Foundations of Data Analysis (MA4800)

Massimo Fornasier

Fakultät für Mathematik
Technische Universität München
massimo.fornasier@ma.tum.de
http://www-m15.ma.tum.de/

Introduction to the Lecture
April 25 2017

Organizational matters

Lecture times to avoid conflicts with other courses:

- Tuesdays, 10:00-11:30 in Garching-Hochbrueck, **room BC2 3.1.08 ONLY on May 2** and room BC2 3.5.06 for the rest of the time
  `http://portal.mytum.de/campus/roomfinder/roomfinder_viewmap?mapid=183&roomid=BC2%203.5.06@8102`
- Wednesdays, 8:30-10:00 in Garching, room MI Hoersaal 3 00.06.011

About the course

Foundations of Data Analysis belongs to the curricula of the Masters of *Mathematics in Data Science* and *Data Engineering and Analytics* of the Technical University of Munich (TUM), starting April 2017 (`http://www.data-master.tum.de`).

Prerequisites: standard courses of analysis, linear algebra, probability, and rudiments of optimization as taught in **Computer Science**.

Goals:

- to refresh and consolidate the knowledge of students (in particular for students of computer science) precisely on the above mentioned basic mathematical topics with emphasis on their interpretation and application in the analysis of large data sets (big data).
- to furnish to students (in particular to students of mathematics) the fundamental mathematical tools to be able to access on a solid basis more advanced lectures of the Master program *Mathematics in Data Science*, in particular the courses of **Mathematical Foundations of Machine Learning**, **Statistical Learning**, **Geometry and Topology for Data Analysis**, and **Probabilistc Methods and Algorithms in Data Analysis**.

Content and literature

We will explore how linear algebra, optimization, and probabiliy need to be fused together for allowing the analysis of large data sets in high-dimension.

Basic references: To help students to keep track of the basic mathematical tools, we include references to standard lecture notes (Skripta) of courses delivered at TUM for **students of computer science**, as well as easy-to-access *online* resources from courses delivered to **students of computer science**, e.g., at MIT and at the Carnegie Mellon Universities (the best programs of computer science in the US)[1].

Website, slides, lecture notes: At the website
`https://www-m15.ma.tum.de/Allgemeines/FDA`
we provide slides, (growing) lecture notes, and references.

[1] https://www.usnews.com/best-graduate-schools/top-science-schools/computer-science-rankings

Learning objectives and EXAM!

Content is certainly of mathematical nature, all the results presented do have practical impact and significant uses in analysis of large data sets, and can be implemented numerically (for instance in Matlab).

Main learning goals:

- Understanding precisely the mathematical notions and the main theoretical results;
- Being able to apply the theoretical results on concrete examples, which might involve some calculations, and being able to detect when such results are not applicable;
- Implementing and testing the main results numerically on data sets and verfying their (practical) validity, for instance by means of Matlab.

All these goals will be tested in the final exam, including the Matlab implementations.

Learning the proofs!

Students with a mathematical background and students of computer science, who are interested to get more solid in theoretical computer science for data analysis are warmly invited to pick their (unique!) chance to study in detail also the proofs of the theoretical results and get familiar with such more theoretical reasoning, fundamental nowadays for standing the international competition in data science research.

Disclamer on the learning

Despite the fact that we are furnishing a truly exaggerated amount of supports to learning (lectures, lecture notes, slides, references to books and online resources, etc.), there is nothing more crucial and fundamental for learning than the old, simple, effective, and galvanizing individual experience of

- attending the lectures!
- taking your own notes during the lectures, without relying on the given supports;
- re-reading, correcting and extending your own lecture notes, while you are studying quietly at home or anywhere else your feel comfortable and focused;
- including in your lecture notes the exercises and home-works and their solutions.

Tips for the success of the learning

We strongly recommend to invest not less than 6 hours/week in the study of the theoretical part of the course.

Anything less than that will lead to a very superficial understanding and likely poor results at the exam.

Being able to focus for a few hours without being disturbed is a sweet and deep form of meditation, which can be facilitated by listening (repeatidly) some pleasant music (like a mantra), and by avoiding distractions (such as internet searches and the use of the mobile phone).

Learning and exercising how to focus is necessary to achieve higher intellectual results. Master it, it will help you to achieve, including better long term careers and opportunities.

Disclamer on the lecture notes - April 23, 2017

The course **Foundations of Data Analysis** has never been taught before at TUM and it is likely one of the first attempts of presenting recent mathematical methods for big data analysis for students for mathematics **and** computer science in Germany.

The lecture notes are a living text currently growing and evolving according to the development of the first editions (2017, 2018, etc.) of the course.

They will likely contain typos, (small) errors, and they will be constantly updated. For this reason, we furnish plenty of more consolidated references from which the content is (partially) taken and re-elaborated.

