Vittorio Casella

DIET – Università di Pavia

email: vittorio.casella@unipv.it

I Minimi Quadrati non lineari
Introduzione

Dispense

Breve introduzione

Nei minimi quadrati lineari, la relazione fra il vettore delle incognite $\mathbf{X}$ e quello delle osservazioni $\mathbf{Y}$ è

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

dove

$$\mathbf{Y} \in M(m,1)$$

$$\mathbf{X} \in M(n,1)$$

$$\mathbf{A} \in M(m,n)$$

$$\mathbf{b} \in M(m,1)$$

Vi sono fenomeni invece, in cui la relazione fra le grandezze incognite e quelle osservate è di tipo non lineare

$$\mathbf{Y} = \mathbf{g}(\mathbf{X})$$

Esempio di MQ non lineari

Viene discusso il problema della determinazione delle coordinate di un punto con misure di sola distanza: tecnica detta *trilaterazione* in topografia classica. La discussione è ambientata nel piano, ma il formalismo può essere banalmente esteso al caso 3D. Si fa notare tra l'altro che il caso 3D del problema considerato è la base del *point positioning* (o soluzione navigazionale) nel GPS.

La trilaterazione - 1

Consideriamo alcuni punti P_i di coordinate note

$$P_i = \left(x_i, y_i\right) \qquad i = 1,2,\ldots,k \tag{1}$$

Consideriamo un punto P di coordinate incognite

$$P = \left(x, y\right) \tag{2}$$

Si conoscono le distanze d_i fra i primi e il secondo. La relazione che lega le grandezze in gioco è data dal teorema di Pitagora

$$d_i = \sqrt{\left(x_i - x\right)^2 + \left(y_i - y\right)^2} \tag{3}$$

Per un i fissato, note d_i e $\left(x_i, y_i\right)$, la (3) è l'equazione di un cerchio avente centro in $\left(x_i, y_i\right)$ e raggio d_i .

La trilaterazione - 2

Se si considera un sistema di equazioni

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} \qquad i = 1,2,\ldots,k \qquad (4)$$

esso rappresenta l'intersezione dei k cerchi. Se il problema non è banale (punti P_i coincidenti), vi è evidentemente un solo punto di intersezione e questo è P.

La soluzione del sistema (4) fornisce dunque le coordinate cercate.

Per motivi pratici si considerano in genere le equazioni (4) al quadrato

$$d_i^2 = (x_i - x)^2 + (y_i - y)^2 \qquad i = 1,2,\ldots,k \qquad (5)$$

che sono equivalenti.

La trilaterazione - 3

Nel casi pratici, gli errori accidentali contenuti nelle distanze d_i (e anche nelle coordinate (x_i, y_i), che qui non vengono considerati) rendono il sistema (5) insolubile nel senso classico, quando $k > 2$. Equivalentemente, dal punto di vista geometrico, i k cerchi (5) non si intersecano esattamente, dunque bisogna cercare una soluzione generalizzata e approssimata.

L'idea per la soluzione ai MQ è trovare un vettore $\hat{\mathbf{x}}$ tale che le distanze

$$\left\| \hat{\mathbf{x}} - \mathbf{x}_i \right\|$$

Siano, se non nulle, minime, per tutti i k punti considerati.

Il formalismo MQ della trilaterazione

Cerchiamo di ricondurre al formalismo dei MQ il sistema (5)

$$\mathbf{Y} = \begin{pmatrix} d_1^2 \\ d_2^2 \\ \vdots \\ d_k^2 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad \mathbf{g}(\mathbf{X}) = \begin{pmatrix} (x_1 - x)^2 + (y_1 - y)^2 \\ (x_2 - x)^2 + (y_2 - y)^2 \\ \vdots \\ (x_k - x)^2 + (y_k - y)^2 \end{pmatrix} \tag{6}$$

