

Metodi di Approssimazione

Riassunto

Alex Foglia

January 8, 2019

Il seguente documento contiene un riassunto del corso di Metodi di Approssimazione tenuto dal prof. Luigi Brugnano durante l'A.A. 2017-18.

Esso contiene una sintesi dei risultati più importanti mostrati a lezione, e le rispettive applicazioni in contesti pratici, senza tuttavia la pretesa di essere un sostituto del materiale ufficiale fornito dal Brugnano. Si utilizzerà un linguaggio quanto più possibile semplice e un formalismo ridotto all'essenziale; con l'obiettivo che il testo risulti il più chiaro possibile a tutti coloro che, pur provenendo da un corso di laurea di stampo scientifico, non sono nè laureati in matematica, nè studenti magistrali di matematica.¹

¹Per mancanza di tempo, non sono riuscito a sintetizzare il capitolo 7, che viene allegato per intero

1 I Sistemi Dinamici

Molti fenomeni studiati nelle applicazioni (per esempio, il meteo) hanno una natura *evolutiva*. Possiamo modellare i fenomeni evolutivi attraverso un processo dinamico che evolve nel tempo. Tali processi dinamici, da un punto di vista matematico, sono descritti attraverso la soluzione di un *problema ai valori iniziali*.

- Se il tempo evolve in un intervallo continuo, avremo un problema ai valori iniziali per *equazioni differenziali*
- Se il tempo evolve in un intervallo discreto, avremo un problema ai valori iniziali per *equazioni alle differenze*

Formalizzando, nel caso continuo avremo un problema della forma:

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

$$y(0) = y_0 \in \mathbb{R}^m$$

Nel caso discreto invece, avremo un problema della forma:

$$y_{n+1} = f(n, y_n)$$

$$y_0 \in \mathbb{R}^m$$

Per esempio, sappiamo che la coordinata y di un punto materiale di massa m , soggetto alla sola forza peso $-mg$, la si può modellare come un processo che evolve nel tempo, un tempo continuo, e la si ricava risolvendo un problema ai valori iniziali:

$$y''(t) = -mg$$

$$y(0) = y_0 \in \mathbb{R}$$

Dove y_0 è la quota y nell'istante in cui iniziamo a studiare il moto del punto (potrebbe essere l'istante in cui lo lasciamo cadere).

La soluzione a questo problema è una funzione $y(t)$ che ci fornisce la coordinata y del punto per ogni istante di tempo $t \in [0, T]$ oppure $t \in [0, \infty)$, che sappiamo essere:

$$y(t) = \frac{1}{2}gt^2 + v_0t + y_0$$

Ossia la legge oraria del moto rettilineo uniformemente accelerato.

La funzione $y(t)$, soluzione del problema continuo, fornisce dunque quella che chiamiamo *traiettoria continua*. Similmente, la *successione* $\{y_n\}$ con $n \geq 0$, soluzione del problema discreto, descrive una *traiettoria discreta*.

Durante il corso ci occuperemo di come un problema continuo possa essere trasformato in un problema discreto preservandone le stesse proprietà formali. Discretizzare un problema continuo è necessario dal momento che per risolvere questi problemi noi utilizziamo, al giorno d'oggi, calcolatori che lavorano in *aritmetica finita*.

1.1 Condizionamento e stabilità

Dato un problema ai valori iniziali continuo, definiamo una *mappa continua* $\phi_t : \mathbb{R} \rightarrow S$ dove S è un insieme di funzioni, la funzione che, applicata a $y_0 \in \mathbb{R}^m$, genera la soluzione $y(t)$ del problema ai valori iniziali che stiamo considerando. Similmente possiamo definire una *mappa discreta* come la funzione $\phi_n : \mathbb{R} \rightarrow S$, dove S stavolta è un insieme di successioni, come la funzione che, applicata a y_0 fornisce la soluzione y_n del problema discreto ai valori iniziali.

Immaginiamo adesso di perturbare leggermente y_0 . Studiare come si comporta la traiettoria soluzione al netto di una piccola perturbazione, significa studiare il condizionamento della soluzione.

Definiamo un *insieme critico* un insieme $P \subset \mathbb{R}^m$ tale per cui, nel caso continuo:

$$\bigcup_{t \geq 0} \phi_t(P) = P$$

Mentre nel caso discreto:

$$\bigcup_{n \geq 0} \phi_n(P) = P$$

Il caso più semplice di insieme critico è un punto di equilibrio \bar{y} tale per cui

$$f(t, \bar{y}) = 0 \quad \forall t \geq 0$$

nel caso continuo. Nel caso discreto invece, un punto di equilibrio è un valore \bar{y} tale per cui

$$f(n, \bar{y}) = \bar{y} \quad \forall n \geq 0$$

Ciò significa che, se imponiamo come condizione iniziale un insieme critico, allora la traiettoria deve rimanere sull'insieme critico. Questi insiemi hanno rilevanza fisica solo se piccole perturbazioni nella dinamica permettono di tornare almeno asintoticamente sull'insieme critico: ci interessa dunque studiare la *stabilità* di un insieme critico. Facciamo un esempio:

$$y'(t, y) = \sin(y)$$

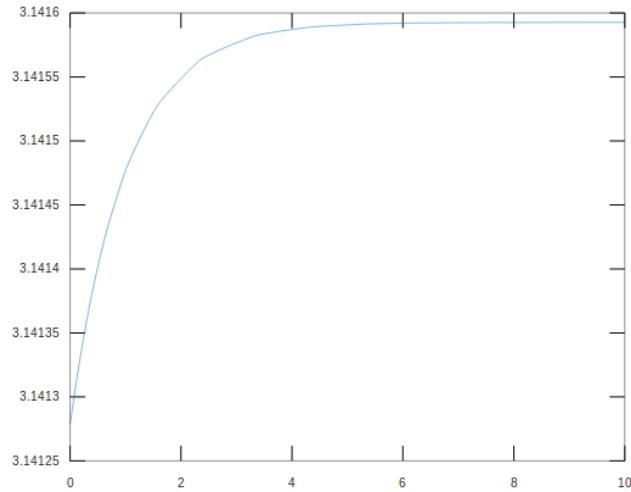
$$y(0) = \pi$$

Notiamo che, per ogni t , $y'(t, k\pi) = 0$ $k \in \mathbb{Z}$, quindi ad esempio $\bar{y} = \pi$ è un punto di equilibrio, infatti $y'(t, \pi) = \sin(\pi) = 0$. Se π è un punto di equilibrio allora porre come condizione iniziale $y(0) = \pi$ deve significare che la soluzione al problema rimane sul punto di equilibrio $\bar{y} = \pi$. Fatto che si verifica sperimentalmente abbastanza velocemente, di fatti la soluzione al problema è

$$y(t) = \pi$$

Ossia la funzione costante di valore π .

Se invece imponessimo come condizione iniziale una piccola perturbazione di π , come ad esempio 0.9999π otteniamo una funzione che, asintoticamente, torna ad assumere il valore π :



La proprietà di stabilità di un insieme critico deve essere mantenuta dai sistemi dinamici discreti indotti dai modelli continui, induzione che avviene attraverso gli opportuni *metodi di approssimazione* che studieremo durante il corso.

2 Equazioni alle differenze: generalità

Le equazioni alle differenze costituiscono la controparte discreta delle equazioni differenziali. Per capire di cosa parleremo, occorre anzitutto definire cosa sia una *funzione discreta*.

Definiamo un *dominio discreto* $\Omega = \{x_0, x_1, \dots\}$ come un insieme di elementi $x_n \in \mathbb{R}$ tali per cui $x_n < x_{n+1}$ con $n = 0, 1, \dots$

Se f è una funzione $f : \Omega \rightarrow V$ con V spazio vettoriale, allora f è una *funzione discreta*. D'ora in avanti adotteremo la notazione $f_n = f(x_n)$.

2.1 Gli operatori Δ, E, I

In questa sezione definiamo tre operatori applicabili a una funzione discreta:

- shift E : $Ef_n = f_{n+1}$
- identità I : $If_n = f_n$
- differenza in avanti Δ : $\Delta f_n = f_{n+1} - f_n$

Tali operatori godono delle seguenti proprietà:

1. Linearità:

$$\Delta(\alpha f_n + \beta g_n) = \alpha \Delta f_n + \beta \Delta g_n$$

$$E(\alpha f_n + \beta g_n) = \alpha Ef_n + \beta Eg_n$$

$$I(\alpha f_n + \beta g_n) = \alpha If_n + \beta Ig_n$$

2. $\Delta = E - I$

3. Commutatività

4. Potenze:

$$E^0 = \Delta^0 = I$$

$$I^k f_n = If_n$$

$$E^k f_n = f_{n+k} \text{ (shift in avanti } k \text{ volte)}$$

$$\Delta^k f_n = (E - I)^k f_n = \sum_{i=0}^k \binom{k}{i} (-1)^{k-i} E^i f_n$$

Osserviamo che Δ è la controparte discreta dell'operatore di derivata nel continuo, infatti se z_n, y_n sono successioni, abbiamo che:

$$\Delta(z_n y_n) = z_n \Delta y_n + \Delta z_n y_n + \Delta z_n \Delta y_n$$

$$\Delta \left(\frac{y_n}{z_n} \right) = \frac{z_n \Delta y_n - \Delta z_n y_n}{z_n z_{n+1}}$$

Sia $\{y_n\}$ una successione nota. Si ha che

$$\sum_{n=n_0}^{N-1} \Delta y_n = y_N - y_{n_0}$$

Il che è un risultato molto simile al teorema fondamentale del calcolo integrale:

$$\int_a^b f(x)dx = F(b) - F(a)$$

In cui vale la relazione $F'(x) = f(x)$.

La somma è quindi la controparte discreta dell'operatore di integrazione.

2.2 Equazioni alle differenze del primo ordine

L'equazione alle differenze più semplice che ci possiamo trovare davanti è un'equazione della forma:

$$\Delta y_n = g_n$$

Con la sola g_n successione nota. Per esempio:

$$\Delta y_n = 2n + 1$$

È un'equazione alle differenze del primo ordine. La chiamiamo del *primo* ordine poichè possiamo riscriverla come:

$$y_{n+1} - y_n = g_n = 2n + 1$$

Vedremo meglio successivamente perchè questa forma di equazione alle differenze si classifica come equazione del *primo* ordine. Intuitivamente, si dice che è del primo ordine perchè la differenza in avanti di ordine massimo che compare è proprio y_{n+1} .

In questa sezione siamo interessati a trattare la soluzione di un'equazione alle differenze del primo ordine da un punto di vista esclusivamente *teorico*, senza addentrarci per il momento nella risoluzione effettiva.

Dunque, se

$$\Delta y_n = g_n$$

allora diciamo formalmente che la soluzione y_n è data da:

$$y_n = \Delta^{-1} g_n$$

Per esempio:

$$\Delta y_n = 2n + 1$$

$$y_n = \Delta^{-1}(2n + 1)$$

L'operatore Δ^{-1} è l'operatore *antidifferenza* per cui vale che:

- $\Delta(\Delta^{-1} g_n) = g_n$
- $\Delta\Delta^{-1} = I$

Conoscendo il caso continuo, possiamo pensare che l'operatore antidifferenza si comporti in maniera simile all'operatore di integrazione, e per quanto visto in precedenza, questo consista di una sommatoria. Vediamo che, infatti, l'operatore di antidifferenza si comporta veramente come l'operatore di integrale.

Sappiamo, dagli studi di Analisi I, che se:

$$\begin{aligned}\frac{dy}{dx} &= f(x) \\ \Downarrow \\ y &= \int f(x)dx = F(x) + c\end{aligned}$$

In cui F è la *primitiva* di f , mentre c è una costante.
Allo stesso modo, nel caso discreto:

$$\begin{aligned}\Delta y_n &= g_n \\ \Downarrow \\ y_n &= \Delta^{-1}g_n + \omega_n\end{aligned}$$

In cui ω_n è una successione costante di valore n , infatti:

$$\Delta \omega_n = \omega_{n+1} - \omega_n = n - n = 0$$

Pertanto:

$$y_n = \Delta^{-1}g_n + \omega_n$$

È ancora soluzione del problema.

In un contesto di risoluzione di un problema ai valori iniziali, la successione ω_n viene univocamente determinata dall'imposizione delle condizioni iniziali. Concludiamo (usando un abuso di notazione) dicendo che $\Delta \Delta^{-1} = I - \omega_n$.
Conoscere a priori l'antidifferenza di una successione, costituisce l'analogico discreto del caso in cui, nel continuo, è facile calcolare l'integrale di una funzione di cui è nota la primitiva.

Abbiamo dunque visto che l'equazione:

$$\Delta y_n = g_n$$

Ammette come soluzione

$$y_n = \Delta^{-1}g_n + \omega_n$$

In cui la successione costante ω_n è univocamente determinata dall'imposizione delle condizioni iniziali, quindi possiamo porre $\omega_n = y_{n_0}$ per comodità di notazione, e scrivere:

$$y_n = \Delta^{-1}g_n + y_{n_0}$$

L'operatore anti-differenza svolge il ruolo dell'operatore di integrale nel contesto continuo, e abbiamo visto che per farlo esso deve rappresentare per forza una sommatoria, essendo questa l'analogico nel discreto dell'integrazione. Scriveremo che, supposta nota la g_n :

$$y_n = \sum_{i=n_0}^{n-1} g_i + y_{n_0} \quad n \geq n_0$$

Supponiamo ora di voler valutare il seguente polinomio in $x = 2$:

$$\sum_{i=0}^2 b_i x^{n-i} \quad b_i = \{1, 1, 1\}$$

Abbiamo:

$$\sum_{i=0}^2 b_i x^{n-i} = x^3 + x^2 + x = p(x)$$

$$p(2) = 2^3 + 2^2 + 2 = 15$$

Il *metodo di Horner* costituisce un metodo per valutare un polinomio risolvendo un problema ai valori iniziali discreto. Infatti, se poniamo:

$$y_0 = b_0 = 1$$

$$y_i = xy_{i-1} + b_i$$

Otteniamo:

$$y_i = 2y_{i-1} + b_i$$

Poichè, a titolo di esempio, vogliamo valutare il polinomio in $x = 2$

$$y_0 = b_0 = 1$$

$$y_1 = 2y_0 + b_1 = 2 + 1 = 3$$

$$y_2 = 2y_1 + b_2 = 6 + 1 = 7$$

$$y_3 = 2y_2 + b_3 = 14 + 1 = 15$$

Risolvendo ricorsivamente la successione abbiamo ottenuto la valutazione del polinomio senza risolvere esplicitamente la sommatoria che definisce il polinomio. Questo succede poichè la stessa sommatoria è la soluzione esplicita del problema ai valori iniziali mostrato.

2.3 Potenze fattoriali

Le potenze fattoriali, o pseudo-potenze, costituiscono una controparte discreta dell'elevamento a potenza nel caso continuo. Controparte discreta in quanto, vedremo, calcolare differenza e antiderivata di una potenza fattoriale è molto simile a calcolare derivata e integrale di un elevamento a potenza standard.

Definiamo la potenza fattoriale, o pseudo-potenza come:

$$x^{(n)} = x(x-1)(x-2)\dots(x-n+1)$$

Ad esempio:

$$10^{(4)} = 10(10-1)(10-2)(10-3) = 5040$$

Vediamo un breve paragone fra potenze e pseudo-potenze:

Potenze	Pseudo-potenze
$\frac{dy}{dx} x^n = nx^{n-1}$	$\Delta x^{(n)} = nx^{(n-1)}$
$\int x^n dx = \frac{x^{n+1}}{n+1}$	$\Delta^{-1} x^{(n)} = \frac{x^{(n+1)}}{n+1}$
$x^{m+n} = x^m x^n$	$x^{(m+n)} = x^{(m)}(x-m)^{(n)}$
$x^0 = 1$	$x^{(0)} = 1$
$x^{-k} = \frac{1}{x^k}$	$x^{(-k)} = \frac{1}{(x+k)^{(k)}}$

Inoltre, vale:

- $x^{(n)}$ è un polinomio monico di variabile x di grado n , le cui radici sono $0, 1, \dots, n-1$
- $k^{(n)} = 0$ se $n > k$
- $k^{(k)} = k^{(k-1)} = k!$
- $\frac{n^{(k)}}{k^{(k)}} = \binom{n}{k}$

Esiste una relazione algebrica che lega le potenze alle pseudo-potenze:

$$x^n = \sum_{i=1}^n S_i^n x^{(i)}$$

Dove gli S_i^n sono chiamati *numeri di Stirling* di seconda specie. Essi sono definiti in modo ricorsivo:

$$\begin{aligned} S_1^n &= S_n^n = 1 \\ S_i^{n+1} &= S_{i-1}^n + iS_i^n \end{aligned}$$

Sia:

$$T_n = \begin{bmatrix} S_1^1 & 0 & 0 & \dots & 0 \\ S_1^2 & S_2^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ S_1^n & \dots & \dots & \dots & S_n^n \end{bmatrix}$$

La matrice triangolare inferiore a diagonale unitaria contenente i numeri di Stirling di seconda specie, e siano

$$\vec{x}^n = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ \vdots \\ x^n \end{bmatrix}$$

$$\vec{x}^{(n)} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ \vdots \\ x^{(n)} \end{bmatrix}$$

I vettori contenenti i primi n elevamenti a potenza di uno scalare x e le prime n potenze fattoriali dello stesso scalare x . Abbiamo che:

$$\vec{x}^n = T_n \vec{x}^{(n)}$$

Dunque

$$\vec{x}^{(n)} = T_n^{-1} \vec{x}^n$$

Il generico elemento $s_{i,j}$ di T_n^{-1} è chiamato numero di Stirling di prima specie, il quale è utilizzato nell'espressione della relazione algebrica, questa volta, tra pseudo-potenze e potenze. Prima abbiamo visto le potenze espresse in funzione

delle pseudo-potenze, ora vediamo il contrario, ossia le pseudo-potenze espresse in funzione delle potenze:

$$x^{(n)} = \sum_{i=1}^n s_i^n x^i$$

Questa relazione è utile a dimostrare il seguente Teorema:

$$p(x) \in \Pi_k \Rightarrow \Delta^j p(x) \in \Pi_{\max\{0, k-j\}}$$

Ossia, similmente a quanto avviene nel caso continuo, l'operatore differenza applicato j volte a un polinomio di grado al più k , genera un polinomio di grado $k-j$ a meno che non sia $j \geq k$, nel qual caso otterremo il polinomio costante nullo.

2.4 Teoremi del confronto

Non sempre è possibile risolvere esplicitamente un'equazione alle differenze, dunque è utile poter discutere opportune *disequazioni* alle differenze che ci consentono comunque di porre dei vincoli sul comportamento delle traiettorie soluzione di un problema ai valori iniziali. A tale scopo, possiamo avvalerci dei *teoremi del confronto*.

- Sia $g(n, y)$ una funzione non decrescente rispetto al secondo argomento tale per cui:

$$y_{n+1} \leq g(n, y_n) \text{ e } u_{n+1} \geq g(n, u_n)$$

Se $y_{n_0} \leq u_{n_0}$ allora $y_n \leq u_n \forall n \geq 0$

- Sia $g(n, i, y)$ una funzione non decrescente rispetto al terzo argomento. Sia inoltre $\{p_n\}$ una successione nota.

Se

$$y_n \leq p_n + \sum_{i=0}^{n-1} g(n, i, y_i) \quad n \geq n_0$$

e

$$u_n \geq p_n + \sum_{i=0}^{n-1} g(n, i, u_i) \quad n \geq n_0$$

Allora:

$$y_n \leq u_n \forall n \geq n_0$$

- Siano $p_n \geq 0, g_n$ successioni note. Se $\{y_n\}$ è una successione tale per cui $y_{n+1} \leq p_n y_n + g_n$ allora:

$$y_n \leq y_{n_0} \prod_{k=n_0}^{n-1} p_k + \sum_{i=n_0}^{n-1} g_i \prod_{k=i+1}^{n-1} p_k$$

- Siano $p_n \geq 0, g_n$ successioni note. Se $\{y_n\}$ è una successione tale per cui

$$y_n \leq y_{n_0} + \sum_{i=0}^{n-1} [p_i y_i + g_i]$$

Allora:²

$$y_n \leq y_{n_0} \exp \left(\sum_{k=n_0}^{n-1} p_k \right) + \sum_{i=n_0}^{n-1} g_i \exp \left(\sum_{k=i+1}^{n-1} p_k \right)$$

- Siano $p_n \geq 0, g_n$ successioni note. Se

$$y_n \leq g_n + \sum_{i=n_0}^{n-1} p_i y_i$$

Allora, per $n \geq n_0$:

$$y_n \leq g_n + \sum_{i=0}^{n-1} p_i g_i \exp \left(\sum_{k=i+1}^{n-1} p_k \right)$$

Questo teorema è noto come *lemma di Gronwall nel discreto*. Esiste infatti anche una controparte continua di questo teorema, detto *lemma di Gronwall nel continuo*³: se

$$y(t) \leq g(t) + \int_{t_0}^t p(s)y(s)ds$$

Allora:

$$y(t) \leq g(t) + \int_{t_0}^t p(s)g(s) \exp \left(\int_s^t p(x)dx \right) ds$$

²la scrittura $\exp(x)$ sta per $e^{(x)}$

³Esistono le controparti continue anche dei teoremi del confronto mostrati sopra

3 Equazioni alle differenze lineari

Finora abbiamo discusso le equazioni alle differenze da un punto di vista teorico, adesso vogliamo darne una classificazione più precisa e introdurre altresì i metodi per risolverle esplicitamente. Quelle viste finora erano equazioni alle differenze lineari del *primo* ordine. In generale, diremo che un'equazione alle differenze lineare di ordine k è un'equazione della forma:

$$\sum_{i=0}^k p_i(n) y_{n+k-i} = g_n$$

Dove sono noti i $p_i(n)$ e la successione g_n . A titolo di esempio, l'equazione:

$$\sum_{i=0}^2 y_{n+k-i} = 0$$

↓

$$y_{n+2} + y_{n+1} + y_n = 0$$

È un'equazione alle differenze di ordine $k = 2$.

Le soluzioni di un'equazione alle differenze di ordine k sono univocamente determinate una volta imposte le k condizioni iniziali:

$$y_{n_0} = c_0, y_{n_1} = c_1, \dots, y_{n_0+k-1} = c_{k-1}$$

Chiamiamo L l'operatore tale per cui:

$$Ly_n = \sum_{i=0}^k p_i(n) y_{n+k-i}$$

Possiamo scrivere un'equazione alle differenze di ordine k con una notazione più compatta:

$$Ly_n = g_n$$

Associata a un'equazione alle differenze, vi è l'equazione omogenea:

$$Ly_n = 0$$

Sia \mathbb{S} lo spazio delle soluzioni dell'equazione omogenea associata. Abbiamo che \mathbb{S} è uno spazio vettoriale di dimensione k .

Per dimostrare tale asserto occorre dimostrare che:

- \mathbb{S} è uno spazio vettoriale
- la dimensione di \mathbb{S} è k

Dagli studi di Algebra Lineare, sappiamo che per dimostrare che un certo insieme sia uno spazio vettoriale occorre dimostrare che:

- L'insieme è chiuso rispetto alle operazioni di somma e moltiplicazione per uno scalare
- Il vettore nullo fa parte dell'insieme

Che \mathbb{S} sia un insieme chiuso rispetto alle operazioni di somma e moltiplicazione per uno scalare lo si dimostra banalmente in quanto L è un operatore lineare. Per dimostrare che il vettore nullo è contenuto in \mathbb{S} basta osservare che la *succezione nulla* $y_n \equiv 0$ è soluzione dell'equazione omogenea associata, quindi fa parte di \mathbb{S} , infatti:

$$L 0 = \sum_{i=0}^k p_i(n)0 = \sum_{i=0}^k 0 = 0$$

Abbiamo dunque completato la dimostrazione che \mathbb{S} sia uno spazio vettoriale.

□

Rimane da dimostrare che la sua dimensione è k . La dimensione di uno spazio vettoriale è pari al numero di vettori linearmente indipendenti che ne costituiscono la *base*. Occorre quindi dimostrare che possiamo prendere k vettori (soluzioni!) linearmente indipendenti in \mathbb{S} per costituire una base per \mathbb{S} .

