begin{array}{cc|c|c} 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c|c} 1 & 0 & 1/2 & 1/2 \\ 0 & 1 & 1/2 & -1/2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

is consistent. Hence, $\text{span}(S) = \text{span}(T)$.

2. $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} \right\}$, $T = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\}$.

To check if $\text{span}(S) \subseteq \text{span}(T)$,

$$\left(\begin{array}{ccc|c|c|c} 1 & 1 & 1 & 1 & 1 & 2 \\ 1 & -1 & 2 & 1 & 2 & 3 \\ 1 & 0 & 1 & 1 & 1 & 2 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c|c|c} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{array} \right)$$

To check if $\text{span}(T) \subseteq \text{span}(S)$,

$$\left(\begin{array}{ccc|c|c|c} 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 2 & 3 & 1 & -1 & 2 \\ 1 & 1 & 2 & 1 & 0 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c|c|c} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right)$$

This shows that $\text{span}(T) \not\subseteq \text{span}(S)$. In particular, $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \notin \text{span}(S)$.

Remark. We have seen that the standard basis $E = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ spans $\mathbb{R}^n$. Since

$$\left(\begin{array}{cccc|c|c|c|c} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n & \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{array} \right)$$

is always consistent, if we want to check if a set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ spans $\mathbb{R}^n$, it suffice to check that

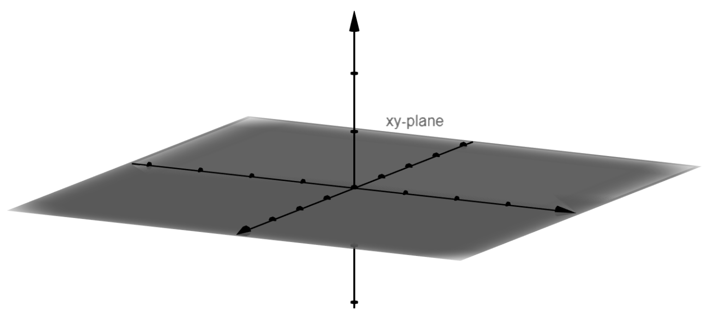
$$\left(\begin{array}{cccc|c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{array} \right)$$

is consistent. But we have a theorem above saying that $\text{span}(S) = \mathbb{R}^n$ if and only if the reduced row-echelon form of $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ has no zero rows. Are these two different algorithms?

This also show that $\mathbf{v} \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ if and only if $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{v}\}$. That is, a vector in the spanning set is a linear combination of the others is and only if it is "reduntant" in the spanning set. This means that if we want the most efficient/no redundancy spanning set, we have to make sure that none of the vectors in the spanning set can be written as a linear combination of the others. We will discuss this in details in lecture 7.

3.6 Subspace

A subspace is a vector space that is contained in another vector space. Here we are restriction ourselves to only subspaces in Euclidean spaces $\mathbb{R}^n$. We will give a precise definition in a while. Consider the xy-plane in $\mathbb{R}^3$.



It looks exactly like $\mathbb{R}^2$. So can we say that $\mathbb{R}^2$ is a subset of $\mathbb{R}^3$? For $\mathbb{R}^2$ to be a subset of $\mathbb{R}^3$, every vector in $\mathbb{R}^2$ must also be a vector in $\mathbb{R}^3$. So for $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, to be a vector in $\mathbb{R}^3$, it needs 3 coordinates,

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix}?$$

Naturally, we can let $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$. Then we can say that we have an image of $\mathbb{R}^2$ in $\mathbb{R}^3$, $\mathbb{R}^2 \hookrightarrow \mathbb{R}^3$ given by this mapping. However, the mapping is not unique. The following mappings also produce images of $\mathbb{R}^2 \hookrightarrow \mathbb{R}^3$,

$$1.\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ x \\ y \end{pmatrix},2.\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ 0 \\ y \end{pmatrix},3.\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ x-y \\ 0 \end{pmatrix},4.\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

However, we do not just want the image to look like $\mathbb{R}^2$. We want the image to have the same behaviour as $\mathbb{R}^2$, that is, in $\mathbb{R}^2$, we have the origin, we can add vectors, and we can take scalar multiple, etc. So one can see for example the image under mapping 4.,

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

does not contain the origin. Moreover, it does not respect vector addition,

$$\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) \mapsto \begin{pmatrix} (x_1+x_2) \\ (y_1+y_2) \\ 1 \end{pmatrix} \neq \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix}.$$

It turns out that if a subset $V \subseteq \mathbb{R}^n$ contains the origin and is closed under linear combinations, then it looks exactly like the image of $\mathbb{R}^k$ in $\mathbb{R}^n$ for some $k \le n$, and have all the desired properties $\mathbb{R}^k$ has. So we have the following definition.

A subset $V \subseteq \mathbb{R}^n$ is a <u>subspace</u> if

(i) (Contains the origin) $\mathbf{0} \in V$, and

(ii) (Closed under linear combination) for any $\mathbf{u}, \mathbf{v} \in V$ and $\alpha, \beta \in \mathbb{R}$,

$$\alpha\mathbf{u} + \beta\mathbf{v} \in V.$$

Remark. In some textbooks, condition (i) is replaced by the condition that V is nonempty. This is equivalent for if V is nonempty, then by (ii), pick any vector $\mathbf{v} \in V$, then $\mathbf{0} = 0\mathbf{v} \in V$, and conversely, certainly if V contains the origin, it is nonempty.

We have seen that a spanning set satisfies these 2 properties, and hence is a subspace. It turns out that for Euclidean spaces, they are actually equivalent. The proof that every subspaces can be written as a spanning set will be given in lecture 8. We will state it as a theorem.

Theorem. *A subset $V \subseteq \mathbb{R}^n$ is a subspace if and only if there exists a finite set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ such that $V = span(S)$.*

Remark. Hence, to show that $V \subseteq \mathbb{R}^n$ is a subspace, we can either check that the definition is satisfied, that is, it contains the origin and is closed under linear combination, or find a finite subset $S \subseteq \mathbb{R}^n$ such that $V = \text{span}(S)$. However, if a subset V is not a subspace, it is impossible to directly show that there exists no $S \subseteq \mathbb{R}^n$ such that $V = \text{span}(S)$, since there are infinitely many S to check.

To show that $V \subseteq \mathbb{R}^n$ is not a subspace, we have to show that it does not satisfies some of the conditions in the definition. That is, show that either

(i) it does not contain the origin $\mathbf{0} \notin V$,

(ii) there is a vector $\mathbf{v} \in V$ and a $\alpha \in \mathbb{R}$ such that $\alpha\mathbf{v} \notin V$, or

(iii) there are vectors $\mathbf{u}, \mathbf{v} \in V$ such that $\mathbf{u} + \mathbf{v} \notin V$.

Example. 1. $V = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = x\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \middle| \, x, y \in \mathbb{R} \right\}$ is a subspace spanned by $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$.

2. $V = \left\{ \begin{pmatrix} x+y \\ x-y \\ 0 \end{pmatrix} = x\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \middle| \, x, y \in \mathbb{R} \right\}$ is a subspace spanned by $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$.

3. The set $\left\{ \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \middle| \, x, y \in \mathbb{R} \right\}$ is not a subspace since it does not contain the origin. It is not possible to write it as a span of some vectors in $\mathbb{R}^3$.

4. The subset $V = \left\{ \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \middle| \, ab = cd \right\}$ is not a subspace since $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \in V$ and $\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \in V$, but

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 0 \end{pmatrix} \notin V.$$

Thus, it is not possible to write it as a span of some vectors in $\mathbb{R}^4$.

5. The subset $V = \left\{ \begin{pmatrix} s \\ s^2 \\ t \end{pmatrix} \middle| \, s, t \in \mathbb{R} \right\}$ is not a subspace since $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \in V$ but $2\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \notin V$, since $2^2 \neq 2$. Thus, It is not possible to write it as a span of some vectors in $\mathbb{R}^3$.

By the theorem above, the set containing only the origin $\{\mathbf{0}\}$ in $\mathbb{R}^n$ is a subspace, since $\{\mathbf{0}\} = \text{span}\{\mathbf{0}\}$. We can also check that it satisfies the conditions of a subspace,

(i) it contains the origin, $\mathbf{0} \in \{\mathbf{0}\}$, and

(ii) the only vector in $\{\mathbf{0}\}$ is $\mathbf{0}$, and for any $\alpha, \beta \in \mathbb{R}$,

$$\alpha\mathbf{0} + \beta\mathbf{0} = \mathbf{0} \in \{\mathbf{0}\}.$$

This space is called the zero space. It is the only subspace that has finitely many (one) vector. Any subspaces V besides the zero space must have infinitely many vectors, since if $\mathbf{v} \in V$ and $\mathbf{v} \neq \mathbf{0}$, then $t\mathbf{v} \in V$ for all $t \in \mathbb{R}$, and they are distinct for different choices of $t \in \mathbb{R}$.

Exercise: Is the set $\left\{ \begin{pmatrix} s^3 \\ t^3 \\ 0 \end{pmatrix} \middle| \; s, t \in \mathbb{R} \right\}$ a subspace?

Subspaces of $\mathbb{R}^2$

(i) Zero space: $\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$.

(ii) Image of $\mathbb{R}$: lines, $L = \text{span}\left\{ \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \right\}$ for some fixed $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

(iii) Whole $\mathbb{R}^2$.

Subspaces of $\mathbb{R}^3$

(i) Zero space: $\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}$.

(ii) Image of $\mathbb{R}$: Lines, $L = \text{span}\left\{ \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \right\}$ for some fixed $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

(iii) Image of $\mathbb{R}^2$: Planes, $P = \text{span}\left\{ \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \right\}$ for some $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$ that are not a scalar multiple of each other.

(iv) Whole $\mathbb{R}^3$.

3.7 Solution Set of Linear Systems and Subspaces

Theorem. *The solution set* $\{ \; \mathbf{u} \mid \mathbf{A}\mathbf{u} = \mathbf{b} \; \}$ *to a linear system* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *is a subspace if and only if* $\mathbf{b} = \mathbf{0}$*, that is, the system is homogeneous.*

Proof. Suppose the solution $\{\ \mathbf{u} \mid \mathbf{Au} = \mathbf{b}\ \}$ is a subspace. Then it must contain the origin. Hence, $\mathbf{b} = \mathbf{A0} = \mathbf{0}$.

Conversely, let $V = \{\ \mathbf{u} \mid \mathbf{Au} = \mathbf{0}\ \}$ be the solution set to a homogeneous system $\mathbf{Ax} = \mathbf{0}$. We will show that V contains the origin and is closed under linear combinations. Clearly it contains the origin since $\mathbf{A0} = \mathbf{0}$. Now suppose $\mathbf{u}$ and $\mathbf{v}$ are in V, that is, they are solutions to the homogeneous system, $\mathbf{Au} = \mathbf{0} = \mathbf{Av}$. Then for any $\alpha, \beta \in \mathbb{R}$,

$$A(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha(\mathbf{Au}) + \beta(\mathbf{Av}) = \alpha\mathbf{0} + \beta\mathbf{0} = \mathbf{0}.$$

Hence, $\alpha\mathbf{u} + \beta\mathbf{v}$ is a solution to the homogeneous system, and thus is in V too. So V is closed under linear combination. $\square$

In general, the solution set of any linear system is known as an affine subspace. A set $W \subseteq \mathbb{R}^n$ is an affine subspace if there is a vector $\mathbf{u} \in \mathbb{R}^n$ and subspace $V \subseteq \mathbb{R}^n$ such that $W = \mathbf{u} + V$, that is, for every $\mathbf{w} \in W$, we can write $\mathbf{w} = \mathbf{u} + \mathbf{v}$ for some $\mathbf{v} \in V$.

Given a solution set

$$W = \{\ \mathbf{w} \mid \mathbf{Aw} = \mathbf{b}\ \}$$

of a linear system $\mathbf{Ax} = \mathbf{b}$, we claim that $W = \mathbf{u} + V$, where

$$V = \{\ \mathbf{v} \mid \mathbf{Av} = \mathbf{0}\ \}$$

is the solution set to the homogeneous system, and $\mathbf{u}$ is a particular solution, $\mathbf{Au} = \mathbf{b}$. This is because for any $\mathbf{v} \in V$, a solution to the homogeneous system, we have

$$\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

This shows that $\mathbf{u} + V \subseteq W$.

Conversely, suppose $\mathbf{w} \in W$ is a solution to the system $\mathbf{Ax} = \mathbf{b}$. Let $\mathbf{v} = \mathbf{w} - \mathbf{u}$. Then

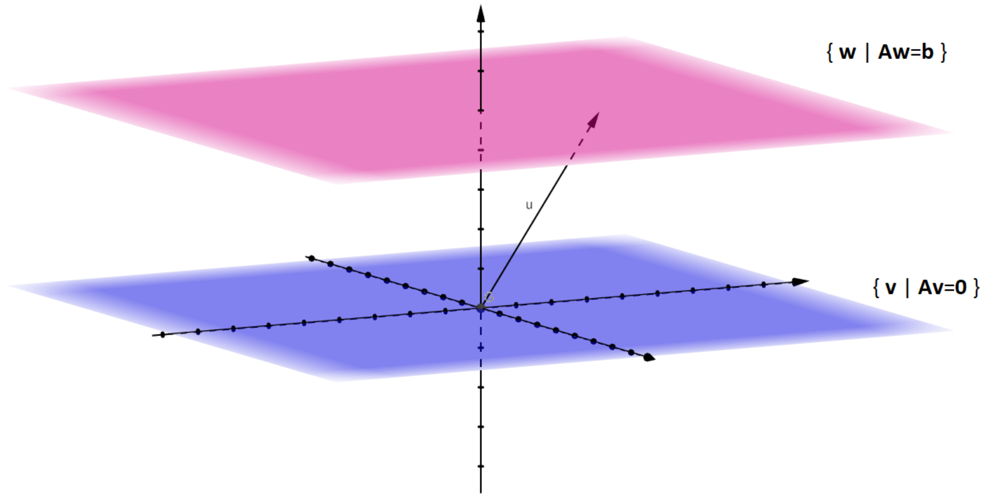
$$\mathbf{Av} = \mathbf{A}(\mathbf{w} - \mathbf{u}) = \mathbf{Aw} - \mathbf{Au} = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

tells us that $\mathbf{v} \in V$ is a solution to the homogeneous system. So we can write $\mathbf{w} = \mathbf{u} + (\mathbf{w} - \mathbf{u}) = \mathbf{u} + \mathbf{v}$ for some $\mathbf{v} \in V$. This shows that every $\mathbf{w} \in W$ can be written as $\mathbf{u} + \mathbf{v}$, and hence $W \subseteq \mathbf{u} + V$.

Hence, we have equality. Thus we have the following theorem.

Theorem. *The solution set* $W = \{\ \mathbf{w} \mid \mathbf{Aw} = \mathbf{b}\ \}$ *of a linear system* $\mathbf{Ax} = \mathbf{b}$ *is given by* $\mathbf{u} + V$*, where* $V = \{\ \mathbf{v} \mid \mathbf{Av} = \mathbf{0}\ \}$ *is the solution space to the associated homogeneous system and* $\mathbf{u}$ *is a particular solution,* $\mathbf{Au} = \mathbf{b}$.

That is, the solution set of a linear system is an affine subspace.


{ w | Aw=b }
u
{ v | Av=0 }

Appendix for Lecture 6

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ be a subset. Now instead of asking if a specific vector $\mathbf{v} \in \mathbb{R}^n$ is in the span, we may ask if all the vectors in $\mathbb{R}^n$ is in the span, that is, whether $\text{span}(S) = \mathbb{R}^n$. This is equivalent to asking if for every $\mathbf{v}$ in $\mathbb{R}^n$, the system $\mathbf{Ax} = \mathbf{v}$ is consistent, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ is the $n \times k$ matrix whose columns are the vectors in S. Let $\mathbf{R}$ be the reduced row-echelon form of $\mathbf{A}$. Then $\mathbf{PA} = \mathbf{R}$, where $\mathbf{P} = \mathbf{E}_r \cdots \mathbf{E}_2\mathbf{E}_1$ is an order n invertible matrix and $\mathbf{E}_i$, $i = 1, ..., r$ are the elementary matrices that reduce $\mathbf{A}$ to $\mathbf{R}$. Let us consider whether $\mathbf{R}$ has zero rows.

(i) Case 1: $\mathbf{R}$ has a zero row, that is, $\mathbf{R}$ is of the form

$$\mathbf{R} = \begin{pmatrix} \mathbf{Q} \\ \mathbf{0}_{1\times n} \end{pmatrix}$$

for some $n - 1 \times k$ matrix $\mathbf{Q}$. Let $\mathbf{v} = \mathbf{P}^{-1}\mathbf{e}_n$, where $\mathbf{e}_n$ is the n-th column of the order n identity matrix. Then

$$\mathbf{Ax} = \mathbf{v} = \mathbf{P}^{-1}\mathbf{e}_n \Rightarrow \mathbf{Rx} = \mathbf{PAx} = \mathbf{e}_n$$

is inconsistent since the augmented matrix of $\mathbf{Rx} = \mathbf{e}_n$ is

$$\left(\begin{array}{ccc|c} & \mathbf{Q} & & \begin{matrix} 0 \\ \vdots \end{matrix} \\ 0 & \cdots & 0 & 1 \end{array}\right).$$

This means that $\text{span}(S) \neq \mathbb{R}^n$, since $\mathbf{v} = \mathbf{P}^{-1}\mathbf{e}_n \notin \text{span}(S)$.

(ii) Case 2: If $\mathbf{R}$ has no zero rows, then for any $\mathbf{v} \in \mathbb{R}^n$,

$$\left(\begin{array}{c|c} \mathbf{A} & \mathbf{v} \end{array}\right) \longrightarrow \left(\begin{array}{c|c} \mathbf{R} & \mathbf{v}' \end{array}\right)$$

is consistent since the pivot columns are always in the left hand side of the augmented matrix. Hence, for any $\mathbf{v} \in \mathbb{R}^n$, the system $\mathbf{Ax} = \mathbf{v}$ is consistent, and therefore $\text{span}(S) = \mathbb{R}^n$.

In summary, we have the following statement.

Corollary. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ be a subset and $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ be a $n \times k$ matrix whose columns are the vectors in S. Then span$(S) = \mathbb{R}^n$ if and only if the reduced row-echelon form $\mathbf{R}$ of $\mathbf{A}$ has no zero rows.*

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 7 Notes

References

1. Elementary Linear Algebra: Application Version, Section 4.3-4.4
2. Linear Algebra with Application, Section 5.2, 6.3

3.8 Linear Independence

Consider $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u}_4 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$. Then

$$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\} = V$$

where V is the xy-plane in $\mathbb{R}^3$. But span$\{\mathbf{u}_1\}$ is the x-axis, which is not V. So it seems like we are allowed to remove some vectors and still span the same subspace. It is then natural to ask if given a set S such that span$(S) = V$, can we find a smallest subset of S that can still do the job of spanning V. In the above example, one can say that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is a smallest set. But note that the choice of smallest set is not unique, we can also choose $\{\mathbf{u}_3, \mathbf{u}_4\}$; in fact, in this case, any choice of two vectors will give us a smallest set to span the xy-plane.

A vector $\mathbf{u}$ is linearly dependent on $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ if it is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$,

$$\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$$

for some $c_1, c_2, ..., c_k \in \mathbb{R}$.

Example.

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$$

So $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$ is linearly dependent on $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$. But observe that

$$\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

So we may also say that $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is linearly dependent on $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$, and $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ is linearly dependent on $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

So it is more accurate to describe the linearly dependency of a set rather than individual vectors. So we might say that a set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly dependent if one of the $\mathbf{u}_i$ is a linear combination of the others. This is easy to understand, but very tedious to check, since by this definition, we have to check for consistency of

$$\left(\begin{array}{ccccc|c} \mathbf{u}_1 & \cdots & \mathbf{u}_{i-1} & \mathbf{u}_{i+1} & \cdots & \mathbf{u}_k & \mathbf{u}_i \end{array}\right)$$

for every single $i = 1, ..., n$. However, observe that if say, without lost of generality,

$$\mathbf{u}_k = c_1\mathbf{u}_1 + \mathbf{u}_2 + \cdots + c_{k-1}\mathbf{u}_{k-1}, \tag{3}$$

then

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_{k-1}\mathbf{u}_{k-1} - \mathbf{u}_k = \mathbf{0},$$

that is, we are able to find a nontrivial (because the coefficient of $\mathbf{u}_k$ is $-1 \neq 0$) linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ to give $\mathbf{0}$. We may replace $\mathbf{u}_k$ in (3) with any of the $\mathbf{u}_i$ and still have the same conclusion that there is a nontrivial combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ to give us $\mathbf{0}$. Hence, we now have a useful definition of linear dependency.

A set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is <u>linearly dependent</u> if there exists $c_1, c_2, ..., c_k \in \mathbb{R}$, not all zero such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$

A set is <u>linearly independent</u> otherwise. Explicitly, a set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly independent if whenever $c_1, c_2, ..., c_k \in \mathbb{R}$ is such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0},$$

necessarily $c_1 = c_2 = \cdots = c_k = 0$. In words, the only linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ to give $\mathbf{0}$ is the trivial one.

Theorem. *A set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly dependent if and only if there is a $i = 1, ..., k$ such that*

$$\mathbf{u}_i = c_1\mathbf{u}_1 + \cdots + c_{i-1}\mathbf{u}_{i-1} + c_{i+1}\mathbf{u}_{i+1} + \cdots + c_k\mathbf{u}_k,$$

that is, $\mathbf{u}_i$ is a linear combination of the rest of the vectors in the set S.

Thus to check if a set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly dependent, we are asking if the (homogeneous) system associated to

$$\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{0} \end{array}\right)$$

has nontrivial solution. We will state it as a theorem.

Theorem. *A set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ is linearly independent if and only if the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ is the matrix whose column is formed by the vectors in S.*

But recall that the system $\mathbf{Ax} = \mathbf{0}$ has nontrivial solutions if and only if the reduced row-echelon form of $\mathbf{A}$ has non-pivot columns (since the number of non-pivot column is equal to the number of parameters required in a general solution).

Theorem. *A set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ is linearly dependent if and only if any row-echelon form of $\mathbf{A}$ has non-pivot columns, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ is the matrix whose column is formed by the vectors in S.*

We will learn later that the theorem above is equivalent to $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ is linearly independent if and only if $\text{rank}(\mathbf{A}) = k$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.

Corollary. *Any subset of $\mathbb{R}^n$ containing more than n vectors must be linearly dependent.*

Proof. If $k > n$ and $S \subseteq \mathbb{R}^n$ has k vectors, the matrix whose columns are vectors in S has size $n \times k$, and must have at least $k - n$ non-pivot columns. $\square$

Example. 1. $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$.

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

is already in reduced row-echelon form, and has 2 non-pivot columns. Hence S is linearly dependent. Alternatively, since S contains 4 vectors in $\mathbb{R}^3$, it must be linearly dependent.

2. $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} \right\}$ is linearly dependent since

$$\begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & -1 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

3. $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is linearly independent since

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

A set $S = \{\mathbf{u}\} \subseteq \mathbb{R}^n$ containing one nonzero vector, $\mathbf{u} \neq \mathbf{0}$, is linearly independent. For the only linear combination of $\mathbf{u}$ is taking scalar multiple, and since $\mathbf{u} \neq \mathbf{0}$, $c\mathbf{u} = \mathbf{0}$ if and only if $c = 0$. However, the set containing the origin $S = \{\mathbf{0}\} \subseteq \mathbb{R}^n$ is linearly dependent, since we can take $c = 1$ and

$$c\mathbf{0} = \mathbf{0}$$

is a nontrivial linear combination (scalar multiple) of $\mathbf{0}$ to get $\mathbf{0}$.

Lemma. *A set $S = \{\mathbf{u}, \mathbf{v}\}$ containing two vectors is linearly independent if and only if the vectors are not a multiple of each other.*

Proof. S is linearly dependent if and only if $\alpha\mathbf{u} + \beta\mathbf{v} = 0$ for some $\alpha, \beta \in \mathbb{R}$ not both zero. If $\alpha \neq 0$, then $\mathbf{u} = (-\beta/\alpha)\mathbf{v}$, and if $\beta \neq 0$, $\mathbf{v} = (-\alpha/\beta)\mathbf{u}$, that is the vectors are multiple of each other. $\square$

Theorem. *Suppose* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ *is linearly dependent. Then for any* $\mathbf{u} \in \mathbb{R}^n$*, the set*
$$\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$$
is linearly dependent.

That is, adding vectors to a linearly dependent set will not make it linearly independent.

This means that any set containing the origin $\mathbf{0}$ must be linearly independent. We can also see this from the fact that a matrix containing a zero column must have non-pivot column in its reduced row-echelon form; the zero column is a non-pivot column.

Converse to the previous theorem, we are allowed to add vectors to linearly independent set and still have a linearly independent set if the added vector is not linearly dependent of the vectors in the original set. We will state it precisely as a theorem.

Theorem. *Suppose* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ *is linearly independent and* $\mathbf{u}$ *is not a linearly combination of* $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$*,* $\mathbf{u} \notin span\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$*. Then the set* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ *is linearly independent.*

Example. 1. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$. Then

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \\ 1 & 1 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

which shows that S is linearly dependent. So for any $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, the set

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\}$$

will be linearly dependent,

$$\begin{pmatrix} 1 & 1 & 0 & x \\ 2 & 3 & 0 & y \\ 1 & 1 & 0 & z \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 3x - y \\ 0 & 1 & 0 & -2x + y \\ 0 & 0 & 0 & -x + z \end{pmatrix}.$$

2. Let $S = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$. Then S is linearly independent,

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

For any $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$,

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & x_1 \\ 2 & 0 & 1 & x_2 \\ 1 & 0 & 1 & x_3 \\ -1 & 1 & 0 & x_4 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{x_1-x_4}{2} \\ 0 & 1 & 0 & \frac{x_1+x_4}{2} \\ 0 & 0 & 1 & 2x_3 - x_2 \\ 0 & 0 & 0 & \frac{-x_1+2x_2-2x_3+x_4}{2} \end{array}\right).$$

Hence, $\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \right\}$ is linearly independent if and only if $-x_1 + 2x_2 - 2x_3 + x_4 \neq 0$, if and only if $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \notin \text{span}(S)$.

3.9 Basis

Recall that a subset $V \subseteq \mathbb{R}^n$ is a subspace if and only if $V = \text{span}(S)$ for some finite subset $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$. This would mean that every vector $\mathbf{v} \in V$ can be expressed as a linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$. Suppose further that the set S is also linearly independent. Then we claim that the coefficient of the linear combination is unique. Indeed, suppose $\mathbf{v} \in V$ is such that

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \text{ and } \mathbf{v} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k$$

for some $c_1, ..., c_k, d_1, ..., d_k \in \mathbb{R}$. Then

$$(c_1 - d_1)\mathbf{u}_1 + (c_2 - d_2)\mathbf{u}_2 + \cdots + (c_k - d_k)\mathbf{u}_k = \mathbf{v} - \mathbf{v} = \mathbf{0}.$$

Since S is linearly independent, necessary the coefficients are zero, $c_1 - d_1 = c_2 - d_2 = \cdots = c_k - d_k = 0$, that is

$$c_1 = d_1,\ c_2 = d_2, ...,\ c_k = d_k.$$

This means that the coefficients of $\mathbf{v}$ as a linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ are unique. So a set S that both spans V and is linearly independent is special in the sense that every vectors $\mathbf{v} \in V$ can be written as a linear combination of the vectors in S uniquely. We call such a special sets a basis.

Let $V \subseteq \mathbb{R}^n$ be a subspace. A set $S \subseteq V$ is a basis for V if

(i) $\text{span}(S) = V$, and

(ii) S is linearly independent.

Note that basis is not unique. We have already seen that both $S = \left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix} \right\}$ and $T = \left\{ \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix} \right\}$ are bases for the xy-plane.

Recall that the solution set to a homogeneous system $\mathbf{Ax} = \mathbf{0}$ is a subspace. Suppose

$$s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k,\ s_1, s_2, ..., s_k \in \mathbb{R}$$

is a general solution to the homogeneous system such that the parameters correspond to the non-pivot columns in the reduced row-echelon form of $\mathbf{A}$. Then observe that in each $\mathbf{u}_i$, there must be a entry where only $\mathbf{u}_i$ has 1 in the entry and $\mathbf{u}_j$ has 0 in this entry for all $j \neq i$. Thus the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly independent. Clearly the set spans the solution space. We thus arrive at the following theorem.

