Corso di Analisi Numerica - AN410

Parte 3: metodi iterativi per sistemi lineari ed equazioni nonlineari

Roberto Ferretti

ROMA TRE
UNIVERSITÀ DEGLI STUDI

DMF

- Filosofia generale dei metodi iterativi
- Metodi iterativi per Sistemi Lineari
- Convergenza dei metodi iterativi per Sistemi Lineari
- Metodi per equazioni scalari: la bisezione
- Metodi per equazioni scalari: iterazioni di punto fisso
- Metodi per equazioni scalari: metodo di Newton e varianti

Filosofia generale dei metodi iterativi

• Sono metodi che non forniscono la soluzione esatta in un numero finito di operazioni, anche supponendo di lavorare in aritmetica esatta.

• Nei metodi iterativi, la soluzione viene cercata tramite una successione di soluzioni approssimate x_k

• Nella maggioranza dei casi, tale successione si costruisce ponendo il sistema $Ax = b$ o l'equazione $f(x) = 0$ in forma di punto fisso:

$$x = T(x)$$

Dal Teorema delle contrazioni è noto che se esiste un insieme invariante U per la trasformazione $T(\cdot)$ e se $T(\cdot)$ è una contrazione su U, allora preso $x_0 \in U$ e definita la successione

$$x_{k+1} = T(x_k) \tag{1}$$

si ha $x_k \to \bar{x}$, dove $\bar{x}$ è l'unica soluzione in U dell'equazione di punto fisso $x = T(x)$.

• Questo permette di definire costruttivamente una successione convergente ad $\bar{x}$, a patto di avere una contrazione a secondo membro di (1).

Poiché è generalmente più facile localizzare approssimativamente una soluzione piuttosto che trovare un insieme invariante per la trasformazione T, può essere conveniente sostituire l'ipotesi di invarianza con l'ipotesi che T sia una contrazione nell'intorno della soluzione: infatti in questo caso se x_k è in un intorno sferico di $\bar{x}$ ed L_T è la costante di Lipschitz di T, allora

$$\|x_{k+1} - \bar{x}\| = \|T(x_k) - T(\bar{x})\| \leq L_T \|x_k - \bar{x}\| < \|x_k - \bar{x}\|$$

ed anche x_{k+1} è nello stesso intorno (che è quindi l'insieme invariante cercato)

La costante di Lipschitz di una trasformazione $T : \mathbb{R}^n \to \mathbb{R}^n$ può essere determinata come

$$L_T = \sup_x \|J_T(x)\|$$

dove $J_T(x)$ è la matrice jacobiana

$$J_T = \begin{pmatrix} \frac{\partial T_1}{\partial x_1} & \cdots & \frac{\partial T_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial T_n}{\partial x_1} & \cdots & \frac{\partial T_n}{\partial x_n} \end{pmatrix}$$

- Nel caso di trasformazioni affini, la jacobiana è costante
- Se $n = 1$, la norma della jacobiana equivale al modulo della derivata

L'errore $\|x_k - \bar{x}\|$ può essere maggiorato in due modi:

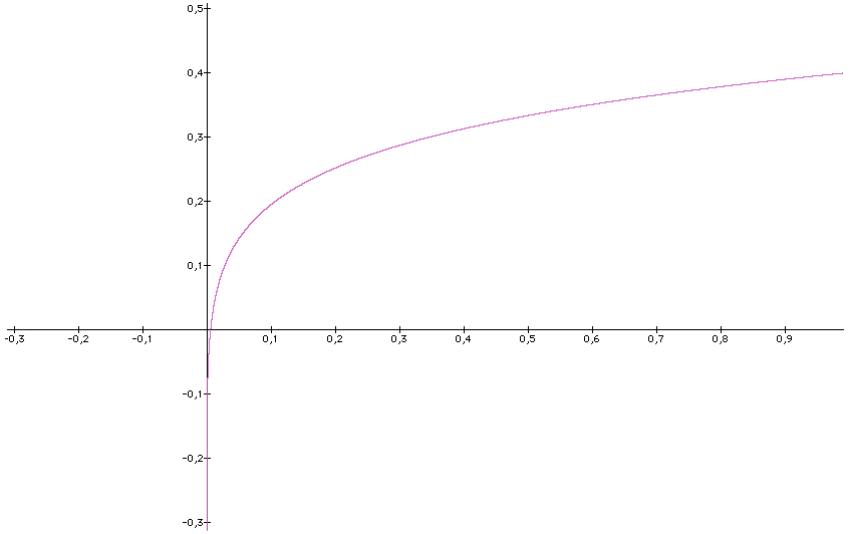
- Se è possibile maggiorare l'errore iniziale $\|x_0 - \bar{x}\|$, allora

$$\|x_k - \bar{x}\| \leq L_T \|x_{k-1} - \bar{x}\| \leq \cdots \leq L_T^k \|x_0 - \bar{x}\|$$

