2. Gaussian Elimination

2.1. Vector and Matrix Norms

- (Linear) Vector Space. Regard the columns (rows) of a matrix \( \{a_1, a_2, ..., a_n\} \) as forming a vector space \( \mathbb{R}^{m \times n} \)
  
  - The key properties of a vector space \( V \) are
    
    i. If \( u, v \in V \) then \( u + v \in V \)
    
    ii. If \( u \in V \) and \( \alpha \in \mathbb{R} \) then \( \alpha u \in V \)
  

  - A *subspace* of \( V \) is a subset that is also a vector space

- Definition 1: A set of vectors \( \{a_1, a_2, ..., a_n\} \in \mathbb{R}^{m \times n} \) is *linearly independent* if the only solution of

\[
\sum_{j=1}^{n} \alpha_j a_j = 0
\]

is \( \alpha_j = 0, j = 1 : n \)
Subspaces

• **Definition 2:** The set of all linear combinations of
  \[
  \{a_1, a_2, ..., a_n\} \in \mathbb{R}^{m \times n}
  \]
is a vector space called the linear span of \(\mathbb{R}^{m \times n}\):
  \[
  \text{span}\left(\{a_1, a_2, ..., a_n\}\right) = \left\{ \sum_{j=1}^{n} \alpha_j a_j | \alpha_j \in \mathbb{R} \right\}
  \]  

  – **Theorem 1:** If \(\{a_1, a_2, ..., a_n\}\) is linearly independent then \(u \in \text{span}(\{a_1, a_2, ..., a_n\})\) is unique

• **Definition 3:** Let \(S_1, S_2 \subset \mathbb{R}^{m \times n}\), then
  i. Their sum is the subspace \(\{a_1 + a_2 | a_1 \in S_1, a_2 \in S_2\}\)
  ii. If \(S_1 \cap S_2 = \emptyset\) then the sum of \(S_1\) and \(S_2\) is called the direct sum and is written as \(S_1 \oplus S_2\)

• **Definition 4:** A subset \(S\) is a maximal linearly independent subset of \(\{a_1, a_2, ..., a_n\}\) if:
  i. it is linearly independent
  ii. it is not properly contained in any linearly independent subset of \(\{a_1, a_2, ..., a_n\}\)

• **Definition 5:** A maximal linearly independent set \(\{b_1, b_2, ..., b_k\}\) forms a basis for \(\{a_1, a_2, ..., a_n\}\)
Range, Null Space, and Rank

• Some properties:
  
i. If \( \{b_1, b_2, ..., b_k\} \) is a basis for \( \{a_1, a_2, ..., a_n\} \) then
  \[
  \text{span}(\{a_1, a_2, ..., a_n\}) = \text{span}(\{b_1, b_2, ..., b_k\})
  \]

  ii. All bases for \( \{a_1, a_2, ..., a_n\} \) have the same number of elements

  iii. This number is the *dimension* of \( \{a_1, a_2, ..., a_n\} \) and is written as \( \text{dim}(\{a_1, a_2, ..., a_n\}) \)

• Definition 6: If \( A \in \mathbb{R}^{m \times n} \) then:
  
i. The *range* of \( A \) satisfies
  \[
  \text{range}(A) = \{y \in \mathbb{R}^m | y = Ax, x \in \mathbb{R}^n\}
  \]

  ii. The *null space* of \( A \) satisfies
  \[
  \text{null}(A) = \{x \in \mathbb{R}^n | Ax = 0\}
  \]

• For \( A \in \mathbb{R}^{m \times n} \)
  
i. \( \text{range}(A) = \text{span}(\{a_1, a_2, ..., a_n\}) \)
  
  ii. The *rank* of \( A \) satisfies \( \text{rank}(A) = \text{dim}(\text{range}(A)) \)
  
  iii. \( \text{rank}(A) = \text{rank}(A^T) \)

     - \( \text{rank}(A) \) is the number of independent rows or columns of \( A \)

  iv. \( \text{dim}(\text{null}(A)) + \text{rank}(A) = n \)
Inverses

- The $n \times n$ identity matrix is

$$I = [e_1, e_2, ..., e_n] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (4)$$

- If $A, X \in \mathbb{R}^{n \times n}$ and

$$AX = I$$

then $X$ is called the inverse of $A$ and is written as $A^{-1}$

- If $A^{-1}$ does not exist then $A$ is called singular, otherwise nonsingular.

- Some properties:

$$\quad (AB)^{-1} = B^{-1}A^{-1} \quad (5a)$$

$$\quad (A^{-1})^T = (A^T)^{-1} := A^{-T} \quad (5b)$$
Determinants

• The determinant of a matrix $A \in \mathbb{R}^{n \times n}$ satisfies

$$\text{det}(A) = \begin{cases} a_{11}, & \text{if } n = 1 \\ \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \text{det}(A_{1j}), & \text{if } n > 1 \end{cases}$$  \hspace{1cm} (6)$$

where $A_{1j}$ is an $(n - 1) \times (n - 1)$ matrix obtained by deleting the first row and column of $A$.

