NUMERO : 210 DATA : 15/02/2012

A P P U N T I

STUDENTE : Alessio

MATERIA : Metodi Numerici e Calcolo Scientifico,
Teoria + Esercizi + Dispense + Temi d'esame
Prof. Puppo

3/10/11

gabriella.puppo@polito.it

su gigapedia ci sono i libri che sono indicati sul portale

Esame

- relazioni facoltative in gruppo
- scritto con appunti (su ultime 2 parti del corso → eq. ellittiche, eq. iperboliche)
- orale (calcolo numerico)

⇓

media tra scritto e orale

NUMERI DI MACCHINA

La macchina memorizza i numeri in base 2. Noi lavoreremo in base 10.

es. $x = 123.5$ → ogni cifra viene associata ad una posizione

$x = 1 \cdot 10^2 + 2 \cdot 10^1 + 3 \cdot 10^0 + 5 \cdot 10^{-1}$

base 10 → 10 cifre (da 0 a 9 compresi)

$x = 1235 \cdot 10^{-1}$ è lo stesso numero, solo scritto diversam.

Il computer però è stupido → il fatto che ci sia ambiguità è un problema → bisogna decidere come dare i numeri al computer → si usa la FORMA NORMALIZZATA:

$$x = \pm 0.\underbrace{a_1 a_2 a_3 \ldots}_{\text{cifre}} \beta^m \qquad a_1 \neq 0, \quad 0 \leq a_i \leq \beta - 1$$

β = base di numerazione

Non si possono rappresentare numeri infiniti in un computer → né numeri troppo grossi, né troppo piccoli, né infiniti.

Avremo quindi NUMERI FLOATING POINT:

$$x = \pm 0.\underbrace{a_1 a_2 \ldots a_t}_{\text{mantissa}} \beta^m \qquad L \leq m \leq U, \quad L<0, \quad U>0$$

t = n° di cifre della mantissa

Tutti i numeri reali compresi tra 2 num. FP vengono approssimati col num FP più vicino.

DISTANZA FRA 2 NUM. FP VICINI:

$x_1 = 0.a_1 a_2 \dots a_t \cdot \beta^m$

$x_2 = 0.a_1 a_2 \dots (a_t+1) \cdot \beta^m$ ($a_t \neq 9$ se no è più complicato)

$\boxed{\Delta x = \beta^{-t}\beta^m}$ → se m è grande → i numeri sono più larghi

⇒ I numeri sono addensati attorno allo zero e si allargano man mano che ci si allontana dallo zero → a seconda del numero commetto errori di approssimaz. più o meno grandi.

⇒ In caso di approssimaz. successive (es. bisezione) devo dire al computer dopo quit fermarsi, e devo tenere conto della spaziatura tra due num. FP.

Dato $x \in \mathbb{R}$ $x = 0.a_1 \dots a_t \beta^m$, l'approssimaz. FP di x è:

$$|\tilde{x} - x| \leq \frac{1}{2}\beta^{-t}\beta^m$$

ERRORE RELATIVO:

$$\frac{|\tilde{x}-x|}{|x|} \leq \frac{\frac{1}{2}\beta^{-t}\beta^m}{|0.a_1 \dots a_t|\beta^m} \leq \frac{1}{2}\frac{\beta^{-t}}{\beta^{-1}} \rightarrow \boxed{\frac{|\tilde{x}-x|}{|x|} \leq \frac{1}{2}\beta^{1-t} = \text{eps}}$$

eps = precisione di macchina

$0.a_1 \dots a_t \geq \beta^{-1}$

L'errore relativo non dipende da quit è grande il numero

In doppia precisione $\text{eps} = 2^{-52} = 2{,}2 \cdot 10^{-16}$

esempio:

Voglio calcolare e^x:

$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ → approssimo e^x con una successione:

$k! = k(k-1)!$ $\qquad S_m = \sum_{k=0}^{m} \frac{x^k}{k!}$ $\qquad e^x = \lim_{m\to\infty} S_m$

$0! = 1$

Però calcoli infiniti non me posso fare → devo scrivere un programma che sappia da solo quando smettere.

$x = 0.123 \cdot 10^{-3}$
$y = 0.123 \cdot 10^{4}$ $\}\rightarrow x + y = y$

I numeri sono troppo diversi per poter essere sommati

Nelle somme e nelle sottrazioni l'errore di macchina si propaga tantissimo.

⇒ per calcolare e^x usando la sommatoria, dopo un po' il mio risultato resta lo stesso

Le operazioni non sono più commutative (se sommo prima i numeri più piccoli e poi quelli più grandi il mio risultato migliora)

SISTEMI LINEARI

$Ax = b$
↑ matrice di rigidezza
↑ vettore dei termini noti

$A \rightarrow m \times m \qquad x \rightarrow m \times 1$

$Ax \rightarrow m \times 1$

$(Ax)_i = \sum_{j=1}^{m} a_{ij} x_j$

$$\left[\ A\ \right]\left[\ x\ \right] = \left[\ Ax\ \right]$$

Il costo di una componente di Ax è:

$(Ax)_i = a_{i1} x_1 + a_{i2} x_2 + \dots + a_{im} x_m$

devo quindi svolgere m (1 prodotto + 1 somma + 2 operazioni di memoria)

→ devo effettuare m flop (flop)

⇒ Tutto Ax costa m^2 flop

Il computer non calcola ogni componente di Ax alla volta, ma ne calcola più di una contemporaneamente, pescando le varie componenti di x una volta sola → BLAS (Basic Linear Algebra Symbolycs)

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 9 di 314

Suppongo che x sia il mio vettore

$\|x\| = 0$ se tutte le componenti di x sono $x_i = 0 \iff x = 0$

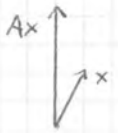
la norma dipende da tutte le operaz. continue del vettore

$\|x\|$ piccola $\iff$ x_i piccole

COME SI CALCOLA LA NORMA DI UNA MATRICE?

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

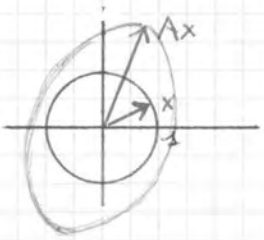
x vettore $\to$ Ax è un vettore di lungh e direz. diverse



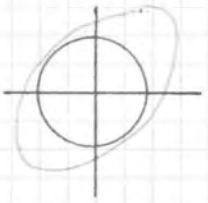
$\frac{x}{\|x\|} \parallel x$ ma $\left\| \frac{x}{\|x\|} \right\| = 1$

$$\|A\|_p = \max_{x \neq 0} \left\| A \frac{x}{\|x\|} \right\| = \max_{\|x\| = 1} \|Ax\|$$

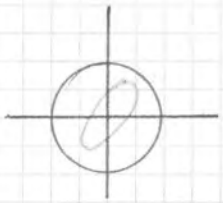
- Matrici 2×2 $(p = 2)$



faccio qst per tutti i pt che sono sulla circonf. di raggio 1 ed ottengo un'ellisse la norma di A è il semidiametro max dell'ellisse



$\|A\| > 1$



$\|A\| < 1$ $\to$ matrice di contraz.

Il cerchio unitario cambia forma a seconda della norma

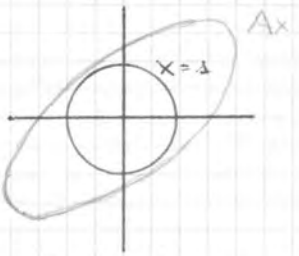
↳

SISTEMA LINEARE

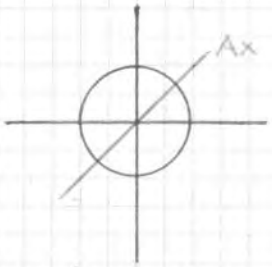
$Ax = b \qquad A \ m \times m \qquad b, x \in \mathbb{R}^m$

Ho un'unica sol. $x = A^{-1}b$ se A è non singolare (cioè è invertibile):

1) A non singolare
2) $\det A \neq 0$
3) $Ax = 0 \Rightarrow x = 0$
4) $\exists A^{-1}$ t.c. $AA^{-1} = I$
5) colonne di A l.i.
6) $\lambda = 0$ non è un autovalore di A



matrice A invertibile



matrice A singolare

A^{-1} non si calcola mai (costa tanto tempo macchina)

NUMERO DI CONDIZIONAMENTO DEL SISTEMA LINEARE

$Ax = b$ $\begin{cases} x = \text{incognita} \\ b = \text{dati (vettore dei dati)} \\ A = \text{operatore} \end{cases}$

$\tilde{b}$ = dati perturbati $\rightarrow$ $A\tilde{x} = \tilde{b}$

Vorrei calcolare k:

$$\frac{\|x - \tilde{x}\|_p}{\|x\|_p} \leq k \frac{\|b - \tilde{b}\|_p}{\|b\|_p}$$

- $Ax - A\tilde{x} = b - \tilde{b} \rightarrow A(x - \tilde{x}) = b - \tilde{b} \rightarrow x - \tilde{x} = A^{-1}(b - \tilde{b})$
  $\rightarrow \|x - \tilde{x}\| = \|A^{-1}(b - \tilde{b})\| \leq \underbrace{\|A^{-1}\|}_{\text{qst andrà a finire in } k} \cdot \|b - \tilde{b}\|$ (1)

- $Ax = b \rightarrow \|b\| = \|Ax\| \leq \|A\| \cdot \|x\| \rightarrow \frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|}$ (2)

Metto insieme (1) e (2) $\rightarrow \frac{\|x - \tilde{x}\|_p}{\|x\|_p} \leq \underbrace{\|A\|_p \cdot \|A^{-1}\|_p}_{k_p(A)} \frac{\|b - \tilde{b}\|_p}{\|b\|_p}$

N.B. • le matrici ortogonali (gli cui le colonne ortogonali → che rappresentano una rotaz.) hanno n° di condizionam. pari ad 1 e si indicano con la lettera Q.

$K_2(Q) = 1$

- Matrici di tipo FEM $K_2(A) \simeq 10^4$
- Matrici di Hilbert (H_m) $K(H_m) \sim e^m$ ← mal condizionata

$$H_{ij} = \frac{1}{i+j-1}$$

es. matrice FEM:

$$\frac{\|x - \tilde{x}\|}{\|x\|} = 10^4 \frac{\|b - \tilde{b}\|}{\|b\|}$$

Se l'errore dei miei dati viene dall'arrotondam. f.p. si ha:

$$\frac{\|b - \tilde{b}\|}{\|b\|} \sim 10^{-16} \Rightarrow \frac{\|x - \tilde{x}\|}{\|x\|} \sim 10^{-12}$$ ← errore piccolo

Se errore viene da una misura sperimentale, allora:

$$\frac{\|b - \tilde{b}\|}{\|b\|} \sim 10^{-3} \Rightarrow \frac{\|x - \tilde{x}\|}{\|x\|} \sim 10$$ ← spazzatura

Su Matlab:
cond(A) ≃ $K_2(A)$
cond(A,p) ≃ $K_p(A)$

RISOLUZIONE DI UN SISTEMA LINEARE

- Cramer richiede n! operazioni (10! ≃ 10^6, 50! ≃ 10^{64})
- Riduzione per righe → n^3 operaz.
  - ↳ la chiameremo FATTORIZZAZIONE LU (eliminazione di Gauss)

Per "molte" matrici A
∃ una matrice triang. inf L (⊠)
∃ " " " sup U (⊠)
t.c. A = LU

→

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 15 di 314

$$A^{(3)} = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 3/2 & -1 \\ 0 & 0 & 5/3 \end{pmatrix} = U$$

Nota che $A^{(3)} = M_2 A^{(2)}$ con $M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/3 & 1 \end{bmatrix}$

Quindi:

$$U = M_2 A^{(2)} = M_2 M_1 A \quad \Rightarrow \quad A = \underbrace{(M_2 M_1)^{-1}}_{L} U$$

$$(M_2 M_1)^{-1} = M_1^{-1} M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -m_{32} & 1 \end{bmatrix} = L$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & -m_{32} & 1 \end{bmatrix}$$

i problemi ci sono quando, calcolando i moltiplicatori, ci troviamo a dover dividere per zero.

6/10/11

Fattorizzaz. LU:

$A = LU$ $\quad A$ $m \times m$

procedo 1 colonna e 1 riga per volta $\rightarrow$ $k = 1, \dots m-1$

- calcolo i moltiplicatori: $m_{ik} = -\dfrac{a_{ik}}{a_{kk}}$ $\quad i = k+1, \dots m$

Aggiorno il resto della matrice sulla riga i:

$$a_{ij} = a_{ij} + m_{ik} a_{kj} \qquad j = k+1, \dots m$$

$A^{(1)} = A$

per $k = 1, \dots m-1$

  per $i = k+1, \dots m$

$$m_{ik} = -\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} \longrightarrow \boxed{a_{kk}^{(k)} \neq 0}$$

  per $j = k+1, \dots m$

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} + m_{ik} a_{kj}^{(k)}$$

  end

end

end

In qst modo i moltiplicatori che calcolo dopo aver effettuato lo scambio rimangono piccoli → algoritmo stabile.

Esempio

(perché è importante che i moltiplicatori siano piccoli)

Sistema fp : $\beta=10 \quad t=4$

- $A = \begin{pmatrix} \varepsilon & 1 \\ 1 & 1 \end{pmatrix}$ $\quad 1 = 0.1 * 10^1$ SENZA PIVOTING

Suppongo $\varepsilon = 10^{-5} \rightarrow \varepsilon = 0.1 * 10^{-4}$

$m_{21} = -\frac{1}{\varepsilon} \Rightarrow$ la 2ª riga diventa $[0, 1 + m_{21} * 1] = [0, 1 - 10^5]$

ora devo memorizzare qst num. nel mio sist. fp.

Chiamo $\hat{x}$ l'approssimaz fp di x

$$A = LU = \begin{bmatrix} 1 & 0 \\ \frac{1}{\varepsilon} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon & 1 \\ 0 & 1 - \frac{1}{\varepsilon} \end{bmatrix}$$

$$\hat{A} = \hat{L}\hat{U} = \begin{bmatrix} 1 & 0 \\ \frac{1}{\varepsilon} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon & 1 \\ 0 & -\frac{1}{\varepsilon} \end{bmatrix} = \begin{bmatrix} \varepsilon & 1 \\ 1 & 0 \end{bmatrix} \neq A$$

- Se invece avessi avuto $A = \begin{pmatrix} 1 & 1 \\ \varepsilon & 1 \end{pmatrix}$ CON PIVOTING

$m_{21} = -\frac{\varepsilon}{1} = -\varepsilon \rightarrow$ 2ª riga $[\varepsilon \ 1] - \varepsilon [1 \ 1] = [0 \ 1-\varepsilon]$

$$A = LU = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1-\varepsilon \end{bmatrix}$$

$$\hat{A} = \hat{L}\hat{U} = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & \hat{\varepsilon+1} \end{bmatrix} \simeq A$$

$\hat{\varepsilon+1} \simeq 1$

⇒ La fattorizz. LU con pivoting è più stabile della fattorizz. senza pivoting, cioè amplifica meno l'effetto degli errori di arrotondam.

⇒ con algoritmo stabile < prob ben condiz. → ok
prob mal condiz. → forse non peggiora troppo le cose.

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 19 di 314

$* = \sqrt{16+9} = \pm 5$ (useremo $*=5$)

$\tilde{Q} = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}$ devo trovare s, c t.c. $\begin{pmatrix} c & -s \\ s & c \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$

$\begin{cases} 4c - 3s = 5 \\ 4s + 3c = 0 \rightarrow s = -\frac{3}{4}c \end{cases} \rightarrow 4c + \frac{9}{4}c = 5 \Leftarrow \begin{matrix} c = \frac{4}{5} \\ s = -\frac{3}{5} \end{matrix}$

$\tilde{Q} = \frac{1}{5}\begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix}$ → ora devo trovare R

$R = \tilde{Q}A = \frac{1}{5}\begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix}\begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 25 & 11 \\ 0 & -2 \end{pmatrix}$

$A = QR \rightarrow Q = \tilde{Q}^{-1} = \tilde{Q}^T \Rightarrow Q = \frac{1}{5}\begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix}$

$A = QR = \frac{1}{5}\begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 5 & 11/5 \\ 0 & -2/5 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix}$

Se ho una matrice più grossa → guardo piccoli blocchi: 2×2 e faccio delle piccole rotaz. in modo da inserire degli zeri nella matrice.

Abbiamo quindi visto 3 modi per risolvere i sistemi lineari fatt. LU, fatt. LU con pivoting, fatt. QR; l'ultimo algoritmo è il più stabile ma è molto più complicato.

15/10/15

Sistemi sovradeterminati

$Ax = b \qquad A = m \times n \quad m > n \quad (m \gg n)$

$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \cdot \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} \\ b \\ \\ \end{bmatrix}$ $\quad x = n \times 1$, $b = m \times 1$

In generale, ~~$\exists$~~ x t.c. $Ax = b$, ma $Ax - b = r$ (residuo) $\forall x$

Bisogna trovare x t.c. $\|r\|_2^2$ sia minima (→ risolvere $Ax = b$ nel senso dei minimi quadrati: $\|r\|_2^2 = \sum r_i^2$)

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 21 di 314

cioè x è s.t. $A^TAx = A^Tb$

$$\begin{bmatrix} A^T \end{bmatrix}\begin{bmatrix} A \end{bmatrix} = \begin{bmatrix} A^TA \end{bmatrix}$$

$$\begin{bmatrix} A^T \end{bmatrix}\begin{bmatrix} b \end{bmatrix} = \begin{bmatrix} A^Tb \end{bmatrix}$$

$A = m \times n$

$A^T = n \times m$

$A^TA = n \times n$

$A^Tb = n \times 1$

$A^TAx = A^Tb$ sistema $n \times n$

$\exists! x$ se A^TA è non singolare $\iff$ le colonne di A sono lin. indip.

Per risolvere x nel senso dei minimi quadrati calcolo x tale che

$A^TAx = A^Tb$ (metodo delle equaz. normali)

→ metodo veloce, ma le colonne di A non sono l.i. ma lo sono quasi → se $\text{cond}(A) \gg 1 \Rightarrow \text{cond}(A^TA) = [\text{cond}(A)]^2 \gg\gg 1$

Esempio

(retta di regressione – 2ª puntata)

$$\|r\|_2^2 = \sum r_i^2 = \sum (Mx_i + Q - f_i)^2 = \varphi(M,Q)$$

$$\nabla\varphi = \begin{bmatrix} \partial_M \varphi \\ \partial_Q \varphi \end{bmatrix}$$

$\partial_M \varphi$ = derivata rispetto ad M di φ

$\partial_Q \varphi$ = derivata rispetto a Q di φ

$$\begin{cases} \partial_M \varphi = \sum_i 2(Mx_i + Q - f_i)\, x_i = 0 \\ \partial_Q \varphi = \sum_i 2(Mx_i + Q - f_i)\cdot 1 = 0 \end{cases}$$

$$\Rightarrow \begin{cases} 2M\sum_i x_i^2 + 2Q\sum x_i - 2\sum_i f_i x_i = 0 \\ 2M\sum_i x_i + 2Q\sum_i 1 - 2\sum_i f_i = 0 \end{cases}$$

$$\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & \sum 1 \end{bmatrix}\begin{bmatrix} M \\ Q \end{bmatrix} = \begin{bmatrix} \sum f_i x_i \\ \sum f_i \end{bmatrix}$$

Sistema delle eq. normali per la retta di regressione

N.B. se x è sol. di $Ax = b \Rightarrow$ è anche sol. di $A^TAx = A^Tb$

→ il residuo è nullo.

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 23 di 314

Riassumendo

Metodo QR:

- risolvo $Ax=b$ nel senso dei minimi quadrati
- calcolo $A=QR$
- risolvo $Rx=Q^Tb$
- $Rx = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \Rightarrow R_1 x = b_1$ (b_1 = prime m componenti di Q^Tb)

Vantaggi: il sistema $Rx=Q^Tb$ ha lo stesso condizionam. di A → sistema più stabile

Svantaggi: è più lento

in matematica esatta i 2 metodi sarebbero uguali, ma $cond(Q)=1$ e non devo fare A^TA

Soluz. di sistemi lineari in Matlab

$Ax=b \Rightarrow x = A\backslash b$ \ = backslash

Qst operaz. significa:

- se sist. quadrato → $m \equiv n$ ⇒ calcola $LU=A$ e risolve $Ly=b$ e $Ux=y$
- se sist. non quadrato → $m>n$ ⇒ calcola $QR=A$ e risolve $Rx=Q^Tb$ nel senso dei min. quadrati

Stampa un warning solo se A è malcondizionata:

$rcond$ piccolo ($\sim 10^{-12}$), $rcond \triangleq \frac{1}{cond(A)}$

⇒ conviene farsi sempre stampare il residuo, così so cosa sta facendo Matlab (se $r=0$ ⇒ sta usando LU)

```
res = norm (ex-b)
```

Oppure:

$x = \underbrace{inv(A)}_{\text{calcola } A^{-1}} * b$ ⇒ il probl ora è come si calcola A^{-1}

Chiamo $[\tilde{a}_1, \tilde{a}_2, \dots \tilde{a}_n] = [A^{-1}]$ le colonne di A^{-1}.

Calcolo le colonne $\tilde{a}_j$ 1 per volta, usando $A \cdot A^{-1} = \mathbb{I}$:

$A \cdot \tilde{a}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ ⇒ Risolvo il sist. lineare e trovo $\tilde{a}_1$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 25 di 314

Oppure:

$[l, u, p] = lu(a)$ ← però grosso spreco di memoria

Se si vuole QR:

$[q, r] = qr(a)$

Calcolo degli autovalori

λ è autoval di A con autovett. w : $Aw = \lambda w \quad w \neq 0$

$\rightarrow (A - \lambda \mathbb{I})\, w = 0 \Rightarrow A - \lambda \mathbb{I}$ è singolare, cioè $\det(A - \lambda \mathbb{I}) = 0$

Quindi, se A è $m \times m$ → $\det(A - \lambda \mathbb{I}) = p^m(\lambda)$ (polinomio di grado m in λ) Trovare gli autovalori vuol dire trovare le radici di $p^m(\lambda) = 0$

⇒ m soluzioni ($\mathbb{R}$ o $\mathbb{C}$ non necessariamente distinte)

Calcolo dell'autovalore di modulo max per matrici diagonalizzabili

(cioè matrici che hanno un sistema completo di autovettori l. i. $\underline{w}_1, \dots \underline{w}_m$)

Supponiamo che l'autoval. che stiamo cercando sia isolato

$\lambda_1 > \lambda_2 \geq \lambda_3 \dots \geq \lambda_m \quad (\lambda_i \in \mathbb{R})$

- Parto con una stima iniziale $\underline{x}_0 = \sum_{i=1}^{m} \alpha_i \underline{w}_i \quad , \alpha_i \in \mathbb{R}$
- Calcolo $A\underline{x}_0 = A\left(\sum_{i=1}^{m} \alpha_i \underline{w}_i\right) =$

$$= \sum_{i=1}^{m} A(\alpha_i \underline{w}_i) = \sum_{i=1}^{m} \alpha_i A \underline{w}_i$$

Poiché $\underline{w}_i$ è autovettore ⇒ $A\underline{x}_0 = \sum_{i=1}^{m} \alpha_i \lambda_i \underline{w}_i$

- Calcolo $A^2\underline{x}_0 = A(A\underline{x}_0) =$

$$= A \sum_{i=1}^{m} \alpha_i \lambda_i \underline{w}_i = \sum_{i=1}^{m} \alpha_i \lambda_i \underbrace{\lambda_i \underline{w}_i}_{A\underline{w}_i} = \sum_{i=1}^{m} \alpha_i \lambda_i^2 \underline{w}_i$$

- Calcolo $A^k \underline{x}_0 = \sum_{i=1}^{m} \alpha_i \lambda_i^k \underline{w}_i$

l'autoval max diventa sempre più grande degli altri

- Isolo il primo termine:

$$A^k \underline{x}_0 = \alpha_1 \lambda_1^k \underline{w}_1 + \sum_{i=2}^{m} \alpha_i \lambda_i^k \underline{w}_i = \lambda_1^k \left[\alpha_1 \underline{w}_1 + \sum_{i=2}^{m} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \underline{w}_i \right]$$

frazione < 1 in modulo, k grande → tende a zero

$A^k \underline{x}_0 \xrightarrow[k \text{ grande}]{} \lambda_1^k \alpha_1 \underline{w}_1$ ho trovato la direzione dell'autovettore di λ_1

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 27 di 314

→ esci se $\|Ax_k - \mu_k x_k\| < tol \cdot \|A\|$

Se μ_k sta effettivam. convergendo ad un autoval di A → esco

N.B. test arresto ha sempre dentro:

- tolleranza (tol) [di solito eps, cmq Matlab fa da solo gli che vuole, ma si può sempre intervenire per cambiarla]
- test relativo ($\leq$)

meglio non usare un ciclo while perché se l'algoritmo non converge rischio di rimbalzare tra 2 posiz. → meglio un ciclo for con k_{max} (se max esce a k_{max} senza convergere)

Qual è la velocità di convergenza?

Dipende da quanto è grande la coda di quelle che butto via.

→ dipende da $\left(\frac{\lambda_2}{\lambda_1}\right)$

METODO DELLE POTENZE INVERSE

Serve per calcolare l'autovalore di modulo minimo

Ipotesi:

A ha n autovett. lin. indip. $\underline{w}_1, \dots \underline{w}_n$ con autovalori

$\lambda_1 \geq \lambda_2 \geq \dots > \lambda_n$

Uso il fatto che se $A\underline{w}_n = \lambda_n \underline{w}_n$ (λ_n autoval min)

$\Rightarrow A^{-1}\underline{w}_n = \frac{1}{\lambda_n}\underline{w}_n$ $\left(\frac{1}{\lambda_n}\right)$ // max di A^{-1}

→ Applico l'algoritmo delle potenze ad A^{-1}:

x_0, $y_0 = x_0 / \|x_0\|$

per $k = 0$: k_{max}

$x_{k+1} = A^{-1} y_k$

$y_{k+1} = x_{k+1} / \|x_{k+1}\|$

$\mu_{k+1} = (y_{k+1}, A^{-1} y_{k+1})$ $\left(\to \frac{1}{\lambda_n}\right)$

test di arresto

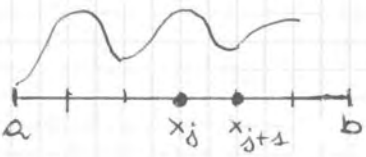
end

Posso evitare di calcolare A^{-1}?

↳

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 29 di 314

* Funzioni polinomiali a tratti



Su ogni intervallo l'interpolante coincide con un polinomio:

$I_m(x) = P_j^k(x) \quad x \in (x_j, x_{j+1})$

* Funzioni wavelets (ondine)

Servono soprattutto per algoritmi di compressione: ho tanti dati da trasmettere di cui molti ridondanti, non significativi → voglio comprimerli senza perdere troppe informazioni. Sono alla base del jpeg e degli mp3

Norma di funzioni

$$\|f\|_p = \left\{ \int |f|^p \right\}^{1/p}$$

$p = 2 \rightarrow \|f\|_2 = \left(\int |f|^2 dx \right)^{1/2}$

$p = 1 \rightarrow \|f\|_1 = \int |f| dx$

$p = \infty \rightarrow \|f\|_\infty = \sup_x |f(x)|$

Se f è continua: $\|f\|_p = 0 \iff f \equiv 0$

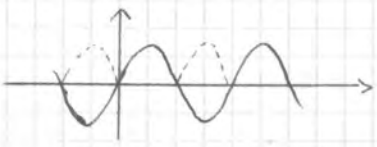
Se f non è continua: $\int |f| = 0 \not\Rightarrow f(x) = 0 \ \forall x$

$\Rightarrow f(x) = 0$ quasi ovunque

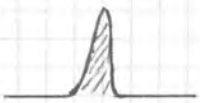
Esempio 1

$f(x) = \sin(x) \rightarrow \|f\|_1 = \int_{-\infty}^{+\infty} |f(x)| = +\infty$

$\rightarrow \|f\|_\infty = 1$



Esempio 2

$f(x) =$ 

$\rightarrow \|f\|_1$ è piccolo

$\rightarrow \|f\|_\infty$ è molto grande

Esempio 3

f è una funz. a gradino → la approssimo con un polinomio p^m continuo →

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 31 di 314

N.B. se $n = 1 \rightarrow$ retta di regressione lineare.

Interpolazione esatta $(m = n+1)$

Il prob è lineare perché lo spazio dei polinomi di grado n è uno spazio lineare:

$$\mathbb{P}^n = \{ v : v \text{ è un pol. di grado } n \}$$

↑ spazio dei polinomi di grado $\leq n$ (alcuni coeff. possono essere nulli)

La dimensione di $\mathbb{P}^n$ è:

$$\dim \mathbb{P}^n = n+1$$ (n° di gradi di lib. delle mie funz. dof = = degrees of freedom)

Quindi ogni elemento $v \in \mathbb{P}^n$ può essere scritto come una combinaz. lineare di una base di $\mathbb{P}^n$, $\phi_1(x), \phi_2(x) \cdots \phi_{n+1}(x)$, cioè $\exists\ \alpha_1, \alpha_2, \cdots \alpha_{n+1}$ $(\in \mathbb{R})$ t.c.

$$v(x) = \sum_{j=1}^{n+1} \alpha_j \phi_j(x) \quad \forall\, v \in \mathbb{P}^n$$

Nell'interpolaz. polinomiale abbiamo usato come base:

$$\phi_1(x) = x^n, \quad \phi_2(x) = x^{n-1}, \quad \cdots \quad \phi_n(x) = x, \quad \phi_{n+1}(x) = x^0 = 1;$$

la matrice V_{nn} diventa:

$$V_n = \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_{n+1}(x_1) \\ \vdots & & & \\ \phi_1(x_{n+1}) & \cdots & \cdots & \phi_{n+1}(x_{n+1}) \end{bmatrix}$$

α_1 = coord. → cambiano a seconda della base utilizzata

$\text{cond}(V_n) \gg 1 \rightarrow$ ≁

Poniamo pensare di cercare un sistema di coordinate α in modo che V_n sia ben condizionata. → cambiamo la base

Base dei poli di lagrange

Ho fissato la griglia $x_1, \cdots x_{n+1}$ e cerco dei polinomi di grado n $\ell_i(x)$ t.c. $\ell_i(x_j) = \begin{cases} 0 & j \neq i \\ 1 & j = i \end{cases}$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 33 di 314

17/10/11

Supponiamo f continua con tutte le derivate continue su $[a, b]$.
Vogliamo scrivere $P^m(x)$ t.c. $P^m(x_j) = f(x_j)$ $j = 1, \dots m+1$

$P^m(x)$ può essere scritto in 2 modi diversi:

$$P^m(x) = a_1 x^{m+1} + a_2 x^m + \dots + a_{m+1}$$

$$P^m(x) = \sum_{j=1}^{m+1} f(x_j)\, \ell_j(x)$$

ℓ_j pol. di Lagrange $\ell_j(x_i) = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$

Matlab usa il primo modo (anche se calcolo dei coeff mal condiz.) perché si divide il probl in 2:

$\underline{a} = \text{polyfit}(\underline{x}, f(\underline{x}), m) \rightarrow$ calcola $a_1, \dots a_{m+1}$

$P = \text{polyval}(\underline{a}, \underline{xx}) \rightarrow$ $\underline{xx}$ = vettore contenente le ascisse su cui voglio calcolare il polinomio P^m

(La 2ª formulaz. dovrebbe ricalcolarsi i polinomi di Lagr. ogni volta che cambio la x)

Cmq anche se coeff mm tanto a posto → il polinomio ne risente pochissimo

Errore nell'interpolazione polinomiale

$$e_m(x) = P^m(x) - f(x)$$

Mi interessa sapere qnt è grosso → norma

$$\| e_m \|_\infty = \| P^m(x) - f(x) \|_\infty \underbrace{\not\to 0}_{m \to \infty}$$

(norma ∞ perché ho tutte le derivate continue)

vogliamo capire perché succede questo.

So che l'errore calcolato sui nodi della griglia è nullo:

$e_m(x_j) = 0$ perché $f(x_j) = P^m(x_j)$

Allora posso scrivere (in forma fattorizzata):

$$e_m(x) = \underbrace{(x-x_1)(x-x_2)\cdots(x-x_{m+1})}_{\text{funz. NODALE}} \underbrace{R_m(x)}_{\text{funz. RESTO}}$$

$\omega_{m+1}(x) = \prod_{j=1}^{m+1} (x - x_j)$

dipende solo dalla griglia (qnt nodi ho e come sono disposti)

non dipende più dalla griglia, dipende solo dalla funz.

ω_{n+1} n dispari → ω_{n+1} pari

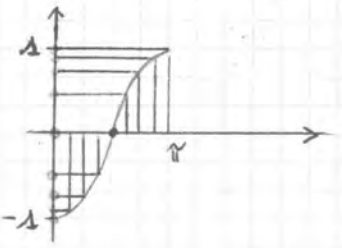
$\|\omega_{n+1}\|_\infty \xrightarrow[n\to\infty]{} \infty$

ho dei corni più alti agli estremi

Per migliorare la situazione pongo infittire la griglia ai bordi → griglie gaussiane

Per esempio: Gauss - Lobatto:

su $[-1,1]$ $\quad x_j = -\cos\left(\frac{\pi(j-1)}{n}\right) \quad j = 1, \dots n+1$

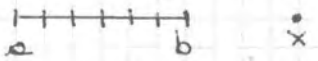


Su un intervallo $[a,b]$: $\quad x_j = \frac{a+b}{2} - \frac{b-a}{2}\left(\cos\left(\frac{\pi(j-1)}{n}\right)\right)$

Si dimostra che:

- $\|\omega_{n+1}\|_\infty \sim 2^{n+1}$ griglia equispaziata → va a ∞ più velocemente
- $\|\omega_{n+1}\|_\infty \sim \log(n+1)$ griglia di Gauss - Lobatto

In ogni caso, $e_n(x)$ sarà grande se $x \notin [a,b]$ (estrapolazione)



(es previsioni del tempo: dati raccolti tra oggi e gli scorsi giorni, poi ovviamente ci allontano da quest intervallo e più l'errore grande. In borsa le previsioni sono ancora peggiori)

Numero di condizionamento

Ho dei dati (valori della funz. sui nodi di griglia) $f(x_j)$

trovo $p^n(x)$

inoltre ho dei dati perturbati $\tilde{f}(x_j)$ → $\delta(x_j) = |f(x_j) - \tilde{f}(x_j)|$

trovo $\tilde{p}^n(x)$

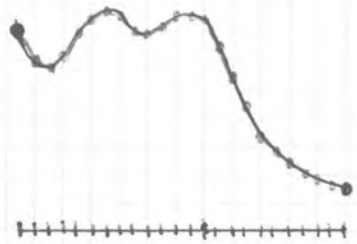
calcolo il n° di cond.:

$$\frac{\|p^n - \tilde{p}^n\|_\infty}{\|p\|_\infty} \leq K(p) \frac{\|f - \tilde{f}\|_\infty}{\|f\|_\infty}$$

d' interpolazione globale funziona bene per:

- f regolare → alti valori di n con griglie qualsiasi
- f poco regolare → bassi valori di n con griglie equispaziate

Grafico di una funzione



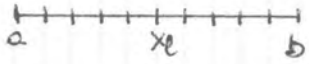
$\underline{x}$, $f(\underline{x})$

plot ($\underline{x}$, $f(\underline{x})$)

interpolazione di grado 1 su ogni intervallo

Interpolazione polinomiale a tratti

tutti polinomi di grado basso messi insieme



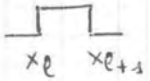
$a = x_1 , \dots x_\ell , \dots x_{m+1} = b$

$I_\ell = (x_\ell , x_{\ell+1})$

Un interpolante polinom. a tratti di grado m è:

$$y^d(x) = \sum_{\ell=1}^{M} p_\ell^m(x)\, \chi_\ell(x)$$

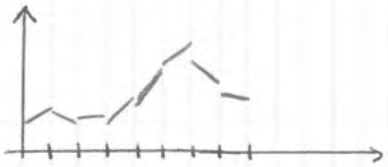
$$\chi_\ell(x) = \begin{cases} 1 & x \in I_\ell \\ 0 & x \notin I_\ell \end{cases}$$



Se $x \in I_\ell \rightarrow \chi_\ell(x) \neq 0$

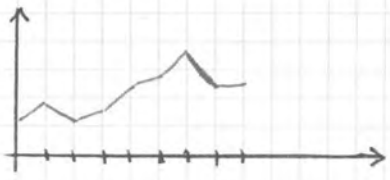
$$y^d(x) = p_\ell^m(x)$$

1) Per m=1 lineare a tratti



(qst si usa per es per le onde d'urto → visto che ci sono discontinuità → avrò un polinomio discontinuo).