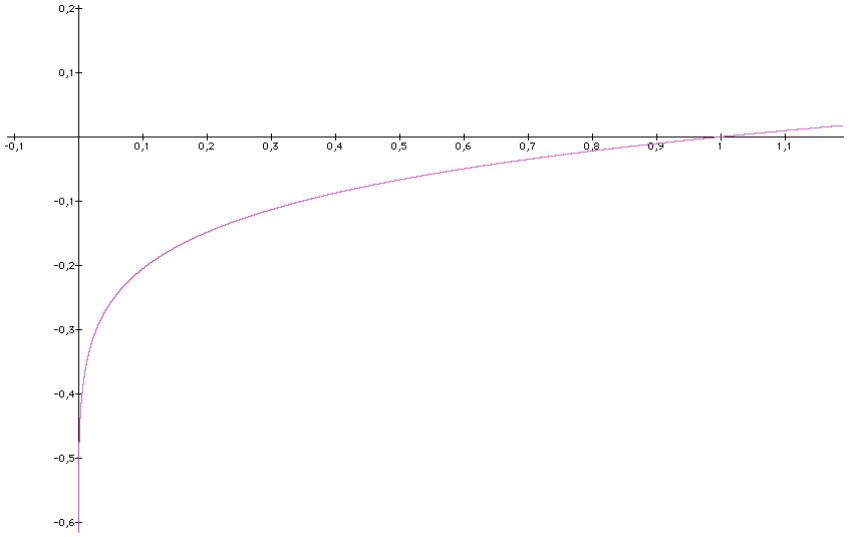
- Altrimenti, dall'aggiornamento $\|x_k - x_{k-1}\|$ si ha la maggiorazione

$$\|x_k - \bar{x}\| \leq \|x_{k+1} - x_k\| + \|x_{k+2} - x_{k+1}\| + \cdots \leq$$
$$\leq L_T \|x_k - x_{k-1}\| + L_T^2 \|x_k - x_{k-1}\| + \cdots \leq \frac{L_T}{1 - L_T} \|x_k - x_{k-1}\|$$

- Nel definire un corretto criterio di arresto delle iterazioni, si deve tenere conto anche dell'entità del residuo ($\|Ax - b\|$ o $|f(x)|$)



$|x_k - \bar{x}|$ piccolo, $|f(x)|$ grande



$|x_k - \bar{x}|$ grande, $|f(x)|$ piccolo

La maggiorazione dell'errore al passo k–esimo in funzione di quello al passo $(k-1)$–esimo,

$$\|x_k - \bar{x}\| \leq L_T \|x_{k-1} - \bar{x}\|$$

in qualche caso può essere migliorata: si definisce ordine di convergenza di un metodo il più grande esponente γ tale che

$$\|x_k - \bar{x}\| \leq C \|x_{k-1} - \bar{x}\|^{\gamma}$$

• Nei metodi basati su una contrazione tipicamente $C = L_T$ e $\gamma = 1$, ma l'interesse è verso metodi con $\gamma > 1$ (un ordine più alto implica di regola un incremento della velocità di convergenza del metodo)

indice

Metodi iterativi per Sistemi Lineari

La forma generale dei metodi iterativi per sistemi lineari si basa su una trasformazione $T : \mathbb{R}^n \to \mathbb{R}^n$ affine:

$$x_{k+1} = T(x_k) = Bx_k + c$$

• Come è facile verificare, la matrice jacobiana della trasformazione è la matrice B (detta anche matrice di iterazione)

• In questo caso, può essere data una condizione di convergenza necessaria e sufficiente, più generale della condizione di contrattività $\|B\| < 1$

Un primo modo di costruire la trasformazione T è dato dal metodo di Jacobi in cui si esplicita la variabile i–sima dalla equazione i–sima:

$$\begin{cases} x_1 = \dfrac{1}{a_{11}} \left(b_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n \right) \\ \quad \vdots \\ x_n = \dfrac{1}{a_{nn}} \left(b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{n,n-1}x_{n-1} \right) \end{cases} \tag{2}$$

• Questo metodo suppone che gli elementi sulla diagonale di A siano non nulli, eventualmente a meno di permutazioni delle righe

• La convergenza può dipendere dall'ordinamento delle righe (o, in altre parole, dall'ordine con cui si esplicitano le variabili)

Il metodo di Jacobi consiste quindi nella iterazione:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j \neq i} a_{ij} x_j^{(k)} \right) \qquad (i = 1, \ldots, n)$$

(in cui il numero di iterazione è stato indicato all'apice, entro parentesi)

• Il metodo richiede di memorizzare due vettori separati $x^{(k)}$ ed $x^{(k+1)}$ (infatti, il vettore non può essere aggiornato fino a che non siano state calcolate tutte le componenti)

Una variante del metodo di Jacobi è il metodo di Gauss–Seidel, in cui le variabili aggiornate vengono utilizzate immediatamente, ottenendo quindi la iterazione:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j<i} a_{ij} x_j^{(k+1)} - \sum_{j>i} a_{ij} x_j^{(k)} \right) \qquad (i = 1, \ldots, n) \quad (3)$$

• Il metodo di Gauss–Seidel richiede quindi di memorizzare un solo vettore che sarà riempito in parte dalla approssimazione $x^{(k)}$ ed in parte da $x^{(k+1)}$

Una ulteriore variante del metodo di Gauss–Seidel è il metodo del sovrarilassamento o SOR (Successive Over Relaxation), in cui viene introdotto un parametro ω detto parametro di rilassamento con l'intento di accelerare la convergenza, ponendo lo schema nella forma:

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \omega x_{i,GS}^{(k+1)} \qquad (i = 1, \ldots, n)$$

in cui $x_{i,GS}^{(k+1)}$ è il secondo membro di (3).

