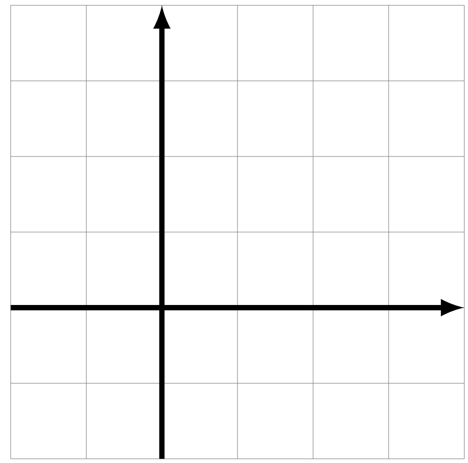
Systems of linear equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

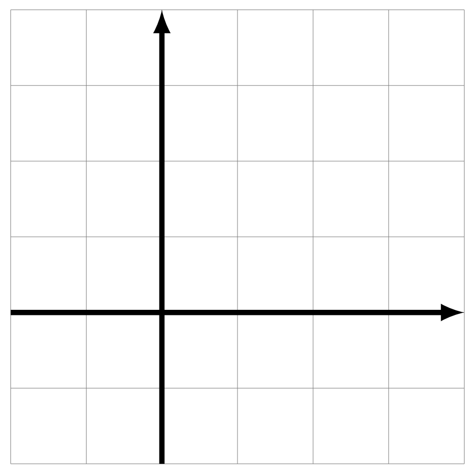
Question: How many solutions a system of linear equations can have?

Example: Systems of equations in 2 variables.

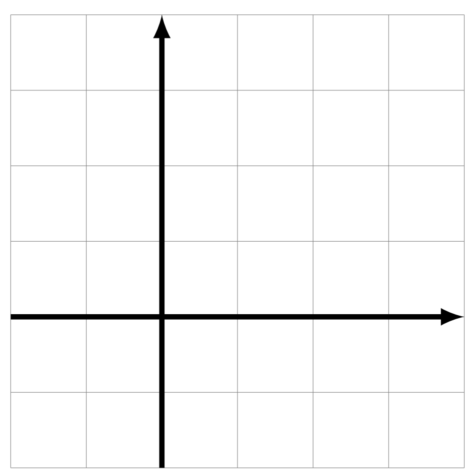
$$\begin{cases} x_1 + x_2 = 1 \\ x_1 - x_2 = 1 \end{cases}$$



$$\begin{cases} x_1 + x_2 = 1 \\ x_1 + x_2 = 2 \end{cases}$$

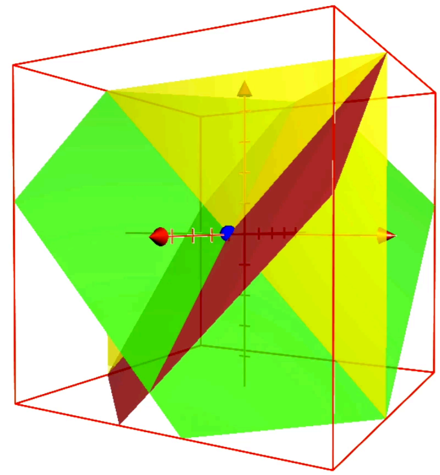


$$\begin{cases} x_1 + x_2 = 1 \\ 2x_1 + 2x_2 = 2 \end{cases}$$

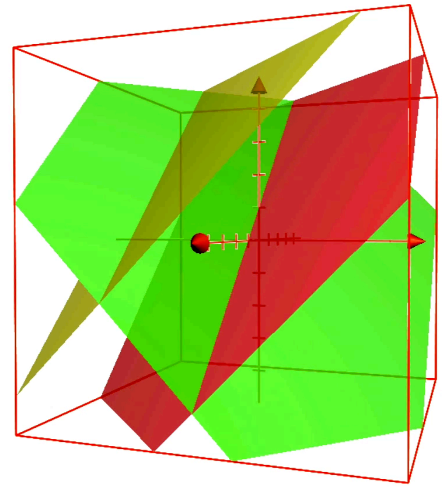


Example: Systems of equations in 3 variables.

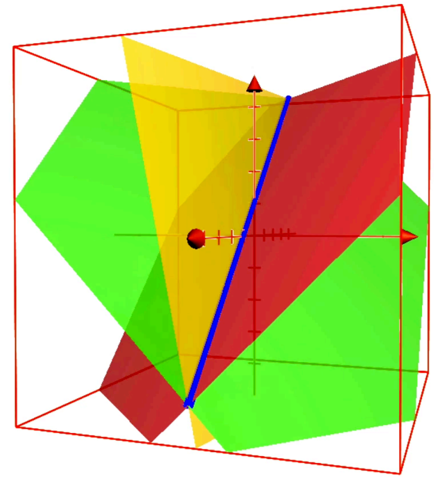
$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 - x_2 + x_3 = 1 \\ x_1 = 1 \end{cases}$$



$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 - x_2 + x_3 = 1 \\ x_1 - x_2 + x_3 = 6 \end{cases}$$



$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 - x_2 + x_3 = 1 \\ x_1 + 5x_2 + x_3 = 1 \end{cases}$$



In general:

A system of linear equations can have either

- no solutions
- exactly one solution
- infinitely many solutions

Definition

If as system of linear equations which has no solutions is called an *inconsistent system*. Otherwise the system is *consistent*.

Next:

How to solve a system of linear equations

system of equations

$$\begin{cases} -x_1 + 2x_2 + 3x_3 = 4 \\ 2x_1 + 6x_3 = 9 \\ 4x_1 - x_2 - 3x_3 = 0 \end{cases}$$

→ *make a matrix* →

[augmented matrix]

↓ *Gauss-Jordan elimination* ↓

[matrix in reduced row echelon form]

← *read off solutions* ←

solutions

$$\begin{cases} x_1 = \dots \\ x_2 = \dots \\ x_3 = \dots \end{cases}$$

Matrices

matrix = rectangular array of numbers

Example.

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

Note

Every system of linear equations can be represented by a matrix.

Example.

$$\begin{cases} -x_1 + 2x_2 + 3x_3 = 4 \\ 2x_1 + 6x_3 = 9 \\ 4x_1 - x_2 - 3x_3 = 0 \end{cases}$$

Elementary row operations:

1) Interchange of two rows.

Example.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 1 \\ 4 & 3 & 0 & 7 \end{bmatrix}$$

2) Multiplication of a row by a non-zero number.

Example.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 1 \\ 4 & 3 & 0 & 7 \end{bmatrix}$$

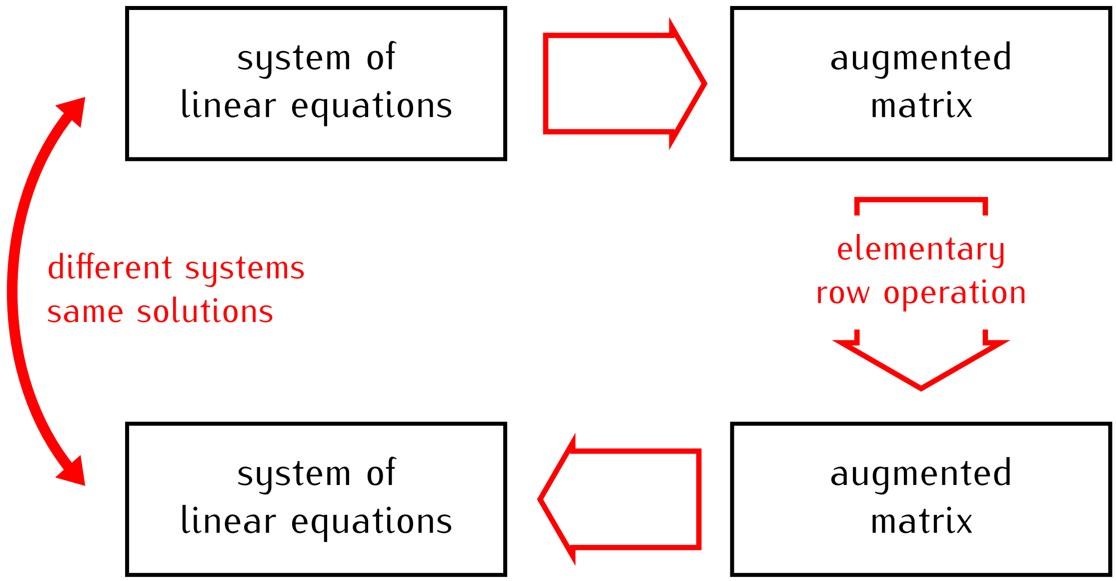
3) Addition of a multiple of one row to another row.

Example.

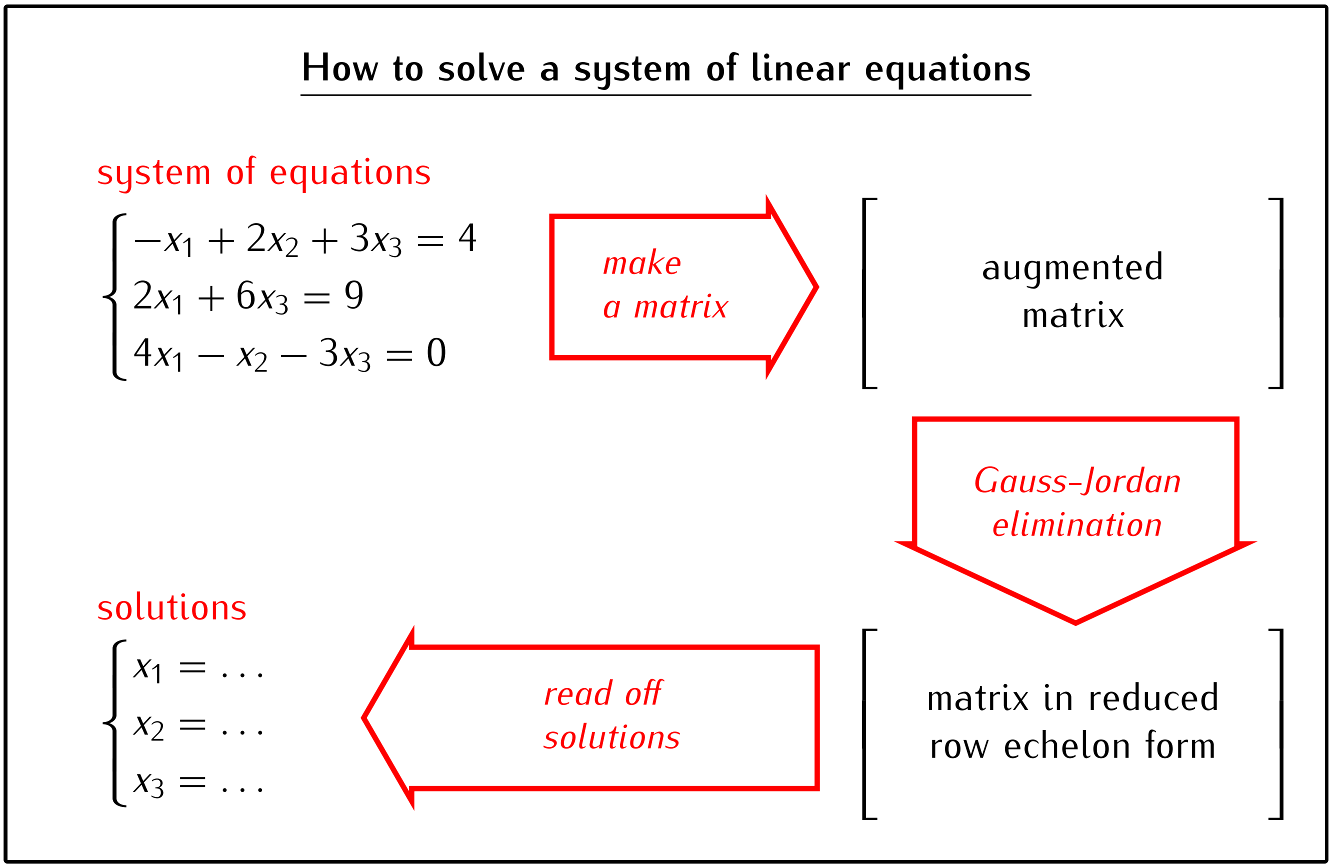
$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 1 \\ 4 & 3 & 0 & 7 \end{bmatrix}$$

Proposition

Elementary row operations do not change solutions of the system of equations represented by a matrix.



Recall:



- Every system of linear equations can be represented by a matrix

- Elementary row operations:
  - interchange of two rows
  - multiplication of a row by a non-zero number
  - addition of a multiple of one row to another row.

- Elementary row operations do not change solutions of systems of linear equations.

Definition

A matrix is in the *row echelon form* if:

1) the first non-zero entry of each row is a 1 ("a leading one");
2) the leading one in each row is to the right of the leading one in the row above it.

A matrix is in the *reduced row echelon form* if in addition it satisfies:

3) all entries above each leading one are 0.

$$\begin{bmatrix} 1 & * & * & * & 0 & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

($* =$ any number)

Example

$$\begin{bmatrix} 1 & 0 & 4 & 0 & 7 & 0 \\ 0 & 1 & 5 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 4 & 6 & 7 & 0 \\ 0 & 1 & 5 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 4 & 0 & 7 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 3 & 6 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Fact

If a system of linear equations is represented by a matrix in the reduced row echelon form then it is easy to solve the system.

Example

$$\left[\begin{array}{cccc|c} 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 7 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Proposition

A matrix in the reduced row echelon form represents an inconsistent system if and only if it contains a row of the form

$$\left[\begin{array}{cccccc} 0 & 0 & 0 \dots & 0 & 1 \end{array}\right]$$

i.e. with the leading one in the last column.

Example

$$\left[\begin{array}{cccc|c} 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 7 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Note

In an augmented matrix in the reduced row echelon form free variables correspond to columns of the coefficient matrix that do not contain leading ones.

Example

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 & 6 \\ 0 & 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 1 & 8 \end{array}\right]$$

Note

A matrix in the reduced row echelon form represents a system of equations with exactly one solution if and only if it has a leading one in every column except for the last one.

Gauss-Jordan elimination process (= row reduction)

[matrix] → *elementary row operations* → [matrix in the reduced row echelon form]

① Interchange rows, if necessary, to bring a non-zero element to the top of the first non-zero column of the matrix.

② Multiply the first row so that its first non-zero entry becomes 1.

③ Add multiples of the first row to eliminate non-zero entries below the leading one.

④ Ignore the first row; apply steps 1-3 to the rest of the matrix.

⑤ Eliminate non-zero entries above all leading ones.

Example.

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

How to solve systems of linear equations: example

$$\begin{cases} 4x_2 - 8x_3 = 4 \\ 2x_1 + 6x_2 - 6x_3 - 2x_4 = -4 \\ 2x_1 + 7x_2 - 8x_3 = -1 \end{cases}$$

$$\left[\begin{array}{cccc|c} 0 & 4 & -8 & 0 & 4 \\ 2 & 6 & -6 & -2 & -4 \\ 2 & 7 & -8 & 0 & -1 \end{array}\right] \xrightarrow{\text{row reduction}} \left[\begin{array}{cccc|c} 1 & 0 & 3 & 0 & -4 \\ 0 & 1 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

Definition

A *pivot position* in a matrix is a position that after row reduction contains a leading one.

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

Theorem

1) A system of linear equations is inconsistent if and only if the last column of its augmented matrix is a pivot column.

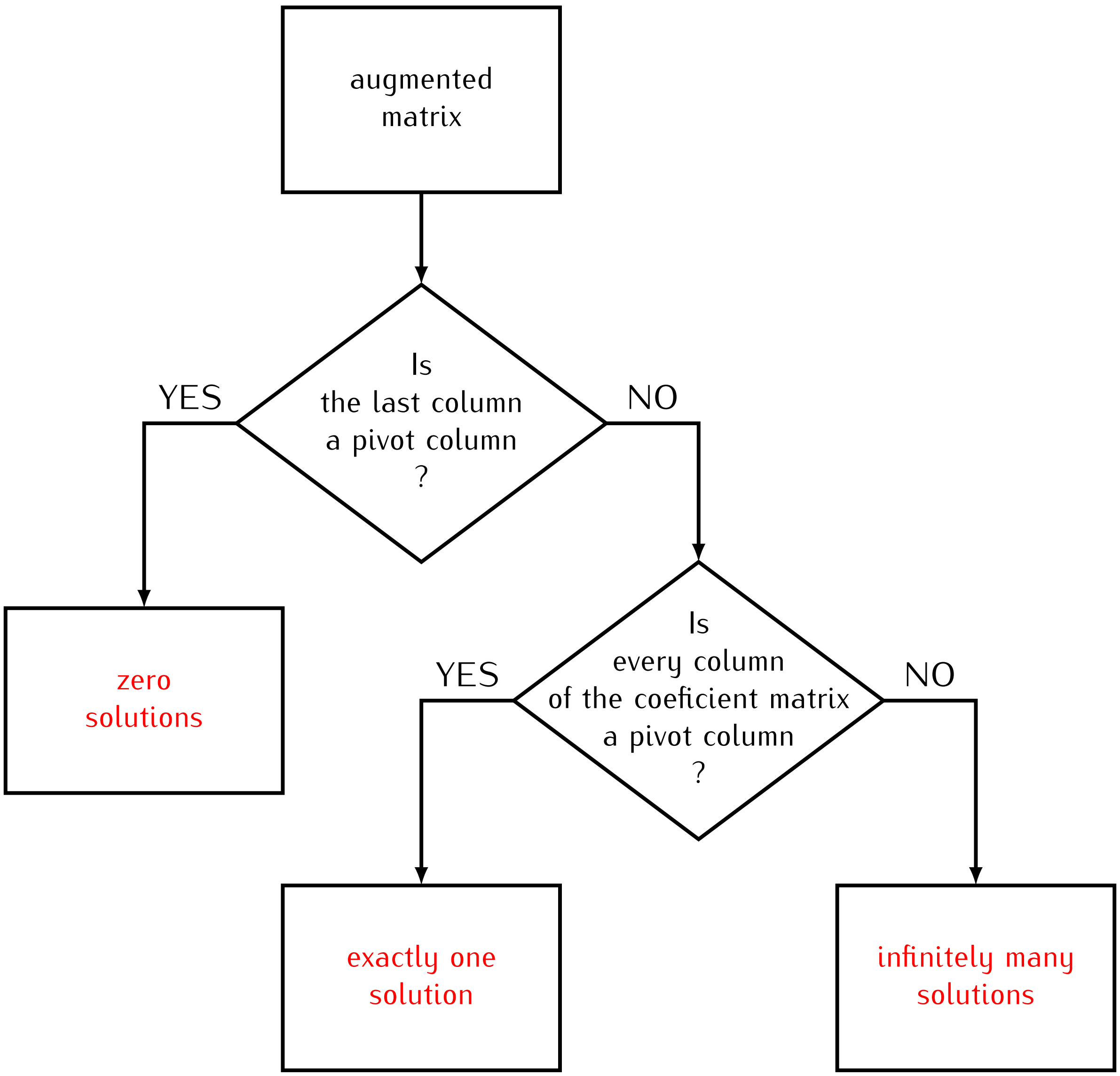
2) Free variables of the system correspond to non-pivot columns of the coefficient matrix.

3) The system has only one solution if and only if every column of its augmented matrix is a pivot column, except for the last column.

Theorem

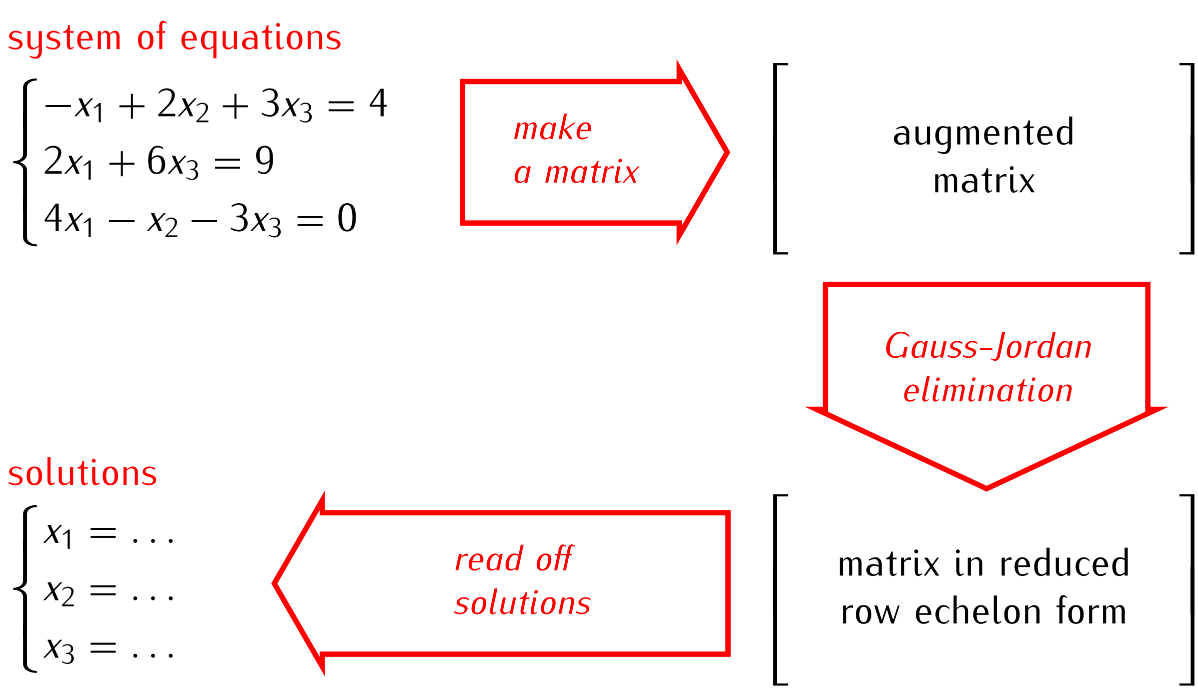
A system of linear equations can have either 0, 1, or infinitely many solutions.

Proof.



Recall:

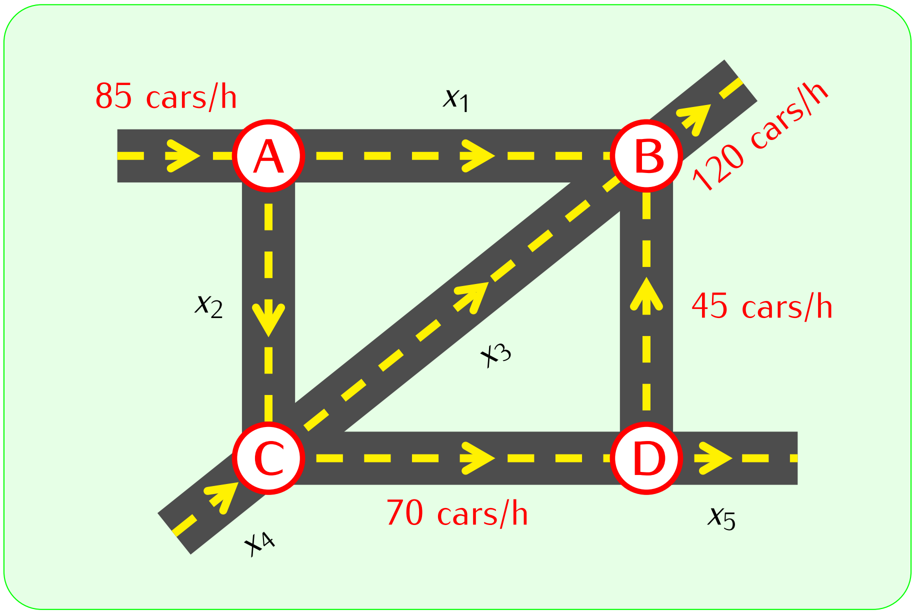
How to solve a system of linear equations



Next: Some applications of systems of linear equations:

- Computations of traffic flow.
- Balancing chemical equations.
- Google PageRank.

Computations of traffic flow



Problem. Find the flow rate of cars on each segment of streets.

Note:

- flow into an intersection = flow out of that intersection
- total flow in = total flow out

Balancing chemical equations

Burning propane:

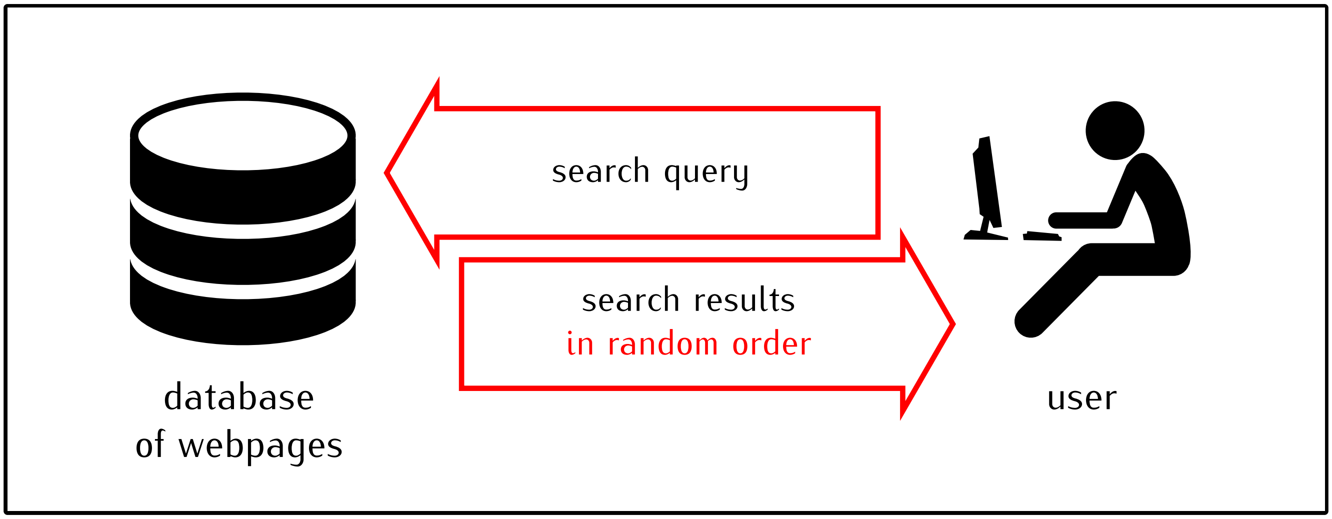
$$x_1 C_3H_8 + x_2 O_2 \longrightarrow x_3 CO_2 + x_4 H_2O$$

Note:

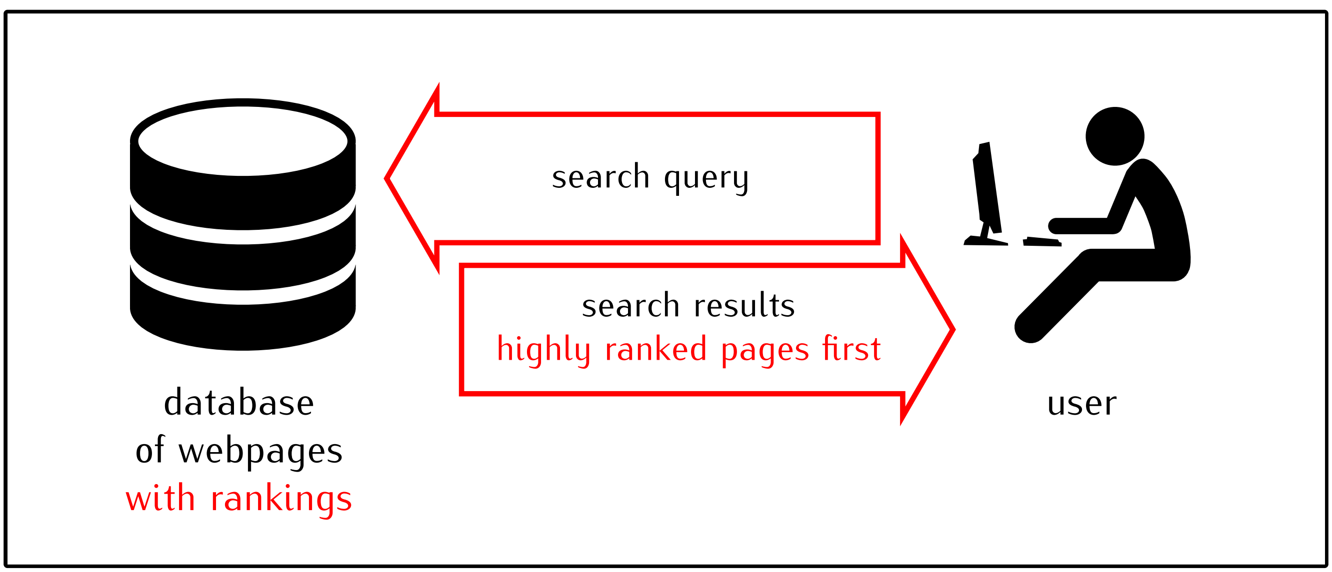
- The numbers x_1, x_2, x_3, x_4 are positive integers.
- The number of atoms of each element on the left side is the same as the number of atoms of that element on the right side.

Google PageRank

Early search engines:



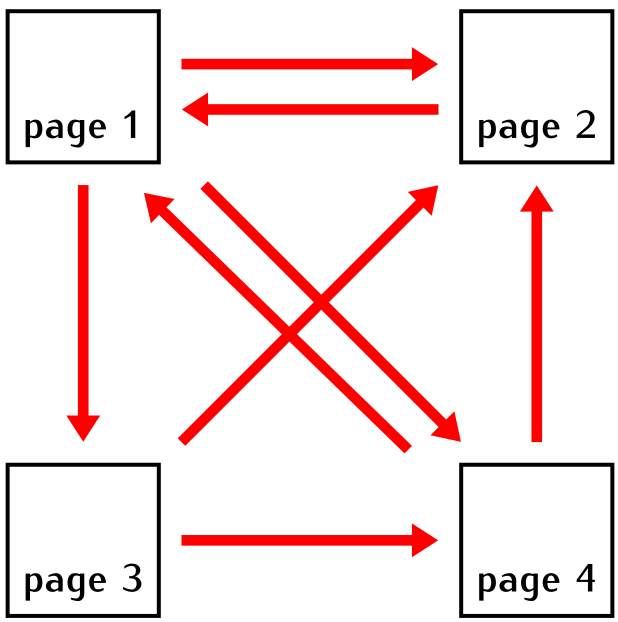
Google search engine:



How to rank webpages?

Very simple ranking:

$$\text{ranking of a page} = \begin{pmatrix} \text{number of links} \\ \text{pointing to that page} \end{pmatrix}$$



Network of web pages.

Problem. This is very easy to manipulate.

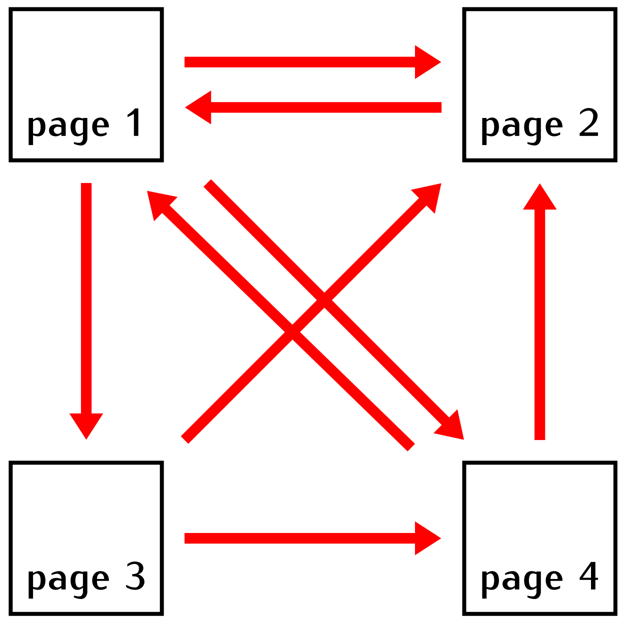
How to rank webpages?

Google PageRank: Links from highly ranked pages are worth more than links from lower ranked pages.

If:

- the rank of a page is x
- the page has n links to other pages

then each link from that page is worth x/n.



Next: From systems of linear equations to vector equations.

$$\begin{cases} x_1 + 2x_2 = 4 \\ 2x_1 + 7x_2 = 9 \\ 4x_1 + x_2 = 0 \end{cases} \quad \Longrightarrow \quad x_1 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 7 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \\ 0 \end{bmatrix}$$

Why vectors and vector equations are useful:

- They show up in many applications (velocity vectors, force vectors etc.)
- They give a better geometric picture of systems of linear equations.

Definition

A *column vector* is a matrix with one column.

Note. Columns of a matrix are column vectors.

Notation

$\mathbb{R}^n$ is the set of all column vectors with n entries.

