NORMS for MATRICES

Prof. Dan A. Simovici

UMB

Matrices having the same format can be added and matrices can be multiplied by a scalar. Therefore, the sets of matrices $\mathbb{R}^{m \times n}$ or $\mathbb{R}^{m \times n}$ can be regarded as linear spaces (over $\mathbb{R}$, or $\mathbb{R}$, respectively).
Norms can be introduces over matrices adopting one of the following points of view:

- a matrix in $\mathbb{R}^{m \times n}$ can be regarded as a real vector with mn components (vectorial norms), or
- a matrix $A \in \mathbb{R}^{m \times n}$ is a transformation h_A of $\mathbb{R}^n$ into $\mathbb{R}^m$ defined by $h_A(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$ (operatorial norms).

Matrix Norms

Definition

A *consistent family of matrix norms* is a family of functions $\mu^{(m,n)} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}_{\geqslant 0}$, where $m, n \in \mathbb{N}$, $m, n \geqslant 1$, that satisfies the following conditions:

- $\mu^{(m,n)}$ is a norm on $\mathbb{R}^{(m,n)}$ for $m, n \in \mathbb{N}$, $m, n \geqslant 1$;
- for every matrix $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ we have

  $$\mu^{(m,p)}(AB) \leqslant \mu^{(m,n)}(A)\mu^{(n,p)}(B)$$

  (the submultiplicative property).

If the format of the matrix A is clear from context or is irrelevant, then we write $\mu(A)$ instead of $\mu^{(m,n)}(A)$.

Example

Consider the vectorial matrix norm μ_1 induced by the vector norm $\| \cdot \|_1$, where $\| \mathbf{x} \|_1 = \sum_{i=1}^{n} |x_i|$. We have $\mu_1(A) = \sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|$ for $A \in \mathbb{R}^{m\times n}$. $\mu_1(A)$ is denoted by $\| A \|_1$.

Actually, this is a matrix norm. Indeed, for $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{p\times n}$ we have:

$$\begin{aligned}
\mu_1(AB) &= \sum_{i=1}^{m} \sum_{j=1}^{n} \left| \sum_{k=1}^{p} a_{ik} b_{kj} \right| \leqslant \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} |a_{ik} b_{kj}| \\
&\leqslant \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k'=1}^{p} \sum_{k''=1}^{p} |a_{ik'}||b_{k''j}| \\
& \quad \text{(because we added extra non-negative terms to the sums)} \\
&= \left(\sum_{i=1}^{m} \sum_{k'=1}^{p} |a_{ik'}| \right) \cdot \left(\sum_{j=1}^{n} \sum_{k''=1}^{p} |b_{k''j}| \right) = \mu_1(A)\mu_1(B).
\end{aligned}$$

Example

The vectorial matrix norm μ_2 induced by the vector norm $\| \cdot \|_2$ is also a matrix norm. Indeed, using the same notations we have:

$$
\begin{aligned}
(\mu_2(AB))^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n}\left|\sum_{k=1}^{p} a_{ik}b_{kj}\right|^2 \leqslant \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{k=1}^{p}|a_{ik}|^2\right)\left(\sum_{l=1}^{p}|b_{lj}|^2\right) \\
& \text{(by Cauchy-Schwarz inequality)} \\
&\leq (\mu_2(A))^2(\mu_2(B))^2.
\end{aligned}
$$

We have $\mu_2(A) = \left(\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|^2\right)^{\frac{1}{2}}$. This norm is usually denoted by $\| A \|_F$ and is known as the Frobenius norm.
If $A \in \mathbb{R}^{n\times n}$, $\| A \|_F^2 = traceA'A = traceAA'$.

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ is orthogonal if $A'A = AA' = I_n$.

Example

The matrix

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$$

is orthogonal.

Theorem

If U is an orthogonal matrix, then $\| U\mathbf{x} \|_2 = \| \mathbf{x} \|_2$ (that is, the length of any vector is invariant under multiplication by U).

Proof If U is orthogonal we have

$$\| U\mathbf{x} \|_2^2 = (U\mathbf{x})' U\mathbf{x} = \mathbf{x}' U' U\mathbf{x} = \| \mathbf{x} \|_2^2$$

because $U'U = I_n$. Thus, $\| U\mathbf{x} \|_2 = \| \mathbf{x} \|_2$.

Properties of Orthogonal Matrices

- the inverse of an orthogonal matrix is $U^{-1} = U'$;
- $\det(U) \in \{-1, 1\}$;
- the Householder matrix $U = I_n - 2\mathbf{v}\mathbf{v}'$ (where $\mathbf{v}$ is a unit vector) is orthogonal.

Frobenius Norm is Invariant

Let U be an orthogonal matrix. We have

$$\parallel UA \parallel_F = \sqrt{trace((UA)'UA)} = \sqrt{trace(A'U'UA)} = \sqrt{trace(A'A)} = \parallel A \parallel_F .$$

Definition

Let ν_m be a norm on $\mathbb{R}^m$ and ν_n be a norm on $\mathbb{R}^n$ and let $A \in \mathbb{R}^{n\times m}$ be a matrix. The operator norm of A is the number $\mu^{(n,m)}(A) = \mu^{(n,m)}(h_A)$, where $\mu^{(n,m)} = N(\nu_m, \nu_n)$.