Sia \vec{c} il vettore contenente le k condizioni iniziali:

$$\vec{c} = (c_0, \dots, c_{k-1})^T$$

Possiamo scrivere \vec{c} come combinazione lineare dei versori degli assi $\vec{e}_1, \dots, \vec{e}_k$. Chiamiamo $y_n(n_0, \vec{e}_i)$ le soluzioni dell'equazione omogenea tali per cui:

$$\sum_{i=1}^k y_n(n_0, \vec{e}_i) c_{i-1} = y_n$$

Le $y_n(n_0, \vec{e}_i)$ sono in tutto k , sono soluzioni quindi stanno in \mathbb{S} e sono linearmente indipendenti in quanto si dimostra banalmente che:

$$\sum_{i=1}^k \alpha_i y_n(n_0, \vec{e}_i) = 0 \Rightarrow \alpha_i = 0 \quad i = 1, \dots, k$$

Ciò conclude la dimostrazione che la dimensione dello spazio vettoriale \mathbb{S} sia effettivamente k .

□

Siano ora $\{f_n^{(i)}\}$ per $i = 1, \dots, k$ le k successioni che risolvono l'equazione omogenea associata, e sia \bar{y}_n una soluzione particolare dell'equazione non omogenea. Allora, ogni altra soluzione dell'equazione non omogenea si può scrivere come:

$$y_n = \bar{y}_n + \sum_{i=1}^k \alpha_i f_n^{(i)}$$

Per determinare gli α_i introduciamo un particolare strumento: la matrice di *Casorati*.

La matrice di Casorati è una matrice associata a un insieme di soluzioni di un'equazione omogenea. Ossia:

$$C(n) = \begin{bmatrix} f_n^1 & \dots & f_n^k \\ f_{n+1}^1 & \dots & f_{n+1}^k \\ \vdots & & \vdots \\ f_{n+k-1}^1 & \dots & f_{n+k-1}^k \end{bmatrix} \in \mathbb{C}^{k \times k}$$

Dunque una matrice di Casorati è una matrice le cui colonne sono le k soluzioni dell'equazione omogenea associata. Vediamo adesso come la si usa per determinare gli α_i :

1. Calcolare la matrice di Casorati in $n = n_0$:

$$C(n_0) = \begin{bmatrix} f_{n_0}^1 & \cdots & f_{n_0}^k \\ f_{n_0+1}^1 & \cdots & f_{n_0+1}^k \\ \vdots & & \vdots \\ f_{n_0+k-1}^1 & \cdots & f_{n_0+k-1}^k \end{bmatrix}$$

2. Risolvere il sistema lineare $C(n_0)\vec{x} = \vec{c}$ dove il vettore \vec{x} è l'incognita (ossia gli α_i), e il vettore \vec{c} è il vettore delle condizioni iniziali.

3.1 Equazioni alle differenze lineari a coefficienti costanti

Un caso molto più semplice da studiare è il caso in cui i coefficienti $p_i(n)$ siano tutti delle costanti. Ossia, formalmente, il caso in cui:

$$p_i(n) \equiv p_i \quad i = 0, \dots, k$$

Perdiamo dunque la dipendenza funzionale da n dei polinomi $p_i(n)$.

Data l'equazione alle differenze lineare a coefficienti costanti:

$$\sum_{i=0}^k p_i y_{n+k-i} = g_n$$

Vediamo come procedere per risolverla:

1. Definire l'equazione omogenea associata:

$$\sum_{i=0}^k p_i y_{n+k-i} = 0$$

2. Cercare una soluzione nella forma $y_n = z^{n-n_0}$ sostituendo tale valore al posto di y_n nell'equazione omogenea:

$$\sum_{i=0}^k p_i z^{n+k-i+n_0}$$

3. Portando fuori dalla sommatoria la costante z^{n-n_0} si ottiene il polinomio caratteristico:

$$z^{n-n_0} \sum_{i=0}^k p_i z^{k-i}$$

Ossia il polinomio definito da quel che rimane della sommatoria:

$$\sum_{i=0}^k p_i z^{k-i}$$

4. Cerchiamo le radici del polinomio caratteristico. Consideriamo per adesso solo il caso in cui queste siano distinte. Se è così, allora la soluzione dell'equazione omogenea sarà:

$$y_n = \sum_{i=1}^k c_i z_i^n$$

Dove le z_i si determinano ponendo il polinomio caratteristico

$$\sum_{i=0}^k p_i z^{k-i} = 0$$

Mentre le costanti c_i sono determinate univocamente dalla scelta delle condizioni iniziali.

Vediamo un esempio pratico. Si consideri l'equazione:

$$y_{n+2} = y_{n+1} + y_n$$

Scriviamola in una notazione che ci consenta di determinare velocemente il polinomio caratteristico:

$$y_{n+2} - y_{n+1} - y_n = 0$$

Il polinomio caratteristico sarà:

$$p(z) = z^2 - z - 1$$

Cerchiamo le sue radici ponendo

$$p(z) = z^2 - z - 1 = 0$$

$$z_{1,2} = \frac{1 \pm \sqrt{1+4}}{2}$$

La soluzione generale sarà pertanto:

$$y_n = c_1 z_1^n + c_2 z_2^n$$

Calcoliamo la matrice di Casorati in n_0

$$C(n_0) = \begin{bmatrix} z_1^0 & z_2^0 \\ z_1^1 & z_2^1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ z_1 & z_2 \end{bmatrix}$$

Imponiamo le seguenti condizioni iniziali:

$$y_{n_0} = 1$$

$$y_{n_1} = 1$$

Risolviamo

$$\begin{bmatrix} 1 & 1 \\ z_1 & z_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} y_{n_0} = 1 \\ y_{n_1} = 1 \end{bmatrix}$$

Ottenendo:

$$c_1 = \frac{z_1}{\sqrt{5}} \quad c_2 = -\frac{z_2}{\sqrt{5}}$$

L'equazione da cui siamo partiti:

$$y_{n+2} = y_{n+1} + y_n$$

È un'equazione già omogenea, pertanto una volta imposte le condizioni iniziali, abbiamo risolto il problema. Non abbiamo dunque la necessità di cercare una soluzione particolare \bar{y}_n per applicare il teorema:

$$y_n = \bar{y}_n + \sum_{i=1}^k \alpha_i f_n^{(i)}$$

Notiamo che, se imponiamo come condizioni iniziali $y_{n_0} = 1, y_{n_1} = 1$, la successione che otteniamo è ben nota ed è la successione dei numeri di Fibonacci.

Quanto visto finora si applica a equazioni alle differenze lineari di ordine k in cui tutte le radici del polinomio caratteristico sono semplici.

Esaminiamo ora, il caso in cui le radici del polinomio caratteristico non siano tutte semplici. In particolare, assumiamo che le radici distinte siano z_1, \dots, z_ν , aventi rispettivamente molteplicità m_1, \dots, m_ν . Osserviamo che:

$$\sum_{i=1}^\nu m_i = k$$

In quanto, al netto delle molteplicità, le radici di un polinomio di grado k devono essere k .

In questo caso, la soluzione generale (soluzione dell'equazione omogenea associata) è data da:

$$y_n = \sum_{i=1}^\nu z_i^{n-n_0} \sum_{j=0}^{m_i-1} c_{ij} (n - n_0)^j \quad n \geq n_0$$

Che è assai differente dalla soluzione generale dell'equazione omogenea nel caso in cui tutte le radici del polinomio caratteristico siano semplici:

$$y_n = \sum_{i=1}^k c_i z_i^n$$

3.2 Stabilità delle soluzioni

Introduciamo ora il concetto di stabilità di una soluzione.

Sia $\{\bar{y}_n\}$ una soluzione di riferimento dell'equazione non omogenea. Data una qualunque altra soluzione della stessa equazione, sia essa $\{y_n\}$, definiamo la successione dell'errore e_n come:

$$e_n = y_n - \bar{y}_n \quad n \geq 0$$

Diremo che la soluzione di riferimento $\{\bar{y}_n\}$ è:

- **stabile**, se e_n è limitata superiormente da una costante: $\sup_{n \geq 0} |e_n| < \infty$
- **asintoticamente stabile** se è stabile e se $\lim_{n \rightarrow +\infty} |e_n| = 0$

- **instabile** se non è stabile (e dunque non può essere neanche asintomaticamente stabile)

Vediamo come classificare una certa soluzione di riferimento $\{\bar{y}_n\}$. Abbiamo che essa è:

- *asintoticamente stabile* se e solo se $|z_i| < 1$ per ogni $i = 1, \dots, \nu$ dove ν è al solito il numero di radici distinte del polinomio caratteristico.
- *stabile* se e solo se $|z_i| \leq 1$ per ogni $i = 1, \dots, \nu$ e, inoltre:

$$|z_i| = 1 \Rightarrow m_i = 1$$

- *instabile* se esiste almeno una radice il cui valore assoluto è superiore a 1, oppure la radice è uguale a 1, ma la molteplicità di tale radice è maggiore di 1.

Definiamo un *Polinomio di Schur* un polinomio che ha tutte le radici nel cerchio aperto unitario del piano complesso, ossia, nel nostro caso, ciò significa un polinomio il cui valore assoluto di tutte le radici è minore di 1.

Definiamo un *Polinomio di von Neumann* un polinomio che ha tutte le radici nel cerchio *chiuso* (quindi comprendono anche il valore 1), ma tutte le radici che si trovano sulla circonferenza (quindi di valore assoluto 1) hanno molteplicità 1.

Con queste definizioni possiamo dire che una soluzione $\{\bar{y}_n\}$ è:

- *asintoticamente stabile* se il polinomio caratteristico è un polinomio di Schur
- *stabile* se il polinomio caratteristico è un polinomio di von Neumann
- *instabile* altrimenti

Vediamo un esempio.

L'equazione alle differenze:

$$y_{n+2} - 2\alpha y_{n+1} + y_n = 2$$

Ammette come soluzione la successione costante:

$$\bar{y} = (1 - \alpha)^{-1}$$

Infatti, se poniamo:

$$\bar{y} - 2\alpha\bar{y} + \bar{y} = 2$$

Otteniamo⁴:

$$\bar{y} = (1 - \alpha)^{-1}$$

Per ogni $\alpha \neq 1$.

Le radici del polinomio caratteristico dell'equazione omogenea associata sono:

$$z^2 - 2\alpha z + 1 = 0 \Rightarrow z_{1,2} = \alpha \pm \sqrt{\alpha^2 - 1}$$

Notiamo che, se $\alpha \in (-1, 1)$, le radici sono complesse e coniugate di modulo 1, pertanto \bar{y} è stabile. Se invece $\alpha = -1$, le radici sono coincidenti e di modulo 1, pertanto \bar{y} è instabile. Alla stessa conclusione si perviene se $|\alpha| > 1$.

⁴Se la successione è costante, allora deve assumere lo stesso valore in $n, n+1, n+2, \dots$

Problema, determinare le radici di un polinomio non è sempre semplice: in questo caso abbiamo un polinomio di secondo grado, e sappiamo risolvere l'equazione $ax^2 + bx + c = 0$, ma in generale non tutti i polinomi sono di secondo grado, quindi non sempre è possibile ricavare esplicitamente le radici di un polinomio di grado arbitrario.

Esistono dei semplici criteri algebrici (criteri di Schur) che caratterizzano un polinomio di Schur o di von Neumann, e che ci consentono pertanto di classificare un polinomio come di Schur o di von Neumann senza bisogno di calcolare esplicitamente le sue radici.

Dato il polinomio caratteristico $p(z)$ di grado k :

$$p(z) = \sum_{i=0}^k p_i z^{k-i}$$

Definiamo il suo *polinomio aggiunto*:

$$q(z) = \sum_{i=0}^k \bar{p}_i z^i$$

In cui la notazione \bar{p}_i indica il complesso e coniugato di p_i ; e definiamo inoltre il polinomio ridotto⁵:

$$p^{(1)}(z) = \frac{q(0)p(z) - p(0)q(z)}{z} = \sum_{i=0}^{k-1} (\bar{p}_0 p_i - p_k \bar{p}_{k-i}) z^{k-i-1} \in \Pi_{k-1}$$

Valgono i seguenti risultati:

- Primo criterio di Schur: il polinomio $p(z)$ è un polinomio di Schur se e solo se $|p_0| > |p_k|$ e, inoltre, $p^{(1)}(z)$ è un polinomio di Schur
- Secondo criterio di Schur: il polinomio $p(z)$ è un polinomio di von Neumann se e solo se:
 - $|p_0| > |p_k|$ e, inoltre, $p^{(1)}(z)$ è un polinomio di von Neumann, oppure:
 - $p^{(1)}(z) = 0$ e, inoltre, $p'(z)$ è un polinomio di Schur

Facciamo un semplice esempio con l'equazione vista in precedenza:

$$y_{n+2} - 2\alpha y_{n+1} + y_n = 2$$

Che ammette come soluzione costante

$$\bar{y} = (1 - \alpha)^{-1}$$

Il polinomio caratteristico dell'equazione omogenea associata abbiamo visto essere:

$$p(z) = z^2 - 2\alpha z + 1$$

Per ottenere il complesso coniugato dovremmo cambiare di segno la parte immaginaria dei coefficienti di z , ma in questo semplice caso, la parte immaginaria è nulla. Dunque, per ogni i , $\bar{p}_i = p_i$. Pertanto il polinomio aggiunto è:

$$q(z) = \sum_{i=0}^{k=2} p_i z^i = 1 - 2\alpha z + z^2 = p(z)$$

⁵Notiamo che abbiamo abbassato il polinomio di un grado

Mentre il polinomio ridotto è:

$$p^{(1)}(z) = \frac{p(z) - p(z)}{z} = 0$$

Possiamo applicare il secondo criterio di Schur, che ci dice:
 $p(z)$ è un polinomio di von Neumann se e solo se:

$$p^{(1)}(z) = 0 \quad (1)$$

e, inoltre, $p'(z)$ è un polinomio di Schur (2).

La condizione (1) è verificata, mentre per la condizione (2) occorre verificare che:

$$p'(z) = 2z - 2\alpha$$

È un polinomio di Schur. Osserviamo che:

$$2z - 2\alpha = 0 \Rightarrow \alpha = z$$

Quindi se $\alpha \in (0, 1)$ allora $p'(z)$ è un polinomio di Schur, poiché tutte le sue radici (una sola) stanno nel cerchio aperto di raggio unitario. Avendo verificato entrambe le condizioni sappiamo che, grazie al secondo criterio di Schur, $p(z)$ è un polinomio di von Neumann se $\alpha \in (0, 1)$. Pertanto se il polinomio caratteristico è di von Neumann, allora la soluzione sarà stabile, e lo sarà appunto se $\alpha \in (0, 1)$. A questa stessa conclusione eravamo giunti discutendo direttamente le radici di $p(z)$. Concludiamo con un'osservazione. Se \bar{y} è un punto di equilibrio *asintoticamente stabile*, allora per $n \rightarrow \infty$, $y_n \rightarrow \bar{y}$. Vediamo un esempio. Il PIL di una nazione è modellabile attraverso la seguente equazione alle differenze:

$$Y_n - \alpha(1 + \rho)Y_{n-1} + \alpha\rho Y_{n-2} = G$$

In cui:

- Y_n è il PIL al tempo n
- $G > 0$ sono le spese di governo
- $\alpha \in (0, 1)$ esprime la relazione lineare che esiste fra i consumi al tempo n e il PIL al tempo $n - 1$ ($\text{Consumi}(n) = \alpha \times Y_{n-1}$)
- $\rho > 0$ esprime similmente la relazione di proporzionalità lineare tra gli investimenti al tempo n e l'incremento dei consumi, ossia la differenza fra i consumi al tempo n e i consumi al tempo $n - 1$.
 $(\text{Investimenti}(n) = \rho \times (\text{Consumi}(n) - \text{Consumi}(n - 1)))$

L'equazione omogenea associata sarà dunque:

$$Y_n - \alpha(1 + \rho)Y_{n-1} + \alpha\rho Y_{n-2} = 0 \quad (*)$$

Il cui polinomio caratteristico è:

$$p(z) = z^2 - \alpha(1 + \rho)z + \alpha\rho$$

Cerchiamo adesso un punto di equilibrio costante \bar{y} tale per cui:

$$\bar{y} - \alpha(1 + \rho)\bar{y} + \alpha\rho\bar{y} = G$$

Ossia, abbiamo sostituito tutti gli Y_i nella (*) con \bar{y} . Raccogliendo \bar{y} si ottiene:

$$\bar{y}(1 - \alpha(1 + \rho) + \alpha\rho) = G$$

Dunque

$$\begin{aligned}\bar{y} &= \frac{G}{1 - \alpha(1 + \rho) + \alpha\rho} \\ \bar{y} &= \frac{G}{1 - \alpha - \alpha\rho + \alpha\rho} = \frac{G}{1 - \alpha}\end{aligned}$$

Affinchè questo punto di equilibrio sia asintoticamente stabile, vogliamo che il polinomio:

$$p(z) = z^2 - \alpha(1 + \rho)z + \alpha\rho$$

Sia un polinomio di Schur. Possiamo applicare il primo criterio di Schur, dunque vogliamo che:

- $|p_0| > |p_k| \Rightarrow 1 > \alpha\rho \Rightarrow \alpha\rho < 1$
- $p^{(1)}(z)$ è un polinomio di Schur

Calcoliamo dunque $p^{(1)}(z)$:

$$p^{(1)}(z) = \dots = (1 + \alpha\rho)z - \alpha(1 + \rho)$$

Abbiamo che $p^{(1)}(z)$ è un polinomio di Schur se e solo se tutte le sue radici stanno nel cerchio aperto di raggio unitario, ossia tutte le radici devono essere in modulo minori di 1. Cerchiamo tali radici:

$$\begin{aligned}p^{(1)}(z) = 0 &\Leftrightarrow (1 + \alpha\rho)z - \alpha(1 + \rho) = 0 \\ z &= \frac{\alpha(1 + \rho)}{1 + \alpha\rho}\end{aligned}$$

Noi vogliamo che $|z| < 1$, dunque:

$$\left| \frac{\alpha(1 + \rho)}{1 + \alpha\rho} \right| < 1 \Rightarrow \rho < \alpha^{-1}$$

Quindi, in sintesi, se $\alpha\rho < 1$ e $\rho < \alpha^{-1}$, allora la soluzione $\bar{y} = \frac{G}{1 - \alpha}$ è asintoticamente stabile, e pertanto $y_n \rightarrow \bar{y}$ se $n \rightarrow \infty$. Vediamo un esempio:

$$\rho = \frac{1}{2} \quad \alpha = \frac{1}{3} \quad G = \frac{2}{3}$$

Con questi parametri,abbiamo che:

$$\alpha\rho = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6} < 1$$

$$\alpha^{-1} = \left(\frac{1}{3}\right)^{-1} = 3 \quad \rho = \frac{1}{2} < \alpha^{-1} = 3$$

Quindi la soluzione

$$\bar{y} = \frac{G}{1 - \alpha} = \frac{\frac{2}{3}}{\frac{1}{2}} = 1$$

Dovrà essere asintoticamente stabile, ossia la successione Y_n dovrà tendere a 1 se n tende a ∞ .

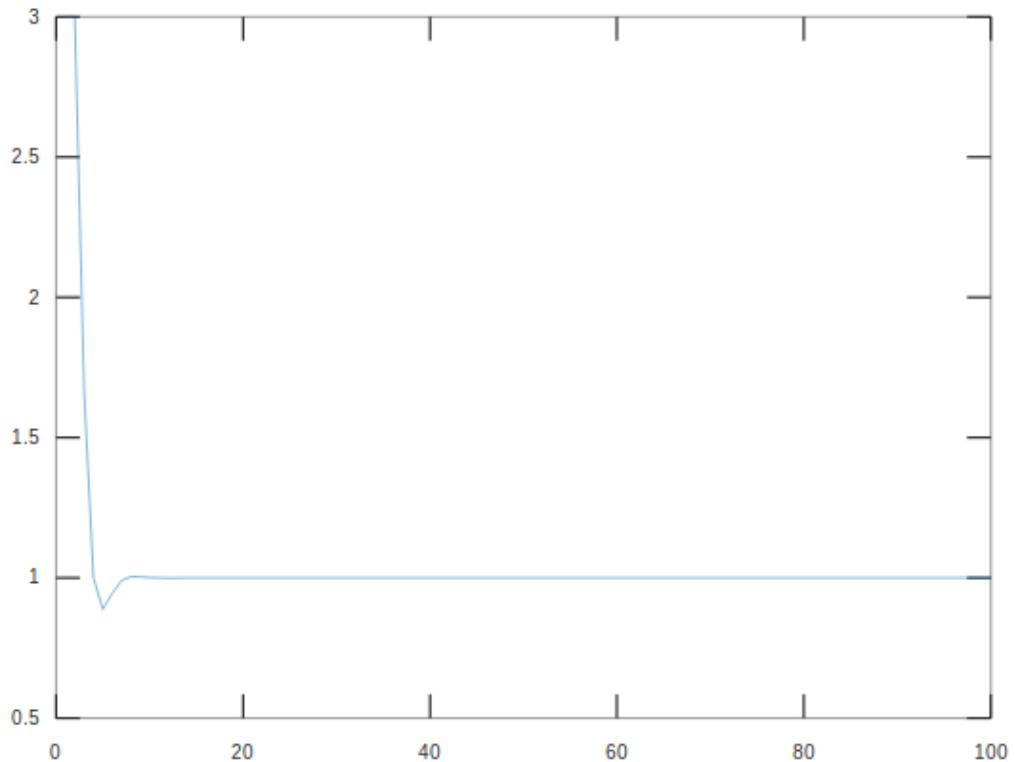
Vediamo un esempio meramente numerico, ponendo come condizioni iniziali $Y_0 = Y_1 = 3$, ricordandoci che

$$Y_n = G + \alpha(1 + \rho)Y_{n-1} - \alpha\rho Y_n - 2$$

E, inoltre:

$$\alpha = \frac{1}{3}, \rho = \frac{1}{2}, G = \frac{2}{3}$$

```
octave:15> for i =3:100
> yn(i) = 2/3 + 1/3*(1+1/2)*yn(i-1) - (1/6)*yn(i-2);
> end
octave:16> plot(yn)
```



4 Metodi lineari multistep

I metodi lineari multistep sono dei metodi per la risoluzione numerica approssimata di problemi ai valori iniziali per equazioni differenziali ordinarie. Sono metodi numerici dunque che *inducono* un problema ai valori iniziali discreto a partire da un problema ai valori iniziali continuo.

Possiamo formalizzare un problema ai valori iniziali continuo come:

$$y'(t) = f(t, y(t)) \quad t \in [t_0, T]$$

$$y(t_0) = y_0$$

Risolvere numericamente un problema del genere significa:

- Definire un dominio discreto $\{t_n\}$
- Discretizzare il problema continuo, definendo un problema discreto su tale dominio
- Risolvere il problema discreto

Per quanto riguarda il primo punto, definiamo un dominio discreto assai semplice:

$$t_n = t_0 + nh$$

$$n = 0, \dots, N$$

$$h = \frac{T - t_0}{N}$$

In questa maniera, i $\{t_n\}$ saranno un dominio discreto contenuto nel dominio continuo iniziale, con cui condivide gli estremi inferiori e superiori. Notiamo tuttavia che la scelta di N è totalmente arbitraria, in generale, più N è grande, più è piccolo h , ossia quello che chiameremo passo di discretizzazione (o integrazione).

Per quanto riguarda il secondo punto, vogliamo discretizzare il problema continuo, mediante un'equazione alle differenze lineare di ordine k .

Sia $f_n \equiv f(t_n, y_n)$ e sia $y_n \equiv y(t_n)$.

