I. Linear Algebra

I.11. Matrix Norms

Lecture based on

https://github.com/gwthomas/math4ml (Garrett Thomas, 2018)

https://mml-book.github.io/ (Deisenroth et al., 2020, Mathematics for Machine Learning)

Prof. Dr. Christoph Lippert

Digital Health & Machine Learning

Matrix Norms

Let $\mathbb{R}^{m\times n}$ denote the **vector space** of all matrices of size $m \times n$ (with m rows and n columns) with entries in $\mathbb{R}$.

A matrix norm is a norm on the vector space $\mathbb{R}^{m\times n}$.

Thus, the matrix norm is a function $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ that must satisfy the following properties:

For all scalars α in $\mathbb{R}$ and for all matrices $\mathbf{A}$ and $\mathbf{B}$ in $\mathbb{R}^{m\times n}$,

i. $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$ (being "absolutely homogeneous")
ii. $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ (being "sub-additive" or satisfying the "triangle inequality")
iii. $\|\mathbf{A}\| \geq 0$ (being "positive-valued")
iv. $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{0}_{m,n}$ (being "definite")

Additionally, some (not all) matrix norms satisfy the following condition:

- $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$ for all matrices $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{B} \in \mathbb{R}^{m\times p}$.

If V and W are vector spaces, then the set of linear maps $T : V \to W$ forms another vector space.

The norms $\|\cdot\|_W$ on W and $\|\cdot\|_V$ on V induce the **operator norm**

$$\|T\|_{\mathrm{op}} = \max_{\substack{\mathbf{x} \in V \\ \mathbf{x} \neq 0}} \frac{\|T(\mathbf{x})\|_W}{\|\mathbf{x}\|_V}$$

If $\mathbf{A} = \mathbb{R}^{m \times n}$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, then the p-norm defines the **matrix p-norm**:

$$\|\mathbf{A}\|_p = \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

For the special cases $p = 1, 2, \infty$, we have

- $\|\mathbf{A}\|_1 = \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_1}{\|\mathbf{x}\|_1} = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |A_{ij}|$
- $\|\mathbf{A}\|_\infty = \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |A_{ij}|$
- $\|\mathbf{A}\|_2 = \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \sigma_1(\mathbf{A})$

These norms have identical values for the Identity matrix

$$\|\mathbf{I}\|_1 = 1 \quad \|\mathbf{I}\|_\infty = 1 \quad \|\mathbf{I}\|_2 = 1$$

By definition, $\|\mathbf{Ax}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{x}\|_p$.

Proposition (The matrix p-norm is submultiplicative.)

$$\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \, \|\mathbf{B}\|_p$$

A matrix norm is **unitary invariant**, if

$$\|\mathbf{Q}_1 \mathbf{A} \mathbf{Q}_2^\top\| = \|\mathbf{A}\|,$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are orhogonal matrices.

- **Spectral norm** $\|\mathbf{A}\|_2 = \sigma_1$
- **Frobenius norm** $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |A_{ij}|^2} = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$
- **Nuclear norm** $\|\mathbf{A}\|_N = \sigma_1 + \cdots + \sigma_r$

These norms have different values for the Identity matrix

$$\|\mathbf{I}\|_2 = 1 \quad \|\mathbf{I}\|_F = \sqrt{n} \quad \|\mathbf{I}\|_N = n$$

Also for an orthogonal matrix $\mathbf{Q}$

$$\|\mathbf{Q}\|_2 = 1 \quad \|\mathbf{Q}\|_F = \sqrt{n} \quad \|\mathbf{Q}\|_N = n$$

Theorem (Eckart-Young Theorem)

Let $\|\dot{}\|$ *be unitary invariant and*

$$\mathbf{A}_k = \mathbf{U_k}\mathbf{\Sigma_k}\mathbf{V_k}^\top = \sigma_1\mathbf{u}_1\mathbf{v}_1^\top + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^\top$$

then, $\|\mathbf{A} - \mathbf{A}_k\| \leq \|\mathbf{A} - \mathbf{B}\|$ *for any* $\mathbf{B} \in \mathbb{R}^{m\times n}$ *with* $\mathrm{rank}(\mathbf{B}) \leq k$.