Lecture 3 - *Norms*

OBJECTIVE:
The notions of size and distance in a vector space are described by norms.
e.g., approximations and convergence are measured by norms.

◇ VECTOR NORMS

Definition 1. *A* norm *is a function* $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ *that assigns a length to a vector.*

For this to make sense, a norm must satisfy the following 3 conditions for all vectors $\mathbf{x}$, $\mathbf{y}$ and scalars α:

1. $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$

2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality)

3. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$

What do these conditions say?

1. the length of a non-zero vector must be positive

2. the length of a vector sum cannot exceed the sum of the individual vector lengths

3. scaling a vector scales its length by the same amount

The most important vector norm in this course is the (unweighted) 2-norm,

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{m} x_i^2 \right)^{\frac{1}{2}} = \sqrt{\mathbf{x}^{\mathbf{T}}\mathbf{x}}$$

But, we will also use

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{m} |x_i|$$

and

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le m} |x_i|$$

◇ MATRIX NORMS INDUCED BY VECTOR NORMS

You could view an $m \times n$ matrix as an mn-dimensional vector and use vector norms to define norms of matrices.

It turns out that there are other special vector norms that can be defined for matrices that are more useful. These are called *induced matrix norms* because the norms are intended for matrices, but they only use vector norms in their definition.

An induced matrix norm is defined in terms of the behaviour of a matrix as an operator between its normed domain and range spaces.

Specifically, let $\|\cdot\|_{(n)}$ be a norm on the domain of $\mathbf{A}$ (since $\mathbf{x} \in \mathbb{R}^n$). Let $\|\cdot\|_{(m)}$ be a norm on the range of $\mathbf{A}$ (since $\mathbf{Ax} \in \mathbb{R}^m$).

Then, the induced matrix norm $\|\mathbf{A}\|_{(m,n)}$ is the smallest number C such that

$$\|\mathbf{Ax}\|_{(m)} \leq C\|\mathbf{x}\|_{(n)} \qquad \text{for all } \mathbf{x} \in \mathbb{R}^n$$

In other words, $\|\mathbf{A}\|_{(m,n)}$ is the maximum value of $\frac{\|\mathbf{Ax}\|_{(m)}}{\|\mathbf{x}\|_{(n)}}$

i.e., the maximum factor by which $\mathbf{A}$ can stretch any vector $\mathbf{x}$.

We say $\|\cdot\|_{(m,n)}$ is the matrix norm induced by the (vector) norms $\|\cdot\|_{(m)}$, $\|\cdot\|_{(n)}$.

Since "stretching factors" are independent of the size of the thing they are stretching, it is equivalent (and convenient!) to defined the induced matrix norm as

$$\|A\|_{(m,n)} = \max_{\substack{\mathbf{x}\in\mathbb{R}^n \\ \mathbf{x}\neq\mathbf{0}}} \frac{\|\mathbf{Ax}\|_{(m)}}{\|\mathbf{x}\|_{(n)}} = \max_{\substack{\mathbf{x}\in\mathbb{R}^n \\ \|\mathbf{x}\|_{(n)}=1}} \|\mathbf{Ax}\|_{(m)}$$

Example 3.1 Consider $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$, which maps $\mathbb{R}^2$ to $\mathbb{R}^2$ (it is convenient to draw pictures!)

matlab demo `Lecture3demo.m`

Example 3.2 NORMS OF A DIAGONAL MATRIX
Let $\mathbf{D}$ be a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \end{bmatrix}$$

$$\|\mathbf{D}\|_2 = \max_{1 \le i \le m} |d_i|, \text{ but same is true for } \|\mathbf{D}\|_1, \|\mathbf{D}\|_\infty$$

Example 3.3 THE 1-NORM OF A MATRIX
If $\mathbf{A} \in \mathbb{R}^{m \times n}$, then

$$\|\mathbf{A}\|_1 = \max\left(\text{absolute column sum}\right) = \max_{1 \le j \le n} \|a_j\|_1$$

Example 3.4 THE ∞-NORM OF A MATRIX
If $\mathbf{A} \in \mathbb{R}^{m \times n}$, then

$$\|\mathbf{A}\|_\infty = \max\left(\text{absolute row sum}\right) = \max_{1 \le i \le m} \|a_i^T\|_1$$

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \end{bmatrix}$.

Then $\|\mathbf{A}\|_1 = 9$, $\|\mathbf{A}\|_\infty = 15$

$\diamond$ CAUCHY-SCHWARTZ INEQUALITY

A useful result to bound the size of inner products: for any vectors $\mathbf{x}$,$\mathbf{y}$

$$|\mathbf{x}^T \mathbf{y}| \le \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

Example 3.5 THE 2-NORM OF A ROW VECTOR

Consider a matrix $\mathbf{A}$ comprised of a single row; i.e., $\mathbf{A} = \mathbf{a}^T$, where $\mathbf{a}$ is a vector.