2) Per m=1 continua → impongo $y^1(x)$ continuo



Interpolazione lineare a tratti

Calcolo una funz. $y^1(x)$ t.c. $y^1(x) = f(x)$ sui nodi della griglia

$$y^1(x_\ell) = f(x_\ell) \qquad \ell = 1, \dots \; m+1$$

$\Rightarrow$ $y^1(x)$ deve essere continua attraverso i nodi

Uso lo spazio delle funz. lineari a tratti continue $\rightarrow \dim(\;) = m+1$

Poiché impongo $m+1$ condizioni:

$$y^1(x_\ell) = f(x_\ell) \qquad \ell = 1, \dots \; m+1$$

$\Rightarrow$ ho una sola funz. interpolante.

Se chiamo $l_j(x)$ la base delle funz. a cappello, ottengo:

$$y^1(x) = \sum_{j=1}^{m+1} f(x_j)\, l_j(x) .$$

Consideriamo una griglia uniforme di spaziatura h:

$x_{\ell+1} - x_\ell = h$

Vorrei valutare l'errore $|f(x) - y^1(x)| = |e(x)|$

Sul singolo intervallo I_ℓ si ha:

$$e_\ell(x) = f(x) - y^1(x)\big|_{I_\ell} = f(x) - P_\ell^1(x)$$

d'errore sul singolo intervallo è:

$$e_\ell(x) = f(x) - P_\ell^1(x) = \omega_2(x)\big|_{I_\ell} \cdot R^1(x)\big|_{I_\ell}$$

I_ℓ = $\overline{\qquad}$ (x_ℓ — $x_{\ell+1}$)

$$\omega_2(x)\big|_{I_\ell} = (x - x_\ell)(x - x_{\ell+1})$$

$$R^1(x)\big|_{I_\ell} = \frac{1}{2!} f''(\xi_\ell) \qquad \xi_\ell \in I_\ell$$

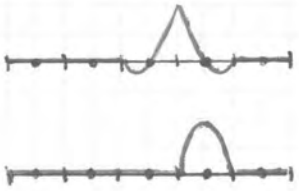
Se chiamo $k = \max_{[a,b]} \|f''\| \Rightarrow \|R^1(x)|_{I_\ell}\| \le k$

$$|\omega_2(x)| = |x - x_\ell| \cdot |x - x_{\ell+1}| \le h \cdot h = h^2$$

Mettendo tutto insieme $\Rightarrow$ $|e_\ell(x)| \le \frac{1}{2!} h^2 \|f''\|_\infty \quad \forall \ell$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 41 di 314

• caso continuo



m+1 funzioni a tenda

m funzioni a bolla

L'interpolante quadratico a tratti di una funz. è:

$$y^2(x) : \begin{cases} y^2(x_e) = f(x_e) & e = 1, \dots m+1 \\ y^2(x_{e\,interni}) = f(x_{e\,interni}) & e_{interni} = 1, \dots m \end{cases}$$

$$x_{e\,interni} = \frac{1}{2}(x_e + x_{e+1})$$

L'errore sarà:

$$e^2(x) = f(x) - y^2(x)$$

Sull'intervallo I_e:

$$e^2(x)|_{I_e} = f(x) - P_e^2(x) = \omega_3(x) R^2(x)|_{I_e}$$

$$|e^2(x)|_{I_e}| = |f(x) - P_e^2(x)| = |\omega_3(x) R^2(x)|_{I_e}| \leq \frac{1}{3!} h^3 \| f'''(x) \|_\infty$$

<u>Interpolazione polinomiale di grado n a tratti</u>

$$\| e^n(x) \|_\infty \leq \frac{1}{(n+1)!} h^{n+1} \| f^{(n+1)} \|_\infty$$

Su ogni intervallo inserisco n-1 nodi intermedi

Funz cost a tratti ($P^0(x)$) → ?

(1) Funz. cubica a tratti ($P^3(x)$) → 2 nodi intermedi

↓ oppure su 1 nodo più condiz.

4m gdl − (vincoli) = 4m − (m−1) = 3m+1 gdl

(2) interpolazione con polinomi di grado 3 continui e con derivata 1ª continua (no pt angolosi)

4m − 2(m−1) = 2m+2 dof

→ 1 solo nodo intermedio per tratto

(3) Interpolaz. spline → polinomi di grado 3 y, y', y'' continue

4m − 3(m−1) = m+3 dof → nessun nodo intermedio

(applicazioni di grafica al computer → no discontinuità visibili ad occhio)

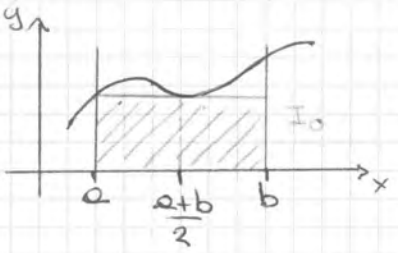
© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 43 di 314

Per polinomi di grado 1:

$$I = \int_a^b x\,dx = \frac{1}{2}(b^2 - a^2)$$

$$I_1 = (b-a)\left(\frac{a+b}{2}\right) = \frac{1}{2}(b^2 - a^2)$$

⇒ la formula del punto medio integra esattam. anche polinomi di grado 1.



se intervallo ab risultati dignitosi

FORMULA DEL PT MEDIO

Per es. $\boxed{m=1}$

Scelgo i nodi $x_1 = a$, $x_2 = b$ → $I_1 = w_1 f(a) + w_2 f(b)$

Impongo che $I = I_1$ per polinomi x^0, x^1

per x^0:

$$\left.\begin{aligned} I &= \int_a^b 1\,dx = b-a \\ I_1 &= w_1 \cdot 1 + w_2 \cdot 1 \end{aligned}\right\} \quad b-a = w_1 + w_2 \quad (1)$$

per x^1:

$$\left.\begin{aligned} I &= \int_a^b x\,dx = \frac{1}{2}(b^2 - a^2) \\ I_1 &= w_1 a + w_2 b \end{aligned}\right\} \quad \frac{1}{2}(b^2 - a^2) = w_1 a + w_2 b \quad (2)$$

Sistema lineare (l'integrale è un operatore lineare):

$$\begin{cases} (1) \\ (2) \end{cases} \rightarrow w_1 = w_2 = \frac{b-a}{2} \Rightarrow I_1 = \frac{b-a}{2}\left[f(a) + f(b)\right]$$

Per polinomi di grado 2:

$$1 = \int_a^b x^2\,dx = \frac{1}{3}(b^3 - a^3)$$

$$I_2 = \frac{b-a}{2}(a^2 + b^2)$$

$\rightarrow I \neq I_2$

La formula integra esattam. solo i polinomi di grado 0 e di grado 1.

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 45 di 314

Per es, se cerco la formula dei trapezi:

$$y_1 = -1 \quad , \quad y_2 = 1 \quad , \quad w_1 = w_2 = 1$$

Esercizio

Trovare pesi e precisione per la formula basata sui nodi $x_1 = \frac{-1}{\sqrt{3}}$ e $x_2 = \frac{1}{\sqrt{3}}$ su $[-1, 1]$ e trasformarla sull'intervallo $[a,b]$

$$x^0 \to I = \int_{-1}^{1} 1\,dx = 2 \qquad I_1 = w_1 + w_2 \qquad \Big\} \; w_1 + w_2 = 2$$

$$x^1 \to I = \int_{-1}^{1} x\,dx = \frac{1}{2}(1-1) = 0 \qquad I_2 = w_1\left(-\frac{1}{\sqrt{3}}\right) + w_2\frac{1}{\sqrt{3}} \qquad \Big\} \; \frac{w_1}{\sqrt{3}} = \frac{w_2}{\sqrt{3}} \to w_1 = w_2$$

$$\Big\} \; w_1 = w_2 = 1$$

$$\int_{-1}^{1} f(x)\,dx = \sum_{j=1}^{2} f(x_j)\,w_j = w_1 f\left(-\frac{1}{\sqrt{3}}\right) + w_2 f\left(\frac{1}{\sqrt{3}}\right) = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)$$

$$\int_a^b f(x)\,dx \simeq \frac{b-a}{2}\sum_{j=1}^{2} w_j f(x(y_j)) = \frac{b-a}{2}\left[w_1 f\left(\frac{a+b}{2} + \frac{b-a}{2}y_1\right) + w_2 f\left(\frac{a+b}{2} + \frac{b-a}{2}y_2\right)\right]$$

$$y_j = \left(x_j - \frac{a+b}{2}\right)\frac{2}{b-a} \to y_2 = \left(-\frac{1}{\sqrt{3}} - \frac{a+b}{2}\right)\frac{2}{b-a} = \frac{-2-a-b}{2\sqrt{3}}\cdot\frac{2}{b-a} = \frac{-2-a-b}{(b-a)\sqrt{3}}$$

$$\to y_1 = \frac{2-a-b}{(b-a)\sqrt{3}}$$

Come si comporta l'errore?

$$e_m = \left| \int_a^b f(x)\,dx - \int_a^b f_m(x)\,dx \right| = \left| \int_a^b [f(x) - f_m(x)]\,dx \right|$$

Per forza bruta:

$$e_m \le \int_a^b |f(x) - f_m(x)|\,dx = \max_x |f(x) - f_m(x)| \int_a^b 1\,dx =$$

$$= \underbrace{\|f - f_m\|_\infty (b-a)}_{\text{errore di interpolaz.}} \le \|\omega_{m+1}\|_\infty \cdot \frac{b-a}{(m+1)!}\|f^{(m+1)}\|_\infty$$

ω_{m+1}: funz. nodale

$$|\omega_{m+1}(x)| \le (b-a)^{m+1} \qquad \omega_{m+1} = \prod_{j=1}^{m+1}(x - x_j)$$

$$\Rightarrow e_m \le \frac{1}{(m+1)!}(b-a)^{m+2}\|f^{(m+1)}\|_\infty \quad \text{dipende da} \begin{cases} m \\ b-a \\ \text{regolarità di } f \end{cases}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 47 di 314

24/ACAS1

FORMULE DI QUADRATURA COMPOSITE

Supponiamo di avere una griglia che copre tutto l'intervallo $[a,b]$:

$$a = x_0 < x_1 \quad \dots \quad < x_m = b$$

$$\int_a^b f\,dx = \sum_{j=1}^{m} \int_{x_{j-1}}^{x_j} f\,dx$$

(↑ formula esatta)

Chiamo $Q_{[x_{j-1}, x_j]}$ la formula di quadratura applicata a $\int_{x_{j-1}}^{x_j} f$

Esempio:

Usiamo la regola dei trapezi:

$$Q_{[x_{j-1}, x_j]} = \frac{x_j - x_{j-1}}{2}\left(f(x_j) + f(x_{j-1})\right)$$

Quindi:

$$\int_a^b f(x)\,dx \approx \sum_{j=1}^{m} Q_{[x_{j-1}, x_j]} = \sum_{j=1}^{m} \frac{x_j - x_{j-1}}{2}\left(f(x_j) + f(x_{j-1})\right)$$

Se la griglia è uniforme $\rightarrow x_j - x_{j-1} \equiv h$, $h = \frac{b-a}{m}$

$$\int_a^b f\,dx \simeq \frac{h}{2}\sum_{j=1}^{m}\left[f(x_j) + f(x_{j-1})\right] = h\sum_{j=1}^{m} f(x_j) - \frac{h}{2}\left[f(x_0) + f(x_m)\right]$$

Errore

$$E = \int_a^b f\,dx - Q = \sum_{j=1}^{m}\left[\int_{x_{j-1}}^{x_j} f\,dx - Q_{[x_{j-1}, x_j]}\right]$$

Su un singolo intervallo l'errore è:

$$E_{x_j} = \int_{x_{j-1}}^{x_j} f\,dx - Q_{[x_{j-1}, x_j]} = \frac{h^3}{12} f''(\xi_j) \qquad \xi_j \in (x_{j-1}, x_j)$$

Quindi, l'errore globale è:

$$|E| = \left|\sum_{j=1}^{m} E_j\right| \le \sum_{j=1}^{m} |E_j| = \sum_{j=1}^{m}\left|\frac{h^3}{12} f''(\xi_j)\right| = \frac{h^3}{12}\sum_{j=1}^{m}\|f''(\xi_j)\| \le$$

$$\le \|f''(\xi_j)\|_\infty \frac{h^3}{12}\sum_{j=1}^{m} 1 \rightarrow |E| \le \|f''\|_\infty \frac{h^3 m}{12}$$

quando h diventa piccolo, m diventa grande → mi ricordo che $h = \frac{b-a}{m}$ e ricavo che:

$$|E| \le \|f''\|_\infty \frac{b-a}{12} h^2 \rightarrow \text{se } h \text{ scende di } \frac{1}{2} \rightarrow E \text{ scende di } \frac{1}{4}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 49 di 314

Infatti, sommando tutti gli E_j:

$$E = \sum |E_j| \leq \frac{\varepsilon}{b-a} \cdot \sum (x_j - x_{j-1}) = \frac{\varepsilon (b-a)}{b-a}$$

somma telescopica → i termini a 2 a 2 si cancellano

Quant'è grande E_j?

Mi serve uno stimatore d'errore, in modo da sfruttare la soluzione numerica stessa per stimare E_j

Stimatore d'errore a posteriori

(stimatore a priori → usa la soluz. esatta)

Chiamo $h_j = x_j - x_{j-1}$

Calcolo $E_j = \int_{x_{j-1}}^{x_j} f\,dx - Q_{[x_{j-1}, x_j]} = \frac{h^3}{12} f''(\xi_j) \quad \xi_j \in (x_{j-1}, x_j)$

Poi inserisco un punto $x_{j-\frac{1}{2}}$ tra x_{j-1} e x_j e calcolo l'integrale come somma di 2 contributi

$$\tilde{E}_j = \int_{x_{j-1}}^{x_j} f\,dx - \left[Q_{[x_{j-1}, x_{j-\frac{1}{2}}]} + Q_{[x_{j-\frac{1}{2}}, x_j]} \right] = \frac{1}{12}\left[\left(\frac{h_j}{2}\right)^3 f''(\xi_j^1) + \left(\frac{h_j}{2}\right)^3 f''(\xi_j^2) \right]$$

Ora devo fare delle ipotesi:

- suppongo che f sia abbastanza regolare da supporre che $f'' \simeq \text{cost}$ su $[x_{j-1}, x_j] \Rightarrow f''(\xi_j) \simeq f''(\xi_j^1) \simeq f''(\xi_j^2)$

$$\Rightarrow E_j = \int_{x_{j-1}}^{x_j} f\,dx - Q_{[x_{j-1}, x_j]} = \frac{1}{12} h_j^3 f''(\xi_j)$$

$$\tilde{E}_j = \int_{x_{j-1}}^{x_j} f\,dx - \left[Q_{[x_{j-1}, x_{j-\frac{1}{2}}]} + Q_{[x_{j-\frac{1}{2}}, x_j]} \right] = \frac{1}{12} h_j^3 f''(\xi_j) \left(\frac{1}{4}\right) \quad \frac{1}{8} + \frac{1}{8}$$

- Sottraggo le 2 equazioni $E_j - \tilde{E}_j$

(Chiamo $Q^{(2)}_{[x_1, x_2]} = Q_{[x_{j-1}, x_{j-\frac{1}{2}}]} + Q_{[x_{j-\frac{1}{2}}, x_j]}$)

$$E_j - \tilde{E}_j \simeq Q^{(2)}_{[x_1, x_2]} - Q_{[x_{j-1}, x_j]} = E_j - \frac{1}{4} E_j = \frac{3}{4} E_j$$

Quindi lo stimatore d'errore è ⟶

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 51 di 314

1) Il difetto di qst algoritmo è che h_j può solo diminuire → se ho una funz. liscia voglio poter aumentare h_j (soddisfando cmq l'accuratezza).
Introduco un nuovo parametro $c < 1$ ($c = \frac{1}{4}$ opp. $c = \frac{1}{10}$ di solito)

⋮

Se s_j ~~$< \frac{\varepsilon h_{j+1}}{b-a}$~~ → sostituisco con

Se $\frac{c\,\varepsilon h_{j+1}}{b-a} < s_j < \frac{\varepsilon h_{j+1}}{b-a}$ ← (la stima è soddisfatta appena, aggiorna l'integrale ma tieni lo stesso passo)

$Q_{[a, x_{j+1}]} = Q_{[a, x_j]} + Q^{(2)}_{[x_j, x_{j+1}]}$

incremento j e go to (+)

Se $s_j < \frac{c\,\varepsilon h_{j+1}}{b-a}$ ← (la stima è soddisfatta abbondantemente, aggiorna l'integrale ma per il futuro prova un passo più grande)

$Q_{[a, x_{j+1}]} = Q_{[a, x_j]} + Q^{(2)}_{[x_j, x_{j+1}]}$

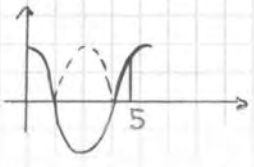
$h_{j+1} = 2 h_{j+1}$

Incremento j e go to (+)

2) Manca ancora la garanzia che si arrivi alla fine.
(Ad intervalli sempre più piccoli rischio di arrivare allo zero di macchina)
Devo introdurre un passo minimo h_{min}

(•) $h_{j+1} = \max(h_{min}, h_{j+1}/2)$
Se $h_{j+1} = h_{min}$ aggiorna l'integrale anche se s_j è troppo grande e vai oltre => stampa un warning (accuratezza non soddisfatta)

Tipicam. qst succede in 2 casi:
- chiedo di calcolare un integrale che non esiste
  es. $\int_0^{\pi/2} \tan x$
- gli faccio attraversare dei punti in cui le derivate della funz. non sono ben definite
  es. $\int_0^5 |\cos x|$



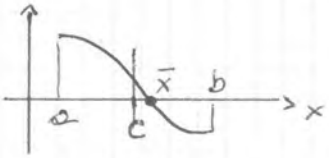
→ gli chiedo:
$\int_0^{\pi/2} |\cos x| + \int_{\frac{\pi}{2}}^{\frac{3}{2}\pi} |\cos x| + \int_{\frac{3}{2}\pi}^{5} |\cos x|$

Inoltre suppongo che lo zero sia unico.

Chiamo $c = \frac{a+b}{2}$

Se $f(a)f(c) > 0 \rightarrow \bar{x} \in (c, b)$

else $\rightarrow \bar{x} \in (a, c)$



Vado avanti finché ad ogni l'intervallo è talmente piccolo che ho trovato lo zero:

$a_0 = a$

$b_0 = b$

per $k = 0, 1, \dots$

$c = (a_k + b_k)/2$

if $f(a_k) f(c) > 0$

$a_{k+1} = c$; $b_{k+1} = b_k$

else

$a_{k+1} = a_k$; $b_{k+1} = c$

end

test di arresto

end

$e_{k+1} = |\bar{x} - x_{k+1}| \leq (b_{k+1} - a_{k+1}) = \frac{1}{2}(b_k - a_k) \rightarrow e_{k+1} = \frac{1}{2} e_k$

$(e_k) = |\bar{x} - x_k| \leq (b_k - a_k) = \frac{1}{2}(b_{k-1} - a_{k-1}) \leq \left(\frac{1}{2}\right)^k (b - a)$

al livello k

il metodo di bisezione converge sempre se funz. continua e con uno zero

test di arresto sulle iterate successive:

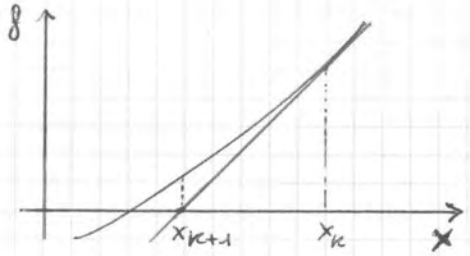
esci se $(b_k - a_k) \leq tol \left(\frac{a_k + b_k}{2}\right)$

test di arresto sul residuo:

esci se $|f(c)| \leq tol - |f_0|$ es. $|f_0| = \frac{1}{2}(|f(a)| + |f(b)|)$

Qst metodo però è lento 😕

→ si può usare il METODO DI NEWTON



retta tan ad f per x_k:

$y = f'(x_k)(x - x_k) + f(x_k)$

$(x_{k+1}, y = 0)$

$0 = f'(x_k)(x_{k+1} - x_k) + f(x_k)$

$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 55 di 314

Qst metodo però nn conv. sempre ☹

⇒ devo usare un mix dei vari metodo visti finora:

- robustezza di bisezione
- circa la stessa velocità di IQI

Parto da (a,b), $f(a)f(b)<0$

$c = \frac{a+b}{2}$ $x_0 = a$, $x_1 = c$, $x_2 = b$

per $k = 2, \dots k_{max}$

$x_{k+1} = IQI(x_k, x_{k-1}, x_{k-2})$

Se $x_{k+1} \in (a_k, b_k)$ continua (ciclo for)

Altrimenti butta via x_{k+1} e applica un passo di bisezione a (a_k, b_k)

N.B. se IQI converge ad una sol. sbagliata, qst algoritmo nn se ne accorge ☹ Per fortuna di solito IQI diverge

Matlab:

fzero(f, x_0) x_0 = pt. di partenza

versione base che rischia di divergere ☹

⇒ fzero(f, [a,b])

dà errore per $f(a)f(b)>0$ (qst routine si aspetta che ci sia uno zero solo e nn per es. 2)

Bisogna controllare cmq cosa succede:

es.

fzero(tan x, 1)

→ dà 1,57... che è $\frac{\pi}{2}$ che è un punto in cui f cambia segno, ma nn è uno zero di f !!!

⇒ mi faccio stampare il residuo

[x, res] = fzero(f, [a,b])

nell'es. dà res = f(x) → mi accorgo dell'errore!

Dall'analisi:

se f è lipshitziana in y e continua in t $\Rightarrow \exists$ 1 solo soluz.

f è' lipshitziana se un insieme $I \times A$ $\quad I \in \mathbb{R}$ (intervallo) $\quad A \in \mathbb{R}^n$

$$\| f(t,y_2) - f(t,y_2) \| \leq L \cdot \| y_2 - y_1 \|$$

$$\forall t \in I \ , \forall y_1, y_2 \in A$$

In particolare, se f è differenziabile con continuità $\Rightarrow$ è lipshitziana

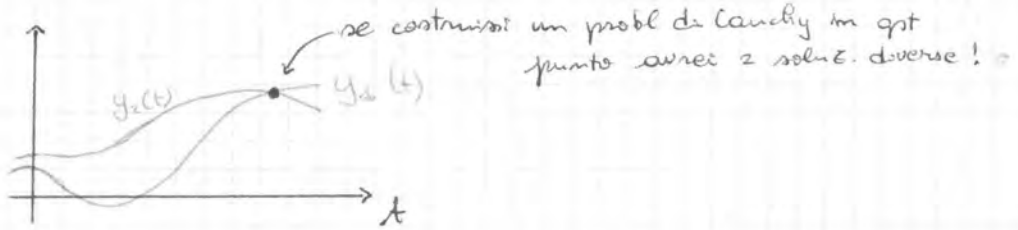
$$| f(t,y_1) - f(t,y_2) | = | f'(t,\xi) | \cdot | y_2 - y_1 |$$
$$\xi \in A$$

Se f' è continua:

$$L = \sup_{y \in A} | f'(t,y) | \qquad t \text{ fissato}$$

f differenziabile $\Rightarrow$ f lipshitziana $\Rightarrow$ f continua

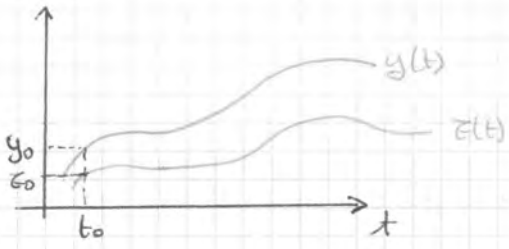
Se la soluzione è unica $\Rightarrow$ 2 soluz. diverse non si possono mai intersecare



Continuità rispetto ai dati iniziali

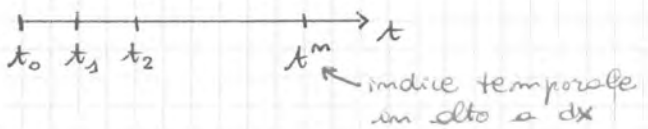
Supponiamo di avere 2 probl. di Cauchy:

$$\begin{cases} y' = f(t,y) \\ y(t_0) = y_0 \end{cases} \qquad \begin{cases} z' = f(t,z) \\ z(t_0) = z_0 \end{cases}$$



le funz. convergono o divergono? Dipende dal problema! Se è ben condizionato, un piccolo errore dei dati cambia di poco le soluzioni, un probl. mal condiz. cambierà di molto le soluzioni

Numericamente introduco la griglia



$y(t)$ soluz. esatta
$u(t)$ soluz. numerica

La soluz. numerica è def solo agli istanti che stanno sulla griglia

$$u^m = u(t^m)$$

Anche per la soluz. esatta, sulla griglia $y^m = y(t^m)$

Chiamo $\Delta t_n = t^{n+1} - t^n$ il passo di integraz.
Per ora $\Delta t = cost$

$y' = f(t,y)$. Conosco y''

$$y' \cong \frac{y(t^m + \Delta t) - y(t^m)}{\Delta t} + O(\Delta t)$$

(O grande)

Per la soluz. numerica:

$$\frac{u(t^m + \Delta t) - u(t^m)}{\Delta t} = f(t^m, u(t^m))$$

$$u(t^m + \Delta t) = u(t^m) + \Delta t\, f(t^m, u(t^m))$$

cioè $\boxed{u^{m+1} = u^m + \Delta t\, f(t^m, u^m)}$ metodo di Eulero esplicito o Forward-Euler

Parto con $n=0 \rightarrow u^0 = y_0$ (dato iniziale)

$u^1 = u^0 + \Delta t\, f(t^0, u^0)$
$u^2 = u^1 + \Delta t\, f(t^1, u^1)$
⋮

Calcolo u^m ad ogni istante t^m
N.B. non posso calcolare una soluz. qualunque senza aver prima calcolato tutte le soluz. precedenti

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 61 di 314

$$\frac{y(t^n + \Delta t) - y(t^n)}{\Delta t} = y'\left(t^n + \frac{\Delta t}{2}\right) + O(\Delta t)^2$$

④ Crank Nicolson (implicito)

$$u^{n+1} = u^n + \frac{\Delta t}{2}\left(f(t^n, u^n) + f(t^{n+1}, u^{n+1})\right)$$

② Punto medio

$$\frac{u(t^n + \Delta t) - u(t^n - \Delta t)}{2\Delta t} = f\left(t^n, u(t^n)\right)$$

$$\rightarrow u^{n+1} = u^{n-1} + 2\Delta t\, f(t^n, u^n)$$

3 metodi di prima erano ad 1 passo $u^n \mapsto u^{n+1}$

Il metodo del pt medio è a 2 passi: $(u^n, u^{n-1}) \mapsto u^{n+1}$

Errore locale

errore introdotto in un unico passo di integrazione $\Rightarrow u(t^n) = y(t^n)$

Calcolo l'errore nel passo successivo:

$$e_{\Delta t} = y(t^n + \Delta t) - u(t^{n+1})$$

Errore locale di troncamento

$$\alpha_{\Delta t} = \frac{e_{\Delta t}}{\Delta t} = \frac{1}{\Delta t}\left[y(t^n + \Delta t) - u(t^{n+1})\right]$$

Quindi posso valutare l'accuratezza del metodo che uso:

- EE $\rightarrow$ $\alpha_{\Delta t} = \frac{1}{\Delta t}\left\{y(t^n + \Delta t) - \left[u^n + \Delta t\, f(t^n, u^n)\right]\right\}$

  uso l'ipotesi $u^n = y(t^n)$

  $$\alpha_{\Delta t} = \frac{1}{\Delta t}\left\{y(t^n + \Delta t) - \left[y(t^n) + \Delta t\, f(t^n, y(t^n))\right]\right\}$$

  $$y(t^n + \Delta t) = y(t^n) + \Delta t\, y'(t^n) + \frac{(\Delta t)^2}{2} y''(\xi)$$

  (in serie)

  $$\xi \in (t^n,\ t^n + \Delta t)$$

  Poiché y è la soluz. esatta $\rightarrow y'(t^n) = f(t^n, y(t^n))$

  $$y'' = \frac{d}{dt}(f)$$

  $$\Rightarrow \alpha_{\Delta t} = \frac{1}{\Delta t}\left\{\frac{(\Delta t)^2}{2}\left.\frac{df}{dt}\right|_\xi\right\} \rightarrow \alpha_{\Delta t} = \frac{(\Delta t)}{2}\left.\frac{df}{dt}\right|_\xi$$

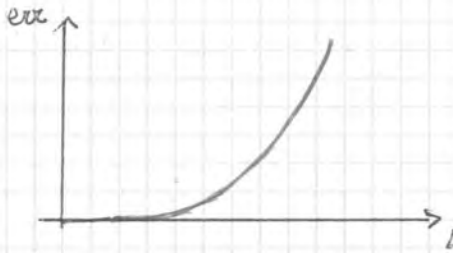
l'errore globale è

$$\|y(t) - u(t)\| \leq C e^{L(t-t_0)} (\Delta t)^p \left\| \frac{d^p f}{dt^p} \right\|_\infty$$

(p: accuratezza del metodo)

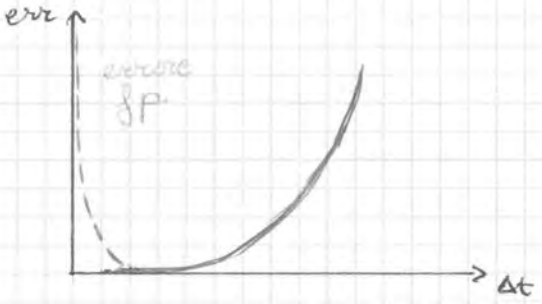
- $\|y(t) - u(t)\|$: errore al tempo $t = n\Delta t$
- L: cost. di Lipshitz
- $(\Delta t)^p$: $\to 0$ per $\Delta t \to \infty$
- $\left\| \frac{d^p f}{dt^p} \right\|_\infty$: regolarità della funz.

l'errore aumenta allontanandosi dal pt di partenza



andam. previsto dalla formula di stima dell'errore

Se però si considera anche l'approssimaz. f.p. si trova un grafico diverso:



se $\Delta t \to 0$ il n° dei passi di integraz. $\to \infty$
$\Rightarrow$ l'effetto dell'errore f.p. diventa importante

l'errore di origine fp dipende dal no di operaz. $\Rightarrow$ tende a ∞ per $n \to \infty$.

Per un metodo ad un passo la convergenza è la stessa cosa della consistenza ($d_{\Delta t} \to 0$ per $\Delta t \to 0$). Ma noi di solito facciamo i conti con Δt finito (così meno tempo e no errore fp). $\to$ Studio il comportamento dello schema per Δt finiti per un probl. modello particolarm. semplice (nn posso usare Taylor)

$$\begin{cases} y' = \lambda y \\ y(0) = y_0 \end{cases}$$

su probl più complicati posso cmq linearizzare (almeno localmente)

$$y(t) = y_0 e^{\lambda t}$$

→

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 65 di 314

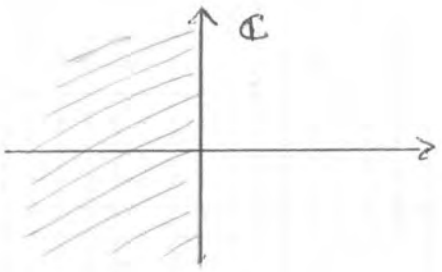
Metodo di Crank-Nicolson:

$$u^{n+1} = u^n + \frac{\Delta t}{2}\left(f(t^n, u^n) + f(t^{n+1}, u^{n+1})\right)$$

$$\Rightarrow u^{n+1} = u^n + \frac{\Delta t}{2}\left(\lambda u^n + \lambda u^{n+1}\right) \rightarrow u^{n+1} = \frac{1 + \frac{\Delta t}{2}\lambda}{1 - \frac{\Delta t}{2}\lambda} u^n$$

Ost metodo e' stabile per:

$$\left|\frac{1 + \frac{\Delta t}{2}\lambda}{1 - \frac{\Delta t}{2}\lambda}\right| = \left|\frac{1+z}{1-z}\right| \leq 1 \qquad z < \frac{\Delta t}{2}\lambda$$



↳ $Re(z) < 0$

⇒ Anche C.N. e' incondizionatam. stabile

⇒ Per sistemi differenziali con m equaz. $y' = Ay$ ⇒ A e' $m \times m$
Per stud. la stabilita' calcolo gli autoval. di A $\lambda_1, \dots \lambda_m$

$\Delta t \lambda_i$ e' regione di stabilita' $\forall i$

la condiz. diventa piu' restrittiva per gli autoval con modulo piu' grande → $|\lambda|_{max}$ (le componenti veloci del sistema decadono rapidam. a zero ma sono quelle che determinano la stabilita'),

Convergenza e stabilita'

Eulero esplicito:

$$\begin{cases} y' = \lambda y \\ y(0) = y_0 \end{cases} \rightarrow y(t) = y_0 e^{\lambda t}$$

$$u^n = (1 + \Delta t \lambda)^n u_0 \qquad t = n\Delta t$$

$$\rightarrow u^n = \left(1 + \frac{t\lambda}{n}\right)^n y_0 \rightarrow \underbrace{\lim_{n\to\infty} u^n}_{\text{condiz. di convergenza}} = \lim_{n\to\infty}\left(1 + \frac{t\lambda}{n}\right)^n y_0 = y_0 e^{\lambda t}$$

↑ il lim. va a finire sulla soluz. esatta

$\Rightarrow u^n \rightarrow y(t)$ per $\Delta t \rightarrow 0$

$|u^n|$ diventa illimitata se Δt non soddisfa la condiz. di stab.

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 67 di 314

Esempio 2 RK 2

tanti metodi, noi vediamo il metodo di Heun (Runge)

0	0	0
1	1	0
	1/2	1/2

$u^{(1)} = u^m$

$u^{(2)} = u^m + \Delta t\, a_{21}\, \lambda u^{(1)} = u^m + \Delta t\, \lambda u^{(1)}$

$u^{m+1} = u^m + \Delta t\left[b_1 \lambda u^{(1)} + b_2 \lambda u^{(2)}\right] = u^m + \frac{\Delta t}{2}\left[\lambda u^m + \lambda\left(u^m + \Delta t \lambda u^m\right)\right] =$

$= \left[1 + \Delta t\lambda + \frac{(\Delta t\lambda)^2}{2}\right] u^m$

sviluppo in serie di Taylor dell'esponenziale

La sol. esatta (dopo 1 passo) è $y^{m+1} = y^m e^{\lambda \Delta t}$. L'errore di troncam. ($y^m = u^m$):

$$d_{\Delta t} = \frac{1}{\Delta t}\left[y^m e^{\lambda \Delta t} - y^m\left(1 + \Delta t\lambda + \frac{(\Delta t\lambda)^2}{2}\right)\right] = \frac{1}{\Delta t}\left(O(\Delta t)^3\right) = O(\Delta t)^2$$

metodo del II ordine

Esempio 3 RK 3

0	0	0	0
1/3	1/3	0	0
2/3	0	2/3	0
	1/4	0	3/4

$c_i = \sum_{k=1}^{2} a_{ik}$

metodo esplicito perché a_{ik} è strettam. triang. inferiore.

