

## Soluzione numerica di

### Equazioni non lineari

Data una funzione  $f : \mathbb{R} \rightarrow \mathbb{R}$  non lineare consideriamo il problema di determinare i valori  $\alpha \in \mathbb{R}$  tali che  $f(\alpha) = 0$ .

Tali valori sono solitamente chiamati **zeri** o **radici** della funzione  $f$ .

**Esempi di equazioni non lineari:**

Equazioni algebriche:  $P_n(x) = \sum_{i=0}^n a_i x^i = 0, \quad n \geq 2$

Altri esempi di equazioni non lineari:

$$x + 4\sin(x)=0 \quad e^x + x^2 = 0, \quad \sqrt{x} - \log(x) = 0$$

Le radici di un'equazione non lineare non possono in generale venire espresse in “forma chiusa” e anche quando ciò è possibile la corrispondente espressione può risultare molto complessa.

Esistono formule esplicite per il calcolo di radici di polinomi di grado minore di 5. L'applicazione di tali formule non è però così immediata come la nota formula per il calcolo di radici di polinomi di secondo grado e risulta sicuramente preferibile disporre di **algoritmi numerici in grado di fornire soluzioni approssimate del problema**. Si ricorre a metodi numerici iterativi approssimanti.

L'approssimazione numerica di una radice  $\alpha$  di  $f(x)$  si basa sull'uso di metodi iterativi che consistono nella costruzione di una successione di iterati  $x_1, x_2, \dots, x_k, \dots$  che tende alla soluzione  $\alpha$  del problema, cioè

$$\lim_{k \rightarrow +\infty} x_k = \alpha \quad (1)$$

E' inoltre necessario **rendere il problema ben posto**, cioè individuare un intervallo  $I$  contenente una sola radice e poi applicare il metodo iterativo fino a convergenza alla soluzione (radice).

## Esempi di applicazioni reali:

### Calcolo del tasso medio di rendita annuale di un fondo di investimento

#### Obiettivo:

Calcolare il tasso medio di rendita annuale ( $r$ ) di un fondo di investimento su un periodo di più anni.

#### Dati:

- Vengono investiti  $v$  euro all'inizio di ogni anno.
- Alla fine dell' $n$ -esimo anno, il montante accumulato è pari a  $M$  euro.

La seguente relazione **lega  $M$  a  $r$** :

$$M = v \frac{1+r}{r} [(1+r)^n - 1]$$

- **$M$ :** Montante totale accumulato alla fine dell' $n$ -esimo anno.
- **$v$ :** Ammontare investito all'inizio di ogni anno.
- **$r$ :** Tasso medio di rendita annuale (incognita da calcolare).
- **$n$ :** Numero di anni considerati.

Si deduce che il tasso medio di rendita annuale è il valore  $r$  che rende uguale il primo termine dell'equazione al secondo, cioè il valore di  $r$  per cui:

$$M - v \frac{1+r}{r} [(1+r)^n - 1] = 0$$

cioè è lo zero dell'equazione non lineare:

$$f(r) = 0$$

con

$$f(r) = M - v \frac{1+r}{r} [(1+r)^n - 1]$$

#### Dati:

- Investimento annuale ( $v$ ): 1000 euro
- Montante dopo  $n=5$  anni ( $M$ ): 6000 euro

Incognita:  **$r$** : Tasso medio di rendita annuale?

## Esempio 2:

### Problema:

Dato un gas reale a una certa temperatura ( $T$ ) e pressione ( $P$ ), determinare **il volume molare** ( $V$ ) del gas.

L'equazione di Beattie-Bridgeman è un'equazione termodinamica usata per descrivere il comportamento di gas reali, ovvero gas che si discostano dal comportamento ideale previsto dalla legge dei gas perfetti. Modellizza la relazione tra la **pressione, il volume e la temperatura** di un gas

$$P = \frac{R \cdot T}{V} + \frac{\beta}{V^2} + \frac{\gamma}{V^3} + \frac{\delta}{V^4}$$

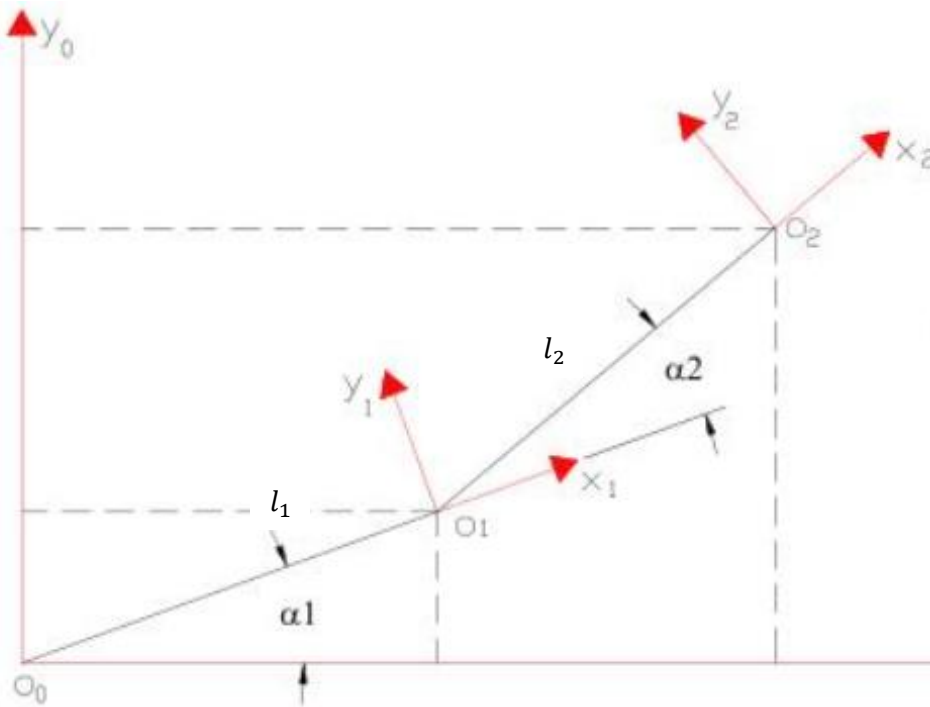
ove  $P$  è la pressione (atm),  $V$  è il volume molare (litro),  $T$  è la temperatura ( $^{\circ}\text{K}$ ),  $\beta, \gamma, \delta$  sono parametri caratteristici di un gas e dipendenti dalla temperatura, e  $R$  è la costante universale dei gas.

L'equazione è *esplicita* nella pressione  $P$ , ma *implicita* nella temperatura  $T$  e nel volume  $V$ . In particolare, per trovare il valore del **volume** corrispondente **a valori assegnati della pressione e della temperatura**, bisogna risolvere l'equazione non lineare

$$f(V) = \frac{R \cdot T}{V} + \frac{\beta}{V^2} + \frac{\gamma}{V^3} + \frac{\delta}{V^4} - P = 0$$

### Esempio 3:

Risolvere il problema di cinematica inverso relativo al seguente braccio meccanico sul piano.



$O_0X_0Y_0$  : sistema di riferimento fisso rispetto a cui vengono definite posizioni ed angoli.

$l_1$  : lunghezza del primo braccio;

$O_1$  : giunto attraverso cui avviene la rotazione;

$l_2$  : lunghezza del secondo braccio ;

$\alpha_1$ : angolo di rotazione del primo braccio rispetto al sistema di riferimento fisso;

$\alpha_2$ : angolo di rotazione del secondo braccio rispetto al primo braccio;

Cerchiamo adesso le equazioni che descrivono la relazione tra gli angoli  $\alpha_1$  ed  $\alpha_2$  e le coordinate cartesiane  $x_2, y_2$  del punto  $O_2$  sull'estremità del secondo braccio.

Il primo braccio ruota di un angolo  $\alpha_1$  rispetto al sistema di riferimento fisso. Le coordinate del punto  $O_1$  sono:

$$x_1 = l_1 \cos(\alpha_1)$$

$$y_1 = l_1 \sin(\alpha_1)$$

Per ottenere le coordinate di  $O_2$ , estremo del secondo braccio, rispetto al sistema di riferimento fisso, bisogna considerare la rotazione totale del secondo braccio ( $\alpha_1 + \alpha_2$ ) e sommare le coordinate di  $O_1$

$$x_2 = l_1 \cos(\alpha_1) + l_2 \cos(\alpha_1 + \alpha_2)$$

$$y_2 = l_1 \sin(\alpha_1) + l_2 \sin(\alpha_1 + \alpha_2)$$

Si tratta di equazioni non lineari a causa della presenza di funzioni trigonometriche.

Possono essere utilizzate per:

- Calcolare la posizione del punto  $O_2$  per valori noti degli angoli articolari (cinematica diretta).
- Calcolare gli angoli articolari necessari per raggiungere una posizione desiderata del punto  $O_2$  (cinematica inversa).