Theorem. *If* $V = \left\{ s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k \;\middle|\; s_1, s_2, ..., s_k \in \mathbb{R} \right\}$ *is a solution space to a homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *such that the parameters correspond to the non-pivot columns in the reduce row-echelon form of* $\mathbf{A}$*, then* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *is a basis for* V*.*

Example. 1. Let $V = \left\{ \begin{pmatrix}x\\y\\z\end{pmatrix} \;\middle|\; x + y - z = 0 \right\}$. It is the solution space to a homogeneous system, so it is a subspace of $\mathbb{R}^3$. Explicitly, $V = \left\{ s\begin{pmatrix}-1\\1\\0\end{pmatrix} + t\begin{pmatrix}1\\0\\1\end{pmatrix} \;\middle|\; s, t \in \mathbb{R} \right\} = \text{span}\left\{ \begin{pmatrix}-1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix} \right\}$. Let $T = \left\{ \begin{pmatrix}-1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix} \right\}$. To show that T is a basis, suffice to show that it is independent.

$$\begin{pmatrix}-1 & 1\\ 1 & 0\\ 0 & 1\end{pmatrix} \longrightarrow \begin{pmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{pmatrix}.$$

Since the reduced row-echelon form has no non-pivot columns, T is linearly independent. Alternatively, one can just observe that the vectors in T cannot be a multiple of each other. Hence, T is a basis.

2. Let V and T as defined in 1. Let $S = \left\{ \begin{pmatrix}-1\\2\\1\end{pmatrix}, \begin{pmatrix}1\\1\\2\end{pmatrix} \right\}$. We will show that S is a basis for V. First we show that $\text{span}(S) = V$. To show that $\text{span}(S) \subseteq V$, suffice to show that $S \subseteq V$. Indeed,

$$(-1) + (2) - (1) = 0 \text{ and } (1) + (1) - 2 = 0.$$

To show that $V \subseteq \text{span}(S)$, we show that $\text{span}(T) \subseteq \text{span}(S)$, that is, suffice to show that $T \subseteq \text{span}(S)$. Indeed,

$$\left(\begin{array}{cc|c|c} -1 & 1 & -1 & 1\\ 2 & 1 & 1 & 0\\ 1 & 2 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c|c} 1 & 0 & 2/3 & 1/3\\ 0 & 1 & -1/3 & 2/3\\ 0 & 0 & 0 & 0 \end{array}\right)$$

is consistent. This shows that $\text{span}(S) = V$. Next, we need to show that S is linearly independent. Indeed, it is clear $\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ cannot be a multiple of each other.

Question: We have seen that the zero space $\{\mathbf{0}\} \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$. What is a basis for the zero space?

We will now discuss when is a subset $S \subseteq \mathbb{R}^n$ a basis for the whole $\mathbb{R}^n$. Recall that if the set S contains less than n vectors, it cannot span $\mathbb{R}^n$ and if it has more than n vectors, it must be linearly dependent. So for S to span $\mathbb{R}^n$ and be linearly independent, it must have exactly n vectors, $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$. Let $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ be the square matrix of order n whose column is formed by the vectors in S. Then for S to span n, the reduced row-echelon form of $\mathbf{A}$ cannot have zero rows, and for S to be linearly independent, the reduced row-echelon form of $\mathbf{A}$ cannot have non-pivot column. But since $\mathbf{A}$ is a square matrix, either of the condition is equivalent to the reduced row-echelon form of $\mathbf{A}$ being the identity matrix. That is, S is a basis for $\mathbb{R}^n$ if and only if $\mathbf{A}$ is invertible. We thus have another equivalent criteria for invertibility.

Theorem. *A subset $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ is a basis for $\mathbb{R}^n$ if and only if $k = n$ and $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ is an invertible matrix.*

Another way of phrasing the theorem.

Theorem. *A square matrix $\mathbf{A}$ of order n is invertible if and only if the columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.*

Taking the transpose and recalling that $\mathbf{A}$ is invertible if and only if its transpose is, we get the following statement.

Theorem. *A square matrix $\mathbf{A}$ of order n is invertible if and only if the rows of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.*

We will add these statements into the list of equivalence of invertibility.

Theorem. *Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.*

(i) *$\mathbf{A}$ is invertible.*

(ii) *$\mathbf{A}$ has a left inverse.*

(iii) *$\mathbf{A}$ has a right inverse.*

(iv) *The reduced row-echelon form of $\mathbf{A}$ is the identity matrix $\mathbf{I}_n$.*

(v) *$\mathbf{A}$ is a product of elementary matrices.*

(vi) *The homogeneous linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution.*

(vii) *The linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.*

(viii) *The determinant of $\mathbf{A}$ is nonzero, $\det(\mathbf{A}) \neq 0$.*

(ix) *The columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.*

(x) The rows of $\mathbf{A}$ *form a basis for* $\mathbb{R}^n$*.*

Recall that we said that subspaces of $\mathbb{R}^n$ are like copies of $\mathbb{R}^k$ inside $\mathbb{R}^n$ for some $k \leq n$. Here we can make this statement precise. We have shown the following statement.

Theorem. *Suppose S is a basis for V. then every vectors $\mathbf{v} \in V$ can be written as a linear combination of vectors in S uniquely.*

So now let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for a subspace $V \subseteq \mathbb{R}^n$. Then given any $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \in \mathbb{R}^k$, we get a unique vector

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \in V.$$

Conversely, for any $\mathbf{v} \in V$, we can find a unique $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \in \mathbb{R}^k$ such that

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \in V.$$

This means that a basis S for V corresponds to an embedding (a copy of) $\mathbb{R}^k$ into $\mathbb{R}^n$. That is,

$$\mathbb{R}^k \xleftrightarrow{\text{via } S} V \subseteq \mathbb{R}^n, \qquad \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \leftrightarrow c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k.$$

We call the unique vector in $\mathbb{R}^k$ corresponding to $\mathbf{v} \in V$ via S the <u>coordinates of $\mathbf{v}$ relative to basis S</u>, and denote it as

$$(\mathbf{v})_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

Example. 1. Let $E = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ be the standard basis. For any vector $\mathbf{v} = (v_i) \in \mathbb{R}^n$, the relative coordinate of $\mathbf{v}$ relative to E is itself, $(\mathbf{v})_E = \mathbf{v}$. That is, the coordinates of a vector is its coordinates relative to the standard basis E.

2. $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for V, the xy-plane in $\mathbb{R}^3$. Then given any $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$,

$$\begin{pmatrix} x \\ y \end{pmatrix} \leftrightarrow x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Hence, for any $\mathbf{v} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \in V$,

$$(\mathbf{v})_S = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Remark. 1. Even though $\mathbf{v} \in V \subseteq \mathbb{R}^n$ has n coordinates, $(\mathbf{v})_S$ has k coordinates if the basis S has k vectors.

2. Note that the correspondence is unique only if S is a basis. If S is not linearly independent, a few vectors in $\mathbb{R}^k$ can map to the same $\mathbf{v} \in V$.

3. The relative coordinates depend to the ordering of the basis, if $S = \left\{ \mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ and $T = \left\{ \mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$, then for $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$,

$$(\mathbf{v})_S = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \neq \begin{pmatrix} 2 \\ 1 \end{pmatrix} = (\mathbf{v})_T.$$

4. The relative coordinates are different for different bases. If $S = \left\{ \mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ and $T = \left\{ \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$, then for $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$,

$$(\mathbf{v})_S = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (\mathbf{v})_T.$$

We will now give an algorithm to find coordinates relative to a basis. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for a subspace $V \subseteq \mathbb{R}^n$. For any $\mathbf{v} \in V$, we are solving for $c_1, c_2, ..., c_k \in \mathbb{R}$ such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{v}.$$

That is, we are solving for

$$\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v} \end{array} \right).$$

Theorem (Algorithm to finding coordinates relative to a basis)**.** *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for a subspace $V \subseteq \mathbb{R}^n$. For any $\mathbf{v} \in V$, $(\mathbf{v})_S$ is the unique solution to $\mathbf{Ax} = \mathbf{v}$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.*

Example. 1. We have seen that $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$ is a basis for the xy-plane V.

For any $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \in V$,

$$\left(\begin{array}{cc|c} 1 & 1 & x \\ 1 & -1 & y \\ 0 & 0 & 0 \end{array} \right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & (x+y)/2 \\ 0 & 1 & (x-y)/2 \\ 0 & 0 & 0 \end{array} \right).$$

So $(\mathbf{v})_S = \begin{pmatrix} (x+y)/2 \\ (x-y)/2 \end{pmatrix}$.

2. $V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \,\middle|\, x_1 - 2x_2 + x_3 = 0, x_2 + x_3 - 2x_4 = 0 \right\}$. We will leave it to the readers to check that $S = \left\{ \begin{pmatrix} -3 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ 0 \\ 1 \end{pmatrix} \right\}$ is a basis for V. For $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \in V$,

$$\left(\begin{array}{cc|c} -3 & 4 & 1 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right).$$

So $(\mathbf{v})_S = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Exercise: Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^n$ is a subspace. Let $\mathbf{v} \in V$.

(i) Suppose there is a non-pivot column in the left side of the reduced row-echelon form of

$$\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v} \end{array} \right).$$

What can you conclude?

(ii) Suppose

$$\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v} \end{array} \right)$$

is inconsistent. What can you conclude?

Appendix to Lecture 7

Theorem. *Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ is linearly dependent. Then for any $\mathbf{u} \in \mathbb{R}^n$, the set*

$$\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$$

is linearly dependent.

That is, adding vectors to a linearly dependent set will not make it linearly independent.

Proof. Since S is linearly dependent, we can find $c_1, c_2, ..., c_k \in \mathbb{R}$ not all zero such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$

To show that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ is linearly dependent, we let $c = 0$ and add $c\mathbf{u}$ to the above equation and obtain

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k + c\mathbf{u} = \mathbf{0}.$$

This is a nontrivial linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}$ to give us $\mathbf{0}$, since one of the c_i, $i = 1, ..., k$ is not zero.. Hence, $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ is also linearly dependent.

Alternatively, since the RREF of $\mathbf{A} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}$ has non-pivot columns, the RREF of $\begin{pmatrix}\mathbf{A} & \mathbf{u}\end{pmatrix} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}\end{pmatrix}$ must also have non-pivot columns (not obvious, details are left to readers). $\square$

Theorem. *Suppose $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq \mathbb{R}^n$ is linearly independent and $\mathbf{u}$ is not a linearly combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$, $\mathbf{u} \notin span\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$. Then the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ is linearly independent.*

Proof. Suppose to the contrary that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ is linearly dependent, that is we can find $c_1, c_2, ..., c_k, c \in \mathbb{R}$ not all zero such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k + c\mathbf{u} = \mathbf{0}.$$

Case 1: $c \neq 0$. Then

$$\mathbf{u} = -\frac{1}{c}(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k)$$

which contradict the hypothesis that $\mathbf{u}$ is not a linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$. So this case cannot be true. Case 2: $c = 0$, and thus one of the $c_i \neq 0$. Then we have

$$\mathbf{0} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k + c\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k.$$

But this would mean that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly dependent, a contradiction. So this cannot happen too.

Hence, $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ must be linearly independent. $\square$

We are now ready to prove that every subspace $V \subseteq \mathbb{R}^n$ can be written a as span of some vectors, $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$. Since V is a subspace, it is not empty. If $V = \{\mathbf{0}\}$, then $V = \text{span}\{\mathbf{0}\}$ and we are done. Suppose not. Then pick a nonzero vector $\mathbf{u}_1 \in V$. Suppose $\text{span}\{\mathbf{u}_1\} = V$, we are done. Otherwise, we can pick a vector in V

but not in $\text{span}\{\mathbf{u}_1\}$, $\mathbf{u}_2 \in V\backslash\text{span}\{\mathbf{u}_1\}$. Then since $\mathbf{u}_2$ is not a linear combination of $\mathbf{u}_1$, $\{\mathbf{u}_1, \mathbf{u}_2\}$ is a linearly independent set and $\text{span}\{\mathbf{u}_1\} \subsetneqq \text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$, $\text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$ is a strictly bigger set.

Proceeding inductively, suppose now we have constructed a linearly independent set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \subseteq V$. If $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, we are done. Otherwise, by the same argument as above, we can find a $\mathbf{u}_{k+1} \in V\backslash\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$. Then $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}\}$ is a linearly independent subset of V and its span is strictly bigger than $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$.

But the process have to stop, since we can at most have n linearly independent vectors in $\mathbb{R}^n$. Hence, there must be a $m \leq n$ such that the constructed set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m\}$ spans V. Hence, $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m\}$ is a linear span of some vectors. In fact, by the construction, $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m\}$ will be a basis for V.

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 8 Notes

References

1. Elementary Linear Algebra: Application Version, Section 4.5, 4.7
2. Linear Algebra with Application, Section 5.2, 5.4, 6.3

3.10 Dimension

Recall that $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$ if

(i) $\text{span}(S) = V$, and

(ii) S is linearly independent.

Now even though the choice of a basis for a subspace is not unique, the number of vectors in any basis must be the same. We will state it as a theorem. The proof of the theorem is beyond the scope of the course; interested readers can refer to the appendix.

Theorem. *Suppose* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *and* $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ *are bases for a subspace* $V \subseteq \mathbb{R}^n$. *Then* $k = m$.

We shall give a name to this intrinsic property of a subspace. The dimension of a subspace $V \subseteq \mathbb{R}^n$ is the number of vectors in any basis, denoted as $\dim(V)$.

Example. 1. The dimension of $\mathbb{R}^n$ is n, since the standard basis $E = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ has n vectors.

2. $\mathbb{R}^3$ is called the 3-dimensional Euclidean space, and $\mathbb{R}^2$ is called the 2-dimensional Euclidean space.

3. The xy-plane in $\mathbb{R}^3$ is a 2-dimensional subspace of $\mathbb{R}^3$.

4. Any k-dimensional subspace of $\mathbb{R}^n$ is an embedding of $\mathbb{R}^k$ into $\mathbb{R}^n$ for $k \leq n$.

Recall that the zero space, $\{\mathbf{0}\}$ is a subspace. What is its dimension? Although the zero space is spanned by $\{\mathbf{0}\}$, it cannot be a basis since it is linearly dependent. It turns out that the basis for the zero space is the empty set, and hence, the dimension of the zero space is 0. Intuitively, the dimension is the least number of directions you will need to be able to travel to any point of the subspace. For example, plane has 2 dimensions, since to be able to go to any point on a plane, you need at least 2 direction. A line is 1-dimensional, since we only need one direction to get to any point along the line, moving forward along the direction or backwards in the negative direction. Since the zero space is a point, we cannot travel anywhere, and hence, we need 0 directions. Readers may refer to the appendix for the formal proof.

Theorem. *The zero space is 0-dimensional. The empty set* $\varnothing = \{\}$ *is the basis for the zero space.*

Remark. Note that the empty set is not $\{\mathbf{0}\}$, and $\{\mathbf{0}\}$ is not the empty set since it contains the origin $\mathbf{0}$.

Recall that any set containing more than n vectors in $\mathbb{R}^n$ must be linearly dependent, and any set containing less than n vectors cannot span $\mathbb{R}^n$. We have similar results for subspaces.

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Any subset of V containing more than k vectors must be linearly dependent.*

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Any subset of V containing less than k vectors cannot span V.*

Theorem. *Let $\mathbf{A}$ be a $m \times n$ matrix. The number of non-pivot columns in the reduced row-echelon form of $\mathbf{A}$ is the dimension of the solution space*

$$\left\{ \ \mathbf{u} \in \mathbb{R}^n \ \middle| \ \mathbf{Au} = \mathbf{0} \ \right\}.$$

Example. Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \ 2x - 3y + z = 0 \ \right\}$ and

$$S = \left\{ \mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ -5 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} \right\}.$$

Let $x = s, y = t$, then $V = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} \right\}$. So $T = \left\{ \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} \right\}$ is a basis for V. This shows that V has 2 dimensions. Hence, S must be linearly dependent since it is a subset of V containing $3 > 2$ vectors. Indeed,

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -5 & 1 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}.$$

3.11 Equivalent Criteria for Basis

Lemma. *Suppose $U, V \subseteq \mathbb{R}^n$ are subspaces of $\mathbb{R}^n$ such that U is a subset of V, $U \subseteq V$. Then $\dim(U) \leq \dim(V)$, with equality if and only if $U = V$.*

The intuitive explanation is as such. Suppose U is a subspace contained in V. Let $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ be a minimum set of directions we need to travel the whole U. Then since $U \subseteq V$, we are allowed to travel these directions $\mathbf{u}_1, ..., \mathbf{u}_k$ in V too. If $U = V$, then the directions $\mathbf{u}_1, ..., \mathbf{u}_k$ are sufficient to travel the entire V too, hence, U and V has the same dimension. Otherwise, we need more directions to travel the whole V, and hence the dimension of V is strictly larger than the dimension of U.

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a k-dimensional subspace. If $S = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is a subset of V containing k number of linearly independent vectors, then S is a basis of V.*

Proof. We need to show that S spans V. Since $S \subseteq V$, we have $\text{span}(S) \subseteq V$. By the previous lemma, they must be equal, for since S is linearly independent, $k = \dim(\text{span}(S)) = \dim(V)$, and thus $\text{span}(S) = V$. $\square$

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Suppose $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ is a set containing k number of vectors and $V \subseteq$ span(S). Then S is a basis for V.*

Proof. We need to show that S is linearly independent. This follows from the previous lemma, since $V \subseteq \text{span}(S)$ and hence $k = \dim(V) \leq \text{span}(S)$. Hence, S must be linearly independent, for otherwise, the subspace spanned by S must have less than k dimensions, which is a contradiction. $\square$

Hence, we have arrived at the equivalent criteria for a set S to be a basis for a subspace V.

Definition	(**B1**)	(**B2**)
(1) $\text{span}(S) = V$ (2) S is L.I.	(1) $\lvert S\rvert = \dim(V)$ (2) $S \subseteq V$ and S is L.I.	(1) $V \subseteq \text{span}(S)$ (2) $\lvert S\rvert = \dim(V)$

By using (B1), we avoid the need to check that S is a spaning set, and by using (B2), we avoid the need to check that S is linearly independent. The other significance of the results above is that (B1) shows that any subset of V containing the maximal number of linearly independent vectors must be a basis, and (B2) shows that any set that spans V containing the minimum number of vectors to do so must be a basis.

Example. 1. Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \ x + y - z = 0 \right\}$ and $S = \left\{ \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right\}$. Then $\dim(V) = 2 = |S|$. By inspection, since the 2 vectors in S are not a multiple of each other, S is linearly independent. Finally, but substituting the vectors in S into the equation defining V, we can see that $S \subseteq V$. Hence, by (B1), S is a basis for V.

2. $V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \middle| \ x_1 - 2x_2 + x_3 = 0, x_2 + x_3 - 2x_4 = 0 \right\}$ and $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ -1 \\ 0 \end{pmatrix} \right\}$.
By inspection, S is linearly independent and $S \subseteq V$. Since the 2 equations defining V are not a multiple of each other and there are 4 variables, any general solution of V must have 2 parameters, and thus $\dim(V) = 2$. Since $|S| = 2$, S is a basis for V.

3. Let $S = \left\{ \begin{pmatrix} 0 \\ 2 \\ 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \\ 1 \\ 4 \\ -3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \\ -1 \end{pmatrix} \right\}$ and $T = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \right\}$.
We will show that S and T are both bases for the same subspace. Firstly, $\text{span}(T) \subseteq$

span(S) follows from

$$\left(\begin{array}{cccc|c|c|c|c} 0 & 0 & 4 & 0 & 1 & 0 & 0 & 0 \\ 2 & 1 & 6 & 2 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 2 & 1 & 4 & 1 & 0 & 0 & 0 & 1 \\ -1 & -1 & -3 & -1 & 0 & 0 & 1 & -1 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c|c|c|c} 1 & 0 & 0 & 0 & -1/4 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & -2 & 2 \\ 0 & 0 & 1 & 0 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1/2 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Now clearly T is linearly independent, and since $|T| = |S|$, by (B2), both S and T are bases for the same subspace.

4 Vector Spaces Associated the a Matrix

4.1 Column and Row Spaces

Let $\mathbf{A}$ be an $m \times n$ matrix,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The rows of $\mathbf{A}$,

$$\mathbf{r}_1 = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}, \mathbf{r}_2 = \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}, ..., \mathbf{r}_m = \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

are vectors in $\mathbb{R}^n$, and the columns of $\mathbf{A}$,

$$\mathbf{c}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{c}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, ..., \mathbf{c}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix},$$

are vectors in $\mathbb{R}^m$.

Define the row space of $\mathbf{A}$, denoted as Row(A), to be the subspace of $\mathbb{R}^n$ spanned by the rows of $\mathbf{A}$,

$$\text{Row}(\mathbf{A}) = \text{span}\left\{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}, \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}, ..., \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}\right\},$$

and the column space of $\mathbf{A}$, denoted as Col(A), to be the subspace of $\mathbb{R}^m$ spanned by the columns of $\mathbf{A}$,

$$\text{Col}(\mathbf{A}) = \text{span}\left\{\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, ..., \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\}.$$

Example. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 1 & 1 & 2 & 2 \end{pmatrix}$. Then the column space of $\mathbf{A}$ is

$$\mathrm{Col}(\mathbf{A}) = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \right\} = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\},$$

and the row space is

$$\begin{aligned} \mathrm{Row}(\mathbf{A}) &= \mathrm{span}\{\begin{pmatrix} 1 & 0 & 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 2 & 2 \end{pmatrix}\} \\ &= \mathrm{span}\{\begin{pmatrix} 1 & 0 & 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 2 \end{pmatrix}\}. \end{aligned}$$

Remark. Note that even though the rows of a $m \times n$ matrix is a $1 \times n$ matrix, one may still use a column vector, that is a $n \times 1$ matrix, to represent it. For example, one may write the row space of the matrix $\mathbf{A}$ in the above example as

$$\mathrm{Row}(\mathbf{A}) = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix} \right\}.$$

Similarly for the column vectors and the column space.

In the example above, we are able to find a basis for the column and row space easily by inspection. We will now discuss how we can find bases for column and row space of a matrix in general. We need the following results.

Theorem (Row operations preserve row space)**.** *Suppose* $\mathbf{A}$ *and* $\mathbf{B}$ *are row equivalent matrices. Then* $\mathrm{Row}(\mathbf{A}) = \mathrm{Row}(\mathbf{B})$.

In words, the theorem says that row operations preserve the row space of a matrix. We will demonstrate this theorem with an example.

Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 2 & 4 & 1 & 0 \\ 1 & 5 & -4 & -9 \end{pmatrix}$. Let $\mathbf{B}$ be obtained from $\mathbf{A}$ via the following row operations.

$$\begin{pmatrix} 1 & 1 & 2 & 3 \\ 2 & 4 & 1 & 0 \\ 1 & 5 & -4 & -9 \end{pmatrix} \xrightarrow{R_2-2R_1} \xrightarrow{R_3-R_1} \xrightarrow{R_3-2R_2} \begin{pmatrix} 1 & 1 & 2 & 3 \\ 0 & 2 & -3 & -6 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbf{B}.$$

We will show that $\mathrm{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \\ -4 \\ -9 \end{pmatrix} \right\} = \mathrm{Row}(\mathbf{A}) = \mathrm{Row}(\mathbf{B}) = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ -3 \\ -6 \end{pmatrix} \right\}$.

Indeed,

$$\left(\begin{array}{ccc|c|c} 1 & 2 & 1 & 1 & 0 \\ 1 & 4 & 5 & 1 & 2 \\ 2 & 1 & -4 & 2 & -3 \\ 3 & 0 & -9 & 3 & -6 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c|c} 1 & 0 & -3 & 1 & -2 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

shows that $\text{Row}(\mathbf{B}) \subseteq \text{Row}(\mathbf{A})$, and

$$\left(\begin{array}{cc|c|c|c} 1 & 0 & 1 & 2 & 1 \\ 1 & 2 & 1 & 4 & 5 \\ 2 & -3 & 2 & 1 & -4 \\ 3 & -6 & 3 & 0 & -9 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c|c|c} 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

shows that $\text{Row}(\mathbf{A}) \subseteq \text{Row}(\mathbf{B})$. Furthermore, this shows that $\left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ -3 \\ -6 \end{pmatrix} \right\}$ is a basis for both row spaces, since it is clearly linearly independent.

Now suppose $\mathbf{R}$ is the reduced row-echelon form of a matrix $\mathbf{A}$. By the theorem above, $\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{R})$. Observe that since $\mathbf{R}$ is in reduced row-echelon form, necessarily the nonzero rows of $\mathbf{R}$ are linearly independent. This is because the leading entries in the nonzero rows are the only entries that are nonzero in its column, and hence the nonzero rows cannot be written as a linear combination of the other rows. Clearly the nonzero rows of $\mathbf{R}$ spans the row space of $\mathbf{R}$. Hence, the nonzero rows of $\mathbf{R}$ form a basis for the row space of $\mathbf{R}$, and thus form a basis for the row space of $\mathbf{A}$.

Theorem (Finding a basis for row space)**.** *For any matrix* $\mathbf{A}$*, the nonzero rows of the reduced row-echelon form of* $\mathbf{A}$ *form a basis for* $\text{Row}(\mathbf{A})$*.*

Theorem (Row operations preserve linear relations)**.** *Let* $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}$ *and* $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$ *be row equivalent* $m \times n$ *matrices, where* $\mathbf{a}_i$ *and* $\mathbf{b}_i$ *are the* i*-th column of* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively, for* $i = 1, ..., n$*. Then for any* $c_1, c_2, ..., c_n \in \mathbb{R}^n$*, if*

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = \mathbf{0}$$

then (for the same coefficients $c_1, c_2, ..., c_n \in \mathbb{R}$*),*

$$c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n = \mathbf{0}.$$

Proof. Tutorial 7. □

In words, the theorem says that row operations preserve the linear relations of the columns of a matrix. That is, if one of the column in $\mathbf{A}$ is a linear combination of the other columns, then the corresponding column in $\mathbf{B}$ linearly depends on the other columns with exactly the same coefficients.

Now suppose $\mathbf{R}$ is the reduced row-echelon form of a matrix $\mathbf{A}$. Then the non-pivot columns of $\mathbf{R}$ can be written as a linear combination as the pivot columns. Hence, we have the same linear dependency in the corresponding columns of $\mathbf{A}$. Also, the pivot columns of $\mathbf{R}$ form a linearly independent set. Thus, the pivot columns of $\mathbf{R}$ form a basis for the column space of $\mathbf{R}$. Thus, the corresponding columns of $\mathbf{A}$ form a basis for the column space of $\mathbf{A}$.

Theorem (Finding basis for column space)**.** *Suppose* $\mathbf{R}$ *is the reduced row-echelon form of a matrix* $\mathbf{A}$*. Then the columns of* $\mathbf{A}$ *corresponding to the pivot columns in* $\mathbf{R}$ *form a basis for the column space of* $\mathbf{A}$*.*

Example. $\mathbf{A} = \begin{pmatrix} 2 & 1 & 4 & 1 & 2 \\ 4 & 2 & 2 & 3 & 2 \\ 2 & 1 & -2 & 2 & 0 \\ 6 & 3 & 6 & 4 & 4 \end{pmatrix} \longrightarrow \mathbf{R} = \begin{pmatrix} 1 & 1/2 & 0 & 5/6 & 1/3 \\ 0 & 0 & 1 & -1/6 & 1/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

Let $\mathbf{r}_i$ and $\mathbf{a}_i$ be the i-th column of $\mathbf{R}$ and $\mathbf{A}$, respectively, for $i = 1, ..., 5$. Then observe that

$$\mathbf{r}_2 = \tfrac{1}{2}\mathbf{r}_1, \quad \mathbf{r}_4 = \tfrac{1}{6}(5\mathbf{r}_1 - \mathbf{r}_3), \quad \mathbf{r}_5 = \tfrac{1}{3}(\mathbf{r}_1 + \mathbf{r}_3), \quad \text{and}$$

$$\mathbf{a}_2 = \tfrac{1}{2}\mathbf{a}_1, \quad \mathbf{a}_4 = \tfrac{1}{6}(5\mathbf{a}_1 - \mathbf{a}_3), \quad \mathbf{a}_5 = \tfrac{1}{3}(\mathbf{a}_1 + \mathbf{a}_3).$$

Hence, $\left\{ \begin{pmatrix} 2 \\ 4 \\ 2 \\ 6 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ -2 \\ 6 \end{pmatrix} \right\}$ form a basis for Col($\mathbf{A}$).

