Calcolo numerico

Indice

1. Lezione 01

1.1. Problema matematico, metodo numerico e condizionamento

Un problema matematico in forma astratta è un problema che chiede di trovare u tale che

$$P(d, u) = 0,$$

con d insieme dei dati, u soluzione e P operatore che esprime la relazione funzionale tra u e d. Le due variabili possono essere numeri, vettori, funzioni, eccetera.

Un metodo numerico per la risoluzione approssimata di un problema matematico consiste nel costruire una successioni di problemi approssimati del tipo

$$P_n(d_n, u_n) = 0 \mid n \geq 1$$

oppure

$$P_h(d_h, u_h) = 0 \mid h > 0$$

che dipendono dai parametri n o h.

Un metodo numerico è convergente se

$$\lim_{n \to \infty} u_n = u$$

oppure

$$\lim_{h \to 0} u_h = u.$$

Il problema matematico $P(d, u) = 0$ è ben posto (o stabile) se, per un certo dato d, la soluzione u esiste ed è unica e dipende con continuità dai dati. Questa ultima proprietà indica che piccole perturbazioni (variazioni) dei dati d producono piccole perturbazioni nella soluzione u.

Per quantificare la dipendenza continua dai dati introduciamo il concetto di numero di condizionamento di un problema.

Consideriamo una funzione $f : [a, b] \longrightarrow \mathbb{R}$ in un punto x_0, ovvero

$$d \coloneqq x_0 \quad u \coloneqq f(x_0) \mid d, u \in \mathbb{R}.$$

Applichiamo lo sviluppo di Taylor di f in x_0, ovvero

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \dots$$

Ma allora

$$\begin{gathered} f(x) - f(x_0) \approx f'(x_0)(x - x_0) \\ \frac{f(x) - f(x_0)}{f(x_0)} \approx \frac{x_0 f'(x_0)}{f(x_0)} \frac{x - x_0}{x_0} \\ \left| \frac{f(x) - f(x_0)}{f(x_0)} \right| \approx \left| \frac{x_0 f'(x_0)}{f(x_0)} \right| \left| \frac{x - x_0}{x_0} \right| \end{gathered}$$

Osserviamo che

$$\Delta f(x_0) \coloneqq \frac{f(x) - f(x_0)}{f(x_0)}$$

e

$$\Delta x_0 := \frac{x - x_0}{x_0}$$

sono le variazioni relative della soluzione $u := f(x_0)$ e del dato $d := x_0$.

Chiamiamo **numero di condizionamento del calcolo di una funzione f in x_0** la quantità

$$K_f(x_0) := \left|\frac{x_0 f'(x_0)}{f(x_0)}\right|.$$

Poiché vale

$$|\Delta f(x_0)| \approx K_f(x_0)|\Delta x_0|$$

diciamo che $K_f(x_0)$ esprime il rapporto tra la variazione relativa subita dalla soluzione e la variazione relativa introdotta nel dato.

Calcolare i numeri di condizionamento nei casi:
- $f(x) = 6$ e $x_0 = 4$;
- $f(x) = e^x$ e $x_0 = 4$;
- $f(x) = 6x - x^3$ e $x_0 = 4$.

Nell'approssimare numericamente un problema fisico si commettono errori di quattro tipi diversi:
1. errori sui dati, riducibili aumentando l'accuratezza nelle misurazioni dei dati;
2. errori dovuti al modello, controllabili nella fase modellistica matematica, quando si passa dal fisico al matematico;
3. errori di troncamento, dovuti al fatto che quando si passa al limite nel calcolatore questi passaggi vengono approssimati, essendo operazioni eseguite nel discreto;
4. errori di arrotondamento, dovuti alla rappresentazione finita dei calcolatori.

L'analisi numerica studia e controlla gli errori 3 e 4.