• In genere per accelerare la convergenza si sceglie $\omega > 1$, ma il valore ottimale è noto solo per alcune classi di matrici

Un'altra possibilità è di aggiornare l'approssimazione $x^{(k)}$ con un multiplo del residuo del sistema. In questo modo si ottiene il metodo di Richardson:

$$x^{(k+1)} = x^{(k)} - \beta \left(Ax^{(k)} - b \right)$$

in cui β è una costante positiva.

• Tale metodo va anche sotto il nome di *metodo del gradiente*, perché se A è definita positiva una iterazione equivale ad uno spostamento nella direzione di $-\nabla f \left(x^{(k)} \right)$, con $f(x) = \frac{1}{2} x^t Ax - b^t x$. Si cerca cioè di ottenere una successione decrescente di valori di f per convergere al minimo (che è la soluzione del sistema)

Complessità dei metodi iterativi:

• Il calcolo di $T(x_k)$ richiede una somma ed un prodotto per ogni elemento non nullo di A. Il numero di operazioni è quindi

$$\begin{cases} O(2n^2) & \text{per matrici piene} \\ O(cost \cdot n) & \text{per matrici sparse} \end{cases}$$

e questo può portare a preferire questi metodi in problemi sparsi

• La possibilità di ottenere efficientemente una soluzione accurata dipende anche dalla costante di contrazione del metodo

Confronto tra metodi diretti e metodi iterativi:

• Nei problemi di piccole dimensioni i metodi diretti forniscono di regola soluzioni accurate con un numero ottimale di operazioni

• Nei problemi di grandi dimensioni i metodi diretti soffrono di instabilità numeriche ed in generale non sono in grado di trarre vantaggio dalla eventuale sparsità del problema

• La situazione più favorevole all'uso di metodi iterativi è quindi quella di problemi sparsi di grandi dimensioni

schema	complessità (probl. pieni)	complessità (probl. sparsi)	occupazione (probl. pieni)	occupazione (probl. sparsi)
MEG	$O\left(\frac{2n^3}{3}\right)$	$O\left(\frac{2n^3}{3}\right)$	$O(n^2)$	$O(n^2)$
LU	$O\left(\frac{2n^3}{3}\right)$, $O(2n^2)$ (*)	$O\left(\frac{2n^3}{3}\right)$, $O(2n^2)$ (*)	$O(n^2)$	$O(n^2)$
QR	$O\left(\frac{4n^3}{3}\right)$, $O(3n^2)$ (*)	$O\left(\frac{4n^3}{3}\right)$, $O(3n^2)$ (*)	$O(n^2)$	$O(n^2)$
iterat.	$O(n^2)$ per it.	$O(n)$ per it.	$O(n^2)$	$O(n)$

(*) in caso di più sistemi con la stessa matrice ma con termini noti diversi

indice

Convergenza dei metodi iterativi per Sistemi Lineari

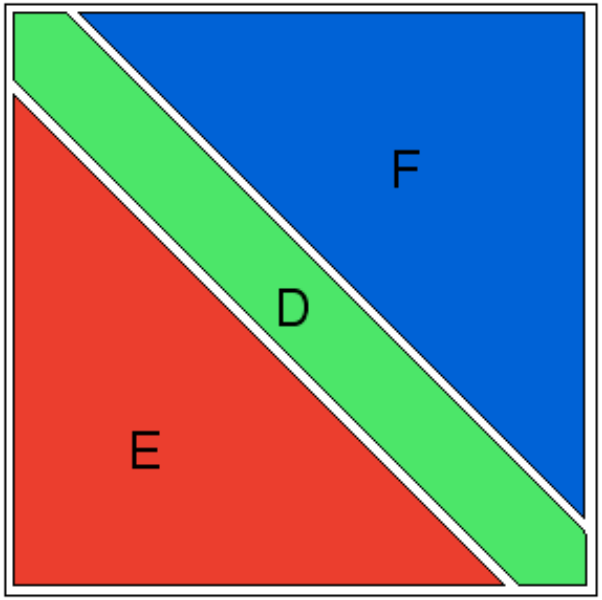
La condizione necessaria e sufficiente di convergenza, per i metodi nella forma

$$x_{k+1} = Bx_k + c \tag{4}$$

è che $\rho(B) < 1$.

• In pratica questa condizione è difficile da verificare e se possibile si sostituisce con la condizione sufficiente $\|B\| < 1$ (che implica l'altra, e coincide con la condizione di contrattività)

Per scrivere esplicitamente la matrice B nei vari casi, si partiziona la matrice A:



$$e_{ij} = \begin{cases} a_{ij} & \text{se } i > j \\ 0 & \text{altrimenti} \end{cases}$$

$$d_{ij} = \begin{cases} a_{ij} & \text{se } i = j \\ 0 & \text{altrimenti} \end{cases}$$

$$f_{ij} = \begin{cases} a_{ij} & \text{se } i < j \\ 0 & \text{altrimenti} \end{cases}$$

Con questa partizione, il metodo di Jacobi può essere riscritto come:

$$Dx_{k+1} = -(E + F)x_k + b$$

che coincide con la forma generale (4) ponendo $B_J = -D^{-1}(E + F)$ e $c_J = D^{-1}b$.

