

UNIVERSITÀ DEGLI STUDI DI CAGLIARI
SCUOLA DI DOTTORATO IN INGEGNERIA INDUSTRIALE

Tesina di Metodi Iterativi per la Risoluzione di Sistemi
Lineari e Non Lineari

Applicazione della TSVD allo studio di una colonna di distillazione

Mariarosa Brundu
Anuta Chindris
Andrea Frau

24 Giugno 2009

Decomposizione a valori singolari (SVD)

La decomposizione a valori singolari (SVD) è un importante strumento per la fattorizzazione di matrici basata sull'uso di autovalori e autovettori, spesso utilizzata nella risoluzione di problemi malcondizionati.

Sia data una matrice $A \in R^{m \times n}$, con $m \geq n$: allora è possibile fattorizzare tale matrice come:

$$A = U\Sigma V^T$$

dove

- $U = [u_1, \dots, u_m] \in R^{m \times m}$ è una matrice unitaria;
- $V = [v_1, \dots, v_n] \in R^{n \times n}$ è una matrice unitaria;
- $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n) \in R^{m \times n}$ è una matrice diagonale di elementi non negativi.

gli elementi σ_i costituenti la matrice diagonale Σ sono i valori singolari di A , mentre i vettori u_i e v_i sono i vettori singolari rispettivamente di sinistra e di destra. Tale fattorizzazione è definita decomposizione a valori singolari (SVD) della matrice A .

Per convenzione si è soliti ordinare i valori singolari in maniera decrescente lungo la diagonale.

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$$

La SVD è definita per qualsiasi valore di m e n ; nel caso in cui $m < n$, la SVD si applica alla matrice A^T e si scambiano U e V .

Il legame tra la matrice A e i valori e vettori singolari derivanti dalla decomposizione è espresso dalle seguenti relazioni:

$$Av = \sigma u$$

$$A^T u = \sigma v$$

Decomposizione a valori singolari troncata (TSVD)

Un'importante applicazione della SVD riguarda l'approssimazione della matrice A con una a rango non pieno, ottenuta ponendo a zero alcuni dei valori singolari, a partire dai più piccoli: in tal caso si parla di SVD troncata (TSVD). Tale metodologia viene frequentemente utilizzata nell'elaborazione di immagini e segnali.

Nel caso di sistemi lineari malcondizionati del tipo $Ax = b$, è possibile ottenere una soluzione approssimata attraverso la TSVD, secondo la formula:

$$\sum_{i=1}^k \frac{u_i^T b}{\sigma_i} v_i$$

dove $k < n$, e u_i , v_i e σ_i derivano dalla decomposizione a valori singolari della matrice A .

Condizionamento di un sistema lineare e matrici di Hilbert

Dato il sistema lineare $Ax = b$, si definisce numero di condizionamento della matrice non singolare A la quantità

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

dove $K(A) \geq 1, \forall A$

Se il numero di condizionamento di una matrice è basso si dice che la matrice è ben condizionata, mentre per valori del numero di condizionamento molto alti ($K(A) > 10^{10}$) si dice che la matrice sia malcondizionata.

Nel caso di un sistema lineare malcondizionato una piccola variazione in uno dei coefficienti della matrice o del vettore dei termini noti provoca una grande variazione nella soluzione del sistema. Assumendo di perturbare con un vettore δb , il termine noto allora anche la soluzione del sistema sarà pari ad $x + \delta x$, e varrà la condizione

$$\frac{\|\delta x\|}{\|x\|} \leq K(A) \frac{\|\delta b\|}{\|b\|}$$

ossia un errore relativo in b anche molto basso porta ad un alto errore relativo nel vettore delle soluzioni nel caso di un condizionamento molto elevato.

Un classico esempio di matrice malcondizionata è la matrice di Hilbert, ossia una matrice quadrata avente come elemento a_{ij} il numero razionale $\frac{1}{i+j-1}$.

Come si può notare in Figura 1, il condizionamento della matrice di Hilbert cresce molto rapidamente all'aumentare della dimensione della matrice (esponenzialmente fino a $i = 13$), e la matrice diventa malcondizionata ($K(A) \approx 10^{10}$) già quando $i = 8$.