$u^{(1)} = u^m$

$u^{(2)} = u^m + \Delta t\, \frac{1}{3}\, \lambda u^{(1)}$

$u^{(3)} = u^m + \Delta t\left(0 + \frac{2}{3}\lambda u^{(2)} + 0\right)$

$u^{m+1} = u^m + \Delta t\left(\frac{1}{4}\lambda u^{(1)} + 0 + \frac{3}{4}\lambda u^{(3)}\right) =$

$= u^m\left(1 + \Delta t\lambda + \frac{1}{2}(\Delta t\lambda)^2 + \frac{1}{6}(\Delta t\lambda)^3\right)$

se < 1 → metodo stabile!

Esempio 4 RK 4

0	0	0	0	0
1/2	1/2	0	0	0
1/2	0	1/2	0	0
1	0	0	1	0
	1/6	1/3	1/3	1/6

finire la cosa

Se uso un metodo di ordine $p+1$, $y(t) - u_{(p+1)}(t) = \tilde{C}(\Delta t)^{p+1} + O(\Delta t)^{p+2}$

u_{p+1} = soluz. numerica ottenuta col metodo di ordine $p+1$

u_p = " " " " " " p

Sottraggo le 2 equaz.:

$$u_{p+1}(t) - u_p(t) = C(\Delta t)^p + O(\Delta t)^{p+1} - \tilde{C}(\Delta t)^{p+1} - \underbrace{O(\Delta t)^{p+2}}_{\text{minuscolo}} \simeq$$

$$\simeq \underbrace{C(\Delta t)^p}_{\text{termine meno piccolo}} + \underbrace{O(\Delta t)^{p+1}}_{\text{tutti gli altri termini}}$$

$$\Rightarrow \boxed{u_{p+1}(t) - u_p(t) \simeq e_p(t)} \quad \text{stimatore d'errore}$$

Quindi ho ottenuto uno stimatore a posteriori dell'errore.

< se $|e_p(t)|$ piccolo → tengo Δt così com'è e vado avanti
< se $|e_p(t)|$ grande → diminuisco Δt (e ripeto il passo) e vado avanti

$\left(|e_p| < tol \cdot |u_p^n| \rightarrow |e_p| \text{ piccolo} \right)$

Ogni passo RK richiede di calcolare f 2 volte.
Uso 2 metodi di RK, uno di accuratezza p, l'altro di accuratezza $p+1$, che abbiamo la stessa matrice a.
Questi si chiamano EMBEDDED RK

c	a
	b → metodo di ordine p
	$\tilde{b}$ → metodo di ordine $p+1$

⇒ gli stage values sono gli stessi per i 2 metodi

$$u_p^{n+1} = u_p^n + \Delta t \sum_{i=1}^{\nu} b_i f(t_i + c\Delta t, u^{(i)}) \qquad e_p = U_p^{n+1} - y(t^{n+1}) = (\Delta t)^p + \dots$$

$$u_{p+1}^{n+1} = u_p^n + \Delta t \sum_{i=1}^{\nu} \tilde{b}_i f(t_i + c\Delta t, u^{(i)}) \qquad e_{p+1} = U_{p+1}^{n+1} - y(t^{n+1}) = \tilde{c}(\Delta t)^{p+1} + \dots$$

$$\Rightarrow u_{p+1}^{n+1} - u_p^{n+1} = \Delta t \sum_{i=1}^{\nu} (\tilde{b}_i - b_i) f(t_i + c\Delta t, u^{(i)}) = C(\Delta t)^p + \dots$$

Calcolo gli stage values e ottengo lo stimatore d'errore ad un costo (solo) leggerm. superiore al calcolo di u_{p+1} (soluz. più accurata)

ode45 → routine MatLab che usa un metodo RK Embedded con $p=4$
(sviluppata da Dortmund - Prince)
Dormand

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 71 di 314

Uso i valori di u^{n+j} per calcolare f_{n+j} e interpolo i dati f_{n+j} con un pol. di grado d

$\Rightarrow$ Ottengo $\int f(\,)\,ds$ con una precisione $O(\Delta t)^{\widetilde{d+1}+1}$ integrando si ottiene un ordine in più

Ho $r+1$ dati, ottengo un polinomio di grado r e dunque uno schema di ordine $r+2$ (metodo implicito)

Per il metodo esplicito ho r dati ($\beta_r = 0$), interpolo con un pol di grado $r-1$ $\Rightarrow$ metodo di ordine $r+1$.

A parità di passi $\rightarrow$ il metodo espl. ha un ordine in meno.

NB per calcolare u^{n+r} ho bisogno di una valutaz. di f per passo, dobbiamo memorizzare i precedenti r passi

Adams meglio di RK ma svantaggi:
- come si fa a partire? (mi servono tot dati)
- è più difficile adattare il passo di integrazione (se a un certo pt è troppo grande non posso più usare i dati di prima) perché il metodo è meno locale.

ode113 usa metodi di A-B (per l'ordine p) combinati con il corrispondente AM (per l'ordine p+1) per $p = 1, \dots, 13$
↳ funziona bene se soluz. prevedibile $\rightarrow$ fenomeno che cambia poco nel tempo, in qst caso è più veloce di ode45 a parità di accuratezza

Problemi stiff (stiff = rigido)

Sono probl. con scale temporali molto diverse

Esempio

$$\begin{cases} y' = \lambda(y - \cos t) - \sin t \\ y(0) = y_0 \end{cases}$$

Soluz. esatta $\rightarrow$ $y(t) = e^{\lambda(t-t_0)}(y_0 - \cos t_0) + \cos t$

- se $\lambda = 0$, $y(0) = 1$ $\rightarrow$ $y(t) = \cos t$ $\rightarrow$ grafico blu
- se $\lambda < 0$ $\rightarrow$ grafico rosso

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 73 di 314

10/11/11

Sistemi iperbolici

$$u_t + f_x(u) = 0 \qquad (1D)$$

- nel caso della gasdinamica:

$$u = \begin{pmatrix} \rho \\ \rho v \\ E \end{pmatrix}(x,t) \qquad f(u) = \begin{pmatrix} \rho v \\ \rho v^2 + p \\ (E+p)v \end{pmatrix}$$

- modelli del traffico
- magnetoidrodinamica
- matematica finanziaria

} campi di utilizzo dei sistemi iperbolici

Partiamo dalle cose più semplici!

Equazione scalare:

$$u_t + f_x(u) = 0 \qquad f = f(u) \text{ reale}$$

f dipende da u in maniera nota ed è una funz. regolare di u, però non è detto che u sia regolare a sua volta (es. densità ρ attraverso un urto)

Modello semplificato

$f(u) = au$ eq. lineare

$u_t + a u_x = 0 \qquad a = \text{cost.}$

Curva caratteristica:

considero una curva nel piano (x,t):

$$\gamma = \{(x(t), t)\}$$

$$u = u(x,t)$$

$$u|_\gamma = u(x(t), t)$$

Voglio calcolare come u varia lungo la curva γ → devo calcolare la derivata:

$$\left|\frac{du}{dt}\right|_\gamma = \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial t}\frac{dt}{dt} = \frac{\partial u}{\partial t} + \frac{dx}{dt}\frac{\partial u}{\partial x}$$

1° argomento

2° argomento

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 75 di 314

Se cambio il pt (x,t) individuo un'altra curva caratteristica con la stessa pendenza ($a = cost$) e dovrò cercare la corrispondente intersez. con l'asse dei valori iniziali

la mia eq. diff. definisce un fascio di linee caratteristiche, in qst caso sono tutte rette // (stessa pendenza) → il segnale si propaga senza cambiare forma

N.B. a è un rapporto tra uno spazio e un tempo → anche dimensionalm. è una velocità.

Sistemi iperbolici:

$$u_t + f'(u)\,u_x = 0$$

$f'(u)$ è lo Jacobiano di f

$$(f'(u))_{ij} = \frac{\partial f_i}{\partial u_j} \quad \text{(matrice)}$$

Gli autovalori di $f'(u)$ dipendono da u e danno le VELOCITÀ CARATTERISTICHE DEL SISTEMA.

Per la gasdinamica ho 3 autovalori

$$\lambda_1 = v - c \qquad \lambda_2 = v \qquad \lambda_3 = v + c$$

c = vel del suono $= \sqrt{\gamma \frac{p}{\rho}}$

v = vel dell'aria

In realtà voglio autovalori reali per avere un modello fisico (in realtà servono anche autoval. linearm. indipendenti)

$u_t + a u_x = 0$ → le caratteristiche sono rette // con pendenza a.

Esercizio

$$u_t + 2u_x = 0$$

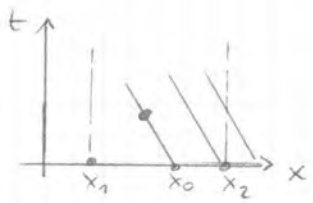
$$u(x, t=0) = u_0(x) = \begin{cases} 3 & x < 0 \\ -1 & x > 0 \end{cases}$$

$a = 2 \quad x - x_0 = 2t \quad \rightarrow x_0 = x - 2t$

$x_0 < 0 \rightarrow u_0(x_0) = 3 \quad \Rightarrow \quad u(x,t) = 3 \quad \text{per } x_0 < 0 \Rightarrow x - 2t < 0$

N.B. la soluzione ha bisogno di dati solo per $(x, t=0)$ e $(x = x_1, t)$!!!

se $a < 0$

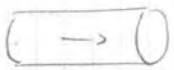


I dati devono essere assegnati

- $u(x, t=0) = u_0(x)$ $\quad x_1 < x < x_2$
- $u(x = x_2, t) = g(t)$ $\quad$ per $t > 0$

In gasdinamica ho 3 caratteristiche

$v - c$
v
$v + c$



$v > 0$

Se il regime è subsonico $(v - c < 0) \rightarrow v,\ v + c > 0$
devo prescrivere 2 dati a sx e 1 dato a dx

Se il regime è supersonico $(v - c > 0) \rightarrow$ devo prescrivere tutti i dati a sx perché $v - c,\ v,\ v + c > 0$

prob non lineari → la caratteristica dipende dalla soluzione → non posso assegnare i dati al bordo finché non trovo la soluzione.

Equazioni lineari del tipo $u_t + a(x) u_x = 0$.

(per es. $a(x) = \frac{1}{x}$)

Prob. ai valori iniziali $u(x, t=0) = u_0(x)$

Caratteristica:

$$\frac{dx}{dt} = a(x) \rightarrow \frac{dx}{dt} = \frac{1}{x}$$

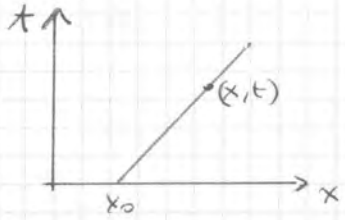
In generale le caratt. non sono più delle rette perché la pendenza a dipende da x. Nell'es. risolviamo per separazione di variabili:

$$\int_{x_0}^{x} x\,dx = \int_{t_0}^{t} dt \rightarrow \frac{x^2}{2} - \frac{x_0^2}{2} = t - t_0$$

Integro l'eq. per le caratt.

$$x - x_0 = a(u)(t - t_0)$$

Se ho un probl. ai valori iniziali



$$u(x, t=0) = u_0(x)$$
$$t_0 = 0$$

La pendenza delle caratt. dipende dalla soluzione.

La solut. sarà:

$$u(x,t) = u_0(x_0) \quad \text{su} \quad x - x_0 = a(u)\,t$$
$$\Rightarrow \quad x - x_0 = a\big(\underbrace{u_0(x_0)}\big)\,t \qquad (*)$$

piede della caratt. — pendenza della caratt.

Trovo x_0 risolvendo $(*)$ e poi calcolo $u(x,t) = u_0(x_0)$

$\Rightarrow$ non posso determ. le caratt. se non conosco la soluzione

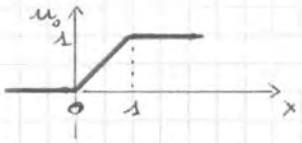
In generale, $(*)$ andrà risolta numericamente perché non è una relaz. semplice

Esempio

($(*)$ semplice)

$f(u) = \frac{1}{2}u^2$ eq di Burgers

$$u_0(x) = \begin{cases} 0 & x < 0 \\ x & 0 < x < 1 \\ 1 & x > 1 \end{cases}$$



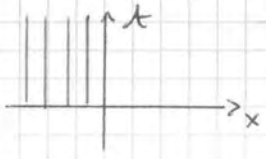
$f'(u) = u \qquad u_t + u u_x = 0$

Le caratt. sono $\frac{dx}{dt} = a(u) = u$

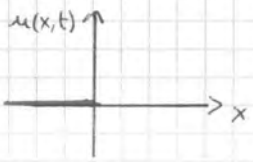
- Se $x_0 < 0 \rightarrow u_0(x_0) = 0$

$$x - x_0 = a\big(u_0(x_0)\big)\,t = u_0(x_0)\,t = 0 \quad \Rightarrow \quad x = x_0$$

quindi $u(x,t) = 0$ per $x < 0$



caratt. verticali



- Se $x_0<0 \Rightarrow u_0(x_0)=1$

  $u(x,t)=1$ per $x-x_0=t \rightarrow x_0=x-t \quad x_0<0 \rightarrow x<t$

  $u(x,t)=1$ per $x<t$

- Se $x_0>1 \Rightarrow u_0(x_0)=0$

  $u(x,t)=0$ per $x-x_0=0\cdot t \rightarrow x_0=x, \; x_0>1 \Rightarrow x>1$

  $u(x,t)=0$ per $x>1$

- Se $0<x_0<1 \Rightarrow u_0(x_0)=1-x_0$

  $x-x_0=(1-x_0)t \rightarrow x-t=(1-t)x_0 \rightarrow x_0=\frac{x-t}{1-t}$

  $u(x,t)=u_0(x_0)=1-x_0=1-\frac{x-t}{1-t}=\frac{1-x}{1-t}$ per $0<x_0<1$

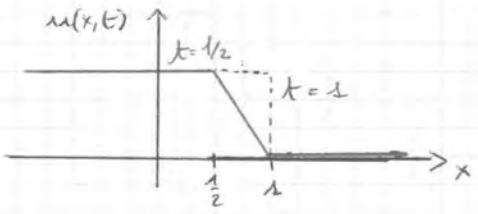
  $\Rightarrow 0<\frac{x-t}{1-t}<1 \rightarrow$ Almeno inizialmente $1-t>0 \quad (t<1)$

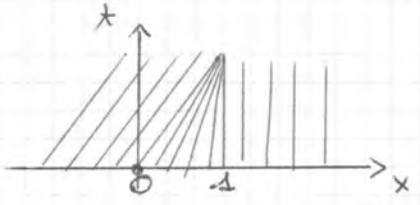
  $u(x,t)=\frac{1-x}{1-t}$ per $0<x-t<1-t \rightarrow t<x<1$

Globalmente:

$$u(x,t)=\begin{cases}1 & x<t\\ \frac{1-x}{1-t} & t<x<1\\ 0 & x>1\end{cases}$$

funz. continua per $t<1$





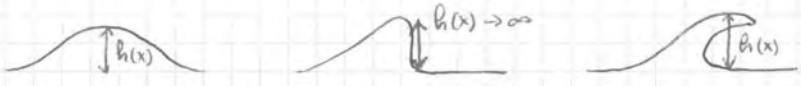
ONDA DI COMPRESSIONE

per $t\rightarrow 1^-$ $u_x\rightarrow-\infty$

$|u_x|$ cresce ↑

Per $t>1$ le caratt si incontrano e trasportano valori diversi di u nello stesso punto x. La soluzione diventa MULTIVALUED.

Se ora è il profilo di un'onda che si avvicina alla spiaggia:

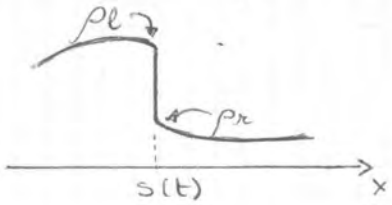


onda di compressione

l'onda si rompe
h(x) assume 3 valori diversi

Se però è la densità di un gas → non posso avere più di una soluzione!!!

Considero una sol. ρ che contiene un gradino ma è differenziabile fuori dal gradino.



Il gradino si trova in posizione $s(t)$ e si muove

$$\rho_l = \lim_{x \to s^-} \rho(x,t)$$

$$\rho_r = \lim_{x \to s^+} \rho(x,t)$$

Abbiamo che: $$\frac{d}{dt}\left[\int_a^{s(t)} \rho(x,t)\,dx + \int_{s(t)}^{b} \rho(x,t)\,dx\right] = f(a,t) - f(b,t)$$

Se f dipende solo da ρ $\Rightarrow$ $f(a,t) = f(\rho(a,t))$
$f(b,t) = f(\rho(b,t))$

Ricordare che:

1) $$\frac{d}{dt}\int_a^b u(x,t)\,dx = \int_a^b \partial_t u(x,t)\,dx$$

2) $$\frac{d}{dt}\int_a^b u(s)\,ds = u(t)$$

3) $$\frac{d}{dt}\int_a^{x(t)} u(s)\,ds = u(x(t))\,\frac{dx}{dt}$$

Quindi, per $\Delta t \to 0$ abbiamo:

$$\int_a^{s(t)} \partial_t \rho + \rho(s^-,t)\,\frac{ds}{dt} + \int_{s(t)}^{b} \partial_t \rho - \rho(s^+,t)\,\frac{ds}{dt} = f(\rho(a,t)) - f(\rho(b,t))$$

Faccio uno zoom sul gradino $\to$ spingo $a \to s^-$ e $b \to s^+$ (operazione di limite):

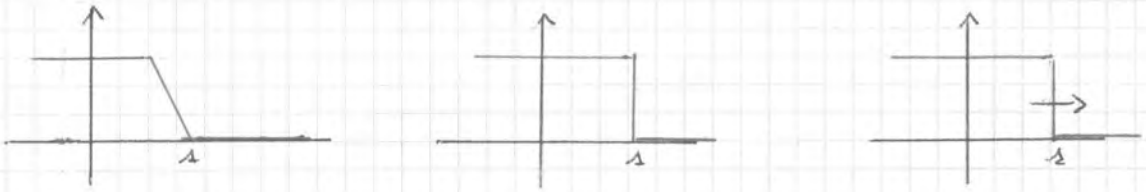
$$\lim_{a \to s^-} \int_a^s \partial_t \rho = 0 \quad , \quad \lim_{b \to s^+} \int_s^b \partial_t \rho = 0$$

$$\lim_{a \to s^-} f(\rho(a,t)) = f(\rho_l) \, , \quad \lim_{b \to s^+} f(\rho(b,t)) = f(\rho_r)$$

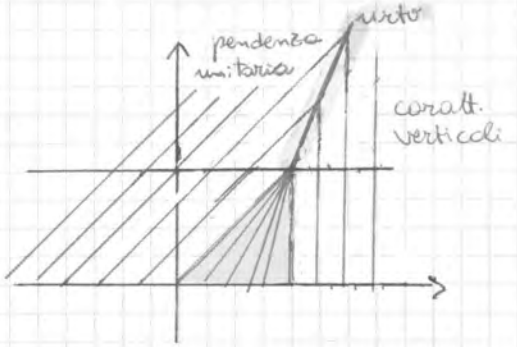
$$\Rightarrow \rho_l(t)\,s' - \rho_r(t)\,s' = f(\rho_l(t)) - f(\rho_r(t))$$

$$(\rho_l(t) - \rho_r(t))\,s' = f(\rho_l(t)) - f(\rho_r(t))$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 85 di 314



per $t<1$ per $t=1$ per $t>1$



Dopo la formaz. dell'onda d'urto, il fenomeno diventa irreversibile

Devo sapere da quale legge di conservazione sono partita (massa, qdm, energia):

$$u = \begin{cases} \rho & \text{massa} \\ \rho v & \text{qdm} \\ E & \text{energia} \end{cases} \Rightarrow \text{R.H.} \quad (\Delta u)s' = \Delta f$$

primi studi

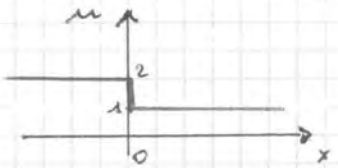
$$u = \begin{cases} \rho & \text{massa} \\ \rho v & \text{qdm} \\ \textcircled{S} & \text{entropia} \end{cases}$$

studi successivi

non si conserva → è errato per i conti con R.H. → $S_t + u S_x = 0$

Esempio

(1) $u(x,t) = \begin{cases} 2 & x<0 \\ 1 & x>0 \end{cases}$ $\qquad u_t + \left(\frac{1}{2}u^2\right)_x = 0$



$u(x,t) = \begin{cases} 2 & x < s't \\ 1 & x > s't \end{cases}$ è una soluzione

$s' = \frac{1}{2}(1) - \frac{1}{2}(1) = \frac{3}{2}$

A sx del gradino $u_t + u u_x = 0$
A dx del gradino $u_t + u u_x = 0$
→ l'eq. è soddisfatta su tutto $\mathbb{R}$

Modello del traffico (Whitham 1967)

ρ = densità di traffico $\left[\frac{n^\circ \text{ auto}}{km}\right]$

$0 < \rho < \rho_M$ $\quad \rho_M = 200$ per es.

$f(\rho) = \rho v(\rho)$

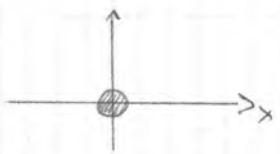
Nel modello più semplice $0 \leq v(\rho) \leq v_M$ $\quad v_M = 130$ km/h

$\begin{cases} v(0) = v_M & \text{qnd non ci è nessuno} \\ v(\rho_M) = 0 & \text{qnd sono tutti fermi in coda} \end{cases} \Rightarrow v(\rho) = v_M\left(1 - \frac{\rho}{\rho_M}\right)$

$f(\rho) = \rho v_M\left(1 - \frac{\rho}{\rho_M}\right)$

Legge di conservazione (delle auto) $\rightarrow$ $\boxed{\rho_t + f_x(\rho) = 0}$

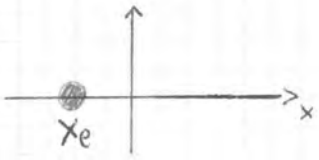
1) Semaforo rosso



$\rho(x,0) = \rho_e$

$\rho(0, t>0) = \rho_M$

2) Semaforo verde



$$\rho(x,0) = \begin{cases} \rho_e & x < x_e \\ \rho_M & x_e < x < 0 \\ 0 & x > 0 \end{cases}$$

15/11/11

Velocità dei segnali:

$f'(\rho) = v(\rho) + \rho v'(\rho) = v_M\left(1 - \frac{\rho}{\rho_M}\right) + \rho v_M\left(-\frac{1}{\rho_M}\right) = v_M\left(1 - 2\frac{\rho}{\rho_M}\right)$

$f'(\rho) = \begin{cases} v_M & \rho = 0 \\ -v_M & \rho = \rho_M \end{cases} \rightarrow -v_M \leq f'(\rho) \leq v_M$

Mentre le auto vanno solo avanti, i segnali vanno sia avanti che indietro.

Quali sono i segnali che io posso vedere?

La vel relativa dei segnali rispetto a me è

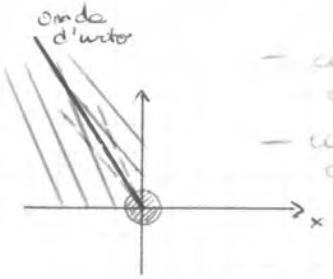
$f'(\rho) - v(\rho) = v_M\left(1 - 2\frac{\rho}{\rho_M}\right) - v_M\left(1 - \frac{\rho}{\rho_M}\right) = -v_M\frac{\rho}{\rho_M} < 0$

quindi io vedo i segnali che ci sono davanti a me

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 89 di 314

Semaforo rosso (in $x=0$)

$$\begin{cases} \rho_t + f_x(\rho) = 0 & \text{in } x<0 \\ \text{Probl con condiz. iniziali} \\ \rho(x, t=0) = \rho_e \\ \text{Condiz. al bordo} \\ \rho(x=0, t) = \rho_M \end{cases}$$



- caratt. inclinate di $-v_M$
- caratt. inclinate di $v_M\left(1-2\frac{\rho_e}{\rho_M}\right)$

in $x=0$: $f'(\rho) = f'(\rho_M) = -v_M < 0$ → caratt. inclinate a sx

in $t=0$: $f'(\rho) = f'(\rho_e) = v_M\left(1-2\frac{\rho_e}{\rho_M}\right) > -v_M$

Quindi le caratt. si incontrano subito → ho un'onda d'urto che si propaga all'indietro con velocità $s' = v_M\left(1-\frac{\rho_e+\rho_R}{\rho_M}\right)$ e ha una pendenza intermedia

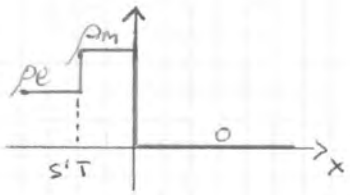
$$\Rightarrow \begin{cases} \rho_e & x < s't \\ \rho_R & x > s't \end{cases}$$ soluzione

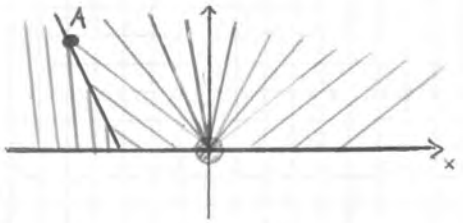
Semaforo verde

All'istante T il semaforo diventa verde

Probl. ai valori iniziali

$$\rho(x, t=T) = \rho_0(x) = \begin{cases} \rho_e & x < s'T \\ \rho_M & s'T < x < 0 \\ 0 & x > 0 \end{cases}$$





- caratt. inclinate di $-v_M$ e v_M
- fan di rarefazione
- caratt. inclinate di $-\rho_e$

} auto che rallentano perché la coda si muove poco alla volta

In (x_A, t_A) ho l'ultima auto ferma → fine della coda

In qst modello del traffico mancano alcune cose:

- multicorsie
- tempo di reazione (freno prima della coda)
- svincoli laterali
- camion → non hanno densità ρ_M perché sono più lunghi delle altre macchine ⇒ modelli multifase

la seconda onda d'urto parte da 3 mm da 0 !!!

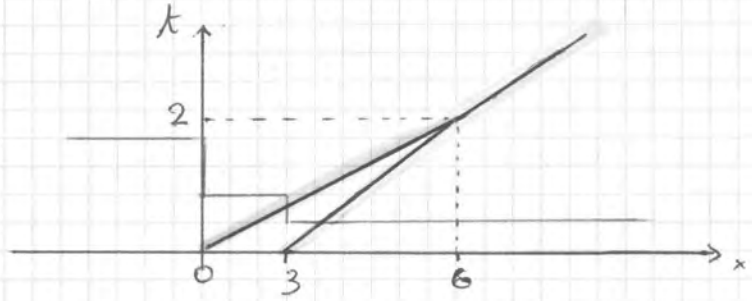
$$u(x,t) = \begin{cases} 4 & x < 3t \\ 2 & 3t < x < 1{,}5t + 3 \\ 1 & x > 1{,}5t + 3 \end{cases}$$

Questo però vale fino a quando le 2 onde d'urto si incontrano!! (collidono)
(e l'intervallo in mezzo sparisce)

le 2 onde d'urto collidono quando sono nello stesso posto:

$$3t = 1{,}5t + 3 \rightarrow \frac{3}{2}t = 3 \rightarrow t = 2 \quad \text{in } x = 6$$

Per $t > 2$ la solue. cambia



Ho un'unica onda d'urto che separa $u_e = 4$ e $u_r = 1$ con velocità $s'_3 = \frac{5}{2}$

$$u(x,t) = \begin{cases} 4 & x < x_5 + s'_3(t)(t - t_5) \\ 1 & x > x_5 + s'_3(t)(t - t_5) \end{cases}$$

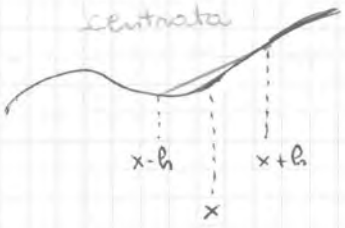
$$u(x,t) = \begin{cases} 4 & x < 6 + \frac{5}{2}(t-2) \\ 1 & x > 6 + \frac{5}{2}(t-2) \end{cases}$$

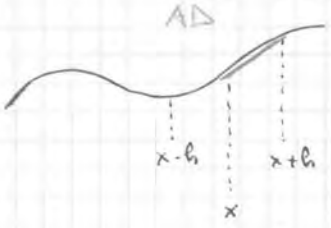
21/11/11

Metodi numerici per equazioni iperboliche lineari a coefficienti costanti e con delle condizioni iniziali

$$\begin{cases} u_t + a u_x = 0 & a = \text{cost} \\ u(x, t=0) = u_0(x) \end{cases}$$

solue. esatta $u(x,t) = u_0(x - at)$ →





Per costruire un metodo numerico trascuro i termini in $O(\Delta t)$ e $O(h)$:

$$\frac{U(x,t+\Delta t) - U(x,t)}{\Delta t} + a\,\frac{U(x+h,t) - U(x-h,t)}{2h} = 0$$

Calcolo la sol.ne numerica sui punti di griglia

$$\left[U_j^{n+1} - U_j^n\right]\frac{1}{\Delta t} + \frac{a}{2h}\left[U_{j+1}^n - U_{j-1}^n\right] = 0$$

$$\Rightarrow \boxed{U_j^{n+1} = U_j^n - \frac{a}{2}\cdot\frac{\Delta t}{h}\left[U_{j+1}^n - U_{j-1}^n\right]}$$

metodo centrato (e^- instabile → non viene usato)

Discretizzo il dato iniziale:

$U_j^0 = u_0(x_j)$ → da qui calcolo U_j^1 $\forall j$, etc etc

N.B. notare la somiglianza con ODE (e^- come se avessi un'equaz. per ogni pt di griglia):

$$\underline{U}^n = \left[\dots, U_{j-1}^n, U_j^n, U_{j+1}^n, \dots\right]^T \leftarrow \text{vettore colonna}$$

Risolvo un sistema di ODE del tipo $\frac{dU}{dt} = F(U)$

I ordine

- AD: $U_j^{n+1} = U_j^n - a\frac{\Delta t}{h}\left(U_{j+1}^n - U_j^n\right)$
- AS: $U_j^{n+1} = U_j^n - a\frac{\Delta t}{h}\left(U_j^n - U_{j-1}^n\right)$
- Metodo di Lax-Friedrichs: $U_j^{n+1} = \frac{1}{2}\left(U_{j+1}^n + U_{j-1}^n\right) - a\frac{\Delta t}{2h}\left(U_{j+1}^n - U_{j-1}^n\right)$

II ordine

- Metodo di Lax-Wendroff: $U_j^{n+1} = U_j^n - \frac{a\Delta t}{2h}\left(U_{j+1}^n - U_{j-1}^n\right) + \frac{1}{2}\left(\frac{a\Delta t}{h}\right)^2\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right)$

Tutti gli ultimi metodi dipendono da $\lambda = \frac{\Delta t}{h}$ (inverso di una velocità)

Sono tutti metodi <u>espliciti</u>

AD, AS sono asimmetrici → ci sono delle direzioni privilegiate, ma ciò è abbastanza logico, infatti la sol.ne esatta è costruita usando le linee caratteristiche..

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 95 di 314

posso distinguere (a) da (b) facendo i conti (algebra) e vedendo se $|a|\Delta t \gtrless h$.

Quindi la cond. CFL indica come scegliere il parametro $\lambda \leq \frac{1}{|a|}$ cioè vale per i metodi centrati. Per gli asimmetrici il discorso è ben diverso.

metodi centrati CFL : $\lambda \leq \frac{1}{|a|}$

metodi upwind CFL : $\lambda \leq \frac{1}{|a|}$ < uso AS per $a>0$ / uso AD per $a<0$

CFL è una condizione necessaria di stabilità:

- se la CFL non è soddisfatta $\Rightarrow$ metodo instabile
- se la CFL è soddisfatta $\Rightarrow$ il metodo può essere stabile

(es. metodo centrato soddisfa CFL per $\lambda \leq \frac{1}{|a|}$ ma è cmq instabile)

Per un sist. iperbolico (come quello della gasdinamica), CFL: $\Delta t \leq \frac{h}{\max(|f'(u)|)}$
dove con $\max(|f'(u)|)$ intendo il max autovalore dello Jf
In gasdin. ho 3 autoval. e quindi $\max(|f'(u)|) = |v| + c|$.

Torniamo all'operatore H che dipende da 3 parametri.
Nei metodi a 3 punti (che sono i metodi più semplici) H agisce su una parte del vettore U^n per costruire la soluz. nel pt di griglia U^{n+1}

$$U^{n+1} = H_h(U^n)$$

Quindi H agisce sul vettore U^n e dà il vettore U^{n+1}. Dal pt di vista matematico:

$$H_h : \mathbb{R}^m \to \mathbb{R}^m \qquad m = n^\circ \text{ dei pti di griglia } x_j$$

Tutti gli schemi numerici visti finora hanno coeff. cost. che non dipendono dalla soluz. $U^n \Rightarrow H_h$ è lineare, cioè è una matrice $\mathcal{H}_h$

$$U^{n+1} = \mathcal{H}_h U^n$$

Per L.F. la riga j di $\mathcal{H}_h$ è $[0, \dots 0, \frac{1}{2}(1+a\lambda)$ o $\frac{1}{2}(1-a\lambda), 0, \dots 0]$
$\rightarrow$ abbiamo 2 sole diagonali $\neq 0$

Posso anche considerare H come operatore lineare che agisce su funzioni
↳

Diremo che il metodo è <u>consistente</u> (in norma p) se:

$$\lim_{\Delta t \to 0} \| d_h(\ , t) \|_p = 0 \quad \forall t \quad \text{per } \lambda \le \lambda_0 \text{ fissato}$$

λ_0 calcolato con la condiz. di stabilità (per Upwind $\lambda_0 = \frac{1}{|a|}$)

$t = n\Delta t$

Poiché λ è fissato se $\Delta t \to 0$ anche $h \to 0$

Per esempio, Upwind è consistente

<u>Accuratezza</u>

Un metodo è accurato di ordine q se

$$\lim_{\Delta t \to 0} \frac{\| d_h \|_p}{\Delta t^q} = \text{cost.}$$

→ il metodo Upwind è accurato di ordine 1

<u>Esercizio</u>

Dimostrare che il metodo di Lax - Wendroff è accurato di ordine 2

$$d_h(x,t) = \frac{1}{\Delta t}\left\{u(x,t+\Delta t) - H_h(u(\ ,t))\right\} = \frac{1}{\Delta t}\left\{u(x,t+\Delta t) - \left(u(x,t) + \right.\right.$$
$$\left.\left. - \frac{a\lambda}{2}\left[u(x+h,t) - u(x-h,t)\right] + \frac{1}{2}a^2\lambda^2\left[u(x+h,t) - 2u(x,t) + u(x-h,t)\right]\right)\right\} =$$
$$= \frac{1}{\Delta t}\left\{ \cancel{u} + \Delta t\, u_t + \frac{1}{2}(\Delta t)^2 u_{tt} - \cancel{u} + \frac{a\lambda}{2}\left[\cancel{u} + h u_x + \frac{1}{2}h^2 u_{xx} - \cancel{u} + h u_x - \frac{1}{2}h^2 u_{xx}\right] + \right.$$
$$\left. - \frac{1}{2}a^2\lambda^2\left[\cancel{u} + \cancel{h u_x} + \frac{1}{2}h^2 u_{xx} - \cancel{2u} + \cancel{u} - \cancel{h u_x} + \frac{1}{2}h^2 u_{xx}\right] + O(\Delta t)^3 + O(h^3)\right\} =$$
$$= u_t + \frac{1}{2}\Delta t\, u_{tt} + \frac{a\lambda}{\Delta t} h u_x - \frac{1}{2}\frac{a^2\lambda^2}{\Delta t} h^2 u_{xx} + O(h^2) = \qquad \lambda = \frac{\Delta t}{h} \to h = \frac{\Delta t}{\lambda}$$
$$= \cancel{u_t} + \frac{1}{2}\Delta t\, u_{tt} + \cancel{a u_x} - \frac{1}{2}a^2 \Delta t\, u_{xx} + O(h^2) \qquad u_t = -a u_x$$
$$d_h = \frac{1}{2}\Delta t\left(a^2 u_{xx} - a^2 u_{xx}\right) + O(h^2)$$
$$d_h = O(h^2)$$

$$u_{ttt} = a^2 \partial_t u_{xx} = a^2 \partial_t \partial_x u_x = a^2 \partial_x \partial_t u_x = a^2 \partial_x \partial_x u_t =$$
$$= a^2 \partial_x \partial_x (-a u_x) = -a^3 \partial_x u_{xx} = -a^3 u_{xxx}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 99 di 314

Se uso LF :

$$u_{new}(ii) = \frac{1}{2}\left(u_{old}(ii+1) + u_{old}(ii-1)\right) - \frac{\lambda a}{2}\left(u_{old}(ii+1) - u_{old}(ii-1)\right)$$

Però c'è un problema !

