Sistemi di equazioni non-lineari

$$F: U \subseteq \mathbb{R}^n \to \mathbb{R}^n$$

es: (n=2)

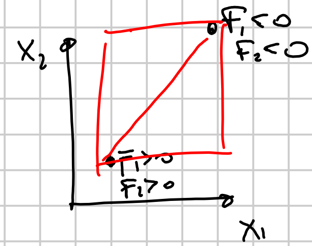
$$F(x) = \begin{bmatrix} 2x_1^2 x_2 + \cos(x_2) \\ \sin(x_1) + 2x_2 - \pi \end{bmatrix} = 0 \qquad \begin{cases} 2x_1^2 x_2 + \cos(x_2) = 0 \\ \sin(x_1) + 2x_2 - \pi = 0 \end{cases}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Vediamo metodi per trovare (approssimativamente) una soluzione di un sistema di questo tipo.

⚠ Più difficili di sistemi lineari, ad es. caratterizzare quante soluzioni ci sono.

Non c'è generalizzazione del metodo di bisezione



Possiamo generalizzare metodi di punto fisso $x_{k+1} = \Phi(x_k)$

e $\boxed{\text{metodo di Newton}}$

Def: matrice Jacobiana di una funzione $F: U \subseteq \mathbb{R}^m \to \mathbb{R}^n$

è definita come

$$J_F = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_m} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1} & \frac{\partial F_n}{\partial x_2} & \cdots & \frac{\partial F_n}{\partial x_m} \end{bmatrix} : U \subseteq \mathbb{R}^m \to \mathbb{R}^{n \times m}$$

ES: $F(x) = \begin{bmatrix} 2x_1^2 x_2 + \cos(x_2) \\ \sin(x_1) + 2x_2 - \pi \end{bmatrix}$ $\qquad J_F(x) = \begin{bmatrix} 4x_1 \cdot x_2 & 2x_1^2 - \sin(x_2) \\ \cos(x_1) & 2 \end{bmatrix}$

Per funzioni suff. regolari $F: U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ $\qquad x, h \in \mathbb{R}^n$

$$F(x+h) = F(x) + J_F(x) \cdot h + O(\|h\|^2)$$

$$[\;] = [\;] + [J_F][\;] + [\;]$$

Sviluppo nel punto x_k, impongo che $x_k + h = x_{k+1}$ soddisfi $F(x_{k+1}) \simeq 0$ nello svil. di Taylor

$$\boxed{0 = F(x_k) + J_F(x_k)\cdot h} \qquad x_{k+1} = x_k + h$$

$$\Updownarrow$$

$$\boxed{h = -J_F(x_k)^{-1}\cdot F(x_k)}$$

Metodo di Newton per sistemi in più variabili:

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ dato} \\ x_{k+1} = x_k - J_F(x_k)^{-1}\cdot F(x_k) \qquad k=0,1,2,\dots \end{cases}$$

Costo computazionale: ad ogni passo: 1 valutazione di F
1 valutazione di J_F
1 soluzione di un sistema lineare $O(n^3)$
(con fatt. LU o varianti)

Convergenza: se partiamo sufficientemente vicino a una soluzione $x_* \in \mathbb{R}^n$, e se $J_F(x_*)$ è invertibile, allora il metodo converge localmente, e la convergenza è di ordine 2 (quadratica), cioè $\dfrac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^2} \to C \in [0,\infty)$

Metodo delle corde:

$$\begin{cases} x_0 \text{ dato} \\ x_{k+1} = x_k - M^{-1} F(x_k) \qquad k=0,1,2,\dots \end{cases}$$

con $M \in \mathbb{R}^{n\times n}$ (o la sua inversa) matrice data

Per esempio, posso scegliere $M = J_F(x_0)$ e tenerla fissa senza cambiarla ad ogni passo.

Vantaggio: ad ogni passo vado a risolvere un sist. lineare con la stessa matrice M, quindi posso calcolare una volta all'inizio $M = LU$ (o un'altra fattorizz.), e ad ogni passo risolvere solo $\begin{cases} Ly = b \\ Ux = y \end{cases}$ (o analoghi per un'altra fattorizzazione)

Costo computazionale: Una volta all'inizio: 1 valutazione di J_F
1 fattorizz. $J_F = LU$ $O(n^3)$

una volta per passo: 1 valutazione di F
1 soluzione di sistema utilizzando la LU già calcolata $O(n^2)$

Convergenza: lineare (quando c'è) anziché quadratica
Li vedremo in pratica.