- The definition is recursive
- Some properties:

$$\text{det}(AB) = \text{det}(A) \text{det}(B)$$ \hspace{1cm} (7a)$$

$$\text{det}(A^T) = \text{det}(A)$$ \hspace{1cm} (7b)$$

$$\text{det}(\alpha A) = \alpha^n \text{det}(A), \quad \alpha \in \mathbb{R}$$ \hspace{1cm} (7c)$$

* $\text{det}(A) \neq 0$ if and only if $A$ is nonsingular
Vector Norms

- Reading: Trefethen and Bau (1997), Lecture 3

- Norms provide a way to measure distances between vectors and matrices
  - They provide a measure of “closeness” that is used to understand convergence

- **Definition 7**: A *vector norm* $|| \cdot ||$ is a functional $(|| \cdot || : \mathbb{R}^n \rightarrow \mathbb{R})$ satisfying
  
  i. $||x|| \geq 0, \quad ||x|| = 0 \leftrightarrow x = 0, \ x \in \mathbb{R}^n$
  
  ii. $||x + y|| \leq ||x|| + ||y||, \ x, y \in \mathbb{R}^n$
  
  iii. $||\alpha x|| = |\alpha||x||, \ \alpha \in \mathbb{R}, \ x, y \in \mathbb{R}^n$

- The $p$-norms are the most useful to us:
  
  $$||x||_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}, \quad p \geq 1 \quad (8)$$
  
  - The most important of these are
    
    $$||x||_1 = |x_1| + |x_2| + \ldots + |x_n| \quad (9a)$$

    $$||x||_2 = (|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2)^{1/2} \quad (9b)$$

    $$||x||_\infty = \max_{1 \leq i \leq n} |x_i| \quad (9c)$$
Vector Norms

- For $\mathbf{x} \in \mathbb{R}^2$

$$||\mathbf{x}||_1 = |x_1| + |x_2|$$

$$||\mathbf{x}||_2 = (|x_1|^2 + |x_2|^2)^{1/2}$$

$$||\mathbf{x}||_p = (|x_1|^p + |x_2|^p)^{1/p}, 1 < p < \infty$$

$$||\mathbf{x}||_\infty = \max(|x_1|, |x_2|)$$
Vector Norms

- Some properties of $p$-norms
  - The Holder inequality:
    \[ |x^T y| \leq \|x\|_p \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1 \quad (10a) \]
  - The Cauchy-Schwartz inequality ($p = q = 2$)
    \[ |x^T y| \leq \|x\|_2 \|y\|_2 \quad (10b) \]
  - Theorem 2: $p$-norms are equivalent in the sense that there exist constants $c, C > 0$ such that
    \[ c\|x\|_\alpha \leq \|x\|_\beta \leq C\|x\|_\alpha \quad (11) \]
    * For example
    \[ \|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty \quad (12a) \]
    \[ \|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty \quad (12b) \]
    \[ \|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2 \quad (12c) \]
Vector Norms

• *Example 1* (cf. Trefethen and Bau (1997)), Problem 3.3). Verify (12)
Matrix Norms

- Matrix norms measure the sensitivity of a matrix to perturbations.
- The definition of a matrix norm is identical to that of a vector norm.
- **Definition 8:** A matrix norm $||A||$, $A \in \mathbb{R}^{m \times n}$, is a functional satisfying
  
  i. $||A|| \geq 0$, $||A|| = 0 \Leftrightarrow A = 0$
  
  ii. $||A + B|| \leq ||A|| + ||B||$
  
  iii. $||\alpha A|| = ||\alpha|| ||A||$, $\alpha \in \mathbb{R}$

- Vector induced matrix norms are defined in terms of $p$-norms of vectors

$$||A||_p = \max_{x \neq 0} \frac{||Ax||_p}{||x||_p}$$  \hspace{1cm} (13a)

- Letting $y = x/||x||_p$, yields the equivalent definition

$$||A||_p = \max_{||x||_p = 1} ||Ax||_p$$  \hspace{1cm} (13b)

- There are also $pq$-norms

$$||A||_{pq} = \max_{x \neq 0} \frac{||Ax||_p}{||x||_q}$$  \hspace{1cm} (13c)
Matrix Norms

• Properties of matrix norms:

  – From (13a), it is clear that

    \[ ||Ax||_p \leq ||A||_p ||x||_p \quad (14a) \]

    and, more generally,

    \[ ||AB||_p \leq ||A||_p ||B||_p, \quad A \in \mathbb{R}^{m \times n}, \quad B \in \mathbb{R}^{n \times r} \quad (14b) \]

    * Matrix norms that satisfy (14b) are called consistent.