**Le useremo per risolvere il secondo problema:**

**Obiettivo:**

**Calcolare gli angoli  $\alpha_1, \alpha_2$  in maniera tale che l'estremo finale del secondo braccio  $O_2$  assuma coordinate  $x_2, y_2$  desiderate.**

Ciò equivale a risolvere un sistema di due equazioni non lineari, per individuare gli angoli incogniti  $\alpha_1, \alpha_2$

$$\begin{cases} f_1(\alpha_1, \alpha_2) = l_1 \cos(\alpha_1) + l_2 \cos(\alpha_1 + \alpha_2) - x_2 = 0 \\ f_2(\alpha_1, \alpha_2) = l_1 \sin(\alpha_1) + l_2 \sin(\alpha_1 + \alpha_2) - y_2 = 0 \end{cases}$$

**Definizione:** Se  $\alpha \in R$  è tale che  $f(\alpha) = 0$  ed  $f'(\alpha) \neq 0$ ,  $\alpha$  viene chiamata **radice semplice**; in generale, se  $f^{(k)}(\alpha) = 0$ ,  $k = 0, \dots, m-1$  e  $f^{(m)}(\alpha) \neq 0$  allora  $\alpha$  detta **radice multipla di molteplicità m**.

## Condizionamento del problema del calcolo degli zeri di una funzione non lineare

Quando si calcola lo zero di una funzione non lineare con un calcolatore, si ottiene un'approssimazione numerica, non il valore esatto. Occorre valutare quanto differisce questa soluzione da quella vera.

Sia  $\alpha$  tale che  $f(\alpha)=0$ ,  $\alpha$  zero semplice.

Il valore calcolato numericamente  $\tilde{\alpha} = \alpha + \delta$ ,  $\delta > 0$ ,  $\delta$  piccolo, può essere visto come la radice dell'equazione

$$\tilde{f}(x) = f(x) + \varepsilon g(x) = 0 \quad (2)$$

$\varepsilon > 0$ ,  $\varepsilon$  piccolo,  $f(x)$ ,  $g(x)$  differenziabili.

dove la differenza  $|\tilde{f}(x) - f(x)| = |\varepsilon g(x)|$  rappresenta la perturbazione sulla funzione originale, quindi sui dati; la differenza  $|\tilde{\alpha} - \alpha| = |\delta|$  è la perturbazione sui risultati.

Si ha quindi che  $\tilde{f}(\tilde{\alpha}) = 0$ ,

Facendo uno sviluppo in serie di Taylor del primo ordine di  $\tilde{f}(x)$  in un intorno di  $\alpha$

$$0 = \tilde{f}(\alpha + \delta) = \tilde{f}(\alpha) + \delta \tilde{f}'(\alpha) + \frac{1}{2} \delta^2 \tilde{f}''(\xi) \quad \xi \in (\alpha, \alpha + \delta)$$

Trascurando in termini del secondo ordine

$$0 = \tilde{f}(\alpha + \delta) \approx \tilde{f}(\alpha) + \delta \tilde{f}'(\alpha)$$

Sostituiamo nella precedente equazione la definizione di  $\tilde{f}(x)$  data in (2), e ricordiamo che  $\tilde{\alpha} = \alpha + \delta$

$$0 \approx f(\alpha) + \varepsilon g(\alpha) + \delta \cdot (f'(\alpha) + \varepsilon g'(\alpha))$$

Nella precedente equazione, per ipotesi  $f(\alpha) = 0$

$$\varepsilon g(\alpha) + \delta f'(\alpha) + \varepsilon \cdot \delta \cdot g'(\alpha) \approx 0$$

Ora, proseguendo con un'analisi al I ordine (ossia considerando  $\varepsilon \cdot \delta$  trascurabile), si ottiene

$$\varepsilon g(\alpha) + \delta f'(\alpha) \approx 0$$

da cui

$$\delta \approx -\frac{\varepsilon g(\alpha)}{f'(\alpha)}$$

ossia

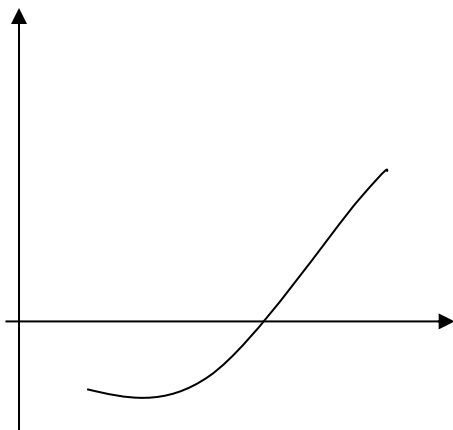
$$|\tilde{\alpha} - \alpha| = |\delta| \approx K |\varepsilon g(\alpha)| \quad (4)$$

dove quantità  $K = \frac{1}{|f'(\alpha)|}$  rappresenta l'indice di condizionamento del problema di calcolare la radice della funzione  $f(x)$ .

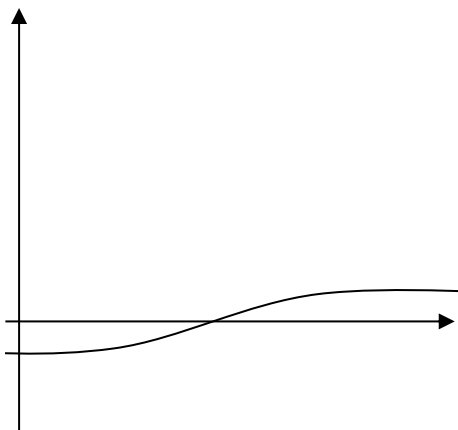
Se  $|f'(\alpha)|$  è molto piccolo, allora il problema è mal condizionato.

Viceversa il problema risulta ben condizionato e  $\tilde{f}(x) = 0$  ha una radice  $\tilde{\alpha}$  che non differisce molto da  $\alpha$

Come mostrano le figure, se  $f'(\alpha)$  è grande, il problema è ben condizionato, se  $f'(\alpha)$  è molto piccolo, vicino allo zero, il problema è mal condizionato.



**Ben Condizionato**



**Mal Condizionato**

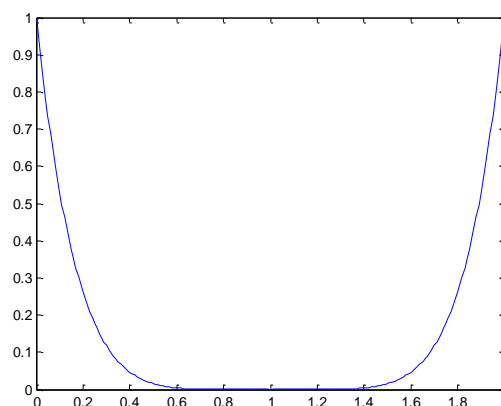
Lo studio precedente ci fa anche capire che il calcolo di radici multiple di molteplicità  $m > 1$  (quelle in cui oltre la funzione si annullano anche le sue derivate fino all'ordine  $m-1$ ) è un problema numericamente molto difficile.

### Esempio

L'equazione:

$$F(x) = (x-1)^6 = x^6 - 6x^5 + 15x^4 - 20x^3 + 15x^2 - 6x + 1 = 0$$

possiede in  $\alpha = 1$  una radice di molteplicità 6, quindi  $F'(1)=0$ , come si vede bene in figura, e il problema della determinazione della radice  $\alpha = 1$  è mal condizionato.





Come anticipato l'approssimazione numerica di una radice  $\alpha$  di  $f(x)$  si basa sull'uso di metodi iterativi che consistono nella costruzione di una successione di iterati  $x_1, x_2, \dots, x_k, \dots$  che tende alla soluzione del problema  $\alpha$ , cioè

$$\lim_{k \rightarrow +\infty} x_k = \alpha$$

**Le problematiche legate ai metodi iterativi per la soluzione di equazioni non lineari sono:**

- 1) **Scelta del valore iniziale  $x_0$  (di innesco) e Convergenza** della successione di iterati ad  $\alpha$
- 2) **Ordine di convergenza** della successione di iterati ad  $\alpha$  (ordine del metodo)
- 3) **Criteri di arresto** del metodo iterativo (questo è un problema numerico dovuto al fatto che lavoriamo con i numeri finiti e dobbiamo fare un numero finito di passi)

**DEF:** Un metodo converge **localmente** ad  $\alpha$  se la convergenza della successione  $\{x^{(k)}\}$  dipende in modo critico dalla vicinanza di  $x_0$  ad  $\alpha$ . Il procedimento è **globalmente** convergente quando la convergenza non dipende da quanto  $x_0$  è vicino ad  $\alpha$ , ossia converge per ogni scelta di  $x_0$ .

Per i metodi a convergenza locale la scelta del punto di innesco è cruciale.

## Definizione: Ordine di Convergenza

Sia data una successione di iterati  $\{x_k\}$ , generata da un metodo numerico, convergente ad un limite  $\alpha$  e sia  $e_k = x_k - \alpha$ . Se esistono due numeri reali  $p \geq 1$  e  $c > 0$ , tali che

$$\lim_{k \rightarrow +\infty} \frac{|e_{k+1}|}{|e_k|^p} = c$$

si dice che la successione ha **ordine di convergenza  $p$  e fattore di convergenza  $c$** .