• In questo caso si dimostra facilmente che la condizione $\|B_J\|_\infty < 1$ equivale alla dominanza diagonale per righe di A, che è quindi una condizione sufficiente di convergenza

• Lo studio della convergenza nella norma $\|\cdot\|_1$ porta ad una condizione simile, ma di verifica meno immediata

A sua volta, il metodo di Gauss–Seidel può essere riscritto come:

$$(D + E)x_{k+1} = -Fx_k + b$$

che si riporta alla forma generale (4) ponendo $B_{GS} = -(D + E)^{-1}F$ e $c_{GS} = (D + E)^{-1}b$.

• In modo piuttosto tecnico si dimostra che due condizioni sufficienti per la convergenza sono la dominanza diagonale e la positività di A

• In teoria, qualsiasi sistema si può riportare a quest'ultimo caso:

$$Ax = b \quad \leftrightarrow \quad A^tAx = A^tb \qquad (A^tA > 0)$$

Ancora ricorrendo a questa partizione di A, il metodo SOR può essere riscritto come:

$$a_{ii}x_i^{(k+1)} = a_{ii}(1-\omega)x_i^{(k)} + \omega b_i - \omega \sum_{j<i} a_{ij}x_j^{(k+1)} - \omega \sum_{j>i} a_{ij}x_j^{(k)}$$

da cui si ottiene la forma

$$(D + \omega E)x_{k+1} = [(1-\omega)D - \omega F]x_k + \omega b$$

la cui matrice di iterazione è $B_{SOR} = (D + \omega E)^{-1}[(1-\omega)D - \omega F]$ e $c_{SOR} = \omega(D + \omega E)^{-1}b$.

• Si dimostra che una condizione sufficiente di convergenza per il metodo SOR è che $A > 0$ ed $0 < \omega < 2$

Nel caso del metodo di Richardson, la matrice di iterazione B_R ed il vettore c_R si scrivono immediatamente come $B_R = I - \beta A$ e $c_R = \beta b$.

- Si dimostra che una condizione sufficiente di convergenza per il metodo di Richardson è che $A > 0$ ed $0 < \beta < \frac{2}{\max_i \lambda_i(A)}$

indice

Metodi per equazioni scalari: la bisezione

Il metodo di bisezione (che *non è nella forma (1)*) è talvolta usato per dimostrare il teorema di esistenza degli zeri. Si parte da un intervallo $[a_0, b_0]$ in cui la funzione cambia di segno, e si costruisce una successione di intervalli $[a_k, b_k]$ in modo che $b_k - a_k = \frac{1}{2}(b_{k-1} - a_{k-1})$ e che la funzione cambi di segno in $[a_k, b_k]$.

- La radice è il limite delle successioni a_k e b_k

- L'algoritmo converge sotto la sola ipotesi di continuità di f

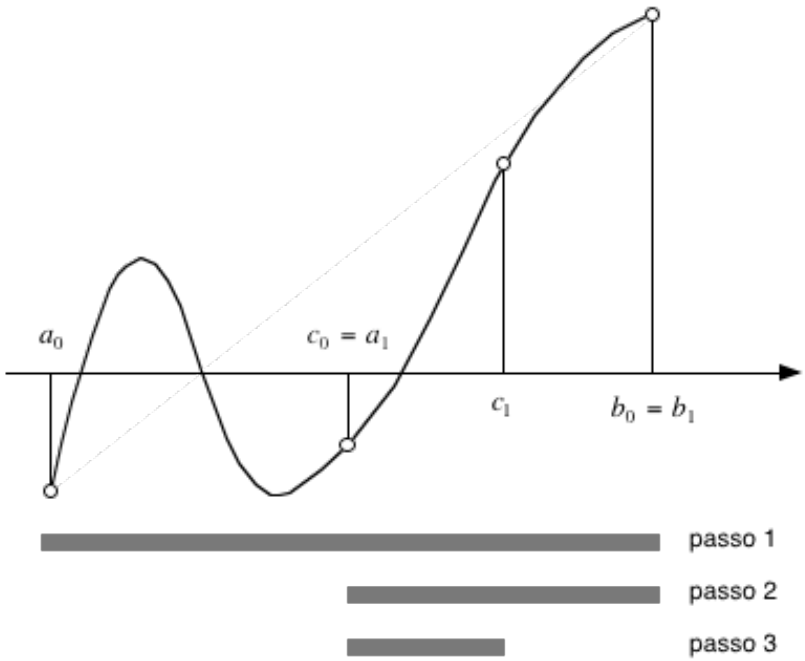
E' dato l'intervallo $[a_0, b_0]$, $k = 0$.

