Unidad I: Conceptos Preliminares

B. Algebra lineal y funciones

- subespacios y subespacios afines, vectores independientes
- matrices, espacio columna, rango, núcleo, inversa
- funciones, gradiente, matriz hessiana
- teorema de Taylor

Subespacios

$\mathcal{S} \subseteq \mathbb{R}^n$ ($\mathcal{S} \neq \emptyset$) se llama un *subespacio* si

$$\mathbf{x}, \mathbf{y} \in \mathcal{S}, \quad \alpha, \beta \in \mathbb{R} \quad \Rightarrow \quad \alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{S}$$

$\alpha\mathbf{x} + \beta\mathbf{y}$ es una combinación lineal de $\mathbf{x}$ e $\mathbf{y}$

ejemplos (en $\mathbb{R}^n$)

- $\mathcal{S} = \mathbb{R}^n, \quad \mathcal{S} = \{0\}$
- $\mathcal{S} = \{\alpha\mathbf{v} \mid \alpha \in \mathbb{R}\}$ donde $\mathbf{v} \in \mathbb{R}^n$ (i.e, una línea que pasa por el origen)
- $\mathcal{S} = \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k) = \{\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k \mid \alpha_i \in \mathbb{R}\}$, donde $\mathbf{v}_i \in \mathbb{R}$
- conjunto de vectores ortogonales a vectores dados $\mathbf{v}_1, \cdots, \mathbf{v}_k$:

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{v}_1^T\mathbf{x} = 0, \cdots, \mathbf{v}_k^T\mathbf{x} = 0\}$$

Vectores independientes

los vectores $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ son *independientes* si y solo si

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_k \mathbf{v}_k = \mathbf{0} \quad \Rightarrow \quad \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

algunas condiciones equivalentes:

- los coeficientes de $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_k \mathbf{v}_k$ son únicamente determinados, i.e.,

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_k \mathbf{v}_k = \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 + \cdots + \beta_k \mathbf{v}_k$$

  implica $\alpha_1 = \beta_1, \alpha_2 = \beta_2, \cdots, \alpha_k = \beta_k$

- ningun vector $\mathbf{v}_i$ puede expresarse como combinación lineal de los demás vectores $\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \cdots, \mathbf{v}_k$

Base y Dimensión

$\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\}$ es una *base* del subespacio $\mathcal{S} \subseteq \mathbb{R}^n$ si

- $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$ generan $\mathcal{S}$, i.e., $\mathcal{S} = \operatorname{span}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k)$
- $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$ son linealmente independientes

equivalentemente: todo $\mathbf{x} \in \mathcal{S}$ puede expresarse en forma única como

$$\mathbf{x} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_k \mathbf{x}_k$$

Para un dado subespacio $\mathcal{S}$, el número de vectores en cualquier base no varia, y se llama *dimensión* de $\mathcal{S}$, denotado $\dim \mathcal{S}$

Conjuntos Afines

Tambien conocidos como *subespacios afines* o *variedades lineales*

$\mathcal{V} \subseteq \mathbb{R}^n$ $(\mathcal{V} \neq \varnothing)$ es un *conjunto afin* si

$$\mathbf{x}, \mathbf{y} \in \mathcal{V}, \;\; \alpha + \beta = 1 \quad \Rightarrow \quad \alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{V}$$

$\alpha\mathbf{x} + \beta\mathbf{y}$ es una *combinación afin* de $\mathbf{x}$ e $\mathbf{y}$

Ejemplos (en $\mathbb{R}^n$)

- subespacios
- $\mathcal{V} = \mathbf{b} + \mathcal{S} = \{\mathbf{x} + \mathbf{b} \mid \mathbf{x} \in \mathcal{S}\}$ donde $\mathcal{S}$ es un subespacio
- $\mathcal{V} = \{\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \mid \alpha_i \in \mathbb{R}, \mathbf{x}_i \in \mathcal{V}, \sum \alpha_i = 1\}$
- $\mathcal{V} = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{A} \in \mathbb{R}^{k\times n}\}$

Todo conjunto afin $\mathcal{V}$ puede escribirse como $\mathcal{V} = \mathbf{x}_0 + \mathcal{S}$, donde $\mathbf{x}_0 \in \mathcal{V}$, y $\mathcal{S}$ es un subespacio

$\dim(\mathcal{V} - \mathbf{x}_0)$ es la dimensión de $\mathcal{V}$

Matrices

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

algunas matrices especiales:

- $\mathbf{A} = \mathbf{0}$ (matriz zero) : $a_{ij} = 0$
- $\mathbf{A} = \mathbf{I}$ (matriz identidad) : $m = n;\ \ a_{ii} = 1,\ i = 1,\cdots,n;\ \ a_{ij} = 0,\ i \neq j$
- $\mathbf{A} = \mathrm{diag}(\mathbf{x})$ donde $\mathbf{x} \in \mathbb{R}^n$ (matriz diagonal) : $m = n$,

$$\mathbf{A} = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix}$$

Operaciones Matriciales

- suma, resta, multiplicación por un escalar
- transpuesta :

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{n\times m}$$

- multiplicación: $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times q}$, $\mathbf{AB} \in \mathbb{R}^{m\times q}$

$$\mathbf{AB} = \begin{bmatrix} \sum_{i=1}^{n} a_{1i}b_{i1} & \sum_{i=1}^{n} a_{1i}b_{i2} & \cdots & \sum_{i=1}^{n} a_{1i}b_{iq} \\ \sum_{i=1}^{n} a_{2i}b_{i1} & \sum_{i=1}^{n} a_{2i}b_{i2} & \cdots & \sum_{i=1}^{n} a_{2i}b_{iq} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} a_{mi}b_{i1} & \sum_{i=1}^{n} a_{mi}b_{i1} & \cdots & \sum_{i=1}^{n} a_{mi}b_{iq} \end{bmatrix}$$

Imagen

la *imagen* de $A \in \mathbb{R}^{\mathrm{m \times n}}$ se define como

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$$

es la *imagen* (también conocida como *alcance* o *recorrido*)
de la transformación lineal $f : \mathbb{R}^n \to \mathbb{R}^m$ dada por $f(\mathbf{x}) = \mathbf{Ax}$

- un subespacio
- conjunto de vectores que pueden ser 'alcanzados' por $\mathbf{y} = \mathbf{Ax}$
- el espacio generado por las columnas de $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_n]$

$$\mathcal{R}(\mathbf{A}) = \mathrm{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

Rango

el *rango* de $A \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ se define como la dimensión del conjunto imagen de $\mathbf{A}$ o como el número de columnas linealmente independientes de $\mathbf{A}$

$$\operatorname{rank}(\mathbf{A}) = \dim \mathcal{R}(\mathbf{A})$$

propiedades (no triviales):

- $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^{\mathrm{T}})$
- $\operatorname{rank}(\mathbf{A})$ es el máximo número de columnas (o filas) independientes de $\mathbf{A}$, luego
$$\operatorname{rank}(\mathbf{A}) \leq \min\{m, n\}$$

decimos que $\mathbf{A} \in \mathbb{R}^{m\times n}$ es de *rango máximo* si

$$\operatorname{rank}(\mathbf{A}) = \min\{m, n\}$$

- para matrices cuadradas, rango máximo implica que $\mathbf{A}$ no es singular

Núcleo

el *núcleo* de $\mathbf{A} \in \mathbb{R}^{\mathrm{m \times n}}$ se define como

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

- un subespacio
- conjunto de vectores ortogonales a todas las filas de $\mathbf{A}$:

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}_1^{\mathrm{T}}\mathbf{x} = \cdots = \mathbf{a}_m^{\mathrm{T}}\mathbf{x} = 0\}$$

donde $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_m]^{\mathrm{T}}$

núcleo cero : $\mathcal{N}(\mathbf{A}) = \{0\} \Leftrightarrow$

- $\mathbf{x}$ se puede determinar de forma única de $\mathbf{y} = \mathbf{Ax}$
- las columnas de $\mathbf{A}$ son independientes

la *nulidad* de $\mathbf{A} \in \mathbb{R}^{\mathrm{m \times n}}$ se define como la dimensión del núcleo de $\mathbf{A}$

$$\mathrm{nulidad}(\mathbf{A}) = n - \mathrm{rank}(\mathbf{A})$$

Inversa

$\mathbf{A} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$ es *invertible* o *no singular* si $\det \mathbf{A} \neq 0$

condiciones equivalentes:

- las columnas de $\mathbf{A}$ son una base para $\mathbb{R}^n$
- las filas de $\mathbf{A}$ son una base para $\mathbb{R}^n$
- $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$
- $\mathcal{R}(\mathbf{A}) = \mathbb{R}^n$
- $\mathbf{y} = \mathbf{A}\mathbf{x}$ tiene una solución única $\mathbf{x}$ para todo $\mathbf{y} \in \mathbb{R}^n$
- $\mathbf{A}$ tiene una inversa $\mathbf{A}^{-1} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$, con $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

