Chapter 1

Vector and Matrix Norms

1.1 Vector Spaces

Let $F$ be a field (such as the real numbers, $\mathbb{R}$, or complex numbers, $\mathbb{C}$) with elements called scalars. A Vector Space, $V$, over the field $F$ is a non-empty set of objects (called vectors) on which two binary operations, (vector) addition and (scalar) multiplication, are defined and satisfy the axioms below.

**Addition:** is a rule which associates a vector $\in V$ with each pair of vectors $x, y \in V$ and that member is called the sum $x + y$.

**Scalar multiplication:** is a rule which associates a vector $\in V$ with each scalar, $\alpha \in F$, and each vector $x \in V$, and it is called the scalar multiple $\alpha x$.

For $V$ to be called a Vector Space the two operations above must satisfy the following axioms $\forall x, y, w \in V$:

1. Vector addition is commutative: $x + y = y + x$;
2. Vector addition is associative: $x + (y + w) = (x + y) + w$;
3. Vector addition has an identity: $\exists$ a vector in $V$, called the zero vector 0, such that $x + 0 = x$;
4. Vector addition has an inverse: for each $x \in V \exists$ an inverse (or negative) element, $(-x) \in V$, such that $x + (-x) = 0$;
5. Distributivity holds for scalar multiplication over vector addition: $\alpha(x + y) = (\alpha x + \alpha y)$, $\alpha \in F$;
6. Distributivity holds for scalar multiplication over field addition: \((\alpha + \beta)x = (\alpha x + \beta x), \alpha, \beta \in F;\)

7. Scalar multiplication is compatible with field multiplication: \(\alpha(\beta x) = (\alpha \beta)x, \alpha, \beta \in F;\)

8. Scalar multiplication has an identity element: \(1x = x,\) where 1 is the multiplicative identity of \(F.\)

Any collection of objects together with two operations which satisfy the above axioms is a vector space. In context of Numerical Analysis a vector space is often called a Linear Space.

**Example 1.1.1** An obvious example of a linear space is \(\mathbb{R}^3\) with the usual definitions of addition of 3D vectors and scalar multiplication (Fig 1.1).

![Figure 1.1: A pictorial example of some vectors belonging to the linear space \(\mathbb{R}^3.\)](image)

**Example 1.1.2** Another example of a linear space is the set of all continuous functions \(f(x)\) on \((-\infty, \infty)\) with the usual definition for the addition of functions and the scalar multiplication of functions, usually denoted by \(C(-\infty, \infty).\)

If \(f(x)\) and \(g(x) \in C(-\infty, \infty)\) then \(f + g\) and \(\alpha f(x)\) are also continuous on \((-\infty, \infty)\) and the axioms are easily verified.

**Example 1.1.3** A sub-space of \(C(-\infty, \infty)\) is \(P_n;\) the space of polynomials of degree at most \(n.\) [We will meet this vector space later in the course.]

Note, it must be of degree “at most” \(n\) so that addition (and subtraction) produce members of \(P_n,\) e.g. \((x - x^2) \in P_2, x^2 \in P_2\) and \((x - x^2) + (x^2) = x \in P_2.\)

### 1.2 The Basis and Dimension of a Vector Space

If \(V\) is a vector space and \(S = \{x_1, x_2, \ldots, x_n\}\) is a set of vectors in \(V\) such that
S is a linearly independent set
A set of vectors is linearly independent if none of them can be written as a linear combination of finitely many other vectors in the set.

and

S spans V
A span is a set of vectors ∈ V for which all other vectors ∈ V can be written as a linear combination of the vectors in the span.

then S is a basis for V.

The dimension of V is n.

Example 1.2.1 A basis for Example 1.1.1, where V = ℝ³, is i = (1, 0, 0), j = (0, 1, 0) and k = (0, 0, 1) and the dimension of the space is 3.

Example 1.2.2 Example 1.1.2 has infinite dimension as no finite set spans C(−∞, ∞).

Example 1.2.3 A basis for Example 1.1.3, where V = Pₙ, is the set of monomials 1, x, x², . . ., xₙ. The dimension of the space is n + 1.

1.3 Normed Linear Spaces

1.3.1 Vector Norm

We require some method to measure the ‘magnitude’ of a vector for error analysis. We generalise the concept of a ‘length’ of a vector in 3D to n dimensions by defining a norm.