1. Poni $c_k = \frac{a_k + b_k}{2}$

2. Se $f(c_k) = 0$ o se è soddisfatta la condizione di arresto, STOP

3. Se $f(a_k)f(c_k) < 0$, poni $a_{k+1} = a_k$, $b_{k+1} = c_k$,
incrementa k e vai a 1

4. Se $f(b_k)f(c_k) < 0$, poni $a_{k+1} = c_k$, $b_{k+1} = b_k$,
incrementa k e vai a 1



- Ogni intervallo $[a_k, b_k]$ per $k \geq 0$ contiene la radice

- E' possibile che si scartino intervalli contenenti radici, se queste sono in numero pari

- Ad ogni passo, l'errore (visto come ampiezza dell'intervallo $[a_k, b_k]$) si dimezza

indice

Metodi per equazioni scalari: iterazioni di punto fisso

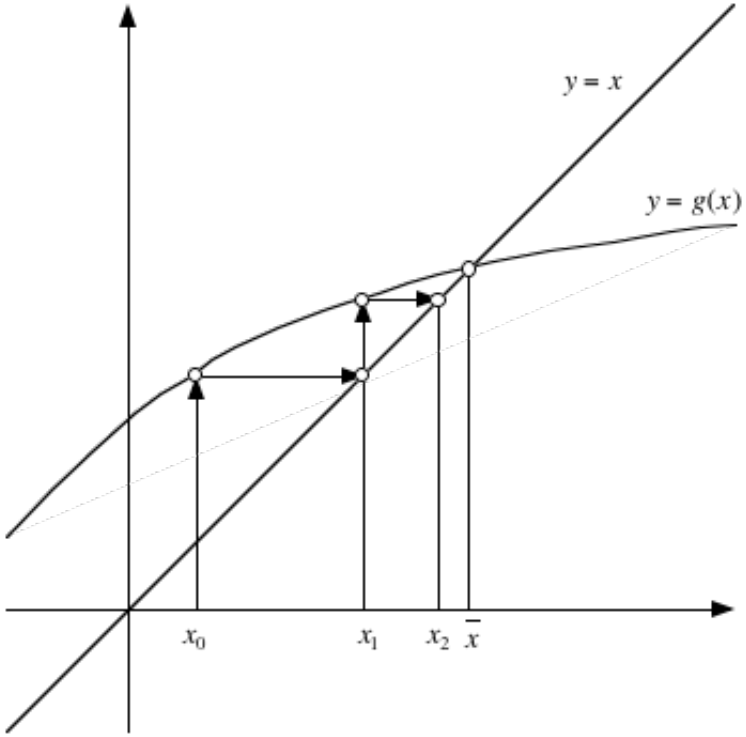
In una dimensione, riscriviamo i metodi in forma (1) come

$$x_{k+1} = g(x_k) \qquad (g : \mathbb{R} \to \mathbb{R}) \tag{5}$$

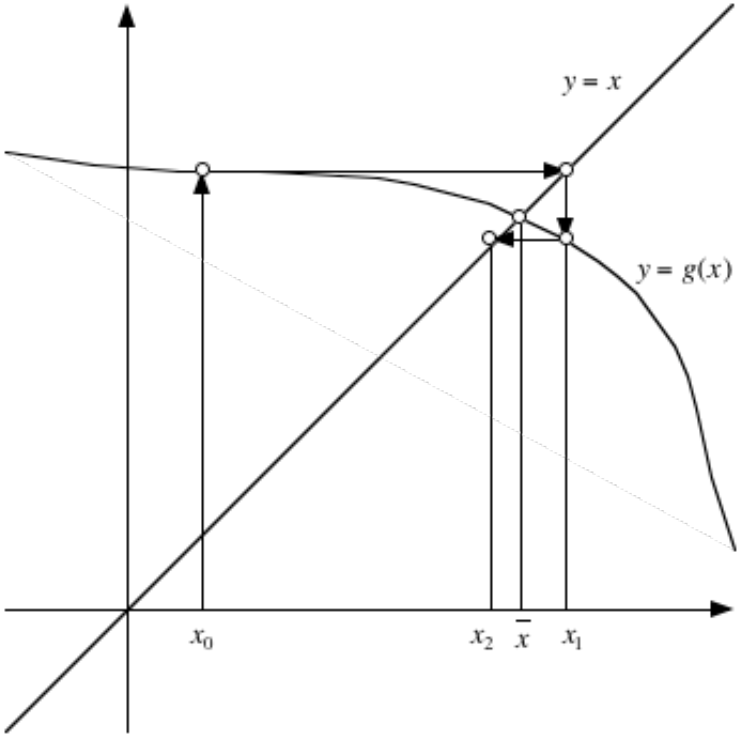
- La condizione di contrattività diviene $|g'(x)| < 1$

- In una dimensione si può dare una interpretazione grafica della costruzione ed eventuale convergenza della successione x_k, basandosi sul fatto che la soluzione è intersezione dei grafici

