ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Matematica

Corso di Laurea in Matematica

Strategie di Pivoting per la fattorizzazione LU

Relatore:
Prof. Valeria Simoncini

Presentata da:
Virginia Nanni

Sessione ottobre 2025

Anno Accademico 2024/2025

Introduzione

Fin dalle prime lezioni di algebra lineare e di matematica numerica viene introdotto il metodo di eliminazione di Gauss, uno degli strumenti più noti per risolvere sistemi lineari e per calcolare il rango di una matrice.

Nel seguente elaborato ci soffermiamo sul suo impiego per la fattorizzazione LU di una matrice: eliminando passo per passo gli elementi sotto la diagonale principale, è possibile ricavare una matrice triangolare inferiore L e una superiore U, il cui prodotto restituisce la matrice originaria.

Questa fattorizzazione, quando esiste, rappresenta uno strumento fondamentale per la risoluzione di sistemi lineari grazie alla sua efficacia e semplicità. Proprio per questo è fondamentale valutarne la stabilità e introdurre strategie che consentano di assicurarla. In particolare, valuteremo le strategie di pivoting, che prevedono la permutazione di righe e colonne della matrice in modo da mantenere sotto controllo la crescita dei valori numerici all'interno della matrice. Definiremo e analizzeremo un'indicatore della stabilità del metodo: il fattore di crescita, che ci mostra, nella fattorizzazione $A = LU$, quanto i termini della matrice U tendano a "crescere" rispetto a quelli della matrice originaria A.

Da circa settant'anni molti matematici hanno iniziato ad occuparsi dell'analisi del fattore di crescita in base alla strategia di pivoting scelta ottenendo diversi risultati, ma ancora oggi persistono interrogativi. Tra i protagonisti

delle ricerche segnalo i nomi di James Wilkinson, L.V. Foster e Alan Edelman, che sono stati tra i primi a occuparsi del fattore di crescita, pubblicando diversi articoli a riguardo e fornendo stime che rimangono tuttora un punto di riferimento.

La seguente tesi si propone di ripercorre, per quanto possibile, la storia dei risultati ottenuti e delle ipotesi fatte (talvolta smentite), analizzando in dettaglio tre strategie di pivoting e le stime dall'alto del fattore di crescita quando queste vengono adottate.

Il Capitolo 1 introduce nozioni di base, richiamando il metodo di eliminazione di Gauss e la fattorizzazione LU, e definendo il fattore di crescita e le strategie di pivoting. Nel Capitolo 2 vengono analizzate tre strategie di pivoting tra le più utilizzate:

- Pivoting parziale;
- Rook pivoting;
- Pivoting completo.

Il capitolo si conclude con un cenno storico riguardo i vari risultati ottenuti e le ipotesi avanzate nel corso degli anni.

Nel Capitolo 3 viene presentata una famiglia di matrici (le randsvd) che mostra fattori di crescita insolitamente elevati per qualsiasi strategia di pivoting, costituendo così un caso limite per la stabilità del metodo.

Il Capitolo 4, infine, mostra risultati sperimentali a conferma di quanto visto nei capitoli precedenti.

Dalle analisi riportate, sia teoriche che sperimentali, emerge che i casi di matrici che presentano un fattore di crescita elevato sono estremamente rari. Tutti gli esperimenti condotti su matrici casuali riportano infatti fattori di crescita contenuti e questo permette di spiegare perché, storicamente, sia stato

difficile trovare esempi di matrici per cui il metodo fosse estremamente instabile anche con l'utilizzo di strategie di pivoting. Possiamo dunque concludere che i casi di instabilità per il metodo di eliminazione di Gauss esistono ma sono estremamente rari, per cui rimane una strategia affidabile per la risoluzione di sistemi lineari.

Indice

Elenco delle figure

Elenco delle tabelle

Capitolo 1

Nozioni preliminari

1.1 Notazione

Nel presente lavoro di tesi adotteremo la seguente notazione:

- Sia $x \in \mathbb{R}^n$. Indichiamo con x_i la i-esima componente del vettore x.
- Sia $A \in \mathbb{R}^{n \times m}$. Indichiamo con $A_{i,j}$ l'elemento nella posizione (i, j), corrispondente all'i-esima riga e alla j-esima colonna della matrice A.
- Sia $A \in \mathbb{R}^{n \times n}$. Indichiamo con δA una perturbazione della matrice A.
- Sia $x \in \mathbb{R}^n$. Indichiamo con $D = \text{diag}(x) \in \mathbb{R}^{n \times n}$ la matrice diagonale con $D_{i,i} = x_i$, per ogni $i = 1, \ldots, n$, e $D_{i,j} = 0$, per ogni $i \neq j$.
- Indichiamo con u l'epsilon machina.
- Indichiamo con I_n la matrice identità di ordine n, ovvero $I_n = \text{diag}(1, \ldots, 1)$.

Inoltre, ricorreremo frequentemente all'utilizzo di norme matriciali, in particolare:

- Norma di Frobenius: Sia $A \in \mathbb{R}^{n\times n}$, $||A||_F := \left(\sum_{i,j=1}^{n} |A_{i,j}|^2\right)^{\frac{1}{2}}$.
  Ricordiamo che $||A||_F^2 = \mathrm{tr}(\mathrm{A^T A})$.
- Sia $A \in \mathbb{R}^{n\times n}$, $||A||_{\max} := \max_{i,j} |A_{i,j}|$.

1.2 Stabilità della fattorizzazione LU

1.2.1 Definizioni e teoremi preliminari

Proposizione 1.2.1 ([16, Prop. 2.5.8])**.** *Sia $A \in \mathbb{R}^{n\times n}$ e siano u, $v \in \mathbb{R}^n$. Se A è non singolare e $1 + v^T A^{-1} u \neq 0$ vale la seguente formula,*

$$(A + uv^T)^{-1} = A^{-1} - A^{-1}u(1 + v^T A^{-1} u)^{-1} v^T A^{-1}. \tag{1.1}$$

Tale formula prende il nome di Formula di Sherman-Morrison.

Definizione 1.2.2. Sia $A \in \mathbb{R}^{n\times n}$. Si dice che la matrice A ammette una fattorizzazione LU se esistono:

- $L \in \mathbb{R}^{n\times n}$ matrice triangolare inferiore con tutti valori unitari sulla diagonale ($L_{i,i} = 1$ per ogni $i = 1, \ldots, n$);
- $U \in \mathbb{R}^{n\times n}$ matrice triangolare superiore

tali che $A = LU$.

Definizione 1.2.3. Sia $\Pi \in \mathbb{R}^{n\times n}$. Π si dice *matrice di permutazione* se è ottenuta da una matrice identità con una permutazione di righe o colonne.

Teorema 1.2.4 ([16, Thm. 2.3.6])**.** *Sia $A \in \mathbb{R}^{n\times n}$. Allora esiste una matrice di permutazione Π per cui è possibile ottenere la fattorizzazione LU di ΠA.*

1.3 Metodo di eliminazione di Gauss

Sia $A \in \mathbb{R}^{n\times n}$. Il metodo di Eliminazione di Gauss è un algoritmo che permette di ottenere una fattorizzazione LU di A.

Lavorando iterativamente su ciascuna colonna della matrice A, il metodo annulla i termini sottostanti la diagonale principale, trasformando così la matrice iniziale in una matrice triangolare superiore (corrispondente a U). Ripercorrendo le operazioni fatte in ciascuna iterazione è poi possibile ricostruire la matrice L, così da ottenere la fattorizzazione cercata.

$$A = \begin{pmatrix} \times & \times & \times & \times & \times \\ \otimes & \times & \times & \times & \times \\ \otimes & \otimes & \times & \times & \times \\ \otimes & \otimes & \otimes & \times & \times \\ \otimes & \otimes & \otimes & \otimes & \times \end{pmatrix} \Rightarrow \begin{pmatrix} \times & \times & \times & \times & \times \\ 0 & * & * & * & * \\ 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & \square & \square \\ 0 & 0 & 0 & 0 & \square \end{pmatrix}$$

In condizioni opportune, il metodo prevede $(n-1)$ iterazioni, in ognuna delle quali viene utilizzato il termine diagonale per annullare gli elementi sottostanti nella matrice, lasciando le righe superiori invariate. Alla fine della k-esima iterazione, la matrice ottenuta avrà le prime k colonne nella forma finale desiderata. Il procedimento si conclude quando tutti i termini sotto la diagonale sono stati eliminati.

Prima colonna: Sia $A^{(1)} = A$, supponendo $A_{11} \neq 0$ definiamo i seguenti moltiplicatori

$$m_{i1} := \frac{A^{(1)}_{i,1}}{A^{(1)}_{1,1}} \qquad i = 1, \ldots, n, \quad \text{e} \quad m_1 := \begin{pmatrix} m_{2,1} \\ m_{3,1} \\ \vdots \\ m_{n,1} \end{pmatrix}.$$

Il moltiplicatore m_1 permette di definire la trasformazione della matrice A nella prima iterazione: la prima riga risulterà inalterata, la prima colonna assumerà la forma cercata, mentre gli altri elementi si ricavano come segue:

$$A_{i,j}^{(2)} = A_{i,j}^{(1)} - m_{i,1}A_{1,j}^{(1)}, \qquad i,j = 2,...,n.$$

Al termine della prima iterazione la matrice $A^{(2)}$ assumerà la forma

$$A^{(2)} = \begin{pmatrix} \times & \times & \times & \times & \times \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{pmatrix}.$$

Proseguendo con le iterazioni avremo, per la k**-esima colonna**,

$$m_{i,k} = \frac{A_{i,k}^{(k)}}{A_{k,k}^{(k)}}, \qquad i = k+1,\dots,n\ ,$$
$$A_{i,j}^{(k+1)} = A_{i,j}^{(k)} - m_{i,k}A_{k,j}^{(k)}, \qquad i,j = k+1,\dots,n\ .$$

Supponendo che l'eliminazione di Gauss sia applicabile fino in fondo, al termine della procedura otteniamo la matrice triangolare superiore $A^{(n)}$. Come anticipato, il metodo può essere utilizzato per ottenere la fattorizzazione LU della matrice A, dove U coincide con la matrice $A^{(n)}$. Ricostruiamo ora la matrice triangolare inferiore L a partire dai moltiplicatori impiegati in ciascuna iterazione, ridefinendoli come vettori colonna in $\mathbb{R}^n$:

$$m_k := (0,\dots,0,m_{k+1,k},\dots,m_{n,k})^T\ .$$

Definiamo quindi la matrice

$$M_1 := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -m_{2,1} & 1 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ -m_{n,1} & 0 & \cdots & 1 \end{pmatrix} = I - \underline{m}_1 e_1^T,$$

e in generale

$$M_k := \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & & \vdots & & \vdots \\ \vdots & & \ddots & & & \\ & & & 1 & & \\ & & & -m_{k+1,k} & & \\ & & & \vdots & \ddots & \\ 0 & 0 & \cdots & -m_{n,k} & \cdots & 1 \end{pmatrix} = I - \underline{m}_k e_k^T.$$

Moltiplicando la matrice M_1 a sinistra della matrice A otteniamo precisamente $A^{(2)}$, ovvero la trasformazione di A a seguito della prima iterazione.

$$M_1 A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ 0 & A_{2,2}^{(2)} & \cdots & A_{2,n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & A_{n,2}^{(2)} & \cdots & A_{n,n}^{(2)} \end{pmatrix} \equiv A^{(2)}.$$

Analogamente, per ogni passo vale

$$M_k A^{(k)} = A^{(k+1)},$$

da cui segue che

$$M_{n-1} \cdot \ldots \cdot M_1 A = A^{(n)} = U. \tag{1.2}$$

$M_{n-1} \cdot \ldots \cdot M_1$ è una matrice triangolare inferiore, poichè prodotto di triangolari inferiori, ed è invertibile, in quanto ciascuna delle matrici M_k ha inversa $M_k^{-1} =$

$I + \underline{m}_k e_k^T$, infatti:

$$(I + \underline{m}_k e_k^T)(I - \underline{m}_k e_k^T) = I + \underline{m} e_k^T - \underline{m} e_k^T - \underline{m}_k e_k^T \underline{m}_k e_k^T = I - \underline{m}_k (e_k^T \underline{m}_k) e_k^T = I.$$

Moltiplicando entrambi i membri dell'equazione (1.2) per l'inversa di $M_1 \cdot \ldots \cdot M_{n-1}$, otteniamo

$$A = (M_{n-1} \cdot \ldots \cdot M_1)^{-1} U = M_1^{-1} \cdot \ldots \cdot M_{n-1}^{-1} U.$$

Definiamo $L := M_1^{-1} \cdot \ldots \cdot M_{n-1}^{-1} = \prod_{k=1}^{n-1} (I + \underline{m}_k e_k^T) = I + \sum_{k=1}^{n-1} \underline{m}_k e_k^T$, ovvero

$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ m_{2,1} & 1 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & 1 \end{pmatrix}.$$

Si ottiene dunque una matrice triangolare inferiore con valori unitari sulla diagonale, che permette di riscrivere la matrice A nella forma $A = LU$.

1.3.1 Stabilità del metodo e strategie di Pivoting

Per valutare la stabilità della fattorizzazione tramite il metodo di eliminazione di Gauss, analizziamo l'errore all'indietro relativo $\frac{||\delta A||_F}{||A||_F}$.

Teorema 1.3.1 ([16, Thm. 2.3.12])**.** *Sia $A \in \mathbb{R}^{n \times n}$ e siano $\tilde{L} = L + \delta L$ e $\tilde{U} = U + \delta U$ le matrici realmente ottenute tramite il metodo di eliminazione di Gauss.*
Allora $\tilde{L}\tilde{U} = A + \delta A$ con

$$||\delta A||_F \leq 2nu(||A||_F + ||\tilde{L}||_F + ||\tilde{U}||_F) + O((nu)^2),$$

dove u è l'epsilon machina.

Dunque la stabilità dipenderà dalle norme $||A||_F$, $||\tilde{L}||_F$ e $||\tilde{U}||_F$: avere norme contenute permette di ottenere un errore nell'ordine dell'epsilon machina. Se la matrice A è ben condizionata, la stabilità del problema dipenderà dalle norme di $\tilde{L}$ (legata ai moltiplicatori del metodo) e di $\tilde{U}$.