Quando io sono ai bordi questo mi chiede 2 celle in più (una a sx e una a dx) che io non ho → dovrei conoscere la soluzione anche in $1-1$ e $m+1$, cioè in $x_1 - h$ e $x_m + h$.

=> Si creano 2 celle extra, dette celle ghost

celle ghost

Devo trovare un modo per definire la sol. anche nelle celle ghost.

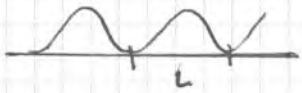
Ho bisogno di condizioni al bordo.

Le condizioni al bordo dipendono dalle caratteristiche. Per $u_t + a u_x = 0$ posso imporre la soluzione in $x_1 - h$ se $a > 0$, in $x_m + h$ se $a < 0$

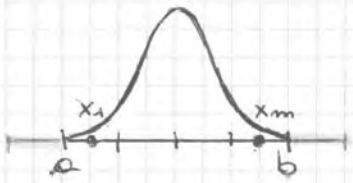
Però in alcuni casi posso fissare le condiz. al bordo indipendentem. dal segno di a :

1) condizioni di periodicità → la sol si ripete con un periodo L

$$u(x,t) = u(x+L,t)$$



Scelgo (a,b) t.c. $b - a = L$



$$u(x_1 - h) = u(x_m)$$
$$u(x_m + h) = u(x_1)$$

Il vettore delle ascisse è : $xx = a - \frac{h}{2} : h : b + \frac{h}{2}$

e contiene anche le 2 celle ghost.

$xx(1)$ è la prima delle ghost

$xx(m+2)$ è la cella ghost di dx

→ $ii = 2 : m+1$ sono gli indici su cui calcolo la sol. (celle interne)

↳

Infatti, per quell' equaz., $d_h = O(\Delta t)^2$

Da qui concludo che la sol. numerica ottenuta con upwind assomiglierà di più alle sol. dell' eq. modificata che non alle sol. dell' eq. di partenza $u_t + a u_x = 0$

⇒ Voglio vedere come sono fatte le sol. dell' eq. modificata rispetto alle sol. dell'eq. di partenza → uso gli sviluppi di Fourier.

Supponiamo che $u_0(x)$ sia periodica (2π per es.)

$$\Rightarrow u(x,t) = \sum_{k=-\infty}^{+\infty} e^{ikx} \underbrace{b_k(t)}_{\text{coeff. dipendenti dal tempo}}, \quad b_k(0) = \frac{1}{2\pi}\int_0^{2\pi} e^{-ikx} u_0(x)$$

Considero il singolo modo di Fourier $u_k(x,t) = b_k(t)\, e^{ikx}$ e lo sostituisco nell' eq. esatta:

$$\partial_t u_k + a\,\partial_x u_k = 0 \quad \forall k$$

$$b_k' e^{ikx} + a\, b_k\, ik\, e^{ikx} = 0 \rightarrow \text{eq. diff. lineare che dipende soltanto dal tempo}$$

$$b_k'(t) = -a\, ik\, b_k(t) \rightarrow b_k(t) = b_k(0) e^{-aikt}$$

⇒ u_k è sol. dell' eq. di partenza se b_k è fatto così

$$u_k(x,t) = b_k(0) e^{ikx} e^{-aikt} = b_k(0) e^{ik(x-at)}$$

per avere l'evol. della sol. pono mettere qst risultato nell' espressione $u(x,t) = \sum \dots$

Ogni numero d' onda k viaggia con velocità a e mantiene la stessa ampiezza nel tempo

Per l' eq. modificata:

$$\partial_t u_k + a\,\partial_x u_k = \nu \Delta t\, \partial_{xx}^2 u_k$$

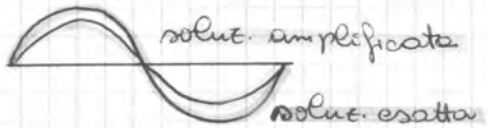
$$b_k' + aik\, b_k = \nu \Delta t\, (ik)^2 b_k \Rightarrow b_k' = (-aik - \nu \Delta t k^2)\, b_k(t)$$

$$\Rightarrow b_k(t) = e^{-aikt} e^{-\nu \Delta t k^2 t} b_k(0)$$

Dunque: (modificazione introdotta dallo schema (*))

$$u_k(t) = \underbrace{e^{ik(x-at)}}_{\text{soluzione esatta}} \; e^{-\nu \Delta t k^2 t} \; b_k(0)$$

Che forma hanno gli effetti spuri qnd il metodo è stabile?
Immagino $u_0(x) = \sin x$



la soluz. numerica è smussata, via via che aumenta il tempo aumenta la smussatura } opt è tipico dei metodi del 1° ordine

I modi che fanno di più assomigliare la soluz. al grading vanno a zero con la viscosità

Qual è l'effetto spurio tipico di un metodo del 2° ordine?

L'errore di troncamento è del tipo:

$$\alpha_h = (\Delta t)^2 \mu u_{xxx} + O(\Delta t)^3$$

L'eq. modificata è $u_t + a u_x = \mu (\Delta t)^2 u_{xxx}$

Sostituendo u_k nell'eq. modificata per un metodo del II ordine

$$b_k' + a i k b_k = \mu (\Delta t)^2 (ik)^3 b_k$$

$$b_k' = (-a i k - i \mu (\Delta t)^2 k^3) b_k$$

$$\rightarrow b_k(t) = b_k(0)\, e^{-aikt}\, e^{-i\mu(\Delta t)^2 k^3 t}$$

quindi: $u_k(t) = b_k(0)\, e^{i(kx - akt - \mu(\Delta t)^2 k^3 t)}$

chiamo $\tilde{a} = a + \mu(\Delta t)^2 k^2 \Rightarrow u_k(t) = b_k(0) e^{ik(x - \tilde{a} t)}$

la perturbazione è in velocità perché u_k si muove con vel. $\tilde{a}$ invece che con vel a.

Se però ho dei modi di Fourier elevati $\tilde{a}$ può diventare grande. Inoltre la correzione in velocità $\tilde{a} - a$ dipende

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 105 di 314

Consideriamo ora la norma $p=1$:

$$\|U^{n+1}\| = \left\| \frac{1}{2}\left[U^n(x+h) + U^n(x-h)\right] - \frac{\lambda a}{2}\left[U^n(x+h) - U^n(x-h)\right] \right\| =$$

$$= \left\| \frac{1}{2}(1-\lambda a)\, U^n(x+h) + \frac{1}{2}(1+\lambda a)\, U^n(x-h) \right\|$$

Vale la disuguaglianza triangolare, quindi:

$$\|U^{n+1}\| \leq \left\| \frac{1}{2}(1-\lambda a)\, U^n(x+h) \right\| + \left\| \frac{1}{2}(1+\lambda a)\, U^n(x-h) \right\|$$

$$\|xx\| = |x| \cdot \|x\| \rightarrow \|U^{n+1}\| \leq \frac{1}{2}|1-\lambda a| \|U^n(x+h)\| + \frac{1}{2}|1+\lambda a| \cdot \|U^n(x-h)\|$$

$$\|U^{n+1}\| \leq \frac{1}{2}\left[|1-\lambda a| + |1+\lambda a|\right] \cdot \|U^n\|$$

se $|\lambda a| < 1 \quad \left(\lambda < \frac{1}{|a|}\right) \Rightarrow 1-\lambda a > 0$ e $1+\lambda a > 0$

posso togliere i moduli $\Rightarrow \|U^{n+1}\| \leq \frac{1}{2}\left[1-\lambda a + 1+\lambda a\right] \cdot \|U^n\|$

quindi $\|U^{n+1}\| \leq \|U^n\|$ se $\boxed{\lambda \leq \frac{1}{|a|}}$ ← condizione sufficiente di stabilità

24/11/11

Sistemi iperbolici lineari

$\underline{u}_t + A\underline{u}_x = 0$ A matrice $(m \times m)$

Il sistema è iperbolico se A ha m autoval reali $\lambda_1, \dots \lambda_m$ ed m autovett. l.i. $\underline{r}_1, \dots \underline{r}_m$

Allora $R^{-1} A R = \Lambda = \text{diag}(\lambda_1, \dots \lambda_m)$ (matrice diagonale)

$R = [\underline{r}_1, \underline{r}_2, \dots \underline{r}_m]$. Poiché gli $\underline{r}_i$ sono l.i. $\Rightarrow \det R \neq 0$ ed $\exists\, R^{-1}$.

Supponiamo che A sia costante $\Rightarrow \Lambda, R, R^{-1}$ sono costanti.

Moltiplichiamo ora il sist. per R^{-1} $\Rightarrow$ ci sarà un cambiam. di coordinate:

$$R^{-1}\partial_t u + R^{-1} A \partial_x u = 0 \rightarrow R^{-1}\partial_t u + \underbrace{R^{-1} A R}_{\text{II}} R^{-1} \partial_x u = 0$$

Chiamo $\underline{v} = R^{-1}\underline{u}$ cambiam. di coord.

$\underline{v}(x,t) = R^{-1}\underline{u}(x,t) \rightarrow \partial_t v = \partial_t R^{-1} u = R^{-1}\partial_t u$

$\partial_x v = R^{-1}\partial_x u$

Quindi trovo $\rightarrow \partial_t \underline{v} + \Lambda \partial_x \underline{v} = 0$

$$R = \underset{\downarrow\ \downarrow}{\begin{matrix} r_1 & r_2 \end{matrix}} \begin{pmatrix} 2 & 2 \\ 1 & -1 \end{pmatrix} \rightarrow R^{-1} = \frac{1}{4}\begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix} \rightarrow R^{-1}AR = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} = \Lambda$$

- Calcolo $R^{-1}u_0$

Suppongo di avere come condiz. iniziale $\underline{u}_0(x) = \begin{cases} \underline{u}_L = \begin{pmatrix} 2 \\ 1 \end{pmatrix} & x<0 \\ \underline{u}_R = \begin{pmatrix} -1 \\ 3 \end{pmatrix} & x>0 \end{cases}$

$$\Rightarrow \underline{w}_0(x) = \begin{cases} R^{-1}\underline{u}_L & x<0 \\ R^{-1}\underline{u}_R & x>0 \end{cases}$$

$$R^{-1}\underline{u}_L = \frac{1}{4}\begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$R^{-1}\underline{u}_R = \frac{1}{4}\begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} -1 \\ 3 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 5 \\ -4 \end{pmatrix}$$

$$\rightarrow \underline{w}_0(x) = \begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix} & x<0 \\ \frac{1}{4}\begin{pmatrix} 5 \\ -4 \end{pmatrix} & x>0 \end{cases}$$

- Componente 1:

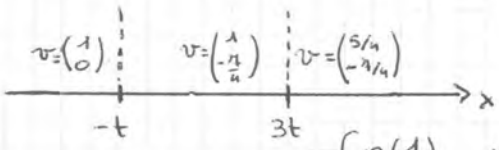
$$v_1|_{t=0} = \begin{cases} 1 & x<0 \\ 5/4 & x>0 \end{cases} \quad \text{viaggia con vel } \lambda_1 = 3$$

$$\Rightarrow v_1(x,t) = \begin{cases} 1 & x<3t \\ 5/4 & x>3t \end{cases}$$

- Componente 2:

$$v_2|_{t=0} = \begin{cases} 0 & x<0 \\ -4/4 & x>0 \end{cases} \quad \text{viaggia con vel } \lambda_2 = -1$$

$$\Rightarrow v_2(x,t) = \begin{cases} 0 & x<-t \\ -4/4 & x>-t \end{cases}$$

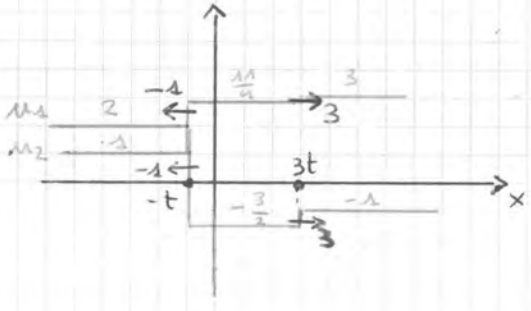


Ogni componente ha un salto solo

$$u(x,t) = Rv(x,t) = \begin{cases} R\begin{pmatrix} 1 \\ 0 \end{pmatrix} & x<-t \\ R\begin{pmatrix} 1 \\ -4/4 \end{pmatrix} & -t<x<3t \\ R\begin{pmatrix} 5/4 \\ -4/4 \end{pmatrix} & x>3t \end{cases} \qquad R = \begin{pmatrix} 2 & 2 \\ 1 & -1 \end{pmatrix}$$

$$u(x,t) = \begin{cases} \begin{pmatrix} 2 \\ 1 \end{pmatrix} & x<-t \\ \frac{1}{4}\begin{pmatrix} -6 \\ -14 \end{pmatrix} & -t<x<3t \\ \begin{pmatrix} -1 \\ 3 \end{pmatrix} & x>3t \end{cases}$$

Le componenti hanno salti sempre



Il gradino iniziale si apre in 2 pezzi e poi non cambia più. Nel mondo delle caratt., i segnali viaggiano uno per volta

$U^n \to V^n = R^{-1} U^n$

$v_\ell^{n+1} = v_\ell^n - \frac{\Delta t}{h} \lambda_\ell (v_\ell^n - v_{\ell-1}^n)$ per $\lambda_\ell > 0$

$v_\ell^{n+1} = v_\ell^n - \frac{\Delta t}{h} \lambda_\ell (v_{\ell+1}^n - v_\ell^n)$ per $\lambda_\ell < 0$

Ho ottenuto $V^{n+1} \Rightarrow U^{n+1} = R V^{n+1}$

Metodo centrale (Lax - Friedrichs)

è simmetrico rispetto alle direz. di propagaz. $\Rightarrow$ non devo passare per le caratteristiche

$$U_j^{n+1} = \frac{1}{2}\left(U_{j+1}^n + U_{j-1}^n\right) - \frac{\lambda A}{2}\left(U_{j+1}^n - U_{j-1}^n\right)$$

Metodo upwind "migliorato"

$$\Lambda = \Lambda^+ + \Lambda^- = \begin{pmatrix} \lambda_1 & & \\ & \ddots \lambda_k & \\ & & 0 \end{pmatrix} + \begin{pmatrix} 0 & & \\ & \lambda_{k+1} & \\ & & \ddots \lambda_m \end{pmatrix}$$

Ricordo $R^{-1} A R = \Lambda \to A = R \Lambda R^{-1} = R(\Lambda^+) R^{-1} + R(\Lambda^-) R^{-1}$

$\Rightarrow A = A^+ + A^-$, dove $A^+ = R(\Lambda^+) R^{-1}$, $A^- = R(\Lambda^-) R^{-1}$

$$U_j^{n+1} = U_j^n - \lambda A^+ \left(U_j^n - U_{j-1}^n\right) - \lambda A^- \left(U_{j+1}^n - U_j^n\right)$$

$\to$ è più semplice da scrivere, ma è più lungo (applico upwind 2 volte per A^- e A^+)

Anche così devo calcolare Λ, R, R^{-1} $\to$ devo passare per le caratteristiche

$\Rightarrow$ metodi UPW sono più complessi, soprattutto se il probl non è lineare.

Problema di Riemann

è un probl. ai valori iniziali con $\underline{u}_0(x) = \begin{cases} \underline{u}_L & x < 0 \\ \underline{u}_R & x > 0 \end{cases}$

Considero UPW scalare ($a > 0$)

$$u_j^{n+1} = u_j^n - \lambda a \left(u_j^n - u_{j-1}^n\right) \qquad \lambda = \frac{1}{a}$$

trovo che $u_j^{n+1} = u_j^n + u_{j-1}^n - u_j^n = u_{j-1}^n$ è la soluz. esatta

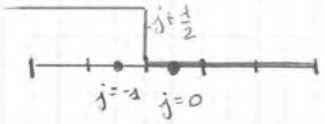
Nelle stesse ipotesi, LF diventa:

$$U_j^{n+1} = \frac{1}{2}\left(U_{j+1}^n + U_{j-1}^n\right) - \frac{\lambda a}{2}\left(U_{j+1}^n - U_{j-1}^n\right) \qquad \lambda = \frac{1}{a}$$

trovo che $U_j^{n+1} = U_{j-1}^n$ è di nuovo la soluz. esatta

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 111 di 314

Nella cella j il flusso entrante è $\lambda U_j U_{j-1}$, quello uscente è $\lambda (U_j)^2$.



Per $j=0$ il flusso entrante è $\lambda U_0 U_{-1} = 0$
Per $j=-1$ il flusso uscente è $\lambda (U_{j=-1})^2 = \lambda$

Morale$_2$: dalla cella $j=-1$ esce della massa che non entra nella cella $j=0$ → lo schema scritto prima non conserva la massa.
Se solut. a gradino → problema grave ☠
Quello che entra in una cella dev'essere uscito dalla cella precedente. → il flusso di massa tra una cella e l'altra deve essere basato sull'interfaccia tra le 2 celle.

L'interfaccia è in una posizione intermedia → $j+\frac{1}{2}$

Metodi conservativi

def: un metodo è conservativo se può essere scritto come:

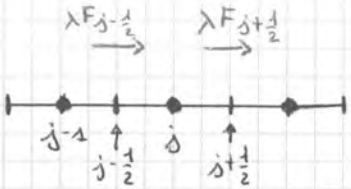
$$U_j^{n+1} = U_j^n - \lambda \left(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} \right)$$

F = funz. di flusso numerico t.c.

1) $F_{j+1/2} = F(U_j, U_{j+1})$
2) F dev'essere "regolare" nei suoi argomenti (Lipshiziana)
3) F dev'essere consistente: $F(U,U) = f(U)$

Nel nostro esempio $f'(U_j) \geq 0$
Con il metodo upwind devo prendere l'informaz. da sx:

$$F_{j+\frac{1}{2}} = F(U_j, U_{j+1}) = f(U_j)$$



ciò evita che si formino pozzi o sorgenti nelle interfacce

Nel nostro esempio: $U_j^0 = \begin{cases} 1 & j<0 \\ 0 & j \geq 0 \end{cases} \rightarrow f'(U_j^0) \geq 0 \quad \forall j$

→

$$U_j^{m+1} = U_j^m - \frac{\lambda}{2}\left[f(U_{j+1}) - f(U_{j-1})\right] + \underbrace{\frac{\alpha\lambda}{2}\left(U_{j+1} - 2U_j + U_{j-1}\right)}_{\text{termine di diffusione numerica che stabilizza il metodo centrato}}$$

Trovo LF per $\alpha = \frac{1}{\lambda}$

Il metodo è stabile per $\lambda \leq \frac{1}{\max|f'(u)|}$ e per $\max|f'(u)| \leq \alpha \leq \frac{1}{\lambda}$

N.B. upt è il metodo conservativo indipendentem. dal segno di f'.

Metodi ai volumi finiti

Parto da $u_t + f_x(u) = 0$

Introduco la media di cella $\bar{u}_j(t) = \frac{1}{h}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x,t)\,dx$

la qntità di massa nella cella j è $M_j = \bar{u}_j h$

Integro l'eq. esatta sulla cella j e divido per h:

$I_j = \left(x_j - \frac{h}{2},\ x_j + \frac{h}{2}\right)$ ↘

$$\frac{1}{h}\int_{I_j} \partial_t u(x,t)\,dx + \frac{1}{h}\int_{I_j} \partial_x f(u(x,t))\,dx = 0$$

$$\partial_t \underbrace{\left[\frac{1}{h}\int_{I_j} u(x,t)\,dx\right]}_{\bar{u}_j(t)} + \frac{1}{h}\left[f\left(u\left(x_{j+\frac{1}{2}},t\right)\right) - f\left(u\left(x_{j-\frac{1}{2}},t\right)\right)\right] = 0$$

$$\Rightarrow \partial_t \bar{u}_j(t) = -\frac{1}{h}\left[f\left(u\left(x_{j+\frac{1}{2}},t\right)\right) - f\left(u\left(x_{j-\frac{1}{2}},t\right)\right)\right] \quad \text{relazione esatta } \forall j$$

Integro nel tempo fra $(t^m,\ t^m + \Delta t)$:

$$\int_{t^m}^{t^m+\Delta t} \partial_t \bar{u}_j(t)\,dt = -\frac{1}{h}\left[\int_{t^m}^{t^m+\Delta t} f\left(u\left(x_{j+\frac{1}{2}},t\right)\right)dt - \int_{t^m}^{t^m+\Delta t} f\left(u\left(x_{j-\frac{1}{2}},t\right)\right)dt\right]$$

$$\bar{u}_j(t^m + \Delta t) - \bar{u}_j(t^m) = -\frac{1}{h}\left[\quad\right]$$

Formulaz. ai valori finiti:

$$\bar{u}_j^{m+1} = u_j^m - \frac{1}{h}\left[\int_{t^m}^{t^m+\Delta t} f\left(u\left(x_{j+\frac{1}{2}},t\right)\right)dt - \int_{t^m}^{t^m+\Delta t} f\left(u\left(x_{j-\frac{1}{2}},t\right)\right)dt\right]$$

N.B. equaz. esatta

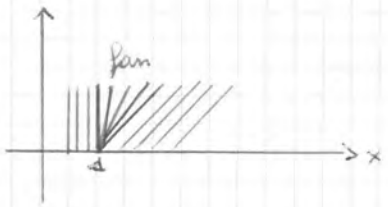
Metodo numerico ai valori finiti:

$$\bar{U}_j^{m+1} = \bar{U}_j^m - \lambda\left[F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}\right] \quad \text{con } F_{j+\frac{1}{2}} \approx \frac{1}{\Delta t}\int_{t^m}^{t^m+\Delta t} f\left(u\left(x_{j+\frac{1}{2}},t\right)\right)dt$$

media nel tempo del flusso all'interfaccia

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 115 di 314

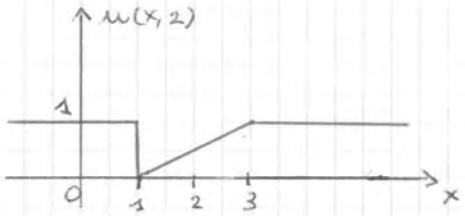
Onda di rarefazione



fan centrato in $x=1$

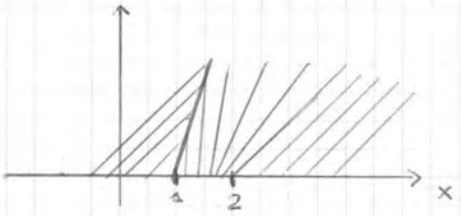
caratt. con pendenza $\frac{x-1}{t} = f'(u) = u$

Im $t=2$



sol. t. dell'istante $t=2$

Per $t>2$



$u_l = 1$

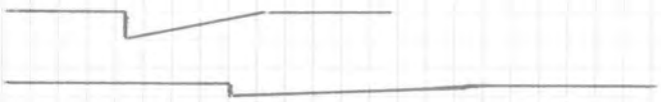
u_r parte da 0 e aumenta

s^1 inizialm. $= \frac{1}{2}$, poi in $t>2$ accelera

fino a $s^1 = 1$

l'onda d'urto accelera mangiando la coda dell'onda di rarefazione, ma non ne raggiunge mai la testa

Im un istante successivo a $t=2$ avrei:



urto sempre più basso con la testa del fan sempre più piatta

Tema d'esame del 12/02/2008 – Es 1

$$u_t + \left(\frac{1}{3}u^3\right)_x = 0 \quad \text{per } x \in \mathbb{R},\ t>0$$

$$u(x,t=0) = u_0(x) = \begin{cases} 2 & x \le 0 \\ 2\sqrt{1-x} & 0<x<1 \\ 0 & x \ge 1 \end{cases}$$

$$s^1 = \frac{\frac{1}{3}u_l^3 - \frac{1}{3}u_r^3}{u_l - u_r} = \frac{1}{3}\left(u_l^2 + u_l u_r + u_r^2\right)$$

* $u_l, u_r > 0$

$f'(u_l) > s^1 > f'(u_r) \rightarrow u_l^2 > \frac{1}{3}\left(u_l^2 + u_l u_r + u_r^2\right) > u_r^2$ (*)

• se $u_l > u_r \rightarrow u_l^2 > u_r^2$, $u_l^2 > u_l u_r \Rightarrow$ la condiz. è soddisfatta

↳

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 117 di 314

Per $0<x_0<1$: $x-x_0 = f'(u_0(x_0))t$

$u(x,t) = u_0(x_0) \rightarrow x-x_0 = u_0(x_0)t \quad x-x_0 = 4(1-x_0)t$

$x_0 = \frac{x-4t}{1-4t} \quad u(x,t) = 2\sqrt{1-x_0} = 2\sqrt{1-\frac{x-4t}{1-4t}} = 2\sqrt{\frac{1-x}{1-4t}}$

qst vale per $1-4t \neq 0$ per t t.c. $1-4t>0$ cioè $t<\frac{1}{4}$

Tema d'esame del 9/07/2008 - es. 2

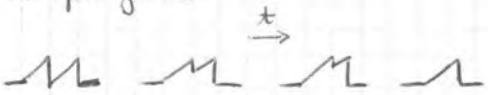
$u_t + \left(\frac{1}{2}u^2\right)_x = 0 \rightarrow$ Burgers

$$u_A(x) = \begin{cases} 0 & x \leq -1 \\ x+1 & -1 < x < 0 \\ x & 0 < x < 1 \\ 0 & x \geq 1 \end{cases} \qquad u_B(x) = \begin{cases} 0 & x \leq -1 \\ -x & -1 < x < 0 \\ 1-x & 0 < x < 1 \\ 0 & x \geq 1 \end{cases}$$

onde d'urto in $t=0$

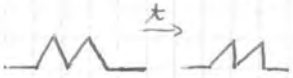
$s'(t=0) = \frac{1}{2}$

(per entrambe le onde), poi l'onda di dx ha $s'(t)=\frac{1}{2}$, quella di sx accelera → le 2 onde d'urto si prendono in un tempo finito



NO onde d'urto in $t=0$

ci sono 2 onde di compressione che si trasformeranno in onde d'urto in un tempo finito



u_B si trasforma in u_A

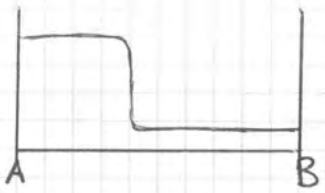
È possibile che $s'=4$?

$s' = \frac{1}{2}(u_l + u_r) \Rightarrow \min(u) \leq s' \leq \max(u) \Rightarrow 0 \leq s' \leq 1$

$\Rightarrow$ non è possibile che ci sia un'onda d'urto con $s'=4$

29/11/11

Conservazione discreta



Vorrei conservare la massa nel tubo

$$M(t) = \int_A^B u(x,t)\,dx, \quad M(t+\Delta t) = \int_A^B u(x,t+\Delta t)\,dx$$

$$\Rightarrow M(t+\Delta t) = M(t) + f(u_A)\Delta t - f(u_B)\Delta t$$

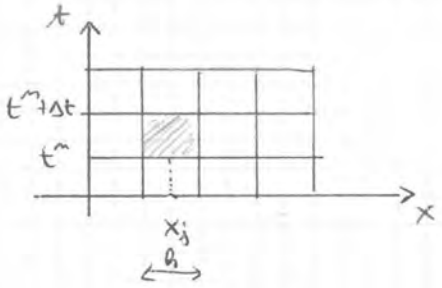
→ deriva le Rankine-Hugoniot

Metodo di Godunov

metodo ai volumi finiti:

$$\bar{U}_j^{n+1} = \bar{U}_j^n - \lambda \left(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} \right) \quad (*)$$

$$F_{j+\frac{1}{2}} \simeq \frac{1}{\Delta t} \int_{t^n}^{t^n + \Delta t} f\left(u(x_{j+\frac{1}{2}}, t)\right) dt$$



$$V_j^n = \left(x_j - \frac{h}{2}, x_j + \frac{h}{2}\right) \times \left[t^n, t^n + \Delta t\right]$$

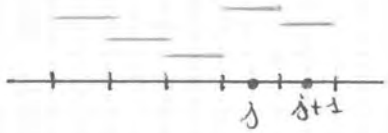
Abbiamo ottenuto (*) integrando la PDE su V_j^n

$$\int_{V_j^n} \left[u_t + f_x(u)\right] dx dt = 0$$

Come passo dalle medie di cella $\bar{U}_j^n$ ai valori nelle interfacce?

1) ricostruzione delle $\left\{ \bar{U}_j^n \right\}$

$$U^n(x) = \bar{U}_j^n \quad x \in I_j = \left(x_j - \frac{h}{2}, x_j + \frac{h}{2}\right)$$ costante a tratti



Ottengo 2 stime $\neq$ per $U^n(x_{j+\frac{1}{2}})$

2) Evoluzione $U^n(x)$ per 1 passo di integrazione

3) Calcolare le nuove $\bar{U}_j^{n+1}$

Ad ogni interfaccia devo risolvere un problema del tipo

$$u^*_{j+\frac{1}{2}}(0) = \begin{cases} \bar{U}_j^n & x < x_{j+\frac{1}{2}} \\ \bar{U}_{j+1}^n & x > x_{j+\frac{1}{2}} \end{cases}$$ dato iniziale } prob. di Riemann

Voglio trovare $u^*(t)$ → come va avanti nel tempo

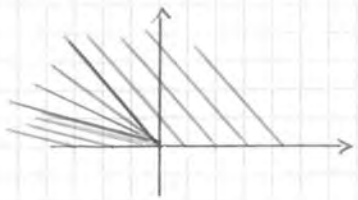
Se conosco $u^*(t)$ ⇒ $F_{j+\frac{1}{2}} = \frac{1}{\Delta t} \int_{t^n}^{t^n + \Delta t} f\left(u^*_{j+\frac{1}{2}}(t)\right) dt$

Per calcolare u^* ho bisogno che tutti i probl. di Riemann restino separati (→ probl a gradino → ok) ⇒ che non interferiscano fra di loro

Al tempo iniziale sono tutti disaccoppiati → anche nei primi istanti essi resteranno tali (vel. di propagazione finite) ⇒ voglio usare un passo temporale abbastanza piccolo da non avere interferenza dei segnali ⇒ la distanza percorsa dal segnale più veloce ($\max|f'(u)|$) deve essere $\leq h$ ⇒ $\Delta t \max|f'(u)| \leq h$ CFL ⇒ $\lambda \leq \frac{1}{\max|f'(u)|}$.

(1147)

caso 2

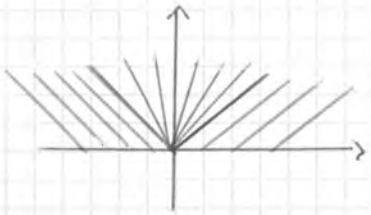


$f(U_j) < f(U_{j+1}) < 0$

se $s' > 0 \quad U^*_{j+\frac{1}{2}} = \bar{U}^n_{j+1}$

Nei primi 2 casi il risultato col metodo non entropico sarebbe stato lo stesso, nel caso 3 no:

caso 3



$u^*_{j+\frac{1}{2}}$ dipende dall'onda di rarefazione

$u^*_{j+\frac{1}{2}}$ t.c. $f'(u^*_{j+\frac{1}{2}}) = 0$

Metodo di Godunov con entropy fix:

1) calcolo Godunov normale:

$$F_{j+\frac{1}{2}} = \begin{cases} f(\bar{U}^n_j) & \text{se } s'_{j+\frac{1}{2}} \geq 0 \\ f(\bar{U}^n_{j+1}) & \text{se } s'_{j+\frac{1}{2}} < 0 \end{cases}$$

2) aggiungo l'entropy fix:

cerco le interfacce $j+\frac{1}{2}$ dove $f'(U_{j+1}) > 0$ e $f'(U_j) < 0$

$\Rightarrow$ trovo U^* t.c. $f'(u^*) = 0$

Calcolo $F_{j+\frac{1}{2}} = f(u^*_{j+\frac{1}{2}})$

Per la gasdinamica le caratteristiche sono $\lambda(f') = \begin{pmatrix} v-c \\ v \\ v+c \end{pmatrix}$

v = vel, c = vel del suono

In che casi posso avere onde d'urto del tipo  (caso 3) ?

and ho onde di rarefaz. che oltrepassano il pt sonico

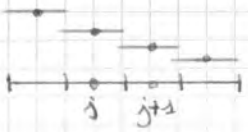
TRANSONIC RAREFACTIONS $\rightarrow$ in qst caso c'è bisogno dell'entropy fix.

Si può rendere Godunov più accurato?