Operations on vectors in $\mathbb{R}^n$

1) Addition of vectors:

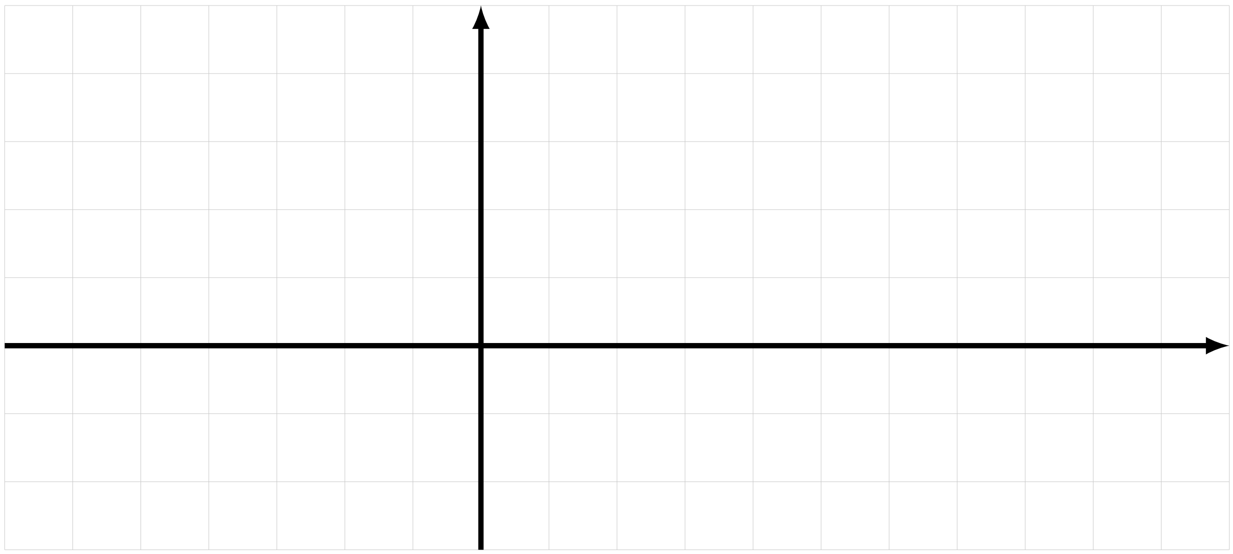
$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix}$$

2) Multiplication by scalars:

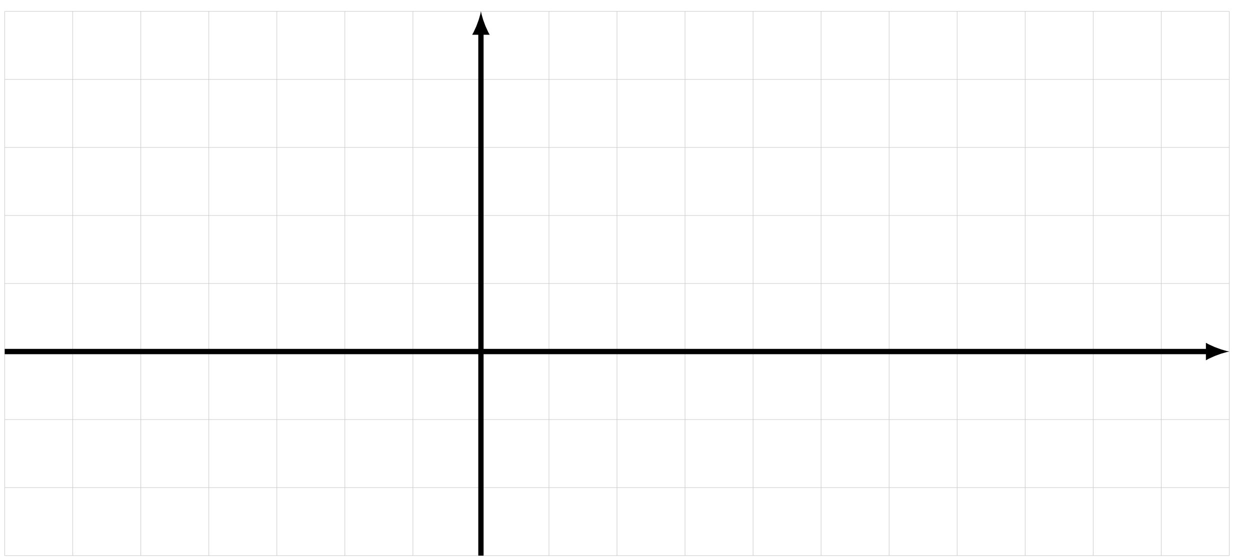
$$c \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} ca_1 \\ \vdots \\ ca_n \end{bmatrix}$$

Geometric interpretation of vectors in $\mathbb{R}^2$

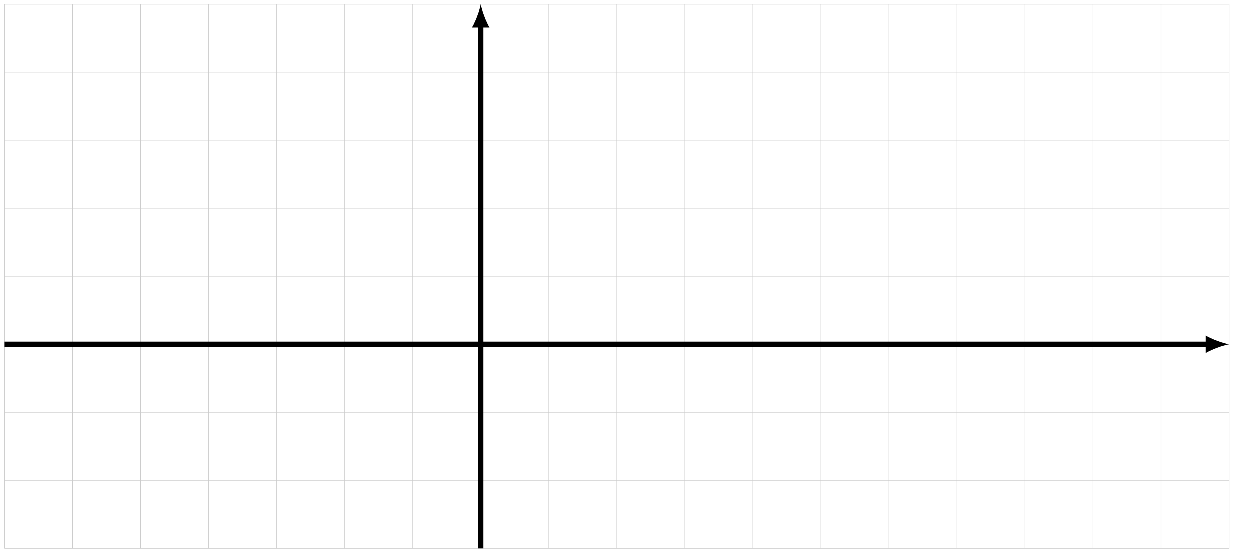
Vector coordinates:



Vector addition:



Scalar multiplication:



Vector equations

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

Example. Solve the following vector equation:

$$x_1\begin{bmatrix} 2 \\ 3 \end{bmatrix} + x_2\begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$$

How to solve a vector equation

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{w}$$

vector of equation

make a matrix

$$\left[\ \mathbf{v}_1 \quad \ldots \quad \mathbf{v}_p \ \middle|\ \mathbf{w}\ \right]$$

augmented matrix

row reduction

[reduced matrix]

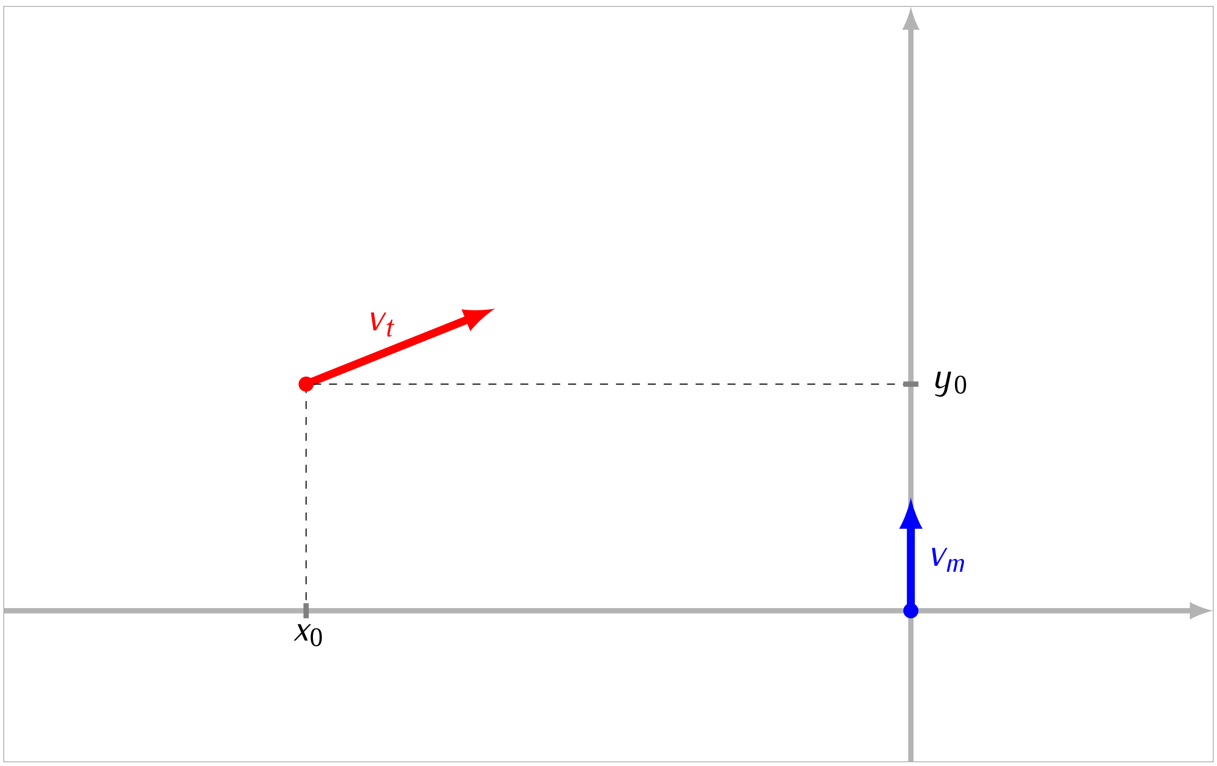
read off solutions

$$\left\{ \begin{array}{ll} x_1 & = \ldots \\ \ldots & \ \ \ \ldots \\ x_p & = \ldots \end{array} \right.$$

solutions

Example: Target shooting.

At time $t = 0$ a target is observed at the position (x_0, y_0) moving in the direction of the vector v_t. The target is moving at such speed, that it travels the length of v_t in one second. A missile is positioned at the point $(0, 0)$. When fired, it will move vertically with such speed, that it will travel the length of the vector v_m in one second. After how many seconds should the missile be fired in order to intercept the target?



Recall:

Vector equations are equivalent to systems of linear equations:

$$x_1 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix} \quad \Longleftrightarrow \quad \begin{cases} 2x_1 + 4x_2 = 7 \\ 3x_1 + 2x_2 = 3 \end{cases}$$

vector equation

system of linear equations

Upshot. A vector equation can have either:

- no solutions
- exactly one solution
- infinitely many solutions

Next:

- When does a vector equation have a solution?
- When does it have exactly one solution?

Definition

A vector $\mathbf{w} \in \mathbb{R}^n$ is a *linear combination* of vectors $\mathbf{v}_1, \ldots \mathbf{v}_p \in \mathbb{R}^n$ if there exists scalars $c_1, \ldots, c_p$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p$$

Equivalently: A vector $\mathbf{w}$ is a linear combination of vectors $\mathbf{v}_1, \ldots \mathbf{v}_p$ is the vector equation

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{w}$$

has a solution.

Example.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 5 \\ 0 \\ 3 \end{bmatrix}$$

Example. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 5 \\ 0 \\ 3 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 9 \\ 3 \\ 6 \end{bmatrix}$$

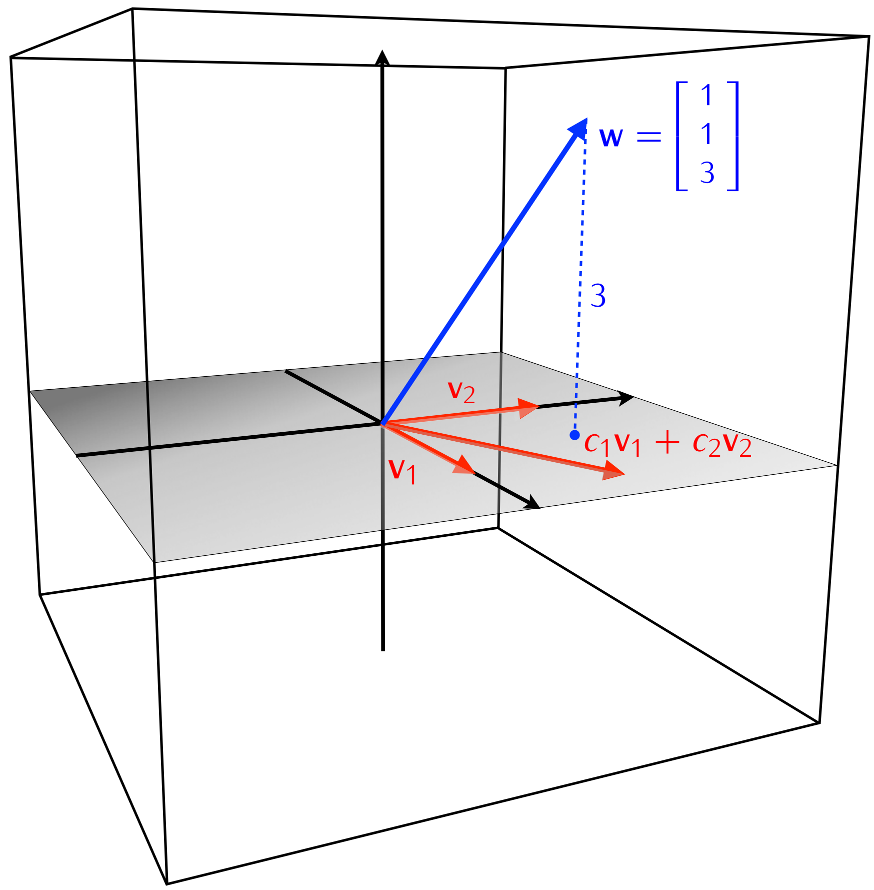
Express $\mathbf{w}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ or show that this is not possible.

Example. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

Express $\mathbf{w}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$ or show that this is not possible.

Geometric picture of the last example



Definition

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are vectors in $\mathbb{R}^n$ then

$$\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p) = \left\{ \begin{array}{c} \text{the set of all} \\ \text{linear combinations} \\ c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p \end{array} \right\}$$

Example.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

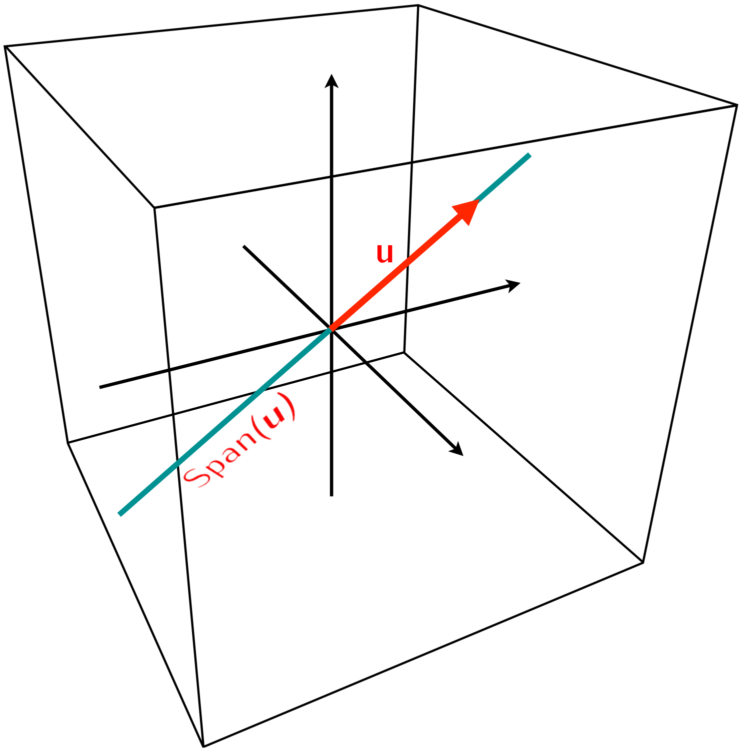
Proposition

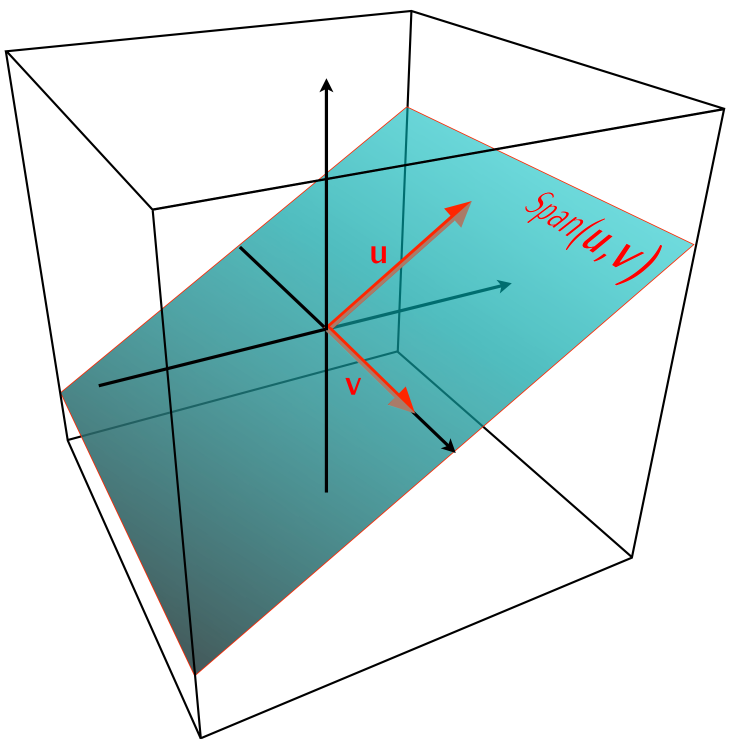
A vector $\mathbf{w}$ is in $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$ if and only if the vector equation

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{w}$$

has a solution.

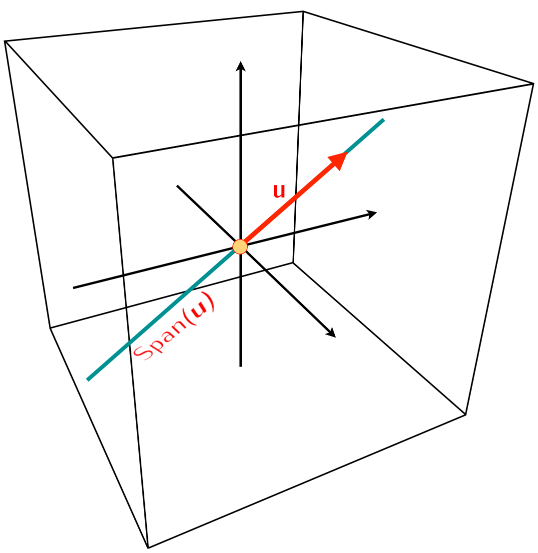
Geometric interpretation of Span

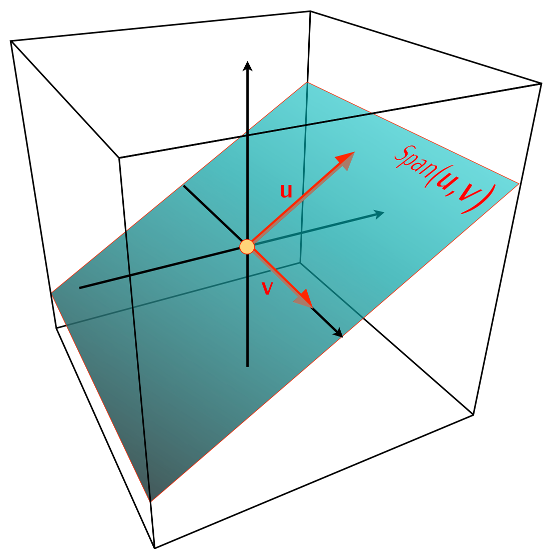




Proposition

For arbitrary vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ the zero vector $\mathbf{0} \in \mathbb{R}^n$ is in $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$.





Definition

A *homogenous vector equation* is a vector equation of the form

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

(i.e. with the zero vector as the vector of constants).

Definition

Let $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$. The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is *linearly independent* if the homogenous equation

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

has only one, trivial solution $x_1 = 0, \ldots, x_p = 0$. Otherwise the set is *linearly dependent*.

Theorem

Let $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$. If the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent then the equation

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{w}$$

has exactly one solution for any vector $\mathbf{w} \in \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$.

If the set is linearly dependent then this equation has infinitely many solutions for any $\mathbf{w} \in \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$.

Example. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 3 \\ -12 \end{bmatrix}$$

Check is the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

Note

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent if and only if every column of the matrix

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_p \end{bmatrix}$$

is a pivot column.

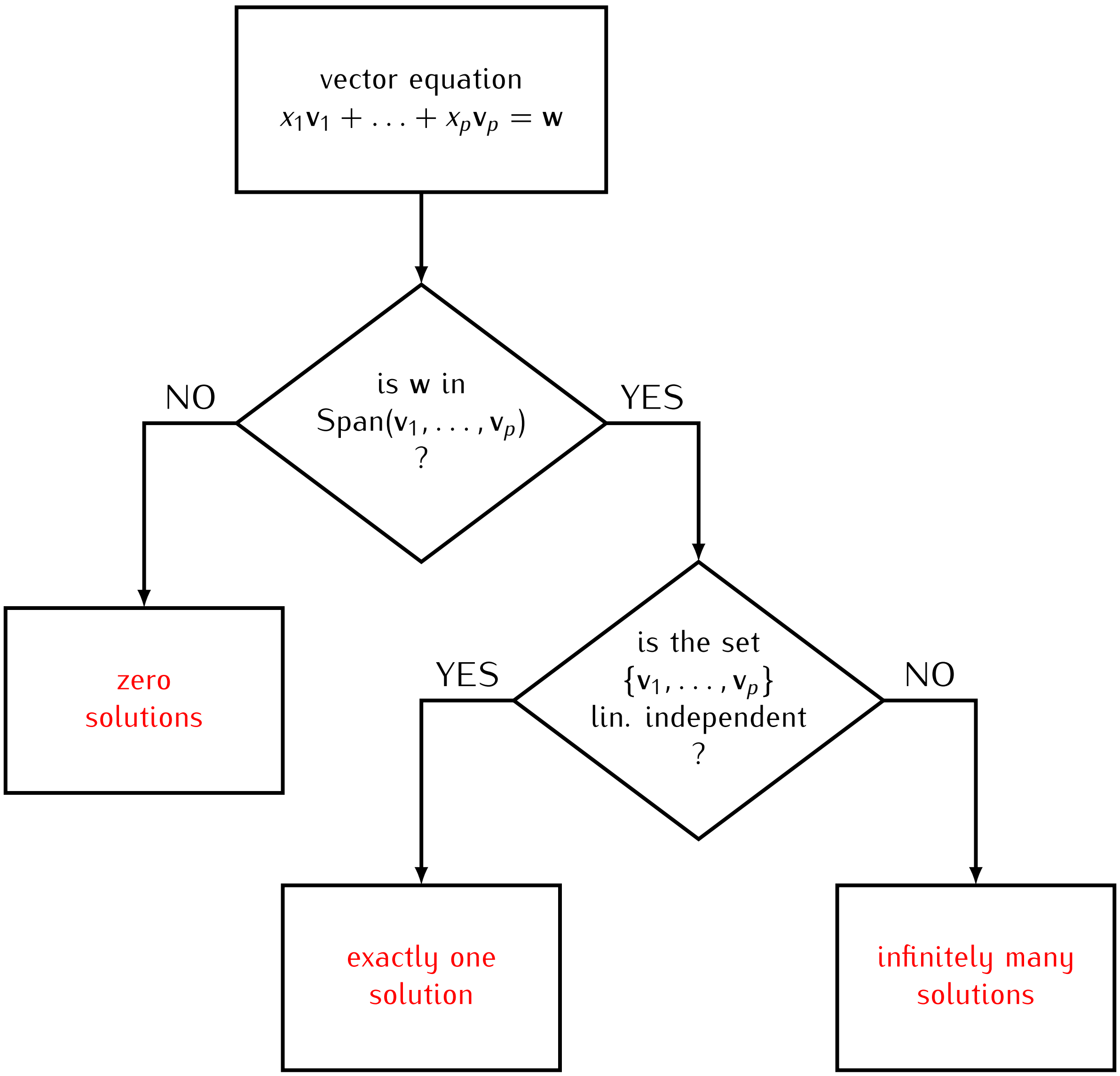
Some properties of linearly (in)dependent sets

1) A set consisting of one vector $\{\mathbf{v}_1\}$ is linearly dependent if and only if $\mathbf{v}_1 = \mathbf{0}$.

2) A set consisting of two vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if and only if one vector is a scalar multiple of the other.

3) If $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a set of p vectors in $\mathbb{R}^n$ and $p > n$ then this set is linearly dependent.

Upshot: how to find the number of solutions of a vector equation



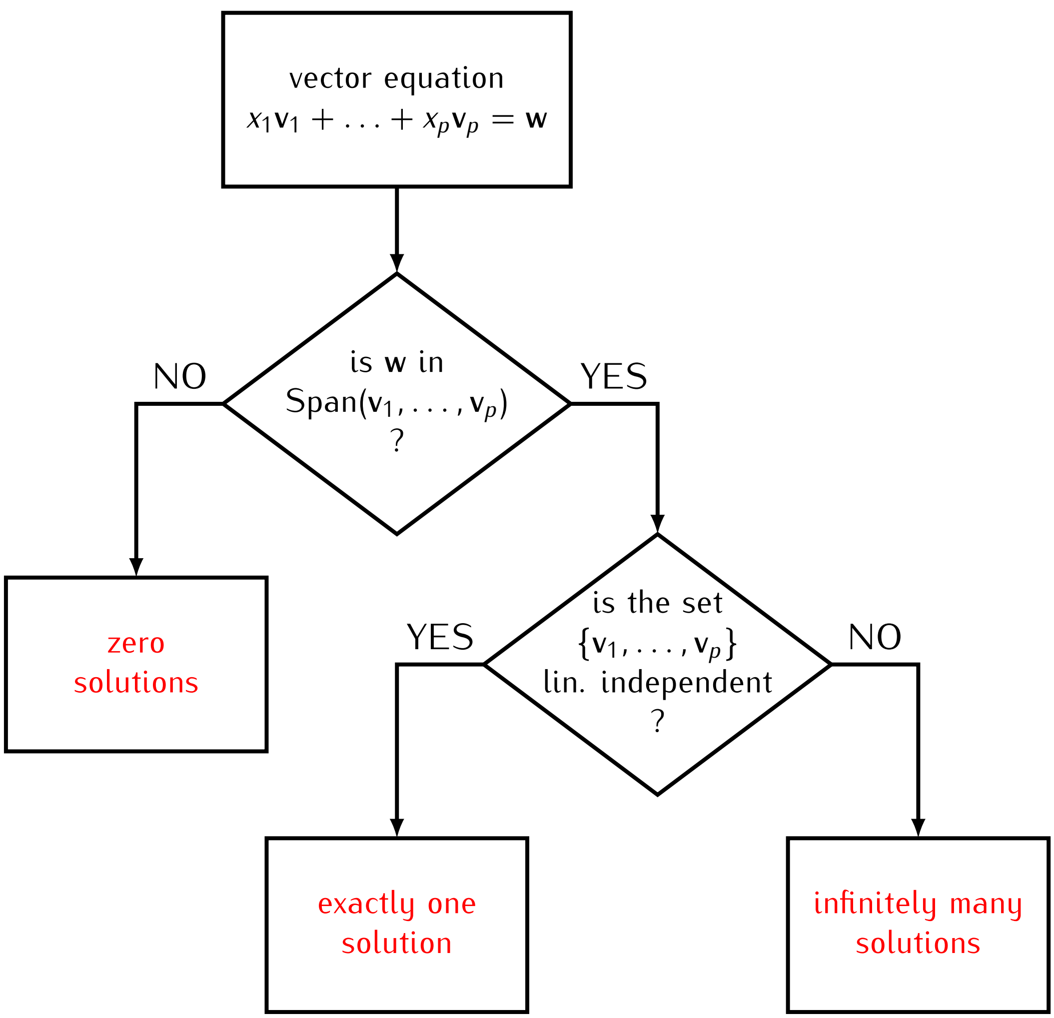
Recall:

1) $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p) = \left\{ \begin{array}{c} \text{the set of all} \\ \text{linear combinations} \\ c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p \end{array} \right\}$

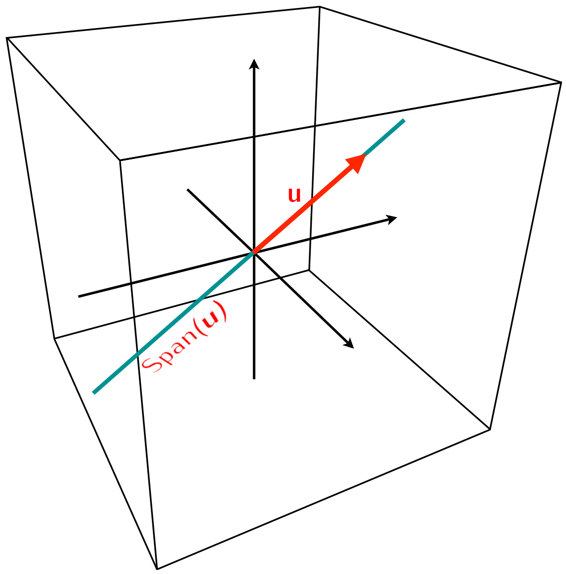
2) A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent if the equation

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

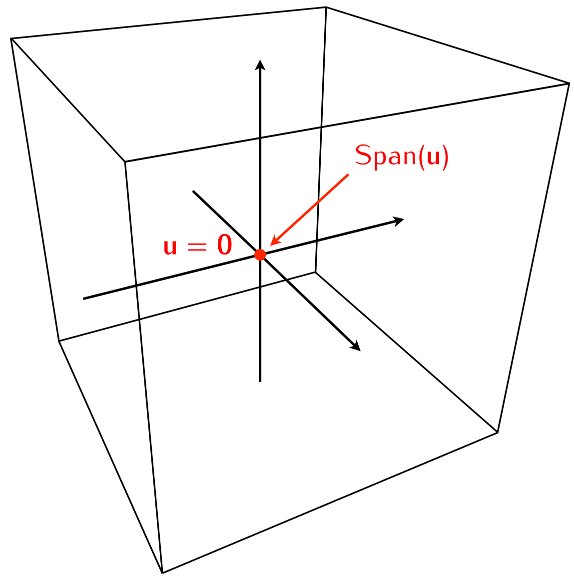
has only one, trivial solution $x_1 = 0, \ldots, x_p = 0$.



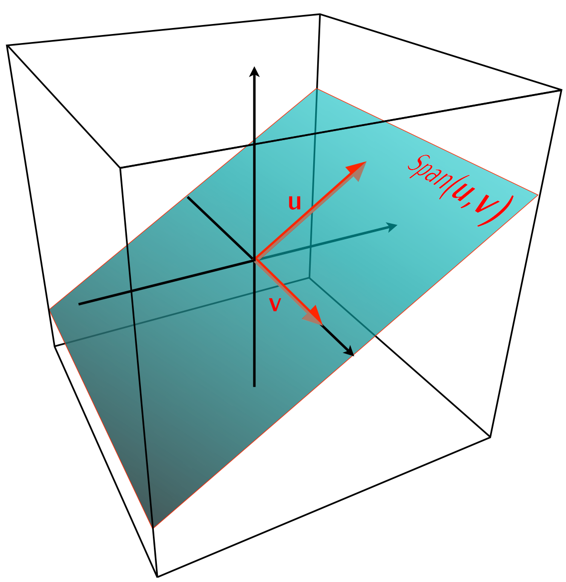
Linear independence vs. Span



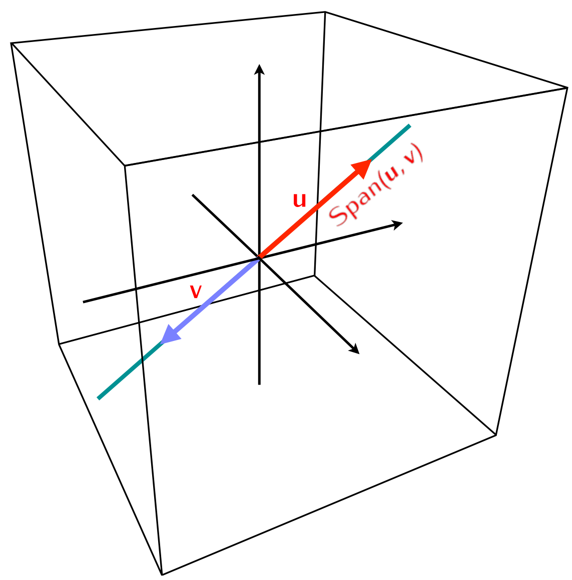
{u} linearly independent



{u} linearly dependent



{u, v} linearly independent



{u, v} linearly dependent

Theorem

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a linearly dependent set of vectors in then:

1) for some $\mathbf{v}_i$ we have $\mathbf{v}_i \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_p)$.

2) for some $\mathbf{v}_i$ we have

$$\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p) = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_p)$$

Example.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

So far:

$$\begin{cases} 2x_1 + 4x_2 + 6x_3 + 3x_4 = 7 \\ 3x_1 + 2x_2 + 2x_3 + 9x_4 = 3 \\ 5x_1 + 8x_2 + 3x_3 + 3x_4 = 9 \end{cases}$$

system of
linear equations

$$\Updownarrow$$

$$x_1 \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 2 \\ 8 \end{bmatrix} + x_3 \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 9 \\ 3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 9 \end{bmatrix}$$

vector equation

Next:

$$\begin{bmatrix} 2 & 4 & 6 & 3 \\ 3 & 2 & 2 & 9 \\ 5 & 8 & 3 & 3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 9 \end{bmatrix}$$

matrix equation

Definition

Let A be an $m \times n$ matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and let $\mathbf{w}$ be a vector in $\mathbb{R}^n$:

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

The product $A\mathbf{w}$ is a vector in $\mathbb{R}^m$ given by

$$A\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$$

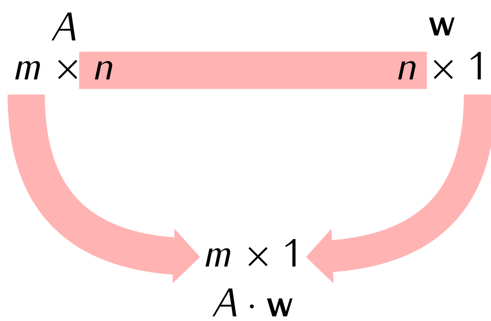
Example.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}$$

Properties of matrix-vector multiplication

1) The product $A\mathbf{w}$ is defined only if

$$(\text{number of columns of } A) = (\text{number of entries of } \mathbf{w})$$



2) $A(\mathbf{v}+\mathbf{w}) = A\mathbf{v} + A\mathbf{w}$

3) If c is a scalar then $A(c\mathbf{w}) = c(A\mathbf{w})$.

Example. Solve the matrix equation

$$\begin{bmatrix} 1 & 1 & -4 \\ 1 & -2 & 3 \\ 3 & -3 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

How to solve a matrix equation

$A\mathbf{x} = \mathbf{b}$

matrix equation

↓

$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$

augmented matrix

↓ *row reduction*

[reduced matrix]

↓ *read off solutions*

$\mathbf{x} = \dots$

solutions

Recall: A vector equation

$$x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n = \mathbf{b}$$

has a solution if and only if $\mathbf{b} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

Definition

If A is a matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$:

$$A = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{bmatrix}$$

then the set $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is called the *column space* of A and it is denoted $\mathrm{Col}(A)$.