Per quanto riguarda la valutazione di $||\tilde{L}||_F$, per rendere il metodo di eliminazione di Gauss più stabile, si può ricorrere a strategie di pivoting. In particolare, tali strategie prevedono, ad ogni iterazione, la scelta di un elemento della matrice tra quelli massimi in modulo per spostarlo, con opportune permutazioni di righe e/o colonne, nella posizione diagonale a_{kk}, con k indice corrispondente all'iterazione: chiamiamo questo elemento pivot.

In questo modo si cerca di evitare che l'elemento diagonale sia troppo piccolo in modulo, poichè ne conseguirebbe una crescita del moltiplicatore, definito da un rapporto con al denominatore il termine a_{kk}.

In base a quanto approfonditamente viene fatta la ricerca dell'elemento più adatto, si distinguono strategie diverse di pivoting. Tra queste, analizzeremo nelle sezioni successive le strategie di partial, rook e complete pivoting. In ogni caso, vedremo che anche la strategia più semplice consente di controllare ogni componente dei moltiplicatori, garantendo $|m_{ij}| \leq 1$ per ogni $i, j = 1, \ldots, n$, da cui segue che $||\tilde{L}||_F \leq n$, infatti:

$$||\tilde{L}||_F = \sqrt{\sum_{i,j=1}^{n} \tilde{L}_{i,j}^2} \leq \sqrt{\sum_{i,j=1}^{n} 1} = n.$$

Resta dunque da valutare la norma $||\tilde{U}||_F$ e, per farlo, introduciamo il *fattore di crescita*

$$\rho := \frac{||\tilde{U}||_{\max}}{||A||_{\max}}.$$

Il fattore di crescita sarà il tema centrale del presente elaborato, in quanto l'analisi dei casi in cui degenera consente di individuare le matrici per cui il

metodo di eliminazione di Gauss risulta inefficace. Infatti, valutare la norma $||\tilde{U}||_F$ equivale a valutare il fattore di crescita:

$$\begin{aligned}\frac{||\tilde{U}||_F}{||A||_F} &\leq \frac{||\tilde{U}||_F}{||A||_{\max}} = \frac{\sqrt{\sum_{i,j} |\tilde{U}_{i,j}|^2}}{||A||_{\max}} = \frac{||\tilde{U}||_{\max}\sqrt{\sum_{ij} \frac{|\tilde{U}_{i,j}|^2}{||\tilde{U}||^2_{\max}}}}{||A||_{\max}} \\ &\leq \frac{||\tilde{U}||_{\max}}{||A||_{\max}}\sqrt{n^2} = n\frac{||\tilde{U}||_{\max}}{||A||_{\max}}.\end{aligned}$$

Possiamo dunque concludere che, nel caso di un fattore di crescita contenuto, la stabilità del metodo dipenderà solo dal buon condizionamento della matrice A. Nelle sezioni successive verranno presentate nel dettaglio le tre strategie di pivoting sopra citate.

Capitolo 2

Strategie di pivoting

2.1 Pivoting parziale

Il pivoting parziale rappresenta la più semplice tra le strategie di pivoting, in quanto prevede solo permutazioni di righe e, ad ogni iterazione, la ricerca del pivot si limita alla colonna corrispondente all'iterazione in corso. Alla k-esima iterazione, il pivot scelto corrisponderà infatti al massimo tra gli elementi della parte ancora non ridotta della k-esima colonna (ovvero gli elementi $A_{i,k}$ per $i \geq k$).

Proposizione 2.1.1. *Sia g_{pp} il fattore di crescita nel caso di fattorizzazione LU con pivoting parziale. Vale la seguente stima dall'alto:*

$$g_{pp} \leq 2^{n-1}.$$

Dimostrazione. Ripercorrendo i passi dell'eliminazione di Gauss

$$A_{i,j}^{(2)} = A_{i,j}^{(1)} - m_{i,1} A_{1,j}^{(1)},$$

da cui, per la disuguaglianza triangolare e poiché $|m_{i1}| \leq 1$, segue che

$$|A_{i,j}^{(2)}| \leq |A_{i,j}^{(1)}| + 1|A_{1,j}^{(1)}| \leq 2||A||_{\max}.$$

Analogamente,

$$A^{(3)}_{i,j} = A^{(2)}_{i,j} - m_{i,2}A^{(2)}_{2,j},$$

da cui

$$|A^{(3)}_{i,j}| \leq |A^{(2)}_{i,j}| + 1|A^{(2)}_{2,j}| \leq 2||A||_{\max} + 2||A||_{\max} = 4||A||_{\max}.$$

Iterando otteniamo

$$||U||_{\max} = ||A^{(n)}||_{\max} \leq 2^{n-1}||A||_{\max},$$

che permette di concludere

$$g_{pp} = \frac{||U||_{\max}}{||A||_{\max}} \leq 2^{n-1}.$$

□

Il precedente risultato mostra che, utilizzando la strategia di pivoting parziale, è possibile avere un aumento del fattore di crescita esponenziale nella dimensione della matrice e una conseguente instabilità nel metodo di eliminazione di Gauss.

La stima è molto pessimistica, e nella maggior parte dei casi non viene raggiunta, ma ci sono alcune matrici per cui il fattore di crescita è esattamente il massimo teorico.

Esempio 2.1.2. Wilkinson [19] ha presentato un esempio di una matrice per cui il limite superiore viene effettivamente raggiunto.

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ -1 & 1 & \cdots & 0 & 1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & -1 & 1 & 1 \\ -1 & -1 & \cdots & -1 & 1 \end{pmatrix}.$$

L'eliminazione di Gauss per questa matrice non richiede pivoting, ma otteniamo $U_{nn} = 2^{n-1}$, da cui segue che $||U||_{\max} = 2^{n-1}$.

2.2 Rook Pivoting

Negli anni Novanta del Novecento, i matematici Neal e Pool [15] hanno proposto una nuova strategia di pivoting, che rappresentasse una via di mezzo tra il partial e il complete pivoting, già ampiamente studiati: il Rook Pivoting. In questo modo hanno reso possibile ottenere una maggiore accuratezza mantenendo costi simili al partial pivoting.

L'algoritmo per la determinazione del pivot si basa su un procedimento di ricerca incrociata tra alcune righe e colonne della matrice.

Alla prima iterazione si seleziona l'elemento di modulo massimo della prima colonna, analogamente al partial pivoting. Una volta determinato, ci si sposta sulla riga corrispondente, continuando la ricerca e individuando all'interno della riga l'elemento massimo in modulo. Se quest'ultimo coincide con quello precedente, la procedura si arresta e il pivot è determinato; altrimenti si prosegue, spostandosi sulla colonna corrispondente al nuovo massimo e si ripetono le stesse operazioni. Il ciclo termina quando, in due iterazioni successive, la scelta del pivot non cambia.

Per la k-esima iterazione si procede analogamente alla prima, limitandosi ad analizzare la sottomatrice di dimensioni $(n-k+1) \times (n-k+1)$ formata dalle righe e colonne nella porzione della matrice non ancora interessata dalle operazioni precedenti.

Alla fine degli anni Novanta, Foster ha mostrato che, utilizzando la strategia di rook pivoting (analogamente al complete pivoting), non si sarebbe verificata una crescita esponenziale del fattore di crescita, dimostrando la seguente stima dall'alto

$$\rho_{rp} \leq 1.5 n^{\frac{3}{4}\log n}. \tag{2.1}$$

2.2.1 Stima dall'alto del fattore di crescita con Rook Pivoting

In questa sezione introduciamo un lemma e un teorema che permetteranno di arrivare al teorema conclusivo con il risultato (2.1).
Per maggiori dettagli riguardo ai risultati riportati rimandiamo al lavoro di L. V. Foster [7].

Lemma 2.2.1. *Siano $p_1, \ldots, p_n$ interi positivi tali che*

- $p_1 \geq p_2 \geq \ldots \geq p_n \geq 0$;
- $\prod_{i=1}^{h} p_i \leq \left[h^{\frac{1}{2}}\left(1 + \sum_{i=h+1}^{n} p_i\right)\right]$ *per* $h = 1, \ldots, n$.

Allora $\max p_1$ *è raggiunto solo se tutte le disuguaglianze nel secondo punto sono uguaglianze.*
Inoltre, in tal caso, definendo s_k come la soluzione positiva di

$$s_k(1 + s_k)^{k-1} = \frac{k^{\frac{k}{2}}}{(k-1)^{\frac{k-1}{2}}} \tag{2.2}$$

con $1 \leq k \leq n$, vale che $p_1 = s_1(1 + s_2) \cdot \ldots \cdot (1 + s_n)$.

Teorema 2.2.2. *Siano $A \in \mathbb{R}^{n \times n}$, s_k definito come sopra per ogni $1 \leq k \leq n$ e ρ_{rp} il fattore di crescita per il metodo di eliminazione di Gauss con Rook pivoting. Allora vale il seguente risultato*

$$\rho_{rp} \leq t_n = s_1(1 + s_2)(1 + s_3) \cdot \ldots \cdot (1 + s_n). \tag{2.3}$$

Dimostrazione. Richiamiamo innanzitutto la disuguaglianza di Hadamard.

Sia $A \in \mathbb{R}^{n \times n}$. Denotando con a_i l'i-esima colonna della matrice A e con $\|a_i\|_2$ la sua norma euclidea, vale la seguente disuguaglianza

$$|det(A)| \leq \prod_{i=1}^{n} ||a_i||_2 \tag{2.4}$$

Senza perdita di generalità facciamo innanzitutto alcune assunzioni:

- La matrice A è già permutata, dunque non sono necessarie ulteriori strategie di pivoting;

- A è normalizzata in modo che $|A_{i,j}| \leq 1$ per $i, j = 1, .., n$ e $||A||_{\max} = 1$.

Possiamo ora scrivere $A = LU$ con:

- L matrice triangolare inferiore con tutti 1 sulla diagonale;

- U matrice tirangolare superiore tale che $|U_{i,j}| \leq |U_{i,i}|$, per $i, j = 1, \ldots, n$.

Definiamo $p_i = |U_{i,i}|$ per $i = 1, 2, \ldots, n$, e notiamo che, per le assunzioni fatte,

$$\rho_{rp} = \max p_i.$$

Denotando con l_i l'i-esima colonna di L e con u_i l'i-esima riga di U, possiamo riscrivere A come somma di matrici di rango 1

$$A = \begin{bmatrix} l_1 & \cdots & l_n \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \sum_{i=1}^{n} l_i u_i. \tag{2.5}$$

Sia $1 \leq h \leq n$, consideriamo due sottoinsiemi di $N = \{1, \ldots, n\}$: $I = \{i_1, \ldots, i_h\}$ contenente h indici e J il suo complementare in N. Possiamo allora spezzare la sommatoria in (2.5):

$$\sum_{i \in I} l_i u_i = A - \sum_{j \in J} l_i u_i \equiv B$$

Definiamo ora $\hat{L}$, $\hat{U}$ e $\hat{B}$ sottomatrici con righe e colonne rispettivamente di L, U e B, relative agli indici in I. Segue che $\hat{B} = \hat{L}\hat{U}$, dove:

- $\hat{L}$ è una matrice tringolare inferiore con tutti 1 sulla diagonale;

- $\hat{U}$ è una matrice triangolare superiore con $|\hat{U}_{k,j}| \leq |\hat{U}_{k,k}| = |U_{r,r}| = p_r$, con $1 \leq j \leq n$ e r il k-esimo indice in I.

Poichè $B = A - \sum_{j \in J} l_i u_i$, segue che

$$|B_{i,j}| \leq |A_{i,j}| + \sum_{j \in J} |L_{j,i}||U_{i,j}| \leq 1 + \sum_{j \in J} p_j. \tag{2.6}$$

Denotando le colonne di $\hat{B}$ con $\hat{b}_k$ e applicando la disuguaglianza di Hadamard, otteniamo il seguente risultato:

$$|det(\hat{B})| \leq \prod_{k=1}^{h} \|\hat{b}_k\|_F \leq \left[h^{\frac{1}{2}}\left(1 + \sum_{j \in J} p_j\right)\right]^h. \tag{2.7}$$

Infatti,

$$\begin{aligned} ||\hat{b}_k||_F = \sqrt{\sum_{i=1}^{h} \hat{B}_{i,k}^2} &\overset{(2.6)}{\leq} \sqrt{\sum_{i=1}^{h} \left(1 + \sum_{j \in J} p_j\right)^2} \\ &\leq \sqrt{h\left(1 + \sum_{j \in J} p_j\right)^2} = h^{\frac{1}{2}}\left(1 + \sum_{j \in J} p_j\right). \end{aligned}$$

Da cui segue la disuguaglianza (2.7).

Dato che $det(\hat{B}) = det(\hat{U})$ concludiamo che per ogni sottoinsieme I di N e J suo complementare, vale che

$$\prod_{i \in I} p_i \leq \left[h^{\frac{1}{2}}\left(1 + \sum_{j \in J} p_j\right)\right]^h. \tag{2.8}$$

Riordinando gli h indici di $\{p_{i_1}, \ldots, p_{i_h}\}$ in modo che $p_{i_1} \geq \ldots \geq p_{i_h}$, possiamo riscrivere la disuguaglianza come

$$\prod_{k=1}^{h} p_{i_k} \leq \left[h^{\frac{1}{2}}\left(1 + \sum_{k=h+1}^{n} p_{i_k}\right)\right]^h. \tag{2.9}$$

Grazie al lemma è possibile concludere. □

Introduciamo ora un'osservazione che useremo nel teorema successivo.

Osservazione 2.2.3. Per ogni $x \in [0,1]$ e per ogni $n \in \mathbb{N}$, vale la relazione

$$(1 - \frac{nx^2}{2})e^{nx} \leq (1+x)^n \leq e^{nx} \tag{2.10}$$

Dimostrazione. Notiamo che il risultato potrebbe essere esteso oltre l'intervallo $[0,1]$, ma non sarà utile ai fini del teorema. Denotando $f_1 = (1 - \frac{nx^2}{2})e^{nx}$, $f_2 = (1+x)^n$ e $f_3 = e^{nx}$, dimostriamo che valgono le due disuguaglianze.

Stima dall'alto: $(1+x)^n \leq e^{nx}$.

