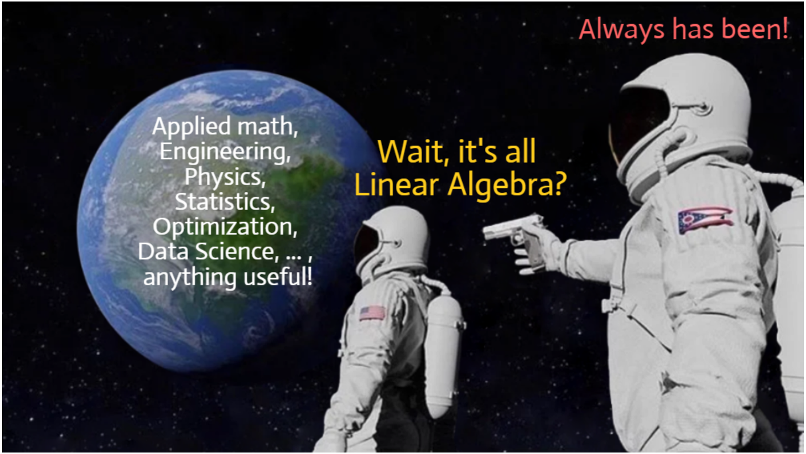

Introduction to Optimization

K. R. Sahasranand

Data Science

sahasranand@iitpkd.ac.in

Vector

An n-dimensional column vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$x_i \sim i$-th component of x

Real vector: $x \in \mathbb{R}^n$

Vector

An n-dimensional column vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$x_i \sim i$-th component of x

Real vector: $x \in \mathbb{R}^n$

Row vector $\widetilde{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]$

Vector

An n-dimensional column vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$x_i \sim i$-th component of x

Real vector: $x \in \mathbb{R}^n$

Row vector $\widetilde{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]$

Transpose:

$$\widetilde{x} = x^T$$

Vector

An n-dimensional column vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$x_i \sim i$-th component of x

Real vector: $x \in \mathbb{R}^n$

Row vector $\widetilde{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]$

Transpose:

$$\widetilde{x} = x^T$$

Equality: $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ are *equal* iff $x_i = y_i$ for all $i \in [n]$.

Vector addition and Scalar multiplication

– The sum of vectors $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ and $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$ denoted $x + y$ is the vector:

$$\begin{bmatrix} x_1 + y_1 & x_2 + y_2 & \cdots & x_n + y_n \end{bmatrix}$$

Vector addition and Scalar multiplication

– The sum of vectors $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ and $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$

denoted $x + y$ is the vector:

$$\begin{bmatrix} x_1 + y_1 & x_2 + y_2 & \cdots & x_n + y_n \end{bmatrix}$$

– The multiplication of a vector $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ by a scalar α

denoted $\alpha \cdot x$ is the vector:

$$\begin{bmatrix} \alpha x_1 & \alpha x_2 & \cdots & \alpha x_n \end{bmatrix}$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)

$$x + (y + z) = (x + y) + z$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)

   $$x + (y + z) = (x + y) + z$$

2. (Commutativity)

   $$x + y = y + x$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)

   $$x + (y + z) = (x + y) + z$$

2. (Commutativity)

   $$x + y = y + x$$

3. (Zero element) $\exists\ \mathbf{0} \in V$:

   $$x + 0 = x$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)

   $$x + (y + z) = (x + y) + z$$

2. (Commutativity)

   $$x + y = y + x$$

3. (Zero element) $\exists\ \mathbf{0} \in V$:

   $$x + 0 = x$$

4. (Additive inverse) $\exists\ -x \in V$:

   $$x + (-x) = \mathbf{0}$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)

$$x + (y + z) = (x + y) + z$$

2. (Commutativity)

$$x + y = y + x$$

3. (Zero element) $\exists\ \mathbf{0} \in V$:

$$x + 0 = x$$

4. (Additive inverse) $\exists\ -x \in V$:

$$x + (-x) = \mathbf{0}$$

1. (Associativity)

$$\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)
$$x + (y + z) = (x + y) + z$$
2. (Commutativity)
$$x + y = y + x$$
3. (Zero element) $\exists\ \mathbf{0} \in V$:
$$x + 0 = x$$
4. (Additive inverse) $\exists\ -x \in V$:
$$x + (-x) = \mathbf{0}$$