Next, clearly $\{\begin{pmatrix} 1 & 1/2 & 0 & 5/6 & 1/3 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 & -1/6 & 1/3 \end{pmatrix}\}$ is a linearly independent set. Hence, it form a basis for Row($\mathbf{A}$).

Remark.

1. Observe that in all the examples above, the dimension of the column space of $\mathbf{A}$ is equal to the dimension of the row space of $\mathbf{A}$.

2. Row operations do not preserve column space. Consider

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then column space of $\mathbf{A}$ is the x-axis in $\mathbb{R}^2$ while the row space of $\mathbf{B}$ is the y-axis in $\mathbb{R}^2$.

3. Row operations do no preserve linear relations of the rows. Consider

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \xrightarrow{R_2 - 2R_1} \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then the second row or $\mathbf{A}$ is twice the first row, but the second row of $\mathbf{B}$ is zero times the first row.

The theorems above are not only useful in finding bases of the row space and column space of a matrix $\mathbf{A}$, but also to find basis of subspaces.

1. Let $S = \left\{ \mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 3 \\ 6 \\ 6 \\ 3 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 4 \\ 9 \\ 9 \\ 5 \end{pmatrix}, \mathbf{u}_4 = \begin{pmatrix} -2 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \mathbf{u}_5 = \begin{pmatrix} 5 \\ 8 \\ 9 \\ 4 \end{pmatrix}, \mathbf{u}_6 = \begin{pmatrix} 4 \\ 2 \\ 7 \\ 3 \end{pmatrix} \right\}$ and $V = \text{span}(S)$.

   Suppose we want to find a subset of S that forms a basis for the subspace V, we use the column method. That is, we arrange the vectors as columns of a matrix $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_6 \end{pmatrix}$ and find a basis for the column space of $\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 4 & -2 & 5 & 4 \\ 2 & 6 & 9 & -1 & 8 & 2 \\ 2 & 6 & 9 & -1 & 9 & 7 \\ 1 & 3 & 5 & 1 & 4 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 0 & -14 & 0 & -37 \\ 0 & 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

   Hence, a basis for $V = \text{Col}(\mathbf{A})$ is $\{\mathbf{u}_1, \mathbf{u}_3, \mathbf{u}_5\}$.

2. Let S and V be as defined in the previous example. Suppose we want a basis of V such that it is easier to find the coordinates of a vectors in V relative to this basis, we use the row method. That is, we arrange the vectors in S (letting them be row vectors) as rows of a matrix $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_6^T \end{pmatrix}$ and find a basis for the row space of $\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 3 & 6 & 6 & 3 \\ 4 & 9 & 9 & 5 \\ -2 & -1 & -1 & 1 \\ 5 & 8 & 9 & 4 \\ 4 & 2 & 7 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\}$ form a basis for V. Now we can easily write any vector in V relative to this basis, for example,

$$\mathbf{u}_6 = 4 \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + 7 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

3. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \end{pmatrix} \right\}$. Suppose we want extend S to form a basis for $\mathbb{R}^4$. We will use the row method.

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix} \longrightarrow \mathbf{R} = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & -2 \end{pmatrix}.$$

Observe that if we add the rows $\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$ to the bottom of the matrix $\mathbf{R}$, we get

$$\begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which is clearly invertible, and thus the rows span $\mathbb{R}^4$. Hence, if we add the same two rows into the original matrix $\mathbf{A}$, then the row space of the resultant matrix is also the whole $\mathbb{R}^4$. Therefore $S \cup \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ form a basis for $\mathbb{R}^4$.

Appendix for Lecture 8

Recall that for each $\mathbf{v} \in V$, the coordinates of $\mathbf{v}$ relative to the basis S is the unique vector

$$(\mathbf{v})_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \in \mathbb{R}^k,$$

where $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$.

Lemma. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for a subspace $V \subseteq \mathbb{R}^n$. For any $\mathbf{u}, \mathbf{v} \in V$ and $\alpha, \beta \in \mathbb{R}$*

$$(\alpha\mathbf{u} + \beta\mathbf{v})_S = \alpha(\mathbf{u})_S + \beta(\mathbf{v})_S$$

Proof. Let $(\mathbf{u})_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ and $(\mathbf{v})_S = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{pmatrix}$. Then

$$\begin{aligned} \alpha\mathbf{u} + \beta\mathbf{v} &= \alpha(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) + \beta(d_1\mathbf{u}_1 + d_c\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k) \\ &= (\alpha c_1 + \beta d_1)\mathbf{u}_1 + (\alpha c_2 + \beta d_2)\mathbf{u}_2 + \cdots + (\alpha c_k + \beta d_k)\mathbf{u}_k. \end{aligned}$$

So

$$(\alpha\mathbf{u} + \beta\mathbf{v})_S = \begin{pmatrix} \alpha c_1 + \beta d_1 \\ \alpha c_2 + \beta d_2 \\ \vdots \\ \alpha c_k + \beta d_k \end{pmatrix} = \alpha \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} + \beta \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{pmatrix} = \alpha(\mathbf{u})_S + \beta(\mathbf{v})_S.$$

□

Lemma. *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$. Then for $\mathbf{0}_{n\times 1} \in \mathbb{R}^n$,*

$$(\mathbf{0}_{n\times 1})_S = \mathbf{0}_{k\times 1}.$$

Proof. Exercise. □

Theorem (Criteria for linear independence of a subset in a subspace)**.** *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$. A subset*

$$\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\} \subseteq V$$

is linearly independent if and only if

$$\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\} \subseteq \mathbb{R}^k$$

is linearly independent.

Proof. The theorem follows from the following equivalence derived from two lemmas above,

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0}_{n\times 1} \Leftrightarrow c_1(\mathbf{v}_1)_S + c_2(\mathbf{v}_2)_S + \cdots + c_m(\mathbf{v}_m)_S = \mathbf{0}_{k\times 1}.$$

Suppose $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ is linearly independent. Then if $c_1(\mathbf{v}_1)_S + c_2(\mathbf{v}_2)_S + \cdots + c_m(\mathbf{v}_m)_S = \mathbf{0}_{k\times 1}$, the left side of the equivalence above tells us that necessarily $c_1 = c_2 = \cdots = c_m = 0$. So $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ is a linearly independent subset of $\mathbb{R}^k$.

Conversely, suppose $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ is linearly independent. Then if $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0}_{n\times 1}$, the right side of the equivalence tells us that necessarily $c_1 = c_2 = \cdots = c_m = 0$. So $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ is a linearly independent subset of $\mathbb{R}^n$ $\square$

Corollary. *Let $V \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Any subset of V containing more than k vectors must be linearly dependent.*

Proof. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$. By the previous theorem, a subset $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ of V is linearly independent if and only if $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ is linearly independent in $\mathbb{R}^k$. Thus if $m > k$, then $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ cannot be linearly independent in $\mathbb{R}^k$. $\square$

Theorem. *Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ are bases for a subspace $V \subseteq \mathbb{R}^n$. Then $k = m$.*

Proof. For $i = 1, ..., m$, write the vectors $\mathbf{v}_i$ in coordinate relative to basis S,

$$T' = \{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\} \subseteq \mathbb{R}^k.$$

Since T is linearly independent, by the previous theorem, T' is a linearly independent set in $\mathbb{R}^k$. This means that necessarily $l \leq k$.

We now exchange the roles of S and T. That is, for $i = 1, ..., k$, write the vectors $\mathbf{u}_i$ in S in coordinates relative to the basis T,

$$S' = \{(\mathbf{u}_1)_T, (\mathbf{u}_2)_T, ..., (\mathbf{u}_k)_T\} \subseteq \mathbb{R}^l.$$

Since S is linearly independent, by the previous theorem, S' is a linearly independent set in $\mathbb{R}^k$. This means that necessarily $k \leq l$.

Therefore, we have equality, $k = l$. $\square$

By the criteria of linear independence of a subset in a subspace, we can conclude that if $V \subseteq \mathbb{R}^n$ is a k-dimensional subspace, then any set containing more than k number of vectors in V must be linearly independent. We have a similar criteria for whether a subset spans V.

Theorem (Criteria for a subset to span a subspace)**.** *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$. A subset*

$$\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\} \subseteq V$$

spans V if and only if

$$\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\} \subseteq \mathbb{R}^k$$

spans $\mathbb{R}^k$.

Proof. ($\Rightarrow$) Suppose $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ spans V. Given any $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{pmatrix} \in \mathbb{R}^k$, let

$$\mathbf{v} = w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + \cdots + w_k\mathbf{u}_k \in V.$$

By construction, $(\mathbf{v})_S = \mathbf{w}$. Since $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ spans V, we can find $d_1, d_2, ..., d_m \in \mathbb{R}$ such that $\mathbf{v} = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m$. Then by a lemma above,

$$d_1(\mathbf{v}_1)_S + d_2(\mathbf{v}_2)_S + \cdots + d_m(\mathbf{v}_m)_S = (\mathbf{v})_S = \mathbf{w}.$$

This shows that any $\mathbf{w} \in \mathbb{R}^k$ can be written as a linear combination of $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ...,$ $(\mathbf{v}_m)_S\}$, and thus the set spans $\mathbb{R}^k$.

($\Leftarrow$) Suppose $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ spans $\mathbb{R}^k$. Now given any $\mathbf{v} \in V$, let

$$\mathbf{v} = w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + \cdots + w_k\mathbf{u}_k \in V,$$

that is $(\mathbf{v})_S = \mathbf{w} = (w_i) \in \mathbb{R}^k$. Then since $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ spans $\mathbb{R}^k$, we can find $d_1, d_2, ..., d_m \in \mathbb{R}$ such that $(\mathbf{v})_S = \mathbf{w} = d_1(\mathbf{v}_1)_S + d_2(\mathbf{v}_2)_S + \cdots + d_m(\mathbf{v}_m)_S$. Then by the same lemma above,

$$d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m = (\mathbf{v})_S = \mathbf{v}.$$

This shows that any $\mathbf{v} \in V$ can be written as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$, and thus the set spans V.

$\square$

Corollary. *Let $V \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Any subset of V containing less than k vectors cannot span V.*

Proof. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is a basis for a subspace $V \subseteq \mathbb{R}^n$. Given a subset $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\} \subseteq V$, $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ spans V if and only if $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ spans $\mathbb{R}^k$. But if $m < k$, then if $\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, ..., (\mathbf{v}_m)_S\}$ cannot span $\mathbb{R}^k$, and hence $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ cannot span V. $\square$

Combining all the results we have thus far, we obtain the following statement.

Corollary. *Let V be a k-dimensional subspace of $\mathbb{R}^n$. Then $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ is a basis for V if and only if $m = k$ and the order k square matrix $\mathbf{A} = \begin{pmatrix}(\mathbf{v}_1)_S & (\mathbf{v}_2)_S & \cdots & (\mathbf{v}_m)_S\end{pmatrix}$ is invertible.*

Lemma. *Suppose $U, V \subseteq \mathbb{R}^n$ are subspaces of $\mathbb{R}^n$ such that U is a subset of V, $U \subseteq V$. Then $\dim(U) \leq \dim(V)$, with equality if and only if $U = V$.*

Proof. Suppose $\dim(U) = k$. Let $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ be a basis for U. Since $U \subseteq V$, S is a linearly independent set in V. Suppose S spans V too. Then $U = \text{span}(S) = V$ and thus $\dim(U) = \dim(V)$. Otherwise, $\text{span}(S) \subsetneq V$ is not the whole V. Then we can find a $v \in V \backslash \text{span}(S)$ that is in V but not in the span of S. This would mean that $\{\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{v}\}$ is a linearly independent subset of V, since v is not a linearly combination of the vectors in S. Thus, $\dim(V) \geq k + 1 > k = \dim(U)$. $\square$

Theorem. *The zero space is 0-dimensional. The empty set $\varnothing = \{\}$ is the basis for the zero space.*

Proof. (i) It is vacuously true that the empty set is linearly independent.

(ii) The more general (accurate) definition of the span of a set S is the smallest subspace containing S. It turns out that if S is a finite set, then the span is indeed all possible linear combination of the vectors in S. Now since every subspace contains the empty set as a subset, the smallest one to do so is the zero space. Hence, by definition, the zero space is the span of the empty set.

□

Theorem (Row operations preserve row space)**.** *Suppose* $\mathbf{A}$ *and* $\mathbf{B}$ *are row equivalent matrices. Then* $\text{Row}(\mathbf{A}) = \text{Row}(\mathbf{B})$.

Proof. Suffice to show for the case that $\mathbf{B} = \mathbf{E}\mathbf{A}$ where $\mathbf{E}$ is an elementary matrix. This is because if it was so, then in general, if $\mathbf{B} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$, then $\text{Row}(\mathbf{B}) = \text{Row}(\mathbf{E}_{k-1} \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}) = \cdots = \text{Row}(\mathbf{E}_2\mathbf{E}_1\mathbf{A}) = \text{Row}(\mathbf{E}_1\mathbf{A}) = \text{Row}(\mathbf{A})$.

Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ matrices. Write $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix}$, where $\mathbf{a}_i$ is the i-th row of $\mathbf{A}$ for $i = 1, ..., m$.

Suppose $\mathbf{E}$ corresponds to multiplying row i by a nonzero constant $c \in \mathbb{R}$, $c \neq 0$,

$$\mathbf{E} \leftrightarrow cR_i.$$

Since $c \neq 0$, it is clear that

$$\begin{aligned} \text{Row}(\mathbf{A}) &= \text{span}\{\mathbf{a}_1, ..., \mathbf{a}_{i-1}, \mathbf{a}_i, \mathbf{a}_{i+1}, ..., \mathbf{a}_m\} \\ &= \text{span}\{\mathbf{a}_1, ..., \mathbf{a}_{i-1}, c\mathbf{a}_i, \mathbf{a}_{i+1}, ..., \mathbf{a}_m\} = \text{Row}(\mathbf{B}). \end{aligned}$$

Suppose $\mathbf{E}$ corresponds to a row swap between the i-th row and the j-th row,

$$\mathbf{E} \leftrightarrow R_i \leftrightarrow R_j.$$

Then it is obvious that the row spaces of $\mathbf{A}$ and $\mathbf{B}$ are equal.

Suppose $\mathbf{E}$ corresponds to adding a times of the j-th row to the i-th row for some $a\mathbb{R}$,

$$\mathbf{E} \leftrightarrow R_i + aR_j.$$

Then clearly

$$\text{Row}(\mathbf{B}) = \text{span}\{\mathbf{a}_1, ..., \mathbf{a}_{i-1}, \mathbf{a}_i + a\mathbf{a}_j, \mathbf{a}_{i+1} ..., \mathbf{a}_m\} \subseteq \text{span}\{\mathbf{a}_1, ..., \mathbf{a}_{i-1}, \mathbf{a}_i, \mathbf{a}_{i+1}, ..., \mathbf{a}_m\}.$$

Note that the j-th row of $\mathbf{A}$ and $\mathbf{B}$ are both $\mathbf{a}_j$, and since $\mathbf{a}_i = (\mathbf{a}_i + a\mathbf{a}_j) - a\mathbf{a}_j$, it shows that $\mathbf{a}_i \in \text{Row}(\mathbf{B})$. Hence, $\text{Row}(\mathbf{A}) \subseteq \text{Row}(\mathbf{B})$ too. □

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 9 Notes

References

1. Elementary Linear Algebra: Application Version, Section 4.8
2. Linear Algebra with Application, Section 5.4, 5.6, 8.1

4.2 Rank

Let $\mathbf{A}$ be a $m \times n$ matrix. Recall that the number of pivot columns in the reduced row-echelon form of $\mathbf{A}$ tells us the dimension of the column space of $\mathbf{A}$, and the number of nonzero rows in the reduced row-echelon form of $\mathbf{A}$ tells us the dimension of the row space of $\mathbf{A}$,

$$\begin{aligned} \#\text{ of pivot columns in RREF of } \mathbf{A} &= \dim(\text{Col}(\mathbf{A})), \\ \#\text{ of nonzero rows in RREF of } \mathbf{A} &= \dim(\text{Row}(\mathbf{A})). \end{aligned}$$

Recall also that the number of pivot columns in the reduced row-echelon form of $\mathbf{A}$ is equals to the number of leading entries, which is equals to the number of nonzero rows,

$$\begin{aligned} \#\text{ of leading entries} &= \#\text{ of pivot columns in RREF of } \mathbf{A} \\ &= \#\text{ of nonzero rows in RREF of } \mathbf{A}. \end{aligned}$$

Hence, the dimension of the column space of $\mathbf{A}$ is equals to the dimension of the row space of $\mathbf{A}$.

Theorem. *For any matrix $\mathbf{A}$, the dimension of the row space of $\mathbf{A}$ is equal to the dimension of the column space of $\mathbf{A}$,*

$$\dim(\text{Row}(\mathbf{A})) = \dim(\text{Col}(\mathbf{A})).$$

Define the <u>rank</u> of a matrix $\mathbf{A}$, denoted as $\text{rank}(\mathbf{A})$, to be the dimension of its column or (and) its row space,

$$\text{rank}(\mathbf{A}) = \dim(\text{Col}(\mathbf{A})) = \dim(\text{Row}(\mathbf{A})).$$

Example.

1. $\text{rank}(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$ the zero matrix.
2. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$. So $\text{rank}(\mathbf{A}) = 3$.
3. $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ -1 & 7 & 5 \\ 1 & 9 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\text{rank}(\mathbf{A}) = 3$.

Now since the number of pivot columns in the reduced row-echelon form is at most the number of columns, the rank of a matrix is at most the number columns,

$$\text{rank} = \#\text{ of pivot columns in RREF} \leq \#\text{ of columns}.$$

Also, the number of nonzero rows in the reduced row-echelon form is at most the number rows, the rank of a matrix is at most the number of rows,

$$\text{rank} = \#\text{ nonzero rows in RREF} \leq \#\text{ of rows}.$$

Hence, we have the following lemma.

Lemma. *For a $m \times n$ matrix $\mathbf{A}$,*

$$\text{rank}(\mathbf{A}) \leq \min\{m, n\}.$$

A $m \times n$ matrix $\mathbf{A}$ is said to be of <u>full rank</u> if equality is attained in the above inequality,

$$\text{rank}(\mathbf{A}) = \min\{m, n\}.$$

Example.

1. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$ is of full rank.

2. $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ -1 & 7 & 5 \\ 1 & 9 & 2 \end{pmatrix}$ is of full rank.

3. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. $\mathbf{A}$ is not full rank.

4. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. $\mathbf{A}$ is not of full rank.

Before you read on, try to prove or disprove the following statements.

1. If the rank of $\mathbf{A}$ is equal to the number of columns, then $\mathbf{A}$ has full rank.
2. If the rank of $\mathbf{A}$ is equal to the number of rows, then $\mathbf{A}$ has full rank.
3. If the reduced row-echelon form of $\mathbf{A}$ has a zero row, then $\mathbf{A}$ cannot be of full rank.
4. If the reduced row-echelon form of $\mathbf{A}$ has a non-pivot column, then $\mathbf{A}$ cannot be of full rank.
5. If the reduced row-echelon form of $\mathbf{A}$ has at least one non-pivot column and one nonzero row, then $\mathbf{A}$ cannot be of full rank.
6. If the reduced row-echelon form of $\mathbf{A}$ has no non-pivot columns, then $\mathbf{A}$ has full rank.
7. If the reduced row-echelon form of $\mathbf{A}$ has no zero rows, then $\mathbf{A}$ has full rank.

4.3 Column Space and Consistency of Linear System

Let $\mathbf{A} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}$ be a $n \times k$ matrix, where $\mathbf{u}_i \in \mathbb{R}^n$ is the i-th column of $\mathbf{A}$, for $i = 1, ..., k$. Recall (lecture 6) that $\mathbf{b} \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = \text{Col}(\mathbf{A})$ if and only if $\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v} \end{array}\right)$ is consistent, which is equivalent to $\mathbf{Ax} = \mathbf{v}$ is consistent.

Theorem. *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix. Then* $\text{Col}(\mathbf{A}) = \{ \ \mathbf{Au} \mid \mathbf{u} \in \mathbb{R}^n \ \}$.

Proof. Suppose $\mathbf{v} \in \text{Col}(\mathbf{A})$. Then $\mathbf{Ax} = \mathbf{v}$ is consistent, and thus there is a $\mathbf{u} \in \mathbb{R}^n$ such that $\mathbf{Au} = \mathbf{v}$. So $\mathbf{v} \in \{ \ \mathbf{Au} \mid \mathbf{u} \in \mathbb{R}^n \ \}$. This shows that $\text{Col}(\mathbf{A}) \subseteq \{ \ \mathbf{Au} \mid \mathbf{u} \in \mathbb{R}^n \ \}$.

Conversely, suppose $\mathbf{v} \in \{ \ \mathbf{Au} \mid \mathbf{u} \in \mathbb{R}^n \ \}$, that is, $\mathbf{v} = \mathbf{Au}$ for some $\mathbf{u} \in \mathbb{R}^n$. Then $\mathbf{Ax} = \mathbf{b}$ is consistent, and thus $\mathbf{b} \in \text{Col}(\mathbf{A})$. This shows that $\{ \ \mathbf{Au} \mid \mathbf{u} \in \mathbb{R}^n \ \} \subseteq \text{Col}(\mathbf{A})$. Therefore, we have equality. $\square$

4.4 Nullspace

Recall that the solution space to a homogeneous system is a subspace. We shall call this space the nullspace of the matrix. The <u>nullspace</u> of a $m \times n$ matrix $\mathbf{A}$ is the solution space to the homogeneous system $\mathbf{Ax} = \mathbf{0}$ with coefficient matrix $\mathbf{A}$. It is denoted as

$$\text{Null}(\mathbf{A}) = \{ \ \mathbf{v} \in \mathbb{R}^n \mid \mathbf{Av} = \mathbf{0} \ \}.$$

The <u>nullity</u> of $\mathbf{A}$ is the dimension of the nullspace of $\mathbf{A}$, denoted as

$$\text{nullity}(\mathbf{A}) = \dim(\text{Null}(\mathbf{A})).$$

Recall that the dimension of the solution space of the homogeneous system $\mathbf{Ax} = \mathbf{0}$ is equals to the number of non-pivot columns in the reduced row-echelon form of $\mathbf{A}$. Since the rank of $\mathbf{A}$ is equal to the number of pivot columns in the reduced row-echelon form of $\mathbf{A}$, we have the rank-nullity theorem.

Theorem (Rank-Nullity Theorem)**.** *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix. Then the sum of the rank and nullity of* $\mathbf{A}$ *is equal to the number of columns of* $\mathbf{A}$*,*

$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n.$$

Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{pmatrix}$.

$$\begin{pmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence,

$$\text{Null}(\mathbf{A}) = \text{span}\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 0 \\ 3 \\ 1 \end{pmatrix} \right\}.$$

Also, $\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = 2 + 2 = 4 = \#$ of columns of $\mathbf{A}$.

4.5 Full Rank Matrices

Let $\mathbf{A}$ be a $m \times n$ matrix. Suppose $m \geq n$. Then $\mathbf{A}$ has full rank (equals the number of columns) if and only if the reduced row-echelon form $\mathbf{R}$ of $\mathbf{A}$ has the form

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{0}_{(m-n)\times n} \end{pmatrix}.$$

Theorem (Rank equals to number of columns)**.** *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\text{rank}(A) = n$.

(ii) The rows of $\mathbf{A}$ *spans* $\mathbb{R}^n$, $\text{Row}(\mathbf{A}) = \mathbb{R}^n$.

(iii) The columns of $\mathbf{A}$ *are linearly independent.*

(iv) The homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$ *has only the trivial solution, that is,* $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

(v) $\mathbf{A}^T\mathbf{A}$ *is an invertible matrix of order* n.

(vi) $\mathbf{A}$ *has a left inverse.*

Now suppose $n \geq m$. Then $\mathbf{A}$ has full rank (equals the number of rows) if and only if the reduced row-echelon form $\mathbf{R}$ of $\mathbf{A}$ has the form

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 & \cdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & \cdots & \cdots & 0 & \cdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & \cdots & \cdots & 0 & \cdots \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \cdots & \vdots & \cdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & \cdots & \cdots & 1 & \cdots \end{pmatrix}$$

Theorem (Rank equals to number of rows)**.** *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\text{rank}(A) = m$.

(ii) The columns of $\mathbf{A}$ *spans* $\mathbb{R}^m$, $\text{Col}(\mathbf{A}) = \mathbb{R}^m$.

(iii) The rows of $\mathbf{A}$ *are linearly independent.*

(iv) $\mathbf{A}\mathbf{A}^T$ *is an invertible matrix of order* m.

(v) $\mathbf{A}$ *has a right inverse.*

(vi) The linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ *is consistent for every* $\mathbf{b} \in \mathbb{R}^m$.

Example. 1. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Then $\text{rank}(\mathbf{A}) = 3$ which is the number of columns of $\mathbf{A}$. We have already seen that therefore $\mathbf{A}$ must fulfill the first 4 statements of the theorem above. Now we will show that $\mathbf{A}^T\mathbf{A}$ is invertible, and thus find a left inverse of $\mathbf{A}$.

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix}, \ (\mathbf{A}^T\mathbf{A})^{-1} = \frac{1}{7}\begin{pmatrix} 5 & -2 & -2 \\ -2 & 5 & -2 \\ -2 & -2 & 5 \end{pmatrix}.$$

Then a left inverse of $\mathbf{A}$ is

$$(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \frac{1}{7}\begin{pmatrix} 3 & 3 & -4 & 1 \\ -4 & 3 & 3 & 1 \\ 3 & -4 & 3 & 1 \end{pmatrix}.$$

Consider now the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$. Suppose $\mathbf{u} \in \mathbb{R}^3$ is a solution, $\mathbf{A}\mathbf{u} = \mathbf{0}$. Then

$$\mathbf{u} = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{A}\mathbf{u} = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{0} = \mathbf{0}.$$

So, the system has only the trivial solution.

2. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$. Then $\text{rank}(\mathbf{A}) = 3$ which is the number of rows. We have already seen that $\mathbf{A}$ fulfills the first 3 statements of the theorem, as well as the last statement. We will now show that $\mathbf{A}\mathbf{A}^T$ is invertible, and find a right inverse of $\mathbf{A}$.

$$\mathbf{A}\mathbf{A}^T = \begin{pmatrix} 7 & 3 & 1 \\ 3 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}, \ (\mathbf{A}\mathbf{A}^T)^{-1} = \frac{1}{12}\begin{pmatrix} 5 & -8 & 1 \\ -8 & 20 & -4 \\ 1 & -4 & 5 \end{pmatrix}.$$

So, a right inverse of $\mathbf{A}$ is

$$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \frac{1}{6}\begin{pmatrix} 3 & -6 & 3 \\ -1 & 4 & 1 \\ 1 & 2 & -1 \\ -2 & 2 & 2 \end{pmatrix}.$$

Now given any $\mathbf{b} \in \mathbb{R}^3$, $\mathbf{u} = (\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1})\mathbf{b}$ is a solution to the system $\mathbf{A}\mathbf{x} = \mathbf{b}$. Indeed,

$$\mathbf{A}\mathbf{u} = \mathbf{A}(\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1})\mathbf{b} = \mathbf{b}.$$

So, the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent.

Remark. 1. The two theorem above also show that a non-square matrix can have at most a left or a right inverse, but not both.

2. If a matrix has more rows than columns, then the matrix cannot have a right inverse. For if $\mathbf{A}$ has a right inverse, then rank of $\mathbf{A}$ must be equal to the number of rows, which cannot happen since the rank is bounded by the number of columns, which is strictly smaller than the number of rows.

3. If a matrix has more columns than rows, then the matrix cannot have a left inverse. For if $\mathbf{A}$ has a left inverse, then rank of $\mathbf{A}$ must be equal to the number of columns, which cannot happen since the rank is bounded by the number of rows, which is strictly smaller than the number of columns.

5 Orthogonal Projection

5.1 Orthogonal and Orthonormal basis

Recall that a set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is orthogonal if

$$\mathbf{u}_i \cdot \mathbf{u}_j = 0$$

for all $i \neq j$, and it is orthonormal if

$$\mathbf{u}_i \cdot \mathbf{u}_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

Suppose now S is an orthogonal set of nonzero vectors. Then all the vectors are perpendicular to each other, and thus geometrically it is clear that they must be linearly independent. In fact, intuitively, each time we add a nonzero vector that is orthogonal to a orthogonal set, the space that contains all the vectors increases its dimension by 1.