Per $x \in [0,1]$ la disequazione $(1+x)^n \leq e^{nx}$ è equivalente a $n\log(1+x) \leq nx$, che è sempre verificata.

Stima dal basso: $(1 - \frac{nx^2}{2})e^{nx} \leq (1+x)^n$.

Se $1 - n\frac{x^2}{2} \leq 0$ allora la stima dal basso è ovvia.

Altrimenti, se $1 - n\frac{x^2}{2} > 0$, passiamo al logaritmo:

$$\log(1 - n\frac{x^2}{2}) + \log(e^{nx}) \leq n\log(1+x),$$

ovvero

$$\log(1 - n\frac{x^2}{2}) + nx - n\log(1+x) \leq 0.$$

Pongo $f(x) = \log(1 - n\frac{x^2}{2}) + nx - n\log(1+x)$. Si ha $f(0) = 0$. Se dimostro che f è decrescente per $x \in [0,1]$, allora si ha $f(x) \leq 0$. Notiamo che

$$\begin{aligned} f'(x) &= \frac{-nx}{1 - n\frac{x^2}{2}} + n - \frac{n}{1+x} \\ &= \frac{-nx - nx^2 + n - n^2\frac{x^2}{2} + nx - n^2\frac{x^3}{2} - n + n^2\frac{x^2}{2}}{(1 - n\frac{x^2}{2})(1+x)} \\ &= nx\left(\frac{-x - n\frac{x^2}{2}}{(1+x)(1 - n\frac{x^2}{2})}\right), \end{aligned}$$

che è negativo per ogni x, quindi f è decrescente. □

Teorema 2.2.4. *Sia $A \in \mathbb{R}^{n\times n}$. Allora*

$$\rho_{rp} \leq 1.5n^{\frac{3}{4}\log n} \tag{2.11}$$

Dimostrazione. Limitiamo la dimostrazione al caso $n > 18$. I risultati per tutte le dimensioni minori sono approfonditi nell'Osservazione 2.2.5.

Sia s_n la soluzione positiva di $x(1+x)^{n-1} = \frac{n^{\frac{n}{2}}}{(n-1)^{\frac{n-1}{2}}}$, vogliamo dimostrare che

$$s_n < \frac{3\log n}{2n}.$$

Per farlo, introduciamo un polinomio che sia crescente per $x \geq 0$, che abbia s_n come zero e sia positivo in $\frac{3\log n}{2n}$, in modo tale da ottenere la disuguaglianza desiderata.

Definiamo il polinomio $p_n(x) := x(1+x)^{n-1} - \frac{n^{\frac{n}{2}}}{(n-1)^{\frac{n-1}{2}}}$, crescente per $x \geq 0$ e con s_n come zero.

Resta dunque da valutarlo in $\frac{3\log n}{2n}$ e, per farlo, maggioriamo i due addendi

$$x(1+x)^{n-1} \qquad \text{e} \qquad -\frac{n^{\frac{n}{2}}}{(n-1)^{\frac{n-1}{2}}},$$

in modo da ottenere un'espressione più semplice da valutare in $\frac{3\log n}{2n}$.
In particolare, vogliamo ottenere la seguente maggiorazione:

$$p_n(x) \geq x(1+x)^{-1}\left(1 - \frac{nx^2}{2}\right)e^{nx} - n^{\frac{1}{2}}e^{\frac{1}{2}} =: g_n(x).$$

Per verificarla, mostriamo che valgono le seguenti disuguaglianze:

$$x(1+x)^{n-1} \geq x(1+x)^{-1}\left(1 - \frac{nx^2}{2}\right) \tag{2.12}$$

$$\frac{n^{\frac{n}{2}}}{(n-1)^{\frac{n-1}{2}}} \leq n^{\frac{1}{2}}e^{\frac{1}{2}} \tag{2.13}$$

Per verificare la disuguaglianza (2.12), consideriamo $x \in [0,1]$ e riprendiamo la disuguaglianza (2.10):

$$(1 - \frac{nx^2}{2})e^{nx}(1+x)^{-1} \leq (1+x)^{n-1},$$

da cui si ottiene

$$x(1+x)^{n-1} \geq \frac{x}{x-1}(1-\frac{nx^2}{2})e^{nx}.$$

Verifichiamo ora la disuguaglianza (2.13): in quanto entrambi i membri sono positivi e la funzione log è monotona crescente in $(0,\infty)$, possiamo passare al logaritmo mantenendo valida la disuguaglianza.

$$\frac{n}{2}\log n - \frac{n-1}{2}\log(n-1) \leq \frac{1}{2}\log n + \frac{1}{2},$$

da cui segue che

$$\frac{n-1}{2}\log(\frac{n}{n-1}) - \frac{1}{2} \leq 0$$
$$\frac{n-1}{2}\log(1+\frac{1}{n-1}) - \frac{1}{2} \leq 0$$

Per n sufficientemente grande, approssimiamo il logaritmo utilizzando gli sviluppi in serie di Taylor:

$$\log(1+\frac{1}{n-1}) \approx \frac{1}{n-1} - \frac{1}{2(n-1)^2},$$

da cui

$$\frac{n-1}{2}\log(1+\frac{1}{n-1}) - \frac{1}{2} \approx \frac{n-1}{2}(\frac{1}{n-1} - \frac{1}{2(n-1)^2}) - \frac{1}{2} = -\frac{1}{4(n-1)}.$$

Dato che $-\frac{1}{4(n-1)} \leq 0$, la disuguaglianza (2.13) è verificata.

Possiamo quindi dire che $p_n(x) \geq g_n(x)$ e valutiamo la disuguaglianza in $x = \frac{3\log n}{2n}$

$$p_n\left(\frac{3\log n}{2n}\right) \geq g_n\left(\frac{3\log n}{2n}\right) = \left[\frac{1.5\log n}{1+1.5\frac{\log n}{n}}\left(1-\frac{9\log^2 n}{8n}\right) - e^{\frac{1}{2}}\right] n^{\frac{1}{2}};$$

resta dunque da dimostrare che il secondo membro sia positivo.

Le funzioni $\log n$ e $n^{\frac{1}{2}}$ sono monotone crescenti per ogni n, mentre le funzioni $\frac{\log n}{n}$ e $\frac{\log^2 n}{n}$ sono monotone decrescenti per $n > e^2$. Ne segue che $g_n(\frac{3\log n}{2n})$ è una

funzione crescente e positiva per n=18 (verificato con operazioni numeriche). Dunque, per $n > 18$ vale $p_n(\frac{3\log n}{2n}) > 0$.

Poichè $p_n(s_n) = 0$ e p_n monotona crescente, possiamo concludere che

$$s_n < \frac{3\log n}{2n}.$$

Riprendiamo ora la definizione di t_n e valutiamo il logarimo della funzione.

$$\begin{aligned}
\log(t_n) &= \log\left(s_1 \prod_{k=2}^{n}(1+s_k)\right) = \log\left(\prod_{k=2}^{n}(1+s_k)\right) \\
&= \sum_{k=2}^{18}\log(1+s_k) + \sum_{k=19}^{n}\log(1+s_k) \leq \sum_{k=2}^{18}\log(1+s_k) + \sum_{k=19}^{n} s_k \\
&\leq \sum_{k=2}^{18}\log(1+s_k) + \sum_{k=19}^{n}\frac{3\log k}{2k} \\
&\leq \sum_{k=2}^{18}\log(1+s_k) + \frac{3\log^2 n - 3\log^2 18}{4}.
\end{aligned}$$

Quindi, per $n \geq 18$, vale che

$$t_n \leq \left[\frac{\prod_{k=1}^{18}(1+s_k)}{18^{\frac{3\log 18}{4}}}\right] n^{\frac{3\log n}{n}} = 0.679 n^{\frac{3\log n}{n}}.$$

Per il Teorema 2.2.2 possiamo affermare che, per $n > 18$

$$\rho_{rp} \leq t_n \leq 0.679 n^{\frac{3\log n}{n}} \tag{2.14}$$

□

Osservazione 2.2.5. Mostriamo empiricamente che per $n \leq 18$ vale

$$t_n \leq 1.5 n^{\frac{3}{4}\log n}, \tag{2.15}$$

dove $t_n = s_1 \cdot (1 + s_2) \cdot ... \cdot (1 + s_n)$.

Nella tabella 2.1 riportiamo, per ciascun $n \leq 18$, la soluzione positiva di $x(x+1)^{n-1} = \dfrac{n^{\frac{n}{2}}}{(n-1)^{\frac{n-1}{2}}}$, calcoliamo t_n e verifichiamo la disuguaglianza.
Tutti i risultati sono stati approssimati alla quarta cifra decimale.

n	x	t_n	$1.5n^{\frac{3}{4}\log n}$	n	x	t_n	$1.5n^{\frac{3}{4}\log n}$
1	1	1	1.5	10	0.3473	$53,7127$	79.9889
2	1	2	2.1507	11	0.3236	$71,0942$	111.9319
3	0.8009	$3,6018$	3.7086	12	0.3033	$92,6571$	153.9358
4	0.6660	$6,0000$	6.3392	13	0.2856	$119,1199$	$208,4478$
5	0.5720	$9,4330$	10.4665	14	0.2702	$151,3061$	$278,3648$
6	0.5033	$14,1806$	16.6642	15	0.2565	$190,1161$	367.0936
7	0.4507	$20,5719$	25.6717	16	0.2443	$236,5615$	478.6172
8	0.4091	$28,9878$	38.4178	17	0.2334	$291,7750$	617.5675
9	0.3753	$39,8669$	56.0538	18	0.2235	$356,9867$	789.3041

Tabella 2.1: Rook pivoting - risultati per $n \leq 18$

Dai risultati ottenuti possiamo notare che la discrepanza tra i due valori aumenta sempre di più con il crescere di n; dunque, da un'immediata osservazione dei dati, possiamo intuire che la disuguaglianza sarà verificata anche per $n > 18$.

2.3 Pivoting completo

Il pivoting completo è, tra le strategie prese in considerazione, la più efficace ma anche la più costosa dal punto di vista computazionale e la più complessa da analizzare in relazione al fattore di crescita.

> "Although complete pivoting is not always the best strategy, we have not been able to construct an example for which it is a very bad strategy".
>
> - J.H. Wilkinson, [18]

Oltre a Wilkinson, anche A. Edelman ha ribadito quanto l'elevato costo computazionale portasse spesso alla scelta di una strategia di pivoting alternativa.

> "Complete pivoting is rarely used, because the improvement in numerical stability over partial pivoting does not justify the time spent searching for the largest element in the submatrix".
>
> - A. Edelman, [5]

Il metodo prevede permutazioni sia di righe che di colonne, selezionando il pivot come il $\max_{i,j} |A_{i,j}|$ con $i, j \geq k$, assicurando così maggiore stabilità rispetto alle altre strategie, ma anche più tempo di ricerca. A differenza del partial pivoting, il comportamento del fattore di crescita non è chiaro: la ricerca di un limite superiore rimane un problema aperto, tanto che, dopo la prima stima ottenuta intorno agli anni Sessanta del Novecento, per circa 60 anni non ci sono stati miglioramenti.

> "Despite its popularity, the worst-case behavior of the growth factor under complete pivoting is poorly understood".

- Bisain, Edelman, Urschel [1]

Il primo risultato per una stima dall'alto del fattore di crescita è stato dimostrato da J.H. Wilkinson [18]:

$$\rho_{cp} \leq 2\sqrt{n}n^{\frac{\log n}{4}}. \tag{2.16}$$

Nello stesso articolo, dopo aver dimostrato il risultato 2.16, Wilkinson ne ha anche fin da subito sottolineato la grossolanità.

> "In our experience [...] no matrix has been encountered in practice for which $\frac{p_1}{p_n}$ was as large as 8".

Ribadita successivamente anche da Foster

> "Furthermore, it is known for complete pivoting that these bounds cannot be attained for $n \geq 3$ and no one has been able to find any examples where the growth factor for complete pivoting is bigger than, for example, $2n$. For these reasons complete pivoting is considered to be numerically stable".

- L.V. Foster [7]

Circa 60 anni dopo, nel 2025, i matematici A. Bisain, A. Edelman e J. Urschel sono riusciti a migliorare la stima [1], ottenendo:

$$\rho_n \leq 2n^{0.25 \ln n + 0.5}. \tag{2.17}$$

2.3.1 Stima dall'alto di Wilkinson

Ad ogni passo del metodo di eliminazione di Gauss, a partire da una matrice $A \in \mathbb{R}^{n \times n}$, si eliminano i termini sottostanti la diagonale corrente e si ottiene una nuova matrice da elaborare in dimensione ridotta. In particolare, dopo la

prima iterazione, la nuova matrice da analizzare sarà di dimensione $(n-1) \times (n-1)$. Procedendo analogamente, indichiamo con $A^{(k)}$ la matrice rimasta da elaborare dopo $(n-k)$ passi nel metodo di eliminazione di Gauss. Denotiamo inoltre con p_k il pivot scelto al passo k, per $k = 1, \ldots, n$. Nel caso del complete pivoting, ad ogni passo selezionamo come pivot il termine massimo in modulo della matrice corrente, dunque avremo $p_k = ||A^{(k)}||_\infty$.

Riscriviamo quindi il fattore di crescita come:

$$\rho = \frac{||U||_{\max}}{||A||_{\max}} = \frac{\max\limits_k p_k}{p_n}.$$

Poichè siamo interessati al massimo fattore di crescita su tutte le matrici $n \times n$ è sufficiente considerare il valore massimo di $\frac{p_1}{p_n}$.

Inoltre, osserviamo che nella fattorizzazione LU, indipendentemente dalla strategia di pivoting utilizzata, i pivot scelti ad ogni iterazione coincidono con l'elemento diagonale corrispondente della matrice triangolare superiore U. Al termine della procedura otteniamo la forma

$$PAR = LU,$$

con $P, R \in \mathbb{R}^{n \times n}$ matrici di permutazione; passando ai determinanti avremo la seguente relazione

$$\det(P)\det(A)\det(R) = \det(L)\det(U);$$

poiché $\det(P) = \det(R) = \det(L) = 1$, possiamo riscriverla come

$$\det(A) = \det(U) = \prod_{i=1}^{n} p_i.$$

Lo stesso discorso può essere fatto per ciascuna matrice $A^{(k)}$ ottenuta dopo $(n-k)$ iterazioni, e vale

$$\det\left(A^{(k)}\right) = \prod_{i=1}^{k} p_i. \tag{2.18}$$

Dimostriamo ora il risultato (2.16) presentando un teorema che riassume i risultati ottenuti da Wilkinson [18].

