Course 1

Linear Algebra for Data Science

Annotation

In the lecture course, we consider some topics of linear algebra beyond the standard first year course which are extremely important for applications. Mostly, these are applications to data analysis and machine learning, as well as to economics and statistics. We begin with inversions of rectangle matrices, that is, we discuss pseudo-inverse matrices (and their connections to the linear regression model). Among others, we discuss iteration methods (and their using in models of random walk on a graph applied to Internet search such as PageRank algorithm), matrix decompositions (such as SVD) and methods of dimension decreasing (with their connection to some image compression algorithms), and the theory of matrix norms and perturbation theory (for error estimates in matrix computations). The course includes also symbolic methods in systems of algebraic equations, approximation problems, Chebyshev polynomials, functions with matrices such as exponents etc. We plan to invite some external lecturers who successfully apply linear algebra in their work. The students are also be invited to give their own talks on additional topics of applied or theoretical linear algebra.

Final Grade

$$\text{GRADE} = \frac{\text{test1} + \text{test2}}{2} + \underbrace{\begin{array}{c}\text{Bonus}\\ \text{for a talk}\end{array}}_{\leq 5} + \underbrace{\begin{array}{c}\text{Bonus}\\ \text{for classes}\end{array}}_{\leq 1\ldots 2}.$$

1. Difference between fundamental and applied linear algebra. Problems with the real data. Pseudoinverse matrices. Skeletonization.

Let's consider some of the standard linear algebra problems, for example, solving the systems of a linear equations. It can be written:

$$A\vec{x} = \vec{b},$$

where $A = M_{m\times n}(F)$ – the matrix of coefficients, $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in F^n$ – unknown vector and $\vec{b} \in F^m$ – known vector. Solving such systems is our goal. In the best situation we can write out the solution:

$$\vec{x} = A^{-1}\vec{b}.$$

Or we have an another method in a more general situation, when the matrix of the coefficients can be rectangular or degenerate, no inverted. We can provide a Gaussian elimination:

After that we can easily express one variable in terms of another one step by step.

But in the real work with the linear models the initial data can be inaccurate due to the observational errors in some physic cases or human reliability in, for example, social or business situations. It can lead to some problems. For example, in Gaussian elimination you need to choose pivot variable, and if it is contains some errors, then other computation will increase them. In situations with high order error such methods cannot be applied. But even if you have the exact formula and enough resources for calculating the inverted matrix you will release that inverse is obtained with some errors. Another problem is rounding. It happens because of precise nature of Gaussian elimination algorithm or other algorithms and machine precision. It is hard and slow to calculate precision of the solution.

Next problem is about speed or complexity of calculations. For example, the complexity for Gaussian elimination is $\mathcal{O}(n^3)$. It is bad for dealing with, for example, video or signals in real time.

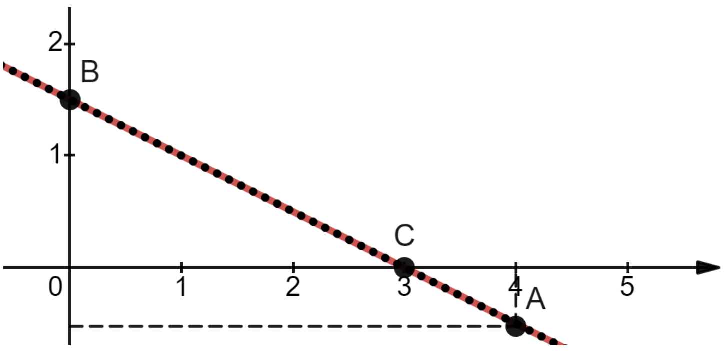
Suppose you have created a linear model, that more or less describe the production or business process. For example, you get the vector of unknowns $\vec{x}$, which means that you should sell x_1 copies of product 1 and x_2 copies of product 2, etc. How much you must produce or sell? It is a possible cases needed to solve. Then you can face with indefinite system because of your initial data is not exact.

Indefinite (indeterminate) system

For example, it means system with less equations than variables or linearly dependent equations. It can be some system of a kind:

$$\begin{cases} x_1 + 2x_2 = 3, \\ 2x_1 + 4x_2 = 5 \end{cases}$$

So we have two lines of the same kind. The problem is how to peek one value of the vector $\vec{x}$.



Example of indefinite system.

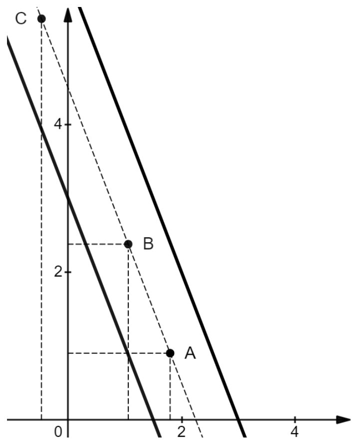
Mathematically we can peek every point from this line, but, for example, negative values (point A), obviously looks strange or even inappropriate for our case. Or, for example, you can peek either point B, or C. Again, mathematically there are solutions of our systems, but it can be no optimum solution. We will understand it on the next lecture. Another problem is about inconsistent system if there is no set of values for the unknown vector $\vec{x}$ that satisfies all of the equations. Or in mathematical terms, it happens when rank of initial coefficients matrix less than rank of augmented matrix (matrix of a kind $[A|b]$). Such systems can be either determined (defined) or underdetermined.

Inconsistent system

For example, you have approximately calculated several coefficients of the initial matrix and after the approximate calculations you obtained the system of equations, for example, a kind of:

$$\begin{cases} x_1 + 2x_2 = 3, \\ 2x_1 + 4x_2 = 5 \end{cases}$$

It looks like 2 parallel lines. Of course, rarely some $\vec{x}$ exists, but we cannot solve this system because of coefficients are not exact (contains some random errors, precisions, rounding after calculations, et cetera). But answers still can be the same. How much we need to sell? We can choose, for example, solutions A or B or even C, but what would be more optimally or correct. We will give an answer for this question right of the bat on the next lecture. Now let's define some new concepts that will help us with it.



Example of inconsistent system.

Pseudoinverse Matrices

Definition

Let $A \in M_{m\times n}(\mathbb{C})$. Then C is called pseudoinverse matrix to A, or Moore-Penrose (pseudo) inverse, if it is satisfies Penrose axioms:

I. $ACA = A$;
II. $CAC = C$;
III. $(AC)^* = AC$;
IV. $(CA)^* = CA$.

Example 1: if $A \in M_{n\times n}$, $\det A \neq 0$, then A^{-1} is a pseudoinverse.

Prop

If a pseudoinverse matrix C to A exists, it is unique.

Proof: Suppose B is some another pseudoinverse matrix to A. Then:

$$AB \overset{I}{=} (AC)(AB) \overset{III}{\Rightarrow} (AC)^*(AB)^* \Rightarrow C^*(A^*B^*A^*) = C^*(ABA)^* = (AC)^* = AC.$$

Similarly, $BA = CA$. Now, $B \overset{II}{=} BAB = BAC = CAC = C$. □

Note

Notation: $C = A^+$.

