Introduction to systems of linear equations

These slides are based on Section 1 in *Linear Algebra and its Applications* by David C. Lay.

Definition 1. A **linear equation** in the variables $x_1, \ldots, x_n$ is an equation that can be written as

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

Example 2. Which of the following equations are linear?

- $4x_1 - 5x_2 + 2 = x_1$ linear: $3x_1 - 5x_2 = -2$
- $x_2 = 2(\sqrt{6} - x_1) + x_3$ linear: $2x_1 + x_2 - x_3 = 2\sqrt{6}$
- $4x_1 - 6x_2 = x_1x_2$ not linear: x_1x_2
- $x_2 = 2\sqrt{x_1} - 7$ not linear: $\sqrt{x_1}$

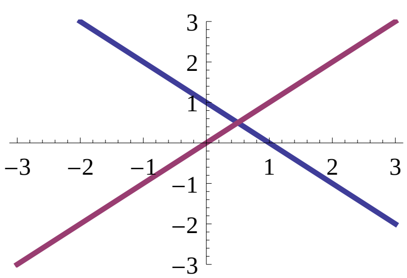
Definition 3.

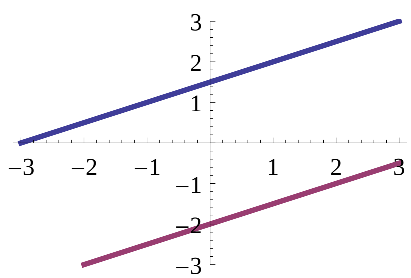
- A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same set of variables, say, $x_1, x_2, \ldots, x_n$.
- A **solution** of a linear system is a list $(s_1, s_2, \ldots, s_n)$ of numbers that makes each equation in the system true when the values $s_1, s_2, \ldots, s_n$ are substituted for $x_1, x_2, \ldots, x_n$, respectively.

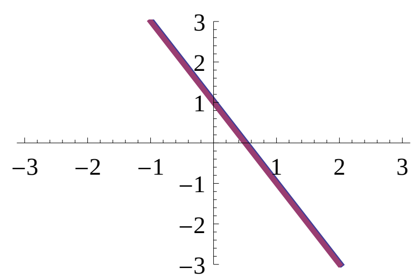
Example 4. (Two equations in two variables)

In each case, sketch the set of all solutions.

$$\begin{array}{rcrcr} x_1 & + & x_2 & = & 1 \\ -x_1 & + & x_2 & = & 0 \end{array} \qquad \begin{array}{rcrcr} x_1 & - & 2x_2 & = & -3 \\ 2x_1 & - & 4x_2 & = & 8 \end{array} \qquad \begin{array}{rcrcr} 2x_1 & + & x_2 & = & 1 \\ -4x_1 & - & 2x_2 & = & -2 \end{array}$$







Theorem 5. A linear system has either

- no solution, or
- one unique solution, or
- infinitely many solutions.

Definition 6. A system is **consistent** if a solution exists.

Armin Straub
astraub@illinois.edu

How to solve systems of linear equations

Strategy: replace system with an equivalent system which is easier to solve

Definition 7. Linear systems are **equivalent** if they have the same set of solutions.

Example 8. To solve the first system from the previous example:

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & 1 \\ -x_1 & + & x_2 & = & 0 \end{array} \overset{R2 \to R2+R1}{\rightsquigarrow} \begin{array}{rcrcl} x_1 & + & x_2 & = & 1 \\ & & 2x_2 & = & 1 \end{array}$$

Once in this **triangular** form, we find the solutions by **back-substitution**:

$$x_2 = 1/2, \qquad x_1 = 1/2$$

Example 9. The same approach works for more complicated systems.

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ -4x_1 & + & 5x_2 & + & 9x_3 & = & -9 \end{array} \quad \Big\downarrow \; R3 \to R3 + 4R1$$

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ & - & 3x_2 & + & 13x_3 & = & -9 \end{array} \quad \Big\downarrow \; R3 \to R3 + \frac{3}{2}R2$$

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ & & & & x_3 & = & 3 \end{array}$$

By back-substitution:

$$x_3 = 3, \qquad x_2 = 16, \qquad x_1 = 29.$$

It is always a good idea to check our answer. Let us check that $(29, 16, 3)$ indeed solves the original system:

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ -4x_1 & + & 5x_2 & + & 9x_3 & = & -9 \end{array} \qquad \begin{array}{rcrcrcr} 29 & - & 2\cdot 16 & + & 3 & \overset{\checkmark}{=} & 0 \\ & & 2\cdot 16 & - & 8\cdot 3 & \overset{\checkmark}{=} & 8 \\ -4\cdot 29 & + & 5\cdot 16 & + & 9\cdot 3 & \overset{\checkmark}{=} & -9 \end{array}$$

Matrix notation

$$\begin{array}{rcrcr} x_1 & - & 2x_2 & = & -1 \\ -x_1 & + & 3x_2 & = & 3 \end{array} \qquad \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix}$$

(coefficient matrix)

$$\left[\begin{array}{rr|r} 1 & -2 & -1 \\ -1 & 3 & 3 \end{array}\right]$$

(augmented matrix)

Armin Straub
astraub@illinois.edu

Definition 10. An **elementary row operation** is one of the following:

- **(replacement)** Add one row to a multiple of another row.
- **(interchange)** Interchange two rows.
- **(scaling)** Multiply all entries in a row by a nonzero constant.

Definition 11. Two matrices are **row equivalent**, if one matrix can be transformed into the other matrix by a sequence of elementary row operations.

Theorem 12. If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Example 13. Here is the previous example in matrix notation.

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ -4x_1 & + & 5x_2 & + & 9x_3 & = & -9 \end{array} \quad \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{array}\right] \quad \Big\downarrow \; R3 \to R3 + 4R1$$

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ & - & 3x_2 & + & 13x_3 & = & -9 \end{array} \quad \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & -3 & 13 & -9 \end{array}\right] \quad \Big\downarrow \; R3 \to R3 + \tfrac{3}{2}R2$$

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & 8x_3 & = & 8 \\ & & & & x_3 & = & 3 \end{array} \quad \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Instead of back-substitution, we can continue with row operations.

After $R2 \to R2 + 8R3$, $R1 \to R1 - R3$, we obtain:

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & & & = & -3 \\ & & 2x_2 & & & = & 32 \\ & & & & x_3 & = & 3 \end{array} \quad \left[\begin{array}{rrr|r} 1 & -2 & 0 & -3 \\ 0 & 2 & 0 & 32 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Finally, $R1 \to R1 + R2$, $R2 \to \frac{1}{2}R2$ results in:

$$\begin{array}{rrrcr} x_1 & & & = & 29 \\ & x_2 & & = & 16 \\ & & x_3 & = & 3 \end{array} \quad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

We again find the solution $(x_1, x_2, x_3) = (29, 16, 3)$.

Row reduction and echelon forms

Definition 14. A matrix is in **echelon form** (or **row echelon form**) if:

(1) Each leading entry (i.e. leftmost nonzero entry) of a row is in a column to the right of the leading entry of the row above it.

(2) All entries in a column below a leading entry are zero.

(3) All nonzero rows are above any rows of all zeros.

Example 15. Here is a representative matrix in echelon form.

$$\begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

($*$ stands for any value, and $\blacksquare$ for any nonzero value.)

Example 16. Are the following matrices in echelon form?

(a) $\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ YES

(b) $\begin{bmatrix} 0 & \blacksquare & * & * & * \\ \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ NOPE (but it is after exchanging the first two rows)

(c) $\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & \blacksquare \\ 0 & 0 & 0 \end{bmatrix}$ YES

(d) $\begin{bmatrix} \blacksquare & 0 & 0 \\ * & \blacksquare & 0 \\ * & 0 & \blacksquare \\ * & 0 & 0 \end{bmatrix}$ NO

Related and extra material

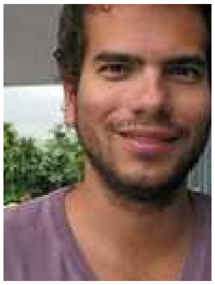
- In our textbook: parts of 1.1, 1.3, 2.2 (just pages 78 and 79)

  However, I would suggest waiting a bit before reading through these parts (say, until we covered things like matrix multiplication in class).

- Suggested practice exercise: 1, 4, 5, 10, 11 from Section 1.3

Pre-lecture trivia

Who are these four?

- Artur Avila, Manjul Bhargava, Martin Hairer, Maryam Mirzakhani
- Just won the **Fields Medal**!
  - analog to Nobel prize in mathematics
  - awarded every four years
  - winners have to be younger than 40
  - cash prize: 15,000 C$

Review

- Each linear system corresponds to an augmented matrix.

$$\begin{array}{rrrcr} 2x_1 & -x_2 & & = & 6 \\ -x_1 & +2x_2 & -x_3 & = & -9 \\ & -x_2 & +2x_3 & = & 12 \end{array} \qquad \left[\begin{array}{rrr|r} 2 & -1 & & 6 \\ -1 & 2 & -1 & -9 \\ & -1 & 2 & 12 \end{array}\right]$$

augmented matrix

- To solve a system, we perform **row reduction**.

$$\overset{R2 \to R2 + \frac{1}{2}R1}{\rightsquigarrow} \left[\begin{array}{rrr|r} 2 & -1 & 0 & 6 \\ 0 & \frac{3}{2} & -1 & -6 \\ 0 & -1 & 2 & 12 \end{array}\right]$$

$$\overset{R3 \to R3 + \frac{2}{3}R2}{\rightsquigarrow} \left[\begin{array}{rrr|r} 2 & -1 & 0 & 6 \\ 0 & \frac{3}{2} & -1 & -6 \\ 0 & 0 & \frac{4}{3} & 8 \end{array}\right]$$

echelon form!

- **Echelon form** in general:

$$\left[\begin{array}{ccccccccccc} 0 & \blacksquare & * & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The leading terms in each row are the **pivots**.

Row reduction and echelon forms, continued

Definition 1. A matrix is in **reduced echelon form** if, in addition to being in echelon form, it also satisfies:

- Each pivot is 1.
- Each pivot is the only nonzero entry in its column.

Example 2. Our initial matrix in echelon form put into reduced echelon form:

$$\begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 0 & \blacksquare & * & 0 & 0 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & \blacksquare & 0 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Note that, to be in reduced echelon form, the pivots $\blacksquare$ also have to be scaled to 1.

Example 3. Are the following matrices in reduced echelon form?

(a) $\begin{bmatrix} 0 & 1 & * & 0 & 0 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \end{bmatrix}$ YES

(b) $\begin{bmatrix} 1 & 0 & 5 & 0 & -7 \\ 0 & 2 & 4 & 0 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ NO

(c) $\begin{bmatrix} 1 & 0 & -2 & 3 & 2 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$ NO

Theorem 4. (Uniqueness of the reduced echelon form) Each matrix is row equivalent to one and only one reduced echelon matrix.

Question. Is the same statement true for the echelon form?

Clearly not; for instance, $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are different row equivalent echelon forms.

Example 5. Row reduce to echelon form (often called **Gaussian elimination**) and then to reduced echelon form (often called **Gauss–Jordan elimination**):

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

Solution.

After $R1 \leftrightarrow R3$, we get: ($R1 \leftrightarrow R2$ would be another option; try it!)

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

Then, $R2 \to R2 - R1$ yields:

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Finally, $R3 \to R3 - \frac{3}{2}R2$ produces the echelon form:

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

To get the reduced echelon form, we first scale all rows:

$$\begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 0 & 1 & -2 & 2 & 1 & -3 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

Then, $R2 \to R2 - R3$ and $R1 \to R1 - 2R3$, gives:

$$\begin{bmatrix} 1 & -3 & 4 & -3 & 0 & -3 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

Finally, $R1 \to R1 + 3R2$ produces the reduced echelon form:

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

Solution of linear systems via row reduction

After row reduction to echelon form, we can easily solve a linear system.
(especially after reduction to reduced echelon form)

Example 6.

$$\left[\begin{array}{ccccc|c} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -8 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{array}\right] \rightsquigarrow \begin{array}{ccccccc} x_1 & +6x_2 & & +3x_4 & & = & 0 \\ & & x_3 & -8x_4 & & = & 5 \\ & & & & x_5 & = & 7 \end{array}$$

- The pivots are located in columns $1, 3, 5$. The corresponding variables x_1, x_3, x_5 are called **pivot variables** (or **basic variables**).
- The remaining variables x_2, x_4 are called **free variables**.
- We can solve each equation for the pivot variables in terms of the free variables (if any). Here, we get:

$$\begin{array}{ccccccc} x_1 & +6x_2 & & +3x_4 & & = & 0 \\ & & x_3 & -8x_4 & & = & 5 \\ & & & & x_5 & = & 7 \end{array} \qquad \left\{\begin{array}{l} x_1 = -6x_2 - 3x_4 \\ x_2 \text{ free} \\ x_3 = 5 + 8x_4 \\ x_4 \text{ free} \\ x_5 = 7 \end{array}\right.$$

- This is the **general solution** of this system. The solution is in parametric form, with parameters given by the free variables.
- Just to make sure: Is the above system consistent? Does it have a unique solution?

Example 7. Find a parametric description of the solution set of:

$$\begin{array}{ccccccc} & 3x_2 & -6x_3 & +6x_4 & +4x_5 & = & -5 \\ 3x_1 & -7x_2 & +8x_3 & -5x_4 & +8x_5 & = & 9 \\ 3x_1 & -9x_2 & +12x_3 & -9x_4 & +6x_5 & = & 15 \end{array}$$

Solution. The augmented matrix is

$$\left[\begin{array}{ccccc|c} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{array}\right].$$

We determined earlier that its reduced echelon form is

$$\left[\begin{array}{ccccc|c} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{array}\right].$$

The pivot variables are x_1, x_2, x_5.

The free variables are x_3, x_4.

Hence, we find the general solution as:

$$\begin{cases} x_1 = -24 + 2x_3 - 3x_4 \\ x_2 = -7 + 2x_3 - 2x_4 \\ x_3 \text{ free} \\ x_4 \text{ free} \\ x_5 = 4 \end{cases}$$

Related and extra material

- In our textbook: still, parts of 1.1, 1.3, 2.2 (just pages 78 and 79)

  As before, I would suggest waiting a bit before reading through these parts (say, until we covered things like matrix multiplication in class).

- Suggested practice exercise:

  Section 1.3: 13, 20; Section 2.2: 2 (only reduce A, B to echelon form)

Review

- We have a standardized recipe to find all solutions of systems such as:

$$\begin{array}{rrrrrcl} & 3x_2 & -6x_3 & +6x_4 & +4x_5 & = & -5 \\ 3x_1 & -7x_2 & +8x_3 & -5x_4 & +8x_5 & = & 9 \\ 3x_1 & -9x_2 & +12x_3 & -9x_4 & +6x_5 & = & 15 \end{array}$$

- The computational part is to start with the **augmented matrix**

$$\left[\begin{array}{rrrrr|r} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{array}\right],$$

  and to calculate its **reduced echelon form** (which is unique!). Here:

$$\left[\begin{array}{rrrrr|r} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{array}\right].$$

- **pivot variables** (or **basic variables**): x_1, x_2, x_5

  free variables: x_3, x_4

- solving each equation for the pivot variables in terms of the free variables:

$$\begin{array}{rrrrcl} x_1 & & -2x_3 & +3x_4 & & = & -24 \\ & x_2 & -2x_3 & +2x_4 & & = & -7 \\ & & & & x_5 & = & 4 \end{array} \qquad \begin{cases} x_1 = -24 + 2x_3 - 3x_4 \\ x_2 = -7 + 2x_3 - 2x_4 \\ x_3 \text{ free} \\ x_4 \text{ free} \\ x_5 = 4 \end{cases}$$

Questions of existence and uniqueness

The question whether a system has a solution and whether it is unique, is easier to answer than to determine the solution set.

All we need is an echelon form of the augmented matrix.

Example 1. Is the following system consistent? If so, does it have a unique solution?

$$\begin{array}{rrrrrcl} & 3x_2 & -6x_3 & +6x_4 & +4x_5 & = & -5 \\ 3x_1 & -7x_2 & +8x_3 & -5x_4 & +8x_5 & = & 9 \\ 3x_1 & -9x_2 & +12x_3 & -9x_4 & +6x_5 & = & 15 \end{array}$$

Solution. In the course of an earlier example, we obtained the echelon form:

$$\left[\begin{array}{rrrrr|r} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{array}\right]$$

Hence, it is consistent (imagine doing back-substitution to get a solution).

Theorem 2. (Existence and uniqueness theorem) A linear system is **consistent** if and only if an echelon form of the augmented matrix has **no** row of the form

$$\left[\begin{array}{ccc|c} 0 & \dots & 0 & b \end{array}\right],$$

where b is nonzero.

If a linear system is consistent, then the solutions consist of either

- a unique solution (when there are no free variables) or
- infinitely many solutions (when there is at least one free variable).

Example 3. For what values of h will the following system be consistent?

$$\begin{array}{rrcl} 3x_1 & -9x_2 & = & 4 \\ -2x_1 & +6x_2 & = & h \end{array}$$

Solution. We perform row reduction to find an echelon form:

$$\left[\begin{array}{rr|r} 3 & -9 & 4 \\ -2 & 6 & h \end{array}\right] \overset{R2 \to R2 + \frac{2}{3}R1}{\rightsquigarrow} \left[\begin{array}{rr|r} 3 & -9 & 4 \\ 0 & 0 & h + \frac{8}{3} \end{array}\right]$$

The system is consistent if and only if $h = -\frac{8}{3}$.

Brief summary of what we learned so far

- Each linear system corresponds to an augmented matrix.
- Using Gaussian elimination (i.e. row reduction to echelon form) on the augmented matrix of a linear system, we can
  - read off, whether the system has no, one, or infinitely many solutions;
  - find all solutions by back-substitution.
- We can continue row reduction to the reduced echelon form.
  - Solutions to the linear system can now be just read off.
  - This form is unique!

Note. Besides for solving linear systems, Gaussian elimination has other important uses, such as computing determinants or inverses of matrices.

A recipe to solve linear systems (Gauss–Jordan elimination)

(1) Write the augmented matrix of the system.

(2) Row reduce to obtain an equivalent augmented matrix in echelon form.
Decide whether the system is consistent. If not, stop; otherwise go to the next step.

(3) Continue row reduction to obtain the reduced echelon form.

(4) Express this final matrix as a system of equations.

(5) Declare the free variables and state the solution in terms of these.

Questions to check our understanding

- On an exam, you are asked to find all solutions to a system of linear equations. You find exactly two solutions. Should you be worried?

  Yes, because if there is more than one solution, there have to be infinitely many solutions. Can you see how, given two solutions, one can construct infinitely many more?

- True or false?
  - There is no more than one pivot in any row.

    True, because a pivot is the first nonzero entry in a row.
  - There is no more than one pivot in any column.

    True, because in echelon form (that's where pivots are defined) the entries below a pivot have to zero.
  - There cannot be more free variables than pivot variables.

    False, consider, for instance, the augmented matrix $\left[\begin{array}{ccc|c} 1 & 7 & 5 & 3 \end{array}\right]$.

Armin Straub
astraub@illinois.edu

The geometry of linear equations

Adding and scaling vectors

Example 4. We have already encountered **matrices** such as

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

Each column is what we call a **(column) vector**.

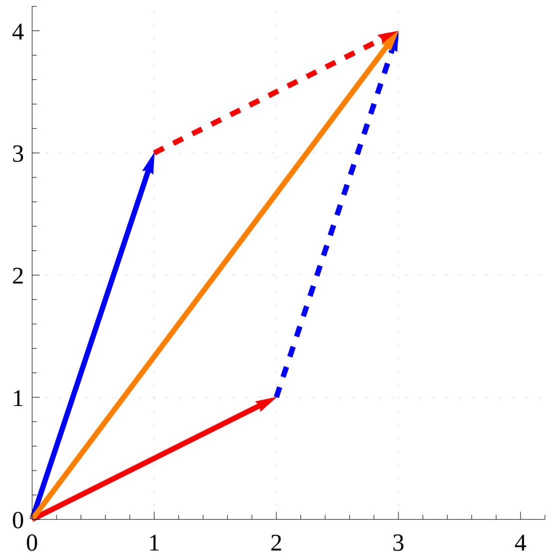
In this example, each column vector has 3 entries and so lies in $\mathbb{R}^3$.

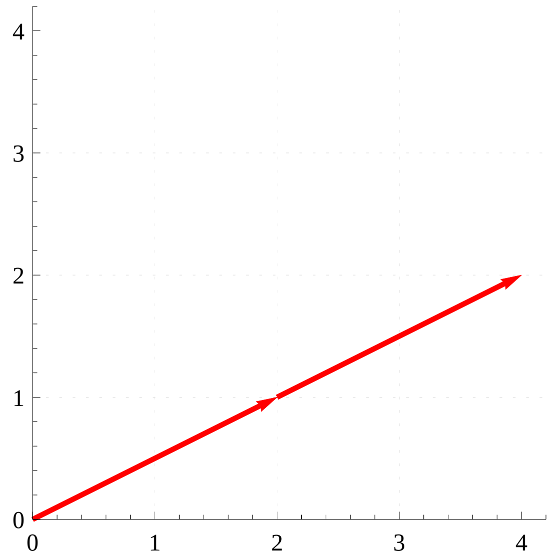
Example 5. A fundamental property of vectors is that vectors of the same kind can be **added** and **scaled**.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 5 \end{bmatrix}, \qquad 7 \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7x_1 \\ 7x_2 \\ 7x_3 \end{bmatrix}.$$

Example 6. (Geometric description of $\mathbb{R}^2$) A vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ represents the point (x_1, x_2) in the plane.

Given $\boldsymbol{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, graph $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{x} + \boldsymbol{y}$, $2\boldsymbol{y}$.





Armin Straub
astraub@illinois.edu

Adding and scaling vectors, the most general thing we can do is:

Definition 7. Given vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ in $\mathbb{R}^n$ and scalars $c_1, c_2, \ldots, c_m$, the vector

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_m\boldsymbol{v}_m$$

is a **linear combination** of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$.

The scalars $c_1, \ldots, c_m$ are the **coefficients** or **weights**.

Example 8. Linear combinations of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ include:

- $3\boldsymbol{v}_1 - \boldsymbol{v}_2 + 7\boldsymbol{v}_3$,
- $\boldsymbol{v}_2 + \boldsymbol{v}_3$,
- $\frac{1}{3}\boldsymbol{v}_2$,
- $\boldsymbol{0}$.

Example 9. Express $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Solution. We have to find c_1 and c_2 such that

$$c_1\begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

This is the same as:

$$\begin{array}{rrcl} 2c_1 & -c_2 & = & 1 \\ c_1 & +c_2 & = & 5 \end{array}$$

Solving, we find $c_1 = 2$ and $c_2 = 3$.

Indeed,

$$2\begin{bmatrix} 2 \\ 1 \end{bmatrix} + 3\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

Note that the augmented matrix of the linear system is

$$\left[\begin{array}{rr|r} 2 & -1 & 1 \\ 1 & 1 & 5 \end{array}\right],$$

and that this example provides a new way of think about this system.

The row and column picture

Example 10. We can think of the linear system

$$\begin{aligned} 2x - y &= 1 \\ x + y &= 5 \end{aligned}$$

in two different geometric ways.

Row picture.

Each equation defines a line in $\mathbb{R}^2$.

Which points lie on the intersection of these lines?

Column picture.

The system can be written as $x\begin{bmatrix} 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$.

Which linear combinations of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ produce $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$?

This example has the unique solution $x = 2$, $y = 3$.

- $(2, 3)$ is the (only) intersection of the two lines $2x - y = 1$ and $x + y = 5$.
- $2\begin{bmatrix} 2 \\ 1 \end{bmatrix} + 3\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is the (only) linear combination producing $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$.

Pre-lecture: the shocking state of our ignorance

Q: *How fast can we solve N linear equations in N unknowns?*

Estimated cost of Gaussian elimination:

$$\begin{bmatrix} \blacksquare & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & & & \vdots \\ 0 & * & * & \cdots & * \end{bmatrix}$$

- to create the zeros below the pivot:
  $\Longrightarrow$ on the order of N^2 operations
- if there is N pivots:
  $\Longrightarrow$ on the order of $N \cdot N^2 = N^3$ op's

- A more careful count places the cost at $\sim\frac{1}{3}N^3$ op's.
- For large N, it is only the N^3 that matters.

  It says that if $N \to 10N$ then we have to work 1000 times as hard.

That's not optimal! We can do better than Gaussian elimination:

- Strassen algorithm (1969): $N^{\log_2 7} = N^{2.807}$
- Coppersmith–Winograd algorithm (1990): $N^{2.375}$
- ... Stothers–Williams–Le Gall (2014): $N^{2.373}$

Is N^2 possible? We have no idea! (better is impossible; why?)

Good news for applications: (will see an example soon)

- Matrices typically have lots of structure and zeros

  which makes solving so much faster.

Organizational

- Help sessions in 441 AH: MW 4-6pm, TR 5-7pm

Review

- A system such as

$$\begin{aligned} 2x - y &= 1 \\ x + y &= 5 \end{aligned}$$

  can be written in **vector** form as

$$x\begin{bmatrix} 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

- The left-hand side is a **linear combination** of the vectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

The row and column picture

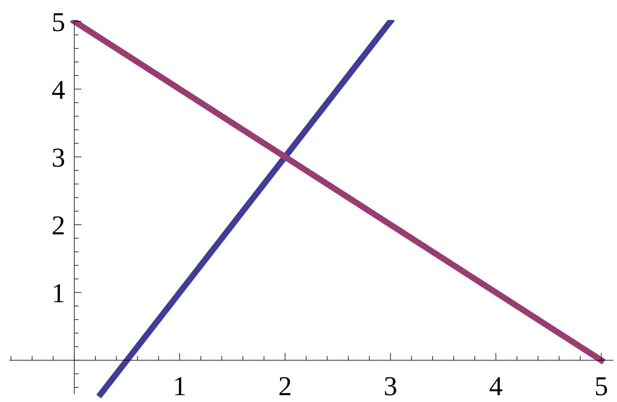
Example 1. We can think of the linear system

$$\begin{aligned} 2x - y &= 1 \\ x + y &= 5 \end{aligned}$$

in two different geometric ways. Here, there is a unique solution: $x = 2$, $y = 3$.

Row picture.

- Each equation defines a line in $\mathbb{R}^2$.
- Which points lie on the intersection of these lines?
- $(2, 3)$ is the (only) intersection of the two lines $2x - y = 1$ and $x + y = 5$.



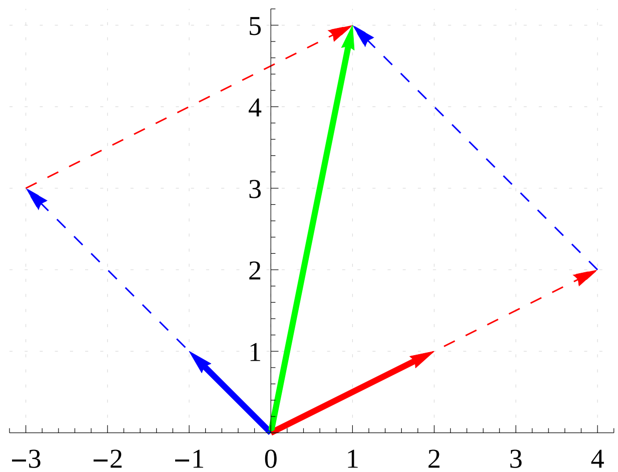
Armin Straub
astraub@illinois.edu

Column picture.

- The system can be written as

  $$x\begin{bmatrix} 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

- Which linear combinations of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ produce $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$?
- $(2,\ 3)$ are the coefficients of the (only) such linear combination.



Example 2. Consider the vectors

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 4 \\ 2 \\ 14 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} 3 \\ 6 \\ 10 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} -1 \\ 8 \\ -5 \end{bmatrix}.$$

Determine if $\boldsymbol{b}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$.

Solution. Vector $\boldsymbol{b}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ if we can find weights x_1, x_2, x_3 such that:

$$x_1\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + x_2\begin{bmatrix} 4 \\ 2 \\ 14 \end{bmatrix} + x_3\begin{bmatrix} 3 \\ 6 \\ 10 \end{bmatrix} = \begin{bmatrix} -1 \\ 8 \\ -5 \end{bmatrix}$$

This vector equation corresponds to the linear system:

$$\begin{array}{rrrcr} x_1 & +4x_2 & +3x_3 & = & -1 \\ & +2x_2 & +6x_3 & = & 8 \\ 3x_1 & +14x_2 & +10x_3 & = & -5 \end{array}$$

Corresponding augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 4 & 3 & -1 \\ 0 & 2 & 6 & 8 \\ 3 & 14 & 10 & -5 \end{array}\right]$$

Note that we are looking for a linear combination of the first three columns which produces the last column.

Such a combination exists $\Longleftrightarrow$ the system is consistent.

Row reduction to echelon form:

$$\begin{bmatrix} 1 & 4 & 3 & -1 \\ 0 & 2 & 6 & 8 \\ 3 & 14 & 10 & -5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 4 & 3 & -1 \\ 0 & 2 & 6 & 8 \\ 0 & 2 & 1 & -2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 4 & 3 & -1 \\ 0 & 2 & 6 & 8 \\ 0 & 0 & -5 & -10 \end{bmatrix}$$

Since this system is consistent, $\boldsymbol{b}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$.

[It is consistent, because there is no row of the form $[\,0 \ \ 0 \ \ 0 \ \ b\,]$ with $b \neq 0$.]

Example 3. In the previous example, express $\boldsymbol{b}$ as a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$.

Solution. The reduced echelon form is:

$$\begin{bmatrix} 1 & 4 & 3 & -1 \\ 0 & 2 & 6 & 8 \\ 0 & 0 & -5 & -10 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 4 & 3 & -1 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 4 & 0 & -7 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

We read off the solution $x_1 = 1$, $x_2 = -2$, $x_3 = 2$, which yields

$$\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} - 2\begin{bmatrix} 4 \\ 2 \\ 14 \end{bmatrix} + 2\begin{bmatrix} 3 \\ 6 \\ 10 \end{bmatrix} = \begin{bmatrix} -1 \\ 8 \\ -5 \end{bmatrix}.$$

Summary

A vector equation

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_m\boldsymbol{a}_m = \boldsymbol{b}$$

has the same solution set as the linear system with augmented matrix

$$\begin{bmatrix} | & | & & | & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_m & \boldsymbol{b} \\ | & | & & | & | \end{bmatrix}.$$

In particular, $\boldsymbol{b}$ can be generated by a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ if and only if this linear system is consistent.

The span of a set of vectors

Definition 4. The **span** of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ is the set of all their linear combinations. We denote it by $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$.

In other words, $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is the set of all vectors of the form

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_m\boldsymbol{v}_m,$$

where $c_1, c_2, \ldots, c_m$ are scalars.

Example 5.

(a) Describe $\operatorname{span}\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$ geometrically.

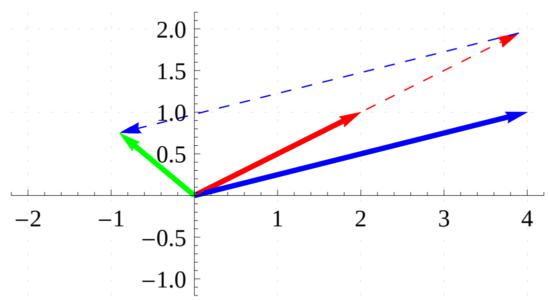
The span consists of all vectors of the form $\alpha \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

As points in $\mathbb{R}^2$, this is a line.

(b) Describe $\operatorname{span}\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix}\right\}$ geometrically.

The span is all of $\mathbb{R}^2$, a plane.

That's because any vector in $\mathbb{R}^2$ can be written as $x_1\begin{bmatrix} 2 \\ 1 \end{bmatrix} + x_2\begin{bmatrix} 4 \\ 1 \end{bmatrix}$.



Let's show this without relying on our geometric intuition: let $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ any vector.

$$\left[\begin{array}{cc|c} 2 & 4 & b_1 \\ 1 & 1 & b_2 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 2 & 4 & b_1 \\ 0 & -1 & b_2 - \frac{1}{2}b_1 \end{array}\right] \quad \text{is consistent}$$

Hence, $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$.

(c) Describe $\operatorname{span}\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \end{bmatrix}\right\}$ geometrically.

Note that $\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Hence, the span is as in (a).

Again, we can also see this after row reduction: let $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ any vector.

$$\left[\begin{array}{cc|c} 2 & 4 & b_1 \\ 1 & 2 & b_2 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 2 & 4 & b_1 \\ 0 & 0 & b_2 - \frac{1}{2}b_1 \end{array}\right] \quad \text{is not consistent for all } \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ is in the span of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$ only if $b_2 - \frac{1}{2}b_1 = 0$ (i.e. $b_2 = \frac{1}{2}b_1$).

So the span consists of vectors $\begin{bmatrix} b_1 \\ \frac{1}{2}b_1 \end{bmatrix} = b_1\begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$.

A single (nonzero) vector always spans a line, two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ usually span a plane but it could also be just a line (if $\boldsymbol{v}_2 = \alpha \boldsymbol{v}_1$).

We will come back to this when we discuss dimension and linear independence.

Example 6. Is $\operatorname{span}\left\{ \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} \right\}$ a line or a plane?

Solution. The span is a plane unless, for some α,

$$\begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} = \alpha \cdot \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}.$$

Looking at the first entry, $\alpha = 2$, but that does not work for the third entry. Hence, there is no such α. The span is a plane.

Example 7. Consider

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 8 \\ 3 \\ 17 \end{bmatrix}.$$

Is $\boldsymbol{b}$ in the plane spanned by the columns of A?

Solution. $\boldsymbol{b}$ in the plane spanned by the columns of A if and only if

$$\left[\begin{array}{cc|c} 1 & 2 & 8 \\ 3 & 1 & 3 \\ 0 & 5 & 17 \end{array}\right]$$

is consistent.

To find out, we row reduce to an echelon form:

$$\left[\begin{array}{cc|c} 1 & 2 & 8 \\ 3 & 1 & 3 \\ 0 & 5 & 17 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 2 & 8 \\ 0 & -5 & -21 \\ 0 & 5 & 17 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 2 & 8 \\ 0 & -5 & -21 \\ 0 & 0 & -4 \end{array}\right]$$

From the last row, we see that the system is inconsistent. Hence, $\boldsymbol{b}$ is not in the plane spanned by the columns of A.

Armin Straub
astraub@illinois.edu

Conclusion and summary

- The **span** of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ is the set of all their **linear combinations**.
- Some vector $\boldsymbol{b}$ is in $\operatorname{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m\}$ if and only if there is a solution to the linear system with augmented matrix

$$\left[\begin{array}{ccccc} | & | & & | & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_m & \boldsymbol{b} \\ | & | & & | & | \end{array}\right].$$

  - Each solution corresponds to the weights in a linear combination of the $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ which gives $\boldsymbol{b}$.
  - This gives a second geometric way to think of linear systems!

Pre-lecture: the goal for today

We wish to write linear systems simply as $A\boldsymbol{x} = \boldsymbol{b}$.

For instance:

$$\begin{array}{rrcl} 2x_1 & +3x_2 & = & b_1 \\ 3x_1 & +x_2 & = & b_2 \end{array} \iff \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Why?

- It's concise.
- The compactness also sparks associations and ideas!
  - For instance, can we solve by *dividing* by A? $\boldsymbol{x} = A^{-1}\boldsymbol{b}$?
  - If $A\boldsymbol{x} = \boldsymbol{b}$ and $A\boldsymbol{y} = 0$, then $A(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{b}$.
- Leads to matrix calculus and deeper understanding.
  - multiplying, inverting, or factoring matrices

Matrix operations

Basic notation

We will use the following notations for an $m \times n$ matrix A (m rows, n columns).

- In terms of the columns of A:

$$A = [\, \boldsymbol{a}_1 \;\; \boldsymbol{a}_2 \;\; \cdots \;\; \boldsymbol{a}_n \,] = \begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix}$$

- In terms of the entries of A:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}, \qquad a_{i,j} = \text{entry in } i\text{-th row, } j\text{-th column}$$

Matrices, just like vectors, are added and scaled componentwise.

Example 1.

(a) $\begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 8 & 3 \end{bmatrix}$

(b) $7\cdot\begin{bmatrix} 2 & 3\\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 21\\ 21 & 7 \end{bmatrix}$

Matrix times vector

Recall that $(x_1, x_2, \ldots, x_n)$ solves the linear system with augmented matrix

$$[\, A \;\; \boldsymbol{b} \,] = \begin{bmatrix} | & | & & | & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n & \boldsymbol{b} \\ | & | & & | & | \end{bmatrix}$$

if and only if

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_n\boldsymbol{a}_n = \boldsymbol{b}.$$

It is therefore natural to define the **product of matrix times vector** as

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_n\boldsymbol{a}_n, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The system of linear equations with augmented matrix $[\, A \;\; \boldsymbol{b} \,]$ can be written in **matrix form** compactly as $A\boldsymbol{x} = \boldsymbol{b}$.

The product of a matrix A with a vector $\boldsymbol{x}$ is a linear combination of the columns of A with weights given by the entries of $\boldsymbol{x}$.

Example 2.

(a) $\begin{bmatrix} 1 & 0\\ 5 & 2 \end{bmatrix}\cdot\begin{bmatrix} 2\\ 1 \end{bmatrix} = 2\begin{bmatrix} 1\\ 5 \end{bmatrix} + 1\begin{bmatrix} 0\\ 2 \end{bmatrix} = \begin{bmatrix} 2\\ 12 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 3\\ 3 & 1 \end{bmatrix}\cdot\begin{bmatrix} 0\\ 1 \end{bmatrix} = \begin{bmatrix} 3\\ 1 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 3\\ 3 & 1 \end{bmatrix}\cdot\begin{bmatrix} x_1\\ x_2 \end{bmatrix} = x_1\begin{bmatrix} 2\\ 3 \end{bmatrix} + x_2\begin{bmatrix} 3\\ 1 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2\\ 3x_1 + x_2 \end{bmatrix}$

This illustrates that linear systems can be simply expressed as $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{array}{rrcl} 2x_1 & +3x_2 & = & b_1 \\ 3x_1 & +x_2 & = & b_2 \end{array} \iff \begin{bmatrix} 2 & 3\\ 3 & 1 \end{bmatrix}\cdot\begin{bmatrix} x_1\\ x_2 \end{bmatrix} = \begin{bmatrix} b_1\\ b_2 \end{bmatrix}$$

(d) $\begin{bmatrix} 2 & 3\\ 3 & 1\\ 1 & -1 \end{bmatrix}\cdot\begin{bmatrix} 1\\ 1 \end{bmatrix} = \begin{bmatrix} 5\\ 4\\ 0 \end{bmatrix}$

Example 3. Suppose A is $m \times n$ and $\boldsymbol{x}$ is in $\mathbb{R}^p$. Under which condition does $A\boldsymbol{x}$ make sense?

We need $n = p$. (Go through the definition of $A\boldsymbol{x}$ to make sure you see why!)

Matrix times matrix

If B has just one column $\boldsymbol{b}$, i.e. $B = [\ \boldsymbol{b}\ ]$, then $AB = [\ A\boldsymbol{b}\ ]$.