Teorema 2.3.1. *Sia $A \in \mathbb{R}^{n\times n}$. Il fattore di crescita per il metodo di eliminazione di Gauss con complete pivoting soddisfa*

$$\rho_{cp} \leq 2n^{\frac{1}{2}+\frac{\log(n)}{4}}. \tag{2.19}$$

Dimostrazione. Ricordando la disuguaglianza di Hadamard

$$\det(A) \leq \prod_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}},$$

segue

$$\det(A^{(k)}) \leq \prod_{i=1}^{k} (k||A^{(k)}||_\infty^2)^{\frac{1}{2}} = \prod_{i=1}^{k} (k^{\frac{1}{2}} p_k) = k^{\frac{k}{2}} p_k^k.$$

e, per l'equazione 2.18, otteniamo

$$\prod_{i=1}^{k} p_i = \det(A^{(k)}) \leq k^{\frac{k}{2}} p_k^k \tag{2.20}$$

e, in particolare,

$$p_1 \cdot \ldots \cdot p_n = \det(A) \leq n^{\frac{n}{2}} p_n^n. \tag{2.21}$$

Passando al logaritmo nell'equazione 2.20 e denotando $q_k \stackrel{\text{def}}{=} \log p_k$, otteniamo

$$\log \prod_{i=1}^{k} p_i \leq \log\left(k^{\frac{k}{2}} p_k\right) = \frac{k}{2}\log k + k q_k,$$

ovvero

$$\sum_{i=1}^{k} q_i \leq \frac{k}{2}\log k + k q_k, \tag{2.22}$$

e

$$\sum_{i=1}^{n} q_i = \log(\det(A)) \leq \frac{n}{2}\log n + n q_n. \tag{2.23}$$

Dividendo entrambi i membri di (2.20) per p_k, otteniamo

$$\prod_{i=1}^{k-1} p_i \leq k^{\frac{k}{2}} p_k^{k-1}.$$

Riprendendo ora l'equazione (2.22) per $k-1$, otteniamo

$$\sum_{i=1}^{k-1} q_i = \log(\prod_{i=1}^{k-1} p_i) \leq \log(k^{\frac{k}{2}} p_k^{k-1}) = \frac{k}{2}\log k + (k-1)q_k. \tag{2.24}$$

Ricordando la somma parziale di una serie telescopica

$$\sum_{i=k}^{n-1}\left(\frac{1}{i-1} - \frac{1}{i}\right) = \frac{1}{k-1} - \frac{1}{n-1} \tag{2.25}$$

otteniamo che

$$\begin{aligned}\frac{1}{k-1} &= \sum_{i=k}^{n-1}\left(\frac{1}{i-1} - \frac{1}{i}\right) + \frac{1}{n-1} = \sum_{i=k}^{n-1}\left(\frac{1}{i(i-1)}\right) + \frac{1}{n-1} = \\ &= \frac{1}{k(k-1)} + \ldots + \frac{1}{(n-1)(n-2)} + \frac{1}{n-1}\end{aligned}$$

Dividendo ciascun membro della disuguaglianza (2.24) per $k(k-1)$ e ciascun membro dell'equazione (2.23) per $(n-1)$, otteniamo:

$$\sum_{i=1}^{k-1} \frac{q_i}{k(k-1)} \leq \frac{1}{2(k-1)}\log k + \frac{1}{k}q_k, \tag{2.26}$$

$$\sum_{i=1}^{n} \frac{q_i}{n-1} = \frac{\log\left(\det(A)\right)}{n-1}. \tag{2.27}$$

Sommiamo ora su k, per $k = 2, \ldots, n-1$, in (2.26):

$$\sum_{k=2}^{n-1}\sum_{i=1}^{k-1} \frac{q_i}{k(k-1)} \leq \sum_{k=2}^{n-1}\left(\frac{1}{2(k-1)}\log k + \frac{1}{k}q_k\right). \tag{2.28}$$

Lavoriamo sul primo membro scambiando le sommatorie:

$$\sum_{i=1}^{n-2}\sum_{k=i+1}^{n-1} \frac{q_i}{k(k-1)} = \sum_{i=1}^{n-2} q_i \sum_{k=i+1}^{n-1} \frac{1}{k(k-1)} = \sum_{i=1}^{n-2} q_i\left(\frac{1}{i} - \frac{1}{n-1}\right) = \tag{2.29}$$

$$= \sum_{i=1}^{n-2}\frac{q_i}{i} - \sum_{i=1}^{n-2}\frac{q_i}{n-1}. \tag{2.30}$$

Tramite l'equazione (2.27), possiamo riscrivere il secondo addendo di (2.30) estraendo gli indici utili:

$$\sum_{i=1}^{n-2} \frac{q_i}{n-1} = \frac{\log\left(\det(A)\right)}{n-1} - \frac{q_n}{n-1} - \frac{q_{n-1}}{n-1}.$$

Possiamo quindi riscrivere l'equazione (2.30) come

$$\sum_{i=1}^{n-2} \sum_{k=i+1}^{n-1} \frac{q_i}{k(k-1)} = \sum_{i=1}^{n-2} \frac{q_i}{i} - \frac{\log\left(\det(A)\right)}{n-1} + \frac{q_n}{n-1} + \frac{q_{n-1}}{n-1}.$$

Lavoriamo ora sul secondo membro dell'equazione (2.28):

$$\sum_{k=2}^{n-1} \left(\frac{\log k}{2(k-1)} + \frac{q_k}{k} \right) = \frac{1}{2} \sum_{k=2}^{n-1} \log k^{\frac{1}{k-1}} + \sum_{k=2}^{n-1} \frac{q_k}{k} \tag{2.31}$$

$$= \frac{1}{2} \log \left(\prod_{k=2}^{n-1} k^{\frac{1}{k-1}} \right) + \sum_{k=2}^{n-1} \frac{q_k}{k}. \tag{2.32}$$

Riprendendo la disequazione (2.28) con i due membri modificati otteniamo:

$$\sum_{i=1}^{n-2} \frac{q_i}{i} + \frac{\log\left((\det A)\right)}{n-1} + \frac{q_n}{n-1} + \frac{q_{n-1}}{n-1} \leq \frac{1}{2} \log(2^1 3^{\frac{1}{2}} \cdot \ldots \cdot (n-2)^{\frac{1}{n-2}}) + \sum_{k=2}^{n-1} \frac{q_k}{k},$$

da cui

$$q_1 + \frac{q_n}{n-1} \leq \log \left(\sqrt{2^1 3^{\frac{1}{2}} \cdot \ldots \cdot (n-2)^{\frac{1}{n-2}}} \right) + \frac{\log\left(\det(A)\right)}{n-1}.$$

Per la disequazione (2.23) otteniamo

$$q_1 + \frac{q_n}{n-1} \leq \log \left(\sqrt{2^1 3^{\frac{1}{2}} \cdot \ldots \cdot (n-2)^{\frac{1}{n-2}}} \right) + \frac{n \log n}{2(n-1)} + \frac{n q_n}{n-1},$$

da cui

$$q_1 - q_n \leq \log \left(\sqrt{2^1 3^{\frac{1}{2}} \cdot \ldots \cdot (n-1)^{\frac{1}{n-2}} n^{\frac{1}{n-1}}} \right) + \frac{1}{2} \log n. \tag{2.33}$$

Denotando $f(n) \stackrel{\text{def}}{=} \sqrt{2^1 3^{\frac{1}{2}} \cdot \ldots \cdot (n-1)^{\frac{1}{n-2}} n^{\frac{1}{n-1}}}$, otteniamo

$$\frac{p_1}{p_n} \leq f(n)\sqrt{n} \leq 2\sqrt{n} n^{\frac{\log n}{4}}. \tag{2.34}$$

□

2.3.2 La recente stima dall'alto

Il risultato ottenuto da Wilkinson è rimasto la miglior stima per oltre 60 anni, quando A. Bisain, A. Edelman e J. Urschel [1] hanno ottenuto una stima dall'alto migliore, dimostrando che

$$\rho_{cp} \leq n^{\frac{\log n}{2[2+(2-\sqrt{2})\log 2]}+0.91} \approx n^{0.2079\log n+0.91}. \tag{2.35}$$

Gli autori hanno riprodotto il risultato di Wilkinson riformulandolo come un problema di ottimizzazione e ottenuto il nuovo bound tramite programmazione lineare.

In particolare, hanno utilizzato le stesse informazioni alla base dello studio di Wilkinson, ma riorganizzandole in modo da evidenziare legami tra i vari pivot. Hanno mostrato che le componenti della matrice non possono crescere molto se la matrice ha valori singolari elevati e sottolineano quanto ulteriori informazioni sulla matrice studiata permettono di ottenere stime migliori.

Inoltre, nel documento specificano che, nonostante le analisi siano state limitate al caso di complete pivoting, queste possono essere eventualmente estese anche alle altre strategie per cui non si è ancora raggiunta una stima definitiva.

Per ulteriori informazioni a riguardo rimandiamo all'articolo [1].

2.3.3 Falsità della congettura di Wilkinson

Le difficoltà legate alla strategia di complete pivoting hanno interessato molti matematici. Tra questi, negli anni Novanta del Novecento, A. Edelman [5] ha presentato un articolo in cui sottolineava i misteri ancora non risolti di questa strategia.

> "The algorithm may be old, but new and unanswered questions continue. Some relate to the practical details of implementing the

> algorithm on new and ever changing architectures. Others concern whether a different algorithm might be more suitable."
>
> - A. Edelman [5]

Tra coloro che hanno cercato di trovare delle risposte, importante è stato Wilkinson, che ha ribadito più volte la difficoltà riscontrata nel costruire una matrice che avesse $\rho(A) > n$, tanto da arrivare a dubitare dell'esistenza effettiva di una matrice del genere. Queste osservazioni hanno condotto Cryer, nel 1968, a presentare la congettura di Wilkinson[2].

Congettura di Wilkinson:

> If Gaussian elimination with complete pivoting is performed on a matrix A, then $\rho_n(A) \leq n$.

Per circa vent'anni non sono stati trovati controesempi e la falsità della congettura è stata dimostrata solo intorno agli anni Novanta. Nel 1991, infatti, Nick Gould [8] ha presentato come controesempio alla congettura una matrice 13×13 per cui $\rho = 13.0205$, aggiungendo che

> "Growth larger than n has also been observed for matrices of orders 14, 15, and 16."
>
> - N. Gould [8]

Importante è sottolineare che, per ottenere questi risultati, Gold ha eseguito calcoli con precisione finita.

L'anno successivo, Alan Edelman, assistito dai suoi studenti Miles Ohlrich e Su-Lin Wu, ha riprodotto i calcoli di Gould in aritmetica esatta per verificarne la correttezza [5].

> "Imagine our surprise when we observed a growth factor of under 7.34 for the matrix that was supposed to give a growth of 13.0205!"
>
> - A. Edelman [5]

Si è così dimostrato che il controesempio di Gould funzionava soltanto per errori dovuti all'arrotondamento della precisione finita. L'iniziale ipotesi di un'effettiva veridicità della congettura è stata subito smentita attraverso una piccola modifica del controesempio di Gould.

Edelman ha infatti notato che al sesto passo dell'eliminazione di Gauss si ottenevano due candidati pivot così vicini da essere indistinguibili nella rappresentazione a precisione finita. Per capire quale fosse il problema, ha creato una funzione che, a partire dalla matrice, restituisse una lista con valore e posizione dei pivot ad ogni passo del metodo di eliminazione di Gauss.

Per capire come intervenire ha confrontato i calcoli in aritmetica esatta con quelli in dimensione finita: in questo modo si è accorto che al sesto passo la scelta del pivot ricadeva in due posizioni diverse. Per risolvere il problema è stato a questo punto sufficiente ridurre il corrispondente elemento nella matrice originale: è bastata infatti la lieve trasformazione dell'elemento in posizione $(11, 10)$ da 1 a $1 - 10^{-7}$ per risolvere il problema.

2.4 Un problema ancora irrisolto

Nelle sezioni precedenti ci siamo occupati di dimostrare i risultati relativi ai limiti superiori del fattore di crescita, che riassumiamo nella Tabella 2.2.

Possiamo osservare che, con la strategia di partial pivoting, il limite superiore del fattore di crescita aumenta esponenzialmente con la dimensione della matrice. Si tratta di una stima molto pessimistica, in quanto i casi peggiori

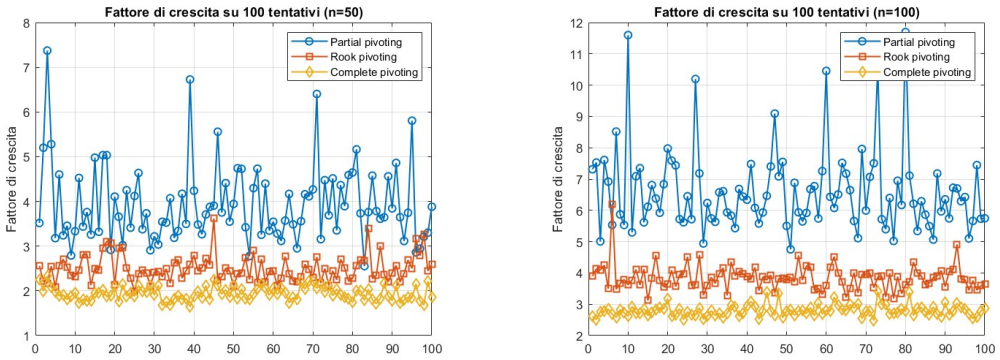
Strategia di Pivoting	Stima superiore
Partial Pivoting	$\rho_n \leq 2^{n-1}$
Rook Pivoting	$\rho_n \leq 1.5n^{\frac{3}{4}logn}$
Complete Pivoting	$\rho_n \leq n^{0.2079logn+0.5}$

Tabella 2.2: Limiti superiori del fattore di crescita con pivoting

sono pochi, e proprio per questo, nella pratica, il partial pivoting è ampiamente utilizzato.

