Linear Algebra Memorandum

1 Vectors and vector spaces

- **Definition**: For any $n > 0$ integer, we cand define $\mathbb{R}^n$, also knows as n-space:

$$\mathbb{R}^n = \{(v_1, v_2, \dots, v_n) | v_1, v_2, \dots, v_n \in \mathbb{R}\}$$

, where $(v_1, v_2, \dots, v_n)$ is a point in n-dimensional space with corresponding **vector** of the form

$$v = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \text{ or } \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix} \in \mathbb{R}^n$$

- **Theorem**: Let $a, b, c \in \mathbb{R}^n$ be vectors and $\alpha, \beta \in \mathbb{R}$ scalars.

  1. $a + b \in \mathbb{R}^n$ Closure under summation
  2. $a + b = b + a$ Commutativity
  3. $(a + b) + c = a + (b + c)$ Associativity
  4. $\exists 0,$ such that $a + 0 = a \quad \forall a$ Additive identity, zero vector
  5. $\forall a \exists -a$ such that $a + (-a) = 0$ Additive inverse
  6. $\alpha a \in \mathbb{R}^n$ Closure under scalar multiplication
  7. $\alpha(a + b) = \alpha a + \alpha b$ Distributivity
  8. $(\alpha + \beta)a = \alpha a + \beta a$ Distributivity
  9. $\alpha(\beta a) = (\alpha\beta)a$ Associativity of multiplication
  10. $1a = a$ Multiplicative identity

- **Definition**: Let

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$$

then the **dot product**, $u \cdot v$,is defined by

$$u \cdot v = u_1 v_1 + u_2 v_2 + \dots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

- **Theorem**: Given $u, v, w \in \mathbb{R}^n$ and scala $c \in \mathbb{R}$, the following properties hold

  1. $u \cdot v = v \cdot u$
  2. $u \cdot (v + w) = u \cdot v + u \cdot w$
  3. $c(u \cdot v) = (cu) \cdot v = u \cdot (cv) = cu \cdot v$
  4. $u \cdot u \geq 0$ and $u \cdot u = 0$ if and only if $u = 0$

- **Definition**: The **length** of a vector $v = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$ is the scalar

$$\|v\| = \sqrt{v \cdot v} = \sqrt{{v_1}^2 + {v_2}^2 + \cdots + {v_n}^2}$$

- **Definition**: A **unit** vector is a vector of length 1. For any vector $v \neq 0$, we have the corresponding unit vector $u = \frac{1}{\|v\|}v$. Finding such a unit vector is called **normalizing**.

- **Cauchy-Schwarz inequality**: Given two vectors $u, v \in \mathbb{R}^n$, then

$$|u \cdot v| \leq \|u\| \, \|v\|$$

- **The triangle inequality**: Given two vectors $u, v \in \mathbb{R}^n$, then

$$\|u + v\| \leq \|u\| + \|v\|$$

- **Definition**: For two vectors $u, v \in \mathbb{R}^n$, we define the **angle** between them by

$$\cos\theta = \frac{u \cdot v}{\|u\| \, \|v\|}$$

  .

- **Definition**: Two vectors $u, v \in \mathbb{R}^n$ are said to be **orthogonal** if $u \cdot v = 0$.

- **Pythagoras theorem**: Given two vectors $u, v \in \mathbb{R}^n$,

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

  if and only if u and v are orthogonal.

- **Definition**: A m by n **matrix** is a rectangular list with m rows and n columns.

- **Definition**: A **row vector** or **row matrix** is a $1 \times n$ matrix:

$$a = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$$

  where $a \in \mathbb{R}^{1 \times n}$

- **Definition**: A **column vector** or **column matrix** is a $1 \times n$ matrix:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix}$$

  where $a \in \mathbb{R}^{n \times 1}$

- **Definition**: Let V be a set on which two operations, **addition** and **scalar multiplication** have been defined. If $u, v \in V$, the addition is denoted by $u + v$ and if α is a scalar, then scalar multiplication is denoted by αu. If the following rules hold for all $u, v \in V$ and for all scalars α, β, then V is called a **vector space** and its elements are called **vectors**:

  1. $u + v \in V$ — Closure under addition
  2. $u + v = v + u$ — Commutativity
  3. $(u + v) + w = u + (v + w)$ — Associativity
  4. There exists an elements (zero vector), denoted 0, such that $u + 0 = u$ — Additive identity

5. For every $u \in V$ there exists an element $-u \in V$ such that $u + (-u) = 0$ Additive inverse
6. $\alpha u \in V$ Closure under scalar multiplication
7. $\alpha(u + v) = \alpha u + \alpha v$ Distributivity
8. $(\alpha + \beta)u = \alpha u + \beta u$ Distributivity
9. $\alpha(\beta u) = (\alpha\beta)u$ Associativity of multiplication
10. $1u = u$ Multiplicative identity

- **Definition**: A **subspace** W, of a vector space V, is a non-empty subset satisfying all the requirements of a vector space under the operations of addition and scalar multiplication defined in V.

- **Theorem**: If $W \subseteq V, V \neq \emptyset$, then W is a subspace of V if and only if $u + v \in W$, $\forall u, v \in W$ and $cw \in W$, $\forall w \in W$ and any scalar c.

2 Independence and orthogonality

- **Definition**: If vector u in a vector space V can be expressed in the form

$$u = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

where $c_i, i = 1, 2, \ldots, n$ are scalars, then u is called a **linear combination** of vectors $v_1, v_2, \ldots, v_n$.

- **Definition**: If $S = \{v_1, v_2, \ldots, v_k\}$ is a set of vectors in a vector space V, the set of all linear combinations of $v_1, v_2, \ldots, v_k$ is called the **span** of $v_1, v_2, \ldots, v_k$ and is denoted by $span(v_1, v_2, \ldots, v_k)$ or $span(S)$. If $V = span(S)$, then S is called a **spanning set** for V and V is said to be **spanned** by S.

- **Definition**: A set of vectors $\{v_1, v_2, \ldots, v_k\}$ are said to be **linearly independent** if

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$$

then $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$. If a set of not all zero $alpha_i$'s exists, then the set is said to be **linearly dependent**.

- **Theorem**: Let $S = \{v_1, v_2, \ldots, v_k\}$ be a set with at least two elements ($k \geq 2$). Then S is linearly dependent if and only if one can express at least one vector as a linear combinations of the other vectors in S.

- **Definition**: A **basis** for a vector space V is a set of linearly independent vectors that span V. Suppose that $S = \{v_1, v_2, \ldots, v_k\}$ is a finite subset of a vector space V. Then S is a basis of V if

  1. The v_i are linearly independent.
  2. $span(S) = V$.

  In this case, V is said to be **finite dimensional**.

- **Theorem**: If $S = \{v_1, v_2, \ldots, v_k\}$ is a basis for a vector space V then any vector in V can be written as a **unique** linear combination of vectors in S.

- **Theorem**: Given that $S = \{v_1, v_2, \ldots, v_n\}$ is a basis for a vector space V then any set containing more than n vectors is linearly dependent.

- **Theorem**: Any two bases for a vector space V contain the same number of vectors.

- **Definition**: If a vector space V has a basis with n vectors, then n is called the **dimension** of V or $dim(V) = n$.

- **Definition**: Let V be a vector space and $u, v, w \in V$. Given any scalar c the **inner product** associates a real number, denoted by $\langle u, v \rangle$ with all pairs of vectors u, v. Furthermore, it satisfies:

  1. $\langle u, v \rangle = \langle v, u \rangle$

2. $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$
3. $c\langle u, v\rangle = \langle cu, v\rangle$
4. $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ if and only if $v = 0$

A vector space with an inner product is often called a **inner product space**.

- **Definition**: Let V be a vector space with an inner product. The **length** or **norm** of a vector $u \in V$ is given by
$$\|u\| = \sqrt{\langle u, u\rangle}$$
.

- **Theorem**: The Cauchy-Schwarz inequality and the triangle inequality still hold for any vector space equipped with an inner product.

- **Definition**: Two subspaces V and W are said to be **orthogonal subspaces** if every vector in V is orthogonal to every vector in W.

- **Definition**: Given a subspace $V \in \mathbb{R}^n$, the space of all vectors orthogonal to V is called the **orthogonal complement** of V. It is denoted by $V^\perp$.