Funciones

$f(\mathbf{x}) \in \mathbb{R}, \quad \mathbf{x} \in \mathbb{R}^n$

la notación $f : \mathcal{S} \to \mathbb{R}$ denota que el dominio de f es $\mathcal{S}$

y que su alcance es un subconjunto de números reales

- $f : \mathbb{R}^n \to \mathbb{R}$ está definida en todas partes

$f : \mathcal{S} \to \mathbb{R}$ es *continua* en $\bar{\mathbf{x}} \in \mathcal{S}$ si para cualquier $\varepsilon > 0$, existe $\delta > 0$

tal que $\mathbf{x} \in \mathcal{S}$ y $\|\mathbf{x} - \bar{\mathbf{x}}\| < \delta$ implica $|f(\mathbf{x}) - f(\bar{\mathbf{x}})| < \varepsilon$

equivalentemente $f : \mathcal{S} \to \mathbb{R}$ es *continua* en $\bar{\mathbf{x}} \in \mathcal{S}$ si para cualquier secuencia

$\{\mathbf{x}_k\} \to \bar{\mathbf{x}}$ tal que $\{f(\mathbf{x}_k)\} \to \bar{f}$, tenemos que $f(\bar{\mathbf{x}}) = \bar{f}$

la función vector $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$ es continua en $\mathbf{x}$

si cada función componente $f_j(\mathbf{x})$, $j = 1, \cdots, m$, es continua en $\mathbf{x}$

lo cual se denota $\mathbf{f} \in \mathbb{C}$

Funciones diferenciables

la notación $\mathbf{f} \in \mathbb{C}^p$ significa que cada función componente de $\mathbf{f}$ tiene derivadas parciales de orden p continuas

si $f \in \mathbb{C}^1$, definida en $\mathbb{R}^n$, $f(\mathbf{x})$; $\mathbf{x} \in \mathbb{R}^n$,
definimos el *gradiente* de f como el vector fila

$$\nabla f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right)$$

si $f \in \mathbb{C}^2$, definida en $\mathbb{R}^n$, $f(\mathbf{x})$; $\mathbf{x} \in \mathbb{R}^n$,
definimos la *matriz hesiana* de f en $\mathbf{x}$ como la matriz

$$\nabla^2 f(\mathbf{x}) = \frac{\partial^2 f}{\partial \mathbf{x}^2}(\mathbf{x}) \in \mathbb{R}^{n \times n}$$

dado que

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

la matriz hesiana es *simétrica*

Teorema del valor medio

si $f \in \mathbb{C}^1$ en una región que contiene el segmento $[\mathbf{x}_1, \mathbf{x}_2]$ luego existe $\lambda \in (0, 1)$ tal que

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla f(\overline{\mathbf{x}})(\mathbf{x}_2 - \mathbf{x}_1),$$

donde $\quad \overline{\mathbf{x}} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$

si $f \in \mathbb{C}^2$ en una región que contiene el segmento $[\mathbf{x}_1, \mathbf{x}_2]$ luego existe $\lambda \in (0, 1)$ tal que

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) + \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)^{\mathrm{T}} \nabla^2 f(\overline{\mathbf{x}})(\mathbf{x}_2 - \mathbf{x}_1),$$

donde $\quad \overline{\mathbf{x}} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$

Teorema de Taylor

si $f \in \mathbb{C}^1$ luego para $\mathbf{x}_2$ próximo a $\mathbf{x}_1$

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) + o\left(\left\|\mathbf{x}_2 - \mathbf{x}_1\right\|\right)$$

si $f \in \mathbb{C}^2$ luego para $\mathbf{x}_2$ próximo a $\mathbf{x}_1$

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) + \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)^{\mathrm{T}} \nabla^2 f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) + o\left(\left\|\mathbf{x}_2 - \mathbf{x}_1\right\|^2\right),$$

Notación o, O:

sea $g(x), x \in \mathbb{R}$, la notación $g(x) = O(x)$ significa que $g(x)$ tiende a cero al menos tan rápido como lo hace x

mas precisamente, existe $K \geq 0$ tal que

$$\left|\frac{g(x)}{x}\right| \leq K \quad \text{cuando} \quad x \to 0$$

la notación $g(x) = o(x)$ significa que $g(x)$ tiende a cero mas rápido que $x, (K = 0)$