Decomposizione a valori singolari troncata e matrici di Hilbert

A causa del forte aumento del numero di condizionamento delle matrici di Hilbert all'aumentare dell'ordine della stessa, queste si sono dimostrate particolarmente utili per testare il metodo della decomposizione a valori singolari troncata prima che venisse utilizzato sul sistema lineare ricavato da un caso reale.

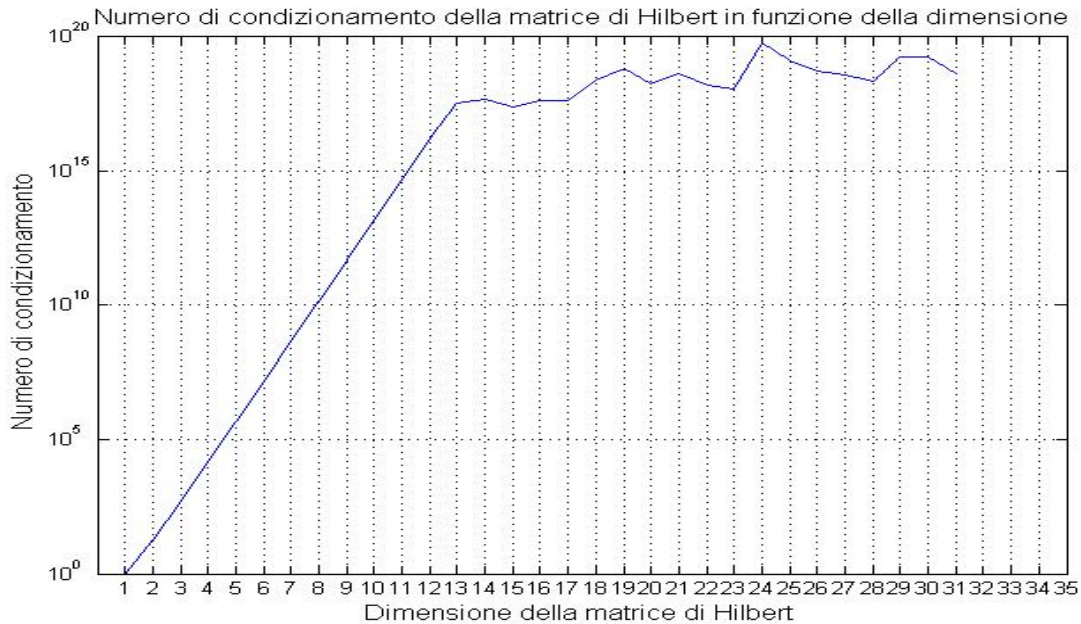


Figura 1: Condizionamento della matrice di Hilbert in funzione della dimensione

In particolare è stato creato un algoritmo che generasse matrici di Hilbert sino al 31-esimo ordine (31 è la dimensione del sistema reale che sarà analizzato nel prossimo capitolo), ed è stato imposto il vettore soluzione allo scopo di valutare l'errore commesso nella risoluzione del sistema con i diversi metodi. Per comodità è stato imposto un vettore soluzione pari a $x = [1 \dots 1]^T$. Attraverso il vettore dei termini noti $b = Ax$ è stato risolto il sistema di equazioni sia attraverso il metodo di Gauss che con il metodo della decomposizione a valori singolari troncata (TSVD). Il grafico dell'errore per la soluzione ottenuta col metodo di Gauss è mostrato in Figura 2, e come si può notare, la norma dell'errore cresce all'aumentare dell'ordine della matrice, ossia all'aumentare del numero di condizionamento. Già per un ordine della matrice pari a 11 la norma dell'errore è circa 10^{-2} . Da ciò appare evidente la necessità di utilizzare metodi alternativi per la risoluzione di problemi fortemente malcondizionati come nel caso della matrice di Hilbert con ordine superiore a 10.

Come precedentemente descritto, tra i diversi metodi disponibili in letteratura per risolvere problemi malcondizionati come questo, è stato scelto il metodo della decomposizione a valori singolari troncata.

