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 introduced 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 $h_A(x) = Ax$ for $x \in \mathbb{R}^n$ (operatorial norms).
Definition

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

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

\[
\mu^{(m,p)}(AB) \leq \mu^{(m,n)}(A)\mu^{(n,p)}(B)
\]

(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 $\mu_1$ induced by the vector norm $\| \cdot \|_1$, where $\| x \|_1 = \sum_{i=1}^{n} |x_i|$. 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:

$$
\mu_1(AB) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left| \sum_{k=1}^{p} a_{ik} b_{kj} \right| 
\leq \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} |a_{ik} b_{kj}|
\leq \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k'=1}^{p} \sum_{k''=1}^{p} |a_{ik'}| |b_{k''j}|
\text{(because we added extra non-negative terms to the sums)}
$$

$$
= \left( \sum_{i=1}^{m} \sum_{k'=1}^{p} |a_{ik'}| \right) \cdot \left( \sum_{j=1}^{n} \sum_{k''=1}^{p} |b_{k''j}| \right) = \mu_1(A)\mu_1(B).
$$
Example

The vectorial matrix norm $\mu_2$ induced by the vector norm $\| \cdot \|_2$ is also a matrix norm. Indeed, using the same notations we have:

$$(\mu_2(AB))^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \left| \sum_{k=1}^{p} a_{ik} b_{kj} \right|^2 \leq \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \sum_{k=1}^{p} |a_{ik}|^2 \right) \left( \sum_{l=1}^{p} |b_{lj}|^2 \right) \tag{by Cauchy-Schwarz inequality}$$

$$\leq (\mu_2(A))^2 (\mu_2(B))^2.$$

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 $A \in \mathbb{R}^{n \times n}$, $\| A \|_F = \text{trace}A' A = \text{trace}AA'$. 
Definition

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

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 $\| Ux \|_2 = \| x \|_2$ (that is, the length of any vector is invariant under multiplication by $U$).

Proof If $U$ is orthogonal we have

$$\| Ux \|_2^2 = (Ux)'Ux = x'U'Ux = \| x \|_2^2$$

because $U'U = I_n$. Thus, $\| Ux \|_2 = \| x \|_2$. 
Properties of Orthogonal Matrices

- the inverse of an orthogonal matrix is $U^{-1} = U'$;
- $\det(U) \in \{-1, 1\}$;
- the Householder matrix $U = I_n - 2vv'$ (where $v$ is a unit vector) is orthogonal.
Let $U$ be an orthogonal matrix. We have
\[
\| UA \|_F = \sqrt{\text{trace}((UA)'UA)} = \sqrt{\text{trace}(A'U'UA)} = \sqrt{\text{trace}(A'A)} = \| A \|_F.
\]
Definition

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

Proof:

\[
\mu^{(n,p)}(AB) = \sup \{ \nu_n((AB)x) \mid \nu_p(x) \leq 1 \}
\]

\[
= \sup \{ \nu_n(A(Bx)) \mid \nu_p(x) \leq 1 \}
\]

\[
= \sup \left\{ \nu_n \left( A \frac{Bx}{\nu_m(Bx)} \right) \nu_m(Bx) \middle| \nu_p(x) \leq 1 \right\}
\]

\[
\leq \mu^{(n,m)}(A) \sup \{ \nu_m(Bx) \middle| \nu_p(x) \leq 1 \}
\]

(because \( \nu_m \left( \frac{Bx}{\nu(Bx)} \right) = 1 \))

\[
= \mu^{(n,m)}(A) \mu^{(m,p)}(B).
\]
Theorem

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

$$
\mu^{(n,m)}(A) = \inf \{ M \in \mathbb{R}_{\geq 0} \mid \nu_n(Ax) \leq M\nu_m(x) \text{ for every } x \in \mathbb{R}^m \}
$$

$$
= \sup \{ \nu_n(Ax) \mid \nu_m(x) \leq 1 \} = \max \{ \nu_n(Ax) \mid \nu_m(x) \leq 1 \}
$$

$$
= \max \{ \nu'(f(x)) \mid \nu(x) = 1 \}
$$

$$
= \sup \left\{ \frac{\nu'(f(x))}{\nu(x)} \mid x \in \mathbb{R}^m - \{0_m\} \right\}.
$$

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 $a_1, \ldots, a_n$ and let $x \in \mathbb{R}^n$ be a vector whose components are $x_1, \ldots, x_n$. Then, $Ax = x_1a_1 + \cdots + x_na_n$, so

$$\|Ax\|_1 = \|x_1a_1 + \cdots + x_na_n\|_1 \leq \sum_{j=1}^{n} |x_j| \|a_j\|_1 \leq \max_j \|a_j\|_1 \sum_{j=1}^{n} |x_j| = \max_j \|a_j\|_1 \cdot \|x\|_1.$$

Thus, $\|A\|_1 \leq \max_j \|a_j\|_1$. 
Example

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

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

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 $x \in \mathbb{R}^n$ we have

$$Dx = \begin{pmatrix} d_1x_1 \\ \vdots \\ d_nx_n \end{pmatrix},$$

so

$$\|D\|_2 = \max\left\{ \|Dx\|_2 \mid \|x\|_2 = 1 \right\}$$

$$= \max\left\{ \sqrt{(d_1x_1)^2 + \cdots + (d_nx_n)^2} \mid x_1^2 + \cdots + x_n^2 = 1 \right\}$$

$$= \max\{ |d_i| \mid 1 \leq i \leq n \}.$$
Operatorial Norms

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

For an orthogonal matrix $U$ we have:

$$\| UA \|_2 = \max \{ \| (UA)x \|_2 | \| x \|_2 = 1 \} = \max \{ \| U(Ax) \|_2 | \| x \|_2 = 1 \} = \max \{ \| Ax \|_2 | \| x \|_2 = 1 \} = A.\|_2.$$