

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

$$\text{GRADE} = \frac{\text{test1} + \text{test2}}{2} + \underbrace{\text{Bonus for a talk}}_{\leq 5} + \underbrace{\text{Bonus for classes}}_{\leq 1 \dots 2}.$$

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:

$$[A|b] \xrightarrow{\text{Gaussian elimination}} \left[\begin{array}{ccc|c} 1 & & * & b_1 \\ & 1 & & b_2 \\ & & \ddots & \vdots \\ & & & 1 & b_m \end{array} \right]$$

(Note: In the original image, there is a circle around the first column and an oval around the last row.)

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.

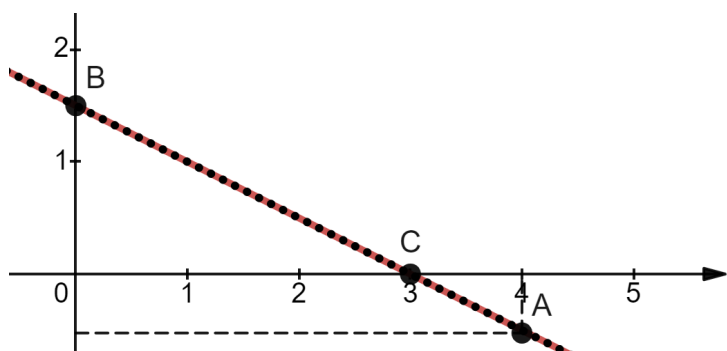
Suppose you have created a linear model, that more or less describe the production or business process. For example, you get the vector of unknowns \vec{x} , which means that you should sell x_1 copies of product 1 and x_2 copies of product 2, etc. How much you must produce or sell? It is a possible 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

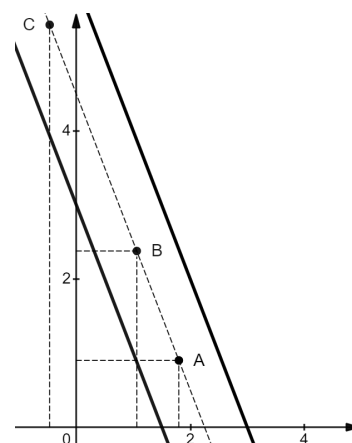
augmented matrix (matrix of a kind $[A|b]$). Such systems can be either determined (defined) or underdetermined.

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 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:

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

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

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, $\text{rank } A = n$. Then: $A^+ = (\underbrace{A^* A}_{n \times n})^{-1} A^*$.

Prop

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

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 $r = \text{rank } A$ is a decomposition:

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

(Then $\text{rank } F = \text{rank } G = r$. 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:

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

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:

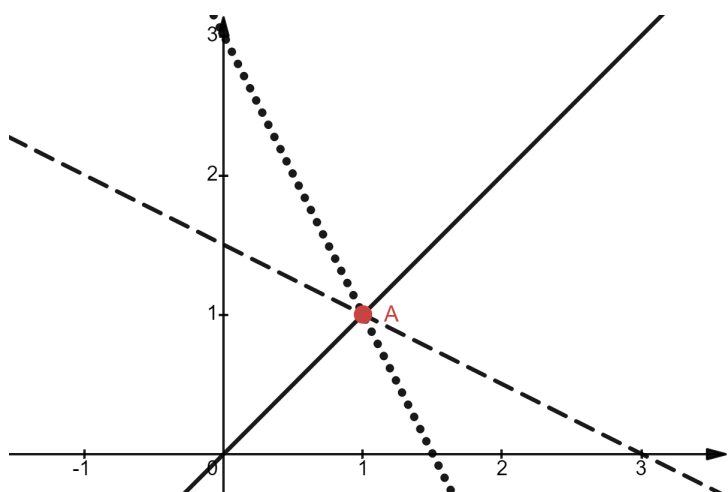
$$A\vec{x} = \vec{b},$$

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 $A \in M_{n \times n}(\mathbb{C})$, $\text{rank } A = n$. In such situation we can easily obtain unique \vec{x} by inverting the matrix of initial coefficients:

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

1. The next easy option is a definite system, when $A \in M_{m \times n}(\mathbb{C})$, $\text{rank } A = n$. 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.

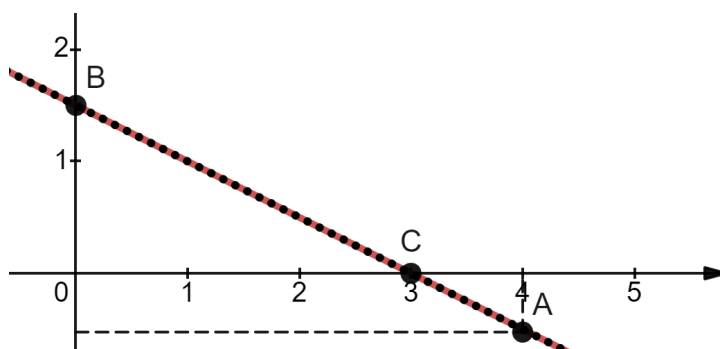
Consider a system of two equations:

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

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.

3.



Example of definite system.

Inconsistent system is a system of a kind:

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

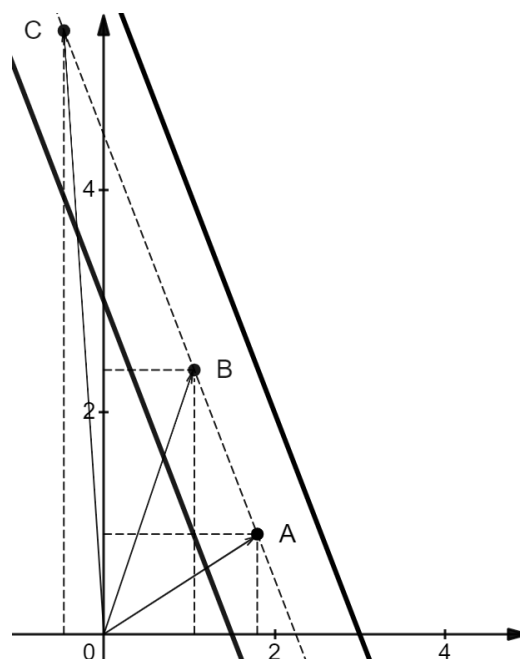
Need to remind, that we want to obtain solution in term of factors:

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

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 $A\vec{x} - \vec{b}$:

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

That is: if $f_x = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} = A\vec{x} - \vec{b}$, then $|f_x|^2 = |f_1|^2 + \dots + |f_n|^2$ 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: $A\hat{x} - \vec{b} = 0 \implies |A\hat{x} - \vec{b}| = 0 = \min |f_x|^2.$

□

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

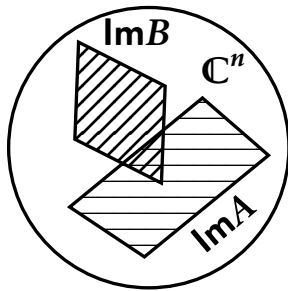
Proof: (Of the theorem) In proof we will use

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

$\text{Im}A \perp \text{Im}B$, where $B = AA^+ - I$.



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

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

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

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

□