Si tratta evidentemente di una relazione non lineare, del tipo

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}) \tag{7}$$

non riconducibile al formalismo MQ lineare classico

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

Idee per la soluzione di problemi MQ non lineari

Ripercorriamo quanto fatto nel caso lineare: è stata considerata la funzione del modello

$$\mathbf{Y}(\mathbf{X}) = \mathbf{A}\mathbf{X} + \mathbf{b} \tag{8}$$

e la soluzione è stata definita

$$\hat{\mathbf{X}} := \min_{\mathbf{X}} \left\| \mathbf{Y}_0 - \mathbf{Y}(\mathbf{X}) \right\|^2 = \min_{\mathbf{X}} \left\| \mathbf{Y}_0 - \mathbf{A}\mathbf{X} - \mathbf{b} \right\|^2 \tag{9}$$

Nel caso non lineare esiste ancora la funzione del modello

$$\mathbf{Y}(\mathbf{X}) = \mathbf{g}(\mathbf{X}) \tag{10}$$

e potremmo definire la soluzione MQ, in analogia con (9), come

$$\hat{\mathbf{X}} := \min_{\mathbf{X}} \left\| \mathbf{Y}_0 - \mathbf{Y}(\mathbf{X}) \right\|^2 = \min_{\mathbf{X}} \left\| \mathbf{Y}_0 - \mathbf{g}(\mathbf{X}) \right\|^2 \tag{11}$$

Ma la linearità di (8) rende il problema (9) facilmente risolubile in forma chiusa

$$\hat{\mathbf{X}} = \left(\mathbf{A}^t \mathbf{A} \right)^{-1} \mathbf{A}^t \left(\mathbf{Y}_0 - \mathbf{b} \right)$$

Idee per la soluzione di problemi MQ non lineari - 2

La non linearità di (10) implica che, in generale, non si conosce la soluzione per (11). Per alcuni problemi si conosce, ma non in generale. Quando non si sa risolvere un problema direttamente, si cerca di approssimarlo con problemi risolubili: è possibile linearizzare la (10)?

Serve una soluzione approssimata $\mathbf{X}_0$. Si può effettuare uno sviluppo in serie della $\mathbf{g}(\mathbf{X})$ in un intorno di $\mathbf{X}_0$. Per fare questo è necessario determinare la matrice jacobiana della funzione

$$\mathbf{g}(\mathbf{X}) \overset{1°}{=} \{\mathbf{J}(\mathbf{g})\}(\mathbf{X}_0)\left(\mathbf{X}-\mathbf{X}_0\right)+\mathbf{g}(\mathbf{X}_0) \tag{12}$$

Si può allora pensare di risolvere il problema

$$\mathbf{Y} = \{\mathbf{J}(\mathbf{g})\}(\mathbf{X}_0)\left(\mathbf{X}-\mathbf{X}_0\right)+\mathbf{g}(\mathbf{X}_0) \tag{13}$$

Invece del problema originario

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}) \tag{14}$$

La soluzione di problemi MQ non lineari - 1

Il problema è lineare è può essere ricondotto al formalismo standard nel modo

$$\mathbf{Y} = \begin{pmatrix} d_1^2 \\ d_2^2 \\ \vdots \\ d_k^2 \end{pmatrix} \qquad \Delta\mathbf{X} = \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \qquad \mathbf{A} = \left\{\mathbf{J(g)}\right\}(\mathbf{X}_0) \qquad \mathbf{b} = \mathbf{g}(\mathbf{X}_0)$$

In modo che la (13) assuma la forma

$$\mathbf{Y} = \mathbf{A}\,\Delta\mathbf{X} + \mathbf{b} \tag{15}$$

Il problema ausiliario (15) ha un vettore delle incognite $\Delta\mathbf{X}$ diverso da $\mathbf{X}$. Tuttavia, una volta trovato $\widehat{\Delta\mathbf{X}}$, si può porre

$$\hat{\mathbf{X}} = \mathbf{X}_0 + \widehat{\Delta\mathbf{X}} \tag{16}$$

La soluzione di problemi MQ non lineari - 2

Domande:

- la soluzione del problema (11) esiste ed è unica? Dipende dal problema?
- La soluzione del problema (15) coincide con quella del (11), se esiste? No dipende dal problema e anche dalla soluzione approssimata scelta

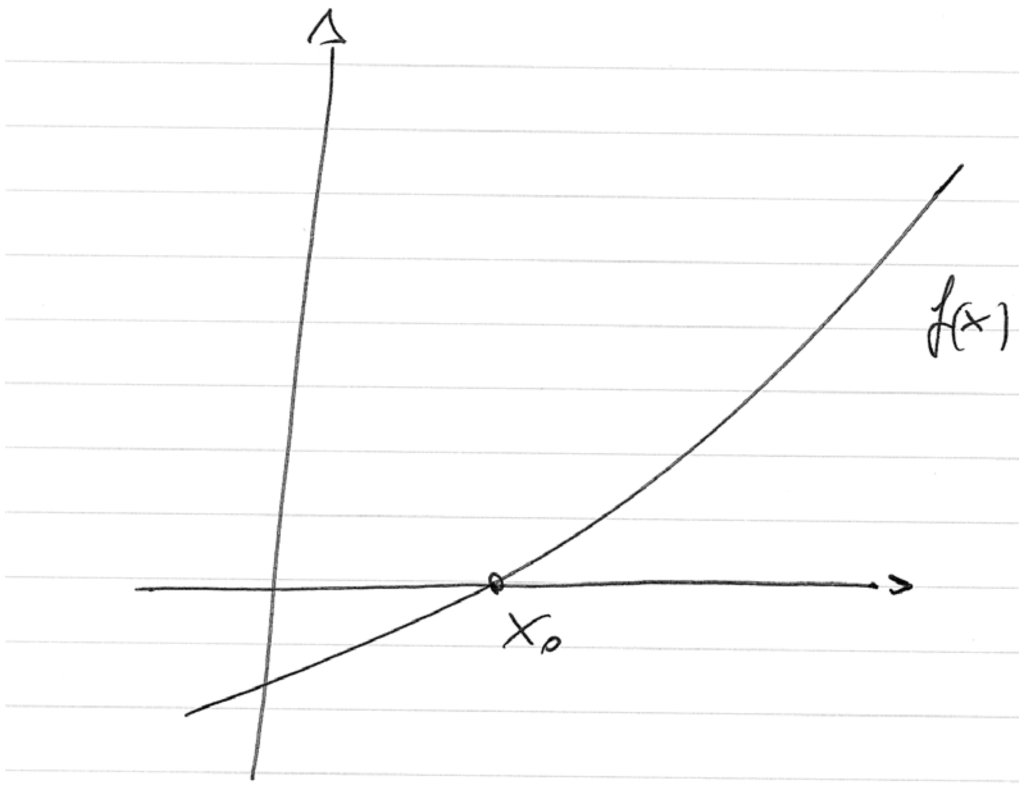
La soluzione del problema originario non è necessariamente unica.

Ammesso che il problema originario abbia soluzione $\mathbf{x}^*$, non detto che la soluzione (16) coincida con essa.

Il metodo di Newton per la soluzione di equazioni – 1

Attenzione, quanto esposto non riguarda direttamente la soluzione del problema (15), anche se vi è stretta connessione. Quanto esposto è invece utile per capire, per analogia, i problemi che si incontrano risolvendo un problema mediante linearizzazione.

Come noto esistono funzioni per cui si sanno trovare in forma chiusa le radici: la retta e la parabola, per esempio. In generale, tuttavia, data una $f(x)$, non si sa calcolare in forma chiusa il valore x_0 per cui $f(x_0)=0$. Omettiamo diversi dettagli: esiste? E' unico?