Ciò significa che esiste un indice  $k_0$ , tale che per  $k > k_0$ , risulta

$$|e_{k+1}| \approx c|e_k|^p$$

Se  $p=1$ , occorre che  $c < 1$  affinché la successione sia convergente. In tal caso si dice convergente linearmente. Risulta:

$$|e_{k+1}| \approx c|e_k| \approx c^2|e_{k-1}| \approx \dots \approx c^{k+1}|e_0|$$

e tanto più piccolo è  $c < 1$  tanto migliore è la convergenza, pur rimanendo lineare.

- se  $p = 1$  la convergenza si dice **lineare**
- se  $1 < p < 2$  la convergenza si dice **superlineare**
- se  $p = 2$  la convergenza si dice **quadratica**

Si dice che il metodo che genera la successione  $\{x_k\}$  ha velocità di convergenza di ordine  $p$  se tale è la successione da esso generata.

## Significato del concetto di ordine di convergenza

Sia  $e_k = x_k - \alpha$ , supponiamo che  $|e_k| \leq \frac{1}{2}10^{-n}$ , cioè la radice  $x_k$  ha  $n$  decimali corretti,

Se il metodo iterativo ha ordine di convergenza  $p$ , allora

$|e_{k+1}| \approx c|e_k|^p \leq c\left(\frac{1}{2}10^{-n}\right)^p = \frac{c}{2^p}10^{-pn}$  cioè la radice  $x_k$  ha  $p \cdot n$  decimali corretti.

Il numero di decimali corretti tende ad essere moltiplicato per  $p$  ad ogni passo solo per  $k \rightarrow \infty$ .

## Criteri di arresto

Un metodo numerico convergente genera una successione  $\{x_k\}$  di iterati che soddisfa la (1). Tuttavia, in un contesto di calcolo computazionale, non è possibile eseguire un numero infinito di passi. Di conseguenza, è necessario stabilire dei criteri **per arrestare il procedimento iterativo**.

Vi sono due possibili criteri:

- Controllo del valore della funzione nel punto  $x_k$
- Controllo dell'incremento che diamo a  $x_{k-1}$  per ottenere  $x_k$ .

Il processo viene arrestato all'iterato  $k$  per il quale si verifica una delle seguenti due condizioni:

### 1. Condizione di arresto basata sul valore della funzione nell'iterato $x_k$ :

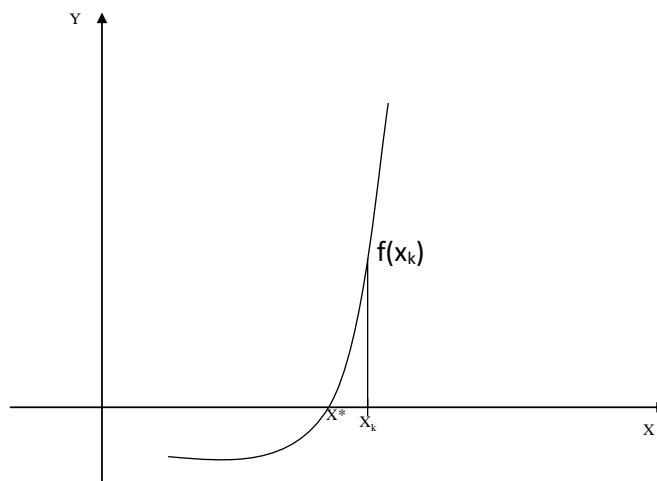
$$|f(x_k)| < \varepsilon$$

## 2. Condizione di arresto basata sull' incremento di $x_k$

$$|x_k - x_{k-1}| < \varepsilon$$

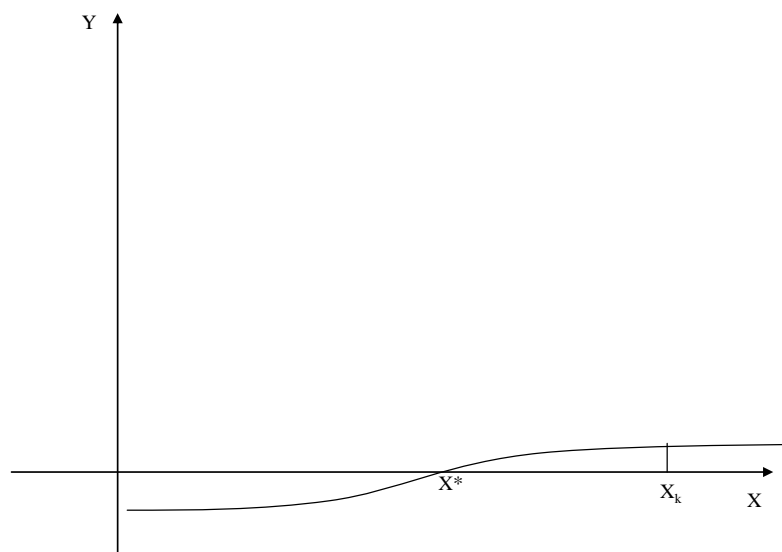
Utilizzando il controllo del valore della funzione nel punto  $x_k$  come criterio di arresto, si possono però presentare due situazioni problematiche:

### Caso Restrittivo



L'iterato  $x_k$  è vicino a  $\alpha$  anche se  $|f(x_k)|$  è grande.

### Caso Ottimistico



Il valore di  $x_k$  è lontano da  $x^*$  ma  $|f(x_k)|$  è piccolo

Si vede quindi che, se la funzione ha una derivata prima alta nell'intorno della soluzione, il test sul valore della funzione può risultare troppo restrittivo, mentre se ha una derivata piccola allora il test può risultare troppo permissivo.

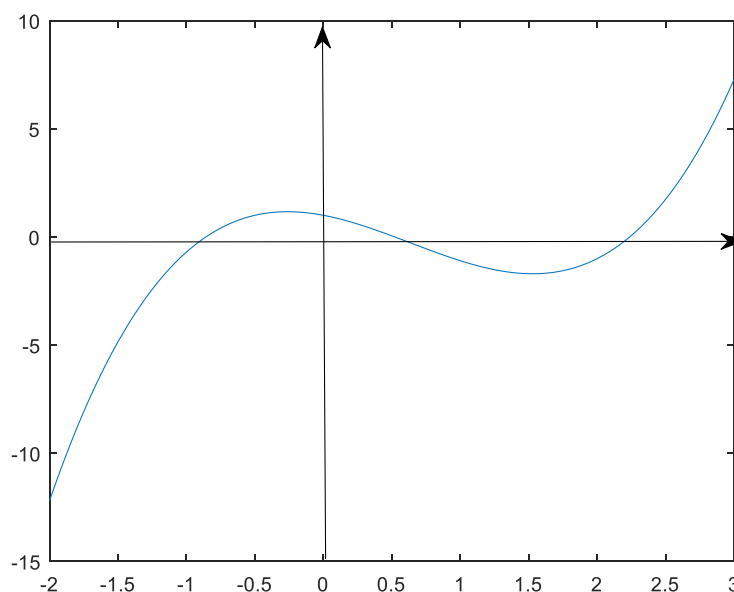
Si può concludere che un criterio d'arresto basato sia sul controllo del valore della funzione sia sul controllo dell'incremento risulta molto più affidabile.

N.B: nella pratica conviene utilizzare il **criterio di arresto relativo**:

$$\frac{|x_k - x_{k-1}|}{|x_k|} < \varepsilon$$

**Localizzare le radici:** determinare il numero delle soluzioni e separare ogni soluzione, cioè individuare, per ogni soluzione, un intervallo che non ne contenga altre. (discretizzando l'intervallo iniziale)

- Applicare per ogni intervallo determinato un metodo iterativo fino alla convergenza ad una soluzione (radice).



## Teorema degli zeri di funzioni continue

Sia  $f(x)$  continua nell'intervallo  $[a, b]$  e sia tale che  $f(a) \cdot f(b) < 0$ , allora  $f$  ammette almeno uno zero in  $(a, b)$ , cioè esiste almeno un punto  $\alpha$  in  $(a, b)$  tale che  $f(\alpha) = 0$ .

- **Metodo di Bisezione**

Il metodo di Bisezione si basa sul teorema degli zeri di funzioni continue.

Sia  $f(a) \cdot f(b) < 0$  e poniamo  $[a_0, b_0] \equiv [a, b]$ .

Il **metodo di bisezione** consiste nel generare una successione di sottointervalli

$I_1 = [a_1, b_1]$ ,  $I_2 = [a_2, b_2]$ , ...,  $I_k = [a_k, b_k]$  tali che racchiudano sempre lo zero cercato, cioè

$$f(a_k) \cdot f(b_k) < 0 \quad e \quad I_k \subset I_{k-1}$$

In particolare si determina

$$c_k = \frac{1}{2}(a_{k-1} + b_{k-1})$$

si calcola  $f(c_k)$  e si determina il sottointervallo che soddisfa le condizioni del teorema.