Preliminaries on Linear Algebra (LA)

We give for granted familiarity with the basics of LA taught in standard courses, in particular,

- vector spaces, spans and linear combinations, linear bases, linear maps and matrices, eigenvalues, complex numbers, scalar products, theory of symmetric matrices.

For more details we refer to the lecture notes (in German)

G. Kemper, *Lineare Algebra fuer Informatik*, TUM, 2017.

of the course taught for Computer Scientsts at TUM, or any other standard international text of LA.

Matrix Notations

The index sets $I, J, L, \dots$ are assumed to be finite.

As soon as the complex conjugate values appear[2], the scalar field is restricted to $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

The set $\mathbb{K}^{I \times J}$ is the vector space of matrices $A \in \mathbb{K}^{I \times J}$, whose entries are denoted by A_{ij}, for $i \in I$ and $j \in J$. Vice versa, numbers $a_{ij} \in \mathbb{K}$ may be used to define $A := (a_{ij})_{i \in I, j \in J} \in \mathbb{K}^{I \times J}$.

If $A \in \mathbb{K}^{I \times J}$, the transposed matrix $A^T \in \mathbb{K}^{J \times I}$, and $A^T_{ji} := A_{ij}$. A matrix $A \in \mathbb{K}^{I \times I}$ is symmetric if $A^T = A$. The Hermitian transposed matrix $A^H \in \mathbb{K}^{J \times I}$ coincides with $\overline{A^T}$. If $\mathbb{K} = \mathbb{R}$ then clearly $A^H = A^T$. A Hermitian matrix satisfies $A^H = A$.

Often, we will need to consider the rows $A^{(i)}$ and the columns $A_{(j)}$ of a matrix, defined respectively as vectors $A^{(i)} = (a_{ij})_{j \in J}$ and $A_{(j)} = (a_{ij})_{i \in I} = (A^T)^{(j)}$.

---

[2] In the case $\mathbb{K} = \mathbb{R}$, then $\alpha = \overline{\alpha}$, for all $\alpha \in \mathbb{K}$.

Matrix Notations

We assume that the standard matrix-vector and matrix-matrix multiplications are done in the usual way: $(Ax)_i = \sum_{j\in J} a_{ij}x_j$ or $(AB)_{i\ell} = \sum_{j\in J} A_{ij}B_{j\ell}$ as usual, for $x \in \mathbb{K}^J$, $A \in \mathbb{K}^{I\times J}$ and $B \in \mathbb{K}^{J\times L}$.

The Kronecker symbol is defined by

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \in I, \\ 0, & \text{otherwise.} \end{cases}$$

The unit vector $e^{(i)} \in \mathbb{K}^I$ is defined by

$$e^{(i)} = (\delta_{ij})_{j\in I}$$

The symbol $I = (\delta_{ij})_{j\in I, j\in I}$ is used for the identity matrix. Since matrices and index sets do not appear in the same place, the simultaneous use of the symbol I do not create confusion (example: $I \in \mathbb{K}^{I\times I}$).

Matrix Notations

The range of matrix $A \in \mathbb{K}^{I \times J}$ is

$$\mathrm{range}(A) = \{Ax : x \in \mathbb{K}^J\} = \mathrm{span}\{A_{(i)}, i \in I\}.$$

Hence, the range of a matrix is a vector space spanned by its columns.

The Euclidean scalar product in $\mathbb{K}^I$ is given by

$$\langle x, y \rangle = y^H x = \sum_{i \in I} x_i \overline{y}_i, \tag{1}$$

where for $\mathbb{K} = \mathbb{R}$ the conjugate sign can be ignored .

It often useful and very important to notice that the matrix-vector Ax and matrix-matrix AB multiplications can also be expressed in terms of scalar products of the rows of A with x and of the rows of A with the colums of B, i.e., according to our terminology of

$$(Ax)_i = \langle A^{(i)}, \overline{x} \rangle, \quad (AB)_{i\ell} = \langle A^{(i)}, \overline{B_{(\ell)}} \rangle. \tag{2}$$

Matrix Notations

Two vectors $x, y \in \mathbb{K}^I$ are orthogonal (and we may write $x \perp y$) if $\langle x, y \rangle = 0$.

We often consider sets which are mutually orthogonal. In this case for $X, Y \subset \mathbb{K}^I$ we say that X is orthogonal to Y and we write $X \perp Y$ if $\langle x, y \rangle = 0$, for all $x \in X$ and $y \in Y$.

When X and Y are linear subspaces their orthogonality can be simply checked by showing that they possess bases which are mutually orthogonal. This is a very important principle we will use often.

A family of vectors $X = \{x_\nu\}_{\nu \in F} \subset \mathbb{K}^I$ is orthogonal if the vectors x_ν are pairwise orthogonal, i.e., $\langle x_\nu, x_{\nu'} \rangle = 0$ for $\nu \neq \nu'$. The family is additionally called orthonormal if $\langle x_\nu, x_\nu \rangle = 1$ for all $\nu \in F$.