In general, the **product of matrix times matrix** is given by

$$AB = [\ A\boldsymbol{b}_1 \ \ A\boldsymbol{b}_2 \ \ \cdots \ \ A\boldsymbol{b}_p\ ], \qquad B = [\ \boldsymbol{b}_1 \ \ \boldsymbol{b}_2 \ \ \cdots \ \ \boldsymbol{b}_p\ ].$$

Example 4.

(a) $\begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 & -3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ 12 & -11 \end{bmatrix}$

because $\begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 5 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 12 \end{bmatrix}$

and $\begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 2 \end{bmatrix} = -3\begin{bmatrix} 1 \\ 5 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ -11 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & 0 \\ 5 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -3 & 1 \\ 12 & -11 & 5 \end{bmatrix}$

Each column of AB is a linear combination of the columns of A with weights given by the corresponding column of B.

Remark 5. The definition of the matrix product is inevitable from the multiplication of matrix times vector and the fact that we want AB to be defined such that $(AB)\boldsymbol{x} = A(B\boldsymbol{x})$.

$$\begin{aligned} A(B\boldsymbol{x}) &= A(x_1\boldsymbol{b}_1 + x_2\boldsymbol{b}_2 + \cdots) \\ &= x_1 A\boldsymbol{b}_1 + x_2 A\boldsymbol{b}_2 + \cdots \\ &= (AB)\boldsymbol{x} \quad \text{if the columns of } AB \text{ are } A\boldsymbol{b}_1, A\boldsymbol{b}_2, \ldots \end{aligned}$$

Example 6. Suppose A is $m \times n$ and B is $p \times q$.

(a) Under which condition does AB make sense?

We need $n = p$. (Go through the boxed characterization of AB to make sure you see why!)

(b) What are the dimensions of AB in that case?

AB is a $m \times q$ matrix.

Basic properties

Example 7.

(a) $\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$

This is the 2×2 **identity matrix**.

Theorem 8. Let A, B, C be matrices of appropriate size. Then:

- $A(BC) = (AB)C$ associative
- $A(B + C) = AB + AC$ left-distributive
- $(A + B)C = AC + BC$ right-distributive

Example 9. However, matrix multiplication is not commutative!

(a) $\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 3 & 1 \end{bmatrix}$

Example 10. Also, a product can be zero even though none of the factors is:

$$\begin{bmatrix} 2 & 0 \\ 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Transpose of a matrix

Definition 11. The **transpose** A^T of a matrix A is the matrix whose columns are formed from the corresponding rows of A. rows $\leftrightarrow$ columns

Example 12.

(a) $\begin{bmatrix} 2 & 0 \\ 3 & 1 \\ -1 & 4 \end{bmatrix}^T = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 1 & 4 \end{bmatrix}$

(b) $[\, x_1 \;\; x_2 \;\; x_3 \,]^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}^T = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$

A matrix A is called **symmetric** if $A = A^T$.

Practice problems

- True or false?
  - AB has as many columns as B.
  - AB has as many rows as B.

The following practice problem illustrates the rule $(AB)^T = B^T A^T$.

Example 13. Consider the matrices

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

Compute:

(a) $AB = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} =$

(b) $(AB)^T = \begin{bmatrix} & & \\ & & \end{bmatrix}$

(c) $B^T A^T = \begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{bmatrix} =$

(d) $A^T B^T = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix} =$

What's that fishy smell?

Review: matrix multiplication

- $A\boldsymbol{x}$ is a linear combination of the columns of A with weights given by the entries of $\boldsymbol{x}$.

$$\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 2 \\ 3 \end{bmatrix} + 1\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

- $A\boldsymbol{x} = \boldsymbol{b}$ is the **matrix form** of the linear system with augmented matrix $\left[\begin{array}{cc} A & \boldsymbol{b} \end{array}\right]$.

$$\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \iff \begin{array}{rrcl} 2x_1 & +3x_2 & = & b_1 \\ 3x_1 & +x_2 & = & b_2 \end{array}$$

- Each column of AB is a linear combination of the columns of A with weights given by the corresponding column of B.

$$\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 7 & 3 \end{bmatrix}$$

- Matrix multiplication is not commutative: usually, $AB \neq BA$.

A comment on lecture notes

My personal suggestion:

- before lecture: have a quick look (15min or so) at the pre-lecture notes to see where things are going
- during lecture: take a minimal amount of notes (everything on the screens will be in the post-lecture notes) and focus on the ideas
- after lecture: go through the pre-lecture notes again and fill in all the blanks by yourself
- then compare with the post-lecture notes

Since I am writing the pre-lecture notes a week ahead of time, there is usually some minor differences to the post-lecture notes.

Armin Straub
astraub@illinois.edu

Transpose of a matrix

Definition 1. The **transpose** A^T of a matrix A is the matrix whose columns are formed from the corresponding rows of A. rows $\leftrightarrow$ columns

Example 2.

(a) $\begin{bmatrix} 2 & 0 \\ 3 & 1 \\ -1 & 4 \end{bmatrix}^T = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 1 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}^T = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$

A matrix A is called **symmetric** if $A = A^T$.

Theorem 3. Let A, B be matrices of appropriate size. Then:

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(AB)^T = B^T A^T$ (illustrated by last practice problems)

Example 4. Deduce that $(ABC)^T = C^T B^T A^T$.

Solution. $(ABC)^T = ((AB)\,C)^T = C^T (AB)^T = C^T B^T A^T$

Back to matrix multiplication

Review. Each column of AB is a linear combination of the columns of A with weights given by the corresponding column of B.

Two more ways to look at matrix multiplication

Example 5. What is the entry $(AB)_{i,j}$ at row i and column j?

The j-th column of AB is the vector $A \cdot (\text{col } j \text{ of } B)$.

Entry i of that is $(\text{row } i \text{ of } A) \cdot (\text{col } j \text{ of } B)$. In other words:

$$(AB)_{i,j} = (\text{row } i \text{ of } A) \cdot (\text{col } j \text{ of } B)$$

Use this **row-column rule** to compute:

$$\begin{bmatrix} 2 & 3 & 6 \\ -1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & -3 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 16 & -3 \\ 0 & 3 \end{bmatrix}$$

$$\begin{array}{cc} & \begin{bmatrix} 2 & -3 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} \\ \begin{bmatrix} 2 & 3 & 6 \\ -1 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 16 & -3 \\ 0 & 3 \end{bmatrix} \end{array}$$

[Can you see the rule $(AB)^T = B^T A^T$ from here?]

Observe the symmetry between rows and columns in this rule!

It follows that the interpretation

> "Each column of AB is a linear combination of the columns of A with weights given by the corresponding column of B."

has the counterpart

> "Each row of AB is a linear combination of the rows of B with weights given by the corresponding row of A."

Example 6.

(a) $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} -1 & -2 & -3 \\ 7 & 8 & 9 \end{bmatrix}$

Armin Straub
astraub@illinois.edu

LU decomposition

Elementary matrices

Example 7.

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ 2d & 2e & 2f \\ g & h & i \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 3a+g & 3b+h & 3c+i \end{bmatrix}$

Definition 8. An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix.

The result of an elementary row operation on A is EA

where E is an elementary matrix (namely, the one obtained by performing the same row operation on the appropriate identity matrix).

Example 9.

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$

We write $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$, but more on inverses soon.

Elementary matrices are **invertible** because elementary row operations are reversible.

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & \\ & \frac{1}{2} & \\ & & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

Armin Straub
astraub@illinois.edu

Practice problems

Example 10. Choose either column or row interpretation to "see" the result of the following products.

(a) $\begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} =$

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix} =$

Review

Example 1. Elementary matrices in action:

(a) $$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} g & h & i \\ d & e & f \\ a & b & c \end{bmatrix}$$

(b) $$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 7g & 7h & 7i \end{bmatrix}$$

(c) $$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 3a+g & 3b+h & 3c+i \end{bmatrix}$$

(d) $$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a+3c & b & c \\ d+3f & e & f \\ g+3i & h & i \end{bmatrix}$$

(e) $$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

LU decomposition, continued

Gaussian elimination revisited

Example 2. Keeping track of the elementary matrices during Gaussian elimination on A:

$$A = \begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix} \quad R2 \to R2 - 2R1$$

$$EA = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 4 & -6 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & -8 \end{bmatrix}$$

Note that:

$$A = E^{-1} \begin{bmatrix} 2 & 1 \\ 0 & -8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & -8 \end{bmatrix}$$

We factored A as the product of a lower and upper triangular matrix!

We say that A has **triangular factorization**.

$A = LU$ is known as the **LU decomposition** of A.

L is lower triangular, U is upper triangular.

Definition 3.

lower triangular

$$\begin{bmatrix} * & 0 & 0 & 0 & 0 \\ \vdots & \ddots & 0 & 0 & 0 \\ * & \cdots & * & 0 & 0 \\ * & * & \cdots & * & 0 \\ * & * & * & \cdots & * \end{bmatrix}$$

upper triangular

$$\begin{bmatrix} * & * & * & \cdots & * \\ & * & * & \cdots & * \\ & & * & \cdots & * \\ & & & \ddots & \vdots \\ & & & & * \end{bmatrix}$$

missing entries are 0

Example 4. Factor $A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$ as $A = LU$.

Solution. We begin with $R2 \to R2 - 2R1$ followed by $R3 \to R3 + R1$:

$$E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$$

$$E_2(E_1 A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix}$$

$$E_3 E_2 E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = U$$

The factor L is given by: note that $E_3 E_2 E_1 A = U \implies A = E_1^{-1} E_2^{-1} E_3^{-1} U$

$$\begin{aligned} L &= E_1^{-1} E_2^{-1} E_3^{-1} \\ &= \begin{bmatrix} 1 & & \\ 2 & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & & \\ 2 & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & -1 & 1 \end{bmatrix} \end{aligned}$$

In conclusion, we found the following LU decomposition of A:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ & -8 & -2 \\ & & 1 \end{bmatrix}$$

Note: The extra steps to compute L were unnecessary! The entries in L are precisely the negatives of the ones in the elementary matrices during elimination. Can you see it?

Once we have $A = LU$, it is simple to solve $A\boldsymbol{x} = \boldsymbol{b}$.

$$\begin{aligned} & A\boldsymbol{x} = \boldsymbol{b} \\ \Longleftrightarrow\ & L(U\boldsymbol{x}) = \boldsymbol{b} \\ \Longleftrightarrow\ & L\boldsymbol{c} = \boldsymbol{b} \quad \text{and} \quad U\boldsymbol{x} = \boldsymbol{c} \end{aligned}$$

Both of the final systems are triangular and hence easily solved:

- $L\boldsymbol{c} = \boldsymbol{b}$ by forward substitution to find $\boldsymbol{c}$, and then
- $U\boldsymbol{x} = \boldsymbol{c}$ by backward substitution to find $\boldsymbol{x}$.

Important practical point: can be quickly repeated for many different $\boldsymbol{b}$.

Example 5. Solve $\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 4 \\ 10 \\ -3 \end{bmatrix}$.

Solution. We already found the LU decomposition $A = LU$:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ & -8 & -2 \\ & & 1 \end{bmatrix}$$

Forward substitution to solve $L\boldsymbol{c} = \boldsymbol{b}$ for $\boldsymbol{c}$:

$$\begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & -1 & 1 \end{bmatrix} \boldsymbol{c} = \begin{bmatrix} 4 \\ 10 \\ -3 \end{bmatrix} \quad \Longrightarrow \quad \boldsymbol{c} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$$

Backward substitution to solve $U\boldsymbol{x} = \boldsymbol{c}$ for $\boldsymbol{x}$:

$$\begin{bmatrix} 2 & 1 & 1 \\ & -8 & -2 \\ & & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} \quad \Longrightarrow \quad \boldsymbol{x} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$$

It's always a good idea to do a quick check:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 4 \\ 10 \\ -3 \end{bmatrix}$$

Triangular factors for any matrix

Can we factor any matrix A as $A = LU$?

Yes, almost! Think about the process of Gaussian elimination.

- In each step, we use a pivot to produce zeros below it.
  The corresponding elementary matrices are lower diagonal!
- The only other thing we might have to do, is a row exchange.
  Namely, if we run into a zero in the position of the pivot.
- All of these row exchanges can be done at the beginning!

Definition 6. A **permutation matrix** is one that is obtained by performing row exchanges on an identity matrix.

Example 7. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ is a permutation matrix.

EA is the matrix obtained from A by permuting the last two rows.

Theorem 8. For any matrix A there is a permutation matrix P such that $PA = LU$.

In other words, it might not be possible to write A as $A = LU$, but we only need to permute the rows of A and the resulting matrix PA now has an LU decomposition: $PA = LU$.

Armin Straub
astraub@illinois.edu

Practice problems

- Is $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ upper triangular? Lower triangular?
- Is $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ upper triangular? Lower triangular?
- True or false?
  - A permutation matrix is one that is obtained by performing column exchanges on an identity matrix.
- Why do we care about LU decomposition if we already have Gaussian elimination?

Example 9. Solve $\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$ using the factorization we already have.

Example 10. The matrix

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

cannot be written as $A = LU$ (so it doesn't have a LU decomposition). But there is a permutation matrix P such that PA has a LU decomposition.

Namely, let $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Then $PA = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

PA can now be factored as $PA = LU$. Do it!!

(By the way, $P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ would work as well.)

Review

- Elementary matrices performing row operations:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d-2a & e-2b & f-2c \\ g & h & i \end{bmatrix}$$

- Gaussian elimination on A gives an LU decomposition $A = LU$:

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 1 \\ & -8 & -2 \\ & & 1 \end{bmatrix}$$

  U is the echelon form, and L records the inverse row operations we did.

- LU decomposition allows us to solve $A\boldsymbol{x} = \boldsymbol{b}$ for many $\boldsymbol{b}$.
- $\begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- Already not so clear: $\begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & b & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ ab & -b & 1 \end{bmatrix}$

  Goal for today: invert these and any other matrices (if possible)

The inverse of a matrix

Example 1. The inverse of a real number a is denoted as a^{-1}. For instance, $7^{-1} = \frac{1}{7}$ and

$$7 \cdot 7^{-1} = 7^{-1} \cdot 7 = 1.$$

In the context of $n \times n$ matrix multiplication, the role of 1 is taken by the $n \times n$ identity matrix

$$I_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

Definition 2. An $n \times n$ matrix A is **invertible** if there is a matrix B such that

$$AB = BA = I_n.$$

In that case, B is the **inverse** of A and we write $A^{-1} = B$.

Example 3. We already saw that elementary matrices are **invertible**.

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

Note.

- The inverse of a matrix is unique. Why? So A^{-1} is well-defined.

  Assume B and C are both inverses of A. Then:

$$C = CI_n = CAB = I_nB = B$$

- Do not write $\frac{A}{B}$. Why?

  Because it is unclear whether it should mean AB^{-1} or $B^{-1}A$.

- If $AB = I$, then $BA = I$ (and so $A^{-1} = B$). Not easy to show at this stage.

Example 4. The matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not invertible. Why?

Solution.

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$$

Example 5. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$ provided that $ad - bc \neq 0$.

Let's check that:

$$\frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} ad - bc & 0 \\ 0 & -cb + ad \end{bmatrix} = I_2$$

Note.

- A 1×1 matrix $\begin{bmatrix} a \end{bmatrix}$ is invertible $\Longleftrightarrow$ $a \neq 0$.
- A 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible $\Longleftrightarrow$ $ad - bc \neq 0$.

We will encounter the quantities on the right again when we discuss determinants.

Solving systems using matrix inverse

Theorem 6. Let A be invertible. Then the system $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

Proof. Multiply both sides of $A\boldsymbol{x} = \boldsymbol{b}$ with A^{-1} (from the left!). □

Example 7. Solve $\begin{array}{rrcl} -7x_1 & +3x_2 & = & 2 \\ 5x_1 & -2x_2 & = & 1 \end{array}$ using matrix inversion.

Solution. In matrix form $A\boldsymbol{x} = \boldsymbol{b}$, this system is

$$\begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Computing the inverse:

$$\begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix}^{-1} = \frac{1}{-1}\begin{bmatrix} -2 & -3 \\ -5 & -7 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}$$

Recall that $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Hence, the solution is:

$$\boldsymbol{x} = A^{-1}\boldsymbol{b} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 17 \end{bmatrix}$$

Recipe for computing the inverse

To solve $A\boldsymbol{x} = \boldsymbol{b}$, we do row reduction on $\left[\begin{array}{c|c} A & \boldsymbol{b} \end{array}\right]$.

To solve $AX = I$, we do row reduction on $\left[\begin{array}{c|c} A & I \end{array}\right]$.

To compute A^{-1}: **Gauss–Jordan method**

- Form the augmented matrix $\left[\begin{array}{c|c} A & I \end{array}\right]$.
- Compute the reduced echelon form. (i.e. Gauss–Jordan elimination)
- If A is invertible, the result is of the form $\left[\begin{array}{c|c} I & A^{-1} \end{array}\right]$.

Example 8. Find the inverse of $A = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, if it exists.

Solution. By row reduction:

$$\left[\begin{array}{cc} A & I \end{array}\right] \rightsquigarrow \left[\begin{array}{cc} I & A^{-1} \end{array}\right]$$

$$\left[\begin{array}{cccccc} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{array}\right]$$

Hence, $A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \\ \frac{3}{2} & 1 & 0 \end{bmatrix}$.

Example 9. Let's do the previous example step by step.

$$\left[\begin{array}{cc} A & I \end{array}\right] \qquad \rightsquigarrow \qquad \left[\begin{array}{cc} I & A^{-1} \end{array}\right]$$

$$\left[\begin{array}{cccccc} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \overset{R2 \to R2 + \frac{3}{2}R1}{\rightsquigarrow} \left[\begin{array}{cccccc} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\overset{\substack{R1 \to \frac{1}{2}R1 \\ R2 \leftrightarrow R3}}{\rightsquigarrow} \left[\begin{array}{cccccc} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{array}\right]$$

Note. Here is another way to see why this algorithm works:

- Each row reduction corresponds to multiplying with an elementary matrix E:

$$[\ A\ |\ I\ ] \rightsquigarrow [\ E_1A\ |\ E_1I\ ] \rightsquigarrow [\ E_2E_1A\ |\ E_2E_1\ ] \rightsquigarrow \ldots$$

- So at each step:

$$[\ A\ |\ I\ ] \rightsquigarrow [\ FA\ |\ F\ ] \qquad \text{with } F = E_r \cdots E_2E_1$$

- If we manage to reduce $[\ A\ |\ I\ ]$ to $[\ I\ |\ F\ ]$, this means

$$FA = I \qquad \text{and hence } A^{-1} = F.$$

Some properties of matrix inverses

Theorem 10. Suppose A and B are invertible. Then:

- A^{-1} is invertible and $(A^{-1})^{-1} = A$.

  Why? Because $AA^{-1} = I$

- A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

  Why? Because $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$ (Recall that $(AB)^T = B^TA^T$.)

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

  Why? Because $(B^{-1}A^{-1})(AB) = B^{-1}IB = B^{-1}B = I$

Review

- The **inverse** A^{-1} of a matrix A is, if it exists, characterized by

$$AA^{-1} = A^{-1}A = I_n.$$

- $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
- If A is invertible, then the system $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.
- Gauss–Jordan method to compute A^{-1}:
  - bring to RREF $[\ A \mid I\ ] \rightsquigarrow [\ I \mid A^{-1}\ ]$
- $(A^{-1})^{-1} = A$
- $(A^T)^{-1} = (A^{-1})^T$
- $(AB)^{-1} = B^{-1}A^{-1}$

  Why? Because $(B^{-1}A^{-1})(AB) = B^{-1}IB = B^{-1}B = I$

Further properties of matrix inverses

Theorem 1. Let A be an $n \times n$ matrix. Then the following statements are equivalent: (i.e., for a given A, they are either all true or all false)

(a) A is invertible.

(b) A is row equivalent to I_n.

(c) A has n pivots. (Easy to check!)

(d) For every $\boldsymbol{b}$, the system $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution.

Namely, $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

(e) There is a matrix B such that $AB = I_n$. (A has a "right inverse".)

(f) There is a matrix C such that $CA = I_n$. (A has a "left inverse".)

Note. Matrices that are not invertible are often called **singular**.

The book uses **singular** for $n \times n$ matrices that do not have n pivots. As we just saw, it doesn't make a difference.

Example 2. We now see at once that $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not invertible.

Why? Because it has only one pivot.

Armin Straub
astraub@illinois.edu

Application: finite differences

Let us apply linear algebra to the **boundary value problem** (BVP)

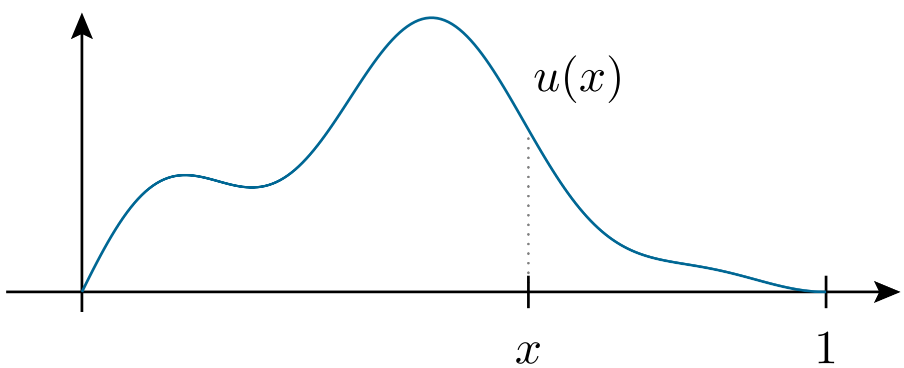
$$-\frac{\mathrm{d}^2u}{\mathrm{d}x^2} = f(x), \quad 0 \leqslant x \leqslant 1, \qquad u(0) = u(1) = 0.$$

$f(x)$ is given, and the goal is to find $u(x)$.

Physical interpretation: models steady-state temperature distribution in a bar ($u(x)$ is temperature at point x) under influence of an external heat source $f(x)$ and with ends fixed at $0°$ (ice cube at the ends?).

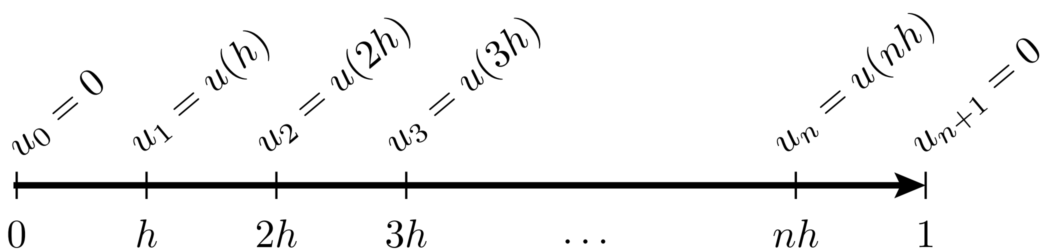
Remark 3. Note that this simple BVP can be solved by integrating $f(x)$ twice. We get two constants of integration, and so we see that the boundary condition $u(0) = u(1) = 0$ makes the solution $u(x)$ unique.

Of course, in the real applications the BVP would be harder. Also, $f(x)$ might only be known at some points, so we cannot use calculus to integrate it.



We will approximate this problem as follows:

- replace $u(x)$ by its values at equally spaced points in $[0, 1]$



- approximate $\frac{\mathrm{d}^2u}{\mathrm{d}x^2}$ at these points (**finite differences**)
- replace differential equation with linear equation at each point
- solve linear problem using Gaussian elimination

Armin Straub
astraub@illinois.edu

Finite differences

Finite differences for first derivative:

$$\frac{\mathrm{d}u}{\mathrm{d}x} \approx \frac{\Delta u}{\Delta x} \begin{array}{l} = \dfrac{u(x+h) - u(x)}{h} \\ \overset{\text{or}}{=} \dfrac{u(x) - u(x-h)}{h} \\ \overset{\text{or}}{=} \dfrac{u(x+h) - u(x-h)}{2h} \\ \text{symmetric and most accurate} \end{array}$$

Note. Recall that you can always use L'Hospital's rule to determine the limit of such quantities (especially more complicated ones) as $h \to 0$.

Finite difference for second derivative:

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} \approx \frac{u(x+h) - 2u(x) + u(x-h)}{h^2}$$

the only symmetric choice involving only $u(x)$, $u(x \pm h)$

Question 4. Why does this approximate $\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}$ as $h \to 0$?

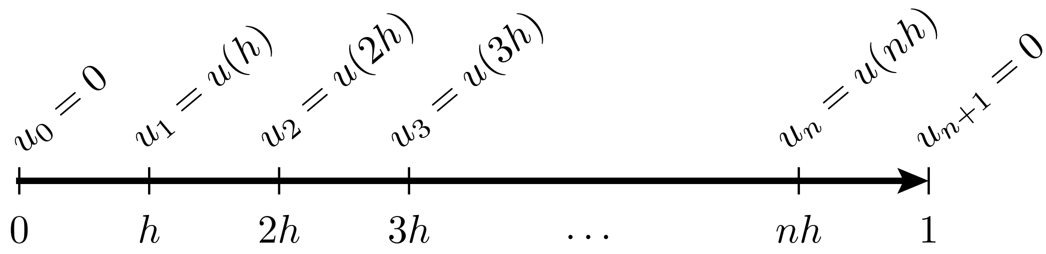
Solution.

$$\begin{aligned} \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} &\approx \frac{\frac{\mathrm{d}u}{\mathrm{d}x}(x+h) - \frac{\mathrm{d}u}{\mathrm{d}x}(x)}{h} \\ &\approx \frac{\frac{u(x+h) - u(x)}{h} - \frac{u(x) - u(x-h)}{h}}{h} \\ &\approx \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} \end{aligned}$$

Armin Straub
astraub@illinois.edu

Setting up the linear equations

$$-\frac{\mathrm{d}^2u}{\mathrm{d}x^2} = f(x), \quad 0 \leqslant x \leqslant 1, \qquad u(0) = u(1) = 0.$$



Using $-\frac{\mathrm{d}^2u}{\mathrm{d}x^2} \approx -\frac{u(x+h) - 2u(x) + u(x-h)}{h^2}$, we get:

at $x = h$: $-\frac{u(2h) - 2u(h) + u(0)}{h^2} = f(h)$

$$\Longrightarrow \; 2u_1 - u_2 = h^2 f(h) \tag{1}$$

at $x = 2h$: $-\frac{u(3h) - 2u(2h) + u(h)}{h^2} = f(2h)$

$$\Longrightarrow \; -u_1 + 2u_2 - u_3 = h^2 f(2h) \tag{2}$$

at $x = 3h$:

$$\Longrightarrow \; -u_2 + 2u_3 - u_4 = h^2 f(3h) \tag{3}$$

$$\vdots$$

at $x = nh$: $-\frac{u((n+1)h) - 2u(nh) + u((n-1)h)}{h^2} = f(nh)$

$$\Longrightarrow \; -u_{n-1} + 2u_n = h^2 f(nh) \tag{n}$$

Example 5. In the case of six divisions ($n = 5$, $h = \frac{1}{6}$), we get:

$$\underbrace{\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} h^2 f(h) \\ h^2 f(2h) \\ h^2 f(3h) \\ h^2 f(4h) \\ h^2 f(5h) \end{bmatrix}}_{\boldsymbol{b}}$$

Armin Straub
astraub@illinois.edu

Such a matrix is called a **band matrix**. As we will see next, such matrices always have a particularly simple LU decomposition.

Gaussian elimination:

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \underset{\begin{bmatrix} 1 & & & & \\ \frac{1}{2} & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}}{\overset{R2 \to R2 + \frac{1}{2}R1}{\rightsquigarrow}} \begin{bmatrix} 2 & -1 & & & \\ 0 & \frac{3}{2} & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

$$\underset{\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & \frac{2}{3} & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}}{\overset{R3 \to R3 + \frac{2}{3}R2}{\rightsquigarrow}} \begin{bmatrix} 2 & -1 & & & \\ 0 & \frac{3}{2} & -1 & & \\ & 0 & \frac{4}{3} & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

$$\underset{\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & \frac{3}{4} & 1 & \\ & & & & 1 \end{bmatrix}}{\overset{R4 \to R4 + \frac{3}{4}R3}{\rightsquigarrow}} \begin{bmatrix} 2 & -1 & & & \\ 0 & \frac{3}{2} & -1 & & \\ & 0 & \frac{4}{3} & -1 & \\ & & 0 & \frac{5}{4} & -1 \\ & & & -1 & 2 \end{bmatrix}$$

$$\underset{\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & \frac{4}{5} & 1 \end{bmatrix}}{\overset{R5 \to R5 + \frac{4}{5}R4}{\rightsquigarrow}} \begin{bmatrix} 2 & -1 & & & \\ 0 & \frac{3}{2} & -1 & & \\ & 0 & \frac{4}{3} & -1 & \\ & & 0 & \frac{5}{4} & -1 \\ & & & 0 & \frac{6}{5} \end{bmatrix}$$

In conclusion, we have the LU decomposition:

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & -\frac{3}{4} & 1 & \\ & & & -\frac{4}{5} & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & & & \\ & \frac{3}{2} & -1 & & \\ & & \frac{4}{3} & -1 & \\ & & & \frac{5}{4} & -1 \\ & & & & \frac{6}{5} \end{bmatrix}$$

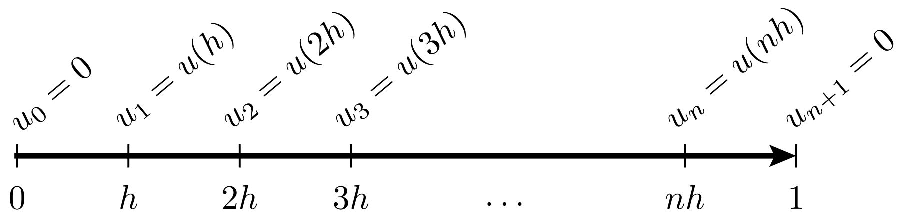
That's how the LU decomposition of band matrices always looks like.

Review

- Goal: solve for $u(x)$ in the **boundary value problem** (BVP)

$$-\frac{\mathrm{d}^2u}{\mathrm{d}x^2} = f(x), \quad 0 \leqslant x \leqslant 1, \qquad u(0) = u(1) = 0.$$

- replace $u(x)$ by its values at equally spaced points in $[0,1]$



- $-\frac{\mathrm{d}^2u}{\mathrm{d}x^2} \approx -\frac{u(x+h) - 2u(x) + u(x-h)}{h^2}$ at these points (**finite differences**)
- get a linear equation at each point $x = h, 2h, \ldots, nh$; for $n = 5$, $h = \frac{1}{6}$:

$$\underbrace{\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} h^2 f(h) \\ h^2 f(2h) \\ h^2 f(3h) \\ h^2 f(4h) \\ h^2 f(5h) \end{bmatrix}}_{\boldsymbol{b}}$$

- Compute the LU decomposition:

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & -\frac{3}{4} & 1 & \\ & & & -\frac{4}{5} & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & & & \\ & \frac{3}{2} & -1 & & \\ & & \frac{4}{3} & -1 & \\ & & & \frac{5}{4} & -1 \\ & & & & \frac{6}{5} \end{bmatrix}$$

That's how the LU decomposition of **band matrices** always looks like.

LU decomposition vs matrix inverse

In many applications, we don't just solve $A\boldsymbol{x} = \boldsymbol{b}$ for a single $\boldsymbol{b}$, but for many different $\boldsymbol{b}$ (think millions).

Note, for instance, that in our example of "steady-state temperature distribution in a bar" the matrix A is always the same (it only depends on the kind of problem), whereas the vector $\boldsymbol{b}$ models the external heat (and thus changes for each specific instance).

- That's why the LU decomposition saves us from repeating lots of computation in comparison with Gaussian elimination on $[\,A\,|\,\boldsymbol{b}\,]$.
- What about computing A^{-1}?

  We are going to see that this is a bad idea. (It usually is.)

Example 1. When using LU decomposition to solve $A\boldsymbol{x} = \boldsymbol{b}$, we employ forward and backward substitution:

$$A\boldsymbol{x} = \boldsymbol{b} \quad \overset{A=LU}{\Longleftrightarrow} \quad L\boldsymbol{c} = \boldsymbol{b} \quad \text{and} \quad U\boldsymbol{x} = \boldsymbol{c}$$

Here, we have to solve, for each $\boldsymbol{b}$,

$$\begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & -\frac{3}{4} & 1 & \\ & & & -\frac{4}{5} & 1 \end{bmatrix}\boldsymbol{c} = \boldsymbol{b}, \quad \begin{bmatrix} 2 & -1 & & & \\ & \frac{3}{2} & -1 & & \\ & & \frac{4}{3} & -1 & \\ & & & \frac{5}{4} & -1 \\ & & & & \frac{6}{5} \end{bmatrix}\boldsymbol{x} = \boldsymbol{c}$$

by forward and backward substitution.

How many operations (additions and multiplications) are needed in the $n \times n$ case?

$2(n-1)$ for $L\boldsymbol{c} = \boldsymbol{b}$, and $1 + 2(n-1)$ for $U\boldsymbol{x} = \boldsymbol{c}$.

So, roughly, a total of $4n$ operations.

On the other hand,

$$A^{-1} = \frac{1}{6}\begin{bmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}.$$

How many operations are needed to compute $A^{-1}\boldsymbol{b}$?

This time, we need roughly $2n^2$ additions and multiplications.

Armin Straub
astraub@illinois.edu

Conclusions

- Large matrices met in applications usually are not random but have some structure (such as band matrices).
- When solving linear equations, we do not (try to) compute A^{-1}.
  - It destroys structure in practical problems.
  - As a result, it can be orders of magnitude slower,
  - and require orders of magnitude more memory.
  - It is also numerically unstable.
  - LU decomposition can be adjusted to not have these drawbacks.

A practice problem

Example 2. Above we computed the LU decomposition for $n=5$. For comparison, here are the details for computing the inverse when $n=3$.

Do it for $n=5$, and appreciate just how much computation has to be done.

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

Solution.

$$\begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \overset{R2 \to R2 + \frac{1}{2}R1}{\rightsquigarrow} \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix}$$

$$\overset{R3 \to R3 + \frac{2}{3}R2}{\rightsquigarrow} \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix}$$

$$\underset{R3 \to \frac{3}{4}R3}{\overset{R1 \to \frac{1}{2}R1}{\underset{R2 \to \frac{2}{3}R2}{\rightsquigarrow}}} \begin{bmatrix} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix}$$

$$\overset{R2 \to R2 + \frac{2}{3}R3}{\underset{R1 \to R1 + \frac{1}{2}R2}{\rightsquigarrow}} \begin{bmatrix} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix}$$

Hence, $\begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix}$.

Vector spaces and subspaces

We have already encountered **vectors** in $\mathbb{R}^n$. Now, we discuss the general concept of vectors.

In place of the space $\mathbb{R}^n$, we think of general **vector spaces**.

Definition 3. A **vector space** is a nonempty set V of elements, called **vectors**, which may be added and scaled (multiplied with real numbers).

The two operations of addition and scalar multiplication must satisfy the following *axioms* for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in V, and all scalars c, d.

(a) $\boldsymbol{u}+\boldsymbol{v}$ is in V

(b) $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$

(c) $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$

(d) there is a vector (called the **zero vector**) $\mathbf{0}$ in V such that $\boldsymbol{u}+\mathbf{0}=\boldsymbol{u}$ for all $\boldsymbol{u}$ in V

(e) there is a vector $-\boldsymbol{u}$ such that $\boldsymbol{u}+(-\boldsymbol{u})=\mathbf{0}$

(f) $c\boldsymbol{u}$ is in V

(g) $c(\boldsymbol{u}+\boldsymbol{v})=c\boldsymbol{u}+c\boldsymbol{v}$

(h) $(c+d)\boldsymbol{u}=c\boldsymbol{u}+d\boldsymbol{u}$

(i) $(cd)\boldsymbol{u}=c(d\boldsymbol{u})$

(j) $1\boldsymbol{u}=\boldsymbol{u}$

tl;dr — A **vector space** is a collection of vectors which can be added and scaled (without leaving the space!); subject to the usual rules you would hope for.

namely: associativity, commutativity, distributivity

Example 4. Convince yourself that $M_{2\times 2}=\left\{\begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \text{ in } \mathbb{R}\right\}$ is a vector space.

Solution. In this context, the zero vector is $\mathbf{0}=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Addition is componentwise:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}+\begin{bmatrix} e & f \\ g & h \end{bmatrix}=\begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

Scaling is componentwise:

$$r\begin{bmatrix} a & b \\ c & d \end{bmatrix}=\begin{bmatrix} ra & rb \\ rc & rd \end{bmatrix}$$

Addition and scaling satisfy the axioms of a vector space because they are defined component-wise and because ordinary addition and multiplication are associative, commutative, distributive and what not.

Important note: we do not use matrix multiplication here!

Note: as a vector space, $M_{2\times 2}$ behaves precisely like $\mathbb{R}^4$; we could translate between the two via

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

A fancy person would say that these two vector spaces are **isomorphic**.

Example 5. Let $\mathbb{P}_n$ be the set of all polynomials of degree at most $n \geqslant 0$. Is $\mathbb{P}_n$ a vector space?

Solution. Members of $\mathbb{P}_n$ are of the form

$$p(t) = a_0 + a_1 t + \ldots + a_n t^n,$$

where $a_0, a_1, \ldots, a_n$ are in $\mathbb{R}$ and t is a variable.

$\mathbb{P}_n$ is a vector space.

Adding two polynomials:

$$\begin{aligned} & [a_0 + a_1 t + \ldots + a_n t^n] + [b_0 + b_1 t + \ldots + b_n t^n] \\ = & [(a_0 + b_0) + (a_1 + b_1) t + \ldots + (a_n + b_n) t^n] \end{aligned}$$

So addition works "component-wise" again.

Scaling a polynomial:

$$\begin{aligned} & r[a_0 + a_1 t + \ldots + a_n t^n] \\ = & [(r a_0) + (r a_1) t + \ldots + (r a_n) t^n] \end{aligned}$$

Scaling works "component-wise" as well.

Again: the vector space axioms are satisfied because addition and scaling are defined component-wise.

As in the previous example, we see that $\mathbb{P}_n$ is isomorphic to $\mathbb{R}^{n+1}$:

$$a_0 + a_1 t + \ldots + a_n t^n \quad \longleftrightarrow \quad \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Example 6. Let V be the set of all polynomials of degree exactly 3. Is V a vector space?

Solution. No, because V does not contain the zero polynomial $p(t) = 0$.

Every vector space has to have a zero vector; this is an easy necessary (but not sufficient) criterion when thinking about whether a set is a vector space.

More generally, the sum of elements in V might not be in V:

$$[1 + 4t^2 + t^3] + [2 - t + t^2 - t^3] = [3 - t + 5t^2]$$

Armin Straub
astraub@illinois.edu

Review

- A **vector space** is a set of vectors which can be added and scaled (without leaving the space!); subject to the "usual" rules.
- The set of all polynomials of degree up to 2 is a vector space.

$$\begin{aligned}[a_0 + a_1 t + a_2 t^2] + [b_0 + b_1 t + b_2 t^2] &= [(a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2] \\ r[a_0 + a_1 t + a_2 t^2] &= [(r a_0) + (r a_1)t + (r a_2)t^2]\end{aligned}$$

  Note how it "works" just like $\mathbb{R}^3$.