*Se  $f(c_k) = 0 \Rightarrow c_k$  è la radice  $\alpha$  cercata  
altrimenti*

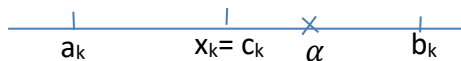
*se  $f(a_{k-1}) \cdot f(c_k) < 0$  si pone  $b_k = c_k$   $a_k = a_{k-1}$*

*se  $f(b_{k-1}) \cdot f(c_k) < 0$  si pone  $a_k = c_k$   $b_k = b_{k-1}$*

In questo modo l'intervallo iniziale viene via via dimezzato. In particolare, dopo  $k$  passi si arriva all'intervallo  $[a_k, b_k]$  di ampiezza

$$(b_k - a_k) = \frac{(b_{k-1} - a_{k-1})}{2} = \dots = \frac{(b_0 - a_0)}{2^k}.$$

La relazione precedente può essere utilizzata per dimostrare che il **metodo converge** con ordine di convergenza  $p=1$  e  $c=1/2$ . Infatti si ha



$$|e_k| = |x_k - \alpha| \leq \frac{1}{2} |b_k - a_k| = \frac{1}{2^{k+1}} |b_0 - a_0|$$

Da cui segue che  $\lim_{k \rightarrow \infty} |e_k| = 0$ , che prova che il metodo di bisezione è convergente.

Il metodo converge linearmente, infatti, poichè

$$|e_{k+1}| = |x_{k+1} - \alpha| \leq \frac{1}{2} |b_{k+1} - a_{k+1}| = \frac{1}{2^{k+2}} |b_0 - a_0|$$

risulta che

$$\frac{|e_{k+1}|}{|e_k|} \approx \frac{1}{2}$$

Il metodo di bisezione ha ordine di convergenza lineare,  $p=1$ , e fattore di convergenza  $c=1/2$ .

Il metodo di bisezione converge globalmente alla soluzione con la sola ipotesi che  $f$  sia continua nell'intervallo  $[a; b]$ .

La convergenza è garantita qualunque sia l'ampiezza dell'intervallo iniziale  $[a, b]$ .

E' un metodo molto lento. Servono fino a 4 iterazioni per ottenere una cifra significativa esatta nella soluzione approssimata.

Supponiamo che ci vogliano  $j$  iterazioni per ottenere una cifra significativa esatta nella soluzione approssimata:

$$|e_{k+j}| \approx \frac{1}{10} |e_k|$$

Vogliamo stimare il numero di iterazioni  $j$ .

$$|e_{k+j}| \leq \frac{1}{2^{k+1+j}} |b_0 - a_0| \approx \frac{1}{10} \frac{1}{2^{k+1}} |b_0 - a_0| \implies \frac{1}{2^j} \approx \frac{1}{10} \implies 2^j \approx 10,$$

da cui segue

$$j \approx \log_2 10 \approx 3.32$$

### Criterio di arresto

Un criterio di arresto del metodo di bisezione è chiedere che l'errore al passo  $k$ ,  $|e_k|$ , sia minore di una precisione fissata  $\varepsilon$ , cioè

$$\left| \frac{b-a}{2^{k+1}} \right| \leq \varepsilon. \quad (2)$$

Questa relazione ci permette di valutare teoricamente quante iterazioni sono necessarie per raggiungere la precisione prefissata. Infatti dalla relazione (2) si ottiene

$$2^{k+1} \geq \frac{b-a}{\varepsilon} \Rightarrow k \geq \log_2 \left( \frac{b-a}{\varepsilon} \right) - 1$$

Poiché il numero di iterazioni deve essere un intero positivo, il primo valore che soddisfa la disuguaglianza deve essere un intero superiore al numero minimo di iterazioni.



Il primo intero positivo per cui risulta  $k \geq \log_2 \left( \frac{b-a}{\varepsilon} \right) - 1$  è  $k = \left\lceil \log_2 \left( \frac{b-a}{\varepsilon} \right) - 1 \right\rceil$

dove la funzione **ceil(x)** (che si indica con  $\lceil x \rceil$ ) arrotonda un numero reale x al più vicino intero superiore.

Il metodo di Bisezione è un metodo di sicura ma lenta convergenza.

**Osservazione:** Nel programmare l'algoritmo è utile, per vedere il segno di  $f(c_k)$  utilizzare la funzione di libreria

$$\text{sign}(x) = \begin{cases} = 1 & \text{se } x > 0 \\ = 0 & \text{se } x = 0 \\ = -1 & \text{se } x < 0 \end{cases}$$

Inoltre quando si va a calcolare il punto  $c_k = \frac{a_{k-1} + b_{k-1}}{2}$  operando con i numeri finiti si può ottenere un risultato fuori dall'intervallo di definizione; infatti ad es. operando con 3 cifre decimali il centro dell'intervallo  $[0.983, 0.986]$  è

$$s_1 = fl(0.983 + 0.986) = fl(1.969) = 0.196 \cdot 10^1$$

$$s_2 = fl\left(\frac{s_1}{2}\right) = fl\left(\frac{0.196 \cdot 10^1}{2}\right) = fl(0.98) = 0.980 \quad \text{esterno all'intervallo}$$

Se invece si utilizza la formula  $c_k = a_{k-1} + \frac{b_{k-1} - a_{k-1}}{2}$

si eseguono le seguenti operazioni:

$$s_1 = fl(0.986 - 0.983) = fl(0.003) = 0.300 \cdot 10^{-2}$$

$$s_2 = fl\left(\frac{s_1}{2}\right) = 0.150 \cdot 10^{-2}$$

$$\begin{aligned} s_3 &= fl(0.983 + s_2) = fl(0.9845) \\ &= 0.984 \quad \text{oppure} \quad 0.985 \quad \text{entrambi interni all'intervallo} \end{aligned}$$

**Pseudocodice dell'algoritmo di bisezione per il calcolo degli zeri di una funzione non lineare:**

**PSEUDOCODICE**

1. Se  $f(a) * f(b) < 0$  si pone  $a_0 := a$  e  $b_0 := b$

2. Finchè non risulta verificato il criterio di arresto:

$$\text{Poni: } x_{k+1} := \frac{a_k + b_k}{2}$$

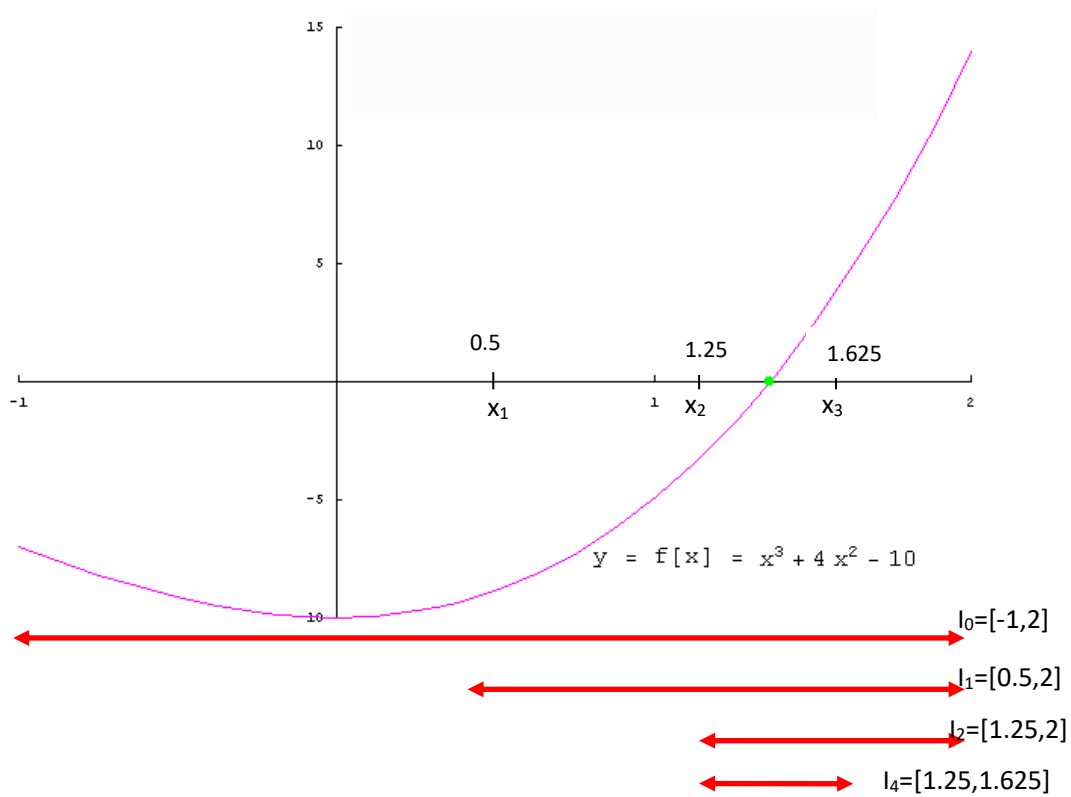
$$a) \text{ Se } f(x_{k+1}) * f(a_k) < 0 \Rightarrow a_{k+1} := a_k; b_{k+1} := x_{k+1}$$

$$b) \text{ Altrimenti se } f(x_{k+1}) * f(b_k) < 0 \Rightarrow a_{k+1} := x_{k+1}; b_{k+1} := b_k$$

$$c) \text{ Altrimenti se } f(x_{k+1}) = 0 \Rightarrow x_{k+1} := \alpha$$

$$d) k = k + 1$$

**Esempio** Determinare lo zero di  $f(x)=x^3 +4x^2-10$  in  $[a,b] = [-1, 2]$ , usando il metodo di Bisezione.



## Metodo della Regula Falsi

Una spiegazione della lenta convergenza del metodo di bisezione può essere ricercata nel fatto che il metodo non trae alcun vantaggio da caratteristiche peculiari della funzione, come la sua derivabilità o la sua forma. Addirittura, il metodo non tiene neanche conto dei valori della funzione, ma soltanto dei segni.