Example 2: $A = O_{m\times n} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}$. Then $A^+ = O_{n\times m}$.

Note

If $A \in M_{m\times n}(\mathbb{C})$, then $C \in M_{n\times m}(\mathbb{C})$.

Example 3: $\begin{bmatrix} 5 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Example 4: $\begin{bmatrix} 1 \\ 2 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \end{bmatrix}$

Example 5: Let $A = \vec{a} \in \mathbb{C}^n$. Then:

$$\vec{a}^+ = \frac{1}{a^* a}\vec{a}^* = \frac{1}{|\vec{a}|^2} = \frac{1}{|a_1|^2 + \dots + |a_n|^2}\vec{a}^*.$$

Prop

Suppose that $A \in M_{m\times n}(\mathbb{C})$ has full column rank, that is, $\operatorname{rank} A = n$. Then: $A^+ = (\underbrace{A^* A}_{n\times n})^{-1} A^*$.

Prop

If $\operatorname{rank} A = m$ (A has full row rank), then: $A^+ = A^*(\underbrace{AA^*}_{m\times m})^{-1}$.

Exercise: Check I-IV axioms for these A^+ **!!!**

Definition: Skeletonization

A full rank decomposition (or skeletonization) of a matrix $A \in M_{m\times n}(\mathbb{C})$ with $r = \operatorname{rank} A$ is a decomposition:

$$A = F \cdot G, \quad \begin{aligned} F &\in M_{m\times r}(\mathbb{C}), \\ G &\in M_{r\times n}(\mathbb{C}). \end{aligned}$$

(Then $\operatorname{rank} F = \operatorname{rank} G = r$. F, G are called matrices of full rank.)

Theorem

For each matrix $A \in M_{m\times n}(\mathbb{C})$, its pseudoinverse matrix A^+ exists. If $A = F \cdot G$ is a full rank decomposition, then:

$$A^+ = G^+ F^+ = G^* \left(G, G^*\right)^{-1} \left(F^*, F\right)^{-1} F^*.$$

2. Pseudosolutions and its applications. Linear regression.

Let's repeat main possible situation for solving linear equations task, which one can be written by the next notation:

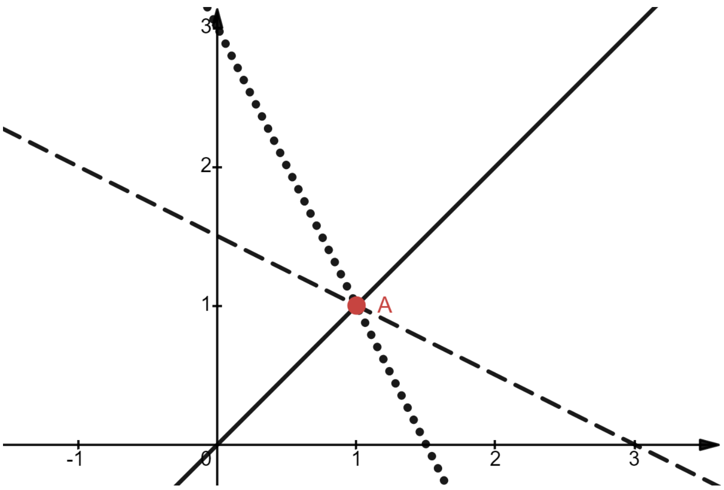
$$A\vec{x} = b,$$

where $A \in M_{m\times n}(\mathbb{C})$, $\vec{b} \in \mathbb{C}^m$, $\vec{x} \in \mathbb{C}^n$.

0. The first case is about square matrix $A \in M_{n\times n}(\mathbb{C})$, $\operatorname{rank} A = n$. In such situation we can easily obtain unique $\overline{x}$ by inverting the matrix of initial coefficients:

$$\overline{x} = A^{-1}\vec{b}.$$

1. The next easy option is a definite system, when $A \in M_{m\times n}(\mathbb{C}), \operatorname{rank} A = n$. Then unique $\hat{x}$ can be expressed by the following ideas.



Example of definite system.

Consider a system of the form:

$$\begin{cases} 2x + y = 3, \\ x + 2y = 3, \\ x - y = 0. \end{cases}$$

It is obviously that system have only one correct solution in the point A and it is a solution of a type: $\hat{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. However, we would like to generalize the method of obtaining a solution in such a way that it looks similar to the first (zero) case, namely:

$$\hat{x} = ? \cdot \vec{b}.$$

And looking ahead we can obtain such a factor to express solution that way. But now let's get a broader generalization.

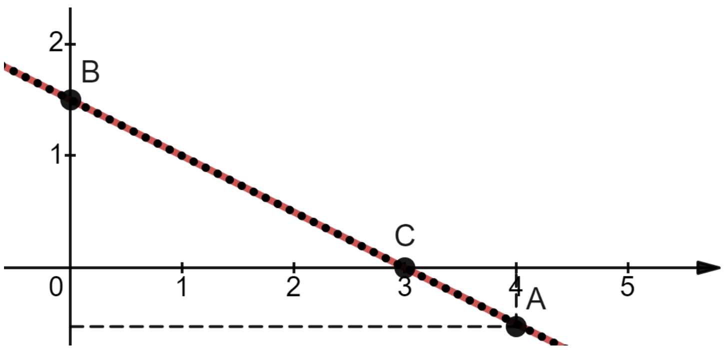
2. Also we can obtain an indefinite solution, that can provide us an infinite amount of solutions.

Consider a system of two equations:

$$\begin{cases} x + 2y = 3, \\ 2x + 4y = 5. \end{cases}$$

It is not so obvious to choose a specific solution here because a whole family of solutions of the following form $\hat{x} = \begin{bmatrix} 3 - 2y \\ y \end{bmatrix}$ is suitable for us.

And now we need to get some understanding about which solution is a kind of optimum. We will discuss it a little bit later, now let's consider one more possible situation.



Example of definite system.

3.