Il rook e il complete pivoting permettono invece di ottenere limiti superiori più contenuti, evitando così la possibile crescita esponenziale del partial pivoting. Il complete pivoting, in particolare, rappresenta la strategia più robusta per garantire la stabilità del metodo, ma spesso è preferibile non utilizzarla nella pratica per il costo computazionale maggiore rispetto alle altre due.

Mostriamo ora che le stime dimostrate sono molto pessimistiche, riportando in figura 2.1 i grafici dei fattori di crescita ottenuti su 100 matrici casuali di dimensioni rispettivamente 50×50 (figura 2.1a) e 100×100 (figura 2.1b) utilizzando le strategie di partial, rook e complete pivoting.



(a) $n = 50$ (b) $n = 100$

Figura 2.1: Fattore di crescita per matrici random

Da entrambe le figure possiamo osservare che il partial pivoting produce fattori di crescita in genere più elevati e variabili rispetto alle altre due strategie. In rook e complete pivoting il fattore di crescita si mantiene infatti più moderato, confermando così una stabilità maggiore. Si osserva inoltre che, con l'aumentare delle dimensioni delle matrici prese in considerazione, aumenta anche il fattore di crescita; tuttavia tale crescita rimane in generale contenuta, non compromettendo la stabilità del metodo.

Se da un lato non è stato un problema trovare un limite superiore per il partial pivoting che fosse definitivo, in quanto abbiamo esempi di matrici per cui questa stima viene effettivamente raggiunta, analizzare lo stesso problema nei casi di rook e complete pivoting si è rivelato ben più complesso. La ricerca di un limite superiore sempre più accurato per il fattore di crescita è infatti un problema tuttora oggetto di studi.

I primi risultati di questa ricerca sono stati raggiunti intorno agli anni Sessanta del Novecento, ma gli studi sono ancora in avanzamento. Basti pensare che il risultato riportato in tabella 2.2 per il complete pivoting è stato ottenuto a febbraio del 2025.

> "Although the growth factor is one of the most well-known quantities in numerical analysis, its behavior when pivoting is used is not completely understood"
>
> - N. J. Higham, D. J. Higham [11]

Nel corso degli anni, molti matematici hanno avanzato ipotesi, in alcuni casi poi smentite e in altri confermate:

> "It is our experience that any substantial increase in size of elements of successive A_n is extremely uncommon even with partial pivoting

> [...]. No example which has arisen naturally has in my experience given an increase by a factor as large as 16".
>
> - James Wilkinson [18]

Inizialmente si pensava che, nonostante fosse possibile trovare casi patologici di matrici con fattori di crescita estremamente elevati, tali situazioni non si sarebbero presentate naturalmente in altri contesti.

Successivamente, altri matematici sono intervenuti nel dibattito. Tra questi, Higham e Higham [11] hanno trovato diversi esempi non forzati con fattore di crescita tra $\frac{n}{2}$ e n. Tuttavia, sebbene fosse stata superata l'esperienza di Wilkinson, i risultati ottenuti si mantenevano ben al di sotto del limite teorico 2^{n-1}.

Alcuni dei risultati più importanti sono stati ottenuti da Wright [20] e Foster [6] che, anche se in modi diversi, hanno presentato degli esempi pratici di crescite esponenziali. Da un lato, Wright ha presentato una classe di matrici sparse, dall'altro, Foster ha mostrato che si sarebbero incontrate matrici del genere cercando di risolvere le equazioni integrali di Volterra con un metodo di quadratura. Il risultato di Foster si avvicinava di più al limite teorico rispetto a quello di Wright, con matrici dense anzichè sparse.

> "[...] although such growth is rare in practice, Foster and Wright have presented examples where exponential growth is achieved for matrices arising from commonly used discretizations of integral and differential equations."
>
> - L.V. Foster [7]

Capitolo 3

Matrici randsvd con fattore di crescita elevato

3.1 Misura di Haar

Introduciamo la misura di Haar che permetterà di definire le matrici di cui ci occuperemo in questo capitolo.

Indicando con $O_n = \{M \in \mathbb{R}^{n\times n} | M^T M = I\}$ l'insieme delle matrici ortogonali reali, possiamo introdurre la misura di Haar generalizzando la misura di Lebesgue sulla sfera.

Ricordando il caso della sfera tridimensionale, sappiamo che possiamo esprimere ciascun punto della sfera in coordinate rispetto a determinati parametri prefissati (ad esempio coordinate cartesiane o cilindriche). La teoria della misura ci dà i mezzi per pesare diverse porzioni della sfera in base alla loro posizione nello spazio e, in questo modo, selezionare casualmente un punto in modo uniforme (tenendo conto della distribuzione dei punti sulla sfera e delle regioni in cui sono più affollati). Così, siamo in grado di descrivere la distribuzione di ogni parametro, in modo da scegliere casualmente un insieme di punti ottenendo un

insieme sufficientemente uniforme.

Lo stesso discorso può essere fatto per gli elementi del gruppo O_n, le matrici ortogonali di ordine n, su cui possiamo fare operazioni, applicare funzioni e campionarle in modo uniforme, esattamente come quanto detto per i punti della sfera. Per questo gruppo considereremo la misura di Haar, che è una misura di probabilità caratterizzata dall'invarianza per moltiplicazioni sinistre e, nel caso specifico del gruppo O_n, anche per moltiplicazioni destre.

Il metodo migliore per campionare matrici distribuite secondo Haar è selezionare una generica matrice $A \in \mathbb{R}^{n\times n}$ con gli elementi $A_{i,j}$ con distribuzione normale standard. Ricavando la fattorizzazione QR di A, $A = QR$, con $Q \in O_n$ e $R \in \mathbb{R}^{n\times n}$ triangolare inferiore con diagonale positiva, ovvero $R_{i,i} > 0$ per ogni $i = 1, \ldots, n$. Allora la matrice Q sarà una matrice ortogonale distribuita secondo Haar.

Per ulteriori approfondimenti riguardo la misura di Haar rimando alle pubblicazioni [3] di Olivia di Matteo e [12] di Michael Howes.

In questo capitolo analizziamo una particolare famiglia di matrici per cui otteniamo fattori di crescita elevati per qualsiasi strategia di pivoting. Si tratta di matrici della forma

$$A = U\Sigma V^T, \tag{3.1}$$

dove $\Sigma = \mathrm{diag}\{1, \ldots, 1, \sigma_n\}$ con $\sigma_n \leq 1$, mentre U e V sono matrici ortogonali dalla distribuzione di Haar. Faremo riferimento a queste matrici con il termine 'randsvd'. Possiamo riscrivere (3.1) come somma di matrici di rango 1:

$$A = U\Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T = \sum_{i=1}^{n-1} u_i v_i^T + u_n \sigma_n v_n^T.$$

Inoltre, aggiungendo e sottraendo $u_n v_n^T$, è possibile interpretare A come perturbazione di una matrice ortogonale dalla distribuzione di Haar.

$$A = \sum_{i=1}^{n} u_i v_i^T + (\sigma_n - 1) u_n v_n^T = UV^T + (\sigma_n - 1) u_n v_n^T. \tag{3.2}$$

3.2 Matrici ortogonali dalla distribuzione di Haar

Consideriamo il caso particolare $\sigma_n = 1$, per il quale otteniamo una matrice della forma $A = UV^T$, distribuita secondo Haar, per invarianza rispetto alla moltiplicazione.

In questa sezione dimostreremo che matrici ortogonali dalla distribuzione di Haar hanno fattori di crescita tipicamente elevati per qualsiasi strategia di pivoting; in particolare, proveremo il seguente risultato

$$\rho(A) \geq \frac{n}{4 \log n}. \tag{3.3}$$

I primi contributi in questa direzione si devono ai matematici Donoho e Huo [4, Thm. VIII.1], che hanno dimostrato che, con elevata probabilità, le matrici ortogonali dalla distribuzione di Haar hanno termine massimo in valore assoluto inferiore a $2\sqrt{\frac{\log n}{n}}$.

Più nel dettaglio, sia $A \in \mathbb{R}^{n \times n}$ distribuita secondo Haar, vale la stima

$$P\left(\max_{i,j} |A_{i,j}| > 2\sqrt{\frac{\log n}{n}}(1 + \epsilon) \right) \to 0. \tag{3.4}$$

Successivamente, un risultato più forte è stato ottenuto da Jiang [13, Prop. 1], che ha dimostrato la convergenza in probabilità

$$\sqrt{\frac{n}{\log n}} \max_{i,j} |A_{i,j}| \xrightarrow[n \to \infty]{P} 2. \tag{3.5}$$

Ci dedichiamo ora a dimostrare il risultato (3.5), introducendo un lemma e un teorema preliminare che useremo nella dimostrazione della proposizione finale. Per tutto il resto della sezione useremo la seguente notazione. Sia $A \in \mathbb{R}^{n\times n}$, indichiamo:

- a_i, con $i = 1, \ldots, n$, l'i-esima colonna di A;
- $b_n := 4\log n - \log(\log n)$;
- $J_x := \sqrt{\frac{1}{2\pi}}e^{-\frac{x}{2}}$, con $x \in \mathbb{R}$;
- $\alpha = \max_{i,j} |A_{i,j}|$, con $i, j = 1, \ldots, n$.

Lemma 3.2.1 (Jiang [13, Lem. 2.2])**.** *Sia $A \in \mathbb{R}^{n\times n}$ una matrice ortogonale dalla distribuzione di Haar. Definiamo l'evento*

$$A_j := \sqrt{n}||a_j||_\infty \geq \sqrt{b_n + x} \quad \textit{per } x > -a_n,\ j = 1, \ldots, n.$$

Allora, per ogni $m \geq 1$, per ogni $x \in \mathbb{R}$ vale

$$\lim_{n\to\infty} n^m P(A_1 \cap \ldots \cap A_m) = \left(\sqrt{\frac{1}{2\pi}}e^{-\frac{x}{2}}\right)^m. \tag{3.6}$$

Teorema 3.2.2. *Sia $A \in \mathbb{R}^{n\times n}$ una matrice ortogonale dalla distribuzione di Haar. Allora, per ogni $x \in \mathbb{R}$, vale*

$$\lim_{n\to\infty} P(n\alpha^2 - b_n \leq x) = e^{-\sqrt{\frac{1}{2\pi}}e^{-\frac{x}{2}}}. \tag{3.7}$$

Dimostrazione. Dimostrare il risultato (3.7) equivale a dimostrare che

$$P(\sqrt{n}\alpha \leq \sqrt{x + b_n}) \to e^{-J_x},$$

ovvero

$$P(\sqrt{n}\alpha > \sqrt{x + b_n}) \to 1 - e^{-J_x}.$$

Per ciascuna colonna, riprendiamo la definizione dell'evento A_j nel Lemma 3.2.1 e, poiché $\alpha = \max_{i,j} |A_{i,j}| = max_j ||a_j||_\infty$, si ha

$$\{\sqrt{n}\alpha > \sqrt{x + b_n}\} = \bigcup_{j=1}^{n} A_j.$$

Dunque il problema si riduce a dimostrare che

$$P(\max_{i \le j \le n} \{\sqrt{n}||a_j||_\infty \ge \sqrt{x + b_n}\}) \to 1 - e^{-J_x},$$

ovvero a valutare la probabilità dell'unione

$$P(\bigcup_{j=1}^{n} A_j) \to 1 - e^{-J_x}.$$

Per la disuguaglianza di Bonferroni (rimando all'Osservazione 3.2.3), fissato $m \ge 1$ tale che $n \ge 2m$, otteniamo come limiti della probabilità dell'unione

$$\sum_{j=1}^{2m} (-1)^{j+1} S_j \le P(\bigcup_{j=1}^{n} A_j) \le \sum_{j=1}^{2m+1} (-1)^{j+1} S_j, \tag{3.8}$$

con $S_j = \sum_{1 \le i_1 \le \ldots \le i_j} P(A_1 \cap \ldots \cap A_j) \quad$ per $1 \le j \le n$.

Per l'invarianza della misura di Haar rispetto a permutazioni di righe e colonne, segue che $P(A_j) = P(A_k)$ per ogni $j, k = 1, \ldots, n$ e $P(A_{j_1} \cap \ldots \cap A_{j_k})$ dipende unicamente da k e non dagli indici scelti. Possiamo quindi riscrivere l'equazione (3.8) nel modo seguente:

$$\sum_{j=1}^{2m} \binom{n}{j} (-1)^{j+1} P(A_1 \cap \ldots \cap A_j) \le P(\bigcup_{j=1}^{n} A_j) \le \sum_{j=1}^{2m+1} \binom{n}{j} (-1)^{j+1} P(A_1 \cap \ldots \cap A_j)$$

Poiché vale che $\dfrac{\binom{n}{i}}{n^i} \to \dfrac{1}{i!}$ e, per il Lemma 3.2.1,

$$n^k P(A_1 \cap \ldots \cap A_k) \to \left(\sqrt{\frac{1}{2\pi}} e^{-\frac{x}{2}}\right)^m,$$

dividendo e moltiplicando il k-esimo addendo per n^k, per ogni $k = 1, \ldots, 2m$, otteniamo come limite per la stima dal basso:

$$\sum_{k=1}^{2m}(-1)^{k+1}n^k\frac{\binom{n}{k}}{n^k}P(A_1 \cap \ldots \cap A_k) \underset{n\to\infty}{\to} \sum_{k=1}^{2m}\frac{(-1)^{k+1}}{k!}\left(\sqrt{\frac{1}{2\pi}}e^{-\frac{x}{2}}\right)^k. \quad (3.9)$$

Analogamente, il limite della stima superiore sarà

$$\sum_{k=1}^{2m+1}(-1)^{k+1}n^k\frac{\binom{n}{k}}{n^k}P(A_1 \cap \ldots \cap A_k) \underset{n\to\infty}{\to} \sum_{k=1}^{2m+1}\frac{(-1)^{k+1}}{k!}\left(\sqrt{\frac{1}{2\pi}}e^{-\frac{x}{2}}\right)^k. \quad (3.10)$$