- The set of all polynomials of degree exactly 2 is not a vector space.

$$\underbrace{[1 + 4t + t^2]}_{\text{degree 2}} + \underbrace{[3 - t - t^2]}_{\text{degree 2}} = \underbrace{[4 + 3t]}_{\text{NOT degree 2}}$$

- An easy test that often works is to check whether the set contains the zero vector. (Works in the previous case.)

Example 1. Let V be the set of all functions $f\colon \mathbb{R} \to \mathbb{R}$. Is V a vector space?

Solution. Yes!

Addition of functions f and g:

$$(f + g)(x) = f(x) + g(x)$$

Note that, once more, this definition is "component-wise".

Likewise for scalar multiplication.

Subspaces

Definition 2. A subset W of a vector space V is a **subspace** if W is itself a vector space.

Since the rules like associativity, commutativity and distributivity still hold, we only need to check the following:

$W \subseteq V$ is a subspace of V if

- W contains the zero vector $\mathbf{0}$,
- W is closed under addition, (i.e. if $\boldsymbol{u}, \boldsymbol{v} \in W$ then $\boldsymbol{u} + \boldsymbol{v} \in W$)
- W is closed under scaling. (i.e. if $\boldsymbol{u} \in W$ and $c \in \mathbb{R}$ then $c\boldsymbol{u} \in W$)

Note that "$\mathbf{0}$ in W" (first condition) follows from "W closed under scaling" (third condition). But it is crucial and easy to check, so deserves its own bullet point.

Example 3. Is $W = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ a subspace of $\mathbb{R}^2$?

Solution. Yes!

- W contains $\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
- $\begin{bmatrix} a \\ a \end{bmatrix} + \begin{bmatrix} b \\ b \end{bmatrix} = \begin{bmatrix} a+b \\ a+b \end{bmatrix}$ is in W.
- $c \begin{bmatrix} a \\ a \end{bmatrix} = \begin{bmatrix} ca \\ ca \end{bmatrix}$ is in W.

Example 4. Is $W = \left\{ \begin{bmatrix} a \\ 0 \\ b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$ a subspace of $\mathbb{R}^3$?

Solution. Yes!

- W contains $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.
- $\begin{bmatrix} a_1 \\ 0 \\ b_1 \end{bmatrix} + \begin{bmatrix} a_2 \\ 0 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 + a_2 \\ 0 \\ b_1 + b_2 \end{bmatrix}$ is in W.
- $c \begin{bmatrix} a \\ 0 \\ b \end{bmatrix} = \begin{bmatrix} ca \\ 0 \\ cb \end{bmatrix}$ is in W.

The subspace W is isomorphic to $\mathbb{R}^2$ (translation: $\begin{bmatrix} a \\ 0 \\ b \end{bmatrix} \leftrightarrow \begin{bmatrix} a \\ b \end{bmatrix}$) but they are not the same!

Example 5. Is $W = \left\{ \begin{bmatrix} a \\ 1 \\ b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$ a subspace of $\mathbb{R}^3$?

Solution. No! Missing $\mathbf{0}$.

Armin Straub
astraub@illinois.edu

Note: $W = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \left\{ \begin{bmatrix} a \\ 0 \\ b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$ is "close" to a vector space.

Geometrically, it is a plane, but it does not contain the origin.

Example 6. Is $W = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$ a subspace of $\mathbb{R}^2$?

Solution. Yes!

- W contains $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.
- $\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is in W.
- $c \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is in W.

Example 7. Is $W = \left\{ \begin{bmatrix} x \\ x+1 \end{bmatrix} : x \text{ in } \mathbb{R} \right\}$ a subspace of $\mathbb{R}^2$?

Solution. No! W does not contain $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

[If $\mathbf{0}$ is missing, some other things always go wrong as well.
For instance, $2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ or $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ are not in W.]

Example 8. Is $W = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \cup \left\{ \begin{bmatrix} x \\ x+1 \end{bmatrix} : x \text{ in } \mathbb{R} \right\}$ a subspace of $\mathbb{R}^2$?

[In other words, W is the set from the previous example plus the zero vector.]

Solution. No! $2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ not in W.

Spans of vectors are subspaces

Review. The **span** of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ is the set of all their linear combinations. We denote it by $\text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$.

In other words, $\text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is the set of all vectors of the form

$$c_1 \boldsymbol{v}_1 + c_2 \boldsymbol{v}_2 + \ldots + c_m \boldsymbol{v}_m,$$

where $c_1, c_2, \ldots, c_m$ are scalars.

Theorem 9. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are in a vector space V, then $\text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ is a subspace of V.

Why?

- $\mathbf{0}$ is in $\text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$
- $[c_1 \boldsymbol{v}_1 + \ldots + c_m \boldsymbol{v}_m] + [d_1 \boldsymbol{v}_1 + \ldots + d_m \boldsymbol{v}_m]$
  $= [(c_1 + d_1) \boldsymbol{v}_1 + \ldots + (c_m + d_m) \boldsymbol{v}_m]$
- $r[c_1 \boldsymbol{v}_1 + \ldots + c_m \boldsymbol{v}_m] = [(r c_1) \boldsymbol{v}_1 + \ldots + (r c_m) \boldsymbol{v}_m]$

Example 10. Is $W = \left\{ \begin{bmatrix} a+3b \\ 2a-b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$ a subspace of $\mathbb{R}^2$?

Solution. Write vectors in W in the form

$$\begin{bmatrix} a+3b \\ 2a-b \end{bmatrix} = \begin{bmatrix} a \\ 2a \end{bmatrix} + \begin{bmatrix} 3b \\ -b \end{bmatrix} = a \begin{bmatrix} 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

to see that

$$W = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right\}.$$

By the theorem, W is a vector space. Actually, $W = \mathbb{R}^2$.

Example 11. Is $W = \left\{ \begin{bmatrix} -a & 2b \\ a+b & 3a \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$ a subspace of $M_{2\times 2}$, the space of 2×2 matrices?

Solution. Write "vectors" in W in the form

$$\begin{bmatrix} -a & 2b \\ a+b & 3a \end{bmatrix} = a \begin{bmatrix} -1 & 0 \\ 1 & 3 \end{bmatrix} + b \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$$

to see that

$$W = \operatorname{span}\left\{ \begin{bmatrix} -1 & 0 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix} \right\}.$$

By the theorem, W is a vector space.

Armin Straub
astraub@illinois.edu

Practice problems

Example 12. Are the following sets vector spaces?

(a) $W_1 = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a + 3b = 0, 2a - c = 1 \right\}$

No, W_1 does not contain $\mathbf{0}$.

(b) $W_2 = \left\{ \begin{bmatrix} a+c & -2b \\ b+3c & c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

Yes, $W_2 = \operatorname{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \right\}$.

Hence, W_2 is a subspace of the vector space $\mathrm{Mat}_{2\times 2}$ of all 2×2 matrices.

(c) $W_3 = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} : ab \geqslant 0 \right\}$

No. For instance, $\begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ is not in W_3.

(d) W_4 is the set of all polynomials $p(t)$ such that $p'(2) = 1$.

No. W_4 does not contain the zero polynomial.

(e) W_5 is the set of all polynomials $p(t)$ such that $p'(2) = 0$.

Yes. If $p'(2) = 0$ and $q'(2) = 0$, then $(p + q)'(2) = 0$. Likewise for scaling.

Hence, W_5 is a subspace of the vector space of all polynomials.

Armin Straub
astraub@illinois.edu

Midterm!

- Midterm 1: Thursday, 7–8:15pm
  - in 23 Psych if your last name starts with A or B
  - in **Foellinger Auditorium** if your last name starts with C, D, ..., Z
  - bring a picture ID and show it when turning in the exam

Review

- A **vector space** is a set V of vectors which can be added and scaled (without leaving the space!); subject to the "usual" rules.
- $W \subseteq V$ is a **subspace** of V if it is a vector space itself; that is,
  - W contains the zero vector $\mathbf{0}$,
  - W is closed under addition, (i.e. if $\boldsymbol{u}, \boldsymbol{v} \in W$ then $\boldsymbol{u} + \boldsymbol{v} \in W$)
  - W is closed under scaling. (i.e. if $\boldsymbol{u} \in W$ and $c \in \mathbb{R}$ then $c\boldsymbol{u} \in W$)
- $\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ is always a subspace of V. ($\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are vectors in V)

Example 1. Is $W = \left\{ \begin{bmatrix} 2a-b & 0 \\ b & 3 \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$ a subspace of $M_{2\times 2}$, the space of 2×2 matrices?

Solution. No, W does not contain the zero "vector".

Example 2. Is $W = \left\{ \begin{bmatrix} 2a-b & 0 \\ b & 3a \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$ a subspace of $M_{2\times 2}$, the space of 2×2 matrices?

Solution. Write "vectors" in W in the form

$$\begin{bmatrix} 2a-b & 0 \\ b & 3a \end{bmatrix} = a \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} + b \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}$$

to see that

$$W = \operatorname{span}\left\{ \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix} \right\}.$$

Like any span, W is a vector space.

Armin Straub
astraub@illinois.edu

Example 3. Are the following sets vector spaces?

(a) $W_1 = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a + 3b = 0, 2a - c = 1 \right\}$

No, W_1 does not contain $\mathbf{0}$.

(b) $W_2 = \left\{ \begin{bmatrix} a + c & -2b \\ b + 3c & c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

Yes, $W_2 = \operatorname{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \right\}$.

Hence, W_2 is a subspace of the vector space $\mathrm{Mat}_{2 \times 2}$ of all 2×2 matrices.

(c) $W_3 = \left\{ \begin{bmatrix} a + c & -2b \\ b + 3c & c + 7 \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$ (more complicated)

We still have $W_3 = \begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix} + \operatorname{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \right\}$.

Hence, W_3 is a subspace if and only if $\begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix}$ is in the span. (We can answer such questions!)

Equivalently (why?!), we have to check whether $\begin{bmatrix} a + c & -2b \\ b + 3c & c + 7 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ has solutions a, b, c.

There is no solution ($-2b = 0$ implies $b = 0$, then $b + 3c = 0$ implies $c = 0$; this contradicts $c + 7 = 0$).

(d) $W_4 = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} : ab \geqslant 0 \right\}$

No. For instance, $\begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ is not in W_4.

(e) W_5 is the set of all polynomials $p(t)$ such that $p'(2) = 1$.

No. W_5 does not contain the zero polynomial.

(f) W_6 is the set of all polynomials $p(t)$ such that $p'(2) = 0$.

Yes. If $p'(2) = 0$ and $q'(2) = 0$, then $(p + q)'(2) = p'(2) + q'(2) = 0$. Likewise for scaling.

Hence, W_6 is a subspace of the vector space of all polynomials.

What we learned before vector spaces

Linear systems

- Systems of equations can be written as $A\boldsymbol{x} = \boldsymbol{b}$.

$$\begin{array}{rcrcr} x_1 & - & 2x_2 & = & -1 \\ -x_1 & + & 3x_2 & = & 3 \end{array} \quad \Longrightarrow \quad \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

  Sometimes, we represent the system by its augmented matrix.

$$\left[\begin{array}{cc|c} 1 & -2 & -1 \\ -1 & 3 & 3 \end{array}\right]$$

- A linear system has either
  - no solution (such a system is called **inconsistent**),

    $\Longleftrightarrow$ echelon form contains row $[\,0\ \ \ldots\ \ 0\,|\,b\,]$ with $b \neq 0$
  - one unique solution,

    $\Longleftrightarrow$ system is consistent and has no free variables
  - infinitely many solutions.

    $\Longleftrightarrow$ system is consistent and has at least one free variable
- We know different techniques for solving systems $A\boldsymbol{x} = \boldsymbol{b}$.
  - Gaussian elimination on $[\ A\ \ \boldsymbol{b}\ ]$
  - LU decomposition $A = LU$
  - using matrix inverse, $\boldsymbol{x} = A^{-1}\boldsymbol{b}$

Matrices and vectors

- A **linear combination** of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ is of the form

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_m\boldsymbol{v}_m.$$

- $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is the set of all such linear combinations.
  - Spans are always vector spaces.
  - For instance, a span in $\mathbb{R}^3$ can be $\{\mathbf{0}\}$, a line, a plane, or $\mathbb{R}^3$.
- The **transpose** A^T of a matrix A has rows and columns flipped.

$$\begin{bmatrix} 2 & 0 \\ 3 & 1 \\ -1 & 4 \end{bmatrix}^T = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 1 & 4 \end{bmatrix}$$

  - $(A + B)^T = A^T + B^T$
  - $(AB)^T = B^T A^T$
- An $m \times n$ **matrix** A has m rows and n columns.
- The product $A\boldsymbol{x}$ of **matrix times vector** is

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_n\boldsymbol{a}_n.$$

- Different interpretations of the product of **matrix times matrix**:
  - column interpretation

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a+3c & b & c \\ d+3f & e & f \\ g+3i & h & i \end{bmatrix}$$

  - row interpretation

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 3a+g & 3b+h & 3c+i \end{bmatrix}$$

  - row-column rule

$$(AB)_{i,j} = (\text{row } i \text{ of } A) \cdot (\text{col } j \text{ of } B)$$

Armin Straub
astraub@illinois.edu

- The **inverse** A^{-1} of A is characterized by $A^{-1}A = I$ (or $AA^{-1} = I$).
  - $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
  - Can compute A^{-1} using Gauss–Jordan method.

$$[\ A \quad I\ ] \quad \overset{\text{RREF}}{\rightsquigarrow} \quad [\ I \quad A^{-1}\ ]$$

  - $(A^T)^{-1} = (A^{-1})^T$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - An $n \times n$ matrix A is invertible

    $\Longleftrightarrow$ A has n pivots

    $\Longleftrightarrow$ $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution (if true for one $\boldsymbol{b}$, then true for all $\boldsymbol{b}$)

Gaussian elimination

- **Gaussian elimination** can bring any matrix into an **echelon form**.

$$\begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

  It proceeds by **elementary row operations**:
  - **(replacement)** Add one row to a multiple of another row.
  - **(interchange)** Interchange two rows.
  - **(scaling)** Multiply all entries in a row by a nonzero constant.
- Each elementary row operation can be encoded as multiplication with an **elementary matrix**.

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e-a & f-b & g-c & h-d \\ i & j & k & l \end{bmatrix}$$

- We can continue row reduction to obtain the (unique) RREF.

Using Gaussian elimination

Gaussian elimination and row reductions allow us:

- solve systems of linear systems

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 4 & -5 \\ 3 & -7 & 8 & 8 & 9 \\ 3 & -9 & 12 & 6 & 15 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & -2 & 0 & -24 \\ 0 & 1 & -2 & 0 & -7 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right] \qquad \begin{cases} x_1 = -24 + 2x_3 \\ x_2 = -7 + 2x_3 \\ x_3 \text{ free} \\ x_4 = 4 \end{cases}$$

- compute the LU decomposition $A = LU$

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ & -8 & -2 \\ & & 1 \end{bmatrix}$$

- compute the inverse of a matrix

to find $\begin{bmatrix} 2 & 0 & 0 \\ -3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \\ \frac{3}{2} & 1 & 0 \end{bmatrix}$, we use Gauss–Jordan:

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{bmatrix}$$

- determine whether a vector is a linear combination of other vectors

$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ if and only if

the system corresponding to $\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 0 & 3 \end{array}\right]$ is consistent.

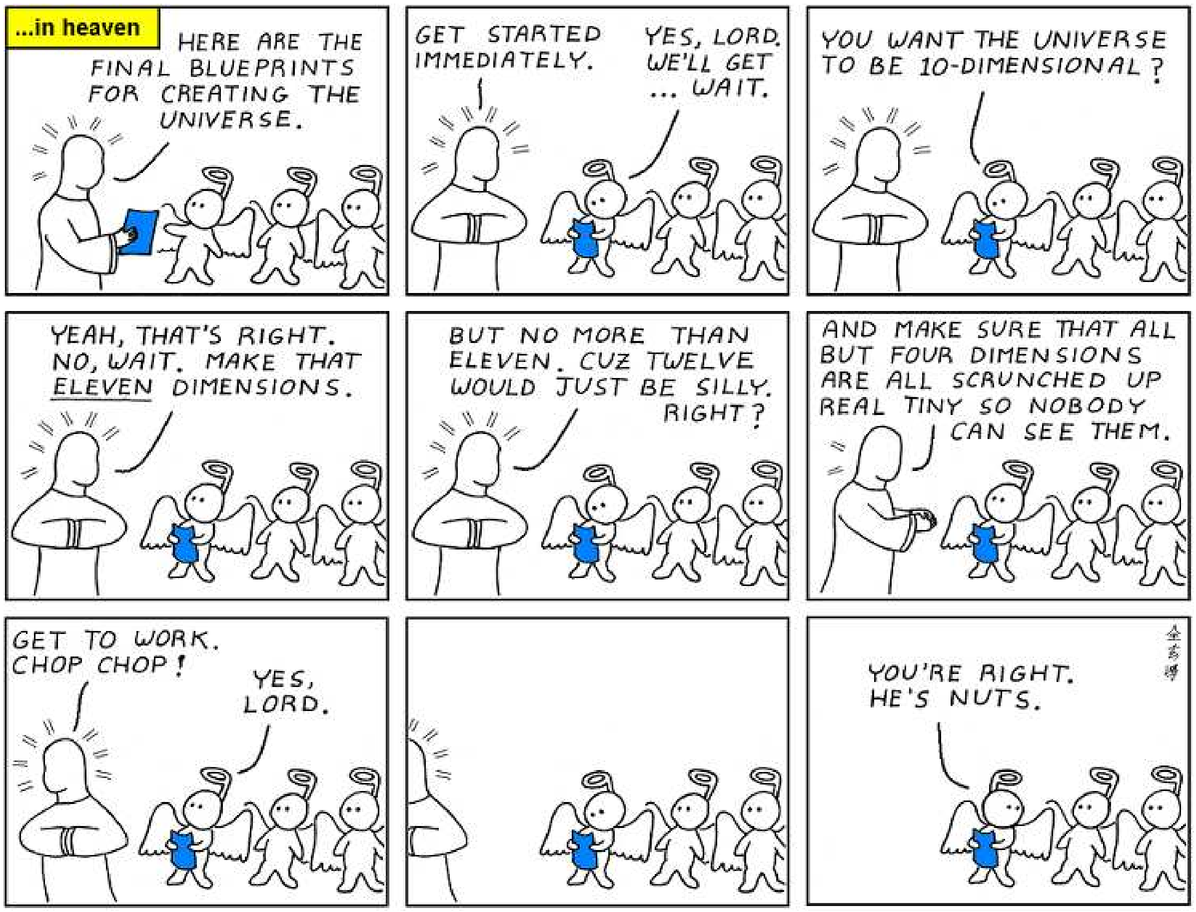
(Each solution $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ gives a linear combination $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$.)

Organizational

- Interested in joining class committee?

  meet ~ 3 times to discuss ideas you may have for improving class

Next: bases, dimension and such



http://abstrusegoose.com/235

Solving $Ax = 0$ and $Ax = b$

Column spaces

Definition 1. The **column space** $\mathrm{Col}(A)$ of a matrix A is the span of the columns of A.

If $A = [\ \boldsymbol{a}_1\ \ \ldots\ \ \boldsymbol{a}_n\ ]$, then $\mathrm{Col}(A) = \mathrm{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$.

- In other words, $\boldsymbol{b}$ is in $\mathrm{Col}(A)$ if and only if $A\boldsymbol{x} = \boldsymbol{b}$ has a solution.

  Why? Because $A\boldsymbol{x} = x_1\boldsymbol{a}_1 + \ldots + x_n\boldsymbol{a}_n$ is the linear combination of columns of A with coefficients given by $\boldsymbol{x}$.

- If A is $m \times n$, then $\mathrm{Col}(A)$ is a subspace of $\mathbb{R}^m$.

  Why? Because any span is a space.

Example 2. Find a matrix A such that $W = \mathrm{Col}(A)$ where

$$W = \left\{ \begin{bmatrix} 2x - y \\ 3y \\ 7x + y \end{bmatrix} : x, y \text{ in } \mathbb{R} \right\}.$$

Solution. Note that

$$\begin{bmatrix} 2x - y \\ 3y \\ 7x + y \end{bmatrix} = x \begin{bmatrix} 2 \\ 0 \\ 7 \end{bmatrix} + y \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}.$$

Hence,

$$W = \mathrm{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 7 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} \right\} = \mathrm{Col}\left(\begin{bmatrix} 2 & -1 \\ 0 & 3 \\ 7 & 1 \end{bmatrix} \right).$$

Null spaces

Definition 3. The **null space** of a matrix A is

$$\mathrm{Nul}(A) = \{\boldsymbol{x} : A\boldsymbol{x} = \boldsymbol{0}\}.$$

In other words, if A is $m \times n$, then its null space consists of those vectors $\boldsymbol{x} \in \mathbb{R}^n$ which solve the **homogeneous** equation $A\boldsymbol{x} = \boldsymbol{0}$.

Theorem 4. If A is $m \times n$, then $\mathrm{Nul}(A)$ is a subspace of $\mathbb{R}^n$.

Proof. We check that $\mathrm{Nul}(A)$ satisfies the conditions of a subspace:

- $\mathrm{Nul}(A)$ contains $\boldsymbol{0}$ because $A\boldsymbol{0} = \boldsymbol{0}$.
- If $A\boldsymbol{x} = \boldsymbol{0}$ and $A\boldsymbol{y} = \boldsymbol{0}$, then $A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y} = \boldsymbol{0}$.

  Hence, $\mathrm{Nul}(A)$ is closed under addition.
- If $A\boldsymbol{x} = \boldsymbol{0}$, then $A(c\boldsymbol{x}) = cA\boldsymbol{x} = \boldsymbol{0}$.

  Hence, $\mathrm{Nul}(A)$ is closed under scalar multiplication.

□

Armin Straub
astraub@illinois.edu

Solving $A\boldsymbol{x} = \mathbf{0}$ yields an *explicit description* of $\operatorname{Nul}(A)$.

By that we mean a description as the span of some vectors.

Example 5. Find an explicit description of $\operatorname{Nul}(A)$ where

$$A = \begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 13 & 0 & 3 \end{bmatrix}.$$

Solution.

$$\begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 13 & 0 & 3 \end{bmatrix} \overset{R2 \to R2 - 2R1}{\rightsquigarrow} \begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 0 & 0 & 1 & -6 & -15 \end{bmatrix}$$
$$\overset{R1 \to \frac{1}{3}R1}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 2 & 1 & 3 \\ 0 & 0 & 1 & -6 & -15 \end{bmatrix}$$
$$\overset{R1 \to R1 - 2R2}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 0 & 13 & 33 \\ 0 & 0 & 1 & -6 & -15 \end{bmatrix}$$

From the RREF we read off a parametric description of the solutions $\boldsymbol{x}$ to $A\boldsymbol{x} = \mathbf{0}$. Note that x_2, x_4, x_5 are free.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2x_2 - 13x_4 - 33x_5 \\ x_2 \\ 6x_4 + 15x_5 \\ x_4 \\ x_5 \end{bmatrix}$$
$$= x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix}$$

In other words,

$$\operatorname{Nul}(A) = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -13 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -33 \\ 0 \\ 15 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Note. The number of vectors in the spanning set for $\operatorname{Nul}(A)$ as derived above (which is as small as possible) equals the number of free variables in $A\boldsymbol{x} = \mathbf{0}$.

Armin Straub
astraub@illinois.edu

Another look at solutions to $A\boldsymbol{x} = \boldsymbol{b}$

Theorem 6. Let $\boldsymbol{x}_p$ be a solution of the equation $A\boldsymbol{x} = \boldsymbol{b}$.

Then every solution to $A\boldsymbol{x} = \boldsymbol{b}$ is of the form $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n$, where $\boldsymbol{x}_n$ is a solution to the **homogeneous** equation $A\boldsymbol{x} = \boldsymbol{0}$.

- In other words, $\{\boldsymbol{x} : A\boldsymbol{x} = \boldsymbol{b}\} = \boldsymbol{x}_p + \mathrm{Nul}(A)$.
- We often call $\boldsymbol{x}_p$ a **particular solution**.

  The theorem then says that every solution to $A\boldsymbol{x} = \boldsymbol{b}$ is the sum of a fixed chosen particular solution and some solution to $A\boldsymbol{x} = \boldsymbol{0}$.

Proof. Let $\boldsymbol{x}$ be another solution to $A\boldsymbol{x} = \boldsymbol{b}$.

We need to show that $\boldsymbol{x}_n = \boldsymbol{x} - \boldsymbol{x}_p$ is in $\mathrm{Nul}(A)$.

$$A(\boldsymbol{x} - \boldsymbol{x}_p) = A\boldsymbol{x} - A\boldsymbol{x}_p = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}$$

□

Example 7. Let $A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 1 \\ 5 \\ 5 \end{bmatrix}$.

Using the RREF, find a parametric description of the solutions to $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{bmatrix} 1 & 3 & 3 & 2 & 1 \\ 2 & 6 & 9 & 7 & 5 \\ -1 & -3 & 3 & 4 & 5 \end{bmatrix} \overset{\substack{R2 \to R2 - 2R1 \\ R3 \to R3 + R1}}{\rightsquigarrow} \begin{bmatrix} 1 & 3 & 3 & 2 & 1 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 6 & 6 & 6 \end{bmatrix}$$

$$\overset{R3 \to R3 - 2R2}{\rightsquigarrow} \begin{bmatrix} 1 & 3 & 3 & 2 & 1 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\overset{R2 \to \frac{1}{3}R2}{\rightsquigarrow} \begin{bmatrix} 1 & 3 & 3 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\overset{R1 \to R1 - 3R2}{\rightsquigarrow} \begin{bmatrix} 1 & 3 & 0 & -1 & -2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Every solution to $A\boldsymbol{x} = \boldsymbol{b}$ is therefore of the form:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2 - 3x_2 + x_4 \\ x_2 \\ 1 - x_4 \\ x_4 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_p} + \underbrace{x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}}_{\text{elements of } \operatorname{Nul}(A)}$$

We can see nicely how every solution is the sum of a particular solution $\boldsymbol{x}_p$ and solutions to $A\boldsymbol{x} = \mathbf{0}$.

Note. A convenient way to just find a particular solution is to set all free variables to zero (here, $x_2 = 0$ and $x_4 = 0$).

Of course, any other choice for the free variables will result in a particular solution.

For instance, $x_2 = 1$ and $x_4 = 1$ we would get $\boldsymbol{x}_p = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 1 \end{bmatrix}$.

Practice problems

- True or false?
  - The solutions to the equation $A\boldsymbol{x} = \boldsymbol{b}$ form a vector space.

    No, with the only exception of $\boldsymbol{b} = \mathbf{0}$.
  - The solutions to the equation $A\boldsymbol{x} = \mathbf{0}$ form a vector space.

    Yes. This is the null space $\operatorname{Nul}(A)$.

Example 8. Is the given set W a vector space?

If possible, express W as the column or null space of some matrix A.

(a) $W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : \ 5x = y + 2z \right\}$

(b) $W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : \ 5x - 1 = y + 2z \right\}$

(c) $W = \left\{ \begin{bmatrix} x \\ y \\ x + y \end{bmatrix} : \ x, y \text{ in } \mathbb{R} \right\}$

Example 9. Find an explicit description of $\operatorname{Nul}(A)$ where

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

Review

- Every solution to $A\boldsymbol{x} = \boldsymbol{b}$ is the sum of a fixed chosen particular solution and some solution to $A\boldsymbol{x} = \boldsymbol{0}$.

  For instance, let $A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 1 \\ 5 \\ 5 \end{bmatrix}$.

  Every solution to $A\boldsymbol{x} = \boldsymbol{b}$ is of the form:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2 - 3x_2 + x_4 \\ x_2 \\ 1 - x_4 \\ x_4 \end{bmatrix} = \underbrace{\begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_p} + \underbrace{x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}}_{\text{elements of Nul}(A)}$$

- Is $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} \right\}$ equal to $\mathbb{R}^3$?

Linear independence

Review.

- $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is the set of all linear combinations

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_m\boldsymbol{v}_m.$$

- $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is a vector space.

Example 1. Is $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} \right\}$ equal to $\mathbb{R}^3$?

Solution. Recall that the span is equal to

$$\left\{ \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} \boldsymbol{x} : \boldsymbol{x} \text{ in } \mathbb{R}^3 \right\}.$$

Hence, the span is equal to $\mathbb{R}^3$ if and only if the system with augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & b_1 \\ 1 & 2 & 1 & b_2 \\ 1 & 3 & 3 & b_3 \end{array}\right]$$

is consistent for all b_1, b_2, b_3.

Gaussian elimination:

$$\left[\begin{array}{rrr|l} 1 & 1 & -1 & b_1 \\ 1 & 2 & 1 & b_2 \\ 1 & 3 & 3 & b_3 \end{array}\right] \rightsquigarrow \left[\begin{array}{rrr|c} 1 & 1 & -1 & b_1 \\ 0 & 1 & 2 & b_2 - b_1 \\ 0 & 2 & 4 & b_3 - b_1 \end{array}\right]$$
$$\rightsquigarrow \left[\begin{array}{rrr|c} 1 & 1 & -1 & b_1 \\ 0 & 1 & 2 & b_2 - b_1 \\ 0 & 0 & 0 & b_3 - 2b_2 + b_1 \end{array}\right]$$

The system is only consistent if $b_3 - 2b_2 + b_1 = 0$.

Hence, the span does not equal all of $\mathbb{R}^3$.

- What went "wrong"? $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} \right\}$

  Well, the three vectors in the span satisfy

$$\begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = -3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

- Hence, $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} \right\} = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}$.

- We are going to say that the three vectors are **linearly dependent** because they satisfy

$$-3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \mathbf{0}.$$

Definition 2. Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ are said to be **linearly independent** if the equation

$$x_1 \boldsymbol{v}_1 + x_2 \boldsymbol{v}_2 + \ldots + x_p \boldsymbol{v}_p = \mathbf{0}$$

has only the trivial solution (namely, $x_1 = x_2 = \ldots = x_p = 0$).

Likewise, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ are said to be **linearly dependent** if there exist coefficients $x_1, \ldots, x_p$, not all zero, such that

$$x_1 \boldsymbol{v}_1 + x_2 \boldsymbol{v}_2 + \ldots + x_p \boldsymbol{v}_p = \mathbf{0}.$$

Example 3.

- Are the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$ independent?
- If possible, find a linear dependence relation among them.

Solution. We need to check whether the equation

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

has more than the trivial solution.

In other words, the three vectors are independent if and only if the system

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$$

has no free variables.

To find out, we reduce the matrix to echelon form:

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Since there is a column without pivot, we do have a free variable.

Hence, the three vectors are not linearly independent.

To find a linear dependence relation, we solve this system.

Initial steps of Gaussian elimination are as before:

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

x_3 is free. $x_2 = -2x_3$, and $x_1 = 3x_3$. Hence, for any x_3,

$$3x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 2x_3 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since we are only interested in one linear combination, we can set, say, $x_3 = 1$:

$$3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Linear independence of matrix columns

- Note that a linear dependence relation, such as

$$3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \mathbf{0},$$

can be written in matrix form as

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

- Hence, each linear dependence relation among the columns of a matrix A corresponds to a nontrivial solution to $A\boldsymbol{x} = \mathbf{0}$.

Theorem 4. Let A be an $m \times n$ matrix.

The columns of A are linearly independent.
$\Longleftrightarrow$ $A\boldsymbol{x} = \mathbf{0}$ has only the solution $\boldsymbol{x} = 0$.
$\Longleftrightarrow$ $\mathrm{Nul}(A) = \{\mathbf{0}\}$
$\Longleftrightarrow$ A has n pivots. (one in each column)

Example 5. Are the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$ independent?

Solution. Put the vectors in a matrix, and produce an echelon form:

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 2 \\ 1 & 3 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 3 \\ 0 & 2 & 4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & -2 \end{bmatrix}$$

Since each column contains a pivot, the three vectors are independent.

Example 6. (once again, short version)

Are the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$ independent?

Solution. Put the vectors in a matrix, and produce an echelon form:

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Since the last column does not contain a pivot, the three vectors are linearly dependent.

Special cases

- A set of a single nonzero vector $\{\boldsymbol{v}_1\}$ is always linearly independent.

  Why? Because $x_1\boldsymbol{v}_1 = 0$ only for $x_1 = 0$.

- A set of two vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is linearly independent if and only if neither of the vectors is a multiple of the other.

  Why? Because if $x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 = 0$ with, say, $x_2 \neq 0$, then $\boldsymbol{v}_2 = -\frac{x_1}{x_2}\boldsymbol{v}_1$.

- A set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ containing the zero vector is linearly dependent.

  Why? Because if, say, $\boldsymbol{v}_1 = \mathbf{0}$, then $\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \ldots + 0\boldsymbol{v}_p = \mathbf{0}$.

- If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. In other words:

  Any set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ of vectors in $\mathbb{R}^n$ is linearly dependent if $p > n$.

  Why?

  Let A be the matrix with columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$. This is a $n \times p$ matrix.

  The columns are linearly independent if and only if each column contains a pivot.

  If $p > n$, then the matrix can have at most n pivots.

  Thus not all p columns can contain a pivot.

  In other words, the columns have to be linearly dependent.

Example 7. With the least amount of work possible, decide which of the following sets of vectors are linearly independent.

(a) $\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 6 \\ 4 \end{bmatrix} \right\}$

Linearly independent, because the two vectors are not multiples of each other.

(b) $\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\}$

Linearly independent, because it is a single nonzero vector.

(c) columns of $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 8 & 7 & 6 \end{bmatrix}$

Linearly dependent, because these are more than 3 (namely, 4) vectors in $\mathbb{R}^3$.

(d) $\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 6 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$

Linearly dependent, because the set includes the zero vector.

Armin Straub
astraub@illinois.edu

Review

- Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ are **linearly dependent** if

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \ldots + x_p\boldsymbol{v}_p = \boldsymbol{0},$$

  and not all the coefficients are zero.
- The columns of A are linearly independent
  $\Longleftrightarrow$ each column of A contains a pivot.
- Are the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$ independent?

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

  So: no, they are dependent! (Coeff's $x_3 = 1$, $x_2 = -2$, $x_1 = 3$)
- Any set of 11 vectors in $\mathbb{R}^{10}$ is linearly dependent.

A basis of a vector space

Definition 1. A set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ in V is a **basis** of V if

- $V = \operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$, and
- the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ are linearly independent.

In other words, $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ in V is a basis of V if and only if every vector $\boldsymbol{w}$ in V can be uniquely expressed as $\boldsymbol{w} = c_1\boldsymbol{v}_1 + \ldots + c_p\boldsymbol{v}_p$.

Example 2. Let $\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Show that $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is a basis of $\mathbb{R}^3$.

It is called the **standard basis**.

Solution.

- Clearly, $\operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\} = \mathbb{R}^3$.
- $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ are independent, because

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  has a pivot in each column.

Definition 3. V is said to have **dimension** p if it has a basis consisting of p vectors.

This definition makes sense because if V has a basis of p vectors, then every basis of V has p vectors. Why? (Think of $V = \mathbb{R}^3$.)

A basis of $\mathbb{R}^3$ cannot have more than 3 vectors, because any set of 4 or more vectors in $\mathbb{R}^3$ is linearly dependent.

A basis of $\mathbb{R}^3$ cannot have less than 3 vectors, because 2 vectors span at most a plane (challenge: can you think of an argument that is more "rigorous"?).

Example 4. $\mathbb{R}^3$ has dimension 3.

Indeed, the standard basis $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ has three elements.

Likewise, $\mathbb{R}^n$ has dimension n.

Example 5. Not all vector spaces have a finite basis. For instance, the vector space of all polynomials has *infinite dimension*.

Its standard basis is $1, t, t^2, t^3, \ldots$

This is indeed a basis, because any polynomial can be written as a unique linear combination: $p(t) = a_0 + a_1 t + \ldots + a_n t^n$ for some n.

Recall that vectors in V form a **basis** of V if they span V and if they are linearly independent. If we know the dimension of V, we only need to check one of these two conditions:

Theorem 6. Suppose that V has dimension d.

- A set of d vectors in V are a basis if they span V.
- A set of d vectors in V are a basis if they are linearly independent.

Why?

- If the d vectors were not independent, then $d - 1$ of them would still span V. In the end, we would find a basis of less than d vectors.
- If the d vectors would not span V, then we could add another vector to the set and have $d + 1$ independent ones.

Example 7. Are the following sets a basis for $\mathbb{R}^3$?

(a) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$

No, the set has less than 3 elements.

(b) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\}$

No, the set has more than 3 elements.

Armin Straub
astraub@illinois.edu

(c) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \right\}$

The set has 3 elements. Hence, it is a basis if and only if the vectors are independent.

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 1 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 5 \end{bmatrix}$$

Since each column contains a pivot, the three vectors are independent.

Hence, this is a basis of $\mathbb{R}^3$.

Example 8. Let $\mathbb{P}_2$ be the space of polynomials of degree at most 2.

- What is the dimension of $\mathbb{P}_2$?
- Is $\{t, 1-t, 1+t-t^2\}$ a basis of $\mathbb{P}_2$?

Solution.

- The standard basis for $\mathbb{P}_2$ is $\{1, t, t^2\}$.

  This is indeed a basis because every polynomial

  $$a_0 + a_1 t + a_2 t^2$$

  can clearly be written as a linear combination of $1, t, t^2$ in a unique way.

  Hence, $\mathbb{P}_2$ has dimension 3.

- The set $\{t, 1-t, 1+t-t^2\}$ has 3 elements. Hence, it is a basis if and only if the three polynomials are linearly independent.

  We need to check whether

  $$\underbrace{x_1 t + x_2(1-t) + x_3(1+t-t^2)}_{(x_2+x_3)+(x_1-x_2+x_3)t-x_3t^2} = 0$$

  has only the trivial solution $x_1 = x_2 = x_3 = 0$.

  We get the equations

  $$\begin{aligned} x_2 + x_3 &= 0 \\ x_1 - x_2 + x_3 &= 0 \\ -x_3 &= 0 \end{aligned}$$

  which clearly only have the trivial solution. (If you don't see it, solve the system!)

  Hence, $\{t, 1-t, 1+t-t^2\}$ is a basis of $\mathbb{P}_2$.

Armin Straub
astraub@illinois.edu

Shrinking and expanding sets of vectors

We can find a basis for $V = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ by discarding, if necessary, some of the vectors in the spanning set.

Example 9. Produce a basis of $\mathbb{R}^2$ from the vectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} -2 \\ -4 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Solution. Three vectors in $\mathbb{R}^2$ have to be linearly dependent.

Here, we notice that $\boldsymbol{v}_2 = -2\boldsymbol{v}_1$.

The remaining vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_3\}$ are a basis of $\mathbb{R}^2$, because the two vectors are clearly independent.

Checking our understanding

Example 10. Subspaces of $\mathbb{R}^3$ can have dimension $0, 1, 2, 3$.