The Cauchy-Schwartz inequality allows us to find $\|\mathbf{A}\|_2$ in this case.

For any $\mathbf{x}$, we have $\|\mathbf{A}\mathbf{x}\|_2 = |\mathbf{a}^T\mathbf{x}| \leq \|\mathbf{a}\|_2\|\mathbf{x}\|_2$

$$\therefore \qquad \|\mathbf{A}\|_2 = \max \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \|\mathbf{a}\|_2$$

Example 3.6 THE 2-NORM OF AN OUTER PRODUCT

Let $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^n$, and $\mathbf{A} = \mathbf{u}\mathbf{v}^T$. (This is called a rank-one outer product.)

Now,

$$\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{u}\mathbf{v}^T\mathbf{x}\|_2 = \|\mathbf{u}\|_2|\mathbf{v}^T\mathbf{x}| \leq \|\mathbf{u}\|_2\|\mathbf{v}\|_2\|\mathbf{x}\|_2$$

$$\therefore \qquad \|\mathbf{A}\|_2 = \|\mathbf{u}\|_2\|\mathbf{v}\|_2$$

$\diamond$ BOUNDING $\|\mathbf{AB}\|$ IN AN INDUCED MATRIX NORM

Let $\|\cdot\|_{(l)}$, $\|\cdot\|_{(m)}$, $\|\cdot\|_{(n)}$ be norms on $\mathbb{R}^l$, $\mathbb{R}^m$, $\mathbb{R}^n$.

Let $\mathbf{A} \in \mathbb{R}^{l\times m}$, $\mathbf{B} \in \mathbb{R}^{m\times n}$.

Then, for all $\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned} \|\mathbf{ABx}\|_{(l)} &\leq \|\mathbf{A}\|_{(l,m)}\|\mathbf{Bx}\|_{(m)} \\ &\leq \|\mathbf{A}\|_{(l,m)}\|\mathbf{B}\|_{(m,n)}\|\mathbf{x}\|_{(n)} \end{aligned}$$

(Treat $\mathbf{Bx}$ as a vector and use definition of induced matrix norm.)

$$\therefore \qquad \|\mathbf{AB}\|_{(l,n)} \leq \|\mathbf{A}\|_{(l,m)}\|\mathbf{B}\|_{(m,n)}$$

$\diamond$ GENERAL MATRIX NORMS

In general, matrix norms do not always have to be induced by vector norms
$\rightarrow$ a matrix norm must merely satisfy the 3 conditions of a vector norm applied in the mn-dimensional vector space of matrices.

i.e.,

1. $\|\mathbf{A}\| \geq 0$, and $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{0}$

2. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$

3. $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$

The most important matrix norm (not induced by a vector norm) is the *Frobenius norm* (also called the Hilbert-Schmidt norm):

$$\|\mathbf{A}\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{\frac{1}{2}}$$

i.e., the square root of the sum of the squares of all the elements
$\leftrightarrow$ the 2-norm of the matrix viewed as an mn-dimensional vector.

Another useful way to write $\|\mathbf{A}\|_F$ is in terms of 2-norms of the rows or columns. e.g.,

$$\|\mathbf{A}\|_F = \left(\sum_{j=1}^{n} \|a_j\|_2^2 \right)^{\frac{1}{2}}$$

This can be expressed compactly as

$$\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}(\mathbf{A}^T\mathbf{A})} = \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}^T)}$$

where tr$(\cdot)$ is the *trace* of a matrix: the sum of its diagonal elements.

It can also be shown (p.23) that

$$\|\mathbf{AB}\|_F^2 \le \|\mathbf{A}\|_F^2 \|\mathbf{B}\|_F^2$$

$\diamond$ INVARIANCE UNDER ORTHOGONAL MULTIPLICATION

Like the vector 2-norm, *the matrix 2-norm is invariant under multiplication by orthogonal matrices*.

The same property holds for the Frobenius norm.

Theorem 1. *For any $\mathbf{A} \in \mathbb{R}^{m\times n}$ and orthogonal $\mathbf{Q} \in \mathbb{R}^{m\times m}$, we have*

$$\|\mathbf{QA}\|_2 = \|\mathbf{A}\|_2 \qquad \textit{and} \qquad \|\mathbf{QA}\|_F = \|\mathbf{A}\|_F$$

Note 1. *This theorem still holds if $\mathbf{Q} \in \mathbb{R}^{p\times m}$ with $p > m$.*
i.e., $\mathbf{Q}$ is made up of m orthonormal columns of size p. The results also hold for multiplication on the right by $\mathbf{Q}$; i.e., for matrices with orthonormal rows.