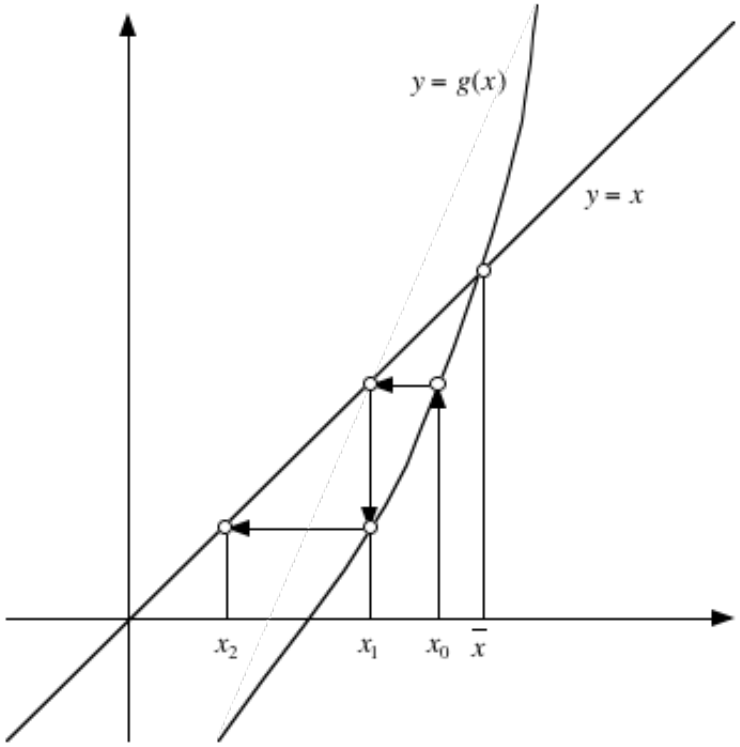
$$\begin{cases} y = x \\ y = g(x) \end{cases}$$



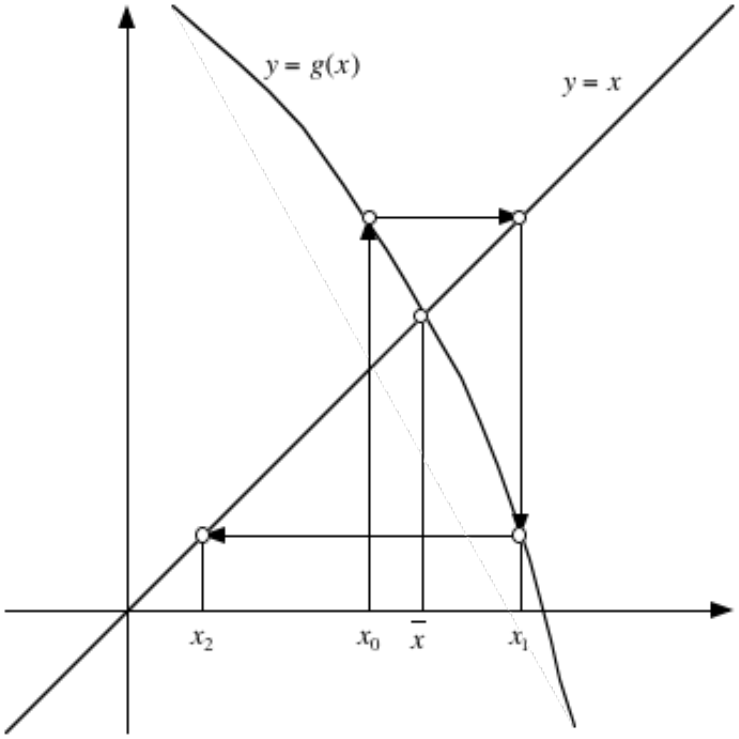
$0 < g'(x) < 1$



$-1 < g'(x) < 0$



$1 < g'(x)$



$g'(x) < -1$

Un modo standard per porre una equazione scalare $f(x) = 0$ in forma di punto fisso è di riscriverla come

$$x = x + \alpha(x) f(x) \tag{6}$$

• Perché questa forma sia equivalente all'equazione originale, la funzione (eventualmente costante) $\alpha(x)$ non deve avere zeri nell'intorno di $\bar{x}$.

• Utilizzando (6) per definire un metodo iterativo, si richiede normalmente che $\bar{x}$ sia una radice semplice (infatti se $\bar{x}$ è una radice multipla, si ha $g'(\bar{x}) = 1$ e quindi g non è una contrazione)

L'ordine di convergenza di un metodo nella forma (5) si può caratterizzare nel modo seguente:

• Se $g \in C^{m+1}$ e $g'(\bar{x}) = \cdots = g^{(m)}(\bar{x}) = 0$, $g^{(m+1)}(\bar{x}) \neq 0$, allora il metodo converge con ordine $m+1$ se x_0 è sufficientemente vicino a $\bar{x}$

• In particolare, se $g(x)$ è nella forma (6) con $\alpha(x) = \alpha$ (costante) allora se $\alpha = -1/f'(\bar{x})$ si ha convergenza quadratica (in generale, con questa struttura del metodo non si può avere un ordine di convergenza superiore a 2)

• Poiché $\bar{x}$ non è noto (ed anche la espressione esplicita di f' può non essere nota), la costante α deve essere una approssimazione del valore ottimale $-1/f'(\bar{x})$