Un modo naturale per migliorare il metodo di bisezione è quello di considerare anche i valori che la funzione assume negli estremi dell'intervallo. Si considera come nuova approssimazione della soluzione l'intersezione dell'asse delle ascisse con la retta passante per  $(a, f(a))$   $(b, f(b))$

$$\begin{cases} y - f(a) = \frac{f(b) - f(a)}{b - a} (x - a) \\ y = 0 \end{cases}$$

da cui si ottiene:

$$x = a - f(a) \frac{(b-a)}{f(b)-f(a)}$$

Il metodo risultante è noto come metodo della **regula falsi o della falsa posizione**

Il metodo genera una successione di intervalli in cui è contenuta la radice: la scelta dell'intervallo in base al segno della funzione, comporta una convergenza globale.

E' più veloce rispetto al metodo di bisezione (**convergenza superlineare**).

In generale l'ampiezza dell'intervallo  $[a_i, b_i]$  non tende a zero pertanto il criterio di arresto basato sull'ampiezza dell'intervallo non è applicabile.

**Pseudocodice dell'algoritmo di falsa posizione per il calcolo degli zeri di una funzione non lineare:**

1. Se  $f(a) * f(b) < 0$  si pone  $a_0 := a$  e  $b_0 := b$

2. Finchè non risulta verificato il criterio di arresto:

$$\text{Poni: } x_{k+1} := a_k - f(a_k) \frac{b_k - a_k}{f(b_k) - f(a_k)}$$

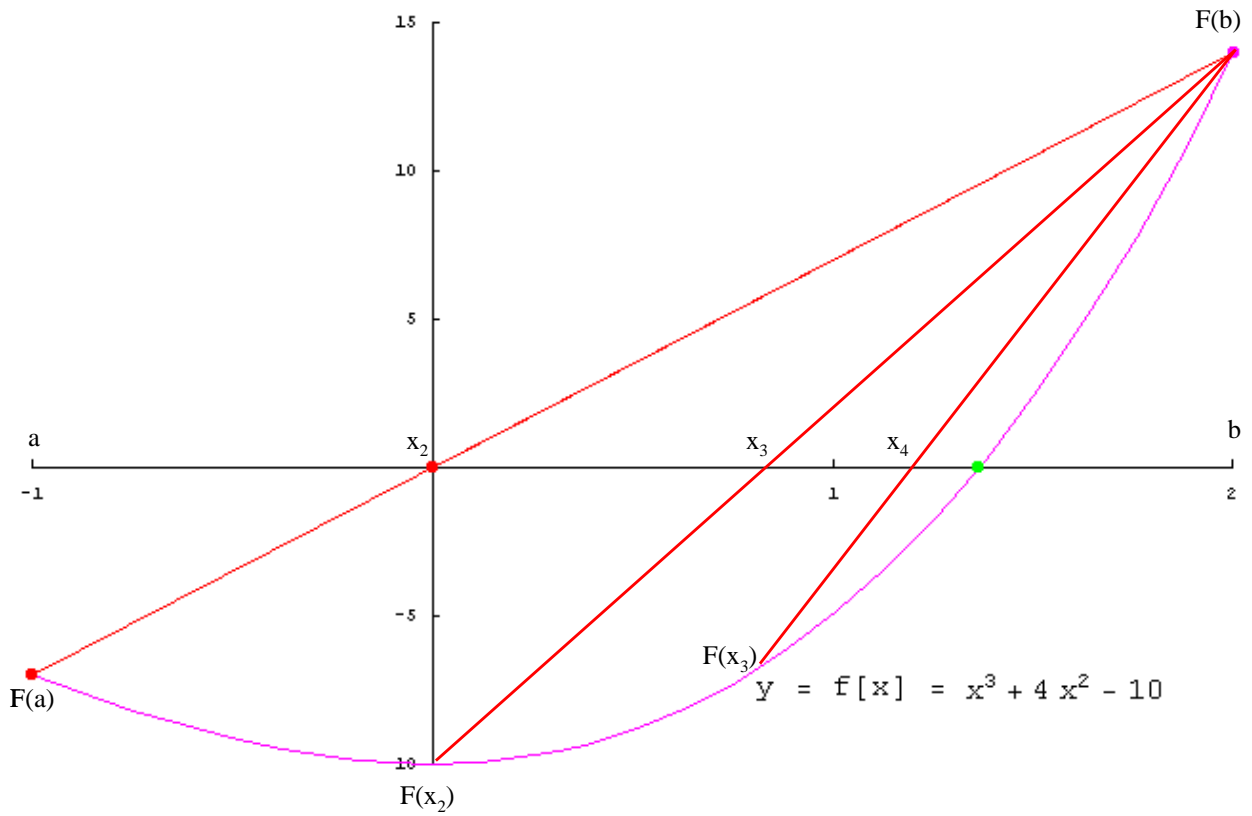
a) Se  $f(x_{k+1}) * f(a_k) < 0 \Rightarrow a_{k+1} := a_k; b_{k+1} := x_{k+1}$

b) Altrimenti se  $f(x_{k+1}) * f(b_k) < 0 \Rightarrow a_{k+1} := x_{k+1}; b_{k+1} := b_k$

c) Altrimenti se  $f(x_{k+1}) = 0 \Rightarrow x_{k+1} := \alpha$

d)  $k = k + 1$

**Esempio:** Determinare lo zero di  $f(x)=x^3+4x^2-10$  in  $[a,b] = [-1, 2]$ , usando il metodo della Regula Falsi



## Metodi di linearizzazione

Data  $f(x), x_0, f(x_0)$ : si approssima la funzione con una retta per  $(x_0, f(x_0))$

$$y = f(x_0) + m(x - x_0)$$

Si ottiene una versione linearizzata del problema  $f(x) = 0$ ,

$$\begin{cases} y = f(x_0) + m(x - x_0) \\ y = 0 \end{cases}$$

da cui

$$x_1 = x_0 - \frac{f(x_0)}{m}$$

In generale

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

A seconda della scelta di  $m_k$  si ottengono:

**metodo delle corde ( $m_k = m = \text{costante}$ ),**

**metodo delle secanti e metodo di Newton**

## Metodo delle corde

Utilizza un valore costante  $m \neq 0$ .