Godunov:

1) ricostruzione
2) evoluzione dei P.R.
3) calcolo nuove medie di cella

1) $\{\bar{U}^n_j\}$



Interpolaz. cost. a tratti $\Rightarrow U^n(x)$

$\Rightarrow$ metodo del I ordine

Per la soluz. esatta:

$$u_t + f_x(u) = 0$$

- per una soluz. regolare $TV(u)$ è cost. nel tempo (uso metodo delle caratteristiche → ho gli stessi dati che avevo all'inizio)
- se onde d'urto → fenomeni irreversibili → ci può essere una perdita di dati trasportati dalle caratt. che finiscono sull'urto nn vengono creati nuovi dati → la variaz. tot. della soluz. esatta decresce nel tempo
- se soluz. numerica oscillante, con oscillazioni spurie, si creano dei valori che nn erano presenti nel dato iniziale → la TV aumenta

Voglio metod che garantiscano TV nn aumentante TVNI (Total Variation Non Increasing) opp. TVD (Total Variation Decreasing)

$$TV(U^{n+1}) \leq TV(U^n)$$

Tutti metodi numerici che abbiamo visto (del primo ordine) sono tutti TVD

Nella scelta del limitatore ϕ devo fare in modo di ottenere uno schema TVD. Abbiamo bisogno di un misuratore di regolarità:

$$\theta_{j+\frac{1}{2}} = \frac{U_j - U_{j-1}}{U_{j+1} - U_j} \simeq \begin{cases} 1 & \text{se } |U_j - U_{j-1}| \simeq |U_{j+1} - U_j| \text{ soluz. regolare} \\ 0 & \text{se } |U_j - U_{j-1}| \ll |U_{j+1} - U_j| \text{ sol nn regolare} \\ \frac{1}{h} & \text{se } |U_j - U_{j-1}| \gg |U_{j+1} - U_j| \quad " \quad " \quad " \\ <0 & \text{se c'è un estremo locale (che potrebbe essere un' oscillazione)} \end{cases}$$

$$\phi(\theta_{j+\frac{1}{2}}) \simeq \begin{cases} 1 & \text{se } \theta_{j+\frac{1}{2}} \simeq 1 \\ 0 & \text{se } \theta_{j+\frac{1}{2}} < 0 \text{ (per togliere eventuali oscillazioni)} \\ ? & \end{cases}$$

Se faccio i conti per avere TV decrescente, si dimostra:

$$0 \leq \phi(\theta) \leq 2 \quad \text{e} \quad 0 \leq \frac{\phi(\theta)}{\theta} \leq 2 \Rightarrow \text{metodo TVD}$$

Un metodo è del II ordine se $\phi(1) = 1$

→

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 125 di 314

Slope limiting

(Stile Godunov)

- ricostruz.
- evoluz.
- calcolo della nuova media

Per aumentare l'accuratezza uso una ricostruz. lineare a tratti

$$U(x) = a_j + b_j (x - x_j) \qquad x \in x_j - \frac{h}{2},\ x_j + \frac{h}{2}$$

(b_j: pendenza)

Conservo la massa:

$$\int_{I_j} U(x) = \bar{U}_j h = a_j h \Rightarrow a_j = \bar{U}_j$$

Per b_j ho diverse scelte:

$$b_j = \frac{1}{h}\left(\bar{U}_{j+1} - \bar{U}_j\right)$$

$$b_j = \frac{1}{h}\left(\bar{U}_j - \bar{U}_{j-1}\right)$$

$$b_j = \frac{1}{2h}\left(\bar{U}_{j+1} - \bar{U}_{j-1}\right)$$

[Usando la prima → evoluz. $u_t + a u_x = 0$]

Conosco $U^n(x)$ (che ho ottenuto con l'algoritmo di ricostruz.)

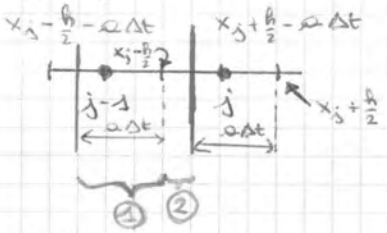
quindi $U^{n+1}(x) = U^n(x - a\Delta t)$

Calcolo delle nuove medie:

$$\bar{U}_j^{n+1} = \frac{1}{h}\int_{x_j - \frac{h}{2}}^{x_j + \frac{h}{2}} U^{n+1}(x)\,dx = \frac{1}{h}\int_{x_j - \frac{h}{2}}^{x_j + \frac{h}{2}} U^n(x - a\Delta t)\,dx$$

chiamo $y = x - a\Delta t \Rightarrow \bar{U}_j^{n+1} = \frac{1}{h}\int_{x_j - \frac{h}{2} - a\Delta t}^{x_j + \frac{h}{2} - a\Delta t} U^n(y)\,dy$

suppongo $a > 0$



si può spezzare in 2 parti:

$$\bar{U}_j^{n+1} = \frac{1}{h}\int_{x_j - \frac{h}{2} - a\Delta t}^{x_j - \frac{h}{2}} \left[a_{j-1} + b_{j-1}(y - x_{j-1})\right] dy + \quad ①$$

$$+ \frac{1}{h}\int_{x_j - \frac{h}{2}}^{x_j + \frac{h}{2} - a\Delta t} \left[a_j + b_j (y - x_j)\right] dy \quad ②$$

$$F_{j+\frac{1}{2}} = \frac{1}{\Delta t}\int_{\Delta t} f\left(u^*_{j+\frac{1}{2}}(t)\right)dt$$

approssimo (per es. regola dei trapezi):

$$\simeq \frac{1}{2}\left[\underbrace{f\left(u^k_{j+\frac{1}{2}}(\Delta t)\right)}_{\text{più complicato}} + \underbrace{f\left(u^k_{j+\frac{1}{2}}(0)\right)}_{\text{sol. Riemann}}\right]$$

- Ben Artzi
- Van Leer
- Colella

non lineare

Oppure: ('90 Shu - Osher)

$$u_t + f_x(u) = 0$$

integro sulle celle $I_j = \left(x_j - \frac{h}{2}, x_j + \frac{h}{2}\right)$ e divido per h → avevo ottenuto:

$$\frac{1}{h}\int_{I_j} \partial_t u = -\frac{1}{h}\left[f\left(u\left(x_{j+\frac{1}{2}},t\right)\right) - f\left(u\left(x_{j-\frac{1}{2}},t\right)\right)\right]$$

$$\underbrace{\partial_t \bar{u}_j(t) = -\frac{1}{h}\left[f\left(u\left(x_{j+\frac{1}{2}},t\right)\right) - f\left(u\left(x_{j-\frac{1}{2}},t\right)\right)\right]}_{\text{formulazione semidiscreta}}$$

eq. esatta che ha solo derivate nel tempo

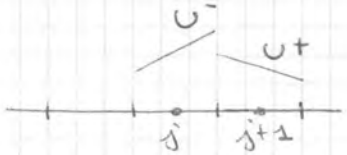
Metodo numerico:

$$\underbrace{\frac{d\bar{U}_j(t)}{dt} = -\frac{1}{h}\left(F_{j+\frac{1}{2}}(t) - F_{j-\frac{1}{2}}(t)\right)}_{\text{formulaz. semidiscreta conservativa}}$$

la formulaz. flusso numerico è:

$$F_{j+\frac{1}{2}} = F\left(U^+_{j+\frac{1}{2}}(t),\ U^-_{j+\frac{1}{2}}(t)\right)$$

boundary extrapolated data



$$U^-_{j+\frac{1}{2}}(t) = \lim_{x\to x^-_{j+\frac{1}{2}}} U^m(x) = \lim_{x\to x^-_{j+\frac{1}{2}}}\left[\bar{U}^m_j + b_j(x - x_j)\right] \Rightarrow U^-_{j+\frac{1}{2}}(t) = \bar{U}^m_j + b_j\frac{h}{2}$$

$$U^+_{j+\frac{1}{2}}(t) = \lim_{x\to x^+_{j+\frac{1}{2}}} U^m(x) = \lim_{x\to x^+_{j+\frac{1}{2}}}\left[\bar{U}^m_{j+1} + b_{j+1}(x - x_{j+1})\right] \Rightarrow U^+_{j+\frac{1}{2}}(t) = \bar{U}^m_{j+1} - b_{j+1}\frac{h}{2}$$

Ad un generico istante t:

$$U^-_{j+\frac{1}{2}}(t) = \bar{U}_j(t) + \frac{h}{2}b_j(t)$$

$$U^+_{j+\frac{1}{2}}(t) = \bar{U}_{j+1}(t) - \frac{h}{2}b_{j+1}(t)$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 129 di 314

3) Chiamo G con i dati $\{\bar{U}^{(1)}\}$
Ottengo $G(\bar{U}^{(1)})$

4) Calcolo $\bar{U}_j^{(2)} = \bar{U}_j^n + \Delta t\, G_j(\bar{U}^{(1)})$

5) Applico le condiz. al bordo

6) Chiamo G con i dati $\{\bar{U}^{(2)}\}$
Ottengo $G(\bar{U}^{(2)})$

7) Assemblo la soluzione $\bar{U}_j^{n+1} = \bar{U}_j^n + \frac{\Delta t}{2}\left(G_j(\bar{U}^{(1)}) + G_j(\bar{U}^{(2)})\right)$

Il grosso vantaggio di qst metodi è che non ci chiedono di risolvere i probl. di Riemann (al max Godunov o Riemann di ordine basso). La parte delicata è il calcolo dei valori estrapolati al bordo, per il quale ho bisogno di algoritmi di ricostruzione che non introducano oscillazioni spurie → devo usare i limitatori sulle pendenze (se uso una ricostruz. lineare a tratti).

Per ordini più elevati si ottengono sistemi ENO / WENO } Shu Jcase '97

Tema d'esame 28/01/2009 — es. 2

$$F_{j+\frac12} = \frac{a^+_{j+\frac12} f(U_j^n) - a^-_{j+\frac12} f(U_{j+1}^n)}{a^+_{j+\frac12} - a^-_{j+\frac12}} + \frac{a^+_{j+\frac12}\, a^-_{j+\frac12}}{a^+_{j+\frac12} - a^-_{j+\frac12}} (U_{j+1}^n - U_j^n)$$

$a^+ = \max(f'(U_j^n), f'(U_{j+1}^n), 0)$
$a^- = \min(\quad " \quad " \quad " \quad)$

a) $u_t + 2u_x = 0$ $\quad f(u) = 2u \quad f'(u) = 2$
b) $u_t - 2u_x = 0$ $\quad f(u) = -2u \quad f'(u) = -2$
c) Burgers $u_0(x) > 0$ $\quad f(u) = \frac12 u^2 \quad f'(u) = u$

caso a)
$a^+ = \max(2,2,0) = 2$
$a^- = \min(2,2,0) = 0$
$\Rightarrow F_{j+\frac12} = \frac{a^+ f(U_j^n) - 0}{a^+ - 0} = f(U_j^n)$ (Upwind)

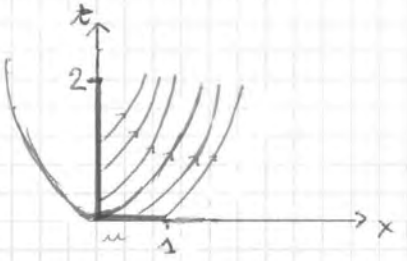
caso b)
$a^+ = \max(-2,-2,0) = 0$
$a^- = \min(-2,-2,0) = -2$
$\Rightarrow F_{j+\frac12} = \frac{0 - a^- f(U_{j+1}^n)}{0 - a^-} = f(U_{j+1}^n)$ (Upw)

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 131 di 314

Tema d'esame 15/06/2009 - es. 2

$$u_t + \frac{1}{x} u_x = 0$$

Probl 1) $u = \begin{cases} 1 & t=0 \quad 0<x<1 \\ 2 & t>0 \quad x=0 \end{cases}$



$u_t + a(x) u_x = 0$

$\frac{dx}{dt} = a(x) = \frac{1}{x} > 0$ per $x>0$

Probl 1 ben posto perché le caratt. sono ~~entranti~~ uscenti

$\frac{dx}{dt} = \frac{1}{x} \quad \int x\,dx = \int dt$

$\frac{x^2}{2} - \frac{x_0^2}{2} = t - t_0$

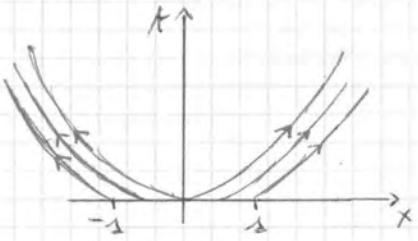
Per $x_0=0$ e $t_0=0$ $\quad t = \frac{x^2}{2} \quad \rightarrow u(x,t) = \begin{cases} 2 & \text{sopra} \rightarrow t \geq \frac{x^2}{2} \\ 1 & \text{sotto} \rightarrow t < \frac{x^2}{2} \end{cases}$

Probl 2) $u = \begin{cases} 1 & t=0 \quad 0<x<1 \\ 2 & t>0 \quad x=1 \end{cases}$

↓

mal posto!

- $t=0 \quad -1 \leq x \leq 1$



quali sono gli x,t t.c. posso calcolare la soluz. $u(x,t)$?

non ho dati lungo l'asse y → posso calcolare la soluz. solo per gli (x,t) sotto ~~x~~ /

$t < \frac{1}{2} x^2$, in più deve essere sopra la caratt. che parte da -1 e 1 (oltre non ho dati)

⇒ sopra : $\begin{cases} \frac{x^2}{2} - \frac{1}{2} = t & x_0 = 1 \\ \frac{x^2}{2} - \frac{1}{2} = t & x_0 = -1 \end{cases}$

d) Metodo Upwind

Oppure:

$$F_{j+\frac{1}{2}} = \frac{1}{2}\left[f(U_j)+f(U_{j+1})\right] - \frac{1}{2}\int_{U_j}^{U_{j+1}} |w|\,dw$$

$$\frac{1}{2}\int_{U_j}^{U_{j+1}} |f'(w)|\,dw \qquad f'(u)<0 \quad U\le w\le 0$$

$$f'(u)>0 \quad 0<w<U_{j+1}$$

$$= \frac{1}{2}\int_{U_j}^{0} -f'(w)\,dw + \frac{1}{2}\int_{0}^{U_{j+1}} f'(w)\,dw$$

Tema d'esame ~~22~~/06/2010 - es 2

1) $$\alpha_h(t) = \frac{1}{\Delta t}\left[u(x,t+\Delta t) - H_{\Delta t}(u;t)\right]$$

$$H_{\Delta t} = u(x,t) - \lambda a\left[\frac{1}{2}u(x-2h,t) - 2u(x-h,t)\right] + \frac{1}{2}\lambda^2 a^3\left[u(x-2h,t) + \right.$$

$$\left. - 2u(x-h,t)\right] + c\,u(x,t)$$

Sviluppo in (x,t) fino al secondo ordine

$$u(x,t+\Delta t) = u + \Delta t\,u_t + \frac{1}{2}(\Delta t)^2 u_{tt} + O(\Delta t)^3$$

$$u(x-2h,t) = u - 2h u_x + \frac{1}{2}(-2h)^2 u_{xx} + O(h^3)$$

$$u(x-h,t) = u - h u_x + \frac{1}{2}(-h)^2 u_{xx} + O(h^3)$$

$$\alpha_h(t) = \frac{1}{\Delta t}\left\{u + \Delta t\,u_t + \frac{1}{2}(\Delta t)^2 u_{tt} - u + \lambda a\left[\frac{1}{2}\left(u - 2h u_x + 2h^2 u_{xx}\right) + \right.\right.$$

$$\left. - 2\left(u - h u_x + \frac{1}{2}h^2 u_{xx}\right)\right] - \frac{1}{2}\lambda^2 a^2\left[u - 2h u_x + 2h^2 u_{xx} - 2\left(u - h u_x + \right.\right.$$

$$\left.\left. + \frac{1}{2}h^2 u_{xx}\right)\right] - cu + O(h^3) + O(\Delta t)^3\Big\}$$

$$\alpha_h(t) = \frac{1}{\Delta t}\left\{\Delta t\,u_t + \frac{1}{2}(\Delta t)^2 u_{tt} + u\left(\frac{\lambda a}{2} - 2\lambda a + \frac{1}{2}\lambda^2 a^2 - c\right) + \right.$$

$$\left. + u_x(\lambda a h) + u_{xx}\left(-\frac{1}{2}\lambda^2 a^2 h^2\right) + O(h^3) + O(\Delta t)^3\right\}$$

$$\alpha_h(t) = \frac{1}{\Delta t}\left\{\Delta t\,u_t + \Delta t\,a\,u_x + u\left(-\frac{3}{2}\lambda a + \frac{1}{2}\lambda^2 a^2 - c\right) - \frac{1}{2}(\Delta t)^2 a^2 u_{xx} + \right.$$

$$\left. + \frac{1}{2}(\Delta t^2 u_{tt}) + O(h^3) + O(\Delta t)^3\right\}$$

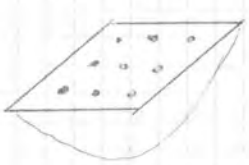
per un metodo del I ordine dovrei avere:

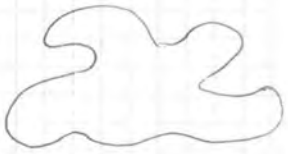
$$\alpha_h = \Delta t \cdot C u_{xx} + O(\Delta t)^2$$

per un metodo del II ordine:

$$\alpha_h = (\Delta t)^2 C u_{xxx} + O(\Delta t)^3$$

Problema della membrana elastica





una griglia uniforme in qst caso funziona male → nn si usano le differenze finite ma gli elementi finiti (FEM)

Ogni punto di griglia mi dà un'equazione, le equaz. sono accoppiate tra loro → ho un sistema lineare.

Nel probl. esatto:

$$\nu u''(x) = f(x) \qquad \forall x$$ Formulazione Forte (FF)

(i punti sono infiniti)

Nel FEM impongo un'equaz. rispetto a determinate funz. peso

Il probl esatto può essere scritto nella forma:

$$L(u) = f \quad \forall x \qquad L = \text{operatore } (L = -\nu \partial^2_{xx} \text{ nel caso del filo elastico}) \qquad (FF)$$

Nel FEM si usa:

$$(L(u) \cdot v) = (f, v) \quad \forall v \in V$$ Formulazione Variazionale (FV)

V = spazio delle funz. test = {tutte le possibili deformaz. del filo}

In molti casi la FV può essere scritta in forma di minimo. Costruisco un funzionale F $\quad F(v) \in \mathbb{R}$

Trovare u:

$$F(u) \leq F(v) \quad \forall v \in V$$

V → mondo di tutte le conformazioni possibili, però una volta che scelgo un carico, il filo assume una sola conformaz., quella che minimizza l'energia.

Nel caso del filo elastico: $F(v)$ = energia del filo = en. elastica + en. potenziale → $F(v) = \frac{1}{2}\nu(v', v') - (f, v)$

$$(v', v') = \int_A^B v' \cdot v' \qquad \text{en. elastica}$$

$$(f, v) = \int_A^B f \cdot v \qquad \text{en. potenziale}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 137 di 314

Invece FM $\Leftrightarrow$ FV , FM = formulaz. di minimo

dim (no all'esame)

So che $a(u',v') = (f,v) \quad \forall v \in V$

Vorrei dim che $F(u) \leq F(v) \quad \forall v \in V$

Scelgo $v \in V$ e calcolo $w = v - u \Rightarrow v = u + w$

e inoltre $w \in V$ perché V è uno spazio lineare

$$F(v) = \frac{1}{2} a(v',v') - (f,v) = \frac{1}{2} a(u'+w', u'+w') - (f, u+w) =$$

$$= \underbrace{\frac{1}{2} a(u',u') + \frac{1}{2} a(u',w') + \frac{1}{2} a(w',u')}_{=} + \frac{1}{2} a(w',w') \underbrace{- (f,u) - (f,w)}_{F(u)}$$

$$F(v) = F(u) + \frac{1}{2} a(w',w') + \underbrace{a(u',w') - (f,w)}_{=0 \text{ perché } w \in V}$$

N.B. $\frac{1}{2} a(w',w') = \frac{1}{2} a \int_A^B (w')^2 > 0 \Rightarrow F(v) > F(u)$ □

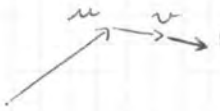
Per qnt riguarda il viceversa:

dim (no all'esame)

Se $F(u) \leq F(v) \quad \forall v \in V \Rightarrow a(u',v') = (f,v) \quad \forall v \in V$

Abbiamo u t.c. $F(u) \leq F(v) \quad \forall v \in V$

Considero una particolare funz. test $w = u + \varepsilon v$, $\varepsilon \in \mathbb{R}$, $v \in V$



$\Rightarrow w \in V$

perché è lineare

Fisso v e lascio variare ε. Ottengo una funz. $g(\varepsilon)$

$g(\varepsilon) = F(u + \varepsilon v) = \frac{1}{2} a(u' + \varepsilon v', u' + \varepsilon v') - (f, u + \varepsilon v) =$

↑ funz. di variabile reale

$= \frac{1}{2} a(u',u') + a\varepsilon(u',v') + \frac{1}{2} a \varepsilon^2 (v',v') + - (f,u) - \varepsilon (f,v)$

$g(\varepsilon)$ ha un minimo in $\varepsilon = 0$

$g'(\varepsilon) = a(u',v') + a\varepsilon(v',v') - (f,v)$

$g'(0) = a(u',v') - (f,v)$

Sappiamo che in $g(0)$ c'è un minimo $\rightarrow g'(0) = 0 \Rightarrow a(u',v') = (f,v)$

$\forall v \in V$

□

Matrice di rigidità A_h:

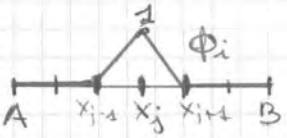
Devo scegliere una base che renda il calcolo di A_h semplice

Gli elementi della matrice non dipendono da h, ma le sue dimensioni sì → uso il pedice A_h

Se cambio base, cambiano A_h, b_h, c, ma non cambia u_h (cambia solo la sua rappresentazione)

Uso la base delle funz. a cappello per lo spazio delle funz. lineari a tratti

$$\Phi_i(x_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$



$$(A_h)_{i,j} = \nu \int_A^B \Phi_j' \Phi_i' \, dx$$

Φ_i lineare a tratti → Φ_i' cost. a tratti

bisogna solo vedere che succede qud $\Phi_i \neq 0$

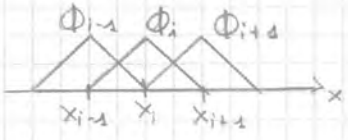
$$\nu \int_A^B \Phi_j' \Phi_i' = \nu \int_{x_{j-1}}^{x_{j+1}} \Phi_j' \Phi_i' \quad \text{perché } \Phi_i = 0 \text{ fuori da } (x_{j-1}, x_{j+1})$$

ogt è $\neq 0$ per $j = i$, $j = i \pm 1$

A_h è tridiagonale ↵

Calcolo gli elementi $\neq 0$

$$\Phi_i' = \begin{cases} \frac{1}{h} & x_{i-1} < x < x_i \\ -\frac{1}{h} & x_i < x < x_{i+1} \\ 0 & \text{altrimenti} \end{cases} \Leftarrow \text{pendenza della funz. } \Phi_i$$



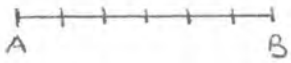
$$\int \Phi_{i-1}' \Phi_i' = -\frac{1}{h} = \int \Phi_{i+1}' \Phi_i' \quad \Leftarrow \int_{x_{i-1}}^{x_i} \Phi_{i-1}' \Phi_i' \, dx = \left(-\frac{1}{h}\right)\left(\frac{1}{h}\right)(x_i - x_{i-1}) = \left(-\frac{1}{h}\right)\left(\frac{1}{h}\right) \cdot h = -\frac{1}{h}$$

$$\int \Phi_i' \Phi_i' = \frac{2}{h} = \int_{x_{i-1}}^{x_{i+1}} = \int_{x_{i-1}}^{x_i} + \int_{x_i}^{x_{i+1}} = \left(-\frac{1}{h}\right)^2 h + \left(\frac{1}{h}\right)^2 h$$

$$A_h = \frac{\nu}{h} \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & 0 \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & 0 & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}$$

matrice del filo elastico

Introduco una griglia (uniforme per ora)



$h = \frac{B-A}{N}$ $\quad x_j = A + jh \quad j = 0, \dots N$

$I_j = (x_j, x_{j+1})$ ho N intervalli $V_h = \{ v : v(A) = v(B) = 0, v|_{I_j} \in \mathbb{P}^1 \}$ continue

N.B. (1) V e V_h sono spazi lineari → sono chiusi rispetto alla somma e al prodotto esterno:

- $\forall x, y \in V \ (\circ V_h) \rightarrow x + y \in V \ (\circ V_h)$
- $\forall x \in V \ (\circ V_h) \quad \forall \alpha \in \mathbb{R} \rightarrow \alpha x \in V (\circ V_h)$

(2) $V_h \subset V$ cioè se $v \in V_h \Rightarrow v \in V$

(3) $v \in V_h$ è def da un numero finito di g.d.l.

Ogni funz. $v \in V_h$ è completam. def. se assegno il valore di v sui nodi interni x_j:

v è completam. def. da $N-1$ valori $\Rightarrow \dim(V_h) = N-1$

quindi $V_h \approx \mathbb{R}^{N-1}$

$v \in V_h \iff [v(x_1) \dots v(x_{N-1})]$

Metodo agli elementi finiti:

Trovare $u_h \in V_h$ t.c. $a(u_h', v') = (f, v) \quad \forall v \in V_h$

Queste funz. lineari a tratti sono ancora infinite → se sfrutto il fatto che V_h è lineare ~~[illegible]~~ ogni $v \in V_h$ può essere scritta come combinaz. lineare di una base $\phi_1, \phi_2, \dots \in V_h$. Cioè $\forall v \in V_h \ \exists$ coeff. $\alpha_1, \alpha_2, \dots$ t.c. $v(x) = \sum_i \alpha_i \phi_i(x)$ (Questo vale sia per V_h che per V)

Poiché $\dim(V_h) = N-1$ ogni base contiene $N-1$ elementi:

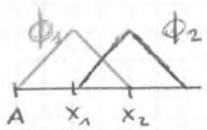
$\forall v \in V_h \quad v(x) = \sum_{i=1}^{N-1} \alpha_i \phi_i(x)$

Il mio probl quindi diventa:

Trovare $u_h \in V_h$ t.c. $a(u_h', \phi_i') = (f, \phi_i) \quad i = 1, \dots N-1$

$\Rightarrow$ le mie equaz. sono finite $(N-1)$ $\Rightarrow$ posso risolvere il sistema

La prima e l'ultima riga sono influenzate dalle condizioni al bordo



$$\int \phi_1' \phi_1' = \frac{2}{h} = a_{11} \leftarrow \int_A^{x_2} \phi_1' \phi_1' dx + \int_{x_1}^{x_2} \phi_1' \phi_1' dx = \left(\frac{1}{h}\right)^2 h + \left(-\frac{1}{h}\right)^2 h = \frac{1}{h} + \frac{1}{h} = \frac{2}{h}$$

$$\int \phi_2' \phi_1' = -\frac{1}{h} = a_{12}$$

$$\int_{x_1}^{x_2} \phi_2' \phi_1' dx = \left(\frac{1}{h}\right)\left(-\frac{1}{h}\right) h = -\frac{1}{h}$$

a sx non ci sono gradi di libertà
→ non ci sono elementi di base da considerare

⇒ è la condiz. di Dirichlet che garantisce l'invertibilità della matrice.

a = zeros (n-1) → matrice piena

c = a\b ~ $\frac{N^3}{3}$ operazioni

a = spdiags ( ) ⇒ c = a\b ~ 5N operazioni

Filo non vincolato

$$\begin{cases} -\nu u'' = f \\ u'(A) = 5 \\ u(B) = 0 \end{cases} \leftarrow \text{condiz. di Neumann} \qquad u(A) = ?$$

Se ho 2 condiz. di Neumann (A e B)
→ il probl non è più def in maniera univoca
perché se u è soluz. anche u + c è soluz.

Moltiplico per $v \in V$ e integro:

$$-\nu \int_A^B u' v = \int f v \qquad \forall v \in V$$

$$-\nu \left[u'(B) v(B) - \underbrace{u'(A)}_{=5} v(A) - \int_A^B u' v' \right] = \int f v$$

In B: $u(B) = 0$ → conosco la soluz., non ho gdl → prendo $v(B) = 0$

In A: devo calcolare u → ho bisogno di 1 gdl in più
⇒ ottengo $v(A) \neq 0$. Conosco $u'(A) = 5$

Devo trovare $u \in V$ t.c.

$$\nu \int u' v' = \int f v - \nu u'(A) v(A)$$

formulaz. debole variazionale del mio problema

$$V = \{ v : v(B) = 0 \quad \int (v')^2 < \infty \}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 145 di 314

se $\exists$ un vettore $\neq 0$ con $A_h c = 0 \Rightarrow A_h$ è singolare ($\rightarrow$ non invertibile)

Termine noto:

$b_i = \int\int \phi_i - \nu \delta \phi_i(A)$ per $i = 0, 1, \dots N-1$

Per le righe $1 \le i \le N-1$ $\int\int \phi_i \simeq h f(x_i)$ e $\phi_i(A) = 0$

1ª riga ($i=0$): $\int\int \phi_0 = \int_A^{A+h} \int \phi_0 = \frac{h}{2} \underbrace{(\int \phi_0)(A)}_{\phi_0(A)=1} + \underbrace{(\int \phi_0)(A+h)}_{\phi_0(A+h)=0} = \frac{h}{2} f(A)$

$\Rightarrow$ resta $f(A)\frac{h}{2}$

Nel termine noto le condiz. al bordo entrano:

$$b = h \begin{bmatrix} \frac{1}{2} f(A) \\ f(A+h) \\ \vdots \\ f(B-h) \end{bmatrix} + \begin{bmatrix} -\nu\alpha \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(se non ci entra $\frac{1}{2}$ spinge molto ai bordi)

$\phi_i(A) = \begin{cases} 1 & i = A \\ 0 & i \neq A \end{cases}$

Condizioni di Dirichlet non omogenee

$$\begin{cases} -\nu u'' = f & \text{filo vincolato non sul piano} \\ u(A) = g_1 & \text{dell'equilibrio, ma da} \\ u(B) = g_2 & \text{un'altra parte} \end{cases}$$

Conosco u in A e in B $\rightarrow$ i gdl sono sui nodi interni.

Moltiplico per una funz. test ed integro.

$-\nu \int u'' v = \int f v \quad \forall v \in V$

$V = \{ v : \underbrace{v(A) = v(B) = 0}_{\text{se non imponessi questo } V \text{ non sarebbe più uno spazio lineare!}} \text{ e } \int (v')^2 < \infty \}$

Se io avessi $V = \{ v : v(A) = 1, \dots \}$ e prendessi $u, w \in V$

$\Rightarrow (v+w)(A) = 1 + 1 = 2 \Rightarrow v + w \notin V$

Come prima, integro per parti:

$-\nu \left[u' v \Big|_A^B - \int u' v' \right] = \int f v \quad \forall v \in V$

($u'v|_A^B = 0$)

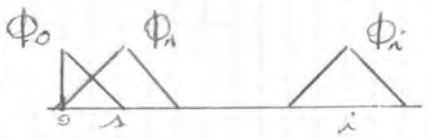
$\Rightarrow \nu \int u' v' = \int f v \quad \forall v \in V$ (come prima)

Stessa matrice di Dirichlet omogenea, quello che cambia è il termine noto:

$$(f, \Phi_i) = h f(x_i) \qquad i = 1, \dots N-1$$

$$\nu g_1 \left(\Phi_0', \Phi_i'\right) \neq 0 \text{ solo per } i = 1$$

$$\nu g_1 \int_A^{A+h} \Phi_0' \Phi_1' = \nu g_1 \int_A^{A+h} \left(-\frac{1}{h}\right)\left(\frac{1}{h}\right) = \frac{-\nu g_1}{h^2} h = -\frac{\nu g_1}{h}$$



$$b = h \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_{N-1}) \end{bmatrix} + \left(-\frac{\nu}{h}\right) \begin{bmatrix} g_1 \\ 0 \\ \vdots \\ 0 \\ g_2 \end{bmatrix}$$

13/12/11

Sappiamo che $V_h \subset V$

Considero la formulaz. variazionale solo su $v \in V_h$

$$\nu(u', v') = (f, v) \quad \forall v \in V_h$$

Sottraggo FEM

$$\nu(u', v') - \nu(u_h', v') = (f, v) - (f, v) \quad \forall v \in V_h$$

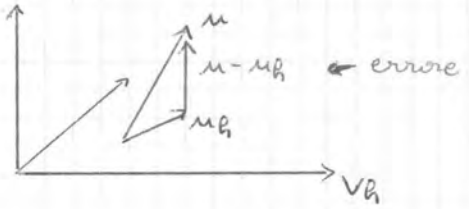
$$\boxed{\nu(u' - u_h', v') = 0 \quad \forall v \in V_h}$$ ortogonalità di Galerkin

$x, y \in \mathbb{R}^m$

$(x, y) = x^T y = \sum x_i y_i$

$x \perp y \Rightarrow (x, y) = 0$

lunghezza di $x = \|x\| = \sqrt{(x, x)}$



$v, w \in V$

$\|v, w\|^2 = (v', w') = \int v' \cdot w'$

$v \perp w \rightarrow \|v, w\| = 0$

lungh. di $v = \sqrt{\nu \|v, v\|} = \sqrt{\nu \int v', v'}$ = energia elastica della conformazione v

$= \|v\|_E$

Stima di errore

Calcolo

$\|u - u_h\|_E^2 = (u' - u_h', u' - u_h')$

Fisso $v \in V_h$ e scelgo $w = v - u_h \Rightarrow$ anche $w \in V_h$

(spazio lineare) → vale l'ortogon. di Galerkin

$(u' - u_h', w') = 0$

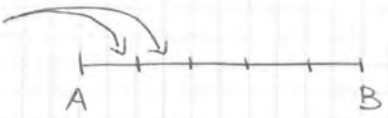
→

elementi P_1 •——• segmenti con 2 g.d.l. agli estremi } per funz. lineari a tratti

elementi P_2 •—•—• segmenti con 3 gdl. } per funz. quadratiche a tratti

Elementi P2

$h = \frac{B-A}{N}$ ogni segmento ha 3gdl



$x_j = A + jh$

i nodi intermedi sono $x_{j+\frac{1}{2}}$

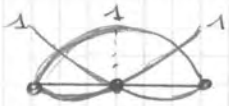
$V_h = \{ v : v(A) = v(B) = 0 \quad v|_{I_j} \in \mathbb{P}^2 \text{ continue} \}$

Trovare $u_h \in V_h$ t.c. $a(u', v') = (f, v) \quad \forall v \in V_h$

$\dim(V_h)$ = n° di parametri necessari a descrivere una funz. di V_h che coincide col n° dei nodi interni (le condiz. al bordo sono fissate) $= (N-1) + N = 2N - 1$

Sull'elemento di riferimento dobbiamo costruire una parabola → la base è data dai polinomi di Lagrange (3)

$$\mathcal{L}_j(x_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$



$$\mathcal{L}_j(x_i) = \prod_{\substack{i=1 \\ j \neq i}}^{n} \left(\frac{x - x_j}{x_i - x_j} \right)$$

$$\Rightarrow \mathcal{L}_1(x) = \frac{(x-\frac{1}{2})(x-1)}{(0-\frac{1}{2})(0-1)} = 2\left(x-\frac{1}{2}\right)(x-1)$$

$$\mathcal{L}_2(x) = \frac{(x-0)(x-1)}{(\frac{1}{2}-0)(\frac{1}{2}-1)} = \dots$$

} base locale sull'elemento di riferimento

La base locale sull'elemento di riferim. la chiameremo:

$\varphi_1(\hat{x}), \varphi_2(\hat{x}), \varphi_3(\hat{x})$

$\varphi_i(\hat{x}) = \mathcal{L}_i(\hat{x}) \qquad i = 1, 2, 3$

- riga pari $(\Phi_j', \Phi_i') \neq 0$ per:
  $j = i$, $j = i \pm 1$, $j = i \pm 2$

⇒ matrice pentadiagonale (almeno sulle righe pari)

Sull'elemento di riferimento ho una matrice di rigidezza locale 3×3:

$$(a^{loc})_{l,k} = \int_0^1 \varphi_k'(\hat{x})\, \varphi_l'(\hat{x})\, d\hat{x} = \frac{1}{3}\begin{pmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{pmatrix}$$

$\varphi_1 =$  $(2\hat{x}-1)(\hat{x}-1)$

$\varphi_2 =$  $-4\hat{x}(\hat{x}-1)$

$\varphi_3 =$  $\hat{x}(2\hat{x}-1)$

09/01/2012

Dobbiamo trasformare

$$\int_{x_{i-1}}^{x_i} \Phi_j' \Phi_i'\, dx = \dots\dots\dots\rightarrow \int_0^1 \frac{d}{d\hat{x}}\varphi_k \frac{d}{d\hat{x}}\varphi_l\, d\hat{x}$$

$x - x_{i-1} = h\hat{x}$

se $x = x_{i-1} \rightarrow \hat{x} = 0$

se $x = x_i \rightarrow \hat{x} = 1$

$$dx = h\, d\hat{x}$$

$$\frac{d}{dx} = \frac{d\hat{x}}{dx}\frac{d}{d\hat{x}} = \frac{1}{h}\frac{d}{d\hat{x}}$$

