Per gli studenti del corso di Elementi di Analisi Numerica

Questo scritto contiene:

- Cose scritte e insegnate nel I semestre A.A.2013-2014 durante il corso di Elementi di Analisi Numerica (EAN), non trattate negli anni passati (per il resto del programma di Elementi di Analisi Numerica, vedi: "Analisi Numerica 3 - Di Fiore" al focal point; metodo QR; deflazione).
- I tre esoneri di EAN 2013-2014, con le rispettive correzioni.

Più precisamente questo scritto contiene:

correzione I esonero EAN

correzione III esonero EAN

testo III esonero EAN

osservazioni su convergenza dei metodi ϕ e sui metodi Runge-Kutta (non presenti negli appunti al focal point)

correzione II esonero EAN

matrice di adiacenza dei ragazzi di EAN (..ha ilcellulare di..)

ciò che ho insegnato di diverso rispetto agli anni scorsi sulla teoria di Perron-Frobenius, metodo delle potenze, pagerank e metodo di Richardson-Eulero (è inclusa la dimostrazione del teorema di Gershgorin rafforzato)

osservazioni su problema di Cauchy e condiz. di Lip. (non presenti negli appunti al focal point)

testo II esonero EAN

domanda di cultura generale

il problema di Bruno Iannazzo, risolto solo in un caso particolare (argomento non presente negli appunti al focalpoint)

testo I esonero EAN

29 Dicembre 2013, Carmine Di Fiore

1) (i) Posto $\mathcal{I} = \int_1^{\sqrt{3}} \frac{1}{1+x^2}\,dx$, trovare le approssimazioni $\mathcal{I}_{\sqrt{3}-1}$ e $\mathcal{I}_{\frac{\sqrt{3}-1}{2}}$ di $\mathcal{I}$ fornite dalla formula dei trapezi e dedurne, tramite il metodo di estrapolazione di Romberg, l'approssimazione di ordine superiore $\tilde{\mathcal{I}}_{\frac{\sqrt{3}-1}{2}}$. (Facoltativo: notare che $\tilde{\mathcal{I}}_{\frac{\sqrt{3}-1}{2}}$ approssima $\mathcal{I}$ molto meglio di $\mathcal{I}_{\frac{\sqrt{3}-1}{4}}$).
(ii) Si può dire che la successione di approssimazioni $\mathcal{I}_{\frac{\sqrt{3}-1}{2^k}}$, $k = 0, 1, 2, 3, \ldots$, è non crescente?

Risoluzione

(i) $\mathcal{I}_{\sqrt{3}-1} = (\sqrt{3}-1)\left[\frac{1}{1+(1)^2}\frac{1}{2} + \frac{1}{1+(\sqrt{3})^2}\frac{1}{2}\right] = (\sqrt{3}-1)\left[\frac{3}{8}\right] = 0.274519..$

$\mathcal{I}_{\frac{\sqrt{3}-1}{2}} = \frac{\sqrt{3}-1}{2}\left[\frac{3}{8} + \frac{1}{1+(\frac{1+\sqrt{3}}{2})^2}\right] = \frac{\sqrt{3}-1}{2}\left[\frac{28+3\sqrt{3}}{8(4+\sqrt{3})}\right] = 0.26497..$

$\tilde{\mathcal{I}}_{\frac{\sqrt{3}-1}{2}} = \frac{1}{2^2-1}(2^2\mathcal{I}_{\frac{\sqrt{3}-1}{2}} - \mathcal{I}_{\sqrt{3}-1}) = \frac{1}{3}\left[(\sqrt{3}-1)2\frac{28+3\sqrt{3}}{8(4+\sqrt{3})} - \frac{3}{8}(\sqrt{3}-1)\right] = \frac{\sqrt{3}-1}{3}\frac{44+3\sqrt{3}}{8(4+\sqrt{3})} = 0.261788..$

$\mathcal{I}_{\frac{\sqrt{3}-1}{4}} = \frac{\sqrt{3}-1}{4}\left[\frac{28+3\sqrt{3}}{8(4+\sqrt{3})} + \frac{1}{1+(1+\frac{\sqrt{3}-1}{4})^2} + \frac{1}{1+(1+3\frac{\sqrt{3}-1}{4})^2}\right] = \frac{\sqrt{3}-1}{4}\left[\frac{28+3\sqrt{3}}{8(4+\sqrt{3})} + \frac{8}{14+3\sqrt{3}} + \frac{8}{22+3\sqrt{3}}\right] = 0.26259...$

$\mathcal{I} = \int_1^{\sqrt{3}} \frac{1}{1+x^2}dx = [\arctan x]_1^{\sqrt{3}} = \pi/12 = 0.261799..$

Dunque $\tilde{\mathcal{I}}_{\frac{\sqrt{3}-1}{2}}$ ha quattro cifre significative in comune con $\mathcal{I}$, mentre $\mathcal{I}_{\frac{\sqrt{3}-1}{4}}$, pur richiendendo due valutazioni di $f(x) = 1/(1+x^2)$ in più (rispetto a quelle sufficienti per ottenere $\tilde{\mathcal{I}}_{\frac{\sqrt{3}-1}{2}}$), ha solo due cifre significative in comune con $\mathcal{I}$.

(ii) La risposta è si. Per provarlo è sufficiente: notare che i nodi della formula dei trapezi $\mathcal{I}_{(\sqrt{3}-1)/2^k}$ sono nodi della formula dei trapezi $\mathcal{I}_{(\sqrt{3}-1)/2^{k+1}}$; notare che le formule dei trapezi $\mathcal{I}_{(\sqrt{3}-1)/2^k}$ e $\mathcal{I}_{(\sqrt{3}-1)/2^{k+1}}$ sono relative allo stesso intervallo $[1,\sqrt{3}]$; dimostrare che f è convessa in $[1,\sqrt{3}]$. In questa spiegazione è di aiuto un disegno in $[1,\sqrt{3}]$ di f e delle funzioni "rette a tratti" $\varphi_k|_{[1,\sqrt{3}]}$ e $\varphi_{k+1}|_{[1,\sqrt{3}]}$ di cui $\mathcal{I}_{(\sqrt{3}-1)/2^k}$ e $\mathcal{I}_{(\sqrt{3}-1)/2^{k+1}}$ rappresentano gli integrali su $[1,\sqrt{3}]$.

Segue una prova più dettagliata.

Considerati i punti $x_i = x_i^{(k)} = 1 + i\frac{\sqrt{3}-1}{2^k}$, $i = 0, 1, \ldots, 2^k$, dell'intervallo $[1,\sqrt{3}]$ e chiamata $\varphi_k(x)$ la funzione definita in $[1,\sqrt{3}]$ dall'identità $\varphi_k(x)|_{[x_i,x_{i+1}]} = f(x_i)\frac{x-x_{i+1}}{x_i-x_{i+1}} + f(x_{i+1})\frac{x-x_i}{x_{i+1}-x_i}$ (retta per $(x_i, f(x_i))$ e $(x_{i+1}, f(x_{i+1}))$, $i = 0, 1, \ldots, 2^k - 1$ (notare che $\varphi_k > 0$ in $[1,\sqrt{3}]$), il numero $\mathcal{I}_{(\sqrt{3}-1)/2^k}$ rappresenta l'integrale di φ_k in $[1,\sqrt{3}]$. Analogamente, $\mathcal{I}_{(\sqrt{3}-1)/2^{k+1}}$ rappresenta l'integrale di φ_{k+1} in $[1,\sqrt{3}]$.

Per $i = 0, 1, \ldots, 2^k - 1$ si ha $\varphi_k(x_i) = f(x_i) = \varphi_{k+1}(x_i)$ e $\varphi_k(x_{i+1}) = f(x_{i+1}) = \varphi_{k+1}(x_{i+1})$, perché $x_i = x_i^{(k)} = x_{2i}^{(k+1)}$ e $x_{i+1} = x_{i+1}^{(k)} = x_{2(i+1)}^{(k+1)}$. Inoltre, φ_k è una retta in $[x_i, x_{i+1}]$ e φ_{k+1} è una retta in $[x_i, (x_i + x_{i+1})/2]$ e in $[(x_i + x_{i+1})/2, x_{i+1}]$ (notare che $x_{2i+1}^{(k+1)} = (x_i + x_{i+1})/2$). Se fosse

$$\varphi_{k+1}(x_{2i+1}^{(k+1)}) = f(x_{2i+1}^{(k+1)}) \le \varphi_k(x_{2i+1}^{(k+1)}) = (f(x_{2i}^{(k+1)}) + f(x_{2i+2}^{(k+1)}))/2, \quad i = 0, 1, \ldots, 2^k - 1, \qquad (*)$$

avremmo che $\varphi_{k+1}|_{[x_i,x_{i+1}]} \le \varphi_k|_{[x_i,x_{i+1}]}$, $i = 0, 1, \ldots, 2^k - 1$, e quindi avremmo la disuguaglianza $\mathcal{I}_{(\sqrt{3}-1)/2^{k+1}} \le \mathcal{I}_{(\sqrt{3}-1)/2^k}$. La disuguaglianza (*) è ad esempio verificata se f è convessa in $[1,\sqrt{3}]$.

Poiché $f(x) = 1/(1+x^2) \Rightarrow f'(x) = -2x/(1+x^2)^2 \Rightarrow f''(x) = 2(3x^2-1)/(1+x^2)^3$ e f'' è positiva in $(1/\sqrt{3}, +\infty)$ e quindi in $[1,\sqrt{3}]$, la nostra f è effettivamente convessa in $[1,\sqrt{3}]$ e di conseguenza vale (*), cioè la successione $\{\mathcal{I}_{(\sqrt{3}-1)/2^k}\}_{k=0}^{+\infty}$ è non crescente.

2) Sia $f_t(x) = \sum_{n=0}^{+\infty} B_n(x)t^n/n!$.
(i) Dimostrare che $f_t(x)$ soddisfa le due condizioni $\frac{d}{dx}f_t(x) = tf_t(x)$ e $\int_0^1 f_t(x)\,dx = 1$.
(ii) Dedurne una formula chiusa per $f_t(x)$.

Risoluzione
(i) Poiché $\int_0^1 B_n(x)\,dx = 0$ e $B_n'(x) = nB_{n-1}(x)$, $n = 1, 2, \ldots$, si ha

$$\begin{aligned}\frac{d}{dx}f_t(x) &= \frac{d}{dx}\sum_{n=0}^{+\infty} B_n(x)\frac{t^n}{n!} = \frac{d}{dx}\Big(\ 1 + \sum_{n=1}^{+\infty} B_n(x)\frac{t^n}{n!}\ \Big)\\ &= \sum_{n=1}^{+\infty} B_n'(x)\frac{t^n}{n!} = \sum_{n=1}^{+\infty} nB_{n-1}(x)\frac{t^n}{n!}\\ &= \sum_{n=1}^{+\infty} B_{n-1}(x)\frac{t^n}{(n-1)!} = \sum_{j=0}^{+\infty} B_j(x)\frac{t^{j+1}}{j!} = tf_t(x),\end{aligned}$$

$$\begin{aligned}\int_0^1 f_t(x)\,dx &= \int_0^1\Big(\ 1 + \sum_{n=1}^{+\infty} B_n(x)\frac{t^n}{n!}\ \Big)\,dx\\ &= 1 + \sum_{n=1}^{+\infty}\frac{t^n}{n!}\int_0^1 B_n(x)\,dx = 1.\end{aligned}$$

(ii) La condizione $\frac{d}{dx}f_t(x) = tf_t(x)$ implica $f_t(x) = c(t)e^{tx}$. Quindi, la condizione $\int_0^1 f_t(x)\,dx = 1$ diventa $1 = \int_0^1 c(t)e^{tx}dx = c(t)\int_0^1 e^{tx}dx = c(t)\frac{1}{t}[e^{tx}]_0^1 = c(t)\frac{1}{t}(e^t - 1)$, da cui l'espressione per la funzione $c(t)$, $c(t) = t/(e^t - 1)$. Dunque, $f_t(x) = e^{tx}t/(e^t - 1)$.

3) Si vuole approssimare $\varsigma(1+\varepsilon) = \sum_{r=1}^{+\infty}\frac{1}{r^{1+\varepsilon}}$ uniformemente per $\varepsilon \in (0, 2]$.
(i) Usando la formula di Eulero-Maclaurin, trovare una funzione $\varphi(\varepsilon, m, k)$, $\varepsilon \in \mathbb{R}^+$, $m \geq 1$, $k \geq 0$ $(m, k \in \mathbb{Z})$, per cui vale l'identità

$$\sum_{r=m}^{+\infty}\frac{1}{r^{1+\varepsilon}} = \varphi(\varepsilon, m, k) + u_{k+1}(\infty),$$

dando una maggiorazione per $|u_{k+1}(\infty)|$ (in funzione di ε, m, k).
(ii) Usare la maggiorazione ottenuta nel punto precedente per trovare una costante $M_{m,k}$ tale che $|u_{k+1}(\infty)| \leq M_{m,k}$ per ogni $\varepsilon \in [1, 2]$.
(iii) Porre $m = 2$ e, valutando la costante $M_{2,k}$ del punto precedente per $k = 0, 1, 2, \ldots$, determinare il minimo valore di k per cui

$$|\{1 + \varphi(\varepsilon, 2, k)\} - \varsigma(1+\varepsilon)| \leq \frac{1}{2^9}, \quad \forall\, \varepsilon \in [1, 2].$$

(Quest'ultimo risultato vale in realtà per ogni $\varepsilon \in (0, 2]$, con lo stesso k).

Risoluzione
La formula di Eulero-Maclaurin

$$\sum_{r=m}^{n} f(r) = \frac{1}{2}(f(m) + f(n)) + \int_m^n f(x)\,dx + \sum_{j=1}^{k}\frac{B_{2j}(0)}{(2j)!}[f^{(2j-1)}(n) - f^{(2j-1)}(m)] + u_{k+1}$$

per $f(x) = 1/x^{1+\varepsilon}$ diventa

$$\sum_{r=m}^{n}\frac{1}{r^{1+\varepsilon}} = \frac{1}{2}\Big(\frac{1}{m^{1+\varepsilon}} + \frac{1}{n^{1+\varepsilon}}\Big) - \frac{1}{\varepsilon n^{\varepsilon}} + \frac{1}{\varepsilon m^{\varepsilon}} + \sum_{j=1}^{k}\frac{B_{2j}(0)}{(2j)!}\Big[\Big(-\frac{1}{n^{2j+\varepsilon}} + \frac{1}{m^{2j+\varepsilon}}\Big)(1+\varepsilon)(2+\varepsilon)\cdots(2j-1+\varepsilon)\Big] + u_{k+1} \qquad \text{(EMform)}$$

($\int_m^n \frac{1}{x^{1+\varepsilon}}dx = [\frac{x^{-\varepsilon}}{-\varepsilon}]_m^n$; $f^{(s)}(x) = (-1)^s(1+\varepsilon)(2+\varepsilon)\cdots(s+\varepsilon)x^{-(s+1+\varepsilon)}$, $f^{(2j-1)}(x) = -(1+\varepsilon)(2+\varepsilon)\cdots(2j-1+\varepsilon)x^{-(2j+\varepsilon)}$). Inoltre, poiché $f^{(2k+2)}(x) > 0$ in $(0,+\infty)$ e quindi non cambia segno in $[m,n]$ se $m \geq 1$, il modulo di u_{k+1} deve essere più piccolo di due volte il modulo del termine successivo della somma,

$$|u_{k+1}| \leq 2\frac{|B_{2k+2}(0)|}{(2k+2)!}|-\frac{1}{n^{2k+2+\varepsilon}}+\frac{1}{m^{2k+2+\varepsilon}}|(1+\varepsilon)(2+\varepsilon)\cdots(2k+1+\varepsilon). \tag{EMerr}$$

Per $n \to +\infty$, le (EMfor) e (EMerr) diventano

$$\sum_{r=m}^{+\infty}\frac{1}{r^{1+\varepsilon}} = \frac{1}{2m^{1+\varepsilon}}+\frac{1}{\varepsilon m^{\varepsilon}}+\sum_{j=1}^{k}\frac{B_{2j}(0)}{(2j)!}\frac{(1+\varepsilon)(2+\varepsilon)\cdots(2j-1+\varepsilon)}{m^{2j+\varepsilon}}+u_{k+1}(\infty) =: \varphi(\varepsilon,m,k)+u_{k+1}(\infty),$$

$$|u_{k+1}(\infty)| \leq 2\frac{|B_{2k+2}(0)|}{(2k+2)!}\frac{(1+\varepsilon)(2+\varepsilon)\cdots(2k+1+\varepsilon)}{m^{2k+2+\varepsilon}}.$$

(ii) In particolare, se $\varepsilon \in [1,2]$, allora

$$|u_{k+1}(\infty)| \leq 2\frac{|B_{2k+2}(0)|}{(2k+2)!}\frac{(3)(4)\cdots(2k+3)}{m^{2k+3}} = \frac{|B_{2k+2}(0)|(2k+3)}{m^{2k+3}} =: M_{m,k},$$

e se $\varepsilon \in (0,1]$, allora

$$|u_{k+1}(\infty)| \leq 2\frac{|B_{2k+2}(0)|}{(2k+2)!}\frac{(2)(3)\cdots(2k+2)}{m^{2k+2}} = \frac{2|B_{2k+2}(0)|}{m^{2k+2}} =: N_{m,k}.$$

(iii) Sappiamo che

$$|1+\varphi(\varepsilon,2,k)-\zeta(1+\varepsilon)| = |\varphi(\varepsilon,2,k)-\sum_{r=2}^{+\infty}\frac{1}{r^{1+\varepsilon}}| \leq \left\{\begin{array}{ll} M_{2,k} & \varepsilon \in [1,2] \\ N_{2,k} & \varepsilon \in (0,1] \end{array}\right. ,$$

dove

$$M_{2,k} = \frac{|B_{2k+2}(0)|(2k+3)}{2^{2k+3}}, \qquad N_{2,k} = \frac{|B_{2k+2}(0)|}{2^{2k+1}}.$$

Quindi, per rispondere alla domanda è sufficiente trovare il minimo k per cui $M_{2,k} \leq 1/2^9$. Vedremo che per lo stesso k si avrà anche $N_{2,k} \leq 1/2^9$. Si ha

$$\begin{array}{ll} M_{2,0} = \frac{1}{6}3\frac{1}{2^3} = \frac{1}{2^4}, & N_{2,0} = \frac{1}{6}\frac{1}{2} = \frac{1}{3}\frac{1}{2^2}, \\ M_{2,1} = \frac{1}{30}5\frac{1}{2^5} = \frac{1}{3}\frac{1}{2^6}, & N_{2,1} = \frac{1}{30}\frac{1}{2^3} = \frac{1}{15}\frac{1}{2^4}, \\ M_{2,2} = \frac{1}{42}7\frac{1}{2^7} = \frac{1}{3}\frac{1}{2^8} < \frac{1}{2^9}, & N_{2,2} = \frac{1}{42}\frac{1}{2^5} = \frac{1}{21}\frac{1}{2^6} < \frac{1}{2^9}, \end{array}$$

dunque il valore di k richiesto è $k = 2$. Ne segue che il numero

$$\begin{array}{rcl} 1+\varphi(\varepsilon,2,2) & = & 1+\frac{1}{2\cdot 2^{1+\varepsilon}}+\frac{1}{\varepsilon 2^{\varepsilon}}+\sum_{j=1}^{2}\frac{B_{2j}(0)}{(2j)!}\frac{(1+\varepsilon)(2+\varepsilon)\cdots(2j-1+\varepsilon)}{2^{2j+\varepsilon}} \\ & = & 1+\frac{1}{2\cdot 2^{1+\varepsilon}}+\frac{1}{\varepsilon 2^{\varepsilon}}+\frac{B_2(0)}{2!}\frac{(1+\varepsilon)}{2^{2+\varepsilon}}+\frac{B_4(0)}{4!}\frac{(1+\varepsilon)(2+\varepsilon)(3+\varepsilon)}{2^{4+\varepsilon}} \\ & = & 1+\frac{1}{2\cdot 2^{1+\varepsilon}}+\frac{1}{\varepsilon 2^{\varepsilon}}+\frac{(1+\varepsilon)}{12\cdot 2^{2+\varepsilon}}-\frac{(1+\varepsilon)(2+\varepsilon)(3+\varepsilon)}{30\cdot 24\cdot 2^{4+\varepsilon}} \end{array}$$

è tale che $|1+\varphi(\varepsilon,2,2)-\zeta(1+\varepsilon)| < 1/2^9$, $\forall\,\varepsilon \in (0,2]$.

4) Siano: A generica $n\times n$ ($a_{ij}\in\mathbb{C}$),
X $n\times k$ ($k<n$) con colonne linearmente indipendenti tale che $AX = XJ$ per una matrice $k\times k$ J, Y $n\times(n-k)$ tale che la matrice $n\times n$ $[X\ Y]$ è non singolare. Dimostrare che
(i) Esiste una matrice $(n-k)\times(n-k)$ B tale che

$$[X\ Y]^{-1}A[X\ Y] = \begin{bmatrix} J & * \\ O & B \end{bmatrix}.$$

(ii) (Facoltativo) Per ogni matrice $k\times n$ W^H tale che $\det(W^HX)\neq 0$ si ha

$$[X\ Y]^{-1}\Big(A - XJ(W^HX)^{-1}W^H\Big)[X\ Y] = \begin{bmatrix} O & *' \\ O & B \end{bmatrix}.$$

(iii) Per ogni matrice $k\times n$ W^H tale che $\det(W^HX)\neq 0$ si ha

$$[X\ Y]^{-1}\Big(A - X(W^HX)^{-1}W^HA\Big)[X\ Y] = \begin{bmatrix} O & *'' \\ O & B \end{bmatrix}.$$

e $W^H\Big(A - X(W^HX)^{-1}W^HA\Big)$ è la matrice nulla $k\times n$.
(iv) Posto

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ a & 0 & 1-a & 0 \\ 0 & b & 0 & 1-b \\ 0 & 0 & 1 & 0 \end{bmatrix},\ X = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix},\ J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

si osserva che $AX = XJ$; dunque lo spettro di A è del tipo $\sigma(A) = \{1,-1,\lambda_3,\lambda_4\}$. Osservare che si può usare il risultato in (iii) per

$$W^H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

e, quindi, introdurre una matrice 2×2 i cui autovalori sono λ_3 e λ_4.

Risoluzione
(i)

$$\begin{aligned}\Big[X\ \Big|\ Y\Big]^{-1}A\Big[X\ \Big|\ Y\Big] &= \Big[X\ \Big|\ Y\Big]^{-1}\Big[AX\ \Big|\ AY\Big] = \Big[X\ \Big|\ Y\Big]^{-1}\Big[XJ\ \Big|\ AY\Big] \\ &= \Big[\Big[X\ \Big|\ Y\Big]^{-1}XJ\ \Big|\ \Big[X\ \Big|\ Y\Big]^{-1}AY\Big] = \Big[\begin{bmatrix} I_k \\ O \end{bmatrix}J\ \Big|\ \Big[X\ \Big|\ Y\Big]^{-1}AY\Big] \\ &= \Big[\begin{bmatrix} J \\ O \end{bmatrix}\ \Big|\ \Big[X\ \Big|\ Y\Big]^{-1}AY\Big] = \begin{bmatrix} J & * \\ O & B \end{bmatrix},\end{aligned}$$

dove si è posto $\begin{bmatrix} * \\ O \end{bmatrix} = \Big[X\ \Big|\ Y\Big]^{-1}AY$.

(ii)

$$\left[X \,\middle|\, Y\right]^{-1}(A - XJ(W^HX)^{-1}W^H)\left[X \,\middle|\, Y\right] = \begin{bmatrix} J & * \\ O & B \end{bmatrix} - \begin{bmatrix} I_k \\ O \end{bmatrix} J(W^HX)^{-1}\left[W^HX \middle| W^HY\right]$$
$$= \begin{bmatrix} J & * \\ O & B \end{bmatrix} - \begin{bmatrix} J \\ O \end{bmatrix}\left[I_k \middle| (W^HX)^{-1}W^HY\right] = \begin{bmatrix} J & * \\ O & B \end{bmatrix} - \begin{bmatrix} J & J(W^HX)^{-1}W^HY \\ O & O \end{bmatrix}$$
$$= \begin{bmatrix} O & *' \\ O & B \end{bmatrix}, \quad *' = * - J(W^HX)^{-1}W^HY.$$

(iii)

$$\left[X \,\middle|\, Y\right]^{-1}(A - X(W^HX)^{-1}W^HA)\left[X \,\middle|\, Y\right] = \begin{bmatrix} J & * \\ O & B \end{bmatrix} - \begin{bmatrix} I_k \\ O \end{bmatrix} (W^HX)^{-1}\left[W^HAX \middle| W^HAY\right]$$
$$= \begin{bmatrix} J & * \\ O & B \end{bmatrix} - \begin{bmatrix} I_k \\ O \end{bmatrix} (W^HX)^{-1}\left[W^HXJ \middle| W^HAY\right]$$
$$= \begin{bmatrix} J & * \\ O & B \end{bmatrix} - \begin{bmatrix} I_k \\ O \end{bmatrix}\left[J \middle| (W^HX)^{-1}W^HAY\right] = \begin{bmatrix} J & * \\ O & B \end{bmatrix} - \begin{bmatrix} J & (W^HX)^{-1}W^HAY \\ O & O \end{bmatrix}$$
$$= \begin{bmatrix} O & *'' \\ O & B \end{bmatrix}, \quad *'' = * - (W^HX)^{-1}W^HAY.$$

(iv) AX e XJ sono entrambe uguali alla matrice $\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, dunque $AX = XJ$. Ne segue, in particolare, che 1 e -1 sono autovalori di A, cioè $\sigma(A) = \{1, -1, \lambda_3, \lambda_4\}$.