Il metodo assume la forma

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

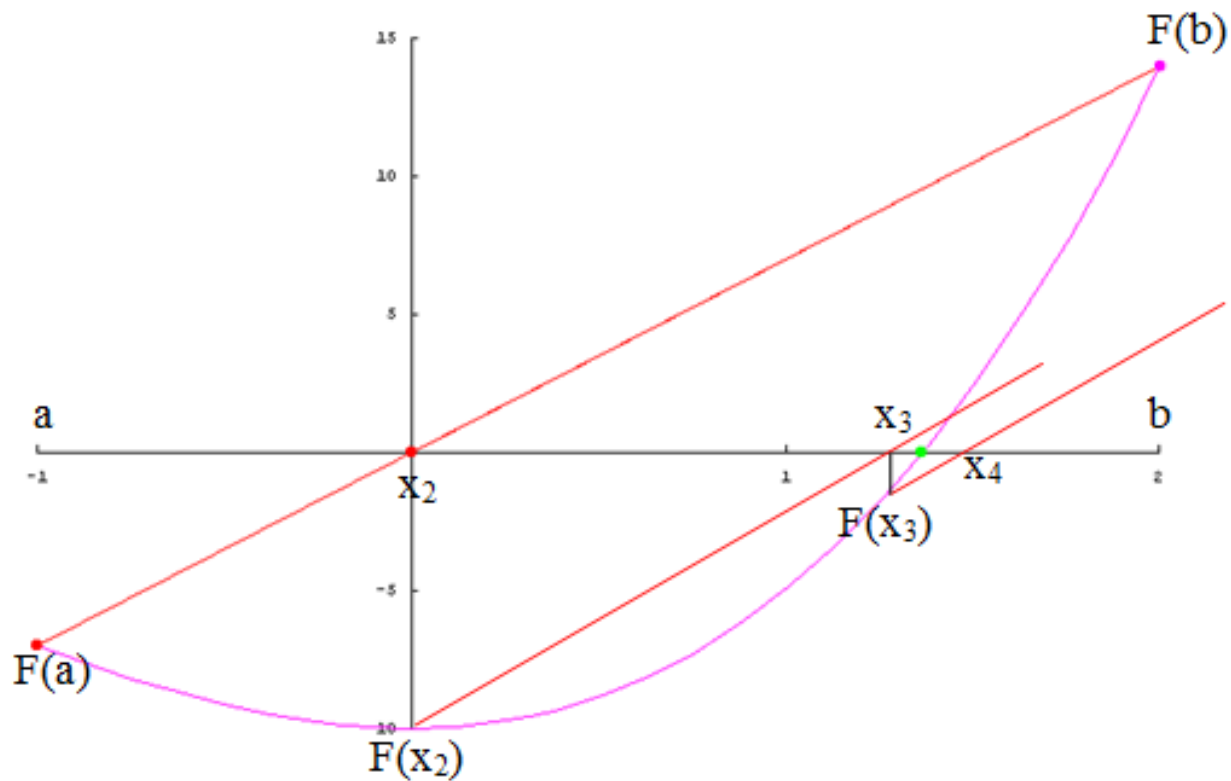
Una scelta classica è quella di utilizzare il coefficiente angolare della retta che congiunge i punti  $(a, f(a))$  e  $(b, f(b))$ .

$$m = \frac{f(b) - f(a)}{b - a}$$

da cui si ottiene la formula ricorsiva

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

✚ Determinare lo zero di  $f(x)=x^3 + 4x^2 - 10$  in  $[a,b] = [-1, 2]$ , usando il metodo delle Corde.





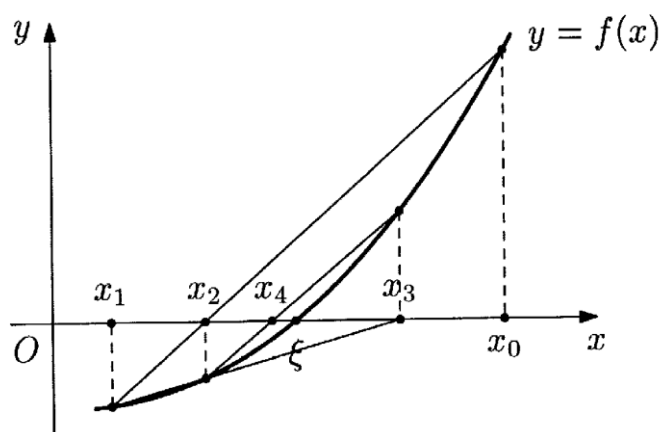
## Metodo delle secanti

Assegnati i due valori iniziali  $x_0, x_1$ , al passo  $k$  l'approssimazione della funzione  $f$  nell'intervallo  $[x_{k-1}, x_k]$  è la retta che passa per i punti  $(x_{k-1}, f(x_{k-1})), (x_k, f(x_k))$  con coefficiente angolare

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

che interseca l'asse  $x$  nel punto di ascissa

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

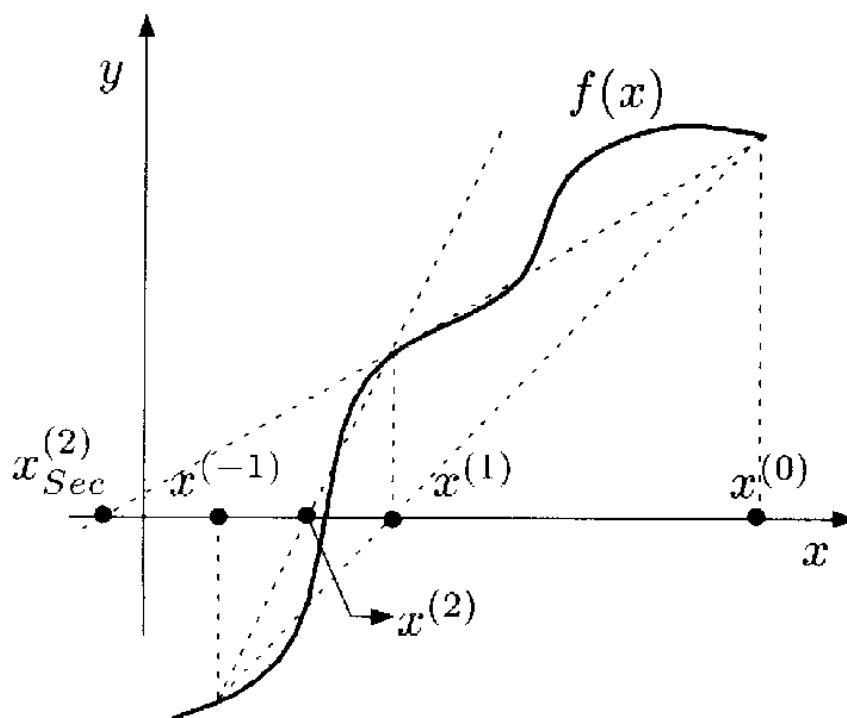
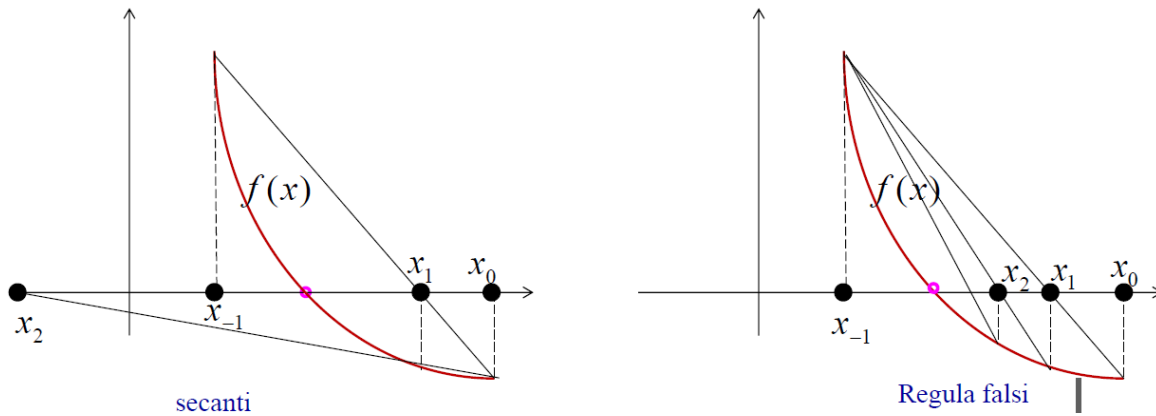


La convergenza del metodo è garantita se le approssimazioni  $x_0$  ed  $x_1$  si scelgono abbastanza vicine alla soluzione: **convergenza locale**. In tal caso la convergenza è

**superlineare** ( $p = \frac{(1+\sqrt{5})}{2} \approx 1.618$ ).

### Confronto tra regola falsi e secanti

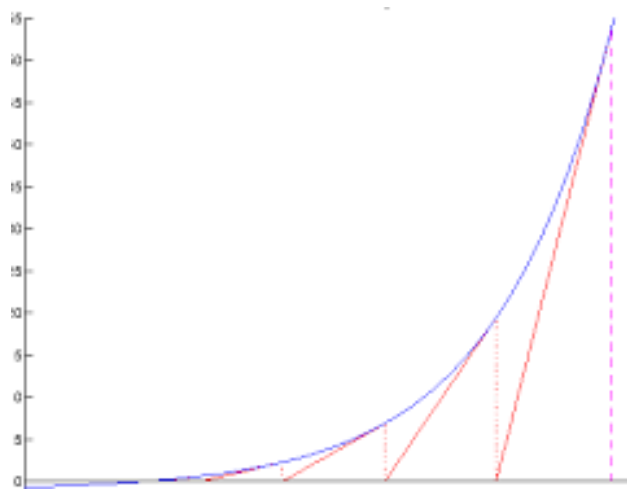
Il metodo delle secanti può essere più veloce ma non converge sempre. Non c'è più la certezza di avere sempre il punto cercato all'interno dell'intervallo.