Attraverso l'algoritmo costruito è stato possibile analizzare diverse opzioni di troncamento al variare delle dimensioni della matrice di Hilbert: in particolare è stato scelto come troncamento ottimale quello che permette di minimizzare la norma dell'errore. Dal grafico in Figura 3, si può notare che per matrici di ordine inferiore a 13 la soluzione che consente di minimizzare la norma

dell'errore è quella che considera tutti i valori singolari, mentre per matrici di dimensioni maggiori, quindi con numero di condizionamento maggiore, il numero di valori singolari da utilizzare è minore della dimensione.

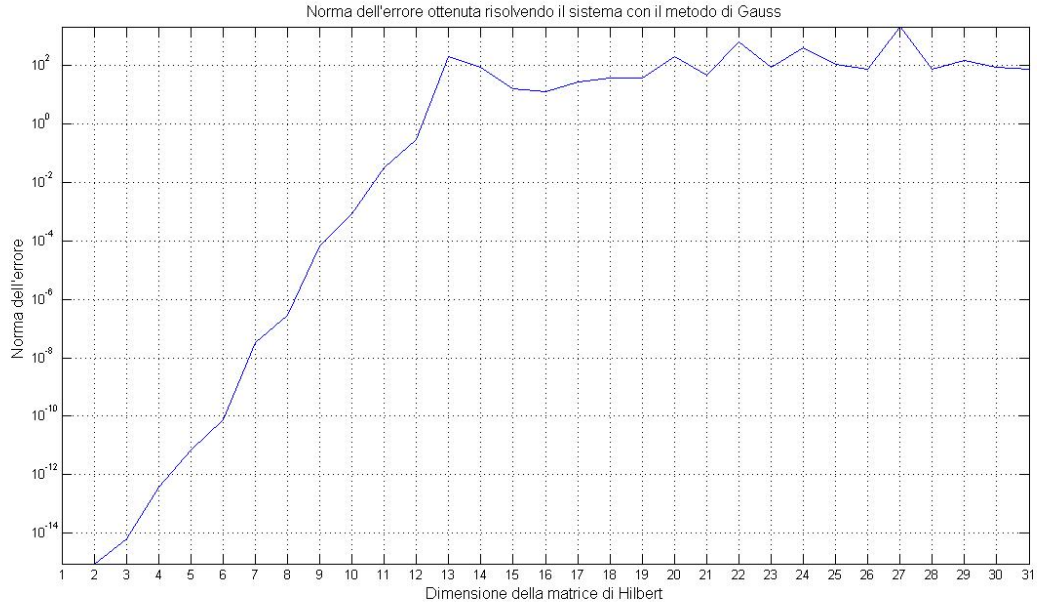


Figura 2: Norma dell'errore in funzione della dimensione della matrice di Hilbert

