Introduction to Linear Algebra I

- Inner products
- Cauchy-Schwarz inequality
- Triangle inequality, reverse triangle inequality
- Vector and matrix norms
- Equivalence of $\ell_p$ norms
- Basic norm inequalities (useful for proofs)
- Matrices
Basics

- Sets, vector space
- $\mathbb{R}^N$: $N$-dimensional **Euclidean space**
- A **vector** $\mathbf{a} \in \mathbb{R}^N$ is an $n$-tuple $[a_1, a_2, \ldots, a_N]$, where $a_i \in \mathbb{R}$. (think of vectors as a column vector or a $N \times 1$ matrix.)

**Inner product:**

$a, b \in \mathbb{R}^N, \langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{N} a_i b_i = \mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a} = \langle \mathbf{b}, \mathbf{a} \rangle$

(Note that $\langle \mathbf{a}, \mathbf{b} \rangle \in \mathbb{R}$.)

**Euclidean norm:** Induced by the inner-product

$a \in \mathbb{R}^N, \| \mathbf{a} \|_2 = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}$

Don’t confuse the norm $\| \mathbf{x} \|_2$ with the absolute value $|x|$
Lemma (Cauchy-Schwarz inequality)

Given \( \mathbf{a}, \mathbf{b} \in \mathbb{R}^N \),

\[
|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\|_2 \|\mathbf{b}\|_2.
\]

- Probably the most important inequality out there!
- There is a book solely devoted to this inequality.
- When does it hold with equality?
- Is used to derive the triangle inequality shown next.
Triangle inequality

Lemma (Triangle inequality)

Given \( \mathbf{a}, \mathbf{b} \in \mathbb{R}^N \),

\[
\|\mathbf{a} + \mathbf{b}\|_2 \leq \|\mathbf{a}\|_2 + \|\mathbf{b}\|_2.
\]

- Proof uses Cauchy-Schwarz inequality (do on board)
- When does this inequality hold with equality?
- Reverse (or inverse) triangle inequalities:

\[
\|\mathbf{a} + \mathbf{b}\|_2 \geq \|\mathbf{a}\|_2 - \|\mathbf{b}\|_2
\]

\[
\|\mathbf{a} + \mathbf{b}\|_2 \geq \|\mathbf{b}\|_2 - \|\mathbf{a}\|_2
\]
What is a norm?

- Assigns a positive number to each non-zero vector
- Is only zero if the vector is an all-zero vector
- Key aspect in proving uniqueness results

Norm properties

- Homogeneity: \( \|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\| \), for \( \mathbf{x} \in \mathbb{R}^N \) and \( \alpha \in \mathbb{R} \)
- Subadditivity: \( \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \), for \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^N \)
- Separability: If and only if \( \|\mathbf{x}\| = 0 \), then \( \mathbf{x} = 0 \)
It’s time to play “IS IT A NORM?!?”

- $\|x\|_2$
- $\|x\|_0$ counts the number of non-zeros in $x$
- $\|x\|_1$ or $|x|$
- $\|\nabla x\|_2$
- $\sqrt{x^T A x}$, for some matrix $A$. 
$\ell_p$ norms

**Definition ($\ell_p$ norms)**

For $p \geq 1$, $a \in \mathbb{R}^N$, $\|a\|_p = \left( \sum_{i=1}^{N} |a_i|^p \right)^{1/p}$

- $\ell_2$ norm: $p = 2$, $\|a\|_2 = \sqrt{\sum_i |a_i|^2}$
- $\ell_1$ norm: $p = 1$, $\|a\|_1 = \sum_i |a_i|$
- $\ell_\infty$ norm: $p = \infty$, $\|a\|_\infty = \max_i |a_i|$

**Lemma (Minkowski’s inequality)**

$1 \leq p \leq \infty$, $\|a + b\|_p \leq \|a\|_p + \|b\|_p$
$\ell_p$-norm balls

**Definition ($\ell_p$ ball)**

$$\epsilon \geq 0, \quad B_{\ell_p}(\epsilon) = B_p(\epsilon) = \{ a \mid \|a\|_p \leq \epsilon \}$$

$B_p(1)$ is referred to as the *unit ball* (i.e., $\epsilon = 1$).
Equivalence of norms

Given any two norms, say \( \ell_p \) and \( \ell_q \), \( \exists \, \alpha, \beta > 0 \) such that

\[
\forall \mathbf{a} \in \mathbb{R}^N, \, \alpha \| \mathbf{a} \|_q \leq \| \mathbf{a} \|_p \leq \beta \| \mathbf{a} \|_q.
\]

- \( \| \mathbf{a} \|_\infty \leq \| \mathbf{a} \|_2 \leq \sqrt{N} \| \mathbf{a} \|_\infty \)
- \( \| \mathbf{a} \|_\infty \leq \| \mathbf{a} \|_1 \leq N \| \mathbf{a} \|_\infty \)
- \( \| \mathbf{a} \|_2 \leq \| \mathbf{a} \|_1 \leq \sqrt{N} \| \mathbf{a} \|_2 \)

This implies that all \( p \)-norms behave—at least in principle—similarly. However, we will show that they have very distinct properties.