1.2. Aritmetica floating point

L'insieme dei numeri macchina è l'insieme

$$\mathcal{F}(\beta, t, L, U) = \left\{\sigma(.a_1 a_2 ... a_t)_\beta \beta^e\right\} \cup \{0\}$$

e con il simbolo

$$\text{float}(x) \in \mathcal{F}(\beta, t, L, U)$$

il generico elemento dell'insieme, cioè il generico numero macchina.

Abbiamo:
- σ segno di $\text{float}(b)$;
- β base della rappresentazione;
- e esponente con $L \leq e \leq U$ con $L > 0$ e $U > 0$;
- t numero di cifre significative;
- $a_1 \neq 0$ e $0 \leq a_i \leq \beta - 1$;
- $m = (.a_1 a_2 ... a_t)_\beta = \frac{a_1}{\beta} + \frac{a_2}{\beta^2} + ... + \frac{a^t}{\beta^t}$ mantissa.

Facciamo un po' di osservazioni:
- $|\text{float}(x)| \in [\beta^{L-1}, (1 - \beta^{-t})\beta^U]$;
- in `MATLAB` si ha $\beta = 2$, $t = 53$, $L = -1021$ e $U = 1024$;
- il risultato di un'operazione fra numeri macchina non è necessariamente un numero macchina.

Preso il numero reale

$$x = \sigma(.a_1 a_2 ... a_t a_{t+1} a_{t+2})_\beta \beta^e \in \mathbb{R}.$$

Distinguiamo i seguenti casi:
- $L \leq e \leq U, a_i = 0 \forall i > t$ allora si ha la rappresentazione esatta di x, ovvero $\text{float}(x) = x$;
- $e < L$ allora si ha underflow, ovvero $\text{float}(x) = 0$;
- $e > U$ allora si ha overflow, ovvero $\text{float}(x) = \infty$
- se $\exists i > t \mid a_i \neq 0$ allora:
  - troncamento:
  $$\text{float}(x) = \sigma(.a_1 a_2 ... a_t)_\beta \beta^e;$$
  - arrotondamento:
  $$\sigma \begin{cases} (.a_1 a_2 ... a_t)_\beta \beta^e \text{ se } 0 \leq a_{t+1} < \frac{\beta}{2} \\ (.a_1 a_2 ... a_t + 1)_\beta \beta^e \text{ se } \frac{\beta}{2} \geq a_{t+1} \leq \beta - 1 \end{cases}.$$

Si può dimostrare che l'errore commesso approssimando un numero reale x con la sua rappresentazione macchina $\text{float}(x)$ è maggiorato da

$$\left| \frac{\text{float}(x) - x}{x} \right| \leq k \beta^{1-t}$$

con $k = 1$ per troncamento e $k = \frac{1}{2}$ per arrotondamento.

La quantità

$$\text{eps} = k \beta^{1-t}$$

è detta precisione macchina nel fissato sistema floating point. La precisione si può caratterizzare come il più piccolo numero macchina per cui vale

$$\text{float}(1 + \text{eps}) > 1.$$

Esercizio: costruire $\mathcal{F}(\beta, t, L, U)$ con $\beta = 2, t = 3, L = -1, U = 2$.

2. Lezione 02

2.1. Vettori e matrici

Una tabella di $m \times n$ numeri reali disposti in m righe e n colonne del tipo

$$A = \begin{bmatrix} a_{11} & \dots & a_1 n \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = (a_{ij}) \mid i = 1, ..., m \quad j = 1, ..., n$$

si chiama matrice di m righe e n colonne. Ogni elemento a_{ij} ha un indice di riga i e un indice di colonna j che indicano riga e colonna di A in cui si trova quell'elemento.

Indichiamo con $\mathbb{R}^{m \times n}$ l'insieme delle matrici $m \times n$.

Chiamiamo **vettore colonna** di dimensione n una matrice $n \times 1$ formata da n righe e una sola colonna. Analogamente, il **vettore riga** è una matrice di dimensione $1 \times n$ formata da una sola riga e n colonne.

