Mathematics for Machine Learning

— Linear Algebra: Basis, Rank, Linear Mappings & Affine Spaces

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,
Tamkang University

Fall 2023

Credits for the resource

- The slides are based on the textbooks:
    - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
    - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.*
- We could partially refer to the monograph:
  Francesco Orabona: A Modern Introduction to Online Learning.
  https://arxiv.org/abs/1912.13213

Outline

Outline

Why linear algebra?

- Crucial in the graduate school entrance examination.

Why linear algebra?

- Crucial in the graduate school entrance examination.
- Matrix operations.

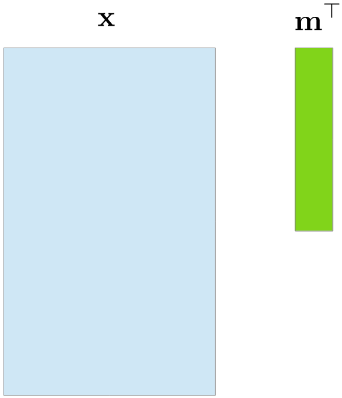
Why linear algebra?

- Crucial in the graduate school entrance examination.
- Matrix operations.
- Vectorization.

Vectorization Example (1/3)

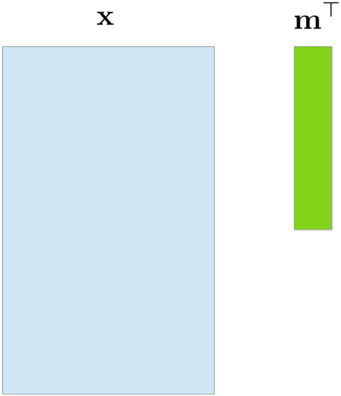
$$
\begin{aligned}
y_i &= \langle \mathbf{m}, \mathbf{x}_i \rangle \\
&= m_1 x_{i,1} + m_2 x_{i,2} + \ldots + m_k x_{i,k}.
\end{aligned}
$$

```
m = np.random.rand(1,5)
x = np.random.rand(5000000,5)
#assume k=5
```



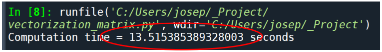
Vectorization Example (2/3)

```
start = time.time()
zer = []
for i in range(0,5000000):
    total = 0
    for j in range(0,5):
        total = total + x[i][j]*m[0][j]
    zer.append(total)
zer = np.array(zer)
end = time.time()
```



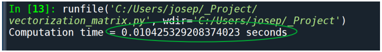
Vectorization Example (2/3)

```
start = time.time()
zer = []
for i in range(0,5000000):
    total = 0
    for j in range(0,5):
        total = total + x[i][j]*m[0][j]
    zer.append(total)
zer = np.array(zer)
end = time.time()
```



Vectorization Example (3/3)

```
start = time.time()
zer = np.matmul(x, m.T)
end = time.time()
```



Outline

Group

Group

Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ defined on $\mathcal{G}$. Then $G : (\mathcal{G}, \otimes)$ is called a group if the following hold:

Group

Group

Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ defined on $\mathcal{G}$. Then $G : (\mathcal{G}, \otimes)$ is called a group if the following hold:

1. $\forall x, y \in \mathcal{G}$, $x \otimes y \in \mathcal{G}$.

Group

Group

Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ defined on $\mathcal{G}$. Then $G : (\mathcal{G}, \otimes)$ is called a group if the following hold:

1. $\forall x, y \in \mathcal{G}$, $x \otimes y \in \mathcal{G}$.
2. $\forall x, y, z \in \mathcal{G}$, $(x \otimes y) \otimes z = x \otimes (y \otimes z)$.

Group

Group

Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ defined on $\mathcal{G}$. Then $G : (\mathcal{G}, \otimes)$ is called a group if the following hold:

1. $\forall x, y \in \mathcal{G},\ x \otimes y \in \mathcal{G}$.
2. $\forall x, y, z \in \mathcal{G},\ (x \otimes y) \otimes z = x \otimes (y \otimes z)$.
3. $\exists e \in \mathcal{G}$ such that $\forall x \in \mathcal{G},\ x \otimes e = e \otimes x = x$.