Lemma (General equivalence of \( \ell_p \) norms)

\[
1 \leq p < q, \quad \| \mathbf{a} \|_q \leq \| \mathbf{a} \|_p \leq N^{1/p - 1/q} \| \mathbf{a} \|_q
\]
Two important inequalities

Hölder’s inequality:

- \( |\langle a, b \rangle| \leq \|a\|_p \|b\|_q \) with \( 1/p + 1/q = 1 \) and \( p, q \in [1, \infty] \)
- \( p \) and \( q \) are so-called dual norms
- Generalization of the Cauchy-Schwarz inequality

Jensen’s inequality:

- Let \( f(x) \) be a convex function with \( x_1, x_2 \in \mathbb{R} \) and for \( t \in [0, 1] \)
  \[
  f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)
  \]
- Also
  \[
  f \left( \frac{\sum_i a_i x_i}{\sum_i a_i} \right) \leq \frac{\sum_i a_i f(x_i)}{\sum_i a_i}
  \]
Collection of vectors, Subspaces

A set of $T$-vectors, $V = \{a_1, a_2, \ldots, a_T\}$

- **Linear combination:** $\sum_{k=1}^{T} \alpha_k a_k, \alpha_k \in \mathbb{R}$
- **Linearly independent:** No vector in $V$ can be written as linear combination of others
- **Span:** $\text{Span}(V) = \{x \mid x = \sum_k \alpha_k a_k, \alpha_k \in \mathbb{R}\}$

**Definition (Subspace)**

A collection of vectors $V \subset \mathbb{R}^N$ is a subspace iff it is closed under linear combinations

$$a, b \in V \implies \alpha a + \beta b \in V, \alpha, \beta \in \mathbb{R}$$

- **Basis of a subspace:** A linearly independent *spanning* set
- **Dimensionality of a subspace:** \#elements in a basis
Matrix

- $A \in \mathbb{R}^{M \times N}$: A matrix of dimension $M \times N$
- $A = [a_{ij}] = [a_1, a_2, \ldots, a_N], a_i \in \mathbb{R}^M$
- $\text{rank}(A) =$ largest number of linearly independent columns
- $\text{rank}(A) = \text{rank}(A^T) \leq \min(M, N)$
- $A$ is full-rank if $\text{rank}(A) = \min(M, N)$.

Matrices are representations of linear operators.

$A : \mathbb{R}^N \rightarrow \mathbb{R}^M$
$x \in \mathbb{R}^N \mapsto Ax \in \mathbb{R}^M$

Examples of linear operators that aren’t matrices?
Matrix norms

Definition (Spectral norm)

\[ \|A\|_{2,2} = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2 = 1} \|Ax\|_2 \]

- The norm used above is the *induced* norm or the \(\ell_2\)-norm.
- Quantifies the maximum increase in length of unit-norm vectors due to the operation of the matrix \(A\)
- \(\|A\|_{2,2}\) is equal to the largest *singular value* of \(A\) (more on this later)
- \(\|Ax\|_2 \leq \|A\|_{2,2}\|x\|_2\) (Question: When is it equal?)

Lemma

\[ \|AB\|_{2,2} \leq \|A\|_{2,2}\|B\|_{2,2} \]

- Can you show this?
Induced matrix norms

**Definition**

\[ \|A\|_{p,q} = \max_{x \neq 0} \frac{\|Ax\|_q}{\|x\|_p} = \max_{\|x\|_p = 1} \|Ax\|_q \]

- \( \|A\|_{2,2} \) the maximum singular value of \( A \)
- \( \|A\|_{1,1} \) : maximum of the absolute column sums
- \( \|A\|_{\infty,\infty} \) : maximum of the absolute row sums

- \( \|Ax\|_q \leq \|A\|_{p,q} \|x\|_p \) (by definition)
- \( \|A\|_{2,2}^2 \leq \|A\|_{1,1} \|A\|_{\infty,\infty} \) (similar to Hölder’s inequality)

Note: We get lazy and write \( \|A\|_2 \) for \( \|A\|_{2,2} \)
Other frequently-used matrix norms