Given a vector/linear space V, then a norm, denoted by ∥x∥ for x ∈ V, is a real number such that

\[
\begin{align*}
\|x\| &> 0, \forall x \neq 0, \ (\|0\| = 0) \\
\|\alpha x\| &= |\alpha|\|x\|, \ \alpha \in \mathbb{R} \\
\|x + y\| &\leq \|x\| + \|y\|.
\end{align*}
\]

The norm is a measure of the size of the vector x where Equation (1.1) requires the size to be positive, Equation (1.2) requires the size to be scaled as the vector is scaled, and Equation (1.3) is known as the triangular inequality and has its origins in the notion of distance in ℝ³.

Any mapping of an n-dimensional Vector Space onto a subset of ℝ that satisfies these three requirements can be called a norm. The space together with a defined norm is called a Normed Linear Space.

Example 1.3.1 For the vector space V = ℝⁿ with x ∈ V given by x = (x₁, x₂, . . ., xₙ) an obvious
definition of a norm is
\[ \|x\| = \left( x_1^2 + x_2^2 + \cdots + x_n^2 \right)^{1/2}. \]

This is just a generalisation of the normed linear space \( V = \mathbb{R}^3 \) with the norm defined as the magnitude of the vector \( x \in \mathbb{R}^3 \).

**Example 1.3.2** Another norm on \( V = \mathbb{R}^n \) is
\[ \|x\| = \max_{1 \leq i \leq n} \{|x_i|\}, \]
and it is easy to verify that the three axioms are obeyed (see Tutorial sheet 1).

**Example 1.3.3** Let \( V = C[a, b] \), the space of all continuous functions on the interval \([a, b]\), and define
\[ \|f\| = \left\{ \int_a^b (f(x))^2 \, dx \right\}^{1/2}. \]
This is also a normed linear space.

### 1.3.2 Inner Product Space

One of the most familiar norms is the magnitude of a vector, \( x = (x_1, x_2, x_3) \in \mathbb{R}^3 \), (Example 1.3.1). This norm is commonly denoted \(|x|\) and equals
\[ \|x\| = |x| = \sqrt{x_1^2 + x_2^2 + x_3^2} = \sqrt{x \cdot x}. \]
i.e. it is equal to the dot product of \( x \). The dot product of a vector is an example of an Inner Product. Some other norms are also inner products.

In §1.3.1 we presented examples of Normed Linear Spaces. Two of these can be obtained from the alternative route of Inner Product Spaces and it is useful to observe this fact as it gives access to an important result, namely the Cauchy-Schwarz Inequality. For such spaces
\[ \|x\| = \{\langle x, x \rangle\}^{1/2} \]
where
\[ \langle x, y \rangle = \sum_{i=1}^n x_i y_i \quad \text{in Example 1.3.1} \]
and
\[ \langle f, g \rangle = \int_a^b f(x)g(x) \, dx \quad \text{in Example 1.3.3}. \]
Hence, inner products give rise to norms but not all norms can be cast in terms of inner products.

The formal definition of an Inner Product is given on Tutorial Sheet 1 and it is easy to show that the above two examples satisfy the stated requirements. The Cauchy-Schwarz Inequality is given by
\[ \langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle \]
and can be used to confirm the triangular inequality for norms. For example, given

\[ \|x\| = \{\langle x, x \rangle \}^{1/2} \]
\[ \|x + y\|^2 = \langle x + y, x + y \rangle \]
\[ = \langle x, x \rangle + \langle y, y \rangle + 2\langle x, y \rangle \]
\[ \leq \langle x, x \rangle + \langle y, y \rangle + 2\sqrt{\langle x, x \rangle \langle y, y \rangle} \quad \text{(using C−S inequality)} \]
\[ \|x + y\|^2 \leq (\|x\| + \|y\|)^2 \quad (1.4) \]
and hence the triangular inequality holds for all such norms. A particular example is given on Tutorial Sheet 1.

1.3.3 Commonly Used Norms and Normed Linear Spaces

In theory, any mapping of an \(n\)-dimensional space onto real numbers that satisfies (1.1)-(1.3) is a norm. In practice we are only concerned with a small number of Normed Linear Spaces and the most frequently used are the following:

- 1. The linear space \(\mathbb{R}^n\) (Euclidean Space) where \(x = (x_1, x_2, \ldots, x_i, \ldots, x_n)\) together with the norm
  \[ \|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \quad p \geq 1, \]
  is known as the \(L_p\) normed linear space. The most common are the one norm, \(L_1\), and the two norm, \(L_2\), linear spaces where \(p = 1\) and \(p = 2\), respectively.