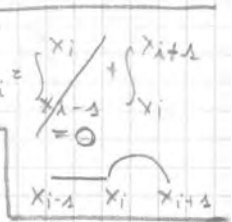
$$\int_{x_{i-1}}^{x_i} \Phi_j' \Phi_i'\, dx = \int_0^1 \frac{1}{h}\frac{1}{h}\frac{d}{d\hat{x}}\varphi_k \frac{d}{d\hat{x}}\varphi_l\, h\, d\hat{x} = \frac{1}{h}\int_0^1 \frac{d}{d\hat{x}}\varphi_k \frac{d}{d\hat{x}}\varphi_l\, d\hat{x}$$

<u>1ª riga</u>: $i = 1$, Φ_i = funz. bolla

$$(A_h)_{1,1} = \int_0^h \Phi_1' \Phi_1' = \frac{1}{h}\int_0^1 \varphi_2' \varphi_2'\, d\hat{x} = \frac{1}{h}\frac{16}{3}$$

$$(A_h)_{1,2} = \int_0^h \Phi_2' \Phi_1' = \frac{1}{h}\int_0^1 \varphi_3' \varphi_2'\, d\hat{x} = \frac{1}{h}\left(-\frac{8}{3}\right)$$

$(A_h)_{i,i} = \int_{x_{i-1}}^{x_i} + \int_{x_i}^{x_{i+1}}$ (il primo integrale $= 0$)



gli altri elementi della 1ª riga sono nulli perché la funz. bolla finisce qua

$(A_h)_{i,i+1} = \int_{x_i}^{x_{i+1}} \Phi_{i+1}' \Phi_i'$

$\|u - u_h\| \leq C h^2 \|f'\|$ (P_2)

$\|u - u_h\| \leq C h^3 \|f''\|$ (P_3)

⋮

FEM in 2D

(Simili agli elementi PIASTRA)

$-\nu \Delta u = f$ in $\Omega \subset \mathbb{R}^2$, $\Delta = \partial^2_{xx} + \partial^2_{yy}$

$u = 0$ su $\partial\Omega$, $\partial\Omega$ = bordo di Ω (continua)

ν scalare (costante elastica)

$u : \mathbb{R}^2 \to \mathbb{R}$ $\quad u(x,y) \in \mathbb{R}$

$\Delta u \in \mathbb{R}$ $\quad f : \mathbb{R}^2 \to \mathbb{R}$

① Teorema di Green.

funz. vettoriale $\vec{g} : \mathbb{R}^d \to \mathbb{R}^d$

$$\int_V \operatorname{div}(\vec{g})\, dv = \int_S \vec{g} \cdot \vec{n}\, ds$$

$\operatorname{div}(\vec{g}) = \vec{\nabla} \cdot \vec{g} = \partial_x g_x + \partial_y g_y$ scalare

in 2D: $V = \Omega$ $\quad S = \partial\Omega$

$$\int_\Omega \operatorname{div}(\vec{g}) = \int_{\partial\Omega} \vec{g} \cdot \vec{n}\, dl$$

② $\Delta u \cdot v = \vec{\nabla}(v \vec{\nabla} u) - (\vec{\nabla} u \cdot \vec{\nabla} v)$ (3 grandezze scalari)

infatti:

$\vec{\nabla}(v \vec{\nabla} u) = \vec{\nabla} v \cdot \vec{\nabla} u + v \vec{\nabla} \cdot (\vec{\nabla} u)$

Formulazione variazionale

$$-\int_\Omega \nu \Delta u \cdot v = \int_\Omega f \cdot v \underset{②}{=} -\nu \int_\Omega \vec{\nabla}(v \nabla u) + \nu \int_\Omega \nabla u \nabla v =$$

$$\underset{①}{=} -\nu \int_{\partial\Omega} v \cdot \vec{\nabla} u \cdot \vec{n}\, ds + \nu \int_\Omega \nabla u \nabla v \qquad \forall v \in \mathbf{V} = \left\{ v : v|_{\partial\Omega} = 0, \underbrace{\int \nabla v \nabla v}_{\text{en. elastica}} < \infty \right\}$$

$\rightarrow$ v_1 su AB è determinata
v_2 su AB " " $\Rightarrow v_1|_{AB} \equiv v_2|_{AB}$

$\rightarrow v$ è continua attraverso AB

Trovare $u_h \in V_h$ t.c. $\nu \int_{\Omega_h} \nabla u_h \nabla v = \int_{\Omega_h} f v \quad \forall v \in V_h$

Se $\partial\Omega$ è poligonale, ogni è equivalente a:

$$\nu \int_\Omega \nabla u_h \nabla v = \int_\Omega f v \quad \forall v \in V_h$$

N.B. V_h è uno spazio lineare perché:
- $v, w \in V_h \rightarrow v + w \in V_h$
- $v \in V_h, \alpha \in \mathbb{R} \rightarrow \alpha v \in V_h$

$\dim(V_h) = M$ = n° dei vertici interni (sul bordo la funz. vale 0)

Base di Lagrange

$\Phi_i(x,y) \in V_h$ t.c. su un vertice (x_j, y_j) $j = 1, \dots M$

$\Phi_i(x_j, y_j) = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$

supporto $(\Phi_i)_j = \cup T_j$
(Castre di vetro)

T_j ha (x_i, y_i) come vertice

Ogni $v \in V_h$ è $v(x,y) = \sum_{i=1}^{M} v_i \Phi_i(x,y) \qquad v_i = v(x_i, y_i)$

Trovare $u_h \in V_h$ t.c.

$$\nu \int_\Omega \nabla u_h \nabla \Phi_i = \int f \Phi_i \quad i = 1, \dots M$$

con $u_h = \sum c_j \Phi_j(x,y) \Rightarrow \nabla u_h = \sum c_j \nabla \Phi_j$

Trovare $[c_1, \dots c_M]^T$ t.c.

$$\sum c_j \nu \int_\Omega \nabla \Phi_i \nabla \Phi_i = \int_\Omega f \Phi_i$$

$(A_h)_{ij} = \int_\Omega \nabla \Phi_j \nabla \Phi_i$, $A_h c = b$, $b_i = \int_\Omega f \Phi_i$, $c = [c_1, \dots c_M]^T$

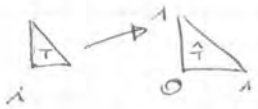
A_h è $M \times M$, simmetrica e definita positiva $\rightarrow$ invertibile

Devo calcolare $\int \nabla \Phi_j \nabla \Phi_i \qquad (A_h)_{i,j} \neq 0$

$\text{supp}(\Phi_i) \cap \text{supp}(\Phi_j) \neq 0$ ↑
◆ = $\text{supp}(\Phi_i) \cup \text{supp}(\Phi_j)$

$A_h \neq 0$ se il nodo i e il nodo j sono vertici di uno stesso triangolo

Considero il vertice (x_i, y_j)



$$\begin{cases} x - x_i = h_x \hat{x} \\ y - y_j = h_y \hat{y} \end{cases}$$

$$\begin{cases} dx = h_x d\hat{x} \\ dy = h_y d\hat{y} \end{cases}$$

$\nabla = (\partial_x, \partial_y)$ $\quad \partial_x = \frac{d\hat{x}}{dx} \partial_{\hat{x}} = \frac{1}{h_x} \partial_{\hat{x}}$, $\quad \partial_y = \frac{d\hat{y}}{dy} \partial_{\hat{y}} = \frac{1}{h_y} \partial_{\hat{y}}$

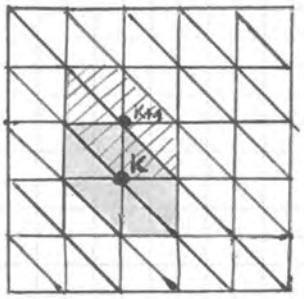
$$\int_T \left[\partial_x \phi_k \partial_x \phi_\ell + \partial_y \phi_k \partial_y \phi_\ell \right] dx\,dy = \int_T \nabla \phi_k \nabla \phi_\ell \, dx\,dy =$$

$$= \int_{\hat{T}} \left[\frac{1}{h_x^2} \partial_{\hat{x}} \varphi_m \partial_{\hat{x}} \varphi_m + \frac{1}{h_y^2} \partial_{\hat{y}} \varphi_m \partial_{\hat{y}} \varphi_m \right] \cdot d\hat{x}\,d\hat{y}\, h_x h_y$$

Se $h_x = h_y \rightarrow = \int_{\hat{T}} \hat{\nabla} \varphi_m \hat{\nabla} \varphi_m \, d\hat{x}\,d\hat{y}$

10/1/12

Per montare la matrice globale:
- dividiamo il nostro quadrato in triangoli uguali
- indice globale $k = (N-1)(i-1) + j$
- considero il nodo k e voglio calcolare $\int \nabla \phi_k \nabla \phi_\ell \neq 0$

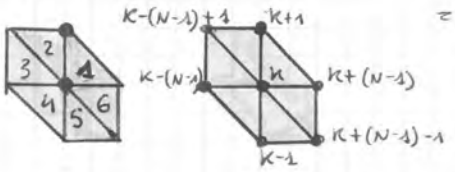


$\int \nabla \phi_k \nabla \phi_\ell \neq 0$ se $\text{supp}\,\phi_k \cap \text{supp}\,\phi_\ell \neq 0$

$\int \nabla \phi_k \nabla \phi_\ell \neq 0$ per $\ell = k+1$

$$\int \nabla \phi_k \nabla \phi_{k+1} = \int_{T_1} \nabla \phi_k \nabla \phi_{k+1} + \int_{T_2} \nabla \phi_k \nabla \phi_{k+1} =$$

$$= \int_{\hat{T}} \hat{\nabla} \varphi_1 \hat{\nabla} \varphi_3 + \int_{\hat{T}} \hat{\nabla} \varphi_3 \hat{\nabla} \varphi_3 = \frac{1}{2}(-1-1) = -1$$



Analogamente $\int \nabla \phi_k \nabla \phi_{k-1} = -1$

$\int \nabla \phi_k \nabla \phi_{k+(N-1)} = -1$

$\int \nabla \phi_k \nabla \phi_{k-(N-1)} = -1$

$\int \nabla \phi_k \phi_{k+(N-1)-1} = \int_{T_5} (\quad) + \int_{T_6} (\quad) = \int_{\hat{T}} \hat{\nabla} \varphi_2 \hat{\nabla} \varphi_3 + \int_{\hat{T}} \hat{\nabla} \varphi_2 \hat{\nabla} \varphi_3 = 0$

$\int \nabla \phi_k \phi_{k-(N-1)+1} = 0$

sulla diagonale $\rightarrow = 4$

Calcolare autoval. e autovett. del probl. esatto vuol dire risolvere:

$\begin{cases} -u'' = \lambda u \\ u(0) = u(1) = 0 \end{cases}$ λ = livelli energetici

sin, cos rispettano la prima equaz.

$u(x) = a \sin kx + b \cos kx$

$u''(x) = -k^2 u \Rightarrow \lambda = +k^2$

Devo considerare anche le condiz. al bordo

$\begin{cases} u(0) = 0 \to b = 0 \quad (\text{nn posso avere la funz. coseno}) \\ u(1) = 0 \to a \sin k = 0 \to k = \ell\pi \quad \ell \neq 0 \text{ intero} \end{cases}$

$\Rightarrow \lambda = k^2 = \pi^2 \ell^2 = \pi^2, 4\pi^2, 9\pi^2, 16\pi^2, \dots$

i valori nn sono continui ma quantizzati

sono illimitati

Gli autovettori sono $u_\ell(x) = \sin \ell\pi x$

Per la matrice, cerco dei vettori u_ℓ t.c. $A_h u_\ell = \lambda_\ell u_\ell$

- riga i-esima:

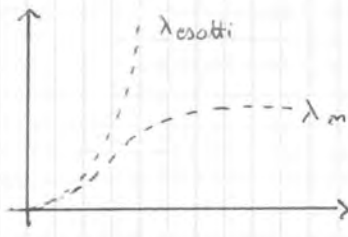
$-(u_\ell)_{i-1} + 2(u_\ell)_i - (u_\ell)_{i+1} = \lambda_\ell (u_\ell)_i$

prove $(u_\ell)_i = \sin \ell\pi (x_i)$ punto di griglia

$\begin{cases} x_{i-1} = x_i - h \\ x_{i+1} = x_i + h \end{cases}$

$\Rightarrow -\sin \ell\pi (x_i - h) + 2 \sin \ell\pi x_i - \sin \ell\pi (x_i + h) = \lambda_\ell \sin \ell\pi x_i$

$\lambda_\ell = 2(1 - \cos \ell\pi h)$



---- parabola discreta (a punti)

λ numerici $\to (u_\ell)_i = \sin \ell\pi x_i$

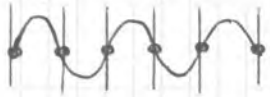
Autovett discreto: $\vec{u}_\ell = [\sin \ell\pi h, \sin \ell\pi 2h, \dots \sin \ell\pi (N-1)h]^T$

Per bassi valori di ℓ: $\lambda_{esatto} \simeq \lambda_{numerico}$

Problema: sembra che ℓ sia illimitata, ma la matrice è $(N-1) \times (N-1)$ → il primo autovett. di troppo è u_ℓ per $\ell = N$:

$(u_\ell)_i = \sin \ell\pi ih = \sin N\pi \frac{ih}{\text{intero}}$

per la griglia che ho scelto $Nh = 1 \to (u_\ell)_i = \sin \pi i = 0$



sulla griglia $\sin \ell\pi x$ con $\ell = N$ coincide con la funz. nulla

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 161 di 314

Per risolverlo con gli elementi finiti, procediamo col solito metodo.

→ formulazione variazionale:

$$\int(-\nu u'' + \beta u')v = \int fv$$

assumo $\nu, \beta = cost$ → $-\nu\int_\Omega u''v + \beta\int_\Omega u'v = \int_\Omega fv$

integrando per parti → $-\nu\,(u'v)|_{\partial\Omega} + \nu\int_\Omega u'v' + \beta\int_\Omega u'v = \int_\Omega fv$

condiz. al contorno di Dirichlet → $v = 0$

Trovare $u \in V$ t.c. $\nu\int_\Omega u'v' + \beta\int_\Omega u'v = \int_\Omega fv \quad \forall v \in V$

$V = \{ v : v|_{\partial\Omega} = 0, \ \int(v')^2 < \infty \}$

se $v|_{\partial\Omega} = 0$ e $\int(v')^2 < \infty \Rightarrow \int v^2 < \infty$

quindi $\int u'v \leq \int(u')^2 \int v^2 < \infty$

FEM.

Introduco una griglia (uniforme) → Ω viene diviso in N intervalli uguali di ampiezza h. Uso elementi P_1:

$$V_h = \{ v : v(0) = v(1) = 0 \quad v|_{I_j} \in \mathbb{P}^1 \text{ continue} \}$$

Base delle funz. a cappello $\{\phi_i\}_{i=1}^{N-1}$

$$u_h(x) = \sum_{j=1}^{N-1} c_j \phi_j(x)$$

Trovare $c = [c_1, \dots, c_{N-1}]$ t.c. $\sum c_j \left[\nu\int\phi_j'\phi_i' + \beta\int\phi_j'\phi_i\right] = \int f\phi_i$ $\quad i = 1, \dots N-1$

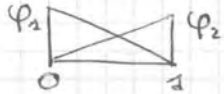
$$(A_h)_{ij} = \underbrace{\nu\int_\Omega \phi_j'\phi_i'}_{\text{opt perché so già com'è fatto}} + \underbrace{\beta\int_\Omega \phi_j'\phi_i}_{\text{termine di convezione (da calcolare)}}$$

Termine di convezione

$(B_{ij}) = \beta\int_\Omega \phi_j'\phi_i$ non simmetrico

è una matrice antisimmetrica: $\beta\int\phi_j'\phi_i = -\beta\int\phi_i\phi_j'$

Base di elementi



$\varphi_1 = 1 - \hat{x}$

$\varphi_2 = \hat{x}$

(a) $|Pe|<1$ → matrice a dominanza diagonale
$2 \geq 2$ → dominanza stretta ai bordi
} A_h invertibile

(b) $|Pe|>1$ → matrice non a dominanza diag. → non so se A_h è invertibile

↓

per altra via si scopre che A_h è comq invertibile

Quello che succede è che:
- A_h è invertibile $\forall Pe$
- se $|Pe|<1$ → gli autoval. sono tutti $\in \mathbb{R}$ e u_h è regolare
- se $|Pe|>1$ → gli autoval. sono tutti $\in \mathbb{C}$ e u_h è oscillante

$Pe = \frac{\beta h}{2\nu}$ $\begin{cases} |Pe|<1 \Rightarrow h < \frac{2\nu}{|\beta|} \\ |Pe|>1 \Rightarrow h > \frac{2\nu}{|\beta|} \end{cases}$ griglia fitta → riesco a riprodurre la realtà → u_h non oscilla

Esempio

$$\begin{cases} \nu u'' + \beta u' = 0 & u = ce^{\lambda x} \quad \text{in } (0,1) \\ u(0) = 0 \quad u(1) = 1 & \beta, \nu > 0 \end{cases}$$

$$\nu\lambda^2 + \beta\lambda = 0 \Rightarrow \begin{cases} \lambda_1 = 0 \\ \lambda_2 = \frac{\beta}{\nu} \end{cases} \rightarrow u = c_1 + c_2 e^{\frac{\beta}{\nu}x}$$

$$u(x) = \frac{e^{(\beta/\nu)x} - 1}{e^{\beta/\nu} - 1}$$

1) $\frac{\beta}{\nu} \ll 1$ → diffusione dominante

$$u(x) \sim \frac{1 + \frac{\beta}{\nu}x - 1}{1 + \frac{\beta}{\nu} - 1} = x$$

2) $\frac{\beta}{\nu} \gg 1$ → convet. dominante

$$u(x) \sim \frac{e^{(\beta/\nu)x}}{e^{\beta/\nu}} = e^{\frac{\beta}{\nu}(x-1)} = \begin{cases} \simeq 0 & \text{se } x \neq 1 \\ 1 & \text{se } x = 1 \end{cases}$$

($x-1 < 0$)

strato limite condiz. al bordo downwind ($\beta > 0$)

la funz. è $\simeq 0$ tranne che nello strato lim.

Nello strato limite δ la solut. salta da 0 a 1 (nel nostro es.)

→ $u_x|_\delta \simeq \frac{1-0}{\delta}$

Per convet. dominante $u' \sim \frac{\beta}{\nu} \frac{e^{\frac{\beta}{\nu}x}}{e^{\frac{\beta}{\nu}}} \rightarrow u'(1) \simeq \frac{\beta}{\nu} = \frac{1}{\delta} \rightarrow \delta \simeq \frac{\nu}{\beta}$

la solut. non oscilla se $Pe < 1 \rightarrow \frac{\beta h}{2\nu} < 1 \rightarrow h < \frac{2\nu}{\beta} = 2\delta$

quindi ho bisogno di 1 pt di griglia all'interno dello strato lim. →

Trovare $u_h \in V_h$ t.c. $\nu \int_\Omega \nabla u_h \nabla v + \int_\Omega \beta \nabla u_h \, v = \int_\Omega f v$

$\forall v \in V_h = \{ v : v|_{\partial\Omega_h} = 0 \,,\; v|_{T_j} \in \mathbb{P}^1 \text{ continue} \}$

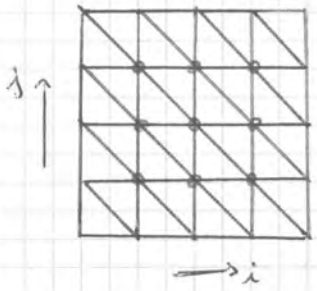
Base ϕ_k funz. a cappello $u_h(x,y) = \sum_k c_k \phi_k(x,y)$

$\sum_k c_k \left[\int_\Omega \nu \nabla \phi_k \nabla \phi_\ell + \int_\Omega \beta \nabla \phi_k \phi_\ell \right] = \int_\Omega f \phi_\ell \qquad \ell = 1, 2, \dots M$

$M = \dim(V_h)$

Esempio

$\Omega = (0,1)^2$ $\quad \Omega_h$ triangolaz. uniforme $h = \frac{1}{N}$



$\dim V_h = (N-1)^2$

$k = (N-1)(i-1) + j$

$i, j = 1, \dots N-1$

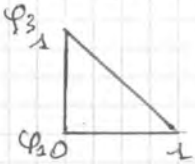
Generico modo:

$(X_i, Y_j) = (ih, jh) \qquad i, j = 1, \dots N-1$

$= (x_i, y_k)$

$(A_h)_{\ell,k} = \underbrace{\nu \int_\Omega \nabla \phi_k \nabla \phi_\ell}_{\text{diffusione}} + \underbrace{\int_\Omega (\beta \nabla \phi_k) \phi_\ell}_{\text{convezione}}$

Sull'elemento di riferimento:



$\varphi_1 = 1 - \hat{x} - \hat{y} \qquad \nabla \varphi_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$\varphi_2 = \hat{x}$

$\varphi_3 = \hat{y}$

$\int_{\hat{T}} \beta \nabla \varphi_m \varphi_m = a_{m,m} \rightarrow$ per es. $a_{1,1} = \int_{\hat{T}} [a \;\; b] \begin{bmatrix} -1 \\ -1 \end{bmatrix} (1 - \hat{x} - \hat{y}) \, d\hat{x} \, d\hat{y} =$

$= (-a - b) \int_0^1 d\hat{x} \int_0^{1-\hat{x}} (1 - \hat{x} - \hat{y}) \, d\hat{y} =$

$= (-a - b) \underbrace{\frac{1}{3} \cdot \frac{1}{2} \cdot 1}_{\text{volume piramide}} = -\frac{a+b}{6}$

$A_{h\,e,e-1} = \int_{⑤}(\quad) + \int_{④} \beta \nabla \Phi_{e-1} \Phi_e = h a_{3,1} - h a_{1,3} = \frac{h}{6}(-a-b) - \frac{h}{6} b = -\frac{h}{6}(a+2b)$

$\int_{⑤} = \int_{\hat{T}} \nabla \varphi_1 \varphi_3 = a_{3,1} \qquad \int_{④} = -\int_{\hat{T}} \beta \nabla \varphi_3 \varphi_1 = a_{1,3}$

⇒ vediamo che A_h effettivam. è antisimmetrica!

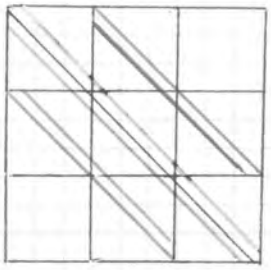
$A_{e,e} = 0$ perché di 6 triangoli < 3 hanno segno ⊕ / " " " ⊖ } si annullano.

$(A_h)_{e,\,e+(N-1)-1} = \int_{⑤} + \int_{⑥} = h a_{3,2} - h a_{2,3} = \frac{h}{6}(a-b)$

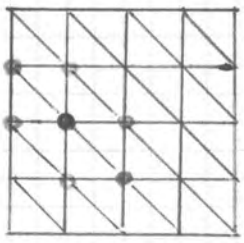
$\int_{⑤} \beta \nabla \Phi_{e+(N-1)-1} \Phi_e = \int \beta \nabla \varphi_2 \varphi_3 = a_{3,2} \qquad \int_{⑥}(\quad) = -\int \beta \nabla \varphi_3 \varphi_2 = a_{2,3}$

Risolvendo tutti i calcoli si ricava:

- A_h è una matrice a blocchi
- ogni blocco ha dim $N-1$
- sulla diagonale ci sono zeri
- sotto la diagonale ci sono gli elementi $e-1$
- matrice antisimm. → $(-) = -(-)$



matrice di convezione



$A_{h\,e,e} = 0$

$A_{h\,e,e+1} = +\frac{bh}{3} + \frac{ah}{6}$

$A_{h\,e,e-1} = -\frac{bh}{3} - \frac{ah}{6}$

$A_{h\,e,e+(N-1)} = \frac{ah}{3} + \frac{bh}{6}$

$A_{h\,e,e-(N-1)} = -\frac{ah}{3} - \frac{bh}{6}$

$A_{h\,e,e-(N-1)+1} = \frac{bh}{6} - \frac{ah}{6}$

$A_{h\,e,e+(N-1)-1} = \frac{ah}{6} - \frac{bh}{6}$

Matrice globale:

$(A_h) = (\text{diffusione}) + (\text{convezione})$

Mi aspetto che, se convez. >> diffusione, potrò avere delle oscillazioni spurie. Calcolo i Peclet: $P_x = \frac{ah}{2\nu}$

$P_y = \frac{bh}{2\nu}$ →

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 169 di 314

Triang. uniforme

$$V_h^1 = \left\{ v : \; v|_{\Gamma_2} = 0 \quad v|_{T_K} \in \mathbb{P}^1 \; \forall T_K \text{ continua} \right\}$$

N° incognite = M = nodi interni + nodi di Neumann =

$$= (N-1)^2 + 2(N-1) = (N-1)(N+1) = \text{dim matrice}$$

$$A_h = \begin{bmatrix} N_e & & \\ & (N-1)^2 \times (N-1)^2 & \\ & & N_e \end{bmatrix}$$

N_e = blocco di Neumann

$N_e = (N-1) \times (N-1)$

funz. a meno cappello

Considero il primo blocco di Neumann

$$\int \nabla\phi_e \nabla\phi_e = ① + ② + ③ = 2$$

$$A_{e,e+1} = \int \nabla\phi_{e+1} \nabla\phi_e = ② = -\frac{1}{2}$$

$$A_{e,e-1} = -\frac{1}{2}$$

$$A_{e,e+(N-1)} = -1$$

$$A_h = \begin{bmatrix} 2 & -\frac{1}{2} & & -1 & & \\ -\frac{1}{2} & & -\frac{1}{2} & & -1 & \\ & -\frac{1}{2} & 2 & & & -1 \\ -1 & & & 4 & -1 & \\ & -1 & & -1 & 4 & -1 \\ & & -1 & & -1 & 4 \\ & & & -1 & & \\ & & & & -1 & \\ & & & & & -1 \end{bmatrix} = \begin{bmatrix} N_e & -I & & \\ -I & G_1 & \ddots & \\ & \ddots & \ddots & -I \\ & & -I & N_e \end{bmatrix}$$

$$N_e = \begin{bmatrix} 2 & -\frac{1}{2} & & \\ -\frac{1}{2} & \ddots & \ddots & \\ & \ddots & \ddots & -\frac{1}{2} \\ & & -\frac{1}{2} & 2 \end{bmatrix} \qquad G_1 = \begin{bmatrix} 4 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 4 \end{bmatrix}$$

A_h è a dominanza diagonale

A_h è " " " stretta dove ci sono i nodi di Dirichlet

Vettore di carico

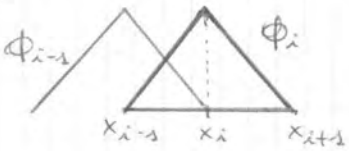
$$b_e = \int_\Omega 4\phi_e + \int_{\Gamma_1} \phi_e$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 171 di 314

$\{\phi_\ell\}$ base di V_h, $\ell = 1, \dots N-1$

Trovare $c = [c_1, \dots c_{N-1}]$ t.c. $\sum_{j=1}^{N-1} c_j \int \nu(x) \phi_j' \phi_i' = \int_1^2 \phi_i \quad i = 1, \dots N-1$

$A_{i,j} = \int_1^2 \nu(x) \phi_j' \phi_i' \neq 0$ per $j = i$, $j = i \pm 1$



$A_{i,i-1} = \int_{x_{i-1}}^{x_i} \times \left(-\frac{1}{h}\right)\left(\frac{1}{h}\right) dx =$ (sale, scende)

$= -\frac{1}{h^2} \frac{x^2}{2} \Big|_{x_{i-1}}^{x_i} = -\frac{1}{2h^2}\left(x_i^2 - x_{i-1}^2\right) = \frac{-1}{2h}(x_i + x_{i-1})$

$A_{i,i+1} = \int_{x_i}^{x_{i+1}} \times \left(-\frac{1}{h}\right)\left(\frac{1}{h}\right) dx = -\frac{1}{2h}(x_i + x_{i+1})$

$A_{i,i} = \int_{x_{i-1}}^{x_i} \times \left(\frac{1}{h}\right)^2 dx + \int_{x_i}^{x_{i+1}} \times \left(+\frac{1}{h}\right)^2 dx = \frac{1}{2h}(x_i + x_{i+1} + x_i + x_{i-1})$

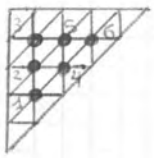
$A_{i,i-1} = A_{i-1,i}$ => simmetrica

inoltre è a dominanza diagonale stretta => invertibile

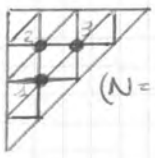
Tema d'esame del 22/06/2010 - es 1

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = 0 & \text{su } \partial\Omega \end{cases}$$

$\Omega =$ ($N=5$)



$h = \frac{1}{N}$



Formulaz. variazionale:

$$\int_\Omega \nabla u \nabla v = \int_\Omega f v \quad \forall v \in V$$

$$V = \{ v : v|_{\partial\Omega} = 0, \int |\nabla v|^2 < \infty \}$$

Nodi interni:

$$M = \frac{(N-1)^2 - (N-1)}{2} = \frac{(N-1)(N-2)}{2}$$

per $N=4$ ho 3 nodi interni

$$\begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{bmatrix}$$

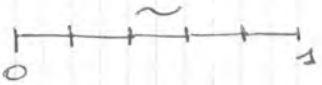
per $N=5$ ho 6 nodi interni

$$\left[\begin{array}{ccc|ccc} 4 & -1 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & -1 & 0 & 0 \\ 0 & -1 & 4 & 0 & -1 & 0 \\ \hline 0 & -1 & 0 & 4 & -1 & 0 \\ 0 & 0 & -1 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & -1 & 4 \end{array}\right]$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 173 di 314

Per procedere con gli elementi finiti devo introdurre una griglia. Devo usare elementi che garantiscano la continuità della derivata prima v' (e quindi di v) → non posso usare elementi $\mathbb{P}_1$

$h = \frac{1}{N}$



Supponiamo di aver calcolato il nostro polinomio → devo poter fissare 2 condiz. a sx e 2 a dx → il grado min. dei polinomi da usare in qst caso è 3 ⇒ polinomi di Hermite : H^3

Sono definiti assegnando $H^3(x)$ t.c. interpoli una funz. $f(x)$ in 2 punti (x_1 e x_2) e interpoli $f'(x)$ negli stessi punti

continuità di f
continuità di f'

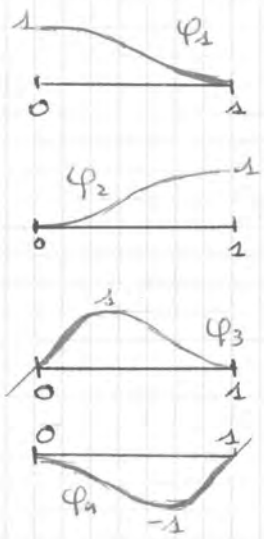
Localmente

$\varphi_1(0) = 1$ $\quad \varphi_1(1) = 0$
$\varphi_1'(0) = 0$ $\quad \varphi_1'(1) = 0$

$\varphi_2(0) = 0$ $\quad \varphi_2(1) = 1$
$\varphi_2'(0) = 0$ $\quad \varphi_2'(1) = 0$

$\varphi_3(0) = 0$ $\quad \varphi_3(1) = 0$
$\varphi_3'(0) = 1$ $\quad \varphi_3'(1) = 0$

$\varphi_4(0) = 0$ $\quad \varphi_4(1) = 0$
$\varphi_4'(0) = 0$ $\quad \varphi_4'(1) = 1$



$\varphi_1(x) = 1 - x^2 + 2x^2(x-1)$
$\varphi_2(x) = x^2 - 2x^2(x-1)$
$\varphi_3(x) = x - x^2 + x^2(x-1)$
$\varphi_4(x) = x^2(x-1)$

$$V_h = \left\{ v : \begin{array}{l} v(0) = v(1) = 0 \\ v'(0) = v'(1) = 0 \end{array} \quad v|_{I_j} \in H^3 \quad v, v' \text{ continue} \right\}$$

⇒ $V_h \subset V$, V_h è uno spazio lineare → dimostriamolo!:

$\dim(V_h) = \underbrace{N^\circ \text{ di } dof}_{4N \text{ perché ho un polinomio di grado 3}} - \underbrace{\text{vincoli}}_{4 \text{ (nodi al bordo)}\ 2(N-1) \text{ (in ogni nodo interno)}} = 4N - 4 - 2(N-1) = 2(N-1)$

171

$$(A_h)_{ii} = \int_0^1 \Phi_i^{IIm}\,\Phi_i^{IIm} = \int_{x_{i-1}}^{x_i} \Phi_i^{IIm}\,\Phi_i^{IIm} + \int_{x_i}^{x_{i+1}} \Phi_i^{IIm}\,\Phi_i^{IIm} =$$

$$= \frac{1}{h^3}\left[\int_0^1 \varphi_2^{II}\,\varphi_2^{II} + \int_0^1 \varphi_1^{II}\,\varphi_1^{II}\right] = \frac{1}{h^3}(12+12) = \frac{24}{h^3}$$

$$(A_h)_{i,i+1} = \int_0^1 \Phi_i^{IIm}\,\Phi_{i+1}^{IIm} = \int_{x_i}^{x_{i+1}} \Phi_i^{IIm}\,\Phi_{i+1}^{IIm} = \frac{1}{h^3}\,a_{12} = -\frac{12}{h^3}$$

$$A_h = \left[\begin{array}{cc|cc} 24 & -12 & 0 & 6 \\ -12 & & -6 & \\ & -12 & & 6 \\ -12 & 24 & -6 & 0 \\ \hline 0 & -6 & 8 & 2 \\ 6 & & 2 & \\ & -6 & & 2 \\ 6 & 0 & 2 & 8 \end{array}\right]$$

A_h è simmetrica definita positiva

$\text{cond}(A_h) \sim \frac{1}{h^4}$

N.B. i polinomi H^3 sono fatti per dare sol. continue → se li uso per probl. che sono discontinui → ci sono oscillaz. spurie. (per es. non posso simulare fratture)

Vettore di carico

uso la regola di Simpson

$$\int_a^b g(x)\,dx \simeq \frac{b-a}{6}\left[g(a) + 4g\left(\frac{a+b}{2}\right) + g(b)\right]$$

$$\int_{x_{i-1}}^{x_{i+1}} f\,\Phi_i^m(x) = \int_{x_{i-1}}^{x_i} f\,\Phi_i^m + \int_{x_i}^{x_{i+1}} f\,\Phi_i^m$$

chiamo $x_{i-1/2} = \frac{1}{2}(x_i + x_{i-1})$

$$x_{i+1/2} = \frac{1}{2}(x_i + x_{i+1})$$

$$\int f\Phi_i^m \simeq \frac{h}{6}\left[\cancel{f\Phi_i^m(x_{i-1})}^{=0} + 4f\Phi_i^m(x_{i-1/2}) + \underbrace{f\Phi_i^m(x_i)}_{=1} + \underbrace{f\Phi_i^m(x_i)}_{=1} + 4f\Phi_i^m(x_{i+1/2}) + \cancel{f\Phi_i^m(x_{i+1})}^{=0}\right]$$

$$\Phi_i^m(x_{i-1/2}) = \varphi_2\left(\frac{1}{2}\right) = \frac{1}{2}$$

$$\int f\Phi_i^m \simeq \frac{h}{6}\left[\frac{4}{2}f(x_{i-1/2}) + 2f(x_i) + \frac{4}{2}f(x_{i+1/2})\right]$$

Il precondiz. M deve:

(1) $\det M \neq 0$ ($\to$ M invertibile)

(2) $M \approx A$

(3) M^{-1} sia facile da calcolare, o meglio $Mx_{k+1} = Mx_k + bAx_k$ dev'essere facile da risolvere (cioè veloce da risolvere)

(2) e (3) tendono ad andare un po' in conflitto $\frac{n}{2}$

(1) Metodo di Richardson

(è lentissimo, ma è facile)

$M_k = M = \frac{1}{\alpha} I \quad \alpha \in \mathbb{R}$ ($M^{-1} = \alpha I \to$ è facile calcolare l'inversa $\to$ soddisfa la (3))

$x_{k+1} = x_k + \alpha \underbrace{(b - Ax_k)}_{r_k}$

r_k residuo (quello che manca per risolvere il mio sistema)

Calcolo l'errore $e_k := x - x_k$

$x - x_{k+1} = x - x_k - \alpha(Ax - Ax_k) \to e_{k+1} = e_k - \alpha A e_k = \underbrace{(I - \alpha A)}_{B_R} e_k$

B_R matrice di iterazione (R = Richardson)

Affinché il metodo converga, B_R deve ridurre l'errore ad ogni passo $\to$ dev'essere una matrice di contraz.: il metodo converge se e solo se B_R è una matrice di contrazione $\to \underbrace{\rho(B_R)}_{\text{raggio spettrale = autovalore di modulo massimo}} < 1$.