Vogliamo discretizzare il problema continuo mediante un'equazione alle differenze di ordine k della forma:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f_{n+i} \quad n = 0, \dots, N - k$$

In cui i coefficienti α, β definiscono univocamente il particolare metodo lineare multistep, o linear multistep formula (LMF), a k passi che stiamo considerando. Come visto in precedenza, possiamo risolvere un'equazione alle differenze di ordine k solamente se sono note k condizioni iniziali, pertanto anche in questo contesto si suppongono note k condizioni iniziali:

$$y_0, \dots, y_{k-1}$$

Notiamo tuttavia che nel caso avessimo tutti gli $\alpha_i = 0$ otterremmo:

$$\begin{aligned} \sum_{i=0}^k \alpha_i y_{n+i} &= h \sum_{i=0}^k \beta_i f_{n+i} \quad n = 0, \dots, N-k \\ &\Downarrow \\ 0 &= h \sum_{i=0}^k \beta_i f_{n+i} \quad n = 0, \dots, N-k \end{aligned}$$

Il che rende evidentemente impossibile la soluzione al problema: non possiamo determinare alcun y_{n+i} .

Per ovviare a questo problema imponiamo una *normalizzazione* di un coefficiente. In generale, potremmo imporre che uno qualsiasi dei coefficienti α_i sia non-nullo, ma nel nostro caso, per convenzione, imponiamo la normalizzazione:

$$\alpha_k = 1$$

Ossia, l'ultimo α lo vogliamo pari a 1. Quindi, partendo da:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f_{n+i} \quad n = 0, \dots, N-k$$

Portiamo fuori dalla sommatoria $\alpha_k y_{n+k}$:

$$\alpha_k y_{n+k} + \sum_{i=0}^{k-1} \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f_{n+i} \quad n = 0, \dots, N-k$$

Ma, avendo imposto $\alpha_k = 1$, otteniamo:

$$y_{n+k} + \sum_{i=0}^{k-1} \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f_{n+i} \quad n = 0, \dots, N-k$$

Portiamo fuori dalla sommatoria di destra anche $\beta_k f_{n+k}$:

$$y_{n+k} + \sum_{i=0}^{k-1} \alpha_i y_{n+i} = h \beta_k f_{n+k} + h \sum_{i=0}^{k-1} \beta_i f_{n+i} \quad n = 0, \dots, N-k$$

Quindi, portando a sinistra tutti i termini fuori dalle sommatorie, e a destra tutti i termini dentro le sommatorie, possiamo esplicitare y_{n+k} :

$$y_{n+k} - h \beta_k f_{n+k} = - \sum_{i=0}^{k-1} \alpha_i y_{n+i} + h \sum_{i=0}^{k-1} \beta_i f_{n+i}$$

Quindi, note le k condizioni iniziali, determineremo le successive y_k, y_{k+1}, \dots, y_N dall'equazione cui sopra. In particolare, se imponiamo $\beta_k = 0$, otteniamo:

$$y_{n+k} = - \sum_{i=0}^{k-1} \alpha_i y_{n+i} + h \sum_{i=0}^{k-1} \beta_i f_{n+i}$$

In questo modo, abbiamo l'espressione esplicita della $y_{n+k} = y(n+k)$. Abbiamo quindi un'espressione della funzione incognita y valutata nell'ascissa $n+k$. Si

parla, in questo caso, di *metodo esplicito*. Qualora non avessimo $\beta_k = 0$, per ottenere y_{n+k} dovremmo conoscere in qualche maniera $f_{n+k} = f(t_{n+k}, y_{n+k})$ ad ogni passo, ma non conosciamo ancora y_{n+k} . Si parla in questo caso di *metodo implicito*.

È evidente che l'implementazione di un metodo esplicito è assai più semplice di quella di un metodo implicito.

Concludiamo definendo i seguenti polinomi:

$$\rho(z) = \sum_{i=0}^k \alpha_i z^i \quad \sigma(z) = \sum_{i=0}^k \beta_i z^i$$

Denominati, rispettivamente, primo e secondo polinomio caratteristico associati a un particolare metodo LMF (identificato univocamente, come già detto, dagli α_i, β_i). Un particolare metodo LMF lo troviamo chiamato spesso con la dicitura metodo (ρ, σ) in quanto gli α_i, β_i identificano univocamente sia il metodo LMF che i suoi polinomi caratteristici.

4.1 Ordine di un metodo LMF

In generale, la soluzione del problema discreto differirà dalla soluzione al problema continuo, ma noi vogliamo che comunque esse siano vicine nel seguente senso: sia $y(n+k)$ la soluzione esatta, valutata sui nodi della mesh. La differenza tra la soluzione esatta $y(n+k)$ e la soluzione approssimata y_{n+k} deve essere un infinitesimo di opportuno ordine del passo h :

$$y(n+k) - y_{n+k} = \tau_{n+k} = O(h^{p+1})$$

In particolare, un metodo si dirà consistente, se $p \geq 1$. Vale il seguente risultato:
Il metodo LMF ha ordine p se:

$$\sum_{i=0}^k \alpha_i = 0$$

E inoltre:

$$\sum_{i=0}^k (i^j \alpha_i - j i^{j-1} \beta_i) = 0 \quad j = 1, \dots, p$$

In poche parole, significa che per avere ordine p un metodo (ρ, σ) deve avere la somma degli α_i nulla, e deve valere $\sum_{i=0}^k (i^j \alpha_i - j i^{j-1} \beta_i) = 0$ per ogni j che va da 1 a p . Possiamo pensare di testare la sommatoria ponendo $j = 1$, poi $j = 2$ e così via finché la sommatoria rimane nulla. Appena troviamo un j che rende la sommatoria non più nulla, significa che l'ordine del metodo è quel j , meno 1. Come semplice corollario, si ottiene che un metodo è consistente se ha ordine almeno $p = 1$, ovvero se:

$$\sum_{i=0}^k \alpha_i = 0$$

E inoltre:

$$\sum_{i=0}^k (i \alpha_i - \beta_i) = 0$$

Un modo veloce per verificare la consistenza di un metodo è quella di valutare se:

$$\rho(1) = 0, \quad \rho'(1) = \sigma(1)$$

Vale, inoltre, il seguente risultato:

L'ordine massimo di un LMF a k passi è $2k$ se il metodo è implicito, o $2k - 1$ se il metodo è esplicito.

4.2 Convergenza

Studiare la convergenza di un LMF significa verificare che la soluzione numerica converga, quindi tenda, alla soluzione esatta. Un LMF si dirà convergente se, definito l'errore:

$$e_n = y(t_n) - y_n$$

Si ha che, all'infinito, la norma massima di e_n è nulla:

$$\lim_{n \rightarrow +\infty} \max_{n=0,\dots,N} \|e_n\| = 0$$

Notiamo adesso che, per avere un LMF consistente (di ordine almeno 1), abbiamo dovuto imporre:

$$\rho(1) = 0, \quad \rho'(1) = \sigma(1)$$

Si dimostra che tale imposizione rende impossibile il fatto che il polinomio $\rho(z)$ sia un polinomio di Schur. Tuttavia questo potrebbe essere un polinomio di von Neumann. Infatti, se $\rho(z)$ è un polinomio di von Neumann, allora definiamo il corrispondente LMF come un metodo *0-stabile*. La definizione di *0-stabilità* è importante perché ci porta all'enunciato del seguente teorema:

Se un metodo di ordine p è *0-stabile*, e le condizioni iniziali hanno accuratezza (ossia si discostano dalle condizioni iniziali esatte) $O(h^p)$, allora l'errore globale e_n sarà a sua volta $O(h^p)$.

Come semplice corollario, si ottiene che un LMF che sia consistente e *0-stabile*, è convergente.

La proprietà di *0-stabilità* di un metodo LMF è importante, ma introduce dei vincoli sui coefficienti α_i che non potranno essere utilizzati per incrementare l'ordine del metodo. Questo argomento si concretizza nell'enunciato della prima barriera di Dahlquist:

L'ordine massimo di un metodo LMF a k passi, che sia *0-stabile*, è:

- $k + 1$ se k è dispari
- $k + 2$ se k è pari

4.3 Esempi di metodi LMF convergenti

Abbiamo detto che se un metodo (ρ, σ) è tale per cui il polinomio $\rho(z)$ è un polinomio di von Neumann, allora il metodo (ρ, σ) è detto 0-stabile. Se, inoltre il metodo è anche di ordine $p \geq 1$, dunque anche consistente, allora il metodo è convergente.

Vediamo alcuni esempio di metodi LMF convergenti.

4.3.1 I metodi di Adams

Per questi metodi, il polinomio $\rho(z)$ è della forma:

$$\rho(z) = z^{k-1}(z - 1)$$

I metodi di Adams si dividono in esplicativi ed impliciti. Vediamoli.

- Esplicativi, ordine $p = k$:

- $k = 1 \rightarrow$ Metodo di Eulero esplicito

$$y_{n+1} - y_n = hf_n$$

- $k = 2 \rightarrow$ Metodo di Adams Bashforth

$$y_{n+2} - y_{n+1} = \frac{h}{2}(3f_{n+1} - f_n)$$

- Impliciti, ordine $p = k + 1$:

- $k = 1 \rightarrow$ Metodo dei trapezi

$$y_{n+1} - y_n = \frac{h}{2}(f_{n+1} + f_n)$$

- $k = 2 \rightarrow$ Metodo di Adams Moulton:

$$y_{n+2} - y_{n+1} = \frac{h}{12}(5f_{n+2} + 8f_{n+1} - f_n)$$

4.3.2 Formule di Newton-Cotes

Sono formule della forma:

$$y_{n+k} - y_n = h \sum_{i=0}^k \beta_i f_{n+i}$$

Il polinomio $\rho(z) = z^k - 1$ è un polinomio di von Neumann e quindi se il metodo ha almeno ordine 1 sarà convergente.

4.3.3 Backward Differentiation Formulae

Anche note come BDF, sono metodi della forma:

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \beta_k f_{n+k}$$

La cosa interessante è che determiniamo i coefficienti $\{\alpha_i\}$ imponendo l'ordine massimo $p = k$. Questi metodi sono impliciti e *0-stabili* fino a $k = 6$

4.4 Stabilità per h fissato

Nonostante il risultato negativo dato dalla prima barriera di Dahlquist, abbiamo visto che esistono metodi *0-stabili* con un ordine di accuratezza arbitrariamente elevato, e pertanto, convergenti.

Nell'analisi su cui si basa tale proprietà, abbiamo assunto di poter far tendere $h \rightarrow 0$, aumentando il numero N dei punti del dominio discreto. Questo però non si può sempre fare, poiché per alcuni problemi $T \rightarrow \infty$ e pertanto non è possibile assumere il passo h infinitesimo, sia per ragioni di efficienza computazionale sia per ragioni legate all'accumulo degli errori in aritmetica finita. Per quanto argomentato, ci troviamo quindi a dover discutere l'*equazione dell'errore* per h fissato. Questo è, in generale, un compito assai arduo, trattandosi di una equazione alle differenze di ordine k che, in generale, non è lineare.

Questo compito però si semplifica notevolmente nel caso in cui il problema continuo sia espresso da una particolare $f(t, y)$: la cosiddetta *equazione test*.

Sia $\lambda \in \mathbb{C}$ uno scalare arbitrario tale per cui la parte reale di λ è negativa. Imponiamo:

$$\begin{aligned} y'(t, y) &= \lambda y \\ y(0) &= y_0 \end{aligned}$$

Risolvendo analiticamente otteniamo:

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

Notiamo che $\lim_{t \rightarrow \infty} y(t) = 0$.

Pertanto, l'origine è un punto di equilibrio asintoticamente stabile. Che è un punto di equilibrio lo si verifica facilmente in quanto $y'(t, 0) = \lambda 0 = 0 \forall t \in \mathbb{R}$, in accordo con quanto abbiamo visto nel primo capitolo.

Intuitivamente, se l'origine fosse ancora un punto di equilibrio asintoticamente stabile della soluzione discreta indetta da un certo metodo (ρ, σ) , ovvero se:

$$y_n \rightarrow 0, \quad n \rightarrow \infty$$

Allora l'errore sarebbe tale per cui:

$$e_n \rightarrow 0, \quad n \rightarrow \infty$$

In tal caso, si dirà che il metodo (ρ, σ) è *assolutamente stabile* in $q = h\lambda$.

Vale il seguente teorema:

Un metodo (ρ, σ) ammette l'origine come punto di equilibrio asintoticamente stabile se e solo se il polinomio di stabilità del metodo:

$$\pi(z, q) = \rho(z) - q\sigma(z)$$

È un polinomio di Schur.

Definiamo la regione del piano complesso

$$\mathbb{D} = \{q \in \mathbb{C} : \pi(z, q) \text{ sia un polinomio di Schur}\}$$

Tale regione è denominata *regione di assoluta stabilità* del metodo (ρ, σ) e, in generale, più essa è ampia, più ho libertà nello scegliere h grande.

4.4.1 A-Stabilità e Boundary Locus

Abbiamo detto che un metodo (ρ, σ) è *0-stabile* se il polinomio $\rho(z)$ è un polinomio di von Neumann.

Una ulteriore classificazione sulla stabilità dei metodi (ρ, σ) è data dalla *A-Stabilità*. Vediamola in dettaglio.

Si dice che un metodo (ρ, σ) sia A-Stabile se $\mathbb{C}^- \subseteq \mathbb{D}$, ossia se il semipiano complesso negativo è contenuto nella regione di assoluta stabilità di un metodo (ρ, σ) . Nel caso in cui i due insiemi coincidano, allora il metodo si dirà perfettamente (o precisamente) A-Stabile.

La soluzione discreta indotta da un metodo perfettamente A-Stabile ha sempre lo stesso comportamento qualitativo di quella continua.

Purtroppo, vale il seguente risultato negativo (seconda barriera di Dahlquist): Non esistono metodi LMF A-stabili espliciti. Inoltre, l'ordine massimo di un LMF A-Stabile è 2. Definiamo l'insieme:

$$\Gamma = \left\{ q(\theta) = \frac{p(e^{i\theta})}{\sigma(e^{i\theta})} : 0 \leq \theta \leq 2\pi \right\}$$

Come il *boundary locus* del metodo (ρ, σ) , ossia il luogo dei punti del piano complesso tali per cui il polinomio di stabilità $\pi(z, q)$ ha almeno una radice di modulo 1, segue che il boundary locus è il confine geometrico tra la regione di assoluta stabilità di un metodo (ρ, σ) e il resto del piano complesso. Osserviamo che studiare il boundary locus è più semplice rispetto a utilizzare direttamente i criteri di Schur sul polinomio $\pi(z, q)$ per determinare quali q rendono il polinomio un polinomio di Schur, e dunque quali q formano la regione di assoluta stabilità di un metodo (ρ, σ) .

Definiamo il *tipo* di un polinomio $p(z) \in \Pi_k$ come la terna (k_1, k_2, k_3) con $k_1, k_2, k_3 \geq 0$ e $k_1 + k_2 + k_3 = k$ tale per cui:

- k_1 è il numero di radici di $p(z)$ di modulo minore di 1 (contate con la loro molteplicità)
- k_2 è il numero di radici di $p(z)$ di modulo uguale a 1 (contate con la loro molteplicità)
- k_3 è il numero di radici di $p(z)$ di modulo maggiore di 1 (contate con la loro molteplicità)

Pertanto:

- I polinomi di tipo $(k, 0, 0)$ sono i polinomi di Schur
- I polinomi di tipo $(k_1, k_2, 0)$ sono i polinomi di von Neumann (se le k_2 radici di modulo 1 sono semplici)

Da quanto su esposto, il boundary locus di un metodo LMF suddivide il piano complesso in un numero di regioni connesse in cui, per ogni q in ciascuna di esse, il tipo del polinomio $\pi(z, q)$ è costante.

L'unione delle regioni in cui il polinomio $\pi(z, q)$ ha tipo $(k, 0, 0)$ è la *regione di assoluta stabilità* del metodo (ρ, σ) .

Andiamo ad esaminare alcune importanti proprietà del boundary locus di metodi consistenti e 0-stabili. Infatti, per un metodo consistente e 0-stabile, deve avversi

$$\rho(1) = 0, \quad \rho'(1) = \sigma(1) \neq 0$$

E, quindi:

$$q(0) = \frac{\rho(1)}{\sigma(1)} = 0 \in \Gamma$$

Pertanto, l'origine del piano complesso appartiene al boundary locus del metodo. Inoltre, si dimostra anche che, se il metodo è consistente e 0-stabile allora l'asse immaginario è tangente al boundary locus.

Infine osserviamo che il boundary locus è una curva simmetrica rispetto all'asse reale.

4.5 L-Stabilità

Concludiamo la trattazione parlando di L-Stabilità.

Se applichiamo l'equazione di test $y'(t, y) = \lambda y$ a un certo metodo LMF, la successione approssimata y_n dipende ovviamente da h e da λ .

Poniamo $q = h\lambda$.

Se, al tendere di q all'infinito, $y_n(q)$ tende a 0, e il metodo LMF a cui applichiamo l'equazione di test è un metodo A-Stabile, allora il particolare metodo LMF si dice essere L-Stabile.

5 Funzioni di Matrici

Fino ad ora abbiamo considerato problemi ai valori iniziali modellati da un'unica funzione incognita. Prima di poter discutere di *sistemi lineari* di equazioni alle differenze o differenziali, occorre introdurre un nuovo strumento metodologico: le funzioni di matrici.

Nel seguito, tenteremo di dare un senso alla scrittura $f(A)$ con $f : \mathbb{C} \rightarrow \mathbb{C}$ e A matrice quadrata di dimensione \bar{m} .

Cominciamo con il caso più semplice, ossia il caso in cui la funzione $f(z)$ sia un polinomio, ovvero $f(z) \equiv p(z) \in \Pi_k$

5.1 Funzioni di Matrici polinomiali

In questo caso abbiamo:

$$f(z) = p(z) = \sum_{i=0}^k p_i z^i$$

In tal caso si ha che, se $A \in \mathbb{C}^{\bar{m} \times \bar{m}}$:

$$f(A) = p(A) = \sum_{i=0}^k p_i A^i \in \mathbb{C}^{\bar{m} \times \bar{m}}$$

Per esempio, siano:

$$p(z) = 3z + 1 \quad A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Allora

$$p(A) = \sum_{i=0}^1 p_i A^i = 1 \times A^0 + 3 \times A^1 = 1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

Inoltre, se $\lambda \in \sigma(A)$ è un autovalore relativo all'autovettore \vec{v} di A (quindi, si trova nello *spettro* di A), ovvero⁶:

$$A\vec{v} = \lambda\vec{v}$$

Allora si dimostra il seguente risultato:

$$A\vec{v} = \lambda\vec{v} \Rightarrow p(A)\vec{v} = p(\lambda)\vec{v}$$

Ossia, calcolare il prodotto fra un polinomio valutato su una matrice e un autovettore della matrice, è equivalente a calcolare il prodotto fra lo stesso polinomio, valutato sul relativo autovalore, e l'autovettore stesso.

Tuttavia, vi sono alcune differenze fondamentali tra i polinomi di una variabile scalare, e i polinomi di una matrice. Infatti, se $z \in \mathbb{C}$, allora

$$z \neq 0 \Leftrightarrow z^n \neq 0 \quad n \geq 0$$

⁶semplice definizione di autovalore relativo a un autovettore

Ossia, posso elevare a potenza uno scalare nonnullo quante volte voglio senza ottenere mai un valore nullo. Questo potrebbe non essere vero per una matrice, ad esempio:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \rightarrow A^2 = O$$

Dove O è la matrice nulla di dimensione appropriata.

Più in generale, vale il seguente risultato (noto come teorema di Cayley-Hamilton):

Sia $p(\lambda) = \det(A - \lambda I)$ il polinomio caratteristico di A . Si ha che $p(A) = O$.

Questo significa che se il polinomio su cui vogliamo valutare la matrice è proprio il suo polinomio caratteristico, allora la valutazione di tale polinomio sulla stessa matrice restituisce la matrice nulla. Vediamo un esempio:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$

$$A - I\lambda = \begin{bmatrix} 1 - \lambda & 1 \\ 0 & 2 - \lambda \end{bmatrix}$$

Pertanto:

$$\det(A - I\lambda) = (1 - \lambda)(2 - \lambda) = \lambda^2 - 3\lambda + 2 = p(\lambda)$$

Dunque:

$$p(A) = 2A^0 - 3A + A^2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 3 & 3 \\ 0 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = O$$

Supponiamo ora che il polinomio caratteristico di A sia:

$$p(\lambda) = \prod_{i=1}^{\nu} (\lambda - \lambda_i)^{\overline{m}_i}, \quad \lambda_i \neq \lambda_j, \quad \text{se } i \neq j$$

Ovvero, i λ_i sono gli autovalori distinti di A aventi molteplicità algebrica, rispettivamente, $\overline{m}_1, \dots, \overline{m}_{\nu}$. Dal teorema di Cayley-Hamilton, segue che esiste un polinomio monico⁷ di grado \overline{m} (la dimensione della matrice) che calcolato in A , vale la matrice nulla (e questo è esattamente il polinomio caratteristico, che come abbiamo visto ha grado pari alla dimensione della matrice A). Potrebbe tuttavia esistere un polinomio di grado minore o uguale alla dimensione della matrice (\overline{m}) che, tuttavia, si annulla quando calcolato in A . Definiamo tale polinomio come il *polinomio minimo* (o minimale) di A , e sia esso $\psi(z)$. Segue che:

$$m \equiv \deg(\psi) \leq \overline{m} = \deg(\det(A - I\lambda))$$

Sia ora $p(z)$ il polinomio caratteristico di A , ossia $\det(A - Iz)$. Vale il seguente risultato:

Ogni radice di $p(z)$ è radice del polinomio minimo $\psi(z)$ e viceversa. Infatti, poiché $m \leq \overline{m}$ è possibile dividere $p(z)$ per $\psi(z)$, ottenendo:

$$p(z) = q(z)\psi(z) + r(z)$$

In cui il grado del polinomio *resto* $r(z)$ è minore del grado del polinomio minima. Dal momento che, per ipotesi $\psi(A) = O$, per il teorema di Cayley-Hamilton $p(A) = O$, otteniamo:

$$O = q(z)O + r(z) \Rightarrow O = r(z)$$

⁷monico significa che ha il coefficiente di grado massimo pari a 1

Quindi $r(z) \equiv 0$, e pertanto, ogni radice di $\psi(z)$ è radice di $p(z)$. Il viceversa è vero, infatti, se

$$\lambda \in \sigma(A)$$

È un autovalore di A relativo all'autovettore \vec{v} , allora dovremmo per forza avere:

$$p(\lambda) = 0, \quad A^j \vec{v} = \lambda^j \vec{v} \quad j \geq 0$$

Per definizione di autovalore e autovettore.

Ne consegue che $\psi(\lambda)\vec{v} = \psi(A)\vec{v} = \vec{0}$ e, quindi, $\psi(\lambda) = 0 \quad \square$.

Sia $p(z)$ il polinomio caratteristico di A, e sia:

$$\psi(z) = \prod_{i=1}^{\nu} (z - \lambda_i)^{m_i}$$

Il suo polinomio minimale. Allora, per ogni i , vale:

$$1 \leq m_i \leq \overline{m_i}$$

Ossia, la molteplicità delle radici del polinomio minimale è minore o uguale alla molteplicità delle radici del polinomio caratteristico.

Osserviamo che, nel caso in cui gli autovalori distinti abbiano tutti molteplicità 1, allora il polinomio minimo e il polinomio caratteristico coincideranno. Quanto enunciato porta alla formulazione del seguente teorema:

Sia $h(z)$ un polinomio di grado maggiore del grado del polinomio minimale di una data matrice A. Se dividiamo $h(z)$ per il polinomio minimale, si ottiene:

$$h(z) = q(z)\psi(z) + r(z)$$

In cui il grado di $r(z)$ è minore del grado del polinomio minimale.

In tal caso, si ha $h(A) = r(A)$.

Ciò significa che preso un qualunque polinomio $h(z)$, per il quale io voglio calcolare $h(A)$, non devo necessariamente applicare la definizione

$$h(A) = \sum_{i=0}^k h_i A^i$$

In cui k è il grado di $h(z)$. Infatti è sufficiente dividere $h(z)$ per l'opportuno $\psi(z)$ e valutare il polinomio residuo $r(z)$ in A. Osserviamo pertanto che due polinomi diversi $h(z)$ e $r(z)$ portano ad ottenere la stessa matrice quando valutati in A. In generale, vale il seguente teorema:

Due polinomi $h(z)$ e $g(z)$ soddisfano la condizione $h(A) = g(A)$ se e solo se essi assumono gli stessi valori sullo spettro di A.