Inconsistent system is a system of a kind:

$$\begin{cases} 2x + y = 3, \\ 2x + y = 6 \end{cases}$$

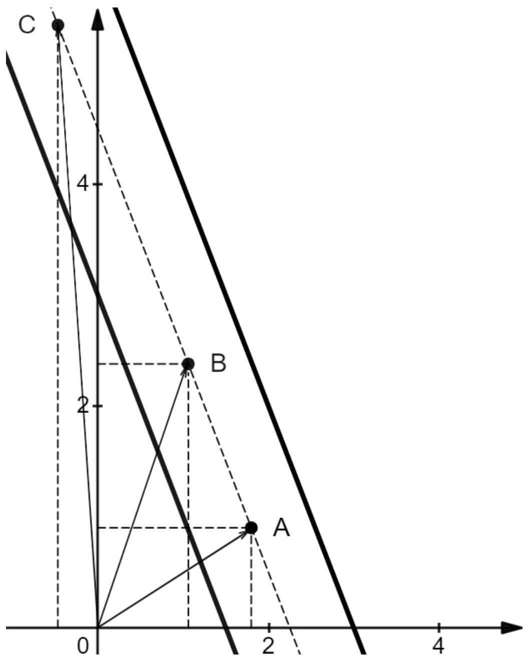
Need to remind, that we want to obtain solution in term of factors:

$$\hat{x} = ? \cdot \vec{b}.$$

There we have matrix and vector of initial values

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

Manually we can understand that the best solution will lie somewhere between two parallel lines, perhaps even exactly in the middle. But it is still a whole family of solutions that can be the answer to the request of the product or business problem. We need a general variant to find the best solution. For this we introduce the definition:



Example of inconsistent system.

Definition

Consider a system of a linear equations $A\vec{x} = \vec{b}$ $(A \in M_{m\times n}(\mathbb{C}))$. A vector $\vec{u} \in \mathbb{C}^n$ is called a pseudosolution or a least square solution, if $\forall \vec{x} \in \mathbb{C}^n$ the length of $A\vec{u} - \vec{b}$ is less or equal to the length of $A\vec{x} - \vec{b}$:

$$\left|A\vec{u} - \vec{b}\right| \leq \left|A\vec{x} - \vec{b}\right|.$$

That is: if $f_x = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} = A\vec{x} - \vec{b}$, then $|f_x|^2 = |f_1|^2 + \ldots + |f_n|^2$ for $\vec{x} = \vec{u}$ is minimal.

Theorem

The vector $\vec{u} = A^{+}\vec{b}$ is a pseudosolution of the system of linear equations $A\vec{x} = \vec{b}$. Moreover, among all pseudosolutions, $\vec{u}$ has the minimal length.

Prop

If $\hat{x}$ is a solution, then it is a pseudosolution.

Proof: $A\hat{x} - \vec{b} = 0 \Longrightarrow \left|A\hat{x} - \vec{b}\right| = 0 = \min |f_x|^2.$ □

Example 6:

Type of a system	Solution
definite	$\vec{u} = \hat{x}$ is the solution
indefinite	$\vec{u} = \hat{x}$ is the solution of minimal length
inconsistent	$\vec{u} = \hat{x}$ is the pseudosolution of minimal length

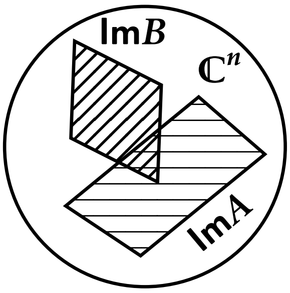
Proof: (Of the theorem) In proof we will use

Theorem: (Pythagoras)

Suppose $\vec{a} \perp \vec{b}$, that is $(\vec{a}, \vec{b}) = 0$. Then for $\vec{c} = \vec{a} + \vec{b}$: $|\vec{c}|^2 = |\vec{a}|^2 + |\vec{b}|^2$. In particular $|\vec{c}| \geq |\vec{a}|$. The equality holds only for $\vec{b} = \vec{0}$.

Lemma

Im$A \perp$ ImB, where $B = AA^+ - I$.



Proof: We should prove: each column A^j of A is orthogonal to the one B^l of B.

$$\text{OR: } \forall l \ \left(B^l, A^j\right) \stackrel{?}{=} 0, \ \text{ or } \ B^{l^*} \cdot A^j \stackrel{?}{=} 0,$$

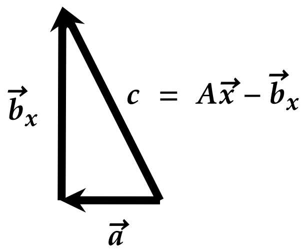
$$\text{or } \ (B^*)_l \cdot A^j \stackrel{?}{=} 0, \ \text{ or } \ B^* A \stackrel{?}{=} 0.$$

$$((AA^+)^* - I^*)\,A = (AA^+ - I)\,A = AA^+A - A = 0.$$

□

We need to prove that $\vec{u}$ is a pseudosolution. Let $\vec{x} \in \mathbb{C}^n$. We need to show:

$$\left|A\vec{x} = \vec{b}\right| \geq \left|A\vec{u} - \vec{b}\right|.$$



Here $\vec{c} = A\vec{x} - \vec{b} = A\vec{x} - A\vec{u} + A\vec{u} - \vec{b} = A(\vec{x} - \vec{u}) + AA^+\vec{b} - \vec{b} =$

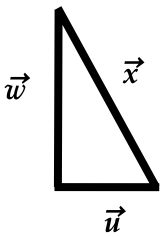
$$= A(\vec{x} - \vec{u}) + (AA^+ - I)\vec{b} = \underbrace{A\,(\vec{x} - \vec{u})}_{\vec{b}_x} + \underbrace{B\vec{b}}_{\vec{a}}.$$

By Lemma, $\vec{b}_x \perp \vec{a}$. By Pythagoras theorem, $|\vec{c}| = \min \iff \vec{b}_x = \vec{0}$. For example, it is so for $\vec{x} = \vec{u}$. So $\vec{u}$ is a pseudosolution.

We have shown, what $\vec{x}$ is a pseudosolution $\iff \vec{b}_x = 0$ or $A(\vec{x} - \vec{u}) = 0$, or $A\vec{x} = A\vec{u}$, or $A\vec{x} = AA^+\vec{b}$.

Suppose $\vec{x}$ is another pseudosolution. We need to prove that $|\vec{u}| \leq |\vec{x}|$. Let $\vec{w} = \vec{x} - \vec{u}$.

If we prove that $\vec{w} \perp \vec{u}$ then $|\vec{x}| \geq |\vec{u}|$. We have



$$\left(\vec{u}, \vec{w}\right) = \vec{u}^* \vec{w} = (A^+\vec{b})^* \vec{w} = b^* A^{+^*} \vec{w}, \ \text{ where } \ A\vec{w} = 0$$

Here $(A^+)* \stackrel{II}{=} (A^+AA^+)^* = A^{+^*}\,(A^+A)^* \stackrel{IV}{=} A^{+^*}A^+A$. So

$$(\vec{u}, \vec{w}) = b^* A^{+^*} A^+ \underbrace{A\vec{w}}_{0} = 0.$$

□

Let's return to our inconsistent system and find pseudosolution:

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Then pseudosolution can be found by the formula:

$$\hat{x} = A^+\vec{b}.$$

Pseudoinverse matrix to A can be obtained by:

$$A^+ = \left(\begin{bmatrix}1\\1\end{bmatrix}\cdot\begin{bmatrix}2 & 1\end{bmatrix}\right)^+ = \begin{bmatrix}2 & 1\end{bmatrix}^+\cdot\begin{bmatrix}1\\1\end{bmatrix}^+ = \frac{1}{2}\cdot\frac{1}{5}\begin{bmatrix}2 & 2\\1 & 1\end{bmatrix} = \frac{1}{10}\begin{bmatrix}2 & 2\\1 & 1\end{bmatrix}$$

Then we can get a pseudosolution:

$$\hat{x} = \frac{1}{10}\begin{bmatrix}2 & 2\\1 & 1\end{bmatrix}\cdot\begin{bmatrix}3\\6\end{bmatrix} = \frac{1}{10}\begin{bmatrix}18 & 9\end{bmatrix}$$

Prop

$\vec{x}$ is a pseudosolution $\iff$ $\vec{x}$ is a solution of the "normal"system of a kind:

$$A^*A\vec{x} = A^*\vec{b}.$$

Prop

All pseudosolutions (solutions) of $A\vec{x} = \vec{b}$ are given by the formula:

$$\vec{x} = A^+\vec{b} - (A^+A - I)\vec{y},$$

where $\vec{y} \in \mathbb{C}^n$ – orbitary vector.

