Review of Linear Algebra
Definition 2.1 (Matrix). With $m, n \in \mathbb{N}$ a real-valued \((m, n)\) matrix \(A\) is an \(m \cdot n\)-tuple of elements \(a_{ij}, i = 1, \ldots, m, j = 1, \ldots, n\), which is ordered according to a rectangular scheme consisting of \(m\) rows and \(n\) columns:

\[
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}, \quad a_{ij} \in \mathbb{R}.
\]  

(2.11)

By convention \((1, n)\)-matrices are called rows and \((m, 1)\)-matrices are called columns. These special matrices are also called \underline{row column vectors}.

\(\mathbb{R}^{m \times n}\) is the set of all real-valued \((m, n)\)-matrices. \(A \in \mathbb{R}^{m \times n}\) can be equivalently represented as \(a \in \mathbb{R}^{mn}\) by stacking all \(n\) columns of the matrix into a long vector; see Figure 2.4.

\[
a_1 = \begin{bmatrix} A_{11} \\
a_{21} \\
\vdots \\
a_{m1}
\end{bmatrix} \quad a_2 = \begin{bmatrix} A_{12} \\
a_{22} \\
\vdots \\
a_{m2}
\end{bmatrix}
\]
- **Associativity:**
  \[ \forall A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}, C \in \mathbb{R}^{p \times q} : (AB)C = A(BC) \]  
  \[ (2.18) \]

- **Distributivity:**
  \[ \forall A, B \in \mathbb{R}^{m \times n}, C, D \in \mathbb{R}^{n \times p} : (A + B)C = AC + BC \]
  \[ A(C + D) = AC + AD \]  
  \[ (2.19a) \]
  \[ (2.19b) \]

- **Multiplication with the identity matrix:**
  \[ \forall A \in \mathbb{R}^{m \times n} : I_m A = A I_n = A \]  
  \[ (2.20) \]

  Note that \( I_m \neq I_n \) for \( m \neq n. \)

\[ AB \in \mathbb{R}^{m \times p} \]
\[ m \times n \rightarrow n \times p \]

\[(AB)_{ij} = \sum_k A_{ik} b_{kj} \]
Definition 2.3 (Inverse). Consider a square matrix \( A \in \mathbb{R}^{n \times n} \). Let matrix \( B \in \mathbb{R}^{n \times n} \) have the property that \( AB = I_n = BA \). \( B \) is called the inverse of \( A \) and denoted by \( A^{-1} \).

Unfortunately, not every matrix \( A \) possesses an inverse \( A^{-1} \). If this inverse does exist, \( A \) is called regular/invertible/nonsingular; otherwise singular/noninvertible. When the matrix inverse exists, it is unique. In See

\[
AB = BA = I \quad \rightarrow \quad B \text{ is called the inverse of } A
\]
Definition 2.4 (Transpose). For $\mathbf{A} \in \mathbb{R}^{m \times n}$ the matrix $\mathbf{B} \in \mathbb{R}^{n \times m}$ with 
\[ b_{ij} = a_{ji} \]

is called the transpose of $\mathbf{A}$. We write $\mathbf{B} = \mathbf{A}^\top$.

In general, $\mathbf{A}^\top$ can be obtained by writing the columns of $\mathbf{A}$ as the rows of $\mathbf{A}^\top$. The following are important properties of inverses and transposes:

\[
(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1} \quad (2.26)
\]
\[
(A + B)^{-1} \neq A^{-1} + B^{-1} \quad (2.28)
\]
\[
(\mathbf{A}^\top)^\top = \mathbf{A} \quad (2.29)
\]
\[
(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top \quad (2.30)
\]
\[
(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top \quad (2.31)
\]

Definition 2.5 (Symmetric Matrix). A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$. 
Note that only \((n, n)\)-matrices can be symmetric. Generally, we call \((n, n)\)-matrices also square matrices because they possess the same number of rows and columns. Moreover, if \(A\) is invertible, then so is \(A^\top\), and
\[
(A^{-1})^\top = (A^\top)^{-1} =: A^{-\top}.
\]
\[
(\overline{A^{-1}})^\top = (A^\top)^{-1}
\]
Solving systems of linear equations

\[ a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \]
\[ \vdots \]
\[ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m, \]

\[ Ax = b, \quad \mathbb{R}^m \]

\[ n \text{ variables} \]
\[ x_1, x_2, \ldots, x_n \in \mathbb{R} \]

\[ m \text{ equations} \]

\[ A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \]
2.3 Solving Systems of Linear Equations