5.2 Funzioni definite sullo spettro di una Matrice

L'ultimo risultato esposto porta a dare la definizione di *funzione definita sullo spettro di A*.

Sia A una matrice, e $\sigma(A)$ il suo spettro. Abbiamo che il generico autovalore

$$\lambda_i \in \sigma(A)$$

Ha molteplicità m_i nel polinomio minimale e molteplicità \bar{m}_i nel polinomio caratteristico. Diciamo che $f(z)$ è definita sullo spettro di A se f è olomorfa in un dominio contenente lo spettro di A . Questo significa che f , per essere definita sullo spettro di A , deve ammettere la derivata fino all'ordine $m_i - 1$ quando valutata sull'autovalore λ_i . Più formalmente, se ν è il numero di autovalori distinti λ_i aventi rispettivamente molteplicità m_i , affinché f sia definita sullo spettro di A , deve essere sempre definito:

$$f^{(j)}(\lambda_i) \quad i = 1, \dots, \nu \quad j = 0, \dots, m_i - 1$$

Fino a questo momento, abbiamo definito solamente cosa signifchi la scrittura $f(A)$ quando f è un polinomio. Più in generale, per funzioni non obbligatoriamente polinomiali, vale il seguente teorema:

Sia $f(z)$ una funzione definita sullo spettro di A , e sia $g(z)$ il polinomio che assume gli stessi valori di $f(z)$ sullo spettro di A . Intuitivamente, possiamo pensare che $g(z)$ sia il *polinomio interpolante* la funzione f valutata sugli autovalori di A . Abbiamo che:

$$f(A) \equiv g(A)$$

Per essere sicuri che il polinomio $g(z)$ esista sempre, e sia unico, lo definiamo come il Polinomio Interpolante di Hermite generalizzato. Esso è definito come segue:

$$g(\lambda) = \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} f^{(j-1)}(\lambda_i) \Phi_{ij}(\lambda)$$

Al solito, ν è il numero di autovalori distinti ed m_i la molteplicità algebrica dell'autovalore λ_i , mentre abbiamo che $\Phi_{ij}(\lambda) = Z_{ij}$ in cui le Z_{ij} sono dette *matrici componenti* di A , sono linearmente indipendenti, e non dipendono in alcun modo da $f(\lambda)$. Vediamo un esempio:

$$f(x) = e^x \quad A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Abbiamo che:

$$A - I\lambda = \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix} \Rightarrow \det(A - I\lambda) = (2 - \lambda)(2 - \lambda) - 1 = \lambda^2 - 4\lambda + 3$$

↓

$$\lambda_1 = 1 \quad \lambda_2 = 3$$

Calcoliamo:

$$\begin{aligned} \Phi_{11}(\lambda) &= \frac{\lambda - \lambda_2}{\lambda_1 - \lambda_2} = \frac{\lambda - 3}{1 - 3} = \frac{1}{2}(3 - \lambda) \\ \Phi_{22}(\lambda) &= \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1} = \frac{\lambda - 1}{3 - 1} = \frac{1}{2}(\lambda - 1) \end{aligned}$$

Le matrici componenti sono date dalla valutazione di A sui polinomi $\Phi(\lambda)$:

$$Z_{11} = \frac{1}{2}(3A^0 - A^1) = \frac{1}{2}(3I - A) = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$Z_{21} = \frac{1}{2}(A^1 - A^0) = \frac{1}{2}(A - I) = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Pertanto:

$$f(A) = g(A) = \sum_{i=1}^2 \sum_{j=1}^1 f^{(j-1)}(\lambda_i) Z_{ij} = \\ f(\lambda_1)Z_{11} + f(\lambda_2)Z_{21} = f(1)Z_{11} + f(3)Z_{22} = \\ \frac{1}{2}f(1)\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{1}{2}f(3)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} =$$

Ma abbiamo:

$$f(x) = e^x$$

Dunque:

$$f(A) = e^A = \frac{e}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{e^3}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \approx \begin{bmatrix} 11.402 & 8.684 \\ 8.684 & 11.402 \end{bmatrix}$$

5.3 Proprietà delle Matrici Componenti

Abbiamo visto che le matrici componenti giocano un ruolo fondamentale nel calcolo di una funzione definita sullo spettro di una Matrice. Vediamo che proprietà hanno.

Osserviamo innanzitutto che esse commutano tra di loro, ossia:

$$Z_{ij}Z_{kr} = Z_{kr}Z_{ij} \quad \forall i, j, k, r$$

in quanto, come visto nell'esempio, sono dei polinomi valutati sulla stessa matrice. Si dimostra che:

$$\begin{aligned} Z_{ij}Z_{kr} &= O \quad i \neq k \\ Z_{i1}Z_{ij} &= Z_{ij} \quad j \geq 1 \\ Z_{i2}Z_{ij} &= jZ_{i,j+1} \quad j \geq 1 \\ Z_{ij} &= O \quad j \geq m_i \end{aligned}$$

Opportune considerazioni su queste proprietà portano a riscrivere:

$$f(A) = \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} f^{(j-1)}(\lambda_i) Z_{ij}$$

In una forma in cui è necessario calcolare solamente Z_{i1} :

$$f(A) = \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} \frac{f^{(j-1)}(\lambda_i)}{j-1!} Z_{i1}(A - \lambda_i I)^{j-1}$$

In cui, al solito, m_i è la molteplicità dell'autovalore λ_i nel polinomio minimale, per cui vale la relazione già esposta:

$$1 \leq m_i \leq \bar{m}_i$$

In cui \bar{m}_i è la molteplicità di λ_i nel polinomio caratteristico. Si dimostra che nell'espressione di $f(A)$ possiamo utilizzare indifferentemente le due molteplicità:

$$f(A) = \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} \frac{f^{(j-1)}(\lambda_i)}{j-1!} Z_{i1}(A - \lambda_i I)^{j-1} = \sum_{i=1}^{\nu} \sum_{j=1}^{\bar{m}_i} \frac{f^{(j-1)}(\lambda_i)}{j-1!} Z_{i1}(A - \lambda_i I)^{j-1}$$

Osserviamo infine che esistono alcune matrici per le quali il polinomio minimale coincide con il polinomio caratteristico. Ricordiamoli:

$$\psi(\lambda) = \prod_{i=1}^{\nu} (\lambda - \lambda_i)^{m_i}$$

$$p(\lambda) = \prod_{i=1}^{\nu} (\lambda - \lambda_i)^{\bar{m}_i}$$

La differenza sostanziale fra questi due polinomi è che il polinomio caratteristico è dato dal determinante di $A - I\lambda$, mentre, per come lo abbiamo definito intuitivamente, il polinomio minimale va cercato (numericamente?). Tuttavia esso sarà della forma sopracitata e i due polinomi saranno evidentemente lo stesso polinomio nel caso in cui:

$$\bar{m}_i = m_i \quad \forall i$$

Un esempio di matrice il cui polinomio caratteristico coincide col suo polinomio minimale è la matrice di compagnia di Frobenius, quella matrice che ci permette di calcolare la matrice di Casorati valutata in $n + 1$, a partire dalla matrice di Casorati valutata in n .

5.4 Successioni di funzioni di matrici

Sia ora assegnata una successione di funzioni $\{f_k(z)\}$ definite sullo spettro di A. Diremo che la successione converge a una certa funzione f sullo spettro di A se:

$$\lim_{k \rightarrow +\infty} f_k^{(j)}(\lambda_i) = f^{(j)}(\lambda_i)$$

In cui al solito, la funzione f_k deve ammettere la derivata in λ_i fino all'ordine molteplicità di λ_i nel polinomio minimale -1, per ogni λ_i autovalore distinto. Naturalmente, abbiamo che se una certa successione di funzioni converge a una certa funzione f sullo spettro di A, allora la successione di funzioni valutata in A, anzichè in z, converge a $f(A)$.

Sia A una matrice avente autovalori di modulo minore di 1. Abbiamo che

$$\sum_{n=0}^{\infty} A^n = (I - A)^{-1}$$

Dunque:

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

Possiamo utilizzare questo fatto per esporre il seguente teorema: Il limite superiore di $\|A^n\| =$

- $\lim_{n \rightarrow +\infty} \|A^n\| = 0$, se e solo se tutti gli autovalori di A hanno modulo minore di 1
- $c < \infty$, se e solo se tutti gli autovalori di A hanno modulo non maggiore di 1
- ∞ altrimenti

Pertanto definiamo una matrice A convergente se il suo raggio spettrale:

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$$

È minore di 1. Quindi A è convergente se, tra tutti i suoi autovalori, l'autovalore che ha modulo massimo (ossia l'autovalore che ne determina il raggio spettrale) ha modulo minore di 1. Semplice corollario:

$$A^n \rightarrow 0, n \rightarrow \infty \Leftrightarrow \rho(A) < 1$$

Sia dato il sistema di equazioni differenziali lineari:

$$y' = Ay$$

$$y(t_0) = y_0$$

La soluzione di tale sistema è data da:

$$y(t) = e^{A(t-t_0)} y_0 \quad t \geq t_0$$

Consideriamo il sistema lineare

$$y' = Ay$$

La soluzione $y = \vec{0}$ è un punto di equilibrio, in quanto:

$$y' = A\vec{0} = \vec{0} \quad \forall t$$

Tale soluzione è:

- asintoticamente stabile, se e solo se tutti gli autovalori di A hanno parte reale negativa
- stabile, se e solo se tutti gli autovalori di A hanno parte reale non positiva
- instabile, altrimenti

Analogamente, nel caso discreto si dimostra il seguente risultato:

La soluzione dell'equazione

$$y_{n+1} = Ay_n \quad n \geq n_0$$

È

$$y_n = A^{n-n_0} y_{n_0}$$

Dove y_{n_0} è la condizione iniziale. La soluzione nulla di tale equazione è un punto di equilibrio, in quanto:

$$A\vec{0} = \vec{0}$$

Ossia, ponendo $\overline{y_n} = \vec{0}$, abbiamo che:

$$f(n, \overline{y_n}) = \vec{0} = \overline{y_n}$$

Tale soluzione è:

- asintoticamente stabile, se e solo se tutti gli autovalori di A hanno parte reale negativa
- stabile, se e solo se tutti gli autovalori di A hanno parte reale non positiva
- instabile, altrimenti

5.5 Forma canonica di Jordan

Sia A una matrice $\overline{m} \times \overline{m}$ a valori reali.

Ci domandiamo se A è *diagonalizzabile* ossia se A è *simile* ad una matrice D diagonale.

Affinchè A sia simile a una matrice diagonale, deve esistere M di analoga dimensione tale per cui:

$$M^{-1}AM = D$$

La richiesta che facciamo sulla matrice D è che essa dovrà contenere sulla sua diagonale principale tutti gli autovalori di A contati con le loro molteplicità algebriche.

Vale il seguente teorema:

A è diagonalizzabile **se e solo se** tutti i suoi autovalori hanno molteplicità algebrica e molteplicità geometrica uguali.

Ricordiamo brevemente che per molteplicità algebrica si intende il numero di volte in cui l'autovalore λ annulla il polinomio caratteristico, mentre per molteplicità geometrica si intende il numero di autovettori linearmente indipendenti relativi all'autovettore λ , ossia, la dimensione dell'*autospazio* di λ . Infatti, se λ è autovalore di A relativo all'autovettore \vec{v}_0 allora:

$$A\vec{v}_0 = \lambda\vec{v}_0$$

È pertanto possibile trovare tutti gli autovettori di λ risolvendo il sistema lineare:

$$A\vec{x} = \lambda\vec{x}$$

In cui \vec{x} è l'autovettore incognito.

Una condizione sufficiente affinchè A sia diagonalizzabile è che A abbia tutti gli autovalori distinti. Infatti, data la relazione:

$$1 \leq mg(\lambda) \leq ma(\lambda)$$

Dove $mg(\lambda)$ è la molteplicità geometrica ed $ma(\lambda)$ la molteplicità algebrica di λ , se tutti gli autovalori sono distinti abbiamo:

$$1 \leq mg(\lambda) \leq 1 \Rightarrow 1 = mg(\lambda) = ma(\lambda)$$

Quindi tutti gli autovalori hanno effettivamente molteplicità algebrica e molteplicità geometrica uguali, e la matrice M che *diagonalizza* è la matrice che ha come colonne gli autovettori linearmente indipendenti di A .

Primo problema: Il fatto che A sia una matrice reale, non implica necessariamente che tutti i suoi autovalori siano reali: una matrice può essere non diagonalizzabile sui reali ma lo potrebbe essere sui complessi.

Secondo problema: se gli autovalori di A non hanno tutti quanti molteplicità algebrica uguale alla molteplicità geometrica, non è possibile diagonalizzare A . Soluzione: *forma canonica di Jordan*, la quale costituisce una *generalizzazione* della diagonalizzazione.

Diciamo che una matrice A è in forma canonica di Jordan se essa è costituita esclusivamente da miniblocchi di Jordan sulla diagonale principale.

Una matrice M è un miniblocco di Jordan se essa è della forma: valore costante sulla diagonale, 1 sotto la diagonale, 0 altrove.

Osserviamo che uno scalare è il caso particolare di un miniblocco di Jordan di

dimensione unitaria.

Teorema: se A è una matrice quadrata definita sul campo dei numeri complessi, allora essa sarà sempre simile a una matrice in forma canonica di Jordan, ossia esisterà sempre una matrice M tale per cui:

$$M^{-1}AM = J$$

In cui i valori costanti lungo la diagonale principale di J sono dati dagli autovalori (contati con le loro molteplicità algebriche e geometriche) di A : infatti ogni autovalore genera g miniblocchi di Jordan, dove g è la molteplicità geometrica dell'autovalore, e all'interno di ciascun miniblocco, l'autovalore è ripetuto a volte sulla diagonale, dove a è la molteplicità algebrica dell'autovalore.

La matrice M che Jordanizza A , in maniera simile alla matrice che diagonalizza, sarà la matrice che ha come colonne gli autovettori linearmente indipendenti e gli *autovettori generalizzati* di A . Ossia, se λ_i è un autovalore di A con molteplicità algebrica \bar{m}_i , abbiamo che esso darà luogo a $k_i \geq 1$ autovettori linearmente indipendenti (dunque k_i è la molteplicità geometrica di λ_i) e dovendo essere $k_i < \bar{m}_i$ ⁸, avremo altri $\bar{m}_i - k_i$ autovettori che chiamiamo autovettori *generalizzati*: questi, assieme ai k_i autovettori linearmente indipendenti, formano le colonne di M .

Osserviamo che se A è una matrice a valori reali, caso particolare di una matrice a valori complessi, essa sarà sempre simile a una matrice in forma canonica di Jordan, il punto è che quest'ultima potrebbe anche contenere al suo interno numeri complessi.

Il vantaggio di poter sempre Jordanizzare una matrice è dato dal fatto che la matrice J è simile alla matrice A di partenza, è più semplice da manipolare, ed inoltre esse condividono lo stesso polinomio caratteristico e lo stesso polinomio minimo.

Inoltre, se $f(z)$ è definita sullo spettro di A , e T è una matrice non singolare avente la stessa dimensione, si ha che:

$$T^{-1}f(A)T = f(T^{-1}AT)$$

⁸se fosse $k_i = m_i$ allora la matrice potrebbe essere diagonalizzabile e quindi gli autovettori generalizzati coincidono con gli autovettori classici, e pertanto la forma canonica di Jordan coinciderà con la matrice diagonale

6 Sistemi lineari di equazioni

Andiamo a discutere la risoluzione di sistemi lineari di equazioni differenziali e alle differenze. Da questo momento in poi, una grandezza vettoriale \vec{y} sarà espressa con la notazione \mathbf{y} .

6.1 Il caso continuo

Si consideri il seguente sistema di equazioni lineari non omogeneo:

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t) + \mathbf{b}(t) \quad t \geq t_0$$

Si definisce un tale sistema *sistema lineare non autonomo* in quanto esso costituisce l'analogo in più dimensioni della seguente equazione differenziale:

$$y'(t) = a(t)y(t) + b(t)$$

Ossia, possiamo notare che la funzione

$$y'(t) = f(t, (y(t))) \equiv a(t)y(t) + b(t)$$

Dipende anche da dal parametro t , infatti, in generale:

$$a(t_1), b(t_1) \neq a(t_2), b(t_2)$$

Pertanto, la $f(t, y)$ dipende sia da t che da y . Quando questo **non** accade, si parla di sistema autonomo.

Iniziamo la trattazione partendo dal caso in cui il sistema lineare sia in generale non autonomo, dunque il caso in cui sia la $A(t)$ che il vettore $\mathbf{b}(t)$ dipendono da t .

Al sistema

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t) + \mathbf{b}(t) \quad t \geq t_0$$

Si associa l'equazione omogenea:

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t) \quad t \geq t_0$$

Ricerchiamo dunque la soluzione generale dell'equazione omogenea che, successivamente, ci permetterà di ottenere la soluzione generale dell'equazione non omogenea.

Supponiamo esista una funzione a valori matriciali $W(t)$ tale per cui:

$$W'(t) = A(t)W(t) \quad t \geq t_0 \quad \det(W(t_0)) \neq 0 \quad (*)$$

Si dimostra che se esiste una tale $W(t)$ allora il suo determinante sarà diverso da 0 per ogni $t \geq t_0$

Definiamo la matrice fondamentale dell'equazione (*) la funzione di matrice:

$$\Phi(t, s) = W(t)W^{-1}(s)$$

Tale matrice è soluzione del problema:

$$\Phi'(t, t_0) = A(t)\Phi(t, t_0) \quad t \geq t_0, \Phi(t_0, t_0) = I$$

Definite W, Φ , ricerchiamo ora la soluzione dell'equazione omogenea soddisfacente la condizione iniziale

$$\mathbf{y}(t_0) = \mathbf{y}_0$$

La cercheremo nella forma

$$\mathbf{y}(t) = W(t)\mathbf{c}$$

Sostituendo si ottiene:

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t) \quad (\text{eq. omogenea})$$

$$\mathbf{y}'(t) = A(t)W(t)\mathbf{c}$$

Quindi tale $\mathbf{y}(t)$ soddisfa l'equazione omogenea.

Imponendo la condizione iniziale

$$\mathbf{y}(t_0) = \mathbf{y}_0$$

Otteniamo che:

$$W(t_0)\mathbf{c} = \mathbf{y}_0 \Rightarrow \mathbf{c} = W^{-1}(t_0)\mathbf{y}_0$$

In sintesi otteniamo che la soluzione del problema ai valori iniziali dato dall'equazione omogenea

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t) \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad t \geq t_0$$

È data da:

$$\mathbf{y}(t) = \Phi(t, t_0)\mathbf{y}_0 \quad t \geq t_0$$

La soluzione del problema espresso dall'equazione non omogenea si dimostra essere invece data da:

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t) + \mathbf{b}(t) \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad t \geq t_0$$

$$\mathbf{y}(t) = \Phi(t, t_0)\mathbf{y}_0 + \int_{t_0}^t \Phi(t, s)\mathbf{b}(s) \, ds$$

Osserviamo che nel semplice caso in cui avessimo un sistema autonomo, quindi nel caso in cui:

$$A(t) \equiv A \quad \mathbf{b}(t) \equiv \mathbf{b}$$

Avremmo:

$$\Phi(t, s) = e^{A(t-s)}$$

Dunque la soluzione generale (dell'eq. omogenea) diventa⁹:

$$\mathbf{y}(t) = e^{A(t-t_0)}\mathbf{y}_0 \quad t \geq t_0$$

Mentre la soluzione dell'equazione non omogenea diventa:

$$\mathbf{y}(t) = e^{A(t-t_0)}\mathbf{y}_0 + \int_0^{t-t_0} e^{As}\mathbf{b} \, ds$$

È utile osservare che nel caso in cui A abbia autovalori tutti a parte reale negativa, dunque non singolare, esiste un punto:

$$\bar{\mathbf{y}} = -A^{-1}\mathbf{b}$$

⁹Questo è vero anche nel caso in cui \mathbf{b} rimanga dipendente dal tempo, poiché nell'equazione omogenea non compare $\mathbf{b}(t)$

In cui la derivata si annulla:

$$f(t, \bar{\mathbf{y}}) = A \bar{\mathbf{y}} b = -A * A^{-1} \mathbf{b} + \mathbf{b} = -\mathbf{b} + \mathbf{b} = \mathbf{0}$$

Pertanto $\bar{\mathbf{y}}$ è un punto di equilibrio per il sistema, e, inoltre esso è anche asintoticamente stabile, poichè, per qualunque scelta delle condizioni iniziali:

$$\mathbf{y}(t) \rightarrow \bar{\mathbf{y}}, \quad t \rightarrow \infty$$

Dunque, riassumendo, dato il problema continuo autonomo:

$$\mathbf{y}'(t) = A\mathbf{y}(t) + \mathbf{b} \quad \mathbf{y}(0) = \mathbf{y}_0$$

Se la matrice A ha tutti gli autovalori a parte reale negativa, allora esiste ed è unico il punto di equilibrio:

$$\bar{\mathbf{y}} = -A^{-1}\mathbf{b}$$

Inoltre, per qualunque scelta della condizione iniziale \mathbf{y}_0 :

$$\mathbf{y}(t) \rightarrow \bar{\mathbf{y}} \quad t \rightarrow \infty$$

Analogamente, nel caso discreto, dato il sistema di equazioni alle differenze autonomo:

$$\mathbf{y}_{n+1} = A\mathbf{y}_n + \mathbf{b}$$

In cui è assegnata \mathbf{y}_0 , se il raggio spettrale di A è minore di 1, allora esiste ed è unico il punto di equilibrio:

$$\bar{\mathbf{y}} = (I - A)^{-1}\mathbf{b}$$

Inoltre, per qualunque scelta della condizione iniziale \mathbf{y}_0 :

$$\mathbf{y}_n \rightarrow \bar{\mathbf{y}} \quad n \rightarrow \infty$$

6.2 Sistemi dinamici nel piano delle fasi

Si consideri il seguente sistema di equazioni differenziali:

$$\mathbf{y}' = A\mathbf{y}$$

Con A matrice 2×2 .

Osserviamo che il punto di equilibrio è stato traslato nell'origine:

$$A^{-1}\mathbf{0} = \mathbf{0}$$

A questo punto, se A è una matrice 2×2 allora i suoi autovalori saranno λ_1, λ_2 , e sono tali per cui:

$$\lambda_1 \neq 0 \neq \lambda_2$$

. Possono avversi le seguenti possibilità

1. $\lambda_1 \neq \lambda_2$
2. $\lambda_1 = \lambda_2 = \lambda$
3. $\lambda_1 = \lambda = \bar{\lambda}_2 \neq \lambda_2$

In cui la scrittura $\bar{\lambda}_2$ indica il complesso e coniugato di λ_2 .

6.2.1 $\lambda_1 \neq \lambda_2$

Nel primo caso, quello più semplice, si ha che è possibile diagonalizzare la matrice A. Questo porta di fatto a trasformare il sistema nel suo equivalente, in cui A è una matrice diagonale:

$$y'_1 = \lambda_1 y_1 \quad y'_2 = \lambda_2 y_2$$

Dunque l'equazione

$$\mathbf{y}' = A\mathbf{y}$$

Si è decomposta in due equazioni che descrivono la derivata della prima e della seconda componente del vettore \mathbf{y} in funzione dei due autovalori distinti. Da questa relazione si ottiene facilmente che:

$$y_2 = cy_1^k$$

In cui c dipende dalle condizioni iniziali, e $k = \frac{\lambda_2}{\lambda_1}$

Si distinguono 2 casi:

- $\lambda_1 \lambda_2 > 0 \Rightarrow k > 0$

In questo caso si ottiene una famiglia di parabole passanti per l'origine.
Se entrambi gli autovalori sono negativi, le traiettorie andranno verso l'origine, e tale famiglia è detta configurazione a nodo stabile.
Se entrambi gli autovalori sono invece positivi, le traiettorie si allontaneranno dall'origine, e quest'ultima configurazione è detta a nodo instabile.

- $\lambda_1 \lambda_2 < 0 \Rightarrow k < 0$

In questo caso si ottiene una famiglia di iperboli che non passa dall'origine.
Si parla di configurazione a sella.