AGGIUNTI ESEMPI DI VETTORI COME PRIMA.

Con il termine vettore indicheremo un vettore colonna, e l'insieme dei vettori di dimensione n lo indichiamo con $\mathbb{R}^n$.

Usiamo vettori e matrici per rappresentare molte grandezze fisiche che non possono essere rappresentate come scalari, ma come vettori (tipo spostamento, velocità, accelerazione, eccetera).

Siano $a = (a_i), b = (b_i) \in \mathbb{R}^n$ due vettori, si chiama vettore somma il vettore $c = (c_i) \in \mathbb{R}^n$ tale che

$$c_i = a_i + b_i \forall i = 1...n.$$

Geometricamente parlando, il vettore somma è la diagonale del parallelogramma avente due lati coincidenti con a e b (regola del parallelogramma).

AGGIUNGI IMMAGINE CARINA.

La somma di vettori gode di alcune proprietà:
- **commutativa**: $\forall a, b \in \mathbb{R}^n \quad a + b = b + a$;
- **associativa**: $\forall a, b, c \in \mathbb{R}^n \quad (a + b) + c = a + (b + c)$;
- **esistenza del neutro**: il vettore $0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ è l'elemento neutro della somma, cioè $\forall a \in \mathbb{R}^n \quad a + 0 = 0 + a = a$;
- **esistenza dell'opposto**: per ogni vettore $a \in \mathbb{R}^n$ esiste un altro vettore $b \in \mathbb{R}^n$ tale che $a + b = 0$; tale vettore b viene detto vettore opposto di a e si indica con $-a$.

Siano $a = (a_i) \in \mathbb{R}^n$ un vettore e $\beta \in \mathbb{R}$ uno scalare. Si chiama prodotto vettore-scalare il vettore $c = (c_i) \in \mathbb{R}^n$ tale che

$$c_i = \beta a_i \forall i = 1, ..., n.$$

Valgono le due proprietà distributive:
- $\forall \alpha \in \mathbb{R} \quad \forall a, b \in \mathbb{R}^n \quad \alpha(a + b) = \alpha a + \alpha b$;
- $\forall \alpha, \beta \in \mathbb{R} \quad \forall a \in \mathbb{R}^n \quad (\alpha + \beta)a = \alpha a + \beta a$.

Siano $a = (a_i), b = (b_i) \in \mathbb{R}^n$ due vettori; si chiama prodotto scalare lo scalare $c = a \cdot b \in \mathbb{R}$ tale che

$$c = a \cdot b = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + ... + a_n b_n.$$

Diciamo che l'applicazione

$$\|\cdot\| : \mathbb{R}^n \longrightarrow \mathbb{R}^+ \cup \{0\}$$

è una norma vettoriale se valgono le seguenti condizioni:
1. $\|x\| \geq 0 \forall x \in \mathbb{R}^n$ e $\|x\| = 0$ se e solo se $x = 0$;
2. $\|\alpha x\| = |\alpha| \|x\| \forall \alpha \in \mathbb{R} \quad \forall x \in \mathbb{R}^n$;
3. $\|x + y\| \leq \|x\| + \|y\| \forall x, y \in \mathbb{R}^n$.

Le norme più famose sono quella euclidea (detta norma 2) tale che

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}} \forall x \in \mathbb{R}^n$$

oppure la norma 1 tale che

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \forall x \in \mathbb{R}^n$$

oppure la norma ∞ (norma del massimo) tale che

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \forall x \in \mathbb{R}^n.$$

Una matrice si dice quadrata (di ordine n) se $m = n$. Una matrice quadrata è triangolare superiore (inferiore) se

$$a_{ij} = 0 \mid i > j (i < j),$$

cioè se sono nulli gli elementi al di sotto (sopra) della diagonale principale a_{ii}.

Se valgono entrambe le definizioni la matrice è detta diagonale.