1. (Associativity)
$$\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x$$
2. (Multiplicative identity) $\exists\ \mathbf{1} \in F$:
$$\mathbf{1} \cdot x = x$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)

$$x + (y + z) = (x + y) + z$$

2. (Commutativity)

$$x + y = y + x$$

3. (Zero element) $\exists\ \mathbf{0} \in V$:

$$x + 0 = x$$

4. (Additive inverse) $\exists\ -x \in V$:

$$x + (-x) = \mathbf{0}$$

1. (Associativity)

$$\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x$$

2. (Multiplicative identity) $\exists\ \mathbf{1} \in F$:

$$\mathbf{1} \cdot x = x$$

3. (Distributivity I)

$$\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$$

Vector space over a field F

A non-empty set V satisfying (for $x, y, z \in V$ and $\alpha, \beta \in F$):

1. (Associativity)

$$x + (y + z) = (x + y) + z$$

2. (Commutativity)

$$x + y = y + x$$

3. (Zero element) $\exists\ \mathbf{0} \in V$:

$$x + 0 = x$$

4. (Additive inverse) $\exists\ -x \in V$:

$$x + (-x) = \mathbf{0}$$

1. (Associativity)

$$\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x$$

2. (Multiplicative identity) $\exists\ \mathbf{1} \in F$:

$$\mathbf{1} \cdot x = x$$

3. (Distributivity I)

$$\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$$

4. (Distributivity II)

$$(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$$

Linear combination and linear independence

– A linear combination of the vectors $u_1, u_2, \ldots, u_k \in V$ is a vector in V of the form

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_k u_k$$

where $\alpha_i \in \mathbb{R},\ i \in [k]$.

Linear combination and linear independence

– A linear combination of the vectors $u_1, u_2, \ldots, u_k \in V$ is a vector in V of the form

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_k u_k$$

where $\alpha_i \in \mathbb{R},\ i \in [k]$.

– A set of vectors $\{u_1, u_2, \ldots, u_k\}$ is said to be linearly independent if the *only* way to obtain a linear combination $\mathbf{0}$ is via

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

Linear combination and linear independence

– A linear combination of the vectors $u_1, u_2, \ldots, u_k \in V$ is a vector in V of the form

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_k u_k$$

where $\alpha_i \in \mathbb{R},\ i \in [k]$.

– A set of vectors $\{u_1, u_2, \ldots, u_k\}$ is said to be linearly independent if the *only* way to obtain a linear combination $\mathbf{0}$ is via

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

– Linearly dependent if not independent

Linear combination and linear independence

– A linear combination of the vectors $u_1, u_2, \ldots, u_k \in V$ is a vector in V of the form

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_k u_k$$

where $\alpha_i \in \mathbb{R},\ i \in [k]$.

– A set of vectors $\{u_1, u_2, \ldots, u_k\}$ is said to be linearly independent if the *only* way to obtain a linear combination $\mathbf{0}$ is via

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

– Linearly dependent if not independent

Theorem – A set of vectors $\{u_1, u_2, \ldots, u_k\}$ is linearly dependent *iff* one of the vectors from the set is a linear combination of the remaining vectors.

Linear combination and linear independence

– A linear combination of the vectors $u_1, u_2, \ldots, u_k \in V$ is a vector in V of the form

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_k u_k$$

where $\alpha_i \in \mathbb{R},\ i \in [k]$.

– A set of vectors $\{u_1, u_2, \ldots, u_k\}$ is said to be linearly independent if the *only* way to obtain a linear combination $\mathbf{0}$ is via

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

– Linearly dependent if not independent

Theorem – A set of vectors $\{u_1, u_2, \ldots, u_k\}$ is linearly dependent *iff* one of the vectors from the set is a linear combination of the remaining vectors.

(Proof left as exercise)

Subspace

A subspace is a non-empty subset of a vector space V that is

closed under addition and scalar multiplication

– Every subspace contains the zero vector $\mathbf{0}$. Proof?