Sistemi sovradeterminati (problemi ai minimi quadrati)

Supponiamo di avere un problema del tipo

$$Ax = b \quad , \quad A \in \mathbb{R}^{m\times n} \quad x \in \mathbb{R}^n , \quad b \in \mathbb{R}^m \qquad m > n \quad (A \text{ alta e stretta})$$

"sovradeterminato"

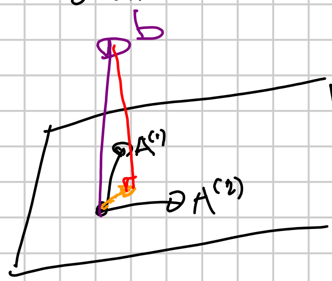
$\boxed{m \text{ equazioni (le righe) in } n \text{ incognite}}$ (le componenti $x_1, x_2, \dots x_n$)

$\Rightarrow$ in generale non possiamo aspettarci di trovare una soluzione

(iperpiano n-dim. in $\mathbb{R}^m$)

Ax può assumere solo valori in $\mathrm{Im}\, A$

b non sta per forza all'interno di questo piano



$m=3, n=2$: $\mathrm{Im}\, A$ è un piano nello spazio,
b non sta per forza nel piano.

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 0 & 0 \end{bmatrix} , \quad b = \begin{bmatrix} 4 \\ 4 \\ 1 \end{bmatrix}$$

$\mathrm{Im}\, A = \{ x_3 = 0 \}$

No sol. di $Ax=b$ in generale, posso però chiedermi qual è il vettore x tale che Ax più si avvicina a b

Chiamiamo residuo il vettore $r = Ax - b$ e mi chiedo qual è la soluzione di

$$(*) \qquad \min_{x \in \mathbb{R}^n} \| \underbrace{Ax - b}_{= r} \|$$

Facile da risolvere se scelgo la norma-2. In questo caso, $(*)$ è equivalente a risolvere

$$\min_{x \in \mathbb{R}^n} \| \underbrace{Ax - b}_{r} \|^2 = \underbrace{r_1^2 + r_2^2 + \dots + r_m^2}_{\text{"problema dei minimi quadrati"}} = r^T r = [r_1 \; r_2 \; \dots \; r_m] \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

ES con le A, b di sopra,
$$\begin{cases} x_1 + x_2 = 4 \\ 2x_1 + x_2 = 4 \\ 0x_1 + 0x_2 = 1 \end{cases}$$
non ha soluzione, ma posso cercare x_1, x_2 che minimizzano

$$\|r^2\| = (x_1 + x_2 - 4)^2 + (2x_1 + x_2 - 4)^2 + (-1)^2$$

Teo: sia $x \in \mathbb{R}^n$ un vettore tale che $\underbrace{A^T}_{A^T \in \mathbb{R}^{n \times m}} \cdot (\underbrace{Ax-b}_{\in \mathbb{R}^m}) = 0$

Allora, x risolve il problema $(*)$.

Dim: Vogliamo dimostrare che dato un qualunque $y \in \mathbb{R}^n$, si ha

$$\|\underbrace{Ax-b}_{\text{residuo di } x}\|^2 \leq \|\underbrace{Ay-b}_{\text{residuo di } y}\|^2$$

$$\underbrace{Ay-b}_{\text{residuo di } y} = Ay - Ax + Ax - b = \underbrace{A(y-x)}_{s \in \mathbb{R}^m} + \underbrace{(Ax-b)}_{\substack{r \in \mathbb{R}^m \\ \text{residuo di } x}}$$

$$\|Ay-b\|^2 = \|s+r\|^2 = (s+r)^T(s+r) = s^Ts + s^Tr + r^Ts + r^Tr.$$

Notiamo che il prodotto scalare $s^Tr = r^Ts$ è uguale a zero: difatti,

$$s^Tr = (A(y-x))^T(Ax-b) = (y-x)^T \underbrace{A^T(Ax-b)}_{=0} = 0$$

Allora $\|Ay-b\|^2 = \underbrace{s^Ts}_{\checkmark} + \underbrace{r^Tr}_{= \|Ax-b\|^2}$ residuo di x

