

Foundations of Data Analysis (MA4800)

Massimo Fornasier



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

Slides of Lecture
April 26 2017

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 Scientists 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 appear¹, 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 $A_{ji}^T := A_{ij}$. 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 $A_{(j)} = (a_{ij})_{i \in I} = (A^T)^{(j)}$.

¹In the case $\mathbb{K} = \mathbb{R}$, then $\alpha = \bar{\alpha}$, 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

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

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

$$\langle x, y \rangle = y^H x = \sum_{i \in I} x_i \bar{y}_i, \quad (1)$$

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 columns of B , i.e., according to our terminology of

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

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 orthogonal 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 inverse 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

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

- ▶ (f) r is maximal with the property that there exists an invertible $r \times r$ submatrix of A ;
- ▶ (g) 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 $r_{max} = \min\{\#I, \#J\}$, 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 turns 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 $\|\cdot\| : V \rightarrow [0, \infty)$ 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 inequality:

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

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 $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$ 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 $\lambda \in \mathbb{K}$;
- ▶ $\langle w, u + \lambda v \rangle = \langle w, u \rangle + \bar{\lambda} \langle w, v \rangle$ 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.

Hilbert spaces

A Hilbert space is a pre-Hilbert space $(V, \langle \cdot, \cdot \rangle)$, whose topology induced by the associated norm $\| \cdot \|$ is complete, i.e., any Cauchy sequence in such vector spaces is convergent.

It is important to stress that every finite dimensional pre-Hilbert space $(V, \langle \cdot, \cdot \rangle)$ is actually complete, hence always a Hilbert space.

Thus, those who are not familiar with topological notions (e.g., completeness, Cauchy sequences, etc.), one should just reason in what follows according to their Euclidean geometric intuition.

Projections

Recall now that a set C in a vector space is convex if, for all $v, w \in C$ and all $t \in [0, 1]$, $tv + (1 - tw) \in C$.

Given a closed convex set C in a Hilbert space $(V, \langle \cdot, \cdot \rangle)$ one defines the projection of any vector v on C as

$$P_C(v) = \arg \min_{w \in C} \|v - w\|.$$

This definition is well-posed, as the projection is actually unique, and an equivalent definition is given by fulfilling the following inequality

$$\langle z - P_C(v), v - P_C(v) \rangle \leq 0,$$

for all $z \in C$. This is left as an exercise.

Projections onto subspaces: Pythagoras-Fourier Theorem

Of extreme importance for us are the orthogonal projections onto subspaces.

In case $C = W \subset V$ is actually a closed linear subspace of V , then the projection onto W can be readily computed as soon as one disposes of an orthonormal basis for W . Let $\{w_\nu\}_{\nu \in F}$ be a (countable) orthonormal basis for W then

$$P_W(v) = \sum_{\nu \in F} \langle v, w_\nu \rangle w_\nu, \text{ for all } v \in V. \quad (3)$$

Moreover, it holds the Pythagoras-Fourier Theorem:

$$\|P_W(v)\|^2 = \sum_{\nu \in F} |\langle v, w_\nu \rangle|^2, \text{ for all } v \in V.$$

Pythagoras-Fourier Theorem and orthonormal expansions

We will use very much this characterization of the orthogonal projections and the Pythagoras-Fourier Theorem, especially for the case where $W = V$.

In this case, obviously, $P_V = I$ and, we have the orthonormal expansion of any vector $v \in V$,

$$v = \sum_{\nu \in F} \langle v, w_\nu \rangle w_\nu,$$

and the norm equivalence

$$\|v\|^2 = \sum_{\nu \in F} |\langle v, w_\nu \rangle|^2.$$

Trace of a matrix

The effort of defining a abstract scalar products and norms allows us now to introduce several norms for matrices.

First of all we need to introduce the concept of trace, which is the map $\text{tr} : \mathbb{K}^{I \times I} \rightarrow \mathbb{K}$ defined by

$$\text{tr}(A) = \sum_{i \in I} A_{ii},$$

i.e., it is the sum of the diagonal elements of the matrix A .

Trace of a matrix: properties

The trace enjoys several properties, which we collect in the following:

Proposition (Properties of the trace)

- (a) $\text{tr}(AB) = \text{tr}(BA)$, for any $A \in \mathbb{K}^{I \times J}$ and $B \in \mathbb{K}^{J \times I}$;
- (b) $\text{tr}(ABC) = \text{tr}(BCA)$, for any A, B, C matrices of compatible size and indexes; this property is called the circularity property of the trace;
- (c) as a consequence of the previous property we obtain the invariance of the trace under unitary transformations, i.e., $\text{tr}(A) = \text{tr}(UAU^H)$ for $A \in \mathbb{K}^{I \times I}$ and any unitary matrix $U \in \mathbb{K}^{I \times I}$;
- (d) $\text{tr}(A) = \sum_{i \in I} \lambda_i$, where $\{\lambda_i : i \in I\}$ is the set of eigenvalues of A .

Matrix norms

The Frobenius norm of a matrix is essentially the Euclidean norm computed over the entries of the matrix (considered as a vector):

$$\|A\|_F := \sqrt{\sum_{i \in I, j \in J} A_{ij}^2}, \quad A \in \mathbb{K}^{I \times J}.$$

It is also known as Schur norm or Hilbert-Schmidt norm. This norm is generated by the scalar product (that's why we made the effort of introducing abstract scalar products!)

$$\langle A, B \rangle_F := \sum_{i \in I} \sum_{j \in J} A_{ij} \overline{B_{ij}} = \operatorname{tr}(AB^H) = \operatorname{tr}(B^H A). \quad (4)$$

In particular,

$$\|A\|_F^2 = \operatorname{tr}(AA^H) = \operatorname{tr}(A^H A) \quad (5)$$

holds.

Matrix norms

Let $\|\cdot\|_X$ and $\|\cdot\|_Y$ be vector norms on the vector spaces $X = K^I$ and $Y = K^J$, respectively. Then the associated matrix norm is

$$\|A\| := \|A\|_{X \rightarrow Y} := \sup_{y \neq 0} \frac{\|Ay\|_X}{\|y\|}, \quad A \in \mathbb{K}^{I \times J}.$$

If both $\|\cdot\|_X$ and $\|\cdot\|_Y$ coincides with the Euclidean norms $\|\cdot\|_2$ on $X = K^I$ and $Y = K^J$, respectively, then the associated matrix norm is called the spectral norm and it is denoted with $\|A\|_\infty$.

As the spectral norm has a central importance in many applications, it is often denoted just by $\|A\|$.

Unitary invariance and submultiplicativity

Both the matrix norms we introduced so far are invariant with respect to unitary transformations, i.e., $\|A\|_F = \|UAV^H\|_F$ and $\|A\|_2 = \|UAV^H\|_2$ for unitary matrices U, V .

Moreover, both are submultiplicative, i.e.,

$\|AB\|_F \leq \|A\|_2 \|B\|_F \leq \|A\|_F \|B\|_F$ and $\|AB\| \leq \|A\| \|B\|$, for matrices A, B of compatible sizes.