Theorem. *Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is an orthogonal set of nonzero vectors. Then S is linearly independent.*

Proof. Suppose there are some coefficients $c_1, c_2, ..., c_k \in \mathbb{R}$ such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$

Then for any $i = 1, ..., k$,

$$\begin{aligned} 0 &= \mathbf{u}_i \cdot \mathbf{0} = \mathbf{u}_i \cdot (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) \\ &= c_1(\mathbf{u}_i \cdot \mathbf{u}_1) + c_2(\mathbf{u}_i \cdot \mathbf{u}_2) + \cdots + c_k(\mathbf{u}_i \cdot \mathbf{u}_k) \\ &= c_i(\mathbf{u}_i \cdot \mathbf{u}_i) \end{aligned}$$

since $\mathbf{u}_i \cdot \mathbf{u}_j = 0$. But since $\mathbf{u}_i \neq \mathbf{0}$, $\mathbf{u}_i \cdot \mathbf{u}_i \neq 0$, and hence, necessarily $c_i = 0$. Therefore the only combination of vectors in S to give the zero vector is the trivial one. This shows that S is linearly independent. □

Corollary. *Every orthonormal set is linearly independent.*

Proof. Follows from the fact that an orthonormal set is orthogonal, and since the vectors are unit vectors, they are nonzero. □

Let $V \subseteq \mathbb{R}^n$ be a subspace. A set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is an <u>orthogonal basis</u> for V if it is a basis for V and is an orthogonal set. Similarly, S is an <u>orthonormal basis</u> for V if it is a basis for V and is an orthonormal set.

Corollary.

1. *Any orthogonal set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ containing n nonzero vectors in $\mathbb{R}^n$ is an orthogonal basis.*
2. *Any orthonormal set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ containing n vectors in $\mathbb{R}^n$ is an orthonormal basis.*

Proof. Follows from the fact that a linearly independent set containing n vectors must be a basis for $\mathbb{R}^n$. □

Corollary. *Suppose $V \subseteq \mathbb{R}^n$ is a k-dimensional subspace.*

1. *Any orthogonal set* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *containing* k *nonzero vectors in* V *is an orthogonal basis for* V*.*

2. *Any orthonormal set* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *containing* k *vectors in* V *is an orthonormal basis for* V*.*

Proof. Exercise. $\square$

The coordinates of a vector relative to an orthogonal or orthonormal basis is rather easy to compute.

Theorem. *Let* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *be an orthogonal basis for a subspace* $V \subseteq \mathbb{R}^n$*. Then for any* $\mathbf{v} \in V$*,*

$$\mathbf{v} = \left(\frac{\mathbf{v}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\right)\mathbf{u}_1 + \left(\frac{\mathbf{v}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}\right)\mathbf{u}_2 + \cdots + \left(\frac{\mathbf{v}\cdot\mathbf{u}_k}{\|\mathbf{u}_k\|^2}\right)\mathbf{u}_k,$$

that is,

$$(\mathbf{v})_S = \begin{pmatrix} (\mathbf{v}\cdot\mathbf{u}_1)/\|\mathbf{u}_1\|^2 \\ (\mathbf{v}\cdot\mathbf{u}_2)/\|\mathbf{u}_2\|^2 \\ \vdots \\ (\mathbf{v}\cdot\mathbf{u}_k)/\|\mathbf{u}_k\|^2 \end{pmatrix}.$$

If further S *is an orthonormal basis, then*

$$\mathbf{v} = (\mathbf{v}\cdot\mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{v}\cdot\mathbf{u}_2)\,\mathbf{u}_2 + \cdots + (\mathbf{v}\cdot\mathbf{u}_k)\,\mathbf{u}_k,$$

that is,

$$(\mathbf{v})_S = \begin{pmatrix} \mathbf{v}\cdot\mathbf{u}_1 \\ \mathbf{v}\cdot\mathbf{u}_2 \\ \vdots \\ \mathbf{v}\cdot\mathbf{u}_k \end{pmatrix}.$$

Proof. The second statement follows immediately from the first. Suppose S is orthogonal. For a $\mathbf{v} \in V$, since S is a basis, we can write $\mathbf{v}$ as a linear combination the vectors in S

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$$

for some $c_1, c_2, ..., c_k \in \mathbb{R}$. Then for each $i = 1, ..., k$,

$$\begin{aligned} \mathbf{u}_i\cdot\mathbf{v} &= \mathbf{u}_i\cdot(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) \\ &= c_1(\mathbf{u}_i\cdot\mathbf{u}_1) + c_2(\mathbf{u}_i\cdot\mathbf{u}_2) + \cdots + c_k(\mathbf{u}_i\cdot\mathbf{u}_k) \\ &= c_i\|\mathbf{u}_i\|^2. \end{aligned}$$

Hence, $c_i = \frac{\mathbf{v}\cdot\mathbf{u}_1}{\|\mathbf{u}_i\|^2}$ for all $i = 1, ..., k$. $\square$

Remark. Note that this only works if $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is an orthogonal or orthonormal basis. For example, let $S = \left\{\mathbf{u}_1 = \begin{pmatrix}1\\1\\0\end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix}1\\0\\0\end{pmatrix}\right\}$, and $\mathbf{w} = \begin{pmatrix}2\\1\\0\end{pmatrix}$. Then

$$\left(\frac{\mathbf{w}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\right)\mathbf{u}_1 + \left(\frac{\mathbf{w}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}\right)\mathbf{u}_2 = \frac{3}{2}\begin{pmatrix}1\\1\\0\end{pmatrix} + 2\begin{pmatrix}1\\0\\0\end{pmatrix} = \frac{1}{2}\begin{pmatrix}7\\3\\0\end{pmatrix} \neq \mathbf{w}.$$

Why is this so?

Example. 1. The standard basis $E = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ is an orthonormal basis. Given any $\mathbf{v} = (v_i) \in \mathbb{R}^n$,

$$\mathbf{e}_i \cdot \mathbf{v} = v_i.$$

Thus, $(\mathbf{v})_E = (\mathbf{e}_i \cdot \mathbf{v}) = (v_i) = \mathbf{v}$.

2. $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$ is an orthogonal basis for V, the xy-plane in $\mathbb{R}^3$. For any $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \in V$,

$$\begin{aligned} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} &= \left(\frac{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}{\left\| \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\|^2} \right) \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \left(\frac{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}}{\left\| \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\|^2} \right) \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\ &= \frac{(x+y)}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \frac{(x-y)}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}. \end{aligned}$$

That is,

$$(\mathbf{v})_S = \frac{1}{2} \begin{pmatrix} x+y \\ x-y \end{pmatrix}.$$

3. Let $S = \left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$ and $V = \text{span}(S)$. S is an orthonormal basis for V. Let $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \in V$. Then

$$\begin{aligned} \mathbf{v} &= \left(\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right) \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right) \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\ &= (\sqrt{3}) \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (\sqrt{2}) \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \end{aligned}$$

That is,

$$(\mathbf{v})_S = \begin{pmatrix} \sqrt{3} \\ \sqrt{2} \end{pmatrix}.$$

Exercise: Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for a subspace V. Let

$$T = \left\{ \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}, \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \right\}$$

be the orthonormal basis for V obtained from normalizing S. Suppose $\mathbf{v} \in V$ is such that

$$(\mathbf{v})_S = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}.$$

What is $(\mathbf{v})_T$?

5.2 Orthogonal Projection

Now suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is an orthonormal basis for a subspace $V \subseteq \mathbb{R}^n$. Let $\mathbf{w} \in \mathbb{R}^n$ be a vector not in V, $\mathbf{w} \notin V$. How do we interpret the vector

$$\mathbf{w}' = (\mathbf{w} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{w} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{w} \cdot \mathbf{u}_k)\mathbf{u}_k?$$

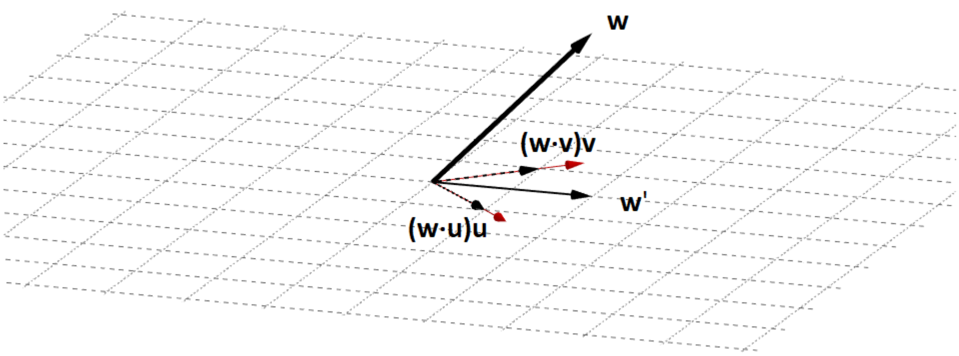
Example. Let $\left\{ \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ be an orthonormal basis for the xy-plane in $\mathbb{R}^3$. Let $\mathbf{w} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Then

$$\mathbf{w}' = (\mathbf{w} \cdot \mathbf{e}_1)\mathbf{e}_1 + (\mathbf{w} \cdot \mathbf{e}_2)\mathbf{e}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Consider now another orthonormal basis $\left\{ \mathbf{v} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$ for the xy-plane. Then again

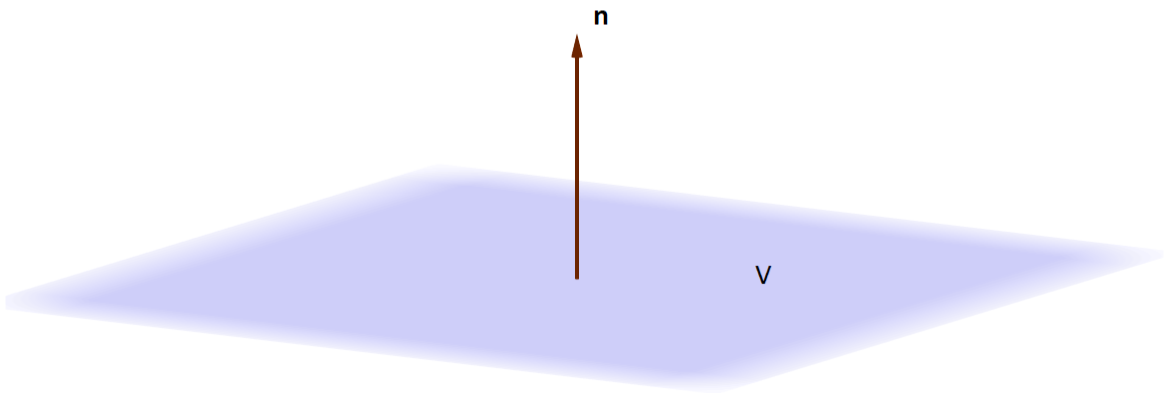
$$\mathbf{w}' = (\mathbf{w} \cdot \mathbf{v})\mathbf{v} + (\mathbf{w} \cdot \mathbf{u})\mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

In both (any) orthonormal basis for the xy-plane, $\mathbf{w}'$ is the same. Geometrically, it is the projection of $\mathbf{w}$ onto the xy-plane.



To show this rigorously, we need to define what is the projection of a vector onto a subspace.

Let $V \subseteq \mathbb{R}^n$ be a subspace. A vector $\mathbf{n} \in \mathbb{R}^n$ is orthogonal to V if for every $\mathbf{v} \in V$, $\mathbf{n} \cdot \mathbf{v} = 0$, that is, $\mathbf{n}$ is orthogonal to every vector in V. We will denote it as $\mathbf{n} \perp V$.



If $\mathbf{n} \neq \mathbf{0}$.

Example. 1. Let $\mathbf{w} = \mathbf{0}$. Then for any subspace $V \subseteq \mathbb{R}^n$, $\mathbf{w} \perp V$.

2. Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \, ax + by + cz = 0 \right\}$. Any vector of the form $s \begin{pmatrix} a \\ b \\ c \end{pmatrix}$, $s \in \mathbb{R}$, is orthogonal to V, since

$$s \begin{pmatrix} a \\ b \\ c \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = s(ax + by + cz) = 0,$$

for any $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in V$. This shows that a plane in $\mathbb{R}^3$ has the expression

$$V = \{ \ \mathbf{v} \in \mathbb{R}^3 \mid \mathbf{v} \cdot \mathbf{n} = 0 \ \}$$

for some $\mathbf{n} \in \mathbb{R}^3$, $\mathbf{n} \neq \mathbf{0}$. $\mathbf{n}$ is said to be normal to the plane V.

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a spanning set for V, span$(S) = V$. Then $\mathbf{w} \in \mathbb{R}^n$ is orthogonal to V if and only if $\mathbf{w} \cdot \mathbf{u}_i = 0$ for all $i = 1, ..., k$.*

Geometrically, this is clear. If $\mathbf{w}$ is the zero vector, then the statement is obvious. Otherwise, since $\mathbf{w}$ is perpendicular to every vector in the spanning set, then it must be pointing in the direction that is perpendicular to the entire space spanned by the spanning set. Refer to the appendix for the detailed proof.

This gives us a means to find vectors that are orthogonal to a subspace $V \subseteq \mathbb{R}^n$.

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for V. Then $\mathbf{w} \perp V$ if and only if $\mathbf{w} \in \text{Null}(\mathbf{A}^T)$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.*

Note that since we write vectors as column vectors, there is a need to take the transpose of $\mathbf{A}$. If we write the vectors in the basis S as row vectors and $\mathbf{A}$ is the matrix whose i-th rows is $\mathbf{u}_i$, then there is no need to take the transpose in the theorem above.

Proof. By the previous theorem, $\mathbf{w} \perp V \Leftrightarrow \mathbf{u}_i^T\mathbf{w} = \mathbf{u}_i \cdot \mathbf{w} = 0$ for all $i = 1, ..., k \Leftrightarrow$ $\mathbf{A}^T\mathbf{w} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_k^T \end{pmatrix} \mathbf{w} = \mathbf{0}$. □

Example. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \right\}$ and $V = \text{span}(S)$. Then $\mathbf{w} \perp V$ if and only if $\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & -1 & 0 \end{pmatrix} \mathbf{w} = 0$.

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & -1 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & -1 & 0 \end{pmatrix} \Rightarrow \mathbf{w} \perp V \Leftrightarrow \mathbf{w} \in \text{span} \left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

We are now ready to define the orthogonal projection of a vector onto a subspace.

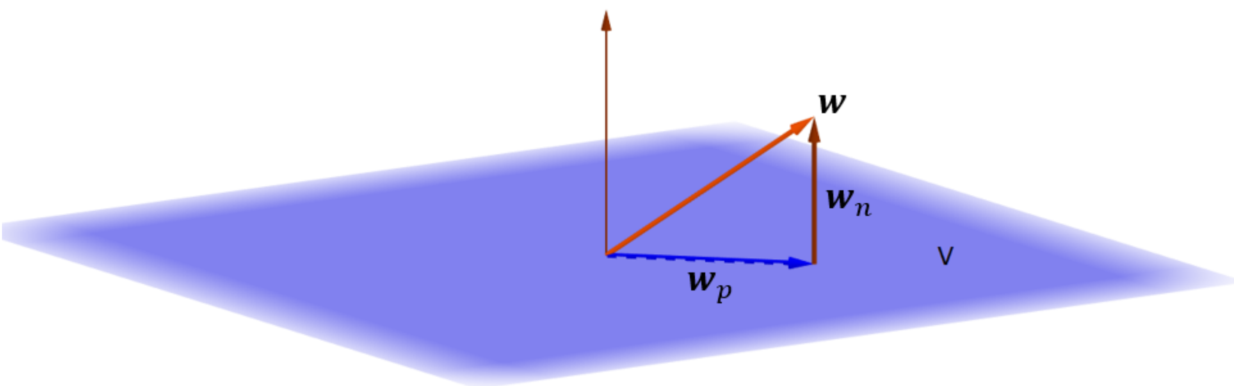
Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace. Every vector $\mathbf{w} \in \mathbb{R}^n$ can be decomposed uniquely as*

$$\mathbf{w} = \mathbf{w}_p + \mathbf{w}_n,$$

where $\mathbf{w}_n$ is orthogonal to V and $\mathbf{w}_p \in V$.

The unique vector $\mathbf{w}_p$ in V is called the orthogonal projection of $\mathbf{w}$ onto V. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be an orthonormal basis for V. Then

$$\mathbf{w}_p = (\mathbf{w} \cdot \mathbf{u}_1)\mathbf{u} + (\mathbf{w} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{w} \cdot \mathbf{u}_k)\mathbf{u}_k.$$



Example. $S = \left\{ \mathbf{v} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$ is an orthonormal basis for the xy-plane in $\mathbb{R}^3$. Let $\mathbf{w} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. Then

$$\mathbf{w}_p = (\mathbf{w} \cdot \mathbf{v})\mathbf{v} + (\mathbf{w} \cdot \mathbf{u})\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \text{ and } \mathbf{w}_n = \mathbf{w} - \mathbf{w}_p = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Appendix to Lecture 9

Lemma. *For any matrix* $\mathbf{A}$*,*

$$\text{Null}(\mathbf{A}^T\mathbf{A}) = \text{Null}(\mathbf{A}).$$

Proof. ($\supseteq$) Suppose $\mathbf{u} \in \text{Null}(\mathbf{A})$. Then $\mathbf{Au} = \mathbf{0}$. Pre-multiplying both sides of the equation by $\mathbf{A}^T$, we get $\mathbf{A}^T\mathbf{Au} = \mathbf{0}$. Thus $\mathbf{u} \in \text{Null}(\mathbf{A}^T\mathbf{A})$.

($\subseteq$) Suppose $\mathbf{u} \in \text{Null}(\mathbf{A}^T\mathbf{A})$. Then $\mathbf{A}^T\mathbf{Au} = \mathbf{0}$. Pre-multiplying both sides of the equation by $\mathbf{u}^T$, we have

$$0 = \mathbf{u}^T\mathbf{A}^T\mathbf{Au} = (\mathbf{Au})^T(\mathbf{Au}) = \|\mathbf{Au}\|^2.$$

By a property of norm, this tells us the that vector $\mathbf{Au}$ must be the zero vector, $\mathbf{Au} = \mathbf{0}$. Hence, $\mathbf{u} \in \text{Null}(\mathbf{A})$.

□

Corollary. *For any matrix* $\mathbf{A}$*,*

$$\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}^T\boldsymbol{A}).$$

Proof. Let $\mathbf{A}$ be a $m \times n$ matrix. From the previous lemma, $\text{nullity}(\mathbf{A}^T\mathbf{A}) = \text{nullity}(\mathbf{A})$. Note that $\mathbf{A}^T\mathbf{A}$ is a $n \times n$ matrix. Then by the rank-nullity theorem,

$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n = \text{rank}(\mathbf{A}^T\mathbf{A}) + \text{nullity}(\mathbf{A}^T\mathbf{A}) = \text{rank}(\mathbf{A}^T\mathbf{A}) + \text{nullity}(\mathbf{A}),$$

and so $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T\mathbf{A})$. □

Theorem (Rank Equals to Number of Columns)**.** *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\text{rank}(A) = n$.

(ii) The rows of $\mathbf{A}$ *spans* $\mathbb{R}^n$, $\text{Row}(\mathbf{A}) = \mathbb{R}^n$.

(iii) The columns of $\mathbf{A}$ *are linearly independent.*

(iv) The homogeneous system $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution, that is,* $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

(v) $\mathbf{A}^T\mathbf{A}$ *is an invertible matrix of order* n.

(vi) $\mathbf{A}$ *has a left inverse.*

Proof. We have already shown the equivalence of the first 4 statements.

Using the result $\text{Null}(\mathbf{A}) = \text{Null}(\mathbf{A}^T\mathbf{A})$, and the fact that a square matrix is invertible if and only if its nullspace is trivial, we have $\mathbf{A}^T\mathbf{A}$ is invertible if and only if $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

Suppose $\mathbf{A}^T\mathbf{A}$ is invertible. Then

$$\mathbf{I} = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{A}) = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{A}$$

shows that $((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)$ is a left inverse of $\mathbf{A}$. So (v) $\Rightarrow$ (vi).

Conversely, suppose $\mathbf{A}$ has a left inverse $\mathbf{C}$. Then if $\mathbf{u} \in \mathbb{R}^n$ is such that $\mathbf{Au} = \mathbf{0}$, we have
$$\mathbf{u} = \mathbf{CAu} = \mathbf{C0} = \mathbf{0}.$$
So the nullspace of $\mathbf{A}$ is trivial, which is equivalent to $\mathbf{A}^T\mathbf{A}$ being invertible. Hence, we have shown (vi) $\Rightarrow$ (v). $\square$

Theorem (Rank Equals to Number of Rows)**.** *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\operatorname{rank}(A) = m$.

(ii) The columns of $\mathbf{A}$ *spans* $\mathbb{R}^m$*,* $\operatorname{Col}(\mathbf{A}) = \mathbb{R}^m$.

(iii) The rows of $\mathbf{A}$ *are linearly independent.*

(iv) $\mathbf{AA}^T$ *is an invertible matrix of order* m.

(v) $\mathbf{A}$ *has a right inverse.*

(vi) The linear system $\mathbf{Ax} = \mathbf{b}$ *is consistent for every* $\mathbf{b} \in \mathbb{R}^m$.

Proof. By replacing $\mathbf{A}$ with $\mathbf{A}^T$ in the previous theorem, we have the equivalence of statements (i) to (v).

Suppose $\mathbf{A}$ has a right inverse $\mathbf{B}$, that is, $\mathbf{AB} = \mathbf{I}$. Then given any $\mathbf{b} \in \mathbb{R}^m$, let $\mathbf{u} = \mathbf{Bb}$, and we have
$$\mathbf{Au} = \mathbf{ABb} = \mathbf{Ib} = \mathbf{b}.$$
So $\mathbf{u}$ is a solution to $\mathbf{Ax} = \mathbf{b}$, and hence the system is consistent. This shows (v) $\Rightarrow$ (vi).

Conversely, suppose $\mathbf{Ax} = \mathbf{b}$ is consistent for every $b \in \mathbb{R}^m$. Let $\mathbf{b}_i$ be a solution to $\mathbf{Ax} = \mathbf{e}_i$, where $\mathbf{e}_i$ is the i-th vector in the standard basis, for $i = 1, ..., m$. Let $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_m \end{pmatrix}$. Then
$$\mathbf{AB} = \begin{pmatrix} \mathbf{Ab}_1 & \mathbf{Ab}_2 & \cdots & \mathbf{Ab}_m \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_m \end{pmatrix} = \mathbf{I}_m.$$
This shows that $\mathbf{A}$ has a right inverse. Thus we have shown (iv) $\Rightarrow$ (v). $\square$

Theorem. *Let* $V \subseteq \mathbb{R}^n$ *be a subspace and* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *be a spanning set for* V*, span*$(S) = V$*. Then* $\mathbf{w} \in \mathbb{R}^n$ *is orthogonal to* V *if and only if* $\mathbf{w} \cdot \mathbf{u}_i = 0$ *for all* $i = 1, ..., k$.

Proof.

($\Rightarrow$) Suppose $\mathbf{w}$ is orthogonal to V. Since $S \subseteq V$, then $\mathbf{w}$ must also be orthogonal to every vector in S, that is, $\mathbf{w} \cdot \mathbf{u}_i = 0$ for all $i = 1, ..., k$.

($\Leftarrow$) Suppose $\mathbf{w}$ is orthogonal to all the vectors in the spanning set S. Given any $\mathbf{v} \in V$, $\mathbf{v}$ can be written as a linear combination of the vectors in S, that is, there are some $c_1, c_2, ..., c_k \in \mathbb{R}$ such that $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$. Then
$$\begin{aligned} \mathbf{w} \cdot \mathbf{v} &= \mathbf{w} \cdot (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) \\ &= c_1(\mathbf{w} \cdot \mathbf{u}_1) + c_2(\mathbf{w} \cdot \mathbf{u}_2) + \cdots + c_k(\mathbf{w} \cdot \mathbf{u}_k) \\ &= c_1 0 + c_2 0 + \cdots + c_k 0 = 0. \end{aligned}$$
This shows that $\mathbf{w}$ is orthogonal to every $\mathbf{v} \in V$. Hence, $\mathbf{w} \perp V$. $\square$

The theorem also shows that if V is a subspace, then the set of all vectors orthogonal to V is also a subspace, since it is the nullspace of some matrix.

Corollary. *Let $V \subseteq \mathbb{R}^n$ be a subspace. The set $\{\ \mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} \perp V\ \}$ is also a subspace of $\mathbb{R}^n$.*

The set in the above corollary is called the orthogonal complement of V, and is denoted as $V^{\perp}$ (pronounced as V perp, short for perpendicular),

$$V^{\perp} = \{\ \mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} \perp V\ \}.$$

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace. Every vector $\mathbf{w} \in \mathbb{R}^n$ can be decomposed uniquely as*

$$\mathbf{w} = \mathbf{w}_p + \mathbf{w}_n,$$

where $\mathbf{w}_n$ is orthogonal to V and $\mathbf{w}_p \in V$.

Proof. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be an orthonormal basis for V. Define

$$\mathbf{w}_p = (\mathbf{w} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{w} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{w} \cdot \mathbf{u}_k)\mathbf{u}_k.$$

Then $\mathbf{w}_n = \mathbf{w} - \mathbf{w}_p$. By construction, $\mathbf{w}_p \in V$. To show that $\mathbf{w}_n$ is orthogonal to V, suffice to show that $\mathbf{w}_n \cdot \mathbf{u}_i = 0$ for all $i = 1, ..., k$. Indeed,

$$\begin{aligned} \mathbf{u}_i \cdot \mathbf{w}_n &= \mathbf{u}_i \cdot (\mathbf{w} - \mathbf{w}_p) \\ &= \mathbf{u}_i \cdot \mathbf{w} - (\mathbf{w} \cdot \mathbf{u}_1)\mathbf{u}_i \cdot \mathbf{u}_1 - \cdots - (\mathbf{w} \cdot \mathbf{u}_i)\mathbf{u}_i \cdot \mathbf{u}_i - \cdots - (\mathbf{w} \cdot \mathbf{u}_k)\mathbf{u}_i \cdot \mathbf{u}_k \\ &= \mathbf{u}_i \cdot \mathbf{w} - \mathbf{u}_i \cdot \mathbf{w} = 0, \end{aligned}$$

since S is an orthonormal set. We will now show that $\mathbf{w}_p$ and $\mathbf{w}_n$ are unique. Suppose $\mathbf{w} = \mathbf{w}'_p + \mathbf{w}'_n$ is another decomposition with $\mathbf{w}'_p \in V$ and $\mathbf{w}'_n \perp V$. Then

$$\mathbf{w}'_p + \mathbf{w}'_n = \mathbf{w}_p + \mathbf{w}_n \Rightarrow \mathbf{w}'_p - \mathbf{w}_p = \mathbf{w}_n - \mathbf{w}'_n.$$

But since $\mathbf{w}'_p$ and $\mathbf{w}_p$ are in V, then so is $\mathbf{w}'_p - \mathbf{w}_p = \mathbf{w}_n - \mathbf{w}'_n$. But since $\mathbf{w}_n$ and $\mathbf{w}'_n$ are orthogonal to V, then so is their linear combination (since the set of all vectors orthogonal to V is a subspace), in particular, $(\mathbf{w}_n - \mathbf{w}'_n) \perp V$. In other words, $\mathbf{w}'_p - \mathbf{w}_p = \mathbf{w}_n - \mathbf{w}'_n$ is a vector that is both in V and orthogonal to V. This can only happen if and only if it is the zero vector. Hence,

$$\mathbf{w}'_p - \mathbf{w}_p = \mathbf{w}_n - \mathbf{w}'_n = \mathbf{0} \Rightarrow \mathbf{w}'_p = \mathbf{w}_p \text{ and } \mathbf{w}_n = \mathbf{w}'_n.$$

This concludes the proof that the decomposition is unique. $\square$

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 10 Notes

References

1. Elementary Linear Algebra: Application Version, Section 5.1, 6.3-6.4
2. Linear Algebra with Application, Section 3.3, 8.1

5.3 Gram-Schmidt Process

We will now provide an algorithm to construct orthogonal and orthonormal basis for a subspace $V \subseteq \mathbb{R}^n$.

