Vector and Matrix Norms

- We are familiar with the norm of a vector \( x \in \mathbb{R}^n \):

\[
\|x\| = \|x\|_2 = \sqrt{x^Tx} = \sqrt{\sum_{i=1}^{n} x_i^2}.
\]

- This is also called the 2-norm, the standard norm, the Euclidean Norm, the magnitude, or the Euclidean length, of \( x \).

- But it is useful to generalize the concept. Henceforth we will think of \( \|x\|_2 \) as only one of infinitely many norms.

- This is a standard generalization technique in mathematics:
  - Decide and list what properties are important.
  - Find other objects with those properties.

- So what are the key properties of a norm?
• **Definition:** A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that associates a number $\|x\|$ with a vector $x$ is called a (vector) **norm** if the following four properties hold for all vectors $x$, $y$ and scalars $k$:

1. $\|x\| \geq 0$

2. $\|x\| = 0 \implies x = 0$

3. $\|kx\| = |k|\|x\|$

4. $\|x + y\| \leq \|x\| + \|y\|$

• Property 4 is called the **triangle inequality**.

• If property 2 is missing then the function $\|x\|$ is called a **semi-norm**.
Examples of Vector Norms

• Throughout let
  \[ x = [x_i]_{i=1}^{n} \in \mathbb{R}^n. \]

• The \( p \)-norm, for \( p \geq 1 \), is defined by
  \[ \|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}. \]
  Special cases of the \( p \)-norm are
  – the \textbf{1-norm}, \( p = 1 \):
    \[ \|x\|_1 = \sum_{i=1}^{n} |x_i| \]
  – the \textbf{standard norm}, etc., \( p = 2 \):
    \[ \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \]
  – the \textbf{infinity or Chebychev norm}, \( p = \infty \):
    \[ \|x\|_\infty = \max_{i=1,...,n} |x_i| \]

\[
\text{(1)}
\]
• **Query:** Why is there no sum in (1)?

• The $p$-norms satisfy the **Hölder Inequality**. Suppose

\[
\frac{1}{p} + \frac{1}{q} = 1
\]

Then

\[
|x^T y| \leq \|x\|_p \|y\|_q
\]

• In the special case that $p = q = 2$ this turns into the **Cauchy-Schwarz Inequality**:

\[
|x^T y| \leq \|x\|_2 \|y\|_2 
\tag{2}
\]

• Of course we know the stronger result that

\[
x^T y = \|x\|_2 \|y\|_2 \cos \theta
\tag{3}
\]

where $\theta$ is the angle formed by $x$ and $y$. Clearly (3) implies (2).

• Another useful norm is a **weighted $p$-norm**. Suppose we have a vector $w$ of weights:

\[
w \in \mathbb{R}^n, \quad w_i > 0, \quad i = 1, \ldots, n.
\]

\[
\|x\|_{w,p} = \left( \sum_{i=1}^{n} w_i |x_i|^p \right)^{1/p}
\]

• A weighted $p$-norm is a special case of the norm $\| \cdot \|_A$ defined by a given norm $\| \cdot \|$ and a non-singular matrix $A$:

\[
\|x\|_A = \|Ax\|
\]
• **Query:** Why does $A$ have to be non-singular?

• **Exercise:** Show that all of the above examples do in fact define a norm.

• **Exercise:** Prove the validity of the Hölder and Cauchy-Schwarz inequalities.

• Choosing among norms is largely a matter of convenience. On $\mathbb{R}^n$, and in fact on all finite-dimensional vector spaces, all norms are equivalent in the sense that if $\| \cdot \|$ and $\| \cdot \|_*$ are two given norms then there exist constants $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 \| x \| \leq \| x \|_* \leq c_2 \| x \|$$

for all $x \in \mathbb{R}^n$.