Unendo i risultati (3.9) e (3.10), segue che

$$\sum_{k=1}^{2m}\frac{(-1)^{k+1}}{k!}\left(\sqrt{\frac{1}{2\pi}}e^{-\frac{x}{2}}\right)^k \leq \lim_{n\to\infty}P(\bigcup_{j=1}^{n}A_j) \leq \sum_{k=1}^{2m+1}\frac{(-1)^{k+1}}{k!}\left(\sqrt{\frac{1}{2\pi}}e^{-\frac{x}{2}}\right)^k.$$

Ora, riprendendo lo sviluppo in serie di Taylor dell'esponenziale

$$e^{-J_x} = 1 + \sum_{i=1}^{\infty}\frac{(-1)^i J_x^i}{i!},$$

vediamo che, facendo tendere m all'infinito, otteniamo

$$\sum_{k=1}^{2m}(-1)^{k+1}\frac{J_x^k}{k!} \to 1 - e^{-J_x},$$

da cui

$$1 - e^{-J_x} \leq \lim_{n\to\infty}P(\bigcup_{j=1}^{n}A_j) \leq 1 - e^{-J_x}.$$

Per il teorema del confronto possiamo concludere che

$$P(\max_{i\leq j\leq n}\{\sqrt{n}||a_j||_\infty \geq \sqrt{x + b_n}\}) \to 1 - e^{-J_x} \quad \text{per ogni } x > 0.$$

□

Osservazione 3.2.3. Nella dimostrazione abbiamo usato la disuguaglianza di Bonferroni [17], che permette di dare un limite inferiore e uno superiore alla

probabilità dell'unione di eventi. In particolare, sia $\{A_j\}_{j=1,\dots,n}$ una famiglia di eventi, definiamo

$$S_1 = \sum_{j=1}^{n} P(A_j),$$
$$S_2 = \sum_{1\le i\le j\le n} P(A_i \cap A_j),$$
$$S_k = \sum_{1\le i_1\le \dots\le i_k} P(A_1 \cap \dots \cap A_k) \quad \text{con } 1 \le k \le n.$$

Allora, per $k \geq 1$ dispari vale

$$P(\bigcup_{i=1}^{n} A_j) \leq \sum_{j=1}^{n} (-1)^{j+1} S_j$$

Mentre per $k \geq 2$ pari vale

$$P(\bigcup_{i=1}^{n} A_j) \geq \sum_{j=1}^{k} (-1)^{j+1} S_j$$

Dimostriamo ora la proposizione che permetterà di ottenere il risultato anticipato in (3.5), in modo da poter affermare che, molto probabilmente, gli elementi della matrice A saranno in modulo inferiori a $2\sqrt{\frac{\log n}{n}}$.

Proposizione 3.2.4. *Siano $A \in \mathbb{R}^{n\times n}$ una matrice ortogonale dalla distribuzione di Haar e $\alpha = \max_{i,j} |A_{i,j}|$ con $i, j = 1, \dots, n$. Allora*

$$\sqrt{\frac{n}{\log n}}\alpha \xrightarrow[n\to\infty]{P} 2. \tag{3.11}$$

Dimostrazione. Dimostrare la convergenza in probabilità, implica dimostrare che

$$P(\left|\sqrt{\frac{n}{\log n}}\alpha - 2\right| \geq \epsilon) \to 0.$$

La disuguaglianza può essere riscritta come

$$\sqrt{\frac{n}{\log n}}\alpha \leq 2-\epsilon \quad \text{o} \quad \sqrt{\frac{n}{\log n}}\alpha \geq 2+\epsilon,$$

ovvero

$$n\alpha^2 \leq (2-\epsilon)^2 \log n \quad \text{o} \quad n\alpha^2 \geq (2+\epsilon)^2 \log n.$$

Da cui segue che

$$P(\left|\sqrt{\frac{n}{\log n}}\alpha - 2\right| \geq \epsilon) = P(n\alpha \leq (2-\epsilon)^2 \log n) + P(n\alpha \geq (2+\epsilon)^2 \log n).$$

Dimostriamo ora che entrambi gli addendi del secondo membro tendono a 0 al crescere di n.

$$\begin{aligned}
&\limsup_{n\to\infty} P(n\alpha^2 \geq (2+\epsilon)^2 \log n) = \\
=&\limsup_{n\to\infty} P\left(n\alpha^2 - 4\log n + \log(\log n) \geq ((2+\epsilon)^2 - 4)\log n + \log(\log n)\right) \leq \\
\leq&\limsup_{n\to\infty} (n\alpha^2 - 4\log n + \log(\log n) > x) \overset{(3.2.2)}{=} 1 - e^{-J_x}, \quad \text{per ogni } x > 0.
\end{aligned}$$

L'ultima disuguaglianza segue dal fatto che $((2+\epsilon)^2-4)\log n + \log(\log n) \overset{n\to\infty}{\to} +\infty$, dunque, per ogni x fissato, esisterà $\hat{n}$ tale che per ogni $n \geq \hat{n}$, è verificato $((2+\epsilon)^2 - 4)\log n + \log(\log n) \geq x$. □

Dimostriamo ora il risultato (3.7) e. per farlo, vediamo il seguente teorema, presentato da D. J. Higham e N. J. Higham [11].

Teorema 3.2.5. *Siano $A \in \mathbb{R}^{n\times n}$ non singolare, $\alpha = \max_{i,j} |A_{i,j}|$, $\beta = \max_{i,j} |(A^{-1})_{i,j}|$ e $\theta = (\alpha\beta)^{-1}$. Allora per qualsiasi scelta di matrici di permutazione Π_c e Π_r tali che la matrice $\Pi_r A \Pi_c$ ammette una fattorizzazione LU, il fattore di crescita soddisfa $\rho(A) \geq \theta$.*

Dimostrazione. Notiamo innanzitutto che $\sum_{j=1}^{n} A_{i,j}(A^{-1})_{i,j} = 1$ (infatti l'i-esima riga di A moltiplicata per l'i-esima colonna di A^{-1} restituisce e_i).

$$\Pi_r A \Pi_c = LU,$$

$$|U^{-1}_{n,n}| = |e_n^T U^{-1} e_n| = |e_n^T U^{-1}(L^{-1}e_n)| = |e_n^T \Pi_c^T A^{-1} \Pi_r^T e_n|.$$

La matrice $e_n^T \Pi_c^T$ seleziona una riga della matrice A^{-1}, mentre la matrice $\Pi_r^T e_n$ una colonna, dunque esisteranno $i, j = 1, \ldots, n$ tali che

$$|U^{-1}_{n,n}| = |(A^{-1})_{i,j}| \leq \beta.$$

Segue che $\max_{i,j} |U_{i,j}| \geq |U_{n,n}| \geq \beta^{-1}$, che permette di cocludere. □

Torniamo ora al caso particolare di matrici ortogonali. Sia $A \in \mathbb{R}^{n\times n}$ matrice ortogonale dalla distribuzione di Haar, vale che $A^{-1} = A^T$ e $\alpha = \beta = 2\sqrt{\frac{\log n}{n}}$. Per la Proposizione 3.2.4 concludiamo che

$$\rho(A) \geq \frac{n}{4 \log n}$$

per n sufficientemente grande.

3.3 Perturbazioni di matrici ortogonali generiche

In questa sezione mostriamo che perturbazioni di rango 1 di matrici ortogonali dalla distribuzione di Haar mantengono il fattore di crescita dello stesso ordine; consideriamo quindi una matrice $A \in \mathbb{R}^{n\times n}$ della forma:

$$A = UV^T + (\sigma_n - 1)u_n v_n^T.$$

3.3.1 Perturbazione di rango 1 generica

Lavoriamo innanzitutto su un caso generale di perturbazione di rango 1, per poi passare ad analizzare i risultati ottenuti nel caso specifico d'interesse. Sia $B \in \mathbb{R}^{n\times n}$, $x, y \in \mathbb{R}^n$. Definiamo $A = B + xy^T$ perturbazione di rango 1 della matrice B. Assumiamo inoltre che esistano le fattorizzazioni LU sia di B che di A, ovvero:

- $B = LU$;
- $A = \tilde{L}\tilde{U}$.

Osservazione 3.3.1. In questo paragrafo utilizzeremo la seguente notazione. Siano $A \in \mathbb{R}^{n\times n}$ e $u \in \mathbb{R}^m$, $v \in \mathbb{R}^k$, $m \leq n$ e $k \leq n$ vettori di indici, indichiamo con:

- $A(u, v)$ la sottomatrice formata dall'intersezione degli indici di riga in u e degli indici di colonna in v;
- A_i la sottomatrice $A(1 : i, 1 : i)$, formata dall'intersezione tra le prime i righe e i colonne.

Otteniamo ora una formula che permetta di esplicitare gli elementi $U_{i,j}$, rifacendoci ai risultati di N. J. Higham [10]. In particolare vogliamo dimostrare che

$$U_{i,j} = \frac{\det(B(1 : i, [1 : i-1, j]))}{\det(B_{i-1})}. \tag{3.12}$$

Per farlo, riscriviamo $B = LDR$, dove R è una matrice triangolare superiore con diagonale unitaria, mentre D è la matrice diagonale $D = \mathrm{diag}(d_1, \ldots, d_n)$ tale che $d_i = U_{i,i}$. Fissato $p \in \mathbb{N}$, $p \leq n$, partizioniamo le matrici come segue.

$$B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix} = \begin{pmatrix} L_1 & \\ L_3 & L_4 \end{pmatrix} \begin{pmatrix} D_1 & \\ & D_4 \end{pmatrix} \begin{pmatrix} R_1 & R_2 \\ & R_4 \end{pmatrix}$$

Con $B_1, L_1, D_1, R_1 \in \mathbb{R}^{p\times p}$, $B_4, L_4, D_4, R_4 \in \mathbb{R}^{(n-p)\times(n-p)}$, $L_3 \in \mathbb{R}^{(n-p)\times p}$ e $R_2 \in \mathbb{R}^{p\times(n-p)}$.

Poiché $U = DR$ segue che

$$U_{i,i} = d_i, \tag{3.13}$$

$$U_{i,j} = d_i R_{i,j}. \tag{3.14}$$

Osservando che $B_1 = L_1 D_1 R_1$, segue che:

$$\det(B_1) = \det(L_1)\det(\mathrm{D}_1)\det(\mathrm{R}_1)\,.$$

Poiché L_1 e R_1 sono matrici triangolari (rispettivamente inferiore e superiore) con diagonale unitaria, segue che $\det(L_1) = \det(R_1) = 1$. Da cui otteniamo

$$\det(\mathrm{B}_1) = \det(\mathrm{D}_1) = d_1 \cdot \ldots \cdot d_p.$$

Poiché $\det(D_p) = d_p \det(D_{p-1}) = d_p \det(B_{p-1})$, si ricava che:

$$U_{i,i} = d_i = \frac{\det(B_p)}{\det(B_{p-1})}. \tag{3.15}$$

Definiamo ora un'espressione esplicita per gli elementi non diagonali di U. Consideriamo la seguente identità

$$\begin{bmatrix} B_p & B_{n-p} \end{bmatrix} = L_p D_p \begin{bmatrix} R_p & R_{n-p} \end{bmatrix}$$

e selezioniamo le prime $p-1$ e la j-esima colonna. Allora si ottiene

$$\begin{aligned} &\det(B(1:p,[1:p-1,j])) = \\ &\det(L_p)\det(D_p)\det\begin{bmatrix} R_p(1:p,1:p-1) & R_{n-p}(1:p,j) \end{bmatrix}. \end{aligned}$$

Poiché $\begin{bmatrix} R_p(1:p,1:p-1) & R_{n-p}(1:p,j) \end{bmatrix}$ è una matrice triangolare superiore con diagonale unitaria ad eccezione dell'ultimo termine $R_{p,j}$, segue che

$$\det\left[\begin{pmatrix} R_p(1:p,1:p-1) & R_{n-p}(1:p,j) \end{bmatrix} = R_{p,j}.$$

Considerato che $\det(L_p) = 1$ e $\det(D_p) = d_p \det(A_{p-1})$, si ricava che

$$\det(B(1:p, [1:p-1, j])) = \det(B_{p-1}) d_p R_{p,j} \overset{(3.14)}{=} \det(B_{p-1}) U_{p,j}.$$

Possiamo quindi concludere con la formula cercata.

$$U_{i,j} = \frac{\det(B(1:i, [1:i-1, j]))}{\det(B_{i-1})}.$$

Tornando al problema iniziale, lavoriamo sulla matrice $A = B + xy^T$ con il fine di trovare una relazione tra i termini $U_{i,j}$ e $\tilde{U}_{i,j}$. Assumiamo che B sia non singolare, in modo da poter riscrivere A come

$$A = B + xy^T = B(I + B^{-1}xy^T),$$

da cui segue che

$$\det(A) = \det(B) \det(I + B^{-1}xy^T) = \det(B)(1 + y^T B^{-1} x). \tag{3.16}$$

Osservazione 3.3.2. La seconda uguaglianza in (3.16) segue dall'identità di Sylvester: date $U \in \mathbb{R}^{m \times n}$ e $V \in \mathbb{R}^{n \times m}$

$$\det(I_m + UV) = \det(I_n + VU). \tag{3.17}$$

A partire dai risultati ottenuti nell'equazione 3.12, si ha

$$\tilde{U}_{i,j} = \frac{\det(A(1:i, [1:i-1, j]))}{\det(A_{i-1})}. \tag{3.18}$$

Consideriamo separatamente numeratore e denominatore.

Per quanto riguarda il numeratore, $\det(A(1:i, [1:i-1, j]))$ coincide con

$$\det(B(1:i, [1:i-1, j]))(1 + y([1:i-1, j])^T B(1:i, [1:i-1, j])^{-1} x(1:i)).$$

Per il denominatore avremo invece

$$\det(A_{i-1}) = \det(B_{i-1})(1 + y(1:i-1)^T B_{i-1} x(1:i)).$$

Ne consegue che

$$\tilde{U}_{i,j} = U_{i,j} \frac{1 + y([1:i-1,j])^T B(1:i,[1:i-1,j])^{-1} x(1:i)}{1 + y(1:i-1)^T B_{i-1} x(1:i)},$$

da cui otteniamo infine

$$\frac{\tilde{U}_{i,j}}{U_{i,j}} = \frac{1 + y([1:i-1,j])^T B(1:i,[1:i-1,j])^{-1} x(1:i)}{1 + y(1:i-1)^T B_{i-1} x(1:i)}. \tag{3.19}$$

3.3.2 Caso particolare

Valutiamo ora il risultato ottenuto in (3.19) per $A \in \mathbb{R}^{n\times n}$, definita come

$$A = W + xy^T,$$

con $W \in \mathbb{R}^{n\times n}$ matrice ortogonale, $x, y \in \mathbb{R}^n$ tali che $||x||_2 \leq 1$ e $||y||_2 \leq 1$.