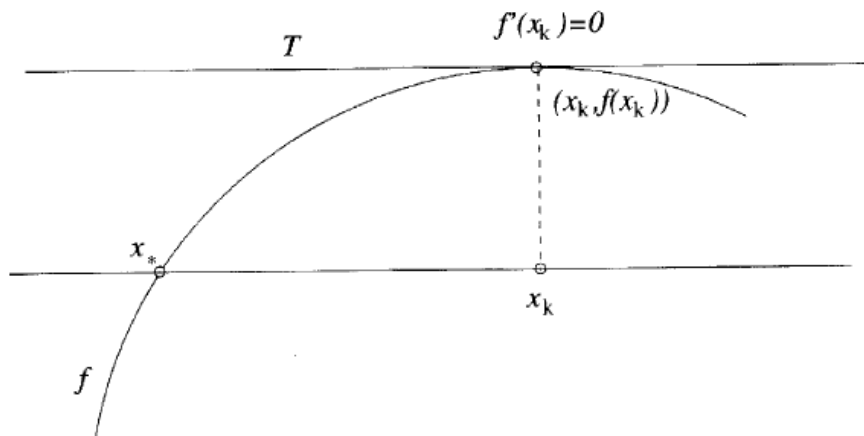
### Metodo di Newton

Nel metodo di Newton, ad ogni passo  $k$ , si considera la retta passante per il punto  $(x_k, f(x_k))$  e tangente alla curva  $f(x)$  e si determina il nuovo iterato come il punto di incontro tra questa retta e l'asse delle  $x$ . Per far ciò, nella formula ricorsiva (4), si pone  $m_k = f'(x_k)$ , ottenendo la formula ricorsiva

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



## Problemi con il metodo di Newton:



## Ordine del metodo di Newton:

Mettiamoci nell'ipotesi in  $\alpha$  sia **una radice semplice di  $f(x)$** , cioè  **$f(\alpha)=0$  ed  $f'(\alpha)\neq 0$** . Consideriamo lo sviluppo del primo ordine di  $f(x)$  in un intorno di  $x_k$

$$f(x) = f(x_k) + (x - x_k)f'(x_k) + \frac{1}{2}(x - x_k)^2 f''(\zeta)$$
per un opportuno  $\zeta$  compreso tra  $x$  ed  $x_k$ .

Valutiamo in  $\alpha$

$$f(\alpha) = 0 = f(x_k) + (\alpha - x_k)f'(x_k) + \frac{1}{2}(\alpha - x_k)^2 f''(\zeta)$$

Dividendo per  $f'(x_k)$ :

$$\frac{f(x_k)}{f'(x_k)} + (\alpha - x_k) + \frac{\frac{1}{2}(\alpha - x_k)^2 f''(\zeta)}{f'(x_k)} = (\alpha - x_{k+1}) + \frac{\frac{1}{2}(\alpha - x_k)^2 f''(\zeta)}{f'(x_k)}$$
$$= 0$$

Ricordiamo che  $e_{k+1} = x_{k+1} - \alpha$ ,

$-e_{k+1} + \frac{1}{2} \frac{e_k^2 f''(\zeta)}{f'(x_k)} = 0 \Rightarrow e_{k+1} = e_k^2 \frac{1}{2} \frac{f''(\zeta)}{f'(x_k)}$  da cui, ricordando che  $\lim_{k \rightarrow \infty} x_k = \alpha$ , segue che

$$\frac{e_{k+1}}{e_k^2} \xrightarrow{k \rightarrow \infty} \frac{1}{2} \frac{f''(\alpha)}{f'(\alpha)} \quad \text{Ordine } p=2, \text{ fattore di convergenza } \frac{1}{2} \frac{f''(\alpha)}{f'(\alpha)}$$

### Osservazione:

Se  $\alpha$  è uno zero di molteplicità  $m$ , cioè se  $f^{(k)}(\alpha) = 0, k = 0, \dots, m-1$  e  $f^{(m)}(\alpha) \neq 0$ , allora il metodo di Newton non ha più convergenza quadratica. Si dimostra che diventa a convergenza lineare del tipo

$$|x_{k+1} - \alpha| \approx c|x_k - \alpha|$$

con

$$c = \frac{m-1}{m}.$$

Per esempio per radici doppie,  $m=2$  e quindi  $c = \frac{1}{2}$ .

Metodo di Newton modificato per radici di molteplicità  $m > 1$

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

Si dimostra che l'ordine di convergenza del metodo di Newton modificato è 2.

### Ordine dei metodi:

Metodo di Newton	$p=2$	convergenza quadratica
Metodo delle secanti	$p=1.618$	convergenza superlineare
Metodo regula falsi		convergenza superlineare
Metodo di bisezione	$p=1$	convergenza lineare

## Metodi a convergenza globale

La convergenza è assicurata per qualsiasi scelta del punto iniziale appartenente all'intervallo che racchiude la radice, cioè  $x_0$  in  $[a, b]$  (Bisezione, Regula Falsi)

## Metodi a convergenza locale

La convergenza è assicurata per  $x_0$  appartenente ad un intorno della soluzione.  
(Secanti, Newton)

## Teorema di convergenza locale:

Se  $f : [a; b] \rightarrow \mathbb{R}$  soddisfa le seguenti ipotesi

i)  $f(a)f(b) < 0$

ii)  $f, f', f''$  sono continue in  $[a; b]$ , ossia  $f \in C^2[a; b]$

iii)  $f'(x) \neq 0 \forall x \in [a, b]$

allora esiste un intorno  $I \subset [a, b]$  dell'unica radice  $\alpha \in (a, b)$  tale che, se  $x \in I$ , allora la successione di Newton  $\{x_i\}_{i \geq 1}$  converge ad  $\alpha$ .

Se la funzione soddisfa alcune condizioni, allora esiste un teorema che garantisce che il metodo di Newton converge globalmente:

## Teorema (di convergenza globale del metodo di Newton)

Sia  $f(x) \in C^2[a, b]$ ,  $[a, b]$  intervallo chiuso e limitato. Se sono verificate le seguenti condizioni

1.  $f(a)f(b) < 0$
2.  $f'(x) \neq 0 \quad \forall x \in [a, b]$
3.  $f''(x) > 0$  oppure  $f''(x) < 0 \quad \forall x \in [a, b]$
4.  $\left| \frac{f(a)}{f'(a)} \right| < b - a \quad \left| \frac{f(b)}{f'(b)} \right| < b - a$

allora il metodo di Newton **converge** all'unica soluzione  $\alpha$  in  $[a, b]$ ,  
**per ogni scelta di  $x_0$  in  $[a, b]$ .**

- La condizione 1. assicura che una radice in  $(a,b)$  esista.
- La condizione 2. assicura che non vi siano tangenti orizzontali. Questo garantisce, insieme ad 1., che vi sia una sola radice interna ad  $(a,b)$ .
- La condizione 3. assicura che la concavità o convessità sia mantenuta su tutto  $[a,b]$ . Questo garantisce, insieme a 2., che gli iterati  $x_k$  siano monotoni per  $k > 1$ .
- La condizione 4. assicura che le tangenti agli estremi intersecano l'asse  $x$  internamente ad  $(a,b)$ . In tal modo tutti gli iterati  $x_k$  (escluso eventualmente  $x_0$ ) risulteranno interni ad  $(a,b)$ .

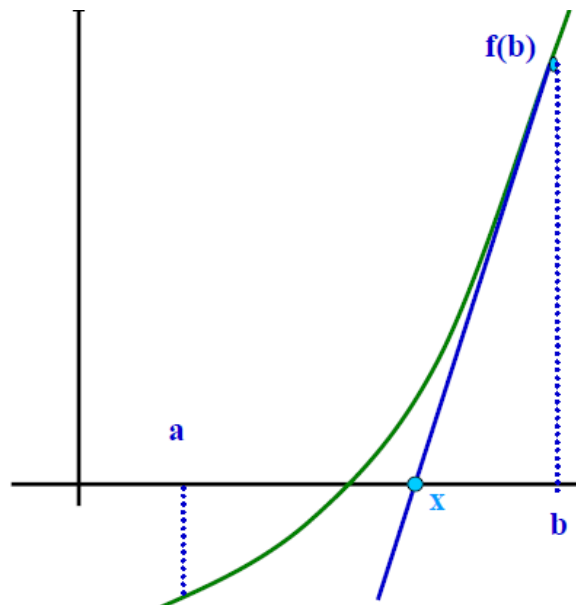
$$\begin{cases} y - f(b) = f'(b)(x - b) \\ y = 0 \end{cases}$$

Infatti, se consideriamo l'intersezione della tangente nell'estremo  $b$  con l'asse  $x$ , si ha

Da cui segue se

$$\left| \frac{f(b)}{f'(b)} \right| = |x - b| < b - a$$

La tangente negli estremi interseca l'asse  $x$  all'interno dell'intervallo  $[a,b]$



### Metodi ibridi

Il metodo di Newton e il metodo delle secanti sono metodi a convergenza locale. La difficoltà pratica sta nel trovare l'intervallo di convergenza, cioè nel trovare un valore iniziale  $x_0$  tale che la successione di iterati converga alla soluzione cercata. Un metodo pratico è quello di far precedere questi metodi da un metodo a convergenza globale come ad esempio il metodo di bisezione. Dopo alcuni passi del metodo globale si innesca quello di ordine superiore. Se esso non converge, si devono fare altri passi del metodo globale.