• **Exercise:** Verify this statement. (This is a bit subtle.)
Errors

- Norms can be used to measure errors. Suppose \( x \) and \( \hat{x} \) are vectors in \( \mathbb{R}^n \) where we think of \( \hat{x} \) as an approximation of \( x \):

\[
\hat{x} \approx x
\]

- Then \( \|x - \hat{x}\| \) is the **absolute error** in \( \hat{x} \) (with respect to the norm \( \| \cdot \| \))

- Similarly, \( \frac{\|x - \hat{x}\|}{\|x\|} \) is the **relative error**.

Convergence

- Norms can be used to define and analyze the convergence of a sequence of vectors. We say that the sequence

\[
x^{(0)}, x^{(1)}, x^{(2)}, x^{(3)}, \ldots \rightarrow \nabla
\]

converges to \( x \) if

\[
\lim_{k \rightarrow \infty} \|x^{(k)} - x\| = 0.
\]

- Because of the equivalence of norms the convergence of the sequence is independent of the choice of the norm used for the analysis.
Matrix Norms

- The space of $m \times n$ matrices is isomorphic to $\mathbb{R}^{mn}$ and we can apply the same definition as we did for vectors:

- **Definition:** A function $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ that associates a number $\|A\|$ with a matrix $A$ is called a **matrix norm** if the following four properties hold for all $m \times n$ matrices $A$, $B$ and scalars $k$:

  1. $\|A\| \geq 0$
  2. $\|A\| = 0 \implies A = 0$
  3. $\|kA\| = |k|\|A\|$ (4)
  4. $\|A + B\| \leq \|A\| + \|B\|$

- For example, the **Frobenius Norm** $\|A\|_F$ of a matrix $A$ is defined by thinking of the matrix as a vector and applying the vector 2-norm:

  $$\|A\|_F = \sqrt{\sum_i \sum_j a_{ij}^2}.$$

The problem with thinking of the matrix...
as a giant vector is that it does not tell us what happens to the norm when we multiply matrices.

- The link between norms and matrix products is provided by the concept of an **induced matrix norm** or **operator norm**:

  - given vector norms in \( \mathbb{R}^m \) and \( \mathbb{R}^n \) we define the induced matrix norm of an \( m \times n \) matrix \( A \) as

    \[
    \|A\| = \max_{\|x\| = 1} \|Ax\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}
    \]

  - Note that the vector norms operate on different spaces if the matrix \( A \) is rectangular.

- **Exercise**: Show that every induced matrix norm satisfies the conditions listed in (4).

- The great advantage of using an induced matrix norm is that we get the additional property that the norm of the product is never larger than the product of the norms. More formally:

- **Product Property**: Let \( \| \cdot \| \) denote both the pertinent vector norms and the induced matrix norms. Then, for all matrices \( A \) and \( B \) and vectors \( x \):

    \[
    \|Ax\| \leq \|A\|\|x\| \quad \text{and} \quad \|AB\| \leq \|A\|\|B\| \tag{5}
    \]
• **Query:** Why do we use the plural “norms” in the above statement?

• The properties in (5) are easy to verify. The first follows straight from the definition:

\[
\|Ax\| = \frac{\|Ax\|}{\|x\|} \|x\| \leq \|A\|\|x\|.
\]

The second follows from the first. We have, for some vector \(x\) with

\[
\|x\| = 1,
\]

that

\[
\|AB\| = \|ABx\|
\]

\[
\leq \|A\|\|Bx\|
\]

\[
\leq \|A\|\|B\|\|x\|
\]

\[
= \|A\|\|B\|
\]

**The Infinity Norm**

• Some induced matrix norms are much easier to compute than others. For example, the infinity- or Chebychev-norm of an \(m \times n\) matrix \(A\) is simply the maximum row sum:

\[
\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty = \max_{i=1...m} \sum_{j=1}^{n} |a_{ij}|.
\]
• Example:

\[
\left\| \begin{bmatrix} 1 & 2 & 3 \\ 4 & -1 & -3 \\ 1 & 1 & 1 \end{bmatrix} \right\|_\infty = \infty
\]

• To see (6) let \( S \) denote the maximum row sum:

\[
S = \max_{i=1,...,m} \sum_{j=1}^{n} |a_{ij}| = \sum_{j=1}^{n} |a_{kj}|
\]

for some \( k \). Thus \( k \) is the index of some row that has the maximum sum. If there are several such rows we pick any particular one of them.

