Mathematics of Data Science

Chapter II: Matrices

Notation: We work with $m \times n$ matrices $A = (A_{ij}) \in \mathbb{R}^{m \times n}$ over the reals or $A \in \mathbb{C}^{m \times n}$ over the complex numbers. We write $A \in \mathbb{K}^{m \times n}$ if $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

Let $A \in \mathbb{C}^{m \times n}$:

transpose: A^T $\qquad (A^T)_{i,j} = A_{j,i}$

adjoint (Hermitian transpose): A^* $\qquad (A^*)_{i,j} = \overline{A_{j,i}}$

columns: $A_{(:j)} = (A_{ij})_{i=1}^{m} \in \mathbb{C}^m$

rows: $A^{(i:)} = (A_{ij})_{j=1}^{n} \in \mathbb{C}^n$

range: $\operatorname{ran}(A) = \{ Ax : x \in \mathbb{K}^n \}$

$$= \operatorname{span}\{ A_{(:j)} : j = 1, \dots, n \} \subset \mathbb{K}^m$$

identity matrix: $I \in \mathbb{R}^{n \times n}$

Euclidean scalar product: for $x, y \in \mathbb{R}^m$

$$\langle x, y \rangle = \sum_{i=1}^{m} x_i \cdot y_i$$

complex scalar product: for $x, y \in \mathbb{C}^n$

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \overline{y_i}$$

x, y are called orthogonal if

$$\langle x, y \rangle = 0$$

A basis $x^1, \ldots, x^n$ of $\mathbb{K}^n$ is called orthonormal if $\langle x^i, x^j \rangle = \delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$

$A \in \mathbb{R}^{n \times n}$ is called orthogonal if

$A^T A = I$ or equivalently if $AA^T = I$.

$A \in \mathbb{C}^{n \times n}$ is called unitary if

$A^* A = I$ or equiv. if $AA^* = I$

In this case the inverse satisfies $A^{-1} = A^*$.

$\operatorname{rank} A = \dim \operatorname{ran} A = \dim \operatorname{ran} (A^T)$

$A \in \mathbb{R}^{m \times n}$ has full rank if $\operatorname{rank} A = \min\{m, n\}$.

Eigenvalues and eigenvectors

For $A \in \mathbb{C}^{m \times m}$, $\lambda \in \mathbb{C}$ is called an eigenvalue of A with corresponding eigenvector $v \in \mathbb{C}^m \setminus \{0\}$ if

$$Av = \lambda v.$$

Eigenvalues are the roots of the characteristic polynomial $\chi_A(\lambda) = \det(A - \lambda I)$, i.e. all $\lambda_i \in \mathbb{C}$ such that $\chi_A(\lambda_i) = 0$.

If $A = A^*$ is Hermitian, then all eigenvalues are real and there exists an orthonormal basis $v_1, \dots, v_n \in \mathbb{C}^m$ of eigenvectors. With $V = (v_1 | \cdots | v_n) \in \mathbb{C}^{m \times m}$ (unitary) we can then write

$$A = V D V^* \quad \text{with } D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

$$= \sum_{j=1}^{n} \lambda_j v_j v_j^*$$

Norms:

Definition 2.1: For a vector space V over $\mathbb{K}$, a norm $\|\cdot\| : V \to \mathbb{R}_+ = \{x \in \mathbb{R}, x \geq 0\}$ is a function satisfying

(i) $\|v\| = 0$ if and only if $v = 0$

(ii) $\|\lambda v\| = |\lambda| \, \|v\|$ for all $v \in V$ and $\lambda \in \mathbb{K}$.

(iii) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$
(triangle inequality)

If (i) is weakened to $\|v\| = 0$ if $v = 0$, then $\|\cdot\|$ is called a semi-norm.

Examples:

a) ℓ_p-norm on $\mathbb{C}^m$: For $1 \leq p < \infty$

$$\|x\|_p := \left(\sum_{j=1}^{m} |x_j|^p \right)^{1/p}, \quad x \in \mathbb{C}^m,$$

$$\|x\|_\infty := \max_{j=1,\dots,m} |x_j|$$

are norms on $\mathbb{C}^m$.