## Sistemi di equazioni non lineari

Un sistema di equazioni non lineari può essere scritto nella forma

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ f_3(x_1, x_2, \dots, x_n) = 0 \\ \dots \\ f_n(x_1, x_2, \dots, x_n) = 0 \end{cases} \quad (1)$$

con  $f_i: R^n \rightarrow R$ ,  $i = 1, \dots, n$  funzioni non lineari continue e differenziabili.

Consideriamo la funzione  $F: R^n \rightarrow R^n$ , **funzione a valori vettoriali**,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ \dots \\ x_n \end{bmatrix} \in R^n \rightarrow F(X) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ f_3(x_1, x_2, \dots, x_n) \\ \dots \\ f_n(x_1, x_2, \dots, x_n) \end{bmatrix},$$

calcolare **la soluzione del sistema (1)**, cioè calcolare il vettore  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T \in R^n$  che annulla contemporaneamente le equazioni **equivale a calcolare  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T \in R^n$  tale che  $F(\alpha)=0$** .

Il gradiente di una funzione  $f: R^n \rightarrow R$ , differenziabile è dato da

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \dots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

Lo **Jacobiano** di  $F(X)$  è la matrice

$$J(X) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

Si ha inoltre che  $\nabla F(X) = J^T(X)$

### Esempio 1:

Calcolare il punto di intersezione tra il cerchio di coordinate  $x_1^2 + x_2^2 = 9$  e la retta  $x_1 + x_2 = 3$ .

Si devono trovare i punti che annullano simultaneamente le funzioni

$$f_1(x_1, x_2) = x_1^2 + x_2^2 - 9 \quad \text{e} \quad f_2(x_1, x_2) = x_1 + x_2 - 3$$

Si tratta quindi di risolvere il sistema non lineare

$$\begin{cases} f_1(x_1, x_2) = x_1^2 + x_2^2 - 9 = 0 \\ f_2(x_1, x_2) = x_1 + x_2 - 3 = 0 \end{cases}$$

## Metodo di Newton Raphson per la soluzione di un sistema di equazioni nonlineari

### Caso n=2

Data

$$F(X) = F(x_1, x_2) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}$$

Individuare il vettore  $\alpha = [\alpha_1, \alpha_2]^T \in R^2$  tale che  $F(\alpha) = 0$ .

Consideriamo lo sviluppo in serie di Taylor troncato al primo ordine di ciascuna delle

due funzioni in un intorno del punto  $X_k = [x_1^{(k)}, x_2^{(k)}]^T \in R^2$

$$\begin{cases} 0 = f_1(X) \approx f_1(X_k) + \frac{\partial f_1}{\partial x_1}(x_1^{(k)})(x_1 - x_1^{(k)}) + \frac{\partial f_1}{\partial x_2}(x_2^{(k)})(x_2 - x_2^{(k)}) \\ 0 = f_2(X) \approx f_2(X_k) + \frac{\partial f_2}{\partial x_1}(x_1^{(k)})(x_1 - x_1^{(k)}) + \frac{\partial f_2}{\partial x_2}(x_2^{(k)})(x_2 - x_2^{(k)}) \end{cases} \quad (2)$$

Indicato con

$$J(X) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}$$

lo jacobiano di  $F(X)$ , la relazione (2) si può esprimere in forma matriciale come

$$F(X) = 0 \approx F(X_k) + J(X_k)(X - X_k)$$

$$J(X_k)X = J(X_k)X_k - F(X_k)$$

Sotto l'ipotesi che  $\det J(X_k) \neq 0$ , si ricava  $X$  premoltiplicando ambo i membri per  $J^{-1}(X_k)$

$$X = X_k - J^{-1}(X_k)F(X_k)$$

e si determina il procedimento iterativo:

$$X_{k+1} = X_k - J^{-1}(X_k)F(X_k)$$

Si osserva che  $-J^{-1}(X_k)F(X_k)$  è la soluzione del sistema lineare

$$J(X_k)s_k = -F(X_k)$$

quindi

$$X_{k+1} = X_k + s_k$$

L'algoritmo di Newton Raphson si può così schematizzare

L'algoritmo si può così schematizzare

Dato  $X_0 \in R^n$  ed  $F$ , per ogni iterazione  $k$

1. Valutare  $J(X_{k-1})$
2. Risolvere il sistema lineare  $J(X_{k-1})s_{k-1} = -F(X_{k-1})$
3. Porre  $X_k = X_{k-1} + s_{k-1}$

**E' un metodo a convergenza locale e ordine di convergenza quadratico.**

**Nel caso**

$$F(X) = F(x_1, x_2) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2^2 - 9 \\ x_1 + x_2 - 3 \end{bmatrix}$$

$$J(X) = \begin{bmatrix} 2x_1 & 2x_2 \\ 1 & 1 \end{bmatrix}$$

## Varianti del Metodo di Newton-Raphson

La valutazione dello Jacobiano richiede di conoscere o poter valutare  $n^2$  derivate parziali.

Alcune varianti al metodo possono migliorarne l'efficienza.

### 1. Approssimazione con rapporti incrementali

Sostituire a  $J(X_{k-1})$ , il cui calcolo esplicito può essere molto costoso, una sua approssimazione ottenuta mediante rapporti incrementali n-dimensionali del tipo

$$\left. \frac{\partial f_j}{\partial X_i} \right|_{X=X_{k-1}} \approx (J^{(k-1)})_{ij} = \frac{f_j(X_{k-1} + e_i h_{ij}) - f_j(X_{k-1})}{h_{ij}}$$

Dove  $e_i$  è l'i-esimo vettore della base canonica  $R^n$

$h_{ij}$  incremento scelto ad ogni passo k.

Il metodo che si ottiene è l'analogo n-dimensionale di quello delle secanti,

### 2. Metodo delle corde.

Si utilizza lo stesso Jacobiano  $J(X_0)$  o una sua approssimazione oppure  $A(X_0)$  per tutte le iterazioni k. Si potrebbe quindi fattorizzare  $J(X_0)=LU$  e utilizzare i medesimi L ed U per ogni iterazione.

### 3. Metodo di Shamanskii

Si valuta lo Jacobiano ogni m iterazioni e quindi lo si utilizza per le m iterazioni successive:

$$J^{k+i}=J^i \quad i=1,\dots,m$$

Giunti al calcolo di  $x_{k+m+1}$  si rivaluta lo Jacobiano.

## Metodo di Newton-Raphson per il calcolo del minimo di una funzione in più variabili.

Data  $f: R^n \rightarrow R$ ,  $f \in C^2$  (differenziabile due volte con continuità), trovare  $x^* \in R^n$  tale che  $x^* = \arg \min_{x \in R^n} f(x)$ .

I punti di stazionarietà locale  $x^*$  sono soluzione del seguente sistema non lineare

$$\nabla f(x^*) = 0$$

Applichiamo il metodo di Newton Raphson al sistema non lineare

$$\nabla f(x) = 0 \Rightarrow \begin{cases} \frac{\partial f(x)}{\partial x_1} = 0 \\ \frac{\partial f(x)}{\partial x_2} = 0 \\ \dots \\ \frac{\partial f(x)}{\partial x_n} = 0 \end{cases}$$

Per verificare poi se tale punto è un massimo, un minimo oppure un punto sella, occorrerà esaminare la matrice hessiana  $H(x) = \nabla^2 f(x)$  in questo punto, dove la matrice hessiana è così definita:

$$(H(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \dots, n$$
$$H(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial^2 x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial^2 x_2} & \dots & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \dots & \frac{\partial^2 f(x)}{\partial^2 x_n} \end{bmatrix}$$

Se il **determinante della matrice Hessiana** calcolata nel punto di **stazionarietà**

- **è positivo e l'elemento in posizione (1,1) è positivo**, si tratta di un minimo locale, ;
- **è positivo e l'elemento in posizione (1,1) è negativo**, si tratta di un massimo locale,
- **è negativo**, si tratta di un punto sella
- **è nullo**, non ci sono informazioni sulla natura del punto di stazionarietà.

Consideriamo uno sviluppo di Taylor del primo ordine di  $\nabla f(x)$  in un intorno di  $x_k \in R^n$ , trascurando il resto

$$\nabla f(x) = 0 \approx \nabla f(x_k) + H(x_k)(x - x_k)$$

$$H(x_k)x = H(x_k)x_k - \nabla f(x_k)$$

**L'algoritmo di Newton-Raphson per la minimizzazione diventa:**

Dato  $x_0 \in R^n$  ed  $F$ , per ogni iterazione  $k$

1. Valutare  $H(x_{k-1})$
2. Risolvere il sistema lineare  $H(x_{k-1})s_{k-1} = -\nabla f(x_{k-1})$
3. Porre  $x_k = x_{k-1} + s_{k-1}$

$s_{k-1}$  definisce una direzione di discesa da  $x_{k-1}$  ad  $x_k$