- The only 0-dimensional subspace is $\{\mathbf{0}\}$.
- A 1-dimensional subspace is of the form $\text{span}\{\boldsymbol{v}\}$ where $\boldsymbol{v} \neq 0$.

  These subspaces are lines through the origin.
- A 2-dimensional subspace is of the form $\text{span}\{\boldsymbol{v}, \boldsymbol{w}\}$ where $\boldsymbol{v}$ and $\boldsymbol{w}$ are not multiples of each other.

  These subspaces are planes through the origin.
- The only 3-dimensional subspace is $\mathbb{R}^3$ itself.

True or false?

- Suppose that V has dimension n. Then any set in V containing more than n vectors must be linearly dependent.

  That's correct.
- The space $\mathbb{P}_n$ of polynomials of degree at most n has dimension $n+1$.

  True, as well. A basis is $\{1, t, t^2, \ldots, t^n\}$.
- The vector space of functions $f: \mathbb{R} \to \mathbb{R}$ is infinite-dimensional.

  Yes. A still-infinite-dimensional subspace are the polynomials.
- Consider $V = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$. If one of the vectors, say $\boldsymbol{v}_k$, in the spanning set is a linear combination of the remaining ones, then the remaining vectors still span V.

  True, $\boldsymbol{v}_k$ is not adding anything new.

Review

- $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is a **basis** of V if the vectors
  - span V, and
  - are independent.
- The **dimension** of V is the number of elements in a basis.
- The columns of A are linearly independent

  $\Longleftrightarrow$ each column of A contains a pivot.

Warmup

Example 1. Find a basis and the dimension of

$$W = \left\{ \begin{bmatrix} a+b+2c \\ 2a+2b+4c+d \\ b+c+d \\ 3a+3c+d \end{bmatrix} : a, b, c, d \text{ real} \right\}.$$

Solution.

First, note that

$$W = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Is $\dim W = 4$? No, because the third vector is the sum of the first two.

Suppose we did not notice...

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

$$\rightsquigarrow \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & -3 & -3 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Not a pivot in every column, hence the 4 vectors are dependent.

Armin Straub
astraub@illinois.edu

[Not necessary here, but:

To get a relation, solve $A\boldsymbol{x} = \mathbf{0}$. Set free variable $x_3 = 1$.

Then $x_4 = 0$, $x_2 = -x_3 = -1$ and $x_1 = -x_2 - 2x_3 = -1$. The relation is

$$-\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 1 \\ 3 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0}.$$

Precisely, what we "noticed" to begin with.]

Hence, a basis for W is $\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $\dim W = 3$.

It follows from the echelon form that these vectors are independent.

Every set of linearly independent vectors can be extended to a basis.

In other words, let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ be linearly independent vectors in V. If V has dimension d, then we can find vectors $\boldsymbol{v}_{p+1}, \ldots, \boldsymbol{v}_d$ such that $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d\}$ is a basis of V.

Example 2. Consider

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

- Give a basis for H. What is the dimension of H?
- Extend the basis of H to a basis of $\mathbb{R}^3$.

Solution.

- The vectors are independent. By definition, they span H.

  Therefore, $\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$ is a basis for H.

  In particular, $\dim H = 2$.

- $\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$ is not a basis for $\mathbb{R}^3$. Why?

  Because a basis for $\mathbb{R}^3$ needs to contain 3 vectors.

  Or, because, for instance, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is not in H.

  So: just add this (or any other) missing vector!

  By construction, $\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$ is independent.

  Hence, this automatically is a basis of $\mathbb{R}^3$.

Armin Straub
astraub@illinois.edu

Bases for column and null spaces

Bases for null spaces

To find a basis for $\operatorname{Nul}(A)$:

- find the parametric form of the solutions to $A\boldsymbol{x}=\mathbf{0}$,
- express solutions $\boldsymbol{x}$ as a linear combination of vectors with the free variables as coefficients;
- these vectors form a basis of $\operatorname{Nul}(A)$.

Example 3. Find a basis for $\operatorname{Nul}(A)$ with

$$A=\begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 15 & 0 & 3 \end{bmatrix}.$$

Solution.

$$\begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 6 & 12 & 15 & 0 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 3 & 6 & 6 & 3 & 9 \\ 0 & 0 & 3 & -6 & -15 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 2 & 1 & 3 \\ 0 & 0 & 1 & -2 & -5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 0 & 5 & 13 \\ 0 & 0 & 1 & -2 & -5 \end{bmatrix}$$

The solutions to $A\boldsymbol{x}=\mathbf{0}$ are:

$$\boldsymbol{x}=\begin{bmatrix} -2x_2-5x_4-13x_5 \\ x_2 \\ 2x_4+5x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4\begin{bmatrix} -5 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} + x_5\begin{bmatrix} -13 \\ 0 \\ 5 \\ 0 \\ 1 \end{bmatrix}$$

Hence, $\operatorname{Nul}(A)=\operatorname{span}\left\{\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -13 \\ 0 \\ 5 \\ 0 \\ 1 \end{bmatrix}\right\}$.

These vectors are clearly independent.

If you don't see it, do compute an echelon form! (permute first and third row to the bottom)

Better yet: note that the first vector corresponds to the solution with $x_2=1$ and the other free variables $x_4=0$, $x_5=0$. The second vector corresponds to the solution with $x_4=1$ and the other free variables $x_2=0$, $x_5=0$. The third vector …

Hence, $\left\{\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -13 \\ 0 \\ 5 \\ 0 \\ 1 \end{bmatrix}\right\}$ is a basis for $\operatorname{Nul}(A)$.

Bases for column spaces

Recall that the columns of A are independent

$\iff$ $A\boldsymbol{x}=\boldsymbol{0}$ has only the trivial solution (namely, $\boldsymbol{x}=0$),

$\iff$ A has no free variables.

A basis for $\mathrm{Col}(A)$ is given by the pivot columns of A.

Example 4. Find a basis for $\mathrm{Col}(A)$ with

$$A=\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix}.$$

Solution.

$$\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & -1 & -5 \\ 0 & 0 & 2 & 10 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivot columns are the first and third.

Hence, a basis for $\mathrm{Col}(A)$ is $\left\{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \end{bmatrix}\right\}$.

Warning: For the basis of $\mathrm{Col}(A)$, you have to take the columns of A, not the columns of an echelon form.

Row operations do not preserve the column space.

[For instance, $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \overset{R1\leftrightarrow R2}{\rightsquigarrow} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ have different column spaces (of the same dimension).]

Review

- $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is a **basis** of V if the vectors span V and are independent.
- To obtain a basis for $\mathrm{Nul}(A)$, solve $A\boldsymbol{x} = \boldsymbol{0}$:

$$\begin{bmatrix} 3 & 6 & 6 & 3 \\ 6 & 12 & 15 & 0 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} -2x_2 - 5x_4 \\ x_2 \\ 2x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -5 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

Hence, $\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 0 \\ 2 \\ 1 \end{bmatrix}$ form a basis for $\mathrm{Nul}(A)$.

- To obtain a basis for $\mathrm{Col}(A)$, take the pivot columns of A.

$$\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence, $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \end{bmatrix}$ form a basis for $\mathrm{Col}(A)$.

- Row operations do not preserve the column space.

  For instance, $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \overset{R1 \leftrightarrow R2}{\rightsquigarrow} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

- On the other hand: row operations do preserve the null space.

  Why? Recall why/that we can operate on rows to solve systems like $A\boldsymbol{x} = 0$!

Dimension of $\mathrm{Col}(A)$ and $\mathrm{Nul}(A)$

Definition 1. The **rank** of a matrix A is the number of its pivots.

Theorem 2. Let A be an $m \times n$ matrix of rank r. Then:

- $\dim \mathrm{Col}(A) = r$

  Why? A basis for $\mathrm{Col}(A)$ is given by the pivot columns of A.

- $\dim \mathrm{Nul}(A) = n - r$ is the number of free variables of A

  Why? In our recipe for a basis for $\mathrm{Nul}(A)$, each free variable corresponds to an element in the basis.

- $\dim \mathrm{Col}(A) + \dim \mathrm{Nul}(A) = n$

  Why? Each of the n columns either contains a pivot or corresponds to a free variable.

Armin Straub
astraub@illinois.edu

The four fundamental subspaces

Row space and left null space

Definition 3.

- The **row space** of A is the column space of A^T.

  $\mathrm{Col}(A^T)$ is spanned by the columns of A^T and these are the rows of A.

- The **left null space** of A is the null space of A^T.

  Why "left"? A vector $\boldsymbol{x}$ is in $\mathrm{Nul}(A^T)$ if and only if $A^T\boldsymbol{x} = \mathbf{0}$.

  Note that $A^T\boldsymbol{x} = \mathbf{0} \Longleftrightarrow (A^T\boldsymbol{x})^T = \boldsymbol{x}^T A = \mathbf{0}^T$.

  Hence, $\boldsymbol{x}$ is in $\mathrm{Nul}(A^T)$ if and only if $\boldsymbol{x}^T A = \mathbf{0}$.

Example 4. Find a basis for $\mathrm{Col}(A)$ and $\mathrm{Col}(A^T)$ where

$$A = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix}.$$

Solution. We know what to do for $\mathrm{Col}(A)$ from an echelon form of A, and we could likewise handle $\mathrm{Col}(A^T)$ from an echelon form of A^T.

But wait!

Instead of doing twice the work, we only need an echelon form of A:

$$\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & -1 & -5 \\ 0 & 0 & 2 & 10 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B$$

Hence, the rank of A is 2.

A basis for $\mathrm{Col}(A)$ is $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \end{bmatrix}$.

Recall that $\mathrm{Col}(A) \neq \mathrm{Col}(B)$. That's because we performed row operations.

However, the row spaces are the same! $\mathrm{Col}(A^T) = \mathrm{Col}(B^T)$

The row space is preserved by elementary row operations.

In particular: a basis for $\mathrm{Col}(A^T)$ is given by $\begin{bmatrix} 1 \\ 2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ -5 \end{bmatrix}$.

Theorem 5. (Fundamental Theorem of Linear Algebra, Part I)

Let A be an $m \times n$ matrix of rank r.

- $\dim \mathrm{Col}(A) = r$ (subspace of $\mathbb{R}^m$)
- $\dim \mathrm{Col}(A^T) = r$ (subspace of $\mathbb{R}^n$)
- $\dim \mathrm{Nul}(A) = n - r$ (subspace of $\mathbb{R}^n$) (# of free variables of A)
- $\dim \mathrm{Nul}(A^T) = m - r$ (subspace of $\mathbb{R}^m$)

In particular:

The column and row space always have the same dimension!

In other words, A and A^T have the same rank. [i.e. same number of pivots]

Easy to see for a matrix in echelon form

$$\begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 7 \end{bmatrix},$$

but not obvious for a random matrix.

Armin Straub
astraub@illinois.edu

Linear transformations

Throughout, V and W are vector spaces.

Definition 6. A map $T: V \to W$ is a **linear transformation** if

$$T(c\boldsymbol{x} + d\boldsymbol{y}) = cT(\boldsymbol{x}) + dT(\boldsymbol{y}) \qquad \text{for all } \boldsymbol{x}, \boldsymbol{y} \text{ in } V \text{ and all } c, d \text{ in } \mathbb{R}.$$

In other words, a linear transformation respects addition and scaling:

- $T(\boldsymbol{x} + \boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$
- $T(c\boldsymbol{x}) = cT(\boldsymbol{x})$

It also sends the zero vector in V to the zero vector in W:

- $T(\mathbf{0}) = \mathbf{0}$ [because $T(\mathbf{0}) = T(0 \cdot \mathbf{0}) = 0 \cdot T(\mathbf{0}) = \mathbf{0}$]

Example 7. Let A be an $m \times n$ matrix.

Then the map $T(\boldsymbol{x}) = A\boldsymbol{x}$ is a linear transformation $T: \mathbb{R}^n \to \mathbb{R}^m$.

Why?

Because matrix multiplication is linear:

$A(c\boldsymbol{x} + d\boldsymbol{y}) = cA\boldsymbol{x} + dA\boldsymbol{y}$

The LHS is $T(c\boldsymbol{x} + d\boldsymbol{y})$ and the RHS is $cT(\boldsymbol{x}) + dT(\boldsymbol{y})$.

Example 8. Let $\mathbb{P}_n$ be the vector space of all polynomials of degree at most n. Consider the map $T: \mathbb{P}_n \to \mathbb{P}_{n-1}$ given by

$$T(p(t)) = \frac{\mathrm{d}}{\mathrm{d}t} p(t).$$

This map is linear! Why?

Because differentiation is linear:

$\frac{\mathrm{d}}{\mathrm{d}t}[ap(t) + bq(t)] = a\frac{\mathrm{d}}{\mathrm{d}t}p(t) + b\frac{\mathrm{d}}{\mathrm{d}t}q(t)$

The LHS is $T(ap(t) + bq(t))$ and the RHS is $aT(p(t)) + bT(q(t))$.

Representing linear maps by matrices

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a basis for V.

A linear map $T: V \to W$ is determined by the values $T(\boldsymbol{x}_1), \ldots, T(\boldsymbol{x}_n)$.

Armin Straub
astraub@illinois.edu

Why?

Take any $\boldsymbol{v}$ in V.

It can be written as $\boldsymbol{v} = c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n$ because $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is a basis and hence spans V.

Hence, by the linearity of T,

$$T(\boldsymbol{v}) = T(c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}) = \ c_1T(\boldsymbol{x}_1) + \ldots + c_nT(\boldsymbol{x}_n).$$

Definition 9. (From linear maps to matrices)

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a basis for V, and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ a basis for W.

The **matrix representing** T with respect to these bases

- has n columns (one for each of the $\boldsymbol{x}_j$),
- the j-th column has m entries $a_{1,j}, \ldots, a_{m,j}$ determined by

$$T(\boldsymbol{x}_j) = a_{1,j}\boldsymbol{y}_1 + \ldots + a_{m,j}\boldsymbol{y}_m.$$

Example 10. Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$. Let T be the linear map such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix}.$$

What is the matrix A representing T with respect to the standard bases?

Solution. The standard bases are

$$\underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_2} \text{ for } \mathbb{R}^2, \qquad \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{y}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{y}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{y}_3} \text{ for } \mathbb{R}^3.$$

$$\begin{aligned} T(\boldsymbol{x}_1) = & \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ & = 1\boldsymbol{y}_1 + 2\boldsymbol{y}_2 + 3\boldsymbol{y}_3 \\ & \Longrightarrow \quad A = \begin{bmatrix} 1 & * \\ 2 & * \\ 3 & * \end{bmatrix} \\ T(\boldsymbol{x}_2) = & \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} = 4\boldsymbol{y}_1 + 0\boldsymbol{y}_2 + 7\boldsymbol{y}_3 \\ & \Longrightarrow \quad A = \begin{bmatrix} 1 & 4 \\ 2 & 0 \\ 3 & 7 \end{bmatrix} \end{aligned}$$

(We did not have time yet to discuss the next example in class, but it will be helpful if your discussion section already meets Tuesdays.)

Example 11. As in the previous example, let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$. Let T be the (same) linear map such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix}.$$

What is the matrix B representing T with respect to the following bases?

$$\underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_1}, \underbrace{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}_{\boldsymbol{x}_2} \text{ for } \mathbb{R}^2, \qquad \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\boldsymbol{y}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{y}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{y}_3} \text{ for } \mathbb{R}^3.$$

Solution. This time:

$$\begin{aligned}
T(\boldsymbol{x}_1) &= T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 10 \end{bmatrix} \\
&= 5\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
\Longrightarrow B &= \begin{bmatrix} 5 & * \\ -3 & * \\ 5 & * \end{bmatrix} \\
T(\boldsymbol{x}_2) &= T\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right) = -T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + 2T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= -\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ -2 \\ 11 \end{bmatrix} \\
&= 7\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 9\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
\Longrightarrow B &= \begin{bmatrix} 5 & 7 \\ -3 & -9 \\ 5 & 4 \end{bmatrix}
\end{aligned}$$

can you see it? otherwise: do it!

Tedious, even in this simple example! (But we can certainly do it.)

A matrix representing T encodes in column j the coefficients of $T(\boldsymbol{x}_j)$ expressed as a linear combination of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$.

Practice problems

Example 12. Suppose $A = \begin{bmatrix} 1 & 2 & 3 & 4 & 1 \\ 2 & 4 & 7 & 8 & 1 \end{bmatrix}$. Find the dimensions and a basis for all four fundamental subspaces of A.

Example 13. Suppose A is a 5×5 matrix, and that $\boldsymbol{v}$ is a vector in $\mathbb{R}^5$ which is not a linear combination of the columns of A.

What can you say about the number of solutions to $A\boldsymbol{x} = \boldsymbol{0}$?

Solution. Stop reading, unless you have thought about the problem!

Existence of such a $\boldsymbol{v}$ means that the 5 columns of A do not span $\mathbb{R}^5$.

Hence, the columns are not independent.

In other words, A has at most 4 pivots.

So, at least one free variable.

Which means that $A\boldsymbol{x} = \boldsymbol{0}$ has infinitely many solutions.

Linear transformations

- A map $T: V \to W$ between vector spaces is **linear** if
  - $T(\boldsymbol{x}+\boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$
  - $T(c\boldsymbol{x}) = cT(\boldsymbol{x})$
- Let A be an $m \times n$ matrix.

  $T: \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\boldsymbol{x}) = A\boldsymbol{x}$ is linear.
- $T: \mathbb{P}_n \to \mathbb{P}_{n-1}$ defined by $T(p(t)) = p'(t)$ is linear.
- The only linear maps $T: \mathbb{R} \to \mathbb{R}$ are $T(x) = \alpha x$.

  Recall that $T(0) = 0$ for linear maps.
- Linear maps $T: \mathbb{R}^2 \to \mathbb{R}$ are of the form $T\begin{pmatrix} x \\ y \end{pmatrix} = \alpha x + \beta y$.

  For instance, $T(x,y) = xy$ is not linear: $T\begin{pmatrix} 2x \\ 2y \end{pmatrix} \neq 2T(x,y)$

Example 1. Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$. Let T be the linear map such that

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}, \quad T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}.$$

- What is $T\left(\begin{bmatrix} 0 \\ 4 \end{bmatrix}\right)$?

  $\begin{bmatrix} 0 \\ 4 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

  $T\left(\begin{bmatrix} 0 \\ 4 \end{bmatrix}\right) = T\left(2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = 2T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) + 2T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix} + \begin{bmatrix} 2 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 8 \end{bmatrix}$

> Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a basis for V.
> A linear map $T: V \to W$ is determined by the values $T(\boldsymbol{x}_1), \ldots, T(\boldsymbol{x}_n)$.

Why?

Take any $\boldsymbol{v}$ in V.

Write $\boldsymbol{v} = c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n$. (Possible, because $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ spans V.)

By linearity of T,

$$T(\boldsymbol{v}) = T(c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}) = c_1T(\boldsymbol{x}_1) + \ldots + c_nT(\boldsymbol{x}_n).$$

Armin Straub
astraub@illinois.edu

Important geometric examples

We consider some linear maps $\mathbb{R}^2 \to \mathbb{R}^2$, which are defined by matrix multiplication, that is, by $\boldsymbol{x} \mapsto A\boldsymbol{x}$.

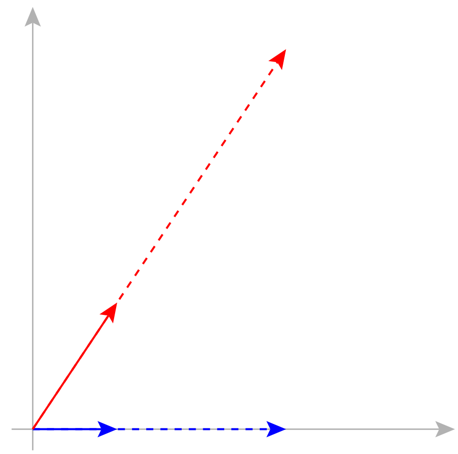
In fact: all linear maps $\mathbb{R}^n \to \mathbb{R}^m$ are given by $\boldsymbol{x} \mapsto A\boldsymbol{x}$, for some matrix A.

Example 2.

The matrix $A = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$

... gives the map $\boldsymbol{x} \mapsto c\boldsymbol{x}$, i.e.

... stretches every vector in $\mathbb{R}^2$ by the same factor c.

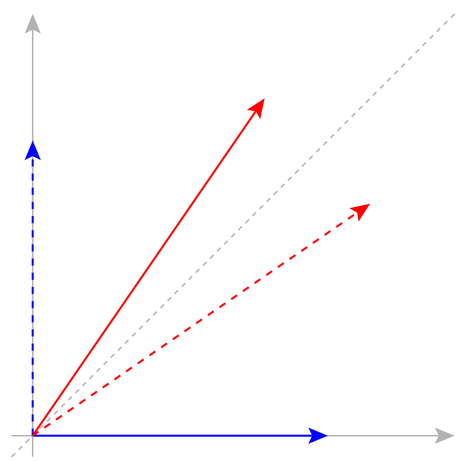


Example 3.

The matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

... gives the map $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} y \\ x \end{bmatrix}$, i.e.

... reflects every vector in $\mathbb{R}^2$ through the line $y = x$.

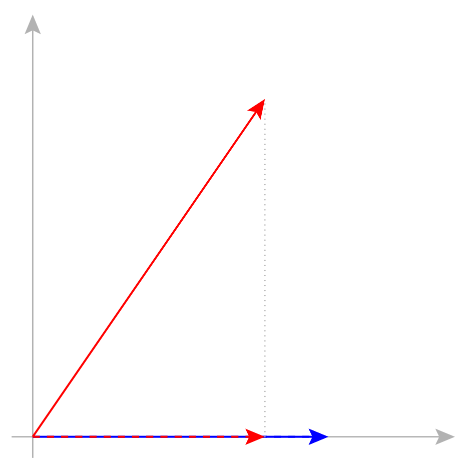


Example 4.

The matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

... gives the map $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ 0 \end{bmatrix}$, i.e.

... projects every vector in $\mathbb{R}^2$ through onto the x-axis.



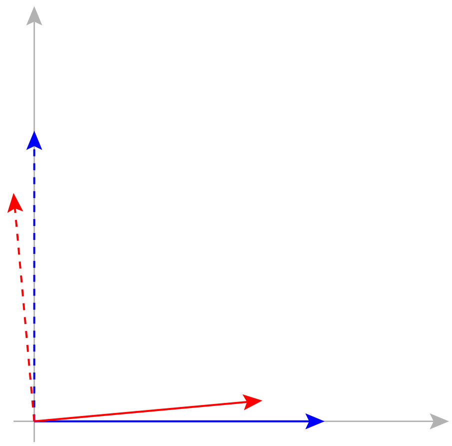
Armin Straub
astraub@illinois.edu

Example 5.

The matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

... gives the map $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} -y \\ x \end{bmatrix}$, i.e.

... rotates every vector in $\mathbb{R}^2$ counter-clockwise by $90°$.



Representing linear maps by matrices

Definition 6. (From linear maps to matrices)

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a basis for V, and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ a basis for W.

The **matrix representing** T with respect to these bases

- has n columns (one for each of the $\boldsymbol{x}_j$),
- the j-th column has m entries $a_{1,j}, \ldots, a_{m,j}$ determined by

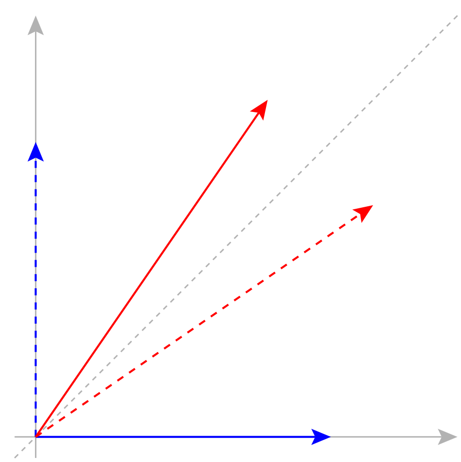
$$T(\boldsymbol{x}_j) = a_{1,j}\boldsymbol{y}_1 + \ldots + a_{m,j}\boldsymbol{y}_m.$$

Example 7.

Recall the map T given by $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} y \\ x \end{bmatrix}$.

(reflects every vector in $\mathbb{R}^2$ through the line $y = x$)

- Which matrix A represents T with respect to the standard bases?
- Which matrix B represents T with respect to the basis $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$?



Solution.

- $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Hence, $A = \begin{bmatrix} 0 & * \\ 1 & * \end{bmatrix}$.

  $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Hence, $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

  If a linear map $T: \mathbb{R}^n \to \mathbb{R}^m$ is represented by the matrix A with respect to the standard bases, then $T(\boldsymbol{x}) = A\boldsymbol{x}$.

Matrix multiplication corresponds to function composition!

That is, if T_1, T_2 are represented by A_1, A_2, then $T_1(T_2(\boldsymbol{x})) = (A_1 A_2)\boldsymbol{x}$.

- $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Hence, $B = \begin{bmatrix} 1 & * \\ 0 & * \end{bmatrix}$.

  $T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Hence, $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Example 8. Let $T: \mathbb{R}^2 \to \mathbb{R}^3$ be the linear map such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix}.$$

What is the matrix B representing T with respect to the following bases?

$$\underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_1}, \underbrace{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}_{\boldsymbol{x}_2} \text{ for } \mathbb{R}^2, \qquad \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\boldsymbol{y}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{y}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{y}_3} \text{ for } \mathbb{R}^3.$$

Solution. This time:

$$\begin{aligned} T(\boldsymbol{x}_1) &= T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 10 \end{bmatrix} \\ &= 5\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ \Longrightarrow B &= \begin{bmatrix} 5 & * \\ -3 & * \\ 5 & * \end{bmatrix} \\ T(\boldsymbol{x}_2) &= T\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right) = -T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + 2T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= -\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ -2 \\ 11 \end{bmatrix} \\ &= 7\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 9\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ \Longrightarrow B &= \begin{bmatrix} 5 & 7 \\ -3 & -9 \\ 5 & 4 \end{bmatrix} \end{aligned}$$

can you see it?
otherwise: do it!

Tedious, even in this simple example! (But we can certainly do it.)

A matrix representing T encodes in column j the coefficients of $T(\boldsymbol{x}_j)$ expressed as a linear combination of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$.

Practice problems

Example 9. Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the map which rotates a vector counter-clockwise by angle θ.

- Which matrix A represents T with respect to the standard bases?
- Verify that $T(\boldsymbol{x}) = A\boldsymbol{x}$.

Solution. Only keep reading if you need a hint!

The first basis vector $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ gets send to $\left[\begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix}\right]$.

Hence, the first column of A is $\ldots$

Review

- A **linear map** $T: V \to W$ satisfies $T(c\boldsymbol{x} + d\boldsymbol{y}) = cT(\boldsymbol{x}) + dT(\boldsymbol{y})$.
- $T: \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\boldsymbol{x}) = A\boldsymbol{x}$ is linear. (A an $m \times n$ matrix)
  - A is the matrix representing T w.r.t. the standard bases

    For instance: $T(\boldsymbol{e}_1) = A\boldsymbol{e}_1 = 1^{\text{st}}$ column of A $\qquad \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$
- Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a basis for V, and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ a basis for W.
  - The matrix representing T w.r.t. these bases encodes in column j the coefficients of $T(\boldsymbol{x}_j)$ expressed as a linear combination of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$.
  - For instance: let $T: \mathbb{R}^3 \to \mathbb{R}^3$ be reflection through the x-y-plane, that is, $(x, y, z) \mapsto (x, y, -z)$.

    The matrix representing T w.r.t. the basis $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & -1 \end{bmatrix}$.

    $T\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 2\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

    $T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$

Example 1. Let $T: \mathbb{P}_3 \to \mathbb{P}_2$ be the linear map given by

$$T(p(t)) = \frac{\mathrm{d}}{\mathrm{d}t} p(t).$$

What is the matrix A representing T with respect to the standard bases?

Solution. The bases are

$$1,\ t,\ t^2,\ t^3 \text{ for } \mathbb{P}_3, \qquad 1,\ t,\ t^2 \text{ for } \mathbb{P}_2.$$

The matrix A has 4 columns and 3 rows.

Armin Straub
astraub@illinois.edu

The first column encodes $T(1) = 0$ and hence is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

For the second column, $T(t) = 1$ and hence it is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

For the third column, $T(t^2) = 2t$ and hence it is $\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$.

For the last column, $T(t^3) = 3t^2$ and hence it is $\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$.

In conclusion, the matrix representing T is

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

Note: By the way, what is the null space of A?

The null space has basis $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. The corresponding polynomial is $p(t) = 1$.

No surprise here: differentation kills precisely the constant polynomials.

Note: Let us differentiate $7t^3 - t + 3$ using the matrix A.

- First: $7t^3 - t + 3$ w.r.t. standard basis: $\begin{bmatrix} 3 \\ -1 \\ 0 \\ 7 \end{bmatrix}$.
- $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 21 \end{bmatrix}$
- $\begin{bmatrix} -1 \\ 0 \\ 21 \end{bmatrix}$ in the standard basis is $-1 + 21t^2$.

Orthogonality

The inner product and distances

Definition 2. The **inner product** (or **dot product**) of $\boldsymbol{v}$, $\boldsymbol{w}$ in $\mathbb{R}^n$:

$$\boldsymbol{v} \cdot \boldsymbol{w} \;=\; \boldsymbol{v}^T \boldsymbol{w} = v_1 w_1 + \ldots + v_n w_n.$$

Because we can think of this as a special case of the matrix product, it satisfies the basic rules like associativity and distributivity.

In addition: $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$.

Example 3. For instance,

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} \;=\; \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} = 1 - 2 - 6 = -7.$$

Definition 4.

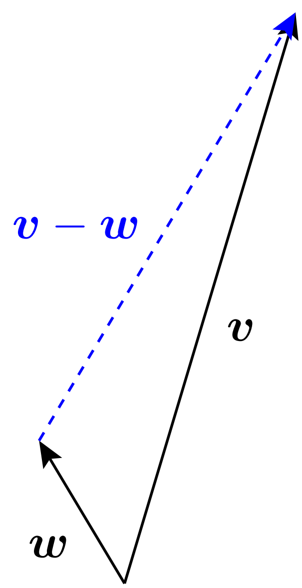
- The **norm** (or **length**) of a vector $\boldsymbol{v}$ in $\mathbb{R}^n$ is

$$\|\boldsymbol{v}\| \;=\; \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + \ldots + v_n^2}.$$

This is the distance to the origin.

- The **distance** between points $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^n$ is

$$\operatorname{dist}(\boldsymbol{v}, \boldsymbol{w}) \;=\; \|\boldsymbol{v} - \boldsymbol{w}\|.$$



Example 5. For instance, in $\mathbb{R}^2$,

$$\operatorname{dist}\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \right) \;=\; \left\| \begin{bmatrix} x_1 - x_2 \\ y_1 - y_2 \end{bmatrix} \right\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Armin Straub
astraub@illinois.edu

Orthogonal vectors

Definition 6. $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^n$ are **orthogonal** if

$$\boldsymbol{v} \cdot \boldsymbol{w} = 0.$$

How is this related to our understanding of right angles?

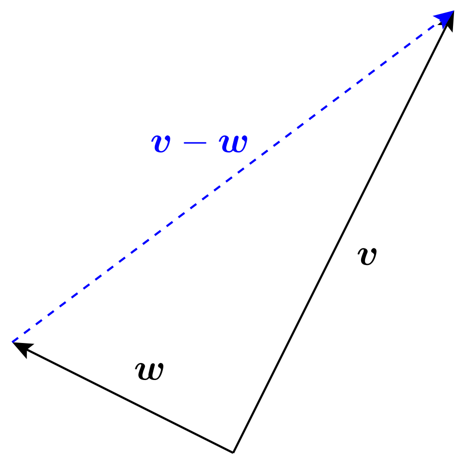
Pythagoras:

$\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal

$\iff \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2$

$\iff \boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{w} \cdot \boldsymbol{w} = \underbrace{(\boldsymbol{v} - \boldsymbol{w}) \cdot (\boldsymbol{v} - \boldsymbol{w})}_{\boldsymbol{v} \cdot \boldsymbol{v} - 2\boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{w}}$

$\iff \boldsymbol{v} \cdot \boldsymbol{w} = 0$



Example 7. Are the following vectors orthogonal?

(a) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \end{bmatrix} = 1 \cdot (-2) + 2 \cdot 1 = 0$. So, yes, they are orthogonal.

(b) $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$

$\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = 1 \cdot (-2) + 2 \cdot 1 + 1 \cdot 1 = 1$. So not orthogonal.

Review

- $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^T \boldsymbol{w} = v_1 w_1 + \ldots + v_n w_n$, the **inner product** of $\boldsymbol{v}$, $\boldsymbol{w}$ in $\mathbb{R}^n$
  - **Length** of $\boldsymbol{v}$: $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + \ldots + v_n^2}$
  - **Distance** between points $\boldsymbol{v}$ and $\boldsymbol{w}$: $\|\boldsymbol{v} - \boldsymbol{w}\|$
- $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^n$ are **orthogonal** if $\boldsymbol{v} \cdot \boldsymbol{w} = 0$.
  - This simple criterion is equivalent to Pythagoras theorem.

Example 1. The vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

- are orthogonal to each other, and
- have length 1.

We are going to call such a basis **orthonormal** soon.

Theorem 2. Suppose that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are nonzero and (pairwise) orthogonal. Then $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are independent.

Proof. Suppose that

$$c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n = \mathbf{0}.$$

Take the dot product of $\boldsymbol{v}_1$ with both sides:

$$\begin{aligned} 0 &= \boldsymbol{v}_1 \cdot (c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n) \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 + c_2 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 + \ldots + c_n \boldsymbol{v}_1 \cdot \boldsymbol{v}_n \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 = c_1 \|\boldsymbol{v}_1\|^2 \end{aligned}$$

But $\|\boldsymbol{v}_1\| \neq 0$ and hence $c_1 = 0$.

Likewise, we find $c_2 = 0$, ..., $c_n = 0$. Hence, the vectors are independent. □

Example 3. Let us consider $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$.

Find $\mathrm{Nul}(A)$ and $\mathrm{Col}(A^T)$. Observe!

Solution.

$\mathrm{Nul}(A) = \mathrm{span}\left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$ can you see it? if not, do it!

$\mathrm{Col}(A^T) = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$

The two basis vectors are orthogonal! $\begin{bmatrix} -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 0$

Armin Straub
astraub@illinois.edu

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

Solution.

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \\ 3 & 6 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & -2 \\ 0 & 0 & -3 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$\operatorname{Nul}(A) = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$

$\operatorname{Col}(A^T) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ The 2 vectors form a basis.

Again, the vectors are orthogonal!

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = 0, \quad \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0.$$

Note: Because $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ is orthogonal to both basis vectors, it is orthogonal to every vector in the row space.

Vectors in $\operatorname{Nul}(A)$ are orthogonal to vectors in $\operatorname{Col}(A^T)$.

The fundamental theorem, second act

Definition 5. Let W be a subspace of $\mathbb{R}^n$, and $\boldsymbol{v}$ in $\mathbb{R}^n$.

- $\boldsymbol{v}$ is **orthogonal** to W, if $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ for all $\boldsymbol{w}$ in W.

  ($\Longleftrightarrow$ $\boldsymbol{v}$ is orthogonal to each vector in a basis of W)

- Another subspace V is **orthogonal** to W, if every vector in V is orthogonal to W.
- The **orthogonal complement** of W is the space $W^\perp$ of all vectors that are orthogonal to W.

  Exercise: show that the orthogonal complement is indeed a vector space.

Example 6. In the previous example, $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \\ 3 & 6 & 0 \end{bmatrix}$.

We found that

$$\operatorname{Nul}(A) = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad \operatorname{Col}(A^T) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

are orthogonal subspaces.

Indeed, $\mathrm{Nul}(A)$ and $\mathrm{Col}(A^T)$ are orthogonal complements.

Why? Because $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ are orthogonal, hence independent, and hence a basis of all of $\mathbb{R}^3$.

Remark 7. Recall that, for an $m \times n$ matrix A, $\mathrm{Nul}(A)$ lives in $\mathbb{R}^n$ and $\mathrm{Col}(A)$ lives in $\mathbb{R}^m$. Hence, they cannot be related in a similar way.

In the previous example, they happen to be both subspaces of $\mathbb{R}^3$:

$$\mathrm{Nul}(A) = \mathrm{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad \mathrm{Col}(A) = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

But these spaces are not orthogonal: $\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \neq 0$

Theorem 8. (Fundamental Theorem of Linear Algebra, Part I)

Let A be an $m \times n$ matrix of rank r.

- $\dim \mathrm{Col}(A) = r$ (subspace of $\mathbb{R}^m$)
- $\dim \mathrm{Col}(A^T) = r$ (subspace of $\mathbb{R}^n$)
- $\dim \mathrm{Nul}(A) = n - r$ (subspace of $\mathbb{R}^n$)
- $\dim \mathrm{Nul}(A^T) = m - r$ (subspace of $\mathbb{R}^m$)

Theorem 9. (Fundamental Theorem of Linear Algebra, Part II)

- $\mathrm{Nul}(A)$ is orthogonal to $\mathrm{Col}(A^T)$. (both subspaces of $\mathbb{R}^n$)

  Note that $\dim \mathrm{Nul}(A) + \dim \mathrm{Col}(A^T) = n$.

  Hence, the two spaces are orthogonal complements in $\mathbb{R}^n$.

- $\mathrm{Nul}(A^T)$ is orthogonal to $\mathrm{Col}(A)$.

  Again, the two spaces are orthogonal complements.

Review

- $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^n$ are **orthogonal** if $\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + \ldots + v_n w_n = 0$.
  - This simple criterion is equivalent to Pythagoras' theorem.
  - Nonzero orthogonal vectors are independent.
- $\operatorname{Nul}\left(\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}\right) = \operatorname{span}\left\{\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right\}$, $\operatorname{Col}\left(\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}^T\right) = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$
- *Fundamental Theorem of Linear Algebra*: (A an $m \times n$ matrix)
  - $\operatorname{Nul}(A)$ is orthogonal to $\operatorname{Col}(A^T)$. (both subspaces of $\mathbb{R}^n$)

    Moreover, $\underbrace{\dim \operatorname{Col}(A^T)}_{= r \text{ (rank of } A)} + \underbrace{\dim \operatorname{Nul}(A)}_{= n - r} = n$

    Hence, they are **orthogonal complements** in $\mathbb{R}^n$.
  - $\operatorname{Nul}(A^T)$ and $\operatorname{Col}(A)$ are orthogonal complements. (in $\mathbb{R}^m$)

$\operatorname{Nul}(A)$ is orthogonal to $\operatorname{Col}(A^T)$.

Why? Suppose that $\boldsymbol{x}$ is in $\operatorname{Nul}(A)$. That is, $A\boldsymbol{x} = \boldsymbol{0}$.
But think about what $A\boldsymbol{x} = \boldsymbol{0}$ means (row-column rule).
It means that the inner product of every row with $\boldsymbol{x}$ is zero.
But that implies that $\boldsymbol{x}$ is orthogonal to the row space.