Applicando il risultato (iii) per $W^H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, si ha che

$$M = A - X(W^HX)^{-1}W^HA$$

ha gli autovalori $0, 0, \lambda_3, \lambda_4$ e le prime due righe di M sono nulle (perché $W^H(\,A - X(W^HX)^{-1}W^HA\,) = O$). Dunque una matrice 2×2 i cui autovalori sono λ_3 e λ_4 è la sottomatrice N 2×2 in basso a destra di M. Per trovarla basta scrivere esplicitamente M:

$$\begin{aligned} M &= A - \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} A = A - \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} A \\ &= A - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ a & 0 & 1-a & 0 \\ 0 & b & 0 & 1-b \\ 0 & 0 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 \\ a & 0 & 1-a & 0 \\ 0 & 1 & 0 & 0 \\ a & 0 & 1-a & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & b-1 & 0 & 1-b \\ -a & 0 & a & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 1-b \\ a & 0 \end{bmatrix}. \end{aligned}$$

Quindi λ_3 e λ_4 sono le radici del polinomio $\lambda^2 + a(b-1)$.

Correzione III esonero di Elementi di Analisi Numerica (Dic. 2013)

(1) Data y funzione sufficientemente regolare in un intorno di x e preso $h > 0$ abbastanza piccolo, trovare una differenza finita $\varphi\Big(y(x-2h), y(x-h), y(x), y(x+h), y(x+2h)\Big)$ tale che

$$y''(x) = \varphi\Big(y(x-2h), y(x-h), y(x), y(x+h), y(x+2h)\Big) + O(h^4).$$

Usare tale differenza finita per introdurre un sistema lineare $A\eta = \mathbf{b}$, A $n\times n$, $\mathbf{b}$ $n\times 1$, la cui soluzione η ha come componente i-esima una approssimazione $\eta(x_i)$ di $y(x_i)$ essendo $y(x)$ la soluzione del problema differenziale $-y''(x) = q(x)$, $x \in (a,b)$, $y(a) = \alpha$, $y(b) = \beta$, e $x_i = a + i(b-a)/(n+1)$, $i = 0, 1, \dots, n, n+1$.

Risoluzione Sommando le identità

$$y(x-h) = y(x) - hy'(x) + \frac{h^2}{2}y''(x) - \frac{h^3}{6}y'''(x) + \frac{h^4}{24}y'^v(x) - \frac{h^5}{120}y^v(x) + O(h^6),$$

$$y(x+h) = y(x) + hy'(x) + \frac{h^2}{2}y''(x) + \frac{h^3}{6}y'''(x) + \frac{h^4}{24}y'^v(x) + \frac{h^5}{120}y^v(x) + O(h^6)$$

si ottiene:

$$y(x-h) + y(x+h) - 2y(x) = h^2 y''(x) + \frac{h^4}{12}y'^v(x) + O(h^6). \qquad \text{(PrimaUguagl)}$$

Sommando le identità

$$y(x-2h) = y(x) - 2hy'(x) + \frac{(2h)^2}{2}y''(x) - \frac{(2h)^3}{6}y'''(x) + \frac{(2h)^4}{24}y'^v(x) - \frac{(2h)^5}{120}y^v(x) + O(h^6),$$

$$y(x+2h) = y(x) + 2hy'(x) + \frac{(2h)^2}{2}y''(x) + \frac{(2h)^3}{6}y'''(x) + \frac{(2h)^4}{24}y'^v(x) + \frac{(2h)^5}{120}y^v(x) + O(h^6)$$

si ottiene:

$$y(x-2h) + y(x+2h) - 2y(x) = 4h^2 y''(x) + \frac{16h^4}{12}y'^v(x) + O(h^6). \qquad \text{(SecondaUguagl)}$$

Moltiplicando per 16 la (PrimaUguagl) ottenuta, si ha l'identità

$$16y(x-h) + 16y(x+h) - 32y(x) = 16h^2 y''(x) + \frac{16h^4}{12}y'^v(x) + O(h^6)$$

e sottraendo a quest'ultima la (SecondaUguagl), si ricava

$$-y(x-2h) + 16y(x-h) - 30y(x) + 16y(x+h) - y(x+2h) = 12h^2 y''(x) + O(h^6).$$

da cui la formula richiesta:

$$y''(x) = \frac{-y(x-2h) + 16y(x-h) - 30y(x) + 16y(x+h) - y(x+2h)}{12h^2} + O(h^4). \qquad \text{(dII5punti)}$$

Se $y(t)$ è la soluzione del problema differenziale $-y''(x) = q(x)$, $x \in (a, b)$, $y(a) = \alpha$, $y(b) = \beta$, allora, considerati i punti $x_i = a + ih$, $h = (b-a)/(n+1)$, $i = 0, 1, \dots, n, n+1$, dell'intervallo $[a, b]$, devono essere verificate le seguenti uguaglianze:

$$-y''(x_i) = q(x_i),\ i = 1, \dots, n, \quad y(x_0) = \alpha,\ y(x_{n+1}) = \beta.$$

A questo punto occorre approssimare le derivate seconde $y''(x_i)$, $i = 1, \dots, n$, con opportune differenze finite. Il testo dell'esercizio suggerisce di usare la formula (dII5punti) sopra ottenuta. Ciò viene fatto bene in Appendice. Infatti in Appendice si mostra che tramite la (dII5punti) si può ottenere un sistema lineare $Ay(\mathbf{x}) = \mathbf{b} + \varepsilon$, $y(\mathbf{x}) = [y(x_1)\ y(x_2)\ \cdots\ y(x_n)]^T$, tale che $(\varepsilon)_i = O(h^6)$ per ogni $i = 1, \dots, n$. Qui invece, per semplicità, otteniamo un sistema lineare $Ay(\mathbf{x}) = \mathbf{b} + \varepsilon$, $y(\mathbf{x}) = [y(x_1)\ y(x_2)\ \cdots\ y(x_n)]^T$, meno preciso e precisamente tale che $(\varepsilon)_i = O(h^6)$, $i = 2, \dots, n-1$, e $(\varepsilon)_i = O(h^4)$, $i = 1, n$.

Usando la (dII5punti) per $x = x_i$, $i = 2, \dots, n-1$, e la differenza finita

$$y''(x) = \frac{y(x-h) - 2y(x) + y(x+h)}{h^2} + O(h^2) \qquad \text{(dII3punti)}$$

(dimostrata a lezione) per $x = x_1$ e $x = x_n$, le equazioni $-y''(x_i) = q(x_i)$, $i = 1, \dots, n$ diventano

$$\begin{aligned}
&-y(x_0) + 2y(x_1) - y(x_2) = h^2 q(x_1) + O(h^4),\\
&y(x_{i-2}) - 16y(x_{i-1}) + 30y(x_i) - 16y(x_{i+1}) + y(x_{i+2}) = 12h^2 q(x_i) + O(h^6), \quad i = 2, \dots, n-1,\\
&-y(x_{n-1}) + 2y(x_n) - y(x_{n+1}) = h^2 q(x_n) + O(h^4),
\end{aligned}$$

$$\begin{bmatrix}
2 & -1 & 0 & 0 & \cdot & & & \cdot & 0\\
-16 & 30 & -16 & 1 & \cdot & & & & \cdot\\
1 & -16 & 30 & & \cdot & \cdot & & & \\
0 & 1 & & & & \cdot & \cdot & & \cdot\\
\cdot & \cdot & \cdot & & & & & 1 & 0\\
 & & \cdot & \cdot & & & 30 & -16 & 1\\
\cdot & & & \cdot & 1 & -16 & & 30 & -16\\
0 & \cdot & & \cdot & 0 & 0 & & -1 & 2
\end{bmatrix}
\begin{bmatrix} y(x_1)\\ y(x_2)\\ y(x_3)\\ \cdot\\ \cdot\\ y(x_{n-2})\\ y(x_{n-1})\\ y(x_n)\end{bmatrix}
=
\begin{bmatrix} \alpha\\ -\alpha\\ 0\\ \cdot\\ \cdot\\ 0\\ -\beta\\ \beta\end{bmatrix}
+12h^2
\begin{bmatrix} \frac{1}{12}q(x_1)\\ q(x_2)\\ q(x_3)\\ \cdot\\ \cdot\\ q(x_{n-2})\\ q(x_{n-1})\\ \frac{1}{12}q(x_n)\end{bmatrix}
+O(h^{4,6})$$

Si definiscono le approssimazioni $\eta(x_i)$ dei valori esatti $y(x_i)$, $i = 1, \dots, n$, sostituendo, nell'uguaglianza vettoriale ottenuta, $y(x_i)$, $i = 1, \dots, n$, con $\eta(x_i)$, $i = 1, \dots, n$, e togliendo gli errori "$O(h^{4,6})$".

(2) Dimostrare che se f è tale che $|f(x, y) - f(x, \tilde{y})| \le L|y - \tilde{y}|$ per ogni $x \in [a, b]$, $y, \tilde{y} \in \mathbb{R}$ $(L \ge 0)$, allora ogni schema di Runge-Kutta

$$\phi(x, y; h) = a_1 f(x, y) + a_2 f(x + p_1 h, y + p_2 h f(x, y))$$

è tale che $|\phi(x, y; h) - \phi(x, \tilde{y}; h)| \le N|y - \tilde{y}|$ per ogni $x \in [a, b]$, $y, \tilde{y} \in \mathbb{R}$, $|h| \le h_0$, dove N dipende da L, h_0 e dai parametri dello schema.

Risoluzione

$$
\begin{aligned}
|\phi(x,y;h)-\phi(x,\tilde{y};h)| &= |a_1 f(x,y)+a_2 f(x+p_1 h, y+p_2 h f(x,y)) - (\ a_1 f(x,\tilde{y})+a_2 f(x+p_1 h,\tilde{y}+p_2 h f(x,\tilde{y}))\ )| \\
&\le |a_1||f(x,y)-f(x,\tilde{y})|+|a_2||f(x+p_1 h,y+p_2 h f(x,y))-f(x+p_1 h,\tilde{y}+p_2 h f(x,\tilde{y}))| \\
&\le |a_1|L|y-\tilde{y}|+|a_2|L|y+p_2 h f(x,y)-(\tilde{y}+p_2 h f(x,\tilde{y}))| \\
&\le |a_1|L|y-\tilde{y}|+|a_2|L\Big(|y-\tilde{y}|+|p_2||h||f(x,y)-f(x,\tilde{y})|\Big) \\
&\le |a_1|L|y-\tilde{y}|+|a_2|L\Big(|y-\tilde{y}|+|p_2||h|L|y-\tilde{y}|\Big) \\
&\le |y-\tilde{y}|\Big[|a_1|L+|a_2|L\Big(1+|p_2|h_0 L\Big)\Big] = |y-\tilde{y}|N.
\end{aligned}
$$

Nota (Martina): In uno dei passaggi si usa il fatto che $x+p_1 h \in [a,b]$. Ciò può essere vero (per ogni h, $|h| \le h_0$) solo se $x \in [c,d] \subset (a,b)$. Dunque la tesi dell'esercizio andrebbe perfezionata.

(3) Dimostrare che lo schema di Runge-Kutta

$$\frac{\eta(x+h)-y}{h} = \frac{1}{4}f(x,y)+\frac{3}{4}f(x+\frac{2}{3}h, y+\frac{2}{3}hf(x,y))$$

è di ordine due ma non di ordine tre. Calcolare con tale schema, ponendo $h=1$, una approssimazione $\eta(0)$ di $y(0)$ essendo $y(t)$ la soluzione del problema di Cauchy $y'(t) = -ty(t)^2$, $y(-1) = 2/3$. Facoltativo: confrontare il risultato con l'approssimazione ottenuta con lo schema di Eulero.

Risoluzione

A lezione si è visto che uno schema di tipo Runge-Kutta

$$\phi(x,y;h) = a_1 f(x,y) + a_2 f(x+p_1 h, y+p_2 h f(x,y))$$

è di ordine due se i parametri a_1, a_2, p_1, p_2 verificano le seguenti condizioni $a_1 + a_2 = 1$, $a_2 p_1 = \frac{1}{2}$, $a_2 p_2 = \frac{1}{2}$. È evidente che le scelte $a_1 = 1/4$, $a_2 = 3/4$, $p_1 = p_2 = 2/3$, che corrispondono allo schema dato, verificano tali condizioni. Quindi lo schema dato ha ordine almeno due. Vedremo che dello sviluppo in potenze di h dello schema dato, anche il coefficiente di h^2 assomiglia, per quanto possibile, al coefficiente di h^2 dello sviluppo del rapporto incrementale esatto $(y(x+h)-y)/h$.

Sia $y(t)$, la soluzione del problema $y'(t) = f(t,y(t))$, $y(x) = y$, sufficientemente regolare nell'intorno di $t = x$. Allora

$$\frac{y(x+h)-y}{h} = y'(x) + \frac{h}{2}y''(x) + \frac{h^2}{6}y'''(x) + O(h^3).$$

Poiché

$$y''(t) = \frac{d}{dt}y'(t) = \frac{d}{dt}f(t,y(t)) = f_x(t,y(t)) + f_y(t,y(t))f(t,y(t)),$$

$$
\begin{aligned}
y'''(t) &= \tfrac{d}{dt}y''(t) = \tfrac{d}{dt}(f_x(t,y(t)) + f_y(t,y(t))f(t,y(t))) \\
&= \Big(f_{xx} + 2f_{xy}f + f_{yy}f^2 + f_y f_x + f_y^2 f\Big)\Big|_{(t,y(t))},
\end{aligned}
$$

si ha

$$
\begin{aligned}
\frac{y(x+h)-y}{h} &= f(x,y) + \frac{h}{2}[f_x(x,y) + f_y(x,y)f(x,y)] + \frac{h^2}{6}\Big[f_{xx}(x,y) + 2f_{xy}(x,y)f(x,y) \\
&\quad + f_{yy}(x,y)f(x,y)^2 + f_y(x,y)f_x(x,y) + f_y(x,y)^2 f(x,y)\Big] + O(h^3).
\end{aligned}
$$

Inoltre,

$$\begin{aligned}
f(x+p_1h,y+p_2hf(x,y)) &= f(x,y)+p_1hf_x(x,y)+p_2hf(x,y)f_y(x,y)\\
&\quad+\frac{1}{2}\Big[p_1^2h^2f_{xx}(x,y)+2p_1p_2h^2f(x,y)f_{xy}(x,y)+p_2^2h^2f(x,y)^2f_{yy}(x,y)\Big]+O(h^3)\\
&= f(x,y)+h[p_1f_x(x,y)+p_2f(x,y)f_y(x,y)]\\
&\quad+h^2\frac{1}{2}\Big[p_1^2f_{xx}(x,y)+2p_1p_2f(x,y)f_{xy}(x,y)+p_2^2f(x,y)^2f_{yy}(x,y)\Big]+O(h^3),
\end{aligned}$$

$$\begin{aligned}
f(x+\frac{2}{3}h,y+\frac{2}{3}hf(x,y)) &= f(x,y)+\frac{2}{3}hf_x(x,y)+\frac{2}{3}hf(x,y)f_y(x,y)\\
&\quad+\frac{1}{2}\Big[\frac{4}{9}h^2f_{xx}(x,y)+2\frac{4}{9}h^2f(x,y)f_{xy}(x,y)+\frac{4}{9}h^2f(x,y)^2f_{yy}(x,y)\Big]+O(h^3)\\
&= f(x,y)+h\Big[\frac{2}{3}f_x(x,y)+\frac{2}{3}f(x,y)f_y(x,y)\Big]\\
&\quad+h^2\Big[\frac{2}{9}f_{xx}(x,y)+\frac{4}{9}f(x,y)f_{xy}(x,y)+\frac{2}{9}f(x,y)^2f_{yy}(x,y)\Big]+O(h^3).
\end{aligned}$$

Quindi

$$\begin{aligned}
&\phi(x,y;h):=\tfrac{1}{4}f(x,y)+\tfrac{3}{4}f(x+\tfrac{2}{3}h,y+\tfrac{2}{3}hf(x,y))\\
&=\tfrac{1}{4}f(x,y)+\tfrac{3}{4}\Big\{f(x,y)+h\Big[\tfrac{2}{3}f_x(x,y)+\tfrac{2}{3}f(x,y)f_y(x,y)\Big]\\
&+h^2\Big[\tfrac{2}{9}f_{xx}(x,y)+\tfrac{4}{9}f(x,y)f_{xy}(x,y)+\tfrac{2}{9}f(x,y)^2f_{yy}(x,y)\Big]+O(h^3)\Big\}\\
&=f(x,y)+\tfrac{1}{2}h[f_x(x,y)+f(x,y)f_y(x,y)]+\tfrac{h^2}{6}\Big[f_{xx}(x,y)+2f(x,y)f_{xy}(x,y)+f(x,y)^2f_{yy}(x,y)\Big]+O(h^3).
\end{aligned}$$

Raccogliendo i risultati ottenuti, si ha

$$\frac{y(x+h)-y}{h}-\phi(x,y;h)=\frac{h^2}{6}[f_y(x,y)f_x(x,y)+f_y(x,y)^2f(x,y)]+O(h^3)$$

cioè lo schema di Runge-Kutta $\phi(x,y;h)$ è di ordine due, ma non di ordine tre, essendo in generale $f_y(x,y)f_x(x,y)+f_y(x,y)^2f(x,y)\neq 0$.

Scriviamo ora il metodo corrispondente: $\eta(x_0)=y(x_0)=y_0$, per $i=0,1,2,\ldots$

$$\eta(x_i+h)=\eta(x_i)+h\Big[\frac{1}{4}f(x_i,\eta(x_i))+\frac{3}{4}f\Big(x_i+\frac{2}{3}h,\eta(x_i)+\frac{2}{3}hf(x_i,\eta(x_i))\Big)\Big],\ x_{i+1}=x_i+h.$$

Tale metodo, applicato al particolare problema di Cauchy $y'(t)=-ty(t)^2$, $y(-1)=2/3$, diventa: $\eta(-1)=y(-1)=2/3$, per $i=0,1,2,\ldots$

$$\eta(x_i+h)=\eta(x_i)+h\Big[\frac{1}{4}(-x_i)\eta(x_i)^2+\frac{3}{4}(-(x_i+\frac{2}{3}h))(\eta(x_i)+\frac{2}{3}h(-x_i)\eta(x_i)^2)^2\Big],\ x_{i+1}=x_i+h.$$

Per $h=1$, facendo un solo passo ($i=0$) si ottiene la seguente approssimazione di $y(0)$:

$$\begin{aligned}
\eta(0;1) &= \eta(-1)+\tfrac{1}{4}(-(-1))\eta(-1)^2+\tfrac{3}{4}(-(-1+\tfrac{2}{3}))(\eta(-1)+\tfrac{2}{3}(-(-1))\eta(-1)^2)^2\\
&= \tfrac{2}{3}+\tfrac{1}{4}\tfrac{4}{9}+\tfrac{3}{4}\tfrac{1}{3}(\tfrac{2}{3}+\tfrac{2}{3}\tfrac{4}{9})^2\\
&= \tfrac{2}{3}+\tfrac{1}{9}+\tfrac{1}{4}\tfrac{4}{9}(1+\tfrac{4}{9})^2\\
&= \tfrac{2}{3}+\tfrac{1}{9}+\tfrac{1}{9}\tfrac{169}{81}\\
&= \tfrac{2}{3}+\tfrac{1}{9}\tfrac{81+169}{81}=\tfrac{736}{729}=1.0096..\ .
\end{aligned}$$

L'approssimazione ottenuta di $y(0) = 1$ è molto migliore di quella fornita dal metodo di Eulero (usando sempre un passo $h = 1$), che era 10/9 (vedi Lezione).

(4) Dimostrare che il metodo: $\eta(x_0) = y_0$, per $i = 0, 1, \ldots$

$$K_1 = f(x_i, \eta(x_i; h)), \quad K_2 = f(x_i + h, \eta(x_i; h) + \frac{1}{2}hK_1 + \frac{1}{2}hK_2),$$

$$\eta(x_i + h; h) = \eta(x_i; h) + \frac{1}{2}h(K_1 + K_2), \quad x_{i+1} = x_i + h,$$

è convergente quando applicato al problema di Cauchy $y'(t) = \mu y(t)$, $t \in \mathbb{R}$, $y(0) = 1$. Dimostrare inoltre che se $\mu < 0$ allora, per h fissato maggiore di zero e opportunamente piccolo, la successione di approssimazioni $\eta(x_i)$ ha lo stesso andamento della successione di valori esatti $y(x_i) = e^{\mu x_i}$, $i = 0, 1, \ldots$.

Nota: il metodo considerato in questo esercizio funziona bene anche se applicato al problema (B) introdotto a lezione (il problema sul quale Eulero per funzionare richiedeva un passo troppo piccolo, considerato l'andamento regolare della soluzione). Verificare questa affermazione (vedi Appendice 2).

Risoluzione

Applichiamo il metodo al problema. Si ha $K_1 = \mu\eta(x_i)$, $K_2 = \mu(\eta(x_i) + \frac{1}{2}h\mu\eta(x_i) + \frac{1}{2}hK_2)$, ovvero $K_2 = \mu\eta(x_i)(1 + \frac{1}{2}h\mu)/(1 - \frac{1}{2}h\mu)$, quindi

$$\eta(x_i + h) = \eta(x_i)\Big[1 + \frac{1}{2}h\mu + \frac{1}{2}h\mu(1 + \frac{1}{2}h\mu)/(1 - \frac{1}{2}h\mu)\Big],$$

$$\eta(x_{i+1}) = \eta(x_i)\Big[(1 + \frac{1}{2}h\mu)/(1 - \frac{1}{2}h\mu)\Big].$$

Quest'ultima identità e la condizione iniziale $\eta(x_0) = \eta(0) = y(0) = 1$ forniscono una formula esplicita per la successione $\eta(x_i) = \eta(x_i; h)$, $x_i = ih$, generata dal metodo:

$$\eta(x_i; h) = \Big(\frac{1 + \frac{1}{2}h\mu}{1 - \frac{1}{2}h\mu}\Big)^i. \qquad (*)$$

Convergenza: siano $x \neq x_0 = 0$ e $h_n = (x - x_0)/n = x/n$; posto $h = h_n$ nel metodo e fatti n passi con tale scelta di h, per la (*) si ottiene la seguente approssimazione in $x = x_n^{(h_n)}$ di $y(x) = e^{\mu x}$:

$$\eta(x; x/n) = \Big(\frac{1 + \frac{1}{2}x\mu/n}{1 - \frac{1}{2}x\mu/n}\Big)^n$$

che, per $n \to +\infty$, tende a $e^{\frac{1}{2}x\mu}/e^{-\frac{1}{2}x\mu} = e^{x\mu} = y(x)$.

Sia $\mu < 0$. Sia $h > 0$. La successione dei valori esatti $y(x_i) = y(ih) = e^{ih\mu}$, $i = 0, 1, 2, \ldots$, è composta da numeri reali positivi convergenti a zero in modo monotono decrescente. Usando la (*) si può dire che anche la successione $\eta(x_i) = \eta(ih)$, $i = 0, 1, 2, \ldots$, è composta da numeri reali positivi convergenti a zero in modo monotono decrescente a patto che il passo di integrazione h soddisfi la condizione $1 + \frac{1}{2}h\mu > 0$, ovvero se $h < -2/\mu$.

(Nota: si richiede un passo molto piccolo solo quando μ è molto negativo; ma questa è una richiesta ammissibile perché quando μ è molto negativo, per accorgersi della forte decrescita iniziale di $y(t)$, è ovviamente necessario scegliere un passo molto piccolo).

(5) La matrice

$$B = \begin{bmatrix} 2 & 1/3 & 1 \\ 3 & -5/3 & 1 \\ 0 & 11/9 & 5/3 \end{bmatrix}$$

ha gli autovalori -2, 1 e 3. Dimostrare che esiste ed è unico un vettore $\mathbf{z}$ tale che $B\mathbf{z} = 3\mathbf{z}$, $\mathbf{z} > \mathbf{0}$, $\|\mathbf{z}\|_1 = 1$ (senza scrivere esplicitamente $\mathbf{z}$!) e che si può utilizzare il metodo delle potenze per il calcolo di $\mathbf{z}$, generando una successione di vettori $\mathbf{z}_k$ ben definiti per ogni k e convergenti a $\mathbf{z}$. (Nota: la coppia di Perron può quindi esistere anche per matrici non non negative).