Span

$u_1, u_2, \ldots, u_k \in \mathbb{R}^n$:

$$\mathsf{span}(u_1, \ldots, u_k) = \left\{ \sum_{i=1}^{k} \alpha_i u_i : \ \alpha_i \in \mathbb{R}, i \in [k] \right\}$$

Span

$u_1, u_2, \ldots, u_k \in \mathbb{R}^n$:

$$\mathsf{span}(u_1, \ldots, u_k) = \left\{ \sum_{i=1}^{k} \alpha_i u_i : \ \alpha_i \in \mathbb{R}, i \in [k] \right\}$$

– If u is a linear combination of $u_1, u_2, \ldots, u_k$,

$$\mathsf{span}(u_1, \ldots, u_k, u) = \mathsf{span}(u_1, \ldots, u_k)$$

Span

$u_1, u_2, \ldots, u_k \in \mathbb{R}^n$:

$$\mathsf{span}(u_1, \ldots, u_k) = \left\{ \sum_{i=1}^{k} \alpha_i u_i : \ \alpha_i \in \mathbb{R}, i \in [k] \right\}$$

– If u is a linear combination of $u_1, u_2, \ldots, u_k$,

$$\mathsf{span}(u_1, \ldots, u_k, u) = \mathsf{span}(u_1, \ldots, u_k)$$

– The span of any set of vectors is a subspace. Proof?

Basis of a vector space V

Any linearly independent set $\{w_1, w_2, \ldots, w_k\}$ such that

$$V = \mathsf{span}(w_1, w_2, \ldots, w_k).$$

Basis of a vector space V

Any linearly independent set $\{w_1, w_2, \ldots, w_k\}$ such that

$$V = \mathsf{span}(w_1, w_2, \ldots, w_k).$$

– All bases contain the same number of vectors: dimension of V

Basis of a vector space V

Any linearly independent set $\{w_1, w_2, \ldots, w_k\}$ such that

$$V = \mathsf{span}(w_1, w_2, \ldots, w_k).$$

– All bases contain the same number of vectors: dimension of V

Proof?

Basis of a vector space V

Any linearly independent set $\{w_1, w_2, \ldots, w_k\}$ such that

$$V = \mathsf{span}(w_1, w_2, \ldots, w_k).$$

– All bases contain the same number of vectors: dimension of V

Proof?

Theorem – If $\{u_1, u_2, \ldots, u_k\}$ is a basis of V, then any element $a \in V$ can be represented *uniquely* as a linear combination

$$a = \alpha_1 u_1 + \ldots + \alpha_k u_k$$

with $\alpha_i \in \mathbb{R},\ i \in [k]$.

Basis of a vector space V

Any linearly independent set $\{w_1, w_2, \ldots, w_k\}$ such that

$$V = \mathsf{span}(w_1, w_2, \ldots, w_k).$$

– All bases contain the same number of vectors: dimension of V

Proof?

Theorem – If $\{u_1, u_2, \ldots, u_k\}$ is a basis of V, then any element $a \in V$ can be represented *uniquely* as a linear combination

$$a = \alpha_1 u_1 + \ldots + \alpha_k u_k$$

with $\alpha_i \in \mathbb{R},\ i \in [k]$.

(Proof left as exercise)

Basis for $\mathbb{R}^n$

A basis for $\mathbb{R}^n$ is given by

$$\{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}\}$$

where $\mathbf{e_i}$ has entry 1 in the i-th position and 0 everywhere else.

Basis for $\mathbb{R}^n$

A basis for $\mathbb{R}^n$ is given by

$$\{\mathbf{e_1}, \mathbf{e_2}, \dots, \mathbf{e_n}\}$$

where $\mathbf{e_i}$ has entry 1 in the i-th position and 0 everywhere else.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ . \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ x_n \end{bmatrix}$$

Basis for $\mathbb{R}^n$

A basis for $\mathbb{R}^n$ is given by

$$\{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}\}$$

where $\mathbf{e_i}$ has entry 1 in the i-th position and 0 everywhere else.

$$\begin{aligned}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ . \\ x_n \end{bmatrix} &= \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ x_n \end{bmatrix} \\
&= x_1 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}
\end{aligned}$$

Basis for $\mathbb{R}^n$

A basis for $\mathbb{R}^n$ is given by

$$\{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}\}$$

where $\mathbf{e_i}$ has entry 1 in the i-th position and 0 everywhere else.

$$\begin{aligned}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ . \\ x_n \end{bmatrix} &= \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ x_n \end{bmatrix} \\
&= x_1 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \\
&= x_1\mathbf{e_1} + x_2\mathbf{e_2} + \cdots x_n\mathbf{e_n}
\end{aligned}$$

Inner product

Inner product or dot product of $x, y \in \mathbb{R}^n$ is given by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x^T y$$

Inner product

Inner product or dot product of $x, y \in \mathbb{R}^n$ is given by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x^T y$$

Properties –

- Positivity

  $\langle x, x \rangle \geq 0$ with equality iff $x = \mathbf{0}$.