- 2. The other standard norm for the space \(\mathbb{R}^n\) is the infinity, or maximum, norm given by
  \[ \|x\|_\infty = \max_{1 \leq i \leq n} (|x_i|). \]
  The vector space \(\mathbb{R}^n\) together with the infinity norm is commonly denoted \(L_\infty\).

**Example 1.3.4** Consider the vector \(x = (3, -1, 2, 0, 4)\), which belongs to the vector space \(\mathbb{R}^5\). Then determine its (i) one norm, (ii) two norm and (iii) infinity norm.

(i) One norm: \(\|x\|_1 = |3| + |-1| + |2| + |0| + |4| = 10.\)

(ii) Two norm: \(\|x\|_2 = \sqrt{|3|^2 + |-1|^2 + |2|^2 + |0|^2 + |4|^2} = \sqrt{30} \)

(iii) Infinity norm: \(\|x\|_\infty = \max(|3|, |-1|, |2|, |0|, |4|) = 4 \)

Note: Each is a different way of measuring the size of a vector \(x \in \mathbb{R}^n\).

- For \(C[a, b]\) (Continuous functions) the corresponding norms are
  \[ \|f(x)\|_p = \left( \int_a^b |f(x)|^p \, dx \right)^{1/p}, \quad p \geq 1, \]
with \( p = 1 \) and \( p = 2 \) the usual cases, plus

\[
\|f(x)\|_{\infty} = \max_{a \leq x \leq b} |f(x)|.
\]

### 1.4 Sub-ordinate Matrix Norms

We are interested in analysing methods for solving linear systems so we need to be able to measure the size of vectors in \( \mathbb{R}^n \) and any associated matrices in a compatible way. To achieve this end, we define a **Sub-ordinate Matrix Norm**.

For the Normed Linear Space \( \{\mathbb{R}^n, \|\cdot\|\} \), where \( \|\cdot\| \) is some norm, we define the norm of the matrix \( A_{n \times n} \) which is sub-ordinate to the vector norm \( \|\cdot\| \) as

\[
\|A\| = \max_{\|x\| \neq 0} \left( \frac{\|Ax\|}{\|x\|} \right).
\]

Note, \( Ax \) is a vector, \( x \in \mathbb{R}^n \Rightarrow Ax \in \mathbb{R}^n \), so \( \|A\| \) is the largest value of the vector norm of \( Ax \) normalised over all non-zero vectors \( x \).

Not surprisingly, the three requirements of a vector norm (1.1 – 1.3) are properties of \( \|A\| \). There are two further properties which are a consequence of the definition for \( \|A\| \). Hence, sub-ordinate matrix norms satisfy the following **five rules**:

1. \( \|A\| > 0, \ A \neq 0, \ (\|0\| = 0) \), where \( 0 \) is the \( n \times n \) zero matrix,
2. \( \|\alpha A\| = |\alpha| \|A\|, \ \alpha \in \mathbb{R} \),
3. \( \|A + B\| \leq \|A\| + \|B\| \),
4. \( \|Ax\| \leq \|A\| \|x\| \),
5. \( \|AB\| \leq \|A\| \|B\| \).

These five rules, together with the three for vector norms (1.1 – 1.3), provide the means for an analysis of a linear system. First, let us justify the above results:

- **Rule 1 (1.5)**: For a matrix \( A_{n \times n} \) where \( A \neq 0 \), \( \exists x \in \mathbb{R}^n \) where \( x \neq 0 \) such that the vector \( Ax \neq 0 \). So both \( \|Ax\| > 0 \) and \( \|x\| > 0 \), thus \( \|A\| > 0 \).

  Also, if \( A = 0 \) then \( Ax = 0 \ \forall x \) and \( \|A\| = 0 \).

- **Rule 2 (1.6)**: Since \( Ax \) is a vector and therefore satisfies (1.2) we can show that

\[
\|\alpha A\| = \max_{\|x\| \neq 0} \left\{ \frac{\|\alpha Ax\|}{\|x\|} \right\} = \max_{\|x\| \neq 0} \left\{ \frac{|\alpha| \|Ax\|}{\|x\|} \right\} = |\alpha| \max_{\|x\| \neq 0} \left\{ \frac{\|Ax\|}{\|x\|} \right\} = |\alpha| \|A\|
\]

\( \square \)
• Rule 3 (1.7):

$$\|A + B\| = \max_{\|x\| \neq 0} \left\{ \frac{\|(A + B)x\|}{\|x\|} \right\} = \frac{\|(A + B)x_m\|}{\|x_m\|}$$

where $x_m$ ($x_m \in \mathbb{R}^n$) maximises the right hand side.