Assumendo che i termini massimi di $\tilde{U}$ e U siano $\tilde{U}_{n,n}$ e $U_{n,n}$, ci limitiamo a valutare il rapporto

$$\frac{\tilde{U}_{n,n}}{U_{n,n}} = \frac{1 + y^T W^T x}{1 + y(1:n-1)^T W_{n-1} x(1:n-1)}. \tag{3.20}$$

L'obiettivo è dimostrare che, generalmente, il rapporto (3.20) ha ordine 1. Per analizzare il comportamento più tipico, scegliamo x e y distribuiti uniformemente sulla sfera unitaria in $\mathbb{R}^n$ e analizziamo i valori attesi del quoziente (3.20) introducendo due lemmi sul valore atteso del prodotto scalare.

Lemma 3.3.3 (Kenney and Laub [14, Thm. 2.1, Lem. 6.1])**.** *Sia $z \in \mathbb{R}^n$ distribuito uniformemente nella sfera n-dimensionale unitaria e sia $g \in \mathbb{R}^n$ un vettore costante. Allora*

$$\mathbb{E}[|z^T g|] = \mu_n ||g||_2, \tag{3.21}$$

con $\mu_n = \left(\frac{2}{\pi(n-\frac{1}{2})}\right)^{\frac{1}{2}} + O(n^{-\frac{3}{2}})$, $\mu_n < n^{-\frac{1}{2}}$.

Lemma 3.3.4. *Siano $x, y \in \mathbb{R}^n$ vettori indipendenti distribuiti uniformemente sulla sfera unitaria n-dimensionale, e sia $B \in \mathbb{R}^{n\times n}$ una matrice costante. Allora*

$$\mathbb{E}(|y^T Bx|) \leq \frac{\mu_n}{n^{\frac{1}{2}}}||B||_F, \tag{3.22}$$

con $\mu_n = \left(\frac{2}{\pi(n-\frac{1}{2})}\right)^{\frac{1}{2}} + O(n^{-\frac{3}{2}})$, $\mu_n < n^{-\frac{1}{2}}$.

Dimostrazione. Poiché x e y sono vettori indipendenti, possiamo riscrivere il valore atteso come

$$\mathbb{E}_{x,y}(|y^T Bx|) = \mathbb{E}_x(\mathbb{E}_y(|y^T Bx|))$$

e, per il Lemma 3.3.3, segue che

$$\mathbb{E}_{x,y}(|y^T Bx|) = \mu_n \mathbb{E}(||Bx||_F).$$

Per i risultati presentati da Gudmundsoon, Kenney e Laub [9, Lem. 2.2] sappiamo che

$$\mathbb{E}(||Bx||_2^2) = \frac{||B||_F^2}{n}.$$

Inoltre, per la disuguaglianza di Jensen,

$$\mathbb{E}(||Bx||_2) \leq \frac{||B||_F}{n^{\frac{1}{2}}},$$

che permette di concludere. □

Ci concentriamo ora a studiare il valore atteso di numeratore e denominatore dell'equazione (3.20).

Numeratore:

$$\mathbb{E}(|1 + y^T W^T x|) \leq 1 + \mathbb{E}(|y^T W^T x|) \leq 1 + \frac{\mu_n}{n^{\frac{1}{2}}} < 1 + \frac{1}{n}.$$

Dunque il numeratore è di ordine 1.

Denominatore:

$$\mathbb{E}(|1 + y(1:n-1)^T W_{n-1}^{-1} x(1:n-1)|) \leq 1 + \mathbb{E}(|y(1:n-1)^T W_{n-1}^{-1} x(1:n-1)|) = \tag{3.23}$$

$$= 1 + \mathbb{E}(|y^T (W_{n-1}^{-1}, 0)x|) \leq 1 + \mu_n \frac{||W_{n-1}^{-1}||_F}{n^{\frac{1}{2}}} \tag{3.24}$$

$$< 1 + n^{-\frac{1}{2}} \frac{||W_{n-1}^{-1}||_F}{n^{\frac{1}{2}}} = 1 + \frac{||W_{n-1}^{-1}||_F}{n}. \tag{3.25}$$

Resta da valutare la norma $||W_{n-1}^{-1}||_F$. Per farlo riscriviamo W come

$$W = \begin{pmatrix} W_{n-1} & a \\ b^T & W_{n,n} \end{pmatrix}$$

dove $W_{n-1} \in \mathbb{R}^{(n-1)\times(n-1)}$, a, $b \in \mathbb{R}^{(n-1)}$ e $W_{n,n} \in \mathbb{R}$. Dall'ortogonalità di W segue che:

$$W^T W = I_n \qquad \text{e} \qquad W W^T = I_n$$

ovvero

$$W^T W = \begin{pmatrix} W_{n-1}^T W_{n-1} + bb^T & W_{n-1}^T a + bW_{n,n} \\ a^T W_{n-1} + W_{n,n} b^T & a^T a + W_{n,n}^2 \end{pmatrix} = \begin{pmatrix} I_{n-1} & \underline{0} \\ \underline{0}^T & 1 \end{pmatrix}$$

e

$$W W^T = \begin{pmatrix} W_{n-1} W_{n-1}^T + aa^T & W_{n-1} b + aW_{n,n} \\ b^T W_{n-1}^T + W_{n,n} a^T & b^T b + W_{n,n}^2 \end{pmatrix} = \begin{pmatrix} I_{n-1} & \underline{0} \\ \underline{0}^T & 1 \end{pmatrix}.$$

Da cui

$$W_{n-1}^T W_{n-1} + bb^T = I_{n-1} \tag{3.26}$$

e

$$||b||_2^2 + W_{n,n}^2 = 1. \tag{3.27}$$

Ricordando che $||A||_F^2 = \mathrm{tr}(A^T A)$, segue che

$$||W_{n-1}^{-1}||^2 = \mathrm{tr}((W_{n-1}^T W_{n-1})^{-1}) = \mathrm{tr}((I_{n-1} - bb^T)^{-1}).$$

Per la formula di Sherman-Morrison riportata nella Proposizione 1.2.1, vale che

$$(I_{n-1} - bb^T)^{-1} = I_{n-1} + I_{n-1}b(1 - b^T I_{n-1} b)^{-1} b^T I_{n-1} = I_{n-1} - \frac{bb^T}{1 - b^T b},$$

da cui

$$\begin{aligned} ||W_{n-1}^{-1}||^2 &= \mathrm{tr}(I_{n-1} - \frac{bb^T}{1 - ||b||^2}) = \mathrm{tr}(I_{n-1}) - \frac{\mathrm{tr}(bb^T)}{1 - (1 - W_{n,n}^2)} \\ &= n - 1 - \frac{\sum_{i=1}^{n-1} b_i^2}{-W_{n,n}^2} = n - 1 + \frac{1 - W_{n,n}^2}{W_{n,n}^2} = \\ &= n - 1 - 1 + \frac{1}{W_{n,n}^2} = n - 2 + W_{n,n}^{-2}. \end{aligned}$$

Riprendendo l'equazione (3.25) vale che

$$\mathbb{E}(|y(1:n-1)^T W_{n-1}^{-1} x(1:n-1)|)^2 < \frac{n - 2 + W_{n,n}^{-2}}{n^2}.$$

Inoltre, poiché $W^T = W^{-1} = L^{-1}U^{-1}$ e $L_{n,n} = 1$, segue che $W_{n,n} = (L_{n,n})^{-1}(U_{n,n})^{-1} = U_{n,n}^{-1}$. Quindi

$$\begin{aligned} \mathbb{E}(|y(1:n-1)^T W_{n-1}^{-1} x(1:n-1)|)^2 &< \frac{1}{n} - \frac{2}{n^2} + \frac{U_{n,n}}{n^2} < \frac{1}{n} + \frac{U_{n,n}}{n^2} \\ &< \frac{1}{n} + \frac{||U||_{\max}}{n^2} \leq \frac{1}{n} + \frac{\rho ||W||_{max}^2}{n^2}. \end{aligned}$$

Dunque la stima del valore atteso del denominatore sarà

$$\mathbb{E}(|1 + y(1:n-1)^T W_{n-1}^{-1} x(1:n-1)|) < 1 + \sqrt{\frac{1}{n} + \frac{\rho ||W||_{max}^2}{n^2}},$$

che permette di concludere che sia numeratore che denominatore del rapporto (3.20) sono di ordine 1, quindi che i fattori di crescita della matrice ortogonale e della sua perturbazione avranno lo stesso ordine.

In particolare, nel caso di W matrice ortogonale dalla distribuzione di Haar, abbiamo mostrato nella sezione precedente che $||W||_{max} \leq \frac{n}{4 \log n}$,

Nei risultati ottenuti abbiamo valutato unicamente il rapporto $\frac{\tilde{U}_{n,n}}{U_{n,n}}$, ma i risultati possono essere generalizzati al rapporto $\frac{\tilde{U}_{i,j}}{U_{i,j}}$ lavorando sulla sottomatrice di ordine $i \times i$ che compare al numeratore.

Capitolo 4

Risultati sperimentali

4.1 Funzione Gallery

Per valutare sperimentalmente i risultati ottenuti nelle sezioni precedenti, utilizziamo la seguente funzione MATLAB:

$$[A_1, \ldots, A_m] = \texttt{gallery('matrixname', P\_1, ... , P\_n)},$$

che genera la famiglia di matrici specificata da 'matrixname' (con gli altri parametri da inserire che dipendono dalla famiglia scelta). In particolare, nelle prove successive analizzeremo matrici dalle famiglie 'orthog' e 'randsvd'.

La funzione `gallery('orthog', n, j)` genera matrici non casuali di ordine n costruite seguendo uno specifico algoritmo determinato dall'indice j. In particolare, useremo come indici $j = 1, 2, 5, 6$, che determinano le seguenti matrici prestrutturate:

- W_1=`gallery('orthog', n, 1)` avrà $W_{i,j} = \sqrt{\frac{2}{n+1}} \sin(ij\frac{\pi}{n+1})$;
- W_2=`gallery('orthog', n, 2)` avrà $W_{i,j} = \frac{2}{\sqrt{2(n+1)}} \sin(2ij\frac{\pi}{2(n+1)})$;
- W_3=`gallery('orthog', n, 5)` avrà $W_{i,j} = \frac{\sin(\frac{2\pi(i-1)(j-1)}{n})}{\sqrt{n}} + \frac{\cos(\frac{2\pi(i-1)(j-1)}{n})}{\sqrt{n}}$;

- W_4=`gallery('orthog', n, 6)` avrà $W_{i,j} = \sqrt{\frac{2}{n}}\cos(\frac{i-1}{2}\frac{j-1}{2}\frac{\pi}{n})$.

Per quanto riguarda invece la famiglia 'randsvd', useremo la seguente sintassi:

$$A = \texttt{gallery('randsvd', n, k, mode, kl, ku)},$$

dove

- `n` rappresenta la dimensione della matrice;
- `k` rappresenta il numero di condizionamento della matrice (se non specificato è $\sqrt{\frac{1}{u}}$);
- `kl` e `ku` permettono di ottenere matrici a banda, dove `kl` indica la banda inferiore, mentre `ku` la banda superiore;
- `mode` definisce i valori singolari preimpostati che scegliamo: utilizzeremo `mode=2`, che crea matrici con tutti i valori singolari unitari ad eccezione dell'ultimo $\sigma_n = \frac{1}{|k|}$.

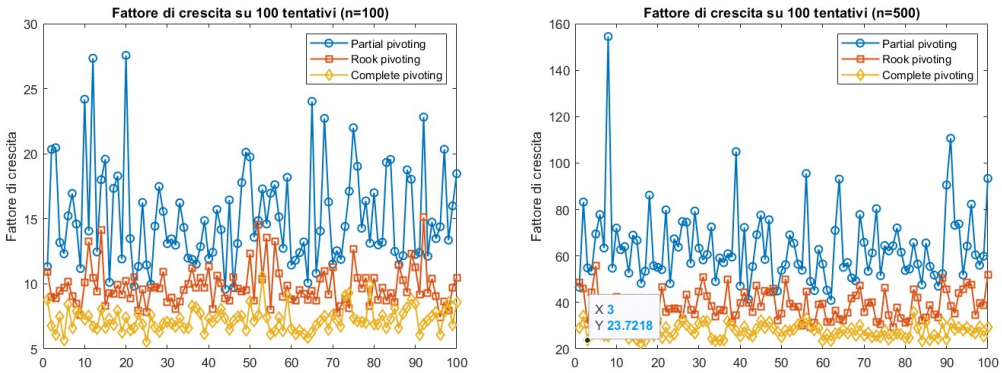
4.2 Matrici randsvd

Consideriamo le matrici della forma

$$A = U\Sigma V^T, \tag{4.1}$$

dove $\Sigma = \text{diag}\{1, \ldots, 1, \sigma_n\}$ con $\sigma_n \leq 1$, U e V matrici ortogonali Haar-distribuite.

La Figura 4.1 mostra, per $n = 100$ (Figura 4.1a) e per $n = 500$ (Figura 4.1b), il fattore di crescita di 100 matrici della forma (4.1), generate con `gallery('randsvd', n, 1e8, 2, [], [])`.



(a) $n = 100$ (b) $n = 500$

Figura 4.1: Fattore di crescita per matrici randsvd

Osserviamo che il partial pivoting produce fattori di crescita notevolmente più elevati e più variabili rispetto alle altre due strategie.

Confrontanto questi risultati con quelli ottenuti in Figura 2.1 per matrici casuali, risulta evidente la presenza di valori nettamente superiori per il fattore di crescita; per $n = 100$, ad esempio, nel primo esperimento abbiamo ottenuto fattori di crescita al più pari a 12, mentre in questo caso abbiamo riscontrato in più esempi fattori di crescita con valori sopra 25.