- Exchange of two equations (rows in the matrix representing the system of equations)
- Multiplication of an equation (row) with a constant $\lambda \in \mathbb{R}\setminus\{0\}$
- Addition of two equations (rows)
Example 2.6
For $a \in \mathbb{R}$, we seek all solutions of the following system of equations:

\[-2x_1 + 4x_2 - 2x_3 - x_4 + 4x_5 = -3 \]
\[4x_1 - 8x_2 + 3x_3 - 3x_4 + x_5 = 2 \]
\[x_1 - 2x_2 + x_3 - x_4 + x_5 = 0 \] \hspace{1cm} (2.44)
\[x_1 - 2x_2 - 3x_4 + 4x_5 = a \]
build the \textit{augmented matrix} (in the form $[A \mid b]$)

\[
\begin{bmatrix}
-2 & 4 & -2 & -1 & 4 \\
4 & -8 & 3 & -3 & 1 \\
1 & -2 & 1 & -1 & 1 \\
1 & -2 & 0 & -3 & 4 \\
\end{bmatrix}
\begin{bmatrix}
-3 \\
2 \\
0 \\
a \\
\end{bmatrix}
\]

Row echelon form

\[
\begin{bmatrix}
1 & -2 & 1 & -1 & 1 & | & 0 \\
0 & 0 & 1 & -1 & 3 & | & -2 \\
0 & 0 & 0 & 1 & -2 & | & 1 \\
0 & 0 & 0 & 0 & 0 & | & a + 1 \\
\end{bmatrix}
\]
Definition 2.6 (Row-Echelon Form). A matrix is in row-echelon form if

- All rows that contain only zeros are at the bottom of the matrix; correspondingly, all rows that contain at least one nonzero element are on top of rows that contain only zeros.
- Looking at nonzero rows only, the first nonzero number from the left (also called the pivot or the leading coefficient) is always strictly to the right of the pivot of the row above it.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>-2</th>
<th>1</th>
<th>-1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>3</td>
<td>-2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>a+1</td>
</tr>
</tbody>
</table>

Remark (Basic and Free Variables). The variables corresponding to the pivots in the row-echelon form are called basic variables and the other variables are free variables. For example, in (2.45), $x_1, x_3, x_4$ are basic variables, whereas $x_2, x_5$ are free variables.
\[\begin{align*}
x_1 - 2x_2 + x_3 - x_4 + x_5 &= 0 \\
x_3 - x_4 + 3x_5 &= -2 \\
x_4 - 2x_5 &= 1 \\
0 &= a + 1
\end{align*}\]

Only for \(a = -1\) this system can be solved. A particular solution is

\[
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_5
\end{bmatrix} =
\begin{bmatrix}
2 \\
0 \\
-1 \\
1 \\
0
\end{bmatrix}.
\tag{2.46}
\]

The general solution, which captures the set of all possible solutions, is

\[
\left\{ \mathbf{x} \in \mathbb{R}^5 : \mathbf{x} = \begin{bmatrix}
2 \\
0 \\
-1 \\
1 \\
0
\end{bmatrix} + \lambda_1 \begin{bmatrix}
2 \\
1 \\
0 \\
0 \\
0
\end{bmatrix} + \lambda_2 \begin{bmatrix}
2 \\
0 \\
-1 \\
2 \\
1
\end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R} \right\}.
\tag{2.47}
\]
Remark (Reduced Row Echelon Form). An equation system is in reduced row-echelon form (also: row-reduced echelon form or row canonical form) if

- It is in row-echelon form.
- Every pivot is 1.
- The pivot is the only nonzero entry in its column.
Complexity of Gaussian elimination: \(O(n^3)\)

of a vector space (Section 2.6.1). Gaussian elimination is an intuitive and constructive way to solve a system of linear equations with thousands of variables. However, for systems with millions of variables, it is impractical as the required number of arithmetic operations scales cubically in the number of simultaneous equations.

