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, matrix functions 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

$$\text{GRADE} = \frac{\text{test1} + \text{test2}}{2} + \underbrace{\substack{\text{Bonus}\\ \text{for a talk}}}_{\le 5} + \underbrace{\substack{\text{Bonus}\\ \text{for classes}}}_{\le 1\ldots 2}$$

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:

$$[A|b] \underset{\text{elimination}}{\overset{\text{Gaussian}}{\rightsquigarrow}} \left[\begin{array}{cccc|c} 1 & & * & & b_1 \\ & 1 & & & b_2 \\ & & \ddots & & \vdots \\ & & & 1 & b_m \\ & \mathbf{0} & & & 0 \\ & & & & \vdots \\ & & & & 0 \end{array}\right]$$

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.

Indefinite system

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 if !!!

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:

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