Inner product

Inner product or dot product of $x, y \in \mathbb{R}^n$ is given by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x^T y$$

Properties –

- Positivity

  $\langle x, x \rangle \geq 0$ with equality iff $x = \mathbf{0}$.

- Symmetry

  $$\langle x, y \rangle = \langle y, x \rangle$$

Inner product

Inner product or dot product of $x, y \in \mathbb{R}^n$ is given by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x^T y$$

Properties –

- Positivity

$$\langle x, x \rangle \geq 0 \text{ with equality iff } x = \mathbf{0}.$$

- Symmetry

$$\langle x, y \rangle = \langle y, x \rangle$$

- Linearity: for $\alpha, \beta \in \mathbb{R}$ and $x, y, z \in \mathbb{R}^n$

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$$

Dot product

Is it $\sum_i x_i y_i$ or is it $\|x\|\|y\|\cos\theta$?

Dot product

Is it $\sum_i x_i y_i$ or is it $\|x\|\|y\|\cos\theta$?

We know that

$$x = \sum_{i=1}^{n} x_i \cdot \mathbf{e_i} \qquad ; \qquad y = \sum_{i=1}^{n} y_i \cdot \mathbf{e_i}$$

Dot product

Is it $\sum_i x_i y_i$ or is it $\|x\|\|y\|\cos\theta$?

We know that

$$x = \sum_{i=1}^{n} x_i \cdot \mathbf{e_i} \qquad ; \qquad y = \sum_{i=1}^{n} y_i \cdot \mathbf{e_i}$$

and

$$\langle \mathbf{e_i}, \mathbf{e_j} \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Dot product

Is it $\sum_i x_i y_i$ or is it $\|x\|\|y\|\cos\theta$?

We know that

$$x = \sum_{i=1}^{n} x_i \cdot \mathbf{e_i} \qquad ; \qquad y = \sum_{i=1}^{n} y_i \cdot \mathbf{e_i}$$

and

$$\langle \mathbf{e_i}, \mathbf{e_j} \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Then, by the "geometric definition",

$$\langle x, \mathbf{e_i} \rangle = \|x\|\|\mathbf{e_i}\| \cos\theta = x_i$$

Dot product

Is it $\sum_i x_i y_i$ or is it $\|x\|\|y\|\cos\theta$?

We know that

$$x = \sum_{i=1}^{n} x_i \cdot \mathbf{e_i} \qquad ; \qquad y = \sum_{i=1}^{n} y_i \cdot \mathbf{e_i}$$

and

$$\langle \mathbf{e_i}, \mathbf{e_j} \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Then, by the "geometric definition",

$$\langle x, \mathbf{e_i} \rangle = \|x\|\|\mathbf{e_i}\| \cos\theta = x_i$$

Hence,

$$\langle x, y \rangle = \langle x, \sum_{i=1}^{n} y_i \cdot \mathbf{e_i} \rangle = \sum_{i=1}^{n} y_i \langle x, \mathbf{e_i} \rangle = \sum_{i=1}^{n} y_i x_i.$$

Orthogonality and norms

– Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal (to each other) if

$$\langle x, y \rangle = 0.$$

Orthogonality and norms

- Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal (to each other) if

$$\langle x, y \rangle = 0.$$

- The Euclidean norm of a vector x is given by

$$\|x\| = \sqrt{\langle x, x \rangle} = \sum_{i=1}^{n} x_i^2$$

Orthogonality and norms

– Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal (to each other) if

$$\langle x, y \rangle = 0.$$

– The Euclidean norm of a vector x is given by

$$\|x\| = \sqrt{\langle x, x \rangle} = \sum_{i=1}^{n} x_i^2$$

– (Cauchy-Schwarz inequality)

$$\langle x, y \rangle \leq \|x\| \cdot \|y\|$$

with equality iff $x = \alpha \cdot y$ for some $\alpha \in \mathbb{R}$.