Data la matrice $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ si chiama matrice trasposta la matrice $A^T = (a_{ij}^T) \in \mathbb{R}^{n \times m}$ ottenuta dallo scambio delle righe e delle colonne di A, ovvero

$$a_{ij} = a_{ji}^T$$

Sia A una matrice quadrata di ordine n, essa si dice simmetrica se $A = A^T$, ovvero $a_{ij} = a_{ji} \forall i, j = 1, ..., n$.

Siano $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m \times n}$ due matrici, si chiama matrice somma la matrice $C = (c_{ij}) \in \mathbb{R}^{m \times n}$ tale che

$$c_{ij} = a_{ij} + b_{ij} \forall i = 1, ..., m \forall j = 1, ..., n.$$

Anche la somma di matrici gode di alcune proprietà:
- **commutativa**: $\forall A, B \in \mathbb{R}^{m \times n} \quad A + B = B + A$;
- **associativa**: $\forall A, B, C \in \mathbb{R}^{m \times n} \quad (A + B) + C = A + (B + C)$;
- **esistenza del neutro**: la matrice $0 = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}$ è l'elemento neutro della somma, cioè $\forall A \in \mathbb{R}^{m \times n} \quad A + 0 = 0 + A = A$;
- **esistenza dell'opposto**: per ogni matrice $A \in \mathbb{R}^n$ esiste un'altra matrice $B \in \mathbb{R}^{m \times n}$ tale che $A + B = 0$; tale matrice B viene detta matrice opposta di A e si indica con $-A$.

Siano $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ una matrice e $\beta \in \mathbb{R}$ uno scalare. Si chiama prodotto matrice-scalare la matrice $C = (c_{ij}) \in \mathbb{R}^{m \times n}$ tale che

$$c_{ij} = \beta a_{ij} \forall i = 1, ..., m \forall j = 1, ..., n.$$

Valgono le due proprietà distributive:

- $\forall \alpha \in \mathbb{R} \quad \forall A, B \in \mathbb{R}^{m \times n} \quad \alpha(A + B) = \alpha A + \alpha B$;
- $\forall \alpha, \beta \in \mathbb{R} \quad \forall A \in \mathbb{R}^{m \times n} \quad (\alpha + \beta)A = \alpha A + \beta A$.

Sia $A = \left(a_{ij}\right) \in \mathbb{R}^{m \times n}$ una matrice e $b = \left(b_i\right) \in \mathbb{R}^n$ un vettore; si chiama prodotto matrice-vettore di A per b il vettore $c = \left(c_i\right) \in \mathbb{R}^m$ tale che

$$c_i = \sum_{j=1}^{n} a_{ij} b_j = a_{i1} b_1 + ... + a_{in} b_n \forall i = 1, ..., m.$$

Siano $A = \left(a_{ij}\right) \in \mathbb{R}^{m \times n}$ e $B = \left(b_{ij}\right) \in \mathbb{R}^{n \times k}$ due matrici; si chiama prodotto matrice-matrice di A per B la matrice $C = \left(c_{ij}\right) \in \mathbb{R}^{m \times k}$ tale che

$$c_{ij} = \sum_{t=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + ... + a_{in} b_{nj} \forall i = 1, ..., m \forall j = 1, ..., k.$$

Il prodotto di matrici in generale non è commutativo, cioè $A \cdot B \neq B \cdot A$.

Si chiama matrice identità di ordine n la matrice quadrata $I = \left(i_{kj}\right)$ di ordine n tale che

$$i_{kj} = \begin{cases} 1 \text{ se } k = j \\ 0 \text{ se } k \neq j \end{cases}.$$

Si può dimostrare che $A \cdot I = I \cdot A = A$.