Group

Group

Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ defined on $\mathcal{G}$. Then $G : (\mathcal{G}, \otimes)$ is called a group if the following hold:

1. $\forall x, y \in \mathcal{G}$, $x \otimes y \in \mathcal{G}$.
2. $\forall x, y, z \in \mathcal{G}$, $(x \otimes y) \otimes z = x \otimes (y \otimes z)$.
3. $\exists e \in \mathcal{G}$ such that $\forall x \in \mathcal{G}$, $x \otimes e = e \otimes x = x$.
4. $\forall x \in \mathcal{G}$, $\exists y \in \mathcal{G}$ such that $x \otimes y = y \otimes x = e$. We denote by x^{-1} the inverse element of x.

Group

Group

Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ defined on $\mathcal{G}$. Then $G : (\mathcal{G}, \otimes)$ is called a group if the following hold:

1. $\forall x, y \in \mathcal{G}$, $x \otimes y \in \mathcal{G}$.
2. $\forall x, y, z \in \mathcal{G}$, $(x \otimes y) \otimes z = x \otimes (y \otimes z)$.
3. $\exists e \in \mathcal{G}$ such that $\forall x \in \mathcal{G}$, $x \otimes e = e \otimes x = x$.
4. $\forall x \in \mathcal{G}$, $\exists y \in \mathcal{G}$ such that $x \otimes y = y \otimes x = e$. We denote by x^{-1} the inverse element of x.

- If G is a group and $\forall x, y \in \mathcal{G}$ we have $x \otimes y = y \otimes x$, then G is an Abelian group.

Examples

- $(\mathbb{Z}, +)$: an Abelain group.
- $(\mathbb{N} \cup \{0\}, +)$ is NOT a group.
- $(\mathbb{Z}, \cdot)$ is NOT a group.
- $(\mathbb{R}, \cdot)$ is NOT a group.
- $(\mathbb{R} \setminus \{0\}, \cdot)$ is an Abelian group.
- $(\mathbb{R}^{m \times n}, +)$ is an Abelian group.

Vector Space

Vector Space

A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations:

$$+ : \mathcal{V} \times \mathcal{V} \mapsto \mathcal{V}$$

$$\cdot : \mathbb{R} \times \mathcal{V} \mapsto \mathcal{V}$$

where

- $(\mathcal{V}, +)$ is an Abelian group.
- Distributivity holds:
  - $\forall \lambda \in \mathbb{R}$, $\mathbf{x}, \mathbf{y} \in \mathcal{V}$: $\lambda \cdot (\mathbf{x} + \mathbf{y}) = \lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}$.
  - $\forall \lambda, \psi \in \mathbb{R}$, $\mathbf{x} \in \mathcal{V}$: $(\lambda + \psi) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \psi \cdot \mathbf{x}$.
- $\forall \lambda, \psi \in \mathbb{R}$, $\mathbf{x} \in \mathcal{V}$: $\lambda \cdot (\psi \cdot \mathbf{x}) = (\lambda\psi) \cdot \mathbf{x}$.
- $\forall \mathbf{x} \in \mathcal{V}$: $1 \cdot \mathbf{x} = \mathbf{x}$.

$\star$ Note: A vector multiplication is not defined.

Vector Subspaces

Vector Subspace

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{U} \subset \mathcal{V}$ and $\mathcal{U} \neq \emptyset$. Then $U = (\mathcal{U}, +, \cdot)$ is called a vector subspace of V if U is a vector space with the operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$ respectively.

- Denote by $U \subseteq V$ a subspace u of V.

Vector Subspaces

Vector Subspace

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{U} \subset \mathcal{V}$ and $\mathcal{U} \neq \emptyset$. Then $U = (\mathcal{U}, +, \cdot)$ is called a vector subspace of V if U is a vector space with the operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$ respectively.

- Denote by $U \subseteq V$ a subspace u of V.

Examples

- The trivial subspace of a vector space V: $\{\mathbf{0}\}$ and V.

Examples

- The trivial subspace of a vector space V: $\{\mathbf{0}\}$ and V.
- The solution set of a homogeneous system of linear equations $A\mathbf{x} = \mathbf{0}$ with n unknowns (i.e., $\mathbf{x} = [x_1, \ldots, x_n]^\top$) is a subspace of $\mathbb{R}^n$.

Examples

- The trivial subspace of a vector space V: $\{\mathbf{0}\}$ and V.
- The solution set of a homogeneous system of linear equations $A\mathbf{x} = \mathbf{0}$ with n unknowns (i.e., $\mathbf{x} = [x_1, \ldots, x_n]^\top$) is a subspace of $\mathbb{R}^n$.
- The intersection of arbitrarily many subspaces is a subspace.