Orthogonality and norms

– Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal (to each other) if

$$\langle x, y \rangle = 0.$$

– The Euclidean norm of a vector x is given by

$$\|x\| = \sqrt{\langle x, x \rangle} = \sum_{i=1}^{n} x_i^2$$

– (Cauchy-Schwarz inequality)

$$\langle x, y \rangle \leq \|x\| \cdot \|y\|$$

with equality iff $x = \alpha \cdot y$ for some $\alpha \in \mathbb{R}$. Proof?

Properties of norms

- Positivity

$$\|x\| \geq 0 \text{ with equality iff } x = \mathbf{0}.$$

Properties of norms

- Positivity

$$\|x\| \geq 0 \text{ with equality iff } x = \mathbf{0}.$$

- Homogeneity: for $\alpha \in \mathbb{R}$

$$\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$$

Properties of norms

- Positivity

$$\|x\| \geq 0 \text{ with equality iff } x = \mathbf{0}.$$

- Homogeneity: for $\alpha \in \mathbb{R}$

$$\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$$

- Triangle inequality

$$\|x + y\| \leq \|x\| + \|y\|$$

Properties of norms

- Positivity

$$\|x\| \geq 0 \text{ with equality iff } x = \mathbf{0}.$$

- Homogeneity: for $\alpha \in \mathbb{R}$

$$\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$$

- Triangle inequality

$$\|x + y\| \leq \|x\| + \|y\|$$

Proof?

(use Cauchy-Schwarz inequality)

Linear function between vector spaces

We say $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if

$$\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y) \qquad \text{for all } x, y \in \mathbb{R}^n$$

$$\mathcal{L}(\alpha x) = \alpha \mathcal{L}(x) \qquad \text{for all } x \in \mathbb{R}^n, \alpha \in \mathbb{R}$$

Linear function between vector spaces

We say $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if

$$\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y) \qquad \text{for all } x, y \in \mathbb{R}^n$$
$$\mathcal{L}(\alpha x) = \alpha \mathcal{L}(x) \qquad \text{for all } x \in \mathbb{R}^n, \alpha \in \mathbb{R}$$

Is a line a linear function?

Linear function between vector spaces

We say $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if

$$\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y) \qquad \text{for all } x, y \in \mathbb{R}^n$$
$$\mathcal{L}(\alpha x) = \alpha \mathcal{L}(x) \qquad \text{for all } x \in \mathbb{R}^n, \alpha \in \mathbb{R}$$

Is a line a linear function?

(explain on board)

Linear function between vector spaces

We say $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if

$$\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y) \qquad \text{for all } x, y \in \mathbb{R}^n$$
$$\mathcal{L}(\alpha x) = \alpha \mathcal{L}(x) \qquad \text{for all } x \in \mathbb{R}^n, \alpha \in \mathbb{R}$$

Is a line a linear function?

(explain on board)

– To specify what $\mathcal{L}$ does to a certain $x \in \mathbb{R}^n$, it suffices to specify what it does to each of $\mathbf{e_1}, \ldots, \mathbf{e_n}$ since

Linear function between vector spaces

We say $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if

$$\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y) \qquad \text{for all } x, y \in \mathbb{R}^n$$
$$\mathcal{L}(\alpha x) = \alpha\mathcal{L}(x) \qquad \text{for all } x \in \mathbb{R}^n, \alpha \in \mathbb{R}$$

Is a line a linear function?

(explain on board)

– To specify what $\mathcal{L}$ does to a certain $x \in \mathbb{R}^n$, it suffices to specify what it does to each of $\mathbf{e_1}, \ldots, \mathbf{e_n}$ since

$$\begin{aligned}\mathcal{L}(x) &= \mathcal{L}\left(x_1\mathbf{e_1} + \cdots + x_n\mathbf{e_n}\right) \\ &= x_1\mathcal{L}(\mathbf{e_1}) + \cdots + x_n\mathcal{L}(\mathbf{e_n}) \\ &= x_1c_1 + \cdots + x_nc_n\end{aligned}$$

Enter matrix

$$A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}$$

Four fundamental subspaces

$$A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \begin{bmatrix} \cdots & \mathbf{r}_1 & \cdots \\ \cdots & \mathbf{r}_2 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{r}_m & \cdots \end{bmatrix}$$