In practice, systems of many linear equations are solved indirectly, by either stationary iterative methods, such as the Richardson method, the Jacobi method, the Gauß-Seidel method, and the successive over-relaxation method, or Krylov subspace methods, such as conjugate gradients, generalized minimal residual, or biconjugate gradients. We refer to the books by Stoer and Burlirsch (2002), Strang (2003), and Liesen and Mehrmann (2015) for further details.
Iterative (approximate) methods

Let $x_*$ be a solution of $Ax = b$. The key idea of these iterative methods is to set up an iteration of the form

$$x^{(k+1)} = Cx^{(k)} + d$$

(2.60)

for suitable $C$ and $d$ that reduces the residual error $\|x^{(k+1)} - x_*\|$ in every iteration and converges to $x_*$. We will introduce norms $\| \cdot \|$, which allow us to compute similarities between vectors, in Section 3.1.

Basic idea: splitting $A$ into $M - N$ where $M$ is invertible. Then

$$Mx = Nx + b$$

Iterate: $x \leftarrow M^{-1}(Nx+b)$
**Definition 2.9 (Vector Space).** A real-valued vector space \( V = (\mathcal{V}, +, \cdot) \) is a set \( \mathcal{V} \) with two operations

\[
+ : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V} \tag{2.62}
\]

\[
\cdot : \mathbb{R} \times \mathcal{V} \rightarrow \mathcal{V} \tag{2.63}
\]

where

1. \((\mathcal{V}, +)\) is an Abelian group
2. Distributivity:
   1. \(\forall \lambda \in \mathbb{R}, x, y \in \mathcal{V} : \lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y\)
   2. \(\forall \lambda, \psi \in \mathbb{R}, x \in \mathcal{V} : (\lambda + \psi) \cdot x = \lambda \cdot x + \psi \cdot x\)
3. Associativity (outer operation): \(\forall \lambda, \psi \in \mathbb{R}, x \in \mathcal{V} : \lambda \cdot (\psi \cdot x) = (\lambda \psi) \cdot x\)
4. Neutral element with respect to the outer operation: \(\forall x \in \mathcal{V} : 1 \cdot x = x\)
Vector Space

\[ \vec{X} \in \mathbb{R}^n, \quad V = \mathbb{R}^2 \]

1. \( \vec{x}, \vec{y} \in V \), then \( \vec{x} + \vec{y} \in V \)
2. \( \lambda \cdot \vec{X} \in V \), for any scalar \( \lambda \).
3. \( \vec{0} + \vec{X} = \vec{X}, \quad \vec{0} \in V \)
4. For \( \vec{X} \in V \), exists \( \vec{V} \) such that \( \vec{V} + (-\vec{V}) = \vec{0} \).
5. \( \lambda \cdot \lambda' \cdot \vec{V} = (\lambda \cdot \lambda') \cdot \vec{V} \)

\[
\begin{bmatrix}
  x_1 \\
  x_2
\end{bmatrix} +
\begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix} =
\begin{bmatrix}
  x_1 + y_1 \\
  x_2 + y_2
\end{bmatrix} \in V
\]

\[
\vec{0} = \begin{bmatrix}
  0 \\
  0
\end{bmatrix}
\]

\[ \lambda \cdot \vec{X} = \begin{bmatrix}
  \lambda x_1 \\
  \lambda x_2
\end{bmatrix} \in V \]

\( (V, +, \cdot) \rightarrow \text{Vector space} \)

\( \vec{0} \)

\( \text{Set operators} \)

\( V = \{ \text{all polynomial functions of order } n \} \)

\( V = \{ \text{all } n \times n \text{ matrices} \} \)
Definition 2.10 (Vector Subspace). Let \( V = (V, +, \cdot) \) be a vector space and \( U \subseteq V, U \neq \emptyset \). Then \( U = (U, +, \cdot) \) is called vector subspace of \( V \) (or linear subspace) if \( U \) is a vector space with the vector space operations \(+\) and \( \cdot \) restricted to \( U \times U \) and \( \mathbb{R} \times U \). We write \( U \subseteq V \) to denote a subspace \( U \) of \( V \).