6.2.2 $\lambda_1 = \lambda_2 = \lambda$

Nel secondo caso, abbiamo che:

$$\lambda_1 = \lambda_2 \in \mathbb{R}$$

Nel caso in cui l'autovalore sia semisemplice, allora possiamo riscrivere il sistema come nel caso precedente:

$$y'_1 = \lambda y_1 \quad y'_2 = \lambda y_2$$

Pertanto si ottengono, nel piano delle fasi¹⁰, le curve:

$$y_1 = cy_2$$

Ovvero un fascio di rette passanti per l'origine. Come prima, si parla di *configurazione a stella stabile* se le traiettorie si dirigono verso l'origine, ossia il caso in cui $\lambda < 0$, oppure si parla di *configurazione a stella instabile* nel caso in cui le traiettorie si allontanino dall'origine, ossia il caso in cui $\lambda > 0$

¹⁰nel piano delle fasi ci sono tutte le possibili traiettorie soluzione, al variare della condizione iniziale

6.2.3 $\lambda_1 = \bar{\lambda}_2$

In quest'ultimo caso, abbiamo che gli autovalori sono ognuno il complesso e coniugato dell'altro. In tal caso, non è possibile diagonalizzare. Tuttavia, abbiamo che:

$$\lambda_1 = \alpha + i\beta, \quad \lambda_2 = \alpha - i\beta$$

In quanto abbiamo detto che gli autovalori sono ognuno il complesso e coniugato dell'altro.

Quindi, definiamo una matrice reale T tale per cui:

$$T^{-1}AT = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \equiv \hat{A}$$

Possiamo trasformare il vettore \mathbf{y} in coordinate polari:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \rho \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$$

Imponendo che sia soddisfatta l'equazione differenziale, si perviene alle equazioni:

$$\rho' = \alpha\rho$$

$$\theta' = \beta$$

La cui soluzione è data da:

$$\rho(t) = \rho(0)e^{\alpha t}$$

$$\theta(t) = \theta(0) + \beta t$$

Queste equazioni individuano una famiglia di traiettorie a spirale che si avvolgono intorno all'origine. Si parla di configurazione *a fuoco*. Come nei casi precedenti, si parla di configurazione a fuoco stabile se le traiettorie spiraleranno verso l'origine, caso che si ottiene quando $\alpha < 0$. Si parla altresì di configurazione a fuoco instabile nel caso in cui le traiettorie si allontaneranno dall'origine, ossia il caso in cui $\alpha > 0$.

Come ultimo caso, se abbiamo $\alpha = 0$, si ottiene una configurazione detta *a centro*, che vede l'origine come punto di equilibrio stabile. Le traiettorie in questo caso sono delle circonferenze.

6.3 Sistemi dinamici discreti nel piano delle fasi

Finora abbiamo discusso il comportamento delle traiettorie soluzione di un sistema dinamico continuo, ossia un sistema di equazioni differenziali.

Nel caso di un sistema dinamico discreto, dunque un sistema di equazioni alle differenze, si ha:

$$\mathbf{y}_{n+1} = A\mathbf{y}_n$$

Rimaniamo sempre nel semplice caso in cui la matrice A sia 2×2 .

Pertanto abbiamo ancora:

$$\sigma(A) = \{\lambda_1, \lambda_2\}$$

Ossia come nel caso precedente, la matrice A non singolare ha 2 autovalori. Possono essenzialmente avversi le seguenti possibilità:

$$\begin{aligned}
\bullet \quad & \lambda_1 \neq \lambda_2 \in \mathbb{R} \Rightarrow \begin{cases} |\lambda_1|, |\lambda_2| < 1 \Rightarrow \text{nodo stabile}; \\ |\lambda_1|, |\lambda_2| > 1 \Rightarrow \text{nodo instabile}; \\ |\lambda_1| < 1 < |\lambda_2| \Rightarrow \text{sella}; \end{cases} \\
\bullet \quad & \lambda_1 = \lambda_2 = \lambda \in \mathbb{R} \Rightarrow \begin{cases} |\lambda| < 1 \quad \text{semisemplice} \Rightarrow \text{nodo a stella stabile}; \\ |\lambda| > 1 \quad \text{semisemplice} \Rightarrow \text{nodo a stella instabile}; \\ |\lambda| < 1 \quad \text{non semisemplice} \Rightarrow \text{nodo degenere stabile}; \\ |\lambda| \geq 1 \quad \text{non semisemplice} \Rightarrow \text{nodo degenere instabile}; \end{cases} \\
\bullet \quad & \lambda_1 = \lambda = \bar{\lambda}_2 \neq \lambda_2 \Rightarrow \begin{cases} |\lambda| < 1 \Rightarrow \text{fuoco stabile}; \\ |\lambda| > 1 \Rightarrow \text{fuoco instabile}; \\ |\lambda| = 1 \Rightarrow \text{centro}. \end{cases}
\end{aligned}$$

6.4 Risoluzione numerica di sistemi di equazioni differenziali

Esaminiamo il caso in cui il sistema di equazioni differenziali autonomo

$$\mathbf{y}' = A\mathbf{y} + \mathbf{b}$$

Ammette un punto di equilibrio

$$A^{-1}\mathbf{b}$$

Asintoticamente stabile, ossia, abbiamo che per qualunque scelta delle condizioni iniziali, la traiettoria risultante converge al punto di equilibrio.

Traslando il punto di equilibrio nell'origine, ci riconduciamo allo studio dell'equazione omogenea

$$\mathbf{y}' = A\mathbf{y}$$

Sappiamo che, per quanto visto nel capitolo 5, affinché il sistema ammetta il punto di equilibrio come asintoticamente stabile, deve essere tale per cui la matrice A ha tutti gli autovalori a parte reale negativa. Infatti, ricordiamolo, nel caso in cui tutti gli autovalori avessero parte reale non positiva, il punto di equilibrio sarebbe solamente stabile, instabile in tutti gli altri casi.

Nel caso che stiamo considerando, possiamo trasformare il problema:

$$\mathbf{y}' = A\mathbf{y} \Rightarrow y' = \lambda y$$

In cui $\lambda \in \sigma(A)$.

Se vogliamo indurre una soluzione discreta che preservi le condizioni di asintotica stabilità della soluzione continua, dovrà aversi:

$$q = h\lambda \in \mathbb{D}$$

E lo si dovrà avere per ogni λ . Ricordiamo che h è al solito il passo di integrazione/discretizzazione.

Notiamo che grazie alla trasformazione del problema, ci troviamo a discutere dell'equazione test vista quando parlavamo di metodi lineari multistep. Per questi sappiamo che se $h\lambda$ appartiene alla regione di assoluta stabilità del metodo LMF, allora la soluzione indotta ammette l'origine come punto di equilibrio asintoticamente stabile.¹¹

La regione di assoluta stabilità è formata da tutti quei q nel piano complesso che rendono il polinomio di stabilità $\pi(z, q)$ un polinomio di Schur.

$$\pi(z, q) = \rho(z) - q\sigma(z)$$

Siano $z_i(q)$ tutte le radici del polinomio di stabilità. Abbiamo che se esse sono tutte distinte, allora ogni soluzione dell'equazione omogenea sarà della forma:

$$\mathbf{y}_n = \sum_{i=1}^k Z_i^n \mathbf{c}_i$$

Dove i vettori \mathbf{c}_i sono determinati dalle condizioni iniziali, k è il numero di passi del metodo LMF, mentre le Z_i sono dette matrici solventi.

Z è una matrice solvente se:

$$\pi(Z, J(q)) = O$$

In cui

$$J(q) = \begin{bmatrix} h\lambda & & \\ 1 & h\lambda & \\ \dots & & \\ & 1 & h\lambda \end{bmatrix}$$

È una matrice con $q = h\lambda$ sulla diagonale principale e 1 sulla sottodiagonale principale. Vale il seguente teorema:

I solventi dell'equazione di test matriciale sono dati da:

$$Z_i = z_i(J(q))$$

In cui le $z_i(q)$ sono le radici del polinomio $\pi(z, q)$. Vediamo un esempio. Consideriamo il metodo di Eulero Esplicito:

$$y_{n+1} = y_n + hf(n, y_n)$$

$$-y_n + y_{n+1} = f(n, y_n)$$

Esso è caratterizzato dai seguenti coefficienti α_i, β_i :

$$\alpha_i = \{-1, 1\} \quad \beta_i = \{1, 0\}$$

Pertanto

$$\rho(z) = \sum_{i=0}^1 \alpha_i z^i = -1 + z = z - 1$$

¹¹ricordiamo che, quando studiamo la stabilità di un metodo per h fissato, facciamo riferimento alla risoluzione dell'equazione di test con quel metodo, la quale ha l'origine come punto di equilibrio asintoticamente stabile. In questo caso, l'intero sistema di equazioni è nella forma dell'equazione di test

$$\sigma(z) = \sum_{i=0}^1 \beta_i z^i = 1$$

Affinchè sia un metodo consistente (quindi, tale per cui ad ogni passo l'errore locale di troncamento è più piccolo di un opportuno infinitesimo di ordine almeno 1 del passo h), dobbiamo verificare che:

$$\begin{aligned}\rho(1) &= 0, \quad \rho'(1) = \sigma(1) \\ \rho(1) &= 1 - 1 = 0 \quad \rho'(1) = 1 \quad \sigma(1) = 1 = \rho'(1)\end{aligned}$$

Abbiamo quindi un metodo consistente e 0-stabile in quanto $\rho(z)$ è un polinomio di von Neumann.

Calcoliamo il polinomio di test

$$\pi(z, q) = \rho(z) - q\sigma(z) = z - 1 - q$$

Il quale ha radici

$$\pi(z, q) = 0 \Leftrightarrow z - 1 - q = 0 \Rightarrow z(q) = 1 + q$$

Per trovare i solventi, in questo caso il solvente, devo calcolare

$$z(J(q))$$

Ossia una funzione polinomiale di matrice. Fissati h ed A posso calcolare i λ e la $J(q)$ e, conseguentemente, anche il solvente.

A conclusione, possiamo quindi asserire che l'analisi di stabilità vista per i metodi LMF, basata sull'equazione test scalare, continua a valere nel caso di sistemi lineari di equazioni.

6.5 Stabilità, condizionamento e *stiffness*

Nella precedente sezione, abbiamo visto che per una corretta approssimazione numerica del problema:

$$\mathbf{y}' = A\mathbf{y} \quad t \in [t_0, T], \mathbf{y}(t_0) = \mathbf{y}_0 \in \mathbb{R}^s, \sigma(A) \equiv \{\lambda_1, \dots, \lambda_s\} \in \mathbb{C}^-$$

È necessario che, fissato h , tutti i prodotti:

$$q_i = h\lambda_i \quad i = 1, \dots, s$$

Siano contenuti nella regione di assoluta stabilità del metodo utilizzato. Nel caso in cui si avesse:

$$\min |\lambda_i| << \max |\lambda_i| \Leftrightarrow \frac{\max |\lambda_i|}{\min |\lambda_i|} >> 1 \quad (*)$$

Ossia, nel caso in cui il rapporto fra il modulo dell'autovalore di modulo massimo e il modulo dell'autovalore di modulo minimo sia molto maggiore di 1 poichè l'autovalore di modulo minimo è molto inferiore all'autovalore di modulo massimo, sarebbe evidente che dovremmo scegliere un h estremamente piccolo per permettere a $h \max |\lambda_i|$ di rimanere nella regione di assoluta stabilità del metodo. Questa è una grave restrizione, in quanto dovremmo scegliere h

molto piccolo per "colpa" di anche un solo autovalore.

Il problema naturalmente si pone solo nei casi in cui il metodo utilizzato abbia una regione di stabilità limitata, come ad esempio nel caso dei metodi esplicativi. In caso avessimo un metodo A-stabile, quindi che comprenda tutto il semipiano complesso negativo nella sua regione di assoluta stabilità, non avremmo restrizioni su h in quanto per ipotesi tutti gli autovalori di A stanno nel semipiano complesso negativo.

Solitamente si definisce stiff un problema per cui vale la (*), ma questa definizione non è molto precisa, in quanto non tiene in considerazione l'ampiezza dell'intervallo di integrazione. Un altro problema è che tale definizione non tiene neanche in considerazione l'ipotesi in cui $s = 1$, quindi il caso in cui stessimo risolvendo un problema scalare: con questa definizione un problema scalare risulterà essere sempre non-stiff.

Per chiarire questo punto, si consideri la solita equazione test. Se la condizione iniziale vale y_0 allora la soluzione è data da:

$$y(t) = y_0 e^{\lambda(t-t_0)}$$

Se la condizione iniziale venisse perturbata di δy_0 , la soluzione subirebbe una perturbazione

$$\delta y(t) = \delta y_0 e^{\lambda(t-t_0)}$$

Studiamo quindi il condizionamento di questo problema, ovvero stabiliamo di quanto si possono amplificare le perturbazioni iniziali, sulla traiettoria risultante. Si ottiene, considerando una opportuna norma di funzioni:

$$\|\delta y\| \leq \|e^{\lambda(\cdot-t_0)}\| |\delta y_0|$$

Chiaramente, $\|e^{\lambda(\cdot-t_0)}\|$ sarà il numero di condizionamento del problema, rispetto alla norma considerata. Una norma che va sicuramente considerata è la norma infinito:

$$\|\delta y\|_\infty = \max_{t_0 \leq t \leq T} |\delta y(t)|$$

Che quantifica il massimo errore commesso. Pertanto, denotando $k = \|e^{\lambda(\cdot-t_0)}\|_\infty$, si ottiene:

$$\|\lambda y\|_\infty \leq k |\delta y_0|$$

Definiamo adesso la media dell'errore come la *norma 1*:

$$\|\delta y\|_1 = \frac{1}{T - t_0} \int_{t_0}^T |\delta y(t)| dt$$

In tal caso, denotando $\gamma = \|e^{\lambda(\cdot-t_0)}\|_1$, si ottiene:

$$\|\lambda y\|_1 \leq \gamma |\delta y_0|$$

Chiaramente, per costruzione, si ha:

$$0 < \gamma \leq k$$

Per quanto riguarda lo studio del condizionamento, si distinguono i seguenti casi:

- $k \approx \gamma \approx O(1)$: in tal caso, il problema è ben condizionato in entrambe le norme
- $k \geq \gamma \gg 1$: in tal caso il problema è malcondizionato in entrambe le norme
- $k \gg \gamma \approx O(1)$: in questo caso il problema è mal condizionato nella norma infinito, ma ben condizionato nella norma 1

Per discutere invece di *stiff* introduciamo il *rappporto di stiffness* come il valore:

$$\sigma = \frac{k}{\gamma}$$

Nel nostro caso, si avrà (supponendo un intervallo di integrazione sufficientemente ampio):

$$k = 1, \quad \gamma \approx [|\lambda|(T - t_0)]^{-1} \Rightarrow \sigma \approx |\lambda|(T - t_0)$$

Notiamo che abbiamo un diverso rapporto di stiff per ciascun autovalore di A :

$$\sigma_i = |\lambda_i|(T - t_0) \quad i = 1, \dots, s$$

Il problema si dirà stiff, se almeno uno di essi è grande.

Chiamiamo ora $|\lambda_{\min}|$ il modulo dell'autovalore di modulo minimo di A , e $|\lambda_{\max}|$ il modulo dell'autovalore di modulo massimo di A .

Si ha che il rapporto di stiffness del problema sarà dato da:

$$\sigma \equiv \max_i \sigma_i = |\lambda_{\max}|(T - t_0)$$

Notiamo che, se si volesse avere una informazione completa sulla dinamica, si dovrebbe integrare fino a che la componente più "lenta", ossia quella legata a λ_{\min} , non si è smorzata. A tal fine, dovrà avversi:

$$T - t_0 \approx \frac{1}{|\lambda_{\min}|}$$

E, pertanto:

$$\sigma = |\lambda_{\max}|(T - t_0) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$$

Ottenendo quindi la definizione classica di stiffness.

È importante sottolineare che se si avesse:

$$\frac{|\lambda_{\max}|}{|\lambda_{\min}|} \gg 1$$

Ma

$$|\lambda_{\max}|(T - t_0) \approx O(1)$$

Allora il problema **non sarebbe** stiff.

6.6 Sistemi dinamici positivi

Importanza particolare assumono, nelle applicazioni, sistemi dinamici in cui la soluzione deve essere positiva. Si parla, in tal caso, di *sistemi dinamici positivi*. Partiamo con l'enunciare il *Teorema di Perron-Frobenius*:

Se A è una matrice a valori reali di dimensione $n \times n$ strettamente positiva (ossia, tale per cui qualunque suo elemento è maggiore strettamente di zero), allora

$$\exists \lambda_0 > 0 \mathbf{v}_0 > 0 \text{ (autovalore dominante e autovettore dominante)}$$

Tali per cui:

- $A\mathbf{v}_0 = \lambda_0 \mathbf{v}_0$
- $\forall \lambda \in \sigma(A) : \lambda \neq \lambda_0$, risulta $|\lambda| < \lambda_0$
- λ_0 è semplice

Come corollario, osserviamo che se $\exists k \in \mathbb{N} : A^k > O$ (ossia, A^k ha tutti gli elementi strettamente maggiori di 0) allora per A vale il teorema di Perron Frobenius.

Del teorema di Perron Frobenius ne esiste anche la cosiddetta forma debole:

Se A è una matrice a valori reali di dimensione $n \times n$ tale per cui $A \geq O$, (ossia, ha tutti i suoi elementi maggiori o uguali di zero) allora

$$\exists \lambda_0 \geq 0 \mathbf{v}_0 \geq 0 \neq 0$$

(Il vettore \mathbf{v}_0 è maggiore o uguale di zero ma diverso dal vettore nullo)

Tali per cui:

- $A\mathbf{v}_0 = \lambda_0 \mathbf{v}_0$
- $\forall \lambda \in \sigma(A)$ risulta $|\lambda| \leq \lambda_0$

6.6.1 Sistemi dinamici positivi discreti

Sia dato il sistema di equazioni alle differenze:

$$\mathbf{y}_{n+1} = A\mathbf{y}_n + \mathbf{b} \quad n \geq n_0$$

Se $A \geq O \neq 0$ e $\mathbf{b} > 0$ parleremo di un sistema dinamico discreto positivo.

Se un sistema dinamico discreto è positivo, allora:

$$\mathbf{y}_{n_0} \geq 0 \Rightarrow \mathbf{y}_n > 0, \quad n > n_0$$

Ossia, ponendo come condizione iniziale un vettore positivo, la traiettoria soluzione rimarrà sempre positiva per ogni altro $n > n_0$ (maggiore perchè abbiamo imposto la condizione iniziale).

Inoltre, se la matrice

$$(I - A)$$

è non singolare, si avrà che il punto:

$$\mathbf{y} = (I - A)^{-1} \mathbf{b}$$

È un punto di equilibrio per il sistema, e per i sistemi dinamici positivi, vale il seguente importante risultato:

Esiste un punto di equilibrio $\mathbf{y} > 0$ se e solo se $\rho(A) < 1$.

In altri termini, per un sistema dinamico positivo discreto, l'esistenza di un punto di equilibrio a componenti positive equivale alla sua asintotica stabilità. Osserviamo che in generale l'esistenza del punto di equilibrio si ha quando la matrice A è non singolare. Abbiamo visto che tale punto è asintoticamente stabile, in generale, quando tutti gli autovalori hanno parte reale negativa. Nel caso di sistemi dinamici discreti positivi, tali per cui il punto di equilibrio è a componenti positive, essi ammettono tale punto come asintoticamente stabile anche se gli autovalori non sono tutti a parte reale negativa, tuttavia per gli autovalori deve essere vero che il raggio spettrale è minore di 1. Quest'ultima osservazione si completa con l'asserzione che raggio spettrale minore di 1 e positività del punto di equilibrio si implicano a vicenda, pertanto se il punto di equilibrio è positivo allora sicuramente il raggio spettrale è minore di 1. Parimenti, se il raggio spettrale è minore di 1, allora il punto di equilibrio è positivo.

6.6.2 Sistemi dinamici positivi continui

Un sistema lineare dinamico positivo è, nel caso continuo, un sistema della forma:

$$\mathbf{y}' = A\mathbf{y} + \mathbf{b} \quad \mathbf{y}(0) = \mathbf{y}_0$$

con

$$\mathbf{b} > 0 \quad A = B - cI \quad B = (b_{ij}) \geq 0 \quad c > 0$$

Ossia, la matrice A deve essere una matrice di *Metzeler* (se vogliamo un sistema dinamico continuo positivo).

Analogamente al caso discreto abbiamo:

$$\mathbf{y}_0 \geq 0 \Rightarrow \mathbf{y}(t) > 0 \quad \forall t > 0$$

Per il teorema di Perron Frobenius in forma debole, abbiamo che, essendo $B \geq 0$:

$$\exists \lambda_0 \in \sigma(B) : \lambda_0 = \rho(B)$$

Nel caso in cui fosse $c > \lambda_0$ allora la matrice A sarebbe nonsingolare, e il sistema ammetterebbe il punto di equilibrio:

$$\bar{\mathbf{y}} = -A^{-1}\mathbf{b} = (cI - B)^{-1}\mathbf{b}$$

Tale punto di equilibrio è anche asintoticamente stabile.

Vale il seguente teorema:

$$\exists \bar{\mathbf{y}} > 0 \Leftrightarrow \sigma(A) \subset \mathbb{C}^-$$

Pertanto, come già visto nel capitolo 5, il punto di equilibrio è asintoticamente stabile quando tutti gli autovalori di A sono a parte reale negativa.

Analogamente al caso discreto, anche per sistemi dinamici positivi continui l'esistenza di un punto di equilibrio a componenti positive equivale alla sua asintotica stabilità.

Nel caso discreto, tuttavia, abbiamo che il punto di equilibrio positivo esiste se

e solo se il raggio spettrale di A è minore di 1, mentre in questo caso il punto di equilibrio positivo esiste se e solo se la matrice ha tutti gli autovalori a parte reale negativa. Concludiamo con una semplice definizione. La matrice

$$A = cI - B \quad B \geq O$$

Si definisce M-Matrice nel caso in cui si abbia $\rho(B) < c$. Le M-Matrici giocano un ruolo fondamentale nei sistemi dinamici lineari positivi, sia continui che discreti. Inoltre, il ruolo che nel discreto è svolto dalle matrici positive, nel continuo è svolto dalle matrici di Metzeler.

Capitolo 7

Sistemi nonlineari di equazioni

I fenomeni reali sono raramente lineari: pertanto i modelli lineari, che finora abbiamo considerato, sono spesso solo delle approssimazioni di modelli più complessi. In questo capitolo vedremo, in forma molto semplificata, le questioni essenziali riguardanti lo studio di sistemi di equazioni, differenziali e alle differenze, nonlineari.

7.1 Il caso discreto

Sia assegnato il problema discreto nonlineare del primo ordine

$$\mathbf{y}_{n+1} = \mathbf{f}(n, \mathbf{y}_n), \quad n \geq n_0, \quad \mathbf{y}_{n_0} \in \mathbb{R}^m \quad \text{assegnato.} \quad (7.1)$$

Un vettore $\bar{\mathbf{y}}$ è detto *punto di equilibrio* o *punto critico* se

$$\bar{\mathbf{y}} = \mathbf{f}(n, \bar{\mathbf{y}}), \quad n \geq n_0. \quad (7.2)$$

Senza perdere in generalità, assumeremo che

$$\bar{\mathbf{y}} = \mathbf{0}. \quad (7.3)$$

Infatti, se così non fosse, ponendo $\mathbf{z}_n = \mathbf{y}_n - \bar{\mathbf{y}}$, si otterrebbe

$$\mathbf{z}_{n+1} = \mathbf{f}(n, \mathbf{z}_n + \bar{\mathbf{y}}) - \mathbf{f}(n, \bar{\mathbf{y}}) \equiv \mathbf{g}(n, \mathbf{z}_n),$$

che avrebbe il punto di equilibrio nell'origine.

Ciò premesso, andiamo a dare le seguenti definizioni di stabilità che generalizzano quelle viste nel caso lineare.