Let $V \subseteq \mathbb{R}^n$ be a subspace, and suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$is a basis for V. Our goal is to first construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ (from S), and then by normalizing, we obtain an orthonormal basis

$$\left\{\mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \mathbf{w}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, ..., \mathbf{w}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}.$$

We will construct the $\mathbf{v}_i$ inductively, for $i = 1, ..., k$. We will first let $\mathbf{v}_1 = \mathbf{u}_1$. Then $\text{span}\{\mathbf{v}_1\} = \text{span}\{\mathbf{u}_1\}$ and $\{\mathbf{v}_1\}$ is an orthogonal set.

Suppose we have constructed an orthogonal set $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\}$ such that $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$. We then find the projection of $\mathbf{u}_i$ onto the subspace $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$. From lecture 9, we know that since $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\}$ is an orthogonal set, the projection is given by

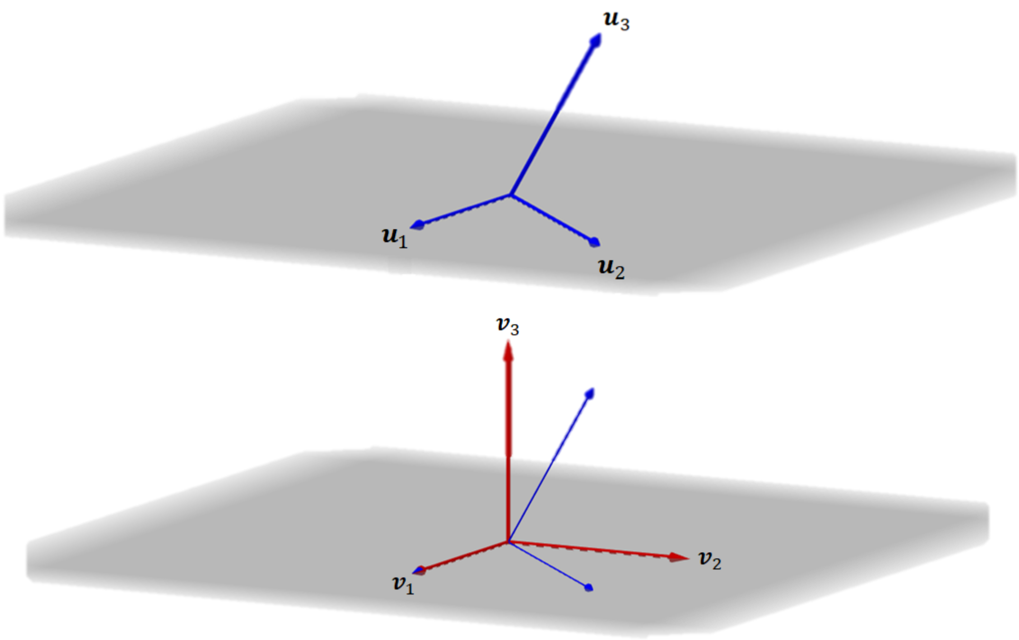
$$\left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_i}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 + \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_i}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 + \cdots + \left(\frac{\mathbf{v}_{i-1} \cdot \mathbf{u}_i}{\|\mathbf{v}_{i-1}\|^2}\right)\mathbf{v}_{i-1}.$$

So, if we let

$$\mathbf{v}_i = \mathbf{u}_i - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_i}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_i}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{i-1} \cdot \mathbf{u}_i}{\|\mathbf{v}_{i-1}\|^2}\right)\mathbf{v}_{i-1} \tag{4}$$

then $\mathbf{v}_i$ is the subtraction of $\mathbf{u}_i$ from its projection onto $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$. Thus, by construction, $\mathbf{v}_i$ is orthogonal to the subspace $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$ and so it is orthogonal to $\mathbf{v}_j$ for all $j < i$. Therefore $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$ is an orthogonal set. By equation (4), $\mathbf{u}_i$ is a linear combination of $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$, we have $\text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}, \mathbf{u}_i\} \subseteq \text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$. But since $\mathbf{v}_j$ is a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_{i-1}$ for $j < i$, from equation (4), we can see that $\mathbf{v}_i$ is a linear combination of $\mathbf{u}_1, ..., \mathbf{u}_{i-1}, \mathbf{u}_i$. Thus $\text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}, \mathbf{u}_i\} = \text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$.

We will give a visualization of the Gram-Schmidt process in $\mathbb{R}^3$.



Theorem (Gram-Schmidt Process). *Let* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *be a linearly independent set. Let*

$$
\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1 \\
\mathbf{v}_2 &= \mathbf{u}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{u}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_3}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_3}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_i &= \mathbf{u}_i - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_i}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_i}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{i-1} \cdot \mathbf{u}_i}{\|\mathbf{v}_{i-1}\|^2}\right)\mathbf{v}_{i-1} \\
&\vdots \\
\mathbf{v}_k &= \mathbf{u}_k - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_k}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_k}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{k-1} \cdot \mathbf{u}_k}{\|\mathbf{v}_{k-1}\|^2}\right)\mathbf{v}_{k-1}.
\end{aligned}
$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ *is an orthogonal set of nonzero vectors. Hence,*

$$\left\{\mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \mathbf{w}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, ..., \mathbf{w}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}$$

is an orthonormal set such that $span\{\mathbf{w}_1, ..., \mathbf{w}_k\} = span\{\mathbf{u}_k, ..., \mathbf{u}_k\}$.

Exercise:

1. Show that $\mathbf{v}_i \neq \mathbf{0}$ for all $i = 1, ..., k$ (otherwise $\frac{1}{\|\mathbf{v}_i\|^2}$ is undefined).
2. What happens if S is not linearly independent?

Example. We will use the Gram-Schmidt process to convert $S = \left\{\begin{pmatrix}1\\2\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\2\end{pmatrix}\right\}$ into an orthonormal basis for $\mathbb{R}^3$.

$$\begin{aligned}
\mathbf{v}_1 &= \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \\
\mathbf{v}_2 &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1+2+1}{1^2+2^2+1^2}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \text{ let } \mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \text{ instead} \\
\mathbf{v}_3 &= \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} - \frac{1+2+2}{1^2+2^2+1^2}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \frac{1-1+2}{1^2+(-1)^2+1^2}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \\
&\quad \text{let } \mathbf{v}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \text{ instead.}
\end{aligned}$$

Why are we allowed to take $\mathbf{v}_2$ and $\mathbf{v}_3$ to be a multiple of the original vector found?

$$\text{So } \left\{ \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} \text{ is an orthonormal basis for } \mathbb{R}^3.$$

5.4 Least Square Approximation

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Recall that $\mathbf{b} \in \mathrm{Col}(\mathbf{A})$ if and only if $\mathbf{Ax} = \mathbf{b}$ is consistent.

More often, in real life application, the data collected or the mathematical model used might not be very accurate, and thus there might not be a solution to the system $\mathbf{Ax} = \mathbf{b}$. However, we might want to find an approximation (vis-à-vis dismissing the model or data) $\mathbf{b}'$ such that $\mathbf{Ax} = \mathbf{b}'$ is consistent, and $\mathbf{b}'$ is the "closest' to $\mathbf{b}$. In other words, we want a $\mathbf{b}' \in \mathrm{Col}(\mathbf{A})$ in the column space of $\mathbf{A}$ that has the shortest distance to $\mathbf{b}$,

$$\|\mathbf{b}' - \mathbf{b}\| = d(\mathbf{b}', \mathbf{b}) \leq d(\mathbf{w}, \mathbf{b}) = \|\mathbf{w} - \mathbf{b}\|$$

for any $\mathbf{w} \in \mathrm{Col}(\mathbf{A})$.

Since $\mathbf{Ax} = \mathbf{b}'$ is consistent, we can find a $\mathbf{u} \in \mathbb{R}^n$ such that $\mathbf{Au} = \mathbf{b}'$. Recall also that $\mathrm{Col}(\mathbf{A}) = \{ \mathbf{Av} \mid \mathbf{v} \in \mathbb{R}^n \}$. So, $\mathbf{Au} = \mathbf{b}'$ being the closest in the column space $\mathrm{Col}(\mathbf{A})$ to $\mathbf{b}$ means

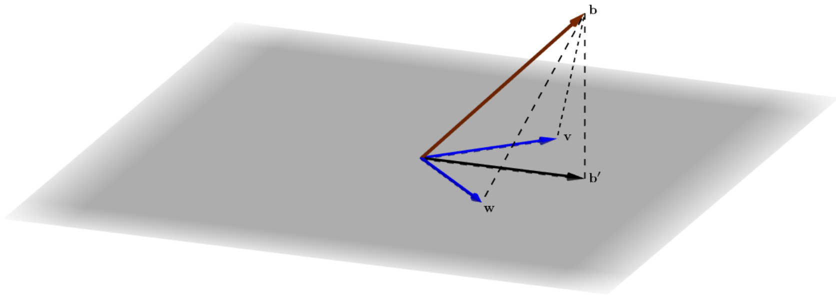
$$\|\mathbf{Au} - \mathbf{b}\| \leq \|\mathbf{Av} - \mathbf{b}\|$$

for any $v \in \mathbb{R}^n$. Then $\mathbf{u}$ is the best approximation to the system $\mathbf{Ax} = \mathbf{b}$. We will now give the formal definition.

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. A vector $\mathbf{u} \in \mathbb{R}^n$ is a <u>least square solution</u> to $\mathbf{Ax} = \mathbf{b}$ if for every vector $\mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{Au} - \mathbf{b}\| \leq \|\mathbf{Av} - \mathbf{b}\|.$$

Geometrically, the vector $\mathbf{b}'$ in a subspace V closest to a vector $\mathbf{b}$ is the projection of $\mathbf{b}$ onto the subspace V.



Theorem. *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix and* $\mathbf{b} \in \mathbb{R}^m$*.* $\mathbf{u} \in \mathbb{R}^n$ *is a least square solution to* $\mathbf{Ax} = \mathbf{b}$ *if and only if* $\mathbf{Au}$ *is the projection of* $\mathbf{b}$ *onto the column space of* $\mathrm{Col}(\mathbf{A})$*.*

The previous theorem gives us an algorithm to find least square solutions. For $\mathbf{u}$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{Au}$ is the projection of $\mathbf{b}$ onto $\mathrm{Col}(\mathbf{A})$, which is equivalent to $(\mathbf{Au} - \mathbf{b}) \perp \mathrm{Col}(\mathbf{A})$. Recall from lecture 9 that $\mathbf{w} \perp \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ if and only if $\mathbf{w} \in \mathrm{Null}(\mathbf{A}^T)$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$. Hence, $\mathbf{u}$ is a least square solution of $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{A}^T(\mathbf{Au} - \mathbf{b}) = \mathbf{0}$, or $\mathbf{A}^T\mathbf{Au} = \mathbf{A}^T\mathbf{b}$. Thus, we arrive at the following theorem.

Theorem. *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix and* $\mathbf{b} \in \mathbb{R}^m$*. A vector* $\mathbf{u} \in \mathbb{R}^n$ *is a least square solution to* $\mathbf{Ax} = \mathbf{b}$ *if and only if* $\mathbf{u}$ *is a solution to* $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$*.*

Remark.
1. Note that even though projection is unique, least square solution may not be unique.
2. The previous theorem seems to suggest that $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ is always consistent. Why is that so?

Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Then

$$\left(\begin{array}{cccc|c} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right).$$

This shows that $\mathbf{b} \notin \mathrm{Col}(\mathbf{A})$. We will find a least square solution of $\mathbf{Ax} = \mathbf{b}$.

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 & 2 \\ 3 & 6 & 3 & 3 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 1 & 2 \end{pmatrix}, \ \mathbf{A}^T\mathbf{b} = \begin{pmatrix} 2 \\ 4 \\ 2 \\ 2 \end{pmatrix}$$

$$\left(\begin{array}{cccc|c} 2 & 3 & 1 & 2 & 2 \\ 3 & 6 & 3 & 3 & 4 \\ 1 & 3 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 & 2 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 2/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

A general solution is $\begin{pmatrix} 0 \\ 2/3 \\ 0 \\ 0 \end{pmatrix} + s\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, $s, t \in \mathbb{R}$. We may choose $s = 0 = t$, then $\mathbf{u} = \begin{pmatrix} 0 \\ 2/3 \\ 0 \\ 0 \end{pmatrix}$ is a least square solution. Then the projection of $\mathbf{b}$ onto the column space is $\mathbf{Au} = \frac{2}{3}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

Notice that $\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ are in the nullspace of $\mathbf{A}$, and hence for any choice of s and t, the projection $\mathbf{Au}$ is unique.

The theorem above also gives us an alternative way to find orthgonal projection without having to (use Gram-Schmidt Process to) construct an orthonormal basis.

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for V. Then the orthogonal projection of a vector $\mathbf{w} \in \mathbb{R}^n$ onto V is $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.*

Proof. Since S is a basis for V, the columns of $\mathbf{A}$ are linearly independent, and thus $\mathbf{A}^T\mathbf{A}$ is invertible. Hence,

$$\mathbf{u} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}$$

is the unique solution to $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{w}$. And hence, $\mathbf{Au} = \mathbf{A}((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w})$ is the projection of $\mathbf{w}$ onto $\mathrm{Col}(\mathbf{A}) = V$. $\square$

Example. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$ and $V = \mathrm{span}(S)$. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 0 \end{pmatrix}$. Then the orthogonal projection of $\mathbf{w} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ onto V is

$$\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x+y \\ x-y \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x+y \\ x-y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Indeed, since V is the xy-plane in $\mathbb{R}^3$.

6 Diagonalization

6.1 Eigenvalues and Eigenvectors

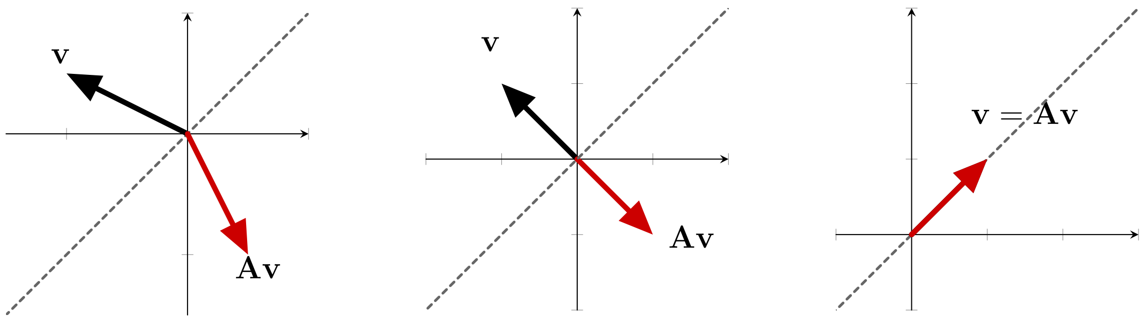
Let $\mathbf{A}$ be a square matrix of order n. Then notice that for any vector $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{Au}$ is also a vector in $\mathbb{R}^n$. So we may think of $\mathbf{A}$ as a map $\mathbb{R}^n \to \mathbb{R}^n$, taking a vector and transforming it to another vector in the same Euclidean space.

Example. 1.

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

Geometrically the matrix $\mathbf{A}$ reflects a vector along the line $x = y$.

$$\mathbf{A}\begin{pmatrix}-2\\1\end{pmatrix} = \begin{pmatrix}1\\-2\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}-1\\1\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}.$$

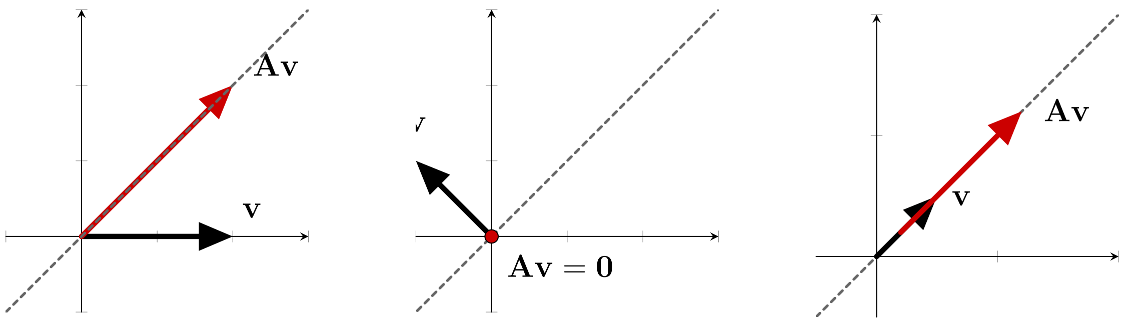


Observe that any vector on the line $x = y$ gets transform back to itself, and any vector along $x = -y$ line get transformed to the negative of itself.

2.
$$\mathbf{A} = \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}$$

The matrix $\mathbf{A}$ takes a vector and maps it to a vector along the line $x = y$ such that both coordinates in $\mathbf{Av}$ are the sum of the coordinates in $\mathbf{v}$.

$$\mathbf{A}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}-1\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix}.$$



Observe that any vector $\mathbf{v}$ along the line $x = y$ is mapped to twice itself, $\mathbf{Av} = 2\mathbf{v}$, and it take any vector $\mathbf{v}$ along the line $x = -y$ to the origin, $\mathbf{Av} = \mathbf{0}$.

Let $\mathbf{A}$ be a square matrix of order n. A real number $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ if there is a nonzero vector $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$, such that $\mathbf{Av} = \lambda\mathbf{v}$. In this case, the nonzero vector $\mathbf{v}$ is called an eigenvector associated to λ. In other words, $\mathbf{A}$ transforms its eigenvectors by scaling it by a factor of the associated eigenvalue.

Remark. Note that for $\mathbf{v}$ to be an eigenvector associate to λ, necessarily $\mathbf{v} \neq \mathbf{0}$. Otherwise, $\mathbf{A0} = \lambda\mathbf{0}$ for any $\lambda \in \mathbb{R}$ and thus every number is an eigenvalue, which makes the definition pointless.

Example. 1. For $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Eigenvalue	Eigenvector
$\lambda = 1$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\lambda = -1$	$\mathbf{v}_\lambda \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

2. For $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Eigenvalue	Eigenvector
$\lambda = 2$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\lambda = 0$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Observe that if $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ associated to eigenvalue λ, then for any $s \in \mathbb{R}$,

$$\mathbf{A}(s\mathbf{v}) = s(\mathbf{A}\mathbf{v}) = s(\lambda\mathbf{v}) = \lambda(s\mathbf{v}).$$

So for any $s \neq 0$, $s\mathbf{v}$ is also an eigenvector associated to eigenvalue λ. Next, $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ if and only if $\mathbf{0} = \lambda\mathbf{v} - \mathbf{A}\mathbf{v} = (\lambda\mathbf{I} - \mathbf{A})\mathbf{v}$. This means that λ is an eigenvalue of $\mathbf{A}$ if and only if the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

Theorem. *Let $\mathbf{A}$ be a square matrix of order n. $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ if and only if the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions.*

Recall that the homogeneous system associated to a square matrix has nontrivial solutions if and only if the matrix is not invertible, which is equivalent to the determinant of the matrix begin 0. Hence, by the theorem above, λ is an eigenvalue of $\mathbf{A}$ if and only if $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$.

Lemma. *Let $\mathbf{A}$ be a square matrix of order n. Then $\det(x\mathbf{I} - \mathbf{A})$ is a polynomial of degree n.*

Example. Let $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an order 2 square matrix. Then the characteristic polynomial of $\mathbf{A}$ is

$$\begin{vmatrix} x - a & -b \\ -c & x - d \end{vmatrix} = (x-a)(x-d) - bc = x^2 - (a+d)x + ad - bc.$$

It is a degree 2 polynomial.

This means that λ is an eigenvalue of $\mathbf{A}$ if and only if it is a root of the polynomial $\det(x\mathbf{I} - \mathbf{A})$. This motivates the following definition.

Let $\mathbf{A}$ be a square matrix of order n, the <u>characteristic polynomial</u> of $\mathbf{A}$, denoted as $\mathrm{char}(\mathbf{A})$, is the degree n polynomial

$$\mathrm{char}(\mathbf{A}) = \det(x\mathbf{I} - \mathbf{A}).$$

So, we have the following theorem.

Theorem. *Let* $\mathbf{A}$ *be a square matrix of order* n. $\lambda \in \mathbb{R}$ *is an eigenvalue of* $\mathbf{A}$ *if and only if* λ *is a root of the characteristic polynomial* $\det(x\mathbf{I} - \mathbf{A})$.

Example.

1. $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x & -1 \\ -1 & x \end{vmatrix} = x^2 - 1.$$

   So the eigenvalues of $\mathbf{A}$ are $\lambda = \pm 1$.

2. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$,

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & -1 \\ -1 & x-1 \end{vmatrix} = (x-1)^2 - 1 = x(x-2).$$

   So the eigenvalues of $\mathbf{A}$ are $\lambda = 0$ and $\lambda = 2$.

3. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 1 \end{pmatrix}$,

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & 0 & 0 \\ 0 & x & 2 \\ 0 & -3 & x-1 \end{vmatrix} = (x-1)[x(x-1)-6)] = (x-1)(x+2)(x-3).$$

   So the eigenvalues of $\mathbf{A}$ are $\lambda = 1$, $\lambda = -2$, and $\lambda = 3$.

Theorem. *A square matrix* $\mathbf{A}$ *is invertible if and only if* 0 *is not an eigenvalue of* $\mathbf{A}$.

Proof. 0 is an eigenvalue of $\mathbf{A}$ $\Leftrightarrow$ 0 is a root of the polynomial $\det(x\mathbf{I} - \mathbf{A})$ $\Leftrightarrow$ $0 = \det(0\mathbf{I} - \mathbf{A}) = \det(\mathbf{A})$ $\Leftrightarrow$ $\mathbf{A}$ not invertible. □

We will add this to our list of equivalent statements for invertibility.

Theorem. *Let* $\mathbf{A}$ *be a square matrix of order* n. *The following statements are equivalent.*

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}$ *has a left inverse.*

(iii) $\mathbf{A}$ *has a right inverse.*

(iv) The reduced row-echelon form of $\mathbf{A}$ *is the identity matrix* $\mathbf{I}_n$.

(v) $\mathbf{A}$ *is a product of elementary matrices.*

(vi) The homogeneous linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ *has only the trivial solution, that is,* $\mathrm{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

(vii) The linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ *has a unique solution for every* $\mathbf{b}$.

(viii) The determinant of $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$.

(ix) *The columns/rows of* $\mathbf{A}$ *are linearly independent.*

(x) *The columns/rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$, $\mathrm{Col}(\mathbf{A}) = \mathbb{R}^n/\mathrm{Row}(\mathbf{A}) = \mathbb{R}^n$.

(xi) 0 *is not an eigenvalue of* $\mathbf{A}$.

Recall that the determinant of a triangular matrix is the product of the diagonal entries. Suppose $\mathbf{A}$ is a triangular matrix. Then $x\mathbf{I} - \mathbf{A}$ is also a triangular matrix. Hence, we have the following statement.

Lemma. *The eigenvaules of a triangular matrix are the diagonal entries.*

Proof. $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$

$$\Rightarrow \begin{vmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ 0 & x - a_{22} & \cdots & -a_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & x - a_{nn} \end{vmatrix} = (x - a_{11})(x - a_{22}) \cdots (x - a_{nn}).$$

□

Let λ be an eigenvalue of $\mathbf{A}$. The algebraic multiplicity of λ is the largest integer r_λ such that

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda)^{r_\lambda} p(x),$$

for some polynomial $p(x)$. Alternatively, r_λ is the positive integer such that in the above equation, λ is not a root of $p(x)$.

Suppose $\mathbf{A}$ is an order n square matrix such that $\det(x\mathbf{I} - \mathbf{A})$ can be factorize into linear factors completely. Then we can write

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

where $r_1 + r_2 + \cdots + r_k = n$, and $\lambda_, \lambda_2, ..., \lambda_k$ are the distinct eigenvalues of $\mathbf{A}$. Then the algebraic multiplicity of λ_i is r_i for $i = 1, ..., k$.

Example.

1. $\mathbf{A} = \mathbf{0}_n$. Then $\det(x\mathbf{I} - \mathbf{0}) = \det(x\mathbf{I}) = x^n$. The multiplicity of the eigenvalue 0 is n, $r_0 = n$.

2. $\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & -2 \\ 0 & 0 & 3 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x - 1)^2(x - 3)$. The eigenvalue 1 has multiplicity 2, while the eigenvalue 3 has multiplicity 1, $r_1 = 2$, $r_3 = 1$. Observe that in general, the algebraic multiplicity of an eigenvalue λ of an triangular matrix $\mathbf{A}$ is the number of diagonal entries that takes the value λ.

3. $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x - 2)^2(x - 4)$. The eigenvalue 2 has multiplicity 2, while the eigenvalue 4 has multiplicity 1, $r_2 = 2$, $r_4 = 1$.

4. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x-1)(x^2+1)$. The only real eigenvalue of $\mathbf{A}$ is 1 with multiplicity 1, $r_1 = 1$. The other eigenvalues are complex numbers (lecture 12).

Now suppose $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of $\mathbf{A}$ associated to the eigenvalue λ. Then for any real numbers $\alpha, \beta \in \mathbb{R}$,

$$\mathbf{A}(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha(\mathbf{A}\mathbf{u}) + \beta(\mathbf{A}\mathbf{v}) = \alpha(\lambda\mathbf{u}) + \beta(\lambda\mathbf{v}) = \lambda(\alpha\mathbf{u} + \beta\mathbf{v}).$$

This means that the set of all vectors that satisfies the relation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ is closed under linearly combination. In particular, if $\alpha\mathbf{u} + \beta\mathbf{v} \neq \mathbf{0}$, then it is also an eigenvector associated to λ. This shows that the collection of all eigenvectors associated to an eigenvalue λ, together with the zero vector, form a subspace. This motivates the following definition.

The eigenspace associated to an eigenvalue λ of $\mathbf{A}$ is

$$E_\lambda = \left\{ \ \mathbf{v} \in \mathbb{R}^n \ \middle| \ \mathbf{A}\mathbf{v} = \lambda\mathbf{v} \ \right\} = \text{Null}(\lambda\mathbf{I} - \mathbf{A}).$$

The geometric multiplicity of an eigenvalue λ is the dimension of its associated eigenspace,

$$\dim(E_\lambda) = \text{nullity}(\lambda\mathbf{I} - \mathbf{A}).$$

Example. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & -1 & 0 \\ -1 & x-1 & -1 \\ 0 & 0 & x-2 \end{vmatrix} = (x-2)((x-1)^2 - 1) = x(x-2)^2.$$

The eigenvalues of $\mathbf{A}$ are 0 and 2, with algebraic multiplicities $r_0 = 1$ and $r_2 = 2$. Since 0 is an eigenvalue of $\mathbf{A}$, $\mathbf{A}$ is not invertible.

Now find the eigenspace.

$\lambda = 0$:

$$0\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0-1 & -1 & 0 \\ -1 & 0-1 & -1 \\ 0 & 0 & 0-2 \end{pmatrix} = \begin{pmatrix} -1 & -1 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

So $E_0 = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$.

$\lambda = 2$:

$$2\mathbf{I} - \mathbf{A} = \begin{pmatrix} 2-1 & -1 & 0 \\ -1 & 2-1 & 0 \\ 0 & 0 & 2-2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So $E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$.

So the geometric multiplicity of $\lambda = 0$ is $\dim(E_0) = 1$ and the geometric multiplicity of $\lambda = 2$ is $\dim(E_2) = 2$.

Appendix to Lecture 10

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace and $\mathbf{b} \in \mathbb{R}^n$. Let $\mathbf{b}_p \in V$ be the orthogonal projection of $\mathbf{b}$ onto V. Then for any $\mathbf{w} \in V$,*

$$\|\mathbf{b}_p - \mathbf{b}\| = d(\mathbf{b}_p, \mathbf{b}) \leq d(\mathbf{w}, \mathbf{b}) = \|\mathbf{w} - \mathbf{b}\|,$$

that is, the distance between $\mathbf{b}$ and the projection $\mathbf{b}_p$ is the minimum distance between $\mathbf{b}$ and any vector $\mathbf{w} \in V$.