Risoluzione

La matrice $A = B + \frac{5}{3}I$ ha gli autovalori $-2 + \frac{5}{3}$, $1 + \frac{5}{3}$, $3 + \frac{5}{3}$, ed è non negativa irriducibile, quindi esiste un unico vettore positivo $\mathbf{z}$ tale che $A\mathbf{z} = \rho(A)\mathbf{z} = (3 + \frac{5}{3})\mathbf{z}$ e $\|\mathbf{z}\|_1 = 1$. Dunque esiste un unico vettore positivo $\mathbf{z}$ tale che $B\mathbf{z} = 3\mathbf{z}$ e $\|\mathbf{z}\|_1 = 1$. Per calcolare $\mathbf{z}$ si può applicare il metodo delle potenze alla matrice A: sia $\mathbf{z}_0 > \mathbf{0}$ tale che $\|\mathbf{z}_0\|_1 = 1$ e, per $k = 1, 2, \ldots$,

$$\mathbf{a}_k = A\mathbf{z}_{k-1}, \quad \varphi_k = \|\mathbf{a}_k\|_1, \quad \mathbf{z}_k = \frac{1}{\|\mathbf{a}_k\|_1}\mathbf{a}_k$$

(Nota: $\mathbf{a}_k > \mathbf{0}$, $\mathbf{z}_k > \mathbf{0}$, $\|\mathbf{z}_k\|_1 = 1$, $\forall\, k$). Poiché $r = \rho(A) = 3 + \frac{5}{3}$ domina gli altri due autovalori di A ($r > \max\{|-2+\frac{5}{3}|, |1+\frac{5}{3}|\}$), si ha che $\varphi_k \to \rho(A) = 3+\frac{5}{3}$ e $\mathbf{z}_k \to \mathbf{z}$. La velocità di convergenza è $O((\max\{|-2+\frac{5}{3}|/r, |1+\frac{5}{3}|/r\})^k) = O((\frac{4}{7})^k)$, essendo la matrice A diagonalizzabile. Questi risultati di convergenza sono veri per il teorema enunciato a lezione sul metodo delle potenze per A non negativa irriducibile, che non richiede alcuna ipotesi tranne la dominanza dell'autovalore semplice $\rho(A)$ di A sugli altri autovalori di A.

Nota. Perché non applicare il metodo delle potenze direttamente a B?

Sappiamo che $B\mathbf{z} = 3\mathbf{z}$. Siano $\mathbf{x}, \mathbf{y}$ non nulli tali che $B\mathbf{x} = \mathbf{x}$, $B\mathbf{y} = -2\mathbf{y}$. Sia $\mathbf{v}$ un qualsiasi vettore tale che nella espressione $\mathbf{v} = \alpha\mathbf{z}+\beta\mathbf{x}+\gamma\mathbf{y}$ il coefficiente α è diverso da zero. Il metodo delle potenze applicato a B pone $\mathbf{v}_0 = \mathbf{v}$, e per $k = 1, 2, \ldots$ pone $\mathbf{a}_k = B\mathbf{v}_{k-1}$, $\varphi_k = \mathbf{u}^H\mathbf{a}_k/\mathbf{u}^H\mathbf{v}_{k-1}$, $\mathbf{v}_k = \mathbf{a}_k/\|\mathbf{a}_k\|$. Tesi: per $k \to +\infty$ si ha $\varphi_k = \mathbf{u}^H B^k\mathbf{v}/\mathbf{u}^H B^{k-1}\mathbf{v} \to 3$, $\mathbf{v}_k = B^k\mathbf{v}/\|B^k\mathbf{v}\| \to \alpha\mathbf{z}/\|\alpha\mathbf{z}\|$.

Tuttavia la tesi appena enunciata è vera solo se sono verificate le seguenti ipotesi non facilmente controllabili $\alpha \neq 0$, $\mathbf{u}^H(\alpha\mathbf{z}) \neq 0$, $\mathbf{u}^H B^{k-1}\mathbf{v} \neq 0$ ($\mathbf{u}^H\mathbf{v}_{k-1} \neq 0$). Vedi il teorema enunciato a lezione sul metodo delle potenze per matrici generiche.

Dim. $\frac{1}{3^k}B^k\mathbf{v} = \alpha\mathbf{z} + \beta(\frac{1}{3})^k\mathbf{x} + \gamma(-\frac{2}{3})^k\mathbf{y} \to \alpha\mathbf{z}$, $\sigma_k = \frac{1}{3^k}\mathbf{u}^H B^k\mathbf{v} = \alpha\mathbf{u}^H\mathbf{z} + \beta(\frac{1}{3})^k\mathbf{u}^H\mathbf{x} + \gamma(-\frac{2}{3})^k\mathbf{u}^H\mathbf{y}$, $\sigma_{k+1}/\sigma_k \to 1$, $\mathbf{u}^H B^{k+1}\mathbf{v}/\mathbf{u}^H B^k\mathbf{v} \to 3$. ....

Nota. $\mathbf{z} = z_2[5/4 \;\; 1 \;\; 11/12]^T$ con z_2 tale che $\|\mathbf{z}\|_1 = 1$.

Appendice

Si usa la differenza finita (dII5punti), per $x = x_i$, per ogni $i = 1, \ldots, n$. In questo caso le equazioni $-y''(x_i) = q(x_i)$, $i = 1, \ldots, n$, ovvero le equazioni $-12h^2y''(x_i) = 12h^2q(x_i)$, $i = 1, \ldots, n$, diventano

$$y(x_{i-2})-16y(x_{i-1})+30y(x_i)-16y(x_{i+1})+y(x_{i+2})=12h^2q(x_i)+O(h^6),\quad i=1,\dots,n.$$

$$\begin{bmatrix} 30 & -16 & 1 & 0 & & & & 0 \\ -16 & 30 & -16 & 1 & & & & \\ 1 & -16 & 30 & & & & & \\ 0 & 1 & & & & & & \\ & & & & & & & 0 \\ & & & & & & -16 & 1 \\ & & & & 1 & -16 & 30 & -16 \\ 0 & & & & 0 & 1 & -16 & 30 \end{bmatrix} \begin{bmatrix} y(x_1) \\ y(x_2) \\ y(x_3) \\ . \\ . \\ y(x_{n-2}) \\ y(x_{n-1}) \\ y(x_n) \end{bmatrix} = \begin{bmatrix} 16\alpha - y(x_{-1}) \\ -\alpha \\ 0 \\ . \\ . \\ 0 \\ -\beta \\ 16\beta - y(x_{n+2}) \end{bmatrix} +12h^2 \begin{bmatrix} q(x_1) \\ q(x_2) \\ . \\ . \\ . \\ . \\ q(x_{n-1}) \\ q(x_n) \end{bmatrix} +O(h^6). \qquad (*)$$

Se si suppone y (prolungabile in modo) sufficientemente regolare anche in un intorno sinistro di a e in un intorno destro di b, si può pensare di sostituire, nella prima e nell'ultima equazione, $y(x_{-1})$ e $y(x_{x_{n+2}})$ rispettivamente con il valore in x_{-1} del polinomio interpolante y in $x_0, x_1, \dots$ e con il valore in x_{n+2} del polinomio interpolante y in $x_{n+1}, x_n, \dots$. (Per richiami sull'interpolazione polinomiale vedi la fine di questa Appendice).

Sia $p_r^{(s)}$ il polinomio interpolante y in $x_0, x_1, \dots, x_r$. È semplice verificare che esiste $\xi_s \in (x_{-1}, x_r)$ tale che $y(x_{-1}) - p_r^{(s)}(x_{-1}) = (-1)^{r+1}h^{r+1}y^{(r+1)}(\xi_s)$. Usando la formula di Lagrange per $p_r^{(s)}(x)$ si ottengono facilmente le seguenti identità:

$$\begin{aligned}
&y(x_{-1}) = p_r^{(s)}(x_{-1}) + O(h^{r+1}) : \\
&r=0;\ y(x_{-1}) = y(x_0) + O(h), \\
&r=1;\ y(x_{-1}) = 2y(x_0) - y(x_1) + O(h^2), \\
&r=2;\ y(x_{-1}) = 3y(x_0) - 3y(x_1) + y(x_2) + O(h^3), \\
&r=3;\ y(x_{-1}) = 4y(x_0) - 6y(x_1) + 4y(x_2) - y(x_3) + O(h^4), \\
&r=4;\ y(x_{-1}) = 5y(x_0) - 10y(x_1) + 10y(x_2) - 5y(x_3) + y(x_4) + O(h^5), \\
&r=5;\ y(x_{-1}) = 6y(x_0) - 15y(x_1) + 20y(x_2) - 15y(x_3) + 6y(x_4) - y(x_5) + O(h^6)
\end{aligned}$$

(i passaggi sono descritti nei dettagli più avanti nei casi $r=3$ e $r=5$). Utilizzando tali espressioni di $y(x_{-1})$ si vede che la prima equazione in (*)

$$30y(x_1) - 16y(x_2) + y(x_3) = 16\alpha - y(x_{-1}) + 12h^2q(x_1) + O(h^6)$$

è equivalente rispettivamente alle seguenti equazioni:

$$\begin{aligned}
&30y(x_1) - 16y(x_2) + y(x_3) = 15\alpha + O(h) + 12h^2q(x_1) + O(h^6), \\
&29y(x_1) - 16y(x_2) + y(x_3) = 14\alpha + O(h^2) + 12h^2q(x_1) + O(h^6), \\
&27y(x_1) - 15y(x_2) + y(x_3) = 13\alpha + O(h^3) + 12h^2q(x_1) + O(h^6), \\
&24y(x_1) - 12y(x_2) = 12\alpha + O(h^4) + 12h^2q(x_1) + O(h^6), \\
&20y(x_1) - 6y(x_2) - 4y(x_3) + y(x_4) = 11\alpha + O(h^5) + 12h^2q(x_1) + O(h^6), \\
&15y(x_1) + 4y(x_2) - 14y(x_3) + 6y(x_4) - y(x_5) = 10\alpha + O(h^6) + 12h^2q(x_1) + O(h^6).
\end{aligned}$$

Nota: dalle identità appena ottenute, ricordando che $q(x_1) = -y''(x_1)$, si ottengono le seguenti differenze finite approssimanti $y''(x_1)$

$$y''(x_1) = \frac{1}{12h^2}\left\{\begin{array}{l} 13y(x_0) - 27y(x_1) + 15y(x_2) - y(x_3) \\ 12y(x_0) - 24y(x_1) + 12y(x_2) \\ 11y(x_0) - 20y(x_1) + 6y(x_2) + 4y(x_3) - y(x_4) \\ 10y(x_0) - 15y(x_1) - 4y(x_2) + 14y(x_3) - 6y(x_4) + y(x_5) \end{array}\right. + \left\{\begin{array}{l} O(h) \\ O(h^2) \\ O(h^3) \\ O(h^4) \end{array}\right.$$

Esempio ($r = 3$): Sostituire $y(x_{-1})$ e $y(x_{n+2})$ con il valore in x_{-1} di

$$\begin{aligned} p_3^s(x) \;=\; & y(x_0)\tfrac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)} + y(x_1)\tfrac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)} \\ & + y(x_2)\tfrac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)} + y(x_3)\tfrac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)} \end{aligned}$$

e il valore in x_{n+2} di

$$\begin{aligned} p_3^d(x) \;=\; & y(x_{n+1})\tfrac{(x-x_n)(x-x_{n-1})(x-x_{n-2})}{(x_{n+1}-x_n)(x_{n+1}-x_{n-1})(x_{n+1}-x_{n-2})} + y(x_n)\tfrac{(x-x_{n+1})(x-x_{n-1})(x-x_{n-2})}{(x_n-x_{n+1})(x_n-x_{n-1})(x_n-x_{n-2})} \\ & + y(x_{n-1})\tfrac{(x-x_{n+1})(x-x_n)(x-x_{n-2})}{(x_{n-1}-x_{n+1})(x_{n-1}-x_n)(x_{n-1}x_{n-2})} + y(x_{n-2})\tfrac{(x-x_{n+1})(x-x_n)(x-x_{n-1})}{(x_{n-2}-x_{n+1})(x_{n-2}-x_n)(x_{n-2}x_{n-1})} : \end{aligned}$$

$y(x_{-1})$ viene sostituito con $y(x_0)4 - y(x_1)6 + y(x_2)4 - y(x_3)$ e $y(x_{n+2})$ viene sostituito con $y(x_{n+1})4 - y(x_n)6 + y(x_{n-1})4 - y(x_{n-2})$

(errori: $|y^{\prime v}(\xi_s)|h^4$, $\xi_s \in (x_{-1}, x_3)$, e $|y^{\prime v}(\xi_d)|h^4$, $\xi_d \in (x_{n-2}, x_{n+2})$). Quindi:

$$\begin{aligned} & y(x_{-1}) = y(x_0)4 - y(x_1)6 + y(x_2)4 - y(x_3) + O(h^4), \\ & y(x_{n+2}) = y(x_{n+1})4 - y(x_n)6 + y(x_{n-1})4 - y(x_{n-2}) + O(h^4). \end{aligned}$$

Nota: si osserva che usando queste espressioni per $y(x_{-1})$ e $y(x_{n+2})$ in (*) si ottiene proprio l'equazione vettoriale già trovata (risolvendo in modo semplice l'esercizio) approssimando $-y''(x_1) = q(x_1)$ e $-y''(x_n) = q(x_n)$ con la (dIItrepunti).

Questa osservazione ci fa capire che ottenendo formule del tipo $y(x_{-1}) = (...) + O(h^6)$ e $y(x_{n+2}) = (...) + O(h^6)$ — ad esempio scrivendo i due polinomi interpolanti y rispettivamente nei sei punti x_i, $i = 0, \ldots, 5$, e nei sei punti x_{n+1-i}, $i = 0, \ldots, 5$ e prendendo $(...)$ = valori di tali polinomi in x_{-1} e x_{n+2} — si otterrebbe allo stesso tempo una approssimazione di $y''(x)$ con errore $O(h^4)$ in termini di $y(x-h), y(x), y(x+h), y(x+2h), y(x+3h), y(x+4h)$.

$r = 5$: Quindi, si vuole sostituire $y(x_{-1})$ e $y(x_{n+2})$ con il valore in x_{-1} di $p_5^s(x)$, il polinomio interpolante la funzione $y(x)$ nei punti x_i, $i = 0, \ldots, 5$, e il valore in x_{n+2} di $p_5^d(x)$, il polinomio interpolante la funzione $y(x)$ nei punti x_{n+1-i}, $i = 0, \ldots, 5$:

$y(x_{-1}) - p_5^s(x_{-1}) = y^{vi}(\xi_s)h^6$, $\xi_s \in (x_{-1}, x_5)$,

$y(x_{n+2}) - p_5^d(x_{n+2}) = y^{vi}(\xi_d)h^6$, $\xi_d \in (x_{n-4}, x_{n+2})$,

$y(x_{-1}) = 6y(x_0) - 15y(x_1) + 20y(x_2) - 15y(x_3) + 6y(x_4) - y(x_5) + O(h^6)$

(quest'ultima formula si ottiene facilmente scrivendo $p_5^s(x) = \sum_{i=0}^5 y(x_i)\frac{\Pi_{j\neq i}(x-x_j)}{\Pi_{j\neq i}(x_i-x_j)}$ per esteso e sostituendo, nell'espressione ottenuta, x con x_{-1}),

$y(x_{n+2}) = 6y(x_{n+1}) - 15y(x_n) + 20y(x_{n-1}) - 15y(x_{n-2}) + 6y(x_{n-3}) - y(x_{n-4}) + O(h^6)$.

Sostituendo l'espressione di $y(x_{-1})$ nella prima equazione del sistema (*), si ottiene

$30y_1 - 16y_2 + y_3 - 15y_1 + 20y_2 - 15y_3 + 6y_4 - y_5 + O(h^6) = 10\alpha + 12h^2 q_1 + O(h^6)$

da cui l'identità: $15y_1 + 4y_2 - 14y_3 + 6y_4 - y_5 = 10\alpha + 12h^2 q_1 + O(h^6)$. Idem con l'espressione di $y(x_{n+2})$.

Dunque il sistema (*) diventa

$$\begin{bmatrix} 15 & 4 & -14 & 6 & -1 & 0 & \cdot & 0 \\ -16 & 30 & -16 & 1 & 0 & & & \cdot \\ 1 & -16 & 30 & & & \cdot & & \\ 0 & 1 & & & & & 0 & \cdot \\ \cdot & 0 & & & & & 1 & 0 \\ & & \cdot & & & & -16 & 1 \\ \cdot & & & 0 & 1 & -16 & 30 & -16 \\ 0 & \cdot & 0 & -1 & 6 & -14 & 4 & 15 \end{bmatrix} \begin{bmatrix} y(x_1) \\ y(x_2) \\ y(x_3) \\ \cdot \\ \cdot \\ y(x_{n-2}) \\ y(x_{n-1}) \\ y(x_n) \end{bmatrix} = \begin{bmatrix} 10\alpha \\ -\alpha \\ 0 \\ \cdot \\ \cdot \\ 0 \\ -\beta \\ 10\beta \end{bmatrix} + 12h^2 \begin{bmatrix} q(x_1) \\ q(x_2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ q(x_{n-1}) \\ q(x_n) \end{bmatrix} + O(h^6).$$

Esercizio. Sia $A\eta = \mathbf{b}$ il sistema lineare che si ottiene da quest'ultima uguaglianza vettoriale sostituendo $y(x_i)$ con $\eta(x_i)$ $(i = 1, \ldots, n)$ e togliendo gli errori $O(h^6)$. La matrice A è invertibile? Quale metodo è più conveniente usare per calcolare $\eta = A^{-1}\mathbf{b}$?

Si osserva inoltre che $15y_1 + 4y_2 - 14y_3 + 6y_4 - y_5 = 10\alpha + 12h^2 q_1 + O(h^6) = 10\alpha - 12h^2 y''(x_1) + O(h^6)$, da cui si deduce che deve valere la seguente uguaglianza:

$$y''(x_1) = \frac{10y_0 - 15y_1 - 4y_2 + 14y_3 - 6y_4 + y_5}{12h^2} + O(h^4),$$

ovvero si è ottenuta la differenza finita non centrata che si cercava.

Esercizio. È possibile ottenere formule del tipo

$$y''(x_1) = \varphi(y_0, y_1, y_2, y_3, h) + O(h^4), \qquad y''(x_1) = \varphi(y_0, y_1, y_2, y_3, y_4, h) + O(h^4) \quad ?$$

Cenni sull'interpolazione polinomiale

(1) Il polinomio $p_r(x) = \sum_{i=0}^r y_i l_i(x)$, $l_i(x) = \left(\Pi_{j=0,\, j\neq i}^r (x - x_j)\right) \Big/ \left(\Pi_{j=0,\, j\neq i}^r (x_i - x_j)\right)$, con x_i reali distinti e y_i reali, ha grado minore o uguale di r ed è tale che $p_r(x_k) = y_k$, $k = 0, 1, \ldots, r$ (dimostrarlo!).

(2) Se p è un polinomio tale che $p(x_k) = y_k$, $\forall\, k$, e $\partial p \leq r$, allora $p = p_r$ (dimostrarlo!).

(3) Il polinomio p_r è detto polinomio interpolante la tabella (x_k, y_k), $k = 0, 1, \ldots, r$. La sua rappresentazione in (1) è detta di Lagrange perché in termini della particolare base di Lagrange $l_i(x)$, $i = 0, 1, \ldots, r$, dell'insieme dei polinomi di grado minore o uguale di r.

(4) Se $y_k = f(x_k)$ con f sufficientemente regolare allora esiste ξ nel minimo intervallo contenente $x, x_0, x_1, \ldots, x_r$ tale che $f(x) - p_r(x) = \frac{1}{(r+1)!} f^{(r+1)}(\xi) \Pi_{j=0}^r (x - x_j)$.

Dimostrazione. Posto $F(t) = (f(t) - p_r(t))\Pi_{j=0}^r (x - x_j) - (f(x) - p_r(x))\Pi_{j=0}^r (t - x_j)$, si osserva che $F(x) = 0$, $F(x_k) = 0\ \forall\, k$. Supponendo $x \neq x_k\ \forall\, k$ (per $x = x_k$ per qualche k la tesi è banalmente vera), tali uguaglianze implicano che esiste ξ nel minimo intervallo contenente $x, x_0, x_1, \ldots, x_r$ tale che $f^{(r+1)}(\xi) = 0$. La tesi segue osservando che $f^{(r+1)}(t) = f^{(r+1)}(t)\Pi_{j=0}^r (x - x_j) - (f(x) - p_r(x))(r+1)!$.

Appendice 2

Applicando il metodo Runge-Kutta implicito introdotto nell'esercizio (4) al problema

$$(B)\ \ y'(t) = \lambda y(t) + (\mu - \lambda)e^{\mu t}, \quad y(0) = 1 \ \ (\lambda < 0,\ \mu < 0)$$

dopo alcuni passaggi si dimostra che

$$\eta((i+1)h) = \eta(ih)\frac{1+\frac{1}{2}h\lambda}{1-\frac{1}{2}h\lambda} + s\,e^{\mu i h}, \quad s = \frac{1}{2}h(\mu-\lambda)\frac{1+e^{\mu h}}{1-\frac{1}{2}h\lambda}.$$

Risolvendo questa equazione alle differenze non omogenea con condizione iniziale $\eta(0) = 1$, si ottiene la seguente formula esplicita per $\eta(x_i) = \eta(ih)$:

$$\eta(ih) = (1-r_h)\Big(\frac{1+\frac{1}{2}h\lambda}{1-\frac{1}{2}h\lambda}\Big)^i + r_h e^{\mu i h}, \quad r_h = \frac{1}{2}\frac{h(\mu-\lambda)(1+e^{\mu h})}{e^{\mu h} - 1 - \frac{1}{2}h\lambda(e^{\mu h}+1)}$$

(Nota: r_h tende a 1 quando $h \to 0$).

Usando questa formula è facile vedere che $\eta(x; x/n) \to y(x) = e^{\mu x}$, se $n \to +\infty$, cioè il metodo è convergente quando usato per risolvere il problema (B).

Inoltre, è evidente che, per h fissato piccolo (in modo che r_h è circa 1), la successione di valori approssimati $\eta(ih)$ ha lo stesso andamento della successione dei valori esatti $y(ih) = e^{\mu i h}$ a patto che $|(1+\frac{1}{2}h\lambda)/(1-\frac{1}{2}h\lambda)| = |1+\frac{1}{2}h\lambda|/(1-\frac{1}{2}h\lambda) < 1$. Ma questa disuguaglianza è sempre verificata (per ogni $h > 0$) nel nostro caso in cui $\lambda < 0$, anche se $\lambda << 0$.

Quindi, come previsto (dato il suo carattere implicito), il metodo dell'esercizio (4) funziona bene sul problema (B), con $\mu < 0$ moderato (ad es. $\mu = -1$) e $\lambda << 0$, dove invece il classico metodo di Eulero (esplicito) fallisce, nel senso che per funzionare richiede un passo troppo piccolo se confrontato con l'andamento regolare della soluzione esatta $e^{\mu t}$.

III esonero di Elementi di Analisi Numerica (19 Dicembre 2013)

(1) Data y funzione sufficientemente regolare in un intorno di x e preso $h > 0$ abbastanza piccolo, trovare una differenza finita $\varphi\Big(y(x-2h), y(x-h), y(x), y(x+h), y(x+2h)\Big)$ tale che

$$y''(x) = \varphi\Big(y(x-2h), y(x-h), y(x), y(x+h), y(x+2h)\Big) + O(h^4).$$

Usare tale differenza finita per introdurre un sistema lineare $A\eta = \mathbf{b}$, A $n\times n$, $\mathbf{b}$ $n\times 1$, la cui soluzione η ha come componente i-esima una approssimazione $\eta(x_i)$ di $y(x_i)$ essendo $y(x)$ la soluzione del problema differenziale $-y''(x) = q(x)$, $x \in (a, b)$, $y(a) = \alpha$, $y(b) = \beta$, e $x_i = a + i(b-a)/(n+1)$, $i = 0, 1, \ldots, n, n+1$.

(2) Dimostrare che se f è tale che $|f(x, y) - f(x, \tilde{y})| \leq L|y - \tilde{y}|$ per ogni $x \in [a, b]$, $y, \tilde{y} \in \mathbb{R}$ ($L \geq 0$), allora ogni schema di Runge-Kutta

$$\phi(x, y; h) = a_1 f(x, y) + a_2 f(x + p_1 h, y + p_2 h f(x, y))$$

è tale che $|\phi(x, y; h) - \phi(x, \tilde{y}; h)| \leq N|y - \tilde{y}|$ per ogni $x \in [a, b]$, $y, \tilde{y} \in \mathbb{R}$, $|h| \leq h_0$, dove N dipende da L, h_0 e dai parametri dello schema.

