Metodi Numerici per l'Informatica

Anthony

28 mar 2023

1 Introduzione

Abbiamo preso familiarità con problemi lineari del tipo $\mathbf{A}\mathbf{x} = \mathbf{b}$ in cui $\mathbf{A}$ non è necessariamente quadrata. Assumiamo da ora che $\mathbf{A}$ sia una matrice quadrata o alta, ovvero dispone abbastanza dati per risolvere il sistema lineare.

1.1 Spazio delle colonne

Possiamo scrivere $\mathbf{b}$ come combinazione lineare delle colonne di $\mathbf{A}$ con i coefficienti di $\mathbf{x}$:

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots \\ | & | & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = x_1 \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix} + x_2 \begin{pmatrix} | \\ \mathbf{a}_2 \\ | \end{pmatrix} + \dots$$

Tutte le possibili combinazioni lineari di $\mathbf{A}$ formano lo *spazio delle colonne*.

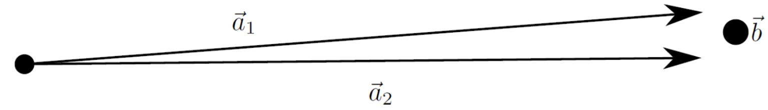
$$\operatorname{col} \mathbf{A} = \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2, \dots)$$

2 Instabilità numerica

Consideriamo il seguente problema:

$$\mathbf{A} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots \\ | & | & \end{pmatrix} = \begin{pmatrix} 0 & 0.0001 & \\ \vdots & \vdots & \dots \\ 1 & 1.0001 & \end{pmatrix}$$

Le due colonne identificano due vettori quasi identici, cioè due vettori quasi paralleli. I vettori sono quasi linearmente dipendenti. Se abbiamo un vettore $\mathbf{b}$ che è una combinazione lineare delle due colonne, esistono due modi che danno approssimativamente un modo per esprimere $\mathbf{b}$. Questo ci porta a una situazione ambigua.



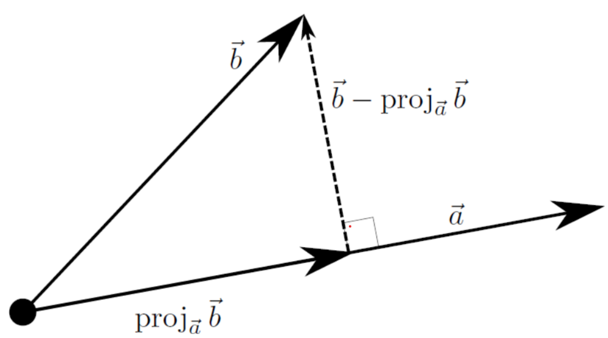
$$\mathbf{b} \approx x_1\mathbf{a}_1 + x_2\mathbf{a}_2$$

$$\mathbf{b} \approx x_2\mathbf{a}_1 + x_1\mathbf{a}_2$$

Questo problema si dice *malcondizionato* per via dell'*instabilità numerica* dei due vettori quasi-paralleli.
In generale, una soluzione per un problema del tipo $\mathbf{Ax} = \mathbf{b}$ potrebbe non esistere o essere unica e dipende proprio dallo spazio delle colonne di $\mathbf{A}$ per cui vogliamo risolvere il problema dell'instabilità numerica.

3 Ortogonalità

Consideriamo due vettori $\mathbf{a}$ e $\mathbf{b}$. Supponiamo di voler proiettare $\mathbf{b}$ verso $\mathbf{a}$.



Tale proiezione possiamo vederla come un multiplo di $\mathbf{a}$, per cui possiamo chiederci per quale scalare c il vettore $\mathbf{a}$ è il più vicino a $\mathbf{b}$:

$$\min_c \|c \cdot \mathbf{a} - \mathbf{b}\|_2^2$$

Sia tale scalare c^*, esso rappresenta la proiezione di $\mathbf{b}$ su $\mathbf{a}$:

$$\mathrm{proj}_{\mathbf{a}}\mathbf{b} \equiv c^*\mathbf{a}$$

Possiamo osservare che il *complemento* $\mathbf{b} - \mathrm{proj}_{\mathbf{a}}\mathbf{b}$ è *ortogonale* a $\mathrm{proj}_{\mathbf{a}}\mathbf{b}$. Tale vettore si chiama *complemento ortogonale*.
Il problema della proiezione equivale a trovare una soluzione ai minimi quadrati in cui $\mathbf{a} \cdot c \approx \mathbf{b}$. Possiamo applicare la stessa soluzione che abbiamo già visto in precedenza: calcoliamo il gradiente, lo poniamo uguale a zero e risolviamo per c.