Special case ℓ_2-norm is Euclidean norm,

$$\|x\|_2 = \sqrt{\langle x, x\rangle}.$$

b) sup-norm on vector space $C[0,1]$: $\{f : [0,1] \to \mathbb{R},\ f \text{ continuous}\}$

$$\|f\|_\infty := \sup_{t\in[0,1]} |f(t)|$$

c) L^p-norm on $C[0,1]$:

$$\|f\|_p := \left(\int_0^1 |f(t)|^p \, dt\right)^{1/p}$$

Trace

Trace: $\operatorname{tr}: \mathbb{C}^{m\times m} \longrightarrow \mathbb{C}$

$$\operatorname{tr}(A) = \sum_{i=1}^{m} A_{ii} \qquad A \in \mathbb{C}^{m\times m}$$

Proposition 2.2: (Properties of the trace)

a) cyclicity: $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for all $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}$

Proof: calculation!

b) invariance under unitary conjugation

$$\operatorname{tr}(UAU^*) = \operatorname{tr}(AU^*U) = \operatorname{tr}(A)$$

c) $\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A (counted with algebraic multiplicities)

Proof for Hermitian matrix $A = A^*$:

Eigenvalue decomposition $A = VDV^*$ with V unitary and $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

$$\operatorname{tr}(A) = \operatorname{tr}(VDV^*) = \operatorname{tr}(DV^*V) = \operatorname{tr}(D)$$
$$= \sum_{i=1}^{n} \lambda_i \qquad \square$$

Frobenius scalar product: For $A, B \in \mathbb{C}^{m \times n}$

$$\langle A, B \rangle_F := \operatorname{Tr}(AB^*) = \operatorname{Tr}(B^*A)$$
$$= \sum_{i=1}^{m} (AB^*)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \overline{B_{ij}}$$

Frobenius norm (Hilbert-Schmidt-norm):

$$\|A\|_F = \sqrt{\langle A, A \rangle_F} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2}$$
$$= \sqrt{\operatorname{Tr}(A^*A)}$$

Euclidean norm when $\mathbb{R}^{m \times n}$ is identified with $\mathbb{R}^{m \cdot n}$.

It holds

$$\|A\|_F = \sqrt{\operatorname{Tr}(A^*A)} = \sqrt{\sum_{i=1}^{m} \lambda_i(A^*A)}$$

Since A^*A is positive semidefinite, the eigenvalues of A^*A satisfy $\lambda_i(A^*A) \geq 0$.

Recall: A Hermitian matrix $A = A^* \in \mathbb{C}^{m\times m}$ is called positive semidefinite if

$$x^*Ax \geq 0 \quad \text{for all } x \in \mathbb{C}^m.$$

It is called positive definite if

$$x^*Ax > 0 \quad \text{for all } x \in \mathbb{C}^m \setminus \{0\}.$$

Fact: A Hermitian matrix $A = A^* \in \mathbb{C}^{m\times m}$ is positive (semi)definite if and only if $\lambda_i(A) \geq 0$ $(\lambda_i > 0)$ for all $i = 1, \dots, m$.

left out in lecture!

Operator norm:

For $X = (\mathbb{R}^n, \|\cdot\|_X)$, $Y = (\mathbb{R}^m, \|\cdot\|_Y)$

the operator norm of $A \in \mathbb{R}^{m \times n}$ $(A : X \to Y)$

is defined as

$$\|A\|_{X \to Y} = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_Y}{\|x\|_X}$$

$$= \sup_{\substack{x \in \mathbb{R}^n \\ \|x\|_X = 1}} \|Ax\|_Y$$

Example: $\|A\|_{\ell^2 \to \ell^2} = \max_j \sqrt{\lambda_j (A^* A)}$

(spectral norm)

Unitary invariance: For all $A \in \mathbb{C}^{m \times n}$ and unitary matrices $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$, it holds

$$\|UAV\|_{\ell^2 \to \ell^2} = \|A\|_{\ell^2 \to \ell^2}$$

$$\|UAV\|_F = \|A\|_F$$

exercise