Examples

- The trivial subspace of a vector space V: $\{\mathbf{0}\}$ and V.
- The solution set of a homogeneous system of linear equations $A\mathbf{x} = \mathbf{0}$ with n unknowns (i.e., $\mathbf{x} = [x_1, \ldots, x_n]^\top$) is a subspace of $\mathbb{R}^n$.
- The intersection of arbitrarily many subspaces is a subspace.
- The solution of an inhomogeneous system of linear equations $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} \neq \mathbf{0}$ is NOT a subspace of $\mathbb{R}^n$.

Linear Combination

Linear Combination

Consider a vector space V and a finite number of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$. Then, every $\mathbf{v} \in V$ of the form

$$\mathbf{v} = \lambda_1 \mathbf{x}_1 + \cdots \lambda_k \mathbf{x}_k = \sum_{i=1}^{k} \lambda_i x_i \in V$$

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$.

- **Question**: How to represent $\mathbf{0}$ as a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_k$?

Linearly Independent

Linear (In)dependence

Consider a vector space V with $k > 0$ vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$.

- If there is a nontrivial linear combination such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i x_i$ with at least one $\lambda_i \neq 0$, then we say $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent.
- If only the trivial solution exists (i.e., $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$), then we say $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent.

Recall some facts

- If at least one of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is $\mathbf{0}$ then they are linearly dependent.

Recall some facts

- If at least one of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is $\mathbf{0}$ then they are linearly dependent.
- Two identical vectors are linearly dependent.

Recall some facts

- If at least one of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is $\mathbf{0}$ then they are linearly dependent.
- Two identical vectors are linearly dependent.
- Write all vectors as rows (or columns) of a matrix and perform Gaussian elimination until the matrix is in row echelon form.

Remark (1/2)

Consider a vector space V with k linear independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k$ and m linear combinations

$$\begin{aligned}
\mathbf{x}_1 &= \sum_{i=1}^{k} \lambda_{i,1} \mathbf{b}_i \\
&\vdots \\
\mathbf{x}_m &= \sum_{i=1}^{k} \lambda_{i,m} \mathbf{b}_i
\end{aligned}$$

Remark (1/2)

Consider a vector space V with k linear independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k$ and m linear combinations

$$\begin{aligned}
\mathbf{x}_1 &= \sum_{i=1}^{k} \lambda_{i,1}\mathbf{b}_i \\
&\vdots \\
\mathbf{x}_m &= \sum_{i=1}^{k} \lambda_{i,m}\mathbf{b}_i
\end{aligned}$$

- Define $\boldsymbol{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_k]$ (i.e., a matrix), then

$$\mathbf{x}_j = \boldsymbol{B}\boldsymbol{\lambda}_j, \text{ for } \boldsymbol{\lambda}_j = \begin{bmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{kj} \end{bmatrix}, \; j = 1, \ldots, m.$$

Remark (2/2)

We want to test whether $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent.

- $\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \mathbf{0}$.

Remark (2/2)

We want to test whether $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent.

- $\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \mathbf{0}$.
- So,

$$\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \sum_{j=1}^{m} \psi_j \boldsymbol{B} \boldsymbol{\lambda}_j = \boldsymbol{B} \sum_{j=1}^{m} \boldsymbol{\lambda}_j.$$

Remark (2/2)

We want to test whether $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent.

- $\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \mathbf{0}$.
- So,

$$\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \sum_{j=1}^{m} \psi_j \boldsymbol{B} \boldsymbol{\lambda}_j = \boldsymbol{B} \sum_{j=1}^{m} \boldsymbol{\lambda}_j.$$

  - Why does the last equality hold?

Remark (2/2)

We want to test whether $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent.

- $\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \mathbf{0}$.
- So,

$$\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \sum_{j=1}^{m} \psi_j \boldsymbol{B} \boldsymbol{\lambda}_j = \boldsymbol{B} \sum_{j=1}^{m} \boldsymbol{\lambda}_j.$$

  - Why does the last equality hold?

- $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ are linearly independent iff $\{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m\}$ are linearly independent.

Remark (2/2)

We want to test whether $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent.