Theorem

Let $\{\nu_n \mid n \geqslant 1\}$ be a family of vector norms, where ν_n is a vector norm on $\mathbb{R}^n$. The family of norms $\{\mu^{(n,m)} \mid n, m \geqslant 1\}$ is consistent.

Proof:

$$
\begin{aligned}
\mu^{(n,p)}(AB) &= \sup\{\nu_n((AB)\mathbf{x}) \mid \nu_p(\mathbf{x}) \leqslant 1\} \\
&= \sup\{\nu_n(A(B\mathbf{x})) \mid \nu_p(\mathbf{x}) \leqslant 1\} \\
&= \sup\left\{\nu_n\left(A\frac{B\mathbf{x}}{\nu_m(B\mathbf{x})}\right)\nu_m(B\mathbf{x}) \middle| \nu_p(\mathbf{x}) \leqslant 1\right\} \\
&\leqslant \mu^{(n,m)}(A)\sup\{\nu_m(B\mathbf{x}) \big| \nu_p(\mathbf{x}) \leqslant 1\} \\
&\quad \left(\text{because } \nu_m\left(\frac{B\mathbf{x}}{\nu(B\mathbf{x})}\right) = 1\right) \\
&= \mu^{(n,m)}(A)\mu^{(m,p)}(B).
\end{aligned}
$$

Theorem

Let ν_n be a norm on $\mathbb{R}^n$ for $n \geqslant 1$. The following equalities hold for $\mu^{(n,m)}(A)$, where $A \in \mathbb{R}^{(n,m)}$.

$$\begin{aligned}
\mu^{(n,m)}(A) &= \inf\{M \in \mathbb{R}_{\geqslant 0} \mid \nu_n(A\mathbf{x}) \leq M\nu_m(\mathbf{x}) \text{ for every } \mathbf{x} \in \mathbb{R}^m\} \\
&= \sup\{\nu_n(A\mathbf{x}) \mid \nu_m(\mathbf{x}) \leqslant 1\} = \max\{\nu_n(A\mathbf{x}) \mid \nu_m(\mathbf{x}) \leqslant 1\} \\
&= \max\{\nu'(f(\mathbf{x})) \mid \nu(\mathbf{x}) = 1\} \\
&= \sup\left\{\frac{\nu'(f(\mathbf{x}))}{\nu(\mathbf{x})} \mid \mathbf{x} \in \mathbb{R}^m - \{\mathbf{0}_m\}\right\}.
\end{aligned}$$

The operator matrix norm induced by the vector norm $\|\cdot\|_p$ is denoted by $|||\cdot|||_p$.

The norm $\|A\|_1$

Example

Let the columns of A be $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and let $\mathbf{x} \in \mathbb{R}^n$ be a vector whose components are $x_1, \ldots, x_n$. Then, $\mathbf{Ax} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$, so

$$\begin{aligned} \| A\mathbf{x} \|_1 &= \| x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \|_1 \leqslant \sum_{j=1}^{n} |x_j| \, \| \mathbf{a}_j \|_1 \\ &\leqslant \max_j \| \mathbf{a}_j \|_1 \sum_{j=1}^{n} |x_j| = \max_j \| \mathbf{a}_j \|_1 \cdot \| \mathbf{x} \|_1 . \end{aligned}$$

Thus, $\|A\|_1 \leqslant \max_j \| \mathbf{a}_j \|_1$.

The norm $\|A\|_1$ (cont'd)

Example

Let $\mathbf{e}_j$ be the vector whose components are 0 with the exception of its j^{th} component that is equal to 1. Clearly, we have $\| \mathbf{e}_j \|_1 = 1$ and $\mathbf{a}_j = A\mathbf{e}_j$. This, in turn implies $\| \mathbf{a}_j \|_1 = \| A\mathbf{e}_j \|_1 \leqslant \|A\|_1$ for $1 \leqslant j \leqslant n$. Therefore, $\max_j \| \mathbf{a}_j \|_1 \leq \|A\|_1$, so

$$\|A\|_1 = \max_j \| \mathbf{a}_j \|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|.$$

In other words, $\|A\|_1$ equals the maximum column sum of the absolute values.

Example

Let $D = \text{diag}(d_1, \ldots, d_n) \in \mathbb{R}^{n\times n}$ be a diagonal matrix. If $\mathbf{x} \in \mathbb{R}^n$ we have

$$D\mathbf{x} = \begin{pmatrix} d_1x_1 \\ \vdots \\ d_nx_n \end{pmatrix},$$

so

$$\begin{aligned} \|D\|_2 &= \max\{\| D\mathbf{x} \|_2 \mid \| \mathbf{x} \|_2 = 1\} \\ &= \max\{\sqrt{(d_1x_1)^2 + \cdots + (d_nx_n)^2} \mid x_1^2 + \cdots + x_n^2 = 1\} \\ &= \max\{|d_i| \mid 1 \leqslant 1 \leqslant n\}. \end{aligned}$$

Norm $\|\cdot\|_2$ is invariant

For an orthogonal matrix U we have:

$$\begin{aligned} \|UA\|_2 &= \max\{\| (UA)\mathbf{x} \|_2 \mid \| \mathbf{x} \|_2 = 1\} = \max\{\| U(A\mathbf{x}) \|_2 \mid \| \mathbf{x} \|_2 = 1\} \\ &= \max\{\| A\mathbf{x} \|_2 \mid \| \mathbf{x} \|_2 = 1\} = \|A\|_2. \end{aligned}$$