Now, let us define $Ax_m = y_m$ and $Bx_m = z_m$ ($y_m, z_m \in \mathbb{R}^n$) so that

$$\|(A + B)x_m\| = \|Ax_m + Bx_m\| = \|y_m + z_m\| \leq \|y_m\| + \|z_m\| = \|Ax_m\| + \|Bx_m\|.$$ 

therefore

$$\|A + B\| \leq \|A\| + \|B\|.$$ 

□

• Rule 4 (1.8): By definition,

$$\|A\| \geq \frac{\|Ax\|}{\|x\|} \text{ for any } x \neq 0.$$ 

Hence

$$\|A\| \|x\| \geq \|Ax\| \quad \forall x \quad (\text{including } x \equiv 0)$$ 

□

• Rule 5 (1.9): Finally, consider $\|AB\|$, where $A$ and $B$ are $n \times n$ matrices. Using a similar argument to that used in rule 3, we have

$$\|AB\| = \max_{\|x\| \neq 0} \left\{ \frac{\|ABx\|}{\|x\|} \right\} = \frac{\|ABx_m\|}{\|x_m\|}$$

where $x_m$ ($x_m \in \mathbb{R}^n$) maximises the right hand side.

If we define $Bx_m = z_m$ ($z_m \in \mathbb{R}^n$) and use rule 4 (1.8), then

$$\|ABx_m\| = \|Az_m\| \leq \|A\| \|z_m\| = \|A\| \|Bx_m\| \leq \|A\| \|B\| \|x_m\|.$$ 

Hence

$$\|AB\| = \frac{\|ABx_m\|}{\|x_m\|} \leq \frac{\|A\| \|B\| \|x_m\|}{\|x_m\|} = \|A\| \|B\|.$$ 

□

1.4.1 Commonly Used Sub-ordinate Matrix Norms

Clearly, the sub-ordinate matrix norm is not easy to calculate direct from its definition as in theory all vectors $x \in \mathbb{R}^n$ should be considered. Here we will consider the commonly used matrix norms and consider practical ways to calculate them.
The easiest matrix norms to compute are matrix norms subordinate to the $L_1$ and $L_\infty$ vector norms. These are

$$\| A \|_1 = \max_{x \neq 0} \left( \frac{\sum_{i=1}^{n} |(Ax)_i|}{\sum_{i=1}^{n} |x_i|} \right)$$

$\| A \|_1$ which is equivalent to the maximum column sum of absolute values of $A$

and

$$\| A \|_\infty = \max_{x \neq 0} \left( \frac{\max_{1 \leq i \leq n} |(Ax)_i|}{\max_{1 \leq i \leq n} |x_i|} \right)$$

$\| A \|_\infty$ which is equivalent to the maximum row sum of absolute values of $A$.

Proof that $\| A \|_1$ is equivalent to the maximum column sum of absolute values of $A$ (proof of the relation for $\| A \|_\infty$ is similar)

The $L_1$ norm is

$$\| x \|_1 = \sum_{i=1}^{n} |x_i| .$$

Let $y = Ax$ ($x, y \in \mathbb{R}^n$ and $A_{n \times n}$) and so $y_i = \sum_{j=1}^{n} a_{ij} x_j$ and

$$\| Ax \|_1 = \| y \|_1 = \sum_{i=1}^{n} |y_i| = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} a_{ij} x_j \right\} .$$

Now

$$\left| \sum_{j=1}^{n} a_{ij} x_j \right| \leq \sum_{j=1}^{n} |a_{ij} x_j| \leq \sum_{j=1}^{n} |a_{ij}| |x_j| .$$

Thus

$$\| Ax \|_1 \leq \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} |a_{ij}| |x_j| \right\} .$$

Changing the order of summation then gives

$$\| Ax \|_1 \leq \sum_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}| |x_j|$$

$$\leq \sum_{j=1}^{n} |x_j| \left\{ \sum_{i=1}^{n} |a_{ij}| \right\}$$

where $\sum_{i=1}^{n} |a_{ij}|$ is the sum of absolute values in column $j$.

Each column will have a sum, $S_j = \sum_{i=1}^{n} |a_{ij}|$, $1 \leq j \leq n$ and $S_j \leq S_m$, where $m$ is the column with the largest sum. So

$$\| Ax \|_1 \leq \sum_{j=1}^{n} |x_j| S_j \leq S_m \sum_{j=1}^{n} |x_j| = S_m \| x \|_1$$

and hence

$$\frac{\| Ax \|_1}{\| x \|_1} \leq S_m .$$
where $S_m$ is the maximum column sum of absolute values. This is true for all non-zero $x$. Hence, $\|A\| \leq S_m$.