• We need to show that \( \|Ax\|_\infty \leq S \) for all vectors \( x \) that satisfy \( \|x\|_\infty = 1 \) and that there exist some particular vector \( x \) with \( \|x\|_\infty = 1 \) such that \( \|Ax\|_\infty \geq S \).

• To see the first inequality let \( x \) be such that

\[
\|x\|_\infty \leq 1,
\]

i.e.,

\[
|x_i| \leq 1, \quad i = 1, \ldots, n.
\]
Then

\[ \|Ax\|_\infty = \max_{i=1,\ldots,m} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \]

\[ \leq \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij} x_j| \]

\[ = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}| |x_j| \]

\[ \leq \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}| \]

\[ = S \]

- On the other hand, with

\[ s = [\text{sign}a_{k,j}] \]

where

\[ \text{sign}z = \begin{cases} 
1 & \text{if } z > 0 \\
-1 & \text{if } z < 0 \\
0 & \text{if } z = 0 
\end{cases} \]
we have

$$\|As\|_\infty = \max_{i=1...n} \left| \sum_{j=1}^{n} a_{ij} \text{sign} a_{kj} \right|$$

$$\geq \left| \sum_{j=1}^{n} a_{kj} \text{sign} a_{kj} \right|$$

$$= \sum_{j=1}^{n} |a_{kj}|$$

$$= S$$

- For example, for the previously considered matrix

$$A = \begin{bmatrix}
1 & 2 & 3 \\
4 & -1 & -3 \\
1 & 1 & 1 \\
\end{bmatrix}$$

the maximum row sum is 8 and occurs in row 2. The vector of signs is $s = [1, -1, -1]^T$ and indeed $\|As\|_\infty = 8$:

$$\|A\|_\infty = 8$$

$$\|A\|_1 = 7$$

- Exercise: Use the same idea to show that the 1-norm of a matrix is the maximum column
sum:
\[ \| A \|_1 = \max_{j=1}^n \sum_{i=1}^m |a_{ij}|. \]

• Compare this with the 2-norm of a matrix. It is given by
\[ \| A \|_2 = \max_{\| x \|_2 = 1} \| Ax \|_2 = \sqrt{\rho(A^T A)} \]

where \( \rho(B) \) is the spectral radius of a square matrix \( B \). The spectral radius is the maximum of the absolute values of the eigenvalues.

• This is much harder to compute than the 1 or infinity norm. A straightforward calculation would require \( O(n^2) \) operations for the 1 or infinity norm and \( O(n^3) \) operations for the computation of the 2-norm of \( A \).

• Exercise: Show that for any square matrix \( A \) and any induced matrix norm \( \| \cdot \| \)
\[ \| A \| \geq \rho(A) \]

• Exercise: Show that the Frobenius norm is not induced by any vector norm.

• Our next task is to use the ideas we discussed today to analyze the role of errors in solving a linear system.
\[ Ax = b \quad \rightarrow \quad ^\wedge x \quad e = x - ^\wedge x \quad \text{error} \]

\[ Ax - b = 0 \]

\[ Ax - b = r \quad \text{residual} \]

\[ r = Ax - b \]

\[ = A^\wedge x - Ax \]

\[ = A(e - x) \]

\[ = Ae \]

\[ Ae = r \]

\[ Ax = b \]

\[ \frac{\|e\|}{\|x\|} \leq \frac{\|r\|}{\|b\|} \]

\[ \frac{2}{\sqrt{2}} \]

\[ Ax = b \quad \|b\| \leq \|A\| \|x\| \]

\[ A^\wedge x = r \quad \|r\| \leq \|A\| \|e\| \]

\[ A^{-1} b = x \quad \|x\| \leq \|A^{-1}\| \|b\| \]

\[ A^{-1} r = e \quad \|e\| \leq \|A^{-1}\| \|r\| \]