Upshot. A matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b} \in \mathrm{Col}(A)$.

Question: What conditions on the matrix A guarantee that the equation $A\mathbf{x} = \mathbf{b}$ has a solution for an arbitrary vector $\mathbf{b}$?

Example.

$$A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \xrightarrow{\textit{row reduction}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Example.

$$A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \\ 2 & 3 & 5 & 7 \end{bmatrix} \xrightarrow{\textit{row reduction}} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Proposition

A matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b}$ if and only if A has a pivot position in every row.

In such case $\text{Col}(A) = \mathbb{R}^m$, where m is the number of rows of A.

Recall: A vector equation

$$x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n = \mathbf{b}$$

has only one solution for each $\mathbf{b} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ if and only if the homogenous equation

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

has only the trivial solution $x_1 = 0, \ldots, x_n = 0$.

Definition

If A is a matrix then the set of solution of the homogenous equation

$$A\mathbf{x} = \mathbf{0}$$

is called the *null space* of A and it is denoted $\mathrm{Nul}(A)$.

Upshot. A matrix equation $A\mathbf{x} = \mathbf{b}$ has only one solution for each $\mathbf{b} \in \mathrm{Col}(A)$ if and only if $\mathrm{Nul}(A) = \{\mathbf{0}\}$.

Example. Find the null space of the matrix

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

Proposition

$\text{Nul}(A) = \{0\}$ if and only if the matrix A has a pivot position in every column.

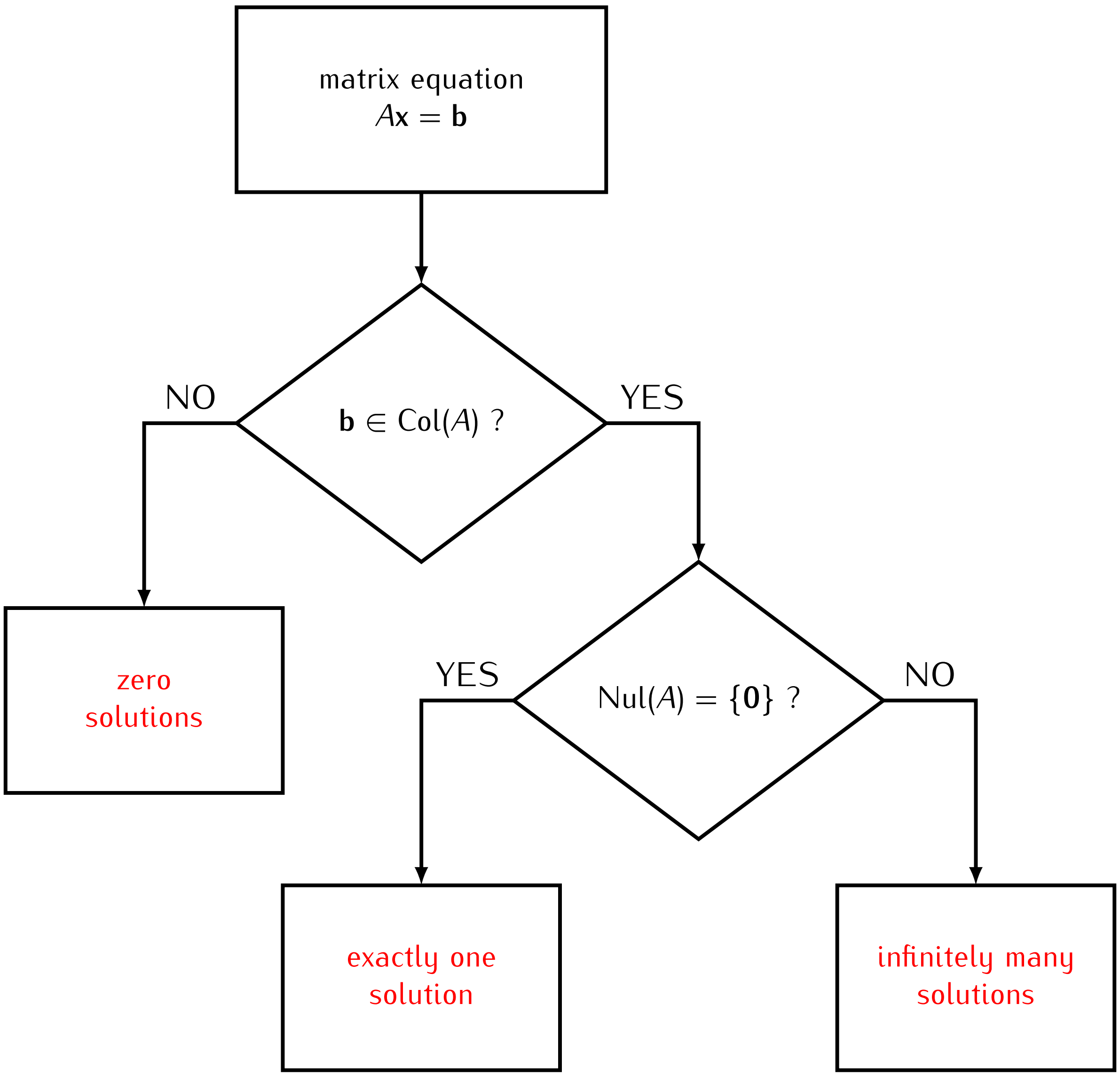
Example. Find the null space of the matrix

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

Note

If A is an $m \times n$ matrix then $\mathrm{Nul}(A)$ can be always described as a span of some vectors in $\mathbb{R}^n$.

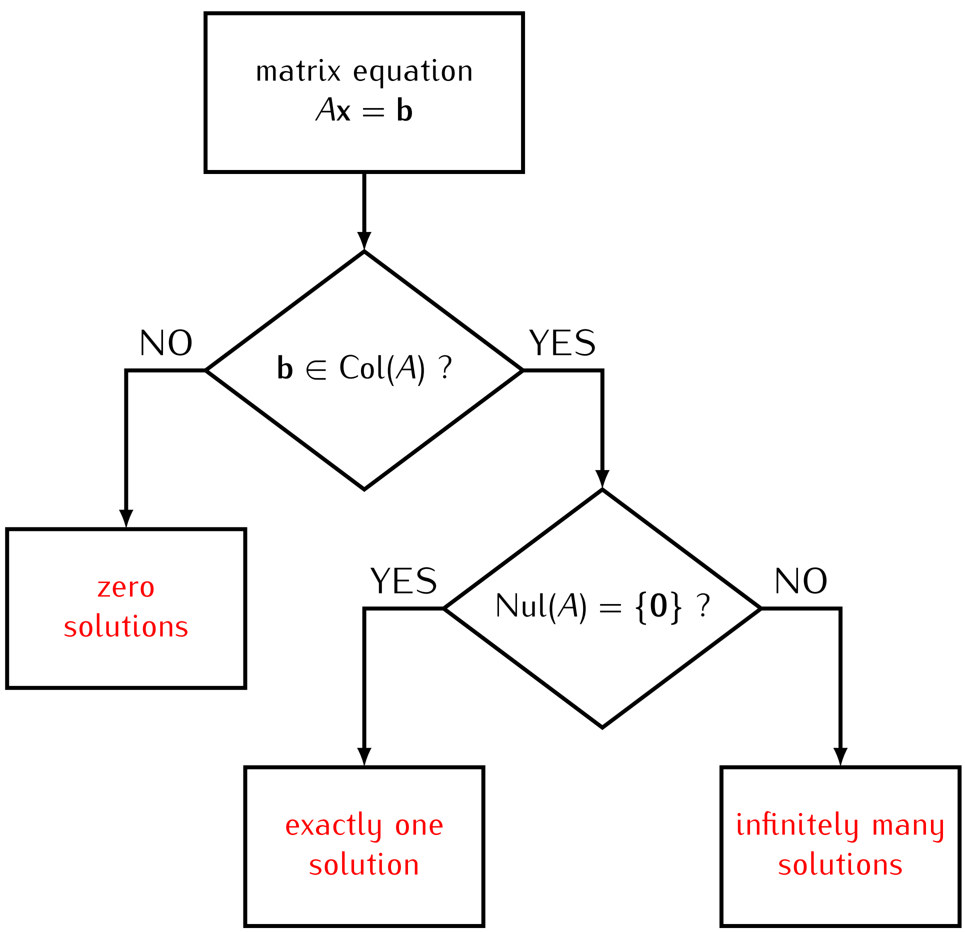
Upshot: how to find the number of solutions of a matrix equation



Recall:

1) We can multiply vectors by matrices.

2) Matrix equation: $A\mathbf{x} = \mathbf{b}$



$\text{Col}(A) = $ (span of column vectors of A)

$\text{Nul}(A) = $ (set of solutions of $A\mathbf{x} = \mathbf{0}$)

Recall: Nul(A) can be always described as a span of some vectors.

Example. Find the null space of the matrix

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

Example. Solve the matrix equation $A\mathbf{x} = \mathbf{b}$ where

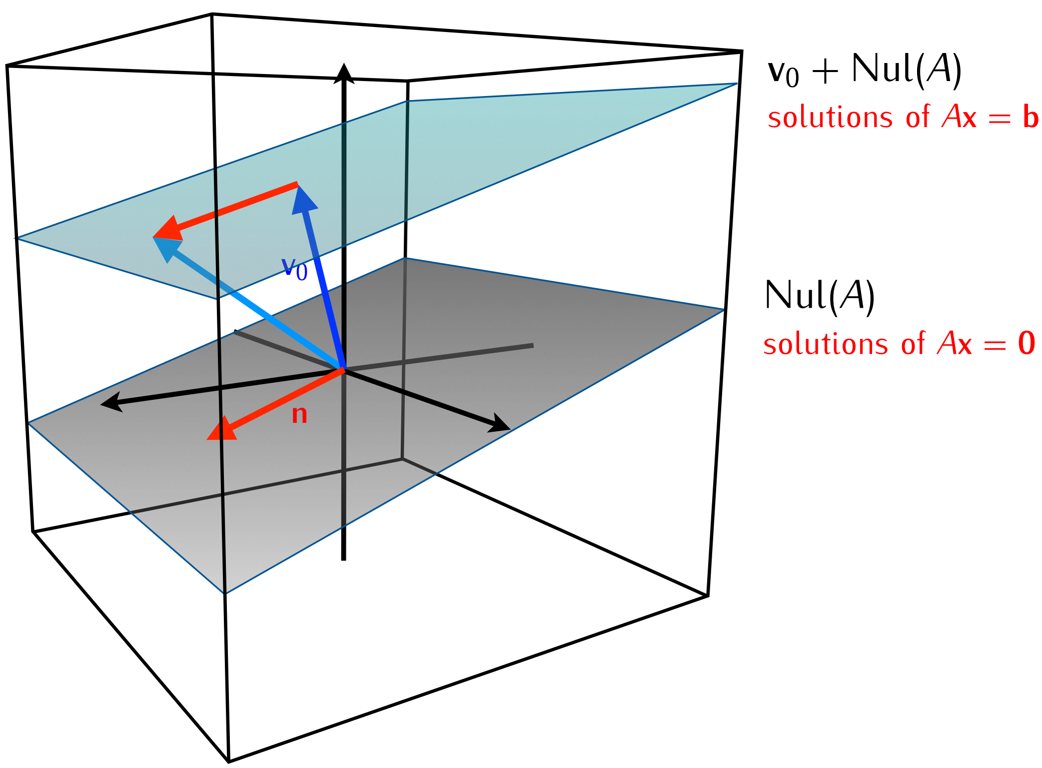
$$A = \begin{bmatrix} 1 & 1 & 0 & 2 \\ -2 & -2 & 1 & -5 \\ 1 & 1 & -1 & 3 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

Proposition

Let $\mathbf{v}_0$ be some chosen solution of a matrix equation $A\mathbf{x} = \mathbf{b}$. Then any other solution $\mathbf{v}$ of this equation is of the form

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{n}$$

where $\mathbf{n} \in \text{Nul}(A)$.



Recall: If A is an $m \times n$ matrix then

$$A \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$$

Definition

If A is an $m \times n$ matrix then the function

$$T_A \colon \mathbb{R}^n \to \mathbb{R}^m$$

given by $T_A(\mathbf{v}) = A\mathbf{v}$ is called the *matrix transformation associated to* A.

Example.

Let $T_A\colon \mathbb{R}^3 \to \mathbb{R}^2$ be the matrix transformation defined by the matrix

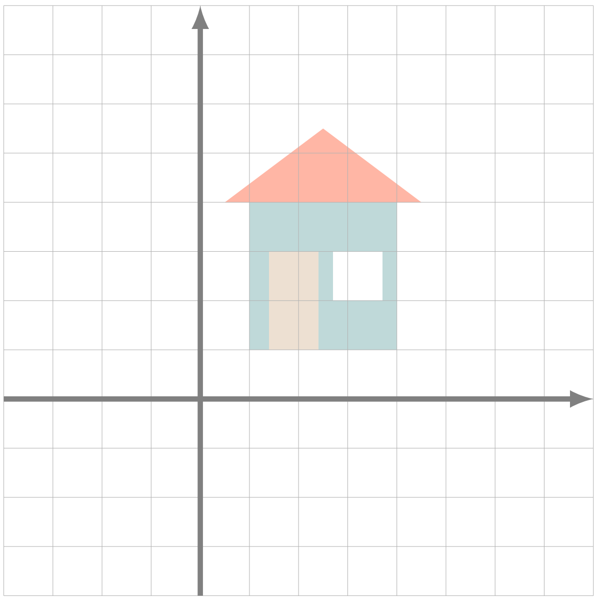
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{bmatrix}$$

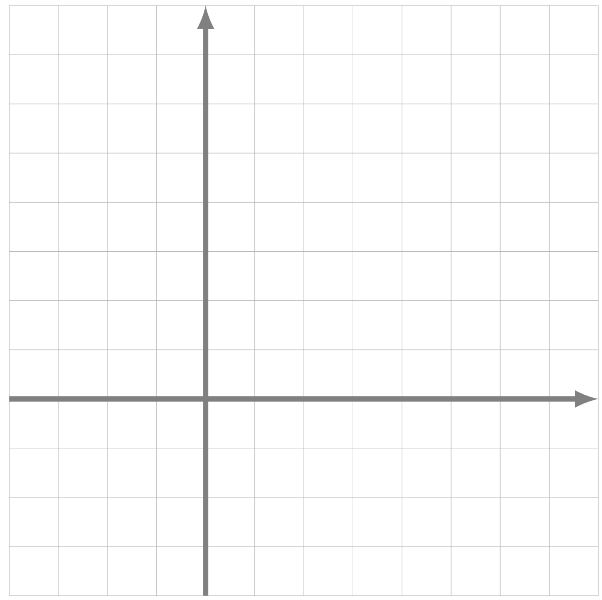
1) Compute $T_A(\mathbf{v})$ where $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

2) Find a vector $\mathbf{v}$ such that $T_A(\mathbf{v}) = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$.

Geometric interpretation of matrix transformations $\mathbb{R}^2 \to \mathbb{R}^2$

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

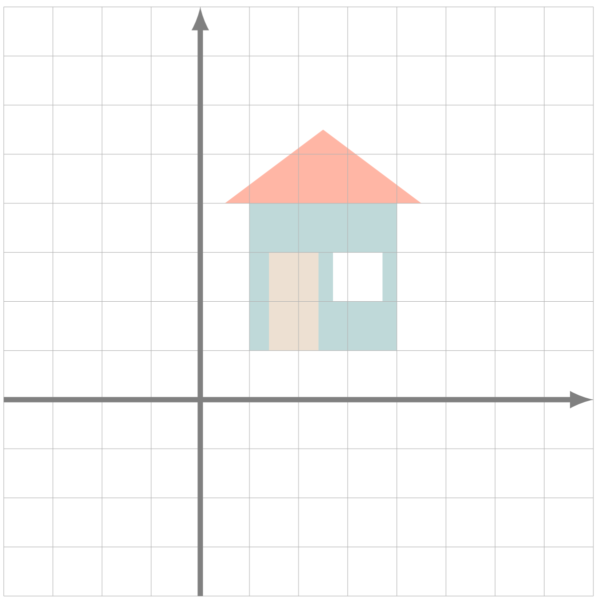


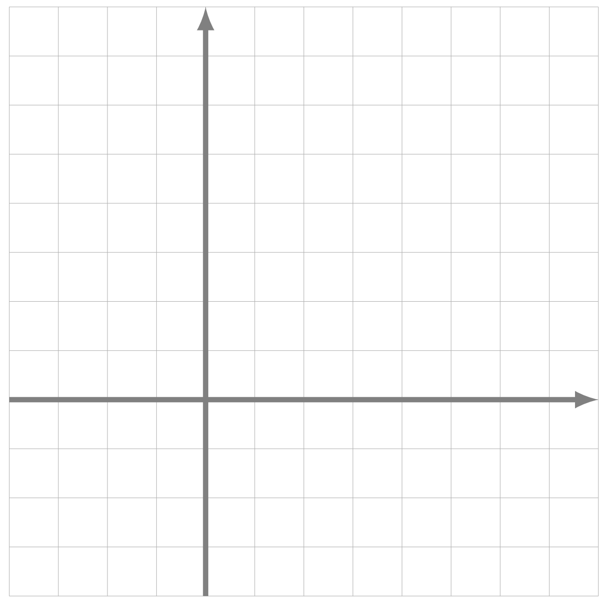


Null spaces, column spaces and matrix transformations

Example.

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





Note

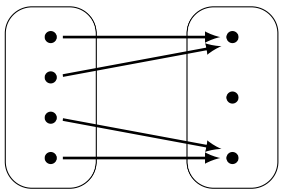
If $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation associated to a matrix A then:

- $\text{Col}(A) =$ the set of values of T_A.
- $\text{Nul}(A) =$ the set of vectors $\mathbf{v}$ such that $T_A(\mathbf{v}) = \mathbf{0}$.
- $T_A(\mathbf{v}) = T_A(\mathbf{w})$ if and only if $\mathbf{w} = \mathbf{v} + \mathbf{n}$ for some $\mathbf{n} \in \text{Nul}(A)$.

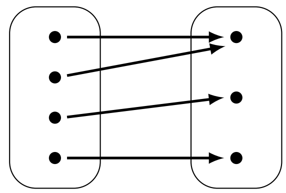
Recall:

A function $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is:

- *onto* if for each $\mathbf{b} \in \mathbb{R}^m$ there is $\mathbf{v} \in \mathbb{R}^n$ such that $F(\mathbf{v}) = \mathbf{b}$;

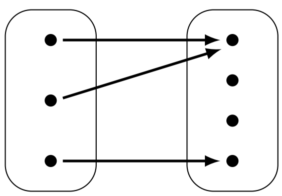


not onto

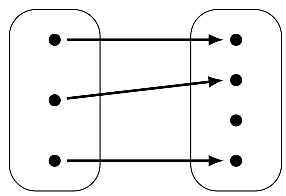


onto

- *one-to-one* if for any $\mathbf{v}_1, \mathbf{v}_2$ such that $\mathbf{v}_1 \neq \mathbf{v}_2$ we have $F(\mathbf{v}_2) \neq F(\mathbf{v}_2)$.



not one-to-one



one-to-one

Proposition

Let A be an $m \times n$ matrix. The following conditions are equivalent:

1) The matrix transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ is onto.

2) $\mathrm{Col}(A) = \mathbb{R}^m$.

3) The matrix A has a pivot position in every row.

Proposition

Let A be an $m \times n$ matrix. The following conditions are equivalent:

1) The matrix transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one.

2) $\mathrm{Nul}(A) = \{0\}$.

3) The matrix A has a pivot position in every column.

Example. For the following 2×2 matrix A check if the matrix transformation $T_A \colon \mathbb{R}^2 \to \mathbb{R}^2$ is onto and if it is one-to-one.

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

Example. For the following 3×4 matrix A check if the matrix transformation $T_A \colon \mathbb{R}^4 \to \mathbb{R}^3$ is onto and if it is one-to-one.

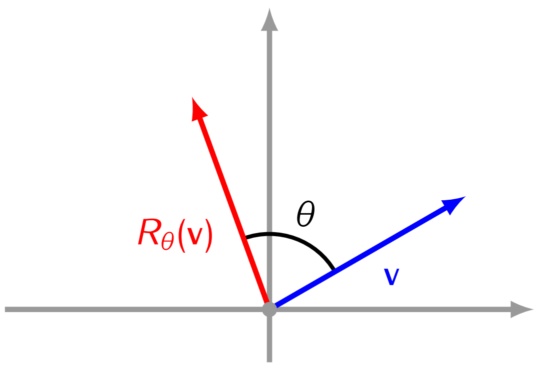
$$A = \begin{bmatrix} 1 & 1 & 0 & 2 \\ -2 & -2 & 1 & -5 \\ 1 & 1 & -1 & 4 \end{bmatrix}$$

Proposition

Let A be an $m \times n$ matrix. If the matrix transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ is both onto and one-to-one then we must have $m = n$ (i.e. A must be a square matrix).

Problem: How to recognize if a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation?

Example. Rotation by an angle θ:



$$R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$$

Definition

A function $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if it satisfies the following conditions:

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

2) $T(c\mathbf{v}) = cT(\mathbf{v})$ for any $\mathbf{v} \in \mathbb{R}^n$ and any scalar c.

Proposition

Every matrix transformation is a linear transformation.

Theorem

Every linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation:

$$T = T_A$$

for some matrix A.

Corollary

If $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation then $T = T_A$ where A is the matrix given by

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \end{bmatrix}$$

This matrix is called the *standard matrix* of T.

Example. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the function given by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ 0 \\ 2x_1 \end{bmatrix}$$

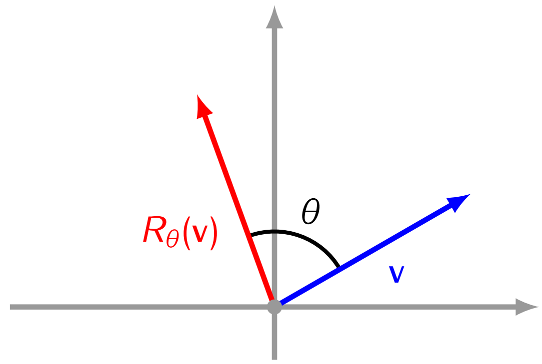
Check if T is a linear transformation. If it is, find its standard matrix.

Example. Let $S\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the function given by

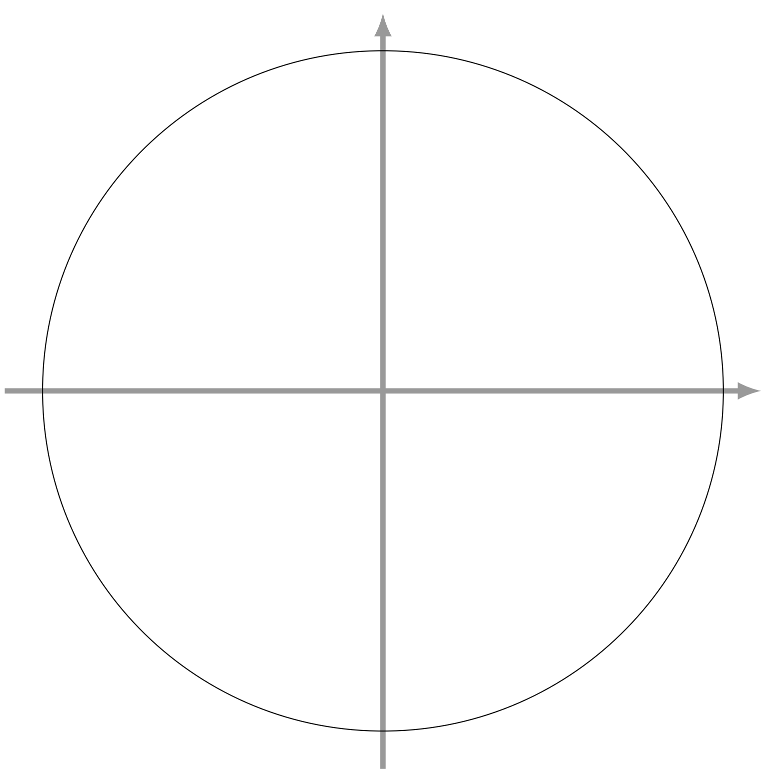
$$S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 + x_2 \\ x_2 \\ 3x_1 \end{bmatrix}$$

Check if S is a linear transformation. If it is, find its standard matrix.

Back to rotations:



$$R_\theta \colon \mathbb{R}^2 \to \mathbb{R}^2$$



Proposition

Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ be the standard basis of of $\mathbb{R}^n$. For any vectors $\mathbf{v}_1, \mathbf{v}_n, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ there exists one and only one linear transformation

$$T \colon \mathbb{R}^n \to \mathbb{R}^m$$

such that

$$T(\mathbf{e}_1) = \mathbf{v}_1 \quad T(\mathbf{e}_2) = \mathbf{v}_2, \quad \ldots, \quad T(\mathbf{e}_n) = \mathbf{v}_n$$

The standard matrix of this linear transformation is given by

$$A = \left[\; \mathbf{v}_1 \quad \mathbf{v}_2 \quad \ldots \quad \mathbf{v}_n \; \right]$$

Recall:

1) If A is an $m \times n$ matrix then the function

$$T_A\colon \mathbb{R}^n \to \mathbb{R}^m$$

defined by $T_A(\mathbf{v}) = A\mathbf{v}$ is called the matrix transformation associated to A.

2) A function $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if

(ii) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
(ii) $T(c\mathbf{v}) = cT(\mathbf{v})$

3) Every matrix transformation is a linear transformation.

4) Every linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation:

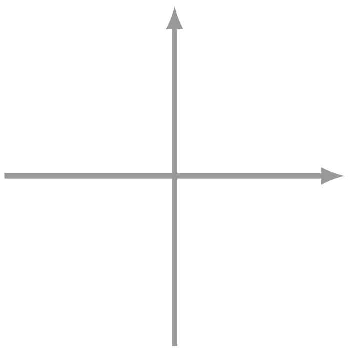
$$T(\mathbf{v}) = A\mathbf{v}$$

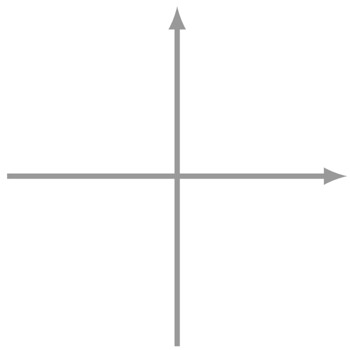
where

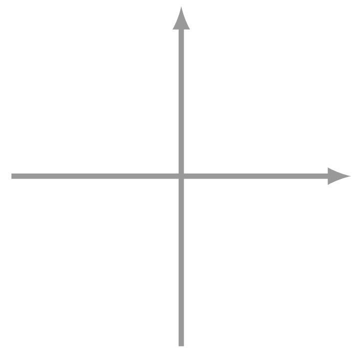
$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \end{bmatrix}$$

The matrix A is called the standard matrix of T.

Composition of linear transformations







Theorem

If $S\colon \mathbb{R}^n \to \mathbb{R}^m$ and $T\colon \mathbb{R}^m \to \mathbb{R}^k$ are linear transformation then the composition

$$T \circ S\colon \mathbb{R}^n \to \mathbb{R}^k$$

is also a linear transformation.

Upshot. The function $T \circ S$ is represented by some matrix C:

$$T \circ S(\mathbf{v}) = C\mathbf{v}$$

Question. Let $S\colon \mathbb{R}^n \to \mathbb{R}^m$ and $T\colon \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations, and let

- B is the standard matrix of S
- A is the standard matrix of T

What if the standard matrix of $T \circ S\colon \mathbb{R}^n \to \mathbb{R}^k$?

Definition

Let

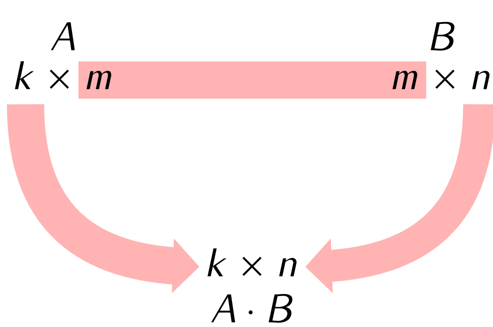
- A be an $k \times m$ matrix
- $B = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}$ be an $m \times n$ matrix

Then $A \cdot B$ is an $k \times n$ matrix given by

$$A \cdot B = \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \dots & A\mathbf{v}_n \end{bmatrix}$$

Note. The product $A \cdot B$ is defined only if

$$(\text{number of columns of } A) \;=\; (\text{number of rows of } B)$$



Example.

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

$$A = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{km} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix}$$

$$AB = \begin{bmatrix} c_{11} & \dots & c_{1m} \\ \vdots & & \vdots \\ c_{k1} & \dots & c_{km} \end{bmatrix}$$

$$c_{ij} = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{im} \end{bmatrix} \cdot \begin{bmatrix} b_{1j} \\ b_{1j} \\ \vdots \\ b_{1j} \end{bmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{im}b_{mj}$$

Example.

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

Example.

- Acme Inc. makes two types of widgets: **WG1** and **WG2**.
- Each widget must go though two processes: **assembly** and **testing**.
- The number of hours required to complete each process is as follows:

	assembly	testing
WG1	3	1
WG2	7	3

- Acme Inc. has three plans in New York, Texas, and Minnesota.
- Hourly cost (in dollars) of each process in each plant is as follows:

	NY	TX	MN
assembly	10	15	12
testing	15	20	15

Problem. What is the cost of producing each type of widgets in each plant?

Other operations on matrices

1) Addition.

If $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$, $B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix}$ are $m \times n$ matrices then

$$A + B = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

Note. The sum $A + B$ is defined only if A and B have the same dimensions.

2) Scalar multiplication.

If $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$, and c is a scalar then

$$cA = \begin{bmatrix} ca_{11} & \dots & ca_{1n} \\ \vdots & & \vdots \\ ca_{m1} & \dots & ca_{mn} \end{bmatrix}$$

Properties of matrix algebra

1) $(AB)C = A(BC)$

2) $(A+B)C = AC + BC$
$A(B+C) = AB + AC$

3) I_n = the $n \times n$ identity matrix:

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

If A is an $m \times n$ matrix then

$$A \cdot I_n = A$$

$$I_m \cdot A = A$$

Non-commutativity of matrix multiplication

1) If AB is defined then BA need not be defined.

2) Even if both AB and BA are both defined then usually

$$AB \neq BA$$

One more operation on matrices: matrix transpose

Definition

The transpose of a matrix A is the matrix A^T such that

$$(\text{rows of } A^T) = (\text{columns of } A)$$

Properties of transpose

1) $(A^T)^T = A$
2) $(A+B)^T = (A^T + B^T)$
3) $(AB)^T = B^T A^T$

Operations on matrices so far:

- addition/subtraction $A \pm B$
- scalar multiplication $c \cdot A$
- matrix multiplication $A \cdot B$
- matrix transpose A^T

Next: How to divide matrices?

Definition

A matrix A is *invertible* if there exists a matrix B such that

$$A \cdot B = B \cdot A = I$$

(where I = the identity matrix). In such case we say that B is the *inverse* of A and we write $B = A^{-1}$.

Example.

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \text{ is invertible,} \quad A^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Matrix inverses and matrix equations

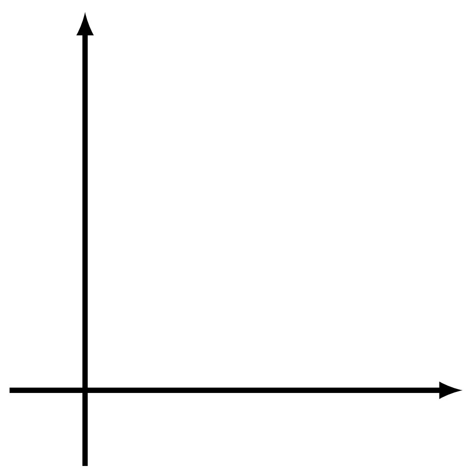
Proposition

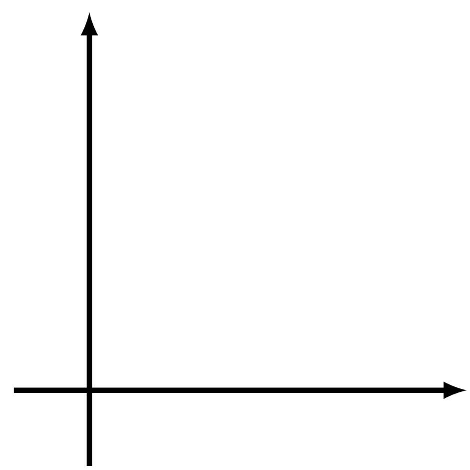
If A is an invertible matrix then for any vector $\mathbf{b}$ the equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution.

Example. Solve the following matrix equation:

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

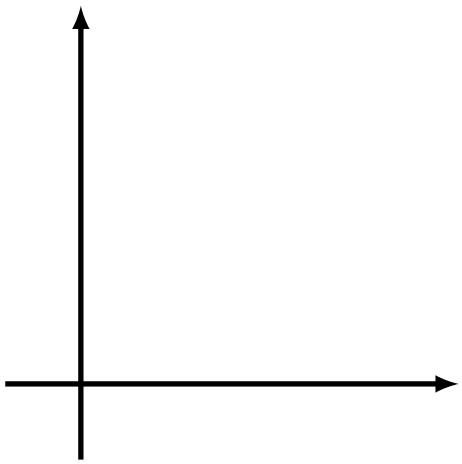
Matrix inverses and matrix transformations

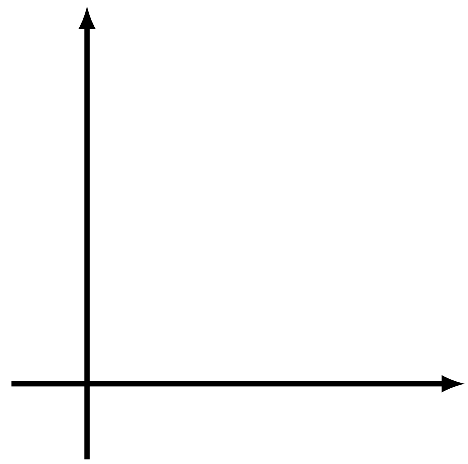




Example.

