Metodi iterativi per sistemi lineari

$A = M - N$ splitting

$$Ax = b \iff x = M^{-1}(b + Nx) \rightsquigarrow \begin{cases} x^{(0)} \in \mathbb{R}^n \text{ dato} \\ x^{(k+1)} = M^{-1}(b + N x^{(k)}) \end{cases}$$

<u>Teo</u>: il metodo converge (per ogni scelta di $x^{(0)}$) se e solo se $\rho(H) < 1$. $\quad H = M^{-1}N$

Metodo di Jacobi: $M = \begin{bmatrix} A_{11} & & 0 \\ & A_{22} & \\ & & \ddots & \\ 0 & & & A_{nn} \end{bmatrix}$

Metodo di Gauss-Seidel: $M = \begin{bmatrix} A_{11} & & & 0 \\ A_{21} & A_{22} & & \\ \vdots & & \ddots & \\ A_{n1} & \cdots & & A_{nn} \end{bmatrix}$

<u>Teo</u>: i metodi di J. e G.S. sono convergenti se A è una matrice predominante diagonale (per righe o per colonne)

Ricordiamo le definizione: A è pred. diag. <u>per righe</u> se

$$\forall i = 1, 2, \dots n \qquad |A_{ii}| > \sum_{j=1,\, j \neq i}^{n} |A_{ij}|$$

e pred. diag. <u>per colonne</u> se

$$\forall j = 1, 2, \dots n \qquad |A_{jj}| > \sum_{i=1,\, i \neq j}^{n} |A_{ij}|$$

Dimostriamolo per pred. diag. per <u>righe</u> (per colonne è analogo).

Mostriamo innanzitutto che i metodi sono <u>applicabili</u>:

$$|A_{ii}| > \sum_{j\neq i} |A_{ij}| \geq 0 \quad \text{quindi} \quad |A_{ii}|>0, \text{ cioè } A_{ii}\neq 0 \quad \forall i=1,\dots n$$

Per la convergenza, vogliamo usare il criterio che $\rho(H)<1$, quindi vogliamo dimostrare che ogni autovalore di H ha modulo strettamente minore di 1.

λ è autovalore di H se e solo se

$$\det(H-\lambda I)=0 \qquad \text{(polinomio caratteristico)}$$

$$\Leftrightarrow \det(M^{-1}N-\lambda \underbrace{I}_{M^{-1}M})=0 \Leftrightarrow \det(M^{-1}(N-\lambda M))=0$$

$$\Leftrightarrow \underbrace{\det(M^{-1})}_{\neq 0 \text{ perché } M^{-1} \text{ è invertibile}}\underbrace{\det(N-\lambda M)}=0$$

gli autovalori di H sono esattamente i λ per cui $\det(N-\lambda M)=0$

Per mostrare che tutti gli autovalori di H hanno $|\lambda|<1$, mi basta far vedere che quando $|\lambda|\geq 1$, $B=N-\lambda M$ è invertibile.

Vogliamo mostrare che se A è predom. diagonale, e $|\lambda|\geq 1$, allora anche B è predom. diagonale.

$$A=\begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn}\end{bmatrix} \qquad \underline{\text{Jacobi}}: \; M=\begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{nn}\end{bmatrix} \qquad N=-\begin{bmatrix} 0 & A_{21} \cdots & A_{1n} \\ A_{21} & \ddots & A_{n-1,n} \\ \vdots & \ddots & \\ A_{n1} & \cdots A_{n,n-1} & 0\end{bmatrix}$$

$$B=-\begin{bmatrix} \lambda A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & \lambda A_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_{n-1,n} \\ A_{n1} & \cdots & A_{n,n-1} & \lambda A_{nn}\end{bmatrix}$$

$$|B_{ii}|=|\lambda||A_{ii}|\geq |A_{ii}| > \sum_{j\neq i}|A_{ij}| = \sum_{j\neq i}|B_{ij}|$$

e questa catena di uguaglianze mostra che anche B è pred. diagonale.

Gauss-Seidel:

$$M=\begin{bmatrix} A_{11} & & & 0 \\ A_{21} & \ddots & & \\ \vdots & & \ddots & \\ A_{n1} & \cdots & A_{n,n-1} & A_{nn} \end{bmatrix} \qquad N=-\begin{bmatrix} 0 & A_{12} & \cdots & A_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & A_{n-1,n} \\ 0 & & & 0 \end{bmatrix}$$

$$B=N-\lambda M=-\begin{bmatrix} \lambda A_{11} & A_{12} & \cdots & A_{1n} \\ \lambda A_{21} & \lambda A_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_{n-1,n} \\ \lambda A_{n,1} & \cdots & \lambda A_{n,n-1} & \lambda A_{nn} \end{bmatrix}$$