Definizione 7.1 *La soluzione nulla di (7.1) si dirà:*

- stabile, se $\forall \varepsilon > 0 \exists \delta = \delta(\varepsilon, n_0)$ tale che $\|\mathbf{y}_{n_0}\| < \delta \Rightarrow \|\mathbf{y}_n\| < \varepsilon, \forall n \geq n_0$;
- uniformemente stabile, se è stabile e, inoltre, $\delta = \delta(\varepsilon)$;
- asintoticamente stabile, se è stabile e, inoltre, $\lim_{n \rightarrow \infty} \mathbf{y}_n = \mathbf{0}$;
- uniformemente asintoticamente stabile, se è uniformemente stabile ed asintoticamente stabile;

- esponenzialmente asintoticamente stabile, se

$$\exists a, \delta > 0 \text{ e } \eta \in (0, 1) \quad \text{tali che} \quad \|\mathbf{y}_{n_0}\| \leq \delta \Rightarrow \|\mathbf{y}_n\| \leq a\delta \eta^{n-n_0}, \forall n \geq n_0.$$

Osservazione 7.1 Osserviamo che la esponenziale asintotica stabilità implica la uniforme asintotica stabilità ma la velocità con cui le soluzioni tendono a 0 è di tipo esponenziale. Ad esempio, l'equazione

$$y_{n+1} = \frac{|y_n|}{|y_n| + 1}, \quad n \geq n_0,$$

ha l'origine come punto di equilibrio uniformemente asintoticamente stabile ma non esponenzialmente asintoticamente stabile. Infatti,

$$y_n \simeq (n - n_0 + 1)^{-1}, \quad n \geq n_0.$$

Il caso lineare

Nel caso in cui il problema sia lineare (ma, in genere, non autonomo),

$$\mathbf{y}_{n+1} = A_n \mathbf{y}_n, \tag{7.4}$$

le precedenti definizioni impongono delle condizioni sulla matrice fondamentale Φ_{n,n_0} che, ricordiamo, è soluzione del problema

$$\Phi_{n+1,n_0} = A_n \Phi_{n,n_0}, \quad n \geq n_0, \quad \Phi_{n_0,n_0} = I,$$

ed è tale che $\mathbf{y}_n = \Phi_{n,n_0} \mathbf{y}_{n_0}$, $n \geq n_0$. Valgono, infatti, i seguenti risultati.

Teorema 7.1 La soluzione nulla di (7.4) è uniformemente stabile se e solo se esiste $M > 0$ tale che:

$$\|\Phi_{n,n_0}\| \leq M, \quad \forall n \geq n_0. \tag{7.5}$$

Dimostrazione. Supponiamo che valga la (7.5). Segue quindi che

$$\|\mathbf{y}_n\| = \|\Phi_{n,n_0} \mathbf{y}_{n_0}\| \leq \|\Phi_{n,n_0}\| \cdot \|\mathbf{y}_{n_0}\| \leq M \|\mathbf{y}_{n_0}\|.$$

Pertanto, fissato $\varepsilon > 0$, ponendo $\delta = \delta(\varepsilon) = M^{-1}\varepsilon$, si ha che

$$\|\mathbf{y}_{n_0}\| < \delta \Rightarrow \|\mathbf{y}_n\| < \varepsilon$$

e, pertanto, la soluzione nulla è uniformemente stabile. Viceversa, supponiamo che la soluzione nulla sia uniformemente stabile, allora $\forall n \geq n_0$

$$\sup_{\|\mathbf{x}\|=1} \|\Phi_{n,n_0} \mathbf{x}\| \equiv \|\Phi_{n,n_0}\|$$

è limitato, ovvero vale la (7.5). \square

Per la uniforme asintotica stabilità, vale il seguente risultato che non dimostriamo, anche se una implicazione era già stata osservata in precedenza.¹

¹ Per la dimostrazione, si veda il Teorema 4.2.2 in [15, pag. 107].

Teorema 7.2 La soluzione nulla di (7.4) è uniformemente asintoticamente stabile se e solo se $\exists a > 0$ e $\eta \in (0, 1)$ tali che:

$$\|\Phi_{n,n_0}\| \leq a\eta^{n-n_0}, \quad n \geq n_0. \quad (7.6)$$

Corollario 7.1 La soluzione nulla di (7.23) è uniformemente asintoticamente stabile se e solo se essa è esponenzialmente asintoticamente stabile.²

Osservazione 7.2 Quindi nel caso lineare generale, le proprietà di stabilità della soluzione di equilibrio dipendono dalle proprietà della matrice fondamentale e, pertanto, indipendenti dalla soluzione di equilibrio stessa.

Processo di linearizzazione

Consideriamo ora sistemi nonlineari del tipo

$$\mathbf{y}_{n+1} = A_n \mathbf{y}_n + \mathbf{g}(n, \mathbf{y}_n), \quad n \geq n_0, \quad \mathbf{y}_{n_0} \in \mathbb{R}^m \text{ dato}, \quad (7.7)$$

in cui A_n è nonsingolare e $\mathbf{g}(n, \mathbf{y})$ è in genere nonlineare e soddisfa

$$\mathbf{g}(n, \mathbf{0}) = \mathbf{0}, \quad \forall n \geq n_0. \quad (7.8)$$

In questo caso, evidentemente, l'origine è ancora un punto critico. Riguardando il termine nonlineare in (7.7) formalmente come un termine noto \mathbf{b}_n , si ottiene, in virtù della (6.18),

$$\mathbf{y}_n = \Phi_{n,n_0} \mathbf{y}_{n_0} + \sum_{i=n_0}^{n-1} \Phi_{n,i+1} \mathbf{g}(i, \mathbf{y}_i), \quad n \geq n_0. \quad (7.9)$$

Da questo si intuisce come, sotto opportune ipotesi, le proprietà di stabilità del punto critico dipendano dalle proprietà di stabilità derivanti dalla parte lineare di (7.7), formalmente data da (7.4). In questo modo, è possibile discutere le proprietà di stabilità del punto critico di (7.1) dove, supponendo che $\mathbf{f}(n, \mathbf{y})$ sia abbastanza regolare rispetto a \mathbf{y} , si ottiene, mediante linearizzazione, che

$$\mathbf{y}_{n+1} = \underbrace{\mathbf{f}(n, \mathbf{0})}_{=0} + \underbrace{J_{\mathbf{f}}(n, \mathbf{0})}_{\equiv A_n} \mathbf{y}_n + \mathbf{g}(n, \mathbf{y}_n) = A_n \mathbf{y}_n + \mathbf{g}(n, \mathbf{y}_n), \quad (7.10)$$

dove abbiamo denotato con $J_{\mathbf{f}}$ la matrice Jacobiana di \mathbf{f} rispetto a \mathbf{y} e, inoltre, supponendo che \mathbf{f} sia sufficientemente regolare (come assumeremo nel seguito),

$$\|\mathbf{g}(n, \mathbf{y}_n)\| = O(\|\mathbf{y}_n\|^2). \quad (7.11)$$

Osservazione 7.3 L'equazione linearizzata nel punto di equilibrio, definisce il cosiddetto problema variazionale. In generale, quest'ultimo può essere definito rispetto ad una qualunque soluzione di riferimento.

Vale il seguente risultato.

Teorema 7.3 Siano dati il problema (7.10) ed il problema omogeneo associato (7.4). Se:

²Ovvero, nel caso lineare, uniforme asintotica stabilità ed esponenziale asintotica stabilità di un punto critico sono proprietà equivalenti.

- $\|\mathbf{g}(n, \mathbf{y}_n)\| \leq L\|\mathbf{y}_n\|$, con L sufficientemente “piccolo”,
 - la soluzione nulla di (7.4) è uniformemente asintoticamente stabile,
- allora la soluzione nulla di (7.10) è esponenzialmente asintoticamente stabile.

Dimostrazione. Essendo l’origine uniformemente asintoticamente stabile per la (7.4) esistono, in virtù del Teorema 7.2 $a > 0$ e $\eta \in (0, 1)$ tali che la matrice fondamentale del problema soddisfa (7.6). Pertanto, dalla (7.9) segue che:

$$\|\mathbf{y}_n\| \leq a\eta^{n-n_0}\|\mathbf{y}_{n_0}\| + aL \sum_{i=n_0}^{n-1} \eta^{n-i-1}\|\mathbf{y}_i\|, \quad n \geq n_0.$$

Moltiplicando membro a membro per η^{-n} , e ponendo $p_n = \eta^{-n}\|\mathbf{y}_n\|$, si ottiene

$$p_n \leq ap_{n_0} + aL\eta^{-1} \sum_{i=n_0}^{n-1} p_i, \quad n \geq n_0.$$

Dal Corollario 2.3 (Teorema di Gronwall discreto, vedi (2.28)) segue quindi che:

$$p_n \leq ap_{n_0} \prod_{i=n_0}^{n-1} (1 + aL\eta^{-1}) = ap_{n_0}(1 + aL\eta^{-1})^{n-n_0}, \quad n \geq n_0.$$

Moltiplicando ambo i membri per η^n , si ottiene, quindi:

$$\|\mathbf{y}_n\| \leq a\|\mathbf{y}_{n_0}\|(\eta + aL)^{n-n_0}, \quad n \geq n_0.$$

Si vede quindi che, se L è sufficientemente “piccolo” da aversi:

$$\eta + aL < 1,$$

allora la soluzione è esponenzialmente asintoticamente stabile. \square

Corollario 7.2 (Stabilità in prima approssimazione, caso discreto) *Se la soluzione nulla di*

$$\mathbf{y}_{n+1} = J_f(n, \mathbf{0})\mathbf{y}_n, \quad n \geq n_0$$

è uniformemente asintoticamente stabile e, con riferimento alla (7.10),

$$\lim_{\|\mathbf{y}\| \rightarrow 0} \frac{\|\mathbf{g}(n, \mathbf{y})\|}{\|\mathbf{y}\|} = 0$$

uniformemente rispetto a n , allora la stessa è esponenzialmente asintoticamente stabile per (7.1)–(7.3), purché $\|\mathbf{y}_{n_0}\|$ sia sufficientemente piccola.

Dimostrazione. Infatti, dalla (7.11), segue che, per $\|\mathbf{y}\|$ sufficientemente piccolo, definendo

$$L(\varepsilon) = \sup_{\|\mathbf{y}\| \leq \varepsilon} \frac{\|\mathbf{g}(n, \mathbf{y})\|}{\|\mathbf{y}\|},$$

si ottiene che $L(\varepsilon)$ è una funzione non decrescente e tale che $\lim_{\varepsilon \rightarrow 0} L(\varepsilon) = 0$. Pertanto,

$$\|\mathbf{g}(n, \mathbf{y})\| \leq L(\|\mathbf{y}\|)\|\mathbf{y}\|,$$

e, di conseguenza, ripercorrendo passi analoghi a quelli visti nella dimostrazione del precedente Teorema 7.3, si otterrà:

$$\|\mathbf{y}_n\| \leq a\|\mathbf{y}_{n_0}\| \prod_{i=n_0}^{n-1} (\eta + aL(\|\mathbf{y}_i\|)).$$

Se fosse $a \leq 1$, la tesi discenderebbe facilmente per induzione, imponendo che sia $\|\mathbf{y}_{n_0}\|$ sufficientemente piccola da aversi $\eta + aL(\|\mathbf{y}_0\|) < 1$. In tal caso, infatti, si otterrebbe $\|\mathbf{y}_{n_0+1}\| \leq \|\mathbf{y}_{n_0}\|(\eta + aL(\|\mathbf{y}_0\|))$ e quindi, per induzione,

$$\|\mathbf{y}_n\| \leq \|\mathbf{y}_{n_0}\| (\eta + aL(\|\mathbf{y}_{n_0}\|))^{n-n_0}, \quad n \geq n_0,$$

da cui segue l'asserto. Nel caso in cui si abbia $a > 1$, ponendo $z_n = \log \|\mathbf{y}_n\|$, si ottiene:

$$z_n \leq z_0 + \log a + \sum_{i=n_0}^{n-1} \log(\eta + aL(e^{z_i})). \quad (7.12)$$

Supponiamo ora che $\|\mathbf{y}_{n_0}\|$ sia sufficientemente piccola, in modo tale che

$$z_0 < z_0 + \log a \leq -N,$$

con N tale che, per un opportuno $\varepsilon > 0$, si abbia

$$\log(\eta + aL(e^{-N})) \leq -\varepsilon.$$

Pertanto,

$$z_{n_0+1} \leq -N + \log(\eta + aL(e^{z_0})) \leq -N + \log(\eta + aL(e^{-N})) \leq -N - \varepsilon.$$

Ragionando per induzione, dalla (7.12) si ottiene, quindi,

$$z_n \leq -N - (n - n_0)\varepsilon, \quad n \geq n_0,$$

ovvero, per ogni $\|\mathbf{y}_{n_0}\| \leq \delta \equiv e^{-N}a^{-1}$, risulta:

$$\|\mathbf{y}_n\| \leq a\|\mathbf{y}_{n_0}\| \prod_{i=n_0}^{n-1} (\eta + aL(\|\mathbf{y}_i\|)) \leq e^{-N}e^{-\varepsilon(n-n_0)} \equiv a\delta \hat{\eta}^{n-n_0}, \quad n \geq n_0,$$

con $\hat{\eta} = e^{-\varepsilon} < 1$, da cui la tesi segue. \square

Il corollario successivo, particolarmente importante per le applicazioni, riguarda il sistema

$$\mathbf{y}_{n+1} = A\mathbf{y}_n + \mathbf{g}(n, \mathbf{y}_n), \quad \mathbf{g}(n, \mathbf{0}) = \mathbf{0}, \quad n \geq n_0. \quad (7.13)$$

Corollario 7.3 (Teorema di Perron, versione discreta) *Sia, nella (7.13), $\rho(A) < 1$ e, inoltre,*

$$\lim_{\|\mathbf{y}\| \rightarrow 0} \frac{\|\mathbf{g}(n, \mathbf{y})\|}{\|\mathbf{y}\|} = 0 \quad (7.14)$$

uniformemente rispetto a n . Allora la sua soluzione nulla è esponenzialmente asintoticamente stabile.

Dimostrazione. Essendo $\rho(A) < 1$, la soluzione nulla è (uniformemente) asintoticamente stabile per il sistema

$$\mathbf{y}_{n+1} = A\mathbf{y}_n, \quad n \geq n_0.$$

La tesi discende, quindi, dal precedente Corollario 7.2. \square

Osservazione 7.4 Chiaramente, se l'origine è instabile per la (7.4), allora lo è anche per la (7.10).

Il teorema di Perron, insieme al suo corrispettivo continuo che esamineremo successivamente, è di fondamentale importanza nelle applicazioni. Infatti esso giustifica l'uso della parte lineare di equazioni più complesse in luogo di queste ultime. Esso, inoltre, è di utilità anche nella modellistica dei fenomeni (come vedremo). Di seguito, riportiamo alcuni esempi di applicazione che riguardano specificatamente l'Analisi Numerica.

Convergenza dei metodi iterativi per la ricerca degli zeri di funzioni

La ricerca di zeri di sistemi di equazioni nonlineari,

$$\mathbf{f}(\mathbf{x}) = \mathbf{0},$$

è effettuata per mezzo di metodi iterativi della forma

$$\mathbf{x}_{n+1} = \phi(\mathbf{x}_n), \quad n \geq 0, \tag{7.15}$$

dove ϕ è la *funzione di iterazione* del metodo. Ad esempio, per il metodo di Newton si ha:

$$\phi(\mathbf{x}) = \mathbf{x} - J_{\mathbf{f}}(\mathbf{x})^{-1}\mathbf{f}(\mathbf{x}),$$

dove $J_{\mathbf{f}}(\mathbf{x})$ è la matrice Jacobiana di $\mathbf{f}(\mathbf{x})$. In questo caso, la soluzione del problema, sia essa $\bar{\mathbf{x}}$, tale che $\mathbf{f}(\bar{\mathbf{x}}) = \mathbf{0}$, diviene un punto fisso (ovvero un punto di equilibrio) per la funzione di iterazione:

$$\bar{\mathbf{x}} = \phi(\bar{\mathbf{x}}).$$

Il metodo (7.15) sarà convergente alla soluzione $\bar{\mathbf{x}}$ se l'errore converge al vettore nullo:

$$\mathbf{e}_n \equiv \mathbf{x}_n - \bar{\mathbf{x}} \rightarrow \mathbf{0}, \quad n \rightarrow \infty.$$

Sviluppando la ϕ in serie di Taylor in $\bar{\mathbf{x}}$ (che, ovviamente, supponiamo essere definita), dalla (7.15) si ottiene:

$$\mathbf{e}_{n+1} = J_{\phi}(\bar{\mathbf{x}})\mathbf{e}_n + \mathbf{g}(\mathbf{e}_n), \quad n \geq 0,$$

dove

$$\|\mathbf{g}(\mathbf{e}_n)\| = O(\|\mathbf{e}_n\|^2).$$

Ponendo $A = J_{\phi}(\bar{\mathbf{x}})$, si ottiene

$$\mathbf{e}_{n+1} = A\mathbf{e}_n + \mathbf{g}(\mathbf{e}_n), \quad n \geq 0, \tag{7.16}$$

che è nella forma (7.13). Dal Teorema di Perron, segue quindi che se $\rho(A) < 1$, allora il procedimento risulterà convergente in un opportuno intorno della soluzione stessa. Questo avverrà sempre, sotto opportune ipotesi di regolarità per \mathbf{f} , e assumendo $J_{\mathbf{f}}(\bar{\mathbf{x}})$ nonsingolare, poiché

$$J_{\phi}(\bar{\mathbf{x}}) = I - \underbrace{J_{\mathbf{f}}(\bar{\mathbf{x}})^{-1}J_{\mathbf{f}}(\bar{\mathbf{x}})}_{=I} - \underbrace{F(\bar{\mathbf{x}})\mathbf{f}(\bar{\mathbf{x}})}_{=0} = O,$$

dove si è indicato con F il tensore con le derivate rispetto a \mathbf{x} di $J_{\mathbf{f}}^{-1}$. Pertanto, la convergenza sarà sempre assicurata, partendo da un punto iniziale sufficientemente vicino alla radice.

Uso dell'equazione test

Il Teorema di Perron, permette di dare una giustificazione rigorosa per l'analisi di stabilità lineare dei metodi numerici per la risoluzione di equazioni differenziali ordinarie. Infatti, sia dato il problema (che supponiamo scalare, per semplicità e brevità di trattazione)

$$y' = f(y), \quad f(0) = 0, \quad (7.17)$$

avente un punto di equilibrio asintoticamente stabile nell'origine. Utilizzando un metodo (ρ, σ) per la sua risoluzione approssimata con passo h , si otterrà l'equazione alle differenze

$$\rho(E)y_n - h\sigma(E)f(y_n) = 0, \quad n \geq 0.$$

Ponendo $\lambda = f'(0)$, si ottiene quindi

$$\rho(E)y_n - q\sigma(E)y_n - h\sigma(E)g(y_n) = 0, \quad n \geq 0, \quad (7.18)$$

dove

$$q = h\lambda \quad \text{e} \quad g(y_n) = O(y_n^2).$$

Se

$$\rho(z) = \sum_{i=0}^k \alpha_i z^i, \quad \sigma(z) = \sum_{i=0}^k \beta_i z^i,$$

questa equazione alle differenze di ordine k può essere trasformata in un sistema lineare del primo ordine di dimensione k . Infatti, definendo i vettori

$$\mathbf{y}_n = \begin{pmatrix} y_n \\ y_{n+1} \\ \vdots \\ y_{n+k-1} \end{pmatrix}, \quad \mathbf{g}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{h}{\alpha_k - q\beta_k} \sigma(E)g(y_n) \end{pmatrix} \in \mathbb{R}^k,$$

e la matrice (in forma di Frobenius)

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -\frac{\alpha_0 - q\beta_0}{\alpha_k - q\beta_k} & \dots & \dots & \dots & -\frac{\alpha_{k-1} - q\beta_{k-1}}{\alpha_k - q\beta_k} \end{pmatrix} \in \mathbb{R}^{k \times k},$$

la (7.18) può essere riscritta come

$$\mathbf{y}_{n+1} = A\mathbf{y}_n + \mathbf{g}_n$$

che è nella forma richiesta dal Teorema di Perron e, pertanto, se $\rho(A) < 1$, allora la soluzione discreta sarà asintoticamente stabile in un opportuno intorno dell'origine. Essendo la matrice in forma di Frobenius, si ricava che, a meno del fattore moltiplicativo $(\alpha_k - q\beta_k)^{-1}$, il suo polinomio caratteristico è

$$\pi(z, q) = \rho(z) - q\sigma(z)$$

e, pertanto,

$$\rho(A) < 1 \quad \Leftrightarrow \quad q \in \mathcal{D},$$

essendo \mathcal{D} la regione di assoluta stabilità del metodo (ρ, σ) .

Osservazione 7.5 In conclusione, il Teorema di Perron giustifica l'analisi di stabilità lineare dei metodi, basata sull'equazione test, nell'intorno di un punto di equilibrio asintoticamente stabile di un sistema autonomo.

Influenza degli errori di macchina

L'ipotesi (7.14) del Teorema di Perron, che vale in aritmetica esatta, è quasi sempre non verificata quando si utilizza l'aritmetica finita di un calcolatore. In questo caso, la (7.14) è più realisticamente sostituita da una diseguaglianza del tipo

$$\|\mathbf{g}(n, \mathbf{y}_n)\| \leq \delta,$$

dove δ è "piccolo" ma finito.³ In questo caso, se consideriamo ad esempio la (7.16), si ottiene (supponendo, per semplicità, che si abbia $\|A\| < 1$, oltre che $\rho(A) < 1$)⁴:

$$\|\mathbf{e}_{n+1}\| \leq \|A\| \|\mathbf{e}_n\| + \delta, \quad n \geq 0.$$

Da questa, considerando l'equazione del confronto, si ricava facilmente che

$$\|\mathbf{e}_n\| \leq \sum_{i=0}^{n-1} \|A\|^i \delta = \frac{1 - \|A\|^n}{1 - \|A\|} \delta < \frac{1}{1 - \|A\|} \delta, \quad n \geq 0.$$

Se ne deduce che l'errore dell'equazione perturbata rimane limitato, sebbene si perda la asintotica stabilità dell'origine (che, anzi, potrebbe non essere più un punto critico dell'equazione perturbata). In questo caso, in cui le perturbazioni non sono infinitesime con l'errore, ma permanenti, si parla di *problema di stabilità totale*. In maggior dettaglio, in questo caso la soluzione nulla dell'equazione

$$\mathbf{e}_{n+1} = A\mathbf{e}_n, \quad n \geq n_0$$

si dirà *totalmente stabile*.

Stabilità totale

Gli argomenti appena visti possono essere estesi ad equazioni più generali della (7.16). Più in generale, se consideriamo le equazioni:

$$\mathbf{y}_{n+1} = f(n, \mathbf{y}_n) + R(n, \mathbf{y}_n), \quad n \geq n_0, \quad (7.19)$$

$$\mathbf{y}_{n+1} = f(n, \mathbf{y}_n), \quad f(n, \mathbf{0}) = \mathbf{0}, \quad (7.20)$$

abbiamo che l'origine è un punto di equilibrio per la (7.20) (ma non necessariamente per la (7.19)).

Definizione 7.2 Diremo che l'origine è totalmente stabile (o stabile rispetto a perturbazioni permanenti) per la (7.20) se, per ogni $\varepsilon > 0$, esistono $\delta_1 = \delta_1(\varepsilon)$ e $\delta_2 = \delta_2(\varepsilon)$ tali che, per la soluzione $\{\mathbf{y}_n\}$ di (7.19):

$$\|\mathbf{y}_{n_0}\| < \delta_1 \quad e \quad \|R(n, \mathbf{y}_n)\| < \delta_2 \quad \Rightarrow \quad \|\mathbf{y}_n\| < \varepsilon, \quad \forall n \geq n_0.$$

A riguardo, è possibile dimostrare il seguente risultato (vedi [15, pag. 137]).

Teorema 7.4 Se l'origine è uniformemente asintoticamente stabile per (7.20), e f è Lipschitziana rispetto a \mathbf{y} , allora l'origine è totalmente stabile.

³Ad esempio, potrebbe avversi $\delta \simeq \varepsilon$, la precisione di macchina dell'aritmetica utilizzata.