$$c = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$$

Osserviamo che al numeratore non abbiamo altro che la norma ℓ_2:

$$c = \frac{\mathbf{a}^T\mathbf{b}}{\|\mathbf{a}\|_2^2}$$

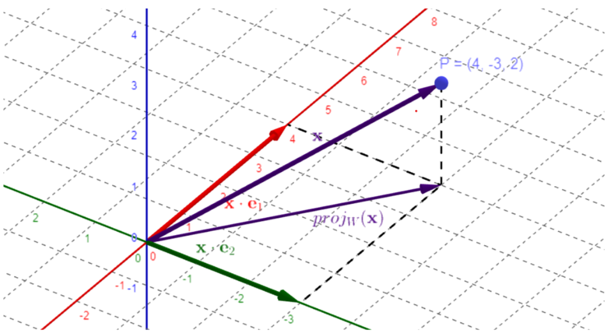
Per cui

$$\text{proj}_{\mathbf{a}}\mathbf{b} = \frac{\mathbf{a}^T\mathbf{b}}{\|\mathbf{a}\|_2^2}$$

In maniera analoga, possiamo proiettare $\mathbf{b}$ su un intero spazio vettoriale. Assumiamo di avere una base *ortonormale* $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$, ovvero una base in cui tutti i vettori sono perpendicolari e con norma pari a 1, allora possiamo proiettare un vettore sul suo span risolvendo il problema:

$$\min_{c_1,\dots,c_k} \|c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k - \mathbf{b}\|_2^2$$

Esempio Assumiamo di essere in $\mathbb{R}^3$ con $k = 2$:



La soluzione è analoga a prima, ma anziché risolvere per uno scalare risolviamo per k scalari. Il problema della proiezione si risolve quindi come segue:

$$\text{proj}_{\text{span}(\mathbf{a}_1,\dots,\mathbf{a}_k)}\mathbf{b} = (\mathbf{a}_1^T\mathbf{b})\mathbf{a}_1 + \cdots + (\mathbf{a}_k^T\mathbf{b})\mathbf{a}_k$$

Oss: non dividiamo per la norma perché abbiamo assunto una base ortonormale.

3.1 Matrici ortogonali

Abbiamo finora esplorato due concetti, ortogonalità e ortonormalità:

- Due vettori si dicono ortogonali se l'angolo che formano è di 90 gradi;
- Due vettori sono ortonormali se sono ortogonali e se hanno norma pari a 1.

Una matrice si dice *ortogonale* se le sue colonne sono ortonormali. Osserviamo ora il seguente problema banale:

$$\mathbf{I}\mathbf{x} = \mathbf{b}$$

In cui $\mathbf{I}$ è la matrice identica. La soluzione ovvia del problema è $\mathbf{x} = \mathbf{b}$. Abbiamo già osservato che l'errore dei minimi quadrati $\mathbf{A}\mathbf{x} \approx \mathbf{b}$ corrisponde al seguente problema lineare:

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

Sia $\mathbf{A}$ una matrice ortogonale, abbiamo che $\mathbf{A}\mathbf{A}^T = \mathbf{I}$, per cui ricadiamo nel caso banale.
Per chiarezza ora, indichiamo con $\mathbf{Q}$ una generica matrice ortogonale. Il prodotto $\mathbf{Q}^T\mathbf{Q}$ ha la forma:

$$\mathbf{Q}^T\mathbf{Q} = \begin{pmatrix} - & \vec{q}_1^T & - \\ - & \vec{q}_2^T & - \\ & \vdots & \\ - & \vec{q}_n^T & - \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \dots & \vec{q}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \dots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \dots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \dots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \dots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

Osserviamo che la diagonale contiene tutte le lunghezze al quadrato delle colonne di $\mathbf{Q}$. Tutte le colonne di $\mathbf{Q}$ sono ortogonali e hanno lunghezza 1 per cui $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Se le colonne di $\mathbf{Q}$ non avessero lunghezza 1, sulla diagonale avremmo solo i quadrati delle lunghezze delle colonne. Le colonne di $\mathbf{Q}$ formano una base ortonormale per lo spazio delle colonne.
Esempi:

1. La base standard è una base ortonormale.
2. La matrice identità $\mathbf{I}$ è ortogonale.
3. Qualsiasi matrice di permutazione, ovvero una matrice che ha un unico 1 su ogni colonna e su ogni riga e 0 altrove, è ortogonale.