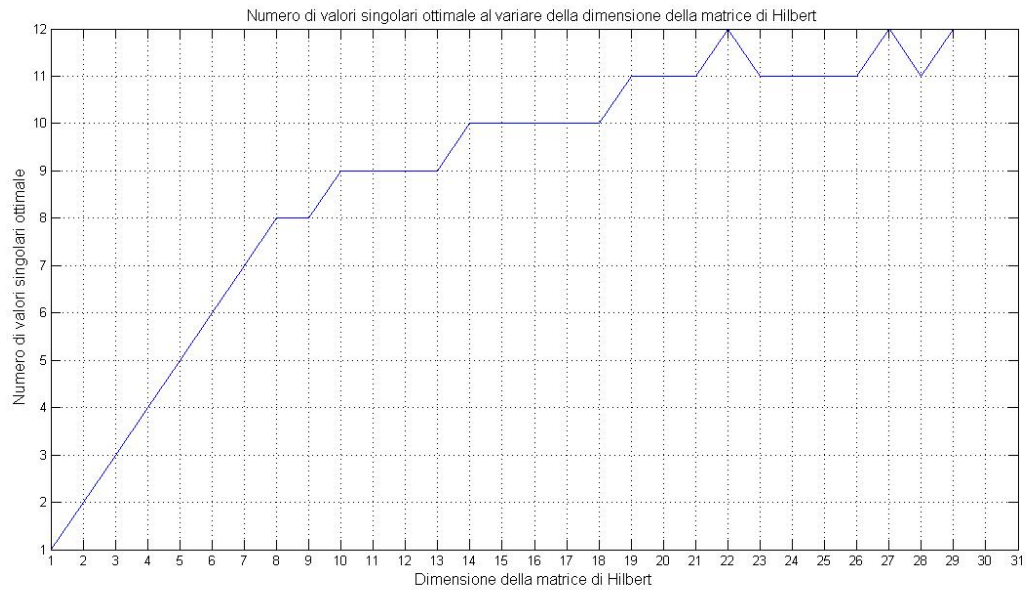


Figura 3: Troncamento ottimale al variare della dimensione della matrice di Hilbert

Nella Tabella 1 è riportato il confronto tra la norma dell'errore ottenuta attraverso il metodo di Gauss e quella ottenuta per la TSVD con il miglior troncamento. Dalla Tabella 1 si evince che per dimensioni della matrice superiori a 9 la TSVD troncata ci consente di ottenere dei risultati migliori rispetto al metodo di Gauss.

Dimensione della matrice	Norma dell'errore della soluzione calcolata con Gauss	Numero di valori singolari necessari per il miglior troncamento	Norma dell'errore della soluzione calcolata con la TSVD per il miglior troncamento
1	0	1	0
2	8.95E-16	2	8.01E-16
3	6.48E-15	3	1.16E-14
4	3.74E-13	4	7.05E-13
5	6.74E-12	5	2.28E-11
6	7.55E-11	6	8.43E-10
7	3.25E-08	7	2.18E-08
8	2.87E-07	8	4.07E-06
9	6.45E-05	8	6.59E-06
10	0.000872	9	1.23E-05
11	0.031676	9	4.39E-06
12	0.283664	9	7.01E-06
13	205.6411	9	1.21E-05
14	84.42127	10	1.42E-05
15	15.39552	10	4.99E-06
16	12.61284	10	6.30E-06
17	26.54445	10	9.13E-06
18	37.80304	10	1.25E-05
19	38.16279	11	2.97E-06
20	201.1761	11	7.35E-06
21	46.89893	11	9.29E-06
22	650.1765	12	7.25E-06
23	88.84155	11	1.02E-05
24	410.799	11	1.06E-05
25	110.7278	11	1.29E-05
26	73.00816	11	1.56E-05
27	2107.571	12	1.86E-05
28	75.86987	11	2.19E-05
29	149.8787	12	1.63E-05
30	85.38205	12	7.40E-06
31	76.982	12	2.31E-05

Tabella 1: Confronto tra la norma dell'errore ottenuta attraverso il metodo di Gauss e quella ottenuta per la TSVD con il miglior troncamento

Analisi di un caso reale

Modellazione dello stato stazionario di una colonna di distillazione per la separazione di una miscela etanolo acqua

La distillazione è un'operazione unitaria che, attraverso lo sfruttamento del calore, consente di separare una miscela di più componenti in base al diverso punto di ebollizione. All'interno della colonna la fase liquida e la fase vapore procedono in controcorrente scambiando materia così che il vapore, procedendo verso l'alto, si arricchisce dei composti più leggeri (più volatili), mentre il liquido, procedendo verso il basso si arricchisce dei composti più pesanti (meno volatili). In uscita dall'apparecchiatura si hanno due correnti, il distillato ed il residuo, che sono rispettivamente la frazione leggera e la frazione pesante della miscela alimentata.