4.3 Perturbazioni di rango 1

Mostriamo in questa sezione che perturbazioni di rango 1 di matrici con elevato fattore di crescita continuano ad avere un fattore alto. In particolare, consideriamo quattro matrici campione, denotate W_i, $i = 1, \ldots, n$, note per avere un fattore di crescita elevato, e otteniamo per ciascuna di queste quaranta matrici di perturbazione, della seguente forma:

$$A = W_i + xy^T, \tag{4.2}$$

dove i vettori x e y hanno distribuzione uniforme in $(0, 1)$ per le prime venti matrici e distribuzione normale standard per le altre venti. Le matrici W_i sono ottenute tramite la funzione `gallery('orthog', 500, j)` con $j = 1, 2, 5, 6$.

La figura 4.2 riporta i fattori di crescita con partial pivoting della matrice originaria W_i e di ciascuna delle 20 pertubazioni ottenute con vettori di distribuzione uniforme. La figura 4.4 riporta lo stesso risultato ma prendendo x e y con distribuzione normale.

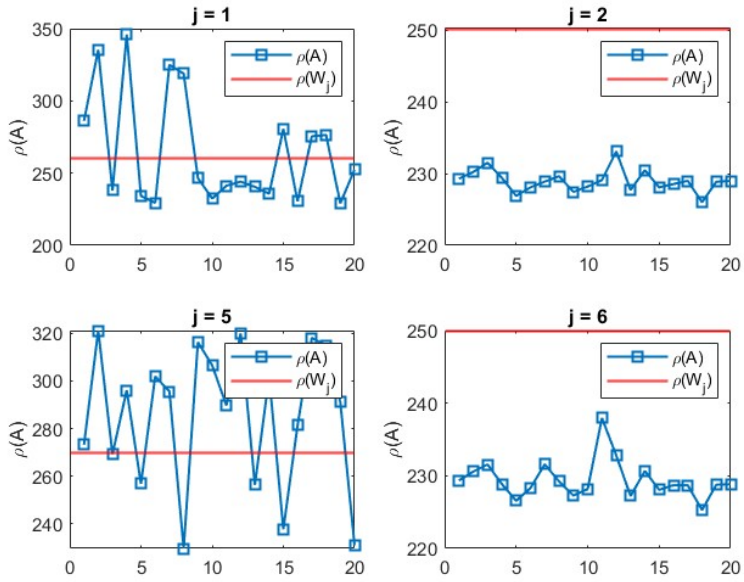


Figura 4.2: Perturbazione con distribuzione uniforme

Possiamo notare che il fattore di crescita delle matrici perturbate si mantiene dello stesso ordine di grandezza delle matrici originarie, a conferma dei risultati teorici ottenuti nel Capitolo 3. Inoltre, il comportamento del fattore di crescita in entrambi gli esperimenti risulta comparabile; possiamo per esempio notare che, sia per la distribuzione uniforme che per la distribuzione normale, per $j = 2$, tutte le matrici generate presentano fattore di crescita inferiore rispetto a quello originario.

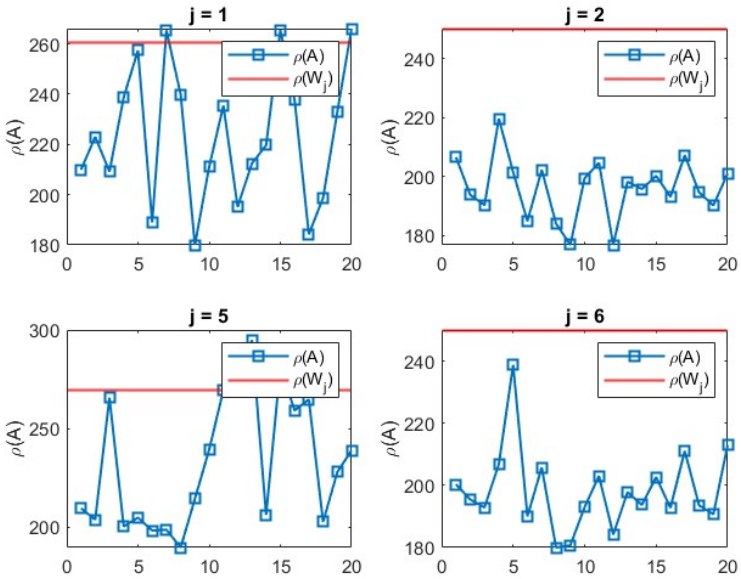


Figura 4.3: Perturbazione con distribuzione normale

Riportiamo nella Tabella 4.1 il fattore di crescita minore, medio e massimo delle matrici così ottenute.

j	$\rho(W_j)$	$\min \rho(A)$	$\text{mean}\, \rho(A)$	$\max \rho(A)$
1	2.61e + 02	2.29e + 02	2.65e + 02	3.46e + 02
2	2.50e + 02	2.26e + 02	2.29e + 02	2.33e + 02
3	2.70e + 02	2.30e + 02	2.85e + 02	3.21e + 02
4	2.50e + 02	2.25e + 02	2.29e + 02	2.38e + 02

Tabella 4.1: Fattore di crescita min, medio e max

Possiamo osservare che il fattore di crescita medio risulta, in generale, vicino a quello della matrice originaria; i valori di minimo e massimo permettono di capire che, nonostante le oscillazioni intorno al valor medio, il fattore di crescita mantiene invariato l'ordine di grandezza.

Consideriamo ora soltanto la matrice

$$W = \texttt{gallery('orthog', n, 1)},$$

al variare di $n = 100 : 100 : 2500$. Per ogni n generiamo 12 matrici della forma

$$A = W + xy^T,$$

con x e y distribuiti uniformemente, e riportiamo in Figura 4.4 il fattore di crescita di W al variare di n e il valor medio del fattore di crescita delle corrispondenti matrici A.

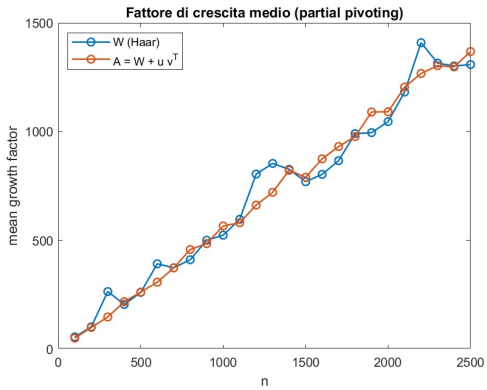


Figura 4.4: Caso specifico - distribuzione uniforme

Osserviamo che il valor medio non si discosta in modo significativo da quello della matrice originaria, a riprova dei risultati ottenuti finora.

Bibliografia

[1] Ankit Bisain, Alan Edelman, and John Urschel. A new upper bound for the growth factor in Gaussian elimination with complete pivoting. *Bulletin of the London Mathematical Society*, 57(5):1369–1387, 2025.

[2] Colin W Cryer. Pivot size in Gaussian elimination. *Numerische Mathematik*, 12(4):335–345, 1968.

[3] Olivia Di Matteo. Understanding the Haar measure. `https://pennylane.ai/qml/demos/tutorial_haar_measure`, March 2021.

[4] David L Donoho, Xiaoming Huo, et al. Uncertainty principles and ideal atomic decomposition. *IEEE transactions on information theory*, 47(7):2845–2862, 2001.

[5] Alan Edelman. The complete pivoting conjecture for Gaussian elimination is false. *The Mathematica Journal*, 3 1992.

[6] Leslie V Foster. Gaussian elimination with partial pivoting can fail in practice. *SIAM Journal on Matrix Analysis and Applications*, 15(4):1354–1362, 1994.

[7] Leslie V Foster. The growth factor and efficiency of Gaussian elimination with rook pivoting. *Journal of Computational and Applied Mathematics*, 86(1):177–194, 1997.

[8] Nick Gould. On growth in Gaussian elimination with complete pivoting. *SIAM Journal on Matrix Analysis and Applications*, 12(2):354–361, 1991.

[9] T. Gudmundsson, C. S. Kenney, and A. J. Laub. Small-sample statistical estimates for matrix norms. *SIAM Journal on Matrix Analysis and Applications*, 16(3):776–792, 1995.

[10] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002.

[11] Nicholas J Higham and Desmond J Higham. Large growth factors in Gaussian elimination with pivoting. *SIAM Journal on Matrix Analysis and Applications*, 10(2):155–164, 1989.

[12] Michael Howes. Haar distributed random matrices. `https://mathstoshare.com/wp-content/uploads/2024/03/random_haar_matrices.pdf`, March 2024. Post su MathToShare.

[13] Tiefeng Jiang. Maxima of entries of haar distributed matrices. *Probability Theory and Related Fields*, 131(1):121–144, 2005.

[14] Charles S Kenney and Alan J Laub. Small-sample statistical condition estimates for general matrix functions. *SIAM Journal on Scientific Computing*, 15(1):36–61, 1994.

[15] George Poole and Larry Neal. The rook's pivoting strategy. *Journal of Computational and Applied Mathematics*, 123(1-2):353–369, 2000.

[16] Davide Palitta Valeria Simoncini. Dispense del corso di calcolo numerico. Lecture notes, Università di Bologna, 2022. Materiale didattico del corso.

[17] Wikipedia contributors. Disuguaglianze di Boole e di Bonferroni. `https://it.wikipedia.org/wiki/Disuguaglianze_di_Boole_e_di_Bonferroni`.

[18] James Hardy Wilkinson. Error analysis of direct methods of matrix inversion. *Journal of the ACM (JACM)*, 8(3):281–330, 1961.

[19] JH Wilkinson. The algebraic eigenvalue problem. In *Handbook for Automatic Computation, Volume II, Linear Algebra*. Springer New York, NY, USA, 1971.

[20] Stephen J Wright. A collection of problems for which Gaussian elimination with partial pivoting is unstable. *SIAM Journal on Scientific Computing*, 14(1):231–238, 1993.

Ringraziamenti

Ed eccoci giunti alla sezione della mia tesi per me più difficile da scrivere ma che effettivamente verrà letta da qualcuno. Innanzitutto ci tengo a ringraziare chi è arrivato fin qui avendo effettivamente letto le altre sezioni, quindi direi, quasi con certezza, solo la mia relatrice Valeria Simoncini ed io. Ringrazio poi chiunque stia in questo momento leggendo questa pagina, che sia perché ci conosciamo o perché si è trovato davanti la mia tesi ed era incuriosito, grazie per l'attenzione.
Posso ora iniziare i ringraziamenti di dovere; un grazie va a mia mamma che, nonostante i rimproveri per il caos che lascio dietro di me, la mia poca presenza e la mia frenesia nel fare le cose, ogni volta che torno da lei a chiedere aiuto è sempre presente e cerca, per quanto possibile, di starmi vicina. Ringrazio poi mio padre che, nonostante il mio bipolarismo e la mia sbadataggine, ancora mi sopporta e mi vizia sempre un sacco volenteroso quando gli chiedo qualcosa (grazie per farmi trovare il caffè pronto ogni mattina e per le lunghe chiacchierate in terrazza). Ringrazio poi mia nonna, che quest'anno è stata probabilmente la persona che ho ammirato di più. Ha affrontato sfide difficilissime senza mollare, mostrando una grinta incredibile che mi rende estremamente fiera di lei. Ringrazio mio nonno e l'Anna: anche voi non avete passato dei periodi facili e tornerò presto a salutarvi per stare di più con voi.
Un ringraziamento speciale va poi alle Wa-Fabeniane e alla mia famiglia allargata che, nonostante non sia necessariamente unita da legami di sangue, è la famiglia più divertente e disagiata (in senso positivo) che io conosca.
Rivolgo anche un pensiero alla mia famiglia più lontana, a mia sorella, ai miei nipoti e in particolare a mio cugino Seba. Ricordo in particolare i miei nonni Jacqueline e Silvio; ogni volta che sento parlare di loro e delle loro avventure rimango incredula

e mi dispiace non poterli avere accanto.
Un immenso grazie va alla Laure e Fede. Siete le uniche persone che mi hanno seguito per quasi tutto il percorso della mia vita, c'eravate nei miei momenti peggiori e nonostante ciò siete ancora qui vicino a me (a proposito, auguri Fede). Non ho ricordi passati senza di voi e non ho intenzione di averne di nuovi. Vi voglio bene.
Grazie poi ad Alice, Anna, Edo, Giulia, Mario e Jacopo perchè mi avete, per quanto possibile, alleggerito la triennale, facendomi venire voglia di studiare solo perché lo facevo con voi e andare a lezione solo perché c'eravate voi. Frequentare ora la magistrale senza di voi ha lasciato un vuoto a lezione, e non voglio neanche pensare a come sarà la sessione. Grazie anche a tutti gli altri ragazzi di matematica, in particolare grazie Pietro, che riesce a mettere ogni giorno alla prova la mia pazienza.
Grazie a tutti i miei compagni di squadra: è la prima volta nella mia vita in cui riesco a giocare spensierata e divertendomi (nonostante rimango comunque una persona piena d'ansia...) ed è tutto merito vostro, che riuscite a capire quando mi abbatto e mi siete sempre vicini. Grazie Tommy per essere costantemente presente per qualsiasi cosa, daltronde alla base di questa tesi c'è la seduta di gelato consolatorio ripetendo calcolo numerico il giorno prima dell'orale. Un grazie speciale anche a Matti, ci hai sempre accolti in casa facendoci sentire benvoluti e per qualsiasi cosa sei sempre pronto e disponibile. Grazie anche a Marti e Giuse, siete le persone più perse che conosca e per questo vi adoro.
Grazie poi a Lu, tu tra tutti sei stato la persona che ha vissuto maggiormente i miei crolli mentali e mi sembra incredibile che riusciremo a laurearci insieme. Quest'anno mi hai insegnato che parlare sinceramente risolve tantissimi problemi e che ogni tanto nella vita bisogna anche fare le cose senza pensarci troppo.
Grazie alla Cate e alla Frenci, voi siete il motivo per cui le Querce occupano quasi metà delle mie giornate, perché so che ogni volta che ci siete voi la giornata sarà molto più leggera e spensierata.
Ringrazio infine Sean e i miei compagni delle superiori, le balotte di palazzo e di Granaglione.