Four fundamental subspaces

$$A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \begin{bmatrix} \cdots & \mathbf{r}_1 & \cdots \\ \cdots & \mathbf{r}_2 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{r}_m & \cdots \end{bmatrix}$$

- Column space

$$\mathsf{Col}(A) = \mathsf{span}\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$$

Four fundamental subspaces

$$A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \begin{bmatrix} \cdots & \mathbf{r}_1 & \cdots \\ \cdots & \mathbf{r}_2 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{r}_m & \cdots \end{bmatrix}$$

- Column space

$$\mathsf{Col}(A) = \mathsf{span}\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$$

- Row space

$$\mathsf{Row}(A) = \mathsf{span}\{\mathbf{r}_1, \dots, \mathbf{r}_m\}$$

Four fundamental subspaces

$$A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \begin{bmatrix} \cdots & \mathbf{r}_1 & \cdots \\ \cdots & \mathbf{r}_2 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{r}_m & \cdots \end{bmatrix}$$

- Row space

$$\mathsf{Row}(A) = \mathsf{span}\{\mathbf{r}_1, \ldots, \mathbf{r}_m\}$$

- Column space

$$\mathsf{Col}(A) = \mathsf{span}\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$$

- Null space

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$$

Four fundamental subspaces

$$A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \begin{bmatrix} \cdots & \mathbf{r}_1 & \cdots \\ \cdots & \mathbf{r}_2 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{r}_m & \cdots \end{bmatrix}$$

- Column space

$$\mathsf{Col}(A) = \mathsf{span}\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$$

- Row space

$$\mathsf{Row}(A) = \mathsf{span}\{\mathbf{r}_1, \ldots, \mathbf{r}_m\}$$

- Left null space

$$\mathcal{N}(A^\top) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y}^\top A = \mathbf{0}\}$$

- Null space

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$$

Orthogonal complement

The orthogonal complement of a subspace $V \subset \mathbb{R}^n$ is given by

$$V^\perp = \left\{x : x^T v = 0 \text{ for all } v \in V\right\}.$$

Orthogonal complement

The orthogonal complement of a subspace $V \subset \mathbb{R}^n$ is given by

$$V^\perp = \left\{x : x^T v = 0 \text{ for all } v \in V\right\}.$$

- $V^\perp$ is a subspace

Orthogonal complement

The orthogonal complement of a subspace $V \subset \mathbb{R}^n$ is given by

$$V^\perp = \left\{ x : x^T v = 0 \text{ for all } v \in V \right\}.$$

- $V^\perp$ is a subspace — Proof?

Orthogonal complement

The orthogonal complement of a subspace $V \subset \mathbb{R}^n$ is given by

$$V^\perp = \left\{x : x^T v = 0 \text{ for all } v \in V\right\}.$$

- $V^\perp$ is a subspace — Proof?
- Every $x \in \mathbb{R}^n$ can be represented uniquely as

$$x = x_1 + x_2 \qquad\qquad x_1 \in V,\ x_2 \in V^\perp.$$

Orthogonal complement

The orthogonal complement of a subspace $V \subset \mathbb{R}^n$ is given by

$$V^\perp = \{x : x^T v = 0 \text{ for all } v \in V\}.$$

- $V^\perp$ is a subspace Proof?
- Every $x \in \mathbb{R}^n$ can be represented uniquely as

$$x = x_1 + x_2 \qquad x_1 \in V,\ x_2 \in V^\perp.$$

Proof?

Orthogonal complement

The orthogonal complement of a subspace $V \subset \mathbb{R}^n$ is given by

$$V^\perp = \left\{ x : x^T v = 0 \text{ for all } v \in V \right\}.$$

- $V^\perp$ is a subspace — Proof?
- Every $x \in \mathbb{R}^n$ can be represented uniquely as

$$x = x_1 + x_2 \qquad\qquad x_1 \in V,\, x_2 \in V^\perp.$$

Proof?

- For a matrix A, $\text{Row}(A)^\perp = \mathcal{N}(A)$

Orthogonal complement

The orthogonal complement of a subspace $V \subset \mathbb{R}^n$ is given by

$$V^\perp = \left\{x : x^T v = 0 \text{ for all } v \in V\right\}.$$

- $V^\perp$ is a subspace — Proof?
- Every $x \in \mathbb{R}^n$ can be represented uniquely as

$$x = x_1 + x_2 \qquad\qquad x_1 \in V,\, x_2 \in V^\perp.$$

Proof?