⁴In realtà, tra tutte le norme indotte su matrice, ce n'è sempre una per cui $\rho(A) = \|A\|$.

7.2 Il caso continuo

Argomenti del tutto analoghi possono essere ripetuti nel caso di sistemi nonlineari di equazioni differenziali,

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}), \quad \mathbf{y}(t_0) = \mathbf{y}_0 \in \mathbb{R}^m, \quad (7.21)$$

in cui, al solito senza perdere in generalità, si suppone che

$$\mathbf{f}(t, \mathbf{0}) = \mathbf{0}, \quad \forall t \geq t_0. \quad (7.22)$$

Ovvero, $\mathbf{0}$ è un *punto di equilibrio*, o *punto critico*, per (7.21). Le seguenti definizioni, costituiscono la controparte continua di quanto visto nella precedente sezione.

Definizione 7.3 *La soluzione nulla di (7.21)-(7.22) si dirà:*

- stabile, se $\forall \varepsilon > 0 \exists \delta = \delta(\varepsilon, t_0)$ tale che $\|\mathbf{y}_0\| < \delta \Rightarrow \|\mathbf{y}(t)\| < \varepsilon, \forall t \geq t_0$;
- uniformemente stabile, se è stabile e, inoltre, $\delta = \delta(\varepsilon)$;
- asintoticamente stabile, se è stabile e, inoltre, $\lim_{t \rightarrow \infty} \mathbf{y}(t) = \mathbf{0}$;
- uniformemente asintoticamente stabile, se è uniformemente stabile ed asintoticamente stabile;
- esponenzialmente asintoticamente stabile, se

$$\exists \alpha, \beta, \delta > 0 \text{ tali che } \|\mathbf{y}_0\| \leq \delta \Rightarrow \|\mathbf{y}(t)\| \leq \alpha \delta e^{-\beta(t-t_0)}, \forall t \geq t_0.$$

Osservazione 7.6 Analogamente al caso discreto, anche ora osserviamo che la esponenziale asintotica stabilità implica la uniforme asintotica stabilità ma la velocità con cui le soluzioni tendono a 0 è di tipo esponenziale. Ad esempio, l'equazione

$$y' = -y^2, \quad t \geq t_0, \quad \text{con condizione iniziale} \quad y(t_0) = a > 0,$$

ha l'origine come punto di equilibrio uniformemente asintoticamente stabile ma non esponenzialmente asintoticamente stabile. Infatti,

$$y(t) = \frac{a}{a(t-t_0) + 1} \simeq t^{-1}, \quad t \gg t_0.$$

Il caso lineare

Nel caso in cui il problema sia lineare (ma, in genere, non autonomo),

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t), \quad (7.23)$$

le precedenti definizioni impongono delle condizioni sulla matrice fondamentale $\Phi(t, t_0)$ che, ricordiamo, è soluzione del problema

$$\Phi'(t, t_0) = A(t)\Phi(t, t_0), \quad t \geq t_0, \quad \Phi(t_0, t_0) = I.$$

Analogamente al caso discreto, valgono, infatti, i seguenti risultati.

Teorema 7.5 La soluzione nulla di (7.23) è uniformemente stabile se e solo se esiste $M > 0$ tale che:

$$\|\Phi(t, t_0)\| \leq M, \quad \forall t \geq t_0.$$

Teorema 7.6 La soluzione nulla di (7.23) è uniformemente asintoticamente stabile se e solo se $\exists \alpha, \beta > 0$ tali che:

$$\|\Phi(t, t_0)\| \leq \alpha e^{-\beta(t-t_0)}, \quad t \geq t_0.$$

Corollario 7.4 La soluzione nulla di (7.23) è uniformemente asintoticamente stabile se e solo se essa è esponenzialmente asintoticamente stabile.⁵

Osservazione 7.7 Quindi nel caso lineare generale, le proprietà di stabilità della soluzione di equilibrio dipendono dalle proprietà della matrice fondamentale e , pertanto, indipendenti dalla soluzione di equilibrio stessa.

Processo di linearizzazione

Analogamente a quanto visto nel caso discreto, consideriamo ora sistemi nonlineari del tipo

$$\mathbf{y}'(t) = A(t)\mathbf{y}(t) + \mathbf{g}(t, \mathbf{y}(t)), \quad t \geq t_0, \quad \mathbf{y}(t_0) = \mathbf{y}_0 \in \mathbb{R}^m, \quad (7.24)$$

in cui $\mathbf{g}(t, \mathbf{y})$ è in genere nonlineare e soddisfa

$$\mathbf{g}(t, \mathbf{0}) = \mathbf{0}, \quad \forall t \geq t_0. \quad (7.25)$$

In questo caso, evidentemente, l'origine è ancora un punto critico. Riguardando il termine nonlineare in (7.24) formalmente come un termine noto $\mathbf{b}(t)$, si ottiene, in virtù della (6.8),

$$\mathbf{y}(t) = \Phi(t, t_0)\mathbf{y}_0 + \int_{t_0}^t \Phi(t, s)\mathbf{g}(s, \mathbf{y}(s))ds, \quad t \geq t_0. \quad (7.26)$$

Analogamente al caso discreto, da questo si intuisce come, sotto opportune ipotesi, le proprietà di stabilità del punto critico dipendano dalle proprietà di stabilità derivanti dalla parte lineare della (7.24), formalmente data da (7.23). In questo modo, è possibile discutere le proprietà di stabilità del punto critico di (7.21) dove, supponendo che $\mathbf{f}(t, \mathbf{y})$ sia abbastanza regolare rispetto a \mathbf{y} , si ottiene, mediante linearizzazione, che

$$\mathbf{y}'(t) = \underbrace{\mathbf{f}(t, \mathbf{0})}_{=0} + \underbrace{J_{\mathbf{f}}(t, \mathbf{0})}_{\equiv A(t)} \mathbf{y}(t) + \mathbf{g}(t, \mathbf{y}(t)) = A(t)\mathbf{y}(t) + \mathbf{g}(t, \mathbf{y}(t)), \quad (7.27)$$

dove abbiamo denotato con $J_{\mathbf{f}}$ la matrice Jacobiana di \mathbf{f} rispetto a \mathbf{y} e, inoltre, supponendo che \mathbf{g} sia sufficientemente regolare,

$$\|\mathbf{g}(t, \mathbf{y}(t))\| = O(\|\mathbf{y}(t)\|^2).$$

Osservazione 7.8 Come nel caso discreto, l'equazione linearizzata nel punto di equilibrio, definisce il cosiddetto problema variazionale. In generale, quest'ultimo può essere definito rispetto ad una qualunque soluzione di riferimento.

⁵Ovvero, anche nel caso continuo, nel caso lineare uniforme asintotica stabilità ed esponenziale asintotica stabilità di un punto critico sono proprietà equivalenti.

Vale il seguente risultato, che si dimostra con argomenti simili a quelli utilizzati nel caso discreto.

Teorema 7.7 *Siano dati il problema (7.27) ed il problema lineare associato (7.23). Se:*

- $\|\mathbf{g}(t, \mathbf{y}(t))\| \leq L\|\mathbf{y}(t)\|$, con L sufficientemente “piccolo”,
- la soluzione nulla di (7.23) è uniformemente asintoticamente stabile,

allora la soluzione nulla di (7.27) è esponenzialmente asintoticamente stabile.

Inoltre, analogamente al caso discreto, si dimostra il seguente risultato.

Corollario 7.5 (Stabilità in prima approssimazione, caso continuo) *Se la soluzione nulla dell’equazione*

$$\mathbf{y}' = J_{\mathbf{f}}(t, \mathbf{0})\mathbf{y}, \quad t \geq t_0,$$

è uniformemente asintoticamente stabile e, con riferimento alla (7.27),

$$\lim_{\|\mathbf{y}\| \rightarrow 0} \frac{\|\mathbf{g}(t, \mathbf{y})\|}{\|\mathbf{y}\|} = 0$$

uniformemente rispetto a t , allora la stessa è esponenzialmente asintoticamente stabile per (7.21)–(7.22), purché $\|\mathbf{y}(t_0)\|$ sia sufficientemente piccola.

Il corollario successivo, riguarda il sistema

$$\mathbf{y}'(t) = A\mathbf{y}(t) + \mathbf{g}(t, \mathbf{y}(t)), \quad \mathbf{g}(t, \mathbf{0}) = \mathbf{0}, \quad t \geq t_0, \quad (7.28)$$

e si dimostra in modo analogo al caso discreto.

Corollario 7.6 (Teorema di Perron, versione continua) *Dato il sistema di equazioni (7.28), se $\sigma(A) \subset \mathbb{C}^-$ e, inoltre,*

$$\lim_{\|\mathbf{y}\| \rightarrow 0} \frac{\|\mathbf{g}(t, \mathbf{y})\|}{\|\mathbf{y}\|} = 0$$

uniformemente rispetto a t , allora la sua soluzione nulla è esponenzialmente asintoticamente stabile.

Osservazione 7.9 *Analogamente al caso discreto, se l’origine è instabile per la parte lineare (7.23), allora lo è anche per la (7.27).*

Osservazione 7.10 *Anche le questioni di stabilità totale si trattano in modo analogo a quanto visto nel discreto, e con risultati del tutto simili.*

Giustificazione dell’equazione test

Consideriamo ancora una volta l’equazione (7.17), avente l’origine come punto critico. Si ottiene, ponendo $\lambda = f'(0)$ e supponendo f sufficientemente regolare:

$$y' = \lambda y + g(y), \quad \text{con} \quad g(y) = O(y^2).$$

Pertanto, se l'origine è asintoticamente stabile per la parte lineare (cioè, se $\Re(\lambda) < 0$), allora essa sarà esponenzialmente asintoticamente stabile per (7.17). In altri termini, il Teorema di Perron giustifica l'utilizzo dell'equazione test

$$y' = \lambda y, \quad \Re(\lambda) < 0,$$

nell'intorno di un punto di equilibrio uniformemente asintoticamente stabile.

Ad esempio, per il problema (4.45), la linearizzazione nel punto di equilibrio $\bar{y} = 5$ fornisce l'equazione linearizzata (traslando \bar{y} nell'origine) $y' = -10y$, per cui l'origine è asintoticamente stabile.

Modello per la diagnosi del diabete mellito

Il diabete mellito è una patologia che si manifesta con una elevata concentrazione del glucosio nel sangue e nelle urine.⁶ La funzione del glucosio è quella di portare energia alle cellule. Esso è presente nel sangue con una concentrazione ottimale che indicheremo con \bar{G} . Tale concentrazione aumenta in seguito alla ingestione di cibo, o può diminuire in seguito ad un aumentato bisogno dell'organismo. A regolare il meccanismo di aumento o diminuzione della concentrazione del glucosio nel sangue vi sono diversi ormoni. Tra questi, si menzionano i seguenti:

Insulina. È secreta dalle cellule β del pancreas. Esso favorisce l'assorbimento del glucosio da parte delle cellule e di conseguenza la diminuzione della sua concentrazione.

Glucagone. È secreto dalle cellule α del pancreas. Un eccesso di glucosio nel sangue viene trasformato in glicogeno ed immagazzinato nel fegato. In caso di ipoglicemia, cioè bassa concentrazione di glucosio, il glucagone favorisce la trasformazione di glicogeno in glucosio e quindi favorisce l'aumento di concentrazione di quest'ultima sostanza.

Adrenalina. È prodotto dalle capsule surrenali e fa aumentare la concentrazione del glucosio nel sangue attivando un processo più rapido e più completo rispetto al glucagone, oltre che inibendo la produzione di insulina, mediante riduzione della funzionalità del pancreas. Questo processo si attiva in casi di elevata ipoglicemia o in situazioni di pericolo ed è finalizzata a rendere disponibile il glucosio da parte dei muscoli.

Cortisolo. Anche questo ormone è prodotto dalle ghiandole surrenali. Esso stimola la glucogenesi epatica, con un conseguente aumento della glicemia.

Il modello che descriviamo considera solo l'interazione glucosio-insulina. Sebbene modelli più completi siano stati formulati, nondimeno questo è assai utilizzato nella pratica clinica. Siano dunque $G(t)$ ed $H(t)$ le concentrazioni del glucosio e dell'insulina nel sangue e \bar{G}, \bar{H} le concentrazioni ottimali. Il modello è il seguente.

$$\begin{aligned} G' &= F_1(G, H), \\ H' &= F_2(G, H). \end{aligned}$$

Se la concentrazione delle due sostanze nel sangue è ai livelli ottimali, il meccanismo di regolazione non viene attivato, cioè si avrà:

$$\begin{aligned} F_1(\bar{G}, \bar{H}) &= 0, \\ F_2(\bar{G}, \bar{H}) &= 0, \end{aligned}$$

⁶Esiste anche una forma senile di diabete, dovuta a deficit metabolici dovuti all'età.

Questo significa che (\bar{G}, \bar{H}) è un punto di equilibrio per il sistema. Il meccanismo di regolazione sarà funzionante se una perturbazione nell'intorno di questo punto, tenderà ad essere smorzata fino a ritornare al punto di equilibrio. Il punto di equilibrio deve dunque essere asintoticamente stabile.⁷

Supporremo che le funzioni F_1 ed F_2 siano sviluppabili in serie di Taylor almeno fino al secondo termine. Introducendo le variabili

$$g(t) = G(t) - \bar{G}, \quad \text{e} \quad h(t) = H(t) - \bar{H},$$

il modello può scriversi nella forma

$$\begin{aligned} g' &= -m_1 g - m_2 h + \gamma_1(g, h), \\ h' &= m_3 g - m_4 h + \gamma_2(g, h), \end{aligned}$$

dove le funzioni $\gamma_1(g, h)$, $\gamma_2(g, h)$ rappresentano i termini di ordine superiore nello sviluppo in serie e $m_i > 0$, $i = 1, \dots, 4$. La parte lineare del modello è stata evidenziata in maniera che i segni dei coefficienti appaiano chiaramente indicati. Convincersi che i segni siano quelli indicati è abbastanza semplice. Infatti se, per esempio, al tempo $t = t_0$ si parte da una situazione in cui vi sia eccesso di glucosio ($g > 0$) ed $h = 0$, il meccanismo di autoregolazione sarà attivato e si avrà, a meno di termini infinitesimi di ordine superiore,

$$g'(t_0) \simeq \frac{\partial F_1}{\partial g}(0, 0) \equiv -m_1 < 0$$

e, inoltre,

$$h'(t_0) \simeq \frac{\partial F_2}{\partial g}(0, 0) \equiv m_3 > 0.$$

Il meccanismo farà, cioè, diminuire il glucosio e aumentare l'insulina. Similmente, se per $t = t_0$ si avesse $g = 0$ e $h > 0$, entrambe le derivate saranno negative. È facile verificare che gli autovalori della matrice

$$A = \begin{pmatrix} -m_1 & -m_2 \\ m_3 & -m_4 \end{pmatrix}$$

sono a parte reale negativa e quindi, per la parte lineare del sistema, l'origine è asintoticamente stabile. Dunque, per il Teorema di Perron, l'origine è asintoticamente stabile per il sistema completo.

Questo modello può essere utilizzato per la diagnosi del diabete mellito. Infatti, se il ritorno al livello di equilibrio è troppo lento, questo si configura come una condizione patologica. Il test per rilevare la correttezza del meccanismo di autoregolazione, si basa sulla verifica della concentrazione del solo glucosio (la cui misura si fa facilmente ed è poco dispendiosa).⁸ Per questo motivo, ricercheremo un'espressione matematica che contenga solo la variabile G . Derivando la prima delle due equazioni rispetto al tempo, e tenendo conto della seconda, si ha, infatti⁹

$$g'' = -m_1 g' - m_2(m_3 g - m_4 h).$$

Ricavando h dalla prima e sostituendo, si ha

$$g'' + (m_1 + m_4)g' + (m_1 m_4 + m_2 m_3)g = 0.$$

⁷Diversamente, si andrebbe in coma per iperglicemia o ipoglicemia.

⁸Esistono addirittura misuratori portatili della glicemia, che utilizzano una goccia di sangue prelevata da un dito.

⁹Al solito, si considera la sola parte lineare del modello.

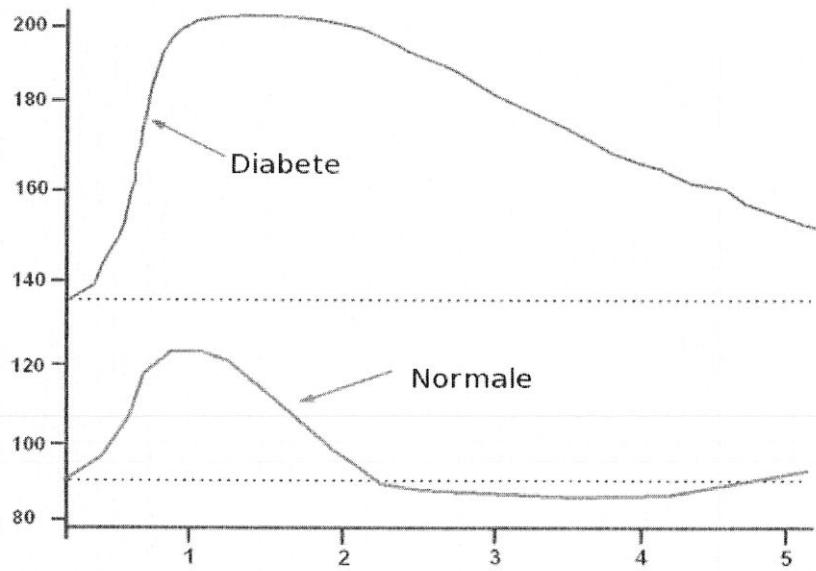


Figura 7.1: Esempi di curva da carico normale (in basso) e con diabete mellito (in alto).

Posti

$$\beta^2 = m_1 m_4 + m_2 m_3, \quad 2\alpha = m_1 + m_4,$$

si ottiene

$$g'' + 2\alpha g' + \beta^2 g = 0.$$

La soluzione, ponendo $t_0 = 0$, è

$$g(t) = A e^{-\alpha t} \cos(\omega t + \phi), \quad \text{dove} \quad \omega^2 = \beta^2 - \alpha^2 > 0.$$

Tornando alla vecchia variabile, si ha

$$G(t) = \bar{G} + A e^{-\alpha t} \cos(\omega t + \phi).$$

Le quantità α , A , ω , ϕ devono essere determinate in base ai risultati dei test. Poiché la prima misura della glicemia si esegue dopo una nottata di digiuno, si assume che essa fornisca il valore di equilibrio \bar{G} . Successivamente, viene fatta assumere al paziente una certa quantità di glucosio (dell'acqua zuccherata, con una quantità di zucchero proporzionale al peso corporeo) e si eseguono m misurazioni G_i , della glicemia ad istanti prefissati t_i , $i = 1, \dots, m$. I parametri incogniti si possono quindi ricavare con il metodo dei minimi quadrati. Si definisce cioè la funzione

$$F(\alpha, \omega, A, \phi) = \sum_{i=1}^m (G_i - \bar{G} - A e^{-\alpha t_i} \cos(\omega t_i + \phi))^2$$

e se ne cerca il minimo rispetto alle sue variabili. In particolare ω è il parametro meno sensibile agli errori sperimentali. Il diabete è diagnosticato quando $\bar{G} > 126 \text{ mg/dl}$ e

$$\frac{T}{2} = \frac{\pi}{\omega} > 3.5 \quad \text{ore.}$$

A titolo di esempio, due curve da carico sono raffigurate in Figura 7.1. Sull'asse orizzontale, il tempo è misurato in *ore*, mentre su quello verticale la glicemia è misurata in *mg/dl*. La curva inferiore è normale e, come si può vedere, il ritorno all'equilibrio si ha in poco più di 2 ore. In quella superiore, a parte il valore molto alto della glicemia iniziale, il ritorno all'equilibrio si ottiene dopo oltre 5 ore.

7.3 Il caso di stabilità marginale

Il Teorema di Perron, nella sua versioni continue e discrete, permettono di studiare la stabilità del punto di equilibrio nel caso in cui tale punto di equilibrio sia asintoticamente stabile (o instabile) per il sistema linearizzato. Nel caso in cui per quest'ultimo il punto di equilibrio sia solo stabile, nulla si può dire sul comportamento del sistema completo, come dimostrato dal seguente esempio. Sia, per $\alpha \in \mathbb{R}$,

$$\begin{aligned} x' &= -y + \alpha x \sqrt{x^2 + y^2} \\ y' &= x + \alpha y \sqrt{x^2 + y^2}. \end{aligned} \tag{7.29}$$

La parte lineare,

$$x' = -y, \quad y' = x,$$

è marginalmente stabile, poiché gli autovalori della matrice

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

sono $\pm i$. Se consideriamo il problema completo, tuttavia, passando in coordinate polari,

$$x = \rho \cos \theta, \quad y = \rho \sin \theta,$$

si ottiene

$$\begin{aligned} \rho' \cos \theta - \rho \sin \theta \theta' &= -\rho \sin \theta + \alpha \rho^2 \cos \theta, \\ \rho' \sin \theta + \rho \cos \theta \theta' &= \rho \cos \theta + \alpha \rho^2 \sin \theta. \end{aligned}$$

Moltiplicando la prima equazione per $\cos \theta$, la seconda per $\sin \theta$ e sommando, si ottiene:

$$\rho' = \alpha \rho^2. \tag{7.30}$$

Similmente, moltiplicando la prima equazione per $-\sin \theta$, la seconda per $\cos \theta$ e sommando, si ottiene:

$$\theta' = 1.$$

La soluzione di quest'ultima equazione, ponendo $\theta(0) = \theta_0$, è

$$\theta(t) = \theta_0 + t, \quad t \geq 0.$$

Pertanto, le traiettorie si avvolgeranno attorno all'origine in senso antiorario. Assumendo, per la (7.30) la condizione iniziale $\rho(0) = \rho_0 > 0$, la soluzione si verifica essere data da

$$\rho(t) = \frac{\rho_0}{1 - \alpha \rho_0 t}, \quad t \geq 0.$$

Pertanto, se $\alpha < 0$ si ottiene un fuoco stabile mentre, nel caso $\alpha > 0$, per $t \rightarrow (\alpha\rho)^{-1}$ la traiettoria diviene illimitata.¹⁰ Poiché la parte lineare non dipende dal parametro α , si comprende come le sue proprietà non diano informazione nel caso di stabilità marginale.

7.4 Funzioni di Lyapunov

Abbiamo già evidenziato che lo studio del comportamento qualitativo delle soluzioni di un sistema nonlineare nell'intorno di un punto di equilibrio mediante l'analisi della parte lineare non è sempre possibile. Anche la tecnica che stiamo per descrivere non è applicabile in tutti i casi. Tuttavia, ove utilizzabile, essa risulta essere molto semplice e potente. Essa fu proposta dal matematico A.M. Lyapunov verso la fine del diciannovesimo secolo. Si consideri, per semplicità, il problema autonomo¹¹

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}), \quad \mathbf{y}(0) = \mathbf{y}_0 \in \mathbb{R}^m, \quad (7.31)$$

che supponiamo ammetta un'unica soluzione per $t \geq 0$, ed ammetta l'origine come unico punto critico:

$$\mathbf{f}(\mathbf{y}) = \mathbf{0} \quad \Rightarrow \quad \mathbf{y} = \mathbf{0}.$$

Supponiamo, ora, che esista una funzione differenziabile e con derivate continue $V : \mathbf{y} \in \mathbb{R}^m \rightarrow \mathbb{R}$ tale che:

$$\begin{aligned} V(\mathbf{y}) &\geq 0, \\ V(\mathbf{y}) &= 0 \quad \Rightarrow \quad \mathbf{y} = \mathbf{0}, \\ V(\mathbf{y}) &\rightarrow \infty \quad \text{per} \quad \mathbf{y} \rightarrow \infty. \end{aligned} \quad (7.32)$$

Inoltre, in un opportuno intorno dell'origine, supporremo inoltre che valga:

$$\nabla V(\mathbf{y})^T \mathbf{f}(\mathbf{y}) \leq 0, \quad \nabla V(\mathbf{y})^T \mathbf{f}(\mathbf{y}) = 0 \quad \Rightarrow \quad \mathbf{y} = \mathbf{0}. \quad (7.33)$$

Osserviamo che, essendo $V(\mathbf{y})$ positiva, ed avendosi $V(\mathbf{y}) \rightarrow \infty$, per $\mathbf{y} \rightarrow \infty$, le regioni

$$\Gamma(c) = \{\mathbf{y} \in \mathbb{R}^m : V(\mathbf{y}) \leq c\}, \quad c \geq 0, \quad (7.34)$$

saranno costituite da componenti connesse chiuse e limitate, di cui una, sia essa $\mathcal{R}(c)$, conterrà l'origine. Infatti, se $c_1 > c_2$, si avrà, evidentemente, $\Gamma(c_2) \subseteq \Gamma(c_1)$ e, inoltre, $\Gamma(0) \equiv \mathcal{R}(0) = \{\mathbf{0}\}$. È interessante osservare che, se $\exists c > 0$ per cui

$$\nabla V(\mathbf{y})^T \mathbf{f}(\mathbf{y}) < 0, \quad \forall \mathbf{y} \in \partial \mathcal{R}(c),$$

questo significa che traiettorie del sistema dinamico indotto da (7.31) che originano da punti su $\partial \mathcal{R}(c)$, saranno dirette verso l'interno di questa regione. Infatti, $\mathbf{y}' = \mathbf{f}(\mathbf{y})$ per le soluzioni di (7.31), mentre $\nabla V(\mathbf{y})$ è un vettore ortogonale alla superficie di livello $\{V(\mathbf{y}) = c\} \equiv \partial \mathcal{R}(c)$, e diretto verso valori crescenti di c . Pertanto, una volta che una traiettoria entra in $\mathcal{R}(c)$, non ne uscirà più. Per questo motivo, diremo che $\mathcal{R}(c)$ è una *regione invariante* per il sistema dinamico. Con questa premessa, è possibile dimostrare il seguente risultato, che è una versione semplificata del risultato originale di Lyapunov.