3.1.1 Isometria

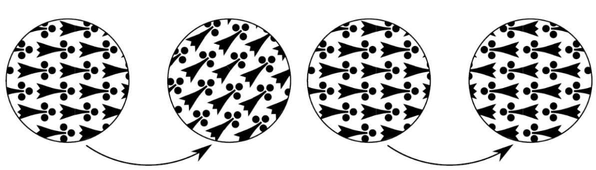
Le matrici ortogonali, in quanto matrici, modellano delle mappe lineari. Le matrici ortogonali hanno la proprietà che, quando applicate a un vettore, la lunghezza del vettore non cambia:

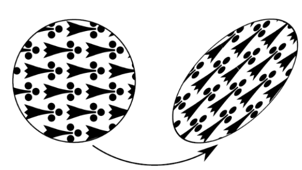
$$\|\mathbf{Q}\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{Q}^T\mathbf{Q}\mathbf{x} = \mathbf{x}^T\mathbf{I}\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|_2^2$$

Le matrici ortogonali, inoltre, preservano l'angolo tra due vettori (prodotto interno):

$$\langle \mathbf{Q}\mathbf{x}, \mathbf{Q}\mathbf{y} \rangle = \mathbf{x}^T\mathbf{Q}^T\mathbf{Q}\mathbf{y} = \mathbf{x}^T\mathbf{I}\mathbf{y} = \mathbf{x}^T\mathbf{y}$$

Per queste proprietà, la mappa $\mathbf{x} \mapsto \mathbf{Q}\mathbf{x}$ è un *isometria* in $\mathbb{R}^n$, ovvero rappresenta una rotazione dello spazio.


(a) Isometric


(b) Not isometric

3.2 Matrici non ortogonali

Quando risolviamo un problema del tipo $\mathbf{Ax} = \mathbf{b}$, tipicamente $\mathbf{A}$ non è ortogonale, le matrici ortogonali, tuttavia, possono aiutarci a semplificare il problema. Idea: possiamo riscrivere $\mathbf{A}$ come prodotto di due matrici tali che $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ è più facile da risolvere. In particolare, supponiamo che possiamo scrivere la matrice $\mathbf{A}$ come segue:

$$\mathbf{A} = \mathbf{QR}$$

In cui $\mathbf{Q}$ è ortogonale e $\mathbf{R}$ è invertibile. Per cui l'operazione

$$\mathbf{AR}^{-1} = \mathbf{Q}$$

data la matrice $\mathbf{A}$, la trasformiamo attraverso la matrice $\mathbf{R}^{-1}$ e ciò fa sì che le colonne di $\mathbf{A}$ diventino ortogonali. Questo significa che possiamo ortogonalizzare qualunque matrice.

4 Decomposizione QR

Qualsiasi matrice $\mathbf{A}$ ammette la fattorizzazione:

$$\mathbf{A} = \mathbf{QR}$$

Ciò significa che possiamo ortogonalizzare le colonne di $\mathbf{A}$. Inoltre, lo spazio delle colonne di $\mathbf{A}$ è identico allo spazio delle colonne di $\mathbf{AR}^{-1}$:

$$\text{col } \mathbf{AR}^{-1} = \text{col } \mathbf{A}$$

Ciò ci suggerisce che ortogonalizzare le colonne di $\mathbf{A}$ non cambia lo spazio che esse spannano. Dato che $\mathbf{Q} = \mathbf{AR}^{-1}$ possiamo affermare che:

$$\text{col } \mathbf{Q} = \text{col } \mathbf{A}$$

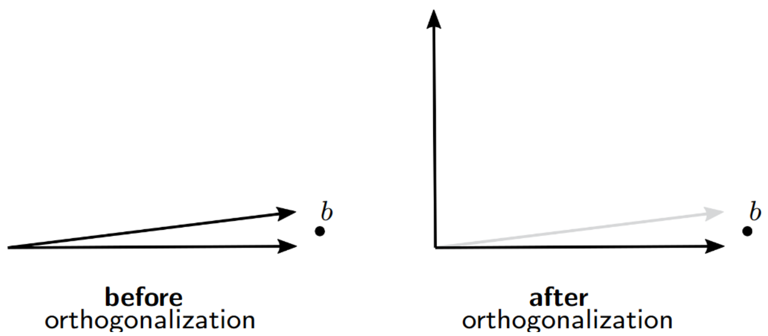
Per cui le colonne di $\mathbf{Q}$ formano una base ortonormale per le col $\mathbf{A}$.

4.1 Colonne quasi-parallele

Riprendendo l'esempio delle colonne *quasi parallele*:

$$\mathbf{A} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots \\ | & | & \end{pmatrix} = \begin{pmatrix} 0 & 0.0001 & \\ \vdots & \vdots & \dots \\ 1 & 1.0001 & \end{pmatrix}$$

Con l'ortogonalizzazione possiamo condizionare meglio il problema.