We need to determine now if $S_m$ is the maximum value of $\|Ax\|_1$, or whether it is merely an upper bound. In other words, is it possible to pick an $x \in \mathbb{R}^n$ for which we reach $S_m$?

Try by choosing the following:

$$x = [0, \ldots, 0, 1, 0, \ldots, 0]^T \quad \text{where } x_j = \begin{cases} 0 & j \neq m \\ 1 & j = m \end{cases}.$$  

Now, substituting this particular $x$ into $Ax = y$ implies $y_i = \sum_{j=1}^{n} a_{ij}x_j = a_{im}$ and thus

$$\|Ax\|_1 = \|y\|_1 = \sum_{i=1}^{n} |y_i| = \sum_{i=1}^{n} |a_{im}| = S_m.$$  

So,

$$\frac{\|Ax\|_1}{\|x\|_1} = \frac{S_m}{1} = S_m.$$  

This means that the bound can be reached with a suitable $x$ and so

$$\|A\|_1 = \max_{\|x\|_1 \neq 0} \left( \frac{\|Ax\|_1}{\|x\|_1} \right) = S_m,$$  

where $S_m$ is the maximum column sum of absolute values. Hence, $S_m$ is not just an upper bound, it is actually the maximum!

**Example 1.4.1:** Determine the matrix norm sub-ordinate to (i) the one norm and (ii) the infinity norm for the following matrices:

$$A = \begin{bmatrix} 3 & -6 & 2 \\ 2 & 5 & 1 \\ -3 & 2 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 2 & 1 \\ 2 & -3 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

- $\|A\|_1 = \max((|3| + |2| + | - 3|), (| - 6| + |5| + |2|), (|2| + |1| + |2|)),$
  - $= \max(8, 13, 5) = 13$.

- $\|B\|_1 = \max((|3| + |2| + |1|), (|2| + | - 3| + |0|), (|1| + |0| + | - 1|))$,  
  - $= \max(6, 5, 2) = 6$.

- $\|A\|_{\infty} = \max((|3| + | - 6| + |2|), (|2| + |5| + |1|), (| - 3| + |2| + |2|))$,  
  - $= \max(11, 8, 7) = 11$.

- $\|B\|_{\infty} = \max((|3| + |2| + |1|), (|2| + | - 3| + |0|), (|1| + |0| + | - 1|))$,  
  - $= \max(6, 5, 2) = 6$.

Note, since $B$ is symmetric, the maximum column sum of absolute values equals the maximum row sum of absolute values. i.e. $\|B\|_1 = \|B\|_{\infty}$.  

12
1.5 Spectral Radius

A useful, and important, quantity associated with matrices is the Spectral Radius of a matrix. An $n \times n$ matrix $\mathbf{A}$ has $n$ eigenvalues $\lambda_i, i = 1, \ldots, n$. The Spectral Radius of $\mathbf{A}$, which is denoted by $\rho(\mathbf{A})$ is defined as

$$\rho(\mathbf{A}) = \max_{1 \leq i \leq n} (|\lambda_i|).$$

Note, that $\rho(\mathbf{A}) \geq 0$ for all $\mathbf{A} \neq \mathbf{0}$. Furthermore,

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|,$$

for all sub-ordinate matrix norms.

**Proof** Let $\mathbf{e}_i$ be an eigenvector of $\mathbf{A}$, so $\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{e}_i$ where $\lambda_i$ is the associated eigenvalue. Hence, $\mathbf{e}_i \neq \mathbf{0}$ and $\lambda_i \neq 0$. Then

$$\|\mathbf{A}\mathbf{e}_i\| = \|\lambda_i \mathbf{e}_i\| = |\lambda_i| \|\mathbf{e}_i\|$$

and

$$\|\mathbf{A}\mathbf{e}_i\| \leq \|\mathbf{A}\| \|\mathbf{e}_i\|,$$

so

$$|\lambda_i| \|\mathbf{e}_i\| \leq \|\mathbf{A}\| \|\mathbf{e}_i\|,$$

or

$$|\lambda_i| \leq \|\mathbf{A}\|.$$

Now this result holds true $\forall \lambda_i$ and hence

$$\rho(\mathbf{A}) = \max_{1 \leq i \leq n} (|\lambda_i|) \leq \|\mathbf{A}\|.$$

Note, the spectral radius is **not** a norm! This can be easily shown.