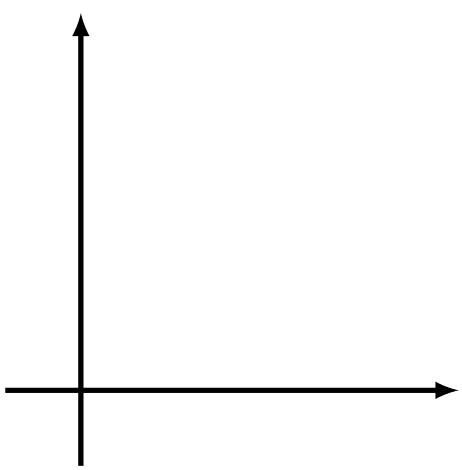
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

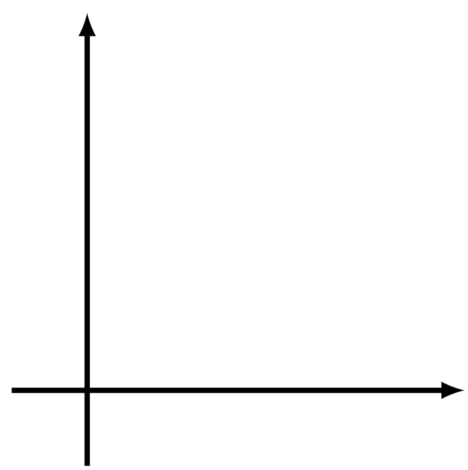




Example.

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





Upshot. If an $m \times n$ matrix A is invertible then the matrix transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ must be one-to-one and onto.

Recall: If A be is $m \times n$ matrix then the matrix transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ is:

- <u>onto</u> if and only if A has a pivot position in every row
- <u>one-to-one</u> if and only if A has a pivot position in every column.

Theorem

If A is not a square matrix then it is not invertible.

If A is a square matrix then the following conditions are equivalent:

1) A is an invertible matrix.

2) The matrix A has a pivot position in every row and column.

3) The reduced row echelon form of A is the identity matrix I_n.

Proposition

If A is an $n \times n$ invertible matrix then

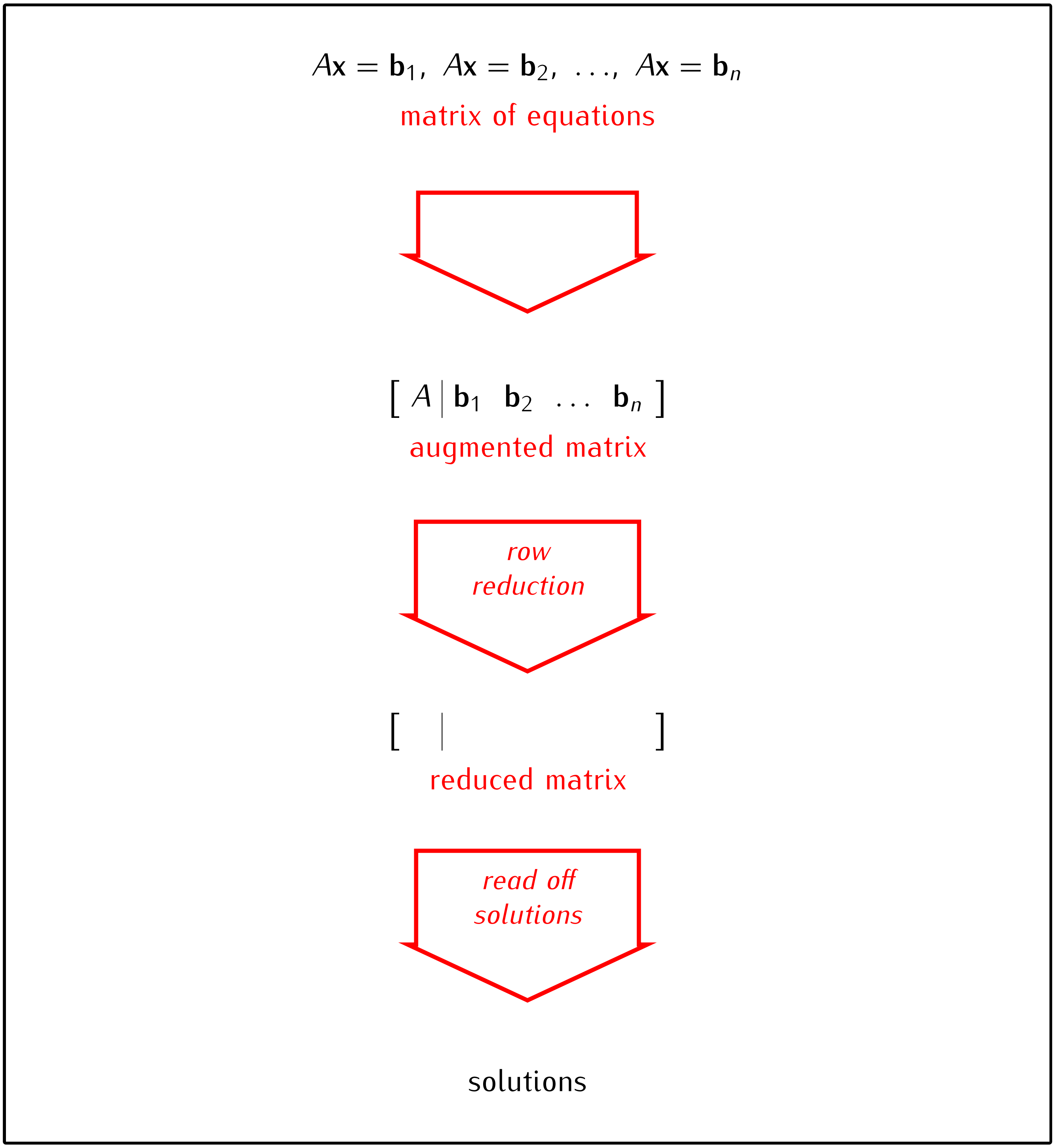
$$A^{-1} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_n \end{bmatrix}$$

where $\mathbf{w}_i$ is the solution of $A\mathbf{x} = \mathbf{e}_i$.

Example.

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

Simplification:
How to solve several matrix equations with the same coefficient matrix at the same time



Example. Solve the vector equations $A\mathbf{x} = \mathbf{e}_1$ and $A\mathbf{x} = \mathbf{e}_2$ where

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \qquad \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Summary:
How to invert a matrix

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

1) Augment A by the identity matrix.

2) Reduce the augmented matrix.

2) If after the row reduction the matrix on the left is the identity matrix, then A is invertible and

$$A^{-1} = \text{the matrix on the right}$$

Otherwise A is not invertible.

Properties of matrix inverses

1) If A is invertible then A^{-1} is invertible and

$$(A^{-1})^{-1} = A$$

2) If A, B are invertible then AB is invertible and

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

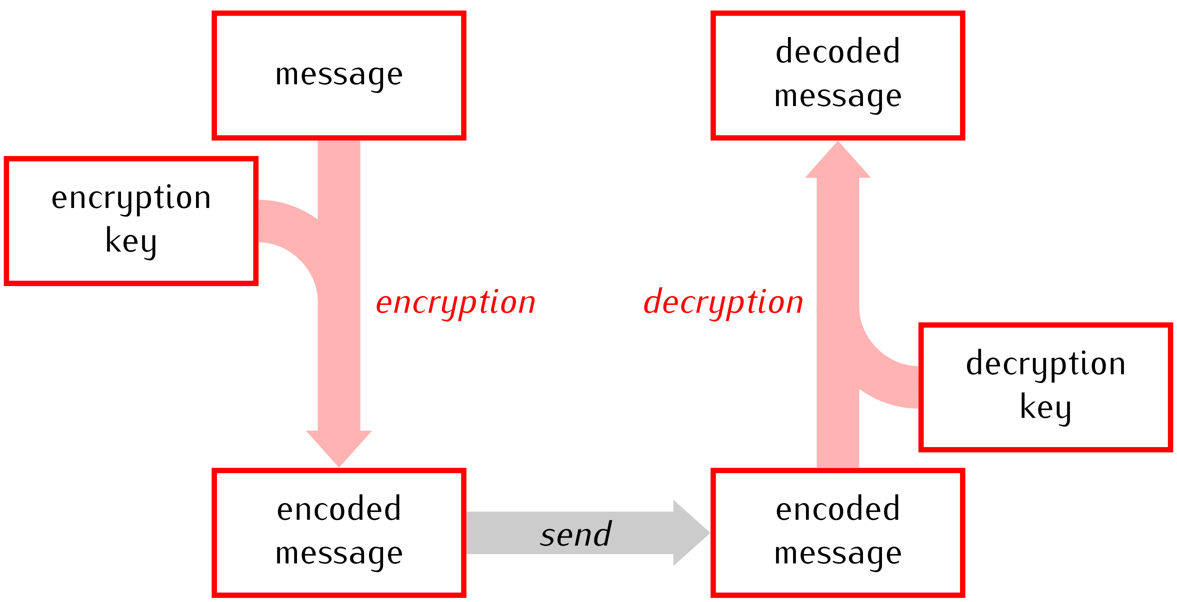
3) If A is invertible then A^T is invertible and

$$(A^T)^{-1} = (A^{-1})^T$$

Ciphers.

Cipher is an algorithm for encrypting and decrypting data to conceal its meaning.

Basic working scheme of ciphers



Substitution cipher: Replace each letter of the alphabet by some other letter.

Example.

encrypt ↓ / decrypt ↑

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
T	V	W	X	Y	S	C	N	O	U	Z	A	B	P	I	M	J	Q	R	K	D	E	F	G	H	L

encryption/decryption key

message: TOP SECRET

Hill cipher: Use matrix multiplication

Example.

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

encryption key
invertible matrix

decryption key
matrix inverse

message: TOP SECRET

Encryption:

1) Replace letters by numbers:

_	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

2) Since the key is a 3×3 matrix split the number sequence numbers in vectors with 3 entries each.

3) Multiply each vector by the encryption matrix A.

4) Write the new vectors as a sequence of numbers.

We can do better, but the next part will not work with an arbitrary invertible matrix A. It will work though e.g. if all entries of A and A^{-1} are integers.

5) Reduce all numbers obtained in step 4 modulo 27. That is, add or subtract from each number a multiple of 27 to get a number between 0 and 26.

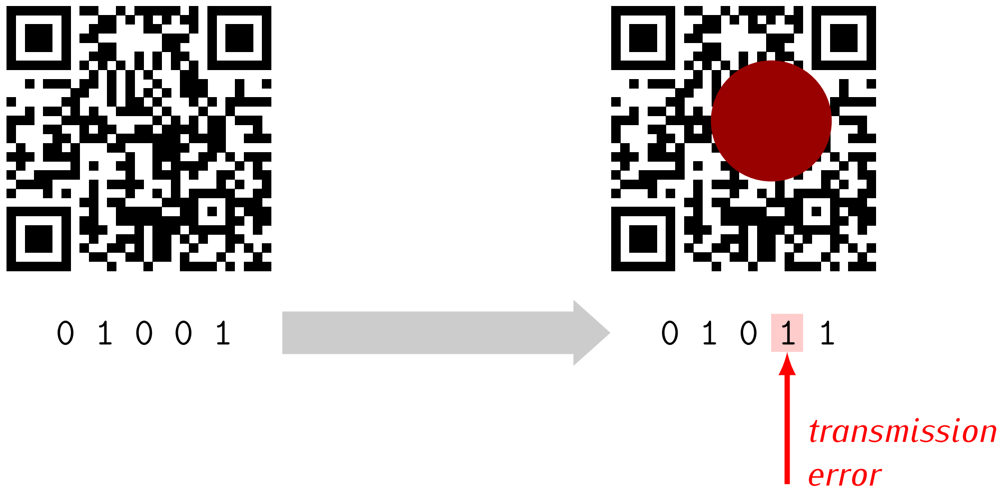
6) Replace numbers by letters.

Decryption.

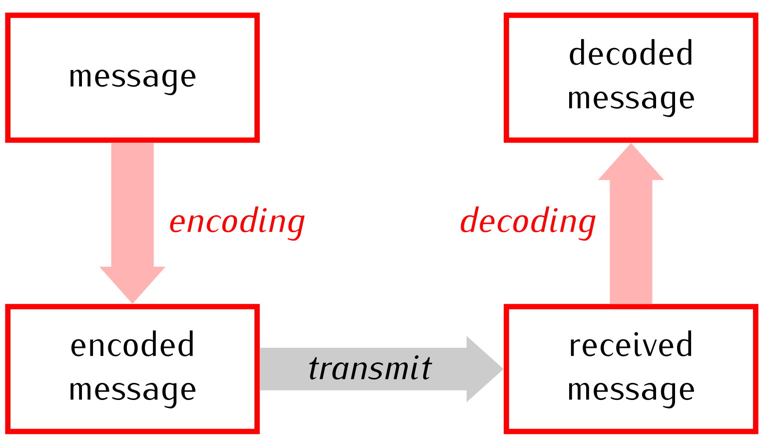
1) Replace letters by numbers, split into vectors, and multiply each vector by A^{-1}

2) Write the new vectors as a sequence of numbers, reduce each number modulo 27.

3) Replace numbers by letters



Basic scheme of error correction



Working assumption for this lecture: We expect at most one transmission error in any message up to 20 bits long.

A simple error correcting code: triple repeat.

message: 1011

Problem: The encoded message is 3 times longer than the original message.

Better error correction: Hamming (7,4) code.

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

encoding matrix

decoding matrix

message: 10111101

Encoding.

1) Split the message into vectors with 4 entries, and multiply each vector by the encoding matrix E.

2) Reduce all numbers obtained in step 1 modulo 2. That is, add or subtract from each number a multiple of 2 to get either 0 or 1.

Encoded message:

Received message:

Decoding. Split the received message into vectors with 7 entries, multiply each vector by the decoding matrix D, and reduce modulo 2.

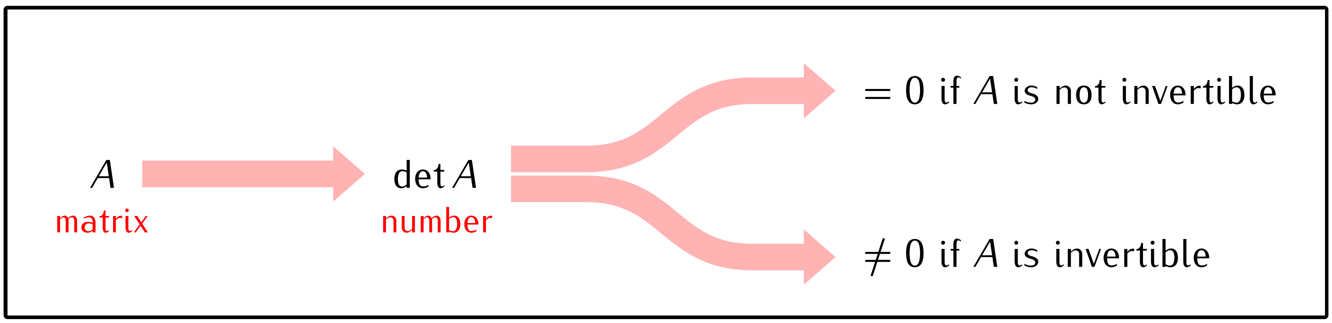
Decoded message:

How the Hamming code works:

Recall: If an $n \times n$ matrix A is invertible then:

- the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$
- the linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^n$, $T_A(\mathbf{v}) = A\mathbf{v}$ has an inverse function.

Determinants recognize which matrices are invertible:



Example: Determinant for a 1×1 matrix.

$$A = \begin{bmatrix} a \end{bmatrix}$$

Example: Determinant for a 2×2 matrix.

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

Definition

If A is an $n \times n$ matrix then for $1 \leq i, j \leq n$ the (i, j)-*minor* of A is the matrix A_{ij} obtained by deleting the i^{th} row and j^{th} column of A.

Example.

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

Definition

Let A be an $n \times n$ matrix

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

1) If $n = 1$, i.e. $A = \begin{bmatrix} a_{11} \end{bmatrix}$, then $\det A = a_{11}$

2) If $n > 1$ then

$$\begin{aligned} \det A = \quad & (-1)^{1+1} a_{11} \cdot \det A_{11} \\ + \quad & (-1)^{1+2} a_{12} \cdot \det A_{12} \\ \dots \quad & \quad \dots \quad \dots \quad \dots \\ + \quad & (-1)^{1+n} a_{1n} \cdot \det A_{1n} \end{aligned}$$

Example. $(n = 2)$

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

Note

If A is a 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then $\det A = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$

Example. (n=3)

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

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

Example (n=4)

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

Note. In order to compute the determinant of an $n \times n$ matrix in this way we need to compute:

$$\begin{array}{rl} n & \text{determinants of } (n-1) \times (n-1) \text{ matrices} \\ n(n-1) & \text{determinants of } (n-2) \times (n-2) \text{ matrices} \\ n(n-1)(n-2) & \text{determinants of } (n-3) \times (n-3) \text{ matrices} \\ \dots \quad \dots \quad \dots \quad \dots & \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ n(n-1)(n-2) \cdot \ldots \cdot 3 & \text{determinants of } 2 \times 2 \text{ matrices} \end{array}$$

E.g. for a 25×25 matrix we would need to compute

$$25 \cdot 24 \cdot 23 \cdot \ldots \cdot 3 = 7,755,605,021,665,492,992,000,000$$

determinants of 2×2 matrices.

<u>**Next:**</u> How to compute determinants faster.

Definition

If A is an $n \times n$ matrix and $1 \leq i, j \leq n$ then the *ij-cofactor of A* is the number

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

Example.

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

Note. By the definition of the determinant we have:

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n}$$

Theorem

Let A be an $n \times n$ matrix.

1) For any $1 \leq i \leq n$ we have

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}$$

(cofactor expansion across the i^{th} row).

2) For any $1 \leq j \leq n$ we have

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$$

(cofactor expansion down the j^{th} column).

Example.

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

Example. Compute the determinant of the following matrix:

$$
\begin{bmatrix}
1 & 0 & 0 & 3 & 0 & 0 & 2 & 0 & 3 & 0 & 0 & 0 & 0 & e & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & \pi & 0 & 0 & 0 & 6 & 0 & 0 & 5 & 6 & 0 & 2 & 0 & 7 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 11 & 0 & 0 & 0 & 0 & 0 & 7 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 2 & 0 & 4 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 2 & 1 & 2 & 3 & 4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 2 & 7 & 0 & -4 & 0 & 0 & 3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5} & 0 & 1 & 0 & 4 & 3 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 8 & 7 & 7 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 2 & 8 & 9 & 0 & 3 & 3 & 2 & 5 & 6 & 3 & 8 & 9 & 2 & 6 & 2 & 2 & 1
\end{bmatrix}
$$

Definition

An square matrix is *upper triangular* is all its entries below the main diagonal are 0.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

Proposition

If A is an upper triangular matrix as above then

$$\det A = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn}$$

Recall: If A is an upper triangular matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

then $\det A = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn}$.

Note. If A is a square matrix then the row echelon form of A is always upper triangular.

Theorem

Let A and B be $n \times n$ matrices.

1) If B is obtained from A by interchanging two rows (or two columns) then

$$\det B = -\det A$$

2) If B is obtained from A by multiplying one row (or one column) of A by a scalar k then

$$\det B = k \cdot \det A$$

2) If B is obtained from A by adding a multiple of one row of A to another row (or adding a multiple of one column to another column) then

$$\det B = \det A$$

Example.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 7 \\ 2 & 5 & 1 \end{bmatrix}$$

Computation of determinants via row reduction

Idea. To compute $\det A$, row reduce A to the row echelon form. Keep track how the determinant changes at each step of the row reduction process.

Example. Compute $\det A$ where

$$A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 2 & 4 & 0 & 10 \\ 3 & 4 & 1 & 7 \\ -2 & 5 & 3 & 0 \end{bmatrix}$$

Theorem

If A is a square matrix then A is invertible if and only if $\det A \neq 0$

Recall: A is invertible if and only if its reduced row echelon form is the identity matrix.

Further properties of determinants

1) $\det(A^T) = \det A$

2) $\det(AB) = (\det A) \cdot (\det B)$

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

Note. In general $\det(A + B) \neq \det A + \det B$.

Recall: If A is square matrix then the *ij-cofactor of A* is the number

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

Definition

If A is an $n \times n$ matrix then the *adjoint* (or *adjugate*) of A is the matrix

$$\mathrm{adj}A = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}^T = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

Theorem

If A is an invertible matrix then

$$A^{-1} = \frac{1}{\det A} \cdot \mathrm{adj}A$$

Example. Compute A^{-1} for

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

Recall: If A is an invertible matrix then the equation $A\mathbf{x} = \mathbf{b}$ has only one solution: $\mathbf{x} = A^{-1}\mathbf{b}$.

Definition

If A is an $n \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$ then $A_i(\mathbf{b})$ is the matrix obtained by replacing the i^{th} column of A with $\mathbf{b}$.

Example.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix}$$

Theorem (Cramer's Rule)

If A is an $n \times n$ invertible matrix and $\mathbf{b} \in \mathbb{R}^n$ then the unique solution of the equation

$$A\mathbf{x} = \mathbf{b}$$

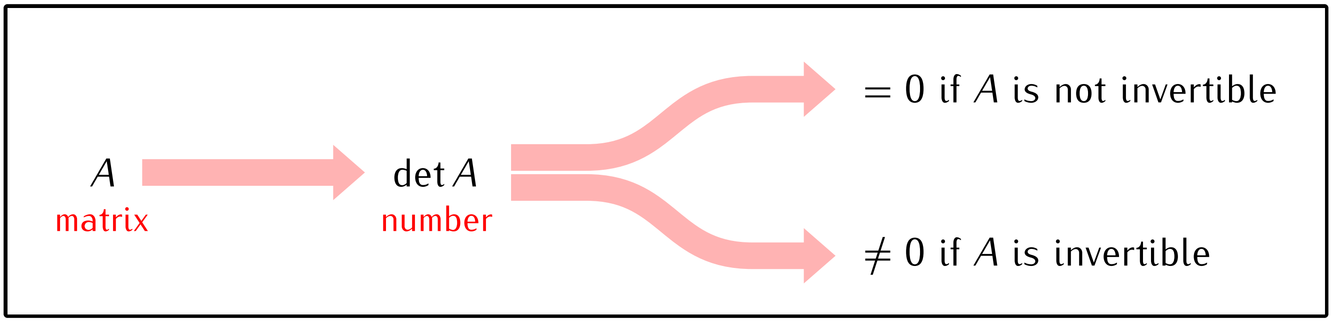
is given by

$$\mathbf{x} = \frac{1}{\det A} \begin{bmatrix} \det A_1(\mathbf{b}) \\ \vdots \\ \det A_n(\mathbf{b}) \end{bmatrix}$$

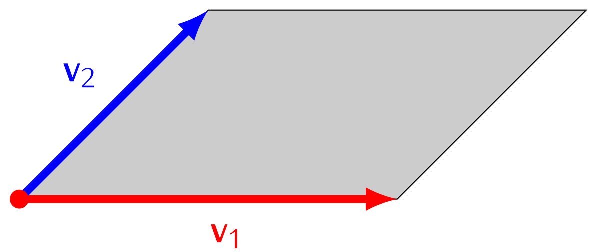
Example. Solve the equation

$$\begin{bmatrix} 1 & 1 & 2 \\ 4 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Recall:



Note. Any two vectors in $\mathbb{R}^2$ define a parallelogram:



Notation

$$\text{area}(\mathbf{v}_1, \mathbf{v}_2) = \begin{pmatrix} \text{area of the parallelogram} \\ \text{defined by } \mathbf{v}_1 \text{ and } \mathbf{v}_2 \end{pmatrix}$$

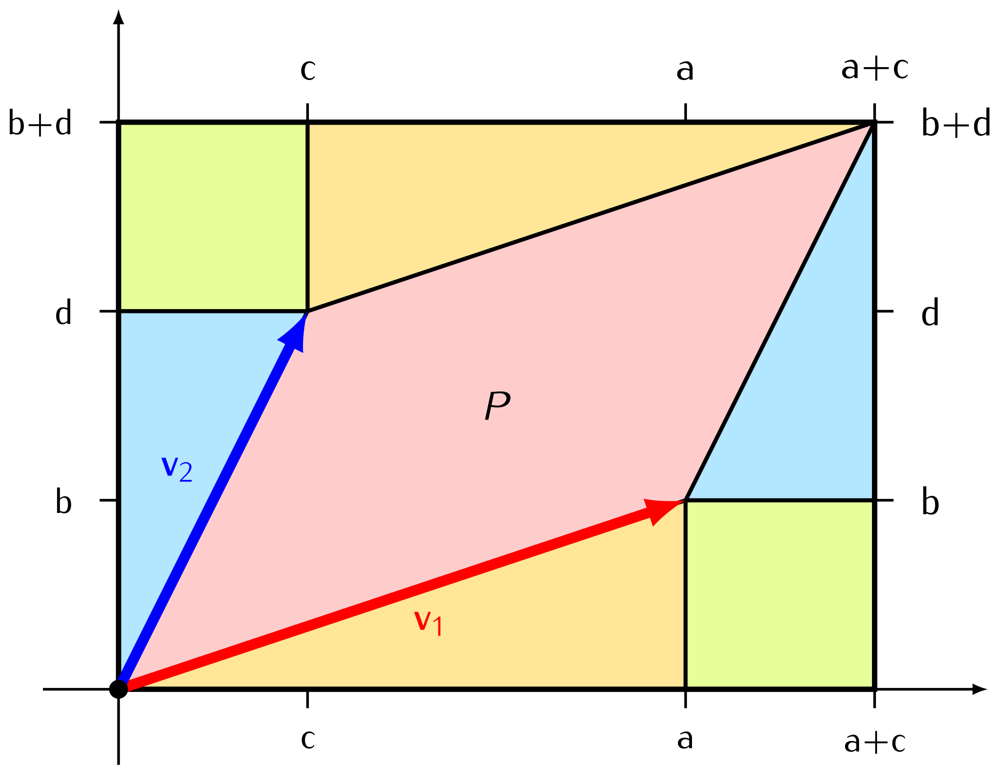
Theorem

If $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ then

$$\text{area}(\mathbf{v}_1, \mathbf{v}_2) = \left|\det \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}\right|$$

Idea of the proof.

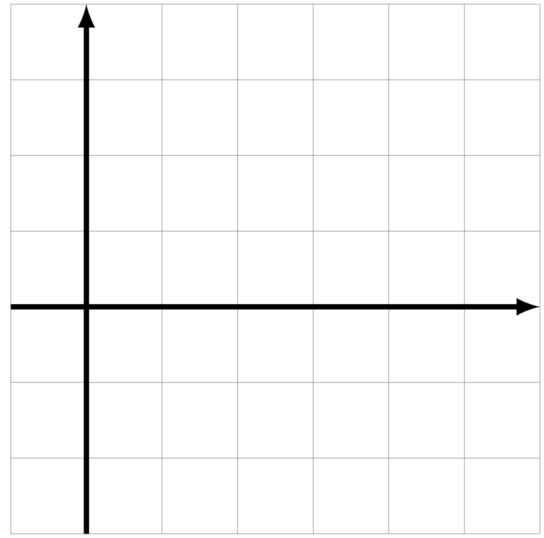
$$\mathbf{v}_1 = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} c \\ d \end{bmatrix}$$



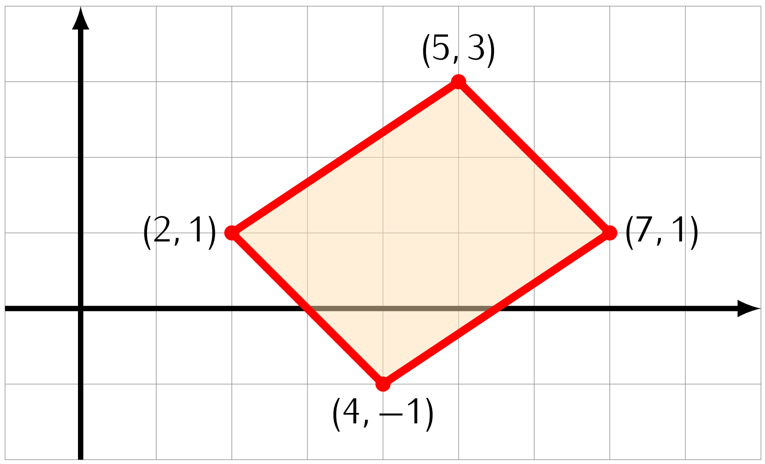
$$\begin{aligned} & \text{area}(P) \\ + & \tfrac{1}{2}ab \\ + & \tfrac{1}{2}ab \\ + & \tfrac{1}{2}cd \\ + & \tfrac{1}{2}cd \\ + & cb \\ + & cb \\ \hline & (a+c)(b+d) \end{aligned}$$

Example.

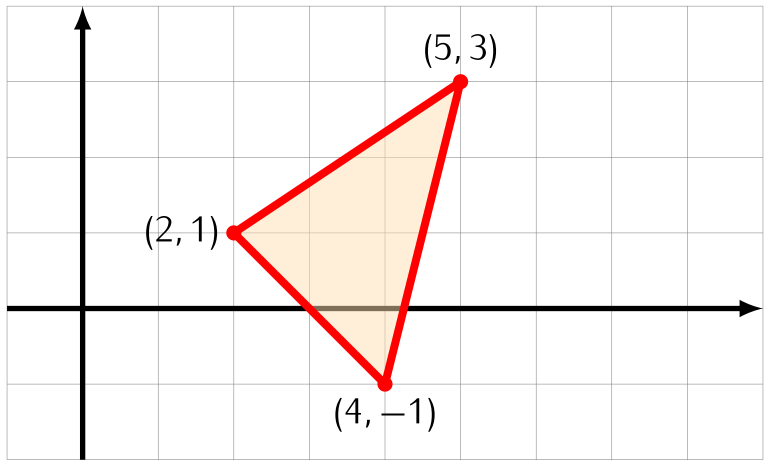
$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$



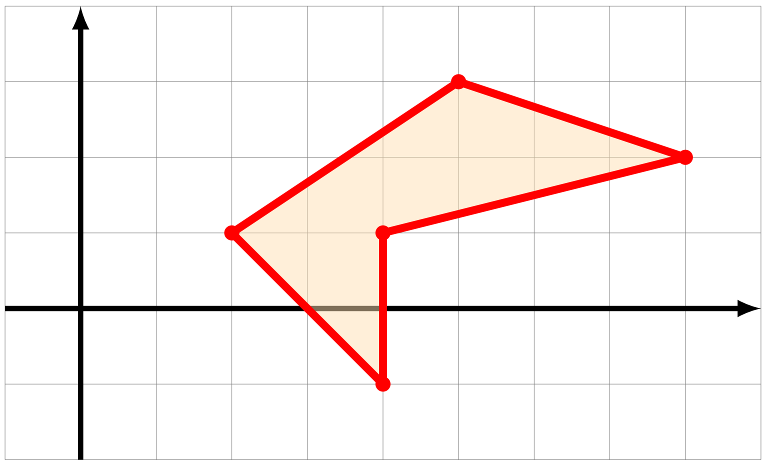
Example. Calculate the area of the parallelogram with vertices at the points $(2,1)$, $(5,3)$, $(7,1)$, $(4,-1)$.



Example. Calculate the area of the triangle with vertices at the points $(2, 1)$, $(5, 3)$, $(4, -1)$.



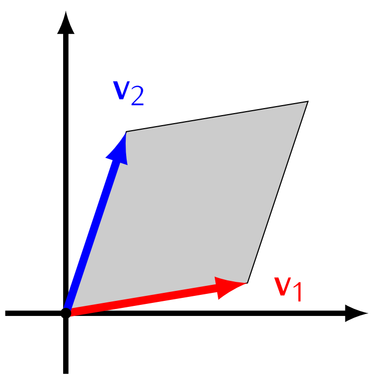
Note. In order to compute areas of other polygons, subdivide them into triangles.

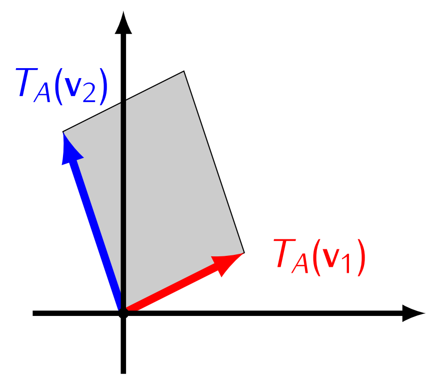


Recall: If A is a 2×2 matrix then it defines a linear transformation

$$T_A \colon \mathbb{R}^2 \to \mathbb{R}^2 \qquad T_A(\mathbf{v}) = A\mathbf{v}$$

Note. T_A maps parallelograms to parallelograms:





Theorem

If A is a 2×2 matrix and $\mathbf{v}_1, \mathbf{v}_1 \in \mathbb{R}^2$ then

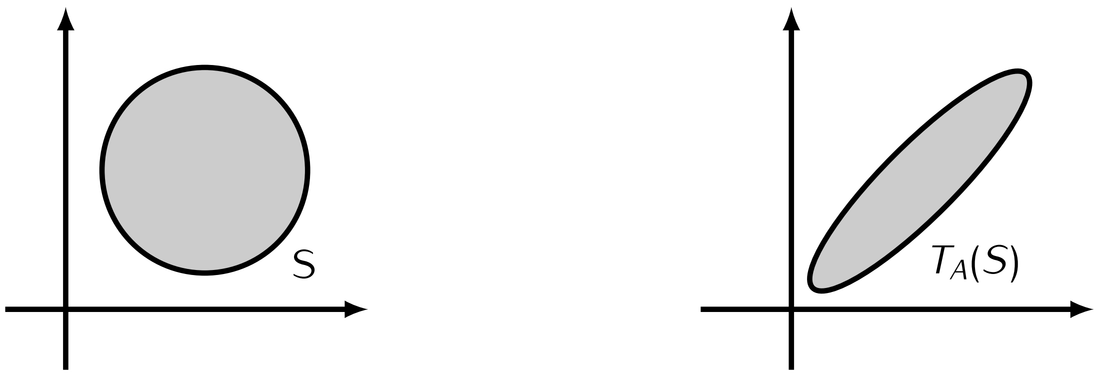
$$\text{area}(T_A(\mathbf{v}_1), T_A(\mathbf{v}_2)) = |\det A| \cdot \text{area}(\mathbf{v}_1, \mathbf{v}_2)$$

Generalization:

Theorem

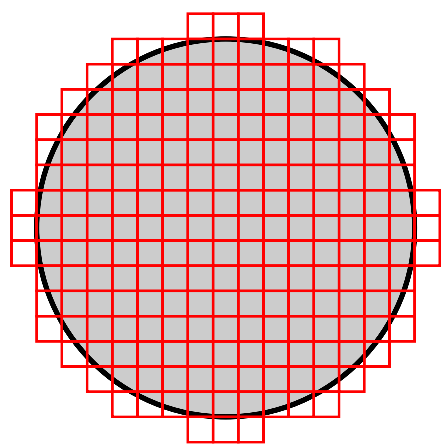
If A is a 2×2 matrix then for any region S of $\mathbb{R}^2$ we have:

$$\text{area}(T_A(S)) = |\det A| \cdot \text{area}(S)$$



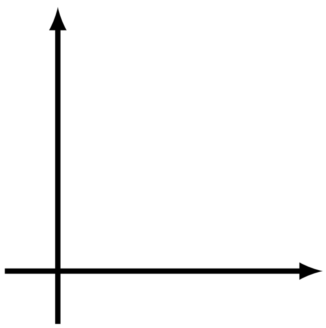
Idea of the proof.