Usando la fattorizzazione QR, l'equazione

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

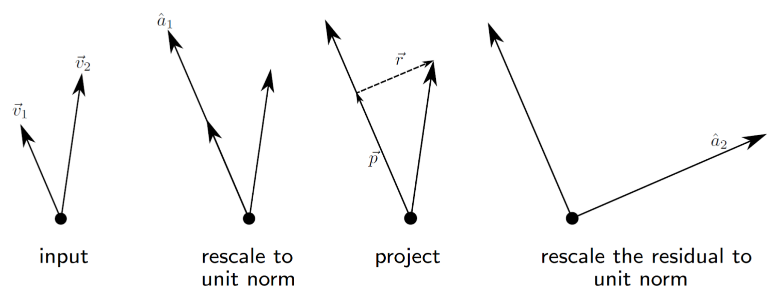
Diventa:

$$\begin{aligned}(\mathbf{QR})^T\mathbf{QR}\mathbf{x} &= (\mathbf{QR})^T\mathbf{b}\\ = \mathbf{R}^T\mathbf{Q}^T\mathbf{QR}\mathbf{x} &= \mathbf{R}^T\mathbf{Q}^T\mathbf{b}\\ = \mathbf{R}^T\mathbf{I}\mathbf{R}\mathbf{x} &= \mathbf{R}^T\mathbf{Q}^T\mathbf{b}\\ = \mathbf{R}\mathbf{x} &= \mathbf{Q}^T\mathbf{b} \quad \text{per invertibilità di } \mathbf{R}\end{aligned}$$

Un'altra proprietà della matrice $\mathbf{R}$ è che è *triangolare superiore*. Ciò rende il problema più facile da risolvere con la *back-substitution*.

4.2 Algoritmo Gram-Schmidt

L'algoritmo, dati due vettori $\vec{v_1}, \vec{v_2}$, attraverso le proiezioni ortogonali, ortogonalizza i due vettori.



Possiamo osservare che:

1. I vettori ortogonalizzati spannano lo stesso spazio di $\vec{v_1}$ e $\vec{v_2}$
2. L'algoritmo funziona in modo incrementale: possiamo ortogonalizzare un qualsiasi numero di vettori uno alla volta
3. Se i vettori in input sono le colonne di $\mathbf{A}$, i vettori in output ortogonalizzati sono le colonne di $\mathbf{Q}$. Per ottenere $\mathbf{R}$, possiamo computare semplicemente $\mathbf{R} = \mathbf{Q}^T\mathbf{A}$

Osserviamo ora l'implementazione in pseudo-codice dell'algoritmo di Gram-Schmidt. Ricordiamo che il residuo non è altro che il complemento ortogonale.

Algorithm 1 Gram-Schmidt

```
function GRAM-SCHMIDT(v⃗1, v⃗2, ..., v⃗k)
▷ Assume che v⃗1, v⃗2, ..., v⃗k sono linearmente indipendenti.
▷ Trova una base ortonormale â1, ..., âk per span {v⃗1, v⃗2, ..., v⃗k}
    â1 ← v⃗1/‖v⃗1‖2                                  ▷ Scaliamo il primo vettore
    for i ← 2, 3, ..., k do
        p⃗ ← 0⃗
        for j ← 1, 2, ..., i − 1 do   ▷ Proiezione di v⃗i sullo span {â1, ..., â(i−1)}
            p⃗ ← p⃗ + (v⃗i · âj)âj             ▷ Proiezione sulla base ortonormale
        end for
        r⃗ ← v⃗i − p⃗                        ▷ Il residuo è ortogonale alla base corrente
        âi ← r⃗/‖r⃗‖2       ▷ Normalizziamo il residuo e aggiungiamolo alla base
    end for
    return {â1, ..., âk}
end function
```

4.3 Varianti di Gram-Schmidt

Nel problema $\mathbf{Ax} = \mathbf{b}$ sappiamo che $\mathbf{A}$ potrebbe essere rettangolare, in particolare alta. Supponiamo essa abbia dimensioni $m \times n$. Che forma devono avere $\mathbf{Q}$ ed $\mathbf{R}$?

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times n}\underbrace{\mathbf{R}}_{n\times n}$$

oppure

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times m}\underbrace{\mathbf{R}}_{m\times n}$$

Questo ci suggerisce che esistono due fattorizzazioni valide per $\mathbf{A}$. In una $\mathbf{R}$ è quadrata, mentre nell'altra $\mathbf{Q}$ è quadrata. L'algoritmo Gram-Schimdt standard implementa la prima equazione, ovvero quella in cui $\mathbf{R}$ è quadrata. Ci sono altre varianti che implementano la seconda equazione, ma dobbiamo osservare che se $\mathbf{Q}$ è quadrata, e $m > n$, ovvero $\mathbf{A}$ è alta, allora significa che $\mathbf{Q}$ ha più colonne di quante ne avevamo in $\mathbf{A}$. Tali colonne, le ultime $n - k$, possono essere ignorate. Esistono diverse varianti dell'algoritmo Gram-Schmidt ed essi barattano l'efficienza per lo spazio e viceversa.