Per Richardson gli autoval di B_R sono $1 - \alpha\lambda(A)$ e quindi:

$|1 - \alpha\lambda(A)| < 1$

Se A è SPD (simm. e def. positiva) $\to \lambda(A) > 0$

$-1 < 1 - \alpha\lambda(A) < 1$

($\alpha > 0$)

$-2 < -\alpha\lambda(A)$

$\alpha < \frac{2}{\lambda(A)} \quad \forall \lambda$

$\alpha < \frac{2}{\lambda_{max}}$

Ma perché ci dev'essere una condiz. di convergenza che dipende da λ? $\to$

Matlab:

routine del gradiente coniugato

$y = cg(a,b)$ → mi dà la soluz. del problema

Calcolo $\|Ay - b\|$ (residuo) → se è piccolo → soluz. buona

nmax = n° max di iterazioni che fa Matlab → se non lo imposto Matlab per default prende nmax = 20 → poche iterazioni !!
→ il residuo sarà enorme!!!

Matrice malcondizionata:

- amplifica l'errore dei dati in input
- se uso un metodo approssim. converge lentamente

Per Richardson:

n° di iterazioni per arrivare a convergenza è $\sim K(A)$.

Per matrici FEM per:

- $-\Delta u = f$ $\quad K(A) \sim N^2$ (N = n° intervallo per unità di lungh)

② Metodo di Jacobi

$M = D = \text{diag}(A)$

$x_{k+1} = x_k + D^{-1}(b - Ax_k)$

$D^{-1} = \text{diag}\left(\frac{1}{a_{11}}, \frac{1}{a_{22}}, \dots \frac{1}{a_{ii}}\right)$

Si trova che la velocità di Jacobi è circa quella di Richardson.

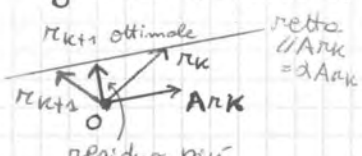
③ Metodo di Ortomin

$x_{k+1} = x_k + \alpha_k (b - Ax_k)$

Scelgo α_k in modo da minimizzare il residuo al passo successivo:

$r_{k+1} = b - Ax_{k+1} = b - Ax_k - \alpha_k Ar_k = r_k - \alpha_k Ar_k$

Geometricamente:



$r_k \neq 0$, Amm singolare $\neq 0 \rightarrow Ar_k \neq 0$

Calcolo $(r_{k+1}, Ar_k) = 0$ (prodotto scalare nullo → vettori ortogonali)

$\downarrow$

$(r_k, Ar_k) - \alpha_k (Ar_k, Ar_k) = 0$

$\alpha_k = \dfrac{(r_k, Ar_k)}{(Ar_k, Ar_k)}$ → $\|r_{k+1}\|$ minimo

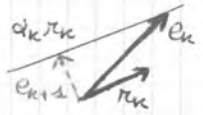
④ Metodo del gradiente

$x_{k+1} = x_k + \alpha_k r_k$ (iterate identica ad ortomin)

È diverso il calcolo di α_k → α_k è tale che e_{k+1} sia minimo. L'errore non è facile da calcolare perché non conosco la soluz. esatta:

per def $e_{k+1} = x - x_{k+1}$, dove x = soluz. esatta

$e_{k+1} = e_k - \alpha_k r_k \Rightarrow e_{k+1} \perp r_k$



siccome non conosco né e_{k+1} né r_k, voglio che $e_{k+1} \perp r_k$ siano $\perp$ nel senso di A, cioè $(e_{k+1}, Ar_k) = 0$ → così $\|e_{k+1}\|_A$ sarà minima.

Vorrei α_k t.c. $(e_{k+1}, Ar_k) = 0$

Sfrutto il fatto che A è simmetrica:

$$(e_{k+1}, Ar_k) = (r_k, Ae_{k+1}) = (r_k, A(x - x_{k+1})) =$$
$$= (r_k, \underbrace{b}_{Ax} - A\underbrace{(x_k + \alpha_k r_k)}_{x_{k+1}}) = (r_k, r_k) - \alpha_k (r_k, Ar_k)$$

$\Rightarrow (r_k, r_k) - \alpha_k (r_k, Ar_k) = 0$ per $\alpha_k = \dfrac{(r_k, r_k)}{(r_k, Ar_k)}$

il fatto che A sia SPD ci assicura che $(r_k, Ar_k) > 0$

(r_k, r_k) è sempre ≥ 0 perché il prodotto scalare di un vettore per se stesso.

$\Rightarrow$ $\boxed{\alpha_k > 0 \quad \forall k}$ il metodo del gradiente converge sempre!

Riassunto:

Stima iniziale x_0 [$x_0 = b$ ad es., Matlab usa $x_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$]

per $k = 0, 1, \dots$

$$x_{k+1} = x_k + \alpha_k r_k \qquad r_k = b - Ax_k$$

dove:

$$\alpha_k = \frac{(r_k, r_k)}{(r_k, Ar_k)} \quad \text{[metodo gradiente]}$$

$$\alpha_k = \frac{(Ar_k, r_k)}{(Ar_k, Ar_k)}$$

end.

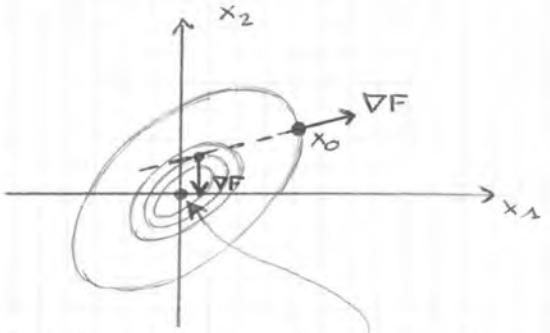
⇒ Manca il test di arresto! → ci sono 2 modi:

1) Sul residuo:

esci se $(r_k, r_k) \leq tol \cdot \|b\|$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 183 di 314

Le curve di livello sono:



$x_0 = (x_1, x_2)_0$
curva di livello $F(x_0)$

Voglio arrivare al centro il più in fretta possibile. Il gradiente ci dà la direzione di max crescita ⇒ la direz. di max discesa è l'opposta ⇒ mi muovo lungo $-\nabla F|_{x_0}$
Visto che non posso muovermi all'infinito, vado fino a quel punto x_1 → lì la curva di livello è tan alla direz. del gradiente
Quindi, mi muovo lungo $-\nabla F|_{x_0}$ con un passo α_k che mi porta tangente alla successiva curva di livello

$$x_{k+1} = x_k + \alpha_k \left(-\nabla F|_{x_k}\right)$$

Calcolo ∇F, $\quad F(x) = \frac{1}{2}(x, Ax) - (b, x)$

$$\nabla F = Ax - b = -r$$

Quindi il metodo $x_{k+1} = x_k + \alpha_k r_k$ coincide col metodo del gradiente.
Il passo α ottimale è determinato imponendo che:

$$F(x_k + \alpha_k r_k) \le F(x_k + \alpha r_k) \quad \forall \alpha$$

cioè voglio trovare il minimo di $g(\alpha) = F(x_k + \alpha x_k) =$

$$= \frac{1}{2}(x_k + \alpha r_k, A(x_k + \alpha r_k)) - (b, x_k + \alpha r_k)$$

$g(\alpha): \mathbb{R} \to \mathbb{R}$ quindi è tutto finito tranne α.
So calcolare i minimi:

$$g(\alpha) = \frac{1}{2}(x_k, Ax_k) + \frac{1}{2}\alpha(r_k, Ax_k) + \frac{1}{2}\alpha(x_k, Ar_k) + \frac{1}{2}\alpha^2(r_k, Ar_k) +$$
$$-(b, x_k) - \alpha(b, r_k) =$$

↖ termini uguali perché A è simmetrica

$$= \frac{1}{2}(x_k, Ax_k) + \alpha(r_k, Ax_k) - \alpha(b, r_k) - (b, x_k) + \frac{1}{2}\alpha^2(r_k, Ar_k)$$

Derivo rispetto ad α:

$$g'(\alpha) = (r_k, Ax_k) - (b, r_k) + \alpha(r_k, Ar_k) =$$
$$= (r_k, Ax_k - b) + \alpha(r_k, Ar_k) = -(r_k, r_k) + \alpha(r_k, Ar_k)$$

Voglio $g'(\alpha) = 0$, quindi:

$$\boxed{\alpha = \frac{(r_k, r_k)}{(r_k, Ar_k)}}$$

giustamente questo valore è uguale a quello che avevamo trovato prima

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 185 di 314

(A) n° incognite = n° nodi interni + 1 nodo sul bordo

=> n° incognite = $(N-1)+1 = N$ => matrice $A_h \in \mathbb{R}^{N\times N}$

Formulaz. variazionale

$$\int_0^1 (-u'' - 2u')v = \int_0^1 fv$$

$$-\underbrace{u'v(1)}_{=0} + \underbrace{u}_{=2}v(0) + \int u'v' - 2\int u'v = \int fv$$

entrambe calcolate in 1

$$V = \{ v : \underbrace{v(1)=0}_{\text{pongo il vincolo solo nel nodo dove conosco la soluzione}} \quad \int (v')^2 < \infty \}$$

Trovare $u \in V$: $\underbrace{\int u'v' - 2\int uv}_{A_h} = \int fv - 2v(0) \quad \forall v \in V$

($\int uv$: termine di convez.)

$$(A_h)_{i,j} = \int \Phi_j' \Phi_i' - 2\int \Phi_j' \Phi_i$$

per $i=1$ => riga di Neumann $\quad \Phi_1 =$ 

$\int \Phi_1' \Phi_1' = \frac{1}{h}$ calcolati a lezione

$\int \Phi_1' \Phi_2' = -\frac{1}{h}$

Dobbiamo calcolare i termini relativi alla convezione:

$-2\int \Phi_1' \Phi_1 = -2\left(-\frac{1}{h}\right)\int \Phi_1 = \frac{2}{h}\frac{h}{2} = 1$ "ingrana" la diag. princ.

$-2\int \Phi_2' \Phi_1 = -2\left(\frac{1}{h}\right)\int \Phi_1 = -\frac{2}{h}\frac{h}{2} = -1$

$$A_h = \frac{1}{h}\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & 0 \\ & -1 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & 0 & & -1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 & & & \\ 1 & 0 & -1 & & 0 \\ & 1 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & 0 & & +1 & 0 \end{bmatrix}$$

Vettore b (1§1) :

$b_i = \int f\Phi_i - 2\Phi_i(0)$ $\quad \rightarrow \quad b_1 = \int f\Phi_1 = \frac{h}{2}f(0) = \frac{3}{2}h$ (mezzo triangolo!)

$\quad \rightarrow \quad b_{i\,(i\neq 1)} = \int f\Phi_i = 3h$

$$b_i = 3h\begin{bmatrix} 1/2 \\ 1 \\ \vdots \\ 1 \end{bmatrix} - 2\Phi_i(0), \quad \text{dove } \Phi_i(0) = \begin{cases} 1 & i=1 \\ 0 & i\neq 1 \end{cases}$$

quindi $\quad b_i = 3h \cdot \begin{bmatrix} 1/2 \\ 1 \\ \vdots \\ 1 \end{bmatrix} - 2 \cdot \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ $\rightarrow$

quindi:

$$b = 3h \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \left(\frac{2}{h} - 2\right) \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad i = 1, \dots N-1$$

N.B. qui devo ricordare l'effetto della condiz. di Dirichlet non omogenea e fare i calcoli attentamente!

Tema d'esame dell'11/02/2011 - es 1

$$\begin{cases} -u'' = 1 & x \in (-1, 1) \\ u(-1) = 0 \\ u(1) = 0 \end{cases}$$

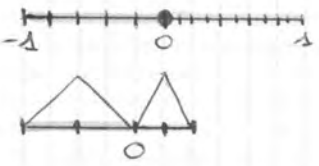
Diffusione pura
→ condiz. di Dirichlet omogenea

$[-1, 0]$ è diviso in N intervalli uguali → $h = \frac{1}{N}$
$[0, 1]$ " " " $2N$ " " → $h_2 = \frac{1}{2N} = \frac{h}{2}$

La griglia globale è data. Devo trovare:

1) n° incognite
2) A_h (in funz. di h!! cioè del parametro della griglia meno fitta)
3) stud. invertibilità
4) vettore di carico in funz. di h

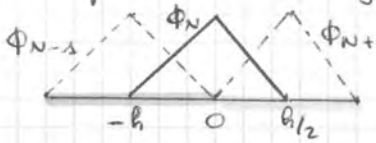
Com'è fatta la griglia?



incognite = modi interni = ~~N-1~~ + 2N-1 + (1) = 3N-1
(1) modo in 0 che è interno

- per i primi N-1 modi → matrice di rigidezza del filo elastico standard con $h = h_1 = 1/N$
- gli ultimi 2N-1 modi hanno una matrice di rigidezza del filo elastico con ampiezza di griglia $h/2$ (dove c'è h metto $\frac{h}{2}$)

→ il problema è la riga N che coincide con $x_N = 0$



la funz. Φ_N è asimmetrica!

- $\int_0^1 \Phi_N' \Phi_N' = \int_{0-h}^{0} \Phi_N' \Phi_N' + \int_0^{h/2} \Phi_N' \Phi_N' = \int_{0-h}^{0} \left(\frac{1}{h}\right)^2 + \int_0^{h/2} \left[\frac{1}{h/2}\right]^2 =$

  pendenza prima dello 0 ↑ ; pendenza dopo lo 0 ↑

$$= \frac{1}{h^2} h + \frac{4}{h^2} \frac{h}{2} = \frac{3}{h}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 189 di 314

Tema d'esame del 7/03/2011 - es 1

Diffusione - reazione

$$\begin{cases} -\Delta u + u = 1 & x \in \Omega = (0,1)^2 \\ u(x,y) = 0 & x \in \partial\Omega \end{cases} \qquad h = 1/N$$

Matrice di reazione di elemento:

$$(Q_R)_{i,j} = \int_{\hat{T}} \varphi_i \varphi_j = \frac{1}{24}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$\varphi_1 = 1 - \hat{x} - \hat{y}$
$\varphi_2 = \hat{x}$
$\varphi_3 = \hat{y}$

$$\int_\Omega (-\Delta u + u) v = \int_\Omega f v$$

già il termine di convezione (di ordine 1) non necessita di essere integrato per parti → questo ancora meno

$$\int_\Omega -v \cancel{\frac{\partial u}{\partial h}}^{=0} + \int_\Omega (\nabla u \nabla v + uv) = \int f v \quad \forall v \in V$$

$$V = \{ v : v|_{\partial\Omega} = 0 \text{ e } \int |\nabla u|^2 < \infty \}$$

Trovare $u \in V$ t.c. $\int_\Omega (\nabla u \nabla v + uv) = \int_\Omega f v \quad \forall v \in V$

Cambia la matrice di rigidità ma non lo spazio delle funzioni test, questo cambia solo se cambia l'ordine della derivata di ordine max (Piastra).

Formulazione FEM:

$$V_h = \{ v : v|_{\partial\Omega} = 0 \quad \in v|_{T_j} \in \mathbb{P}^1 \text{ continue} \}$$

Trovare $u_h \in V_h$ t.c.:

$$\int \nabla u_h \nabla v + \int uv = \int f v \quad \forall v \in V_h$$

$$u_h(x,y) = \sum_{\ell=1}^{(N-1)^2} c_\ell \, \phi_\ell(x,y)$$

$$A_h c = b$$

$$(A_h)_{k,\ell} = \int_\Omega (\nabla \phi_k \nabla \phi_\ell + \phi_k \phi_\ell)$$

$$b = \int f \phi_k$$

Matrice di reazione $A^R_{k,\ell} = \int \phi_k \phi_\ell$

$$\int_T \phi_k \phi_\ell \, dx\, dy = h^2 \int_{\hat{T}} \varphi_m \varphi_m \, d\hat{x}\, d\hat{y}$$

$dx = h\, d\hat{x}$
$dy = h\, d\hat{y}$

→

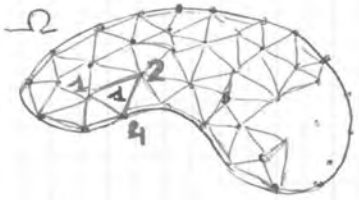
© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 191 di 314

24/01/2012

Come funziona il metodo FEM?

$$\begin{cases} -\nu \Delta u = f & \Omega \in \mathbb{R}^2 \\ u|_{\partial\Omega} = g \end{cases}$$

1) input: $\nu, f, g, \Omega, \partial\Omega$, Tol
2) costruisce la triangolaz.
3) calcola le matrici di rigidità e il vettore di carico di elemento per ogni triangolo a^j, — b^j $j = 1, \dots N$ (N = n° triangoli
4) assembla la matrice di rigidità e il vettore di carico globali
5) risolve il sistema lineare
6) calcola uno stimatore di errore e modifica la griglia se necessario (⇒ torna a 2)
7) stampa i risultati



Su Ω il sistema mette vari punti più o meno equidistanti. Usando la triangolazione di Delaunay si costruisce triangoli il più possibile equilateri △ (massimizza il cerchio inscritto)

Su ogni vertice finiscono vari triangoli → numero i vertici: N_i $i = 1, \dots N$ ed i triangoli: T_j $j = 1, \dots M$.
Costruisco un array che contenga l'indirizzo di ogni nodo

array per i nodi:
$E(2, i)$ $i = 1, \dots N$
$E(1, i) = x_i$
$E(2, i) = y_i$
(x_i, y_i) coordinate del nodo i

Per i triangoli:
array per i triangoli:
$T(3, j)$ $j = 1, \dots M$
$T(1, j)$ = vertice 1 del triangolo j
$T(2, j)$ = 〃 2 〃 〃 〃
$T(3, j)$ = 〃 3 〃 〃 〃

con riferim. alla fig:
$T(1,1) = 1$
$T(2,1) = 2$
$T(3,1) = 4$

Gradiente → N^2 (matrice del laplaciano, se trave elastica N^4)

Gradiente coniugato → N

" " precondizionato → $\sqrt{N}$

Multigrid → $C = cost$ (non dipende da N)

valgono tutti per matrice del laplaciano → SPD

Se matrice non SPD (es. convez. - diffusione) → c'è solo GMRES $<<N^2$

(però ogni iterazione diventa più pesante e più lunga da fare)

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 195 di 314

Contents

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 197 di 314

Chapter 1

Numbers, errors and norms

The real numbers are infinite, and of course we cannot think of representing (and storing) all of them on a computer. So the computer uses only a finite subset of the real numbers. This subset is composed of a finite number of numbers. Not only, but each number in this set must contain a finite information, and therefore it must have a finite number of digits. For this part, see [2, chapters 1, and 2].

1.1 Floating point numbers

Numbers can be represented in the form:

$$x = a_1 a_2 \dots a_t \dots \beta^n$$

where a_1, a_2 etc are the digits composing the number (which can be infinite), β is an integer and it is the basis of the numeration being used, and n is the exponent. In our usual computations $\beta = 10$. For a computer $\beta = 2$. The digits are integer numbers which go from 0 to $\beta - 1$, both of them included. This representation is not unique: $x = 123$ is the same as $x = 12.3 \times 10^1$. To eliminate this ambiguity, we choose to represent numbers in normalized form:

$$x = 0.a_1 a_2 \dots a_t \dots \beta^n, \qquad a_1 \neq 0$$

The floating point number system is obtained keeping only the numbers in normalized form that have a finite number t of digits and a value of the exponent n such that $L \leq n \leq U$, with $L < 0$ and $U > 0$, both of them finite. The sequence of digits $a_1 a_2 \dots a_t$ is called *mantissa*.

The total number of floating point numbers is roughly $2\beta^t(U - L + 1)$. A floating point system is determined when one specifies t, β, U and L. Another way of characterizing floating point numbers is by giving the amount of memory each number needs, and how it is used. Two floating point systems are common:

- Single precision: they use 32 bits. One bit is used for the sign, 8 bits for the exponent and 23 bits for the mantissa.
- Double precision: they use 64 bits. One bit is used for the sign, 11 bits for the exponent and 52 bits for the mantissa.

In scientific computing we can only hope to be able to bound the relative error between the computed and the exact result. This implies that everytime we wish to stop a process when the computed answer is close enough to the exact result, we must set up the stopping criterion using the relative error between the computed and the exact result.

In double precision, $\beta = 2$, $t = 52$, thus the bound on the relative error is:

$$\frac{|x - \tilde{x}|}{|x|} \leq \frac{1}{2}\beta^{1-t} = 2^{-52} = 2.2 \cdot 10^{-16} = \texttt{eps} \tag{1.1}$$

The number on the right is called *machine precision* or in short eps, and it is characteristic of the number system used on the computer. It represents the maximum relative precision we may hope to attain performing operations on the computer. Note also that relative errors do not depend on the units of measure used during the computation.

1.2 Measuring vectors

In many scientific applications the output of our program is a list with an impressive amount of numbers (typically, millions, and for 3D computations there can be billions of numbers). It is clear that looking at those numbers will not help much in deciding how the solution will look like. It is important therefore to process those numbers, for instance with graphical tools. But this usually only gives qualitative information.

Suppose we can compute the exact solution for some problem and we wish to compare it with the computed solution. Since the computed solution is known as a list of numbers, we actually want to compare two lists of numbers. We can compute the error, subtracting these two lists of numbers. The advantage is that now we have a single, huge list of numbers. We can look at all of them, and decide whether they are small or not. But this is time consuming, and not very decisive. What we want is to express how large the list of numbers is expressing its "length" with a single number.

So we need to associate a single number to our list, and this number should tell us how "large" the list is. To express this mathematically, we arrange the numbers in the list in a vector, and each number of the list will be a component of the vector. So the problem now is how to compute the length of a vector.

It turns out there are several reasonable ways of doing it. In particular, at the very least, since we are interested in measuring errors,

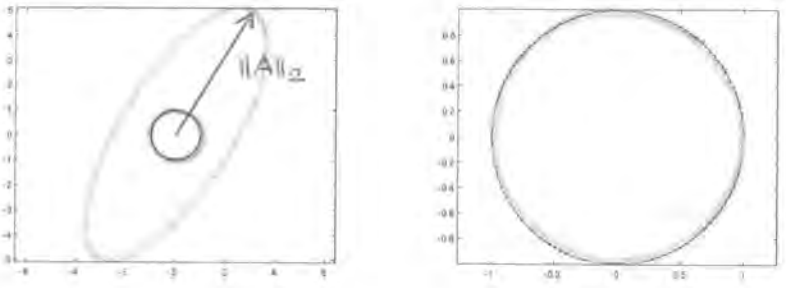
1. We want the global error to be exactly zero only if all the components of the error are zero. And viceversa.

2. We want the global error to be small only if all or most of the components of the error are small. That is we want some sort of continuity between the length of a vector and its components.

The concept of **norm** in mathematics satisfies these requests and more. The norm of a vector is a function, applied to the vector and returning a real, non negative number. It is quite a general, but very precise concept, as often happens in mathematics.

The definition of norm is the following:

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 201 di 314



(A_1) (A_2)

Figure 1.1: Transformation of the unit circle (blue line) under the action of matrices A_1, and A_2 (green markers). The norm of A in the $p = 2$ case is the maximum distance of the green curve from the origin. Since the matrix A_2 is orthogonal, it just rotates the unit circle, and the green markers are superposed on the blue line.

The norm of a matrix is defined as the maximum amplification that a matrix operates on a vector. In symbols:

$$|||A|||_p = \max_{x \neq 0} \left(\frac{||Ax||_p}{||x||_p} \right) \tag{1.5}$$

Note that $x/||x||_p$ is a vector with p norm equal to 1. Therefore, using the properties of a linear operator, we can re-write the definition above in the following equivalent form:

$$|||A|||_p = \max_{x:||x||_p=1} \left(||Ax||_p \right). \tag{1.6}$$

Let us illustrate the concept of matrix norm using 2×2 matrices, for which we can visualize the results. In the following, I will show results only for the 2-norm. Let us consider the four matrices:

$$A_1 = \begin{bmatrix} 2 & 3 \\ -1 & 5 \end{bmatrix}, \quad A_2 = \begin{bmatrix} \frac{1}{2}\sqrt{3} & -0.5 \\ 0.5 & \frac{1}{2}\sqrt{3} \end{bmatrix}, \quad A_3 = \begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0.4 & -0.1 \\ 0.4 & 0.9 \end{bmatrix}$$

The matrix A_1 is a generic matrix with two complex eigenvalues. Its norm is $|||A_1||| = 5.83$. The matrix A_2 is an orthogonal matrix. It is a rotation matrix, in the sense that $A_2 x$ is a vector, which is as long as x, and it is rotated by an angle θ with respect to x. All two by two orthogonal matrices have the form:

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

In our example, A_2 is an orthogonal matrix that rotates by an angle $\theta = \pi/6$. Recall that for all orthogonal matrices $Q^{-1} = Q^T$, because Q^T represents a rotation by the angle $-\theta$. Therefore $QQ^T = I$, where I is the identity matrix. Since orthogonal matrices do not modify the length of a vector, in the 2-norm, $|||Q|||_2 = 1$. We find the effect of the matrices A_1

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 203 di 314

Therefore, to compute the 1-norm of a matrix, we choose a column, and we compute the sum of all the absolute values of the elements of A on that column. We repeat this computation for all the columns of the matrix, and we pick the maximum number we obtain in this fashion. We apply the same process to the rows of a matrix to compute the infinity norm of a matrix. For the 2-norm, we have that:

$$|||A|||_2 = \max_j \sigma_j, \tag{1.10}$$

where σ_j is the generic singular value of the matrix A. If the matrix is symmetric, the singular values coincide with the absolute values of the eigenvalues and therefore we find that this expression coincides with (1.7).

Chapter 2

Linear conservation laws

In this chapter we will begin our study of hyperbolic problems, starting from the simple case of linear equations. Linear conservation laws have the form:

$$u_t + a(x,t)u_x = 0, \tag{2.1}$$

where $a(x,t)$ is a known function, and $u(x,t)$ is the unknown. The equation is linear, because the following two conditions are satisfied:

1. if $u(x,t)$ and $v(x,t)$ are two solutions of (2.1), the sum $u(x,t)+v(x,t)$ is also a solution;
2. if $u(x,t)$ is a solution of (2.1), and k is a real number, than $k\,u(x,t)$ is also a solution of (2.1).

2.1 We get by with the help of characteristic curves

To compute solutions of (2.1), we introduce the notion of *characteristic curve*. Consider a curve γ in the plane (x,t), parametrized with the variable t. Then any point on the curve has coordinates $(x(t),t)$, for a given function $x(t)$. We now consider the unknown function u, and we compute its derivative along the curve γ:

$$\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_\gamma = \frac{\partial u}{\partial x}\left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_\gamma + \frac{\partial u}{\partial t} = \frac{\partial u}{\partial t} + \left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_\gamma \frac{\partial u}{\partial x}$$

Comparing the expression just computed with the equation we're trying to solve, we see that the linear conservation law can be written as:

$$\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_\gamma = 0, \qquad \text{on} \quad \frac{\mathrm{d}x}{\mathrm{d}t} = a(x,t). \tag{2.2}$$

This means that the solution u remains constant along the curve γ, defined by $\frac{\mathrm{d}x}{\mathrm{d}t} = a(x,t)$. The curves defined in this way are called characteristics, and the equation states that u remains constant along characteristics.

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 207 di 314

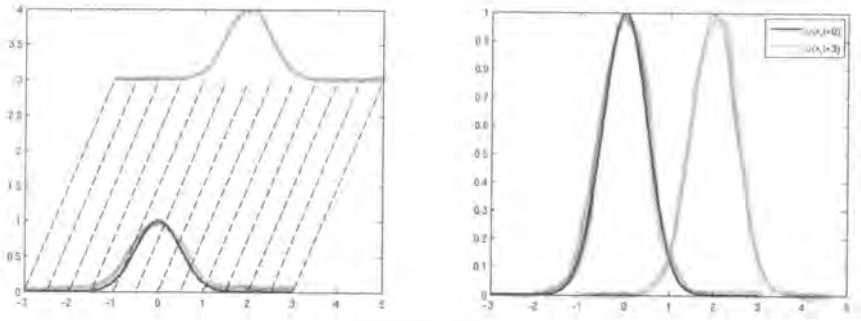


Figure 2.2: Signal moving with speed a for the linear advection equation, with a constant.

Solution. The exact solution is $u(x,t) = e^{-2(x-\frac{2}{3}t)^2}$. Fig. 2.2 shows the initial data (in blue) which move along the characteristics (dashed lines). At time $t = 3$, the signal has moved (green curve) without changing its shape. The right part of the figure shows the solution at $t = 0$ (in blue) and the solution at $t = 3$ (in green).

■

2.2 Initial and boundary value problems

In many applications, we are given an initial-boundary value problem. Usually, this happens when the phenomenon we are studying occurs on a finite domain. In this case data are prescribed not only at $t = 0$, but also on some portion of the boundary. Again, the problem can be solved with the method of characteristics. This means that we compute the equation for the characteristics, than we pick a point (x,t), which selects a single characteristic curve, and we prolong the characteristic backward in time, until we intersect the boundary along which data are prescribed. We illustrate this problem with an example.

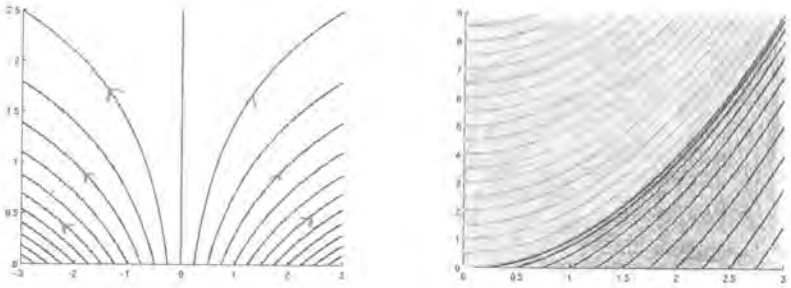
Example 2.2.1. *Consider the initial boundary value problem*

$$\begin{aligned} &u_t + au_x = 0 \\ &u(x, t = 0) = u_0(x) \quad \text{for } x \in [x_1, x_2] \\ &u(x = x_1, t) = g(t). \end{aligned}$$

with $a > 0$.

Solution. Again all characteristics are straight lines, with the same slope. Choose a point (x,t) and prolong the corresponding characteristic backward in time. There are two possibilities. One possibility is that the characteristic intersects the axis $t = 0$ in the point $x_0 \in [x_1, x_2]$. In this case, the solution is again $u(x,t) = u_0(x - at)$. The other possibility is that $x_0 \notin [x_1, x_2]$. Since $a > 0$, we have that $x_0 < x_1$. For this value of x_0 we do not have prescribed data, but certainly the characteristic will intersect the axis $x = x_1$ in some point $t_0 > 0$, and the point t_0 is given by $x - x_1 = a(t - t_0)$. Therefore, in this second case, the solution will be $u(x,t) = g(t_0)$ with $t_0 = t - (x - x_1)/a$.

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 209 di 314



$(u_t + xu_x = 0)$ $(u_t + \frac{1}{x}u_x = 0)$

Figure 2.4: Initial-boundary value problems with non constant a. On the left, characteristic field for $u_t + xu_x = 0$. In this case, the solution is completely determined by the initial data. On the right, $u_t + \frac{1}{x}u_x = 0$. In this case, boundary data must be prescribed on $t > 0$.

by separation of variables:

$$\frac{\mathrm{d}x}{x} = \mathrm{d}t \quad \Rightarrow \ln(|x|) - \ln(|x_0|) = t - t_0.$$

The characteristic field is shown in Fig. 2.4 for $-3 < x < 3$, on the left. Note that all characteristics exit from the axis of initial data and enter the two boundaries $x \equiv -3$ and $x \equiv 3$. Thus the only data that determine the solution are prescribed in $(-3 < x < 3, t = 0)$. Data cannot be prescribed on $x \equiv -3$ and $x \equiv 3$: they would be carried by the flow outside the computational region. Therefore $t_0 = 0$, and the value where each characteristic intersects the initial value axis is $x_0 = xe^{-t}$, and the solution is $u(x,t) = u_0(x_0) = u_0(xe^{-t})$.
■

Example 2.2.3. *Consider the initial value problem:*

$$\begin{aligned} &u_t + \tfrac{1}{x}u_x = 0 \qquad in \quad (0 < x < 3) \times (t \geq 0) \\ &u(x, t = 0) = 2 \quad 0 < x < 3 \\ &u(x = 0, t) = -1 \quad t \geq 0. \end{aligned}$$

Solution. Again we solve the equation for the characteristics by separation of variables.

$$x\mathrm{d}x = \mathrm{d}t \quad \Rightarrow x^2 - x_0^2 = t - t_0.$$

The characteristic field is shown in Fig. 2.4 for $0 < x < 3$, on the right. The blue characteristics denote the curves which intersect the axis of initial data, which occurs for the curves that lie below the characteristic passing through the origin, i.e. $t = x^2$, while the green curves are characteristics exiting form the boundary data curve, in this example, $(x = 0, t \geq 0)$. So for $x > t^2$, we are in the region where the solution is determined by the initial data, while for $x < t^2$, the solution depends on the boundary data. Therefore:

$$u(x,t) = \begin{cases} -1 & x \leq t^2 \\ 2 & x > t^2 \end{cases}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 211 di 314

Chapter 3

Non linear scalar conservation laws

In this chapter, we study equations of the form:

$$u_t + f_x(u) = 0, \tag{3.1}$$

where f is a smooth function of $u(x,t)$. The function f is called flux, and the equation is non linear if f depends on u non linearly. In fact the linear problem with constant a can be written in the form (3.1) with $f(u) = au$. In non linear problems, the solution of the partial differential equation becomes much more complex than in the linear case. The main difficulty is due to the fact that solutions of (3.1) can develop singularities in a finite time, and this will force us to extend the notion of solution.