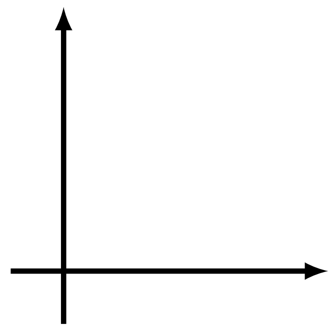
The area of S can be approximated by the sum of small squares covering S.



Example.

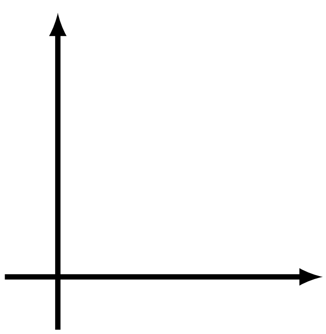
$$A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

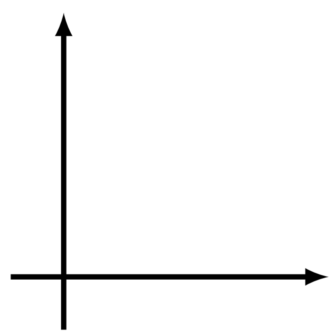




Example.

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





Theorem

If A is a 2×2 matrix then the linear transformation $T_A \colon \mathbb{R}^2 \to \mathbb{R}^2$ preserves orientation if $\det A > 0$ and reverses orientation if $\det A < 0$.

Linear Algebra	Calculus
$\mathbb{R}^n = \begin{pmatrix}\text{set of all column vectors} \\ \text{with } n \text{ entries}\end{pmatrix}$	$C^\infty(\mathbb{R}) = \begin{pmatrix}\text{set of all smooth} \\ \text{functions } f\colon \mathbb{R} \to \mathbb{R}\end{pmatrix}$
Column vectors can be added and multiplied by real numbers.	**Functions** can be added and multiplied by real numbers.
Linear transformation is a function $$T\colon \mathbb{R}^n \to \mathbb{R}^m, \quad T(\mathbf{v}) = A\mathbf{v}$$ It satisfies: • $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ • $T(c\mathbf{v}) = cT(\mathbf{v})$	**Differentiation** is a function $$D\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}), \quad D(f) = f'$$ It satisfies: • $D(f+g) = D(f) + D(g)$ • $D(cf) = cD(f)$
Typical problem: given a vector $\mathbf{b}$ find all vectors $\mathbf{x}$ such that $$T(\mathbf{x}) = \mathbf{b}$$ (i.e solve the equation $A\mathbf{x} = \mathbf{b}$).	**Typical problem:** given a function g find all functions f such that $$D(f) = g$$ (i.e find antiderivatives of g).
Fact: Such vectors x are of the form $$\mathbf{x} = \mathbf{v}_0 + \mathbf{n}$$ where: • $\mathbf{v}_0$ is some distinguished solution of $A\mathbf{x} = \mathbf{b}$; • $\mathbf{n} \in \text{Nul}(A)$ (i.e. $\mathbf{n}$ is a solution of $A\mathbf{x} = \mathbf{0}$).	**Fact:** Such functions f are of the form $$f = F + C$$ where: • F is some distinguished antiderivative of g; • C is a constant function (i.e. C is a solution of $D(f) = 0$).

Definition

A *(real) vector space* is a set V together with two operations:

- addition

$$\begin{aligned} V \times V &\longrightarrow V \\ (\mathbf{u}, \ \mathbf{v}) &\longmapsto \mathbf{u} + \mathbf{v} \end{aligned}$$

- multiplication by scalars

$$\begin{aligned} \mathbb{R} \times V &\longrightarrow V \\ (c, \ \mathbf{v}) &\longmapsto c \cdot \mathbf{v} \end{aligned}$$

Moreover the following conditions must be satisfied:

1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

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

3) there is an element $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$ for any $\mathbf{u} \in V$

4) for any $\mathbf{u} \in V$ there is an element $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

5) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

6) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

7) $(cd)\mathbf{u} = c(d\mathbf{u})$

8) $1\mathbf{u} = \mathbf{u}$

Elements of V are called *vectors*.

Theorem

If V is a vectors space then:

1) $c \cdot \mathbf{0} = \mathbf{0}$ where $c \in \mathbb{R}$ and $\mathbf{0} \in V$ is the zero vector;

2) $0 \cdot \mathbf{u} = \mathbf{0}$ where $0 \in \mathbb{R}$, $\mathbf{u} \in V$ and $\mathbf{0}$ is the zero vector;

3) $(-1) \cdot \mathbf{u} = -\mathbf{u}$

Examples of vector spaces.

Defitnition

Let V be a vector space. A *subspace* of V is a subset $W \subseteq V$ such that

1) $\mathbf{0} \in W$
2) if $\mathbf{u}, \mathbf{v} \in W$ then $\mathbf{u} + \mathbf{v} \in W$
3) if $\mathbf{u} \in W$ and $c \in \mathbb{R}$ then $c\mathbf{u} \in W$.

Example.

Recall: $\mathbb{P}$ = the vector space of all polynomials.

Proposition

Let V be a vector space and $W \subseteq V$ is a subspace then W is itself a vector space.

Example.

Recall: $\mathcal{F}(\mathbb{R})$ = the vector space of all functions $f\colon \mathbb{R} \to \mathbb{R}$

Some interesting subspaces of $\mathcal{F}(\mathbb{R})$:

1) $C(\mathbb{R})$ = the subspace of all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$

2) $C^n(\mathbb{R})$ = the subspace of all functions $f\colon \mathbb{R} \to \mathbb{R}$ that are differentiable n or more times.

3) $C^\infty(\mathbb{R})$ = the subspace of all smooth functions $f\colon \mathbb{R} \to \mathbb{R}$ (i.e. functions that have derivatives of all orders: f', f'', f''', $\ldots$).

Note. If V is a vector space then:

1) the biggest subspace of V is V itself;

2) the smallest subspace of V is the subspace $\{\mathbf{0}\}$ consisting of the zero vector only;

3) if a subspace of V contains a non-zero vector, then it contains infinitely many vectors.

Definition

Let V, W be vector spaces A *linear transformation* is a function

$$T\colon V \to W$$

which satisfies the following conditions:

1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$

2) $T(c\mathbf{v}) = cT(\mathbf{v})$ for any $\mathbf{v} \in V$ and any scalar c.

Note. If $T\colon V \to W$ is a linear transformation then for any vector $\mathbf{b} \in W$ we can consider the equation

$$T(\mathbf{x}) = \mathbf{b}$$

Definition

If $T\colon V \to W$ is a linear transformation then:

1) The *kernel* of T is the set

$$\operatorname{Ker}(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}$$

2) The *image* of T is the set

$$\operatorname{Im}(T) = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}$$

Proposition

If $T\colon V \to W$ is a linear transformation then:

1) $\mathrm{Ker}(T)$ is a subspace of V

2) $\mathrm{Im}(T)$ is a subspace of W

Theorem

If $T\colon V \to W$ is a linear transformation and v_0 is a solution of the equation

$$T(\mathbf{x}) = \mathbf{b}$$

then all other solutions of this equation are vectors of the form

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{n}$$

where $\mathbf{n} \in \mathrm{Ker}(T)$.

Example.

$$\begin{aligned} D\colon C^\infty(\mathbb{R}) &\longrightarrow C^\infty(\mathbb{R}) \\ f &\longmapsto f' \end{aligned}$$

Recall:

• A vector space is a set V equipped with operations of addition and multiplication by scalars. These operations must satisfy some properties.

• Some examples of vector spaces:

1) $\mathbb{R}^n$ = the vector space of column vectors.

2) $\mathcal{F}(\mathbb{R})$ = the vector space of all functions $f\colon \mathbb{R} \to \mathbb{R}$.

3) $C(\mathbb{R})$ = the vector space of all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$.

4) $C^\infty(\mathbb{R})$ = the vector space of all smooth functions $f\colon \mathbb{R} \to \mathbb{R}$.

5) $M_{m,n}(\mathbb{R})$ = the vector space of all $m \times n$ matrices.

6) $\mathbb{P}$ = the vector space of all polynomials.

7) $\mathbb{P}_n$ = the vector space of polynomials of degree $\leq n$.

• If V, W are vector spaces then a linear transformation is a function $T\colon V \to W$ such that

1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

2) $T(c\mathbf{v}) = cT(\mathbf{v})$

• Many problems involving $\mathbb{R}^n$ can be easily solved using row reduction, matrix multiplication etc.

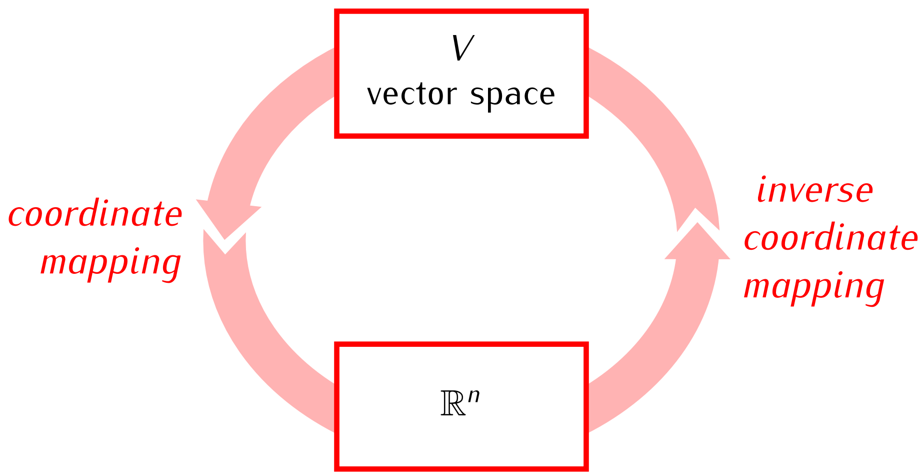
• The same types of problems involving other vector spaces can be difficult to solve.

Next goal:

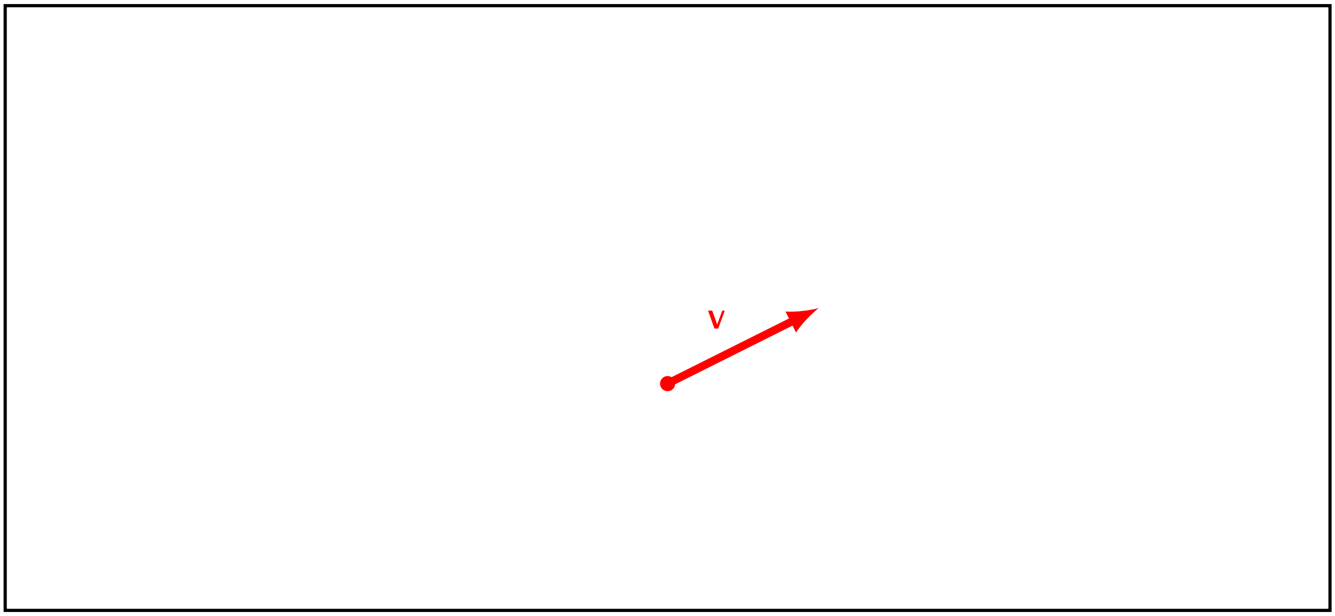
If V is a *finite dimensional* vector space then we can construct a *coordinate mapping*

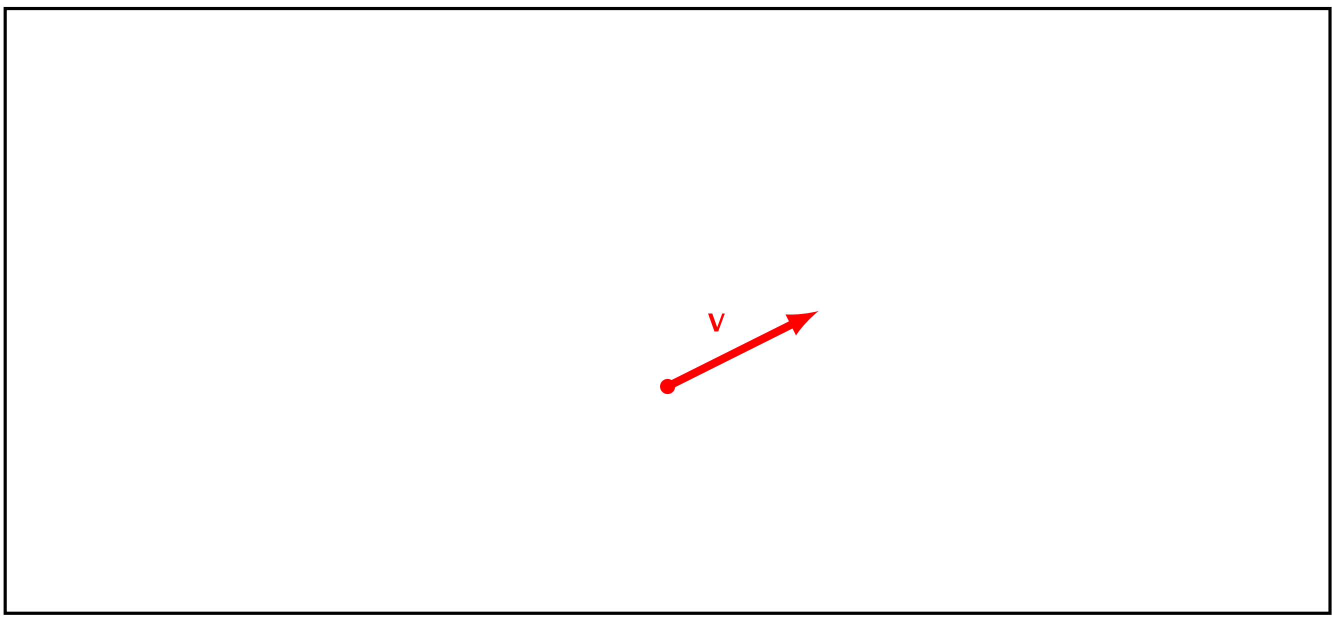
$$V \to \mathbb{R}^n$$

which lets us turn computations in V into computations in $\mathbb{R}^n$.



Motivation: How to assign coordinates to vectors





Definition

If V is a vector space then vector $\mathbf{w} \in V$ is a *linear combination* of vectors $\mathbf{v}_1, \dots \mathbf{v}_p \in V$ if there exist scalars $c_1, \dots, c_p$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + \dots + c_p\mathbf{v}_p$$

Definition

If V is a vector space and $\mathbf{v}_1, \dots, \mathbf{v}_p$ are vectors in V then

$$\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_p) = \left\{ \begin{array}{c} \text{the set of all} \\ \text{linear combinations} \\ c_1\mathbf{v}_1 + \dots + c_p\mathbf{v}_p \end{array} \right\}$$

Definition

If V is a vector space and $\mathbf{v}_1, \dots, \mathbf{v}_p$ are vectors in V such that

$$V = \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_p)$$

the the set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is called the *spanning set* of V.

Example.

Definition

If V is a vector space and $\mathbf{v}_1, \ldots, \mathbf{v}_p \in V$ then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is *linearly independent* if the homogenous equation

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

has only one, trivial solution $x_1 = 0, \ldots, x_p = 0$. Otherwise the set is *linearly dependent*.

Theorem

Let V be a vector space, and let $\mathbf{v}_1, \ldots, \mathbf{v}_p \in V$. If the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent then the equation

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{w}$$

has exactly one solution for any vector $\mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$.

Example.

Recall: $\mathcal{F}(\mathbb{R})$ = the vector space of all functions $f\colon \mathbb{R} \to \mathbb{R}$. Let $f, g, h \in \mathcal{F}(\mathbb{R})$ be the following functions:

$$f(t) = \sin(t), \quad g(t) = \cos(t), \quad h(t) = \cos^2(t)$$

Check if the set $\{f, g, h\}$ is linearly independent.

Example.

Let $f, g, h \in \mathcal{F}(\mathbb{R})$ be the following functions:

$$f(t) = \sin^2(t), \quad g(t) = \cos^2(t), \quad h(t) = \cos 2t$$

Check if the set $\{f, g, h\}$ is linearly independent.

Definition

A *basis* of a vector space V is an ordered set of vectors

$$\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$$

such that

1) $\text{Span}(\mathbf{b}_1, \ldots, \mathbf{b}_n) = V$

2) The set $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is linearly independent.

Theorem

A set $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis of a vector space V if any only if for each $\mathbf{v} \in V$ the vector equation

$$x_1\mathbf{b}_1 + \ldots + x_n\mathbf{b}_n = \mathbf{v}$$

has a unique solution.

Definition

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis of a vector space V. For $\mathbf{v} \in V$ let $c_1, \ldots, c_n$ be the unique numbers such that

$$c_1\mathbf{b}_1 + \ldots + c_n\mathbf{b}_n = \mathbf{v}$$

Then the vector

$$\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$$

is called the *coordinate vector of* $\mathbf{v}$ *relative to the basis* $\mathcal{B}$ and it is denoted by $\big[\mathbf{v}\big]_{\mathcal{B}}$.

Example. Let $\mathcal{E} = \{1, t, t^2\}$ be the standard basis of $\mathbb{P}_2$, and let

$$p(t) = 3 + 2t - 4t^2$$

Find the coordinate vector $\big[p\big]_{\mathcal{E}}$.

Example. Let $\mathcal{B} = \{1, 1+t, 1+t+t^2\}$. One can check that $\mathcal{B}$ is a basis of $\mathbb{P}_2$. Let

$$p(t) = 3 + 2t - 4t^2$$

Find the coordinate vector $\big[p\big]_{\mathcal{B}}$.

Recall:

- A basis of a vector space V is a set of vectors $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ such that

  1) $\mathrm{Span}(\mathbf{b}_1, \ldots, \mathbf{b}_n) = V$

  2) The set $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is linearly independent.

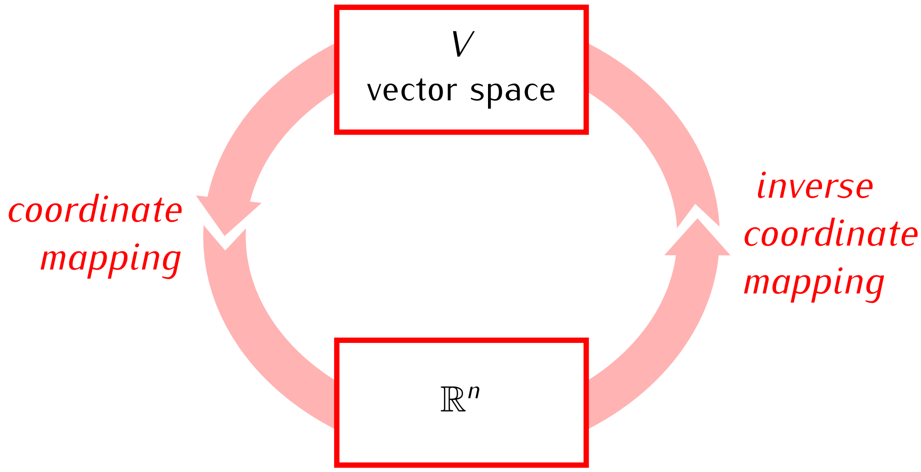
- For $\mathbf{v} \in V$ let $c_1, \ldots, c_n$ be the unique numbers such that

$$c_1\mathbf{b}_1 + \ldots + c_n\mathbf{b}_n = \mathbf{v}$$

The vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$$

is called the *coordinate vector of* $\mathbf{v}$ *relative to the basis* $\mathcal{B}$.



Theorem

Let $\mathcal{B}$ be a basis of a vector space V. If $\mathbf{v}_1, \dots \mathbf{v}_p, \mathbf{w} \in V$ then:

1) Solutions of the equation $x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{w}$ are the same as solutions of the equation $x_1\left[\mathbf{v}_1\right]_{\mathcal{B}} + \ldots + x_p\left[\mathbf{v}_p\right]_{\mathcal{B}} = \left[\mathbf{w}\right]_{\mathcal{B}}$.

2) The set of vectors $\{\mathbf{v}_1, \dots \mathbf{v}_p\}$ is linearly independent if and only if the set $\left\{\left[\mathbf{v}_1\right]_{\mathcal{B}}, \dots, \left[\mathbf{v}_p\right]_{\mathcal{B}}\right\}$ is linearly independent.

3) $\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_p) = V$ if any only if $\mathrm{Span}\left(\left[\mathbf{v}_1\right]_{\mathcal{B}}, \dots, \left[\mathbf{v}_p\right]_{\mathcal{B}}\right) = \mathbb{R}^n$.

4) $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis of V if and only if $\left\{\left[\mathbf{v}_1\right]_{\mathcal{B}}, \dots, \left[\mathbf{v}_p\right]_{\mathcal{B}}\right\}$ is a basis of $\mathbb{R}^n$.

Example. Recall that $\mathbb{P}_2$ is the vector space of polynomials of degree ≤ 2. Consider the following polynomials in $\mathbb{P}_2$:

$$p_1(t) = 1 + 2t + t^2$$
$$p_2(t) = 3 + t + 2t^2$$
$$p_3(t) = 1 - 8t - t^2$$

Check if the set $\{p_1, p_2, p_3\}$ is linearly independent.

Theorem

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be vectors in $\mathbb{R}^n$. The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a basis of $\mathbb{R}^n$ if and only if the matrix

$$A = \left[\begin{array}{ccc} \mathbf{v}_1 & \ldots & \mathbf{v}_p \end{array} \right]$$

has a pivot position in every row and in every column (i.e. if A is an invertible matrix).

Corollary

Every basis of $\mathbb{R}^n$ consists of n vectors.

Theorem

Let V be a vector space. If V has a basis consisting of n vectors then every basis of V consists of n vectors.

Definition

A vector space has *dimension* n if V has a basis consisting of n vectors. Then we write $\dim V = n$.

Example.

Theorem

Let V be a vector space such that $\dim V = n$, and let $\mathbf{v}_1, \dots \mathbf{v}_p \in V$.

1) If $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a spanning set of V then $p \geq n$.

2) If $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a linearly independent set then $p \leq n$.

Corollary

Let V be a vector space such that $\dim V = n$. If W be a subspace of V then $\dim W \leq n$. Moreover, if $\dim W = n$ then $W = V$.

Note.

1) One can show that every vector space has a basis.

2) Some vector spaces have bases consisting of infinitely many vectors. If V is such vector space then we write $\dim V = \infty$.

Example.

Recall:

If $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ is an $m \times n$ matrix then:

1) $\text{Col}(A) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$

2) $\text{Nul}(A) = \{\mathbf{v} \in \mathbb{R}^m \mid A\mathbf{v} = \mathbf{0}\}$

Construction of a basis of Col(A)

Lemma

Let V be a vector space, and let $\mathbf{v}_1, \ldots, \mathbf{v}_p \in V$. If a vector $\mathbf{v}_i$ is a linear combination of the other vectors then

$$\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p) = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_p)$$

Upshot. One can construct a basis of a vector space V as follows:

- Start with a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ such that $\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_p) = V$.
- Keep removing vectors without changing the span, until you get a linearly independent set.

Example. Find a basis of Col(A) where A is the following matrix:

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

Example. Find a basis of Col(A) where A is the following matrix:

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

Construction of a basis of Nul(A)

Example. Find a basis of Nul(A) where A is the following matrix:

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

Upshot. If A is matrix then:

$\dim \operatorname{Col}(A)$ = the number of pivot columns of A

$\dim \operatorname{Nul}(A)$ = the number of non-pivot columns of A

Definition

If A is a matrix then:

- the dimension of $\operatorname{Col}(A)$ is called the *rank* of A and it is denoted $\operatorname{rank}(A)$
- the dimension of $\operatorname{Nul}(A)$ is called the *nullity* of A.

The Rank Theorem

If A is an $m \times n$ matrix then

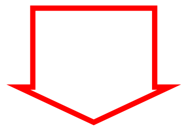
$$\operatorname{rank}(A) + \dim \operatorname{Nul}(A) = n$$

Example. Let A be a 100×101 matrix such that $\dim \mathrm{Nul}(A) = 1$. Show that the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b} \in \mathbb{R}^{100}$.

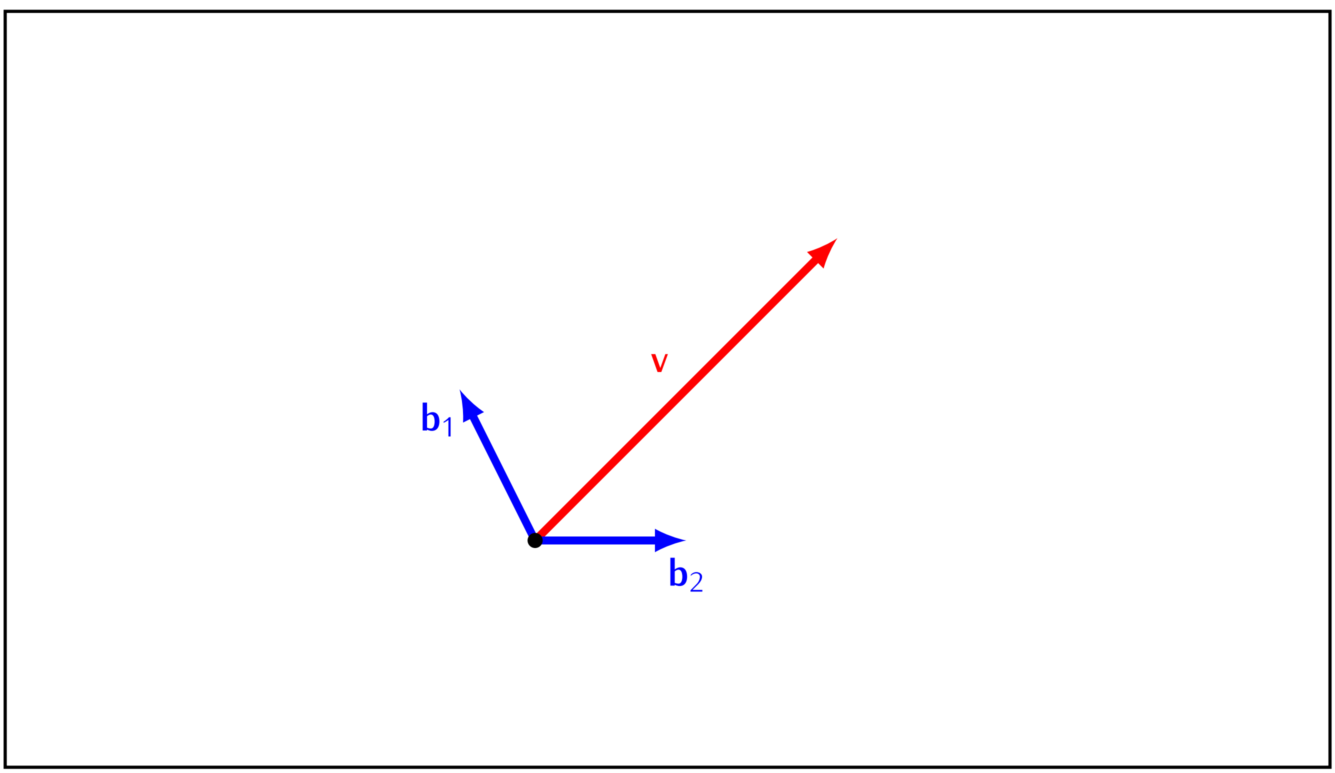
Example. Let A be a 5×9. Can the null space of A have dimension 3?

Recall: Any basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of a vector space V defines a coordinate system:

$$\mathbf{v} = c_1\mathbf{b}_1 + \ldots + c_n\mathbf{b}_n = \mathbf{v}$$

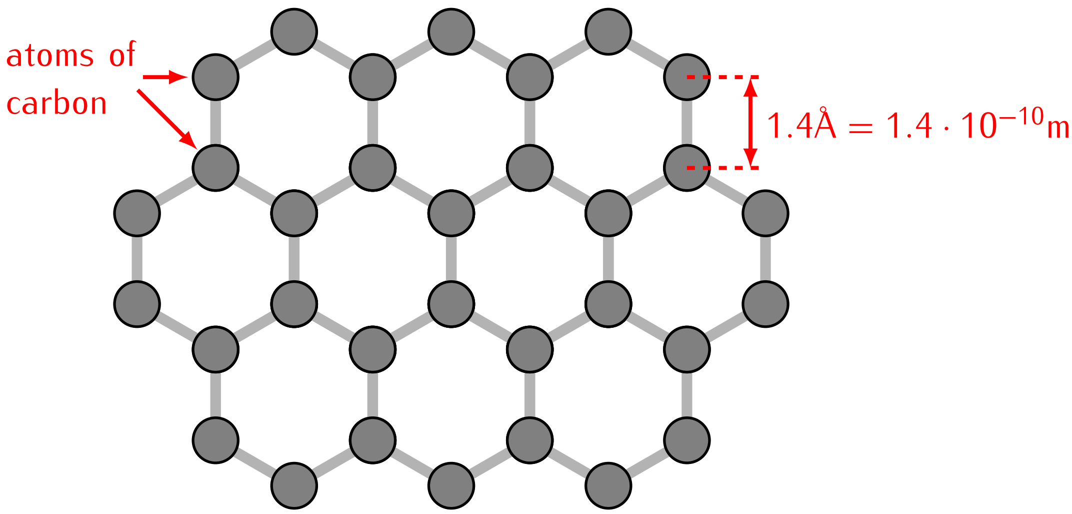


$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$



Note. Choosing a convenient basis can simplify computations.

Example. Graphene lattice.



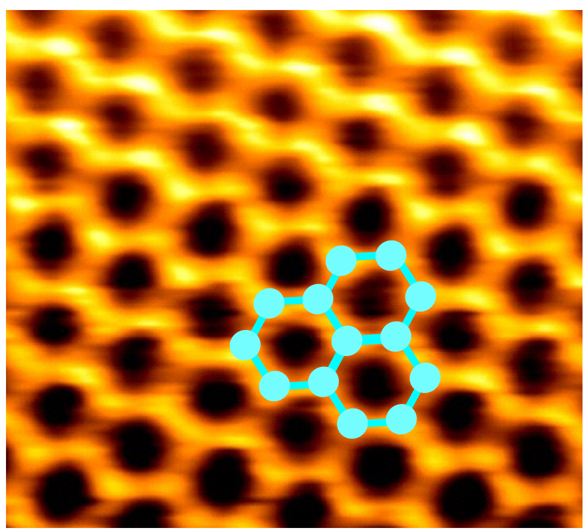


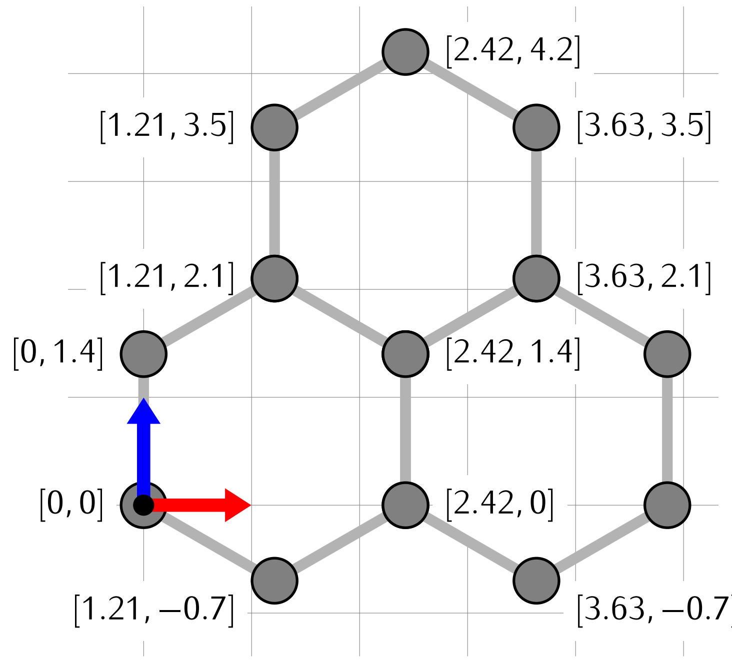
Image of graphene taken with an atomic force microscope.
© University of Augsburg, Experimental Physics IV.