Proof. Write $\mathbf{b} = \mathbf{b}_p + \mathbf{b}_n$, where $\mathbf{b}_p \in \text{Col}(\mathbf{A})$ is the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$, and $\mathbf{b}_n$ is orthogonal to $\text{Col}(\mathbf{A})$. Let $\mathbf{w}$ be any vector in V. Since $\mathbf{w}$ and $\mathbf{b}_p$ are in the subspace V and $\mathbf{b}_n \perp V$, $(\mathbf{w} - \mathbf{b}_p) \cdot \mathbf{b}_n = 0$. So

$$\begin{aligned} \|\mathbf{w} - \mathbf{b}\|^2 &= (\mathbf{w} - \mathbf{b}_p - \mathbf{b}_n) \cdot (\mathbf{w} - \mathbf{b}_p - \mathbf{b}_n) \\ &= (\mathbf{w} - \mathbf{b}_p) \cdot (\mathbf{w} - \mathbf{b}_p) - 2(\mathbf{w} - \mathbf{b}_p).\mathbf{b}_n + \mathbf{b}_n \cdot \mathbf{b}_n \\ &= \|\mathbf{w} - \mathbf{b}_p\|^2 + \|\mathbf{b}_n\|^2. \end{aligned} \tag{5}$$

So $\|\mathbf{w} - \mathbf{b}\| \geq \|\mathbf{b}_n\| = \|\mathbf{b}_p - \mathbf{b}\|$. □

Theorem. *Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. $\mathbf{u} \in \mathbb{R}^n$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{Au}$ is the projection of $\mathbf{b}$ onto the column space of $\text{Col}(\mathbf{A})$.*

Proof. ($\Rightarrow$) Suppose $\mathbf{u}$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$. Since the projection $\mathbf{b}_p \in \text{Col}(\mathbf{A})$ is in the column space, there must be a $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{Av} = \mathbf{b}_p$. Then by equation (5),

$$\|\mathbf{Au} - \mathbf{b}_p\|^2 + \|\mathbf{b}_n\|^2 = \|\mathbf{Au} - \mathbf{b}\|^2 \leq \|\mathbf{Av} - \mathbf{b}\|^2 = \|\mathbf{b}_p - \mathbf{b}\|^2 = \|\mathbf{b}_n\|^2,$$

which happens if and only if $\mathbf{Au} = \mathbf{b}_p$, that is, $\mathbf{Au}$ is the projection of $\mathbf{b}$ onto $\text{Col}(\mathbf{A})$.

($\Leftarrow$) Suppose $\mathbf{Au} = \mathbf{b}_p$ is the projection of $\mathbf{b}$ onto $\text{Col}(\mathbf{A})$. Then by the theorem above, and the fact that $\text{Col}(\mathbf{A}) = \{\ \mathbf{Av} \mid \mathbf{v} \in \mathbb{R}^n\ \}$,

$$\|\mathbf{Au} - \mathbf{b}\| = \|\mathbf{b}_p - \mathbf{b}\| \leq \|\mathbf{Av} - \mathbf{b}\|$$

for any $\mathbf{v} \in \mathbb{R}^n$, that is, $\mathbf{u}$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$.

□

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 11 Notes

References

1. Elementary Linear Algebra: Application Version, Section 5.2, 5.4, 7.1-7.2
2. Linear Algebra with Application, Section 3.3-3.4, 5.5, 8.2

6.2 Diagonalization

A square matrix $\mathbf{A}$ is said to be diagonalizable if there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ is a diagonal matrix.

Remark. The statement above is equivalent to being able to express $\mathbf{A}$ as $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ for some invertible $\mathbf{P}$ and diagonal matrix $\mathbf{D}$.

Example.

1. Any square zero matrix is diagonalizable, $\mathbf{0} = \mathbf{I}\mathbf{0}\mathbf{I}^{-1}$.
2. Any diagonal matrix $\mathbf{D}$ is diagonalizable, $\mathbf{D} = \mathbf{I}\mathbf{D}\mathbf{I}^{-1}$.
3. $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix}$ is diagonalizable, with $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}$.

Suppose $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, for some matrix $\mathbf{D}$ and invertible matrix $\mathbf{P}$. Then the characteristic polynomial of $\mathbf{A}$ is

$$\begin{aligned} \det(x\mathbf{I} - \mathbf{A}) &= \det(x\mathbf{I} - \mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \det(\mathbf{P}x\mathbf{P}^{-1} - \mathbf{P}\mathbf{D}\mathbf{P}^{-1}) \\ &= \det(\mathbf{P}(x\mathbf{I} - \mathbf{D})\mathbf{P}^{-1}) = \det(\mathbf{P})\det(x\mathbf{I} - \mathbf{D})\det(\mathbf{P}^{-1}) \\ &= \det(\mathbf{P})\det(\mathbf{P})^{-1}\det(x\mathbf{I} - \mathbf{D}) = \det(x\mathbf{I} - \mathbf{D}). \end{aligned} \tag{6}$$

In other words, the characteristics polynomail of $\mathbf{A}$ is the same as $\mathbf{D}$. Furthermore, if $\mathbf{D}$ is a diagonal matrix, then the eigenvalues of $\mathbf{A}$ are exactly the diagonal entries of $\mathbf{D}$, with algebraic multiplicity equals to the number of diagonal entries taking the value. Hence, if $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, then the diagonal entries of the matrix $\mathbf{D}$ are exactly the eigenvalues of $\mathbf{A}$, appearing their algebraic multiplicities times. The invertible matrix $\mathbf{P}$ is constructed from the eigenvectors of $\mathbf{A}$.

Theorem. *Let $\mathbf{A}$ be a square matrix of order n. $\mathbf{A}$ is diagonalizable if and only if tdiag()here exists a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}$.*

The invertible matrix $\mathbf{P}$ that diagonalizes $\mathbf{A}$ have the form $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, where $\mathbf{u}_i$ are eigenvectors of $\mathbf{A}$, and the diagonal matrix is $\mathbf{D} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$, where λ_i is the eigenvalue associated to eigenvector $\mathbf{u}_i$. In other words, the i-th column of the invertible matrix $\mathbf{P}$ is an eigenvector of $\mathbf{A}$ with eigenvalue the i-th diagonal entry of $\mathbf{D}$. For the details of the proof, readers may refer to the appendix.

Theorem (Geometric Multiplicity is no greater than Algebraic multiplicity)**.** *The geometric multiplicity of an eigenvalue λ of a square matrix $\mathbf{A}$ is no greater than the algebraic multiplicity, that is,*

$$1 \leq \dim(E_\lambda) \leq r_\lambda.$$

Theorem. *Suppose* $\mathbf{A}$ *is a square matrix such that its characteristic polynomial can be written as a product of linear factors,*

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

where r_{λ_i} *is the algebraic multiplicity of* λ_i*, for* $i = 1, ..., k$*, and the eigenvalues are distinct,* $\lambda_i \neq \lambda_j$ *for all* $i \neq j$*. Then* $\mathbf{A}$ *is diagonalizable if and only if for each eigenvalue of* $\mathbf{A}$*, its geometric multiplicity is equal to its algebraic multiplicity,*

$$\dim(E_{\lambda_i}) = r_{\lambda_i}$$

for all eigenvalues λ_i *of* $\mathbf{A}$*.*

In other words, there are two obstructions to a matrix $\mathbf{A}$ being diagonalizable.

(i) The characteristic polynomial of $\mathbf{A}$ do not split into linear factors.

(ii) There is an eigenvalue of $\mathbf{A}$ where the geometric multiplicity is strictly less than the algebraic multiplicity, $\dim(E_{\lambda_i}) < r_{\lambda_i}$.

For in either cases, there will not be enough linearly independent eigenvectors to form a basis for $\mathbb{R}^n$.

Corollary. *If* $\mathbf{A}$ *is a square matrix of order* n *with* n *distinct eigenvalues, then* $\mathbf{A}$ *is diagonalizable.*

Proof. If $\mathbf{A}$ has n distinct eigenvalues, then the algebraic multiplicity of each eigenvalue must be 1. Thus

$$1 \leq \dim(E_\lambda) \leq r_\lambda = 1 \Rightarrow \dim(E_\lambda) = 1 = r_\lambda$$

for every eigenvalue λ of $\mathbf{A}$. Therefore $\mathbf{A}$ is diagonalizable. $\square$

Algorithm to diagonalization

(i) Compute the characteristic polynomial of $\mathbf{A}$

$$\det(x\mathbf{I} - \mathbf{A}).$$

If it cannot be factorized into linear factors, then $\mathbf{A}$ is not diagonalizable.

(ii) Otherwise, write

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

where r_{λ_i} is the algebraic multiplicity of λ_i, for $i = 1, ..., k$, and the eigenvalues are distinct, $\lambda_i \neq \lambda_j$ for all $i \neq j$. For each eigenvalue λ_i of $\mathbf{A}$, $i = 1, ..., k$, find a basis for the eigenspace, that is, find the solution space of the following linear system,

$$(\lambda_i\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}.$$

If there is a i such that $\dim(E_{\lambda_i}) < r_{\lambda_i}$, that is, if the number of parameters in the solution space of the above linear system is not equal to the algebraic multiplicity, then $\mathbf{A}$ is not diagonalizable.

(iii) Otherwise, find a basis S_{λ_i} of the eigenspace E_{λ_i} for each eigenvalue λ_i, $i = 1, ..., k$. Necessarily $|S_{\lambda_i}| = r_{\lambda_i}$ for all $i = 1, ..., k$. Let $S = \bigcup_{i=1}^{k} S_{\lambda_i}$. Then

$$|S| = \sum_{i=1}^{k} |S_{\lambda_i}| = \sum_{i=1}^{k} r_{\lambda_i} = n,$$

and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

(iv) Let

$$\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \operatorname{diag}(\mu_1, \mu_2, ..., \mu_n) = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where μ_i is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$. Then

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

Example. 1. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. It has eigenvalues 0 and 2 with multiplicity $r_0 = 1$ and $r_2 = 2$, respectively. Also, $\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for E_0 and $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is a basis for E_2. Then $\dim(E_0) = 1 = r_0$ and $\dim(E_2) = 2 = r_2$. Hence, $\mathbf{A}$ is diagonalizable, with

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}.$$

2. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$. $\mathbf{A}$ is a triangular matrix, hence the diagonal entries, $1, 2, 3$ are the eigenvalues, each with algebraic multiplicity 1. Therefore $\mathbf{A}$ is diagonalizable. We will need to find a basis for each of the eigenspace.

$\lambda = 1$: $\begin{pmatrix} 1-1 & -1 & -1 \\ 0 & 1-2 & -2 \\ 0 & 0 & 1-3 \end{pmatrix} = \begin{pmatrix} 0 & -1 & -1 \\ 0 & -1 & -2 \\ 0 & 0 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$ is a basis for E_1.

$\lambda = 2$: $\begin{pmatrix} 2-1 & -1 & -1 \\ 0 & 2-2 & -2 \\ 0 & 0 & 2-3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & -2 \\ 0 & 0 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for E_2.

$\lambda = 3$: $\begin{pmatrix} 3-1 & -1 & -1 \\ 0 & 3-2 & -2 \\ 0 & 0 & 3-3 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 0 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} \right\}$ is a basis for E_3.

$$\Rightarrow \mathbf{A} = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix}^{-1}.$$

3. $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. $\lambda = 1$ is the only eigenvalue with algebraic multiplicity $r_1 = 2$.

$\lambda = 1$: $\begin{pmatrix} 0 & -2 \\ 0 & 0 \end{pmatrix}$. There is only one non-pivot columns, hence $\dim(E_1) = 1 < 2 = r_1$.

This shows that $\mathbf{A}$ is not diagonalizable.

6.3 Orthogonal Diagonalization

A square matrix $\mathbf{A}$ of order n is an orthogonal matrix if $\mathbf{A}^T = \mathbf{A}^{-1}$, equivalently, $\mathbf{A}^T\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^T$.

Theorem. *Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.*

(i) $\mathbf{A}$ is orthogonal.

(ii) The columns of $\mathbf{A}$ forms an orthonormal basis for $\mathbb{R}^n$.

(iii) The rows of $\mathbf{A}$ forms an orthonormal basis for $\mathbb{R}^n$.

Proof. Write

$$\mathbf{A} = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{pmatrix},$$

where for $i = 1, ..., n$, $\mathbf{c}_i$ and $\mathbf{r}_i$ are the columns of and rows of $\mathbf{A}$, respectively. Then

$$\begin{aligned} \mathbf{A}^T\mathbf{A} &= \begin{pmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_n^T \end{pmatrix} \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1^T\mathbf{c}_1 & \mathbf{c}_1^T\mathbf{c}_2 & \cdots & \mathbf{c}_1^T\mathbf{c}_n \\ \mathbf{c}_2^T\mathbf{c}_1 & \mathbf{c}_2^T\mathbf{c}_2 & \cdots & \mathbf{c}_2^T\mathbf{c}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{c}_n^T\mathbf{c}_1 & \mathbf{c}_n^T\mathbf{c}_2 & \cdots & \mathbf{c}_n^T\mathbf{c}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{c}_1 \cdot \mathbf{c}_1 & \mathbf{c}_1 \cdot \mathbf{c}_2 & \cdots & \mathbf{c}_1 \cdot \mathbf{c}_n \\ \mathbf{c}_2 \cdot \mathbf{c}_1 & \mathbf{c}_2 \cdot \mathbf{c}_2 & \cdots & \mathbf{c}_2 \cdot \mathbf{c}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{c}_n \cdot \mathbf{c}_1 & \mathbf{c}_n \cdot \mathbf{c}_2 & \cdots & \mathbf{c}_n \cdot \mathbf{c}_n \end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}\mathbf{A}^T &= \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^T & \mathbf{r}_2^T & \cdots & \mathbf{r}_n^T \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1\mathbf{r}_1^T & \mathbf{r}_1\mathbf{r}_2^T & \cdots & \mathbf{r}_1\mathbf{r}_n^T \\ \mathbf{r}_2\mathbf{r}_1^T & \mathbf{r}_2\mathbf{r}_2^T & \cdots & \mathbf{r}_2\mathbf{r}_n^T \\ \vdots & & \ddots & \vdots \\ \mathbf{r}_n\mathbf{r}_1^T & \mathbf{r}_n\mathbf{r}_2^T & \cdots & \mathbf{r}_n\mathbf{r}_n^T \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{r}_1 \cdot \mathbf{r}_1 & \mathbf{r}_1 \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_1 \cdot \mathbf{r}_n \\ \mathbf{r}_2 \cdot \mathbf{r}_1 & \mathbf{r}_2 \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_2 \cdot \mathbf{r}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{r}_n \cdot \mathbf{r}_1 & \mathbf{r}_n \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_n \cdot \mathbf{r}_n \end{pmatrix}. \end{aligned}$$

So $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ if and only if $\mathbf{c}_i \cdot \mathbf{c}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, and $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ if and only if $\mathbf{r}_i \cdot \mathbf{r}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ □

An order n square matrix $\mathbf{A}$ is orthogonally diagonalizable if

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

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

Supppose $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ for some orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$. Then

$$\mathbf{A}^T = (\mathbf{P}\mathbf{D}\mathbf{P}^T)^T = (\mathbf{P}^T)^T\mathbf{D}^T\mathbf{P}^T = \mathbf{P}\mathbf{D}\mathbf{P}^T = \mathbf{A},$$

since $\mathbf{D}$ is diagonal, and hence symmetric. This shows that if $\mathbf{A}$ is orthogonally diagonalizable, it is symmetric. The converse is also true, but the proof is beyond the scope of this course.

Theorem. *An order n square matrix is orthogonally diagonalizable if and only if it is symmetric.*

The algorithm to orthogonally diagonalize a matrix is the same as the usual diagonalization, except until the last step, instead of using a basis of eigenvectors to form the matrix $\mathbf{P}$, we have to turn it into an orthonormal basis of eigenvectors, that is, to use the Gram-Schmidt process. However, we do not need to use the Gram-Schmidt process for the whole basis, but only among those eigenvectors that belong to the same eigenspace. This follows from the fact that the eigenspaces are already orthogonal to each other.

Theorem. *If $\mathbf{A}$ is orthogonally diagonalizable, then the eigenspaces are orthogonal to each other. That is, suppose λ_1 and λ_2 are distinct eigenvalues of a symeetric matrix $\mathbf{A}$, $\lambda_1 \neq \lambda_2$. Let E_{λ_i} denote the eigenspace associated to eigenvalue λ_i, for $i = 1, 2$. Then for any $\mathbf{v}_1 \in E_{\lambda_1}$ and $\mathbf{v}_2 \in E_{\lambda_2}$, $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.*

Algorithm to orthogonal diagonalization

Follow step (i) to (iii) in algorithm to diagonalization.

(iv) Apply Gram-Schmidt process to each basis S_{λ_i} of the eigenspace E_{λ_i} to obtain an orthonormal basis T_{λ_i}. Let $T = \bigcup_{i=1}^{k} T_{\lambda_i}$, it is an orthonormal set. Similarly, we have $|T_{\lambda_i}| = r_{\lambda_i}$, and so

$$|T| = \sum_{i=1}^{k} |T_{\lambda_i}| = \sum_{i=1}^{k} r_{\lambda_i} = n,$$

which shows that $T = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

(v) Let

$$P = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \operatorname{diag}(\mu_1, \mu_2, ..., \mu_n) = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where μ_i is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$. Then $\mathbf{P}$ is an orthogonal matrix, and

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T.$$

Example. Let $\mathbf{A} = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$. $\mathbf{A}$ is symmetric, hence it is orthogonally diagonalizable.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-5 & 1 & 1 \\ 1 & x-5 & 1 \\ 1 & 1 & x-5 \end{vmatrix} = (x-3)(x-6)^2.$$

$\mathbf{A}$ has eigenvalues $\lambda = 3, 6$, with miltiplicity $r_3 = 1$, $r_6 = 2$. Let us now compute the eigenspaces.

$\lambda = 3$: $\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a basis for E_3.

$\lambda = 6$: $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for E_6.

Observe that $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$, but the set $\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is not orthogonal. So we only need to apply the Gram-Schmidt process to the set $\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$.

$$\begin{aligned} \mathbf{v}_1 &= \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \\ \mathbf{v}_2 &= \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}. \end{aligned}$$

Indeed, $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\}$ is an orthogonal basis. Normalizing, we get an orthonormal basis

$$\left\{ \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\}.$$

So

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}^T.$$

6.4 Application of Diagonalization: Recursion Formula

Lemma. *Suppose* $\mathbf{A} = \mathbf{PDP}^{-1}$. *Then* $\mathbf{A}^m = \mathbf{PD}^m\mathbf{P}^{-1}$.

Proof. Exercise $\square$

Lemma. *If* $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$, *then* $\mathbf{D}^m = \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^m \end{pmatrix}$.

Proof. Exercise $\square$

Combining the two lemmas, we have the following statement. Suppose $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{PDP}^{-1}$ with $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$. Then

$$\mathbf{A}^m = \mathbf{P} \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^m \end{pmatrix} \mathbf{P}^1.$$

Linear Recurrence

Consider the Fibonacci sequence

$$a_0 = 0, a_1 = 1, a_2 = 1, a_3 = 2, a_4 = 3, a_5 = 5, a_6 = 8, ...$$

where $a_{n+1} = a_n + a_{n-1}$. We can represent the linear recurrence relation as a matrix equation,

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

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Observe that

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, \begin{pmatrix} a_2 \\ a_3 \end{pmatrix} = \mathbf{A} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \mathbf{A}^2 \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, ..., \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \mathbf{A}^n \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \mathbf{A}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence, by computing $\mathbf{A}^n$, we are able to find the general formula a_n for the Fibonacci sequence. We will diagonalize $\mathbf{A}$.

$$\begin{vmatrix} x & -1 \\ -1 & x-1 \end{vmatrix} = x^2 - x - 1.$$

The eigenvalues are $\frac{-1\pm\sqrt{5}}{2}$. Let $\lambda_\alpha = \frac{-1+\sqrt{5}}{2}$ and $\lambda_\beta = \frac{-1-\sqrt{5}}{2}$. We will now compute the eigenvectors. For each $\lambda = \lambda_\alpha, \lambda_\beta$,

$$\lambda\mathbf{I} - \mathbf{A} = \begin{pmatrix} \lambda & -1 \\ -1 & \lambda - 1 \end{pmatrix}.$$

So the eigenvector is $\mathbf{v} = \begin{pmatrix} 1 \\ \lambda \end{pmatrix}$, for $\lambda = \lambda_\alpha, \lambda_\beta$. Hence,

$$\mathbf{A} = \mathbf{PDP}^{-1} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha & 0 \\ 0 & \lambda_\beta \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1}.$$

Then

$$\mathbf{A}^n = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n & 0 \\ 0 & \lambda_\beta^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1}.$$

So,

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n & 0 \\ 0 & \lambda_\beta^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

We can avoid computing $\begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1}$ but compute $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ directly, $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, that is, we only need to solve the linear system.

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ \lambda_\alpha & \lambda_\beta & 1 \end{array}\right) \xrightarrow{R_2 - \lambda_\alpha R_1} \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & \lambda_\beta - \lambda_\alpha & 1 \end{array}\right) \xrightarrow{\frac{1}{\lambda_\beta - \lambda_\alpha} R_2} \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & \frac{1}{\lambda_\beta - \lambda_\alpha} \end{array}\right)$$
$$\xrightarrow{R_1 - R_2} \left(\begin{array}{cc|c} 1 & 0 & \frac{1}{\lambda_\alpha - \lambda_\beta} \\ 0 & 1 & \frac{1}{\lambda_\beta - \lambda_\alpha} \end{array}\right)$$

And so $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/(\lambda_\alpha - \lambda_\beta) \\ 1/(\lambda_\beta - \lambda_\alpha) \end{pmatrix} = \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, which gives us

$$\begin{aligned} \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} &= \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n & 0 \\ 0 & \lambda_\beta^n \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n \\ -\lambda_\beta^n \end{pmatrix} = \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} \lambda_\alpha^n - \lambda_\beta^n \\ \lambda_\alpha^{n+1} - \lambda_\beta^{n+1} \end{pmatrix}. \end{aligned}$$

Reading off the first coordinates, we have

$$a_n = \frac{\lambda_\alpha^n - \lambda_\beta^n}{\lambda_\alpha - \lambda_\beta} = \frac{(-1 + \sqrt{5})^n - (-1 - \sqrt{5})^n}{2^n \sqrt{5}}.$$

Markov Chain

Sheldon only patronizes three stalls in the school canteen, the mixed rice, noodle, and mala hotpot stall for lunch everyday. He never buys from same stall two days in a row. If he buys from the mixed rice stall on a certain day, there is a 40% chance he will eat from the noodles stall the next day. If he buys from the noodle stall on a certain day, there is a 50% chance he will eat mala hotpot the next day. If he buys eats mala hotpot on a certain day, there is a 60% chance he will buy mixed rice the next day.

(a.) Suppose Sheldon had noodles today. What is the probability that he patronizes each of the stalls 3 days later?

(b.) Show that regardless of what he had today, in the long run, he is most likely to patronize the mixed rice and mala hotpot stall equally.

Let a_n, b_n, and c_n be the probability that Sheldon eats from the mixed rice, noodles, and mala hotpot stalls n days later, respectively. Then the probabilities a_n, b_n, c_n depends on the probabilities for the previous day as such

$$\begin{array}{rcrcrcr} a_n & = & & & 0.5b_{n-1} & + & 0.6c_{n-1} \\ b_n & = & 0.4a_{n-1} & & & + & 0.4c_{n-1} \\ c_n & = & 0.6a_{n-1} & + & 0.5b_{n-1} & & \end{array},$$

or,

$$\begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} = \mathbf{A} \begin{pmatrix} a_{n-1} \\ b_{n-1} \\ c_{n-1} \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} 0 & 0.5 & 0.6 \\ 0.4 & 0 & 0.4 \\ 0.6 & 0.5 & 0 \end{pmatrix}$.

We shall now diagonalize $\mathbf{A}$.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x & -0.5 & -0.6 \\ -0.4 & x & -0.4 \\ -0.6 & -0.5 & x \end{vmatrix} = (x-1)(x+0.6)(x+0.4).$$

The eigenvalues of $\mathbf{A}$ are $\lambda = 1, -0.6, -0.4$, with algebraic multiplicity 1 each.

$$\lambda = 1\text{:} \quad \begin{pmatrix} 1 & -0.5 & -0.6 \\ -0.4 & 1 & -0.4 \\ -0.6 & -0.5 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -0.8 \\ 0 & 0 & 0 \end{pmatrix}. \text{ Let } \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0.8 \\ 1 \end{pmatrix}.$$

$$\lambda = -0.6\text{:} \quad \begin{pmatrix} -0.6 & -0.5 & -0.6 \\ -0.4 & -0.6 & -0.4 \\ -0.6 & -0.5 & -0.6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \text{ Let } \mathbf{v}_{-0.6} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

$$\lambda = -0.4\text{:} \quad \begin{pmatrix} -0.4 & -0.5 & -0.6 \\ -0.4 & -0.4 & -0.4 \\ -0.6 & -0.5 & -0.4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}. \text{ Let } \mathbf{v}_{-0.4} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

So $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, where $\mathbf{P} = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}$ and $\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.6 & 0 \\ 0 & 0 & -0.4 \end{pmatrix}$, and

$$\begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} = \mathbf{A}^n \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix}.$$

(a.) Since he had noodles today, $\begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. So, the probabilities 3 days later will be

$$\mathbf{A}^3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-0.6)^3 & 0 \\ 0 & 0 & (-0.4)^3 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.38 \\ 0.24 \\ 0.38 \end{pmatrix}.$$

(b.) In the long run, $\mathbf{A}^k = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-0.6)^k & 0 \\ 0 & 0 & (-0.4)^k \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1}$ tends to

$$\begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1} = \frac{1}{14} \begin{pmatrix} 5 & 5 & 5 \\ 4 & 4 & 4 \\ 5 & 5 & 5 \end{pmatrix}.$$

Notice that the columns are the same, so the probabilities in the long run is the same regardless of what he had today. Moreover, he is most likely, with $5/14$ chance of patronizing the mixed rice and mala hotpot stall.

Appendix to Lecture 11

Theorem. *Let* $\mathbf{A}$ *be a square matrix of order* n*.* $\mathbf{A}$ *is diagonalizable if and only if there exists a basis* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *of* $\mathbb{R}^n$ *of eigenvectors of* $\mathbf{A}$*.*

Proof.

($\Rightarrow$) Suppose $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{PDP}^{-1}$ for some diagonal matrix $\mathbf{D}$ and invertible matrix $\mathbf{P}$. Write $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$ and $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, where $\mathbf{u}_i$ is the i-th column of $\mathbf{P}$, for $i = 1, ..., n$. Since $\mathbf{P}$ is invertible, $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ forms a basis for $\mathbb{R}^n$. Then we have

$$\begin{aligned}
&\mathbf{A} = \mathbf{PDP}^{-1} \Leftrightarrow \mathbf{AP} = \mathbf{PD} \\
&\Leftrightarrow \mathbf{A}\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \qquad (7) \\
&\Leftrightarrow \begin{pmatrix} \mathbf{Au}_1 & \mathbf{Au}_2 & \cdots & \mathbf{Au}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \cdots & \lambda_n\mathbf{u}_n \end{pmatrix} \\
&\Leftrightarrow \mathbf{Au}_i = \lambda_i\mathbf{u}_i \text{ for all } i = 1, ..., n.
\end{aligned}$$

The third equivalence follows from that fact that $\text{diag}(\lambda_1, \lambda_2, ..., \lambda_n) = \text{diag}(\lambda_1, 1, ..., 1)$ $\text{diag}(1, \lambda_2, ..., 1) \cdots \text{diag}(1, 1, ..., \lambda_n)$, and post-multiplying a matrix by $\text{diag}(1, ..., \lambda_i, ..., 1)$ is multiplying the i-th column of the matrix by λ_i. Since $\mathbf{P}$ is invertible, $\mathbf{u}_i \neq \mathbf{0}$ for all $i = 1, ..., n$. This shows that $\mathbf{u}_i$ are eigenvectors of $\mathbf{A}$ associated to eigenvalues λ_i. So $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.