(3) Dimostrare che lo schema di Runge-Kutta

$$\frac{\eta(x+h) - y}{h} = \frac{1}{4} f(x, y) + \frac{3}{4} f(x + \frac{2}{3}h, y + \frac{2}{3} h f(x, y))$$

è di ordine due ma non di ordine tre. Calcolare con tale schema, ponendo $h = 1$, una approssimazione $\eta(0)$ di $y(0)$ essendo $y(t)$ la soluzione del problema di Cauchy $y'(t) = -ty(t)^2$, $y(-1) = 2/3$. Facoltativo: confrontare il risultato con l'approssimazione ottenuta con lo schema di Eulero.

(4) Dimostrare che il metodo: $\eta(x_0) = y_0$, per $i = 0, 1, \ldots$

$$K_1 = f(x_i, \eta(x_i; h)), \quad K_2 = f(x_i + h, \eta(x_i; h) + \frac{1}{2} h K_1 + \frac{1}{2} h K_2),$$

$$\eta(x_i + h; h) = \eta(x_i; h) + \frac{1}{2} h (K_1 + K_2), \quad x_{i+1} = x_i + h,$$

è convergente quando applicato al problema di Cauchy $y'(t) = \mu y(t)$, $t \in \mathbb{R}$, $y(0) = 1$. Dimostrare inoltre che se $\mu < 0$ allora, per h fissato maggiore di zero e opportunamente piccolo, la successione di approssimazioni $\eta(x_i)$ ha lo stesso andamento della successione di valori esatti $y(x_i) = e^{\mu x_i}$, $i = 0, 1, \ldots$.

(5) La matrice

$$B = \begin{bmatrix} 2 & 1/3 & 1 \\ 3 & -5/3 & 1 \\ 0 & 11/9 & 5/3 \end{bmatrix}$$

ha gli autovalori -2, 1 e 3. Dimostrare che esiste ed è unico un vettore $\mathbf{z}$ tale che $B\mathbf{z} = 3\mathbf{z}$, $\mathbf{z} > \mathbf{0}$, $\|\mathbf{z}\|_1 = 1$ (senza scrivere esplicitamente $\mathbf{z}$!) e che si può utilizzare il metodo delle potenze per il calcolo di $\mathbf{z}$, generando una successione di vettori $\mathbf{z}_k$ ben definiti per ogni k e convergenti a $\mathbf{z}$. (Nota: la coppia di Perron può quindi esistere anche per matrici non non negative).

Completamento di quanto scritto nei miei appunti al Focal Point sulla risoluz. numerica di problemi diff. di Cauchy

Sul teorema di convergenza di un metoso Φ:

Le ipotesi del Teorema erano: $\Phi(x, y; h)$ Lip. rispetto a y e Φ di ordine p. La prima ipotesi è di solito soddisfatta senza richiedere nessuna regolarità su f. Infatti l'unica ipotesi richiesta su f è quella di Lip. rispetto al suo secondo argomento, e tale ipotesi è comunque essenziale affinché il problema di Cauchy abbia una unica soluzione [Ricorda l'esempio $y'(t) = \sqrt{|t|}$, $y(0) = 0$ insegnato ad Analisi]:

Esercizio: verificare che se $f(x, y)$ è Lip. rispetto a y sulla striscia $[a, b] \times \mathbb{R}$, cioè

$$|f(x, y) - f(x, \tilde{y})| \leq L|y - \tilde{y}|$$

per ogni $x \in [a, b]$, $y, \tilde{y} \in \mathbb{R}$, allora $\Phi(x, y; h)$ definita come nei metodi Runge-Kutta (di ordine minore o uguale a due) è Lip. rispetto a y sulla striscia $[a, b] \times \mathbb{R}$, cioè esiste N tale che

$$|\Phi(x, y; h) - \Phi(x, \tilde{y}; h)| \leq N|y - \tilde{y}|$$

per ogni $x \in [a, b]$, $\forall\, y, \tilde{y} \in \mathbb{R}$, $\forall\, h$ tale che $|h| \leq h_0$. E se Φ è definita come nei metodi di Taylor?

La seconda ipotesi, Φ di ordine p, è effettivamente verificata solo se Φ viene applicato a un problema di Cauchy dove f è sufficientemente regolare (ad es. la Φ che definisce il metodo di Eulero è effettivamente di ordine uno se la f, la f_x e la f_y sono tutte e tre, oltre che ben definite, limitate sulla striscia $[a, b] \times \mathbb{R}$). Tuttavia, se si va a guardare la dimostrazione del Teorema di Convergenza, se la Φ fosse solo tale che

$$|\Delta(x, y; h) - \Phi(x, y; h)| = \left|\frac{y(x + h) - y}{h} - \Phi(x, y; h)\right| \;\to\; 0, \;\; se\ h \to 0,$$

per ogni $x \in (a, b)$, $y \in \mathbb{R}$, h tale che $x + h \in [a, b]$ (per ogni f continua e Lip. in $[a, b] \in \mathbb{R}$), dove $y(t)$ è la soluzione (definita univocamente) del problema $y'(t) = f(t, y(t))$, $t \in [a, b]$, $y(x) = y$, allora si avrebbe lo stesso il risultato di convergenza per Φ, non si avrebbero però risultati sulla rapidità di convergenza. Infatti, si avrebbe

$$|\xi_{i+1}| \leq |\xi_i|(1 + |h|N) + |g(h)|, \;\; \lim_{h\to 0} \frac{g(h)}{h} = 0,$$

($\xi_i = \eta(x_i; h) - y(x_i)$) e, dal Lemma visto a lezione, la disuguaglianza

$$|\xi_i| \leq \frac{e^{i|h|N} - 1}{|h|N}|g(h)|,$$

che per $h = h_n = (x - x_0)/n$ e $i = n$ implica

$$|\eta(x; h_n) - y(x)| \leq \frac{e^{|x-x_0|N} - 1}{|h_n|N}|g(h_n)|.$$

Dunque $|\eta(x; h_n) - y(x)| \to 0$ se $n \to +\infty$.

Osservazione. È comunque meglio che un metodo sia effettivamente di ordine p (cioè sia abbastanza preciso) su problemi con f regolare, piuttosto che non sia di ordine p su nessun problema. A meno che non si trovi un modo per introdurre Φ precisi su qualsiasi problema; per farlo occorrerebbe innanzitutto definire cosa si intende per "precisi".

Inoltre, data una Φ definita in termini di parametri è chiaro che questi parametri vanno scelti in modo che il metodo Φ (oltre ad essere convergente) sia più preciso possibile, almeno su una classe di problemi. Un modo è imporre che Φ sia di ordine p (secondo la definizione di "ordine p" data a lezione); in tal caso il metodo fornirà la massima precisione quando applicato a problemi la cui soluzione $y(t)$ è derivabile almeno $p+1$ volte (ovvero, f derivabile parzialmente p volte). Ma non è affatto escluso che ci possono essere altri modi per scegliere i parametri, che magari garantiscono la precisione di Φ anche su una classe di problemi di Cauchy dove la f non è regolare.

Ancora su Runge-Kutta:

Ricapitolando, i metodi Runge-Kutta per essere applicati e per convergere non richiedono che la f sia regolare più di quanto serve affinché la soluzione del problema di Cauchy, definito in termini di f, esista e sia unica (e quindi abbia senso cercare di calcolarla). Se però vogliamo avere anche informazioni sulla loro rapidità di convergenza, allora (almeno per quanto visto a lezione) le possiamo avere solo se la f del problema al quale li applichiamo è piu regolare del necessario. Infatti il risultato visto a lezione, $|\eta(x;h_n)-y(x)| \leq C_x|h_n|^p$, vale solo se il metodo è effettivamente di ordine p, quando applicato al problema di Cauchy di cui $y(t)$ è soluzione, e cioè quando $y(t)$ è derivabile fino all'ordine $p+1$.

Come visto a lezione, altra caratteristica importante dei metodi Runge-Kutta è che possono essere applicati facilmente nella risoluzione di problemi di Cauchy relativi a sistemi di equazioni differenziali.

Infine, una caratteristica dei metodi Runge-Kutta, comune a tutti i metodi one-step, è che usandoli, quando si vuole, si può modificare senza problemi il "passo di integrazione" h (ciò non è vero, come è facilmente intuibile, per i metodi multi-step, dove $\eta(x_i+h)$ è definito oltre che in termini di $\eta(x_i)$ anche in termini di $\eta(x_{i-1}), \ldots$).

È importante poter cambiare il passo facilmente, perché questo va fatto non appena ci si accorge che le $\eta(x_i)$, prodotte dal metodo, stanno approssimando troppo bene i valori esatti $y(x_i)$ (in tal caso, si aumenta h), oppure non stanno approssimando bene gli $y(x_i)$ (in quest'ultimo caso si diminuisce il passo h). Ci si può accorgere di questa troppa o poca precisione mediante criteri che permettono di stimare l'errore $|y(x_i)-\eta(x_i)|$. Ad esempio, periodicamente (ogni tot passi) si può applicare per un solo passo (contemporaneamente al metodo Φ) un metodo $\tilde{\Phi}$ molto più preciso, di ordine maggiore, di Φ, e usare la quantità $|\eta(x_i)-\tilde{\eta}(x_i)|$ come stima di $|y(x_i)-\eta(x_i)|$.

Riportiamo qui sotto per completezza la struttura di un generico metodo Φ Runge-Kutta:

$$K_1 = f(x,y), \qquad K_2 = f(x+p_1h, y+p_2hK_1),$$

$$K_3 = f(x+p_3h, y+p_4hK_1+p_5hK_2), \qquad K_4 = f(x+p_6h, y+p_7hK_1+p_8hK_2+p_9hK_3), \quad \ldots$$

$$\Phi(x,y;h) = \Phi^{RK}(x,y;h) := \sum_{j=1}^{s} a_jK_j,$$

$$\frac{\eta(x+h)-y}{h} = \Phi(x,y;h).$$

I parametri, come abbiamo detto sopra, andranno scelti in modo che $\Phi(x,y;h)$ approssimi $\Delta(x,y;h)$ (ovvero $\eta(x+h)$ approssimi $y(x+h)$) il meglio possibile, almeno su una classe di problemi di Cauchy.

Il caso $s = 2$ è stato esaminato a lezione. È stato dimostrato che ci sono infiniti valori dei parametri che rendono Φ di ordine due. Si può dimostrare che non esistono valori dei parametri per cui Φ, $s = 2$, è di ordine tre.

Notiamo che anche per $s = 4$ ci sono tante scelte dei parametri che rendono Φ di ordine quattro, dunque esistono tanti metodi Runge-Kutta di ordine quattro. Tra questi c'è il famoso metodo Runge-Kutta di ordine quattro (quello che ha dato il nome a tutti gli altri), riportato qui sotto:

$$\begin{aligned}
K_1 &= f(x,y),\\
K_2 &= f(x+\frac{1}{2}h, y+\frac{1}{2}hK_1),\\
K_3 &= f(x+\frac{1}{2}h, y+\frac{1}{2}hK_2),\\
K_4 &= f(x+h, y+hK_3),\\
\Phi(x,y;h) &= \frac{1}{6}\Big(K_1+2K_2+2K_3+K_4\Big),\\
\eta(x+h) &= y+h\Phi(x,y;h).
\end{aligned}$$

Notare che con questo metodo, con sole quattro valutazioni di f per passo (in quattro punti diversi), si ottiene la stessa precisione che si ottiene con un metodo di Taylor di ordine quattro, che però richiede la valutazione (sempre nello stesso punto) di f, f_x, f_y, f_{xx}, f_{xy}, f_{yy}, f_{xyy}, f_{xxy}, f_{xxx}, f_{yyy}.

Esercizio. Scrivere il metodo Runge-Kutta di ordine quattro nella forma $\eta(x_i+h) = \eta(x_i) + h\Phi(x_i, \eta(x_i); h)$. Applicarlo a problemi di Cauchy relativi a sistemi di due equazioni differenziali.

Esercizio. Applicare i metodi di Taylor e di Runge-Kutta di ordine due con $h = 1$ e $h = 1/2$ per ottenere approssimazioni di $y(0)$ essendo $y(t)$ la soluzione del problema $y'(t) = -ty(t)^2$, $y(-1) = 2/3$. Confrontare tra loro i risultati. Valutare allo stesso modo $y(2)$ essendo $y(t)$ la soluzione del problema $y'(t) = y(t)^2$, $y(1) = -1$.

Esercizio (facoltativo). Scrivere il sistema delle equazioni che devono verificare i parametri di un metodo Φ di Runge-Kutta affinché esso sia di ordine almeno tre (scegliere $s = 3$). Notare che il sistema è non lineare e che il numero delle sue equazioni è minore del numero dei parametri.

Sull'argomento vedi i libri di Stoer, Lambert, Butcher. In particolare Butcher, molto specialistico, sviluppa una teoria che permette di ottenere facilmente le equazioni che devono essere soddisfatte da un schema Φ di tipo Runge-Kutta affinché esso sia di un dato ordine.

Dicembre 2013 (Ho scritto queste tre pagine anche per rispondere a domande di Fabio, Giulio, Gianbattista)

Correzione II esonero Elementi di Analisi Numerica 2013-2014

(1) Dimostrare che, per ogni $a, b, c \in \mathbb{R}$, almeno un autovalore della matrice

$$A = \begin{bmatrix} 1 & -2 & 1 & -\frac{1}{2} \\ -2 & 4 & -11/6 & 5/6 \\ 1 & -11/6 & a & b\mathbf{i} \\ -\frac{1}{2} & 5/6 & -b\mathbf{i} & c \end{bmatrix}$$

è nell'intervallo $(-1/9, 1/9)$. (Suggerimento: due colonne di A sono quasi linearmente dipendenti).

Risoluzione

A è hermitiana, dunque normale. Ne segue che per ogni $\mathbf{x} \in \mathbb{C}^n$ $\mathbf{x} \neq \mathbf{0}$ il cerchio di centro $\mathbf{x}^H A\mathbf{x}/\mathbf{x}^H\mathbf{x}$ e raggio $R_{\mathbf{x}} = \sqrt{(\|A\mathbf{x}\|_2/\|\mathbf{x}\|_2)^2 - |\mathbf{x}^H A\mathbf{x}/\mathbf{x}^H\mathbf{x}|^2}$ contiene almeno un autovalore di A. Se riusciamo a scegliere un $\mathbf{x}$ per cui $\|A\mathbf{x}\|_2/\|\mathbf{x}\|_2$ è piccolo, avremo localizzato bene almeno un autovalore di A. A tal scopo è naturale sfruttare il fatto che le prime due colonne di A sono quasi dipendenti, cioè esiste $\mathbf{x} \neq \mathbf{0}$ per cui $A\mathbf{x}$ è vicino al vettore nullo. Scegliendo dunque $\mathbf{x} = [2\ 1\ 0\ 0]^T$ si ottiene

$$A\mathbf{x} = [0\ 0\ 1/6\ \ -1/6]^T,\ \|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^H\mathbf{x}} = \sqrt{5},\ \|A\mathbf{x}\|_2 = \sqrt{1/18},$$
$$\mathbf{x}^H A\mathbf{x}/\mathbf{x}^H\mathbf{x} = 0, \quad R_{\mathbf{x}} = \|A\mathbf{x}\|_2/\|\mathbf{x}\|_2 = \sqrt{1/(18 \cdot 5)}.$$

Quindi nel cerchio in $\mathbb{C}$ di centro $(0, 0)$ e raggio $\sqrt{1/90}$ c'è almeno un autovalore di A. Ma gli autovalori di A sono reali (perché A è hermitiana), quindi si conclude che nell'intervallo $[-\sqrt{1/90}, \sqrt{1/90}] \subset (-1/9, 1/9)$ c'è almeno un autovalore di A.

Risoluzione di Fabio (fornisce un risultato migliore)

$$\begin{bmatrix} 2 & 1 & & \\ -1 & 2 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} A \begin{bmatrix} 2 & -1 & & \\ 1 & 2 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \frac{1}{5} = \begin{bmatrix} 0 & 0 & \frac{1}{6} & -\frac{1}{6} \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} 2 & -1 & & \\ 1 & 2 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \frac{1}{5} = \begin{bmatrix} 0 & 0 & \frac{1}{30} & -\frac{1}{30} \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

Quest'ultima matrice è hermitiana perché ottenuta tramite trasformazione per similitudine unitaria di una matrice hermitiana.

Applicando allora Weinstein a questa matrice ho che almeno un suo autovalore, e quindi almeno un autovalore di A, deve essere nel cerchio in $\mathbb{C}$ di centro $(0, 0)$ e raggio $\sqrt{2/30^2}$. Dunque almeno un autovalore di A è nell'intervallo $[-\sqrt{2}/30, \sqrt{2}/30] \subset (-1/9, 1/9)$ (A ha autovalori reali perché è hermitiana).

(2) Sia A la seguente matrice:

$$A = \begin{bmatrix} 3 & \sqrt{3}/2 & \frac{1}{2} \\ \sqrt{3}/2 & 2 & \sqrt{3} \\ \frac{1}{2} & \sqrt{3} & 0 \end{bmatrix}.$$

(i) Mediante una trasformazione per similitudine unitaria, trasformare A in una matrice del tipo

$$B = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0}^T & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & \delta \\ \delta & 3 \end{bmatrix}, \; \delta \in \mathbb{R}, \; |\delta| = 1$$

(usare il metodo di Givens che trasforma una matrice piena in una di Hessenberg). Dunque -1 è autovalore di A e gli altri due autovalori di A sono quelli di C.

(ii) Sia $M = B + x(\mathbf{e}_3\mathbf{e}_2^T + \mathbf{e}_2\mathbf{e}_3^T)$, $x \in \mathbb{R}$. Esiste un valore di x per cui $\det(\lambda I - M)$ ha due radici reali e coincidenti?

(iii) Calcolare gli autovalori di C effettuando un passo del metodo di Jacobi.

(iv) Sia $C_0 = C$ e siano M_0, N_0 2×2 tali che $C_0 = M_0 N_0$. Posto $C_1 = N_0 M_0$ si osserva che $C_1 = M_0^{-1} C_0 M_0$ e dunque C_1 ha gli stessi autovalori di C_0. Confrontare tra loro le tre matrici $C_1^{(LU)}$, $C_1^{(LL^H)}$, $C_1^{(QR)}$ ottenute utilizzando le decomposizioni LU, LL^H, QR di C_0, e prevedere la struttura di $C_\infty^{(LU)}$, $C_\infty^{(LL^H)}$, $C_\infty^{(QR)}$. (Facoltativo: notare che $C_1^{(QR)} = C_2^{(LL^H)}$).

(v) Introdurre A_0 2×2 tale che esistono L_0, U_0 per cui $A_0 = L_0 U_0$, ma non esistono L_1, U_1 tali che $A_1 := U_0 L_0 = L_1 U_1$, cioè anche se A $n \times n$ ammette la decomposizione LU, il generico passo del metodo LU per il calcolo degli autovalori di A può non essere ben definito.

Risoluzione

(i)

$$AS = \begin{bmatrix} 3 & \sqrt{3}/2 & \frac{1}{2} \\ \sqrt{3}/2 & 2 & \sqrt{3} \\ \frac{1}{2} & \sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & -\beta & \alpha \end{bmatrix} = \begin{bmatrix} 3 & \alpha\sqrt{3}/2 - \frac{1}{2}\beta & \beta\sqrt{3}/2 + \frac{1}{2}\alpha \\ \sqrt{3}/2 & 2\alpha - \sqrt{3}\beta & 2\beta + \sqrt{3}\alpha \\ \frac{1}{2} & \alpha\sqrt{3} & \sqrt{3}\beta \end{bmatrix}$$

Se $\alpha = \sqrt{3}/2$ e $\beta = -\frac{1}{2}$, allora

$$AS = \begin{bmatrix} 3 & 1 & 0 \\ \sqrt{3}/2 & 3\sqrt{3}/2 & \frac{1}{2} \\ \frac{1}{2} & 3/2 & -\sqrt{3}/2 \end{bmatrix},$$

$$S^T AS = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & \frac{1}{2} \\ 0 & -\frac{1}{2} & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} 3 & 1 & 0 \\ \sqrt{3}/2 & 3\sqrt{3}/2 & \frac{1}{2} \\ \frac{1}{2} & 3/2 & -\sqrt{3}/2 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0}^T & -1 \end{bmatrix}.$$

(ii) Si chiede se la matrice

$$M = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & x \\ 0 & x & -1 \end{bmatrix}, \; x \in \mathbb{R},$$

può avere due autovalori reali e coincidenti. Dimostriamo che la risposta è no.

Se $x \neq 0$ allora M è 1) tridiagonale, 2) reale, 3) tale che $M_{i,i+1}M_{i+1,i} > 0$ per ogni i. Ma, per un Corollario del Teorema di Sturm, ogni matrice $n \times n$ che soddisfa le proprietà 1),2),3) deve avere n autovalori reali e distinti. Dunque, se $x \neq 0$, la nostra M ha tre autovalori reali e distinti.

(Nota: se $x \neq 0$ la matrice $\lambda I - M$, per λ autovalore di M, ha rango esattamente 2, quindi la molteplicità geometrica di ogni autovalore di M è uno. Inoltre, essendo M diagonalizzabile (M è simmetrica reale!) i suoi autovalori devono avere molteplicità algebriche e geometriche coincidenti. Ne segue che per $x \neq 0$ ogni autovalore di M ha molteplicità algebrica esattamente uno, cioè M per $x \neq 0$ ha tre autovalori reali e distinti).

Se $x = 0$, allora gli autovalori di M sono -1, λ_2 , λ_3 essendo λ_2, λ_3 gli autovalori di C. I due autovalori di C, λ_2 e λ_3, sono reali e distinti tra loro (per lo stesso Corollario di cui sopra) e sono diversi da -1 perché contenuti (per Gershgorin) nell'intervallo $[2, 4]$. Quindi anche per $x = 0$ la matrice M ha tre autovalori reali e distinti.

(iii) Il primo passo del metodo di Jacobi applicato alla matrice C:

$$S^T C S = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} = \begin{bmatrix} 3(\alpha^2+\beta^2) - 2\alpha\beta & \alpha^2 - \beta^2 \\ \alpha^2 - \beta^2 & 3(\alpha^2+\beta^2) + 2\alpha\beta \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

(si è scelto $\alpha = \beta = 1/\sqrt{2}$ in modo che
1) $\alpha^2 + \beta^2 = 1$, quindi $S^T = S^{-1}$ e dunque $S^T C S$ è simile a C ;
2) $[S^T C S]_{12} = \alpha^2 - \beta^2 = 0$).
Essendo $S^T C S$ simile a C ed essendo i suoi autovalori 2 e 4, si conclude che gli autovalori di C sono 2 e 4. (iv)

$$(LU) \quad \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} y & w \\ 0 & z \end{bmatrix} = \begin{bmatrix} y & w \\ xy & xw+z \end{bmatrix} \Leftrightarrow \begin{array}{l} y = 3 \\ w = 1 \\ x = 1/y = 1/3 \\ z = 3 - xw = 8/3 \end{array} \Rightarrow M_0 = \begin{bmatrix} 1 & 0 \\ \frac{1}{3} & 1 \end{bmatrix},\ N_0 = \begin{bmatrix} 3 & 1 \\ 0 & \frac{8}{3} \end{bmatrix}$$

$$\Rightarrow C_1^{(LU)} = N_0 M_0 = \begin{bmatrix} 3 & 1 \\ 0 & \frac{8}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1}{3} & 1 \end{bmatrix} = \begin{bmatrix} \frac{10}{3} & 1 \\ \frac{8}{9} & \frac{8}{3} \end{bmatrix}. \text{ Si intuisce che } C_k^{(LU)} = \begin{bmatrix} \diamondsuit_k & 1 \\ \triangle_k & \square_k \end{bmatrix},\ \triangle_k \to 0,\ C_\infty^{(LU)} = \begin{bmatrix} 4 & 1 \\ 0 & 2 \end{bmatrix}.$$

$$(LL^H) \quad \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} x & 0 \\ y & w \end{bmatrix} \begin{bmatrix} x & y \\ 0 & w \end{bmatrix} = \begin{bmatrix} x^2 & xy \\ xy & y^2 + w^2 \end{bmatrix} \Leftrightarrow \begin{array}{l} x^2 = 3 \\ y = 1/x \\ w^2 = 3 - 1/x^2 = 8/3 \end{array} \Rightarrow \left\{ \begin{array}{l} M_0 = \begin{bmatrix} \sqrt{3} & 0 \\ \frac{1}{\sqrt{3}} & \sqrt{8/3} \end{bmatrix}, \\ N_0 = M_0^T \quad \Rightarrow\ C_0 = M_0 N_0 \end{array} \right. \Rightarrow C_1^{(LL^H)} = N_0 M_0 = \begin{bmatrix} \sqrt{3} & \frac{1}{\sqrt{3}} \\ 0 & \sqrt{8/3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ \frac{1}{\sqrt{3}} & \sqrt{8/3} \end{bmatrix} = \begin{bmatrix} \frac{10}{3} & \sqrt{8}/3 \\ \sqrt{8}/3 & \frac{8}{3} \end{bmatrix}.$$

$$(QR) \quad \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3\alpha - \beta & \alpha - 3\beta \\ 3\beta + \alpha & \beta + 3\alpha \end{bmatrix} = \begin{bmatrix} \sqrt{10} & 6/\sqrt{10} \\ 0 & 8/\sqrt{10} \end{bmatrix} \ (\alpha = \frac{3}{\sqrt{10}},\ \beta = \frac{-1}{\sqrt{10}}) \ \Rightarrow C_0 = M_0 N_0,$$

$$M_0 = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix},\ N_0 = \frac{1}{\sqrt{10}} \begin{bmatrix} 10 & 6 \\ 0 & 8 \end{bmatrix} \ \Rightarrow\ C_1^{(QR)} = N_0 M_0 = \frac{1}{5} \begin{bmatrix} 18 & 4 \\ 4 & 12 \end{bmatrix}.$$

$C_k^{(LL^H)}$ e $C_k^{(QR)}$ sono simmetriche reali $\forall\, k$ e $[C_k^{(LL^H)}]_{ij}, [C_k^{(QR)}]_{ij} \to 0$ se $i \neq j$, $\Rightarrow$ $C_\infty^{(LL^H)} = C_\infty^{(QR)} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$.

$$(LL^H\ 2^o \text{ passo}) \quad \begin{bmatrix} \frac{10}{3} & \frac{\sqrt{8}}{3} \\ \frac{\sqrt{8}}{3} & \frac{8}{3} \end{bmatrix} = M_1 N_1,\ M_1 = \begin{bmatrix} x & 0 \\ y & w \end{bmatrix},\ N_1 = M_1^T \ \Leftarrow \begin{array}{l} x = \sqrt{\frac{10}{3}} \\ y = \frac{2\sqrt{3}}{3\sqrt{5}} \\ w = \sqrt{\frac{12}{5}} \end{array}.\ \ C_2^{(LL^H)} = N_1 M_1 = \begin{bmatrix} \frac{18}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{12}{5} \end{bmatrix}.$$

Nota. È sbagliato cercare α, β tali che $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} = R$, ovvero trovare la fattorizzazione RQ di C_0 con R triangolare superiore, perché tale procedimento non si può generalizzare al caso $n > 2$.