Matrix Notations

A matrix $A \in \mathbb{K}^{I \times J}$ is called orthogonal, if the columns of A are orthonormal, equivalently if

$$A^H A = I \in \mathbb{K}^{J \times J}.$$

An orthogoal square matrix $A \in \mathbb{K}^{I \times I}$ is called unitary. Differently from just orthogonal matrices, unitary matrices satisfy

$$A^H A = A A^H = I,$$

i.e, $A^H = A^{-1}$ is the invese matrix of A.

Assume that the index sets satisfy either $I \subset J$ or $J \subset I$. Then a (rectangular) matrix $A \in \mathbb{K}^{I \times J}$ is diagonal if $A_{ij} = 0$ for all $i \neq j$.

Matrix Rank

Proposition

Let $A \in \mathbb{K}^{I \times J}$. The following statements are equivalent and may be all used as a definition of matrix rank $r = \text{rank}(A)$.

- *(a)* $r = \dim \text{range}(A)$;
- *(b)* $r = \dim \text{range}(A^H)$;
- *(c) r is the maximal number of linearly independent rows of A;*
- *(d) r is the maximal number of linearly independent columns of A;*
- *(e) r is minimal with the property*

$$A = \sum_{i=1}^{r} a_i b_i^H, \text{ where } a_i \in \mathbb{K}^I, b \in \mathbb{K}^J;$$

- *(e) r is maximal with the property that there exists an invertible $r \times r$ submatrix of A;*
- *(f) r is the number of positive singular values (soon!).*

Matrix Rank

The rank is bounded by the maximal rank, which, for matrices, is always given by $r_{max} = \max\{\#I, \#J\}$, and this bound is attained by full-rank matrices.

As linear independence may depend on the field $\mathbb{K}$ one may question whether the rank of a real-valued matrix depends on considering it in $\mathbb{R}$ or in $\mathbb{C}$. For matrices it tunrs out that it does not matter: the rank is the same.

We will often work with matrices of bounded rank $r \leq k$. We denote accordingly with $\mathcal{R}_k = \{A \in \mathbb{K}^{I \times J} : \text{range}(A) \leq k\}$ such a set. Notice that this set is not a vector space (exercise!).

Norms

In the following we consider an abstract vector space V over the field $\mathbb{K}$. As a typical example we may keep in mind the Euclidean plane $V = \mathbb{R}^2$ endowed with the classical Euclidean norm.

We recall below the axioms of an abstract norm on V: a norm is a map $\|\cdot\| : V \to [0, \infty)$ with the following properties

- $\|v\| = 0$ if and only if $v = 0$;
- $\|\lambda v\| = |\lambda|\|v\|$ for all $v \in V$ and $\lambda \in \mathbb{K}$;
- $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$ (triangle inequality).

A norm is always continuous as a consequence of the (inverse) triangle inequalty:

$$|\|v\| - \|w\|| \leq \|v - w\|, \text{ for all } v, w \in V.$$

A vector space V endowed with a norm $\|\cdot\|$, and we write the pair $(V, \|\cdot\|)$ to indicate it, is called a normed vector space. As mentioned above a typical example is the Euclidean plane.

Scalar products and (pre)-Hilbert spaces

A normed vector space $(V, \|\cdot\|)$ is a pre-Hilbert space if its norm is defined by

$$\|v\| = \sqrt{\langle v, v\rangle}, \quad v \in V,$$

where $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{K}$ is a scalar product on V, i.e., it fulfills the properties

- $\langle v, v\rangle > 0$ for $v \neq 0$;
- $\langle v, w\rangle = \overline{\langle w, v\rangle}$, for $v, w \in V$;
- $\langle u + \lambda v, w\rangle = \langle u, w\rangle + \lambda\langle v, w\rangle$ for $u, v, w \in V$ and $\lambda \in \mathbb{K}$;
- $\langle w, u + \lambda v\rangle = \langle w, u\rangle + \overline{\lambda}\langle w, v\rangle$ for $u, v, w \in V$ and $\lambda \in \mathbb{K}$.

The triangle inequality for the norm follows from the Schwarz inequality

$$|\langle v, w\rangle| \leq \|v\|\|w\|, \quad v, w \in V.$$

Scalar products and (pre)-Hilbert spaces

We describe the pre-Hilbert space also by the pair $(V, \langle\cdot,\cdot\rangle)$. A typical example of scalar product over $\mathbb{K}^I$ is the one we introduced in (1), which generates the Euclidean norm on $\mathbb{K}^I$:

$$\|v\|_2 = \sqrt{\sum_{i \in I} |v_i|^2}, \quad v \in \mathbb{K}^I.$$

As before one can define orthogonality between vectors, orthogonal, and orthonormal sets of vectors.