• Una possibilità è quella di sostituire $f'(\bar{x})$ con il rapporto incrementale di f calcolato su un intorno (sufficientemente piccolo) $[a, b]$ di $\bar{x}$: in questo modo si ha il cosiddetto metodo delle corde:

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

convergente se $f \in C^1$, $f'(\bar{x}) \neq 0$ e a, b, x_0 sono abbastanza vicini a $\bar{x}$

indice

Metodi per equazioni scalari: metodo di Newton e varianti

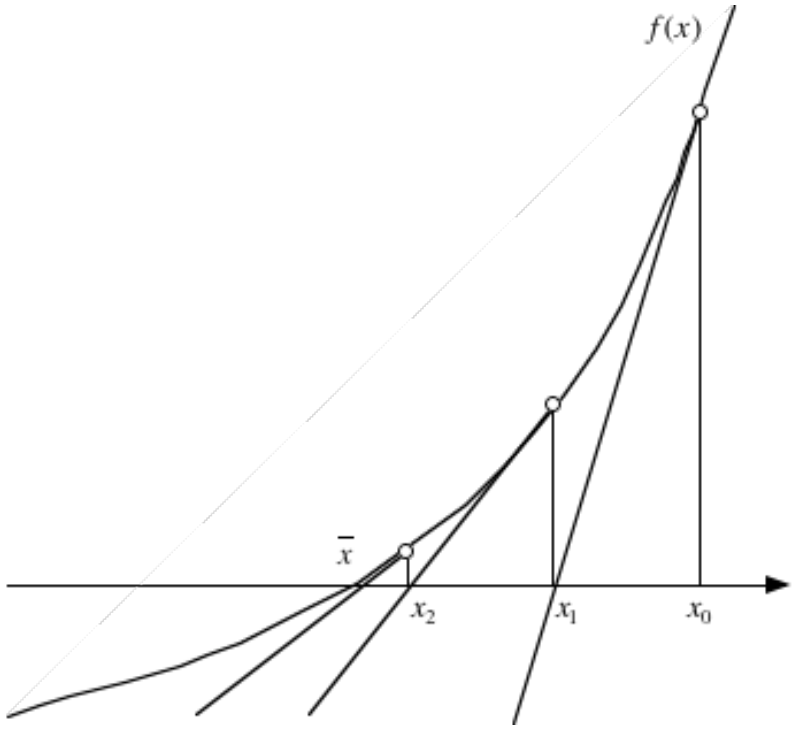
Il metodo di Newton si ottiene utilizzando la forma (6) con $\alpha(x) = -1/f'(x)$

- Si suppone quindi di conoscere l'espressione esplicita della derivata

- Poiché, se $f \in C^2$ e $f'(\bar{x}) \neq 0$, si ha

$$g'(\bar{x}) = 1 - \frac{f'(\bar{x})^2 - f(\bar{x})f''(\bar{x})}{f'(\bar{x})^2} = 0$$

il metodo converge con ordine quadratico se x_0 è abbastanza vicino a $\bar{x}$



- L'approssimazione x_{k+1} è lo zero della tangente al grafico di f in $(x_k, f(x_k))$

- Va evitato che la derivata si annulli nell'intorno di $\bar{x}$ (in questo caso la tangente diventerebbe parallela all'asse delle ascisse)

• Si può dimostrare convergenza monotona, globale dello schema in uno dei seguenti casi:

$$\begin{cases} x_0 > \bar{x}, & \begin{cases} f(x) \text{ crescente convessa} \\ f(x) \text{ decrescente concava} \end{cases} \\ x_0 < \bar{x}, & \begin{cases} f(x) \text{ crescente concava} \\ f(x) \text{ decrescente convessa} \end{cases} \end{cases}$$

• In caso di radici di molteplicità $m > 1$, il metodo non è più quadratico, ma converge invece con ordine quadratico il metodo

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

Un metodo di Newton approssimato si può ottenere sostituendo il calcolo di con $f'(x)$ con quello del rapporto incrementale di passo h nel punto x_k, ottenendo lo schema:

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

- Non è richiesta l'espressione esplicita della derivata, ma si calcola f due volte per iterazione (in x_k ed in $x_k + h$). Per evitare perdita di cifre significative, h non va scelto troppo piccolo

- In generale $g'(\bar{x}) \neq 0$ e quindi si ha convergenza lineare, ma il coefficiente di contrazione può essere abbastanza vicino a zero

Nel metodo delle secanti si sostituisce il calcolo di con $f'(x)$ con quello del rapporto incrementale tra i punti x_{k-1} e x_k, ottenendo quindi lo schema (che non è nella forma (6)):