Coordinates of atoms in the graphene lattice

In the standard basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

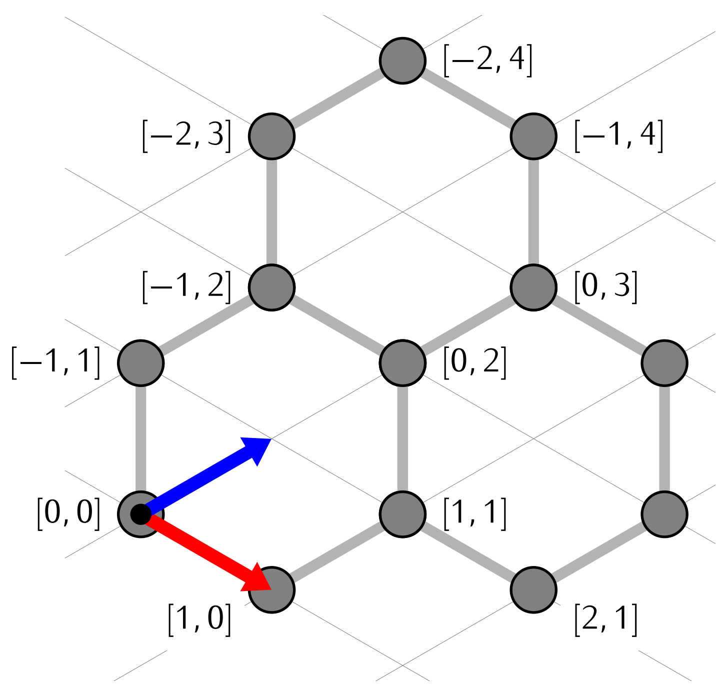
$$\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



In a more convenient basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$:

$$\mathbf{b}_1 = \begin{bmatrix} 1.21 \\ -0.7 \end{bmatrix}$$

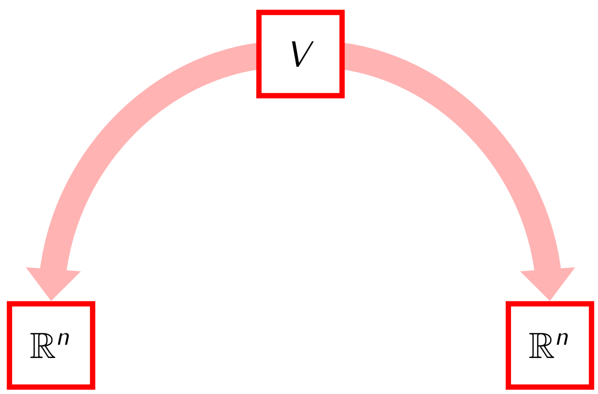
$$\mathbf{b}_2 = \begin{bmatrix} 1.21 \\ 0.7 \end{bmatrix}$$



Problem Let

$$\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}, \quad \mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_1\}$$

be two bases of a vector space V, and let $\mathbf{v} \in V$. Assume that we know $\big[\mathbf{v}\big]_{\mathcal{B}}$. What is $\big[\mathbf{v}\big]_{\mathcal{D}}$?



Definition

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_1\}$ be two bases of a vector space V. The matrix

$$P_{\mathcal{D}\leftarrow\mathcal{B}} = \left[\ \left[\mathbf{b}_1\right]_{\mathcal{D}} \ \ \left[\mathbf{b}_2\right]_{\mathcal{D}} \quad \ldots \quad \left[\mathbf{b}_n\right]_{\mathcal{D}}\ \right]$$

is called the *change of coordinates matrix* from the basis $\mathcal{B}$ to the basis $\mathcal{D}$.

Propostion

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_1\}$ be two bases of a vector space V. For any vector $\mathbf{v} \in V$ we have

$$\left[\mathbf{v}\right]_{\mathcal{D}} = P_{\mathcal{D}\leftarrow\mathcal{B}} \cdot \left[\mathbf{v}\right]_{\mathcal{B}}$$

Example. Let $\mathbb{P}_2$ be the vector space of polynomials of degree ≤ 2. Consider two bases of $\mathbb{P}_2$:

$$\mathcal{B} = \{1, 1+t, 1+t+t^2\}$$
$$\mathcal{D} = \{1+t, 1-5t, 2+t^2\}$$

1) Compute the change of coordinates matrix $P_{\mathcal{D}\leftarrow\mathcal{B}}$.

2) Let $p \in \mathbb{P}_2$ be a polynomial such that

$$\big[p\big]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

Compute $\big[p\big]_{\mathcal{D}}$.

Proposition

If $\mathcal{B}, \mathcal{D}, \mathcal{E}$ are three bases of a vector space V then:

1) $P_{\mathcal{B}\leftarrow\mathcal{D}} = (P_{\mathcal{D}\leftarrow\mathcal{B}})^{-1}$
2) $P_{\mathcal{E}\leftarrow\mathcal{B}} = P_{\mathcal{E}\leftarrow\mathcal{D}} \cdot P_{\mathcal{D}\leftarrow\mathcal{B}}$

What we want:

What we have:

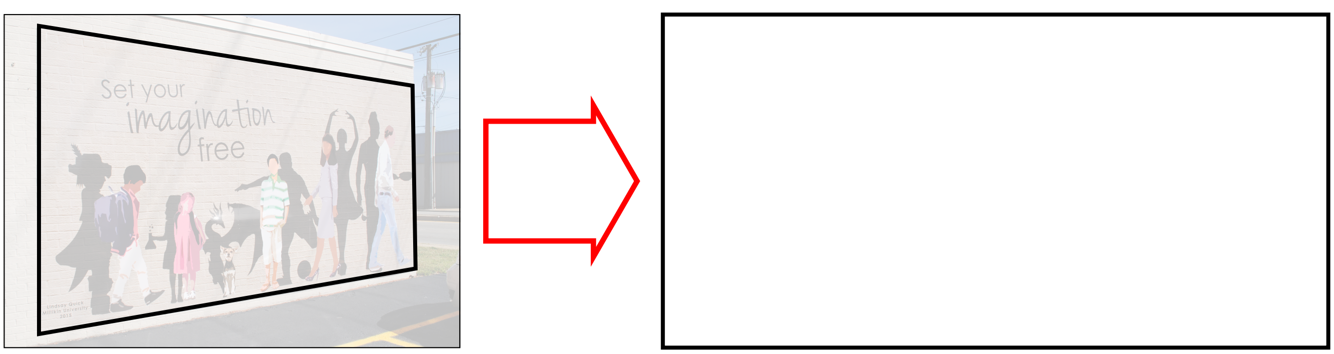
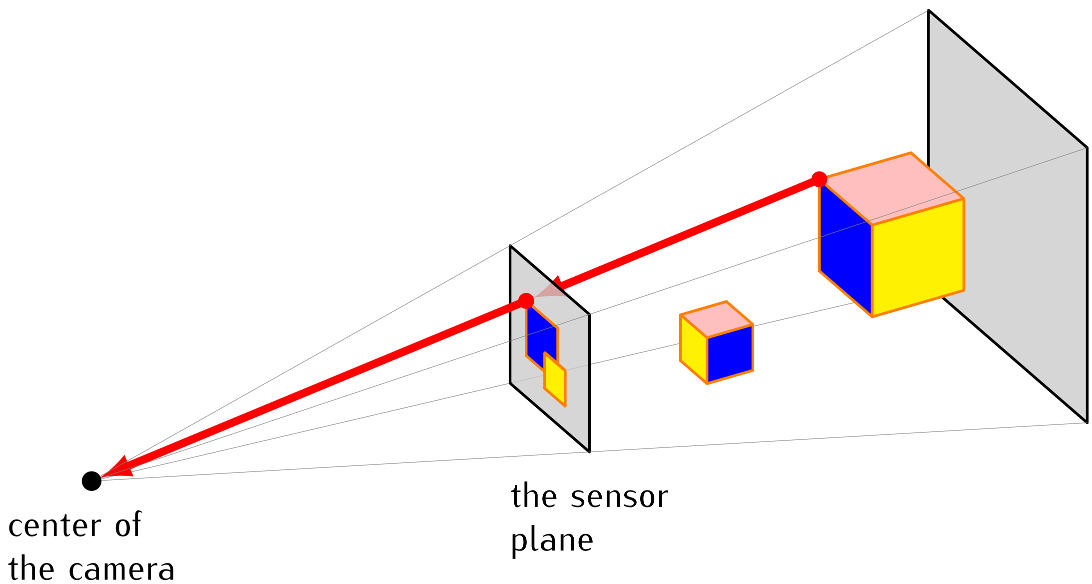
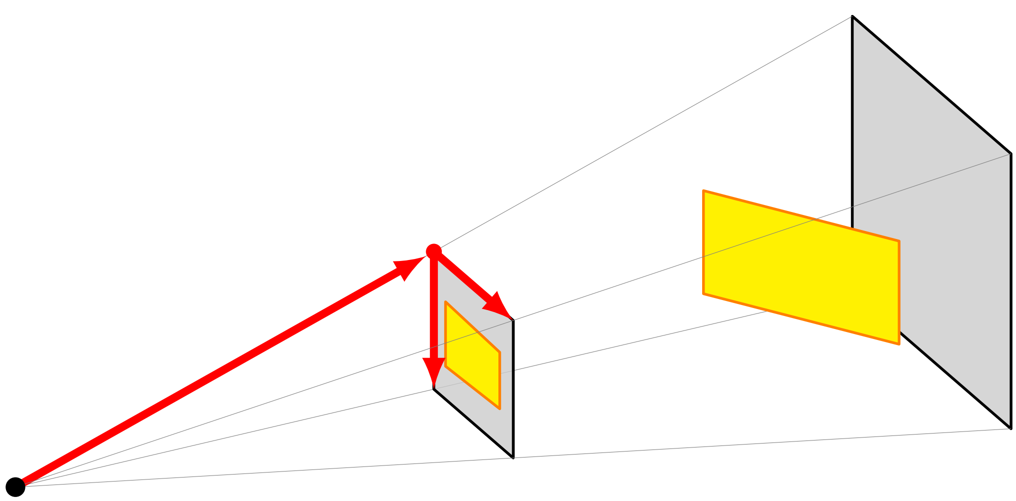
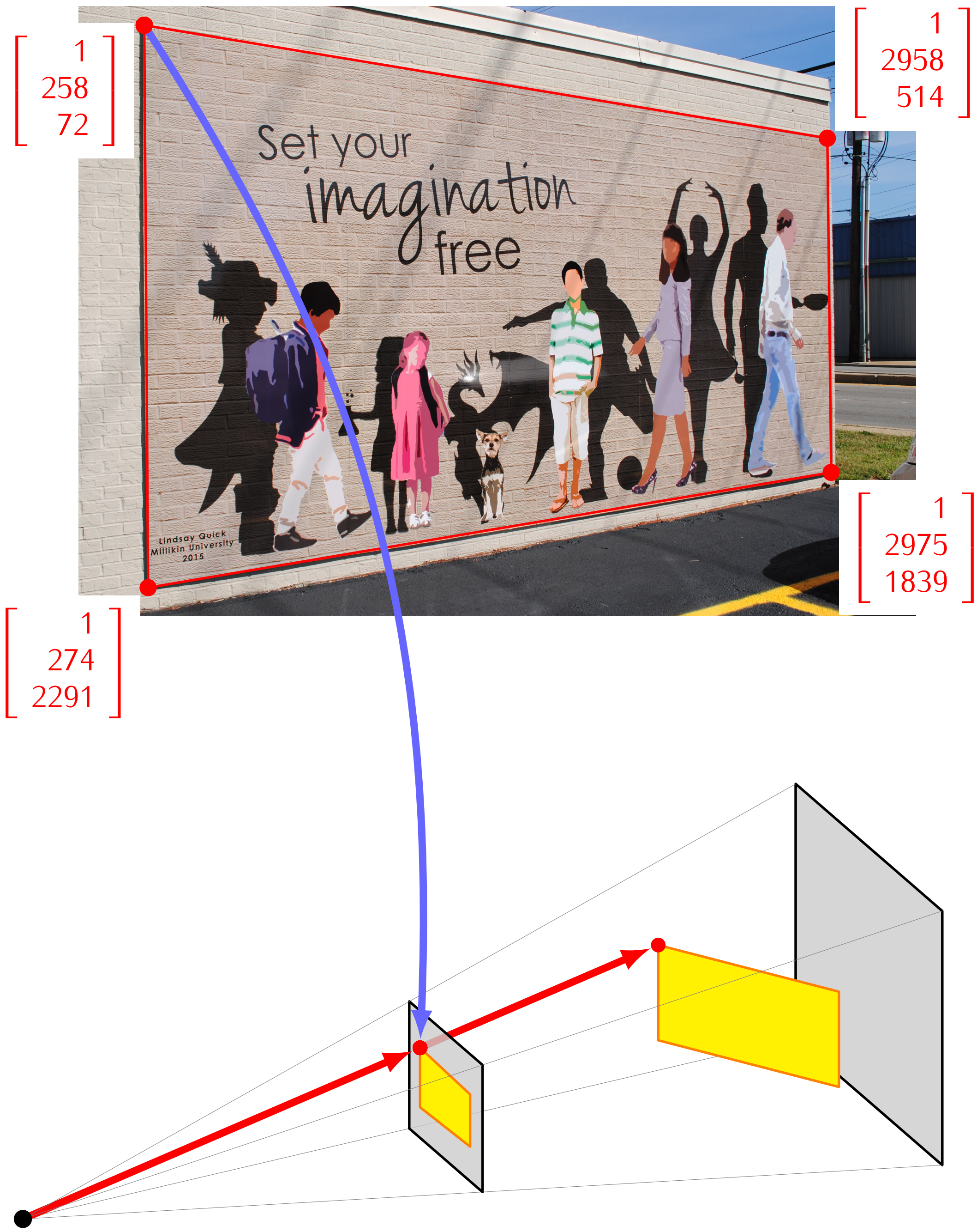


Image formation in a camera

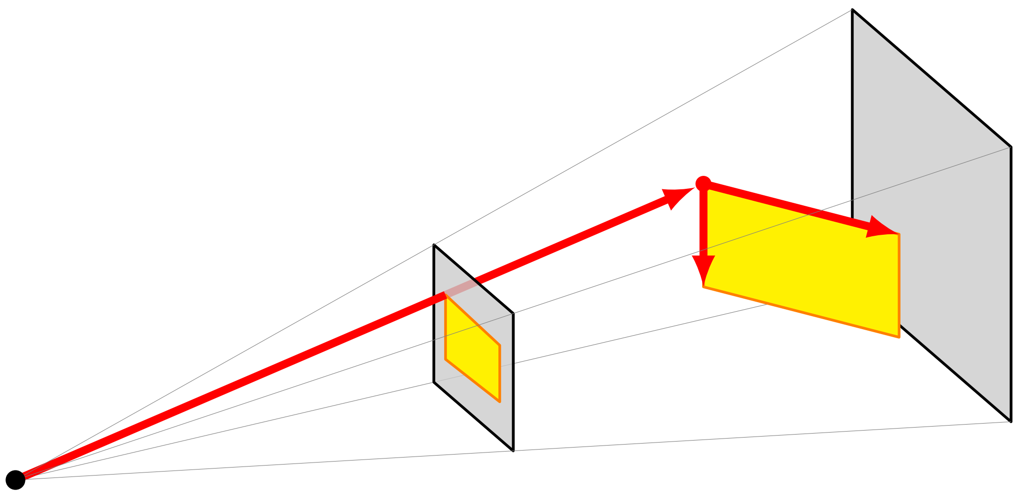


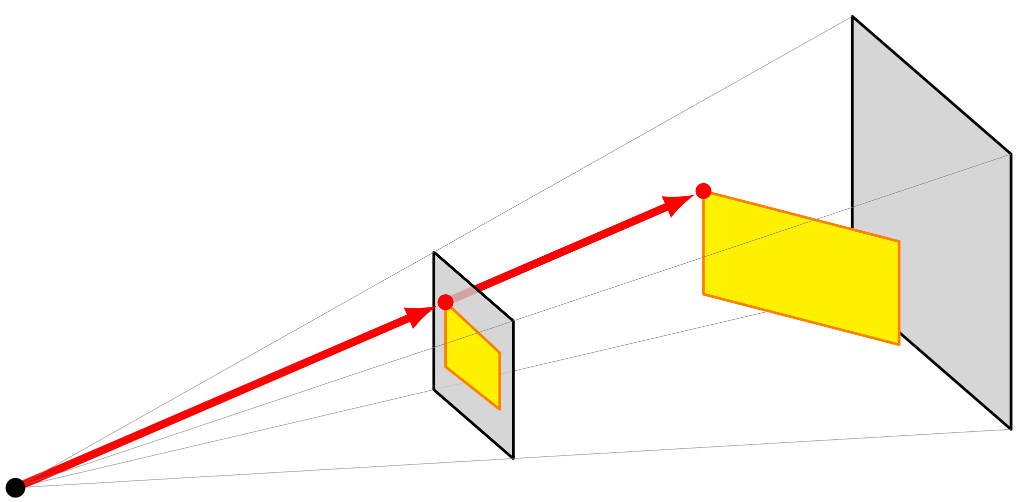
The camera coordinate system $\mathcal{C}$

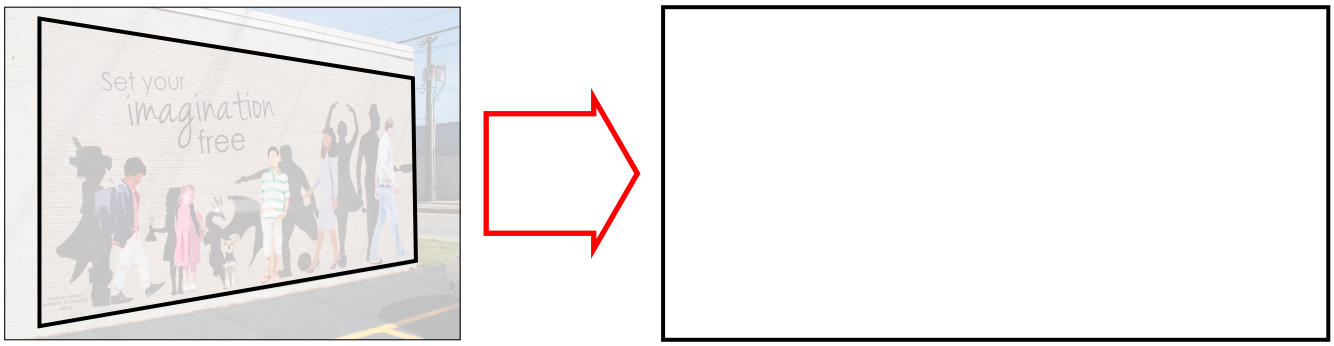



1
258
72
1
2958
514
1
2975
1839
1
274
2291
Set your imagination free
Lindsay Quick
Millikin University
2015

The mural coordinate system $\mathcal{M}$

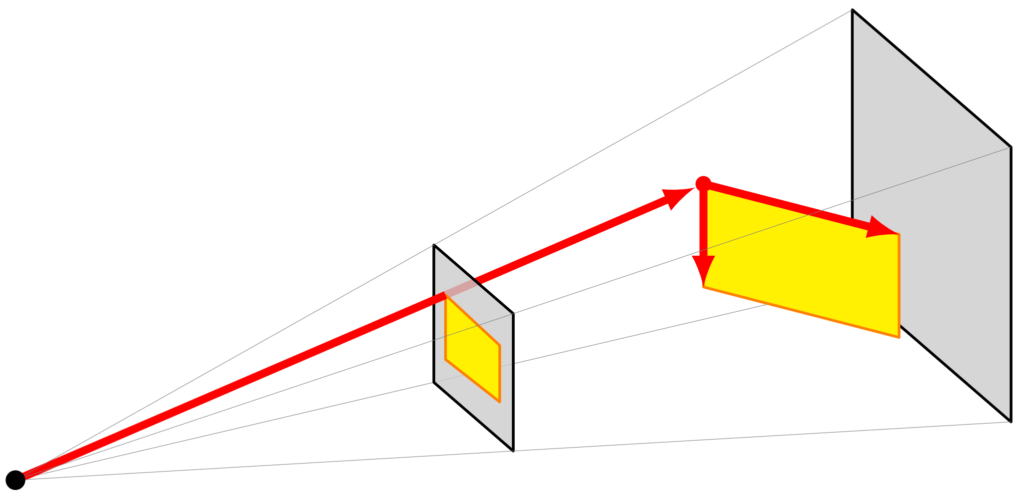





Set your
imagination
free

From mural coordinates to camera coordinates

$$P_{\mathcal{C}\leftarrow\mathcal{M}} = \left[\ [\mathbf{m}_1]_{\mathcal{C}}\ \ [\mathbf{m}_2]_{\mathcal{C}}\ \ [\mathbf{m}_3]_{\mathcal{C}}\ \right]$$



Problem: What are the numbers a, b, c?

Definition

If

$$\mathbf{u} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

are vectors in $\mathbb{R}^n$ then the *inner product* (or *dot product*) of $\mathbf{u}$ and $\mathbf{v}$ is the number

$$\mathbf{u} \cdot \mathbf{v} = a_1 b_1 + \ldots + a_n b_n$$

Properties of the dot product:

1) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$
4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

Definition

If $\mathbf{u} \in \mathbb{R}^n$ then the *length* (or the *norm*) of $\mathbf{u}$ is the number

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

Note. If $\mathbf{u} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ then $||\mathbf{u}|| = \sqrt{a_1^2 + \ldots + a_n^2}$.

Properties of the norm:

1) $||u|| \geq 0$ and $||u|| = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

2) $||cu|| = |c| \cdot ||u||$

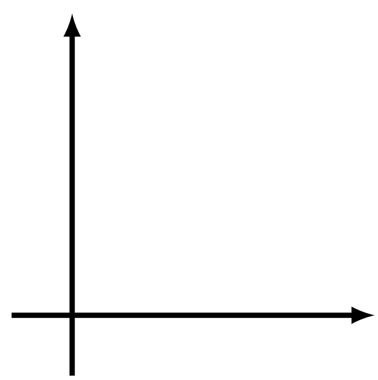
Definition

A vector $\mathbf{u} \in \mathbb{R}^n$ is an *unit vector* if $||\mathbf{u}|| = 1$.

Definition

If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ then the *distance* between $\mathbf{u}$ and $\mathbf{v}$ is the number

$$\text{dist}(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$$



Note. If $\mathbf{u} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ then

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \sqrt{(a_1 - b_1)^2 + \ldots + (a_n - b_n)^2}$$

Definition

Vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are *orthogonal* if $\mathbf{u} \cdot \mathbf{v} = 0$.

Pythagorean Theorem

Vectors $\mathbf{u}, \mathbf{v}$ are orthogonal if and only if

$$||\mathbf{u}||^2 + ||\mathbf{v}||^2 = ||\mathbf{u} + \mathbf{v}||^2$$

Definition

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is an *orthogonal set* if each pair each pair of vectors in this set is orthogonal, i.e.

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0$$

for all $i \neq j$.

Example.

$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is an orthogonal set in $\mathbb{R}^3$.

Example.

$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -5 \\ 3 \end{bmatrix} \right\}$ is another orthogonal set in $\mathbb{R}^3$.

Proposition

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set of non-zero vectors in $\mathbb{R}^n$ then this set is linearly independent.

Recall: Any linearly independent set of n vectors in $\mathbb{R}^n$ is a basis of $\mathbb{R}^n$.

Corollary

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal set of n non-zero vectors in $\mathbb{R}^n$ then this set is a basis of $\mathbb{R}^n$.

Definition

If V is a subspace of $\mathbb{R}^n$ then we say that a set $\{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is an *orthogonal basis* of V if

1) $\{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is a basis of V and

2) $\{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is an orthogonal set.

Recall. If $\mathcal{B} = \{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is a basis of a vector space V and $\mathbf{w} \in V$ then the coordinate vector of $\mathbf{w}$ relative to $\mathcal{B}$ is the vector

$$[\,\mathbf{w}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

where $c_1, \ldots, c_k$ are scalars such that $c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k = \mathbf{w}$.

Propostion

If $\mathcal{B} = \{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is an orthogonal basis of V and $\mathbf{w} \in V$ then

$$[\,\mathbf{w}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

where $c_i = \dfrac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} = \dfrac{\mathbf{w} \cdot \mathbf{v}_i}{||\mathbf{v}_i||^2}$

Example. Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -5 \\ 3 \end{bmatrix} \right\}, \qquad \mathbf{w} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

The set $\mathcal{B}$ is an orthogonal basis of $\mathbb{R}^3$. Compute $\begin{bmatrix} \mathbf{w} \end{bmatrix}_{\mathcal{B}}$.

Theorem (Gram-Schmidt Process)

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a basis of V. Define vectors $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ as follows:

$$\mathbf{w}_1 = \mathbf{v}_1$$

$$\mathbf{w}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{w}_1 \cdot \mathbf{v}_2}{\mathbf{w}_1 \cdot \mathbf{w}_1}\right)\mathbf{w}_1$$

$$\mathbf{w}_3 = \mathbf{v}_3 - \left(\frac{\mathbf{w}_1 \cdot \mathbf{v}_3}{\mathbf{w}_1 \cdot \mathbf{w}_1}\right)\mathbf{w}_1 - \left(\frac{\mathbf{w}_2 \cdot \mathbf{v}_3}{\mathbf{w}_2 \cdot \mathbf{w}_2}\right)\mathbf{w}_2$$

$$\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots$$

$$\mathbf{w}_k = \mathbf{v}_k - \left(\frac{\mathbf{w}_1 \cdot \mathbf{v}_k}{\mathbf{w}_1 \cdot \mathbf{w}_1}\right)\mathbf{w}_1 - \left(\frac{\mathbf{w}_2 \cdot \mathbf{v}_k}{\mathbf{w}_2 \cdot \mathbf{w}_2}\right)\mathbf{w}_2 - \dots - \left(\frac{\mathbf{w}_{k-1} \cdot \mathbf{v}_k}{\mathbf{w}_{k-1} \cdot \mathbf{w}_{k-1}}\right)\mathbf{w}_{k-1}$$

Then the set $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ is an orthogonal basis of V.

Example. In $\mathbb{R}^4$ take

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 7 \\ 4 \\ 3 \\ -3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 5 \\ 7 \\ 7 \\ 8 \end{bmatrix}$$

The set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis of some subspace $V \subseteq \mathbb{R}^4$. Find an orthogonal basis of V.

Definition

An orthogonal basis $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ of V is called an *orthonormal basis* if $||\mathbf{w}_i|| = 1$ for $i = 1, \ldots, k$.

Propostion

If $\mathcal{B} = \{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is an orthonormal basis of V and $\mathbf{w} \in V$ then

$$[\ \mathbf{w}\ ]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

where $c_i = \mathbf{w} \cdot \mathbf{v}_i$.

Note. If $\mathcal{B} = \{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is an orthogonal basis of V then

$$\mathcal{C} = \left\{ \frac{\mathbf{v}_1}{||\mathbf{v}_1||}, \ldots, \frac{\mathbf{v}_k}{||\mathbf{v}_k||} \right\}$$

is an orthonormal basis of V.

Recall:

1) If

$$\mathbf{u} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

are vectors in $\mathbb{R}^n$ then:

- $\mathbf{u} \cdot \mathbf{v} = a_1 b_1 + \ldots + a_n b_n$
- $||\mathbf{u}|| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$
- $\text{dist}(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$

2) Vectors $\mathbf{u}, \mathbf{v}$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$.

3) Pythagorean theorem: $\mathbf{u}, \mathbf{v}$ are orthogonal if and only if

$$||\mathbf{u}||^2 + ||\mathbf{v}||^2 = ||\mathbf{u} + \mathbf{v}||^2$$

4) If $V \subseteq \mathbb{R}^n$ is a subspace then an orthogonal basis of V is a basis which consists of vectors that are orthogonal to one another.

5) If $\mathcal{B} = \{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ is an orthogonal basis of V and $\mathbf{w} \in V$ then

$$[\ \mathbf{w}\ ]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

where $c_i = \dfrac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$.

6) Gram-Schmidt process:

a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of $V \subseteq \mathbb{R}^n$ → *G-S process* → an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ of V

$$\mathbf{w}_1 = \mathbf{v}_1$$

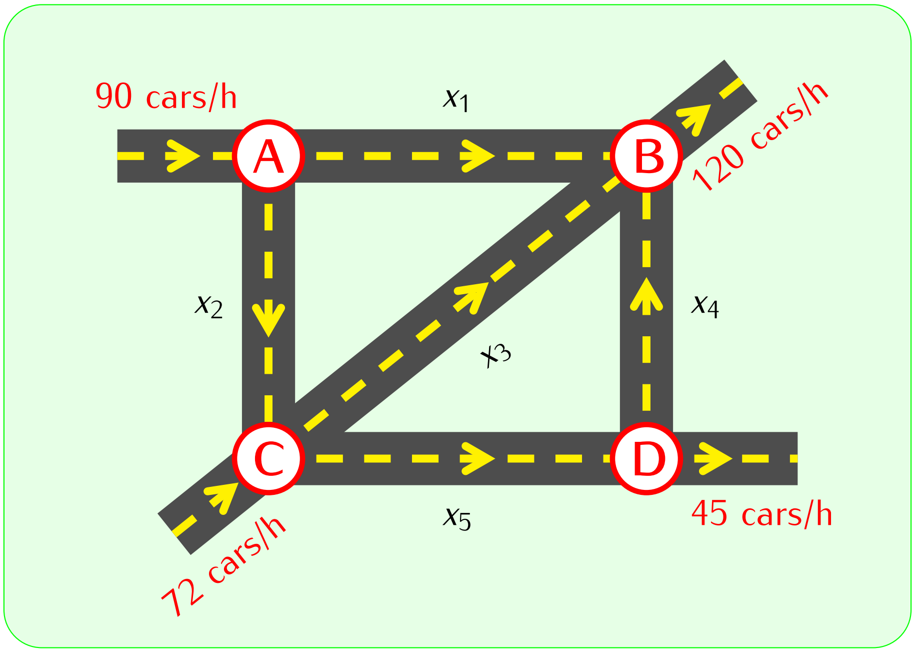
$$\mathbf{w}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{w}_1 \cdot \mathbf{v}_2}{\mathbf{w}_1 \cdot \mathbf{w}_1}\right)\mathbf{w}_1$$

$$\mathbf{w}_3 = \mathbf{v}_3 - \left(\frac{\mathbf{w}_1 \cdot \mathbf{v}_3}{\mathbf{w}_1 \cdot \mathbf{w}_1}\right)\mathbf{w}_1 - \left(\frac{\mathbf{w}_2 \cdot \mathbf{v}_3}{\mathbf{w}_2 \cdot \mathbf{w}_2}\right)\mathbf{w}_2$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\mathbf{w}_k = \mathbf{v}_k - \left(\frac{\mathbf{w}_1 \cdot \mathbf{v}_k}{\mathbf{w}_1 \cdot \mathbf{w}_1}\right)\mathbf{w}_1 - \left(\frac{\mathbf{w}_2 \cdot \mathbf{v}_k}{\mathbf{w}_2 \cdot \mathbf{w}_2}\right)\mathbf{w}_2 - \ldots - \left(\frac{\mathbf{w}_{k-1} \cdot \mathbf{v}_k}{\mathbf{w}_{k-1} \cdot \mathbf{w}_{k-1}}\right)\mathbf{w}_{k-1}$$

Problem. Find the flow rate of cars on each segment of streets:

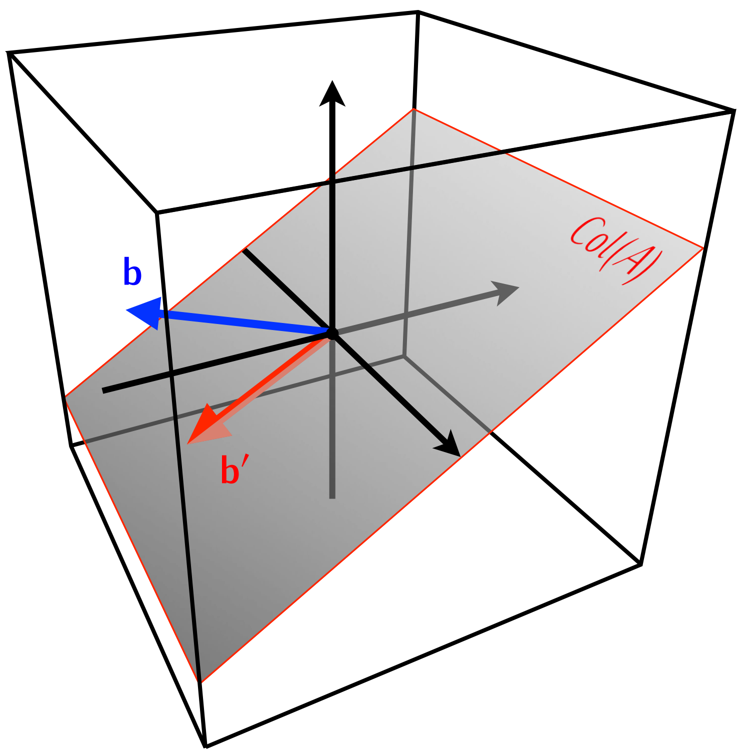


Upshot.

- Recall: a matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b} \in \text{Col}(A)$.

- In practical applications we may obtain a matrix equation that has no solutions, i.e. where $\mathbf{b} \notin \text{Col}(A)$.