Now let's find all pseudosolutions for example with inconsistent system. We have already obtained one pseudosolution:

$$\hat{x} = \frac{1}{10}\begin{bmatrix}18\\9\end{bmatrix}$$

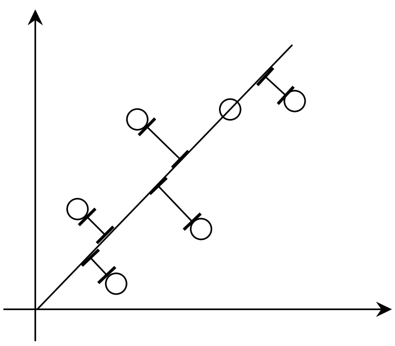
Now we need to obtain $A^+A - I$:

$$A^+A - I = \frac{1}{10}\begin{bmatrix}2 & 2\\1 & 1\end{bmatrix}\cdot\begin{bmatrix}2 & 1\\2 & 1\end{bmatrix} - \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} = \frac{1}{10}\begin{bmatrix}8 & 4\\4 & 2\end{bmatrix} - \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}-0.2 & 0.4\\0.4 & -0.8\end{bmatrix}$$

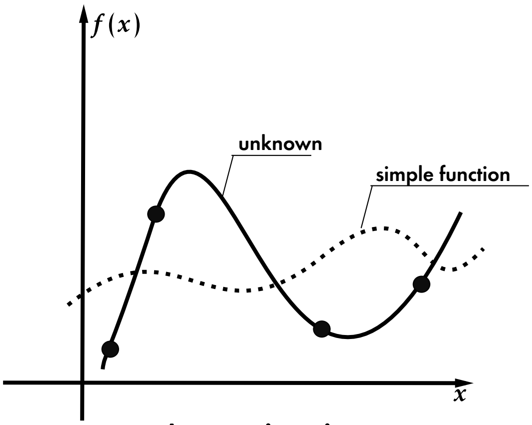
Finally,

$$\hat{x} = \frac{1}{10}\begin{bmatrix}18\\9\end{bmatrix} - \begin{bmatrix}-0.2 & 0.4\\0.4 & -0.8\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix} = \begin{bmatrix}1.8 + 0.2y_1 - 0.4y_2\\0.9 - 0.4y_1 + 0.8y_2\end{bmatrix}$$

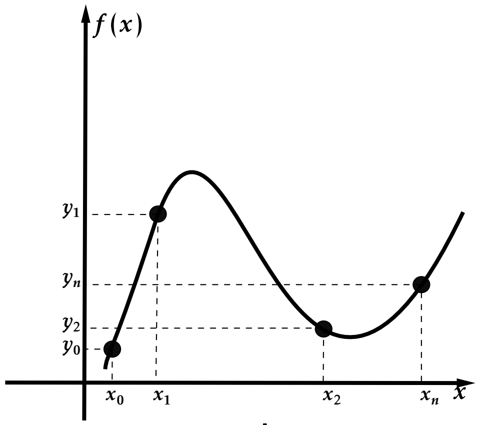
Linear regression problem



3. Approximation. Interpolation problem. Polynomial interpolation. Hermitian interpolation. Splines. Bézier curves and splines.



Approximation



Interpolation

Polynomial interpolation

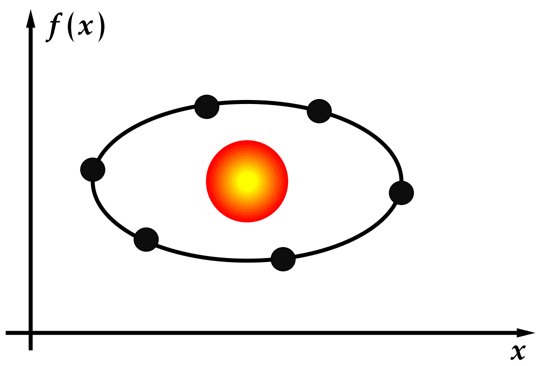
Classical problem of polynomial interpolation

Problem: Suppose some function $f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1} + a_nx^n$ is a polynomial of degree $\leq n$. Given the values of function $f(x)$:

$$\begin{cases} f(x_0) = y_0, \\ \ldots \\ f(x_n) = y_n \end{cases}$$

Recover $f(x)$ is our goal. That is the main problem to find a vector $\vec{a} = \begin{bmatrix} a_0 & \ldots & a_n \end{bmatrix}^\top = ?$

CRAMER,1750



Some algebraic equations $f(x, y) = 0$. We need $\frac{n \cdot (n+3)}{2}$ observations to recover an orbit equation of degree equal to n (generally). From the system we have system of equations:

$$\begin{cases} a_0 + a_1x_0 + \ldots + a_nx_0^n = y_0 \\ a_0 + a_1x_1 + \ldots + a_nx_1^n = y_1 \\ \ldots \\ a_0 + a_1x_n + \ldots + a_nx_n^n = y_n \end{cases}$$

We can rewrite it by matrix product:

$$V\vec{a} = \vec{y},$$

where $a = \begin{bmatrix} a_0 & a_1 & \ldots & a_n \end{bmatrix}^\top$, $y = \begin{bmatrix} y_0 & y_1 & \ldots & y_n \end{bmatrix}^\top$ and V is a Vandermonde matrix:

$$V = \begin{bmatrix} 1 & x_0 & \ldots & x_0^n \\ 1 & x_1 & \ldots & x_1^n \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \ldots & x_n^n \end{bmatrix}$$

Note

Vendermonde determinant:

$$v = v(x_0, \dots, x_n) = \det V = (x_1 - x_0)(x_2 - x_0) \dots (x_2 - x_1) \dots (x_n - x_{n-1}) =$$
$$= \prod_{0 \le i < j \le n} (x_j - x_i).$$

Example 7: $\det V = v(x_0, x_1) = \begin{vmatrix} 1 & x_0 \\ 1 & x_1 \end{vmatrix} = x_1 - x_0.$

Example 8: $\det V = v(x_0, x_1, x_2) = \begin{vmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{vmatrix} = (x_2 - x_0)(x_1 - x_0)(x_2 - x_1).$

Corollary: if all $x_0, \dots, x_n$ are different $det V = v \neq 0$. Then $\vec{a} = V^{-1}\vec{y}$ is the unique solution.