- **Definition**: A set of vectors forms an **orthogonal** set $\{q_1, q_2, \ldots, q_k\}$ if $q_i \cdot q_j = 0$ for all $i \neq j$.

- **Theorem**: The vectors of an orthogonal set are linearly independent.

- **Definition**: A set of vectors $\{q_1, q_2, \ldots, q_k\}$ are orthonormal if they are orthogonal and $q_i \cdot q_i = 1$ for all $i = 1, 2, \ldots, k$. In this case, $\{q_1, q_2, \ldots, q_k\}$ is called an **orthonormal** set.

- **Decomposition into orthogonal components** Assume that $\{q_1, q_2, \ldots, q_k\}$ is an orthogonal set of vectors and u is an arbitrary vector. Then
$$v = u - \frac{{q_1}^T u}{{q_1}^T q_1} q_1 - \cdots - \frac{{q_k}^T u}{{q_k}^T q_k} q_k$$
is orthogonal to $\{q_1, q_2, \ldots, q_k\}$. We can write $proj_{q_i} u = \frac{{q_i}^T u}{{q_i}^T q_i} q_i$ for the projection of u along q_i.

- **Gram-Schmidt orthogonalization** Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of linearly independent vectors in $\mathbb{R}^m$ and $m \geq n$. For obtaining an orthonormal set of vectors $Q = \{q_1, q_2, \ldots, q_n\}$, such that $span(Q) = span(X)$, we can use the following formulae:
$$q_j = \frac{x_j - \sum_{i=1}^{j-1} ({q_i}^T x_j) q_i}{\left\| x_j - \sum_{i=1}^{j-1} ({q_i}^T x_j) q_i \right\|}$$

- **Definition**: A matrix with orthonormal columns will be denoted by Q and is called an **orthonormal matrix**.

3 Matrices

- **Matrix addition**: Two matrices in M_{mn} can be added together as follows:
$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \ldots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \ldots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \ldots & a_{nn}+b_{nn} \end{bmatrix}$$

- **Scalar matrix multiplication**: We can multiply $\alpha \in \mathbb{R}$ with a matrix in M_{mn} as follows:

$$\alpha \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \dots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \dots & \alpha a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{n1} & \alpha a_{n2} & \dots & \alpha a_{nn} \end{bmatrix}$$

- **Matrix matrix multiplication**: If $A \in M_{mn}$ and $B \in M_{np}$, then $AB = C \in M_{mp}$ can be computed as follows:

$$AB = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{1i}b_{i1} & \sum_{i=1}^n a_{1i}b_{i2} & \dots & \sum_{i=1}^n a_{1i}b_{ip} \\ \sum_{i=1}^n a_{2i}b_{i1} & \sum_{i=1}^n a_{2i}b_{i2} & \dots & \sum_{i=1}^n a_{2i}b_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n a_{mi}b_{i1} & \sum_{i=1}^n a_{mi}b_{i2} & \dots & \sum_{i=1}^n a_{mi}b_{ip} \end{bmatrix} C$$

- **Definition**: The **transpose** of a matrix A, denoted A^T, is the matrix having the i-th column of A as it's i-th row.

- **Definition**: A matrix A is **symmetric** if $A = A^T$.

- **Definition**: The n by n **identity matrix**, denoted I_n is given by

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

.

- **Definition**: Let A be an m by n matrix, the **column space**, or **range**, denoted by $C(A)$, contains all linear combinations of the columns of A and is a subspace of $\mathbb{R}^m$.

- **Definition**: Let A be an m by n matrix, the **row space**, denoted by $R(A)$, contains all linear combinations of the rows of A and is a subspace of $\mathbb{R}^n$.

- **Theorem**: Let A be an m by n matrix. The system $Ax = b, x \in \mathbb{R}^n, b \in \mathbb{R}^m$ is solvable if and only if $b \in C(A)$.

- **Definition**: The **nullspace** of an m by n matrix A consists of all vectors x such that $Ax = 0$. It is denoted by $N(A)$ and is a subspace of $\mathbb{R}^n$.

- **Theorem**: The columns of A are linearly independent exactly when $N(A) = 0$.

- **Theorem**: The row space $R(A)$ is orthogonal to the nullspace $N(A)$. The column space $C(A)$ is orthogonal to the left nullspace $N(A^T)$.