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix}$$

which gives

$$\rho(\mathbf{A}) = \max(|1|, |0|) = 1 \quad \text{and} \quad \rho(\mathbf{B}) = \max(|0|, |1|) = 1.$$

However,

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad \rho(\mathbf{A} + \mathbf{B}) = \max(|1|, |3|) = 3$$

hence $\rho(\mathbf{A} + \mathbf{B}) \ngeq \rho(\mathbf{A}) + \rho(\mathbf{B})$.

So the spectral radius does not satisfy (1.3), rule 3 for a norm.
### 1.6 The Condition Number and Ill-Conditioned Matrices

Suppose we are interested in solving

\[
Ax = b \quad \text{where } A \sim n \times n, \ x \sim n \times 1 \ & \ b \sim n \times 1
\]

and the elements of \( b \) are subject to small errors \( \delta b \) so that instead of finding \( x \), we find \( \tilde{x} \) where

\[
A\tilde{x} = b + \delta b.
\]

How big is the disturbance in \( x \) in relation to the disturbance in \( b \)? (i.e. How sensitive is this system of linear equations?)

\[
Ax = b \quad \text{and} \quad A\tilde{x} = b + \delta b
\]

so,

\[
A(\tilde{x} - x) = \delta b
\]

and hence the critical change in \( x \) is given by

\[
(\tilde{x} - x) = A^{-1}\delta b.
\]

We are interested in the size of the errors, so let us take the norm:

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

for some norm. This gives a measure of the actual change in the solution. However, a more useful measure is the relative change and so we also consider,

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

so

\[
\frac{1}{\|x\|} \leq \frac{\|A\| \|A^{-1}\|}{\|b\|}.
\]

Hence, the relative change in \( x \), given by

\[
\frac{\|\tilde{x} - x\|}{\|x\|} \leq \|A\| \|A^{-1}\| \frac{\|\delta b\|}{\|b\|},
\]

is at worst the relative change in \( b \) times \( \|A\| \|A^{-1}\| \).

The quantity \( \|A\| \|A^{-1}\| = K(A) \) is called the condition number of \( A \) with respect to the norm chosen. The condition number gives an indication of the potential sensitivity of a linear system of equations. The smaller the conditional number, \( K(A) \), the smaller the change in \( x \). If \( K(A) \) is very large, the solution \( \tilde{x} \), is considered unreliable.

We would like the condition number to be small, but it is easy to see that it will never be \( < 1 \).
Note, \( AA^{-1} = I \) and thus \( \|AA^{-1}\| = \|I\| \). Using Rule 5 (1.9), we find
\[ \|AA^{-1}\| = \|I\| \leq \|A\| \|A^{-1}\| = K(A) \]
and using the definition of the subordinate matrix norm:
\[ \|I\| = \max_{\|x\| \neq 0} \left( \frac{\|Ix\|}{\|x\|} \right) = \max \left( \frac{\|x\|}{\|x\|} \right) = 1. \]
Hence,
\[ K(A) \geq 1. \]
If \( K(A) \) is modest then we can be confident that small changes to \( b \) do not seriously affect the solution. However, if \( K(A) \gg 1 \) (very large) then small changes to \( b \) “may” produce large changes in \( x \). When this happens, we say the system is ill-conditioned.

How large is 'large'? This all depends on the system. Remember, the condition number helps in finding the relative error:
\[ \frac{\|\tilde{x} - x\|}{\|x\|} \leq \|A\| \|A^{-1}\| \|\delta b\| \|b\|. \]
If the relative error of \( \|\delta b\|/\|b\| = 10^{-6} \) and we wish to know \( x \) to within a relative error of \( b = 10^{-4} \) then provided the condition number \( K(A) < 100 \) then it would be considered small.

### 1.6.1 An Example of Using Norms to Solve a Linear System of Equations

**Example 1.6.1:** We will consider an innocent looking set of equations that result in a large condition number.

Consider the linear system \( Wx = b \), where \( W \) is the Wilson Matrix,
\[
W = \begin{bmatrix}
10 & 7 & 8 & 7 \\
7 & 5 & 6 & 5 \\
8 & 6 & 10 & 9 \\
7 & 5 & 9 & 10
\end{bmatrix}
\]
and the vector \( x = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \).