- In such cases we may look for approximate solutions as follows:
  - replace $\mathbf{b}$ by a vector $\mathbf{b}'$ such that $\mathbf{b}' \in \text{Col}(A)$ and $\text{dist}(\mathbf{b}, \mathbf{b}')$ is a as small as possible.
  - then solve $A\mathbf{x} = \mathbf{b}'$



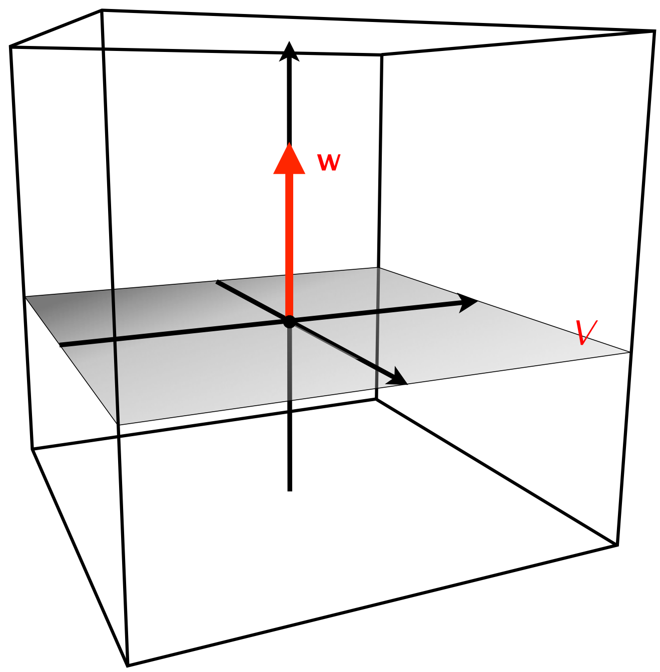
Definition

Given $\mathbf{b}' \in \text{Col}(A)$ as above we will say that a vector $\mathbf{v}$ is a *least square solution* of the equation $A\mathbf{x} = \mathbf{b}$ if $\mathbf{v}$ is a solution of the equation $A\mathbf{x} = \mathbf{b}'$.

Next: How to find the vector $\mathbf{b}'$?

Definition

Let V be a subspace of $\mathbb{R}^n$. A vector $\mathbf{w} \in \mathbb{R}^n$ is *orthogonal to* V if $\mathbf{w} \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in V$.



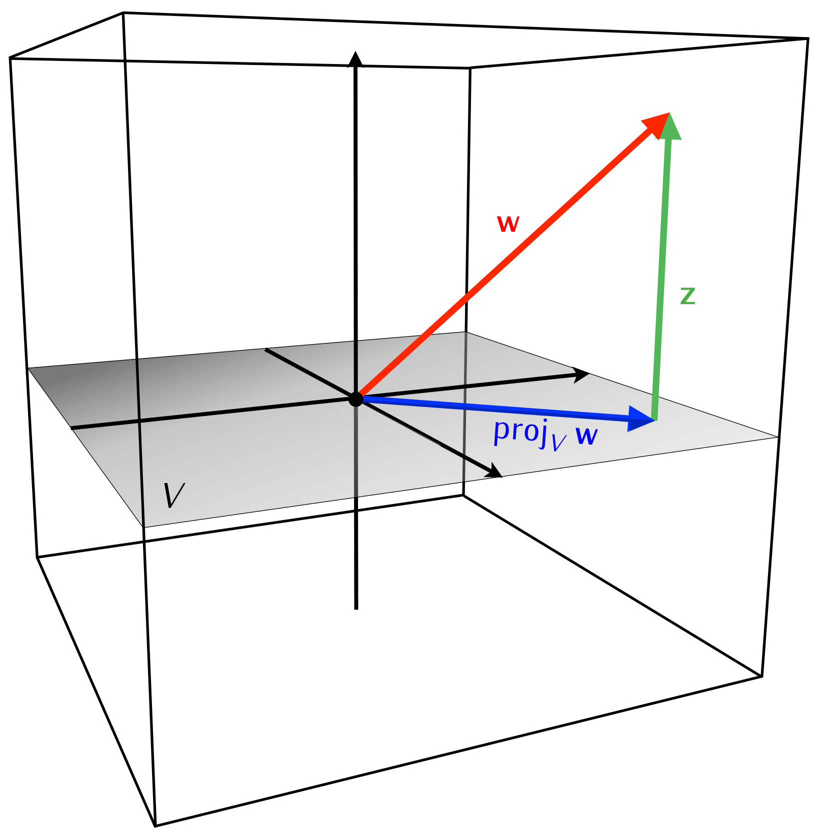
Proposition

If $V = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ then a vector $\mathbf{w} \in \mathbb{R}^n$ is orthogonal to V if and only if $\mathbf{w} \cdot \mathbf{v}_i = 0$ for $i = 1, \ldots, k$.

Definition

Let V be a subspace of $\mathbb{R}^n$ and let $\mathbf{w} \in \mathbb{R}^n$ the *orthogonal projection of* $\mathbf{w}$ *onto* V is a vector $\text{proj}_V \mathbf{w}$ such that

1) $\text{proj}_V \mathbf{w} \in V$
2) the vector $\mathbf{z} = \mathbf{w} - \text{proj}_V \mathbf{w}$ is orthogonal to V.



The Best Approximation Theorem

If V is a subspace of $\mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ then $\text{proj}_V \mathbf{w}$ is a vector in V which is closest to $\mathbf{w}$:

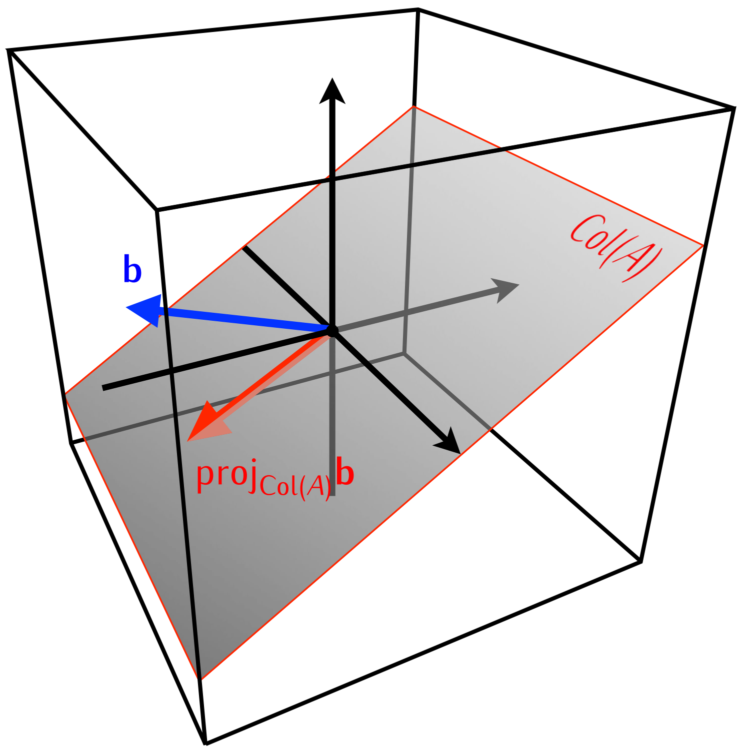
$$\text{dist}(\mathbf{w}, \text{proj}_V \mathbf{w}) \leq \text{dist}(\mathbf{w}, \mathbf{v})$$

for all $\mathbf{v} \in V$.

Corollary

The least square solutions of a matrix equation $A\mathbf{x} = \mathbf{b}$ are solutions of the equation

$$A\mathbf{x} = \text{proj}_{\text{Col}(A)}\mathbf{b}$$



Next: If V is a subspace of $\mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ how to compute $\text{proj}_V\mathbf{w}$?

Theorem

If V is a subspace of $\mathbb{R}^n$ with an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\mathbf{w} \in \mathbb{R}^n$ then

$$\text{proj}_V \mathbf{w} = \left(\frac{\mathbf{w} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \right) \mathbf{v}_1 + \ldots + \left(\frac{\mathbf{w} \cdot \mathbf{v}_k}{\mathbf{v}_k \cdot \mathbf{v}_k} \right) \mathbf{v}_k$$

Corollary

If V is a subspace of $\mathbb{R}^n$ with an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\mathbf{w} \in \mathbb{R}^n$ then

$$\text{proj}_V \mathbf{w} = (\mathbf{w} \cdot \mathbf{v}_1)\, \mathbf{v}_1 + \ldots + (\mathbf{w} \cdot \mathbf{v}_k)\, \mathbf{v}_k$$

Example. Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ 5 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 0 \\ -2 \end{bmatrix} \right\}, \qquad \mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}$$

The set $\mathcal{B}$ is an orthogonal basis of some subspace V of $\mathbb{R}^4$. Compute $\mathrm{proj}_V \mathbf{w}$.

Note. In general if V is a subspace of $\mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ then in order to find $\text{proj}_V \mathbf{w}$ we need to do the following:

1) find a basis of V.

2) use the Gram-Schmidt process to get an orthogonal basis of V

3) use the orthogonal basis to compute $\text{proj}_V \mathbf{w}$.

Example. Consider the following matrix A and vector $\mathbf{u}$:

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 3 & 4 & 2 \\ 2 & 6 & 3 & -1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$$

Compute $\text{proj}_{\text{Col}(A)} \mathbf{u}$.

Example. Find least square solutions of the matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 90 \\ 120 \\ 72 \\ 45 \end{bmatrix}$$

Recall:

1) The least square solutions of a matrix equation $A\mathbf{x} = \mathbf{b}$ are the solutions of the equation

$$A\mathbf{x} = \text{proj}_{\text{Col}(A)}\mathbf{b}$$

2) If $A\mathbf{x} = \mathbf{b}$ is a consistent equation, then $\mathbf{b} \in \text{Col}(A)$, and $\text{proj}_{\text{Col}(A)}\mathbf{b} = \mathbf{b}$. In such case the least square solutions of $A\mathbf{x} = \mathbf{b}$ are just the ordinary solutions.

3) If $A\mathbf{x} = \mathbf{b}$ is inconsistent, then the least square solutions are the best substitute for the (nonexistent) ordinary solutions.

4) If $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an orthogonal basis of a subspace V of $\mathbb{R}^n$ then

$$\text{proj}_V\mathbf{w} = \left(\frac{\mathbf{w} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 + \dots + \left(\frac{\mathbf{w} \cdot \mathbf{v}_k}{\mathbf{v}_k \cdot \mathbf{v}_k}\right)\mathbf{v}_k$$

5) If $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an arbitrary basis of V then we can use the Gram-Schmidt process to obtain an orthogonal basis of V.

How to compute least square solutions of $A\mathbf{x} = \mathbf{b}$

(version 1.0)

1) Compute a basis of $\text{Col}(A)$.

2) Use the Gram-Schmidt process to get an orthogonal basis of $\text{Col}(A)$.

3) Use the orthogonal basis to compute $\text{proj}_{\text{Col}(A)}\mathbf{b}$.

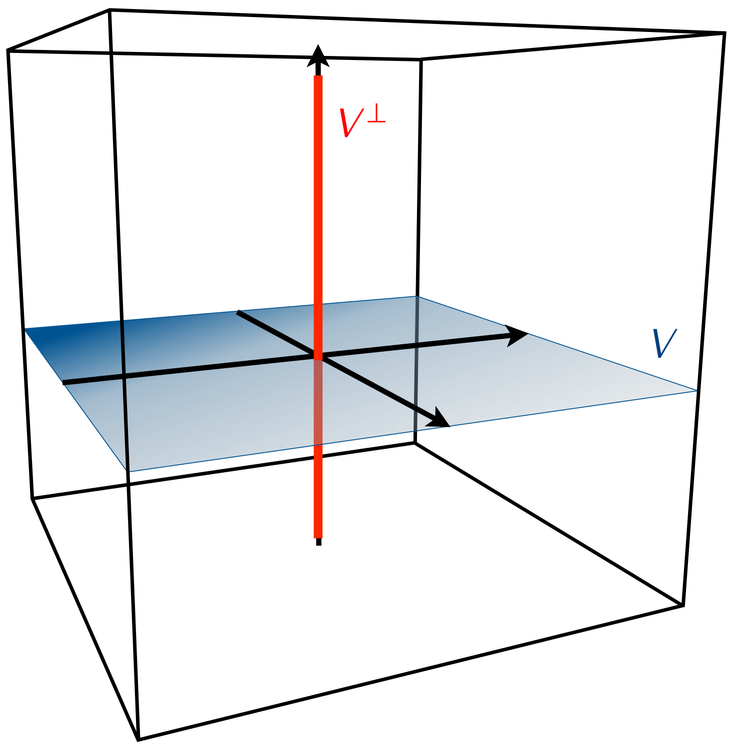
4) Compute solutions of the equation $A\mathbf{x} = \text{proj}_{\text{Col}(A)}\mathbf{b}$.

Next goal: Simplify this.

Definition

If V is a subspace of $\mathbb{R}^n$ then the *orthogonal complement* of V is the set $V^\perp$ of all vectors orthogonal to V:

$$V^\perp = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \in V\}$$



Proposition

If V is a subspace of $\mathbb{R}^n$ then:

1) $V^\perp$ is also a subspace of $\mathbb{R}^n$.
2) For each vector $\mathbf{w} \in \mathbb{R}^n$ there exist unique vectors $\mathbf{v} \in V$ and $\mathbf{z} \in V^\perp$ such that $\mathbf{w} = \mathbf{v} + \mathbf{z}$.

Definition

If A is an $m \times n$ matrix then the *row space* of A is the subspace $\text{Row}(A)$ of $\mathbb{R}^n$ spanned by rows of A.

Example

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

Proposition

If A is a matrix then

$$\text{Row}(A)^\perp = \text{Nul}(A)$$

Corollary

If A is a matrix then

$$\mathrm{Col}(A)^{\perp} = \mathrm{Nul}(A^T)$$

Back to least square solutions

Theorem

A vector $\hat{\mathbf{x}}$ is a least square solution of a matrix equation

$$A\mathbf{x} = \mathbf{b}$$

if and only if $\hat{\mathbf{x}}$ is an ordinary solution of the equation

$$(A^T A)\mathbf{x} = A^T\mathbf{b}$$

Definition

The equation

$$(A^T A)\mathbf{x} = A^T\mathbf{b}$$

is called the *normal equation* of $A\mathbf{x} = \mathbf{b}$.

How to compute least square solutions of $A\mathbf{x} = \mathbf{b}$

(version 2.0)

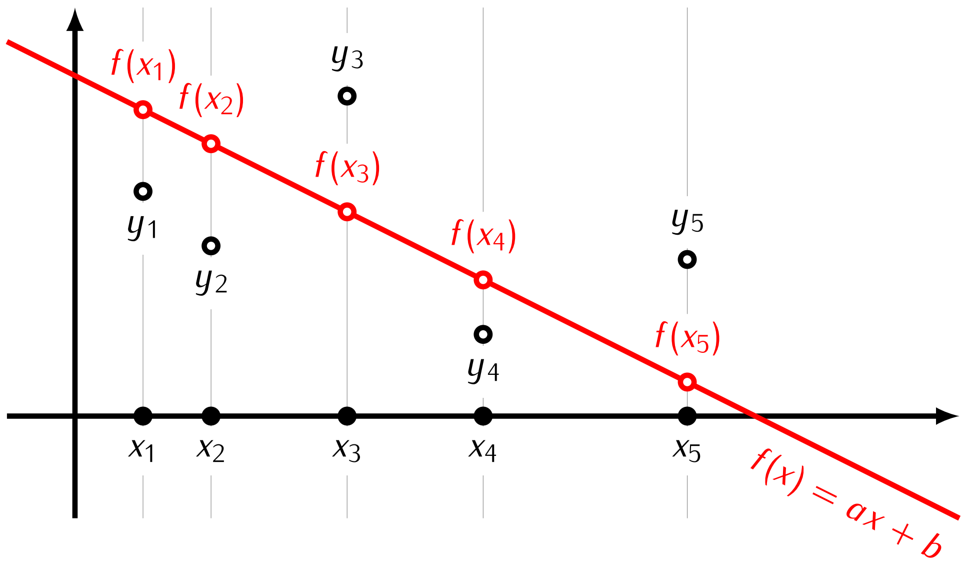
1) Compute $A^T A$, $A^T\mathbf{b}$.

2) Solve the normal equation $(A^T A)\mathbf{x} = A^T\mathbf{b}$.

Example. Compute least square solutions of the following equation:

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Application: Least square lines



Definition

If $(x_1, y_1), \ldots, (x_p, y_p)$ are points on the plane then the *least square line* for these points is the line given by an equation $f(x) = ax + b$ such that the number

$$\operatorname{dist}\left(\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_p) \end{bmatrix}\right) = \sqrt{(y_1 - f(x_1))^2 + \ldots + (y_p - f(x_p))^2}$$

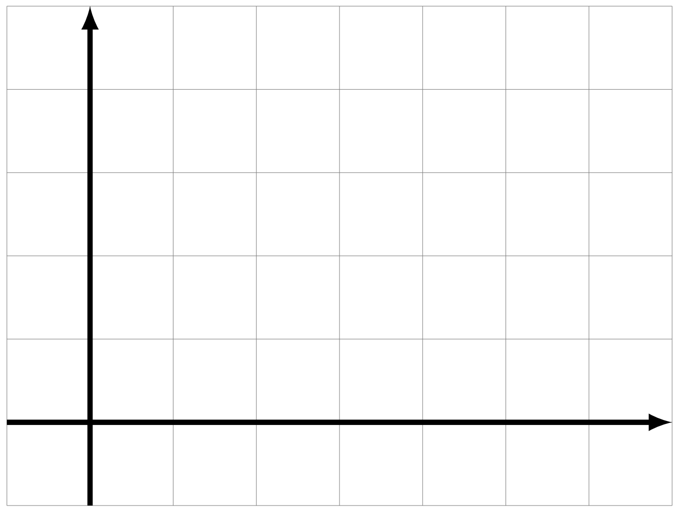
is the smallest possible.

Proposition

The line $f(x) = ax + b$ is the least square line for points $(x_1, y_1), \ldots, (x_p, y_p)$ if the vector $\begin{bmatrix} a \\ b \end{bmatrix}$ is the least square solution of the equation

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_p & 1 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

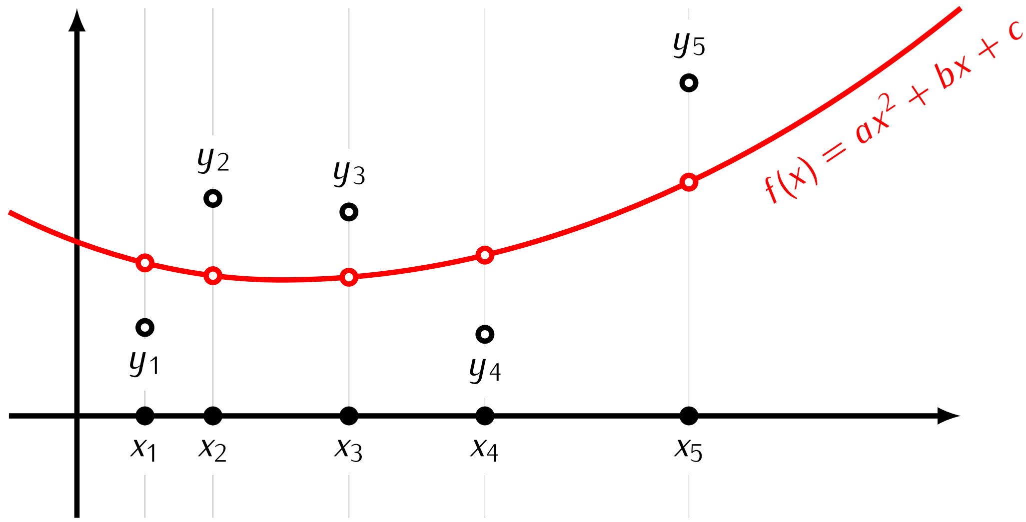
Example. Find the equation of the least square line for the points $(0, 0)$, $(1, 1)$, $(3, 1)$, $(5, 3)$.



Application: Least square curves

The above procedure can be used to determine curves other than lines that fit a set of points in the least square sense.

Example: Least square parabolas



Definition

If $(x_1, y_1), \ldots, (x_p, y_p)$ are points on the plane then the *least square parabola* for these points is the parabola given by an equation $f(x) = ax^2 + bx + c$ such that the number

$$\operatorname{dist}\left(\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_p) \end{bmatrix}\right) = \sqrt{(y_1 - f(x_1))^2 + \ldots + (y_p - f(x_p))^2}$$

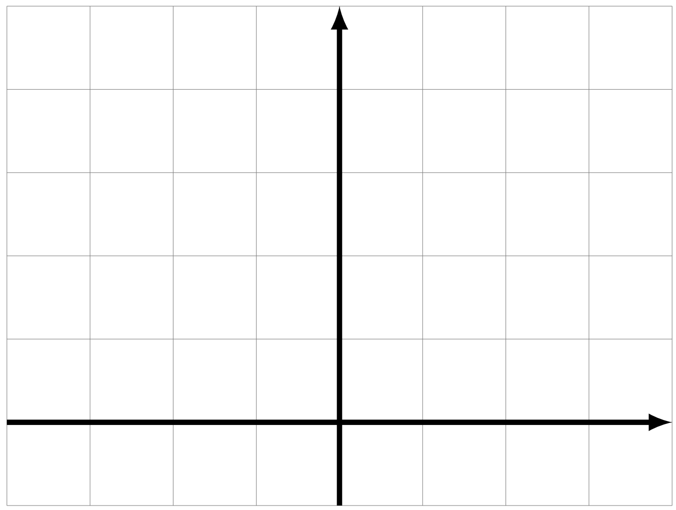
is the smallest possible.

Proposition

The parabola $f(x) = ax^2 + bx + c$ is the least square parabola for points $(x_1, y_1), \ldots, (x_p, y_p)$ if the vector $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is the least square solution of the equation

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \\ x_p^2 & x_p & 1 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

Example. Find the equation of the least square parabola for the points $(-2, 2)$, $(0, 0)$, $(1, 1)$, $(2, 3)$.



Recall:

1) The dot product in $\mathbb{R}^n$:

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1b_1 + a_2b_2 + \ldots a_nb_n$$

2) Properties of the dot product:

a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$

d) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

2) Using the dot product we can define:

- length of vectors
- distance between vectors
- orthogonality of vectors
- orthogonal and orthonormal bases
- orthogonal projection of a vector onto a subspace of $\mathbb{R}^n$
- ...

Next: Generalization to arbitrary vector spaces.

Definition

Let V be a vector space. An *inner product* on V is a function

$$\begin{aligned} V \times V &\longrightarrow \mathbb{R} \\ \mathbf{u}, \ \mathbf{v} &\longmapsto \langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

such that:

a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

b) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

c) $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$

d) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

Definition

Let V be a vector space with an inner product $\langle \ , \ \rangle$.

1) The *length* (or *norm*) of a vector $\mathbf{v}$ is the number

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

2) The *distance* between vectors $\mathbf{u}, \mathbf{v} \in V$ is the number

$$\text{dist}(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$$

3) Vectors $\mathbf{u}, \mathbf{v} \in V$ are *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Example. The dot product is an inner product in $\mathbb{R}^n$.

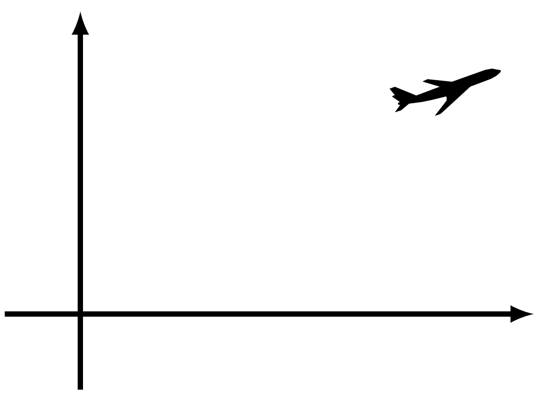
Example. Let $p_1, \ldots, p_n$ be any positive numbers. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$

$$\mathbf{u} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

define:

$$\langle \mathbf{u}, \mathbf{v} \rangle = p_1(a_1 b_1) + p_2(a_2, b_2) + \ldots + p_n(a_n b_n)$$

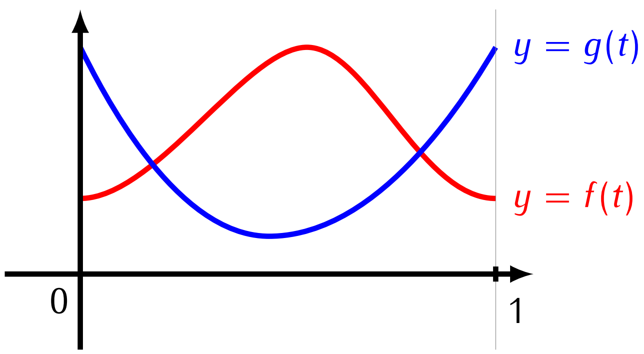
This gives an inner product in $\mathbb{R}^n$.



Example. Let $C[0,1]$ be the vector space of continuous functions $f\colon [0,1] \to \mathbb{R}$. Define:

$$\langle f, g\rangle = \int_0^1 f(t)g(t)dt$$

This is an inner product on $C[0,1]$.



Example. Compute the length of the function

$$f(t) = 1 + t^2$$

in $C[0,1]$.

Definition

Let V be a vector space with an inner product $\langle\ ,\ \rangle$, and let W be a subspace of V. A vector $\mathbf{v} \in V$ is *orthogonal to* W if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} \in W$.

Definition

Let V be a vector space with an inner product $\langle\ ,\ \rangle$, and let W be a subspace of V. The *orthogonal projection of a vector* $\mathbf{v} \in V$ *onto* W is a vector $\text{proj}_W\mathbf{v}$ such that

1) $\text{proj}_W\mathbf{v} \in W$
2) the vector $\mathbf{z} = \mathbf{v} - \text{proj}_W\mathbf{v}$ is orthogonal to W.

Best Approximation Theorem

If V is a vector space with an inner product $\langle\ ,\ \rangle$, W is a subspace of V, and $\mathbf{v} \in V$, then $\text{proj}_W\mathbf{v}$ is the vector of V which is the closest to $\mathbf{v}$:

$$\text{dist}(\mathbf{v}, \text{proj}_W\mathbf{v}) \leq \text{dist}(\mathbf{v}, \mathbf{w})$$

for all $\mathbf{w} \in W$.

Theorem

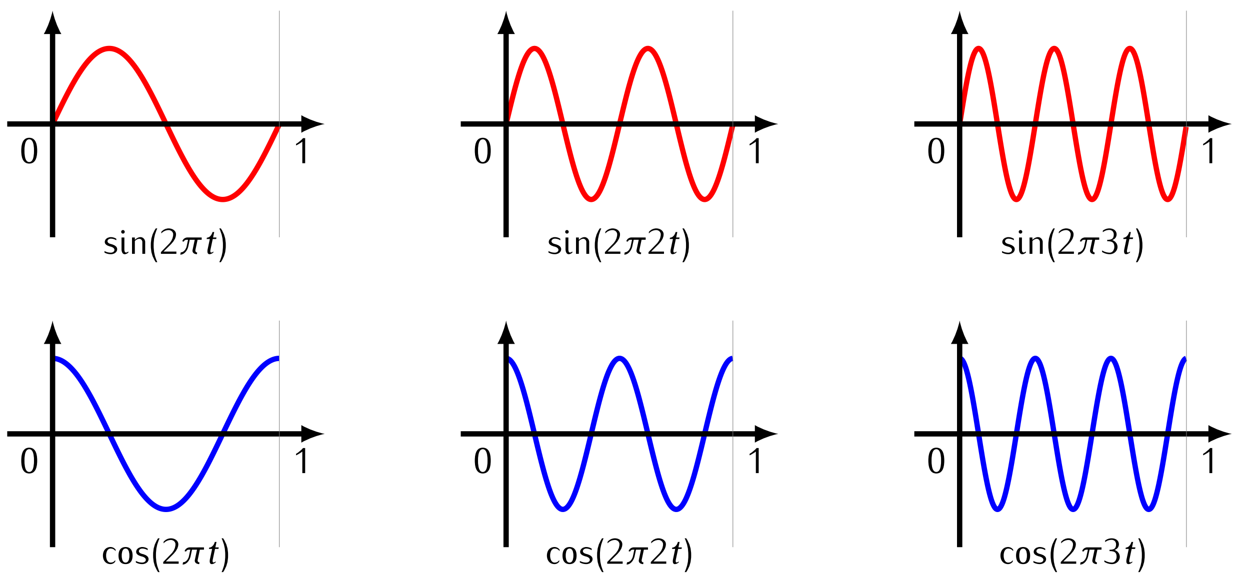
Let V is a vector space with an inner product $\langle\ ,\ \rangle$, and let W be a subspace of V. If $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ is an orthogonal basis of W (i.e. a basis such that $\langle \mathbf{w}_i, \mathbf{w}_j \rangle = 0$ for all $i \neq j$) then for $\mathbf{v} \in V$ we have:

$$\text{proj}_W\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \dots + \frac{\langle \mathbf{v}, \mathbf{w}_k \rangle}{\langle \mathbf{w}_k, \mathbf{w}_k \rangle}\mathbf{w}_k$$

Application: Fourier approximations.

Goal: Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function. Find the best possible approximation of f of the form

$$\begin{aligned} P(t) = \ & a_0 \\ & + a_1 \sin(2\pi t) \ + \ b_1 \cos(2\pi t) \\ & + a_2 \sin(2\pi 2t) \ + \ b_2 \cos(2\pi 2t) \\ & \cdots \ \cdots \ \cdots \ \cdots \ \cdots \ \cdots \ \cdots \\ & + a_n \sin(2\pi n t) \ + \ b_n \cos(2\pi n t) \end{aligned}$$



Note: Let W_n be a subspace of $C[0,1]$ given by:

$$W_n = \text{Span}(1, \sin(2\pi t), \cos(2\pi t), \ldots, \sin(2\pi n t), \cos(2\pi n t))$$

By the Best Approximation Theorem, the best approximation of f is obtained if we take $P(t) = \text{proj}_{W_n} f(t)$.

Theorem

The set

$$\{1, \sin(2\pi t), \cos(2\pi t), \ldots, \sin(2\pi n t), \cos(2\pi n t)\}$$

is an orthogonal basis of W_n.

Corollary

If $f \in C[0,1]$ then

$$\begin{aligned} \mathrm{proj}_{W_n} f(t) = \quad & a_0 \\ & + a_1 \sin(2\pi t) \quad + \quad b_1 \cos(2\pi t) \\ & + a_2 \sin(2\pi 2t) \;+\; b_2 \cos(2\pi 2t) \\ & \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ & + a_n \sin(2\pi n t) \;+\; b_n \cos(2\pi n t) \end{aligned}$$

where:

$$a_0 = \frac{\langle f, 1\rangle}{\langle 1, 1\rangle} = \int_0^1 f(t)dt$$

and for $k > 0$:

$$a_k = \frac{\langle f, \sin(2\pi k t)\rangle}{\langle \sin(2\pi k t), \sin(2\pi k t)\rangle} = 2\int_0^1 f(t) \cdot \sin(2\pi k t)dt$$

$$b_k = \frac{\langle f, \cos(2\pi k t)\rangle}{\langle \cos(2\pi k t), \cos(2\pi k t)\rangle} = 2\int_0^1 f(t) \cdot \cos(2\pi k t)dt$$

Example. Compute $\text{proj}_{W_n} f(t)$ for the function $f(t) = t$.

Application: Polynomial approximations.

Goal: Let $f\colon [0,1] \to \mathbb{R}$ be a continuous function. Find the best possible approximation of f given by a polynomial $P(t)$ of degree $\leq n$:

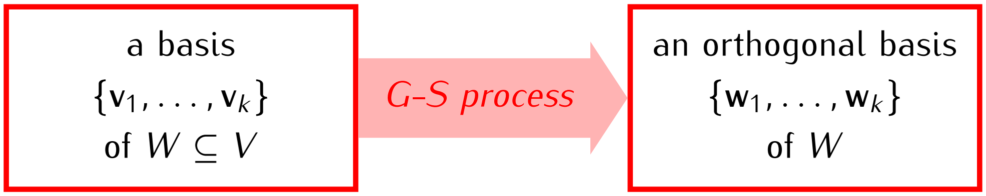
$$P(t) = a_0 + a_1 t + \ldots + a_n t^n$$

Note: Let $\mathbb{P}_n$ be the subspace of $C[0,1]$ consisting of all polynomials of degree $\leq n$:

$$\mathbb{P}_n = \{a_0 + a_1 t + \ldots + a_n t^n \mid a_k \in \mathbb{R}\}$$

By the Best Approximation Theorem, the best approximation of f is obtained if we take $P(t) = \mathrm{proj}_{\mathbb{P}_n} f(t)$.

Gram-Schmidt process:



Theorem (Gram-Schmidt Process)

Let V be a vector space with an inner product $\langle\ ,\ \rangle$, and let W be a subspace of V. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of W. Define vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ as follows:

$$\mathbf{w}_1 = \mathbf{v}_1$$

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{w}_1, \mathbf{v}_2 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1$$

$$\mathbf{w}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{w}_1, \mathbf{v}_3 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{w}_2, \mathbf{v}_3 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\mathbf{w}_k = \mathbf{v}_k - \frac{\langle \mathbf{w}_1, \mathbf{v}_k \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{w}_2, \mathbf{v}_k \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \ldots - \frac{\langle \mathbf{w}_{k-1}, \mathbf{v}_k \rangle}{\langle \mathbf{w}_{k-1}, \mathbf{w}_{k-1} \rangle}\mathbf{w}_{k-1}$$

Then the set $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is an orthogonal basis of W.

Example. Find an orthogonal basis of the subspace $\mathbb{P}_2$ of the vector space $C[0, 1]$.

Example. Compute $\text{proj}_{\mathbb{P}_2} f(t)$ for $f(t) = \sqrt{t}$.

Recall: An $n \times n$ matrix A defines a linear transformation

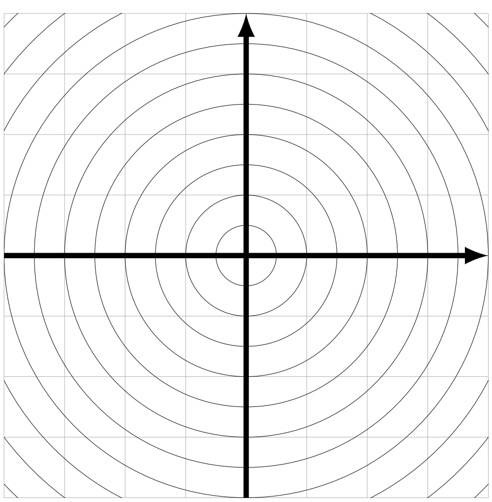
$$T_A : \mathbb{R}^n \to \mathbb{R}^n$$

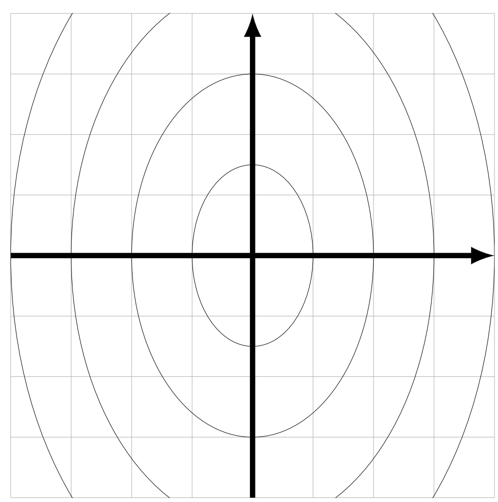
given by $T_A(\mathbf{v}) = A\mathbf{v}$.