Il funzionamento di una colonna di distillazione può essere schematizzato attraverso una successione di stadi di equilibrio liquido-vapore. Il distillato non è altro che una frazione del vapore in uscita dall'ultimo stadio di equilibrio (quello più in alto), mentre il residuo è una frazione del liquido in uscita dal primo stadio di equilibrio (quello più in basso). Parte del vapore in uscita dal primo stadio viene condensata e inviata in colonna come riflusso. La restante parte, che costituisce il distillato, può essere prelevata in fase vapore o in fase liquida a seconda delle esigenze, e la colonna avrà quindi un condensatore totale o parziale a seconda dei diversi casi. Nel caso in cui il condensatore sia parziale questo rappresenta l'ultimo stadio di equilibrio della colonna. Un parametro molto importante per la progettazione e l'esercizio di una colonna di distillazione è il rapporto di riflusso ossia il rapporto tra il riflusso ed il distillato. Analogamente, parte del liquido in uscita dal primo stadio viene vaporizzata e rimandata in colonna. Mentre la condensazione totale o parziale del vapore in testa (parte alta della colonna) è una scelta progettuale e dipende dalle esigenze del caso, la presenza di una vaporizzazione totale del liquido nel fondo (parte bassa della colonna) è meno comune, quindi il ribollitore rappresenta quasi sempre il primo stadio di equilibrio di una colonna di distillazione.

Parametri molto importanti durante la progettazione della colonna sono:

- Condizioni dell'alimentazione
- Pressione di esercizio
- Purezza desiderata
- Numero di stadi
- Rapporto di riflusso

- Posizione dell'alimentazione lungo la colonna

I composti da separare e la pressione di esercizio della colonna sono strettamente connessi alla difficoltà nell'effettuare la separazione della miscela poiché condizionano l'equilibrio. In generale si può dire che ad una separazione più difficile corrisponde un più elevato numero di stadi di equilibrio a parità di purezza desiderata. Una volta fissate le condizioni termodinamiche e la separazione desiderata, il numero di stadi di equilibrio necessari per effettuare tale separazione è strettamente connesso al rapporto di riflusso utilizzato e viceversa. In sede di progetto, a parità di difficoltà di separazione, minore è il rapporto di riflusso maggiore sarà il numero di stadi di equilibrio necessari. In fase di esercizio invece, poiché si lavora con un'apparecchiatura esistente, quindi con un numero di stadi di equilibrio fissati, si è soliti utilizzare il rapporto di riflusso per regolare la composizione entro certi limiti.

Modello dinamico nonlineare

Uno dei possibili modelli non lineari in variabili di stato utilizzati per simulare il comportamento dinamico della nostra colonna di distillazione binaria con N stadi, è il seguente:

Ribollitore ($i = 1$)

$$\dot{x}_1 = \frac{L_S x_2 - V_S y_1 - B x_1}{M_1}$$

Sezione di esaurimento ($i = 2, \dots, N_F - 1$)

$$\dot{x}_i = \frac{L_S(x_{i+1} - x_i) - V_S(y_i - y_{i-1})}{M_i}$$

Stadio di alimentazione ($i = N_F$)

$$\dot{x}_{N_F} = \frac{L_E x_{N_F+1} - L_S x_{N_F} - V_E y_{N_F} + V_S y_{N_F-1} + Fz}{M_i}$$

Sezione di arricchimento ($i = N_F + 1, \dots, N - 1$)

$$\dot{x}_i = \frac{L_E(x_{i+1} - x_i) - V_E(y_i - y_{i-1})}{M_i}$$

Condensatore ($i = N$)

$$\dot{x}_{N_T} = \frac{V_E y_{N_T-1} - L_E x_{N_T} - D x_{N_T}}{M_{N_T}}$$

dove $i = 1, \dots, N$ indica l' i -esimo stadio, x_i, y_i e M_i sono rispettivamente la frazione molare di etanolo in fase liquida, la frazione molare di etanolo in fase vapore e l'holdup nello stadio i -esimo, L_S e V_S sono le portate molari di liquido e vapore nella sezione di esaurimento (L_E e V_E quelle nella

sezione di arricchimento), F e z sono la portata e la frazione molare di etanolo nell'alimentazione (supposta in fase di liquido saturo), D e B sono le portate molari di distillato e residuo.