- For a matrix A, $\text{Row}(A)^\perp = \mathcal{N}(A)$ — Proof?

Rank of a matrix

$\text{rank}(A) =$ maximal number of linearly independent columns of A

Rank of a matrix

$$\begin{aligned}\text{rank}(A) &= \text{maximal number of linearly independent columns of } A \\ &= \dim(\text{Col}(A))\end{aligned}$$

Rank of a matrix

$$\begin{aligned}\text{rank}(A) &= \text{maximal number of linearly independent columns of } A \\ &= \dim(\text{Col}(A))\end{aligned}$$

Row rank = Col rank

Rank of a matrix

$$\begin{aligned} \mathsf{rank}(A) &= \text{maximal number of linearly independent columns of } A \\ &= \dim(\mathsf{Col}(A)) \end{aligned}$$

Row rank = Col rank   Proof?

Rank of a matrix

$$\begin{aligned}\text{rank}(A) &= \text{maximal number of linearly independent columns of } A \\ &= \dim(\text{Col}(A))\end{aligned}$$

$$\text{Row rank} = \text{Col rank}$$ Proof?

- If $A \in \mathbb{R}^{m \times n}$, then $\text{rank}(A) \le \min\{m, n\}$.

Rank of a matrix

$$\text{rank}(A) = \text{maximal number of linearly independent columns of } A$$
$$= \dim(\text{Col}(A))$$

Row rank = Col rank Proof?

- If $A \in \mathbb{R}^{m\times n}$, then $\text{rank}(A) \leq \min\{m, n\}$. Proof?

Linear equations

Suppose that we are given m equations in n unknowns of the form

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

Linear equations

Suppose that we are given m equations in n unknowns of the form

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

In vector form,

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

Linear equations

Suppose that we are given m equations in n unknowns of the form

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

In vector form,

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

In matrix form,

$$Ax = b.$$

Linear equations

Suppose that we are given m equations in n unknowns of the form

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

In vector form,

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

In matrix form,

$$Ax = b.$$

- The system $Ax = b$ has a solution iff rank $A =$ rank $[A,\ b]$.

Linear equations

Suppose that we are given m equations in n unknowns of the form

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

In vector form,

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

In matrix form,

$$Ax = b.$$

- The system $Ax = b$ has a solution iff rank A = rank $[A,\ b]$. Proof?

Invertible matrix

A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if

- $\text{rank}(A) = n$
- columns of A form a basis for $\mathbb{R}^n$
- there is a matrix $B \in \mathbb{R}^{n \times n}$ such that $AB = BA = I_n$
- $\cdots$

Invertible matrix

A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if

- $\text{rank}(A) = n$
- columns of A form a basis for $\mathbb{R}^n$
- there is a matrix $B \in \mathbb{R}^{n \times n}$ such that $AB = BA = I_n$
- $\cdots$
- 0 is not an eigenvalue of A

Eigenvalues and Eigenvectors

$$Ax = \lambda x$$

- How do you find these λ's and x's?

Eigenvalues and Eigenvectors

$$Ax = \lambda x$$

- How do you find these λ's and x's?
- If $A \in \mathbb{R}^{n\times n}$, then A has n eigenvalues

Eigenvalues and Eigenvectors

$$Ax = \lambda x$$

- How do you find these λ's and x's?
- If $A \in \mathbb{R}^{n \times n}$, then A has n eigenvalues — Why?

Example: 2×2 matrix

Let

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

Example: 2×2 matrix

Let

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

Find the eigenvalues:

$$\det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{vmatrix} = 0$$

Example: 2×2 matrix

Let

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

Find the eigenvalues:

$$\det(A - \lambda I) = \begin{vmatrix} 4 - \lambda & 1 \\ 2 & 3 - \lambda \end{vmatrix} = 0$$

Compute determinant:

$$(4 - \lambda)(3 - \lambda) - 2 = \lambda^2 - 7\lambda + 10 = 0$$

Example: 2×2 matrix

Let

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

Find the eigenvalues:

$$\det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{vmatrix} = 0$$

Compute determinant:

$$(4-\lambda)(3-\lambda) - 2 = \lambda^2 - 7\lambda + 10 = 0$$

Solve:

$$\lambda = 2, \quad \lambda = 5$$

Example: 2×2 matrix

Let

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

Find the eigenvalues:

$$\det(A - \lambda I) = \begin{vmatrix} 4 - \lambda & 1 \\ 2 & 3 - \lambda \end{vmatrix} = 0$$

Compute determinant:

$$(4 - \lambda)(3 - \lambda) - 2 = \lambda^2 - 7\lambda + 10 = 0$$

Solve:

$$\lambda = 2, \quad \lambda = 5$$

Eigenvectors?