[esempio_radice_equazione.png]

Il metodo di Newton per la soluzione di equazioni – 2

Il metodo di Newton approssima un problema

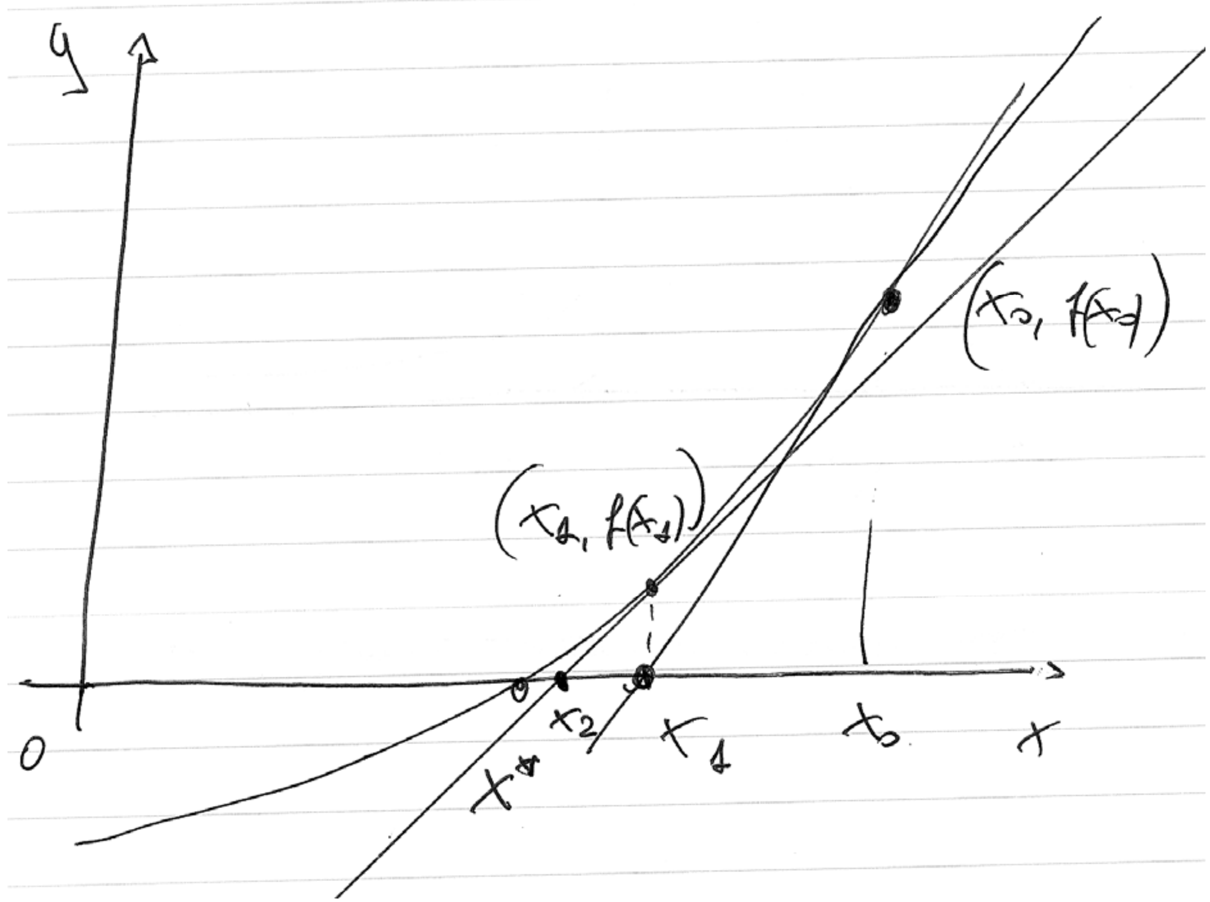
$$f(x)=0$$

che non si sa risolvere (avente soluzione x^*) con una successione di problemi che sono risolubili in forma chiusa.