Lagrange form of interpolation polynomial

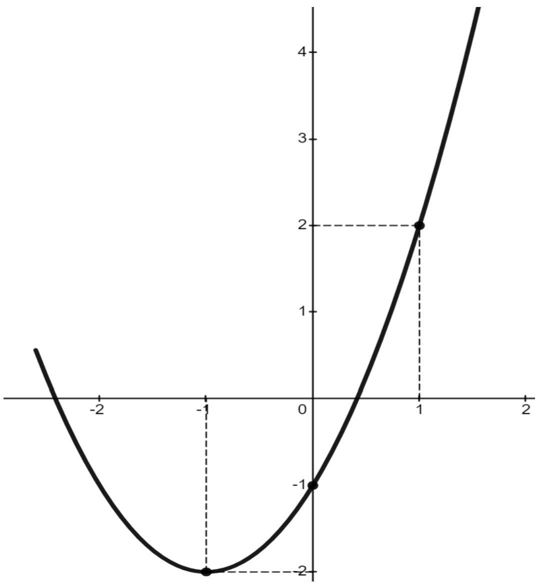
$$f(x) = \sum_{i=0}^{n} \frac{v(x_0, \dots, x, \dots, x_n)}{v(x_0, \dots, x_i, \dots x_n)} y_i = \sum_{i=0}^{n} y_i \frac{(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}.$$

Example 9: Let

$$x_0 = -1;\ y_0 = -2$$
$$x_1 = 0;\ y_1 = -1$$
$$x_2 = 1;\ y_2 = 2.$$

Let's apply Lagrange form of interpolation polynomial to calculate it:

$$f(x) = -2 \cdot \frac{(x-0)(x-1)}{(-1-0)(-1-1)} -$$
$$-1 \cdot \frac{(x+1)(x-1)}{(0+1)(0-1)} + 2 \cdot \frac{(x+1)(x-0)}{(1+1)(1-0)} =$$
$$= -x(x-1) + x^2 - 1 + x(x+1) = x^2 + 2x - 1.$$



Example of Lagrange form of interpolation polynomial.

Hermitian interpolation or interpolation with multiple knots

Definition

A number x_1 is a root of a polynomial $f(x)$ with multiplicity d if

$$f(x) = (x - x_1)^d \cdot g(x)$$

for some polynomial $g(x)$ such that $g(x_1) \neq 0$

Lemma

x_1 is a root of multiplicity d for a polynomial $f(x)$ if and only if:

$$\begin{cases} f(x_1) = 0 \\ f'(x_1) = 0 \\ \vdots \\ f^{(d-1)}(x_1) = 0 \\ f^{d}(x_1) \neq 0 \end{cases}$$

Proof: x_1 is a root of $f(x)$ of multiplicity $d \geq 1$, if and only if:

$$f'(x) = d\cdot(x-x_1)^{d-1}\cdot g(x) + (x-x_1)^d g'(x) = \\ = (x-x_1)^{d-1}\underbrace{(dg(x) + (x-x_1)g'(x))}_{h(x)},$$

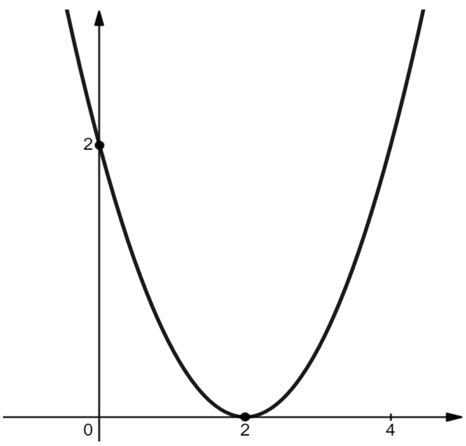
where $h(x_1) = dg(x_1) \neq 0 \Leftrightarrow x_1$ is a root of $f'(x)$ with multiplicity $d-1$. □

Problem (Brief): find $f(x)$ by m knots with multiplicities $h_1, h_2, \ldots, h_m$.

Complete: to find a polynomial $f(x)$ of degree $\leq n-1$, such that for some different $\underbrace{x_1, x_2, \ldots, x_m}_{\text{knots}} \in \mathbb{R}$ and $\underbrace{h_1, \ldots, h_m}_{\text{multiplicities}} \in \mathbb{N}$ with $h_1 + h_2 + \ldots + h_m = n$, and $y_1, y_1^{(1)}, \ldots, y_m^{h_m-1}$;

$$f(x_1) = y_1;\ f'(x_1) = y_1^{(1)}, \ldots, f(x_1)^{(h_1-1)} \\ \ldots \\ f(x_m) = y_m, \ldots, f^{(h_m-1)}(x_m) = y_m^{(h_m-1)}.$$

Prop This problem always has a unique solution.



Example of interpolation with multiple knots

Example 10:

$$f(x) = ax^2 + bx + c; \\ f'(x) = 2ax + b \\ x_1 = 0;\quad f(0) = 2; x_2 = 2;\quad f(2) = 0;\ f'(2) = 0;$$

$$\begin{cases} f(0) = a0^2 + b0 + c = 2; & n = 3, \deg f \leq 2 \\ f(2) = 4a + 2b + c = 0; & x_1 = 0,\ h_1 = 1 \\ f'(2) = 4a + b = 0; & x_2 = 2,\ h_2 = 2. \end{cases}$$

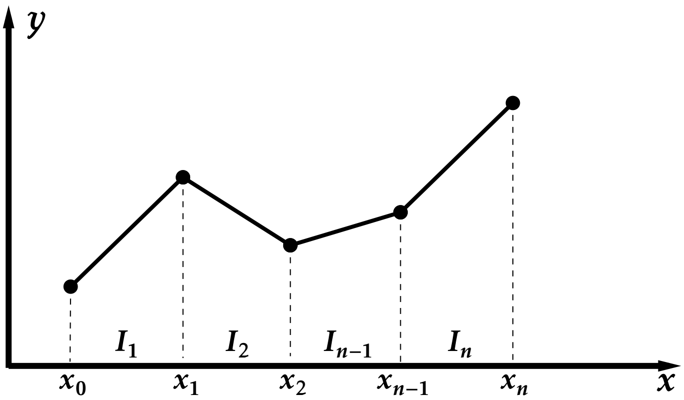
Solving equations we can obtain next results:

$$\begin{cases} c = 2 \\ b = -2 \\ a = \frac{1}{2} \end{cases} \Longrightarrow f(x) = \frac{1}{2}x^2 - 2x + 2.$$

Splines

An idea: to interpolate a "smooth" function $f(x)$ with knots $x_0, \ldots, x_n$, on each $I_i = [x_{i-1}, x_i]$ put $f(x) = f_i(x)$.