  – As a further consequence of (14b)

    \[ ||A^k||_p \leq ||A||^k_p \]

• Matrix \( p \) norms corresponding to vector \( p \)-norms are

  \[ ||A||_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| \quad (15a) \]

  \[ ||A||_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \quad (15b) \]

  \[ ||A||_2 = [\rho(A^T A)]^{1/2} \quad (15c) \]

  – The spectral radius \( \rho(A) \) is the magnitude of the largest eigenvalue of a matrix \( A \in \mathbb{R}^{m \times n} \).
Frobenius Matrix Norm

- $p$ norms
  - The 1-norm is the maximum absolute column sum
  - The $\infty$-norm is the maximum absolute row sum
  - The 2-norm requires an eigenvalue analysis

- The Frobenius norm:

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

(16a)

- It is not a $p$-norm
- It can be viewed as the 2-norm of a vector in $\mathbb{R}^{mn}$
- It satisfies (14b)
- It has the equivalent definition

\[
\|\mathbf{A}\|_F = [\text{tr}(\mathbf{A}^T \mathbf{A})]^{1/2} = [\text{tr}(\mathbf{A} \mathbf{A}^T)]^{1/2}
\]  

(16b)

where

\[
\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}, \quad \mathbf{A} \in \mathbb{R}^{n \times n}
\]  

(16c)
Fun with Norms

- Example 2. Calculate the 1, 2, \(\infty\), and \(F\) norms of

\[
A = \begin{bmatrix}
1 & -3 & 6 \\
4 & 7 & -3 \\
-2 & 5 & 9 \\
3 & -2 & 4
\end{bmatrix}
\]

- The column sums are

\[
\sum_{i=1}^{4} |a_{i1}| = , \quad \sum_{i=1}^{4} |a_{i2}| = , \quad \sum_{i=1}^{4} |a_{i3}| =
\]

Thus,

\[||A||_1 = \]

- The row sums are

\[
\sum_{j=1}^{3} |a_{1j}| = , \quad \sum_{j=1}^{3} |a_{2j}| = \\
\sum_{j=1}^{3} |a_{3j}| = , \quad \sum_{j=1}^{3} |a_{4j}| =
\]

Thus,

\[||A||_{\infty} = \]
More Fun

• The product of $\mathbf{A}$ and $\mathbf{A}^T$ is

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 30 & 9 & -12 \\ 9 & 87 & -2 \\ -12 & -2 & 142 \end{bmatrix}$$

• The eigenvalues of $\mathbf{A}^T \mathbf{A}$ are 88.1293, 27.4447, and 143.4260

• Thus,

$$\|\mathbf{A}\|_2 = \sqrt{143.4260} = 11.9761$$

• Also $\|\mathbf{A}\|_F = 16.0935$

• **Theorem 3**: If $\mathbf{A} \in \mathbb{R}^{m \times n}$ then

$$\|\mathbf{A}\|_2^2 \leq \|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty$$

(17)

- *Proof*: $\mu = \|\mathbf{A}\|_2$ satisfies

$$\mathbf{A}^T \mathbf{A} \mathbf{z} = \mu^2 \mathbf{z}$$

$\mu^2$ is the largest eigenvalue of $\mathbf{A}^T \mathbf{A}$ and $\mathbf{z}$ is the corresponding eigenvector

- Take the 1 norm and use (14b) and (15a,b)

$$\mu^2 \|\mathbf{z}\|_1 = \|\mathbf{A}^T \mathbf{A} \mathbf{z}\|_1 \leq \|\mathbf{A}^T\|_1 \|\mathbf{A}\|_1 \|\mathbf{z}\|_1$$

$$= \|\mathbf{A}\|_\infty \|\mathbf{A}\|_1 \|\mathbf{z}\|_1$$

- Divide by $\|\mathbf{z}\|_1$

• For Example 2, $\|\mathbf{A}\|_2 \leq \sqrt{(22)(16)} = 18.7617$
Spectral Radii and Norms

- **Theorem 4:** For $A \in \mathbb{R}^{n \times n}$

\[
\rho(A) \leq ||A||_p \tag{18}
\]

- Let $\lambda$ be the eigenvalue of $A$ corresponding to the spectral radius and $z$ be the corresponding eigenvector

  \[ * \rho(A) = |\lambda| \]

- Consider

\[ Az = \lambda z \]

- Take a $p$ norm

\[ ||Az||_p = \rho(A)||z||_p \]

or

\[
\rho(A) = \frac{||Az||_p}{||z||_p} \leq \max_{x \neq 0} \frac{||Ax||_p}{||x||_p} = ||A||_p
\]