¹⁰Chiaramente, per $\alpha = 0$ si ha un centro.

¹¹La trattazione del caso non autonomo risulta essere più complessa.

Teorema 7.8 (Teorema di Lyapunov) Se $\exists c > 0$ tale che, per ogni $\mathbf{y} \in \mathcal{R}(c)$ valga (7.33), allora l'origine è asintoticamente stabile per (7.31).

Dimostrazione. Sia $\mathbf{y}(t)$ la traiettoria soluzione di (7.31), dove \mathbf{y}_0 è un generico punto di $\mathcal{R}(c)$. Per quanto detto innanti, varrà:

$$\mathbf{y}(t) \in \mathcal{R}(c), \quad \forall t \geq 0.$$

Evidentemente, essendo

$$V(\mathbf{y}(t)) \geq 0, \quad \text{e} \quad \frac{d}{dt}V(\mathbf{y}(t)) = \nabla V(\mathbf{y}(t))^T \mathbf{f}(\mathbf{y}(t)) \leq 0, \quad t \geq 0,$$

$V(\mathbf{y}(t))$ ammetterà un limite, sia esso $V_\infty \geq 0$ per $t \rightarrow \infty$. Se questo limite fosse 0, la tesi sarebbe dimostrata, perché questo implicherebbe che $\mathbf{y}(t) \rightarrow \mathbf{0}$, per $t \rightarrow \infty$. Se per assurdo così non fosse, allora $\exists \bar{\mathbf{y}} \in \mathcal{R}(c)$, $\bar{\mathbf{y}} \neq \mathbf{0}$, tale che

$$V(\bar{\mathbf{y}}) = V_\infty, \quad \text{e} \quad \mathbf{y}(t) \rightarrow \bar{\mathbf{y}}, \quad \text{per } t \rightarrow \infty.$$

Considerando la traiettoria $\mathbf{y}(t; 0, \bar{\mathbf{y}})$, soluzione di (7.31), che parte da $\bar{\mathbf{y}}$, si avrebbe dunque

$$V(\mathbf{y}(t; 0, \bar{\mathbf{y}})) \equiv V_\infty > 0, \quad \forall t \geq 0,$$

ma questo è assurdo, poiché, essendo $\bar{\mathbf{y}} \neq \mathbf{0}$, risulta:

$$\frac{d}{dt}V(\mathbf{y}(0; 0, \bar{\mathbf{y}})) = \nabla V(\bar{\mathbf{y}})^T \mathbf{f}(\bar{\mathbf{y}}) < 0. \square$$

Definizione 7.4 Una funzione che soddisfi le (7.32)-(7.33), si dice funzione di Lyapunov per (7.31).

Argomenti del tutto analoghi possono essere utilizzati nel caso del sistema di equazioni alle differenze del primo ordine:

$$\mathbf{y}_{n+1} = \mathbf{f}(\mathbf{y}_n), \quad \mathbf{y}_{n_0} \in \mathbb{R}^m \quad \text{assegnato.} \quad (7.35)$$

Supporremo, al solito, che l'origine sia l'unico punto critico per (7.35):

$$\mathbf{f}(\mathbf{y}) = \mathbf{0} \quad \Rightarrow \quad \mathbf{y} = \mathbf{0}.$$

Definizione 7.5 Se esiste una funzione derivabile e con derivate continue $V : \mathbf{y} \in \mathbb{R}^m \rightarrow \mathbb{R}$, soddisfacente (7.32) e, per ogni \mathbf{y} appartenente ad un opportuno intorno dell'origine,

$$\Delta V(\mathbf{y}) \equiv V(\mathbf{f}(\mathbf{y})) - V(\mathbf{y}) \leq 0, \quad \Delta V(\mathbf{y}) = 0 \quad \Rightarrow \quad \mathbf{y} = \mathbf{0}, \quad (7.36)$$

diremo che V è una funzione di Lyapunov per (7.35).

Definendo, come nel caso continuo, le regioni (7.34), ed indicando con $\mathcal{R}(c)$ la componente connessa contenente l'origine, vale il seguente risultato, che è la versione discreta del Teorema 7.8, la cui dimostrazione si ottiene con argomenti simili.

Teorema 7.9 (Teorema di Lyapunov, versione discreta) Se $\exists c > 0$ tale che, per ogni $\mathbf{y} \in \mathcal{R}(c)$ valga (7.36), allora l'origine è asintoticamente stabile per (7.35).

Esempio.

Consideriamo nuovamente il sistema di equazioni (7.29), per il quale l'origine è solo marginalmente stabile per il problema linearizzato, e definiamo la funzione:

$$V(x, y) = x^2 + y^2.$$

Si verifica facilmente che essa soddisfa le condizioni (7.32). Inoltre, dalla (7.29) si ottiene:

$$\nabla V(x, y)^T \begin{pmatrix} x' \\ y' \end{pmatrix} = \alpha (x^2 + y^2)^{\frac{3}{2}} < 0 \quad \Leftrightarrow \quad \alpha < 0 \quad \text{e} \quad (x, y)^T \neq (0, 0)^T.$$

Pertanto, per $\alpha < 0$, si conclude che sono soddisfatte anche le (7.33). Di conseguenza, V è una funzione di Lyapunov per (7.29), e possiamo concludere che l'origine è asintoticamente stabile quando $\alpha < 0$, che è esattamente quello che avevamo visto in precedenza.

7.5 Ancora sul concetto di *stiffness*

Approfondiamo ulteriormente la nozione di *stiffness* introdotta in Sezione 6.5. La nostra analisi generalizza quella fatta per il caso lineare al caso nonlineare. Dato, quindi, il problema nonlineare (che supporremo per semplicità autonomo)

$$\mathbf{y}'(t) = \mathbf{f}(\mathbf{y}(t)), \quad t \in [t_0, T], \quad \mathbf{y}(t_0) = \mathbf{y}_0, \quad (7.37)$$

perturbando la condizione iniziale con un vettore $\delta\mathbf{y}_0 \equiv \boldsymbol{\eta}$ “infinitesimo”, si ha che la differenza tra la soluzione originaria, e quella del problema perturbato, sia essa $\delta\mathbf{y}(t)$, soddisferà il *problema variazionale*

$$\delta\mathbf{y}'(t) = J_{\mathbf{f}}(\mathbf{y}(t))\delta\mathbf{y}(t), \quad t \in [t_0, T], \quad \delta\mathbf{y}(0) = \boldsymbol{\eta}, \quad (7.38)$$

in cui

$$J_{\mathbf{f}}(\mathbf{y}(t)) \equiv A(t),$$

è la matrice Jacobiana di $\mathbf{f}(\mathbf{y}(t))$. Se la perturbazione non è infinitesima, (7.38) costituisce la parte lineare del problema linearizzato lungo $\mathbf{y}(t)$ e, utilizzando argomenti analoghi a quelli visti per il Teorema 7.7, l'asintotica stabilità dell'origine per la soluzione di (7.38) ne implica l'asintotica stabilità per il problema completo. Considerando, quindi, la corrispondente matrice fondamentale:

$$\Phi'(t, t_0) = A(t)\Phi(t, t_0), \quad \Phi(t_0, t_0) = I,$$

se ne deduce che la soluzione del problema (7.38) è data da

$$\delta\mathbf{y}(t) = \Phi(t, t_0)\boldsymbol{\eta}, \quad t \in [t_0, T].$$

Come nel caso lineare, consideriamo le seguenti due norme dell'errore:

$$\|\delta\mathbf{y}\|_{\infty} = \max_{t_0 \leq t \leq T} \|\delta\mathbf{y}(t)\| = \frac{\max_{t_0 \leq t \leq T} \|\delta\mathbf{y}(t)\|}{\|\boldsymbol{\eta}\|} \|\boldsymbol{\eta}\| \equiv \kappa(\boldsymbol{\eta})\|\boldsymbol{\eta}\|,$$

$$\|\delta\mathbf{y}\|_1 = \frac{1}{T - t_0} \int_{t_0 \leq t \leq T} \|\delta\mathbf{y}(t)\| dt = \frac{\int_{t_0 \leq t \leq T} \|\delta\mathbf{y}(t)\| dt}{(T - t_0)\|\boldsymbol{\eta}\|} \|\boldsymbol{\eta}\| \equiv \gamma(\boldsymbol{\eta})\|\boldsymbol{\eta}\|,$$

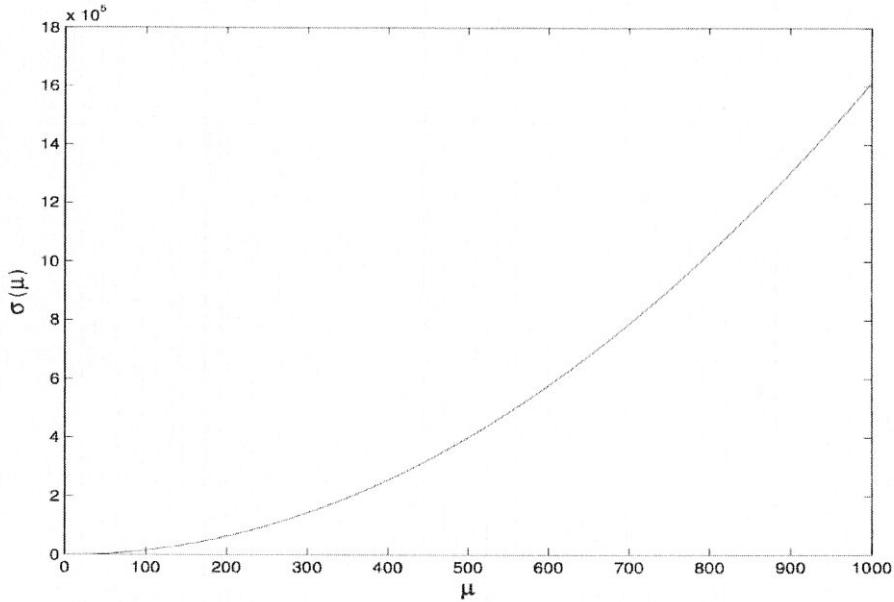


Figura 7.2: *stiffness* del problema di van der Pol (7.40) al crescere di μ .

che misurano, come visto in Sezione 6.5, l'errore massimo e l'errore medio dovuto alla perturbazione iniziale η . Possiamo quindi definire il *rapporto di stiffness*

$$\sigma = \sup_{\eta} \sigma(\eta), \quad \text{con} \quad \sigma(\eta) = \frac{\kappa(\eta)}{\gamma(\eta)}. \quad (7.39)$$

Definizione 7.6 *Diremo che il problema (7.37) è stiff se $\sigma \gg 1$.*

Osservazione 7.11 È opportuno ribadire che il concetto di stiffness si definisce per un problema, e non per un'equazione. Infatti, la stessa equazione definisce problemi diversi, a seconda del punto iniziale considerato. Al contrario del caso lineare, in cui la stiffness non dipende dalla condizione iniziale, per equazioni nonlineari, alcune condizioni iniziali possono dare origine a problemi stiff, ed altri originare problemi non stiff. La ragione di questo deriva dal fatto che il problema variazionale (7.38) è definito lungo la soluzione di riferimento e, quindi, dipende dal punto iniziale in (7.37).

7.5.1 Alcuni esempi

Riportiamo, nel seguito, alcuni esempi che illustrano adeguatamente la precedente definizione. In tutti i casi, il problema dipende da un parametro scalare positivo che rende il problema “stiff” quando questo tende a 0 o a ∞ :

- il problema di Van der Pol, $\mu \rightarrow \infty$;
- il problema di Robertson, $T \rightarrow \infty$;
- il problema di Kreiss e sua modifica, $\varepsilon \rightarrow 0$.

Il problema di van der Pol

Consideriamo il seguente problema, dipendente da un parametro $\mu > 0$:

$$\begin{aligned} x' &= y, & t \in [0, 2\mu], \\ y' &= -x + \mu y(1 - x^2), & x(0) = 2, \quad y(0) = 0, \end{aligned} \quad (7.40)$$

la cui soluzione descrive un'orbita chiusa attrattiva¹² nel piano delle fasi (il periodo della soluzione risulta essere $\simeq 2\mu$, per $\mu \gg 1$.) In Figura 7.2, grafichiamo il rapporto di *stiffness* (7.39) al crescere di μ . Da essa si evince come il problema diventi più *stiff* al crescere del parametro μ . Questo è infatti noto dalla pratica computazionale, in cui si sperimenta una crescente difficoltà nella risoluzione numerica del problema, per il quale i metodi esplicativi sono del tutto inefficaci, anche per modesti valori del parametro μ .

Il problema di Robertson

Questo problema, molto noto nella letteratura del settore, è dato da:

$$\begin{aligned} y'_1 &= -0.04y_1 + 10^4 y_2 y_3, \\ y'_2 &= 0.04y_1 - 10^4 y_2 y_3 - 3 \cdot 10^7 y_2^2, \\ y'_3 &= 3 \cdot 10^7 y_2^2, & t \in [0, T], \\ y_1(0) &= 1, \quad y_2(0) = y_3(0) = 0. \end{aligned} \quad (7.41)$$

In Figura 7.3, è graficato il rapporto di *stiffness* (7.39) in funzione dell'ampiezza T dell'intervallo di integrazione. In questo caso, il rapporto di *stiffness* graficato è stato ottenuto considerando una perturbazione della forma $(0, \varepsilon, -\varepsilon)^T$, con $\varepsilon \approx 0$. Anche in questo caso, si vede che la *stiffness* del problema aumenta al crescere di T , confermando la difficoltà che si incontra, nella pratica computazionale, quando si integra questo problema su intervalli molto ampi.

Esercizio 7.1 Dimostrare che, per il problema (7.41), la quantità $y_1(t) + y_2(t) + y_3(t)$ è costante per $t \geq 0$. Essa costituisce, quindi, un invariante lineare per il corrispondente sistema dinamico. Dimostrare che ogni metodo consistente preserva questo invariante nella soluzione discreta.

Il problema di Kreiss

Questo problema, denominato *problema di Kreiss*, è lineare ma non autonomo:¹³

$$\mathbf{y}'(t) = Q^T(t) \begin{pmatrix} -1 \\ -\varepsilon^{-1} \end{pmatrix} Q(t) \mathbf{y}(t), \quad t \in [0, 4\pi], \quad \mathbf{y}(0) = \mathbf{y}_0, \quad (7.42)$$

in cui \mathbf{y}_0 è dato¹⁴ e

$$Q(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

Considerando una perturbazione iniziale della forma $(-\varepsilon, 1)^T$, si ottiene il rapporto di *stiffness* in funzione di ε raffigurato in Figura 7.4, che dimostra come esso si comporti come $O(\varepsilon^{-1})$, quando $\varepsilon \rightarrow 0$.

¹²Si parla, in questo caso, di *ciclo limite*.

¹³Ovvero, la matrice di trasformazione dipende dal t .

¹⁴Il problema è lineare e, quindi, la sua *stiffness* sarà indipendente dalla condizione iniziale.

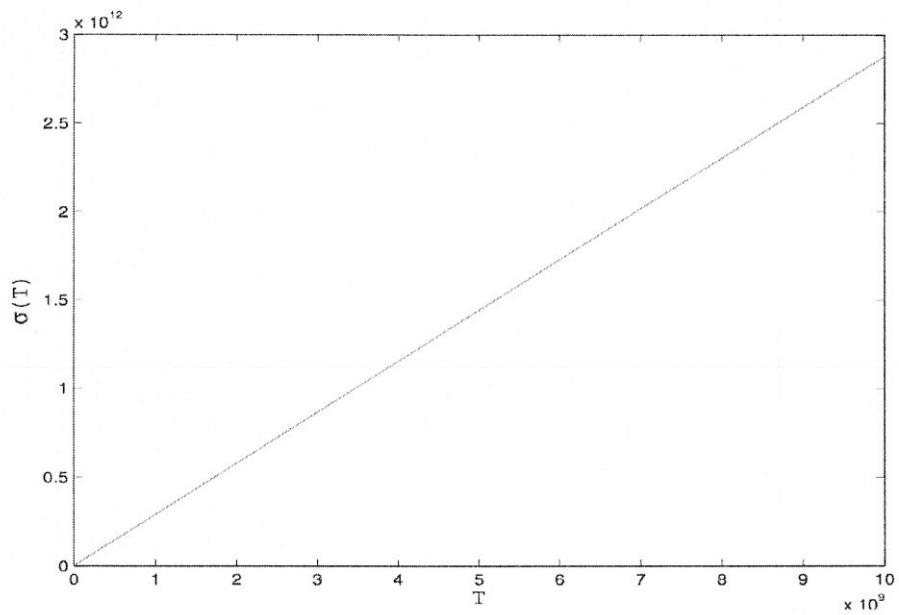


Figura 7.3: *stiffness* del problema di Robertson (7.41) in funzione dell'ampiezza T dell'intervalllo di integrazione.

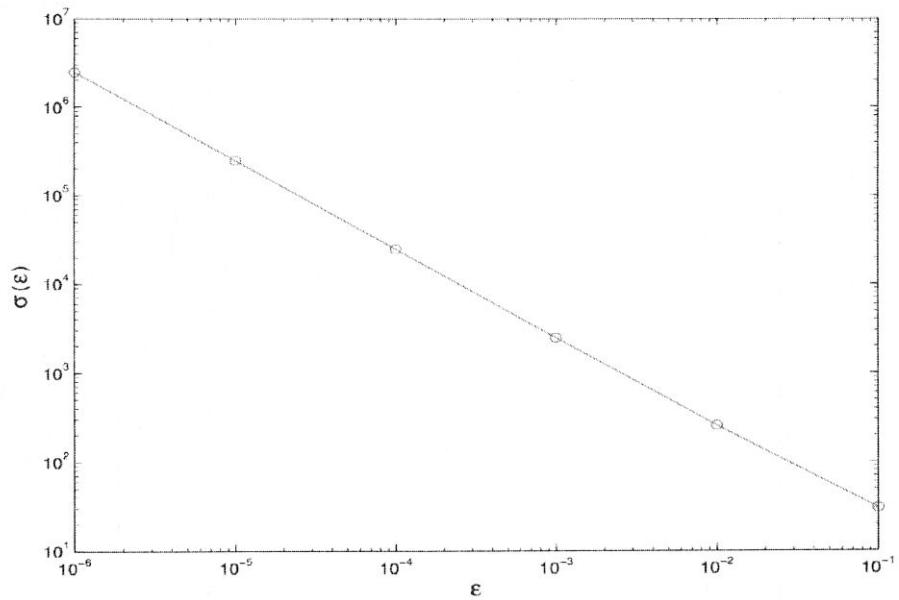
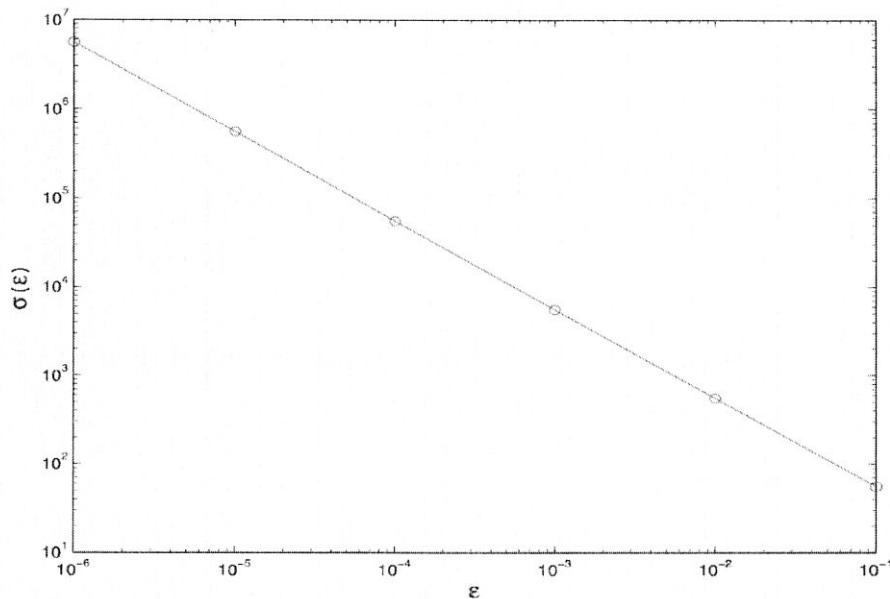


Figura 7.4: *stiffness* del problema (7.42) in funzione di ε .

Figura 7.5: *stiffness* del problema (7.43) in funzione di ε .

Consideriamo, ora, la seguente modifica del problema (7.42),

$$\mathbf{y}'(t) = Q_\varepsilon^{-1}(t)P^{-1} \begin{pmatrix} -1 \\ -\varepsilon^{-1} \end{pmatrix} PQ_\varepsilon(t)\mathbf{y}(t), \quad t \in [0, 4\pi], \quad \mathbf{y}(0) = \mathbf{y}_0, \quad (7.43)$$

in cui:

$$Q_\varepsilon(t) = \begin{pmatrix} 1 & \varepsilon \\ e^{\sin t} & e^{\cos t} \end{pmatrix}, \quad P = \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Anche ora, considerando una perturbazione iniziale della forma $(-\varepsilon, 1)^T$, si ottiene il rapporto di *stiffness* in funzione di ε raffigurato in Figura 7.5, che dimostra come anch'esso si comporti come $O(\varepsilon^{-1})$, per $\varepsilon \rightarrow 0$.

In entrambi i casi, i risultati riguardo alla stiffness rispecchiano la crescente difficoltà nella risoluzione numerica dei due problemi, quando $\varepsilon \rightarrow 0$.

Una osservazione

È interessante sottolineare che, in tutti i precedenti esempi, numericamente si è osservato che la perturbazione $\boldsymbol{\eta}$ che massimizza il rapporto $\sigma(\boldsymbol{\eta})$ in (7.39) è ottenuta considerando un vettore proporzionale all'autovettore dominante dello Jacobiano $J_{\mathbf{f}}(\mathbf{y}(t))$ per $t \approx t_0$ (vedi (7.38)). Osserviamo che, nel caso di un problema lineare autonomo, questa scelta è proprio quella che “attiva” l'autovalore dominante, sia esso λ_{\max} . Pertanto, nel caso in cui la matrice sia diagonalizzabile, il rapporto di *stiffness* (7.39) si riduce all'espressione

$$\sigma = |\lambda_{\max}|(T - t_0),$$

che coincide con la (6.33) vista in Sezione 6.5.