- $\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \mathbf{0}$.
- So,

$$\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \sum_{j=1}^{m} \psi_j \boldsymbol{B} \boldsymbol{\lambda}_j = \boldsymbol{B} \sum_{j=1}^{m} \boldsymbol{\lambda}_j.$$

  - Why does the last equality hold?

- $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ are linearly independent iff $\{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m\}$ are linearly independent.
- **Note:** m linear combinations of k vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly *dependent* if $m > k$.

Outline

Basis

Spanning/Generating

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and a set $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subseteq \mathcal{V}$.

If every vector $\mathbf{v} \in \mathcal{V}$ can be expressed as a linear combination of vectors in $\mathcal{A}$, then $\mathcal{A}$ is called a spanning set (or generating set) of V.

- $\mathcal{A}$ spans V; $\mathrm{span}(\mathcal{A}) = V$.

Basis

Spanning/Generating

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and a set $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subseteq \mathcal{V}$.

If every vector $\mathbf{v} \in \mathcal{V}$ can be expressed as a linear combination of vectors in $\mathcal{A}$, then $\mathcal{A}$ is called a spanning set (or generating set) of V.

- $\mathcal{A}$ spans V; $\text{span}(\mathcal{A}) = V$.

Basis

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and a set $\mathcal{A} \subseteq \mathcal{V}$. Then if one of the following condition holds, we say that $\mathcal{A}$ is a basis of V.

- $\mathcal{A}$ is a minimal generating set of V.

Basis

Spanning/Generating

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and a set $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subseteq \mathcal{V}$.

If every vector $\mathbf{v} \in \mathcal{V}$ can be expressed as a linear combination of vectors in $\mathcal{A}$, then $\mathcal{A}$ is called a spanning set (or generating set) of V.

- $\mathcal{A}$ spans V; $\text{span}(\mathcal{A}) = V$.

Basis

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and a set $\mathcal{A} \subseteq \mathcal{V}$. Then if one of the following condition holds, we say that $\mathcal{A}$ is a basis of V.

- $\mathcal{A}$ is a minimal generating set of V.

  No smaller set $\mathcal{A}' \subsetneq \mathcal{A} \subseteq \mathcal{V}$ that spans V.
- $\mathcal{A}$ spans V and is also linearly independent.

Dimension

Dimension

The number of basis vectors of a vector space V is the *dimension* of V and denoted by $\dim(V)$.

- For $U \subset V$ a subspace of V, $\dim(U) \leq \dim(V)$

Exercise

Given $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix}$.

Find a basis of $\operatorname{span}(\{\mathbf{x}_1, \ldots, \mathbf{x}_4\})$.

Rank

Rank

Rank: the number of linearly independent columns of a matrix $\boldsymbol{A} = \mathbb{R}^{m\times n}$.

Rank

Rank

Rank: the number of linearly independent columns of a matrix $\boldsymbol{A} = \mathbb{R}^{m \times n}$. This equals the number of linearly independent rows of $\boldsymbol{A}$.

Denote by $\text{rank}(\boldsymbol{A})$ the rank of $\boldsymbol{A}$.

Important Properties

- $\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}^\top)$.
- For all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{A}$ is invertible if and only if $\mathrm{rank}(\boldsymbol{A}) = n$.
- $\mathrm{null}(\boldsymbol{A}) = n - \mathrm{rank}(\boldsymbol{A})$, where $\mathrm{null}(\boldsymbol{A})$ is the subspace of $\mathbb{R}^n$ which solutions for $\boldsymbol{A}\mathbf{x} = \mathbf{0}$.
- If $\mathrm{rank}(\boldsymbol{A}) = \min\{m, n\}$ for a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then we say $\boldsymbol{A}$ has full rank.

Outline

Linear Mappings/Linear Transformation

A mapping $\Phi : V \mapsto W$ preserves the structure of the vector space if

- $\Phi(\mathbf{x} + \mathbf{y}) = \Phi(\mathbf{x}) + \Phi(\mathbf{y})$
- $\Phi(\lambda\mathbf{x}) = \lambda\Phi(\mathbf{x})$

for all $\mathbf{x}, \mathbf{y} \in V$ and $\lambda \in \mathbb{R}$.

Linear Mappings/Linear Transformation

A mapping $\Phi : V \mapsto W$ preserves the structure of the vector space if

- $\Phi(\mathbf{x} + \mathbf{y}) = \Phi(\mathbf{x}) + \Phi(\mathbf{y})$
- $\Phi(\lambda \mathbf{x}) = \lambda \Phi(\mathbf{x})$

for all $\mathbf{x}, \mathbf{y} \in V$ and $\lambda \in \mathbb{R}$.