Example 1. Find all vectors orthogonal to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

Solution. (FTLA, no thinking) In other words:

find the orthogonal complement of $\operatorname{Col}\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}\right)$.

FTLA: this is $\operatorname{Nul}\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}^T\right) = \operatorname{Nul}\left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}\right)$,

which has basis: $\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$.

$\operatorname{span}\left\{\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}\right\}$ are the vectors orthogonal to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

Armin Straub
astraub@illinois.edu

Solution. (a little thinking) The FTLA is not magic!

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \boldsymbol{x} = 0 \text{ and } \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \cdot \boldsymbol{x} = 0 \iff \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\iff \boldsymbol{x} \text{ in } \operatorname{Nul}\left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}\right)$$

This is the same null space we obtained from the FTLA.

Example 2. Let $V = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a + b = 2c \right\}$.

Find a basis for the orthogonal complement of V.

Solution. (FTLA, no thinking) We note that $V = \operatorname{Nul}([\,1 \;\; 1 \;\; -2\,])$.

FTLA: the orthogonal complement is $\operatorname{Col}([\,1 \;\; 1 \;\; -2\,]^T)$.

Basis for the orthogonal complement: $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$

Solution. (a little thinking) $a + b = 2c \iff \begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = 0$.

So: V is actually defined as the orthogonal complement of $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}$.

A new perspective on $A\boldsymbol{x} = \boldsymbol{b}$

$A\boldsymbol{x} = \boldsymbol{b}$ is solvable

$\iff$ $\boldsymbol{b}$ is in $\operatorname{Col}(A)$ ("direct" approach)

$\iff$ $\boldsymbol{b}$ is orthogonal to $\operatorname{Nul}(A^T)$ ("indirect" approach)

The indirect approach means: if $\boldsymbol{y}^T A = \boldsymbol{0}$ then $\boldsymbol{y}^T \boldsymbol{b} = 0$.

Example 3. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}$. For which $\boldsymbol{b}$ does $A\boldsymbol{x} = \boldsymbol{b}$ have a solution?

Solution. (old)

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

So, $A\boldsymbol{x} = \boldsymbol{b}$ is consistent $\iff -3b_1 + b_2 + b_3 = 0$.

Solution. (new) $A\boldsymbol{x} = \boldsymbol{b}$ solvable $\iff$ $\boldsymbol{b}$ orthogonal to $\mathrm{Nul}(A^T)$

to find $\mathrm{Nul}(A^T)$: $\begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix}$

We conclude that $\mathrm{Nul}(A^T)$ has basis $\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$.

$A\boldsymbol{x} = \boldsymbol{b}$ is solvable $\iff \boldsymbol{b} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$. As above!

Motivation

Example 4. Not all linear systems have solutions.

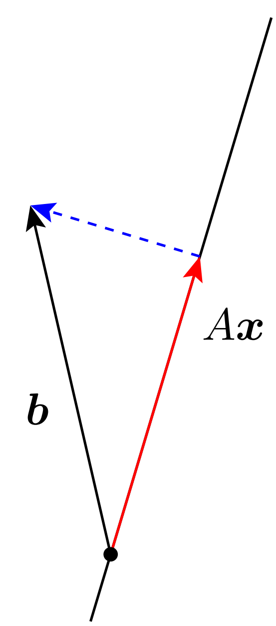
In fact, for many applications, data needs to be fitted and there is no hope for a perfect match.

For instance, $A\boldsymbol{x} = \boldsymbol{b}$ with

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

has no solution:

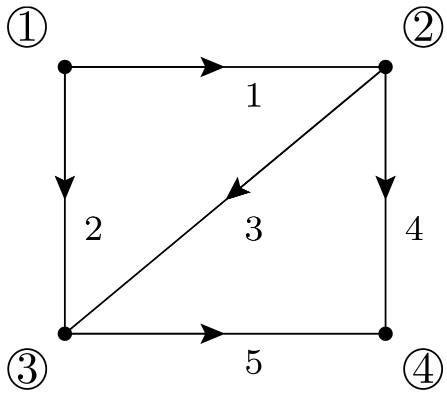
- $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ is not in $\mathrm{Col}(A) = \mathrm{span}\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$
- Instead of giving up, we want the $\boldsymbol{x}$ which makes $A\boldsymbol{x}$ and $\boldsymbol{b}$ as close as possible.



- Such $\boldsymbol{x}$ is characterized by $A\boldsymbol{x}$ being **orthogonal** to the error $\boldsymbol{b} - A\boldsymbol{x}$ (see picture!)

Armin Straub
astraub@illinois.edu

Application: directed graphs



- Graphs appear in network analysis (e.g. internet) or circuit analysis.
- arrow indicates direction of flow
- no edges from a node to itself
- at most one edge between nodes

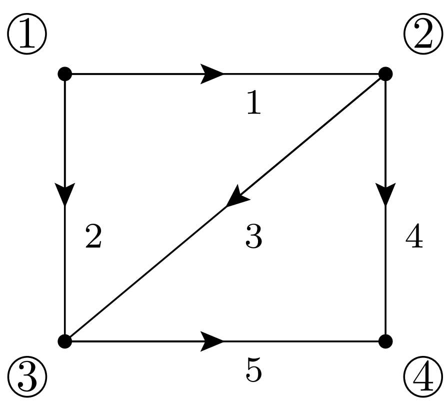
Definition 5. Let G be a graph with m edges and n nodes.

The **edge-node incidence matrix** of G is the $m \times n$ matrix A with

$$A_{i,j} = \begin{cases} -1, & \text{if edge } i \text{ leaves node } j, \\ +1, & \text{if edge } i \text{ enters node } j, \\ 0, & \text{otherwise.} \end{cases}$$

Example 6. Give the edge-node incidence matrix of our graph.

Solution.



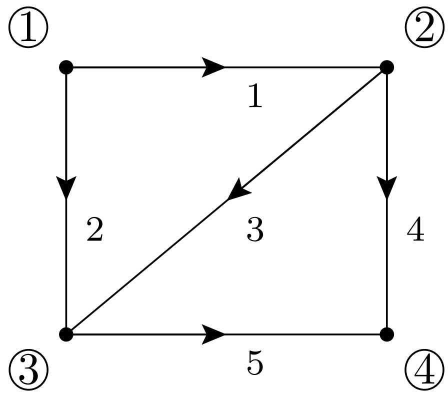
$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

- each column represents a node
- each row represents an edge

Armin Straub
astraub@illinois.edu

Review

- A graph G can be encoded by the **edge-node incidence matrix**:



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

  - each column represents a node
  - each row represents an edge

- If G has m edges and n nodes, then A is the $m \times n$ matrix with

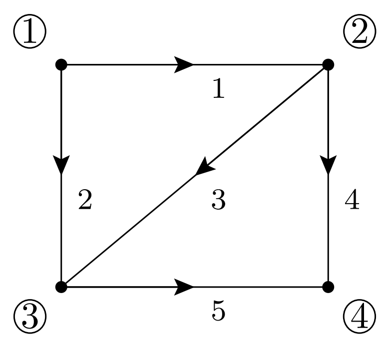
$$A_{i,j} = \begin{cases} -1, & \text{if edge } i \text{ leaves node } j, \\ +1, & \text{if edge } i \text{ enters node } j, \\ 0, & \text{otherwise.} \end{cases}$$

Meaning of the null space

The $\boldsymbol{x}$ in $A\boldsymbol{x}$ is assigning values to each node.

You may think of assigning **potentials** to each node.

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -x_1 + x_2 \\ -x_1 + x_3 \\ -x_2 + x_3 \\ -x_2 + x_4 \\ -x_3 + x_4 \end{bmatrix}$$



So: $A\boldsymbol{x} = \boldsymbol{0}$

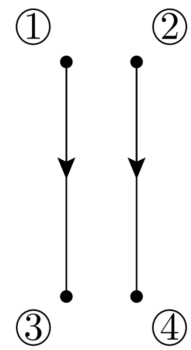
$\iff$ nodes connected by an edge are assigned the same value

Armin Straub
astraub@illinois.edu

For our graph: $\mathrm{Nul}(A)$ has basis $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

This always happens as long as the graph is **connected**.

Example 1. Give a basis for $\mathrm{Nul}(A)$ for the following graph.



Solution. If $A\boldsymbol{x} = \mathbf{0}$ then $x_1 = x_3$ (connected by edge) and $x_2 = x_4$ (connected by edge).

$\mathrm{Nul}(A)$ has the basis: $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$.

Just to make sure: the edge-node incidence matrix is:
$A = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$

In general:
$\dim \mathrm{Nul}(A)$ is the number of connected subgraphs.

For large graphs, disconnection may not be apparent visually.
But we can always find out by computing $\dim \mathrm{Nul}(A)$ using Gaussian elimination!

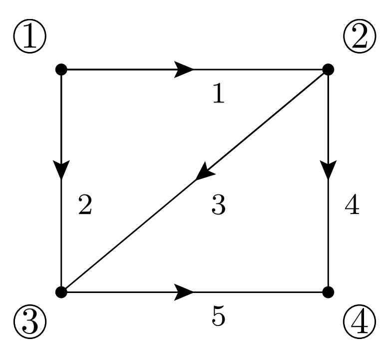
Meaning of the left null space

The $\boldsymbol{y}$ in $\boldsymbol{y}^T A$ is assigning values to each edge.

You may think of assigning **currents** to each edge.

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad A^T = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} -y_1 - y_2 \\ y_1 - y_3 - y_4 \\ y_2 + y_3 - y_5 \\ y_4 + y_5 \end{bmatrix}$$



Armin Straub
astraub@illinois.edu

So: $A^T \boldsymbol{y} = \mathbf{0}$

$\Longleftrightarrow$ at each node, (directed) values assigned to edges add to zero

When thinking of currents, this is **Kirchhoff's first law**.
(at each node, incoming and outgoing currents balance)

What is the simplest way to balance current?

Assign the current in a **loop**!

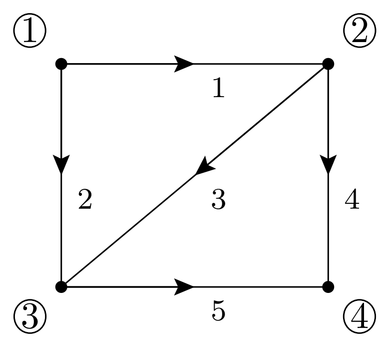
Here, we have two loops: $\text{edge}_1, \text{edge}_3, -\text{edge}_2$ and $\text{edge}_3, \text{edge}_5, -\text{edge}_4$.

Correspondingly, $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ are in $\text{Nul}(A^T)$. Check!

Example 2. Suppose we did not "see" this.

Let us solve $A^T \boldsymbol{y} = \mathbf{0}$ for our graph:

$$\begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



The parametric solution is $\begin{bmatrix} y_3 - y_5 \\ -y_3 + y_5 \\ y_3 \\ -y_5 \\ y_5 \end{bmatrix}$.

So, a basis for $\text{Nul}(A^T)$ is $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$.

Observe that these two basis vectors correspond to loops.

Note that we get the "simpler" loop $\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ as $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$.

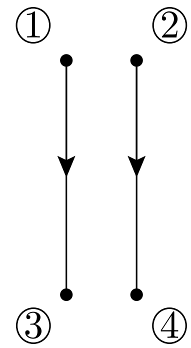
In general:
$\dim \text{Nul}(A^T)$ is the number of (independent) loops.

For large graphs, we now have a nice way to computationally find all loops.

Armin Straub
astraub@illinois.edu

Practice problems

Example 3. Give a basis for $\mathrm{Nul}(A^T)$ for the following graph.



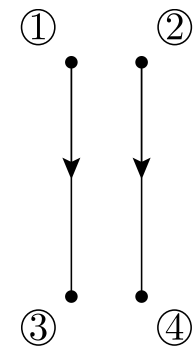
Example 4. Consider the graph with edge-node incidence matrix

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

(a) Draw the corresponding directed graph with numbered edges and nodes.

(b) Give a basis for $\mathrm{Nul}(A)$ and $\mathrm{Nul}(A^T)$ using properties of the graph.

Solutions to practice problems



Example 5. Give a basis for $\mathrm{Nul}(A^T)$ for the following graph.

Solution. This graph contains no loops,

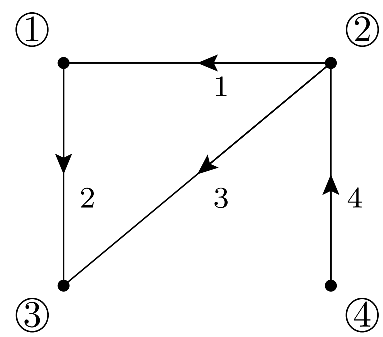
so $\mathrm{Nul}(A^T) = \{\mathbf{0}\}$.

$\mathrm{Nul}(A^T)$ has the empty set as basis (no basis vectors needed).

For comparison: the edge-node incidence matrix

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

indeed has $\mathrm{Nul}(A^T) = \{\mathbf{0}\}$.



Example 6. Consider the graph with edge-node incidence matrix

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

Give a basis for $\mathrm{Nul}(A)$ and $\mathrm{Nul}(A^T)$.

Solution.

If $A\boldsymbol{x} = \mathbf{0}$, then $x_1 = x_2 = x_3 = x_4$ (all connected by edges).

$\mathrm{Nul}(A)$ has the basis: $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

The graph is connected, so only 1 connected subgraph and $\dim \mathrm{Nul}(A) = 1$.

The graph has one loop: $\mathrm{edge}_1, \mathrm{edge}_2, -\mathrm{edge}_3$

Assign values $y_1 = 1$, $y_2 = 1$, $y_3 = -1$ along the edges of that loop.

$\mathrm{Nul}(A^T)$ has the basis: $\begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}$.

The graph has 1 loop, so $\dim \mathrm{Nul}(A^T) = 1$.

Armin Straub
astraub@illinois.edu

Review for Midterm 2

- As of yet unconfirmed:
  - final exam on Friday, December 12, 7–10pm
  - conflict exam on Monday, December 15, 7–10pm

Directed graphs

- Go from directed graph to edge-node incidence matrix A and vice versa.
- Basis for $\mathrm{Nul}(A)$ from connected subgraphs.

  For each connected subgraph, get a basis vector $\boldsymbol{x}$ that assigns 1 to all nodes in that subgraph, and 0 to all other nodes.

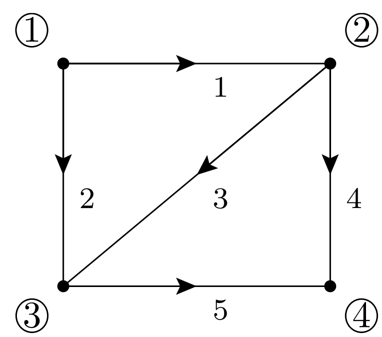
- Basis for $\mathrm{Nul}(A^T)$ from (independent) loops.

  For each (independent) loop, get a basis vector $\boldsymbol{y}$ that assigns 1 and -1 (depending on direction) to the edges in that loop, and 0 to all other edges.

Example 1.

Basis for $\mathrm{Nul}(A)$: $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

Basis for $\mathrm{Nul}(A^T)$: $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$



Fundamental notions

- Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are **independent** if the only linear relation

  $$c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n = \boldsymbol{0}$$

  is the one with $c_1 = c_2 = \ldots = c_n = 0$.

  How to check for independence?

  The columns of a matrix A are independent $\Longleftrightarrow$ $\mathrm{Nul}(A) = \{\boldsymbol{0}\}$.

- Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in V are a **basis** for V if
  - they span V, that is $V = \mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, and
  - they are independent.

  In that case, V has **dimension** n.

- Vectors $\boldsymbol{v}, \boldsymbol{w}$ in $\mathbb{R}^m$ are **orthogonal** if $\boldsymbol{v} \cdot \boldsymbol{w} = v_1w_1 + \ldots + v_mw_m = 0$.

Armin Straub
astraub@illinois.edu

Subspaces

- From an echelon form of A, we get bases for:
  - $\operatorname{Nul}(A)$ — by solving $A\boldsymbol{x} = \mathbf{0}$
  - $\operatorname{Col}(A)$ — by taking the pivot columns of A
  - $\operatorname{Col}(A^T)$ — by taking the nonzero rows of the echelon form

Example 2.

$$A = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 4 & -1 & 3 \\ 3 & 6 & 2 & 22 \\ 4 & 8 & 0 & 16 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Basis for $\operatorname{Col}(A)$: $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \end{bmatrix}$

Basis for $\operatorname{Col}(A^T)$: $\begin{bmatrix} 1 \\ 2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 5 \end{bmatrix}$

Basis for $\operatorname{Nul}(A)$: $\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ -5 \\ 1 \end{bmatrix}$

Dimension of $\operatorname{Nul}(A^T)$: 2

- The solutions to $A\boldsymbol{x} = \boldsymbol{b}$ are given by $\boldsymbol{x}_p + \operatorname{Nul}(A)$.
- The **fundamental theorem** states that
  - $\operatorname{Nul}(A)$ and $\operatorname{Col}(A^T)$ are orthogonal complements

    So: $\dim \operatorname{Nul}(A) + \dim \operatorname{Col}(A^T) = n$ (number of columns of A)
  - $\operatorname{Nul}(A^T)$ and $\operatorname{Col}(A)$ are orthogonal complements

    So: $\dim \operatorname{Nul}(A^T) + \dim \operatorname{Col}(A) = m$ (number of rows of A)
  - In particular, if $r = \operatorname{rank}(A)$ (nr of pivots):
    - $\dim \operatorname{Col}(A) = r$
    - $\dim \operatorname{Col}(A^T) = r$
    - $\dim \operatorname{Nul}(A) = n - r$
    - $\dim \operatorname{Nul}(A^T) = m - r$

Example 3. Consider the following subspaces of $\mathbb{R}^4$:

(a) $V = \left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : a + 2b = 0,\ a + b + d = 0 \right\}$

Armin Straub
astraub@illinois.edu

(b) $V = \left\{ \begin{bmatrix} a+b-c \\ b \\ 2a+3c \\ c \end{bmatrix} : \ a, b, c \text{ in } \mathbb{R} \right\}$

In each case, give a basis for V and its orthogonal complement.

Try to immediately get an idea what the dimensions are going to be!

Solution.

- First step: express these subspaces as one of the four subspaces of a matrix.

  (a) $V = \operatorname{Nul}\left(\begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \right)$

  (b) $V = \operatorname{Col}\left(\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 2 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix} \right)$

- Give a basis for each.

  (a) row reductions: $\begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \end{bmatrix}$

  basis for V: $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$

  (b) row reductions: $\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 2 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & -2 & 5 \\ 0 & 0 & 1 \end{bmatrix}$

  (no need to continue; we already see that the columns are independent)

  basis for V: $\begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 3 \\ 1 \end{bmatrix}$

- Use the fundamental theorem to find bases for the orthogonal complements.

  (a) $V^\perp = \operatorname{Col}\left(\begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}^T \right)$

  note the two rows are clearly independent.

  basis for $V^\perp$: $\begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$

  (b) $V^\perp = \operatorname{Nul}\left(\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 2 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix}^T \right) = \operatorname{Nul}\left(\begin{bmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 3 & 1 \end{bmatrix} \right)$

  row reductions: $\begin{bmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 3 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 5 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & -2/5 \\ 0 & 1 & 0 & 2/5 \\ 0 & 0 & 1 & 1/5 \end{bmatrix}$

  basis for $V^\perp$: $\begin{bmatrix} 2/5 \\ -2/5 \\ -1/5 \\ 1 \end{bmatrix}$

Armin Straub
astraub@illinois.edu

Example 4. What does it mean for $A\boldsymbol{x} = \boldsymbol{b}$ if $\mathrm{Nul}(A) = \{\mathbf{0}\}$?

Solution. It means that if there is a solution, then it is unique.

That's because all solutions to $A\boldsymbol{x} = \boldsymbol{b}$ are given by $\boldsymbol{x}_p + \mathrm{Nul}(A)$.

Linear transformations

Example 5. Let $T: \mathbb{R}^2 \to \mathbb{R}^3$ be the linear map represented by the matrix

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 0 \end{bmatrix}$$

with respect to the bases $\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ of $\mathbb{R}^2$ and $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ of $\mathbb{R}^3$.

(a) What is $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$?

(b) Which matrix represents T with respect to the standard bases?

Solution.

The matrix tells us that:

$$T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 1\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

$$T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = 0\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

(a) Note that $\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Hence, $T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = 2T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) + T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = 2\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 11 \end{bmatrix}$.

(b) Note that $\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Hence, $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) + T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$.

We already know that $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$.

So, T is represented by $\begin{bmatrix} 4 & 3 \\ 4 & 4 \\ 6 & 5 \end{bmatrix}$ with respect to the standard bases.

Check your understanding

Think about why each of these statements is true!

- $A\boldsymbol{x} = \boldsymbol{b}$ has a solution $\boldsymbol{x}$ if and only if $\boldsymbol{b}$ is in $\mathrm{Col}(A)$.

  That's because $A\boldsymbol{x}$ are linear combinations of the columns of A.

- A and A^T have the same rank.

  Recall that the rank of A (number of pivots of A) equals $\dim \mathrm{Col}(A)$.

  So this is another way of saying that $\dim \mathrm{Col}(A) = \dim \mathrm{Col}(A^T)$.

- The columns of an $n \times n$ matrix are independent if and only if the rows are.

  Let r be the rank of A, and let A be $m \times n$ for now.

  The columns are independent $\Longleftrightarrow$ $r = n$ (so that $\dim \mathrm{Nul}(A) = 0$).

  But also: the rows are independent $\Longleftrightarrow$ $r = m$.

  In the case $m = n$, these two conditions are equivalent.

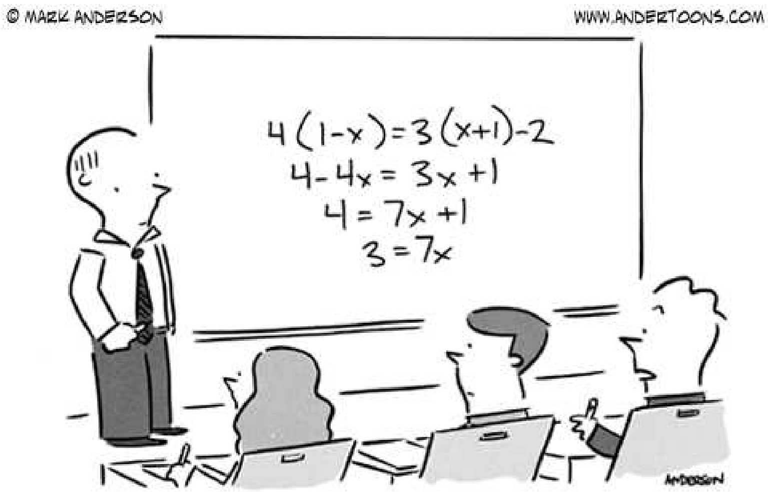
- $A\boldsymbol{x} = \boldsymbol{b}$ has a solution $\boldsymbol{x}$ if and only if $\boldsymbol{b}$ is orthogonal to $\mathrm{Nul}(A^T)$.

  This follows from "$A\boldsymbol{x} = \boldsymbol{b}$ has a solution $\boldsymbol{x}$ if and only if $\boldsymbol{b}$ is in $\mathrm{Col}(A)$"
  together with the fundamental theorem, which says that $\mathrm{Col}(A)$ is the orthogonal complement of $\mathrm{Nul}(A^T)$.

- The rows of A are independent if and only if $\mathrm{Nul}(A^T) = \{\mathbf{0}\}$.

  Recall that elements of $\mathrm{Nul}(A)$ correspond to linear relations between the columns of A.

  Likewise, elements of $\mathrm{Nul}(A^T)$ correspond to linear relations between the rows of A.



"Wouldn't it be more efficient to just find who's complicating equations and ask them to stop?"

- We can deal with "complicated" linear systems, but what to do if there is no solutions and we want a "best" approximate solution?

  This is important for many applications, including fitting data.

- Suppose $A\boldsymbol{x} = \boldsymbol{b}$ has no solution. This means $\boldsymbol{b}$ is not in $\mathrm{Col}(A)$.

  Idea: find "best" approximate solution by replacing $\boldsymbol{b}$ with its projection onto $\mathrm{Col}(A)$.

- Recall: if $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are (pairwise) orthogonal:

$$\boldsymbol{v}_1 \cdot (c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n) \;=\; c_1\boldsymbol{v}_1 \cdot \boldsymbol{v}_1$$

  Implies: the $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are independent (unless one is the zero vector)

Orthogonal bases

Definition 1. A basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of a vector space V is an **orthogonal basis** if the vectors are (pairwise) orthogonal.

Example 2. The standard basis $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is an orthogonal basis for $\mathbb{R}^3$.

Example 3. Are the vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ an orthogonal basis for $\mathbb{R}^3$?

Solution.

Armin Straub
astraub@illinois.edu

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0$$
$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0$$
$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0$$

So this is an orthogonal basis.

Note that we do not need to check that the three vectors are independent. That follows from their orthogonality.

Example 4. Suppose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is an orthogonal basis of V, and that $\boldsymbol{w}$ is in V. Find $c_1, \ldots, c_n$ such that

$$\boldsymbol{w} = c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n.$$

Solution. Take the dot product of $\boldsymbol{v}_1$ with both sides:

$$\begin{aligned} \boldsymbol{v}_1 \cdot \boldsymbol{w} &= \boldsymbol{v}_1 \cdot (c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n) \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 + c_2 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 + \ldots + c_n \boldsymbol{v}_1 \cdot \boldsymbol{v}_n \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 \end{aligned}$$

Hence, $c_1 = \dfrac{\boldsymbol{v}_1 \cdot \boldsymbol{w}}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}$. In general, $c_j = \dfrac{\boldsymbol{v}_j \cdot \boldsymbol{w}}{\boldsymbol{v}_j \cdot \boldsymbol{v}_j}$.

If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is an orthogonal basis of V, and $\boldsymbol{w}$ is in V, then

$$\boldsymbol{w} = c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n \qquad \text{with} \quad c_j = \frac{\boldsymbol{w} \cdot \boldsymbol{v}_j}{\boldsymbol{v}_j \cdot \boldsymbol{v}_j}.$$

Example 5. Express $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ in terms of the basis $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Solution.

$$\begin{aligned} \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} &= c_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \frac{-4}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{10}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{4}{1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

Armin Straub
astraub@illinois.edu

Definition 6. A basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of a vector space V is an **orthonormal basis** if the vectors are orthogonal and have length 1.

Example 7. The standard basis $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is an orthonormal basis for $\mathbb{R}^3$.

If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is an orthonormal basis of V, and $\boldsymbol{w}$ is in V, then

$$\boldsymbol{w} = c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n \qquad \text{with} \quad c_j = \boldsymbol{v}_j \cdot \boldsymbol{w}.$$

Example 8. Express $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ in terms of the basis $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Solution. That's trivial, of course:

$$\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 7 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

But note that the coefficients are

$$\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 3, \quad \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 7, \quad \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 4.$$

Example 9. Is the basis $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ orthonormal? If not, normalize the vectors to produce an orthonormal basis.

Solution.

$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} = \sqrt{2} \implies$ normalized: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} = \sqrt{2} \implies$ normalized: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}} = 1 \implies$ is already normalized: $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

The corresponding orthonormal basis is $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Armin Straub
astraub@illinois.edu

Example 10. Express $\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ in terms of the basis $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

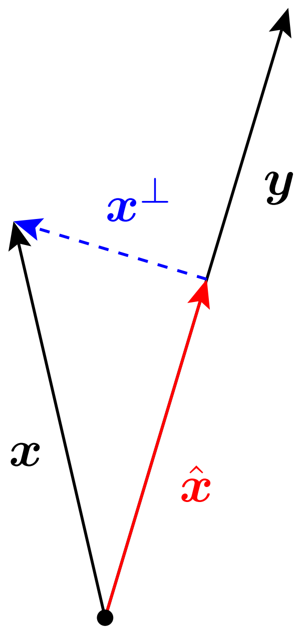
Solution.

$$\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \frac{-4}{\sqrt{2}}, \quad \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{10}{\sqrt{2}}, \quad \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 4.$$

Hence, just as in Example 5:

$$\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} = \frac{-4}{\sqrt{2}} \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{10}{\sqrt{2}} \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Orthogonal projections



Definition 11. The **orthogonal projection** of vector $\boldsymbol{x}$ onto vector $\boldsymbol{y}$ is

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\boldsymbol{y} \cdot \boldsymbol{y}} \boldsymbol{y}.$$

- The vector $\hat{\boldsymbol{x}}$ is the closest vector to $\boldsymbol{x}$, which is in $\operatorname{span}\{\boldsymbol{y}\}$.
- Characterized by: the "error" $\boldsymbol{x}^{\perp} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ is orthogonal to $\operatorname{span}\{\boldsymbol{y}\}$.
- To find the formula for $\hat{\boldsymbol{x}}$, start with $\hat{\boldsymbol{x}} = c\boldsymbol{y}$.

  $(\boldsymbol{x} - \hat{\boldsymbol{x}}) \cdot \boldsymbol{y} = (\boldsymbol{x} - c\boldsymbol{y}) \cdot \boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} - c\boldsymbol{y} \cdot \boldsymbol{y} \overset{\text{wanted}}{=} 0$

  It follows that $c = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\boldsymbol{y} \cdot \boldsymbol{y}}$.

$\boldsymbol{x}^{\perp}$ is also called the **component of $\boldsymbol{x}$ orthogonal to $\boldsymbol{y}$**.

Example 12. What is the orthogonal projection of $\boldsymbol{x} = \begin{bmatrix} -8 \\ 4 \end{bmatrix}$ onto $\boldsymbol{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$?

Solution.

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\boldsymbol{y} \cdot \boldsymbol{y}} \boldsymbol{y} = \frac{-8 \cdot 3 + 4 \cdot 1}{3^2 + 1^2}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = -2\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -6 \\ -2 \end{bmatrix}$$

The component of $\boldsymbol{x}$ orthogonal to $\boldsymbol{y}$ is

$$\boldsymbol{x} - \hat{\boldsymbol{x}} = \begin{bmatrix} -8 \\ 4 \end{bmatrix} - \begin{bmatrix} -6 \\ -2 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}.$$

(Note that, indeed $\begin{bmatrix} -2 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ are orthogonal.)

Armin Straub
astraub@illinois.edu

Example 13. What are the orthogonal projections of $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ onto each of the vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$?

Solution.

$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ on $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$: $\frac{2 \cdot 1 + 1 \cdot (-1) + 1 \cdot 0}{1^2 + (-1)^2 + 0^2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ on $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$: $\frac{2 \cdot 1 + 1 \cdot 1 + 1 \cdot 0}{1^2 + 1^2 + 0^2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ on $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$: $\frac{2 \cdot 0 + 1 \cdot 0 + 1 \cdot 1}{0^2 + 0^2 + 1^2} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Note that these sum up to $\frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$!

That's because the three vectors are an orthogonal basis for $\mathbb{R}^3$.

Recall: If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is an orthogonal basis of V, and $\boldsymbol{w}$ is in V, then

$$\boldsymbol{w} = c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n \qquad \text{with} \quad c_j = \frac{\boldsymbol{w} \cdot \boldsymbol{v}_j}{\boldsymbol{v}_j \cdot \boldsymbol{v}_j}.$$

$\rightsquigarrow$ $\boldsymbol{w}$ decomposes as the sum of its projections onto each basis vector

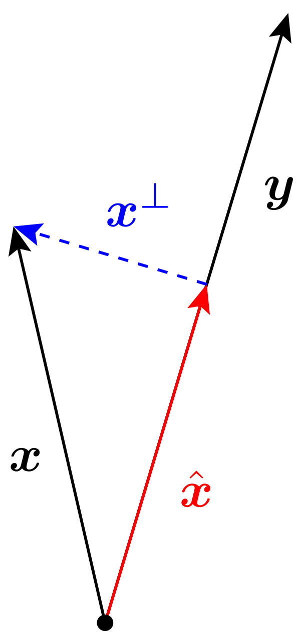
Comments on midterm

Suppose V is a vector space, and you are asked to give a basis.

- CORRECT: V has basis $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$
- CORRECT: V has basis $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$
- OK: $V = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$

  (but you really should point out that the two vectors are independent)
- INCORRECT: $V = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$
- INCORRECT: $\text{basis} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

Review



- **Orthogonal projection** of $\boldsymbol{x}$ onto $\boldsymbol{y}$:

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\boldsymbol{y} \cdot \boldsymbol{y}} \boldsymbol{y}.$$

  "Error" $\boldsymbol{x}^{\perp} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ is orthogonal to $\boldsymbol{y}$.

- If $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ is an **orthogonal basis** of V, and $\boldsymbol{x}$ is in V, then

$$\boldsymbol{x} = c_1 \boldsymbol{y}_1 + \ldots + c_n \boldsymbol{y}_n \qquad \text{with} \quad c_j = \frac{\boldsymbol{x} \cdot \boldsymbol{y}_j}{\boldsymbol{y}_j \cdot \boldsymbol{y}_j}.$$

$\boldsymbol{x}$ decomposes as the sum of its projections onto each vector in the orthogonal basis.

Example 1. Express $\underbrace{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}}_{\boldsymbol{x}}$ in terms of the basis $\underbrace{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}_{\boldsymbol{y}_1}, \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{y}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{y}_3}$.

Solution. Note that $\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3$ is an orthogonal basis of $\mathbb{R}^3$.

Armin Straub
astraub@illinois.edu

$$\begin{bmatrix}2\\1\\1\end{bmatrix} = c_1\begin{bmatrix}1\\-1\\0\end{bmatrix} + c_2\begin{bmatrix}1\\1\\0\end{bmatrix} + c_3\begin{bmatrix}0\\0\\1\end{bmatrix}$$

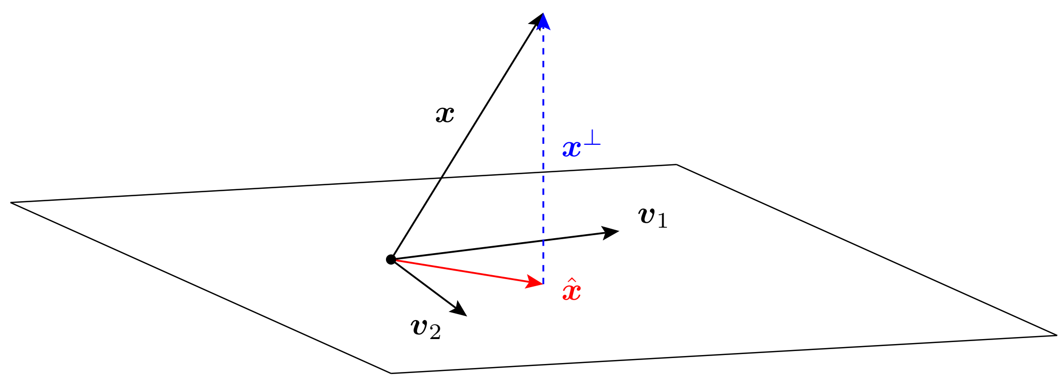
$$= \underbrace{\frac{\begin{bmatrix}2\\1\\1\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\0\end{bmatrix}}{\begin{bmatrix}1\\-1\\0\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\0\end{bmatrix}}\begin{bmatrix}1\\-1\\0\end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{y}_1} + \underbrace{\frac{\begin{bmatrix}2\\1\\1\end{bmatrix}\cdot\begin{bmatrix}1\\1\\0\end{bmatrix}}{\begin{bmatrix}1\\1\\0\end{bmatrix}\cdot\begin{bmatrix}1\\1\\0\end{bmatrix}}\begin{bmatrix}1\\1\\0\end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{y}_2} + \underbrace{\frac{\begin{bmatrix}2\\1\\1\end{bmatrix}\cdot\begin{bmatrix}0\\0\\1\end{bmatrix}}{\begin{bmatrix}0\\0\\1\end{bmatrix}\cdot\begin{bmatrix}0\\0\\1\end{bmatrix}}\begin{bmatrix}0\\0\\1\end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{y}_3}$$

$$= \frac{1}{2}\begin{bmatrix}1\\-1\\0\end{bmatrix} + \frac{3}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} + \begin{bmatrix}0\\0\\1\end{bmatrix}$$

Orthogonal projection on subspaces

Theorem 2. Let W be a subspace of $\mathbb{R}^n$. Then, each $\boldsymbol{x}$ in $\mathbb{R}^n$ can be uniquely written as

$$\boldsymbol{x} = \underbrace{\hat{\boldsymbol{x}}}_{\text{in } W} + \underbrace{\boldsymbol{x}^\perp}_{\text{in } W^\perp}.$$



- $\hat{\boldsymbol{x}}$ is the **orthogonal projection** of $\boldsymbol{x}$ onto W.
  $\hat{\boldsymbol{x}}$ is the point in W closest to $\boldsymbol{x}$. For any other $\boldsymbol{y}$ in W, $\operatorname{dist}(\boldsymbol{x}, \hat{\boldsymbol{x}}) < \operatorname{dist}(\boldsymbol{x}, \boldsymbol{y})$.
- If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ is an orthogonal basis of W, then

$$\hat{\boldsymbol{x}} = \left(\frac{\boldsymbol{x}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\right)\boldsymbol{v}_1 + \ldots + \left(\frac{\boldsymbol{x}\cdot\boldsymbol{v}_m}{\boldsymbol{v}_m\cdot\boldsymbol{v}_m}\right)\boldsymbol{v}_m.$$

  Once $\hat{\boldsymbol{x}}$ is determined, $\boldsymbol{x}^\perp = \boldsymbol{x} - \hat{\boldsymbol{x}}$.
  (This is also the orthogonal projection of $\boldsymbol{x}$ onto $W^\perp$.)

Example 3. Let $W = \operatorname{span}\left\{\begin{bmatrix}3\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$, and $\boldsymbol{x} = \begin{bmatrix}0\\3\\10\end{bmatrix}$.

- Find the orthogonal projection of $\boldsymbol{x}$ onto W.
  (or: find the vector in W which is closest to $\boldsymbol{x}$)
- Write $\boldsymbol{x}$ as a vector in W plus a vector orthogonal to W.

Armin Straub
astraub@illinois.edu

Solution.

Note that $\boldsymbol{w}_1 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$ and $\boldsymbol{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ are an orthogonal basis for W.

[We will soon learn how to construct orthogonal bases ourselves.]

Hence, the orthogonal projection of $\boldsymbol{x}$ onto W is:

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} \cdot \boldsymbol{w}_1}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} \boldsymbol{w}_1 + \frac{\boldsymbol{x} \cdot \boldsymbol{w}_2}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2} \boldsymbol{w}_2 = \frac{\begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \frac{10}{10} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$$

$\hat{\boldsymbol{x}}$ is the vector in W which best approximates $\boldsymbol{x}$.

Orthogonal projection of $\boldsymbol{x}$ onto the orthogonal complement of W:

$\boldsymbol{x}^{\perp} = \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 9 \end{bmatrix}$. Hence, $\boldsymbol{x} = \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix} = \underbrace{\begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}}_{\text{in } W} + \underbrace{\begin{bmatrix} -3 \\ 0 \\ 9 \end{bmatrix}}_{\text{in } W^{\perp}}$.