1. \( U \neq \emptyset \), in particular: \( 0 \in U \)
2. Closure of \( U \):
   a. With respect to the outer operation: \( \forall \lambda \in \mathbb{R}, \forall x \in U : \lambda x \in U \).
   b. With respect to the inner operation: \( \forall x, y \in U : x + y \in U \).
Subspace
Definition 2.11 (Linear Combination). Consider a vector space $V$ and a finite number of vectors $x_1, \ldots, x_k \in V$. Then, every $v \in V$ of the form

$$v = \lambda_1 x_1 + \cdots + \lambda_k x_k = \sum_{i=1}^{k} \lambda_i x_i \in V$$  

(2.65)

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $x_1, \ldots, x_k$. 

**Definition 2.12 (Linear (In)dependence).** Let us consider a vector space $V$ with $k \in \mathbb{N}$ and $x_1, \ldots, x_k \in V$. If there is a non-trivial linear combination, such that $0 = \sum_{i=1}^{k} \lambda_i x_i$ with at least one $\lambda_i \neq 0$, the vectors $x_1, \ldots, x_k$ are *linearly dependent*. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$ the vectors $x_1, \ldots, x_k$ are *linearly independent*. 
2.6.1 Generating Set and Basis

Definition 2.13 (Generating Set and Span). Consider a vector space $V = \langle V, +, \cdot \rangle$ and set of vectors $A = \{x_1, \ldots, x_k\} \subseteq V$. If every vector $v \in V$ can be expressed as a linear combination of $x_1, \ldots, x_k$, $A$ is called a generating set of $V$. The set of all linear combinations of vectors in $A$ is called the span of $A$. If $A$ spans the vector space $V$, we write $V = \text{span}[A]$ or $V = \text{span}[x_1, \ldots, x_k]$.

Any $v \in V$ can be written down as

$$v = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$
Definition 2.14 (Basis). Consider a vector space \( V = (\mathcal{V}, +, \cdot) \) and \( \mathcal{A} \subseteq \mathcal{V} \). A generating set \( \mathcal{A} \) of \( V \) is called minimal if there exists no smaller set \( \tilde{\mathcal{A}} \subseteq \mathcal{A} \subseteq \mathcal{V} \) that spans \( V \). Every linearly independent generating set of \( V \) is minimal and is called a basis of \( V \).
Let $V = (V, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq V, \mathcal{B} \neq \emptyset$. Then, the following statements are equivalent:

- $\mathcal{B}$ is a basis of $V$.
- $\mathcal{B}$ is a minimal generating set.
- $\mathcal{B}$ is a maximal linearly independent set of vectors in $V$, i.e., adding any other vector to this set will make it linearly dependent.
- Every vector $x \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique, i.e., with

$$x = \sum_{i=1}^{k} \lambda_i b_i = \sum_{i=1}^{k} \psi_i b_i$$  \hspace{1cm} (2.77)

and $\lambda_i, \psi_i \in \mathbb{R}, b_i \in \mathcal{B}$ it follows that $\lambda_i = \psi_i, \ i = 1, \ldots, k$. 
Example 2.16

- In $\mathbb{R}^3$, the canonical/standard basis is

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}. \quad (2.78)$$

- Different bases in $\mathbb{R}^3$ are

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}, \quad \mathcal{B}_2 = \left\{ \begin{bmatrix} 0.5 \\ 0.8 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 1.8 \\ 0.3 \\ 0.3 \end{bmatrix}, \begin{bmatrix} -2.2 \\ -1.3 \\ 3.5 \end{bmatrix} \right\}. \quad (2.79)$$

- The set

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\} \quad (2.80)$$

is linearly independent, but not a generating set (and no basis) of $\mathbb{R}^4$: For instance, the vector $[1, 0, 0, 0]^T$ cannot be obtained by a linear combination of elements in $\mathcal{A}$.
Remark. Every vector space \( V \) possesses a basis \( B \). The preceding examples show that there can be many bases of a vector space \( V \), i.e., there is no unique basis. However, all bases possess the same number of elements, the basis vectors.