3.1 Conservation laws in integral form

A good starting point is to try to have an idea of what is the origin of these equations. We consider a highly simplified example, which however contains a lot of information from the general case. Suppose you are given a long pipe, with a uniform cross section of area A. Let x denote the coordinate parallel to the axis of the pipe. In this pipe, some gas is contained with density u, and we will suppose that the density of the gas depends only on x and t. We will also suppose that the gas moves along the pipe with a speed v which may depend on u, and suppose we know the functional dependence of v on u. Now, we consider just a section of the pipe, with extrema $x = a$ and $x = b$. The total mass of the gas contained in this section of the pipe at time t is:

$$M(t) = A \int_a^b u(x,t)\, \mathrm{d}x.$$

Note that the total mass depends only on time. Since the gas flows in the pipe, the mass contained in the section we are studying may change with time. But it should be clear that the mass can change only if there is mass flowing in or out from the end points of the slice of pipe we are considering. No mass is created or destroyed within the pipe. In a small interval

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 213 di 314

only if g is continuous. Thus, the equation (3.3) is equivalent to (3.2) only if u admits *continuous* partial derivatives. On the other hand, (3.2) admits solutions which do not necessarily satisfy such strict requirements of smoothness. Since the steps we have performed to get (3.3) starting from (3.2) can be performed backward, we have proved that all solutions of (3.3) are also solutions of (3.2), but the converse is not true. In other words, the integral form of the continuity equation admits more solutions than the differential form. If we compare the set of solutions of an equation to a club, we can say that the club composed of solutions of (3.2) is less picky than the club formed by solutions of (3.3). In mathematical terms, we say that a solution of eq. (3.2) is a *weak solution*, while solutions of eq.(3.3) are *strong solutions*. In fact, the differential equation (3.3) means that the expression $\partial_t u + \partial_x f(u)$ must be zero at *all possible points* (x,t), while the integral form is not so picky. We will see below that the integral form of the conservation law accepts solutions with jumps, but it dictates the speed at which such jumps move.

3.2 Characteristics again

In this section, we will suppose that the solution u of (3.1) is smooth. Thus, using the chain rule to compute $\partial_x f(u)$, we can rewrite the equation in the form:

$$u_t + a(u)u_x = 0 \qquad \text{where} \quad a(u) = f'(u). \tag{3.4}$$

This equation is of the form (2.2), except that now the speed a depends on the solution u. We still have that

$$\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_\gamma = 0, \qquad \text{with } \gamma: \quad \frac{\mathrm{d}x}{\mathrm{d}t} = a(u) \tag{3.5}$$

and again u remains constant along the characteristic, but now the slope of the characteristic depends on u. Still, since u remains constant along γ, also $a(u)$ remains constant. Thus the slope of the characteristic is fixed, and therefore the characteristic is a straight line, as in the linear case for constant a. Suppose we are given the initial value problem

$$\begin{cases} u_t + f_x(u) = 0 \\ u(x, t=0) = u_0(x), \end{cases} \tag{3.6}$$

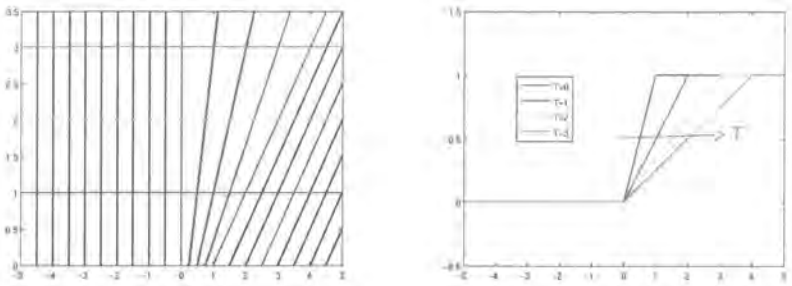
Therefore, the solution of the initial value problem is

$$u(x,t) = u_0(x_0) \qquad \text{on} \quad x - x_0 = a(u)t = a(u_0(x_0))t. \tag{3.7}$$

This equation still defines x_0, and therefore permits to solve the conservation law, determining u. However, the equation (3.7) in general will be non linear and it gives x_0 implicitly. Therefore it is not possible to compute x_0 in closed form, except in a few very particular cases.

The slope of the characteristic depends on $u_0(x_0)$, which in general is not constant. Thus, the characteristics are straight lines, but their slope changes. The first consequence of this important fact is that each value $u_0(x_0)$ travels with a different speed. Thus, the initial signal changes shape with time, that is the signal is *distorted*. This is a typical non linear effect.

There are two main possibilities: either the characteristics diverge, in which case the distance between two values of $u_0(x_0)$ increases with time, or the characteristics tend to



(Characteristic diagram) (Solution)

Figure 3.1: Rarefaction wave resulting from an upward wedge initial data, from example 3.2.1. Left: characteristic diagram with the rarefaction fan (blue). The horizontal lines correspond to the instants in which the solution is shown (right)

and the solution here is

$$u(x,t) = 1 \quad \text{for} \quad x - t < 0.$$

If $x_0 > 1$:

$$u_0(x_0) = 0 \quad \Rightarrow \quad x - x_0 = 0 \quad \Rightarrow x_0 = x,$$

and the solution in this region is

$$u(x,t) = 0 \quad \text{for} \quad x > 1.$$

In the intermediate region, $0 \le x \le 1$,

$$u_0(x_0) = 1 - x_0 \quad \Rightarrow \quad x - x_0 = (1 - x_0)t \quad \Rightarrow \quad x_0 = \frac{x-t}{1-t}.$$

Here

$$u(x,t) = 1 - x_0 = \frac{1-x}{1-t} \quad \text{for} \quad 0 \le x_0 \le 1 \quad \Rightarrow \quad t \le x \le 1.$$

Note that to solve the inequality $0 \le x_0 \le 1$, I assumed $1 - t > 0$, which makes sense, because initially $t = 0$. The global solution is:

$$u(x,t) = \begin{cases} 1 & x < t \\ \frac{1-x}{1-t} & t \le x \le 1 \\ 0 & x > 1. \end{cases}$$

This solution is still continuous, but it is defined only for $t < 1$. As $t \to 1^-$, $\partial_x u|_{(1,t)} \to -\infty$ and the whole construction breaks down, because all characteristics bring to the point $(1, 1)$ conflicting values. Before breakdown, the solution is well defined, the values of $|\partial_x u|$ grow with time in the intermediate region, which is a compression wave. The characteristic diagram and the solution at different times are shown in Fig. 3.2. ■

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 217 di 314

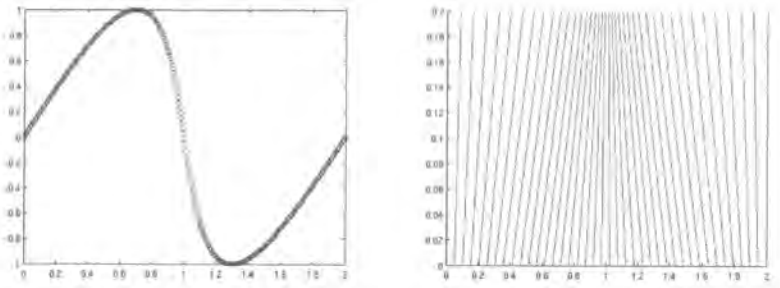


Figure 3.3: Solution of Burgers equation with $u_0(x) = \sin(\pi x)$ on [0,2] at $t = 0.2$ (left) and corresponding characteristic field for $0 \leq t \leq 0.2$ (right)

breaks down for $u_0' > 0$, and if that happens $\partial_x u \to +\infty$. It is much harder to predict the time of breakdown if the flux can change concavity, as it happens for multiphase flows.

What happens when characteristics intersect? Surely in the first instant in which $|\partial_x u(x,t)|$ becomes unbounded, the equation $u_t + f'(u)u_x = 0$ cannot be interpreted in the usual sense, because this equation requires well defined first derivatives. However the construction $u(x,t) = u_0(x_0)$ with x_0 given implicitly by the equation $x - x_0 = f'(u_0(x_0))t$ can still be carried out. The only problem is that when two or more characteristics cross, this equation has more than one solution for x_0 and consequently $u(x,t)$ becomes multiple-valued. We illustrate what happens with an example.

Example 3.2.3. *Consider the initial value problem for Burgers' equation:*

$$\begin{aligned} &\partial_t u + \partial_x(\tfrac{1}{2}u^2) = 0 \qquad on \quad [0,2] \\ &u_0(x) = \sin(\pi x) \end{aligned}$$

with periodic boundary conditions.

Solution. Using (3.8), we see that the x derivative of u becomes unbounded for $t = -1/(u_0' f'')$, and the first instant at which this occurs is $t_s = \min_x(-1/u_0')$, since in this case $f'' = 1$. For the particular initial data I have chosen $t_s = 1/\pi \simeq 0.318$. Fig. 3.3. shows the solution at $t = 0.2$, and the corresponding characteristic field up to that time. It is already apparent that the initial sine function is steepening at $x = 1$, and in fact the characteristics thicken around $x = 1$. As $t \to t_s$, the profile gets even steeper, and the characteristics tend to cross, see Fig 3.4. If instead $t > t_s$, as in Fig. 3.5 the solution obtained with the characteristics becomes multiple valued, and the characteristic diagram shows the intersecting characteristics which determine the multiple-valued solution. The vertical green line signals the initial breakdown position. Note that the two regions formed by the green line have the same area. ■

In some situations, the solution shown in Fig. 3.5 makes sense. Think for instance of a water wave breaking on the shore. Just before breaking, the wave becomes steeper and

3.3 Shock formation

To continue the solution of a conservation law beyond the breakdown of the method of characteristics, we consider again the conservation law in integral form (3.2), and we suppose that the solution contains a single jump discontinuity which at time t is located at $x = s(t)$, with $a < s(t) < b$. This means that the solution is smooth except for this single jump. Then eq. (3.2) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_a^{s(t)} u(x,t)\,\mathrm{d}x + \int_{s(t)}^b u(x,t)\,\mathrm{d}x\right) = f(u(a,t)) - f(u(b,t)). \tag{3.9}$$

Let us denote with $u_L(t)$ and $u_R(t)$ the left and the right values of u across the jump, i.e.

$$u_L(t) = \lim_{x \to s^-(t)} u(x,t)$$
$$u_R(t) = \lim_{x \to s^+(t)} u(x,t)$$

To procede, we should remember how to differentiate under the integral sign. In particular, remember that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u(x,t)\,\mathrm{d}x = \int_a^b \partial_t u(x,t)\,\mathrm{d}x, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{s(t)} u(x)\,\mathrm{d}x = u(s(t))\frac{\mathrm{d}}{\mathrm{d}t}s(t) = u(s(t))s'(t).$$

In our case, we have both contributions, because u depends on (x,t) and one of the extrema of each integral depends on t. Computing the derivatives on the left hand side of (3.9), we find:

$$\int_a^{s(t)} \partial_t u(x,t)\,\mathrm{d}x + u(s^-(t),t)s'(t) + \int_{s(t)}^b \partial_t u(x,t)\,\mathrm{d}x - u(s^+(t),t)s'(t) = f(u(a,t)) - f(u(b,t)). \tag{3.10}$$

Now, we wish to zoom on the discontinuity, so we let $a \to s^-(t)$ and $b \to s^+(t)$. Then the two integrals on the left go to zero, because they are the integrals of a smooth function over a vanishing interval. Since the flux is a smooth function of u, we will have:

$$\lim_{a \to s^-(t)} f(u(a,t)) = f(u_L(t))$$
$$\lim_{b \to s^+(t)} f(u(b,t)) = f(u_R(t)).$$

Therefore zooming on the discontinuity, equation (3.9) reduces to

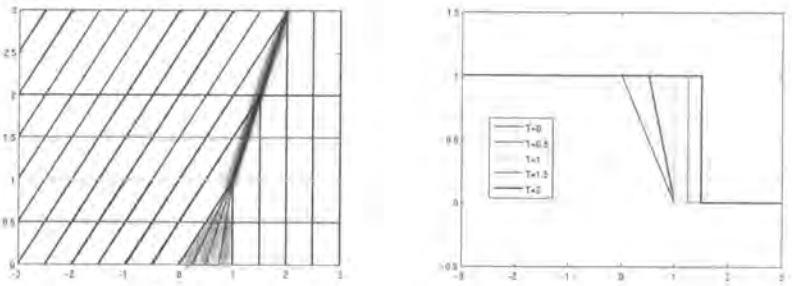
$$u_L(t)s'(t) - u_R(t)s'(t) = f(u_L(t)) - f(u_R(t)). \tag{3.11}$$

Rearranging terms, we see that the equation in integral form (3.9) accepts jumps between two states u_L and u_R, as long as these jumps move with speed

$$s'(t) = \frac{f(u_L(t)) - f(u_R(t))}{u_L(t) - u_R(t)}. \tag{3.12}$$

Therefore, we use the method of characteristics as long as the solution remains single valued, which means that characteristics do not cross. Just before characteristics intersect, $|\partial_x u|$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 221 di 314



(Characteristic diagram) (Solution)

Figure 3.7: Compression wave and shock resulting from a downward wedge initial data, from example 3.3.1. Left: characteristic diagram with the compression fan (blue) and the shock trajectory (blue, thick line). The horizontal lines correspond to the instants in which the solution is shown (right)

The solution together with the characteristic diagram is shown in Fig. 3.7, for $t \in [0, 3]$. Recall that the time of breakdown is $t = 1$. ■

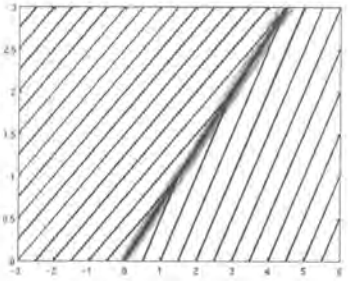
3.4 Entropy conditions

In the previous section, we have constructed solutions with discontinuities which go beyond the classical solutions built with the method of characteristics, for conservation laws of the form (3.3). In fact, since discontinuous solutions are not differentiable, we cannot substitute such solutions in the differential equation for the simple fact that we cannot compute the derivative of a function with a jump. Thus, admitting discontinuous solutions, we have actually *enlarged* the set of possible solutions of the conservation law. Mathematically, one says that discontinuous solutions satisfying the Rankine-Hugoniot condition are *weak solutions* of the conservation law. The problem is that if we consider solutions which are piecewise smooth, with jumps satisfying the Rankine-Hugoniot conditions, we end up with too many solutions, because one can construct more than one weak solution for the same initial value problem. In other words, weak solutions are not unique. Therefore we need extra conditions to get rid of those weak solutions which do not have a physical meaning. In fact, from a physical point of view, we expect that each well posed initial value problem will have just one solution.

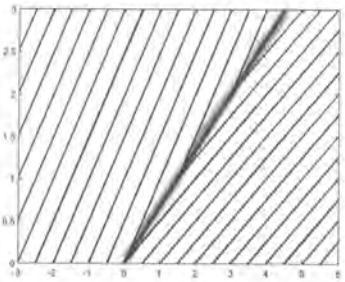
A couple of examples will clarify this point.

Example 3.4.1. Entropic and non entropic shocks
Consider the following initial value problems for Burgers' equation

$$1) \quad u(x,t=0) = \begin{cases} 2 & x < 0 \\ 1 & x > 0 \end{cases} \qquad 2) \quad u(x,t=0) = \begin{cases} 1 & x < 0 \\ 2 & x > 0 \end{cases}$$



(Entropic shock)



(Non-entropic shock)

Figure 3.9: Chracteristic diagrams for two step-like initial data, from example 3.4.1. The shock trajectory is the thick blue line. Left, characteristic diagram for an entropic shock: the characteristics *enter into* the shock. Right, characteristic diagram for a non-entropic shock: the characteristics *leave from* the shock.

The problem of the existence of more than one solution then occurs only for the initial data 2): how should we choose the correct weak solution in this case? Again, we can think of perturbing slightly the initial data, adding a thin layer separating the two states u_L and u_R. In this way the initial data is continuous and we can use the method of characteristics. The situation is very similar to example 3.2.1, and we would obtain a rarefaction wave. No way of obtaining a shock in this case, because the characteristics tend to separate.

Thus the rarefaction wave 3) computed in the example above is stable under small perturbations of the initial data, while the shock wave 2) is not, because it disappears under small perturbations in the initial data. For this reason, we choose as unique solution of Problem 2), in the example above, the rarefaction wave, and we discard the shock. The unique solution for Problem 2) is shown in figure 3.8, together with its characteristic diagram.

It is very easy to know which jumps will give rise to stable shocks. If we look at the characteristic diagrams for the stable and the unstable shock shown in Fig. 3.9, we see that the stable shock has characteristics *entering into* the shock, while in the other case the characteristics *exit from* the shock. In the first case, the characteristics tend to cross, and this is avoided inserting the shock. In the second case, the characteristics tend to diverge, and the method of characteristics must be used.

A stable shock is called an *entropic shock*, and it is a kind of solution which is actually observed in experiments. An unstable shock is called *non entropic* and will not be observed in experiments, because, since it is unstable, it would immediately degenerate becoming a rarefaction wave.

It is also very easy to characterize entropic shocks with an algebraic condition. From Fig. 3.9 we clearly see that the shock is entropic provided that:

$$f'(u_L(t)) \geq s'(t) \geq f'(u_R(t)). \tag{3.13}$$

This condition is called *entropy condition*. Note that when the solution obtained with the method of characteristics breaks down, because characteristics cross, the entropy condition

Chapter 4

Methods for linear advection equations

Integrating a linear conservation law is harder than you would think. In this chapter, I want to illustrate the main difficulties one encounters when discretizing a linear conservation law of the type:

$$u_t + a u_x = 0, \tag{4.1}$$

where a is a given constant. Moreover, in the non linear case

$$u_t + f_x(u) = 0 \tag{4.2}$$

we find mostly the same difficulties, so that many of the techniques we will describe for the linear case carry over to the more general case. What we will miss in the general case are the proofs we are able to carry out here. Most of the material in this chapter can be found in [?].

The need to account for discontinuous solutions is one of the main sources of difficulties in the numerical integration of conservation laws. This problem can also be found in the linear case. If the initial data is smooth, the solution remains smooth for all times, because it does not change its shape. However, an initial solution presenting discontinuities propagates these discontinuities along characteristics. So, even in the linear case, we need numerical techniques that can deal with discontinuous solutions.

4.1 Elementary schemes

We cover the computational domain with points of the form (x_j, t^n), with $x_j = jh$ and $t^n = n\Delta t$, where h and Δt are the mesh widths in space and time respectively, which we will suppose constant for simplicity. The point $x_{j+1/2} = x_j + h/2$ is the location of the cell interface, while $I_j = (x_{j-1/2}, x_{j+1/2})$ is the cell centered around the grid point x_j. In the following, U will always denote the numerical solution, while u will be the exact solution. We will write $u_j^n = u(x_j, t^n)$.

Elementary numerical schemes are derived substituting the partial derivatives appearing

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 227 di 314

In the applications, we will actually use a mix of these two formulas, constituting the Upwind scheme:

$$U_j^{n+1} = U_j^n - a\frac{\Delta t}{h}\begin{cases} U_j^n - U_{j-1}^n & \text{if } a > 0 \\ U_{j+1}^n - U_j^n & \text{if } a < 0. \end{cases} \tag{4.7}$$

In the Upwind scheme, the space derivatives are computed using information from the left if information comes from the left, that is if the propagation speed $a > 0$. If $a < 0$ instead, information comes from the right, and the scheme modifies the differencing formula accordingly.

All schemes written above are first order schemes. To improve accuracy, the discretization of the time derivative must be improved. In Lax Wendroff scheme, this is obtained exploiting the equation itself. Since $u_t = -au_x$, $u_{tt} = a^2 u_{xx}$. So:

$$\begin{aligned} \frac{u_j^{n+1} - u_j^n}{\Delta t} &= \partial_t u|_j^n + \frac{\Delta t}{2}\partial_{tt}^2 u|_j^n + O(\Delta t)^2 \\ &= \partial_t u|_j^n + \frac{\Delta t}{2}a^2\partial_{xx}^2 u|_j^n + O(\Delta t)^2 \\ &= \partial_t u|_j^n + \frac{\Delta t}{2h^2}a^2\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right) + O(\Delta t)^2 + O(h)^2, \end{aligned}$$

where the second derivative $\partial_{xx}^2 u$ has been approximated with a central formula. Using the formula above to discretize the time derivative, and (4.3) to approximate the convective term $a\partial_x u$, we can rewrite the conservation law as:

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = -a\frac{1}{2h}\left(u_{j+1}^n - u_{j-1}^n\right) + \frac{\Delta t}{2h^2}a^2\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right) + O(\Delta t)^2 + O(h)^2.$$

Disregarding the high order error terms, we get the second order accurate (in space and time) Lax Wendroff scheme:

$$U_j^{n+1} = U_j^n - a\frac{\Delta t}{2h}(U_{j+1}^n - U_{j-1}^n) + \frac{\Delta t^2}{2h^2}a^2\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right). \tag{4.8}$$

Applying the same ideas to schemes based on one-sided derivatives, we find the second order (in space and time) Beam-Warming scheme:

$$U_j^{n+1} = U_j^n - a\frac{\lambda}{2}\begin{cases} 3U_j^n - 4U_{j-1}^n + U_{j-2}^n - \lambda a(U_j^n - 2U_{j-1}^n + U_{j-2}^n) & \text{if } a > 0 \\ -3U_j^n + 4U_{j+1}^n - U_{j+2}^n - \lambda a(U_j^n - 2U_{j+1}^n + U_{j+2}^n) & \text{if } a < 0. \end{cases} \tag{4.9}$$

Lax Friedrichs and Lax Wendroff schemes are symmetric with respect to the propagation speed a. They are the prototypes of *Central schemes*. Upwind and Beam Warming are based on asymmetric formulas for the space derivatives: they are the prototypes of *Upwind schemes*.

In real computations, the computational domain is finite, and we must assign boundary conditions. For simplicity, in what follows we will use periodic and free flow boundary conditions. Even in these simplified cases, some care should be used. In real problems, we will always deal with a finite computational domain, which we will denote with the interval $[a, b]$. Suppose the grid has amplitude h, then we have $m = (b - a)/h$ grid points. I place the first point x_1 in $a + \frac{h}{2}$. Thus the last point on which the numerical solution will be computed is $x_m = b - \frac{h}{2}$.

which denotes the product of the j-th row of the matric $\mathcal{H}$ with the vector $\mathbf{U}^n$. Note that for all schemes mentioned, the matrix $\mathcal{H}$ is very sparse. This reflects the finite speed of propagation of the underlying equation. Since information travels with a finite speed in the exact model, the numerical solution at one grid node at the time $t^n + \Delta t$ will depend only on its close neighbors at the time t^n, if Δt is small enough.

We now wish to study the convergence of these schemes. To compare the numerical solution, which is defined only on the grid, with the exact solution which is a function defined everywhere, we extend the numerical solution in order to transform it into a function:

$$U_h(x,t) := U_j^n \quad \text{for} \quad (x,t) \in V_j^n = (x_j - \tfrac{h}{2}, x_j + \tfrac{h}{2}) \times [t^n, t^{n+1}). \tag{4.13}$$

In this fashion, $U(x,t)$ is a piecewise constant function with jumps at the cell edges and along the lines $t = t^n$. To compare the exact and the numerical solutions, we also need to extend the operator $\mathcal{H}_h$ so that it can act on a function. It is easier than it sounds. The expression

$$U_h(x, t + \Delta t) = \mathcal{H}_h\left(U_h(\cdot, t); x\right)$$

means that the numerical solution $U_h(x, t + \Delta t)$ at the time $t + \Delta t$ in the point x has been obtained applying the numerical scheme to the function U_h at the time t, and evaluating the result in the point x. Now we can apply $\mathcal{H}_h$ to any function. For the Lax-Friedrichs' scheme the action of $\mathcal{H}_h$ on the numerical solution U_h can be written as

$$\begin{aligned} U_h(x, t + \Delta t) &= \mathcal{H}_h\left(U_h(\cdot, t); x\right) \\ &= \frac{1}{2}\left(U_h(x+h, t) + U_h(x-h, t)\right) - \frac{\lambda a}{2}\left(U_h(x+h, t) - U_h(x-h, t)\right). \end{aligned} \tag{4.14}$$

With this machinery, we can now introduce the local truncation error, with the same approach usually used in Ordinary Differential Equations. Namely, we suppose that at the time t the numerical and the exact solutions coincide (localization assumption: $u(x,t) = U_h(x,t)$), and we compute the error introduced in a single time step, dividing by Δt:

$$\begin{aligned} \mathcal{L}_h(t) &= \frac{1}{\Delta t}\left(u(x, t + \Delta t) - U_h(x, t + \Delta t)\right) \\ &= \frac{1}{\Delta t}\left[u(x, t + \Delta t) - \mathcal{H}_h\left(U_h(\cdot, t); x\right)\right] \\ &= \frac{1}{\Delta t}\left[u(x, t + \Delta t) - \mathcal{H}_h\left(u(\cdot, t); x\right)\right]. \end{aligned} \tag{4.15}$$

Thanks to the localization assumption, the local truncation error depends only on the exact solution. Suppose the exact solution is smooth. Then we can use Taylor expansion to evaluate how large the truncation error is. We illustrate this technique computing the truncation error for the Lax-Friedrichs' scheme.

Example 4.2.1. Truncation error for the Lax-Friedrichs' scheme.
Compute the truncation error for the Lax-Friedrichs' scheme.

Solution. Suppose the exact solution is smooth. We substitute the expression for the Lax-Friedrichs' scheme (4.4) in the definition of the truncation error (4.15), using the form

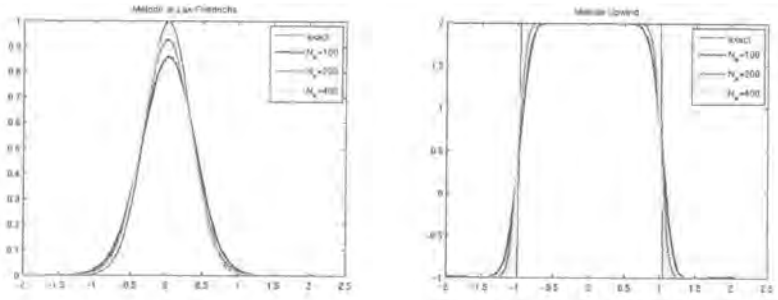


Figure 4.1: Convergence of numerical solutions under grid refinement. Left: smooth case; right: non smooth case

large a function is. In particular, we will be interested in knowing how large the error is, i.e. the difference between the exact and the numerical solutions:

$$e_h(t) = ||u(\cdot, t) - U(\cdot, t)|| \tag{4.19}$$

where $||\cdot||$ denotes a reasonable norm, and the notation $||u(\cdot, t)||$ means that the norm is computed along x, for a fixed time t. For instance, one could have the infinity norm:

$$||u(\cdot, t) - U(\cdot, t)||_\infty = \mathrm{Sup}_x |u(x, t) - U(x, t)|$$

which measures the size of the largest discrepancy between exact and numerical solutions. An alternative is the L^1 norm:

$$||u(\cdot, t) - U(\cdot, t)||_1 = \int |u(x, t) - U(x, t)| \, dx$$

which measures the area of the region enclosed by the graphs of the exact and the numerical solutions. When the exact solution is smooth, both norms give reasonable results, but when discontinuities may occur, the infinity norm tipically does not decrease refining the mesh, so that one does not observe convergence in the infinity norm. On the other hand, the area of the region between the two solutions decreases, and thus convergence can be obtained in the L^1 norm. For this reason, usually the L^1 norm is more useful when dealing with the convergence of schemes for conservation laws.

For instance in Fig 4.1, we can see the convergence of two numerical solutions under grid refinement. If the exact solution is smooth (left of the figure) the maximum distance between the exact and the numerical solutions, and the region enclosed by the graph of the two functions, go to zero, as N_x grows. We say that the numerical solution converges both in the L^1 and in the L^∞ norms. If the exact solution is not smooth (right plot in the figure) the maximum distance between exact and numerical solutions remains approximately constant under grid refinement, and it is approximately equal to a half of the size of the jump. In this case, the numerical solution does not converge to the exact solution in the L^∞ norm. However, we still have convergence in the L^1 norm.

We can now quantify the main properties of the elementary numerical schemes we have introduced. We first define the notion of convergence.

which finally yields

$$\mathcal{L}_h(t) = \frac{1}{6}\Delta t^2 \left(-a^3 + \frac{a}{\lambda^2}\right) u_{xxx} + O(h)^3. \tag{4.21}$$

Since the leading term of the error is $O(\Delta t)^2$, we conclude that the Lax-Wendroff scheme is second order accurate. ■

4.3 Stability and convergence

For one-step numerical schemes for ODE's, consistency is enough to ensure convergence. In that case, we introduced the notion of stability to ensure that the numerical solution is well behaved also for finite time steps, but we know that in any case the scheme will become stable, if the time step is small enough, and will eventually converge when the time step goes to zero.

Here the situation is more complicated, because we can have schemes that are unconditionally unstable, even when the time step goes to zero, and therefore will never converge.

Definition 4.3.1. Stability
A numerical scheme is stable if:

$$||U(\cdot, t + \Delta t)||_p \leq ||U(\cdot, t)||_p, \qquad \forall t = n\Delta t \text{ fixed}, \qquad \lambda \leq \lambda_0. \tag{4.22}$$

Note that the stability condition needs not be satisfied $\forall \Delta t$, but in general will be satisfied $\forall \Delta t \leq \Delta t_0$, that is the time step Δt (or the mesh ratio λ) must be below a certain threshold value t_0 (or λ_0) which gives the *stability condition* for our scheme.

The notion of stability just introduced is quite general, and can be applied also to non linear methods and to non linear equations. Below, we give an example of how we can prove that a scheme is stable, using this definition.

Example 4.3.2. Stability condition for the Lax-Friedrichs' scheme
Compute the stability condition of the Lax-Friedrichs' scheme for an initial value problem with initial condition $u_0(x)$ of compact support.

Solution. A function u_0 has compact support if:

$$u_0(x) \equiv 0, \qquad \forall x \notin [a, b]$$

for some compact interval $[a, b]$. An interval is compact if it is closed and bounded. We can define the support of a function as the set of points on which the function is non-zero. Since in conservation laws signals travel with a finite speed of propagation, if the initial condition has compact support, we know that the solution will have a compact support for all finite time. This in particular ensures that for all finite t:

$$||U(x + h, t)||_p = \left\{ \int_{-\infty}^{\infty} |U(x + h, t)|^p \, \mathrm{d}x \right\}^{1/p} = \left\{ \int_{-\infty}^{\infty} |U(x, t)|^p \, \mathrm{d}x \right\}^{1/p} = ||U(x, t)||_p,$$

that is, each p norm is invariant under traslation. It is important to note that this is not

We are now ready to study the evolution of the global error. At time $t + \Delta t$, the global error is given by:

$$e(x, t + \Delta t) = u(x, t + \Delta t) - U(x, t + \Delta t).$$

From the definition of the local truncation error (4.15), we can rewrite the exact solution as:

$$u(x, t + \Delta t) = \Delta t \mathcal{L}_h(t) + \mathcal{H}_h\left(u(\cdot, t); x\right).$$

On the other hand, the numerical solution $U(x, t + \Delta t)$ is obtained evolving the previous numerical solution $U(x, t)$ by one time step:

$$U(x, t + \Delta t) = \mathcal{H}_h\left(U(\cdot, t); x\right).$$

Substituting both expressions in the equation for the error, we find

$$e(x, t + \Delta t) = \Delta t \mathcal{L}_h(t) + \mathcal{H}_h\left(u(\cdot, t); x\right) - \mathcal{H}_h\left(U(\cdot, t); x\right).$$

We now use the linearity of the scheme, to find

$$\begin{aligned} e(x, t + \Delta t) &= \Delta t \mathcal{L}_h(t) + \mathcal{H}_h\left(u(\cdot, t) - U(\cdot, t); x\right) \\ &= \Delta t \mathcal{L}_h(t) + \mathcal{H}_h\left(e(\cdot, t); x\right). \end{aligned}$$

From this equation, we see that the error is small if 1) the error introduced in the last time step, $\Delta t \mathcal{L}_h(t)$ is small *and* 2) the error inherited from the previous time steps, $e(x, t)$ is small *and* 3) the error inherited by the previous time steps it is not amplified. The first statement is linked to consistency, the second statement is obtained applying by recursion the formula above to all previous time steps, the third statement is guaranteed by stability. I don't want to work out all details (you can find them on LeVeque's book [1]), but I want to stress that to write the error in the form we have obtained above, it is crucial that the scheme be linear. All these results can be gathered in the following theorem.

Theorem 4.3.3. Lax' equivalence theorem
Consider a linear and consistent numerical method $\mathcal{H}_h$. Then stability is a necessary and sufficient condition for convergence.

We end this section giving a *necessary* condition for stability, the CFL condition (from Courant, Friedrichs, Lewy). This means that the CFL condition does not guarantee that a scheme is stable, but if it is not satisfied, then the method will be unstable.

We first introduce the notion of *stencil*. The stencil of a scheme is the set of grid nodes needed to update the solution at the next time step, at a given node. For the Lax Friedrichs scheme, for instance, the stencil for the node x_j is $\{x_{j-1}, x_j, x_{j+1}\}$. The CFL condition states that the stencil to compute the solution at x_j at time t^{n+1} must contain the intersection of the characteristic through (x_j, t^{n+1}) with the axis of the previous time step, $t = t^n$. This is illustrated in Fig. 4.2. In both cases, the distance between the green point and x_j is $|a|\Delta t$. The green point must be contained in the stencil of the scheme. Thus, in both cases, we need $|a|\Delta t \leq h$, where $h = x_j - x_{j-1}$. Therefore, $\lambda \leq \frac{1}{|a|}$.

The CFL condition holds also in the non linear case. Here, the speed of propagation varies, and it is equal to $f'(u(x, t))$. Therefore, the CFL condition at a fixed time t becomes:

$$\lambda \leq \frac{1}{\max_x |f'(u(x, t))|}. \tag{4.23}$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 237 di 314

where ν is characteristic of the particular scheme chosen and depends on λ and a. For the Lax Friedrichs' scheme, $\nu = -\frac{1}{2}a^2\left(1 - \frac{1}{a^2\lambda^2}\right)$.

Suppose now that we apply the first order scheme to the equation:

$$u_t + au_x = \Delta t \nu u_{xx}. \tag{4.25}$$

This equation is called the *modified equation* for the particular scheme under study. If we compute the local truncation error that the scheme will produce on the equation above, we find $\mathcal{L}_h(t) = O(h)^2$. This means that the scheme we are considering is a first order scheme for the equation $u_t + au_x = 0$, but it is a *second* order method for equation (4.25). For this reason, we expect that our scheme will produce a numerical solution which will be more similar to the solutions of (4.25) than to the solutions of the original conservation law we are interested in.

Therefore, we study the qualitative behavior of the solutions of (4.25) to gain insight in what will be the evolution of our numerical solution. To do this we will study the time dependence of the Fourier modes of the modified equation, and compare it to the evolution of the Fourier modes of the exact solution. For simplicity, we will suppose that the problem we are studying is periodic, with period 2π.

Since our problem is linear, the solution can be written as;

$$u(x,t) = \sum_{k=-\infty}^{+\infty} b_k(t)e^{ikx},$$

where the coefficients $b_k(t)$ depend on the equation we are solving, but at the initial time

$$u(x, t=0) = u_0(x) = \sum_{k=-\infty}^{+\infty} b_k(0)e^{ikx}.$$

Thus at the initial time, the Fourier coefficients $b_k(0)$ are the Fourier transform of the initial data, that is we can compute them as

$$b_k(0) = \frac{1}{2\pi}\int_0^{2\pi} u_0(x)e^{-ikx}.$$

Fig. 4.3 shows the amplitude of the first 200 Fourier coefficients for a smooth function (a bell shaped Gaussian) and for a non-smooth function (a square wave). The figure shows clearly that the Fourier modes of the smooth function decrease very fast with k. Actually, in the case shown, they are negligible (below machine precision) for $k > 20$. On the other hand, the Fourier coefficients of the square wave have an amplitude that decreases very slowly with time (note the difference of the vertical scale for the two plots). Thus, in this case, the evolution in time of the high Fourier modes will be important to determine the overall behavior of the solution.

To compute the time evolution of the coefficients b_k, we substitute the generic Fourier mode

$$u_k(x,t) = b_k(t)e^{ikx}$$

in the equation we are solving. For the linear advection equation $u_t + au_x = 0$, the Fourier modes must satisfy:

$$b_k'(t)e^{ikx} + iakb_k(t)e^{ikx} = 0,$$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 239 di 314