Alcune delle grandezze sopra descritte sono legate tra loro da equazioni algebriche lineari:

$$\begin{aligned}L_E &= R \\L_S &= L_E + F \\V_S &= V_E = V \\D &= V_E - L_E \\F &= D + B\end{aligned}$$

dove R è la portata molare di riflusso e V è la portata molare di vapore prodotta dal ribollitore.

Il modello è nonlineare in quanto è tale la relazione tra y_i e x_i per ogni $i = 1, \dots, N$, dato che

$$y_i = \frac{x_i \exp\left(A_{ET} - \frac{B_{ET}}{C_{ET} + T_i}\right)}{P_i}$$

e T_i [K] dipende da x_i secondo l'equazione implicita

$$x_i \exp\left(A_{ET} - \frac{B_{ET}}{C_{ET} + T_i}\right) + (1 - x_i) \exp\left(A_W - \frac{B_W}{C_W + T_i}\right) - P_i = 0$$

dove P_i [mmHg] è la pressione nello stadio i -esimo mentre A_{ET} , B_{ET} e C_{ET} sono le costanti di Antoine per l'etanolo e A_W , B_W e C_W sono quelle per l'acqua.

Condizioni operative

I parametri e le condizioni operative per la colonna in questione sono riportate in Tabella 2 e Tabella 3.

N (stadi)	37
N_F (stadio di alimentazione)	8
P [mmHg] (pressione)	760
R [mol/s] (portata di riflusso)	0.77
V [mol/s] (portata di vapore)	1.45
F [mol/s] (portata di alimentazione)	1.78
z (frazione molare di etanolo nell'alimentazione)	0.14
M [mol] (holdups lungo la colonna)	$[20 \ 2 \ \dots \ 2 \ 20] \cdot 10^3$

Tabella 2: Parametri

	<i>A</i>	<i>B</i>	<i>C</i>
Etanolo	18.9119	3803.98	-41.68
Acqua	18.3036	3816.44	-46.13

Tabella 3: Costanti di Antoine

Tali parametri determinano univocamente le condizioni operative della colonna.

Modello statico lineare

Allo scopo di ottenere un modello statico lineare, è stata inizialmente effettuata una linearizzazione del modello, assumendo $y_i = \alpha_i x_i$ per ogni $i = 1, \dots, N$, dove

$$\alpha_i = \frac{dy_i}{dx_i}(x_{i,SS})$$

e $x_{i,SS}$ è la soluzione calcolata allo stazionario (cioè quando $\dot{x}_i = 0$).

In questo modo, ponendo $y_i = \alpha_i x_i$ e $\dot{x}_i = 0$ per ogni $i = 1, \dots, N$, si ottiene un sistema lineare statico nella forma $Ax = b$, dove $x = [x_1 \dots x_N]^T$, $b = [0 \dots 0 -Fz \ 0 \dots 0]^T$ (con $-Fz$ elemento N_F -esimo del vettore b), mentre A è la matrice tridiagonale riportata in Figura 8.

Condizionamento del problema e soluzione

Dato il sistema lineare $Ax = b$, è possibile calcolare la soluzione del problema come $x = A \setminus b$ (risoluzione del sistema lineare attraverso il metodo di Gauss) se il condizionamento di A non è molto alto. Infatti in questo caso, a meno di grossi errori nella matrice A o nel vettore b , la soluzione del problema inverso sarà pressoché esatta o quantomeno non affetta da grossi errori.

Nel nostro caso si ha

$$\text{cond}(A) = \|A\| \|A^{-1}\| = 2.018 \cdot 10^3$$

Il numero di condizionamento della matrice A è basso, e perciò possiamo calcolare la soluzione del problema come $x = A \setminus b$, ottenendo la soluzione illustrata in Figura 4.

Come illustrato in Figura 4, le due soluzioni sono quasi identiche (l'errore massimo è minore di 10^{-6}).

Affinché vi siano errori rilevanti nel profilo di composizione di etanolo lungo la colonna, è necessario che ci sia una perturbazione nella matrice A o nel vettore dei termini noti b . Questo caso è tutt'altro che irrealistico in quanto le varie grandezze sono solitamente affette da errori di misura più o meno rilevanti.