so that
\[
Wx = b = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix}
\]
Now suppose we actually solve
\[
W\tilde{x} = b + \delta b = \begin{bmatrix} 32.1 \\ 22.9 \\ 33.1 \\ 30.9 \end{bmatrix}
\]
hence \( \delta b = \begin{bmatrix} +0.1 \\ -0.1 \\ +0.1 \\ -0.1 \end{bmatrix} \).
First, we must find the inverse of $W$,

$$W^{-1} = \begin{bmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{bmatrix}.$$ 

Then from $W\tilde{x} = b + \delta b$, we find

$$\tilde{x} = W^{-1}b + W^{-1}\delta b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 8.2 \\ -13.6 \\ 3.5 \\ -2.1 \end{bmatrix}.$$ 

It is clear that the system is sensitive to changes: a small change to $b$ has had a very large effect on the solution. So the Wilson Matrix is an example of an ill-conditioned matrix and we would expect the condition number to be very large.

To evaluate the condition number, $K(W)$ for the Wilson Matrix we need to select a particular norm. First, we select the 1-norm (maximum column sum of absolute values) and estimate the error

$$\|W\|_1 = 33 \quad \text{and} \quad \|W^{-1}\|_1 = 136,$$

and hence,

$$K_1(W) = \|W\|_1 \times \|W^{-1}\|_1 = 33 \times 136 = 4488,$$

which of course is considerably bigger than 1.

Remember that

$$\frac{\|\tilde{x} - x\|_1}{\|x\|_1} \leq K_1(W) \frac{\|\delta b\|_1}{\|b\|_1},$$

such that the error in $x \leq 4488 \times \frac{(0.1 + |0.1| + 0.1 + |-0.1|)}{(|0.1| + |0.1| + |0.1| + |0.1|)} = 4488 \times \frac{0.4}{119} \approx 15.$$

So far, we have used the 1-norm but it is more natural to look at the $\infty$-norm, which deals with the maximum values.

Since $W$ is symmetric, $\|W\|_1 = \|W\|_\infty$ and $\|W^{-1}\|_1 = \|W^{-1}\|_\infty$ so,

$$\frac{\|\tilde{x} - x\|_\infty}{\|x\|_\infty} \leq K_\infty(W) \frac{\|\delta b\|_\infty}{\|b\|_\infty} \leq 4488 \times \frac{\max(|0.1|, |0.1|)}{\max(|0.1|, |0.1|, |0.1|, |0.1|)} = 4488 \times \frac{0.1}{33} \approx 13.6.$$
1.7 Some Useful Results for Determining the Condition Number

1.7.1 Finding Norms of Inverse Matrices

To estimate condition numbers of matrices we need to find norms of inverses. To find an inverse, we have to solve a linear system and if the linear system is sensitive, how reliable is the value we get for $A^{-1}$? Ideally, we wish to estimate $\|A^{-1}\|$ without needing to find $A^{-1}$. One possible result that might help is the following:

Suppose $B$ is a non-singular matrix, $BB^{-1} = I$.

Write $B = I + A$ so

$$\begin{align*}
(I + A)(I + A)^{-1} &= I, \\
(I + A)^{-1} + A(I + A)^{-1} &= I,
\end{align*}$$

so

$$(I + A)^{-1} = I - A(I + A)^{-1}. \quad (1.7)$$

Taking norms and, using Rules 3 & 5 (1.7) and (1.9), we find

$$\begin{align*}
\|(I + A)^{-1}\| &= \|I - A(I + A)^{-1}\| \\
&\leq \|I\| + \|A(I + A)^{-1}\| \\
&\leq 1 + \|A\| \|(I + A)^{-1}\|,
\end{align*}$$

so

$$\|(I + A)^{-1}\| (1 - \|A\|) \leq 1. \quad (1.10)$$

If $\|A\| \leq 1$ then

$$\|B^{-1}\| = \|(I + A)^{-1}\| \leq \frac{1}{1 - \|A\|},$$

where $B = I + A$.

**Example 1.7.1:** Consider the matrix

$$B = \begin{bmatrix}
1 & 1/4 & 1/4 \\
1/4 & 1 & 1/4 \\
1/4 & 1/4 & 1
\end{bmatrix}$$
then
\[ A = B - I = \begin{bmatrix} 0 & 1/4 & 1/4 \\ 1/4 & 0 & 1/4 \\ 1/4 & 1/4 & 0 \end{bmatrix} . \]

Using the infinity norm, \( \|A\|_\infty = \frac{1}{2} < 1 \), so
\[ \|B^{-1}\|_\infty \leq \frac{1}{1 - 1/2} = 2 , \]
and \( \|B\|_\infty = 1.5 \). Hence,
\[ K_\infty(B) = \|B\|_\infty\|B^{-1}\|_\infty \leq 1.5 \times 2 \]
or \( K_\infty(B) \leq 3 \), which is quite modest.