Linear Mapping

For two vector spaces V, W, a mapping $\Phi : V \mapsto W$ is a linear mapping if

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

Transformation Matrix

Transformation Matrix

Given vector spaces V, W with corresponding bases $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $C = (\mathbf{c}_1, \ldots, \mathbf{c}_m)$. Consider a linear mapping $\Phi : V \mapsto W$. For $1 \leq j \leq n$,

$$\Phi(\mathbf{b}_j) = \alpha_{1,j}\mathbf{c}_1 + \cdots \alpha_{m,j}\mathbf{c}_m = \sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i$$

is the unique representation of $\Phi(\mathbf{b}_j)$ w.r.t. C (i.e., coordinate). Then, we call the $m \times n$ matrix $\boldsymbol{A}_\Phi$, whose elements are $A_\Phi(i,j) = \alpha_{ij}$, the transformation matrix of Φ.

- If $\hat{\mathbf{x}}$ is the coordinate of $\mathbf{x} \in V$ w.r.t. B and $\hat{\mathbf{y}} = \Phi(\mathbf{x}) \in W$ w.r.t. C, then

$$\hat{\mathbf{y}} = \boldsymbol{A}_\Phi(\hat{\mathbf{x}}).$$

Example

Consider a linear mapping $\Phi : V \mapsto W$ and ordered bases $B = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ of V and $C = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4)$ of W. Assume that

$$\begin{aligned}
\Phi(\mathbf{b}_1) &= \mathbf{c}_1 - \mathbf{c}_2 + 3\mathbf{c}_3 - \mathbf{c}_4 \\
\Phi(\mathbf{b}_2) &= 2\mathbf{c}_1 + \mathbf{c}_2 + 7\mathbf{c}_3 + 2\mathbf{c}_4 \\
\Phi(\mathbf{b}_3) &= 3\mathbf{c}_2 + \mathbf{c}_3 + 4\mathbf{c}_4.
\end{aligned}$$

The transformation matrix $\boldsymbol{A}_\Phi$ w.r.t. B and C satisfying $\Phi(\mathbf{b}_k) = \sum_{i=1}^{4} \alpha_{ik}\mathbf{c}_i$ for $k = 1, 2, 3$ is

Example

Consider a linear mapping $\Phi : V \mapsto W$ and ordered bases $B = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ of V and $C = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4)$ of W. Assume that

$$\begin{aligned}
\Phi(\mathbf{b}_1) &= \mathbf{c}_1 - \mathbf{c}_2 + 3\mathbf{c}_3 - \mathbf{c}_4 \\
\Phi(\mathbf{b}_2) &= 2\mathbf{c}_1 + \mathbf{c}_2 + 7\mathbf{c}_3 + 2\mathbf{c}_4 \\
\Phi(\mathbf{b}_3) &= 3\mathbf{c}_2 + \mathbf{c}_3 + 4\mathbf{c}_4.
\end{aligned}$$

The transformation matrix $\boldsymbol{A}_\Phi$ w.r.t. B and C satisfying $\Phi(\mathbf{b}_k) = \sum_{i=1}^{4} \alpha_{ik}\mathbf{c}_i$ for $k = 1, 2, 3$ is

$$\boldsymbol{A}_\Phi = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3] = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}.$$

Basis Change (1/5)

- $[I]_B^{B'}$: a transformation matrix that maps coordinates w.r.t. B onto coordinates w.r.t. B'.

Basis Change (1/5)

- $[I]_B^{B'}$: a transformation matrix that maps coordinates w.r.t. B onto coordinates w.r.t. B'.
- For example, let $B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, B' = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$

Basis Change (1/5)

- $[I]_B^{B'}$: a transformation matrix that maps coordinates w.r.t. B onto coordinates w.r.t. B'.
- For example, let $B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, B' = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$
  - $[I]_{B'}^{B} =$

Basis Change (1/5)

- $[I]_B^{B'}$: a transformation matrix that maps coordinates w.r.t. B onto coordinates w.r.t. B'.
- For example, let $B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, B' = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$
  - $[I]_{B'}^{B} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

Basis Change (1/5)

- $[I]_B^{B'}$: a transformation matrix that maps coordinates w.r.t. B onto coordinates w.r.t. B'.
- For example, let $B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, B' = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$
  - $[I]_{B'}^{B} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.
  - What about $[I]_B^{B'}$?