$s^Ts = s_1^2 + s_2^2 + \dots + s_m^2 \geq 0$

Questo mostra che $\|Ay-b\|^2 \geq \|Ax-b\|^2$. $\square$

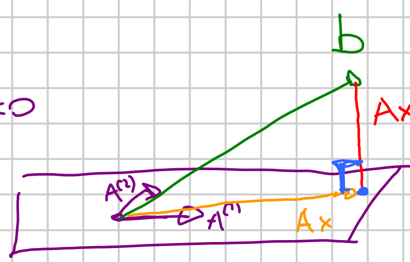
Quindi il problema di minimo $(*)$ è risolto trovando x che verifica

$$\underbrace{A^T(Ax-b) = 0}_{\text{"sistema delle equazioni normali"}} \qquad \boxed{\underbrace{A^TA}\,x = \underbrace{A^Tb}} \sim \text{sistema lineare per } x$$

$$\underset{n \times n}{[\ \ ]}\ \underset{\mathbb{R}^n}{[\,]} = \underset{\mathbb{R}^n}{[\,]}$$

Si può dimostrare che A^TA è SPD (simmetrica e pos. definita), e quindi è invertibile, tutte le volte che le colonne di A sono lin. indipendenti.
$\Rightarrow$ posso risolvere il sistema con la fattorizzazione LDL / Cholesky.

Idea geometrica:
le equazioni $A^T(Ax-b)=0$ dicono che le colonne di A hanno prodotto scalare 0 con $Ax-b=r$



Se le colonne di A sono lin. indipendenti, A^TA è invertibile e possiamo risolvere il sistema.

⚠ possibile problema: la matrice A^TA può essere mal condizionata.
Il condizionamento di questo sist. lineare $(A^TA)x = A^Tb$ può essere peggiore di quello del problema iniziale

$$K(A^TA) = \|A^TA\|\cdot\|(A^TA)^{-1}\|.$$

Metodo alternativo per risolvere il problema: basandosi sulla <u>fattorizzazione QR</u>.

<u>Teo:</u> Per ogni $A\in\mathbb{R}^{m\times n}$ (anche con $m\neq n$, matrice rettangolare), esistono $Q\in\mathbb{R}^{m\times m}$ <u>ortogonale</u> (cioè $Q^TQ=I$), e $R\in\mathbb{R}^{m\times n}$ triangolare superiore (cioè, tutti gli R_{ij} con $i>j$ sono 0) tali che $A=Q\cdot R$,

$$\underset{m\times n}{\square} = \underset{m\times m}{\square}\cdot\underset{m\times n}{\square}$$

Se $m>n$ (A alta e stretta),

$$R = \begin{array}{c} n \\ \left[\begin{array}{ccccc} x & x & x & \cdots & x \\ 0 & x & x & & \\ 0 & 0 & x & & \\ 0 & 0 & 0 & \ddots & \\ \vdots & \vdots & \vdots & 0 & x \\ \hline & & & & 0 \\ & & & & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{array}\right] \end{array} \begin{array}{l} \}\, n \\ \\ \}\, m-n \text{ righe di zeri} \end{array} = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

$R_1\in\mathbb{R}^{n\times n}$ triang. sup.

$$A = Q\cdot R = \left[\; \overset{n}{Q_1} \;\middle|\; \overset{m-n}{Q_2} \;\right]\cdot\begin{bmatrix} R_1 \\ \hline 0 \end{bmatrix}\begin{array}{l} \}\,n \\ \}\,m-n \end{array} = Q_1 R_1$$

si scontrano con zeri quando faccio il prodotto

$$\underset{m\times n}{\boxed{Q_1}}\;\underset{n\times n}{\boxed{R_1}}$$

Facendo alcune manipolazioni algebriche, possiamo mostrare che

$x = (A^TA)^{-1}(A^Tb) = R_1^{-1}Q_1^Tb$, cioè, la soluzione x del problema $(*)$ è la soluzione del sist. lineare $\boxed{R_1x = Q_1^Tb}$

Algoritmo alternativo per $(*)$:

1) Calcolo Q,R tali che $A=QR$, fattorizz. QR di A (o anche solo Q_1, R_1)
2) Risolvo il sistema $\underbrace{R_1}_{\text{triang. sup.}}x = \underbrace{Q_1^Tb}_{\text{vettore da calcolare}}$

Entrambi gli algoritmi (equazioni normali, QR) costano $O(mn^2)$

lineare nella dim. più grande

quadratici nella dim. più piccola