(v) Siano

$$L = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix}, \; U = \begin{bmatrix} y & w \\ 0 & z \end{bmatrix}.$$

Allora

$$A = LU = \begin{bmatrix} y & w \\ xy & xw + z \end{bmatrix}, \; B = UL = \begin{bmatrix} y + wx & w \\ zx & z \end{bmatrix}$$

Imponendo $[UL]_{11} = y + wx = 0$ otteniamo una matrice B che non ammette la decomposizione LU. Ad esempio per $y = x = 1$, $w = -1$, si ha

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 + z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & z \end{bmatrix}, \; B = \begin{bmatrix} 0 & -1 \\ z & z \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & z \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Per $z \neq 0$ è semplice vedere che l'identità

$$B = \begin{bmatrix} 0 & -1 \\ z & z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ l & 1 \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix}$$

è impossibile (comunque si scelgano l, α, β, γ). Quindi $A_0 = A$ verifica la richiesta per ogni $z \neq 0$.

(3) (i) Sia A la seguente matrice $n \times n$:

$$A = \mu I + \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}, \; \mu \in \mathbb{R}.$$

Sia $A(\varepsilon)$ la perturbazione di A definita dall'identità $A(\varepsilon) = A + \varepsilon \mathbf{e}_i \mathbf{e}_j^T$, $\varepsilon > 0$, e sia $\sigma(A(\varepsilon))$ l'insieme degli autovalori di $A(\varepsilon)$. Calcolare, in funzione dei valori di i, j, le possibili espressioni di

$$\max_{\lambda(\varepsilon) \in \sigma(A(\varepsilon))} |\lambda(\varepsilon) - \mu|.$$

(ii) Nel punto (i) si è visto che gli autovalori di una matrice A con molteplicità algebrica maggiore della geometrica possono essere molto sensibili a variazioni degli elementi di A.

In questo punto si mostra che una perturbazione $O(\varepsilon)$ degli elementi di A implica una perturbazione dei suoi autovalori semplici che non può essere di ordine maggiore di $O(\varepsilon)$.

Siano $A \in \mathbb{R}^{n \times n}$, λ un autovalore semplice reale di A, e $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ non nulli tali che $A\mathbf{x} = \lambda \mathbf{x}$ e $\mathbf{y}^T A = \lambda \mathbf{y}^T$. (Facoltativo: dimostrare che $\mathbf{y}^T \mathbf{x} \neq 0$, ad esempio utilizzando la decomposizione di Jordan di A).

Sia $C \in \mathbb{R}^{n \times n}$. Per $\varepsilon \in \mathbb{R}$ siano $\lambda(\varepsilon) \in \mathbb{R}$ e $\mathbf{x}(\varepsilon) \in \mathbb{R}^n$ l'autovalore e l'autovettore della matrice $A + \varepsilon C$ che per $\varepsilon = 0$ coincidono rispettivamente con λ e $\mathbf{x}$. È noto che $\lambda(\varepsilon)$ è una funzione analitica di ε in un intorno di 0.

Usando l'identità $(A + \varepsilon C)\mathbf{x}(\varepsilon) = \lambda(\varepsilon)\mathbf{x}(\varepsilon)$, dimostrare che

$$\lambda(\varepsilon) - \lambda(0) = L\varepsilon + O(\varepsilon^2), \quad L = \frac{\mathbf{y}^T C \mathbf{x}}{\mathbf{y}^T \mathbf{x}}$$

e, nel caso in cui A e C sono simmetriche, dare una limitazione per $|L|$.

Risoluzione

(i) Se $i < j$ allora la matrice $A(\varepsilon)$ è triangolare superiore e i suoi elementi diagonali sono, come quelli di A, tutti uguali a μ. Dunque gli autovalori $\lambda(\varepsilon)$ di $A(\varepsilon)$ sono tutti uguali a μ e $\max_{\lambda(\varepsilon)\in\sigma(A(\varepsilon))} |\lambda(\varepsilon) - \mu| = 0$.

Se $i = j$ allora la matrice $A(\varepsilon)$ è triangolare superiore, $n - 1$ dei suoi elementi diagonali sono uguali a μ, e il rimanente elemento diagonale, l'elemento di posizione (i, i), è uguale a $\mu+\varepsilon$. Dunque $n-1$ autovalori $\lambda(\varepsilon)$ di $A(\varepsilon)$ sono uguali a μ e il rimanente autovalore $\lambda(\varepsilon)$ è uguale a $\mu+\varepsilon$. Quindi $\max_{\lambda(\varepsilon)\in\sigma(A(\varepsilon))} |\lambda(\varepsilon) - \mu| = \varepsilon$.

Se $i > j$ allora la matrice $A(\varepsilon)$ è triangolare superiore a blocchi con un blocco diagonale di ordine $(i - j + 1) \times (i - j + 1)$ e del tipo

$$\begin{bmatrix} \mu & 1 & & \\ & \mu & \ddots & \\ & & \ddots & 1 \\ \varepsilon & & & \mu \end{bmatrix}. \tag{BL}$$

Gli altri blocchi diagonali, che possono essere due (se $n > i > j > 1$), uno (se $n > i > j = 1$ o se $n = i > j > 1$), oppure nessuno (se $n = i > j = 1$), sono triangolari superiori e hanno gli elementi diagonali tutti uguali a μ.

Gli autovalori di $A(\varepsilon)$ sono l'unione degli autovalori dei suoi blocchi diagonali. Poiché gli autovalori della matrice (BL) sono del tipo $\mu+\lambda$, con λ tale che $\lambda^{i-j+1} = \varepsilon$, cioè $\lambda = \varepsilon^{1/(i-j+1)} e^{\mathbf{i}2\pi k/(i-j+1)}$, $k = 0, 1, \dots, i - j$, si può allora concludere che

$$\max_{\lambda(\varepsilon)\in\sigma(A(\varepsilon))} |\lambda(\varepsilon) - \mu| = \max_{\lambda(\varepsilon)\, autov\, di\, BL} |\lambda(\varepsilon) - \mu| = \max_k |\mu + \varepsilon^{1/n} e^{\mathbf{i}2\pi k/(i-j+1)} - \mu| = \varepsilon^{1/(i-j+1)}.$$

(ii) Nell'intorno di $\varepsilon = 0$ la funzione $\lambda(\varepsilon)$ è sviluppabile in serie di potenze. Dunque

$$\lambda(\varepsilon) = \lambda(0) + L\varepsilon + O(\varepsilon^2),\ L = \lambda'(0).$$

Inoltre, derivando rispetto a ε l'espressione $(A + \varepsilon C)\mathbf{x}(\varepsilon) = \lambda(\varepsilon)\mathbf{x}(\varepsilon)$ si ottiene

$$C\mathbf{x}(\varepsilon) + (A + \varepsilon C)\mathbf{x}'(\varepsilon) = \lambda'(\varepsilon)\mathbf{x}(\varepsilon) + \lambda(\varepsilon)\mathbf{x}'(\varepsilon).$$

Ponendo $\varepsilon = 0$ e ricordando che $\mathbf{x}(0) = \mathbf{x}$ si ha:

$$C\mathbf{x} + A\mathbf{x}'(0) = \lambda'(0)\mathbf{x} + \lambda(0)\mathbf{x}'(0) \ \Rightarrow\ \mathbf{y}^T C\mathbf{x} + \mathbf{y}^T A\mathbf{x}'(0) = \lambda'(0)\mathbf{y}^T\mathbf{x} + \lambda(0)\mathbf{y}^T\mathbf{x}'(0) \ \Rightarrow$$

$$\mathbf{y}^T C\mathbf{x} + \lambda(0)\mathbf{y}^T\mathbf{x}'(0) = \lambda'(0)\mathbf{y}^T\mathbf{x} + \lambda(0)\mathbf{y}^T\mathbf{x}'(0) \ \Rightarrow\ \mathbf{y}^T C\mathbf{x} = \lambda'(0)\mathbf{y}^T\mathbf{x},\quad L = \lambda'(0) = \frac{\mathbf{y}^T C\mathbf{x}}{\mathbf{y}^T\mathbf{x}}.$$

Se A è simmetrica, allora $\mathbf{y} = \mathbf{x}$, quindi $L = \frac{\mathbf{y}^T C\mathbf{x}}{\mathbf{y}^T\mathbf{x}} = \frac{\mathbf{x}^T C\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$. Se anche C è simmetrica, allora

$$L = \frac{\mathbf{x}^T C\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \in [\lambda_{\min}(C), \lambda_{\max}(C)]$$

(si ricorda infatti che se M è una matrice $n \times n$ normale allora il numero $\mathbf{z}^H M\mathbf{z}/\mathbf{z}^H\mathbf{z}$, al variare di $\mathbf{z} \in \mathbb{C}^n$ $\mathbf{z} \neq \mathbf{0}$, descrive in $\mathbb{C}$ il poligono convesso di area minima contenente gli autovalori di M; se in particolare M è hermitiana esso descrive l'intervallo della retta reale $[\lambda_{\min}(M), \lambda_{\max}(M)]$; la nostra matrice C è hermitiana, essendo reale simmetrica).

Dunque, se A e C sono simmetriche si ha $|L| \leq \max\{|\lambda_{\min}(C)|, |\lambda_{\max}(C)|\} = \rho(C)$.

Dicembre 2013: I ragazzi del corso di Elementi di Analisi Numerica

L =matrice di adiacenza
$L_{ij} = 1$ se i ha il cellulare di j, $L_{ij} = 0$ altrimenti

(la matrice di transizione P si ottiene da quella di adiacenza moltiplicando per $1/\deg(i)$ le sue righe i per cui $\deg(i) > 0$, essendo $\deg(i)$ il numero dei ragazzi di cui i ha il cellulare)

luigi	0	1		1	1	1							1			1			1	1					
matteo ba		0		1							1		1						1	1			1		
vincenzo			0										1	1											
daniele	1	1		0									1						1						
federico	1				0	1		1	1																
mara	1				1	0	1	1	1								1								
simone						1	0				1														
alessandra								0																	
arianna									0																
matteo br										0									1				1		
dario											0														
roberta												0			1										
fabio	1	1	1	1									0	1	1			1	1	1					1
cristiano														0											
gianbattista												1	1		0					1					
giuseppe	1		1		1											0									
margherita																	0								
giulio													1					0			1	1			
francesca	1	1	1	1	1					1	1		1	1		1		1	0						
diego																				0					
martina																		1			0	1			
jacopo																		1			1	0			
francesco																							0		
andrea			1																1					0	
guido													1	1						1			1		0

Da metà di una pagina degli appunti di Roberta Piersimoni mi sono discostato da tali appunti e ho insegnato le seguenti cose (le cose precedenti della teoria di Perron-Frobenius le ho insegnate come gli altri anni):

Risultato (senza dim): $r = \rho(A) = \sup_{x \geq \mathbf{0}, \mathbf{x} \neq \mathbf{0}} \min_{i:\, x_i \neq 0} \frac{(A\mathbf{x})_i}{x_i}$ è semplice, come autovalore di A non negativa irriducibile (la dim la vedremo solo nel caso $A > O$).

Quindi, riassumendo, se A è non negativa irriducibile allora $\rho(A)$ è positivo, è autovalore semplice di A, ed esiste ed è unico $\mathbf{z} > \mathbf{0}$, $\|\mathbf{z}\|_1 = 1$ tale che $A\mathbf{z} = \rho(A)\mathbf{z}$. Coppia di Perron: $(\rho(A), \mathbf{z})$.

Dunque, siano $r, \lambda_2, \ldots, \lambda_n$ gli autovalori di A non negativa irriducibile. Sappiamo che $\lambda_j \neq r$, $j = 2, \ldots, n$.

Quando $|\lambda_j| < r$, $j = 2, \ldots, n$? "$r = \rho(A)$ dominante"

Non sempre, infatti la matrice A 2×2 le cui righe sono [0 1] e [1 0] è non negativa e irriducibile ma A ha oltre 1 anche -1 come autovalore di modulo $r = \rho(A)$.

Invece, per la matrice A 3×3 con righe [2 1 1], [1 2 1], [1 1 2] si ha effettivamente che $r = \rho(A) = 4$ è autovalore di A e domina gli altri due autovalori di A (1 con molteplicità algebrica due).

CS affinché $r = \rho(A)$ dominante (di A non negativa irriducibile):
(i) $A > 0$
(ii) esiste i tale che $a_{ii} > 0$

Il fatto che (ii) è una CS per r dominante non viene dimostrato qui del tutto (nel senso che si dimostrerà che $|\lambda| = \rho(A) \Rightarrow \lambda = \rho(A)$ e la conclusione seguira' dal fatto, non dimostrato, che r è semplice come autovalore di A).

Si dimostrerà invece completamente che, nell'ipotesi $A > O$, la somma delle molteplicità algebriche degli autovalori di A di modulo minore di r deve essere $n-1$. Dunque, in particolare, seguirà il carattere semplice di r come autovalore di A.

Si capisce quindi perché si evidenzia il risultato (i) anche se è molto piu' debole del risultato (ii). Il risultato (i) si evidenzia anche perché è il primo risultato ottenuto (Perron, 1906), tra i risultati della teoria di Perron-Frobenius.

Lemma. $A \geq O$, A irriducibile è simile ad $M \geq O$, M irriducibile, M $\rho(A)$-stocastica per righe (per colonne).

Dimostrazione.
$A\mathbf{z} = \rho(A)\mathbf{z}$, $Ad(\mathbf{z})\mathbf{e} = \rho(A)\mathbf{z}$, $d(\mathbf{z})^{-1}Ad(\mathbf{z})\mathbf{e} = \rho(A)\mathbf{e}$.
$A^T\mathbf{y} = \rho(A)\mathbf{y}$, $A^Td(\mathbf{y})\mathbf{e} = \rho(A)\mathbf{y}$, $d(\mathbf{y})^{-1}A^Td(\mathbf{y})\mathbf{e} = \rho(A)\mathbf{e}$ ($\mathbf{y} > \mathbf{0}$, $\|\mathbf{y}\|_1 = 1$).
$M = d(\mathbf{z})^{-1}Ad(\mathbf{z})$ è non negativa, irriducibile, $\rho(A)$-stocastica per righe. $M = d(\mathbf{y})Ad(\mathbf{y})^{-1}$ è non negativa, irriducibile, $\rho(A)$-stocastica per colonne. $\square$

Osservazione. I cerchi di Gershgorin di $M = d(\mathbf{z})^{-1}Ad(\mathbf{z})$ hanno centro in $a_{ii} \geq 0$ e raggio $\rho(A) - a_{ii}$ ($\forall\, i$). Quindi passano tutti per il punto $(\rho(A), 0)$ e sono tutti contenuti nel cerchio di centro $(0, 0)$ e raggio $\rho(A)$.

Proposizione. $A \geq O$, A irriducibile, ed esiste i tale che $a_{ii} > 0$. Se λ è autovalore di A e $|\lambda| = \rho(A)$, allora $\lambda = \rho(A)$.

Dimostrazione. Sia λ una autovalore di A tale che $|\lambda| = \rho(A)$. Sappiamo che λ deve essere autovalore di $M = d(\mathbf{z})^{-1}Ad(\mathbf{z})$, che è irriducibile, dunque, per il Teorema di Gershgorin rafforzato

(vedi più avanti),
$$\lambda \in (\cup_i \mathring{K}_i(d(\mathbf{z})^{-1}Ad(\mathbf{z}))) \cup (\cap_i \partial K_i(d(\mathbf{z})^{-1}Ad(\mathbf{z}))).$$
Dal fatto che $|\lambda| = \rho(A)$ segue che $\lambda \in \cap_i \partial K_i(d(\mathbf{z})^{-1}Ad(\mathbf{z}))$. Ma quest'ultimo insieme nell'ipotesi che per un i si ha $a_{ii} > 0$ si riduce al solo punto $(\rho(A), 0)$. $\square$

Proposizione. $A > O$. Se λ è autovalore di A e $|\lambda| = \rho(A)$, allora $\lambda = \rho(A)$ e $\rho(A)$ è semplice come autovalore di A.

Dimostrazione. Siano $r = \rho(A), \lambda_2, \ldots, \lambda_n$ gli autovalori di A (si noti che qui alcuni dei λ_j, $j = 2, \ldots, n$, potrebbero essere uguali a r, cioè facciamo finta di non sapere che r è semplice). Sia $\mathbf{z}$ tale che $\mathbf{z} > \mathbf{0}$, $\|\mathbf{z}\|_1 = 1$, $A\mathbf{z} = \rho(A)\mathbf{z}$. Sia $W = A - s\mathbf{z}\mathbf{e}^T$ e X una matrice non singolare tale che $X\mathbf{e}_1 = \mathbf{z}$. Si dimostra che
$$X^{-1}AX = \begin{bmatrix} \rho(A) & \\ \mathbf{0} & B \end{bmatrix}, \quad X^{-1}WX = \begin{bmatrix} \rho(A) - s & \\ \mathbf{0} & B \end{bmatrix}.$$
Dunque $p_A(\lambda) = (\lambda - \rho(A))p_B(\lambda)$ e $p_W(\lambda) = (\lambda - (\rho(A) - s))p_B(\lambda)$, cioè gli autovalori di W sono $r - s, \lambda_2, \ldots, \lambda_n$.

Sia ora $s = \min a_{ij}$. Poiché $A > O$, s è un numero positivo. Inoltre $W = |W| < A$, dunque le condizioni $|W| \leq A$ e $|W| \neq A$ sono soddisfatte, e quindi, per un risultato ottenuto precedentemente,
$$\max_{j=2,\ldots,n} |\lambda_j| \leq \rho(W) < \rho(A) = r.$$
Cioè, r è semplice, come autovalore di A, e domina gli altri $n-1$ autovalori di A.

CALCOLO DI $\rho(A)$ e $\mathbf{z}$ ($A\mathbf{z} = \rho(A)\mathbf{z}$, $\mathbf{z} > \mathbf{0}$, $\|\mathbf{z}\|_1 = 1$, $\rho(A) > 0$ semplice)

RICORDA IL SEGUENTE RISULTATO (dim. vista solo per A diagonalizzabile e $t = 1$):

$A \in \mathbb{C}^{n\times n}$, $\lambda_1, \lambda_2, \ldots, \lambda_n$ i suoi autovalori, $\lambda_1 = \ldots = \lambda_t$, $t \geq 1$, $t = m_a(\lambda_1) = m_g(\lambda_1)$. $X = [\mathbf{x}_1 \cdots \mathbf{x}_n]$ la matrice che trasforma A in forma di Jordan
$$X^{-1}AX = J = \begin{bmatrix} \lambda_1 I_t & O \\ O & * \end{bmatrix}.$$
SE $|\lambda_1| > |\lambda_j|$, $j > t$, $\mathbf{v} \in \mathbb{C}^n$ ($\mathbf{v}$ si potrebbe scegliere tale che $\|\mathbf{v}\| = 1$) tale che nell'espressione $\mathbf{v} = \sum_{j\leq t} \alpha_j \mathbf{x}_j + \sum_{j>t} \alpha_j \mathbf{x}_j$, il vettore $\mathbf{x} = \sum_{j\leq t} \alpha_j \mathbf{x}_j$ è non nullo (Nota: $A\mathbf{x} = \lambda_1 \mathbf{x}$)

$\mathbf{v}_0 = \mathbf{v}$. Per $k = 1, 2, \ldots$
$$\mathbf{a}_k = A\mathbf{v}_{k-1}, \ \varphi_k = \frac{\mathbf{u}^H \mathbf{a}_k}{\mathbf{u}^H \mathbf{v}_{k-1}}, \ \mathbf{v}_k = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|}$$
($\mathbf{u}$ tale che $\mathbf{u}^H \mathbf{x} \neq 0$, $\mathbf{u}^H \mathbf{v}_{k-1} \neq 0$, $\forall\, k$)

TESI: $\mathbf{v}_k = A^k\mathbf{v}/\|A^k\mathbf{v}\| \to \mathbf{x}/\|\mathbf{x}\|$, $\varphi_k = \mathbf{u}^H A^k \mathbf{v}/\mathbf{u}^H A^{k-1}\mathbf{v} \to \lambda_1$

(Dimostrazione: si dimostra che $\frac{1}{\lambda_1^k} A^k \mathbf{v} \to \mathbf{x}$; si dimostra che $\mathbf{u}^H A^k \mathbf{v}/\mathbf{u}^H A^{k-1}\mathbf{v} \to \lambda_1$)

CASO PARTICOLARE: $A \geq O$, A irriducibile (si scelgono $\|\ \|$ e $\mathbf{u}$ opportune)

$\rho(A) = r$ ha molteplicità algebrica e geometrica coincidenti con $t = 1$. $X = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ la matrice che trasforma A in forma di Jordan

$$X^{-1}AX = J = \begin{bmatrix} r & 0 \\ 0 & * \end{bmatrix}$$

SE $\rho(A) = r > |\lambda_j|$, $j > 1$, $\mathbf{v} \in (\mathbb{R}^n)^+$, $\|\mathbf{v}\|_1 = 1$, tale che in $\mathbf{v} = \alpha\mathbf{z} + \sum_{j>1} \alpha_j \mathbf{x}_j$ il vettore $\mathbf{x} = \alpha\mathbf{z}$ sia non nullo (Nota: $A\mathbf{x} = \rho(A)\mathbf{x}$)

NOTA: per come è stato scelto $\mathbf{v}$ il numero α deve essere necessariamente positivo (e si risponde all'obiezione di Giulio che $\mathbf{x}/\|\mathbf{x}\|$ non farebbe $\mathbf{z}$ se α fosse negativo) e quindi il vettore $\mathbf{x}$ risulta effettivamente non nullo. Questo è dimostrato rigorosamente dopo, ma è già intuibile visto che $\mathbf{v}$, essendo positivo, non può non avere componente lungo $\mathbf{z}$ che è anche positivo.

$\mathbf{v}_0 = \mathbf{v}$. Per $k = 1, 2, \ldots$

$$\mathbf{a}_k = A\mathbf{v}_{k-1}, \ \varphi_k = \|\mathbf{a}_k\|_1, \ \mathbf{v}_k = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|_1}$$

(Nota: $\mathbf{a}_k > \mathbf{0}$; $\mathbf{v}_k > \mathbf{0}$, $\|\mathbf{v}_k\|_1 = 1$)

TESI: $\mathbf{v}_k \to \mathbf{z}$, $\varphi_k \to r = \rho(A)$

Domanda: $\varphi_{k+1} \leq \varphi_k$?

NOTE SUI DUE RISULTATI SOPRA ENUNCIATI

(1) Nel secondo risultato, per come è scelto $\mathbf{v}$, l'ipotesi $\mathbf{x} \neq \mathbf{0}$ è verificata.