Note: Indeed, $\begin{bmatrix} -3 \\ 0 \\ 9 \end{bmatrix}$ is orthogonal to $\boldsymbol{w}_1 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$ and $\boldsymbol{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Definition 4. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ be an orthogonal basis of W, a subspace of $\mathbb{R}^n$. Note that the projection map $\pi_W: \mathbb{R}^n \to \mathbb{R}^n$, given by

$$\boldsymbol{x} \mapsto \hat{\boldsymbol{x}} = \left(\frac{\boldsymbol{x} \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \right) \boldsymbol{v}_1 + \ldots + \left(\frac{\boldsymbol{x} \cdot \boldsymbol{v}_m}{\boldsymbol{v}_m \cdot \boldsymbol{v}_m} \right) \boldsymbol{v}_m$$

is linear. The matrix P representing π_W with respect to the standard basis is the corresponding **projection matrix**.

Example 5. Find the projection matrix P which corresponds to orthogonal projection onto $W = \operatorname{span}\left\{ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ in $\mathbb{R}^3$.

Solution. Standard basis : $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

The first column of P encodes the projection of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$:

$\frac{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{3}{10} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$. Hence $P = \begin{bmatrix} \frac{9}{10} & * & * \\ 0 & * & * \\ \frac{3}{10} & * & * \end{bmatrix}$.

Armin Straub
astraub@illinois.edu

The second column of P encodes the projection of $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$:

$$\frac{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$ Hence $P = \begin{bmatrix} \frac{9}{10} & 0 & * \\ 0 & 1 & * \\ \frac{3}{10} & 0 & * \end{bmatrix}$.

The third column of P encodes the projection of $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$:

$$\frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}.$$ Hence $P = \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix}$.

Example 6. (again)

Find the orthogonal projection of $\boldsymbol{x} = \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix}$ onto $W = \operatorname{span}\left\{ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$.

Solution. $\hat{\boldsymbol{x}} = P\boldsymbol{x} = \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 10 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$, as in the previous example.

Example 7. Compute P^2 for the projection matrix we just found. Explain!

Solution.

$$\begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix} \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix} = \begin{bmatrix} \frac{9}{10} & 0 & \frac{3}{10} \\ 0 & 1 & 0 \\ \frac{3}{10} & 0 & \frac{1}{10} \end{bmatrix}$$

Projecting a second time does not change anything anymore.

Practice problems

Example 8. Find the closest point to $\boldsymbol{x}$ in $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$, where

$$\boldsymbol{x} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -2 \end{bmatrix}, \quad \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

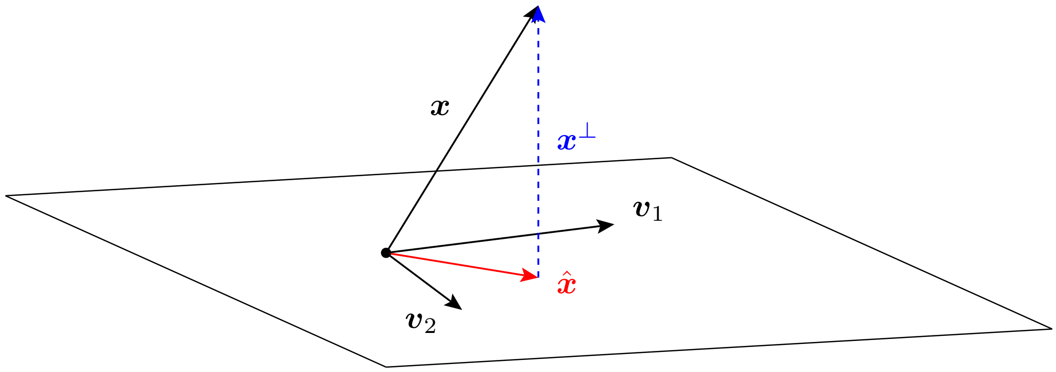
Solution. This is the orthogonal projection of $\boldsymbol{x}$ onto $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$.

...

...

Review

Let W be a subspace of $\mathbb{R}^n$, and $\boldsymbol{x}$ in $\mathbb{R}^n$ (but maybe not in W).



Let $\hat{\boldsymbol{x}}$ be the **orthogonal projection** of $\boldsymbol{x}$ onto W.

(vector in W as close as possible to $\boldsymbol{x}$)

- If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ is an orthogonal basis of W, then

$$\hat{\boldsymbol{x}} = \underbrace{\left(\frac{\boldsymbol{x}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\right)\boldsymbol{v}_1}_{\text{proj. of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_1} + \ldots + \underbrace{\left(\frac{\boldsymbol{x}\cdot\boldsymbol{v}_m}{\boldsymbol{v}_m\cdot\boldsymbol{v}_m}\right)\boldsymbol{v}_m}_{\text{proj. of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_m}.$$

- The decomposition $\boldsymbol{x} = \underbrace{\hat{\boldsymbol{x}}}_{\text{in } W} + \underbrace{\boldsymbol{x}^\perp}_{\text{in } W^\perp}$ is unique.

Least squares

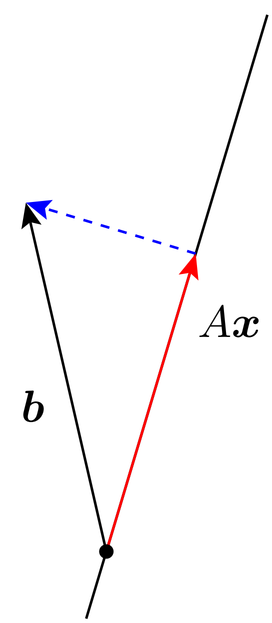
Definition 1. $\hat{\boldsymbol{x}}$ is a **least squares solution** of the system $A\boldsymbol{x} = \boldsymbol{b}$ if $\hat{\boldsymbol{x}}$ is such that $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible.

- If $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, then a least squares solution $\hat{\boldsymbol{x}}$ is just an ordinary solution.

  (in that case, $A\hat{\boldsymbol{x}} - \boldsymbol{b} = \boldsymbol{0}$)

- Interesting case: $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent.

  (in other words: the system is overdetermined)



Idea. $A\boldsymbol{x} = \boldsymbol{b}$ is consistent $\iff$ $\boldsymbol{b}$ is in $\mathrm{Col}(A)$

So, if $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, we

- replace $\boldsymbol{b}$ with its projection $\hat{\boldsymbol{b}}$ onto $\mathrm{Col}(A)$,
- and solve $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$. (consistent by construction!)

Armin Straub
astraub@illinois.edu

Example 2. Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

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

Solution. Note that the columns of A are orthogonal.

[Otherwise, we could not proceed in the same way.]

Hence, the projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\mathrm{Col}(A)$ is

$$\hat{\boldsymbol{b}} = \frac{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

We have already solved $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$ in the process: $\hat{\boldsymbol{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$.

The normal equations

The following result provides a straightforward recipe (thanks to the FTLA) to find least squares solutions for any matrix.

[The previous example was only simple because the columns of A were orthogonal.]

Theorem 3. $\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$
$\iff$ $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ (the **normal equations**)

Proof.

$\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$

$\iff$ $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible

$\iff$ $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is orthogonal to $\mathrm{Col}(A)$

$\overset{\text{FTLA}}{\iff}$ $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is in $\mathrm{Nul}(A^T)$

$\iff$ $A^T(A\hat{\boldsymbol{x}} - \boldsymbol{b}) = \boldsymbol{0}$

$\iff$ $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ □

Example 4. (again) Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

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

Solution.

$$A^T A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$A^T \boldsymbol{b} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

The normal equations $A^T A \hat{\boldsymbol{x}} = A^T \boldsymbol{b}$ are

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \hat{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Solving, we find (again) $\hat{\boldsymbol{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$.

Example 5. Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

What is the projection of $\boldsymbol{b}$ onto $\mathrm{Col}(A)$?

Solution.

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T \boldsymbol{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

The normal equations $A^T A \hat{\boldsymbol{x}} = A^T \boldsymbol{b}$ are

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \hat{\boldsymbol{x}} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

Solving, we find $\hat{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

The projection of $\boldsymbol{b}$ onto $\mathrm{Col}(A)$ is $A\hat{\boldsymbol{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

Just to make sure: why is $A\hat{\boldsymbol{x}}$ the projection of $\boldsymbol{b}$ onto $\mathrm{Col}(A)$?
Because, for a least squares solution $\hat{\boldsymbol{x}}$, $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible.

The projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\mathrm{Col}(A)$ is

$$\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}, \quad \text{with } \hat{\boldsymbol{x}} \text{ such that } A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b}.$$

If A has full column rank, this is (columns of A independent)

$$\hat{\boldsymbol{b}} = A(A^TA)^{-1}A^T\boldsymbol{b}.$$

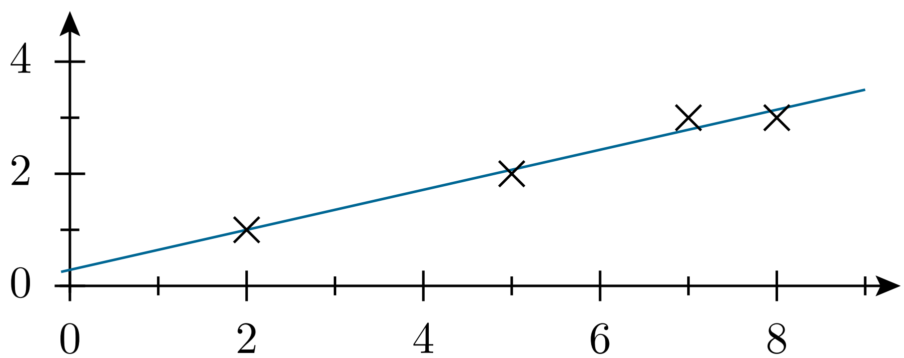
Hence, the projection matrix for projecting onto $\mathrm{Col}(A)$ is

$$P = A(A^TA)^{-1}A^T.$$

Application: least squares lines

Experimental data: (x_i, y_i)

Wanted: parameters β_1, β_2 such that $y_i \approx \beta_1 + \beta_2 x_i$ for all i



This approximation should be so that

$$\underbrace{\mathrm{SS}_{\mathrm{res}} = \sum_i [y_i - (\beta_1 + \beta_2 x_i)]^2}_{\text{residual sum of squares}} \text{ is as small as possible.}$$

Example 6. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2,1)$, $(5,2), (7,3), (8,3)$.

Solution. The equations $y_i = \beta_1 + \beta_2 x_i$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$$

Armin Straub
astraub@illinois.edu

Here, we need to find a least-squares solution to

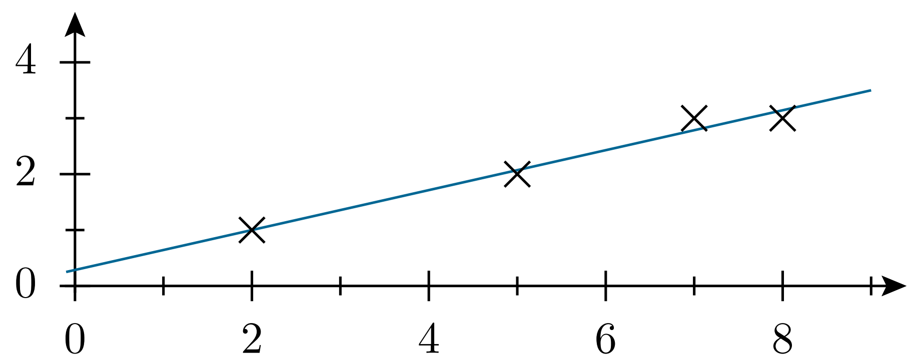
$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

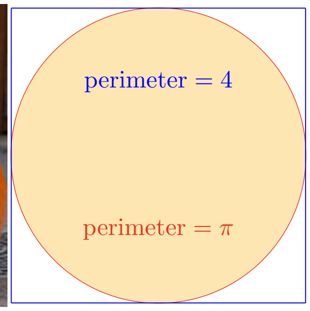
$$X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

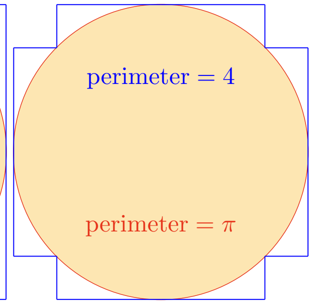
Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$.

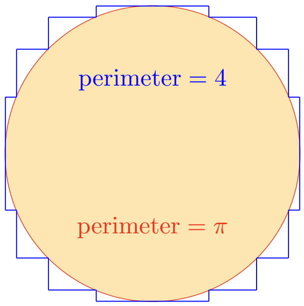
Hence, the least squares line is $y = \frac{2}{7} + \frac{5}{14} x$.

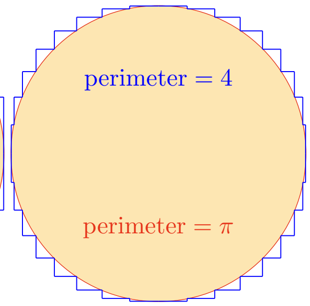


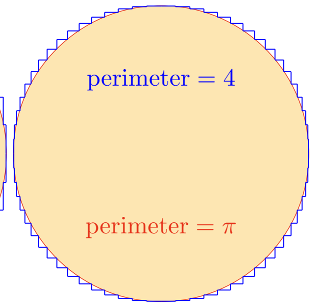
Armin Straub
astraub@illinois.edu

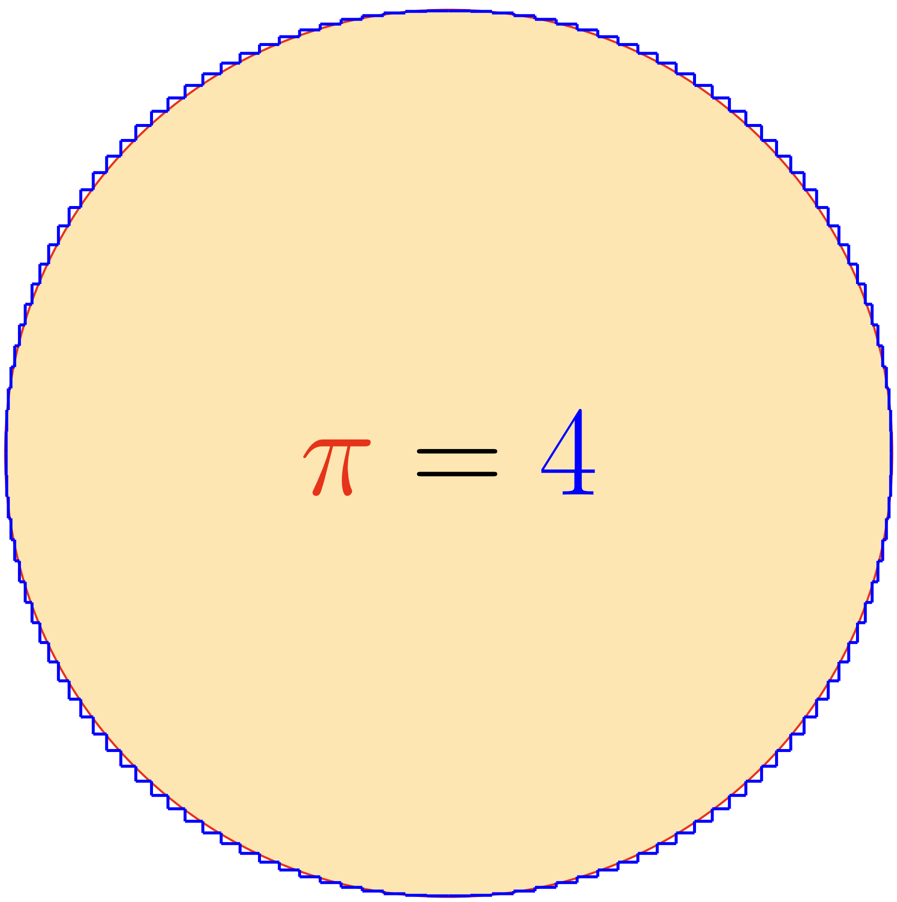












Happy Halloween!

Armin Straub
astraub@illinois.edu

Review

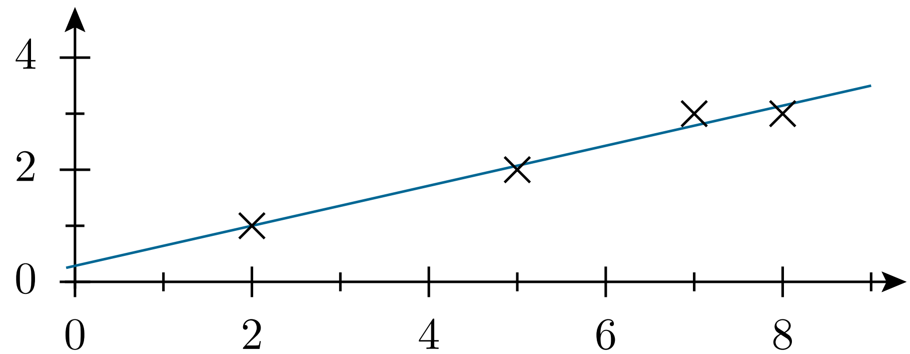
- $\hat{\boldsymbol{x}}$ is a **least squares solution** of the system $A\boldsymbol{x} = \boldsymbol{b}$

  $\Longleftrightarrow$ $\hat{\boldsymbol{x}}$ is such that $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible

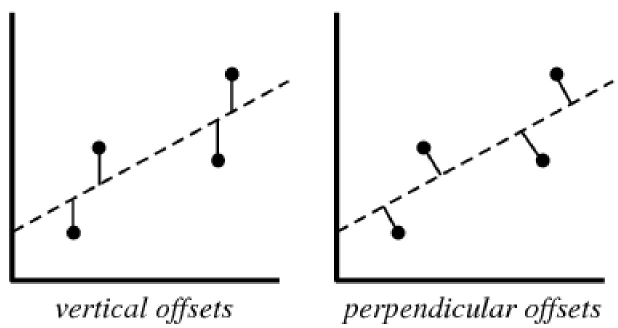
  $\overset{\text{FTLA}}{\Longleftrightarrow}$ $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ (the **normal equations**)

Application: least squares lines

Example 1. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2,1)$, $(5,2), (7,3), (8,3)$.



Comment. As usual in practice, we are minimizing the (sum of squares of the) vertical offsets:



http://mathworld.wolfram.com/LeastSquaresFitting.html

Solution. The equations $y_i = \beta_1 + \beta_2 x_i$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$$

Here, we need to find a least squares solution to

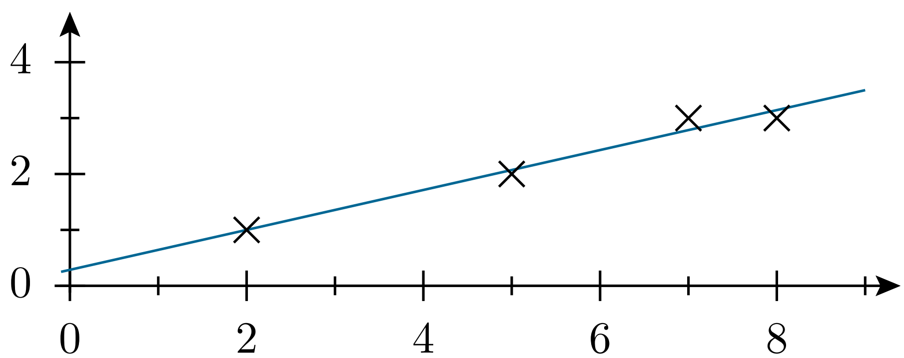
$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

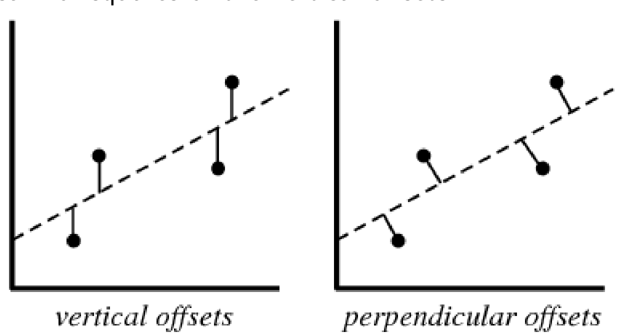
Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$.

Hence, the least squares line is $y = \frac{2}{7} + \frac{5}{14}x$.



How well does the line fit the data $(2,1), (5,2), (7,3), (8,3)$?

How small is the sum of squares of the vertical offsets?



Armin Straub
astraub@illinois.edu

- **residual sum of squares**: $\mathrm{SS}_{\mathrm{res}} = \sum \underbrace{(y_i - (\beta_1 + \beta_2 x_i))^2}_{\text{error at } (x_i, y_i)}$

  The choice of β_1, β_2 from least squares, makes $\mathrm{SS}_{\mathrm{res}}$ as small as possible.

- **total sum of squares**: $\mathrm{SS}_{\mathrm{tot}} = \sum (y_i - \bar{y})^2$,

  where $\bar{y} = \frac{1}{n} \sum y_i$ is the mean of the observed data

- **coefficient of determination**: $R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}}$

  General rule: the closer R^2 is to 1, the better the regression line fits the data.

  Here, $\bar{y} = 9/4$: $(2,1), (5,2), (7,3), (8,3)$

$$R^2 = 1 - \frac{\left(1 - \left(\frac{2}{7} + \frac{5}{14}2\right)\right)^2 + \left(2 - \left(\frac{2}{7} + \frac{5}{14}5\right)\right)^2 + \left(3 - \left(\frac{2}{7} + \frac{5}{14}7\right)\right)^2 + \left(3 - \left(\frac{2}{7} + \frac{5}{14}8\right)\right)^2}{\left(1 - \frac{9}{4}\right)^2 + \left(2 - \frac{9}{4}\right)^2 + \left(3 - \frac{9}{4}\right)^2 + \left(3 - \frac{9}{4}\right)^2}$$

$$= 1 - \frac{0.075}{2.75} = 0.974$$

  very close to $1 \implies$ good fit

Other curves

We can also fit the experimental data (x_i, y_i) using other curves.

Example 2. $y_i \approx \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ with parameters $\beta_1, \beta_2, \beta_3$.

The equations $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$$

Given data (x_i, y_i), we then find the least squares solution to $X\beta = \boldsymbol{y}$.

Multiple linear regression

> *In statistics, **linear regression** is an approach for modeling the relationship between a scalar dependent variable and one or more explanatory variables.*
>
> *The case of one explanatory variable is called simple linear regression.*
>
> *For more than one explanatory variable, the process is called multiple linear regression.*
>
> *http://en.wikipedia.org/wiki/Linear_regression*

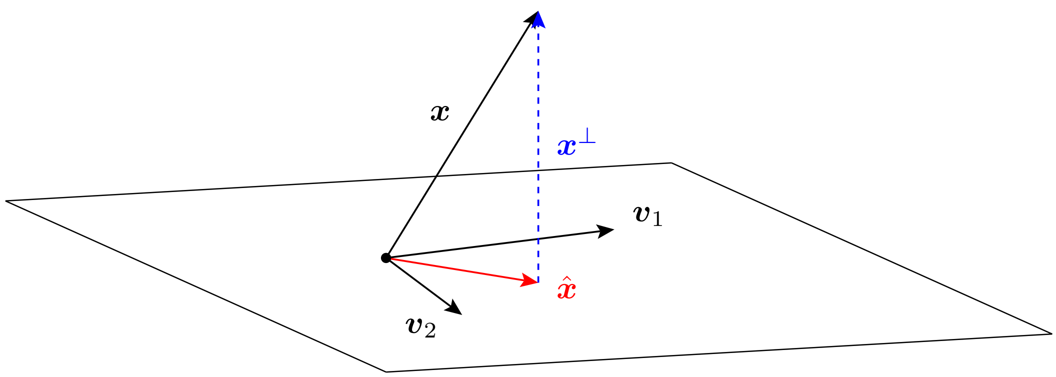
The experimental data might be of the form (v_i, w_i, y_i), where now the dependent variable y_i depends on two explanatory variables v_i, w_i (instead of just one x_i).

Example 3. Fitting a linear relationship $y_i \approx \beta_1 + \beta_2 v_i + \beta_3 w_i$, we get:

$$\underbrace{\begin{bmatrix} 1 & v_1 & w_1 \\ 1 & v_2 & w_2 \\ 1 & v_3 & w_3 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector}}}$$

And we again proceed by finding a least squares solution.

Review



- Suppose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ is an orthonormal basis of W.
  The **orthogonal projection** of $\boldsymbol{x}$ onto W is:

$$\hat{\boldsymbol{x}} = \underbrace{\langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \boldsymbol{v}_1}_{\text{proj. of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_1} + \ldots + \underbrace{\langle \boldsymbol{x}, \boldsymbol{v}_m \rangle \boldsymbol{v}_m}_{\text{proj. of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_m}.$$

(To stay agile, we are writing $\langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle = \boldsymbol{x} \cdot \boldsymbol{v}_1$ for the inner product.)

Gram–Schmidt

Example 4. Find an orthonormal basis for $V = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Recipe. (Gram–Schmidt orthonormalization)

Given a basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, produce an orthonormal basis $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$.

$$\begin{aligned} \boldsymbol{b}_1 &= \boldsymbol{a}_1, & \boldsymbol{q}_1 &= \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} \\ \boldsymbol{b}_2 &= \boldsymbol{a}_2 - \langle \boldsymbol{a}_2, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1, & \boldsymbol{q}_2 &= \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|} \\ \boldsymbol{b}_3 &= \boldsymbol{a}_3 - \langle \boldsymbol{a}_3, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1 - \langle \boldsymbol{a}_3, \boldsymbol{q}_2 \rangle \boldsymbol{q}_2, & \boldsymbol{q}_3 &= \frac{\boldsymbol{b}_3}{\|\boldsymbol{b}_3\|} \\ &\vdots \end{aligned}$$

Example 5. Find an orthonormal basis for $V = \operatorname{span}\left\{ \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} \right\}$.

Solution.

$$\boldsymbol{b}_1 = \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, \qquad \boldsymbol{q}_1 = \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}$$

$$\boldsymbol{b}_2 = \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix} - \left\langle \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 = \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}, \qquad \boldsymbol{q}_2 = \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}$$

$$\boldsymbol{b}_3 = \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} - \left\langle \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 - \left\langle \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix}, \boldsymbol{q}_2 \right\rangle \boldsymbol{q}_2 = \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}, \qquad \boldsymbol{q}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}$$

We have obtained an orthonormal basis for V:

$$\begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}.$$

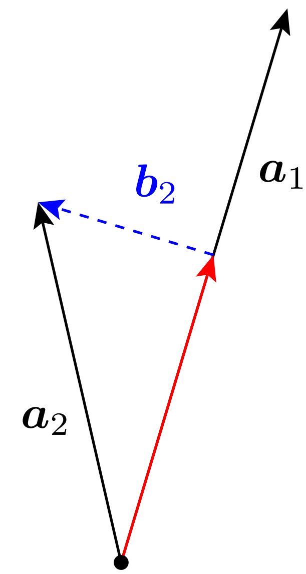
Review

- Vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are orthonormal if
$$\boldsymbol{q}_i^T \boldsymbol{q}_j = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$$
- **Gram–Schmidt** orthonormalization:

Input: basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ for V.

Output: orthonormal basis $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ for V.

$$\begin{aligned} \boldsymbol{b}_1 &= \boldsymbol{a}_1, & \boldsymbol{q}_1 &= \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} \\ \boldsymbol{b}_2 &= \boldsymbol{a}_2 - \langle \boldsymbol{a}_2, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1, & \boldsymbol{q}_2 &= \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|} \\ \boldsymbol{b}_3 &= \boldsymbol{a}_3 - \langle \boldsymbol{a}_3, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1 - \langle \boldsymbol{a}_3, \boldsymbol{q}_2 \rangle \boldsymbol{q}_2, & \boldsymbol{q}_3 &= \frac{\boldsymbol{b}_3}{\|\boldsymbol{b}_3\|} \\ &\vdots \end{aligned}$$



Example 1. Apply Gram–Schmidt to the vectors $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Solution.

$$\boldsymbol{b}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \qquad \boldsymbol{q}_1 = \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{b}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right\rangle \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \qquad \boldsymbol{q}_2 = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$$

$$\boldsymbol{b}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 - \left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{q}_2 \right\rangle \boldsymbol{q}_2 = \ldots = \frac{1}{9}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}, \qquad \boldsymbol{q}_3 = \frac{1}{3}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

We obtained the orthonormal vectors $\frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$.

Theorem 2. The columns of an $m \times n$ matrix Q are orthonormal
$\iff$ $Q^T Q = I$ (the $n \times n$ identity)

Proof. Let $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ be the columns of Q.

They are orthonormal if and only if $\boldsymbol{q}_i^T \boldsymbol{q}_j = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$

Armin Straub
astraub@illinois.edu

All these inner products are packaged in $Q^TQ = I$:

$$\begin{bmatrix} — & \boldsymbol{q}_1^T & — \\ — & \boldsymbol{q}_2^T & — \\ & \vdots & \end{bmatrix} \begin{bmatrix} | & | & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

□

Definition 3. An **orthogonal matrix** is a square matrix Q with orthonormal columns.

It is historical convention to restrict to square matrices, and to say orthogonal matrix even though "orthonormal matrix" might be better.

An $n \times n$ matrix Q is orthogonal $\iff$ $Q^TQ = I$

In other words, $Q^{-1} = Q^T$.

Example 4. $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is orthogonal.

In general, all permutation matrices P are orthogonal.

Why? Because their columns are a permutation of the standard basis.

And so we always have $P^TP = I$.

Example 5. $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Q is orthogonal because:

- $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ is an orthonormal basis of $\mathbb{R}^2$

  Just to make sure: why length 1? Because $\left\| \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \right\| = \sqrt{\cos^2\theta + \sin^2\theta} = 1.$

- Alternatively: $Q^TQ = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Example 6. Is $H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ orthogonal?

No, the columns are orthogonal but not normalized.

But $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ is an orthogonal matrix.

Just for fun: a $n \times n$ matrix with entries ± 1 whose columns are orthogonal is called a *Hadamard matrix* of size n.

A size 4 example: $\begin{bmatrix} H & H \\ H & -H \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$

Continuing this construction, we get examples of size $8, 16, 32, \ldots$

It is believed that Hadamard matrices exist for all sizes $4n$.

But no example of size 668 is known yet.

The QR decomposition (flashed at you)

- Gaussian elimination in terms of matrices: $A = LU$
- Gram–Schmidt in terms of matrices: $A = QR$

Let A be an $m \times n$ matrix of rank n. (columns independent)

Then we have the **QR decomposition** $A = QR$,

- where Q is $m \times n$ and has orthonormal columns, and
- R is upper triangular, $n \times n$ and invertible.

Idea: Gram–Schmidt on the columns of A, to get the columns of Q.

Example 7. Find the QR decomposition of $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$.

Solution. We apply Gram–Schmidt to the columns of A:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \boldsymbol{q}_1$$

$$\begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} - \left\langle \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}, \qquad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{q}_2$$

$$\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} - \left\langle \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 - \left\langle \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \boldsymbol{q}_2 \right\rangle \boldsymbol{q}_2 = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \boldsymbol{q}_3$$

Hence: $Q = [\ \boldsymbol{q}_1 \ \ \boldsymbol{q}_2 \ \ \boldsymbol{q}_3\ ] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

To find R in $A = QR$,

note that $Q^T A = Q^T Q R = R$.

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

Summarizing, we have

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

Recipe. In general, to obtain $A = QR$:

- Gram–Schmidt on (columns of) A, to get (columns of) Q.
- Then, $R = Q^T A$.

The resulting R is indeed upper triangular, and we get:

$$\begin{bmatrix} | & | & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} | & | & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots \\ | & | & \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{a}_1 & \boldsymbol{q}_1^T\boldsymbol{a}_2 & \boldsymbol{q}_1^T\boldsymbol{a}_3 & \cdots \\ & \boldsymbol{q}_2^T\boldsymbol{a}_2 & \boldsymbol{q}_2^T\boldsymbol{a}_3 & \\ & & \boldsymbol{q}_3^T\boldsymbol{a}_3 & \\ & & & \ddots \end{bmatrix}$$

It should be noted that, actually, no extra work is needed for computing R: all the inner products in R have been computed during Gram–Schmidt.

(Just like the LU decomposition encodes the steps of Gaussian elimination, the QR decomposition encodes the steps of Gram–Schmidt.)

Practice problems

Example 8. Complete $\frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}$ to an orthonormal basis of $\mathbb{R}^3$.

(a) by using the FTLA to determine the orthogonal complement of the span you already have

(b) by using Gram–Schmidt after throwing in an independent vector such as $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

Example 9. Find the QR decomposition of $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

Review

- Let A be an $m\times n$ matrix of rank n. (columns independent)

  Then we have the **QR decomposition** $A = QR$,

  - where Q is $m\times n$ with orthonormal columns, and
  - R is upper triangular and invertible.

- To obtain

$$\begin{bmatrix} | & | & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} | & | & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots \\ | & | & \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{a}_1 & \boldsymbol{q}_1^T\boldsymbol{a}_2 & \boldsymbol{q}_1^T\boldsymbol{a}_3 & \cdots \\ & \boldsymbol{q}_2^T\boldsymbol{a}_2 & \boldsymbol{q}_2^T\boldsymbol{a}_3 & \\ & & \boldsymbol{q}_3^T\boldsymbol{a}_3 & \\ & & & \ddots \end{bmatrix}$$

  - Gram–Schmidt on (columns of) A, to get (columns of) Q.
  - Then, $R = Q^T A$. (actually unnecessary!)

Example 1. The QR decomposition is also used to solve systems of linear equations. (we assume A is $n\times n$, and A^{-1} exists)

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} \quad &\Longleftrightarrow \quad QR\boldsymbol{x} = \boldsymbol{b} \\ &\Longleftrightarrow \quad R\boldsymbol{x} = Q^T\boldsymbol{b} \end{aligned}$$

The last system is triangular and is solved by back substitution.

QR is a little slower than LU but makes up in numerical stability.

If A is not $n\times n$ and invertible, then $R\boldsymbol{x} = Q^T\boldsymbol{b}$ gives the least squares solutions!

Example 2. The QR decomposition is very useful for solving least squares problems:

$$\begin{aligned} A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b} \quad &\Longleftrightarrow \quad \underbrace{(QR)^TQR}_{=R^TQ^TQR}\hat{\boldsymbol{x}} = (QR)^T\boldsymbol{b} \\ &\Longleftrightarrow \quad R^TR\hat{\boldsymbol{x}} = R^TQ^T\boldsymbol{b} \\ &\Longleftrightarrow \quad R\hat{\boldsymbol{x}} = Q^T\boldsymbol{b} \end{aligned}$$

Again, the last system is triangular and is solved by back substitution.

$\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$
$\Longleftrightarrow \quad R\hat{\boldsymbol{x}} = Q^T\boldsymbol{b}$ (where $A = QR$)

Application: Fourier series

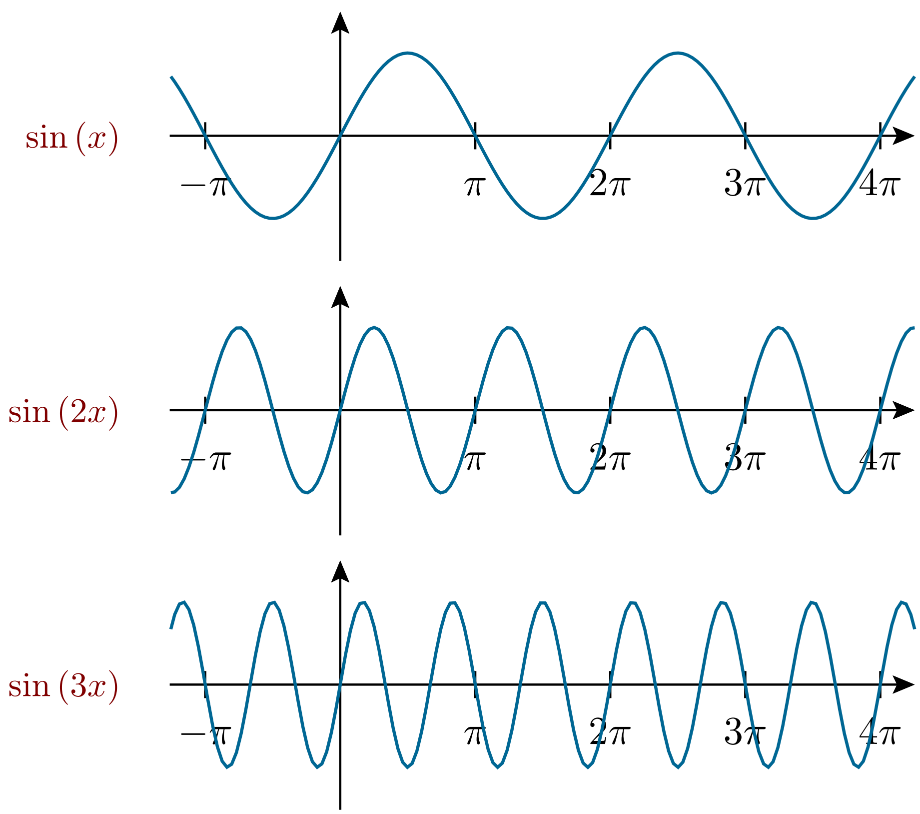
Review. Given an orthogonal basis $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$, we express a vector $\boldsymbol{x}$ as:

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots, \qquad \underbrace{c_i\boldsymbol{v}_i = \frac{\langle \boldsymbol{x}, \boldsymbol{v}_i \rangle}{\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle}\boldsymbol{v}_i}_{} \text{ projection of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_i$$

A **Fourier series** of a function $f(x)$ is an infinite expansion:

$$f(x) = a_0 + a_1\cos(x) + b_1\sin(x) + a_2\cos(2x) + b_2\sin(2x) + \cdots$$

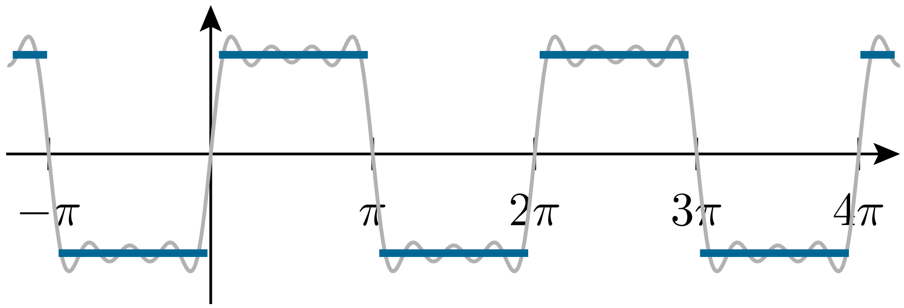
Example 3.



Armin Straub
astraub@illinois.edu

Example 4. (just a preview)

$$\text{blue function} = \frac{4}{\pi}\left(\sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \frac{1}{7}\sin(7x) + \ldots\right)$$



- We are working in the vector space of functions $\mathbb{R} \to \mathbb{R}$.
  - More precisely, "nice" (say, piecewise continuous) functions that have period 2π.
  - These are infinite dimensional vector spaces.
- The functions

$$1, \cos(x), \sin(x), \cos(2x), \sin(2x), \ldots$$

are a basis of this space. In fact, an **orthogonal basis**!