Basis Change (2/5)

Basis Change

Consider a transformation matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

w.r.t. the standard basis (canonical basis) in $\mathbb{R}^2$.

Basis Change (2/5)

Basis Change

Consider a transformation matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

w.r.t. the standard basis (canonical basis) in $\mathbb{R}^2$. Define a new basis $B = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$

Then, what about the transformation matrix $\tilde{\boldsymbol{A}}$ w.r.t. B?

Basis Change (3/5)

Basis Change

Given

- a linear mapping $\Phi : V \mapsto W$, ordered bases

$$B = (\mathbf{b}_1, \ldots, \mathbf{b}_n),\ \tilde{B} = (\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_n) \text{ of } V$$

$$C = (\mathbf{c}_1, \ldots, \mathbf{c}_m),\ \tilde{C} = (\tilde{\mathbf{c}}_1, \ldots, \tilde{\mathbf{c}}_m) \text{ of } W.$$

- a transformation matrix $\boldsymbol{A}_\Phi$ of Φ w.r.t. B and C.

Then, the corresponding transformation matrix $\tilde{\boldsymbol{A}}_\Phi$ w.r.t. $\tilde{B}$ and $\tilde{C}$ is

$$\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{T}^{-1}\boldsymbol{A}_\Phi\boldsymbol{S}.$$

where $\boldsymbol{S} = [I]_{\tilde{B}}^{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{T} = [I]_{\tilde{C}}^{C} \in \mathbb{R}^{m\times m}$.

Proof (1/2)

$\tilde{\boldsymbol{b}}_j = s_{1j}\mathbf{b}_1 + \cdots s_{n,j}\mathbf{b}_n = \sum_{i=1}^{n} s_{ij}\mathbf{b}_i, \quad j = 1, \ldots, n.$

$\tilde{\boldsymbol{c}}_k = t_{1k}\mathbf{c}_1 + \cdots t_{m,k}\mathbf{c}_m = \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell, \quad k = 1, \ldots, m.$

Let $\boldsymbol{S} = ((s_{ij})) = [I]_{\tilde{B}}^{B} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{T} = ((t_{\ell k})) = [I]_{\tilde{C}}^{C} \in \mathbb{R}^{m \times m}$.

Proof (1/2)

$\tilde{\boldsymbol{b}}_j = s_{1j}\mathbf{b}_1 + \cdots s_{n,j}\mathbf{b}_n = \sum_{i=1}^{n} s_{ij}\mathbf{b}_i, \quad j = 1, \ldots, n.$

$\tilde{\boldsymbol{c}}_k = t_{1k}\mathbf{c}_1 + \cdots t_{m,k}\mathbf{c}_m = \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell, \quad k = 1, \ldots, m.$

Let $\boldsymbol{S} = ((s_{ij})) = [I]_{\tilde{B}}^{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{T} = ((t_{\ell k})) = [I]_{\tilde{C}}^{C} \in \mathbb{R}^{m\times m}$.

- Applying the mapping Φ, we get that for all $j = 1, \ldots, n$,

$$\Phi(\tilde{\boldsymbol{b}}_j) = \sum_{k=1}^{m} \underbrace{\tilde{a}_{kj}\tilde{\mathbf{c}}_k}_{\in W}$$

Proof (1/2)

$\tilde{\boldsymbol{b}}_j = s_{1j}\mathbf{b}_1 + \cdots s_{n,j}\mathbf{b}_n = \sum_{i=1}^{n} s_{ij}\mathbf{b}_i, \quad j = 1, \ldots, n.$

$\tilde{\boldsymbol{c}}_k = t_{1k}\mathbf{c}_1 + \cdots t_{m,k}\mathbf{c}_m = \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell, \quad k = 1, \ldots, m.$

Let $\boldsymbol{S} = ((s_{ij})) = [I]_{\tilde{B}}^{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{T} = ((t_{\ell k})) = [I]_{\tilde{C}}^{C} \in \mathbb{R}^{m\times m}$.