E' necessario disporre di una soluzione approssimata x_0.

Si considera la retta tangente la curva nel punto di ascissa x_0.

Essa interseca l'asse delle ascisse in un punto x_1.



[metodo_newton_1.png]

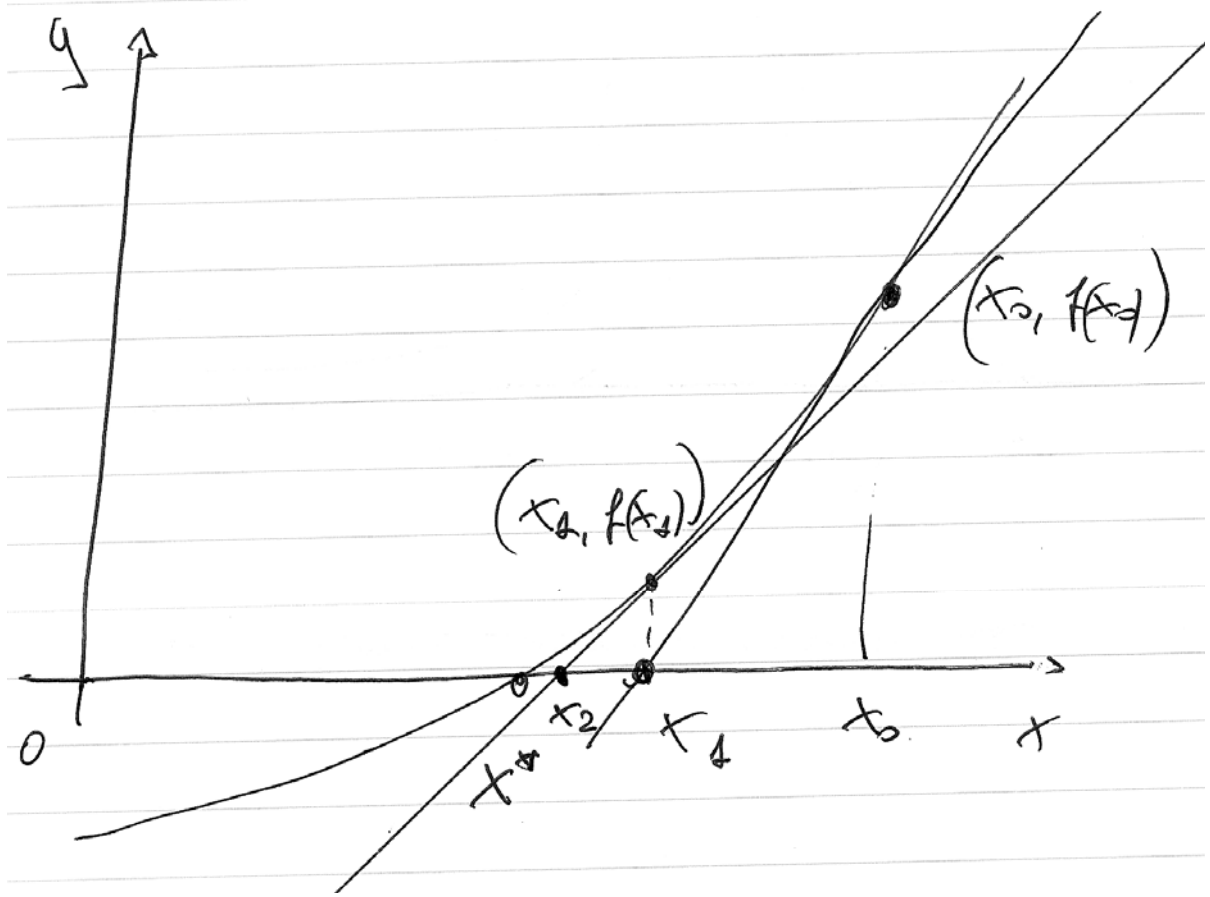
Il metodo di Newton per la soluzione di equazioni – 3

Il punto x_1 coincide con la soluzione x^*? No ma, se non siamo sfortunati, sarà più vicino alla soluzione di x_0.