($\Leftarrow$) Suppose there is a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}$. Let λ_i be the eigevalues of $\mathbf{A}$ associated to $\mathbf{u}_i$ for $i = 1, ..., n$, that is, $\mathbf{Au}_i = \lambda\mathbf{u}_i$ for $i = 1, ..., n$. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ and $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$. Then by (7), $\mathbf{A}$ is diagonalizable. $\square$

Recall we have shown that if $\mathbf{A}$ is diagonalizable, then the algebraic multiplicity of an eigenvalue λ of $\mathbf{A}$ is the number of diagonal entries taking the value λ. Suppose the $i_1, i_2, ..., i_k$ diagonal entries of $\mathbf{D}$ take the value λ, that is, $\lambda_{i_1} = \lambda_{i_2} = \cdots = \lambda_{i_k} = \lambda$. This would mean that the algebraic multiplicity r_λ of λ is k, $r_\lambda = k$. Then the corresponding columns of $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ must be eigenvectors associated to λ, that is, $\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, ..., \mathbf{u}_{i_k}$ are all eigenvectors associated to λ. Since the columns of $\mathbf{P}$ are linearly independent, this means that $\{\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, ..., \mathbf{u}_{i_k}\}$ is a linearly independent subset of the eigenspace E_λ associated to λ. Hence, the geometric multiplicity of λ is less than or equals to its algebraic multiplicity, $\dim(E_\lambda) \leq r_\lambda$.

In general, for a square matrix $\mathbf{A}$, regardless of whether it is diagonalizable, the geometric multiplicity of an eigenvalue is less than or equals to its algebraic multiplicity.

Theorem (Geometric multiplicity is no greater than algebraic multiplicity)**.** *The geometric multiplicity of an eigenvalue* λ *of a square matrix* $\mathbf{A}$ *is no greater than the algebraic multiplicity, that is,*

$$1 \leq \dim(E_\lambda) \leq r_\lambda.$$

Proof. Suppose $\mathbf{A}$ has order n. Let λ be an eigenvalue of $\mathbf{A}$ and E_λ be the associated eigenspace. Suppose $\dim(E_\lambda) = k$. Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for the eigenspace E_λ.

Extend this set to be a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$. Let $\mathbf{Q} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, it is an invertbile matrix. Note that

$$\begin{aligned} \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix} &= \mathbf{I} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{Q}^{-1} \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Q}^{-1}\mathbf{u}_1 & \mathbf{Q}^{-1}\mathbf{u}_2 & \cdots & \mathbf{Q}^{-1}\mathbf{u}_n \end{pmatrix}, \end{aligned}$$

that is, $\mathbf{Q}^{-1}\mathbf{u}_i = \mathbf{e}_i$ for all $i = 1, ..., n$. Let $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$. Then

$$\begin{aligned} \mathbf{B} &= \mathbf{Q}^{-1}\mathbf{A} \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} = \mathbf{Q}^{-1} \begin{pmatrix} \mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \cdots & \mathbf{A}\mathbf{u}_n \end{pmatrix} \\ &= \mathbf{Q}^{-1} \begin{pmatrix} \lambda\mathbf{u}_1 & \cdots & \lambda\mathbf{u}_k & \mathbf{A}\mathbf{u}_{k+1} \cdots & \mathbf{A}\mathbf{u}_n \end{pmatrix} \\ &= \begin{pmatrix} \lambda\mathbf{Q}^{-1}\mathbf{u}_1 & \cdots & \lambda\mathbf{Q}^{-1}\mathbf{u}_k & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1} \cdots & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n \end{pmatrix} \\ &= \begin{pmatrix} \begin{matrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{matrix} & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1} \cdots & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n \end{pmatrix}. \end{aligned}$$

This means that $\det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x)$ for some polynomial $p(x)$. By (6),

$$\det(x\mathbf{I} - \mathbf{A}) = \det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x).$$

This means that the algebraic multiplicity of the eigenvalue λ of $\mathbf{A}$ is no less than k, that is,

$$r_\lambda \geq k = \dim(E_\lambda).$$

□

Lemma. *Suppose $V_i \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$, for $i = 1, 2$ with trivial intersection, $V_1 \cap V_2 = \{\mathbf{0}\}$. Let $S_i \subseteq V_i$ be a linearly independent subset for $i = 1, 2$. Then $S_1 \cup S_2$ is linearly independent.*

Proof. Let $S_1 = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ and $S_2 = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$. Suppose $c_1, c_2, ..., c_k, d_1, d_2, ..., d_m \in \mathbb{R}$ are real numbers such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m = \mathbf{0}.$$

Let us consider cases.

Case 1: $c_1 = c_2 = \cdots = c_k = 0$. Then the equation is reduced to

$$d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m = \mathbf{0},$$

and hence by independence of S_2, necessarily $d_1 = d_2 = \cdots = d_m = 0$.

Case 2: $d_1 = d_2 = \cdots = d_m$. Then the equation is reduced to

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0},$$

and hence by independence of S_1, necessarily $c_1 = c_2 = \cdots = c_k = 0$.

Case 3: There is an $i = 1, ..., k$ and $j = 1, ..., m$ such that $c_i \neq 0$ and $d_j \neq 0$. Let

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = -(d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m).$$

By assumption, $\mathbf{v} \neq \mathbf{0}$, and $\mathbf{v} \in \text{span}(S_i) \subseteq V_i$ for $i = 1, 2$. But since $V_1 \cap V_2 = \{\mathbf{0}\}$, this means that $\mathbf{v} = \mathbf{0}$, a contradiction. Therefore case 3 is not possible.

Hence, $c_1 = c_2 = \cdots = c_k = d_1 = d_2 = \cdots = d_m = 0$ in every possible cases. This shows that $S_1 \cup S_2$ is linearly independent. $\square$

Theorem (Intersection of distinct eigenspaces is trivial)**.** *Let λ_1 and λ_2 be distinct eigenvalues, $\lambda_1 \neq \lambda_2$, of a square matrix $\mathbf{A}$. Let E_{λ_i} be the associated eigenspace for λ_i, $i = 1, 2$. Then the intersection of the eigenspaces is trivial,*

$$E_{\lambda_1} \cap E_{\lambda_2} = \{\mathbf{0}\}.$$

Proof. Let $\mathbf{v} \in E_{\lambda_1} \cap E_{\lambda_2}$. Then

$$\lambda_1\mathbf{v} = \mathbf{A}\mathbf{v} = \lambda_2\mathbf{v},$$

where the first equality follows from $\mathbf{v} \in E_{\lambda_1}$ and the second equality follows from $\mathbf{v} \in E_{\lambda_2}$. This means that

$$(\lambda_1 - \lambda_2)\mathbf{v} = \mathbf{0}.$$

But since $\lambda_1 \neq \lambda_2$, necessarily $\mathbf{v} = \mathbf{0}$. $\square$

Together with the previous lemma, the theorem shows that if λ_1 and λ_2 are two distinct eigenvalues of $\mathbf{A}$, $\lambda_1 \neq \lambda_2$ and E_{λ_i} is the eigenspace associated to λ_i, for $i = 1, 2$, then for any linearly independent subsets $S_1 \subseteq E_{\lambda_1}$ and $S_2 \subseteq E_{\lambda_2}$, $S_1 \cup S_2$ is linearly independent. In particular, if S_i is a basis for E_{λ_i}, $i = 1, 2$, then $S_1 \cup S_2$ is a basis for $E_{\lambda_1} + E_{\lambda_2}$.

Theorem. *Suppose $\mathbf{A}$ is a square matrix such that its characteristic polynomial can be written as a product of linear factors,*

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

where r_{λ_i} is the algebraic multiplicity of λ_i, for $i = 1, ..., k$, and the eigenvalues are distinct, $\lambda_i \neq \lambda_j$ for all $i \neq j$. Then $\mathbf{A}$ is diagonalizable if and only if for each eigenvalue of $\mathbf{A}$, its geometric multiplicity is equal to its algebraic multiplicity,

$$\dim(E_{\lambda_i}) = r_{\lambda_i}$$

for all eigenvalues λ_i of $\mathbf{A}$.

Proof. Firstly, note that $r_{\lambda_1} + r_{\lambda_2} + \cdots + r_{\lambda_k} = n$. Let $m_{\lambda_i} = \dim(E_{\lambda_i})$ and S_{λ_i} be a basis of E_{λ_i}, for $i = 1, ..., k$. Then since geometric multiplicity is no greater than the algebraic multiplicity, we have $|S_{\lambda_i}| = m_{\lambda_i} \leq r_{\lambda_i}$. Let $S = \bigcup_{i=1}^{k} S_{\lambda_i}$ be the union of all the basis for the eigenspaces. Then

$$|S| = \sum_{i=1}^{k} |S_{\lambda_i}| = \sum_{i=1}^{k} m_{\lambda_i} \leq \sum_{i=1}^{k} r_{\lambda_i} = n.$$

By the previous lemma, S is linearly independent. So S is a basis for $\mathbb{R}^n$ if and only if $|S| = n$, which is equivalent to $\dim(E_{\lambda_i}) = m_{\lambda_i} = r_{\lambda_i}$ for all $i = 1, ..., k$. $\square$

MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 12 Notes

References

1. Elementary Linear Algebra: Application Version, Section 5.3-5.4, Appendix B
2. Linear Algebra with Application, Section 3.5, 6.6, 8.7, Appendix A

7 System of Linear Differential Equations

A $m \times n$ matrix with function entries, or function-valued matrix (with variable t) has the form

$$\mathbf{A}(t) = (a_{ij}(t))_{m\times n},$$

where for each $i = 1, ..., m$, $j = 1, ..., n$, $a_{ij}(t)$ is a function. The domain of the function-valued matrix $\mathbf{A}(t)$ is the intersection of all the domains of the functions $a_{ij}(t)$, $i = 1, ..., m$, $j = 1, ..., n$.

Example. Consider the following function-valued matrix

$$\mathbf{A}(t) = \begin{pmatrix} 1/t & t^2 \\ t & \sqrt{t} \end{pmatrix}.$$

The domain of $\mathbf{A}(t)$ is all the positive t, $t > 0$.

An n-vector with function entries, or function-valued n-vector is a $n \times 1$ matrix with function entries.

Example. $\mathbf{v}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$. Domain of $\mathbf{v}(t)$ is all the real numbers, $t \in \mathbb{R}$.

If there is no need to specify the variable, or if it is clear in the context, we may just use $\mathbf{v}$ to denote a function-valued vector. Similarly for function-valued matrices.

A function-valued vector $v(t) = \begin{pmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_n(t) \end{pmatrix}$ is differentiable if each $v_i(t)$ is differentiable for $i = 1, ..., n$. For a differentiable vector v, the derivative is defined as

$$\mathbf{v}'(t) = \begin{pmatrix} v_1'(t) \\ v_2'(t) \\ \vdots \\ v_n'(t) \end{pmatrix}, \text{ where } v_i'(t) = \frac{d}{dt}v_i(t), i = 1, ..., n.$$

Example. The derivative of $\mathbf{v}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$ is

$$\mathbf{v}'(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}.$$

7.1 First Order Linear System of Differential Equations

A first order linear system of differential equations (with variable t) can be written as

$$\left\{\begin{array}{rcccccccc} y_1'(t) & = & a_{11}(t)y_1(t) & + & \cdots & + & a_{1n}(t)y_n(t) & + & g_1(t) \\ y_2'(t) & = & a_{21}(t)y_1(t) & + & \cdots & + & a_{2n}(t)y_n(t) & + & g_2(t) \\ & \vdots & & & & & & & \\ y_n'(t) & = & a_{n1}(t)y_1(t) & + & \cdots & + & a_{nn}(t)y_n(t) & + & g_n(t) \end{array}\right. ,$$

where $y_i(t)$, $g_i(t)$, $a_{ij}(t)$ are all functions, for $i, j = 1, ..., n$. The above system is equivalent to the following function-valued matrix equation

$$\begin{pmatrix} y_1'(t) \\ y_2'(t) \\ \vdots \\ y_n'(t) \end{pmatrix} = \begin{pmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & \cdots & a_{2n}(t) \\ \vdots & \ddots & \vdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} + \begin{pmatrix} g_1(t) \\ g_2(t) \\ \vdots \\ g_n(t) \end{pmatrix},$$

or

$$\mathbf{y}'(t) = \mathbf{A}(t)\mathbf{y}(t) + \mathbf{g}(t),$$

where $\mathbf{A}(t) = (a_{ij}(t))$, $\mathbf{y} = (y_i(t))$, and $\mathbf{g} = (g_i(t))$. Typically, our task is given the functions a_{ij} and g_i, $i, j = 1, ..., n$, solve for the unknown functions $y_i(t)$, $i = 1, ..., n$ such that the above system is satisfied. A differential equation is linear when the unknown functions are acted upon by multiplying by the known functions and adding them up. It is first order if the highest derivative is the first derivative. It is an ordinary differential equation if the derivative is taken with respect to only one variable (as opposed to partial differential equation). A system of first order linear ordinary differential equations (ODE) is a finite collection of first order linear ODE.

A first order linear system of differential equations is homogeneous if $g_i(t) = 0$ for all $i = 1, .., n$, that is, if it has the form

$$\mathbf{y}'(t) = \mathbf{A}(t)\mathbf{y}(t),$$

and it is of constant coefficient if $a_{ij} \in \mathbb{R}$ are constants, that is, $\mathbf{A}$ is a real-valued matrix. So, a first order homogeneous linear system of differential equations with constant coefficient is of the form

$$\mathbf{y}'(t) = \mathbf{A}\mathbf{y}(t).$$

Finally, an initial condition for the linear system of differential equations is

$$\mathbf{y}(t_0) = \begin{pmatrix} y_1(t_0) \\ y_2(t_0) \\ \vdots \\ y_n(t_0) \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n.$$

Example. 1. $\left\{\begin{array}{rcccccc} y_1' & = & \cos(t)y_1 & - & \sin(t)y_2 & + & t^2 \\ y_2' & = & \sin(t)y_1 & + & \cos(t)y_2 & + & 2t \end{array}\right.$, or

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} t^2 \\ 2t \end{pmatrix}.$$

This is a first order non-homogeneous linear system of differential equations with non-constant coefficient. The associated homogeneous first order linear system of differential equations is $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$.

2. $\left\{ \begin{array}{rcrcr} y_1' & = & y_1 & + & y_2 \\ y_2' & = & 2y_1 & - & y_2 \end{array} \right.$, or

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

This is a first order homogeneous linear system of differential equations with constant coefficient.

In this module, we will only be discussing techniques in solving homogeneous first order linear system of differential equations with constant coefficients. Throughout the rest of this section, the coefficient matrix $\mathbf{A}$ will be constant.

7.2 Solutions to System of Different Equations

A function-valued vector $\mathbf{x}(t)$ is a solution to the differential system

$$\mathbf{y}'(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{g}(t)$$

if

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{g}(t).$$

If further $\mathbf{y}(t_0) = \mathbf{a} \in \mathbb{R}^n$ is an initial condition for the system, then $\mathbf{x}(t)$ is a solution to the initial value problem if

$$\mathbf{x}(t_0) = \mathbf{a}.$$

Consider the differential equation $\frac{dy}{dt} = \lambda y$. It is known that the function $y = e^{\lambda t}$ is a solution. Consider now the following system of differential equations $y_1' = \lambda_1 y_1$, $y_2' = \lambda_2 y_2$,...,$y_n' = \lambda_n y_n$, or

$$\mathbf{y}' = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \mathbf{y},$$

where $\mathbf{y} = (y_i)$. It has a solution $y_1 = e^{\lambda_1 t}$, $y_2 = e^{\lambda_2 t}$,..., $y_n = e^{\lambda_n t}$. That is

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} e^{\lambda_1 t}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} e^{\lambda_2 t}, ..., \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} e^{\lambda_n t}$$

are solutions to the system of differential equation. Notice that λ_i are the eigenvalues of the matrix $\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$ with associated eigenvector $\mathbf{e}_i$, the i-th vector in the standard basis. This is also true in general for any first order homogeneous linear system of differential equations with constant coefficient.

Theorem. *Suppose* $\mathbf{v} \in \mathbb{R}^n$ *is an eigenvector associated to the eigenvalue* λ *of a matrix* $\mathbf{A}$. *Then* $\mathbf{v}e^{\lambda t}$ *is a solution to the first order homogeneous linear ODE with constant coefficient*

$$\mathbf{y}' = \mathbf{A}\mathbf{y}.$$

Proof. Since λ is an eigenvalue of $\mathbf{A}$ and $\mathbf{v}$ is an associated eigenvector, we have $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Let $\mathbf{y} = \mathbf{v}e^{\lambda t}$. Then

$$\mathbf{y}' = \frac{d}{dt}\mathbf{v}e^{\lambda t} = \lambda\mathbf{v}e^{\lambda t} = \mathbf{A}\mathbf{v}e^{\lambda t} = \mathbf{A}\mathbf{y}.$$

Hence, $\mathbf{y} = \mathbf{v}e^{\lambda t}$ is indeed a solution to the differential system. $\square$

Theorem (Superposition principle)**.** *Suppose* $\mathbf{x}_1(t)$ *and* $\mathbf{x}_2(t)$ *are solutions to* $\mathbf{y}' = \mathbf{A}\mathbf{y}$. *For any* $\alpha, \beta \in \mathbb{R}$,

$$\alpha\mathbf{x}_1(t) + \beta\mathbf{x}_2(t)$$

is also a solution to the system of differential equations. That is, linear combinations of solutions to a first order homogeneous linear system of differential equations with constant coefficient is a solution.

A function-valued vector $\mathbf{v}(t)$ is zero if it is the constant $\mathbf{0}$ vector, that is, for every t in its domain, $\mathbf{v}(t) = \mathbf{0} \in \mathbb{R}^n$. A set $\{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_k(t)\}$ of function-valued vector is linearly independent if whenever $c_1, c_2, ..., c_k \in \mathbb{R}$ are real numbers such that

$$c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \cdots + c_k\mathbf{x}_k(t) = \mathbf{0}$$

for all t in the domain, necessarily $c_1 = c_2 = \cdots = c_k = 0$. That is, the only linear combination to obtain the constant zero is the trivial one. Otherwise, we say that it is linearly dependent.

Let $S = \{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_n(t)\}$ be a set containing n functioned-valued vectors with n coordinates. Define the Wronskian of S to be

$$W(\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_n(t)) = \det(\begin{pmatrix}\mathbf{x}_1(t) & \mathbf{x}_2(t) & \cdots & \mathbf{x}_n(t)\end{pmatrix}).$$

This is a real-valued function. Then $S = \{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_n(t)\}$ is linearly independent if the Wronskian is not the constant zero function, $W(\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_n(t)) \neq 0$ (as functions).

Example.

1. Let $S = \left\{\begin{pmatrix}\cos(t)\\ t\end{pmatrix}, \begin{pmatrix}\sin(t)\\ t\end{pmatrix}\right\}$. Then the Wronskian is

$$W\left(\begin{pmatrix}\cos(t)\\ t\end{pmatrix}, \begin{pmatrix}\sin(t)\\ t\end{pmatrix}\right) = \begin{vmatrix}\cos(t) & \sin(t)\\ t & t\end{vmatrix} = t(\cos(t) - \sin(t)),$$

which is not the constant zero function. Hence, S is linearly independent.

2. Let $S = \left\{\begin{pmatrix}e^t\\ 2e^t\\ -e^t\end{pmatrix}, \begin{pmatrix}-e^t\\ 0\\ e^t\end{pmatrix}, \begin{pmatrix}e^{2t}\\ e^{2t}\\ 0\end{pmatrix}\right\}$. Then the Wronskian is

$$W\left(\begin{pmatrix}e^t\\ 2e^t\\ -e^t\end{pmatrix}, \begin{pmatrix}-e^t\\ 0\\ e^t\end{pmatrix}, \begin{pmatrix}e^{2t}\\ e^{2t}\\ 0\end{pmatrix}\right) = \begin{vmatrix}e^t & -e^t & e^{2t}\\ 2e^t & 0 & e^{2t}\\ -e^t & e^t & 0\end{vmatrix} = (e^t)(e^t)(e^{2t})\begin{vmatrix}1 & -1 & 1\\ 2 & 0 & 1\\ -1 & 1 & 0\end{vmatrix} = 2e^{4t},$$

which is not the constant zero function. Hence, S is linearly independent. Note that we may factorize out the common factors in the columns in computing the determinant.

Remark. Warning: A set $S = \{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_n(t)\}$ of function-valued vectors can be linearly independent even though the Wronskian is the constant zero function. For example, consider

$$S = \left\{ \begin{pmatrix} t \\ t \end{pmatrix}, \begin{pmatrix} t^2 \\ t^2 \end{pmatrix} \right\}.$$

Then S is linearly independent since

$$c_1 \begin{pmatrix} t \\ t \end{pmatrix} + c_2 \begin{pmatrix} t^2 \\ t^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ for all } t \Leftrightarrow t(c_1 + c_2 t) = 0 \text{ for all } t \Leftrightarrow c_1 = 0 = c_2.$$

But the Wronskian is

$$W\left(\begin{pmatrix} t \\ t \end{pmatrix}, \begin{pmatrix} t^2 \\ t^2 \end{pmatrix} \right) = \begin{vmatrix} t & t^2 \\ t & t^2 \end{vmatrix} = t^3 - t^3 = 0$$

for all t.

A set $S = \{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_n(t)\}$ of solutions to the differential system $\mathbf{y}' = \mathbf{A}\mathbf{y}$ is called a fundamental set of solutions if its Wronskian is nonzero. A general solution to $\mathbf{y}' = \mathbf{A}\mathbf{y}$ captures all possible linear combinations of the function-valued vectors in a fundamental set of solutions, that is, it is of the form

$$\mathbf{x} = c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \cdots + c_n\mathbf{x}_n(t), c_1, c_2, ..., c_n \in \mathbb{R}.$$

Suppose $\mathbf{y}(t_0) = \mathbf{a} \in \mathbb{R}^n$ is an initial condition for the system, then we are able to find (the) solution to the initial value problem by solving for $c_1, c_2, ..., c_n$ in the equation

$$c_1\mathbf{x}_1(t_0) + c_2\mathbf{x}_2(t_0) + \cdots + c_n\mathbf{x}_n(t_0) = \mathbf{a}.$$

Theorem. *Suppose* $\mathbf{A}$ *is diagonalizable. Let* $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ *be* n *linearly independent eigenvectors associated to (real) eigenvalues* $\lambda_1, \lambda_2, ..., \lambda_n$ *(not necessarily distinct). Then*

$$\left\{\mathbf{v}_1 e^{\lambda_1 t}, \mathbf{v}_2 e^{\lambda_2 t}, ..., \mathbf{v}_n e^{\lambda_n t}\right\}$$

is a fundamental set of solutions to the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ *with constant coefficients, and*

$$\mathbf{x} = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t} + \cdots + c_n\mathbf{v}_n e^{\lambda_n t}$$

is a general solution.

Proof. We have already shown that if $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ associated to eigenvalue λ, then $\mathbf{v}e^{\lambda t}$ is a solution to the differential system $\mathbf{y}' = \mathbf{A}\mathbf{y}$. Hence, $\mathbf{v}_i e^{\lambda_i t}$ is a solution for all $i = 1, ..., n$. The Wronskian of the set is

$$W(\mathbf{v}_1 e^{\lambda_1 t}, \mathbf{v}_2 e^{\lambda_2 t}, ..., \mathbf{v}_n e^{\lambda_n t}) = e^{\lambda_1 t} e^{\lambda_2 t} \cdots e^{\lambda_n t} \begin{vmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{vmatrix} \neq 0$$

since $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ is linearly independent, and thus $\begin{vmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{vmatrix} \neq 0$ □

Example. Solve the differential system

$$\begin{cases} y_1' &= y_1 \\ y_2' &= y_1 + 2y_2 \end{cases}$$

with initial conditions: $y_1(0) = 1$, $y_2(0) = 1$. Write

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The eigenvalues of $\mathbf{A}$ are $\lambda = 1, 2$. Since $\mathbf{A}$ has 2 distinct eigenvalues, it is diagonalizable. We will now compute the eigenvectors.

$\lambda = 1$: $\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0 & 0 \\ -1 & -1 \end{pmatrix}$. So $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is an associated eigenvector.

$\lambda = 2$: $2\mathbf{I} - \mathbf{A} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$. So $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is an associated eigenvector.

By the theorem above,

$$\left\{ e^t \begin{pmatrix} 1 \\ -1 \end{pmatrix}, e^{2t} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

is a fundamental set of solutions. Indeed, its Wronskian is

$$W\left(\begin{pmatrix} e^t \\ -e^t \end{pmatrix}, \begin{pmatrix} 0 \\ e^{2t} \end{pmatrix} \right) = \begin{vmatrix} e^t & 0 \\ -e^t & e^{2t} \end{vmatrix} = e^{3t} \neq 0.$$

So a general solution of the differential system is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = c_1 e^t \begin{pmatrix} 1 \\ -1 \end{pmatrix} + c_2 e^{2t} \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

that is, $y_1 = c_1 e^t$ and $y_2 = -c_1 e^t + c_2 e^{2t}$.

We will now find the solution satisfying the initial condition. Substituting the initial conditions, we have

$$\begin{aligned} 1 &= y_1(0) = c_1 \\ 1 &= y_2(0) = -c_1 + c_2 \end{aligned}$$

Solving the system, we get

$$\left(\begin{array}{cc|c} 1 & 0 & 1 \\ -1 & 1 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array} \right),$$

that is, $c_1 = 1$ and $c_2 = 2$. So the (unique) solution to the initial value problem is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} e^t \\ -e^t + 2e^{2t} \end{pmatrix},$$

or $y_1 = e^t$ and $y_2 = 2e^{2t} - e^t$.

Recall that there are two obstructions to diagonalization. Roughly speaking, either there are not enough (real) eigenvalues, or not enough linearly independent eigenvectors. To be precise, either

(i) the characteristic polynomial does not factorize into real linear factors,

(ii) the geometric multiplicity of an eigenvalue is strictly less than the algebraic multiplicity.

The solutions are

(i) allow complex eigenvalues,

(ii) use generalized eigenvectors.

In the next few sections, we will discuss the two solutions above.

7.3 Complex Eigenvalues and Eigenvetors

Readers may refer to the appendix for an introduction to complex numbers.

An <u>n-complex vector</u> is a collection of n ordered complex numbers,

$$\mathbf{v} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}, z_j \in \mathbb{C} \text{ for all } j = 1, ..., n.$$

The collection of all n-complex vectors is denoted as $\mathbb{C}^n$. Given any complex vector $\mathbf{v} \in \mathbb{C}^n$, we can split it into its real and imaginary parts,

$$\mathbf{v} = \begin{pmatrix} x_1 + iy_1 \\ x_2 + iy_2 \\ \vdots \\ x_n + iy_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + i \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = Re(\mathbf{v}) + iIm(\mathbf{v}),$$

where $Re(\mathbf{v}), Im(\mathbf{v}) \in \mathbb{R}^n$.

Let $\mathbf{A}$ be an order n square matrix. A complex number $\lambda \in \mathbb{C}$ is an <u>eigenvalue</u> of $\mathbf{A}$ if there is a nonzero vector $\mathbf{v} \in \mathbb{C}^n$, $\mathbf{v} \neq \mathbf{0}$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

In this case, $\mathbf{v}$ is called an <u>eigenvector</u> associated to λ.

Theorem. *Let $\mathbf{A}$ be an order n square matrix with real entries.*

(i) Then complex eigenvalues of $\mathbf{A}$ comes in conjugate pairs, that is, if $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$, then $\overline{\lambda}$ is also an eigenvalue of $\mathbf{A}$.

(ii) If $\mathbf{v} \in \mathbb{C}^n$ is an eigenvector associated to eigenvalue λ, then $\overline{\mathbf{v}}$ is an eigenvectors associated to eigenvalue $\overline{\lambda}$.

Proof. Suppose $\lambda \in \mathbb{C}$ is a complex eigenvalue of $\mathbf{A}$. Let $\mathbf{v} \in \mathbb{C}^n$ be an eigenvector associated to eigenvalue λ, that is $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Then since $\mathbf{A}$ has real entries, $\overline{\mathbf{A}} = \mathbf{A}$ and so

$$\mathbf{A}\overline{\mathbf{v}} = \overline{\mathbf{A}\mathbf{v}} = \overline{\lambda\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

Since $\overline{\mathbf{v}} \neq \mathbf{0}$, $\overline{\mathbf{v}}$ is a witness to $\overline{\lambda}$ being and eigenvalue of $\mathbf{A}$. □

In words, this means that complex eigenvalues and complex eigenvectors come in conjugate pairs.

The algorithm to find the complex eigenvalues and eigenvectors are analogous to that of their real counterparts.