That's the reason for the success of Fourier series.

But what is the inner product on the space of functions?

- Vectors in $\mathbb{R}^n$: $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v_1 w_1 + \ldots + v_n w_n$
- Functions: $\langle f, g \rangle = \int_0^{2\pi} f(x) g(x) \mathrm{d}x$

Why these limits? Because our functions have period 2π.

Example 5. Show that $\cos(x)$ and $\sin(x)$ are orthogonal.

Solution.

$$\langle \cos(x), \sin(x) \rangle = \int_0^{2\pi} \cos(x) \sin(x) \mathrm{d}x = \left[\frac{1}{2}(\sin(x))^2\right]_0^{2\pi} = 0$$

More generally, $1, \cos(x), \sin(x), \cos(2x), \sin(2x), \ldots$ are all orthogonal to each other.

Armin Straub
astraub@illinois.edu

Example 6. What is the norm of $\cos(x)$?

Solution.

$$\langle \cos(x), \cos(x)\rangle = \int_0^{2\pi} \cos(x)\cos(x)\mathrm{d}x = \pi$$

Why? There's many ways to evaluate this integral. For instance:

- you could use integration by parts,
- you could use a trig identity,
- here's a simple way:
  - $\int_0^{2\pi} \cos^2(x)\mathrm{d}x = \int_0^{2\pi} \sin^2(x)\mathrm{d}x$ ($\cos$ and $\sin$ are just a shift apart)
  - $\cos^2(x) + \sin^2(x) = 1$
  - So: $\int_0^{2\pi} \cos^2(x)\mathrm{d}x = \frac{1}{2}\int_0^{2\pi} 1\mathrm{d}x = \pi$

Hence, $\cos(x)$ is not normalized. It has norm $\|\cos(x)\| = \sqrt{\pi}$.

Example 7. The same calculation shows that $\cos(kx)$ and $\sin(kx)$ have norm $\sqrt{\pi}$ as well.

> Fourier series of $f(x)$:
>
> $$f(x) = a_0 + a_1\cos(x) + b_1\sin(x) + a_2\cos(2x) + b_2\sin(2x) + \cdots$$

Example 8. How do we find a_1?

Or: how much cosine is in a function $f(x)$?

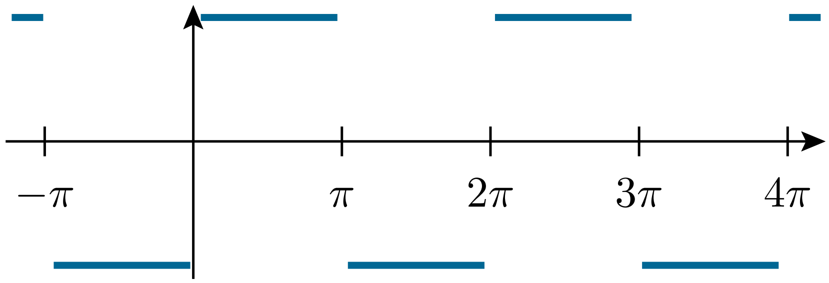
Solution.

$$a_1 = \frac{\langle f(x), \cos(x)\rangle}{\langle \cos(x), \cos(x)\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(x)\mathrm{d}x$$

> $f(x)$ has the Fourier series
>
> $$f(x) = a_0 + a_1\cos(x) + b_1\sin(x) + a_2\cos(2x) + b_2\sin(2x) + \cdots$$
>
> where
>
> $$a_k = \frac{\langle f(x), \cos(kx)\rangle}{\langle \cos(kx), \cos(kx)\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(kx)\mathrm{d}x,$$
> $$b_k = \frac{\langle f(x), \sin(kx)\rangle}{\langle \sin(kx), \sin(kx)\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(kx)\mathrm{d}x,$$
> $$a_0 = \frac{\langle f(x), 1\rangle}{\langle 1, 1\rangle} = \frac{1}{2\pi}\int_0^{2\pi} f(x)\mathrm{d}x.$$

Example 9. Find the Fourier series of the 2π-periodic function $f(x)$ defined by

$$f(x) = \begin{cases} -1, & \text{for } x \in (-\pi, 0), \\ +1, & \text{for } x \in (0, \pi). \end{cases}$$

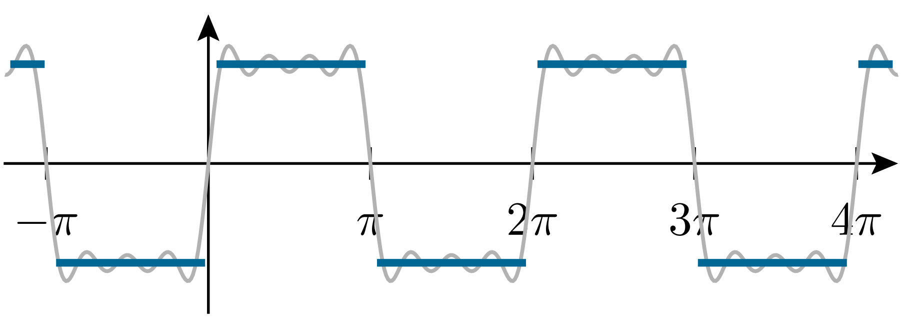


Solution. Note that $\int_0^{2\pi}$ and $\int_{-\pi}^{\pi}$ are the same here. (why?!)

$$\begin{aligned}
a_0 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{d}x = 0 \\
a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\mathrm{d}x \\
&= \frac{1}{\pi}\left[-\int_{-\pi}^{0}\cos(nx)\mathrm{d}x + \int_0^{\pi}\cos(nx)\mathrm{d}x\right] = 0 \\
b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\mathrm{d}x \\
&= \frac{1}{\pi}\left[-\int_{-\pi}^{0}\sin(nx)\mathrm{d}x + \int_0^{\pi}\sin(nx)\mathrm{d}x\right] \\
&= \frac{2}{\pi}\left[\int_0^{\pi}\sin(nx)\mathrm{d}x\right] \\
&= \frac{2}{\pi}\left[-\frac{1}{n}\cos(nx)\right]_0^{\pi} \\
&= \frac{2}{\pi n}[1 - \cos(n\pi)] \\
&= \frac{2}{\pi n}[1 - (-1)^n] = \begin{cases} \frac{4}{\pi n} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}
\end{aligned}$$

In conclusion,

$$f(x) = \frac{4}{\pi}\left(\sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \frac{1}{7}\sin(7x) + \ldots\right).$$



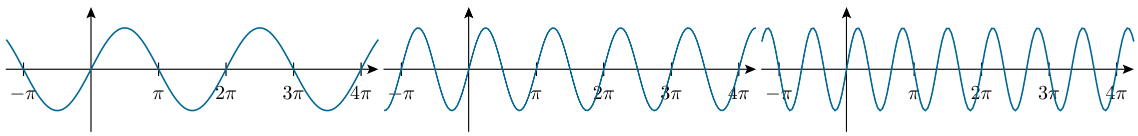
Armin Straub
astraub@illinois.edu

Review

- An inner product for 2π-periodic functions:

  $\langle f, g\rangle = \int_0^{2\pi} f(x) g(x) \mathrm{d}x$ (in $\mathbb{R}^n$: $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = v_1 w_1 + \ldots + v_n w_n$)

- $1, \cos(x), \sin(x), \cos(2x), \sin(2x), \ldots$ are orthogonal



- An expansion in that basis is a **Fourier series**:

$$f(x) = a_0 + a_1 \cos(x) + b_1 \sin(x) + a_2 \cos(2x) + b_2 \sin(2x) + \cdots$$

where

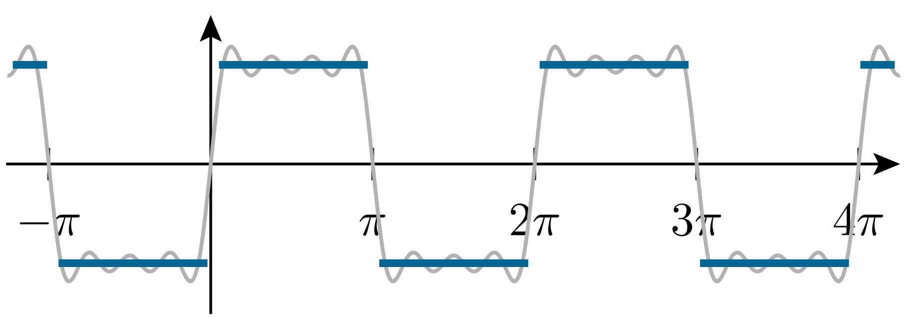
$$a_k = \frac{\langle f(x), \cos(kx)\rangle}{\langle \cos(kx), \cos(kx)\rangle} = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) \mathrm{d}x,$$

$$b_k = \frac{\langle f(x), \sin(kx)\rangle}{\langle \sin(kx), \sin(kx)\rangle} = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) \mathrm{d}x,$$

$$a_0 = \frac{\langle f(x), 1\rangle}{\langle 1, 1\rangle} = \frac{1}{2\pi} \int_0^{2\pi} f(x) \mathrm{d}x.$$

Example 1.

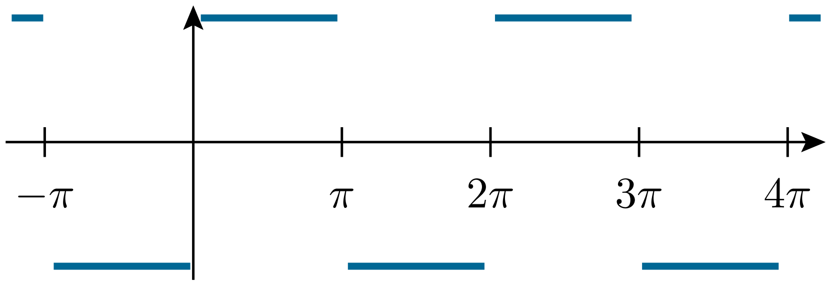
$$\text{blue function} = \frac{4}{\pi}\left(\sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \frac{1}{7}\sin(7x) + \ldots\right)$$



Armin Straub
astraub@illinois.edu

Example 2. Find the Fourier series of the 2π-periodic function $f(x)$ defined by

$$f(x) = \begin{cases} -1, & \text{for } x \in (-\pi, 0), \\ +1, & \text{for } x \in (0, \pi). \end{cases}$$



Solution. Note that $\int_0^{2\pi}$ and $\int_{-\pi}^{\pi}$ are the same here. (why?!)

$$\begin{aligned}
a_0 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{d}x = 0 \\
a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\mathrm{d}x \\
&= \frac{1}{\pi}\left[-\int_{-\pi}^{0} \cos(nx)\mathrm{d}x + \int_0^{\pi} \cos(nx)\mathrm{d}x\right] = 0 \\
b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\mathrm{d}x \\
&= \frac{1}{\pi}\left[-\int_{-\pi}^{0} \sin(nx)\mathrm{d}x + \int_0^{\pi} \sin(nx)\mathrm{d}x\right] \\
&= \frac{2}{\pi}\left[\int_0^{\pi} \sin(nx)\mathrm{d}x\right] \\
&= \frac{2}{\pi}\left[-\frac{1}{n}\cos(nx)\right]_0^{\pi} \\
&= \frac{2}{\pi n}[1 - \cos(n\pi)] \\
&= \frac{2}{\pi n}[1 - (-1)^n] = \begin{cases} \frac{4}{\pi n} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}
\end{aligned}$$

In conclusion,

$$f(x) = \frac{4}{\pi}\left(\sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \frac{1}{7}\sin(7x) + \ldots\right).$$

Armin Straub
astraub@illinois.edu

Determinants

For the next few lectures, all matrices are square!

Recall that

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

The **determinant** of

- a 2×2 matrix is $\det\left(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]\right) = ad-bc$,
- a 1×1 matrix is $\det([\ a\ ]) = a$.

Goal: A is invertible $\iff$ $\det(A)\neq 0$

We will write both $\det\left(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]\right)$ and $\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right|$ for the determinant.

Definition 3. The **determinant** is characterized by:

- the normalization $\det I = 1$,
- and how it is affected by elementary row operations:
  - **(replacement)** Add one row to a multiple of another row.
    Does not change the determinant.
  - **(interchange)** Interchange two rows.
    Reverses the sign of the determinant.
  - **(scaling)** Multiply all entries in a row by s.
    Multiplies the determinant by s.

Example 4. Compute $\begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{vmatrix}$.

Solution.

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{vmatrix} \overset{R2\to\frac{1}{2}R2}{=} 2\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{vmatrix} \overset{R3\to\frac{1}{7}R3}{=} 14\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 14$$

Example 5. Compute $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 7 \end{vmatrix}$.

Solution.

$$\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 7 \end{vmatrix} \overset{R2\to\frac{1}{2}R2}{=} 2\begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 7 \end{vmatrix} \overset{R3\to\frac{1}{7}R3}{=} 14\begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix}$$

$$\overset{\substack{R1\to R1-3R3 \\ R2\to R2-2R3}}{=} 14\begin{vmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \overset{R1\to R1-2R2}{=} 14\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 14$$

Armin Straub
astraub@illinois.edu

The determinant of a triangular matrix is
the product of the diagonal entries.

Example 6. Compute $\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$.

Solution.

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix} \overset{\substack{R2\to R2-3R1 \\ R3\to R3-2R1}}{=} \begin{vmatrix} 1 & 2 & 0 \\ 0 & -7 & 2 \\ 0 & -4 & 1 \end{vmatrix}$$

$$\overset{R3\to R3-\frac{4}{7}R2}{=} \begin{vmatrix} 1 & 2 & 0 \\ 0 & -7 & 2 \\ 0 & 0 & -\frac{1}{7} \end{vmatrix}$$

$$= 1\cdot(-7)\cdot\left(-\frac{1}{7}\right) = 1$$

Example 7. Discover the formula for $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

Solution.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \overset{R2\to R2-\frac{c}{a}R1}{=} \begin{vmatrix} a & b \\ 0 & d-\frac{c}{a}b \end{vmatrix} = a\left(d-\frac{c}{a}b\right) = ad-bc$$

Example 8. Compute $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{vmatrix}$.

Solution.

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{vmatrix} \overset{R4\to R4-\frac{3}{2}R3}{=} \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & \frac{7}{2} \end{vmatrix} = 1\cdot 2\cdot 2\cdot\frac{7}{2} = 14$$

The following important properties follow from the behaviour under row operations.

- $\det(A) = 0 \iff A$ is not invertible

  Why? Because $\det(A) = 0$ if only if, in an echelon form, a diagonal entry is zero (that is, a pivot is missing).

- $\det(AB) = \det(A)\det(B)$
- $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det(A^T) = \det(A)$

Armin Straub
astraub@illinois.edu

Example 9. Recall that $AB = \mathbf{0}$, then it does not follow that $A = \mathbf{0}$ or $B = \mathbf{0}$. However, show that $\det(A) = 0$ or $\det(B) = 0$.

Solution. Follows from $\det(AB) = \det(\mathbf{0}) = 0$,
and $\det(AB) = \det(A)\det(B)$.

A "bad" way to compute determinants

Example 10. Compute $\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$ by **cofactor expansion**.

Solution. We expand by the first row:

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

$$\overset{\text{i.e.}}{=} 1 \cdot \begin{vmatrix} -1 & 2 \\ 0 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} + 0 \cdot \begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix} = 1 \cdot (-1) - 2 \cdot (-1) + 0 = 1$$

Each term in the cofactor expansion is ± 1 times an entry times a smaller determinant (row and column of entry deleted).

The ± 1 is assigned to each entry according to $\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$.

Solution. We expand by the second column:

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

$$= -2 \cdot (-1) + (-1) \cdot 1 - 0 = 1$$

Solution. We expand by the third column:

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

$$= 0 - 2 \cdot (-4) + 1 \cdot (-7) = 1$$

Practice problems

Problem 1. Compute $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 0 & 0 \\ 2 & 7 & 6 & 10 \\ 2 & 9 & 7 & 11 \end{vmatrix}$.

Solution. The final answer should be -10.

Armin Straub
astraub@illinois.edu

Review

- The **determinant** is characterized by $\det I = 1$ and the effect of row op's:
  - replacement: does not change the determinant
  - interchange: reverses the sign of the determinant
  - scaling row by s: multiplies the determinant by s
- $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{vmatrix} = 1 \cdot 2 \cdot 2 \cdot 3 = 12$
- $\det(A) = 0 \iff A$ is not invertible
- $\det(AB) = \det(A)\det(B)$
- $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det(A^T) = \det(A)$
- What's **wrong**?!

  $\det(A^{-1}) = \det \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc}(da - (-b)(-c)) = 1$

  The corrected calculation is:

  $\det \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{(ad - bc)^2}(da - (-b)(-c)) = \frac{1}{ad - bc}$

  This is compatible with $\det(A^{-1}) = \frac{1}{\det(A)}$.

Example 1. Suppose A is a 3×3 matrix with $\det(A) = 5$. What is $\det(2A)$?

Solution. A has three rows.

Multiplying all 3 of them by 2 produces $2A$.

Hence, $\det(2A) = 2^3 \det(A) = 40$.

Armin Straub
astraub@illinois.edu

Example 2. Compute $\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$ by **cofactor expansion**.

Solution. We expand by the second column:

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix} = -2 \cdot \begin{vmatrix} \blacksquare & - & \blacksquare \\ 3 & \blacksquare & 2 \\ 2 & \blacksquare & 1 \end{vmatrix} + (-1) \cdot \begin{vmatrix} 1 & \blacksquare & 0 \\ \blacksquare & + & \blacksquare \\ 2 & \blacksquare & 1 \end{vmatrix} - 0 \cdot \begin{vmatrix} 1 & \blacksquare & 0 \\ 3 & \blacksquare & 2 \\ \blacksquare & - & \blacksquare \end{vmatrix}$$

$$= -2 \cdot (-1) + (-1) \cdot 1 - 0 = 1$$

Solution. We expand by the third column:

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix} = 0 \cdot \begin{vmatrix} \blacksquare & \blacksquare & + \\ 3 & -1 & \blacksquare \\ 2 & 0 & \blacksquare \end{vmatrix} - 2 \cdot \begin{vmatrix} 1 & 2 & \blacksquare \\ \blacksquare & \blacksquare & - \\ 2 & 0 & \blacksquare \end{vmatrix} + 1 \cdot \begin{vmatrix} 1 & 2 & \blacksquare \\ 3 & -1 & \blacksquare \\ \blacksquare & \blacksquare & + \end{vmatrix}$$

$$= 0 - 2 \cdot (-4) + 1 \cdot (-7) = 1$$

Why is the method of cofactor expansion not practical?

Because to compute a large $n \times n$ determinant,

- one reduces to n determinants of size $(n-1) \times (n-1)$,
- then $n(n-1)$ determinants of size $(n-2) \times (n-2)$,
- and so on.

In the end, we have $n! = n(n-1) \cdots 3 \cdot 2 \cdot 1$ many numbers to add.

WAY TOO MUCH WORK! Already $25! = 15511210043330985984000000 \approx 1.55 \cdot 10^{25}$.

Context: today's fastest computer, Tianhe-2, runs at 34 pflops ($3.4 \cdot 10^{16}$ op's per second).

By the way: "fastest" is measured by computed LU decompositions!

Example 3.

First off, say hello to a new friend: i, the **imaginary unit**

It is infamous for $\boxed{i^2 = -1}$.

$$\begin{vmatrix} 1 \end{vmatrix} = 1$$

$$\begin{vmatrix} 1 & i \\ i & 1 \end{vmatrix} = 1 - i^2 = 2$$

$$\begin{vmatrix} 1 & i & \\ i & 1 & i \\ & i & 1 \end{vmatrix} = 1 \begin{vmatrix} 1 & i \\ i & 1 \end{vmatrix} - i \begin{vmatrix} i & 0 \\ i & 1 \end{vmatrix} = 2 - i^2 = 3$$

$$\begin{vmatrix} 1 & i & & \\ i & 1 & i & \\ & i & 1 & i \\ & & i & 1 \end{vmatrix} = 1 \begin{vmatrix} 1 & i & \\ i & 1 & i \\ & i & 1 \end{vmatrix} - i \begin{vmatrix} i & 0 & \\ i & 1 & i \\ & i & 1 \end{vmatrix} = 3 - i^2 \begin{vmatrix} 1 & i \\ i & 1 \end{vmatrix} = 5$$

Armin Straub
astraub@illinois.edu

$$\begin{vmatrix} 1 & i & & & \\ i & 1 & i & & \\ & i & 1 & i & \\ & & i & 1 & i \\ & & & i & 1 \end{vmatrix} = 1 \begin{vmatrix} 1 & i & & \\ i & 1 & i & \\ & i & 1 & i \\ & & i & 1 \end{vmatrix} - i^2 \begin{vmatrix} 1 & i & \\ i & 1 & i \\ & i & 1 \end{vmatrix} = 5 + 3 = 8$$

The Fibonacci numbers!

Do you know about the connection of Fibonacci numbers and rabbits?

Eigenvectors and eigenvalues

Throughout, A will be an $n \times n$ matrix.

Definition 4. An **eigenvector** of A is a nonzero $\boldsymbol{x}$ such that

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

The scalar λ is the corresponding **eigenvalue**.

In words: eigenvectors are those $\boldsymbol{x}$, for which $A\boldsymbol{x}$ is parallel to $\boldsymbol{x}$.

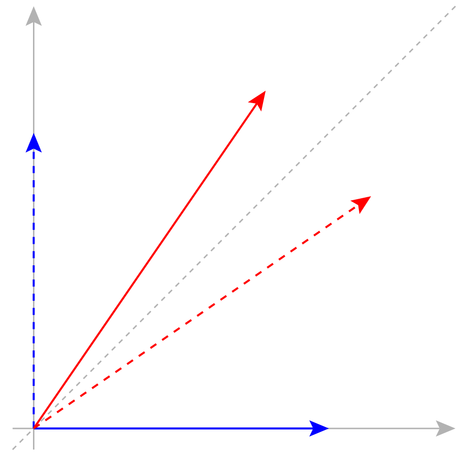
Example 5. Verify that $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}$.

Solution.

$A\boldsymbol{x} = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ -8 \end{bmatrix} = 4\boldsymbol{x}$

Hence, $\boldsymbol{x}$ is an eigenvector of A with eigenvalue 4.

Example 6. Use your geometric understanding to find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.



Solution. $A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$

i.e. multiplication with A is reflection through the line $y = x$.

- $A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

  So: $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 1$.

- $A\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

  So: $\boldsymbol{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = -1$.

Practice problems

Problem 1. Let A be an $n \times n$ matrix.

Express the following in terms of $\det(A)$:

- $\det(A^2) =$
- $\det(2A) =$

  Hint: (unless $n = 1$) this is not just $2\det(A)$

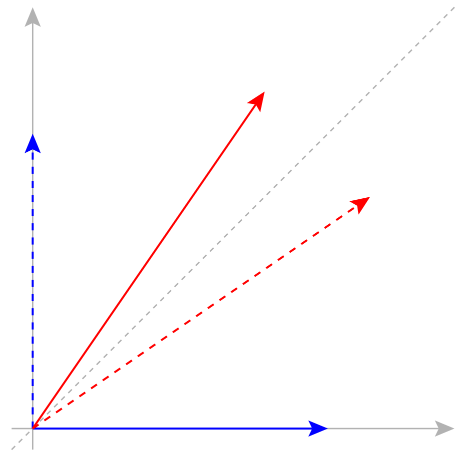
Review

- If $A\boldsymbol{x} = \lambda\boldsymbol{x}$, then $\boldsymbol{x}$ is an **eigenvector** of A with **eigenvalue** λ.
- EG: $\boldsymbol{x} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}$ with eigenvalue 4

  because $A\boldsymbol{x} = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ -8 \end{bmatrix} = 4\boldsymbol{x}$
- Multiplication with $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is reflection through the line $y = x$.
  - $A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

    So: $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 1$.
  - $A\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

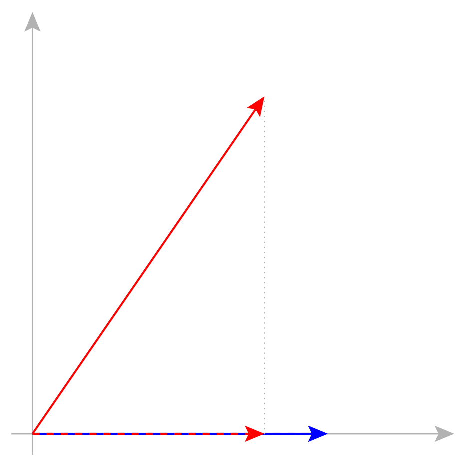
    So: $\boldsymbol{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = -1$.



Example 1. Use your geometric understanding to find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Solution. $A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$

i.e. multiplication with A is projection onto the x-axis.



- $A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

  So: $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 1$.
- $A\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

  So: $\boldsymbol{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 0$.

Armin Straub
astraub@illinois.edu

Example 2. Let P be the projection matrix corresponding to orthogonal projection onto the subspace V. What are the eigenvalues and eigenvectors of P?

Solution.

- For every vector $\boldsymbol{x}$ in V, $P\boldsymbol{x} = \boldsymbol{x}$.

  These are the eigenvectors with eigenvalue 1.

- For every vector $\boldsymbol{x}$ orthogonal to V, $P\boldsymbol{x} = \boldsymbol{0}$.

  These are the eigenvectors with eigenvalue 0.

Definition 3. Given λ, the set of all eigenvectors with eigenvalue λ is called the **eigenspace** of A corresponding to λ.

Example 4. (continued) We saw that the projection matrix P has the two eigenvalues $\lambda = 0, 1$.

- The eigenspace of $\lambda = 1$ is V.
- The eigenspace of $\lambda = 0$ is $V^{\perp}$.

How to solve $A\boldsymbol{x} = \lambda\boldsymbol{x}$

Key observation:

$$\begin{aligned} & A\boldsymbol{x} = \lambda\boldsymbol{x} \\ \Longleftrightarrow \quad & A\boldsymbol{x} - \lambda\boldsymbol{x} = \boldsymbol{0} \\ \Longleftrightarrow \quad & (A - \lambda I)\boldsymbol{x} = \boldsymbol{0} \end{aligned}$$

This has a nonzero solution $\Longleftrightarrow$ $\det(A - \lambda I) = 0$

Recipe. To find eigenvectors and eigenvalues of A.

- First, find the eigenvalues λ using:

  λ is an eigenvalue of A $\Longleftrightarrow$ $\det(A - \lambda I) = 0$

- Then, for each eigenvalue λ, find corresponding eigenvectors by solving $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$.

Example 5. Find the eigenvectors and eigenvalues of

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

Solution.

- $A-\lambda I=\begin{bmatrix} 3 & 1\\ 1 & 3 \end{bmatrix}-\lambda\begin{bmatrix} 1 & 0\\ 0 & 1 \end{bmatrix}=\begin{bmatrix} 3-\lambda & 1\\ 1 & 3-\lambda \end{bmatrix}$
- $\det(A-\lambda I)=\begin{vmatrix} 3-\lambda & 1\\ 1 & 3-\lambda \end{vmatrix}=(3-\lambda)^2-1$

  $=\lambda^2-6\lambda+8=0 \implies \lambda_1=2,\ \lambda_2=4$

  This is the **characteristic polynomial** of A. Its roots are the eigenvalues of A.
- Find eigenvectors with eigenvalue $\lambda_1=2$:

  $A-\lambda_1 I=\begin{bmatrix} 1 & 1\\ 1 & 1 \end{bmatrix}$ $\quad\left(A=\begin{bmatrix} 3 & 1\\ 1 & 3 \end{bmatrix}\right)$

  Solutions to $\begin{bmatrix} 1 & 1\\ 1 & 1 \end{bmatrix}\boldsymbol{x}=\mathbf{0}$ have basis $\begin{bmatrix} -1\\ 1 \end{bmatrix}$.

  So: $\boldsymbol{x}_1=\begin{bmatrix} -1\\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda_1=2$.

  All other eigenvectors with $\lambda=2$ are multiples of $\boldsymbol{x}_1$.

  $\operatorname{span}\left\{\begin{bmatrix} -1\\ 1 \end{bmatrix}\right\}$ is the **eigenspace** for eigenvalue $\lambda=2$.
- Find eigenvectors with eigenvalue $\lambda_2=4$:

  $A-\lambda_2 I=\begin{bmatrix} -1 & 1\\ 1 & -1 \end{bmatrix}$ $\quad\left(A=\begin{bmatrix} 3 & 1\\ 1 & 3 \end{bmatrix}\right)$

  Solutions to $\begin{bmatrix} -1 & 1\\ 1 & -1 \end{bmatrix}\boldsymbol{x}=\mathbf{0}$ have basis $\begin{bmatrix} 1\\ 1 \end{bmatrix}$.

  So: $\boldsymbol{x}_2=\begin{bmatrix} 1\\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda_2=4$.

  The eigenspace for eigenvalue $\lambda=4$ is $\operatorname{span}\left\{\begin{bmatrix} 1\\ 1 \end{bmatrix}\right\}$.

Example 6. Find the eigenvectors and the eigenvalues of

$$A=\begin{bmatrix} 3 & 2 & 3\\ 0 & 6 & 10\\ 0 & 0 & 2 \end{bmatrix}.$$

Solution.

- The characteristic polynomial is:

$$\det(A-\lambda I)=\begin{vmatrix} 3-\lambda & 2 & 3\\ 0 & 6-\lambda & 10\\ 0 & 0 & 2-\lambda \end{vmatrix}=(3-\lambda)(6-\lambda)(2-\lambda)$$

- A has eigenvalues $2,3,6$. $\quad A=\begin{bmatrix} 3 & 2 & 3\\ 0 & 6 & 10\\ 0 & 0 & 2 \end{bmatrix}$

  The eigenvalues of a triangular matrix are its diagonal entries.
- $\lambda_1=2$:

  $(A-\lambda_1 I)\boldsymbol{x}=\begin{bmatrix} 1 & 2 & 3\\ 0 & 4 & 10\\ 0 & 0 & 0 \end{bmatrix}\boldsymbol{x}=\mathbf{0} \implies \boldsymbol{x}_1=\begin{bmatrix} 2\\ -5/2\\ 1 \end{bmatrix}$

Armin Straub
astraub@illinois.edu

- $\lambda_2 = 3$:

  $$(A - \lambda_2 I)\boldsymbol{x} = \begin{bmatrix} 0 & 2 & 3 \\ 0 & 3 & 10 \\ 0 & 0 & -1 \end{bmatrix}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

- $\lambda_3 = 6$:

  $$(A - \lambda_3 I)\boldsymbol{x} = \begin{bmatrix} -3 & 2 & 3 \\ 0 & 0 & 10 \\ 0 & 0 & -4 \end{bmatrix}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$

- In summary, A has eigenvalues $2, 3, 6$ with corresponding eigenvectors $\begin{bmatrix} 2 \\ -5/2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2/3 \\ 1 \\ 0 \end{bmatrix}$.

  These three vectors are independent. By the next result, this is always so.

Theorem 7. If $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are eigenvectors of A corresponding to different eigenvalues, then they are independent.

Why?

Suppose, for contradiction, that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are dependent.

By kicking out some of the vectors, we may assume that there is (up to multiples) only one linear relation: $c_1\boldsymbol{x}_1 + \ldots + c_m\boldsymbol{x}_m = \boldsymbol{0}$.

Multiply this relation with A:

$A(c_1\boldsymbol{x}_1 + \ldots + c_m\boldsymbol{x}_m) = c_1\lambda_1\boldsymbol{x}_1 + \ldots + c_m\lambda_m\boldsymbol{x}_m = \boldsymbol{0}$

This is a second independent relation! Contradiction.

Practice problems

Example 8. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}$.

Example 9. What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 3 & 0 & 0 \\ -1 & 1 & 3 & 0 \\ 0 & 1 & 2 & 4 \end{bmatrix}$?

No calculations!

Review

- If $Ax = \lambda x$, then x is an **eigenvector** of A with **eigenvalue** λ.

  All eigenvectors (plus $\mathbf{0}$) with eigenvalue λ form the **eigenspace** of λ.

- λ is an eigenvalue of $A \iff \underbrace{\det(A - \lambda I)}_{\textbf{characteristic polynomial}} = 0.$

  Why? Because $Ax = \lambda x \iff (A - \lambda I)x = \mathbf{0}$.

  By the way: this means that the eigenspace of λ is just $\mathrm{Nul}(A - \lambda I)$.

- E.g., if $A = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 6 & 10 \\ 0 & 0 & 2 \end{bmatrix}$ then $\det(A - \lambda I) = (3 - \lambda)(6 - \lambda)(2 - \lambda)$.

- Eigenvectors $x_1, \ldots, x_m$ of A corresponding to different eigenvalues are independent.

- By the way:
  - product of eigenvalues = determinant
  - sum of eigenvalues = "trace" (sum of diagonal entries)

Example 1. Find the eigenvalues of A as well as a basis for the corresponding eigenspaces, where

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

Solution.

- The characteristic polynomial is:

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 2-\lambda & 0 & 0 \\ -1 & 3-\lambda & 1 \\ -1 & 1 & 3-\lambda \end{vmatrix} \\ &= (2-\lambda)\begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} \\ &= (2-\lambda)[(3-\lambda)^2 - 1] \\ &= (2-\lambda)(\lambda-2)(\lambda-4) \end{aligned}$$

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

  Since $\lambda = 2$ is a double root, it has **(algebraic) multiplicity** 2.

- $\lambda_1 = 2$:

  $(A - \lambda_1 I)x = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} x = \mathbf{0}$

Armin Straub
astraub@illinois.edu

Two independent solutions: $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$

In other words: the eigenspace for $\lambda = 2$ is $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right\}$.

- $\lambda_2 = 4$:

  $(A - \lambda_2 I)\boldsymbol{x} = \begin{bmatrix} -2 & 0 & 0 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

- In summary, A has eigenvalues 2 and 4:
  - eigenspace for $\lambda = 2$ has basis $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$,
  - eigenspace for $\lambda = 4$ has basis $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

An $n \times n$ matrix A has up to n different eigenvalues.

Namely, the roots of the degree n characteristic polynomial $\det(A - \lambda I)$.

- For each eigenvalue λ, A has at least one eigenvector.

  That's because $\operatorname{Nul}(A - \lambda I)$ has dimension at least 1.

- If λ has multiplicity m, then A has up to m (independent) eigenvectors for λ.

Ideally, we would like to find a total of n (independent) eigenvectors of A.

Why can there be no more than n eigenvectors?!

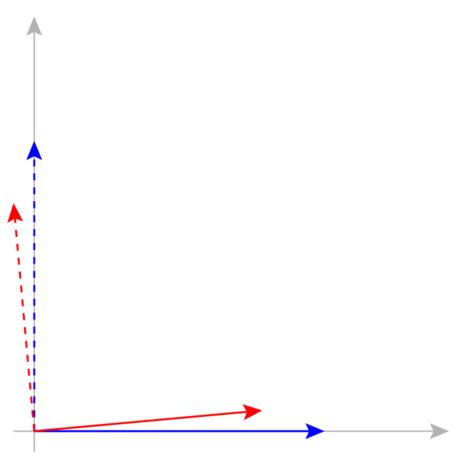
Two sources of trouble: eigenvalues can be

- complex numbers (that is, not enough real roots), or
- repeated roots of the characteristic polynomial.

Example 2. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Geometrically, what is the trouble?

Solution. $A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$

i.e. multiplication with A is rotation by $90°$ (counter-clockwise).



Which vector is parallel after rotation by $90°$? Trouble.

Armin Straub
astraub@illinois.edu

Fix: work with complex numbers!

- $\det(A-\lambda I) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2+1$

  So, the eigenvalues are $\lambda_1 = i$ and $\lambda_2 = -i$.

- $\lambda_1 = i$: $\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \boldsymbol{x}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}$

  Let us check: $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ i \end{bmatrix} = i\begin{bmatrix} i \\ 1 \end{bmatrix}$

- $\lambda_2 = -i$: $\begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \boldsymbol{x}_2 = \begin{bmatrix} -i \\ 1 \end{bmatrix}$

Example 3. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. What is the trouble?

Solution.

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

  So: $\lambda = 1$ is the only eigenvalue (it has multiplicity 2).

- $(A-\lambda I)\boldsymbol{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

  So: the eigenspace is $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$. Only dimension 1!

- Trouble: only 1 independent eigenvector for a 2×2 matrix

  This kind of trouble cannot really be fixed.

  We have to lower our expectations and look for *generalized eigenvectors*.

  These are solutions to $(A-\lambda I)^2\boldsymbol{x} = \boldsymbol{0}$, $(A-\lambda I)^3\boldsymbol{x} = \boldsymbol{0}$, ...

Practice problems

Example 4. Find the eigenvectors and eigenvalues of $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -5 & 0 \\ 1 & 8 & 1 \end{bmatrix}$.

Review

- **Eigenvector** equation: $A\boldsymbol{x} = \lambda\boldsymbol{x} \iff (A - \lambda I)\boldsymbol{x} = \mathbf{0}$

  λ is an **eigenvalue** of $A \iff \underbrace{\det(A - \lambda I)}_{\text{characteristic polynomial}} = 0.$

- An $n \times n$ matrix A has up to n different eigenvalues λ.
  - The **eigenspace** of λ is $\mathrm{Nul}(A - \lambda I)$.

    That is, all eigenvectors of A with eigenvalue λ.
  - If λ has **multiplicity** m, then A has up to m eigenvectors for λ.

    At least one eigenvector is guaranteed (because $\det(A - \lambda I) = 0$).

- Test yourself! What are the eigenvalues and eigenvectors?
  - $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\lambda = 1, 1$ (ie. multiplicity 2), eigenspace is $\mathbb{R}^2$
  - $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ $\lambda = 0, 0$, eigenspace is $\mathbb{R}^2$
  - $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ $\lambda = 2, 2$, eigenspace is $\mathrm{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$

Diagonalization

Diagonal matrices are very easy to work with.

Example 1. For instance, it is easy to compute their powers.

If $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$, then $A^2 = \begin{bmatrix} 2^2 & & \\ & 3^2 & \\ & & 4^2 \end{bmatrix}$ and $A^{100} = \begin{bmatrix} 2^{100} & & \\ & 3^{100} & \\ & & 4^{100} \end{bmatrix}$

Example 2. If $A = \begin{bmatrix} 6 & -1 \\ 2 & 3 \end{bmatrix}$, then $A^{100} = ?$

Solution.

- Characteristic polynomial: $\begin{vmatrix} 6-\lambda & -1 \\ 2 & 3-\lambda \end{vmatrix} = \ldots = (\lambda - 4)(\lambda - 5)$
  - $\lambda_1 = 4$: $\begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix}\boldsymbol{v} = \mathbf{0} \implies$ eigenvector $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
  - $\lambda_2 = 5$: $\begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \implies$ eigenvector $\boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- Key observation: $A^{100}\boldsymbol{v}_1 = \lambda_1^{100}\boldsymbol{v}_1$ and $A^{100}\boldsymbol{v}_2 = \lambda_2^{100}\boldsymbol{v}_2$

  For A^{100}, we need $A^{100}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $A^{100}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Armin Straub
astraub@illinois.edu

- $\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 \boldsymbol{v}_1 + c_2 \boldsymbol{v}_2 = -\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

  $\Longrightarrow \quad A^{100}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = A^{100}\left(-\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = -4^{100}\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2 \cdot 5^{100}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

  $\Longrightarrow \quad A^{100} = \begin{bmatrix} 2 \cdot 5^{100} - 4^{100} & * \\ 2 \cdot 5^{100} - 2 \cdot 4^{100} & * \end{bmatrix}$
- We find the second column of A^{100} likewise. Left as exercise!