Ripetiamo in x_1 quanto fatto in x_0: approccio iterativo.

Dopo un certo numero di iterazioni la soluzione x_{i+1} non sarà significativamente diversa da x_i. E' il momento di arrestare le iterazioni e porre

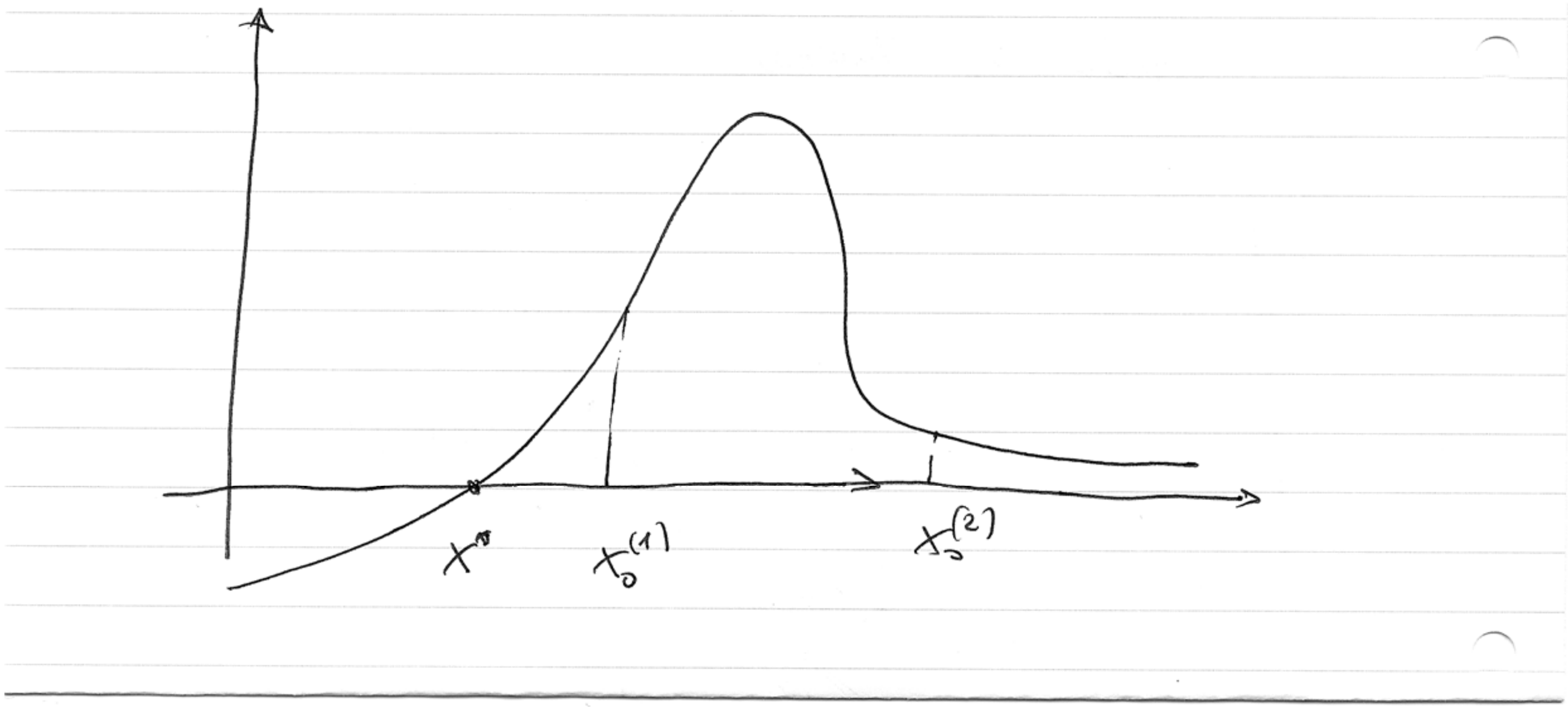
$$\hat{x}^* := x_i$$



Si parla di avvenuta convergenza della successione.