Example 11: Given the values $f(x_i) = y_i$, let $f(x_i)$ be a linear function:



Example of a linear spline

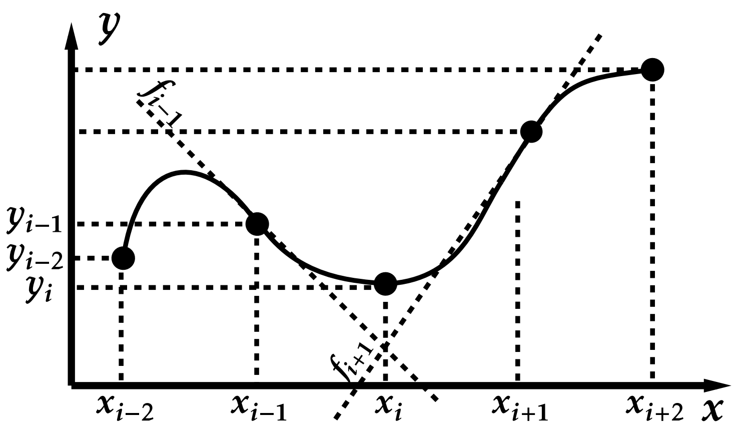
Quadratic Spline

Let $f(x)$ have a continuous derivative $f'(x)$:

$$f(x_i) = y_i; \quad i = 0, 1, \dots, n.$$

Then put $f_i(x) = a_i x^2 + b_i x + c_i$ and let

$$f(x) = \begin{cases} f_1(x); \ x \in I_1 \\ f_2(x); \ x \in I_2 \\ \vdots \\ f_n(x); \ x \in I_n. \end{cases}$$



For each f_i, $i > 1$:

$$\begin{cases} f_i(x_{i-1}) = y_{i-1} \\ f_i'(x_{i-1}) = f_{i-1}'(x_{i-1}) \\ f_i(x_i) = y_i \end{cases}$$

For $i = 1$, we add $f_1'(x_0) = 0$ or $f_1'(x_0) = x$ (if it is known) or $f_1'(x_n) = f_n'(x_n)$ (for periodic processing).

Cubic spline

$f(x)$ has continuous $f''(x)$: for $i = 1$, two initial conditions $f'(x_0) = f''(x_0) = 0$

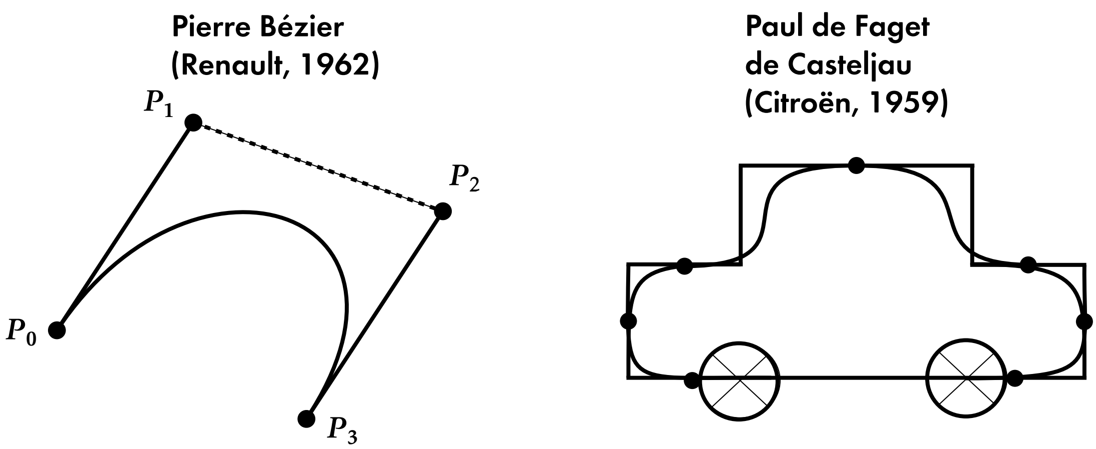
$$\begin{cases} f_i(x_{i-1}) = y_{i-1} \\ f_i(x_{i-1}) = f_{i-1}(x_{i-1}) \\ f_i''(x_{i-1}) = f_{i-1}''(x_{i-1}) \\ f_i(x_i) = y_i \end{cases}$$

Bézier curves and splines

<u>Problem</u>: Given $A_0A_1 \dots A_n \in \mathbb{R}^m$ approximate the path

$$A_0 \to A_1 \to A_2 \to \dots \to A_n$$

by a smooth curve.



Applications of interpolation splines and curves

Explicit formula for Bézier curve:

$$B(t) = \sum_{i=0}^{n} A_i \cdot b_{n,i}(t),$$

where $b_{n,i}$ – Bernstein polynomials of a kind:

$$b_{n,i} = C_n^i (1-t)^{n-i} t^i$$

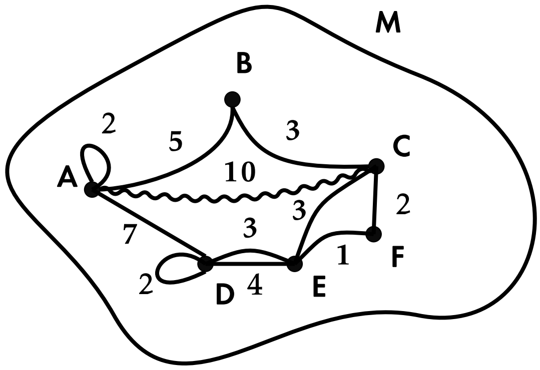
4. Metric axioms. Metric spaces. Norms. Normed linear spaces

Definition: (Metric space)

A metric space is a set X with a metric $\rho : X \times X \to [0, \infty)$ (or it can be valid notation d, in that way we can call it by 'distance') such that $\forall x, y, z \in X$, ρ satisfies the following properties:

1. Positive definite:
$$\rho(x,y) \geq 0, \;\; \forall x \neq y$$
$$\rho(x,y) = 0 \Longleftrightarrow x = y; \; \rho(x,x) = 0.$$
2. Symmetric:
$$\rho(x,y) = \rho(y,x).$$
3. Triangle Inequality:
$$\rho(x,z) \leq \rho(x,y) + \rho(y,z).$$

Example 12:



Let define some metric space with metric $\rho(x,y)$ is equal to the legth if the shortest path. Then $\dim(M) = 10$ and, e.g.:

$$\rho(A,A) = 0;$$
$$\rho(A,D) = 7;$$
$$\rho(A,C) = 8$$

Also we can show, for example, open ball on this example (we will define it little bit later):