1.7.2 Limits on the Condition Number

The Spectral Radius of a matrix states that sub-ordinate matrix norms are always greater than (or equal to) the absolute values of the eigenvalues. This is true for all eigenvalues. Suppose
\[ |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{n-1}| \geq |\lambda_n| > 0 \]
i.e. there is a largest and a smallest eigenvalue. Then, the spectral radius for all norms implies that,\[ \|A\| \geq |\lambda_1| = \max \{|\text{eigenvalues}\| = \rho(A) . \]

Furthermore, if the eigenvalues of \( A \) are \( \lambda_i \) and \( A \) is non-singular, then from \( A e_i = \lambda e_i \) we have
\[ \frac{1}{\lambda_i} e_i = A^{-1} e_i , \]
so the eigenvalues of \( A^{-1} \) are \( \lambda_i^{-1} \) (\( A \) non-singular \( \Rightarrow \lambda \neq 0 \)). Hence, the spectral radius implies
\[ \|A^{-1}\| \geq \max_i \left| \frac{1}{\lambda_i} \right| \]
or
\[ \|A^{-1}\| \geq \frac{1}{|\lambda_n|} . \]

Thus, the condition number for \( A \),
\[ K(A) = \|A\|\|A^{-1}\| \geq \frac{|\lambda_1|}{|\lambda_n|} , \]
where \( \frac{|\lambda_1|}{|\lambda_n|} \) is the ratio of the largest to the smallest eigenvalue. This ratio gives an indication of just how big the condition number can be, i.e. if the ratio is large, then \( K(A) \) will be large (and the matrix is ill-conditioned). Also, this result illustrates another simple result, namely that scaling a matrix does not affect the condition number
\[ K(\alpha A) = K(A) . \]
1.8 Examples of Ill-Conditioned Matrices

Example 1.8.1: An almost singular matrix is ill-conditioned.

Take

\[ A = \begin{bmatrix} 1 & 1 + 10^{-7} \\ 1 & 1 \end{bmatrix} \]

then \( \text{det}(A) = -10^{-7} \), which is very small. The eigenvalues of \( A \) are given by

\[
(1 - \lambda)^2 - (1 + 10^{-7}) = 0,
\]

or, using the expansion \((1 + x)^{1/2} \approx 1 + \frac{1}{2}x + \frac{1}{8}x^2 + \ldots\),

\[
(1 - \lambda) = \pm(1 + 10^{-7})^{1/2} \approx \pm(1 + \frac{1}{2}10^{-7}).
\]

Hence, we find eigenvalues

\[
\lambda_1 \approx 2 \quad \text{and} \quad \lambda_2 \approx -\frac{1}{2}10^{-7}
\]

and hence the condition number \( K(A) = \frac{|\lambda_1|}{|\lambda_n|} \geq 4 \times 10^7 \). So as you might expect, an almost singular matrix is ill-conditioned.

Example 1.8.2: The Hilbert Matrix is notorious for being ill-conditioned

Consider the \( n \times n \) Hilbert matrix

\[
H = \begin{bmatrix} 1 & 1/2 & 1/3 & \ldots & 1/n \\ 1/2 & 1/3 & 1/4 & \ldots & \vdots \\ 1/3 & 1/4 & 1/5 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/n & \ldots & \ldots & 1/(2n-1) \end{bmatrix}_{n \times n}
\]

The condition numbers for various \( H_{n \times n} \) matrices are given in the table below.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( K_\infty(H_{n \times n}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>748</td>
</tr>
<tr>
<td>4</td>
<td>( 2.8 \times 10^4 )</td>
</tr>
<tr>
<td>5</td>
<td>( 9.4 \times 10^5 )</td>
</tr>
<tr>
<td>6</td>
<td>( 2.9 \times 10^7 )</td>
</tr>
</tbody>
</table>

Typical single precision, floating point accuracy on a computer is approximately \( 0.6 \times 10^{-7} \). Apply Gaussian Elimination on a \( 6 \times 6 \) Hilbert matrix on a computer with single precision arithmetic and the result is likely to be subject to significant error! Such error can be avoided with packages like
Maple which allows you to perform calculations using exact arithmetic, i.e. all digits will be carried in the calculations. However, error can still occur if the matrix is not specified properly. If the elements are not generated exactly, the wrong result can still be obtained.