Dimostrazione. Si può supporre che la prima colonna di X sia proprio $\mathbf{z}$. Siccome $X^{-1}A = JX^{-1}$, si ha $(\mathbf{e}_1^T X^{-1})A = \rho(A)(\mathbf{e}_1^T X^{-1})$, dunque $\mathbf{e}_1^T X^{-1} = \beta\mathbf{y}^T$ dove $\mathbf{y}^T A = \rho(A)\mathbf{y}^T$ ($\mathbf{y} > \mathbf{0}$, $\|\mathbf{y}\|_1 = 1$). Ma $1 = \mathbf{e}_1^T X^{-1} X \mathbf{e}_1 = \beta\mathbf{y}^T\mathbf{z}$, da cui la formula per β: $\beta = 1/\mathbf{y}^T\mathbf{z}$ e quindi per la prima riga di X^{-1}: $\mathbf{e}_1^T X^{-1} = (1/\mathbf{y}^T\mathbf{z})\mathbf{y}^T$.

Si nota che se $j > 1$ allora $0 = (\mathbf{e}_1^T X^{-1})(X\mathbf{e}_j) = (1/\mathbf{y}^T\mathbf{z})\mathbf{y}^T\mathbf{x}_j$, da cui il risultato $\mathbf{y}^T\mathbf{x}_j = 0$, se $j > 1$. Dunque, dall'identità $\mathbf{v} = \alpha\mathbf{z} + \sum_{j>1} \alpha_j \mathbf{x}_j$ segue che $\mathbf{y}^T\mathbf{v} = \alpha\mathbf{y}^T\mathbf{z} + \sum_{j>1} \alpha_j \mathbf{y}^T\mathbf{x}_j = \alpha\mathbf{y}^T\mathbf{z}$, che implica $\alpha = \mathbf{y}^T\mathbf{v}/\mathbf{y}^T\mathbf{z}$. Dunque $\mathbf{v} = (\mathbf{y}^T\mathbf{v}/\mathbf{y}^T\mathbf{z})\mathbf{z} + \sum_{j>1} \alpha_j \mathbf{x}_j$ con $(\mathbf{y}^T\mathbf{v}/\mathbf{y}^T\mathbf{z})\mathbf{z}$ non nullo SE $\mathbf{y}^T\mathbf{v} \neq 0$.

Poiché $\mathbf{v} \in (\mathbb{R}^n)^+$, si ha $\mathbf{y}^T\mathbf{v} > 0$, quindi $\mathbf{x} = (\mathbf{y}^T\mathbf{v}/\mathbf{y}^T\mathbf{z})\mathbf{z}$ è non nullo e $\mathbf{x} > \mathbf{0}$. Ne segue che $\alpha > 0$ e $\mathbf{x}/\|\mathbf{x}\|_1 = \mathbf{z}$.

(2) Ogni passo del metodo delle potenze illustrato (nel caso generale e nel caso particolare $A \geq O$, A irriducibile) ha un costo minimo, dell'ordine del costo del prodotto matrice A per vettore. Inoltre, minima è la memoria richiesta per implementarlo.

(3) La velocità di convergenza del metodo è, nei due casi, rispettivamente

$$\max_{j>t} \left|\frac{\lambda_j}{\lambda_1}\right|^k |p^{(j)}(k)|, \quad \max_{j>1} |\frac{\lambda_j}{r}|^k |p^{(j)}(k)|$$

dove il grado del polinomio $p^{(j)}(k)$ è uguale all'ordine del blocco di Jordan di ordine massimo relativo all'autovalore λ_j meno uno. Dunque, se A è diagonalizzabile, la velocità di convergenza del metodo diventa, nei due casi:

$$\max_{j>t} \left|\frac{\lambda_j}{\lambda_1}\right|^k, \quad \max_{j>1} |\frac{\lambda_j}{r}|^k.$$

(Dunque, se A è diagonalizzabile il metodo è più rapido).

(4) Ci si può chiedere se $\varphi_k \geq \varphi_{k+1}$, cioè se la successione di scalari generata dal metodo delle potenze convergente a $\rho(A)$, vi converge dall'alto. Ciò equivale a chiedersi se

$$\mathbf{e}^T A\mathbf{a}_k \leq (\mathbf{e}^T\mathbf{a}_k)^2,\ \ \mathbf{a}_k = A\mathbf{v}_{k-1}\ (\mathbf{v}_{k-1} = \text{ddp, cioè } \mathbf{v}_{k-1} > \mathbf{0},\ \|\mathbf{v}_{k-1}\|_1 = 1).$$

Conoscere una approssimazione dall'alto di $\rho(A)$ avrebbe interesse perché il metodo delle potenze inverse per il calcolo di $\mathbf{z}$, che può essere più veloce del metodo delle potenze, richiede per essere ben definito ed efficiente la conoscenza di una approssimazione dall'alto di $\rho(A)$ (vedi dopo).

(5) Se $\mathbf{v}$, nell'enunciato del secondo Risultato, fosse scelto semplicemente in $\mathbb{R}^n$ tale che $\mathbf{y}^T\mathbf{v} \neq 0$, allora $\mathbf{x}$ sarebbe ancora non nullo, ma la successione $\mathbf{a}_k$ non soddisfarrebbe più le sue belle proprietà (e $\mathbf{u} = \mathbf{e}$ non sarebbe forse più una scelta opportuna, $\mathbf{u} = \mathbf{y}$ forse lo sarebbe ma $\mathbf{y}$ non si conosce...).

Premessa al successivo Risultato: se A fosse anche stocastica per colonne, il problema del calcolo di $\rho(A)$ non sussiste perché $\rho(A) = 1$. Permane il problema del calcolo di $\mathbf{z}$ tale che $A\mathbf{z} = \mathbf{z}$, $\mathbf{z} > \mathbf{0}$, $\|\mathbf{z}\|_1 = 1$. (È invece ovvio qual'è il vettore $\mathbf{y}$ tale che $\mathbf{y}^T A = \mathbf{y}^T$, $\mathbf{y} > \mathbf{0}$, $\|\mathbf{y}\|_1 = 1$: $\mathbf{y} = \frac{1}{n}\mathbf{e}$).

CASO ANCORA PIÙ PARTICOLARE: $A \geq O$, A irriducibile, A stocastica per colonne (si scelgono $\|\ \|$ e $\mathbf{u}$ come prima, $\|\ \| = \|\ \|_1$ e $\mathbf{u} = \mathbf{e}$)

$\rho(A) = 1$ ha molteplicità algebrica e geometrica coincidenti e uguali a uno. $X = [\mathbf{x}_1\ \cdots\ \mathbf{x}_n]$ che trasforma A in forma di Jordan

$$X^{-1}AX = J = \begin{bmatrix} 1 & 0 \\ 0 & * \end{bmatrix}$$

SE $\rho(A) = 1 > |\lambda_j|$, $j > 1$, $\mathbf{v} \in (\mathbb{R}^n)^+$, $\|\mathbf{v}\|_1 = 1$, tale che in $\mathbf{v} = \alpha\mathbf{z} + \sum_{j>1}\alpha_j\mathbf{x}_j$ il vettore $\mathbf{x} = \alpha\mathbf{z}$ è non nullo (lo è perché per quanto detto prima $\alpha = 1$, vedi Nota (1))

$\mathbf{v}_0 = \mathbf{v}$. Per $k = 1, 2, \ldots$

$$\mathbf{v}_k = \mathbf{a}_k = A\mathbf{v}_{k-1},\ \varphi_k = \|\mathbf{a}_k\|_1 = 1$$

(Nota: $\mathbf{a}_k > \mathbf{0}$, $\|\mathbf{a}_k\|_1 = \mathbf{e}^T A\mathbf{v}_{k-1} = \mathbf{e}^T\mathbf{v}_{k-1} = 1 \Rightarrow \mathbf{v}_k = \mathbf{a}_k$)

TESI: $\mathbf{v}_k \to \mathbf{z} = A\mathbf{z}$, $\varphi_k = 1 \to 1 = \rho(A)$

Dimostrazione. $\rho(A) \leq \|A\|_1 = 1$ (perché $A \geq O$, A stocastica per colonne). Inoltre, $A^T\mathbf{e} = \mathbf{e}$, dunque 1 è autovalore di A. Dunque $1 = r = \rho(A)$. Il fatto che $\alpha = 1$ segue dall'epressione esplicita trovata per α nel precedente teorema usando l'osservazione che $\mathbf{y} = \frac{1}{n}\mathbf{e}$. (Nota: non serve più il discorso fatto gli altri anni: $A\mathbf{v} = \lambda\mathbf{v} \Rightarrow (1-\lambda)\mathbf{e}^T v = 0 \Rightarrow \mathbf{e}^T\mathbf{v} = 0$ se $\lambda \neq 1 \ldots \Rightarrow \alpha = 1$). $\square$

Esempio. Discutere al variare di $a \in [0, 1]$ esistenza, unicità e calcolo dell'autovettore di Perron della matrice A 3×3 con righe $[0\ 0\ 1]$, $[a\ 0\ 0]$, e $[1-a\ 1\ 0]$.

Per $a = 1$ l'algoritmo non converge (è evidente perché l'operazione $\mathbf{v}_k = A\mathbf{v}_{k-1}$ non fa altro che permutare gli elementi di $\mathbf{v}_{k-1}$; inoltre l'ipotesi $|\lambda_j| < 1$, $j = 2, 3$, non è soddisfatta)

Per $a = 0$ non è ben definita la coppia di Perron (esiste $\mathbf{z} = A\mathbf{z}$, ma una delle sue componenti è nulla; infatti A è riducibile)

Per $a \in (0, 1)$ l'algoritmo converge (A è non negativa irriducibile e calcolando i due autovalori di A diversi dall'autovalore semplice 1 si vede che hanno entrambi modulo minore di uno; ad es. per $a = 1/4$ la matrice A ha $-\frac{1}{2}$ come autovalore e $m_a(-\frac{1}{2}) = 2 > m_g(-\frac{1}{2}) = 1$)

Coppia di Perron calcolata con il METODO DELLE POTENZE INVERSE

RISULTATO GENERALE

$A \in \mathbb{C}^{n\times n}$, $\lambda_1, \lambda_2, \ldots, \lambda_n$ i suoi autovalori, $m_a(\lambda_i) = m_g(\lambda_i) = t$, $X = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ che trasforma A in forma di Jordan

$$X^{-1}AX = J = \begin{bmatrix} \lambda_i I_t & 0 \\ 0 & * \end{bmatrix}$$

SE $0 < |\lambda_i^* - \lambda_i| < |\lambda_j - \lambda_i^*|$, $\forall\, \lambda_j \neq \lambda_i$,

$\mathbf{v} \in \mathbb{C}^n$ (si può prendere tale che $\|\mathbf{v}\| = 1$) tale che in $\mathbf{v} = \sum_{j\leq t} \alpha_j \mathbf{x}_j + \sum_{j>t} \alpha_j \mathbf{x}_j$ il vettore $\mathbf{x} = \sum_{j\leq t} \alpha_j \mathbf{x}_j$ sia non nullo (Nota: $A\mathbf{x} = \lambda_i \mathbf{x}$)

$\mathbf{v}_0 = \mathbf{v}$. Per $k = 1, 2, \ldots$

$$(\lambda_i^* I - A)\mathbf{a}_k = \mathbf{v}_{k-1}, \quad \mathbf{v}_k = \frac{1}{\|\mathbf{a}_k\|}\mathbf{a}_k$$

TESI: $\mathbf{v}_k \to \mathbf{x}/\|\mathbf{x}\|$ con ordine di convergenza $O(\max_{j:\,\lambda_j \neq \lambda_i} \left|\frac{\lambda_i - \lambda_i^*}{\lambda_j - \lambda_i^*}\right|^k |p^{(j)}(k)|)$

Nota. La convergenza è molto rapida se $|\lambda_i^* - \lambda_i| << |\lambda_j - \lambda_i^*|$

Dimostrazione. Applicare il Teorema delle Potenze a $(\lambda_i^* I - A)^{-1}$. $\square$

CASO PARTICOLARE: $A \geq O$, irriducibile

$\rho(A), \lambda_2, \ldots, \lambda_n$ gli autovalori di A, $m_a(\rho(A)) = m_g(\rho(A)) = 1$, $X = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$ che trasforma A in forma di Jordan

$$X^{-1}AX = J = \begin{bmatrix} \rho(A) & 0 \\ 0 & * \end{bmatrix}$$

SE $0 < |\widetilde{\rho(A)} - \rho(A)| < |\widetilde{\rho(A)} - \lambda_j|$, $\forall\, \lambda_j \neq \rho(A)$ ($\Leftrightarrow \forall\, j > 1$),

$\mathbf{v} \in (\mathbb{R}^n)^+$, $\|\mathbf{v}\|_1 = 1$, tale che in $\mathbf{v} = \alpha\mathbf{z} + \sum_{j>1} \alpha_j \mathbf{x}_j$ il vettore $\mathbf{x} = \alpha\mathbf{z}$ sia non nullo (Nota: $A\mathbf{x} = \rho(A)\mathbf{x}$)

$\mathbf{v}_0 = \mathbf{v}$. Per $k = 1, 2, \ldots$

$$(\widetilde{\rho(A)} I - A)\mathbf{a}_k = \mathbf{v}_{k-1}, \quad \mathbf{v}_k = \frac{1}{\|\mathbf{a}_k\|_1}\mathbf{a}_k$$

Nota: se fosse $\widetilde{\rho(A)} > \rho(A)$ si avrebbe $\mathbf{a}_k > \mathbf{0}$, $\mathbf{v}_k > \mathbf{0}$ e $\|\mathbf{v}_k\|_1 = 1$

TESI: $\mathbf{v}_k \to \mathbf{x}/\|\mathbf{x}\|_1 = \mathbf{z}$ con ordine di convergenza $O(\max_{\lambda_j \neq \rho(A)} \left|\frac{\rho(A) - \widetilde{\rho(A)}}{\lambda_j - \widetilde{\rho(A)}}\right|^k |p^{(j)}(k)|)$.

Domanda: come calcolo $\widetilde{\rho(A)}$? Con il metodo delle potenze ?

CASO ANCORA PIÙ PARTICOLARE: $A \geq O$, irriducibile, stocastica per colonne

$\widetilde{\rho(A)} = \tilde{1} = 1 + \varepsilon$, $\varepsilon > 0$.

Ho quindi insegnato, come gli altri anni, le pagine 132–138 degli appunti di Roberta Piersimoni (su applicazione della teoria di Perron-Frobenius a matrice di transizione di grafo orientato: vettore vertexrank (pagerank), calcolo di tale vettore col metodo delle potenze, ...). Poi mi sono discostato di nuovo e ho insegnato

SU UN ALTRO METODO PER IL CALCOLO DEL VETTORE DI PERRON (PAGERANK)

PREMESSA: come nasce il metodo di Richardson-Eulero

Integrale generale dell'equazione differenziale $z'(t) = -\lambda z(t) + q$: integrale generale dell'equazione omogenea associata, $z'(t) = -\lambda z(t)$, più soluzione particolare dell'equazione. Dunque l'insieme delle soluzioni dell'equazione $z'(t) = -\lambda z(t) + q$ è: $\{Ce^{-\lambda t} + q/\lambda : C \in \mathbb{R}\}$.

(Nota. Mentre insegnavo agli studenti di AN1 come nasce il metodo di Rich.Eul., non riuscivo a trovare una soluzione particolare dell'equazione $z_i'(t) = -\lambda_i z_i(t) + (Q^T\mathbf{b})_i$ (VEDI SOTTO A COSA SERVIVA). Bisognava cercarla di tipo "costante", cioè imporre che $z_i(t) = a$, con a reale, fosse soluzione, cioè imporre l'identità: $0 = -\lambda_i a + (Q^T\mathbf{b})_i$ da cui la formula per a: $a = (Q^T\mathbf{b})_i/\lambda_i$).

Tale integrale generale serve nella dimostrazione del seguente risultato sull'andamento (per $t \to +\infty$) della soluzione di un particolare problema differenziale di Cauchy:

Siano A $n \times n$ reale definita positiva, e $\mathbf{b}$ e $\mathbf{y}_0$ vettori $n \times 1$, e si consideri il problema

$$\mathbf{y}'(t) = \mathbf{b} - A\mathbf{y}(t), \quad \mathbf{y}(0) = \mathbf{y}_0. \tag{$*$}$$

Per $t \to +\infty$ il vettore $y(t)$ di n funzioni soluzione di (*) tende al vettore $A^{-1}\mathbf{b}$, per ogni scelta di $\mathbf{y}_0$.

(Nota. Tale risultato è un caso particolare del risultato noto come Teorema di Perron: $\mathbf{y}'(t) = \mathbf{h}(\mathbf{y}(t))$, $\mathbf{y}(0) = \mathbf{y}_0$, $\mathbf{h}(\mathbf{y}^*) = \mathbf{0}$, $Re(\lambda(\mathrm{Jac}\,\mathbf{h}(\mathbf{y}^*))) < 0$, $\mathbf{h} \in C^1$ in intorno di $\mathbf{y}^*$. Tesi: esiste $\delta > 0$ tale che $\|\mathbf{y}_0 - \mathbf{y}^*\| < \delta \Rightarrow \mathbf{y}(t) \to \mathbf{y}^*$).

Dimostrazione. Osservando che $A = Q\Lambda Q^T$, $\Lambda = \mathrm{diag}\,(\lambda_i,\, i = 1, \ldots, n)$, $\lambda_i > 0$, il sistema differenziale diventa $Q^T\mathbf{y}'(t) = Q^T\mathbf{b} - \Lambda Q^T\mathbf{y}(t)$ e ponendo $\mathbf{z}(t) = Q^T\mathbf{y}(t)$, si ottiene $\mathbf{z}'(t) = -\Lambda\mathbf{z}(t) + Q^T\mathbf{b}$ il cui integrale generale ora possiamo dire che è

$$\mathbf{z}(t) = \mathrm{diag}\,(C_i, i = 1, \ldots, n)(e^{-\lambda_i t})_{i=1}^n + \Lambda^{-1}Q^T\mathbf{b}.$$

dove C_i sono costanti generiche. Dunque $\mathbf{y}(t) = Q\mathbf{z}(t) = Q\,\mathrm{diag}\,(C_i, i = 1, \ldots, n)(e^{-\lambda_i t})_{i=1}^n + Q\Lambda^{-1}Q^T\mathbf{b}$. È evidente che $\mathbf{y}(t)$, per ogni scelta delle C_i, tende a $Q\Lambda^{-1}Q^T\mathbf{b} = A^{-1}\mathbf{b}$ quando t tende a $+\infty$ (questo perché i λ_i sono positivi).

In conclusione, se A è reale definita positiva, per ogni scelta di $\mathbf{y}_0$, la soluzione $\mathbf{y}(t)$ del problema differenziale (*) viene attratta, per $t \to +\infty$, dalla soluzione del sistema lineare $A\mathbf{u} = \mathbf{b}$, cioè da $A^{-1}\mathbf{b}$. $\square$

Dunque, ogni metodo numerico che risolve il problema differenziale (*) su $[0, +\infty)$ è anche un metodo numerico che risolve il sistema lineare $A\mathbf{u} = \mathbf{b}$.

In aula ho illustrato il noto metodo di Eulero per risolvere (*), che consiste nel sostituire $\mathbf{y}'(t)$ con un rapporto incrementale, $(\mathbf{y}(t+\delta t) - \mathbf{y}(t))/\delta t = \mathbf{b} - A\mathbf{y}(t) + O(\delta t)$, scrivere l'identità ottenuta solo per $t = t_k := k * \delta t$, $k = 0, 1, 2, 3, \ldots$:

$(\mathbf{y}(t_k + \delta t) - \mathbf{y}(t_k))/\delta t = \mathbf{b} - A\mathbf{y}(t_k) + O(\delta t)$,

$(\mathbf{y}(t_{k+1}) - \mathbf{y}(t_k))/\delta t = \mathbf{b} - A\mathbf{y}(t_k) + O(\delta t)$

e, a partire dal vettore η_k approssimante il vettore soluzione esatto $\mathbf{y}(t_k)$, definire il vettore η_{k+1} approssimante il vettore soluzione esatto $\mathbf{y}(t_{k+1})$ tramite l'identità:

$(\eta_{k+1} - \eta_k)/\delta t = \mathbf{b} - A\eta_k$, $k = 0, 1, 2, 3, \ldots$

ovvero tramite l'identità

$$\eta_{k+1} = \eta_k + \delta t(\mathbf{b} - A\eta_k), \;\; k = 0, 1, 2, 3, \dots \; . \qquad (+)$$

Ponendo $\eta_0 = \mathbf{y}(0) = \mathbf{y}_0$, vorremmo che i vettori η_k generati con (+) fossero buone approssimazioni dei vettori esatti $\mathbf{y}(t_k)$. In particolare vorremmo che, come $\mathbf{y}(t_k) \to A^{-1}\mathbf{b}$, si ha anche che $\eta_k \to A^{-1}\mathbf{b}$, quando $k \to +\infty$.

Teorema (RE) [caso reale]. Dato un sistema lineare generico $M\mathbf{u} = \mathbf{c}$, $M \in \mathbb{R}^{n\times n}$, $\det(M) \neq 0$, $\mathbf{c} \in \mathbb{R}^n$, esso è equivalente all'equazione di punto fisso $\mathbf{u} = \mathbf{u} + \omega(\mathbf{c} - M\mathbf{u})$, con $\omega \in \mathbb{R}$ $\omega \neq 0$ parametro fissato. È quindi ben definita la successione di vettori

$$\mathbf{u}_0 \in \mathbb{R}^n, \quad \mathbf{u}_{k+1} = \mathbf{u}_k + \omega(\mathbf{c} - M\mathbf{u}_k) = (I - \omega M)\mathbf{u}_k + \omega\mathbf{c}, \quad k = 0, 1, 2, \dots \; . \qquad \text{(RE)}$$

Si osserva che $\mathbf{u} - \mathbf{u}_k = (I - \omega M)^k(\mathbf{u} - \mathbf{u}_0)$. Se $\mathbf{u} - \mathbf{u}_k \to \mathbf{0}$, $\forall\, \mathbf{u}_0$, allora si dice che (RE) è convergente.

(i) (RE) è convergente se e solo se $\rho(I - \omega M) < 1$ (risultato ben noto dalla teoria generale sui metodi iterativi stazionari per la risoluzione di sistemi lineari)

(ii) (RE) è convergente se e solo se $Re(\lambda_i(M)) > 0$ (oppure $Re(\lambda_i(M)) < 0$) $\forall\, i$. Nel caso $Re(\lambda_i(M)) > 0$ (oppure $Re(\lambda_i(M)) < 0$) $\forall\, i$, tutti e soli i valori di ω per cui (RE) converge sono in un intervallo del tipo $(0, \omega_*)$ (oppure $(-\omega_*, 0)$) per un $\omega_* > 0$.

(iii) Se M è definita positiva (cioè $A = A^T$ e $\mathbf{v}^T M\mathbf{v} > 0$ $\forall\, \mathbf{v} \in \mathbb{R}^n$ non nullo), allora (RE) converge se e solo se $\omega \in (0, 2/\max_i \lambda_i(M))$ e $\omega_{ott} := 2/(\min_i \lambda_i(M) + \max_i \lambda_i(M))$ è tale che $\rho(I - \omega_{ott}M) < \rho(I - \omega M)$ $\forall\, \omega \neq \omega_{ott}$.

Dimostrazione. Lasciata al lettore. $\square$

Il Teorema ci dice che:

1) Eulero applicato a (*) non è altro che (RE) applicato ad $A\mathbf{u} = b$, con $\mathbf{u}_0 = \mathbf{y}_0$;

2) Eulero applicato a (*) genera una successione η_k convergente ad $A^{-1}\mathbf{b}$ a patto che δt sia sufficientemente piccolo. Più precisamente, $\eta_k \to A^{-1}\mathbf{b}$ se e solo se $\delta t \in (0, 2/\max_i \lambda_i(A))$.

Il metodo (RE) fu introdotto da Richardson come metodo per risolvere sistemi lineari. Abbiamo visto che esso può essere ottenuto applicando il metodo di integrazione numerica di Eulero al problema (*). Per questi motivi esso è noto come metodo di Richardson-Eulero.

(RE) PER IL CALCOLO DI $\mathbf{p} = (P'')^T\mathbf{p}$, $\mathbf{p} > \mathbf{0}$, $\|\mathbf{p}\|_1 = 1$

Sia G un grafo orientato di vertici $1, 2, \dots, n$. Sia $\deg(i)$ il numero di archi uscenti dal vertice i di G ($i = 1, \dots, n$). Sia P la matrice di transizione di G: $P_{ij} = 1/\deg(i)$ se c'è un arco da i a j, e $P_{ij} = 0$ altrimenti. Sia $P' = P + \mathbf{d}\mathbf{v}^T$, $d_j = \delta_{\deg(j),0}$ ($\mathbf{d}$ =dangling page indicator), $\mathbf{v} > \mathbf{0}$, $\|\mathbf{v}\|_1 = 1$ ($\mathbf{v}$ =personalization vector). Sia $P'' = cP' + (1 - c)\mathbf{e}\mathbf{v}^T$, $c \in (0, 1)$, c vicino a 1.