Poiché A è dom. diagonale:

$$|A_{ii}| > \sum_{j=1}^{i-1} |A_{ij}| + \sum_{j=i+1}^{n} |A_{ij}| \qquad i=1,2,3,\dots n$$

Quindi

$$|\lambda A_{ii}| > \sum_{j=1}^{i-1} |\lambda A_{ij}| + \sum_{j=i+1}^{n} |\lambda A_{ij}| \underset{|\lambda|\geq 1}{\geq} \sum_{j=1}^{i-1} |\lambda A_{ij}| + \sum_{j=i+1}^{n} |A_{ij}|$$

$$|B_{ii}| > \quad \dots \quad \sum_{j=1}^{i-1} |B_{ij}| + \sum_{j=i+1}^{n} |B_{ij}|$$

Quindi, ricapitolando: se $|\lambda|\geq 1$, $B=N-\lambda M$ è pred. diagonale e quindi invertibile

$\Rightarrow$ $\det(M^{-1})\det(N-\lambda M)\neq 0$

$\Rightarrow$ $\det(H-\lambda I)\neq 0$ ogni volta che $|\lambda|\geq 1$

$\Rightarrow$ tutti gli autovalori di H hanno modulo <1

$\Rightarrow$ i metodi di Jacobi e Gauss-Seidel convergono. □

Sistemi lineari sovradeterminati

Sono sistemi lineari in cui ho più equazioni che incognite:

$$Ax=b \qquad A\in\mathbb{R}^{m\times n} \quad m>n \quad , \quad x\in\mathbb{R}^{n} \quad b\in\mathbb{R}^{m}$$

es: $\begin{cases} x_1+x_2=2 & \text{I} \\ x_1-x_2=1 & \text{II} \\ x_2=1 & \text{III} \end{cases}$ $\qquad A=\begin{bmatrix}1 & 1\\ 1 & -1\\ 0 & 1\end{bmatrix} \quad x=\begin{bmatrix}x_1\\ x_2\end{bmatrix} \quad b=\begin{bmatrix}2\\ 1\\ 1\end{bmatrix}$

Se prendiamo x che risolve I, II: $x=\begin{bmatrix}1.5\\ 0.5\end{bmatrix}$ $\qquad Ax-b=\begin{bmatrix}0\\ 0\\ -0.5\end{bmatrix}$

Se prendiamo x che risolve II, III: $x=\begin{bmatrix}2\\ 1\end{bmatrix}$ $\qquad Ax-b=\begin{bmatrix}1\\ 0\\ 0\end{bmatrix}$

Però, per il primo vettore $\|Ax-b\|$ è più piccolo.

Questo suggerisce di risolvere questo problema:

per quale vettore x $\|r(x)\|=\|Ax-b\|$ è più piccolo possibile?

distanza dal centro

ES: $P \approx Mx_1 + Dx_2$

prezzo di una casa

numero di m^2 superficie

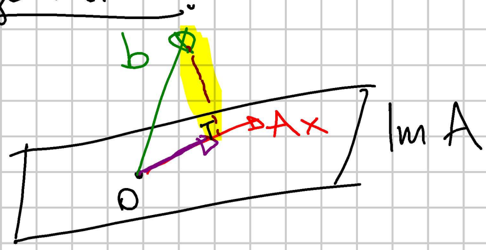
Un dataset con $m>2$ case di cui so prezzo, superficie, distanza mi dà m equazioni, tipicamente non tutte soddisfatte dallo stesso $\begin{bmatrix}x_1\\ x_2\end{bmatrix}$.

Però posso trovare $\begin{bmatrix}x_1\\ x_2\end{bmatrix}$ che rende più piccolo possibile la differenza tra prezzi esatti e stimati.

Questo mi permette di "apprendere" a valutare case sulla base dei dati nel mio dataset

$\rightarrow$ <u>regressione lineare</u> $\rightarrow$ machine learning.

Idea geometrica:



Voglio trovare il vettore dentro Im A che più si avvicina al vettore b

Oss: questo problema dipende dalla scelta della norma:

cerco x t.c. $\|Ax-b\|$ è più piccola possibile.

Diamo una soluzione per la norma-2:

$$r(x) = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \qquad \|r(x)\|_2 = \sqrt{r_1^2 + r_2^2 + \dots + r_m^2}$$

(detto anche: problema dei minimi quadrati)

Nell'esempio sopra: $\sqrt{(x_1+x_2-2)^2 + (x_1-x_2-1)^2 + (x_2-1)^2} = \|r(x)\|_2$

Minimizzare questa norma è la stessa cosa che minimizzare il suo quadrato, togliendo la radice.