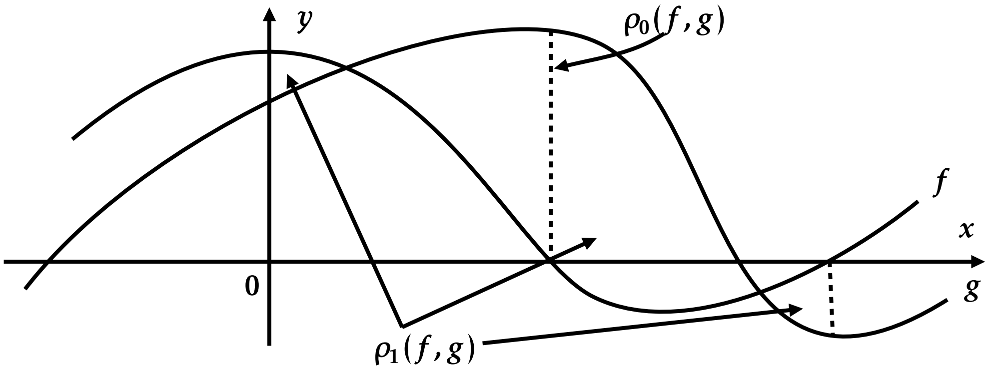
$$B_3(A) = \{A\}$$
$$B_9(A) = \{A, B, C, D, E\}$$

Example 13: Given a set X:

- The discrete metric ρ on X is defined by:
$$\rho(x,y) = \begin{cases} 1, & x \neq y \\ 0, & x = y. \end{cases}$$

- Metric on continuous functions:
Let $X = C[0,1] = \{$continuous functions $f : [0,1] \to \mathbb{R}\}$. Then we can define metrics:
$$\rho_0(f,g) = \max_{x \in [0,1]} |f(x) - g(x)|.$$
$$\rho_1(f,g) = \int_0^1 |f(x) - g(x)| dx$$

Or:
$$\rho(f,g) = \rho_0(f,g) + |f(1) - g(1)|.$$



- Let $x = \mathbb{R}$. Possible metric:

$$\rho(x,y) = \left|e^x - e^y\right|.$$

- Another example of metric:

$$\rho(x,y) = \begin{cases} 1, & x - y \in \mathbb{Q} \\ 2, & x - y \notin \mathbb{Q} \\ 0, & x = y \end{cases}$$

Definition: (Continuous function)

The function is called continuous iff:

$$\lim_{x \to x_0} f(x) = y_0,$$

that is $\forall \varepsilon > 0\ \exists \delta :$

$$f\left[B_\delta(x_0)\right] \subset B_\varepsilon\left[f(x_0)\right] \equiv B_\varepsilon(y_0)$$

Definition: (Open ball)

An open ball of radius $r > 0$ centered at the pont $x_0 \in X$ is the set:

$$B(x_0, r) = \{x \in X |\ \rho(x, x_0) < r\}$$

Definition: (Closed ball)

A closed ball of radius $r > 0$ centered at the point $x_0 \in X$ is the set:

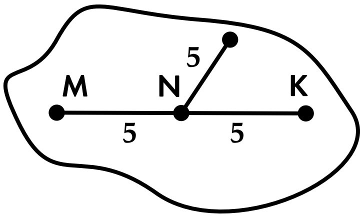
$$\overline{B}(x_0, r) = \{x \in X |\ f(x, x_0) \leq r\}.$$

Note

A ball centered at the point A of radius r in some metric space X"

$$B_r(A) = \{x \in X |\ f(A, x) \leq r\}.$$

Example 14: Let define space:



$$B_8(M) = \{M, N\}$$
$$B_6(N) = \{M, N, K\}$$

Definition: (Normed space)

A (complex or real) vector space V is called normed space if a function ('norm') $\nu : V \to \mathbb{R}$, denoted for $v \in V$ $||v||$, which satisfies the following axioms:

1. Positive definite property:
$$\nu(\vec{x}) > 0;$$
2. Homogeneity
$$\nu(\alpha\vec{x}) = |\alpha|\nu(\vec{x});$$
3. Triangle inequality $\forall x, y \in V$:
$$\nu(\vec{x}+\vec{y}) \le \nu(\vec{x}) + \nu(\vec{y}).$$

Example 15: Euclidean norm:

$$|\vec{x}|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

Prop

In a normed space V, the function $\rho(\vec{x}, \vec{y}) = \nu(\vec{y} - \vec{x})$ is a metric.

Proof: For proving positive definition of function $\nu(\vec{x}, \vec{y})$ we need a lemma:

Lemma

$\nu(\vec{0}) = 0$

Proof: $\nu(0 \cdot \vec{0}) = 0 \cdot \nu(\vec{0}) = 0.$ □

1. Positive definition:
$$\rho(\vec{x}, \vec{y}) = \nu(\vec{y} - \vec{x}) > 0$$
$$\rho(\vec{x}, \vec{x}) = \nu(\vec{x} - \vec{x}) = \nu(\vec{0}) = 0$$
2. Symmetric:
$$\rho(\vec{x}, \vec{y}) = \nu(\vec{y} - \vec{x}) = |-1|\nu(\vec{x} - \vec{y}) = \nu(\vec{x} - \vec{y}) = \rho(\vec{y}, \vec{x})$$
3. Triangle inequality:
$$\rho(\vec{x}, \vec{y}) + \rho(\vec{y}, \vec{z}) = \nu(\vec{y} - \vec{x}) + \nu(\vec{z} - \vec{y}) \ge$$
$$\ge \nu(\vec{y} - \vec{x} + \vec{z} - \vec{y}) = \nu(\vec{z} - \vec{x}) = \rho(\vec{x}, \vec{z})$$

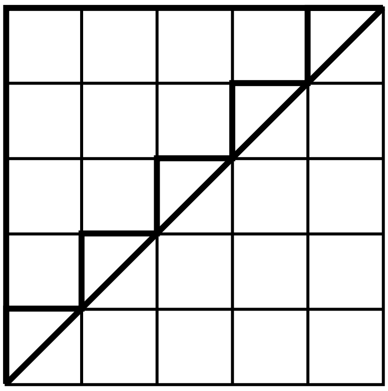
□

Note

Each normed space is a metric space.

Example 16: In the vector space $\mathbb{R}^n(\mathbb{C}^n)$ the following three norms are in common use:

- Manhattan norm (Taxicab norm):



This is the norm of the following kind:

$$||x||_1 = \sum_{i=1}^{n} |x_i|$$

Let's prove that this function is a norm.