Sia $\mathbf{p}$ il vettore vertexrank (pagerank), cioè l'unico vettore che soddisfa le proprietà

$$\mathbf{p} > \mathbf{0}, \;\; \|\mathbf{p}\|_1 = 1, \quad \mathbf{p} = (P'')^T\mathbf{p}.$$

È semplice verificare che $\mathbf{p}$ si ottiene risolvendo un sistema lineare. Infatti,

$$\text{dato } \mathbf{a} \text{ tale che } (I - cP^T)\mathbf{a} = \mathbf{v}, \quad c \in (0, 1), \; c \approx 1, \; \mathbf{v} > \mathbf{0}, \; \|\mathbf{v}\|_1 = 1, \quad \text{si ha } \mathbf{p} = \frac{\mathbf{a}}{\|\mathbf{a}\|_1}.$$

Per risolvere il sistema $(I - cP^T)\mathbf{a} = \mathbf{v}$ si può usare (RE), infatti gli autovalori della matrice dei coefficienti $I - cP^T$ hanno parte reale positiva (perché $\|P\|_\infty \leq 1$). Quindi (RE) è alternativo al metodo delle potenze nel calcolo di $\mathbf{p}$.

Lezione 2 Dicembre 2013

Inizialmente un riassunto:

$\mathbf{b} \in \mathbb{R}^n$, A reale definita positiva

I affermazione: $\mathbf{y}(t)$ tale che (1) $\mathbf{y}'(t) = \mathbf{b} - A\mathbf{y}(t)$, $\mathbf{y}(0) = \mathbf{y}_0$, converge ad $A^{-1}\mathbf{b}$ quando $t \to +\infty$ ($\forall\, \mathbf{y}_0$)

II affermazione: Il metodo di integrazione numerica "di Eulero" applicato a (1)

((2) $\eta_0 = \mathbf{y}(0) = \mathbf{y}_0$, $\eta_{k+1} = \eta_k + \delta t(\mathbf{b} - A\eta_k)$, $k = 0, 1, 2, \dots$ ($\eta_k = \eta(k\delta t) \approx \mathbf{y}(t_k)$))

genera una successione di vettori $\{\eta_k\}$ che converge ad $A^{-1}\mathbf{b}$ se e solo se $\delta t \in (0, \omega^*)$, con $\omega^* > 0$ opportuno, cioè se δt è abbastanza piccolo.

I affermazione: segue dall'espressione esplicita di $\mathbf{y}(t)$.

II affermazione: (2) coincide con (RE) applicato ad $A\mathbf{u} = \mathbf{b}$, che converge se e solo se $\omega \in (0, 2/\max_i \lambda_i(A))$.

NOTA: Le affermazioni I e II sono ancora vere nell'ipotesi più debole $Re(\lambda_i(A)) > 0$, $\forall\, i$.

Dimostrazione. I affermazione: segue dal Teorema di Perron, oppure direttamente dall'espressione esplicita di $\mathbf{y}(t)$ che si ottiene utilizzando la rappresentazione di Jordan $A = XJX^{-1}$ di A. II affermazione: il metodo (RE) applicato ad $A\mathbf{u} = \mathbf{b}$ converge se $Re(\lambda_i(A)) > 0$, $\forall\, i$ (vedi il Teorema (RE)). $\square$

Il vettore pagerank $\mathbf{p}$ risolve sistemi lineari:

$(I - cP^t)\mathbf{a} = \mathbf{v}$, $\mathbf{p} = \frac{1}{\|\mathbf{a}\|_1}\mathbf{a}$ (Nota: $\mathbf{a} > \mathbf{0}$, P^T è quasi-stocastica per colonne)

Esercizio: dimostrare che $(I - c(P')^T)\mathbf{p} = (1-c)\mathbf{v}$ (Nota: 1 è autovalore di $(P')^T$ con molteplicità algebrica e geometrica coincidenti ≥ 1 (vedi nostro Libro); $(P')^T$ è stocastica per colonne con autovalori 1, λ'_j, $j = 2, \dots, n$, $|\lambda'_j| \leq 1$)

$\mathbf{p}$ come limite di successioni di ddp (distr. discr. di prob.):

Potenze a $(P'')^T$

$\mathbf{p}^0 > \mathbf{0}$, $\|\mathbf{p}^0\|_1 = 1$, $\mathbf{p}^{k+1} = (P'')^T\mathbf{p}^k$, $k = 0, 1, \dots$ ($\|\mathbf{p}^k - \mathbf{p}\| = O(c^k|p(k)|)$).

Potenze inverse a $(P'')^T$ ($1 + \varepsilon$ è una approssimazione di 1 !)

$\mathbf{v}_0 > \mathbf{0}$, $\|\mathbf{v}_0\|_1 = 1$, $(I - \frac{1}{1+\varepsilon}(P'')^T)\mathbf{a}_k = \frac{1}{1+\varepsilon}\mathbf{v}_{k-1}$, $\mathbf{v}_k = \mathbf{a}_k/\|\mathbf{a}_k\|_1$, $k = 1, 2, \dots$

($\varepsilon > 0$; $\|\mathbf{v}_k - \mathbf{p}\| = O(|\varepsilon/(1 - c + \varepsilon)|^k|p(k)|)$, $p(k)$ lo stesso polinomio dell'altra successione; il fattore $|p(k)|$ non c'è se P'' è diagonalizzabile; $(P'')^T$ è stocastica per colonne positiva con autovalori 1, $\lambda''_j = c\lambda'_j$, $j = 2, \dots, n$, $|\lambda''_j - (1 + \varepsilon)| \geq 1 - c + \varepsilon$)

$\Rightarrow$ In tre casi occorre saper risolvere sistemi del tipo $(I - \tau A)\mathbf{x} = \mathbf{y}$, con $\mathbf{y} > \mathbf{0}$, $\tau < 1$ τ vicino a 1, $\rho(A) \leq 1$, $A \geq O$, A quasi-stocastica oppure stocastica per colonne, A tale che A per vettore costa poco. Nota: $\mathbf{x} > \mathbf{0}$.

In generale non conviene usare metodi diretti, soprattutto se n è grande. Se ad esempio la matrice A ha $O(n)$ elementi che la definiscono non ha senso con una decomposizione di tipo LU o QR di $I - \tau A$ andare a produrre altri $O(n^2)$ numeri che la definiscono (per poi dover risolvere due sistemi di complessità $O(n^2)$), anche perché il calcolo di tali decomposizioni può richiedere troppe operazioni, fino a $O(n^3)$. Il fatto che l'operazione "A per vettore" costa poco, suggerisce invece l'utilizzo di metodi iterativi per la risoluzione di sistemi del tipo $(I - \tau A)\mathbf{x} = \mathbf{y}$. Infatti, lo spazio richiesto per la loro implementazione può essere minimo, dell'ordine del numero dei parametri che definiscono A; e minimo può essere anche il costo di ogni iterazione, dell'ordine del costo dell'operazione "A per vettore". Se si usano metodi iterativi, però, e importante che questi convergano abbastanza rapidamente, cioè che non richiedano troppe iterazioni per produrre una buona approssimazione della soluzione.

Esempio. $Re(\lambda_i(I - \tau A)) > 0 \Rightarrow$ per risolvere $(I - \tau A)\mathbf{x} = \mathbf{y}$ usare (RE) con $\omega \in (0, \omega^*)$:

$\mathbf{x}_0 \in \mathbb{R}^n$, $\mathbf{x}^{j+1} = \mathbf{x}^j + \omega(\mathbf{y} - (I - \tau A)\mathbf{x}^j) = (I - \omega(I - \tau A))\mathbf{x}^j + \omega\mathbf{y} = H(\omega)\mathbf{x}^j + \omega\mathbf{y}$, $j = 0, 1, \dots$.

Si osserva che $\rho(H(1)) \leq \tau < 1$. Dunque $\omega^* > 1$. Inoltre si può dimostrare che $\rho(H(1)) < \rho(H(\omega))$ $\forall\, \omega \neq 1$, cioè la scelta $\omega = 1$ è ottimale. Quindi la massima rapidità di convergenza ottenibile con (RE) è $O(\tau^k |p(k)|)$ (perché c'è il fattore $p(k)$?)

Domande: $\mathbf{x}^0$ come si sceglie nei tre casi? Per quale delle tre A il polinomio $p(k)$ ha grado minimo?

Nel caso $A = P^T$ o $A = (P')^T$, tale rapidità di convergenza non è migliore di quella del metodo delle potenze applicato a $(P'')^T$.

Ma se ω potesse essere una matrice, forse si potrebbe rendere piu piccolo $\rho(H(\omega))$ ... Vedi la Lezione di Francesco del 5 Dicembre 2013.

Ho assegnato, come gli anni scorsi, l'esercizio di Langville e Meyer, un grafo di 6 vertici (ho citato un loro libro sull'argomento).

Esercizio: matrice di transizione associata a voi (... ha il cellulare di ...)

Ho insegnato, come gli anni scorsi, le pagine 139 ($(P'')^T \cdot \mathbf{z}$ costa quanto $P^T \cdot \mathbf{z}$) e 140 (a che serve $\mathbf{p}$, Google come usa $\mathbf{p}$?) degli appunti di Roberta Piersimoni.

OSSERVAZIONE

Discretizzando problemi differenziali di tipo ellittico con condizioni al bordo, si ottengono in particolare due sistemi lineari $M\mathbf{x} = \mathbf{c}$ dove M è rispettivamente

$$M = \begin{bmatrix} a_1 & -1 & 0 & \dots & .. & 0 \\ -1 & a_2 & -1 & 0.. & .. & 0 \\ 0 & -1 & a_3 & -10 & .. & 0 \\ . & . & & & & \\ . & . & . & & & \\ 0 & .. & 0 & -1 & a_{n-1} & -1 \\ 0 & .. & & 0 & -1 & a_n \end{bmatrix} \quad \text{(tridiagonale } n \times n) \tag{1}$$

con elementi diagonali a_i numeri reali maggiori o uguali di 2, e

$$M = \begin{bmatrix} B & -I & O & \dots & O \\ -I & B & O & \dots & O \\ .. & .. & & \dots & O \\ & & -I & B & -I \\ O & .. & O & -I & B \end{bmatrix} \quad \text{(tridiagonale } n \times n \text{ a blocchi)} \tag{2}$$

con

$$B = \begin{bmatrix} 4 & -1 & O & ... & O \\ -1 & 4 & & & \\ .. & & & & 0 \\ & & -1 & 4 & -1 \\ O & .. & O & -1 & 4 \end{bmatrix} \text{ (tridiagonale } n \times n)$$

(Nota: quest'ultima M è $n^2 \times n^2$).

Entrambi tali sistemi sono equivalenti a un sistema del tipo $(I - A)x = y$ con la matrice A non negativa, quasi-stocastica per colonne (se per quasi-stocastica per colonne si intende che la somma degli elementi di ogni colonna fa un numero compreso tra 0 e 1), e tale che $\rho(A) < 1$.

Dimostrazione. Per il primo sistema, basta moltiplicarne la prima riga per $1/a_1$, la seconda per $1/a_2$, ... , l'ultima per $1/a_n$; per il secondo sistema è ancora più facile, basta moltiplicarlo per $1/4$. $\square$

Appendice: Teorema di Gershgorin rafforzato (tratto dal nostro libro)

Ad ogni matrice $A \in \mathbb{C}^{n\times n}$ si possono associare i seguenti n sottoinsiemi di $\mathbb{C}$, detti *cerchi di Gershgorin*:

$$K_i = \{z \in \mathbb{C} : \, |z - a_{ii}| \leq \sum_{j:\, j\neq i} |a_{ij}|\}, \quad i = 1, \dots, n.$$

Notiamo che i cerchi K_i sono chiusi, quindi indichiamo con $\mathring{K}_i = K_i \backslash (\partial K_i)$ la loro parte interna.

Teorema 3.1 (Gershgorin) Sia $A \in \mathbb{C}^{n\times n}$ e λ un generico autovalore di A. Sia $\mathbf{x}$ un autovettore di A corrispondente all'autovalore λ, e $\mathcal{I} \subset \{1, 2, \dots, n\}$ l'insieme degli indici i per cui $|x_i| = \|\mathbf{x}\|_\infty$. Allora valgono le seguenti affermazioni:

1. $\lambda \in K_i$ per ogni $i \in \mathcal{I}$.
2. Se A è irriducibile e $\mathcal{I} \neq \{1, 2, \dots, n\}$, allora $\exists\, r \in \mathcal{I}$ tale che $\lambda \in \mathring{K}_r$.

Quindi, si ha sempre che $\lambda \in \cup_i K_i$, e, nel caso particolare in cui A è irriducibile, si ha più precisamente che $\lambda \in (\cup_i \mathring{K}_i) \cup (\cap_i \partial K_i)$.

Dimostrazione. Dall'identità $A\mathbf{x} = \lambda \mathbf{x}$ segue che $\sum_j a_{ij}x_j = \lambda x_i$, $\sum_{j:\, j\neq i} a_{ij}x_j = (\lambda - a_{ii})x_i$,

$$|\lambda - a_{ii}||x_i| \leq \sum_{j:\, j\neq i} |a_{ij}||x_j|, \quad i = 1, 2, \dots, n. \tag{3.1}$$

Per dimostrare il punto 1 basta osservare che scegliendo $i \in \mathcal{I}$ in (3.1) si ottengono le disuguaglianze $|\lambda - a_{ii}| \leq \sum_{j:\, j\neq i} |a_{ij}|$, $i \in \mathcal{I}$. Per 2, si nota che, poiché A è irriducibile, deve necessariamente esistere almeno una coppia di indici $r \in \mathcal{I}$ e $k \in \{1, \dots, n\}\backslash\mathcal{I}$ per cui $a_{rk} \neq 0$. Quindi (3.1) con $i = r$ implica

$$\begin{aligned} |\lambda - a_{rr}||x_r| &\leq \sum_{j:\, j\neq r} |a_{rj}||x_j| = \sum_{j\in\mathcal{I}:\, j\neq r} |a_{rj}||x_j| + \sum_{j\notin\mathcal{I}:\, j\neq k} |a_{rj}||x_j| + |a_{rk}||x_k| \\ &< \sum_{j\in\mathcal{I}:\, j\neq r} |a_{rj}||x_r| + \sum_{j\notin\mathcal{I}:\, j\neq k} |a_{rj}||x_r| + |a_{rk}||x_r| = |x_r| \sum_{j:\, j\neq r} |a_{rj}|, \end{aligned}$$

dove la diseguaglianza stretta segue dal fatto che $|x_k| < |x_r|$ e $a_{rk} \neq 0$. $\square$

Problema di Cauchy:

Data una funzione di due variabili $f(t,u)$ e due numeri reali x_0, y_0 tali che f è ben definita e continua in un intorno di (x_0, y_0), si può considerare il problema di chiedersi se esiste, se è unica, e se è possibile determinare una funzione $y(t)$, $y \in C^1$ in un intorno $\mathcal{I}$ di x_0, tale che

$$y'(t) = f(t, y(t)), \ t \in \mathcal{I}, \quad y(x_0) = y_0.$$

Esempio. Si consideri il seguente problema di Cauchy:

$$y'(t) = \sqrt{y(t)}, \ t \in \mathbb{R}, \quad y(0) = 0. \tag{$*$}$$

Esiste un numero infinito di funzioni $y(t) \in C^1(\mathbb{R})$ soluzioni di (*). Esse sono

$$y(t) = \left\{ \begin{cases} 0 & t < -c \\ (\frac{t+c}{2})^2 & t \geq -c \end{cases} : \ c \in (-\infty, 0] \right\}.$$

Si consideri il seguente problema di Cauchy:

$$y'(t) = \sqrt{-y(t)}, \ t \in \mathbb{R}, \quad y(0) = 0. \tag{$*$}$$

Esiste un numero infinito di funzioni $y(t) \in C^1(\mathbb{R})$ soluzioni di (*). Esse sono

$$y(t) = \left\{ \begin{cases} 0 & t > -c \\ -(\frac{t+c}{2})^2 & t \leq -c \end{cases} : \ c \in [0, +\infty) \right\}.$$

Si consideri il seguente problema di Cauchy:

$$y'(t) = \sqrt{|y(t)|}, \ t \in \mathbb{R}, \quad y(0) = 0. \tag{$*$}$$

Esiste un numero infinito di funzioni $y(t) \in C^1(\mathbb{R})$ soluzioni di (*). Esse sono

$$y(t) = \left\{ \begin{cases} (\frac{t+c'}{2})^2 & t \geq -c' \\ 0 & -c' > t > -c \\ -(\frac{t+c}{2})^2 & t \leq -c \end{cases} : \ c' \in (-\infty, 0], \ c \in [0, +\infty) \right\}.$$

(Dimostrare queste affermazioni!). Quindi ci sono problemi di Cauchy la cui soluzione non è univocamente definita.

Si noti che le funzioni $f(t,u)$ che definiscono le tre equazioni differenziali dell'esempio, cioè $f(t,u) = \sqrt{u}$, $\sqrt{-u}$ $\sqrt{|u|}$, pur essendo continue ovunque, non sono Lip., rispettivamente, nell'intorno destro, nell'intorno sinistro, e nell'intorno circolare di $y_0 = 0$ (per la Definizione di Lip. vedi più avanti). Dunque non sono soddisfatte tutte le ipotesi del ben noto Teorema di Esistenza e Unicità della soluzione di un Problema di Cauchy enunciato qui sotto. L'esempio di cui sopra mostra anche che per avere l'unicità della soluzione di un problema di Cauchy, è necessaria l'ipotesi di Lip. di $f(t,u)$ rispetto ad u in un intorno di $t = x_0$ e $u = y_0$.

Teorema. Sia $f : \mathbb{R}^2 \to \mathbb{R}$ continua in $[a,b] \times \mathbb{R}$ e Lip. rispetto al suo secondo argomento, cioè

$$|f(t,u) - f(t,v)| \leq L|u - v|, \ \forall\, t \in [a,b], \ \forall\, u, v \in \mathbb{R}.$$

Sia $(x_0, y_0) \in [a, b] \times \mathbb{R}$. Allora $\exists\,!\ y \in C^1([a, b])$ tale che

$$y'(t) = f(t, y(t)),\ t \in [a, b],\ y(x_0) = y_0.$$

Inoltre, se in tale funzione si mette in evidenza la sua dipendenza da y_0, allora, per $t \in [a, b]$, si ha

$$|y(t, y_0) - y(t, \tilde{y}_0)| \leq e^{L|t-x_0|}|y_0 - \tilde{y}_0|.$$

Ne segue che, se la costante di Lip. è grande, una piccola variazione del dato y_0 può comportare una grossa variazione di $y(t, y_0)$.

(Nota: i risultati enunciati continuano a valere con l'insieme $[a, b] \times \mathbb{R}$ ovunque sostituito da un insieme di tipo $K = [a, b] \times [c, d]$; però, in questo caso, la funzione $y(t)$ è ben definita, è C^1 e soddisfa l'equazione differenziale in un intorno di x_0 (contenuto in $[a, b]$) che può non essere prolungabile a tutto $[a, b]$).

Lip.

Una funzione $g(\mathbf{x})$, $g : \mathbb{R}^n \to \mathbb{R}^m$ si dice Lip. in $K \subset \mathbb{R}^n$ (si può immaginare K come la chiusura di un aperto convesso limitato) se esiste una costante $L \geq 0$ per cui $\|g(\mathbf{x}) - g(\tilde{\mathbf{x}})\| \leq L\|\mathbf{x} - \tilde{\mathbf{x}}\|$ *per ogni* $\mathbf{x}, \tilde{\mathbf{x}} \in K$. (Notare che g Lip. in K implica g continua in K, anzi uniformemente continua, visto che K è un compatto).

Una funzione $g \in C^1(K)$ è Lip. in K. Dimostriamolo solo nel caso $n = m = 1$, in cui K è un intervallo chiuso e limitato della retta reale: si ha

$$|g(x) - g(\tilde{x})| = |g'(\xi)(x - \tilde{x})| \leq \max_{w \in K} |g'(w)||x - \tilde{x}|,$$

$\xi \in$minimo intervallo contenente x e $\tilde{x}$; quindi basta porre $L = \max_{w \in K} |g'(w)|$.

Non è vero invece che g Lip. in K implica $g \in C^1(K)$. Vediamolo di nuovo solo nel caso $n = m = 1$. Ad esempio $g(x) = |x - 1|$ è Lip. in $[0, 2]$ (dimostrare che è Lip. in $\mathbb{R}$!), ma ovviamente non è C^1 in $[0, 2]$.

Sempre nel caso $n = m = 1$ diamo un esempio di funzione g continua in K che non è Lip. in K. Sia $g(x) = \sqrt{|x|}$. In $[0, 1]$ (o in $[-1, 0]$ o in $[-1, 1]$) la funzione g è ovviamente continua, ma non è Lip.. Ad esempio, la condizione $|g(x) - g(\tilde{x})| \leq L|x - \tilde{x}|$ non può essere soddisfatta per $\tilde{x} = 0$ e $x = 1/n$, al variare di $n \in \mathbb{N}$; infatti le disuguaglianze

$$|g(\frac{1}{n}) - g(0)| = \frac{1}{\sqrt{n}} \leq L|\frac{1}{n} - 0| = L\frac{1}{n},\ n \in \mathbb{N},$$

comunque grande si scelga $L \geq 0$, non possono essere tutte contemporaneamente verificate.

Le funzioni $f(t, u)$ dell'Esempio $f(t, u) = \sqrt{u}$, $\sqrt{-u}$ $\sqrt{|u|}$ non soddisfano la condizione

$$|f(t, u) - f(t, v)| \leq L|u - v|,\ \forall\, t \in \mathcal{I},\ \forall\, u, v \in \mathcal{I}'$$

(Lip. di f rispetto al suo secondo argomento), comunque si scelga $\mathcal{I}$ intorno di $x_0 = 0$, e comunque piccolo si scelga $\mathcal{I}'$, rispettivamente intorno destro, intorno sinistro, e intorno circolare di $y_0 = 0$.

Questo risultato, per quanto visto sopra, è evidente. Tuttavia lo riotteniamo qui sotto, facendo anche altre considerazioni.

Sia $f(t,u)$ la funzione definita su $\mathbb{R}^2$ dall'identità:

$$f(t,u) = \sqrt{|u|}.$$

Studiando il grafico della superficie f, si osserva che essa è continua su tutto $\mathbb{R}^2$, e che però $f \notin C^1(\mathbb{R}^2)$. Più precisamente si vede che $f \in C^1(\mathbb{R}^2 \backslash \{(t,0) : t \in \mathbb{R}\})$.

Ci si chiede ora se f è Lip. in un insieme K contenente il punto $(0,0)$. Sappiamo che se fosse $f \in C^1(K)$, f sarebbe Lip. in K (vedi sopra), ma $f \notin C^1(K)$. Dobbiamo allora vedere direttamente se, per qualche $L \geq 0$, è verificata la condizione:

$$|f(t,u) - f(\tilde{t},\tilde{u})| \leq L\|(t,u) - (\tilde{t},\tilde{u})\|, \ \forall\,(t,u),(\tilde{t},\tilde{u}) \in K.$$

Notiamo che quest'ultima condizione, per il fatto che la nostra $f(t,u)$ non dipende da t, diventa

$$|\sqrt{|u|} - \sqrt{|\tilde{u}|}| \leq L\|(t,u) - (\tilde{t},\tilde{u})\|, \ \forall\,(t,u),(\tilde{t},\tilde{u}) \in K,$$

che si riduce a

$$|\sqrt{|u|} - \sqrt{|\tilde{u}|}| \leq L\|u - \tilde{u}\|, \ \forall\,u,\tilde{u} \in [c,d] \tag{**}$$

($[c,d]$ =ordinate dei punti di K). Si noti che $0 \in [c,d]$ e che dunque, per quanto visto sopra, (**) non può essere soddisfatta per nessun $L \geq 0$.

Nota. Se f fosse stata la funzione $f(t,u) = |u|$, allora f sarebbe stata Lip. in K contenente $(0,0)$. In tal caso, bastava scegliere $L = 1$ (dimostrarlo!).

Se f fosse stata la funzione $f(t,u) = u^2$, allora f sarebbe stata Lip. in K contenente $(0,0)$. In tal caso, bastava scegliere $L = \max_{u,\tilde{u}\in[c,d]} |u + \tilde{u}|$ (dimostrarlo!).

Esercizio. Calcolare la soluzione dei problemi di Cauchy $y'(t) = -200ty(t)^2$, $y(-1) = 1/101$, e $y'(t) = y(t)^2$, $y(1) = -1$.