The key idea of the previous example was to work with respect to a basis given by the eigenvectors.

- Put the eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ as columns into a matrix P.

$$A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i \quad \Longrightarrow \quad A\begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- In summary: $AP = PD$

Suppose that A is $n \times n$ and has independent eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.
Then A can be **diagonalized** as $A = PDP^{-1}$.

- the columns of P are the eigenvectors
- the diagonal matrix D has the eigenvalues on the diagonal

Such a diagonalization is possible if and only if A has enough eigenvectors.

Armin Straub
astraub@illinois.edu

Example 3.

Fibonacci numbers: $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$

By the way: "not a universal law but only a fascinatingly prevalent tendency" — Coxeter

Did you notice: $\frac{13}{8} = 1.625$, $\frac{21}{13} = 1.615$, $\frac{34}{21} = 1.619$, ...

The **golden ratio** $\varphi = 1.618\ldots$ Where's that coming from?

By the way, this φ is the *most irrational* number (in a precise sense).

- $F_{n+1} = F_n + F_{n-1} \implies \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix}$
- Hence: $\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}$ $\qquad \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- But we know how to compute $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n$ or $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}$!

Solution. (Exercise to fill in all details!)

- The characteristic polynomial of $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ is $\lambda^2 - \lambda - 1$.
- The eigenvalues are $\lambda_1 = \frac{1 + \sqrt{5}}{2} \approx 1.618$ (the golden mean!) and $\lambda_2 = \frac{1 - \sqrt{5}}{2} \approx -0.618$.
- Corresponding eigenvectors: $\boldsymbol{v}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$
- Write $\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 \boldsymbol{v}_1 + c_1 \boldsymbol{v}_2$. $\qquad$ ($c_1 = \frac{1}{\sqrt{5}}$, $c_2 = -\frac{1}{\sqrt{5}}$)
- $\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = A^n \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_1^n c_1 \boldsymbol{v}_1 + \lambda_2^n c_2 \boldsymbol{v}_2$
- Hence, $F_n = \lambda_1^n c_1 + \lambda_2^n c_2 = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right]$.

  That's **Binet's formula**.
- But $|\lambda_2| < 1$, and so $F_n \approx \lambda_1^n c_1 = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^n$.

  In fact, $F_n = \operatorname{round}\left(\frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^n \right)$. Don't you feel powerful!?

Practice problems

Problem 1. Find, if possible, the diagonalization of $A = \begin{bmatrix} 0 & -2 \\ -4 & 2 \end{bmatrix}$.

Armin Straub
astraub@illinois.edu

Review for Midterm 3

- Bring a **number 2 pencil** to the exam!
- **Extra help session**: today and tomorrow, **4–7pm**, in AH 441
- Room assignments for Thursday, Nov 20, 7-8:15pm:
  - if your last name starts with A-E: 213 Greg Hall
  - if your last name starts with F-L: 100 Greg Hall
  - if your last name starts with M-Sh: 66 Library
  - if your last name starts with Si-Z: 103 Mumford Hall
- Big topics:
  - Orthogonal projections
  - Least squares
  - Gram–Schmidt
  - Determinants
  - Eigenvalues and eigenvectors

Orthogonal projections

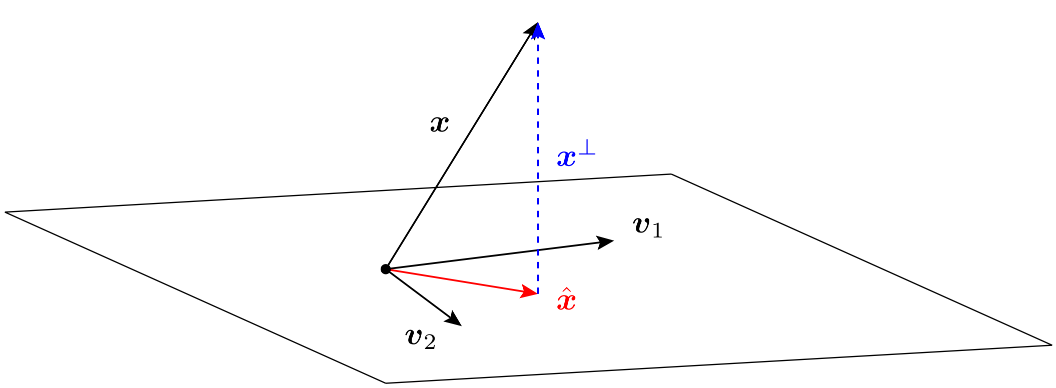
- If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is an **orthogonal basis** of V, and $\boldsymbol{x}$ is in V, then

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n \qquad \text{with} \quad c_j = \frac{\langle \boldsymbol{x}, \boldsymbol{v}_j \rangle}{\langle \boldsymbol{v}_j, \boldsymbol{v}_j \rangle}.$$

- Suppose that V is a subspace of W, and $\boldsymbol{x}$ is in W, then the **orthogonal projection** of $\boldsymbol{x}$ onto V is given by

$$\hat{\boldsymbol{x}} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n \qquad \text{with} \quad c_j = \frac{\langle \boldsymbol{x}, \boldsymbol{v}_j \rangle}{\langle \boldsymbol{v}_j, \boldsymbol{v}_j \rangle}.$$

  - The basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ has to be orthogonal for this formula!!
  - This decomposes $\boldsymbol{x} = \underbrace{\hat{\boldsymbol{x}}}_{\text{in } V} + \underbrace{\boldsymbol{x}^\perp}_{\text{in } V^\perp}$, where the error $\boldsymbol{x}^\perp$ is orthogonal to V. (this decomposition is unique)



Armin Straub
astraub@illinois.edu

- The corresponding **projection matrix** represents $\boldsymbol{x} \mapsto \hat{\boldsymbol{x}}$ with respect to the standard basis.

Example 1.

(a) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ onto $\mathrm{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$?

Solution: The projection is $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

(b) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ onto $\mathrm{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$?

Solution: The projection is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Wrong approach!! $\dfrac{\left\langle \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\rangle}{\left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\rangle} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \dfrac{\left\langle \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\rangle}{\left\langle \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\rangle} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

This is wrong because $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ are not orthogonal. (See next example!)

(c) What is the orthogonal projection of $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ onto $\mathrm{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$?

Solution: The projection is $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$.

Wrong!! $\dfrac{\left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\rangle}{\left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\rangle} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \dfrac{\left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\rangle}{\left\langle \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\rangle} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

Corrected: $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (for instance, using Gram–Schmidt)

$$\dfrac{\left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\rangle}{\left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\rangle} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \dfrac{\left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\rangle}{\left\langle \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\rangle} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

(d) What is the projection matrix corresponding to orthogonal projection onto $\mathrm{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$?

Solution: The projection matrix is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

What would Gram–Schmidt do? $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

(e) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $\mathrm{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$?

Armin Straub
astraub@illinois.edu

Solution: The projection is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

- The space of all nice functions with period 2π has the natural inner product $\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\mathrm{d}x$. [in $\mathbb{R}^n$: $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = x_1y_1 + \ldots + x_ny_n$]
- The functions

$$1, \cos(x), \sin(x), \cos(2x), \sin(2x), \ldots$$

  are an orthogonal basis for this space.
- Expanding a function $f(x)$ in this basis produces its **Fourier series**

$$f(x) = a_0 + a_1\cos(x) + b_1\sin(x) + a_2\cos(2x) + b_2\sin(2x) + \cdots$$

Example 2. How can we compute b_2?

Solution.

$b_2\sin(2x)$ is the orthogonal projection of f onto the span of $\sin(2x)$.

Hence:

$$b_2 = \frac{\langle f(x), \sin(2x)\rangle}{\langle \sin(2x), \sin(2x)\rangle} = \frac{\int_0^{2\pi} f(x)\sin(2x)\mathrm{d}x}{\int_0^{2\pi} \sin^2(2x)\mathrm{d}x}$$

Least squares

- $\hat{\boldsymbol{x}}$ is a **least squares solution** of the system $A\boldsymbol{x} = \boldsymbol{b}$.

  $\iff$ $\hat{\boldsymbol{x}}$ is such that $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible.

  $\iff$ $A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ (the **normal equations**)

Example 3. Find the least squares line for the data points $(2,1), (5,2), (7,3), (8,3)$.

Solution.

Looking for β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data.

The equations $y_i = \beta_1 + \beta_2 x_i$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$$

Here, we need to find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$.

Gram–Schmidt

Recipe. (Gram–Schmidt orthonormalization)

Given a basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, produce an orthonormal basis $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$.

$$\begin{aligned} \boldsymbol{b}_1 &= \boldsymbol{a}_1, & \boldsymbol{q}_1 &= \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} \\ \boldsymbol{b}_2 &= \boldsymbol{a}_2 - \langle \boldsymbol{a}_2, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1, & \boldsymbol{q}_2 &= \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|} \\ \boldsymbol{b}_3 &= \boldsymbol{a}_3 - \langle \boldsymbol{a}_3, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1 - \langle \boldsymbol{a}_3, \boldsymbol{q}_2 \rangle \boldsymbol{q}_2, & \boldsymbol{q}_3 &= \frac{\boldsymbol{b}_3}{\|\boldsymbol{b}_3\|} \\ &\vdots \end{aligned}$$

- An **orthogonal matrix** is a square matrix Q with orthonormal columns. Equivalently, $Q^T Q = I$ (also true for non-square matrices).
- Apply Gram–Schmidt to the (independent) columns of A to obtain the **QR decomposition** $A = QR$.
  - Q has orthonormal columns (the output vectors of Gram–Schmidt)
  - $R = Q^T A$ is upper triangular

Example 4. Find the QR decomposition of $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}$.

Solution. We apply Gram–Schmidt to the columns of A:

$$\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \boldsymbol{q}_1$$

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \left\langle \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{5}}\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 4/5 \\ -2/5 \\ 1 \end{bmatrix}, \qquad \frac{1}{\sqrt{9/5}}\begin{bmatrix} 4/5 \\ -2/5 \\ 1 \end{bmatrix} = \boldsymbol{q}_2$$

Hence: $Q = [\ \boldsymbol{q}_1 \ \ \boldsymbol{q}_2\ ] = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{4}{\sqrt{45}} \\ \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} \\ 0 & \frac{5}{\sqrt{45}} \end{bmatrix}$

And: $R = Q^T A = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0 \\ \frac{4}{\sqrt{45}} & -\frac{2}{\sqrt{45}} & \frac{5}{\sqrt{45}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ 0 & \frac{9}{\sqrt{45}} \end{bmatrix}$

Determinants

- A is invertible $\iff$ $\det(A) \neq 0$
- $\det(AB) = \det(A)\det(B)$
- $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det(A^T) = \det(A)$
- The **determinant** is characterized by:
  - the normalization $\det I = 1$,
  - and how it is affected by elementary row operations:
    - **(replacement)** Add one row to a multiple of another row.

      Does not change the determinant.
    - **(interchange)** Interchange two rows.

      Reverses the sign of the determinant.
    - **(scaling)** Multiply all entries in a row by s.

      Multiplies the determinant by s.

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{vmatrix} \overset{R4 \to R4 - \frac{3}{2}R3}{=} \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & \frac{7}{2} \end{vmatrix} = 1 \cdot 2 \cdot 2 \cdot \frac{7}{2} = 14$$

- **Cofactor expansion** is another way to compute determinants.

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

$$= -2\cdot(-1) + (-1)\cdot 1 - 0 = 1$$

Example 5. What is $\begin{vmatrix} 1 & 1 & 1 & 4 \\ -1 & 2 & 2 & 5 \\ 0 & 3 & 3 & 1 \\ 2 & 0 & 0 & 5 \end{vmatrix}$?

Solution. The determinant is 0 because the matrix is not invertible (second and third column are the same).

Eigenvalues and eigenvectors

- If $A\boldsymbol{x} = \lambda\boldsymbol{x}$, then $\boldsymbol{x}$ is an **eigenvector** of A with **eigenvalue** λ.
- λ is an eigenvalue of A $\iff$ $\underbrace{\det(A - \lambda I)}_{\text{characteristic polynomial}} = 0$.

  Why? Because $A\boldsymbol{x} = \lambda\boldsymbol{x} \iff (A - \lambda I)\boldsymbol{x} = \mathbf{0}$.
- The **eigenspace** of λ is $\operatorname{Nul}(A - \lambda I)$.

  It consists of all eigenvectors (plus $\mathbf{0}$) with eigenvalue λ.
- Eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ of A corresponding to different eigenvalues are independent.
- Useful for checking: sum of eigenvalues $=$ sum of diagonal entries

Welcome back!

- The final exam is on Friday, December 12, 7-10pm

  If you have a conflict (overlapping exam, or more than 2 exams within 24h), please email Mahmood until Sunday to sign-up for the Monday conflict.

- What are the eigenspaces of $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$?
  - $\lambda = 1$ has eigenspace $\operatorname{Nul}\left(\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix}\right) = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$
  - $\lambda = 3$ has eigenspace $\operatorname{Nul}\left(\begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix}\right) = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$
  - INCORRECT: eigenspace $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$

Transition matrices

Powers of matrices can describe transition of a system.

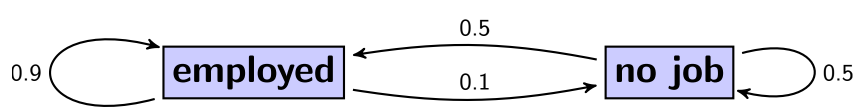
Example 1. (review)

- Fibonacci numbers F_n: $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$
- $F_{n+1} = F_n + F_{n-1} \quad \Longrightarrow \quad \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix}$
- Hence: $\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}$

Example 2. Consider a fixed population of people with or without a job. Suppose that, each year, 50% of those unemployed find a job while 10% of those employed loose their job.

What is the unemployment rate in the long term equilibrium?

Solution.



x_t: proportion of population employed at time t (in years)

y_t: proportion of population unemployed at time t

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 0.9x_t + 0.5y_t \\ 0.1x_t + 0.5y_t \end{bmatrix} = \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix}$$

The matrix $\begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix}$ is a **Markov matrix**. Its columns add to 1 and it has no negative entries.

Armin Straub
astraub@illinois.edu

$\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix}$ is an equilibrium if $\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix} = \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix} \begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix}$.

In other words, $\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix}$ is an eigenvector with eigenvalue 1.

Eigenspace of $\lambda = 1$: $\text{Nul}\left(\begin{bmatrix} -0.1 & 0.5 \\ 0.1 & -0.5 \end{bmatrix} \right) = \text{span}\left\{ \begin{bmatrix} 5 \\ 1 \end{bmatrix} \right\}$

Since $x_\infty + y_\infty = 1$, we conclude that $\begin{bmatrix} x_\infty \\ y_\infty \end{bmatrix} = \begin{bmatrix} 5/6 \\ 1/6 \end{bmatrix}$.

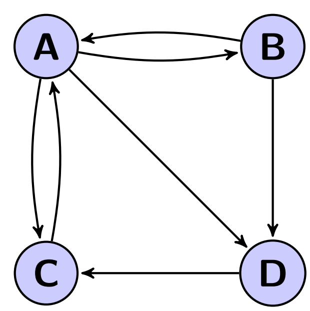
Hence, the unemployment rate in the long term equilibrium is $1/6$.

Page rank

Google's success is based on an algorithm to rank webpages, the **Page rank**, named after Google founder Larry Page.

The basic idea is to determine how likely it is that a web user randomly gets to a given webpage. The webpages are then ranked by these probabilities.

Example 3. Suppose the internet consisted of only the four webpages A, B, C, D linked as in the following graph:



Imagine a surfer following these links at random.

For the probability $\text{PR}_n(A)$ that she is at A (after n steps), we add:

- the probability that she was at B (at exactly one time step before), and left for A, (that's $\text{PR}_{n-1}(B) \cdot \frac{1}{2}$)
- the probability that she was at C, and left for A,
- the probability that she was at D, and left for A.

- Hence: $\text{PR}_n(A) = \text{PR}_{n-1}(B) \cdot \frac{1}{2} + \text{PR}_{n-1}(C) \cdot \frac{1}{1} + \text{PR}_{n-1}(D) \cdot \frac{0}{1}$
- $\begin{bmatrix} \text{PR}_n(A) \\ \text{PR}_n(B) \\ \text{PR}_n(C) \\ \text{PR}_n(D) \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{PR}_{n-1}(A) \\ \text{PR}_{n-1}(B) \\ \text{PR}_{n-1}(C) \\ \text{PR}_{n-1}(D) \end{bmatrix}$
- The **PageRank vector** $\begin{bmatrix} \text{PR}(A) \\ \text{PR}(B) \\ \text{PR}(C) \\ \text{PR}(D) \end{bmatrix} = \begin{bmatrix} \text{PR}_\infty(A) \\ \text{PR}_\infty(B) \\ \text{PR}_\infty(C) \\ \text{PR}_\infty(D) \end{bmatrix}$ is the long-term equilibrium.

  It is an eigenvector of the Markov matrix with eigenvalue 1.
- $\begin{bmatrix} -1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -1 & 0 & 0 \\ \frac{1}{3} & 0 & -1 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -\frac{2}{3} \\ 0 & 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$

  $\Longrightarrow$ eigenspace of $\lambda = 1$ spanned by $\begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix}$

Armin Straub
astraub@illinois.edu

$$\Longrightarrow \begin{bmatrix} \mathrm{PR}(A) \\ \mathrm{PR}(B) \\ \mathrm{PR}(C) \\ \mathrm{PR}(D) \end{bmatrix} = \frac{3}{16}\begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix}$$ This the PageRank vector.

- The corresponding ranking of the webpages is A, C, D, B.

Practice problems

Problem 1. Can you see why 1 is an eigenvalue for every Markov matrix?

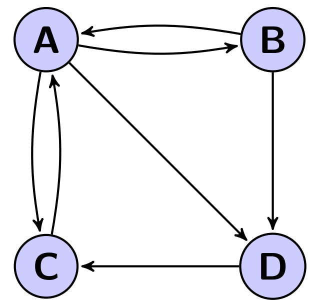
Problem 2. (just for fun) The real web contains pages which have no outgoing links. In that case, our random surfer would get "stuck" (the transition matrix is not a Markov matrix). Do you have an idea how to deal with this issue?

Armin Straub
astraub@illinois.edu

Review

- We model a surfer randomly clicking webpages.

Let $\mathrm{PR}_n(A)$ be the probability that he is at A (after n steps).



$\mathrm{PR}_n(A) = \mathrm{PR}_{n-1}(B) \cdot \frac{1}{2} + \mathrm{PR}_{n-1}(C) \cdot \frac{1}{1} + \mathrm{PR}_{n-1}(D) \cdot \frac{0}{1}$

$$\begin{bmatrix} \mathrm{PR}_n(A) \\ \mathrm{PR}_n(B) \\ \mathrm{PR}_n(C) \\ \mathrm{PR}_n(D) \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}}_{=T} \begin{bmatrix} \mathrm{PR}_{n-1}(A) \\ \mathrm{PR}_{n-1}(B) \\ \mathrm{PR}_{n-1}(C) \\ \mathrm{PR}_{n-1}(D) \end{bmatrix}$$

- The transition matrix T is a **Markov matrix**.

  Its columns add to 1 and it has no negative entries.

- The **Page rank** of page A is $\mathrm{PR}(A) = \mathrm{PR}_\infty(A)$. (assuming the limit exists)

  It is the probability that the surfer is at page A after n steps (with $n \to \infty$).

- The **PageRank vector** $\begin{bmatrix} \mathrm{PR}(A) \\ \mathrm{PR}(B) \\ \mathrm{PR}(C) \\ \mathrm{PR}(D) \end{bmatrix}$ satisfies $\begin{bmatrix} \mathrm{PR}(A) \\ \mathrm{PR}(B) \\ \mathrm{PR}(C) \\ \mathrm{PR}(D) \end{bmatrix} = T \begin{bmatrix} \mathrm{PR}(A) \\ \mathrm{PR}(B) \\ \mathrm{PR}(C) \\ \mathrm{PR}(D) \end{bmatrix}$.

  It is an eigenvector of the transition matrix T with eigenvalue 1.

- $\begin{bmatrix} -1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -1 & 0 & 0 \\ \frac{1}{3} & 0 & -1 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 \end{bmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -\frac{2}{3} \\ 0 & 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$

  $\Longrightarrow$ eigenspace of $\lambda = 1$ spanned by $\begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix}$

  $\Longrightarrow \begin{bmatrix} \mathrm{PR}(A) \\ \mathrm{PR}(B) \\ \mathrm{PR}(C) \\ \mathrm{PR}(D) \end{bmatrix} = \frac{3}{16} \begin{bmatrix} 2 \\ \frac{2}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix}$ This the PageRank vector.

- The corresponding ranking of the webpages is A, C, D, B.

Remark 1. In practical situations, the system might be too large for finding the eigenvector by elimination.

- Google reports having met about 60 trillion webpages

  Google's search index is over 100,000,000 gigabytes

  Number of Google's servers secret; about 2,500,000

  More than 1,000,000,000 websites (i.e. hostnames; about 75% not active)

Armin Straub
astraub@illinois.edu

- Thus we have a gigantic but very sparse matrix.

An alternative to elimination is the **power method**:

If T is an (acyclic and irreducible) Markov matrix, then for any $\boldsymbol{v}_0$ the vectors $T^n\boldsymbol{v}_0$ converge to an eigenvector with eigenvalue 1.

Here: $T = \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$

$$\begin{bmatrix} \mathrm{PR}(A) \\ \mathrm{PR}(B) \\ \mathrm{PR}(C) \\ \mathrm{PR}(D) \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix}$$

$$T\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 3/8 \\ 1/12 \\ 1/3 \\ 5/24 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.083 \\ 0.333 \\ 0.208 \end{bmatrix}$$

Note that the ranking of the webpages is already A, C, D, B if we stop here.

$$T^2\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.125 \\ 0.333 \\ 0.167 \end{bmatrix}$$

$$T^3\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.396 \\ 0.125 \\ 0.292 \\ 0.188 \end{bmatrix}$$

Remark 2.

- If all entries of T are positive, then the power method is guaranteed to work.
- In the context of PageRank, we can make sure that this is the case, by replacing T with

$$(1-p)\cdot\begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} + p\cdot\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

  Just to make sure: still a Markov matrix, now with positive entries

  Google used to use $p = 0.15$.

- Why does $T^n\boldsymbol{v}_0$ converge to an eigenvector with eigenvalue 1?

  Under the assumptions on T, its other eigenvalues λ satisfy $|\lambda| < 1$.

  Now, think in terms of a basis $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of eigenvectors:

  $T^m(c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n) = c_1\lambda_1^m\boldsymbol{x}_1 + \ldots + c_n\lambda_n^m\boldsymbol{x}_n$

  As m increases, the terms with λ_i^m for $\lambda_i \neq 1$ go to zero, and what is left over is an eigenvector with eigenvalue 1.

Linear differential equations

Example 3. Which functions $y(t)$ satisfy the **differential equation** $y' = y$?

Solution: $y(t) = e^t$ and, more generally, $y(t) = Ce^t$. (And nothing else.)

Recall from Calculus the Taylor series $e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots$

Example 4. The differential equation $y' = ay$ with **initial condition** $y(0) = C$ is solved by $y(t) = Ce^{at}$. (This solution is unique.)

Why? Because $y'(t) = aCe^{at} = ay(t)$ and $y(0) = C$.

Example 5. Our goal is to solve (systems of) differential equations like:

$$\begin{array}{rcrrr} y_1' & = & 2y_1 & & \\ y_2' & = & -y_1 & +3y_2 & +y_3 \\ y_3' & = & -y_1 & +y_2 & +3y_3 \end{array} \qquad \begin{array}{rcl} y_1(0) & = & 1 \\ y_2(0) & = & 0 \\ y_3(0) & = & 2 \end{array}$$

In matrix form:

$$\boldsymbol{y}' = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 1 \\ -1 & 1 & 3 \end{bmatrix} \boldsymbol{y}, \qquad \boldsymbol{y}(0) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

Key idea: to solve $\boldsymbol{y}' = A\boldsymbol{y}$, introduce e^{At}

Review of diagonalization

- If $A\boldsymbol{x} = \lambda\boldsymbol{x}$, then $\boldsymbol{x}$ is an **eigenvector** of A with **eigenvalue** λ.
- Put the eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ as columns into a matrix P.

$$A\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i \quad \Longrightarrow \quad A\begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n \\ | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- In summary: $AP = PD$

Let A be $n \times n$ with independent eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.
Then A can be **diagonalized** as $A = PDP^{-1}$.
- the columns of P are the eigenvectors
- the diagonal matrix D has the eigenvalues on the diagonal

Example 6. Diagonalize the following matrix, if possible.

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

Solution.

- A has eigenvalues 2 and 4. (We did that in an earlier class!)
  - $\lambda = 2$: $\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \Longrightarrow$ eigenspace $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$
  - $\lambda = 4$: $\begin{bmatrix} -2 & 0 & 0 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \Longrightarrow$ eigenspace $\operatorname{span}\left\{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right\}$
- $P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 4 \end{bmatrix}$
- $A = PDP^{-1}$

  For many applications, it is not needed to compute P^{-1} explicitly.
- We can check this by verifying $AP = PD$:

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

Review

Let A be $n \times n$ with independent eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

Then A can be **diagonalized** as $A = PDP^{-1}$.

- the columns of P are the eigenvectors
- the diagonal matrix D has the eigenvalues on the diagonal

Why? We need to see that $AP = PD$:

$$A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i \quad \Longrightarrow \quad A\begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ | & & | \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- The **differential equation** $y' = ay$ with **initial condition** $y(0) = C$ is solved by $y(t) = Ce^{at}$.

  Recall from Calculus the Taylor series $e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots$

- Goal: similar treatment of systems like:

$$\boldsymbol{y}' = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 1 \\ -1 & 1 & 3 \end{bmatrix}\boldsymbol{y}, \qquad \boldsymbol{y}(0) = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

Definition 1. Let A be $n \times n$. The **matrix exponential** is

$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

Then: $\frac{\mathrm{d}}{\mathrm{d}t}e^{At} = Ae^{At}$

Why?

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{At} = \frac{\mathrm{d}}{\mathrm{d}t}\left(I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \cdots\right) = A + \frac{1}{1!}A^2t + \frac{1}{2!}A^3t^2 + \cdots = Ae^{At}$$

The solution to $\boldsymbol{y}' = A\boldsymbol{y}$, $\boldsymbol{y}(0) = \boldsymbol{y}_0$ is $\boldsymbol{y}(t) = e^{At}\boldsymbol{y}_0$.

Why? Because $\boldsymbol{y}'(t) = Ae^{At}\boldsymbol{y}_0 = A\boldsymbol{y}(t)$ and $\boldsymbol{y}(0) = e^{0A}\boldsymbol{y}_0 = \boldsymbol{y}_0$.

Example 2. If $A = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$, then:

$$e^A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} + \frac{1}{2!}\begin{bmatrix} 2^2 & 0 \\ 0 & 5^2 \end{bmatrix} + \cdots = \begin{bmatrix} e^2 & 0 \\ 0 & e^5 \end{bmatrix}$$

$$e^{At} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2t & 0 \\ 0 & 5t \end{bmatrix} + \frac{1}{2!}\begin{bmatrix} (2t)^2 & 0 \\ 0 & (5t)^2 \end{bmatrix} + \cdots = \begin{bmatrix} e^{2t} & 0 \\ 0 & e^{5t} \end{bmatrix}$$

Clearly, this works to obtain e^D for any diagonal matrix D.

Armin Straub
astraub@illinois.edu

Example 3. Suppose $A = PDP^{-1}$. Then, what is A^n?

Solution.

First, note that $A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1}$.

Likewise, $A^n = PD^nP^{-1}$.

(The point being that D^n is trivial to compute because D is diagonal.)

Theorem 4. Suppose $A = PDP^{-1}$. Then, $e^A = Pe^DP^{-1}$.

Why? Recall that $A^n = PD^nP^{-1}$.

$$\begin{aligned} e^A &= I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots \\ &= I + PDP^{-1} + \frac{1}{2!}PD^2P^{-1} + \frac{1}{3!}PD^3P^{-1} + \cdots \\ &= P\left(I + D + \frac{1}{2!}D^2 + \frac{1}{3!}D^3 + \cdots\right)P^{-1} = Pe^DP^{-1} \end{aligned}$$

Example 5. Solve the differential equation

$$\boldsymbol{y}' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\boldsymbol{y}, \qquad \boldsymbol{y}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Solution. The solution to $\boldsymbol{y}' = A\boldsymbol{y}$, $\boldsymbol{y}(0) = \boldsymbol{y}_0$ is $\boldsymbol{y}(t) = e^{At}\boldsymbol{y}_0$.

- Diagonalize $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$:
  - $\begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 1$, so the eigenvalues are ± 1
  - $\lambda = 1$ has eigenspace $\mathrm{Nul}\left(\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\right) = \mathrm{span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$
  - $\lambda = -1$ has eigenspace $\mathrm{Nul}\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = \mathrm{span}\left\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$
  - Hence, $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.
- Compute the solution $\boldsymbol{y} = e^{At}\boldsymbol{y}_0$:

$$\begin{aligned} \boldsymbol{y} &= Pe^{Dt}P^{-1}\boldsymbol{y}_0 \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} e^t \\ -e^{-t} \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} e^t + e^{-t} \\ e^t - e^{-t} \end{bmatrix} \end{aligned}$$

Armin Straub
astraub@illinois.edu

Example 6. Solve the differential equation

$$\boldsymbol{y}' = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 1 \\ -1 & 1 & 3 \end{bmatrix} \boldsymbol{y}, \qquad \boldsymbol{y}(0) = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

Solution.

- Recall that the solution to $\boldsymbol{y}' = A\boldsymbol{y}$, $\boldsymbol{y}(0) = \boldsymbol{y}_0$ is $\boldsymbol{y} = e^{At}\boldsymbol{y}_0$.
- A has eigenvalues 2 and 4. (We did that in an earlier class!)
  - $\lambda = 2$: $\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \Longrightarrow$ eigenspace $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$
  - $\lambda = 4$: $\begin{bmatrix} -2 & 0 & 0 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \Longrightarrow$ eigenspace $\operatorname{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$
- $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 4 \end{bmatrix}$
- Compute the solution $\boldsymbol{y} = e^{At}\boldsymbol{y}_0$:

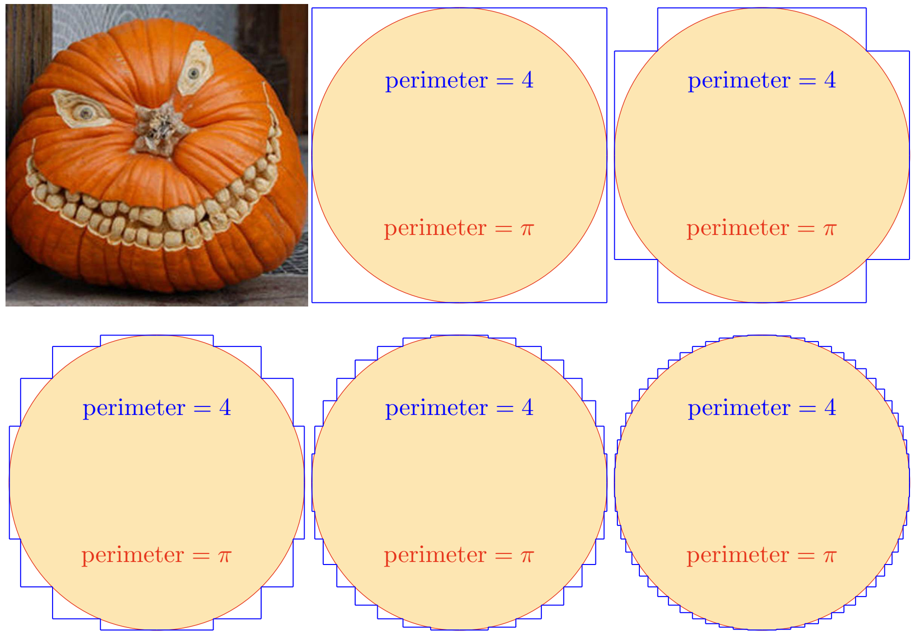
$$\begin{aligned} \boldsymbol{y} = e^{At}\boldsymbol{y}_0 &= Pe^{Dt}P^{-1}\boldsymbol{y}_0 \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} e^{2t} & & \\ & e^{2t} & \\ & & e^{4t} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} e^{2t} & & \\ & e^{2t} & \\ & & e^{4t} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} e^{2t} \\ 0 \\ e^{4t} \end{bmatrix} = \begin{bmatrix} e^{2t} \\ e^{2t} + e^{4t} \\ e^{4t} \end{bmatrix} \end{aligned}$$

Check (optional) that $\boldsymbol{y} = \begin{bmatrix} e^{2t} \\ e^{2t} + e^{4t} \\ e^{4t} \end{bmatrix}$ indeed solves the original problem:

$$\boldsymbol{y}' = \begin{bmatrix} 2e^{2t} \\ 2e^{2t} + 4e^{4t} \\ 4e^{4t} \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & 1 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} e^{2t} \\ e^{2t} + e^{4t} \\ e^{4t} \end{bmatrix}$$

Remark 7. The matrix exponential shares many other properties of the usual exponential:

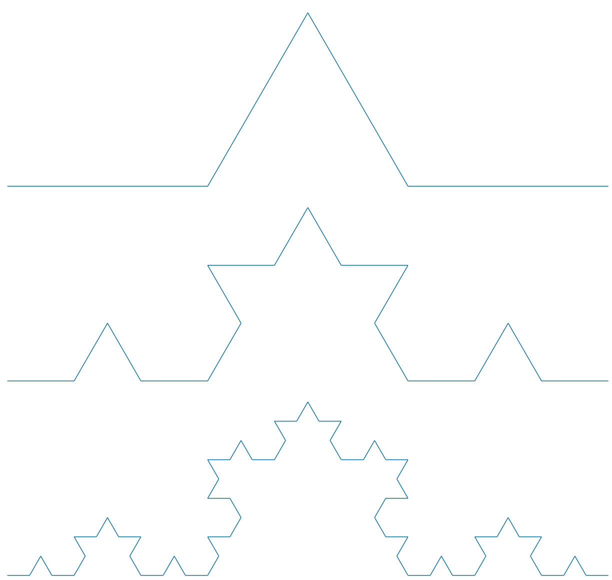
- e^A is invertible and $(e^A)^{-1} = e^{-A}$
- $e^A e^B = e^{A+B} = e^B e^A$ if $AB = BA$

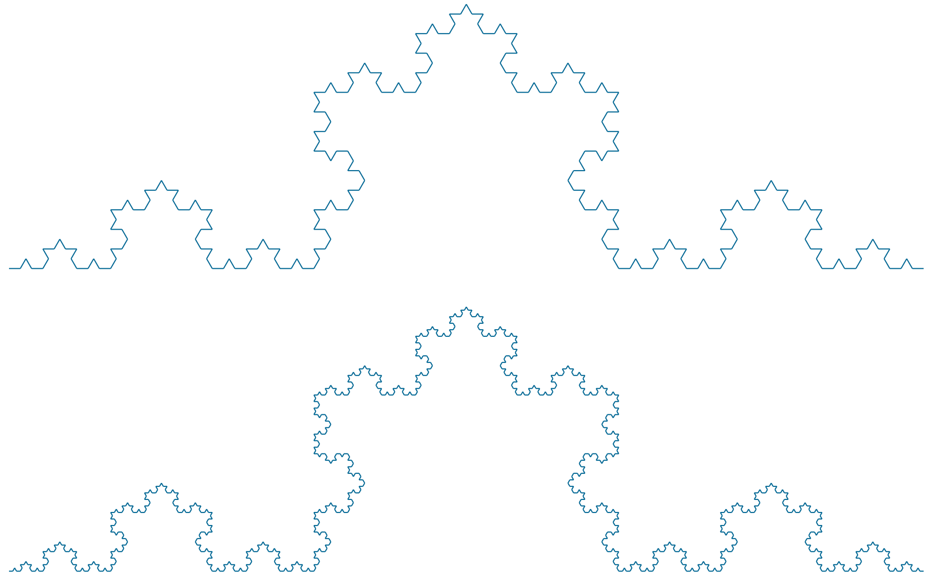


- Length of the graph of $y(x)$ on $[a, b]$ is $\int_a^b \sqrt{1 + y'(x)^2}\mathrm{d}x$.
- While the blue curve does converge to the circle,
  its derivative does not converge!
- In the language of functional analysis:
  The linear map $D\colon y \mapsto y'$ is not continuous!

  (That is, two functions can be close without their derivatives being close.)

Even more extreme examples are provided by **fractals**. The **Koch snowflake**:



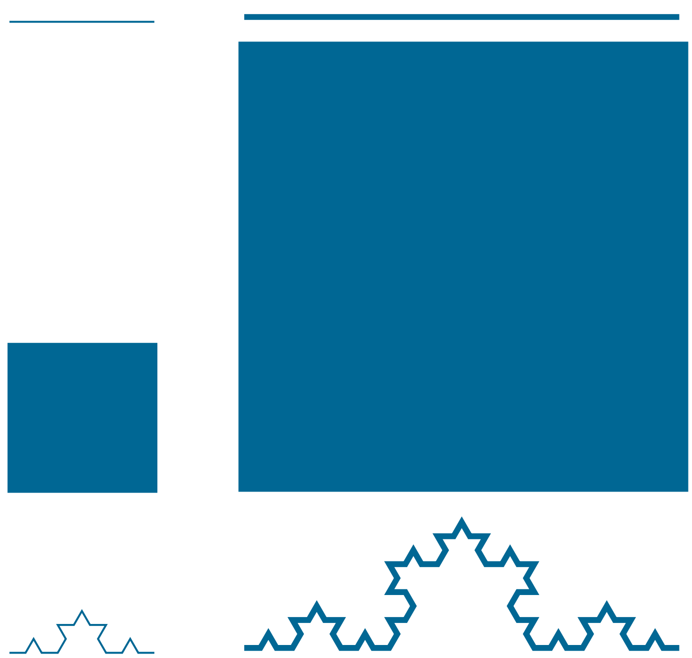


- Its perimeter is infinite!

  Why? At each iteration, the perimeter gets multiplied by $4/3$.

- Its boundary has dimension $\log_3(4) \approx 1.262$!!

the effect of zooming in by a factor of 3



$\times 3$ $\quad d = 1 = \log_3(3)$

$\times 9$ $\quad d = 2 = \log_3(3)$

$\times 4$ $\quad d = \log_3(4)$

- Such fractal behaviour is also observed when attempting to measure the length of a coastline: the measured length increases by a factor when using a smaller scale.

  See: `http://en.wikipedia.org/wiki/Coastline_paradox`