Next goal: Understand this linear transformation better.

Example.

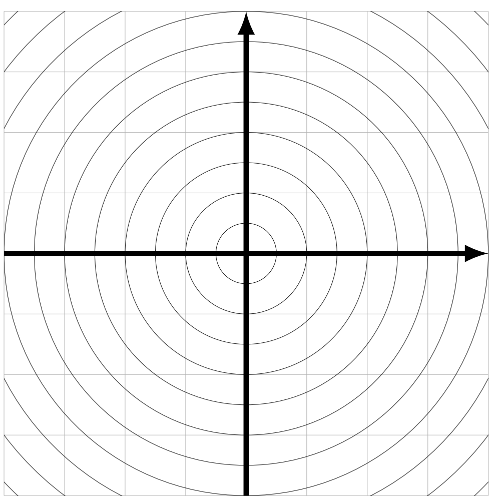
$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$

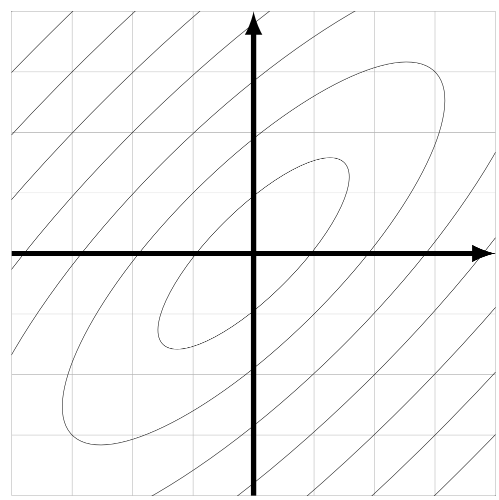




Example.

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





Definition

Let A be an $n \times n$ matrix. If $\mathbf{v} \in \mathbb{R}^n$ is a non-zero vector and λ is a scalar such that

$$A\mathbf{v} = \lambda\mathbf{v}$$

then we say that

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

Example.

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

Example.

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

Computation of eigenvalues

Recall: $I_n = n \times n$ identity matrix:

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Propostiton

If A be an $n \times n$ matrix then $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if the matrix equation

$$(A - \lambda I_n)\mathbf{x} = \mathbf{0}$$

has a non-trivial solution.

Propostiton

If B is an $n \times n$ matrix then equation

$$B\mathbf{x} = \mathbf{0}$$

has a non-trivial solution if and only of the matrix B is not invertible.

Propostiton

If A be an $n \times n$ matrix then $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if

$$\det(A - \lambda I_n) = 0$$

Example. Find all eigenvalues of the following matrix:

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

Definition

If A is an $n \times n$ matrix then

$$P(\lambda) = \det(A - \lambda I_n)$$

is a polynomial of degree n. $P(\lambda)$ is called the *characteristic polynomial* of the matrix A.

Upshot

If A is a square matrix then

$$\text{eigenvalues of } A = \text{roots of } P(\lambda)$$

Example.

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

Corollary

An $n \times n$ matrix can have at most n distinct eigenvalues.

Computation of eigenvectors

Proposition

If λ is an eigenvalue of an $n \times n$ matrix A then

$$\left\{\begin{matrix}\text{eigenvectors of } A\\ \text{corresponding to } \lambda\end{matrix}\right\} = \left\{\begin{matrix}\text{vectors in}\\ \text{Nul}(A - \lambda I_n)\end{matrix}\right\}$$

Corollary/Definition

If A is an $n \times n$ matrix and λ is an eigenvalue of A then the set of all eigenvectors corresponding to λ is a subspace of $\mathbb{R}^n$.

This subspace is called the *eigenspace* of A corresponding to λ.

Proposition

If λ is an eigenvalue of an $n \times n$ matrix A then

$$\left\{ \begin{matrix} \text{eigenspace of } A \\ \text{corresponding to } \lambda \end{matrix} \right\} = \mathrm{Nul}(A - \lambda I_n)$$

Example. Consider the following matrix:

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

Recall that eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 5$. Compute bases of eigenspaces of A corresponding to these eigenvalues.

Solution.

$\underline{\lambda_1 = 1}$

$\underline{\lambda_2 = 5}$

Recall:

1) Let A be an $n \times n$ matrix. If $\mathbf{v} \in \mathbb{R}^n$ is a non-zero vector and λ is a scalar such that

$$A\mathbf{v} = \lambda\mathbf{v}$$

then

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

2) The characteristic polynomial of an $n \times n$ matrix A is the polynomial given by the formula

$$P(\lambda) = \det(A - \lambda I_n)$$

where I_n is the $n \times n$ identity matrix.

3) If A is a square matrix then

$$\text{eigenvalues of } A = \text{roots of } P(\lambda)$$

4) If λ is an eigenvalue of an $n \times n$ matrix A then

$$\left\{\begin{matrix}\text{eigenvectors of } A \\ \text{corresponding to } \lambda\end{matrix}\right\} = \left\{\begin{matrix}\text{vectors in} \\ \text{Nul}(A - \lambda I_n)\end{matrix}\right\}$$

Motivating example: Fibonacci numbers

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ 89,\ 144,\ 233,\ 377,\ \ldots$$

Problem. Find a formula for the n-th Fibonacci number F_n.

General Problem. If A is a square matrix how to compute A^k quickly?

Easy case:

Definition

A square matrix D is *diagonal matrix* if all its entries outside the main diagonal are zeros:

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Proposition

If D is a diagonal matrix as above then

$$D^k = \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix}$$

Definition

A square matrix A is a *diagonalizable* if A is of the form

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

where D is a diagonal matrix and P is an invertible matrix.

Example.

$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ is a diagonalizable matrix:

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

Proposition

If A is a diagonalizable matrix, $A = PDP^{-1}$, then

$$A^k = PD^kP^{-1}$$

Example.

Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. Compute A^{10}.

Diagonalization Theorem

1) An $n \times n$ matrix A is a diagonalizable if and only if it has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

2) In such case $A = PDP^{-1}$ where :

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}$$

$$D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

λ_1 = eigenvalue corresponding to $\mathbf{v}_1$
λ_2 = eigenvalue corresponding to $\mathbf{v}_2$
... ...
λ_n = eigenvalue corresponding to $\mathbf{v}_n$

Example. Diagonalize the following matrix if possible:

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

Note. Not every matrix is diagonalizable.

Example. Check if the following matrix is diagonalizable:

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

Proposition

If A is an $n \times n$ matrix with n distinct eigenvalues then A is diagonalizable.

Back to Fibonacci numbers:

$$\begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{n-1} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Recall:

1) A square matrix A is diagonalizable if there exists an invertible matrix P and a diagonal matrix D such that

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

2) If A is diagonalizable then it is easy to compute powers of A:

$$A^k = PD^kP^{-1}$$

3) An $n \times n$ matrix A is a diagonalizable if and only if it has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. In such case we have:

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}$$

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

λ_1 = eigenvalue corresponding to $\mathbf{v}_1$
λ_2 = eigenvalue corresponding to $\mathbf{v}_2$
... ...
λ_n = eigenvalue corresponding to $\mathbf{v}_n$

4) Not every square matrix is diagonalizable.

Definition

A square matrix A is *symmetric* if $A^T = A$

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

Theorem

Every symmetric matrix is diagonalizable.

Theorem

If A is a symmetric matrix and λ_1, λ_2 are two different eigenvalues of A, then eigenvectors corresponding to λ_1 are orthogonal to eigenvectors corresponding to λ_2.

Note. If $\mathbf{v}$, $\mathbf{w}$ are vectors in $\mathbb{R}^n$ then

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w}$$

Example.

$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$

Theorem

If A is an $n \times n$ symmetric matrix then A has n orthogonal eigenvectors.

Example.
a) Find three orthogonal eigenvectors of the following symmetric matrix:

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

b) Use these eigenvectors to diagonalize this matrix.

Upshot. How to find n orthogonal eigenvectors for a symmetric $n \times n$ matrix A:

1) Find eigenvalues of A.

2) Find a basis of the eigenspace for each eigenvalue.

3) Use the Gram-Schmidt process to find an orthogonal basis of each eigenspace.

Definition

A square matrix $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \end{bmatrix}$ is an *orthogonal matrix* if $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is an orthonormal set of vectors, i.e.:

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

Theorem

If Q is an orthogonal matrix then Q is invertible and $Q^{-1} = Q^T$.

Note. If $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}$ is a matrix with orthogonal columns, then

$$Q = \begin{bmatrix} \frac{\mathbf{v}_1}{||\mathbf{v}_1||} & \frac{\mathbf{v}_2}{||\mathbf{v}_2||} & \dots & \frac{\mathbf{v}_n}{||\mathbf{v}_n||} \end{bmatrix}$$

is an orthogonal matrix.

Theorem

If A is a symmetric matrix then A is *orthogonally diagonalizable*. That is, there exists an orthogonal matrix Q and a diagonal matrix D such that

$$A = QDQ^{-1} = QDQ^T$$

Example. Find an orthogonal diagonalization of the following symmetric matrix:

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

Note. We have seen that any symmetric matrix is orthogonally diagonalizable. The converse statement is also true:

Proposition

If a matrix A is orthogonally diagonalizable then A is a symmetric matrix.

Recall:

1) An orthogonal matrix $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \end{bmatrix}$ is a square matrix such that $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is an orthonormal set of vectors, i.e.:

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

2) If Q is an orthogonal matrix then $Q^{-1} = Q^T$

3) A square matrix A is orthogonally diagonalizable if there exist an orthogonal matrix Q and a diagonal matrix D such that

$$A = QDQ^{-1} = QDQ^T$$

4) A matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix (i.e. $A^T = A$).

Yet another view of matrix multiplication

Note. If C is an $n \times 1$ matrix and D is an $1 \times n$ matrix then CD is an $n \times n$ matrix.

Propostion

Let A be an $n \times n$ matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and B be an $n \times n$ matrix with rows $\mathbf{w}_1, \ldots, \mathbf{w}_n$:

$$A = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{bmatrix} \qquad B = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_n \end{bmatrix}$$

Then

$$AB = \mathbf{v}_1\mathbf{w}_1 + \mathbf{v}_2\mathbf{w}_2 + \ldots + \mathbf{v}_n\mathbf{w}_n$$

Example.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 1 \\ 7 & 2 \end{bmatrix}$$

Theorem

Let A be a symmetric matrix with orthogonal diagonalization

$$A = QDQ^T$$

If

$$Q = \left[\begin{array}{ccc} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{array} \right] \quad \text{and} \quad D = \left[\begin{array}{ccc} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & 0 \\ 0 & \ldots & \lambda_n \end{array} \right]$$

then

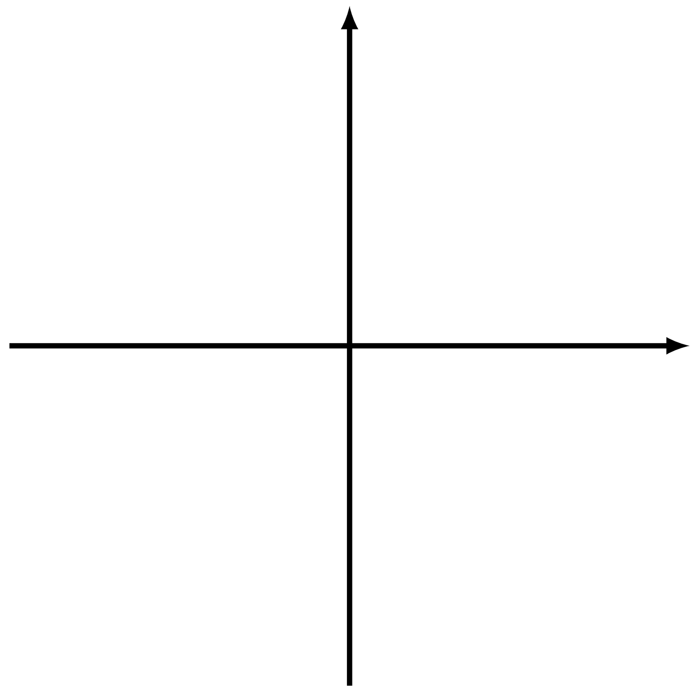
$$A = \lambda_1(\mathbf{u}_1\mathbf{u}_1^T) + \lambda_2(\mathbf{u}_2\mathbf{u}_2^T) + \ldots + \lambda_n(\mathbf{u}_n\mathbf{u}_n^T)$$

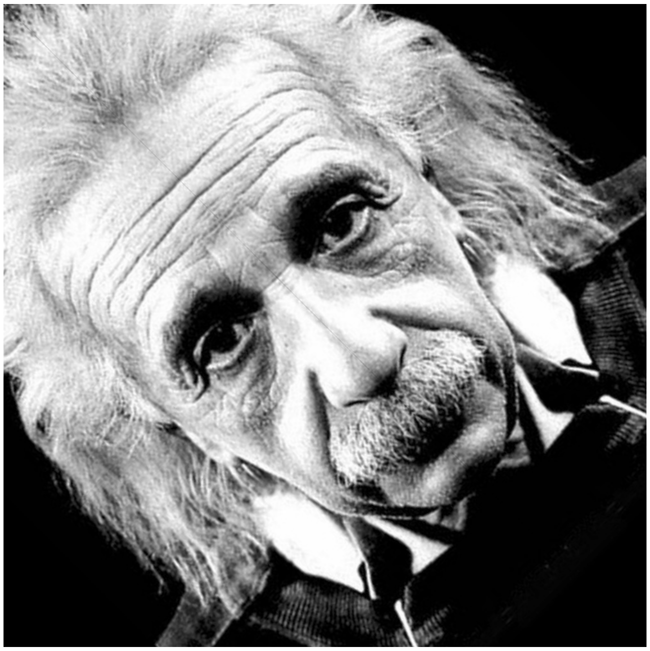
Note. The above formula is called the *spectral decomposition* of the matrix A.

Example.

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

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





- The size of this image is 1000×1000 pixels.
- The color of each pixel is represented by an integer between 0 (black) and 255 (white).
- The whole image is described by a (symmetric) matrix A consisting of $1000 \times 1000 = 1,000,000$ numbers
- Each number is stored in 1 byte, so the image file size is $1,000,000$ bytes (≈ 1 MB).

How to make the image file smaller:

1) Find the spectral decomposition of the matrix A:

$$A = \lambda_1(\mathbf{u}_1\mathbf{u}_1^T) + \lambda_2(\mathbf{u}_2\mathbf{u}_2^T) + \ldots + \lambda_{1000}(\mathbf{u}_{1000}\mathbf{u}_{1000}^T)$$

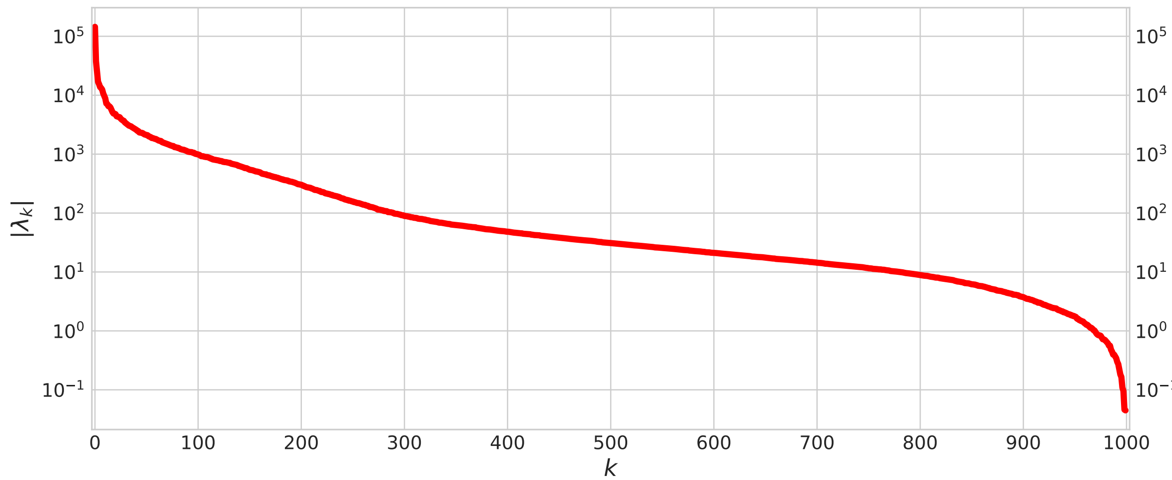
where $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_{1000}|$.

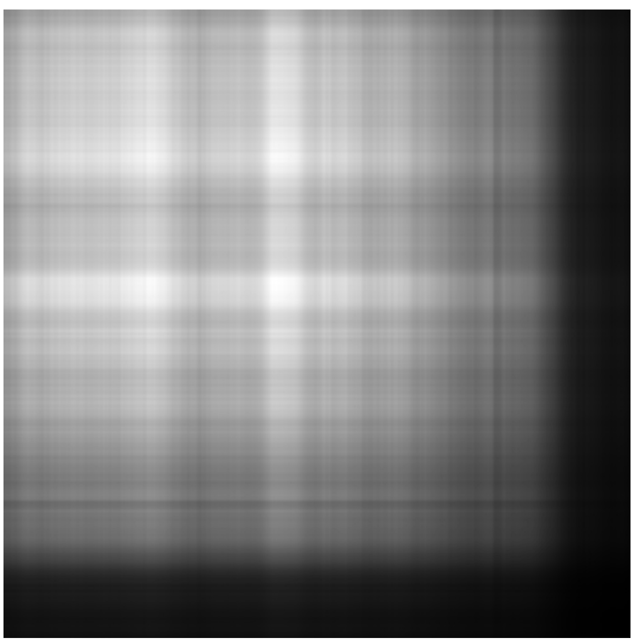
2) For $k = 1, \ldots, 1000$ define:

$$B_k = \lambda_1(\mathbf{u}_1\mathbf{u}_1^T) + \lambda_2(\mathbf{u}_2\mathbf{u}_2^T) + \ldots + \lambda_k(\mathbf{u}_k\mathbf{u}_k^T)$$

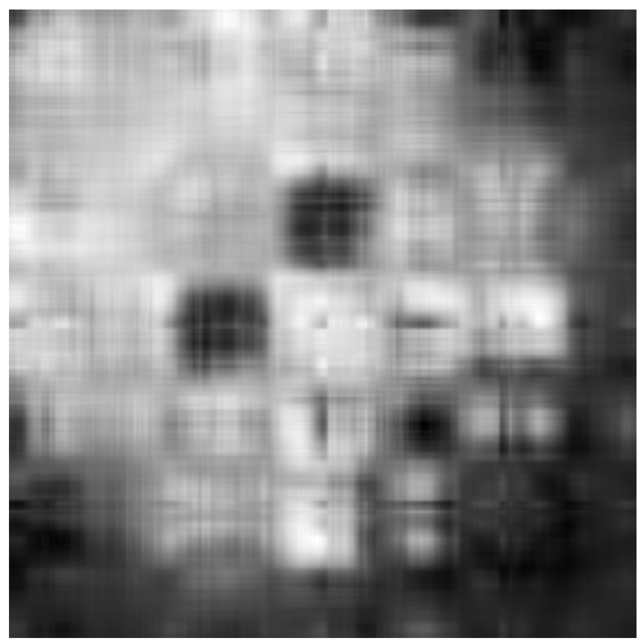
This matrix approximates the matrix A and can be stored using $k \cdot (1000 + 1)$ numbers (i.e. $k \cdot (1000 + 1)$ bytes).

Eigenvalues of the matrix A

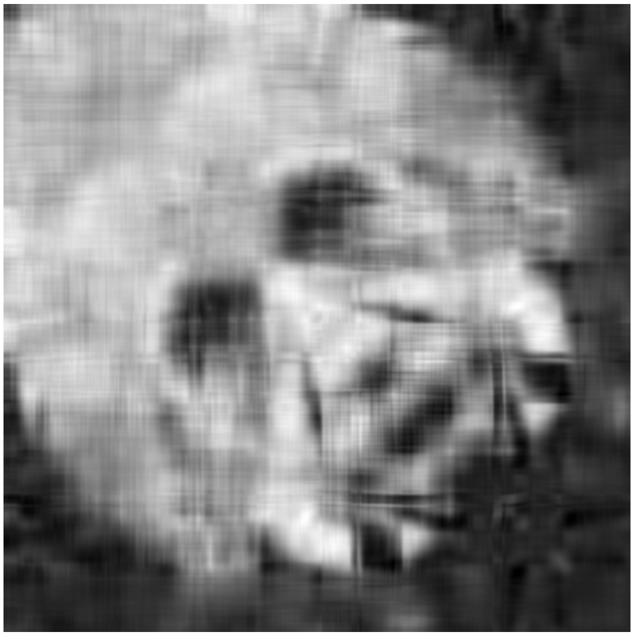




matrix B_1
1001 bytes
compression 1000:1

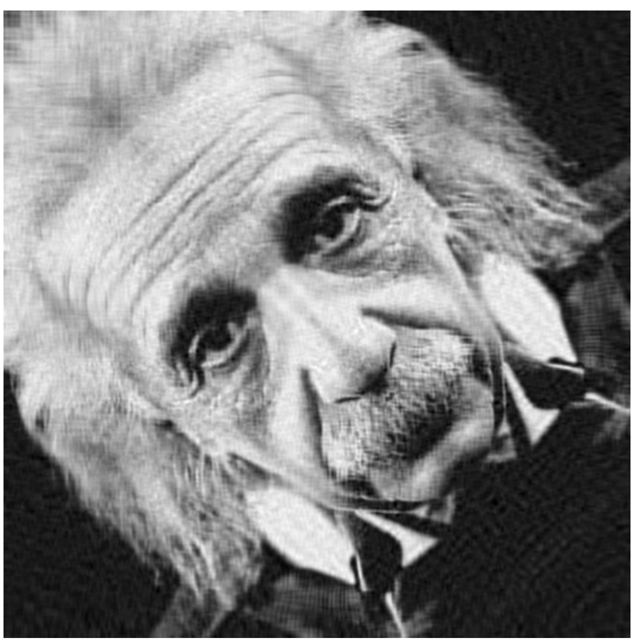


matrix B_5
5005 bytes
compression 200:1

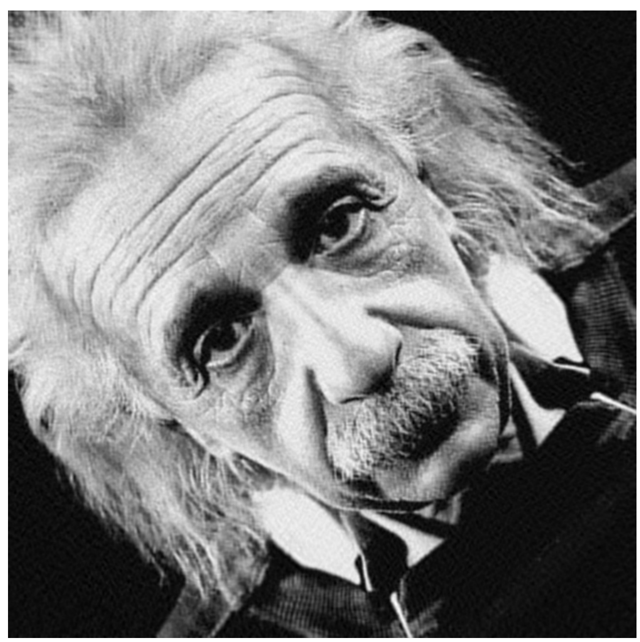


matrix B_{10}
10,010 bytes
compression 100:1

matrix B_{20}
20,020 bytes
compression 50:1



matrix B_{50}
50,050 bytes
compression 20:1



matrix B_{100}
100,100 bytes
compression 10:1

Theorem

Any A an $m \times n$ matrix can be written as a product

$$A = U\Sigma V^T$$

where:

- $U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix}$ is an $m \times m$ orthogonal matrix.
- $V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ is an $n \times n$ orthogonal matrix.
- Σ is an $m \times n$ matrix of the following form:

$$\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{bmatrix}$$

(if $n \leq m$) (if $n \geq m$)

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$.

Note.

- The numbers $\sigma_1, \sigma_2, \ldots$ are called *singular values* of A.
- The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are called *left singular vectors* of A.
- Then the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are called *right singular vectors* of A.
- The formula $A = U\Sigma V^T$ is called a *singular value decomposition (SVD)* of A.
- The matrix Σ is uniquely determined, but U and V depend on some choices.

Theorem

Let A be a matrix with a singular value decomposition

$$A = U\Sigma V^T$$

If

$$U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix} \qquad V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$$

and $\sigma_1, \dots, \sigma_r$ are singular values of A then then

$$A = \sigma_1(\mathbf{u}_1\mathbf{v}_1^T) + \sigma_2(\mathbf{u}_2\mathbf{v}_2^T) + \dots + \sigma_r(\mathbf{u}_r\mathbf{v}_r^T)$$

Application: **Image compression**

- The size of this image is 800×700 pixels.
- The color of each pixel is represented by an integer between 0 (black) and 255 (white).
- The whole image is described by a matrix A consisting of $800 \times 700 = 560,000$ numbers.
- Each number is stored in 1 byte, so the image file size is $560,000$ bytes (≈ 0.53 MB).

How to make the image file smaller:

1) Compute SVD of the matrix A:

$$A = U\Sigma V^T$$

where

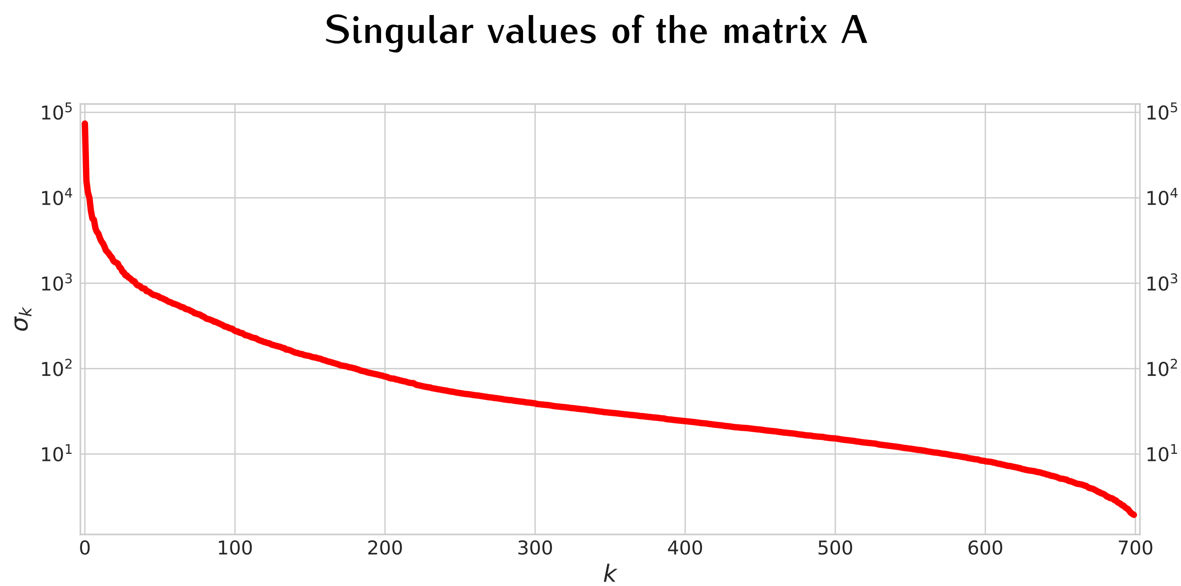
$$U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix} \qquad V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$$

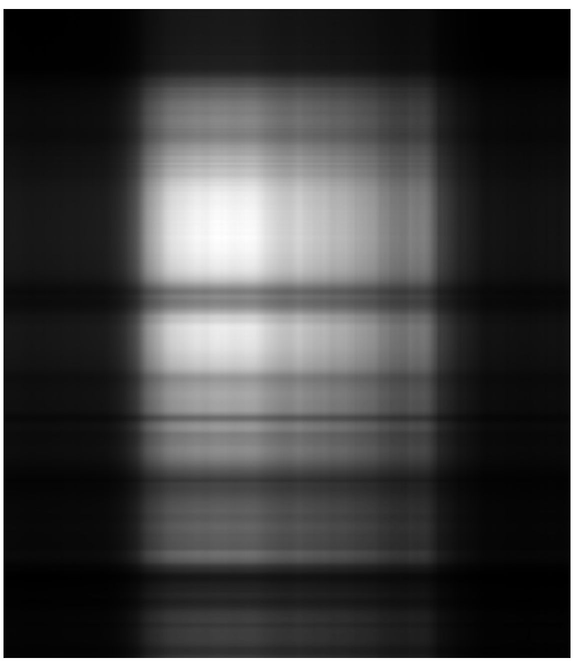
and $\sigma_1, \dots, \sigma_r$ are singular values of A.

2) Replace A by the matrix

$$B_k = \sigma_1(\mathbf{u}_1\mathbf{v}_1^T) + \ldots + \sigma_k(\mathbf{u}_k\mathbf{v}_k^T)$$

for some $1 \leq k \leq 700$. This matrix can be stored using $k \cdot (800 + 700 + 1)$ numbers.





matrix B_1
1501 bytes
compression 374:1

matrix B_5
7905 bytes
compression 75:1

matrix $\mathbf{B}_{10}$
15,010 bytes
compression 37:1

matrix $\mathbf{B}_{20}$
30,020 bytes
compression 18:1

matrix $\mathbf{B}_{50}$
75,050 bytes
compression 7:1

matrix $\mathbf{B}_{100}$
150,100 bytes
compression 4:1

How to compute SVD of a matrix A

How to compute SVD of a matrix A

1) Compute an orthogonal diagonalization of the symmetric $n \times n$ matrix $A^T A$:

$$A^T A = QDQ^T$$

such that eigenvalues on the diagonal of the matrix D are arranged from the largest to the smallest. We set $V = Q$.

2) If

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

then $\sigma_i = \sqrt{\lambda_i}$. This gives the matrix Σ.

Note: if $n > m$ then we use only $\lambda_1, \dots, \lambda_m$. The remaining eigenvalues $\lambda_{m+1}, \dots, \lambda_n$ of D will be equal to 0 in this case.

3) Let $V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$, and let $\sigma_1, \dots, \sigma_r$ be non-zero singular values of A. The first r columns of the matrix $U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix}$ are given by

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

The remaining columns $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ can be added arbitrarily so that U is an orthogonal matrix (i.e.$\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$) is an orthonormal basis of $\mathbb{R}^m$.

Example. Find SVD of the following matrix:

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

Recall:
Let A be a matrix with a singular value decomposition

$$A = U\Sigma V^T$$

If

$$U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix} \qquad V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$$

and $\sigma_1, \dots, \sigma_r$ are singular values of A then then

$$A = \sigma_1(\mathbf{u}_1\mathbf{v}_1^T) + \sigma_2(\mathbf{u}_2\mathbf{v}_2^T) + \dots + \sigma_r(\mathbf{u}_r\mathbf{v}_r^T)$$

Example: Movie ratings:

	Matrix	**Amelie**	**Alien**	**Casablanca**	**Interstellar**
user_1	5	0	5	0	4
user_2	5	0	3	0	5
user_3	0	5	0	5	1
user_4	1	5	0	4	0
user_5	4	0	4	0	3
user_6	0	5	0	4	0
user_7	3	0	3	0	2

Singular value decomposition of the matrix of movie ratings:

$$U = \begin{bmatrix} -0.6 & 0.1 & -0.3 & -0.2 & 0.2 & -0.7 & -0.2 \\ -0.5 & 0.1 & 0.8 & 0.2 & 0.1 & 0.1 & 0.1 \\ -0.1 & -0.6 & 0.2 & -0.7 & -0.4 & 0.0 & 0.0 \\ -0.1 & -0.5 & -0.1 & 0.7 & -0.4 & -0.1 & -0.2 \\ -0.5 & 0.1 & -0.3 & -0.1 & -0.1 & 0.7 & -0.4 \\ -0.1 & -0.6 & -0.1 & 0.0 & 0.8 & 0.1 & 0.2 \\ -0.3 & 0.1 & -0.3 & 0.0 & -0.3 & 0.1 & 0.8 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 13.6 & 0 & 0 & 0 & 0 \\ 0 & 11.4 & 0 & 0 & 0 \\ 0 & 0 & 1.9 & 0 & 0 \\ 0 & 0 & 0 & 1.1 & 0 \\ 0 & 0 & 0 & 0 & 0.3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.6 & 0.1 & 0.0 & 0.7 & -0.4 \\ -0.1 & -0.7 & -0.1 & 0.3 & 0.6 \\ -0.5 & 0.1 & -0.7 & -0.4 & 0.2 \\ -0.1 & -0.6 & 0.0 & -0.4 & -0.7 \\ -0.5 & 0.1 & 0.7 & -0.4 & 0.3 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 0 & 5 & 0 & 4 \\ 5 & 0 & 3 & 0 & 5 \\ 0 & 5 & 0 & 5 & 1 \\ 1 & 5 & 0 & 4 & 0 \\ 4 & 0 & 4 & 0 & 3 \\ 0 & 5 & 0 & 4 & 0 \\ 3 & 0 & 3 & 0 & 2 \end{bmatrix} \approx \begin{bmatrix} -0.6 & 0.1 \\ -0.5 & 0.1 \\ -0.1 & -0.6 \\ -0.1 & -0.5 \\ -0.5 & 0.1 \\ -0.1 & -0.6 \\ -0.3 & 0.1 \end{bmatrix} \cdot \begin{bmatrix} 13.6 & 0 \\ 0 & 11.4 \end{bmatrix} \cdot \begin{bmatrix} -0.6 & -0.1 & -0.5 & -0.1 & -0.5 \\ 0.1 & -0.7 & 0.1 & -0.6 & 0.1 \end{bmatrix}$$

Problem. A new movie “Captive State” was rated by the seven users as follows: 4, 4, 0, 1, 4, 0, 0. What kind of movie it is?