Example. Let $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The characteristic polynomial is

$$\begin{vmatrix} x & 1 \\ -1 & x \end{vmatrix} = x^2 + 1 = (x+i)(x-i).$$

The matrix $\mathbf{A}$ does not have real eigenvalues, its complex eigenvalues are $\lambda = \pm i$. The eigenvalues are indeed conjugate pairs. We will now compute the eigenvectors. Since the eigenvectors come in conjugate pairs, suffice to compute an eigenvector associated to one of the eigenvalue, say $\lambda = i$.

$$i\mathbf{I} - \mathbf{A} = \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix}$$

Since the two rows are necessarily linearly dependent, because $(i\mathbf{I} - \mathbf{A})$ cannot be invertible, we may use any of the rows to read of a general solution of the homogeneous system. Reading off the second row, we obtain that $\mathbf{v} = \begin{pmatrix} i \\ 1 \end{pmatrix}$ is an eigenvector associated to eigenvalue i. Indeed,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ i \end{pmatrix} = i \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

Then $\overline{\mathbf{v}} = \begin{pmatrix} -i \\ 1 \end{pmatrix}$ is an eigenvector associated to eigenvalue $\lambda = \overline{i} = -i$. Indeed,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -i \end{pmatrix} = -i \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Consider the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ with constant coefficient. Suppose $\lambda \in \mathbb{C}$ is a complex eigenvalue of $\mathbf{A}$ associated to complex eigenvector $\mathbf{v} \in \mathbb{C}^n$. Decompose λ and $\mathbf{v}$ into their real and imaginary parts,

$$\lambda = \lambda_r + i\lambda_i, \mathbf{v} = \mathbf{v}_r + i\mathbf{v}_i, \lambda_r, \lambda_i \in \mathbb{R}, \mathbf{v}_r, \mathbf{v}_i \in \mathbb{R}^n.$$

Then

$$\begin{aligned} e^{\lambda t}\mathbf{v} &= e^{(\lambda_r + i\lambda_i)t}(\mathbf{v}_r + i\mathbf{v}_i) \\ &= e^{\lambda_r t}(\cos(\lambda_i t) + i\sin(\lambda_i t))(\mathbf{v}_r + i\mathbf{v}_i) \\ &= e^{\lambda_r t}(\cos(\lambda_i t)\mathbf{v}_r - \sin(\lambda_i t)\mathbf{v}_i) + ie^{\lambda_r t}(\sin(\lambda_i t)\mathbf{v}_r + \cos(\lambda_i t)\mathbf{v}_i) \\ &= \mathbf{x}_r(t) + i\mathbf{x}_i(t), \end{aligned}$$

where

$$Re(e^{\lambda t}\mathbf{v}) = \mathbf{x}_r(t) = e^{\lambda_r t}(\cos(\lambda_i t)\mathbf{v}_r - \sin(\lambda_i t)\mathbf{v}_i)$$

and

$$Im(e^{\lambda t}\mathbf{v}) = \mathbf{x}_i(t) = e^{\lambda_r t}(\sin(\lambda_i t)\mathbf{v}_r + \cos(\lambda_i t)\mathbf{v}_i)$$

are real function-valued vectors.

Theorem. *Both $\mathbf{x}_r(t)$ and $\mathbf{x}_i(t)$ are linearly independent real solutions to the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ with constant coefficient.*

Corollary. *Suppose $\mathbf{A}$ is an order 2 square matrix and $\lambda \in \mathbb{C}$ is a nonreal complex eigenvalue $\lambda \notin \mathbb{R}$. Let $\mathbf{v} \in \mathbb{C}^2$ be an associated eigenvector. Write $e^{\lambda t}\mathbf{v} = \mathbf{x}_r(t) + i\mathbf{x}_i(t)$. Then $\{\mathbf{x}_r(t), \mathbf{x}_i(t)\}$ is a fundamental set of solutions for the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ with constant coefficients.*

Example. Solve the following differential system

$$\begin{aligned} y_1 &= \quad - y_2 \\ y_2 &= y_1 \end{aligned}$$

with initial conditions $y_1(0) = 1 = y_2(0)$.

Let $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. We have computed that $\lambda = i$ is an eigenvalue with eigenvector $\begin{pmatrix} i \\ 1 \end{pmatrix}$. Then we say that λ

$$\lambda_r = 0, \lambda_i = 1, \mathbf{v}_r = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mathbf{v}_i = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

and so

$$\begin{aligned} \mathbf{x}_r(t) &= e^{0t}(\cos(t)\begin{pmatrix} 0 \\ 1 \end{pmatrix} - \sin(t)\begin{pmatrix} 1 \\ 0 \end{pmatrix}) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix} \\ \mathbf{x}_i(t) &= e^{0t}(\sin(t)\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \cos(t)\begin{pmatrix} 1 \\ 0 \end{pmatrix}) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \end{aligned}$$

are solutions to the differential equations. The Wronskian is

$$\begin{vmatrix} -\sin(t) & \cos(t) \\ \cos(t) & \sin(t) \end{vmatrix} = -\sin^2(t) - \cos^2(t) \neq 0$$

for any $t \in \mathbb{R}$. Hence

$$\left\{ \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}, \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \right\}$$

is a fundamental set of solution, and a general solution is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = c_1 \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix} + c_2 \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$

Now substituting the initial conditions into the general solution, we have

$$1 = y_1(0) = c_2, 1 = y_2(0) = c_1.$$

Therefore, the solution to the initial value problem is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix} + \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix},$$

or

$$\begin{aligned} y_1 &= \cos(t) - \sin(t) \\ y_2 &= \cos(t) + \sin(t). \end{aligned}$$

7.4 Repeated Eigenvalue and Generalized Eigenvector

In this section, we will restriction our attention only to order 2 square matrices. Readers may refer to the appendix for the discussion for general square matrices.

Suppose $\mathbf{A}$ is an order 2 square matrix and λ is an eigenvalue of $\mathbf{A}$ with algebraic multiplicity $r_\lambda = 2$ and geometric multiplicity $\dim(E_\lambda) = 1$. Then we say that λ is a repeated eigenvalue. Let $\mathbf{v}_1 \in \mathbb{R}^2$ be an eigenvector assocaited to λ. A vector $\mathbf{v}_2 \in \mathbb{R}^2$ is a generalized eigenvector if it is a solution to

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{v}_1.$$

Let

$$\begin{aligned} \mathbf{x}_1(t) &= e^{\lambda t}\mathbf{v}_1 \\ \mathbf{x}_2(t) &= e^{\lambda t}(t\mathbf{v}_1 + \mathbf{v}_2). \end{aligned}$$

Then $\{\mathbf{x}_1(t), \mathbf{x}_2\}$ is a linearly independent set of real solutions to the differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$.

Remark. Note that when we are computing eigenvectors, it does not matter if we are solving for $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ or $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$. However, in solving for generalized eigenvectors, we are looking for solutions to $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{v}_1$. If we exchange the position of $\lambda\mathbf{I}$ and $\mathbf{A}$, then we are solving for $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = -\mathbf{v}_1$ instead. Readers may refer to the appendix for details.

Theorem. *Let $\mathbf{A}$ be an order 2 square matrix and λ be a repeated eigenvalue. Suppose $\mathbf{v}_1$ is an eigenvector and $\mathbf{v}_2$ a generalized eigenvector associated to λ. Then $\{\mathbf{x}_1(t), \mathbf{x}_2(t)\}$ is a fundamental set of solutions for the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ with constant coefficients.*

Example. Solve the differential system

$$\begin{aligned} y_1' &= -y_1 + y_2 \\ y_2' &= -4y_1 + 3y_2 \end{aligned}$$

with initial conditions $y_1(0) = 2$, $y_2(0) = 1$.

Let $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ -4 & 3 \end{pmatrix}$. Compute the eigenvalues.

$$\begin{vmatrix} x+1 & -1 \\ 4 & x-3 \end{vmatrix} = (x-1)^2.$$

The eigenvalue is $\lambda = 1$ with algebraic multiplicity $r_1 = 2$. Compute the eigenspace.

$$\mathbf{I} - \mathbf{A} = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix},$$

reading off the first row, we have $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. This shows that the geometric multiplicity is 1. We will need to compute the generalized eigenvector, that is, to solve for $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{v}_1$.

$$\left(\begin{array}{cc|c} -2 & 1 & 1 \\ -4 & 2 & 2 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & -1/2 & -1/2 \\ 0 & 0 & 0 \end{array}\right).$$

A general solution is $\begin{pmatrix} (1/2)(s-1) \\ s \end{pmatrix}$, $s \in \mathbb{R}$. Choose $s = 1$, then $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is a generalized eigenvector. So

$$\begin{aligned} \mathbf{x}_1(t) &= e^t \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ \mathbf{x}_2(t) &= e^t \left(t \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \end{aligned}$$

are solutions to the differential equations. The Wronskian is

$$\begin{vmatrix} e^2 & t \\ 2e^t & e^t(2t+1) \end{vmatrix} = e^{2t} \begin{vmatrix} 1 & t \\ 2 & 2t+1 \end{vmatrix} = e^{2t} \neq 0$$

for any $t \in \mathbb{R}$. Hence,

$$\left\{ e^t \begin{pmatrix} 1 \\ 2 \end{pmatrix}, e^t \left(t \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \right\}$$

is a fundamental set of solutions, and a general solution is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = c_1 e^t \begin{pmatrix} 1 \\ 2 \end{pmatrix} + c_2 e^t \left(t \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right), c_1, c_2 \in \mathbb{R}.$$

Now substitute the initial condition into the general solution, we have

$$2 = y_1(0) = c_1, 1 = y_2(0) = 2c_1 + c_2$$

and so $c_1 = 0$ and $c_2 = -1$. Therefore, the solution to the initial value problem is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 2e^t \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3e^t \left(t \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = e^t \begin{pmatrix} 2-3t \\ 1-6t \end{pmatrix},$$

or

$$\begin{aligned} y_1 &= e^t(2-3t) \\ y_2 &= e^t(1-6t) \end{aligned}$$

Remark. Note that when solving for generalized eigenvector, we will get a general solution. In the example above, the general solution is $\begin{pmatrix} (1/2)(s-1) \\ s \end{pmatrix}$. In finding the fundamental set of solutions or general solution for the differential system, we are not to include the parameter s; the function-valued vectors are functions in variable t. We must always pick a particular solution for the generalized eigenvector when forming the fundamental set of solutions or general solution to the differential system.

Appendix to Lecture 12

Solutions to First order homogeneous differential system with constant coefficient

Theorem. *Suppose* $\mathbf{v} \in \mathbb{R}^n$ *is an eigenvector associated to the eigenvalue* λ *of a matrix* $\mathbf{A}$. *Then* $\mathbf{v}e^{\lambda t}$ *is a solution to the first order homogeneous linear ODE with constant coefficient*

$$\mathbf{y}' = \mathbf{A}\mathbf{y}.$$

Proof. Since λ is an eigenvalue of $\mathbf{A}$ and $\mathbf{v}$ is an associated eigenvector, we have $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Let $\mathbf{y} = \mathbf{v}e^{\lambda t}$. Then

$$\mathbf{y}' = \frac{d}{dt}\mathbf{v}e^{\lambda t} = \lambda\mathbf{v}e^{\lambda t} = \mathbf{A}\mathbf{v}e^{\lambda t} = \mathbf{A}\mathbf{y}.$$

Hence, $\mathbf{y} = \mathbf{v}e^{\lambda t}$ is indeed a solution to the differential system. □

Theorem (Superposition principle)**.** *Suppose* $\mathbf{x}_1(t)$ *and* $\mathbf{x}_2(t)$ *are solutions to* $\mathbf{y}' = \mathbf{A}\mathbf{y}$. *For any* $\alpha, \beta \in \mathbb{R}$,

$$\alpha\mathbf{x}_1(t) + \beta\mathbf{x}_2(t)$$

is also a solution to the system of differential equations. That is, linear combinations of solutions to a first order homogeneous linear system of differential equations with constant coefficient is a solution.

Proof.

$$\begin{aligned} \frac{d}{dt}(\alpha\mathbf{x}_1(t) + \beta\mathbf{x}_2(t)) &= \alpha\frac{d}{dt}\mathbf{x}_1(t) + \beta\frac{d}{dt}\mathbf{x}_2(t) \\ &= \alpha\mathbf{A}\mathbf{x}_1(t) + \beta\mathbf{A}\mathbf{x}_2(t) \\ &= \mathbf{A}(\alpha\mathbf{x}_1(t) + \beta\mathbf{x}_2(t)). \end{aligned}$$

□

Introduction to Complex Numbers

Define the imaginary number i to be such that $i^2 = -1$, or $i = \sqrt{-1}$. It is not a real number, $i \notin \mathbb{R}$. A complex number can be written as

$$z = x + iy,$$

for some $x, y \in \mathbb{R}$. The collection of all complex numbers is denoted as

$$\mathbb{C} = \{\ z = x + iy \mid x, y \in \mathbb{R}\ \}.$$

We can think of the set of real numbers as a subset of the set of all complex numbers, $\mathbb{R} \subseteq \mathbb{C}$, via

$$x \mapsto x + i0.$$

For a complex number $z = x + iy$, $Re(z) = x$ is called the real part and $Im(z) = y$ is called the imaginary part. A complex number z is purely imaginary if $z = 0 + iy$, that is $Re(z) = 0$, denoted as $z \in i\mathbb{R}$; it is non-real if the imaginary part is nonzero, $z = x + iy \in \mathbb{C}\backslash\mathbb{R}$, $y \neq 0$.

Here are some operations for complex numbers

(i) Addition: $(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2)$

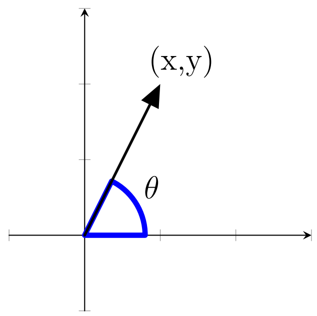
(ii) Multiplication: $(x_1 + iy_1)(x_2 + iy_2) = (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)$

(iii) Conjugation: $\overline{x + iy} = x - iy$.

(iv) Norm: $|x + iy| = \sqrt{x^2 + y^2} = \sqrt{(x + iy)\overline{(x + iy)}}$

Exercise: Write $\overline{i(1 + 2i)}$ as $x + iy$ for some $x, y \in \mathbb{R}$.

Every complex number $z = x + iy$ can be represented by a point (x, y) on the xy-plane, $\mathbb{R}^2$. The argument of a complex number $z = x + iy$ is the value θ such that $z = r(\cos\theta + i \sin\theta)$, where $r = |z|$. Equivalently, if $z = x + iy$ corresponds to (x, y) on $\mathbb{R}^2$, then the argument θ of z is such that $(x, y) = (r\cos\theta, r\sin\theta)$. The expression $z = r(\cos\theta + i\sin\theta)$ is called the polar form of z.



Recall that we have the Euler's formula

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

Thus every complex number z can be written as $z = re^{i\theta}$, where $r = |z|$ and θ is the argument of z. This is called the exponential form of z.

Complex Eigenvalues and Eigenvectors

Recall that if λ is an eigenvalue of a matrix $\mathbf{A}$ with associated eigenvector $\mathbf{v}$, then $e^{\lambda t}\mathbf{v}$ is a solution to the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ with constant coefficient. This is true also for complex eigenvalues and eigenvectors. The proof is similar to that for the theorem for real eigenvalues and eigenvectors.

Theorem. *Suppose $\lambda \in \mathbb{C}$ is a complex eigenvalue of $\mathbf{A}$ and $\mathbf{v} \in \mathbb{C}^n$ is an associated eigenvector. Then $e^{\lambda t}\mathbf{v}$ is a complex solution to the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ with constant coefficient.*

It turns out that the real and imaginary parts of the complex solutions are real solution to the differential system. Write $Re(e^{\lambda t}\mathbf{v}) = \mathbf{x}_r(t)$ and $Im(e^{\lambda t}\mathbf{v}) = \mathbf{x}_i(t)$. Then we have the following theorem.

Theorem. *Both $\mathbf{x}_r(t)$ and $\mathbf{x}_i(t)$ are linearly independent real solutions to the first order homogeneous system of differential equations $\mathbf{y}' = \mathbf{A}\mathbf{y}$ with constant coefficient.*

Proof. Since $e^{\lambda t}\mathbf{v}$ is a (complex) solution to the differential system,

$$\begin{aligned}\mathbf{A}\mathbf{x}_r(t)+i\mathbf{A}\mathbf{x}_i(t) &= \mathbf{A}(\mathbf{x}_r(t)+i\mathbf{x}_i(t)) = \mathbf{A}(e^{\lambda t}\mathbf{v}) \\ &= \frac{d}{dt}(e^{\lambda t}\mathbf{v}) = \frac{d}{dt}\mathbf{x}_r(t)+i\frac{d}{dt}\mathbf{x}_i(t)\end{aligned}$$

and since $\mathbf{A}$ has real entries and $\mathbf{x}_i(t)$ and $\mathbf{x}_i(t)$ are real function-valued vectors, by compairing the real and imaginary parts, we have

$$\mathbf{A}\mathbf{x}_r(t) = \mathbf{x}_r'(t) \text{ and } \mathbf{A}\mathbf{x}_i(t) = \mathbf{x}_i'(t).$$

Hence, $\mathbf{x}_r(t)$ and $\mathbf{x}_i(t)$ are real solutions to $\mathbf{y}' = \mathbf{A}\mathbf{y}$.

Next, we need to show that $\{\mathbf{x}_r(t), \mathbf{x}_i(t)\}$ is linearly independent. Suppose $c_1, c_2 \in \mathbb{R}$ are such that

$$c_1 e^{\lambda_r t}(\cos(\lambda_i t)\mathbf{v}_r - \sin(\lambda_i t)\mathbf{v}_i) + c_2 e^{\lambda_r t}(\sin(\lambda_i t)\mathbf{v}_r + \cos(\lambda_i t)\mathbf{v}_i) = \mathbf{0}.$$

Let $t=0$ and $t=\frac{2\pi}{\lambda_i}$, we have

$$\begin{array}{rcrcl} c_1\mathbf{v}_r & + & c_2\mathbf{v}_i & = & \mathbf{0} \\ -c_1\mathbf{v}_i & + & c_2\mathbf{v}_r & = & \mathbf{0} \end{array}$$

Since it cannnot be that both $\mathbf{v}_r = \mathbf{0} = \mathbf{v}_i$, it is clear that if $\mathbf{v}_r = \mathbf{0}$ or $\mathbf{v}_i = \mathbf{0}$, necessarily $c_1 = c_2 = 0$. If $\mathbf{v}_r$ is a multiple of $\mathbf{v}_i$, then again it is clear that necessarily $c_1 = c_2 = 0$. Otherwise, $\{\mathbf{v}_r, \mathbf{v}_i\}$ is linearly independent, and thus $c_1 = c_2 = 0$. This shows that $\{\mathbf{x}_r(t), \mathbf{x}_i(t)\}$ is linearly independent. □

Now suppose $\lambda \in \mathbb{C}$ is a complex eigenvalue of $\mathbf{A}$ and $\mathbf{v} \in \mathbb{C}^n$ is a complex eigenvector. From the complex solution $e^{\lambda t}\mathbf{v}$ to the differential system $\mathbf{y}' = \mathbf{A}\mathbf{y}$, we obtain the real solutions $Re(e^{\lambda t}\mathbf{v}) = \mathbf{x}_r(t)$ and $Im(e^{\lambda t}\mathbf{v}) = \mathbf{x}_i(t)$. Now consider the conjugate pair of complex eigenvalue and eigenvector $\overline{\lambda}$ and $\overline{\mathbf{v}}$. From the complex solution $e^{\overline{\lambda}t}\overline{\mathbf{v}}$, we obtain another set of real solutions $Re(e^{\overline{\lambda}t}\overline{\mathbf{v}}) = \mathbf{w}_r(t)$ and $Im(e^{\overline{\lambda}t}\overline{\mathbf{v}}) = \mathbf{w}_i(t)$. However, observe that

$$\begin{aligned}\mathbf{w}_r(t) &= Re(e^{\overline{\lambda}t}\overline{\mathbf{v}}) = Re(\overline{e^{\lambda t}\mathbf{v}}) = Re(e^{\lambda t}\mathbf{v}) = \mathbf{x}_r(t) \\ \mathbf{w}_i(t) &= Im(e^{\overline{\lambda}t}\overline{\mathbf{v}}) = Im(\overline{e^{\lambda t}\mathbf{v}}) = -Im(e^{\lambda t}\mathbf{v}) = -\mathbf{x}_i(t)\end{aligned}$$

and thus $\text{span}\{\mathbf{x}_r(t), \mathbf{x}_i(t)\} = \text{span}\{\mathbf{w}_r(t), \mathbf{w}_i(t)\}$. Therefore, it suffice to extract the real solutions from either the pair λ and $\mathbf{v}$ or $\overline{\lambda}$ and $\overline{\mathbf{v}}$.

Generalized Eigenvectors

Let $\mathbf{A}$ be an order n square matrix and λ an eigenvalue of $\mathbf{A}$. A nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is a generalized eigenvector of $\mathbf{A}$ associated to eigenvalue λ if there is a positive integer $k \in \mathbb{Z}$, $k > 0$, such that

$$(\mathbf{A} - \lambda\mathbf{I})^k\mathbf{v} = \mathbf{0}.$$

If furthermore $(\mathbf{A} - \lambda\mathbf{I})^{k-1}\mathbf{v} \neq \mathbf{0}$, we say that the generalized eigenvector has rank k. Note that rank 1 generalized eigenvectors are the usual eigenvectors. Let $\mathbf{v}_k$ be a rank k generalized eigenvector associated to eigenvalue λ. Define inductively

$$\mathbf{v}_i = (\mathbf{A} - \lambda\mathbf{I})\mathbf{v}_{i+1}$$

for $i = 1, ..., k-1$.

Lemma. $\mathbf{v}_i$ *is a generalized eigenvector of rank* i.

Proof. We will prove be induction on $i \geq 1$. For $i = 1$, $\mathbf{v}_1$ is a usual eigenvector and thus $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$. Also $(\mathbf{A} - \lambda\mathbf{I})^{1-1}\mathbf{v} = \mathbf{I}\mathbf{v} \neq \mathbf{0}$ since $\mathbf{v}$ is an eigenvector. This proves the base case.

Now suppose the statement is true for $i - 1$. Then

$$(\mathbf{A} - \lambda\mathbf{I})^i\mathbf{v}_i = (\mathbf{A} - \lambda\mathbf{I})^{i-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{v}_i = (\mathbf{A} - \lambda\mathbf{I})^{i-1}\mathbf{v}_{i-1} = \mathbf{0}$$

by definition of $\mathbf{v}_{i-1}$and the induction hypothesis. Also

$$(\mathbf{A} - \lambda\mathbf{I})^{i-1}\mathbf{v}_i = (\mathbf{A} - \lambda\mathbf{I})^{i-2}\mathbf{v}_{i-1} \neq \mathbf{0}$$

by the induction hypothesis. Hence, $\mathbf{v}_i$ is indeed a generalized eigenvector of rank i. □

The set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ is called a Jordan chain.

Lemma. *The set* $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ *is linearly independent.*

Proof. First observe that $\text{Null}((\mathbf{A} - \lambda\mathbf{I})^i) \subseteq \text{Null}((\mathbf{A} - \lambda\mathbf{I})^{i+1})$. Next, by definition of $\mathbf{v}_i$, we have

$$\mathbf{v}_i \in \text{Null}((\mathbf{A} - \lambda\mathbf{I})^i) \setminus \text{Null}((\mathbf{A} - \lambda\mathbf{I})^{i-1}).$$

This shows that $\mathbf{v}_i \notin \text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{i-1}\} \subseteq \text{Null}((\mathbf{A} - \lambda\mathbf{I})^{i-1})$ for all $i = 1, ..., k$. And so $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ is linearly independent. □

Let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ be a Jordan chain of generalized eigenvectors of $\mathbf{A}$ associated to eigenvalue λ. Define

$$\begin{aligned}
\mathbf{x}_1(t) &= e^{\lambda t}\mathbf{v}_1, \\
\mathbf{x}_2(t) &= e^{\lambda t}(t\mathbf{v}_1 + \mathbf{v}_2), \\
\mathbf{x}_3(t) &= e^{\lambda t}\left(\frac{t^2}{2}\mathbf{v}_1 + t\mathbf{v}_2 + \mathbf{v}_2\right) \\
&\vdots \\
\mathbf{x}_k(t) &= e^{\lambda t}\left(\frac{t^{k-1}}{(k-1)!}\mathbf{v}_1 + \cdots + \frac{t^2}{2}v_{k-2} + t\mathbf{v}_{k-1} + \mathbf{v}_k\right).
\end{aligned}$$

Lemma. *The set* $\{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_k(t)\}$ *form a linearly independent set of solutions to the differential system* $\mathbf{y}' = \mathbf{A}\mathbf{y}$.

Proof. Write $\mathbf{x}_l(t) = e^{\lambda t}\sum_{i=1}^{l}\frac{t^{l-i}}{(l-i)!}\mathbf{v}_i$ for $l = 1, ..., k$. Then

$$\begin{aligned}
\mathbf{x}_l'(t) &= \lambda e^{\lambda t}\sum_{i=1}^{l}\frac{t^{l-i}}{(l-i)!}\mathbf{v}_i + e^{\lambda t}\sum_{i=1}^{l-1}(l-i)\frac{t^{l-i-1}}{(l-i)!}\mathbf{v}_i \\
&= \lambda e^{\lambda t}\sum_{i=1}^{l}\frac{t^{l-i}}{(l-i)!}\mathbf{v}_i + e^{\lambda t}\sum_{i=1}^{l}\frac{t^{l-i}}{(l-i)!}\mathbf{v}_{i-1},
\end{aligned}$$

where $\mathbf{v}_0 = \mathbf{0}$. Since $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v}_i = \mathbf{v}_{i-1}$, or $\mathbf{A}\mathbf{v}_i = \mathbf{v}_{i-1} + \lambda\mathbf{v}_i$,

$$\begin{aligned}\mathbf{A}\mathbf{x}_l(t) &= e^{\lambda t}\sum_{i=1}^{l}\frac{t^{l-i}}{(l-i)!}\mathbf{A}\mathbf{v}_i \\ &= e^{\lambda t}\sum_{i=1}^{l}\frac{t^{l-i}}{(l-i)!}\mathbf{v}_{i-1} + \lambda e^{\lambda t}\sum_{i=1}^{l}\frac{t^{l-i}}{(l-i)!}\mathbf{v}_i \\ &= \mathbf{x}_l'(t).\end{aligned}$$

This shows that indeed $\mathbf{x}_l(t)$ is a solution, for $l = 1, ..., k$.

Next, we will check that $\{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_k(t)\}$ is indeed linearly independent. Suppose

$$c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \cdots + c_k\mathbf{x}_k(t) = \mathbf{0}.$$

Let $t = 0$, we have $\mathbf{x}_1(0) = \mathbf{v}_1, \mathbf{x}_2(0) = \mathbf{v}_2, ..., \mathbf{x}_k(0) = \mathbf{v}_k$, and since they are linearly independent, necessarily $c_1 = c_2 = \cdots = c_k = 0$. Hence, the set of solutions $\{\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_k(t)\}$ is indeed linearly independent. $\square$

Now suppose $\mathbf{A}$ is an order 2 square matrix and λ is an eigenvalue. Then we can at most have a Jordan chain with 2 generalized eigenvectors, they form a linearly independent set. Here $\mathbf{v}_1$ is an eigenvector λ, and $\mathbf{v}_2$ is a solution to

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{v}_1.$$

Then $\{\mathbf{x}_1(t) = e^{\lambda t}\mathbf{v}_1, \mathbf{x}_2(t) = e^{\lambda t}(t\mathbf{v}_1 + \mathbf{v}_2)\}$ is a linearly independent set of solutions to the differential system $\mathbf{y}' = \mathbf{A}\mathbf{y}$, and thus is a fundamental set of solutions. Thus, we obtained the following theorem.

Theorem. *Let* $\mathbf{A}$ *be an order* 2 *square matrix and* λ *be a repeated eigenvalue. Suppose* $\mathbf{v}_1$ *is an eigenvector and* $\mathbf{v}_2$ *a generalized eigenvector associated to* λ*. Then* $\{\mathbf{x}_1(t), \mathbf{x}_2(t)\}$ *is a fundamental set of solutions for the first order homogeneous system of differential equations* $\mathbf{y}' = \mathbf{A}\mathbf{y}$ *with constant coefficients.*

References

1. Elementary Linear Algebra: Application Version - Anton, Rorres

2. Linear Algebra with Application - Nicholson