- **Definition**: The **column rank** of a matrix is the dimension of its column space and the **row rank** of a matrix is teh dimension of its row space. The column rank and row rank are always equal:

$$dim(C(A)) = dim(R(A))$$

  We can refer to this number as the **rank** of matrix A, denoted $rank(A)$.

- **Definition**: The **inverse** of a matrix $A \in M_{nn}$, denoted by A^{-1}, satisfies

$$A^{-1}A = AA^{-1} = I_n$$

.

- **Theorem**: If the inverse exists, it is unique.

- **Definition**: An n by n **elementary matrix** is any matrix that can be obtained by performing an elementary row operation on I_n. There are three types of elementary matrices:

  - $E_{r_2 \leftarrow sr_2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix}$ scalig of a row
  - $E_{r_1 \leftrightarrow r_3} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ interchanging two rows
  - $E_{r_3 \leftarrow r_3 + cr_2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix}$ adding a multiple of one row to another

- **Row reduction by elementary matrix transformations**: By successively applying elementary row operations, every matrix can be transformed into the following forms:

  1. A matrix is in **row echelon form (REF)** if:
     - all nonzero rows are above any rows of all zero.
     - the leading coefficient (the first nonzero number from the left, also called the **pivot**) of a nonzero row is always strictly to the right of the leading coefficient of the row above it.
  2. A matrix is in **reduced row echelon form (REF)** if:
     - it is in row echelon form.
     - every leading coefficient is 1 and the only nonzero entry in its column.

  The process of transforming the matrix into REF is called **Gaussian elimination**, while the one that leads to RREF if called **Gauss-Jordan elimination**.

- **Theorem**: A matrix is invertible if and only if its RREF is the identity matrix I_n and its inverse is the sequence of elementary row operations what were applied to it to reduce it to RREF, applied to I_n.

- **Theorem**: Let A be an n by n matrix. The followin statements are equivalent:

  1. A is invertible.
  2. $Ax = b$ has a unique solution for every $b \in \mathbb{R}^n$.
  3. $Ax = 0$ has only the trivial solution.
  4. Using Gauss-Jordan elimination we may reduce the matrix to I_n.
  5. A is the product of elementary matrices.
  6. $rank(A) = n$.
  7. $N(A) = \{0\}$.
  8. The column vectors of A are linearly independent.
  9. The row vectors of A are linearly independent.

4 Systems of linear equations

- **Definition**: A system of m equations with n variables:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be represented as

$$Ax = b,$$

where $A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^m$.

- **Definition**: A system is called **homogeneous** if $b = 0 \iff Ax = 0$.
- **Definition**: A system is called **overdetermined** if $A \in \mathbb{R}^{m\times n}$ with $m > n$.
- **Definition**: A system is called **underdetermined** if $A \in \mathbb{R}^{m\times n}$ with $m < n$.
- **Definition**: A system is called **consistent** if it has at least one solution and **inconsistent** if it has no solutions.
- **Theorem**: If A is an m by n matrix,

$$n = rank(A) + dim(N(A))$$

5 Determinants and elementary matrix factorizations

- **Definition**: A permutation $\sigma \in S_n$ is said to have an **inversion** (i, j) if $\sigma(i) > \sigma(j)$.
- **Definition**: A permutation is said to be **even/odd** if the total number of inversions it has is even/odd. We can define the **sign** of the permutation to be $(-1)^{N(\sigma)}$, where $N(\sigma)$ is the number of inversions in σ.
- **Definition**: The **determinant** of a n by n matrix A is

$$det(A) = \sum_{\sigma\in S_n} sgn(\sigma)A_{1,\sigma(1)}A_{2,\sigma(2)}\dots A_{n,\sigma(n)}$$

- **Theorem**: Let A be a $n \times n$ matrix.
If A is of size 1×1 then $det(A) = A_{11}$.
Let D_{ij} be the $(N-1)\times(N-1)$ matrix obtained from A by deleting row i and column j. Set $\Delta_{ij} = (-1)^{i+j}det(D_{ij})$. Δ_{ij} is called the **cofactor** of A_{ij}.
Then, for any fixed i,

$$det(A) = \sum_{j=1}^{n} A_{ij}\Delta_{ij}$$

Alternatively, for any fixed j,

$$det(A) = \sum_{i=1}^{n} A_{ij}\Delta_{ij}$$

- **Definition**: A matrix A is called **(non-)singular** if its determinant is (not) 0.
- **Theorem**: Determinants have the following properties:

  1. $det(I_n) = 1$.
  2. $det(A_{r_i\leftrightarrow r_j}) = -det(A)$, for $i \neq j$.
  3. $det(A_{r_i\leftarrow \lambda r_i}) = \lambda det(A)$
  4. $det(A_{r_i\leftarrow r_i+v}) = det(A) + det(A_{r_i\leftarrow v})$, where v is a vector.
  5. $det(A) = 0$ if A has a row of zeros.
  6. $det(A) = A_{11}A_{22}\dots A_{nn}$ if A is lower or upper triangular.
  7. $det(A) = 0$ if A has two identical rows.
  8. $det(A_{r_i\leftarrow r_i+\lambda r_j}) = det(A)$
  9. $det(AB) = det(A)det(B)$

10. A square matrix is invertible if and only if $det(A) \neq 0$.
11. $det(A) = det(A^T)$.

- **LU factorization**: Given a matrix $A \in M_{nn}$, we can proceed as follows to reduce it to the form $A = LU$, where L is a lower triangular matrix and U an upper triangular one: use elementary row operations to reduce it A to an upper triangular matrix $E_k \dots E_2 E_1 A = U$, and then write $A = {E_1}^{-1}{E_2}^{-1} \dots {E_k}^{-1} U$. Since we have only used operations of adding a row j to a row i where $j < i$, we know that E_i is lower triangular, $i = 1, 2, \dots, k$. Hence, ${E_i}^{-1}$ is also lower triangular, and so is their product, so let $L = {E_1}^{-1}{E_2}^{-1} \dots {E_k}^{-1}$, such that $A = LU$.

- **Solving triangular systems of equations**: If L is a lower triangular matrix and we want to solve $Ly = b$:
$$\begin{bmatrix} L_{11} & 0 & \dots & 0 \\ L_{21} & L_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & \dots & L_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$
. This may be solved writing
$$y_1 = \frac{1}{L_{1,1}} b_1$$
$$y_2 = \frac{1}{L_{2,2}} (b_2 - L_{2,1} y_1)$$
$$\vdots$$
$$y_n = \frac{1}{L_{n,n}} (b_n - L_{n,1} y_1 - L_{n,2} y_2 - \dots - L_{n,n-1} y_{n-1})$$

- **Solving systems with LU factorizations**: If we want to solve $Ax = b$, A non-singular, write $A = LU$ and solve $LUx = b$. Writing $y = Ux$ allows us to write the above equation as a pair of 2 linear systems:
$$Ly = b \quad \text{and} \quad Ux = y$$
. Since both L and U are triangular, we can solve the systems with the above mentioned method.

6 Linear transformations

- **Definition**: A **linear transformation** from a vector space U to a vector space V is a mapping $T : U \to V$, such that, for all $u, v \in U$ and for all scalars c:
  1. $T(u) \in V$
  2. $T(u + v) = T(u) + T(v)$
  3. $T(cu) = cT(u)$

- **Theorem**: For any linear transformation $T : U \to V$, $T(0) = 0$.

- **Definition**: Let $u_1, u_2, \dots, u_n$ be a basis for a vector space U. Any vector $u \in U$ may then be written
$$u = \sum_{i=1}^{n} \beta_i u_i$$
with respect to the basis $u_1, u_2, \dots, u_n$. We then say that u **has the coordinates** $\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$ with respect to the basis $u_1, u_2, \dots, u_n$.

- **Definition**: Suppose U and V are vector spaces and $T: U \to V$ is a linear transformation. Given bases for U and V, there always exists a matrix A such that $T(u) = Au$ for all $u \in U$. Let $u_1, u_2, \ldots, u_m$ be a basis for U and $v_1, v_2, \ldots, v_n$ be a basis for V. Then express $T(u_j)$ as a linear combination of the basis vectors of V:

$$T(u_j) = \sum_{i=1}^{n} A_{i,j} v_i$$

. By gathering these constants in a matrix, we obtain the transformation matrix A.

- **Definition**: Suppose U and V are vector spaces and $T: U \in V$ is a linear transformation.