L'applicazione

$$\|\cdot\| : \mathbb{R}^{n \times n} \longrightarrow \mathbb{R}^+ \cup \{0\}$$

è una norma matriciale se valgono le seguenti condizioni:

1. $\|A\| \geq 0 \forall A \in \mathbb{R}^{n \times n}$ e $\|A\|$ se e solo se $A = 0$;
2. $\|\alpha A\| = |\alpha| \|A\| \forall \alpha \in \mathbb{R} \forall A \in \mathbb{R}^{n \times n}$;
3. $\|A + B\| \leq \|A\| + \|B\| \forall A, B \in \mathbb{R}^{n \times n}$;
4. $\|A \cdot B\| \leq \|A\| \cdot \|B\| \forall A, B \in \mathbb{R}^{n \times n}$.

Definiamo la norma matriciale indotta dalla norma vettoriale come

$$\|A\| = \sup\left\{ \frac{\|A_x\|}{\|x\|}, \forall x \in \mathbb{R}^n / \{0\} \right\}.$$

Abbiamo alcuni casi particolari:

- norma 1 (calcolata colonna per colonna), calcolata come

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} \left|a_{ij}\right|;$$

- norma ∞ (calcolata per riga), calcolata come

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} \left|a_{ij}\right|.$$

3. Lezione 03

3.1. Determinante, inversa e rango di matrici

Sia A una matrice quadrata di ordine due, ovvero

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Si chiama determinante di A il numero reale

$$\det(A) := a_{11}a_{22} - a_{12}a_{21} \in \mathbb{R}.$$

Ora vediamo determinanti per matrici di ordine maggiore.

Siano A matrice quadrata di ordine n e a_{ij} il generico elemento; si chiama complemento algebrico di a_{ij} il numero reale

$$\operatorname{compl}(a_{ij}) := (-1)^{i+j} \det(A_{ij}),$$

dove la matrice A_{ij} è la matrice quadrata di ordine $n-1$ ottenuta da A eliminando la riga i e la colonna j.

Sia A una matrice quadrata di ordine n, fissata una qualunque riga o colonna di A, il determinante di A si ottiene sommando il prodotto di ogni elemento di tale riga o colonna per il suo complemento algebrico.

Il calcolo del determinante è indipendente dalla riga o colonna scelta, quindi conviene fissare la riga o colonna con il maggior numero di zeri.

Il determinante gode di alcune proprietà:
- se A è triangolare allora $\det(A) = a_{11}a_{22}...a_{nn}$;
- se A ha una riga o una colonna di soli zeri allora $\det(A) = 0$;
- se A ha due righe o colonne uguali allora $\det(A) = 0$;
- vale il Teorema di Binet, ovvero se A, B sono due matrici quadrate dello stesso ordine allora $\det(A \cdot B) = \det(A) \cdot \det(B)$.

Sia A una matrice quadrata di ordine n, si dice che A è invertibile se esiste una matrice A^{-1} detta matrice inversa di A quadrata di ordine n tale che $A \cdot A^{-1} = A^{-1} \cdot A = I_n$.

Teorema: sia A una matrice quadrata di ordine n, allora A è invertibile se e solo se $\det(A) \neq 0$.

Teorema: sia A una matrice quadrata di ordine due, cioè

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

e supponiamo $\det(A) \neq 0$, allora

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Sia A una matrice $m \times n$ e $k \in \mathbb{N}$ con $k \leq \min(m, n)$. Si chiama minore di ordine k estratto da A il determinante di una qualunque sottomatrice quadrata di ordine k di A, ottenuta prendendo gli elementi comuni a k righe di k colonne di A. Si chiama caratteristica o rango di A ($\operatorname{rk}(A)$) l'ordine massimo dei minori non nulli che si possono estrarre da A.

In altre parole, $\mathrm{rk}(A) = r$ se esiste un minore di ordine r diverso da zero e se tutti i minori di ordine $r + 1$ sono nulli.

Sia A una matrice non nulla, allora $\mathrm{rk}(A) \geq 1$. Inoltre, $\mathrm{rk}(A) \leq \min(m, n)$.