- **Frobenius norm:**
  - Definition: \( \| A \|_F = \sqrt{\sum_{i,j} |A_{i,j}|^2} \)
  - Alternative definition: \( \| A \|_F = \sqrt{\text{trace}(A^T A)} = \sqrt{\text{trace}(AA^T)} \)
  - The Frobenius norm is not an induced norm

- **Nuclear norm:**
  - Definition: \( \| A \|_* = \text{trace}(\sqrt{A^T A}) = \sum_{i=1}^{\min\{M,N\}} \sigma_i \)
  - With \( \sigma_i \) being the singular values of the matrix \( A \)
  - The nuclear norm is not an induced norm

- **ALL matrix norms are also equivalent → Wikipedia**
Eigenvectors and eigenvalues

- Let $A$ be a $N \times N$ square matrix
- $x$ is an eigenvector and $\lambda$ is an eigenvalue of $A$ is

$$Ax = \lambda x$$

**Intuition:** eigenvectors are vectors in $\mathbb{R}^N$ whose direction is preserved under action of $A$; however, length may change

**Eigen-decomposition:** $A = UDU^{-1}$
Spectral Theorem

Theorem

If $A = A^H$, then

- The matrix is “symmetric”
- all eigenvalues are real
- eigenvectors with different eigenvalues are perpendicular
- there exists a complete orthogonal basis of eigenvectors.
Singular value decomposition (SVD)

Definition (SVD)

Any matrix $A \in \mathbb{R}^{M \times N}$ can be written as

$$A = U \Sigma V^T,$$

where $U \in \mathbb{R}^{M \times M}$ and $V \in \mathbb{R}^{N \times N}$ are unitary and $\Sigma \in \mathbb{R}^{M \times N}$ is diagonal.

- Diagonal entries of $\Sigma = \{\sigma_i\}$ are called the singular values; they are positive and real. Typically, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$
- Singular values are the eigenvalues of $\sqrt{A^T A}$ and $\sqrt{AA^T}$.
- If $A = A^T$, singular values are same as the eigenvalues.
- Geometric picture and other properties, read Wikipedia.
- Very useful matrix decomposition!
Singular value decomposition (SVD)

**Definition (SVD)**

A matrix $A \in \mathbb{R}^{M \times N}$ can be written as

$$A = U \Sigma V^T,$$

where $U \in \mathbb{R}^{M \times M}$ and $V \in \mathbb{R}^{N \times N}$ are unitary and $\Sigma \in \mathbb{R}^{M \times N}$ is diagonal.

- If $A^{-1}$ exists, then $A^{-1} = V \Sigma^{-1} U^T$.
- Even if $A$ is singular, we can define a pseudo-inverse $A^\dagger$ as follows:

  $$A^\dagger = V \hat{\Sigma}^{-1} U^T,$$

  where $\hat{\Sigma}^{-1}$ has the diagonal terms $1/\sigma_i$ if $\sigma_i \neq 0$, and zero otherwise.
- The ratio of the largest to smallest singular value is the so-called condition number of $A$. 


Solving $y = Ax$ (square case)

Scenario: $A$ is **full-rank**, $M = N$ (square matrix)
(full rank implies that $A^{-1}$ exists)

Given $y$, the *unique* solution $x$ is

$$\hat{x} = A^{-1}y$$

Geometric picture: $A$ is a *one-to-one, onto* map from $\mathbb{R}^N$ to $\mathbb{R}^M = \mathbb{R}^N$
Block Inversion Formulas

\[
\begin{pmatrix}
A & B \\
C & D
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix} =
\begin{pmatrix}
b \\
c
\end{pmatrix}
\]

We can solve using elimination...

\[
\begin{pmatrix}
I & A^{-1}B \\
C & D
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix} =
\begin{pmatrix}
b \\
c
\end{pmatrix}
\]

\[
\begin{pmatrix}
I & A^{-1}B \\
C & D
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix} =
\begin{pmatrix}
A^{-1}b \\
c
\end{pmatrix}
\]

\[
\begin{pmatrix}
I & A^{-1}B \\
0 & D - CA^{-1}B
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix} =
\begin{pmatrix}
A^{-1}b \\
c - CA^{-1}b
\end{pmatrix}
\]

\[
y = (D - CA^{-1}B)^{-1}(c - CA^{-1}b)
\]
The Schur Complement

\[
\begin{pmatrix}
A & B \\
C & D
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix} =
\begin{pmatrix}
b \\
c
\end{pmatrix}
\]

“You Schur look great today!”

\[S = (D - CA^{-1}B)\]

The Schur complement

- \(S^{-1}\) is a diagonal entry in the matrix inverse
- The block matrix is invertible iff \(S\) is invertible
- Block matrix is PSD iff \(A, S\) are PSD