Teo: sia $x \in \mathbb{R}^n$ un vettore che soddisfa $A^T(Ax-b)=0$.
Allora, x risolve il problema di minimo

$$\min \|Ax-b\|_2$$

Dim: vogliamo mostrare che per ogni $z \neq 0$, $x+z$ ha residuo maggiore o uguale di quello di x:

$$\|r(x+z)\|_2 \geq \|r(x)\|_2 \qquad r(x) = Ax-b$$

$$r(x+z) = A(x+z)-b = Ax+Az-b = \underbrace{Ax-b}_{r(x)} + Az$$

$r(x) =: r$

$Az =: s$

$$\|r+s\| \geq \|r\|$$

$$\|r\|_2^2 = r^T r = r_1 \cdot r_1 + r_2 \cdot r_2 + r_3 \cdot r_3 + \dots + r_m \cdot r_m$$

$$\|r+s\|_2^2 = (r+s)^T(r+s) = r^T r + r^T s + s^T r + s^T s$$

$r^T s = s^T r = r_1 s_1 + r_2 s_2 + \dots + r_m s_m$ sono uguali,

e in particolare sono uguali a zero: $s = Az$ $r = Ax - b$

$$s^T r = (Az)^T (Ax-b) = z^T \underbrace{A^T(Ax-b)}_{\substack{= \\ 0 \text{ per ipotesi}}}$$

$$\underbrace{\|r+s\|_2^2}_{\|r(x+z)\|_2^2} = r^T r + s^T s = \|r\|_2^2 + \underbrace{s_1 \cdot s_1 + s_2 \cdot s_2 + \dots + s_m \cdot s_m}_{\text{somme di quadrati, } \geq 0} \geq \underbrace{\|r\|_2^2}_{\|r(x)\|_2^2}$$

$$\|r(x+z)\|_2^2 \geq \|r(x)\|_2^2 \quad \square$$

Quindi la soluzione del problema è data da quel vettore x che soddisfa

$$A^T(Ax-b) = 0 \iff A^T A x = A^T b$$

Notiamo che $\underset{n\times m \cdot m \times n}{A^T A} \in \mathbb{R}^{n\times n}$, quadrata, e $A^T b \in \mathbb{R}^n$, quindi questo è un sistema lineare con una matrice <u>quadrata</u>. (che già sappiamo risolvere)

Si può dimostrare (non lo vediamo) che $A^T A$ è invertibile se e solo se le colonne di A sono lin. indipendenti.

Algoritmo per risolvere il problema dei minimi quadrati

Input: $A \in \mathbb{R}^{m \times n}$ $m > n$, $b \in \mathbb{R}^m$ (equazioni normali)

Output: x t.c. $\|Ax - b\|_2$ è più piccolo possibile

1. Calcolo $A^T A$ $\quad n \times m \quad m \times n \quad O(mn^2)$
2. Calcolo $A^T b$ $\quad n \times m \quad m \times 1 \quad O(mn)$
3. Risolvo il sistema $(A^T A)x = A^T b$ $\quad O(n^3)$

Il termine maggiore (asintoticamente) è il primo, $O(mn^2)$.

Stabilità: purtroppo, esistono casi in cui il condizionamento di risolvere $(A^T A)x = A^T b$ (cioè $K(A^T A) = \|A^T A\| \cdot \|(A^T A)^{-1}\|$) è molto più grande del condizionamento del problema originale. $\to$ algoritmo instabile.

Vediamo ora qualche cenno a un altro algoritmo più stabile.

FATTORIZZAZIONE QR: ogni $A \in \mathbb{R}^{m \times n}$ si scrive come prodotto di $Q \in \mathbb{R}^{m \times m}$ ortogonale $Q^T Q = I$

$R \in \mathbb{R}^{m \times n}$ triangolare superiore $r_{ij} = 0$ $i > j$

$$\boxed{A} = \boxed{Q} \cdot \boxed{R}$$

$$R = \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

R_1 quadrata, tr. superiore

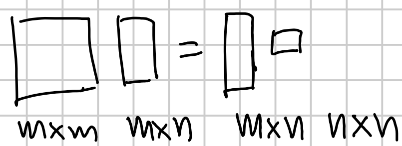
0 blocco di zeri

$$= m \overset{\quad n \qquad m-n}{\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}} \cdot \overset{n}{\begin{bmatrix} R_1 \\ 0 \end{bmatrix}} \begin{matrix} n \\ m-n \end{matrix}$$

Gli elementi nelle $m-n$ colonne più a destra della Q

(il blocco Q_2) vengono moltiplicati per $\emptyset$ quando facciamo il prodotto.

Quindi: $A = Q \cdot R = Q_1 R_1$.



È possibile mostrare che la soluzione del problema dei minimi quadrati è

$$x = R_1^{-1}\left(Q_1^T b\right)$$

$$R_1 x = Q^T b$$

e questo è un algoritmo stabile!

In MatLab: $x = A \setminus b$