- Applying the mapping Φ, we get that for all $j = 1, \ldots, n$,

$$\Phi(\tilde{\boldsymbol{b}}_j) = \sum_{k=1}^{m} \underbrace{\tilde{a}_{kj}\tilde{\mathbf{c}}_k}_{\in W} = \sum_{k=1}^{m} \tilde{a}_{kj} \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell$$

Proof (1/2)

$\tilde{\boldsymbol{b}}_j = s_{1j}\mathbf{b}_1 + \cdots s_{n,j}\mathbf{b}_n = \sum_{i=1}^{n} s_{ij}\mathbf{b}_i, \quad j = 1, \ldots, n.$

$\tilde{\boldsymbol{c}}_k = t_{1k}\mathbf{c}_1 + \cdots t_{m,k}\mathbf{c}_m = \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell, \quad k = 1, \ldots, m.$

Let $\boldsymbol{S} = ((s_{ij})) = [I]_{\tilde{B}}^{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{T} = ((t_{\ell k})) = [I]_{\tilde{C}}^{C} \in \mathbb{R}^{m\times m}$.

- Applying the mapping Φ, we get that for all $j = 1, \ldots, n$,

$$\Phi(\tilde{\boldsymbol{b}}_j) = \sum_{k=1}^{m} \underbrace{\tilde{a}_{kj}\tilde{\mathbf{c}}_k}_{\in W} = \sum_{k=1}^{m} \tilde{a}_{kj} \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell = \sum_{\ell=1}^{m} \left(\sum_{k=1}^{m} t_{\ell k}\tilde{a}_{kj} \right) \mathbf{c}_\ell.$$

Proof (1/2)

$\tilde{\boldsymbol{b}}_j = s_{1j}\mathbf{b}_1 + \cdots s_{n,j}\mathbf{b}_n = \sum_{i=1}^{n} s_{ij}\mathbf{b}_i, \quad j = 1, \ldots, n.$

$\tilde{\boldsymbol{c}}_k = t_{1k}\mathbf{c}_1 + \cdots t_{m,k}\mathbf{c}_m = \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell, \quad k = 1, \ldots, m.$

Let $\boldsymbol{S} = ((s_{ij})) = [I]_{\tilde{B}}^{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{T} = ((t_{\ell k})) = [I]_{\tilde{C}}^{C} \in \mathbb{R}^{m\times m}$.

- Applying the mapping Φ, we get that for all $j = 1, \ldots, n$,

$$\Phi(\tilde{\boldsymbol{b}}_j) = \sum_{k=1}^{m} \underbrace{\tilde{a}_{kj}\tilde{\mathbf{c}}_k}_{\in W} = \sum_{k=1}^{m} \tilde{a}_{kj} \sum_{\ell=1}^{m} t_{\ell k}\mathbf{c}_\ell = \sum_{\ell=1}^{m} \left(\sum_{k=1}^{m} t_{\ell k}\tilde{a}_{kj} \right) \mathbf{c}_\ell .$$

- Alternatively,

$$\begin{aligned} \Phi(\tilde{\boldsymbol{b}}_j) &= \Phi\left(\sum_{i=1}^{n} s_{ij}\mathbf{b}_i \right) = \sum_{i=1}^{n} s_{ij}\Phi(\mathbf{b}_i) = \sum_{i=1}^{n} s_{ij} \sum_{\ell=1}^{m} a_{\ell i}\mathbf{c}_\ell \\ &= \sum_{\ell=1}^{m} \left(\sum_{i=1}^{n} a_{\ell i}s_{ij} \right) \mathbf{c}_\ell \end{aligned}$$

Proof (2/2)

Hence,

$$\sum_{k=1}^{m} t_{\ell k} \tilde{a}_{kj} = \sum_{i=1}^{n} a_{\ell i} s_{ij},$$

and it means that

Proof (2/2)

Hence,

$$\sum_{k=1}^{m} t_{\ell k}\tilde{a}_{kj} = \sum_{i=1}^{n} a_{\ell i}s_{ij},$$

and it means that

$$\boldsymbol{T}\tilde{\boldsymbol{A}}_{\Phi} = \boldsymbol{A}_{\Phi}\boldsymbol{S} \in \mathbb{R}^{m\times n},$$

such that

$$\tilde{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}_{\Phi}\boldsymbol{S}.$$

Basis Change (4/5)

- The theorem tells us that

Basis Change (4/5)

