VECTOR & MATRIX NORMS

- Inner products
- Vector norms
- Matrix norms
- Introduction to singular values
- Expressions of some matrix norms.

Inner products and Norms

**Inner product of 2 vectors**

- Inner product of 2 vectors \( x \) and \( y \) in \( \mathbb{R}^n \):
  \[ x_1y_1 + x_2y_2 + \cdots + x_ny_n \]

Notation: \((x, y)\) or \(y^T x\)

- For complex vectors
  \[ (x, y) = x_1\bar{y}_1 + x_2\bar{y}_2 + \cdots + x_n\bar{y}_n \]

Note: \((x, y) = y^H x\)

- On notation: Sometimes you will find \(\langle.,.\rangle\) for \((.,.)\) and \(A^*\) instead of \(A^H\)

**Properties of Inner Product:**

- \((x, y) = (y, x)\).
- \((\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)\) [Linearity]
- \((x, x) \geq 0\) is always real and non-negative.
- \((x, x) = 0\) iff \(x = 0\) (for finite dimensional spaces).
Given $A \in \mathbb{C}^{m \times n}$ then

$$(Ax, y) = (x, A^H y) \quad \forall \ x \in \mathbb{C}^n, \forall \ y \in \mathbb{C}^m$$

- **Vector norms**

Norms are needed to measure lengths of vectors and closeness of two vectors. Examples of use: Estimate convergence rate of an iterative method; Estimate the error of an approximation to a given solution; ...

- A vector norm on a vector space $X$ is a real-valued function on $X$, which satisfies the following three conditions:

  1. $\|x\| \geq 0, \ \forall \ x \in X$, and $\|x\| = 0$ iff $x = 0$.
  2. $\|\alpha x\| = |\alpha| \|x\|, \ \forall \ x \in X, \ \forall \alpha \in \mathbb{C}$.
  3. $\|x + y\| \leq \|x\| + \|y\|, \ \forall \ x, y \in X$.

- Third property is called the triangle inequality.

- **Important example: Euclidean norm** on $X = \mathbb{C}^n$,

$$\|x\|_2 = (x, x)^{1/2} = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2}$$

- Show that when $Q$ is orthogonal then $\|Qx\|_2 = \|x\|_2$
Most common vector norms in numerical linear algebra: special cases of the Hölder norms (for $p \geq 1$):

$$\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$ 

(A) Find out (online search) how to show that these are indeed norms for any $p \geq 1$ (Not easy for 3rd requirement!)

Property: Limit of $\|x\|_p$ when $p \to \infty$ exists:

$$\lim_{p \to \infty} \|x\|_p = \max_{i=1}^{n} |x_i|.$$ 

Defines a norm denoted by $\|\cdot\|_\infty$.

The cases $p = 1$, $p = 2$, and $p = \infty$ lead to the most important norms $\|\cdot\|_p$ in practice. These are:

- $\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$, 
- $\|x\|_2 = \left[ |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 \right]^{1/2}$, 
- $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$.

The Cauchy-Schwartz inequality (important) is:

$$|(x, y)| \leq \|x\|_2 \|y\|_2.$$ 

When do you have equality in the above relation?

Expand $(x + y, x + y)$. What does the Cauchy-Schwartz inequality imply?

The Hölder inequality (less important for $p \neq 2$) is:

$$|(x, y)| \leq \|x\|_p \|y\|_q,$$ with $\frac{1}{p} + \frac{1}{q} = 1$.

Proof moved to supplement set #2.

Second triangle inequality: $\|x\| - \|y\| \leq \|x - y\|$.

Consider the metric $d(x, y) = \max_i |x_i - y_i|$. Show that any norm in $\mathbb{R}^n$ is a continuous function with respect to this metric.
**Equivalence of norms:**

In finite dimensional spaces \((\mathbb{R}^n, \mathbb{C}^n, \ldots)\) all norms are ‘equivalent’: if \(\phi_1\) and \(\phi_2\) are two norms then there exists positive constants \(\alpha, \beta\) such that:

\[
\beta \phi_2(x) \leq \phi_1(x) \leq \alpha \phi_2(x).
\]

How can you prove this result? [Hint: Show for \(\phi_2 = \|\cdot\|_{\infty}\)]

**Convergence of vector sequences**

A sequence of vectors \(x^{(k)}, k = 1, \ldots, \infty\) converges to a vector \(x\) with respect to the norm \(\|\cdot\|\) if, by definition,

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

Important point: because all norms in \(\mathbb{R}^n\) are equivalent, the convergence of \(x^{(k)}\) w.r.t. a given norm implies convergence w.r.t. any other norm.

Notation:

\[
\lim_{k \to \infty} x^{(k)} = x
\]

**Example:** The sequence

\[
x^{(k)} = \left( \frac{1 + 1/k}{k} \right)
\]

converges to

\[
x = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}
\]

Note: Convergence of \(x^{(k)}\) to \(x\) is the same as the convergence of each individual component \(x_i^{(k)}\) of \(x^{(k)}\) to the corresponding component \(x_i\) of \(x\).
**Matrix norms**

- Can define matrix norms by considering \(m \times n\) matrices as vectors in \(\mathbb{R}^{mn}\). These norms satisfy the usual properties of vector norms, i.e.,

1. \(\|A\| \geq 0, \forall A \in \mathbb{C}^{m \times n}\), and \(\|A\| = 0\) iff \(A = 0\)
2. \(\|\alpha A\| = |\alpha|\|A\|, \forall A \in \mathbb{C}^{m \times n}, \forall \alpha \in \mathbb{C}\)
3. \(\|A + B\| \leq \|A\| + \|B\|, \forall A, B \in \mathbb{C}^{m \times n}\).