Il metodo di Newton per la soluzione di equazioni – 4

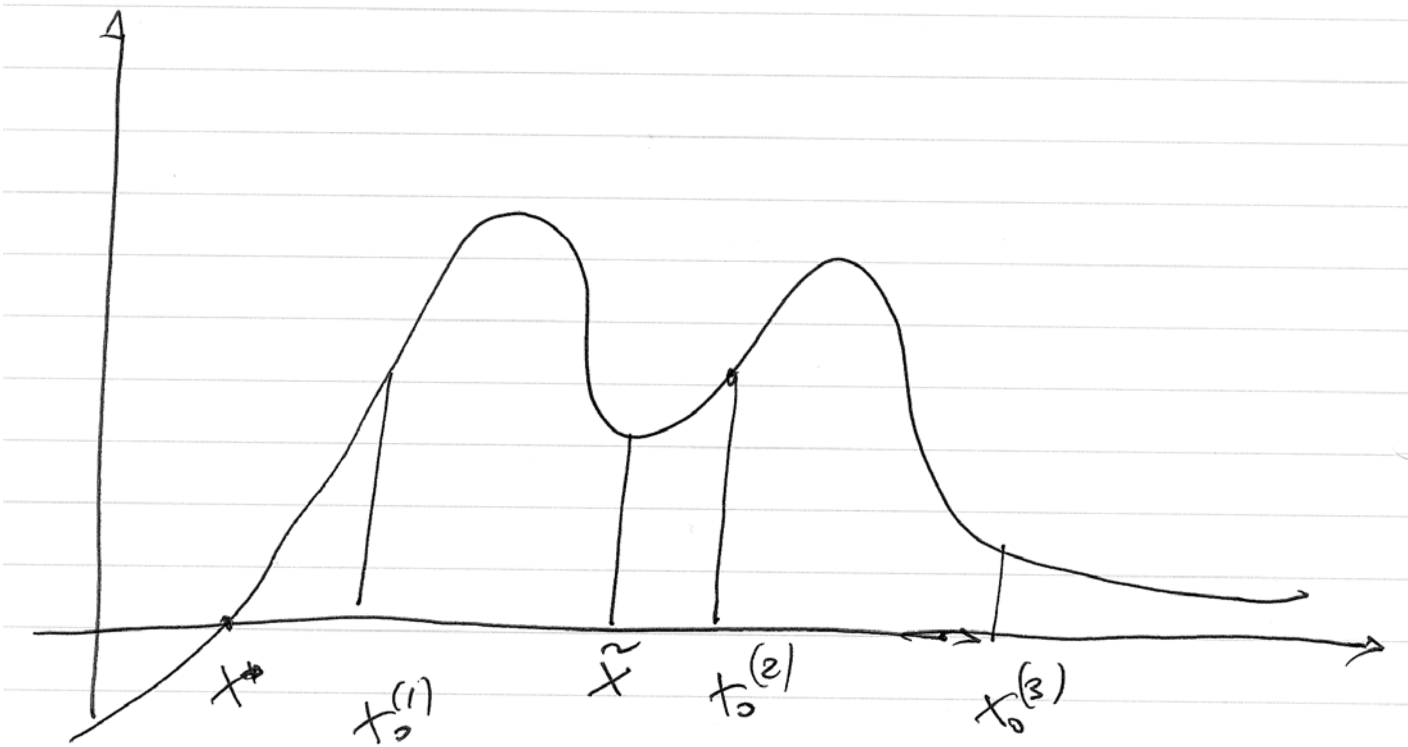
Problemi. Se si adotta come soluzione approssimata $x_0^{(1)}$ la successione di Newton converge alla soluzione vera. Se si sceglie $x_0^{(2)}$ la soluzione di Newton si allontanerà indefinitamente verso $+\infty$