  - The **kernel** or **nullspace** of T, $Ker(T)$, is the set of vectors $u \in U$ such that $T(u) = 0$.
  - The **range** or **image** of T, $Im(T)$ is the set of vectors $v \in V$ such that $v = T(u)$ for some $u \in U$.

- **Theorem**: The kernel and range form vector spaces. The dimension of the kernel is known as the **nullity** of the transformation. The dimension of the range is known as the **rank** of the transformation. Also, it holds that:

$$dim(Ker(T)) + dim(Im(T)) = dim(U)$$

7 Eigenvalues and eigenvectors

- **Definition** Let A be an $n \times n$ matrix and suppose that

$$Au = \lambda u$$

for some scalar λ and non-zero vector u. We then say that λ is an **eigenvalue** of A, with corresponding eigenvector u.

- **Definition**: To find the eigenvalues of a matrix, we use the characteristic equation:

$$det(A - \lambda I) = 0$$

, which will be a polynomial of degree n in λ. This mean A can have at most n eigenvalues.

- **Theorem**: Suppose A has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ with corresponding eigenvectors $v_1, v_2, \ldots, v_n$. There are 2 useful properties:

  1. The eigenvectors $v_1, v_2, \ldots, v_n$ are linearly independent.
  2. Define the matrices S and D by

$$S = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix}, \quad D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

  We may then write $D = S^{-1}AS$ and we say A **can be diagonalized**.

- **Definition**: If A and B are $n \times n$ matrices and there exists another $n \times n$ matrix X such that

$$A = X^{-1}BX$$

, then A and B are knows as **similar matrices**.

- **Theorem**: Let A be a real, symmetric matrix of size $n \times n$. The matrix A will then have the following properties:

  1. A will have n real eigenvalues (possibly including repeated eigenvalues).
  2. Eigenvectors corresponding to different eigenvalues are orthogonal.

3. A diagonal matrix D and an orthogonal matrix P exist such that

$$D = P^T AP$$

where the columns of P are the normalized eigenvectors of A.

- **Theorem**: Let A be a matrix. If A can be diagonalized, then we can easily compute the powers of A:

$$A^n = SD^n S^{-1}$$

- **Power method**: For estimating the eigenvalues with the largest absolute value and its corresponding eigenvector, we can use the following iterative method starting with an initial guess for the eigenvector, x_0:

$$y = Ax_{k-1} \qquad M = max|y_i|$$

$$x_k = \frac{1}{M} y$$

As $k \to \infty$, x_k approaches the eigenvector corresponding to the larges eigenvalue in absolute value. This eigenvalue may then be estimated by the **Raleigh quotient**:

$$\lambda_{max} = \frac{x_k^T A x_k}{x_k^T x_k}$$

- **Perron's theorem**: Let A be a $n \times n$ matrix with all entries positive. We will refer to this as a **positive matrix**. The matrix A has a real eigenvalue λ_1 satisfying:

1. $\lambda_1 > 0$.
2. λ_1 has a corresponding positive eigenvector.
3. For any other eigenvalue λ of A, $|\lambda| \leq \lambda_1$.

- **Definition**: A matrix A is called **reducible** if, subjet to some permutations of the rows, P, and the same permutation of the columns, P^T, A can be written in the block form

$$PAP^T = \begin{bmatrix} B & C \\ O & D \end{bmatrix}$$

where B and D are square matrices. Otherwise, the matrix is said to be **irreducible**.

- **Perron-Frobenius theorem**: Let A be a $n \times n$ non-negative irreducible matrix. Then A has a unique real eigenvalue λ_1 satisfying:

1. $\lambda_1 > 0$.
2. λ_1 has a corresponding positive eigenvector.
3. For any other eigenvalue λ of A, $|\lambda| \leq \lambda_1$.
4. If $|\lambda| = \lambda_1$ then λ is a complex root of the equation $\lambda^n - \lambda_1^n = 0$.

8 Iterative methods for solving linear systems

- **Definition**: A **sparse** matrix is a matrix that has very large N, but few non-zero entries per row.
- **Iterative methods**: When using iterative methods to solve systems of equations, it will be useful to write

$$A = D - L - U$$

, where

  - $-L$ contains the entries of A strictly below the diagonal.