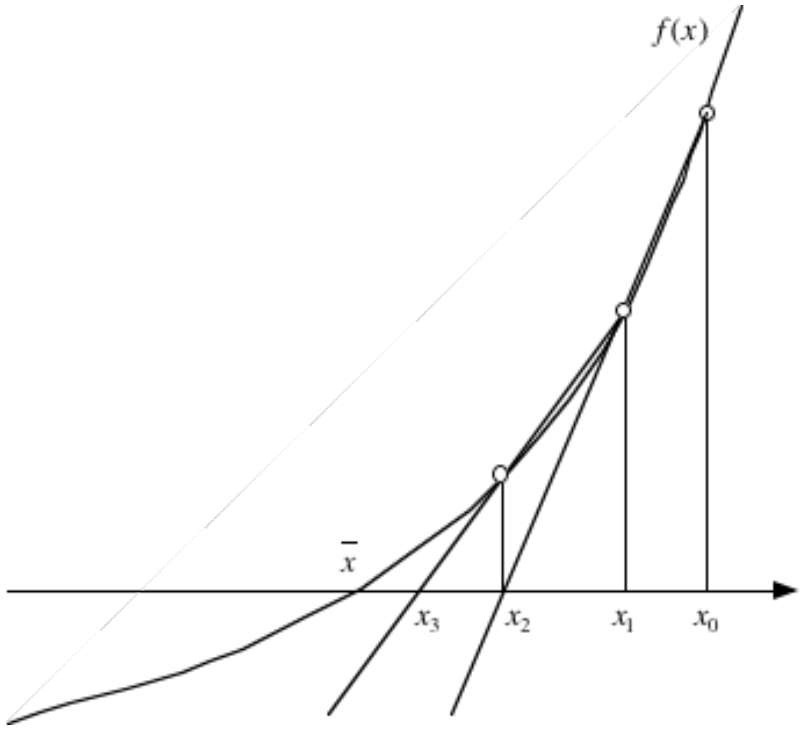
$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k)$$

- Non è richiesta l'espressione esplicita della derivata, ed inoltre si calcola f una sola volta per iterazione

- Se $f \in C^2$ e $f'(\bar{x}) \neq 0$, la convergenza è sopralineare, con esponente

$$\gamma = \frac{1+\sqrt{5}}{2} \approx 1.618$$

se x_0 ed x_1 sono abbastanza vicini a $\bar{x}$



- L'approssimazione x_{k+1} è lo zero della retta passante per i punti $(x_k, f(x_k))$ e $(x_{k-1}, f(x_{k-1}))$

- Come negli altri casi, si richiede che $\bar{x}$ sia una radice semplice

- Si hanno risultati di convergenza monotona analoghi a quelli del metodo di Newton

Nel metodo di Steffensen il calcolo di con $f'(x)$ viene approssimato con quello del rapporto incrementale tra i punti x_k e $x_k + f(x_k)$, ottenendo uno schema che è di nuovo nella forma (6)):

$$x_{k+1} = x_k - \frac{f(x_k)^2}{f(x_k) - f(x_k + f(x_k))}$$

- Non è richiesta l'espressione esplicita della derivata, ma si calcola f due volte per iterazione

- Se $f \in C^2$, $f'(\bar{x}) \neq 0$, e x_0 è abbastanza vicino a $\bar{x}$, la convergenza è quadratica

Nel metodo di Newton cubico si estende l'idea del metodo di Newton definendo x_{k+1} come uno zero del polinomio di Taylor di secondo ordine di f calcolato in x_k:

$$x_{k+1} = x_k - \frac{f'(x_k) \pm \sqrt{f'(x_k)^2 - 2f(x_k)f''(x_k)}}{f''(x_k)}.$$

- La ambiguità di segno si risolve scegliendo la radice più vicina ad x_k

- Il metodo converge con ordine 3 se $f \in C^3$ e x_0 è abbastanza vicino a $\bar{x}$

- Il metodo può convergere a radici complesse (situazione utile nel caso di equazioni algebriche)

Nel metodo di Muller si utilizza l'idea del metodo di Newton cubico, ma si definisce x_{k+1} come lo zero del polinomio (detto *interpolatore*) di secondo grado di f passante per i punti $(x_k, f(x_k))$, $(x_{k-1}, f(x_{k-1}))$ e $(x_{k-2}, f(x_{k-2}))$

• Anche in questo caso, tra due radici reali si sceglie la radice più vicina ad x_k, e si può convergere a radici complesse

• Il metodo converge con ordine $\gamma \approx 1.84$ se $f \in C^3$ e x_0, x_1 e x_2 sono abbastanza vicini a $\bar{x}$

Confronto tra i vari metodi iterativi:

- Per applicare i metodi di Newton occorre conoscere l'espressione analitica della funzione

- Quando l'espressione di f non è nota, i metodi più efficienti sono quelli delle secanti e di Muller, a patto che la funzione f sia sufficientemente regolare

- In mancanza di regolarità della f, l'unico metodo utilizzabile sotto la sola ipotesi di continuità è la bisezione

schema	complessità	ordine	regolarità
bisezione	calcolo $f(x)$	$\gamma = 1$	C^0
corde	calcolo $f(x)$	$\gamma = 1$	C^1
secanti	calcolo $f(x)$	$\gamma \approx 1.62$	C^2
Steffensen	calcolo $f(x)$, $f(x + f(x))$	$\gamma = 2$	C^2
Muller	calcolo $f(x)$	$\gamma \approx 1.84$	C^3
Newton	calcolo $f(x), f'(x)$	$\gamma = 2$	C^2
Newton cubico	calcolo $f(x), f'(x), f''(x)$	$\gamma = 3$	C^3