Esercizio. Calcolare la soluzione del seguente problema di Cauchy del secondo ordine $y''(t) - 4y'(t) + 3y(t) = t$, $y(0) = 4/9$, $y'(0) = 7/3$. (Nota: le equazioni differenziali del secondo ordine di solito hanno un insieme di soluzioni dipendente da due parametri). Scrivere il problema nella forma di un problema di Cauchy del primo ordine ma relativo a un sistema di due equazioni differenziali, cioè nella forma:

$$z_1'(t) = f_1(t, z_1(t), z_2(t)), \quad z_2'(t) = f_2(t, z_1(t), z_2(t)), \quad z_1(0) = z_1^0, \quad z_2(0) = z_2^0,$$

ovvero $\mathbf{z}'(t) = \mathbf{f}(t, \mathbf{z}(t))$, $\mathbf{z}(0) = \mathbf{z}^0$.

II esonero di Elementi di Analisi Numerica (Novembre 2013)

(1) Dimostrare che, per ogni $a, b, c \in \mathbb{R}$, almeno un autovalore della matrice

$$A = \begin{bmatrix} 1 & -2 & 1 & -\frac{1}{2} \\ -2 & 4 & -11/6 & 5/6 \\ 1 & -11/6 & a & b\mathbf{i} \\ -\frac{1}{2} & 5/6 & -b\mathbf{i} & c \end{bmatrix}$$

è nell'intervallo $(-1/9, 1/9)$. (Suggerimento: due colonne di A sono quasi linearmente dipendenti).

(2) Sia A la seguente matrice:

$$A = \begin{bmatrix} 3 & \sqrt{3}/2 & \frac{1}{2} \\ \sqrt{3}/2 & 2 & \sqrt{3} \\ \frac{1}{2} & \sqrt{3} & 0 \end{bmatrix}.$$

(i) Mediante una trasformazione per similitudine unitaria, trasformare A in una matrice del tipo

$$B = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0}^T & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & \delta \\ \delta & 3 \end{bmatrix}, \ \delta \in \mathbb{R}, \ |\delta| = 1$$

(usare il metodo di Givens che trasforma una matrice piena in una di Hessenberg). Dunque -1 è autovalore di A e gli altri due autovalori di A sono quelli di C.

(ii) Sia $M = B + x(\mathbf{e}_3\mathbf{e}_2^T + \mathbf{e}_2\mathbf{e}_3^T)$, $x \in \mathbb{R}$. Esiste un valore di x per cui $\det(\lambda I - M)$ ha due radici reali e coincidenti?

(iii) Calcolare gli autovalori di C effettuando un passo del metodo di Jacobi.

(iv) Sia $C_0 = C$ e siano M_0, N_0 2×2 tali che $C_0 = M_0 N_0$. Posto $C_1 = N_0 M_0$ si osserva che $C_1 = M_0^{-1} C_0 M_0$ e dunque C_1 ha gli stessi autovalori di C_0. Confrontare tra loro le tre matrici $C_1^{(LU)}$, $C_1^{(LL^H)}$, $C_1^{(QR)}$ ottenute utilizzando le decomposizioni LU, LL^H, QR di C_0, e prevedere la struttura di $C_\infty^{(LU)}$, $C_\infty^{(LL^H)}$, $C_\infty^{(QR)}$. (Facoltativo: notare che $C_1^{(QR)} = C_2^{(LL^H)}$).

(v) Introdurre A_0 2×2 tale che esistono L_0, U_0 per cui $A_0 = L_0 U_0$, ma non esistono L_1, U_1 tali che $A_1 := U_0 L_0 = L_1 U_1$, cioè anche se A $n \times n$ ammette la decomposizione LU, il generico passo del metodo LU per il calcolo degli autovalori di A può non essere ben definito.

(3) (i) Sia A la seguente matrice $n \times n$:

$$A = \mu I + \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}, \ \mu \in \mathbb{R}.$$

Sia $A(\varepsilon)$ la perturbazione di A definita dall'identità $A(\varepsilon) = A + \varepsilon \mathbf{e}_i \mathbf{e}_j^T$, $\varepsilon > 0$, e sia $\sigma(A(\varepsilon))$ l'insieme degli autovalori di $A(\varepsilon)$. Calcolare, in funzione dei valori di i, j, le possibili espressioni di

$$\max_{\lambda(\varepsilon) \in \sigma(A(\varepsilon))} |\lambda(\varepsilon) - \mu|.$$

(ii) Nel punto (i) si è visto che gli autovalori di una matrice A con molteplicità algebrica maggiore della geometrica possono essere molto sensibili a variazioni degli elementi di A.

In questo punto si mostra che una perturbazione $O(\varepsilon)$ degli elementi di A implica una perturbazione dei suoi autovalori semplici che non può essere di ordine maggiore di $O(\varepsilon)$.

Siano $A \in \mathbb{R}^{n\times n}$, λ un autovalore semplice reale di A, e $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ non nulli tali che $A\mathbf{x} = \lambda\mathbf{x}$ e $\mathbf{y}^T A = \lambda\mathbf{y}^T$. (Facoltativo: dimostrare che $\mathbf{y}^T\mathbf{x} \neq 0$, ad esempio utilizzando la decomposizione di Jordan di A).

Sia $C \in \mathbb{R}^{n\times n}$. Per $\varepsilon \in \mathbb{R}$ siano $\lambda(\varepsilon) \in \mathbb{R}$ e $\mathbf{x}(\varepsilon) \in \mathbb{R}^n$ l'autovalore e l'autovettore della matrice $A+\varepsilon C$ che per $\varepsilon = 0$ coincidono rispettivamente con λ e $\mathbf{x}$. È noto che $\lambda(\varepsilon)$ è una funzione analitica di ε in un intorno di 0.

Usando l'identità $(A + \varepsilon C)\mathbf{x}(\varepsilon) = \lambda(\varepsilon)\mathbf{x}(\varepsilon)$, dimostrare che

$$\lambda(\varepsilon) - \lambda(0) = L\varepsilon + O(\varepsilon^2), \quad L = \frac{\mathbf{y}^T C\mathbf{x}}{\mathbf{y}^T\mathbf{x}}$$

e, nel caso in cui A e C sono simmetriche, dare una limitazione per $|L|$.

(4) La principessa erede al trono, l'ultimo giorno della sua visita alla città di Roma, salutando i giornalisti in una maestosa sala di un palazzo romano, si accorge che tra di essi vi sono anche le due persone che il giorno prima l'hanno accompagnata in giro per Roma, alle quali aveva nascosto la sua vera identità e da una delle quali si era separata con profondo dolore.

Conservandosi calma continua il suo discorso; è allo stesso tempo commossa, per avere ancora di fronte a se la persona amata, e timorosa, che quelle due persone potessero anche metterla in ridicolo riferendosi, in un loro eventuale intervento, alle cose che lei aveva fatto con loro in giro per Roma il giorno prima, cose non molto opportune per una principessa.

La persona amata effettivamente interviene con una domanda, ma dalle parole di tale domanda risulta alla principessa evidente che essi avrebbero tenuto segreto il loro incontro del giorno prima e così ella si tranquillizza.

Al termine della conferenza stampa ella esprime il desiderio di salutare personalmente alcuni dei giornalisti, confidando che in questo modo avrebbe potuto salutare da vicino, ancora una volta, la persona che il giorno prima l'aveva fatta sentire tanto bene.

Quando lei arriva presso i due giornalisti suoi conoscenti, essi approfittano dell'occasione per restituirle, in una busta da lettera, delle cose che il giorno prima le avevano sottratto a sua insaputa.

Cosa le avevano sottratto?

(B. Iannazzo) A $n \times n$, $A \geq O$, A irreducible, $\alpha > \rho(A)$.

$M = \alpha I - A$, $M_{\pm} = \begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix} M$.

$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \Rightarrow M = \begin{bmatrix} \alpha I - A_{11} & -A_{12} \\ -A_{21} & \alpha I - A_{22} \end{bmatrix}$, $M_{\pm} = \begin{bmatrix} \alpha I - A_{11} & -A_{12} \\ A_{21} & A_{22} - \alpha I \end{bmatrix}$

Prove that

(i) $M_{\pm}$ has k eigenvalues $\mu_1, \dots, \mu_k$ with real part greater than $\alpha - \rho(A)$, and $n-k$ eigenvalues $\mu_{k+1}, \dots, \mu_n$ with real part smaller than $\rho(A) - \alpha$.

(ii) Let μ_k, μ_{k+1} be such that $|\mu_k| = \min_{j=1,\dots,k} |\mu_j|$, $|\mu_{k+1}| = \min_{j=k+1,\dots,n} |\mu_j|$. Then μ_k and μ_{k+1} are real.

Proof of (i).

By P.F. theory, $\rho(A)$ is positive, $\rho(A)$ is a simple eigenvalue of A, and there exists $\mathbf{z} > \mathbf{0}$ such that $A\mathbf{z} = \rho(A)\mathbf{z}$. Set $\mathbf{e} = [1\,1 \dots 1]^T$ and $D = d(\mathbf{z})$. Then $AD\mathbf{e} = \rho(A)D\mathbf{e}$ and therefore $D^{-1}AD\mathbf{e} = \rho(A)\mathbf{e}$. So we have

$$D^{-1}\frac{1}{\rho(A)}AD\mathbf{e} = \mathbf{e},$$

i.e. $D^{-1}\frac{1}{\rho(A)}AD$ is a non negative, irreducible, stochastic by rows matrix. Then

$$\frac{1}{\rho(A)}D^{-1}MD = \frac{\alpha}{\rho(A)}I - D^{-1}\frac{1}{\rho(A)}AD = \delta I - R$$

with $R \geq O$, irreducible, stochastic by rows, and $\delta = \frac{\alpha}{\rho(A)} > 1 = \rho(R)$.

The ith Gershgorin circle of $\delta I - R = \frac{1}{\rho(A)}D^{-1}MD$ is $C(\delta - r_{ii}, 1 - r_{ii})$. So, all Gershgorin circles of $\delta I - R = \frac{1}{\rho(A)}D^{-1}MD$ pass through the point $\delta - 1$. Note that $\delta - 1$ is an eigenvalue of $\delta I - R = \frac{1}{\rho(A)}D^{-1}MD$.

The ith Gershgorin circle of $* = \begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix}(\delta I - R) = \begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix}\frac{1}{\rho(A)}D^{-1}MD$ is $C(\delta - r_{ii}, 1 - r_{ii})$, $i = 1, \dots, k$, and $C(r_{ii} - \delta, 1 - r_{ii})$, $i = k+1, \dots, n$. So the first k of such Gershgorin circles pass through the point $\delta - 1$, and the remaining $n-k$ pass through the point $1 - \delta$. Note that $\delta - 1$ and $1 - \delta$ are not eigenvalues (if $1 \leq k \leq n-1$).

Then, by the three Gershgorin theorems, k eigenvalues of * have real part greater than $\alpha/\rho(A) - 1$ and the remaining $n-k$ have real part smaller than $1 - \alpha/\rho(A)$.

Note that $* = \frac{1}{\rho(A)}D^{-1}\begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix}MD$.

Thus k eigenvalues of $\frac{1}{\rho(A)}\begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix}M$ have real part greater than $\alpha/\rho(A) - 1$ and the remaining $n-k$ have real part smaller than $1 - \alpha/\rho(A)$.

So k eigenvalues of $M_{\pm} = \begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix}M$ have real part greater than $\alpha - \rho(A)$ and the remaining $n-k$ have real part smaller than $\rho(A) - \alpha$. $\square$

Example. If $R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then the eigenvalues of $\delta I - R$ are $\delta - 1$ and $\delta + 1$.

If $\delta > 1$, the eigenvalues of $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} (\delta I - R)$ are $\pm\sqrt{\delta^2 - 1}$. Note that they are real, and that one of them is greater than $\delta - 1$ ($\sqrt{\delta^2-1}$) and the other one is smaller than $1 - \delta$ ($-\sqrt{\delta^2-1}$).

Proof of (ii).

R $n \times n$, $R \geq O$, R irreducible, R stochastic by rows, $\delta > \rho(R) = 1$.

$M = \delta I - R$, $M_\pm = \begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix} M$.

$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \Rightarrow M = \begin{bmatrix} \delta I - R_{11} & -R_{12} \\ -R_{21} & \delta I - R_{22} \end{bmatrix}$, $M_\pm = \begin{bmatrix} \delta I - R_{11} & -R_{12} \\ R_{21} & R_{22} - \delta I \end{bmatrix}$

We know that

(i) $M_\pm$ has k eigenvalues $\lambda_1^+, \dots, \lambda_k^+$ with real part greater than $\delta - 1$, and $n - k$ eigenvalues $\lambda_{k+1}^-, \dots, \lambda_n^-$ with real part smaller than $1 - \delta$.

We want to prove the following assertion:

(ii) Let $\lambda_k^+, \lambda_{k+1}^-$ be such that $|\lambda_k^+| = \min_{j=1,\dots,k} |\lambda_j^+|$, $|\lambda_{k+1}^-| = \min_{j=k+1,\dots,n} |\lambda_j^-|$. Then λ_k^+ and λ_{k+1}^- are real.

Introduce the notation

$$\begin{bmatrix} \delta I - R_{11} & -R_{12} \\ R_{21} & R_{22} - \delta I \end{bmatrix} \begin{bmatrix} \lceil X_+ \rceil & \lceil X_- \rceil \\ \lfloor X_+ \rfloor & \lfloor X_- \rfloor \end{bmatrix} = \begin{bmatrix} \lceil X_+ \rceil & \lceil X_- \rceil \\ \lfloor X_+ \rfloor & \lfloor X_- \rfloor \end{bmatrix} \begin{bmatrix} J_+ & O \\ O & J_- \end{bmatrix},$$

$$M_\pm X_+ = X_+ J_+, \; M_\pm X_- = X_- J_-,$$

where on the diagonal of J_+ there are $\lambda_1^+, \dots, \lambda_k^+$, and on the diagonal of J_- there are $\lambda_{k+1}^-, \dots, \lambda_n^-$.

Then the eigenvalues of the matrix

$$M_\pm - X_- \lfloor X_- \rfloor^{-1} [O \; I_{n-k \times n-k}] M_\pm = \begin{bmatrix} *_{k \times k} & * \\ O & O_{n-k \times n-k} \end{bmatrix} \tag{1}$$

are the eigenvalues of J_+ plus 0 with multiplicity $n - k$; and the eigenvalues of the matrix

$$M_\pm - X_+ \lceil X_+ \rceil^{-1} [I_{k \times k} \; O] M_\pm = \begin{bmatrix} O_{k \times k} & O \\ * & *_{n-k \times n-k} \end{bmatrix} \tag{2}$$

are the eigenvalues of J_- plus 0 with multiplicity k. (Proof: apply the result (iii) of Exercise 4 of I esonero EAN for $X = X_-$ and $W^H = [O \; I]$, and for $X = X_+$ and $W^H = [I \; O]$, respectively).

It follows that the upper left $k \times k$ submatrix of (1),

$$\delta I - (R_{11} + \lceil X_- \rceil \lfloor X_- \rfloor^{-1} R_{21})$$

has the eigenvalues of J_+; and the lower right $n - k \times n - k$ submatrix of (2),

$$-\delta I + (R_{22} + \lfloor X_+ \rfloor \lceil X_+ \rceil^{-1} R_{12})$$

has the eigenvalues of J_-.

If $R_{11} + \lceil X_- \rceil \lfloor X_- \rfloor^{-1} R_{21}$ and $R_{22} + \lfloor X_+ \rfloor \lceil X_+ \rceil^{-1} R_{12}$ are well defined (i.e. $\lfloor X_- \rfloor$ and $\lceil X_+ \rceil$ are invertible) and nonnegative (!!!we will show that these assumptions are verified when $k = n-1$!!!), then we would have

$$\lambda_k^+ = \delta - \rho(R_{11} + \lceil X_- \rceil \lfloor X_- \rfloor^{-1} R_{21}) \ and \ \lambda_{k+1}^- = -\delta + \rho(R_{22} + \lfloor X_+ \rfloor \lceil X_+ \rceil^{-1} R_{12}),$$

and $Re(\lambda_j^+) > \lambda_k^+$, $\forall \lambda_j^+ \neq \lambda_k^+$, and $Re(\lambda_j^-) < \lambda_{k+1}^-$, $\forall \lambda_j^- \neq \lambda_{k+1}^-$, i.e. we would have the thesis.

Case $k = n-1$:

In this case J_- is 1×1 and is the only eigenvalue of $M_\pm$ with real part less than $1-\delta$. Moreover, it J_- must be real because otherwise also its conjugate would be eigenvalue of $M_\pm$ ($M_\pm$ has real entries!) and thus $M_\pm$ would have two eigenvalues with real part less than $1-\delta$, against the II Gershgorin Theorem.

Moreover, $\lfloor X_- \rfloor \neq 0$, in fact $\lfloor X_- \rfloor = 0$ would imply $(\delta I - R_{11}) \lceil X_- \rceil = \lceil X_- \rceil J_-$ with $\lceil X_- \rceil \neq \mathbf{0}$, i.e. J_- would be an eigenvalue of $\delta I - R_{11}$ and this is not possible since the eigenvalues of $\delta I - R_{11}$ have real part greater than $\delta - 1$.

Also, $\lceil X_- \rceil$ must verify the equation $(\delta I - R_{11}) \lceil X_- \rceil - R_{12} \lfloor X_- \rfloor = \lceil X_- \rceil J_-$, thus $\lceil X_- \rceil = ((\delta - J_-)I - R_{11})^{-1} R_{12} \lfloor X_- \rfloor$ with $\delta - J_- > \delta > \rho(R) = 1 \geq \rho(R_{11})$. So

$$\begin{aligned} \lceil X_- \rceil &= ((\delta - J_-)I - R_{11})^{-1} R_{12} \lfloor X_- \rfloor \\ &= (\delta - J_-)^{-1} (I - \frac{1}{\delta - J_-} R_{11})^{-1} R_{12} \lfloor X_- \rfloor \\ &= (\delta - J_-)^{-1} \sum_{k=0}^{+\infty} \frac{1}{(\delta - J_-)^k} R_{11}^k R_{12} \lfloor X_- \rfloor. \end{aligned}$$

Since $\delta I - (R_{11} + \lceil X_- \rceil \lfloor X_- \rfloor^{-1} R_{21})$ has the eigenvalues of J_+, and since $R_{11} + \lceil X_- \rceil \lfloor X_- \rfloor^{-1} R_{21} = R_{11} + (\delta - J_-)^{-1} \sum_{k=0}^{+\infty} \frac{1}{(\delta - J_-)^k} R_{11}^k R_{12} R_{21}$ is a nonnegative matrix we can say that an eigenvalue of J_+ is $\delta - \rho\Big(R_{11} + (\delta - J_-)^{-1} \sum_{k=0}^{+\infty} \frac{1}{(\delta - J_-)^k} R_{11}^k R_{12} R_{21}\Big)$, call it λ_k^+, and thus the eigenvalues of J_+ different from λ_k^+ have real part greater than λ_k^+.

Dal I esonero di EAN (dell'Ottobre 2013)

4) Siano: A generica $n \times n$ ($a_{ij} \in \mathbb{C}$),
X $n \times k$ ($k < n$) con colonne linearmente indipendenti tale che $AX = XJ$ per una matrice $k \times k$ J,
Y $n \times (n-k)$ tale che la matrice $n \times n$ $[X\ Y]$ è non singolare. Dimostrare che
(i) Esiste una matrice $(n-k) \times (n-k)$ B tale che

$$[X\ Y]^{-1} A [X\ Y] = \begin{bmatrix} J & * \\ O & B \end{bmatrix}.$$

(ii) (Facoltativo) Per ogni matrice $k \times n$ W^H tale che $\det(W^H X) \neq 0$ si ha

$$[X\ Y]^{-1} \Big(A - XJ(W^H X)^{-1} W^H\Big) [X\ Y] = \begin{bmatrix} O & *' \\ O & B \end{bmatrix}.$$

(iii) Per ogni matrice $k \times n$ W^H tale che $\det(W^H X) \neq 0$ si ha

$$[X\ Y]^{-1} \Big(A - X(W^H X)^{-1} W^H A\Big) [X\ Y] = \begin{bmatrix} O & *'' \\ O & B \end{bmatrix}$$

e $W^H \Big(A - X(W^H X)^{-1} W^H A\Big)$ è la matrice nulla $k \times n$.

I esonero di Elementi di Analisi Numerica (Ottobre 2013)

1) (i) Posto $\mathcal{I} = \int_1^{\sqrt{3}} \frac{1}{1+x^2}\,dx$, trovare le approssimazioni $\mathcal{I}_{\sqrt{3}-1}$ e $\mathcal{I}_{\frac{\sqrt{3}-1}{2}}$ di $\mathcal{I}$ fornite dalla formula dei trapezi e dedurne, tramite il metodo di estrapolazione di Romberg, l'approssimazione di ordine superiore $\tilde{\mathcal{I}}_{\frac{\sqrt{3}-1}{2}}$. (Facoltativo: notare che $\tilde{\mathcal{I}}_{\frac{\sqrt{3}-1}{2}}$ approssima $\mathcal{I}$ molto meglio di $\mathcal{I}_{\frac{\sqrt{3}-1}{4}}$).
(ii) Si può dire che la successione di approssimazioni $\mathcal{I}_{\frac{\sqrt{3}-1}{2^k}}$, $k = 0, 1, 2, 3, \ldots$, è non crescente?

2) Sia $f_t(x) = \sum_{n=0}^{+\infty} B_n(x) t^n/n!$.
(i) Dimostrare che $f_t(x)$ soddisfa le due condizioni $\frac{d}{dx} f_t(x) = t f_t(x)$ e $\int_0^1 f_t(x)\,dx = 1$.
(ii) Dedurne una formula chiusa per $f_t(x)$.

3) Si vuole approssimare $\varsigma(1+\varepsilon) = \sum_{r=1}^{+\infty} \frac{1}{r^{1+\varepsilon}}$ uniformemente per $\varepsilon \in (0, 2]$.
(i) Usando la formula di Eulero-Maclaurin, trovare una funzione $\varphi(\varepsilon, m, k)$, $\varepsilon \in \mathbb{R}^+$, $m \geq 1$, $k \geq 0$ ($m, k \in \mathbb{Z}$), per cui vale l'identità

$$\sum_{r=m}^{+\infty} \frac{1}{r^{1+\varepsilon}} = \varphi(\varepsilon, m, k) + u_{k+1}(\infty),$$

dando una maggiorazione per $|u_{k+1}(\infty)|$ (in funzione di ε, m, k).
(ii) Usare la maggiorazione ottenuta nel punto precedente per trovare una costante $M_{m,k}$ tale che $|u_{k+1}(\infty)| \leq M_{m,k}$ per ogni $\varepsilon \in [1, 2]$.
(iii) Porre $m = 2$ e, valutando la costante $M_{2,k}$ del punto precedente per $k = 0, 1, 2, \ldots$, determinare il minimo valore di k per cui

$$|\{1 + \varphi(\varepsilon, 2, k)\} - \varsigma(1+\varepsilon)| \leq \frac{1}{2^9}, \quad \forall \varepsilon \in [1, 2].$$

(Quest'ultimo risultato vale in realtà per ogni $\varepsilon \in (0, 2]$, con lo stesso k).

4) Siano: A generica $n \times n$ ($a_{ij} \in \mathbb{C}$),
X $n \times k$ ($k < n$) con colonne linearmente indipendenti tale che $AX = XJ$ per una matrice $k \times k$ J,
Y $n \times (n-k)$ tale che la matrice $n \times n$ $[X\ Y]$ è non singolare. Dimostrare che
(i) Esiste una matrice $(n-k) \times (n-k)$ B tale che

$$[X\ Y]^{-1} A [X\ Y] = \begin{bmatrix} J & * \\ O & B \end{bmatrix}.$$

(ii) (Facoltativo) Per ogni matrice $k \times n$ W^H tale che $\det(W^H X) \neq 0$ si ha

$$[X\ Y]^{-1} \Big(A - XJ(W^H X)^{-1} W^H \Big) [X\ Y] = \begin{bmatrix} O & *' \\ O & B \end{bmatrix}.$$

(iii) Per ogni matrice $k \times n$ W^H tale che $\det(W^H X) \neq 0$ si ha

$$[X\ Y]^{-1} \Big(A - X(W^H X)^{-1} W^H A \Big) [X\ Y] = \begin{bmatrix} O & *'' \\ O & B \end{bmatrix}.$$

e $W^H \Big(A - X(W^H X)^{-1} W^H A \Big)$ è la matrice nulla $k \times n$.

(iv) Posto

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ a & 0 & 1-a & 0 \\ 0 & b & 0 & 1-b \\ 0 & 0 & 1 & 0 \end{bmatrix}, \ X = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, \ J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

si osserva che $AX = XJ$; dunque lo spettro di A è del tipo $\sigma(A) = \{1, -1, \lambda_3, \lambda_4\}$. Osservare che si può usare il risultato in (iii) per

$$W^H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

e, quindi, introdurre una matrice 2×2 i cui autovalori sono λ_3 e λ_4.