- However, these will lack (in general) the right properties for composition of operators (product of matrices).

- The case of \(\|\cdot\|_2\) yields the Frobenius norm of matrices.

**Consistency / sub-multiplicativity of matrix norms**

- A fundamental property of matrix norms is consistency

\[
\|AB\|_p \leq \|A\|_p \|B\|_p.
\]

[Also termed "sub-multiplicativity"]

- Consequence: (for square matrices) \(\|A^k\|_p \leq \|A\|_p^k\)

- \(A^k\) converges to zero if any of its \(p\)-norms is \(< 1\)

[Note: sufficient but not necessary condition]

**Frobenius norms of matrices**

- The Frobenius norm of a matrix is defined by

\[
\|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2\right)^{1/2}.
\]

- Same as the 2-norm of the column vector in \(\mathbb{C}^{mn}\) consisting of all the columns (respectively rows) of \(A\).

- This norm is also consistent [but not induced from a vector norm]
Compute the Frobenius norms of the matrices
\[
\begin{pmatrix}
1 & 1 \\
1 & 0 \\
3 & 2
\end{pmatrix}
\quad \begin{pmatrix}
1 & 2 & -1 \\
-1 & \sqrt{5} & 0 \\
-1 & 1 & \sqrt{2}
\end{pmatrix}
\]

Prove that the Frobenius norm is consistent [Hint: Use Cauchy-Schwarz]

Define the 'vector 1-norm' of a matrix \( A \) as the 1-norm of the vector of stacked columns of \( A \). Is this norm a consistent matrix norm?

[Hint: Result is true – Use Cauchy-Schwarz to prove it.]

Compute the \( p \)-norm for \( p = 1, 2, \infty, F \) for the matrix \( A = \begin{pmatrix} 0 & 2 \\ 0 & 1 \end{pmatrix} \)

Show that \( \rho(A) \leq \|A\| \) for any matrix norm.

Is \( \rho(A) \) a norm?

Expressions of standard matrix norms

Recall the notation: (for square \( n \times n \) matrices)
\[
\rho(A) = \max |\lambda_i(A)|; \quad \text{Tr}(A) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i(A)
\]
where \( \lambda_i(A), i = 1, 2, \ldots, n \) are all eigenvalues of \( A \)

\[
\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^m |a_{ij}|, \\
\|A\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^n |a_{ij}|, \\
\|A\|_2 = \left[\rho(A^H A)^{1/2} = \left[\rho(A A^H)^{1/2} \right]^2 \right], \\
\|A\|_F = \left[\text{Tr}(A^H A)^{1/2} = \left[\text{Tr}(A A^H)^{1/2} \right]^2 \right].
\]

1. \( \rho(A) = \|A\|_2 \) when \( A \) is Hermitian \( (A^H = A) \). True for this particular case...

2. ... However, not true in general. For \( A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \), we have \( \rho(A) = 0 \) while \( A \neq 0 \). Also, triangle inequality not satisfied for the pair \( A \) and \( B = A^T \). Indeed, \( \rho(A + B) = 1 \) while \( \rho(A) + \rho(B) = 0 \).

Given a function \( f(t) \) (e.g., \( e^t \)) how would you define \( f(A) \)?

[Was seen earlier. Here you need to fully justify answer. Assume \( A \) is diagonalizable]
**Singular values and matrix norms**

- Let $A \in \mathbb{R}^{m \times n}$ or $A \in \mathbb{C}^{m \times n}$
- Eigenvalues of $A^H A$ & $AA^H$ are real $\geq 0$. Show this.
- Let
  
  \[ \sigma_i = \sqrt{\lambda_i(A^H A)} \quad i = 1, \cdots, n \text{ if } n \leq m \\
  \sigma_i = \sqrt{\lambda_i(AA^H)} \quad i = 1, \cdots, m \text{ if } m < n \]

  The $\sigma_i$'s are called singular values of $A$.
  
  Note: a total of $\min(m, n)$ singular values.
  
  Always sorted decreasingly: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_k \geq \cdots$

  We will see a lot more on singular values later

More generally: Schatten $p$-norm ($p \geq 1$) defined by

\[ \|A\|_{*,p} = \left[ \sum_{i=1}^{r} \sigma_i^p \right]^{1/p} \]

Note: $\|A\|_{*,p}$ is $p$-norm of vector $[\sigma_1; \sigma_2; \cdots ; \sigma_r]$.

In particular: $\|A\|_{*,1} = \sum \sigma_i$ is called the nuclear norm and is denoted by $\|A\|_\ast$. (Common in machine learning).

**A few properties of the 2-norm and the F-norm**

- Let $A = uv^T$. Then $\|A\|_2 = \|u\|_2 \|v\|_2$

  Prove this result.

- In this case $\|A\|_F = ??$

  For any $A \in \mathbb{C}^{m \times n}$ and unitary matrix $Q \in \mathbb{C}^{m \times m}$ we have $\|QA\|_2 = \|A\|_2$; $\|QA\|_F = \|A\|_F$. 
Show that the result is true for any orthogonal matrix $Q$ (where $Q$ has orthonomal columns), i.e., when $Q \in \mathbb{C}^{p \times m}$ with $p > m$.

Let $Q \in \mathbb{C}^{n \times n}$, unitary. Do we have $\|AQ\|_2 = \|A\|_2$?

$\|AQ\|_F = \|A\|_F$? What if $Q \in \mathbb{C}^{n \times p}$, with $p < n$ (and $Q^HQ = I$)?