- D contains the entries of A on the diagonal.
- $-U$ contains the entries of A strictly above the diagonal.

Also, let u_0 be the initial guess of the iterative algorithm, u_n the guess after n steps and $e_n = u_n - u$ the error after n steps.

- **The Jacobi method**: Iterative step:

$$Du_n = (L + U)u_{n-1} + b$$

. Let

$$G = D^{-1}(L + U)$$

- **The Gauss-Seidel method**: Iterative step:

$$(D - L)u_n = Uu_{n-1} + b$$

. Let

$$G = (D - L)^{-1}U$$

- **The successive over relaxation (SOR) method**: Iterative step:

$$(D - \omega L)u_n = ((1 - \omega)D + \omega U)u_{n-1} + \omega b$$

where $\omega \in (0, 1]$ is a parameter chosen by the user. Let

$$G = (D - \omega L)^{-1}((1 - \omega)D + \omega U)$$

- **Error analysis**: In all three cases, it can be proven that $e_n = Ge_{n-1}$, which leads to

$$e_n = G^n e_0$$

. For $e_n \to 0$, we need $|\lambda_{max}| < 1$, where λ_{max} is the largest modulus eigenvalue of G. For the rate of convergence to be higher, we want $|\lambda_{max}|$ to be as small as possible. Thus, in the SOR method, we can choose ω that minimizes $|\lambda_{max}|$.

- **Definition**: A symmetric matrix A is said to be **positive definite** if $x^T Ax > 0$ for all $x \neq 0$.
- **Theorem**: A symmetric matrix is positive definite if and only if it has positive eigenvalues.
- **Theorem**: Let A be a real, symmetric, positive definite matrix of size $n \times n$. Define

$$F(x) = \frac{1}{2}x^T Ax - x^T b$$

. Then, the solution u to the system $Au = b$ is equal to the unique minimum of $F(x)$.

- **Method of steepest decent**: Iterative step:

$$r_{n-1} = b - Au_{n-1}$$

$$u_n = u_{n-1} + \frac{r_{n-1}^T r_{n-1}}{r_{n-1}^T Ar_{n-1}} r_{n-1}$$

- **Definition**: An over-determined system is a syset $Au = b$ where:
  - A is a $N \times M$ matrix with $N > M$.
  - u is a vector of length M.
  - b is a vector of length N.

- **Definition**: Since we probably can't satisfy all the equations of an over-determined system, we seek the **least squares** solution, the one that minimizes $\|r\|^2 = \|Au - b\|^2$

- **Theorem**: The least-squares solution satisfies the **normal equation**

$$A^T A u = A^T b$$

. This is a linear system of size $M \times M$.

- **QR factorization**: If $A = \begin{pmatrix} v_1 & v_2 & \dots & v_M \end{pmatrix}$ where $v_1, v_2, \dots, v_M$ are linearly independent vectors, then we can write A as

$$A = \begin{pmatrix} q_1 & q_2 & \dots & q_M \end{pmatrix} \begin{pmatrix} q_1 \cdot v_1 & q_1 \cdot v_2 & \dots & q_1 \cdot v_M \\ 0 & q_2 \cdot v_2 & \dots & q_2 \cdot v_M \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_M \cdot v_M \end{pmatrix} = QR$$

. This is called the **QR factorization** of A, where Q is orthogonal and R is upper triangular. We can use this in the normal equation, transforming it to

$$Ru = Q^T b$$

, which is an upper triangular system that may be solved easily.

- **Definition**: Let A be a non-singular $N \times N$ matrix and let v be a vector of length N. We define the **norm** of A by

$$\|A\| = \max_{v \neq 0} \frac{\|Av\|}{\|v\|}$$

- **Theorem**: For a matrix A, we have:
  - $\|A\| = \sqrt{\lambda_{max}}$
  - $\|A^{-1}\| = \sqrt{\frac{1}{\lambda_{min}}}$

- **Definition**: For a non-singular matrix M, suppose we want to solve $Mu = c$. Let δc be the error in c and δu be the error we obtain in u. Then,

$$\frac{\|\delta u\|}{\|u\|} = k(M) \frac{\|\delta c\|}{\|c\|}$$

, where $k(M)$ is the **condition number** of M,

$$k(M) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$