1. Positive definite property: Let $x \in \mathbb{R}^n$ or $x \in \mathbb{C}^n$. Obviously $||x||_1 \geq 0$. Also $||x||_1 = 0$ iff $x = 0$.
2. Homogeneity property:
$$\forall c \in \mathbb{R}: \ ||c \cdot x||_1 = \sum_{i=1}^{n} |c \cdot x_i| = |c| \cdot \sum_{i=1}^{n} |x_i| = |c| \cdot ||x||_1.$$
3. Triangle inequality $\forall x, y \in \mathbb{R}^n$:

$$||x + y||_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} (|x_i| + |y_i|) =$$
$$= \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| = ||x||_1 + ||y||_1.$$

- Maximum norm (Infinity norm):
$$||x||_\infty = \max_{1 \leq i \leq n} |x_i|.$$

1. The function $||x||_\infty$ is positive since it is the maximum over a set of positive terms $|x_i|$.
2. Homogeneity property:
$$||\alpha \cdot x||_\infty = \max_{1 \leq i \leq n} |\alpha \cdot x_i| = \max_{1 \leq i \leq n} |\alpha| \cdot |x_i| = |\alpha| \cdot \max_{1 \leq i \leq n} = |\alpha| \cdot ||x||_\infty.$$
3. Triangle inequality:
$$||x + y||_\infty = \max_{1 \leq i \leq n} |x_i + y_i| \leq \max_{1 \leq i \leq n} (|x_i| + |y_i|) \leq \max_{1 \leq i \leq n} |x_i| + \max_{1 \leq i \leq n} |y_i| = ||x||_\infty + ||y||_\infty.$$

Example 17: For the vector $x = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 5 \end{bmatrix}$ we have:

$$||x||_1 = 11; \ ||x||_2 = \sqrt{39}; \ ||x||_\infty = 5,$$

whereas for the vector $x = \begin{bmatrix} 1+i \\ 2-3i \\ 4 \end{bmatrix}$,

$$||x||_1 = \sqrt{2} + \sqrt{13} + 4;\ ||u||_2 = \sqrt{31};\ ||u||_\infty = 4.$$

The notations $||\cdot||_1, ||\cdot||_2$ and $||\cdot||_\infty$ are justified because of the fact that all these norms are special cases of the general Minkovskiy $p-$norm:

$$||x||_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p},\ p \geq 1$$

Similarly, in the vector space of real-valued continuous functions $C[a,b]$, the following three norms are frequently used:

$$||f||_1 = \int_a^b |f(x)|dx;\ ||f||_2 = \sqrt{\int_a^b f^2(x)dx};\ ||f||_\infty = \max_{x\in[a,b]} |f(x)|$$

And Minkovskiy norm:

$$||f||_p = \sqrt[p]{\int_a^b |f(x)|^p dx}$$

For example, for $V = C[0,1]$:

$$||f||_p = \sqrt[p]{\int_0^1 |f(x)|^p dx}$$

$$||f||_1 = \int_0^1 |f(x)|^p dx$$

$$||f||_\infty = ||f||_0 = \max_{x\in[0,1]} |f(x)|.$$

Some weighted norms:
Let $V = C[0,1]$; $\omega \geq 0$:

$$||f||_p^\omega = \left(\int_0^1 |f(x)|^p \cdot \omega(x)dx\right)^{\frac{1}{p}}$$

$$||f||_\infty^\omega = ||f||_0^\omega = \max_{x\in[0,1]} |f(x)\cdot\omega(x)|.$$

Balls in normed space

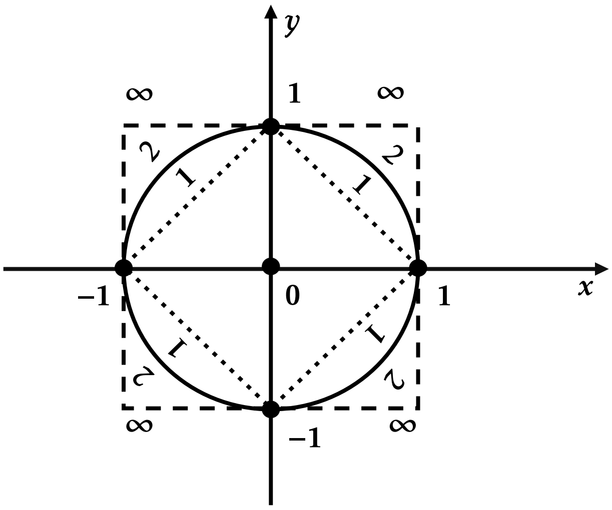
Prop

Let V be a normed space

- All balls of the same radius R are equal geometrically: a parallel translation of $B_R(\vec{x})$ by a vector $\vec{v} = \vec{y} - \vec{x}$ makes $B_R(\vec{y})$.
- For any two balls: $B = B_r(0)$ and $B' = B_R(0)$, there is a homotety $x \longmapsto \lambda\vec{x}$, which transfers B

onto B', where $\lambda = \dfrac{R}{r}$. Also $B_1^p(0)$ = unit ball for 0-norm.

Proof:



1.
$$B_R(\vec{x}) + V = \left\{\vec{a} \mid \nu\left(\vec{a} - \vec{x}\right) \le R\right\} + \left(\vec{y} - \vec{x}\right) =$$
$$= \left\{\vec{a} + \vec{y} - \vec{x} \mid \nu\left(\vec{a} - \vec{x}\right) \le R\right\} = \left\{\vec{b} \mid \nu\left(\vec{b} - \vec{y}\right) \le R\right\} =$$
$$= B_R(\vec{y}),\ \text{where: } \vec{a} + \vec{y} - \vec{x} = \vec{b};\ \vec{a} = \vec{b} - \vec{y} + \vec{x}.$$

2. $\dfrac{R}{r} \cdot \vec{a} = \vec{b}; \vec{a} = \dfrac{r}{R} \cdot \vec{b}.$

$$\lambda \cdot B = \frac{R}{r} \cdot \left\{\vec{a} \mid \nu\left(\vec{a}\right) \le r\right\} = \left\{\frac{R}{r} \cdot \vec{a} \mid \nu\left(\vec{a}\right) \le r\right\} =$$
$$= \left\{\vec{b} \mid \nu\left(\frac{r}{R} \cdot \vec{b}\right) \le r\right\} = \left\{\vec{b} \mid \frac{r}{R} \cdot \nu\left(\vec{b}\right) \le r\right\} =$$
$$= \left\{\vec{b} \mid \nu\left(\vec{b}\right) \le \frac{r \cdot R}{r}\right\} = \left\{\vec{b} \mid \nu\left(\vec{b}\right) \le R\right\} = B'.$$

□

5. Norms. Minkovski's theorem. Euclidian space

Definition: (Norm)

Let V be a vector space $\in \mathbb{R}^n$ or $\mathbb{C}^n$. Then function $\nu : V \to \mathbb{R}$ is a norm if:

1. $\nu(\vec{x}) > 0 \ \forall \vec{x} \neq 0$;
2. $\nu(\alpha \vec{x}) = |\alpha| \nu(\vec{x})$;
3. $\nu(\vec{x} + \vec{y}) \leq \nu(\vec{x}) + \nu(\vec{y})$.

Theorem

6. Chebyshev polynomials of the first kind. Chebyshev polynomials of the second kind