Symmetric matrices

A is symmetric ($A = A^T$)

- All its eigenvalues are real
- Its n eigenvectors are orthogonal

Symmetric matrices

A is symmetric ($A = A^T$)

- All its eigenvalues are real
- Its n eigenvectors are orthogonal

proof for the case with distinct eigenvalues on board

Diagonalizability

Matrix A is said to be diagonalizable if we can write

$$A = VDV^{-1}$$

where D is a diagonal matrix.

Diagonalizability

Matrix A is said to be diagonalizable if we can write

$$A = VDV^{-1}$$

where D is a diagonal matrix.

$$AV = VD \iff Av_i = \lambda_i v_i \text{ for } i \in [n] \text{ and } \{v_i\}_{i \in [n]} \text{ form a basis}$$

Diagonalizability

Matrix A is said to be diagonalizable if we can write

$$A = VDV^{-1}$$

where D is a diagonal matrix.

$$AV = VD \iff Av_i = \lambda_i v_i \text{ for } i \in [n] \text{ and } \{v_i\}_{i \in [n]} \text{ form a basis}$$

- A symmetric matrix is diagonalizable.

Diagonalizability

Matrix A is said to be diagonalizable if we can write

$$A = VDV^{-1}$$

where D is a diagonal matrix.

$$AV = VD \iff Av_i = \lambda_i v_i \text{ for } i \in [n] \text{ and } \{v_i\}_{i\in[n]} \text{ form a basis}$$

- A symmetric matrix is diagonalizable. Proof?

Diagonalizability

Matrix A is said to be diagonalizable if we can write

$$A = VDV^{-1}$$

where D is a diagonal matrix.

$$AV = VD \iff Av_i = \lambda_i v_i \text{ for } i \in [n] \text{ and } \{v_i\}_{i\in[n]} \text{ form a basis}$$

◦ A symmetric matrix is diagonalizable. Proof?

◦ (Rayleigh's inequalities) For a symmetric matrix Q,

$$\lambda_{\min} \leq x^T Q x \leq \lambda_{\max}$$

for any x with $\|x\| = 1$.

$\lambda_{\min}$: smallest eigenvalue of Q

$\lambda_{\max}$: largest eigenvalue of Q

Diagonalizability

Matrix A is said to be diagonalizable if we can write

$$A = VDV^{-1}$$

where D is a diagonal matrix.

$$AV = VD \iff Av_i = \lambda_i v_i \text{ for } i \in [n] \text{ and } \{v_i\}_{i\in[n]} \text{ form a basis}$$

- A symmetric matrix is diagonalizable. Proof?

- (Rayleigh's inequalities) For a symmetric matrix Q,

$$\lambda_{\min} \leq x^T Q x \leq \lambda_{\max}$$

for any x with $\|x\| = 1$.

$\lambda_{\min}$: smallest eigenvalue of Q

$\lambda_{\max}$: largest eigenvalue of Q

(Proof left as exercise)

Quadratic form

A symmetric matrix Q is said to be positive definite if the **quadratic form**

$$x^T Q x > 0 \quad \text{for all } x \neq \mathbf{0}.$$

Quadratic form

A symmetric matrix Q is said to be positive definite if the **quadratic form**

$$x^T Q x > 0 \quad \text{for all } x \neq \mathbf{0}.$$

If Q is symmetric, then

Q is positive definite $\iff$ all eigenvalues of Q are positive

Proof?

Quadratic form

A symmetric matrix Q is said to be positive definite if the **quadratic form**

$$x^T Qx > 0 \quad \text{for all } x \neq \mathbf{0}.$$

If Q is symmetric, then

Q is positive definite $\Longleftrightarrow$ all eigenvalues of Q are positive

Proof?

Positive semidefinite : $\geq$ instead of $>$