We only consider finite-dimensional vector spaces \( V \). In this case, the dimension of \( V \) is the number of basis vectors of \( V \), and we write \( \dim(V) \). If \( U \subseteq V \) is a subspace of \( V \), then \( \dim(U) \leq \dim(V) \).
Remark. A basis of a subspace $U = \text{span}[x_1, \ldots, x_m] \subseteq \mathbb{R}^n$ can be found by executing the following steps:

1. Write the spanning vectors as columns of a matrix $A$.
2. Determine the row-echelon form of $A$.
3. The spanning vectors associated with the pivot columns are a basis of $U$. 
The number of linearly independent columns of a matrix $A \in \mathbb{R}^{m \times n}$ equals the number of linearly independent rows and is called the rank of $A$ and is denoted by $\text{rk}(A)$. 
Rank Theorem

For any m x n matrix A:

\[ \text{rk}(A) + \dim \text{Null}(A) = n \]

Num of pivot columns + num of nonpivot columns = num of columns
\[ \text{rk}(A) = \text{rk}(A^\top), \] i.e., the column rank equals the row rank.

The columns of \( A \in \mathbb{R}^{m \times n} \) span a subspace \( U \subseteq \mathbb{R}^m \) with \( \dim(U) = \text{rk}(A) \). Later we will call this subspace the image or range. A basis of \( U \) can be found by applying Gaussian elimination to \( A \) to identify the pivot columns.

The rows of \( A \in \mathbb{R}^{m \times n} \) span a subspace \( W \subseteq \mathbb{R}^n \) with \( \dim(W) = \text{rk}(A) \). A basis of \( W \) can be found by applying Gaussian elimination to \( A^\top \).

For all \( A \in \mathbb{R}^{n \times n} \) it holds that \( A \) is regular (invertible) if and only if \( \text{rk}(A) = n \).

For all \( A \in \mathbb{R}^{m \times n} \) and all \( b \in \mathbb{R}^m \) it holds that the linear equation system \( Ax = b \) can be solved if and only if \( \text{rk}(A) = \text{rk}(A|b) \), where \( A|b \) denotes the augmented system.

For \( A \in \mathbb{R}^{m \times n} \) the subspace of solutions for \( Ax = 0 \) possesses dimension \( n - \text{rk}(A) \). Later, we will call this subspace the kernel or the null space.

A matrix \( A \in \mathbb{R}^{m \times n} \) has full rank if its rank equals the largest possible rank for a matrix of the same dimensions. This means that the rank of a full-rank matrix is the lesser of the number of rows and columns, i.e., \( \text{rk}(A) = \min(m, n) \). A matrix is said to be rank deficient if it does not have full rank.
Definition 2.15 (Linear Mapping). For vector spaces $V, W$, a mapping $\Phi : V \to W$ is called a linear mapping (or vector space homomorphism/linear transformation) if

$$\forall x, y \in V \forall \lambda, \psi \in \mathbb{R} : \Phi(\lambda x + \psi y) = \lambda \Phi(x) + \psi \Phi(y). \quad (2.87)$$

- **Isomorphism:** $\Phi : V \to W$ linear and bijective
- **Endomorphism:** $\Phi : V \to V$ linear
- **Automorphism:** $\Phi : V \to V$ linear and bijective
- We define $\text{id}_V : V \to V$, $x \mapsto x$ as the identity mapping or identity automorphism in $V$.

Theorem 2.17 (Theorem 3.59 in Axler (2015)). Finite-dimensional vector spaces $V$ and $W$ are isomorphic if and only if $\dim(V) = \dim(W)$. 

Definition 2.23 (Image and Kernel).
For $\Phi : V \rightarrow W$, we define the kernel/null space
\[ \ker(\Phi) := \Phi^{-1}(0_W) = \{ v \in V : \Phi(v) = 0_W \} \] (2.122)
and the image/range
\[ \text{Im}(\Phi) := \Phi(V) = \{ w \in W | \exists v \in V : \Phi(v) = w \}. \] (2.123)

We also call $V$ and $W$ also the domain and codomain of $\Phi$, respectively.

Intuitively, the kernel is the set of vectors in $v \in V$ that $\Phi$ maps onto the neutral element $0_W \in W$. The image is the set of vectors $w \in W$ that can be "reached" by $\Phi$ from any vector in $V$. An illustration is given in Figure 2.12.
- $\Phi$ is injective (one-to-one) if and only if $\ker(\Phi) = \{0\}$. 
Fundamental theorem of linear mapping

**Theorem 2.24** (Rank-Nullity Theorem). *For vector spaces $V, W$ and a linear mapping $\Phi : V \to W$ it holds that*

$$\dim(\ker(\Phi)) + \dim(\text{Im}(\Phi)) = \dim(V). \quad (2.129)$$