- The theorem tells us that

  With

  - a basis change in V (i.e., $B \to \tilde{B}$) and
  - a basis change in W (i.e., $C \to \tilde{C}$),

  the transformation matrix $\boldsymbol{A}_\Phi$ of a linear mapping $\Phi : V \mapsto W$ is replaced by an equivalent matrix $\tilde{\boldsymbol{A}}_\Phi$ with

$$\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{T}^{-1}\boldsymbol{A}_\Phi\boldsymbol{S}.$$

Image and Kernel

Image & Kernel

For $\Phi : V \mapsto W$, we define

$$\ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{\mathbf{v} \in V \mid \Phi(\mathbf{v}) = \mathbf{0}_W\}$$

and

$$\mathrm{Image}(\Phi) := \Phi(V) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V \mid \Phi(\mathbf{v}) = \mathbf{w}\}.$$

- V: domain of Φ
- W: codomain of Φ

Remark

For vector spaces V and W and a linear mapping $\Phi : V \mapsto W$:

- $\Phi(\mathbf{0}_V)) = \mathbf{0}_W$ so $\mathbf{0} \in \ker(\Phi)$.
- $\text{Image}(\Phi) \subseteq W$ is a subspace of W
- $\ker(\Phi) \subseteq V$ is a subspace of V.
- Φ is injective (i.e., one-to-one) if and only if $\ker(\Phi) = \{\mathbf{0}\}$.
- $\text{Image}(\Phi) = \{\boldsymbol{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} = \{\sum_{i=1}^n x_i \mathbf{a}_i \mid x_1, \ldots, x_n \in \mathbb{R}\} = \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \subseteq \mathbb{R}^m$.
- $\text{rank}(\Phi) = \dim(\text{Image}(\Phi))$.

⋆ $\dim(\ker(\Phi)) + \dim(\text{Image}(\Phi)) = \dim(V)$.
  - $\text{null}(\boldsymbol{A}) + \text{rank}(\boldsymbol{A}) =$ number of columns of A.

- If $\dim(V) = \dim(W)$, then Φ is injective, surjective and bijective.

Outline

Affine Spaces

- Spaces that are offset from the origin.
- They are NO LONGER vector (sub)spaces.

Affine Spaces

- Spaces that are offset from the origin.
- They are NO LONGER vector (sub)spaces.

Affine Subspace

Let V be a vector space, $\mathbf{x}_0 \in V$, and $U \subseteq V$ be a subspace. Then,

$$\begin{aligned} L &= \mathbf{x}_0 + U := \{\mathbf{x}_0 + \mathbf{u} \mid \mathbf{u} \in U\} \\ &= \{\mathbf{v} \in V \mid \exists \mathbf{u} \in U : \mathbf{v} = \mathbf{x}_0 + \mathbf{u}\} \subseteq V \end{aligned}$$

is called affine subspace (or linear manifold) of V.

- U: direction space.
- $\mathbf{x}_0$: support point.

Remark

- An affine subspace excludes $\mathbf{0}$ if $\mathbf{x}_0 \notin U$.
- Examples: points, lines, and planes in $\mathbb{R}^3$ which do not go through the origin.

Remark

- An affine subspace excludes $\mathbf{0}$ if $\mathbf{x}_0 \notin U$.
- Examples: points, lines, and planes in $\mathbb{R}^3$ which do not go through the origin.
- One-dimensional affine subspaces:

$$\mathbf{y} = \mathbf{x}_0 + \lambda \mathbf{b}_1$$

  for $\lambda \in \mathbb{R}$ and $U = \text{span}(\mathbf{b}_1)$ is a one-dimensional subspace of $\mathbb{R}^n$.
- Two-dimensional affine subspaces:

$$\mathbf{y} = \mathbf{x}_0 + \lambda_1 \mathbf{b}_1 + \lambda_2 \mathbf{b}_2$$

  for $\lambda_1, \lambda_2 \in \mathbb{R}$ and $U = \text{span}(\{\mathbf{b}_1, \mathbf{b}_2\})$ is a two-dimensional subspace of $\mathbb{R}^n$.
- $\vdots$

Affine Mappings

Affine Mappings

Given two vector spaces V, W, a linear mapping $\Phi : V \mapsto W$, and $\mathbf{a} \in W$, the mapping $\phi : V \mapsto W$ with

$$\phi(\mathbf{x}) = \mathbf{a} + \Phi(\mathbf{x})$$

is called an affine mapping from V to W. The vector $\mathbf{a}$ is called the translation vector of ϕ.

Discussions