[metodo_newton_2.png]

Il metodo di Newton per la soluzione di equazioni – 5

Problemi. Se si adotta come soluzione approssimata $x_0^{(1)}$ la successione di Newton converge alla soluzione vera. Se si sceglie $x_0^{(2)}$ la soluzione di Newton tenderà a un valore $\tilde{x}$ scorretto. Se si sceglie $x_0^{(3)}$ la soluzione di Newton si allontanerà indefinitamente verso $+\infty$.



[metodo_newton_3.png]

Sintesi sui problemi non lineari

Soluzione ottenuta iterativamente

Può essere che la soluzione converga a un valore scorretto

Può essere che la soluzione non converga

Dipende dal tipo di problema e dalla soluzione approssimata.

Tuttavia, in genere le cose funzionano.

Soluzione iterativa per il problema della trilaterazione

Cerchiamo la jacobiana della funzione

$$\mathbf{g}(\mathbf{X}) = \begin{pmatrix} (x_1 - x)^2 + (y_1 - y)^2 \\ (x_2 - x)^2 + (y_2 - y)^2 \\ \vdots \\ (x_k - x)^2 + (y_k - y)^2 \end{pmatrix} \qquad (17)$$

Si ha

$$\{\mathbf{J}(\mathbf{g})\}((x,y)) = -2 \begin{pmatrix} x_1 - x & y_1 - y \\ x_2 - x & y_2 - y \\ \vdots & \vdots \\ x_k - x & y_k - y \end{pmatrix}$$

Soluzione iterativa per il problema della trilaterazione - 2

Da cui

$$\mathbf{J}_0 = \{\mathbf{J}(\mathbf{g})\}\big((x_0, y_0)\big) = -2\begin{pmatrix} x_1 - x_0 & y_1 - y_0 \\ x_2 - x_0 & y_2 - y_0 \\ \vdots & \vdots \\ x_k - x_0 & y_k - y_0 \end{pmatrix}$$

Il problema deve essere risolto iterativamente: la prima soluzione $\widehat{\Delta\mathbf{X}_1}$ consente di stimare

$$\hat{\mathbf{X}}_1 = \mathbf{X}_0 + \widehat{\Delta\mathbf{X}_1}$$

Il vettore $\hat{\mathbf{X}}_1$ può essere usato per effettuare una ulteriore linearizzazione. E così via…

La matrice jacobiana - 1

Sia $\mathbf{g}:\mathbb{R}^n \to \mathbb{R}^m$ funzione a valori vettoriali. La sua matrice jacobiana è una matrice

$$\left\{\mathbf{J}(\mathbf{g})\right\}:\mathbb{R}^n \to \mathbb{R}^{m\times n}$$

che è funzione di $\mathbf{x}\in\mathbb{R}^n$ ed è così definita

$$\mathbf{J}=\left\{\mathbf{J}(\mathbf{g})\right\}(\mathbf{x})=\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}) & \frac{\partial g_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial g_2}{\partial x_1}(\mathbf{x}) & \frac{\partial g_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial g_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1}(\mathbf{x}) & \frac{\partial g_n}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial g_n}{\partial x_n}(\mathbf{x}) \end{pmatrix} \tag{18}$$

La matrice jacobiana - 2

Indicando con $J_{ij}\left(\mathbf{x}\right)$ la componente $\left(i,j\right)$ della matrice $\left\{\mathbf{J}(\mathbf{g})\right\}(\mathbf{x})$, si può riassumere la (18) con

$$J_{ij}\left(\mathbf{x}\right)=\frac{\partial g_i}{\partial x_j}\left(\mathbf{x}\right)$$