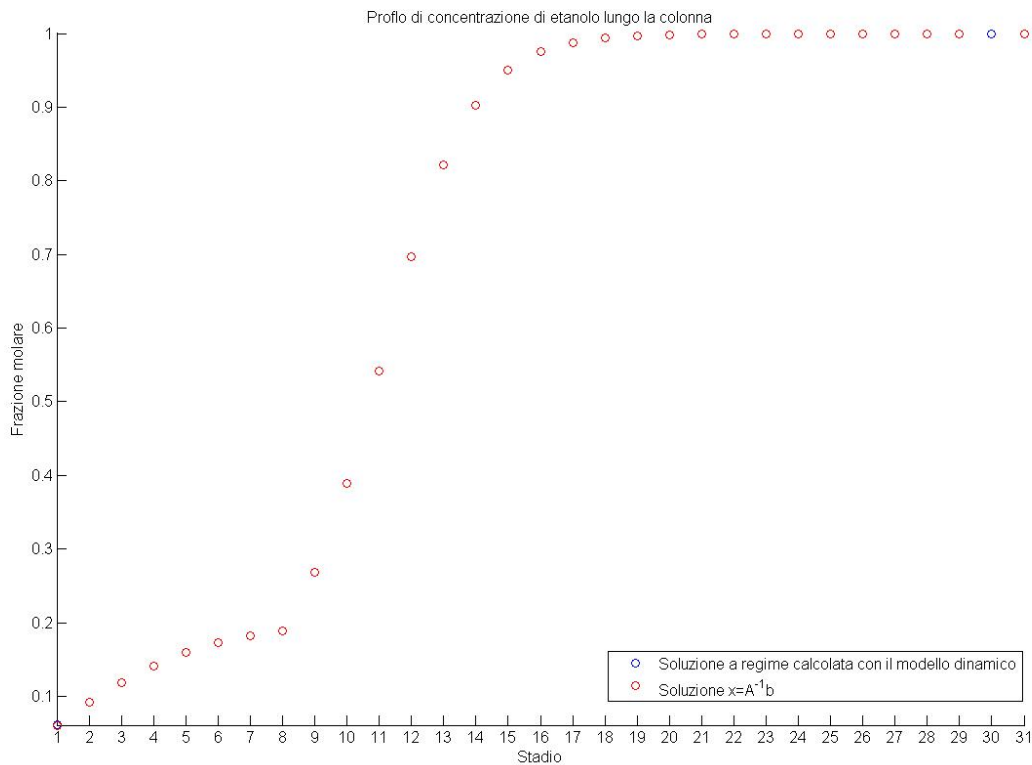


Figura 4: Profilo di concentrazione di etanolo lungo la colonna

Se consideriamo il vettore dei termini noti b , possiamo notare come la grandezza $-Fz$ (unico elemento del vettore b non nullo) sia data dal prodotto (cambiato di segno) di una misura di portata per una di composizione. Queste due misure sono sicuramente affette da errore in quanto i relativi sensori di misura presentano un certo grado di inaccuratezza e danno quindi luogo ad errori di misura. Nel caso in questione, le accuratze dei sensori sono dell'ordine dell'1% (cioè l'errore relativo massimo è dell'1%), per cui possiamo assumere

$$\delta F \approx 10^{-2}$$

$$\delta z \approx 10^{-3}$$

dove δF e δz sono gli errori assoluti di portata e composizione. Possiamo quindi ricavare l'errore e_{Fz} sulla quantità Fz . Si ha

$$Fz + e_{Fz} = (F + \delta F)(z + \delta z) = Fz + F\delta z + z\delta F + \delta F\delta z \approx Fz + F\delta z$$

e quindi

$$e_{Fz} = F\delta z \approx 10^{-2}$$

Abbiamo inizialmente deciso di perturbare non solo il termine $-Fz$ ma tutto il vettore b con un rumore gaussiano δb avente media nulla e deviazione standard e_{Fz} , ottenendo quindi una soluzione perturbata pari a $x + \delta x$. Il confronto di tale soluzione con quella non perturbata è mostrato in Figura 5

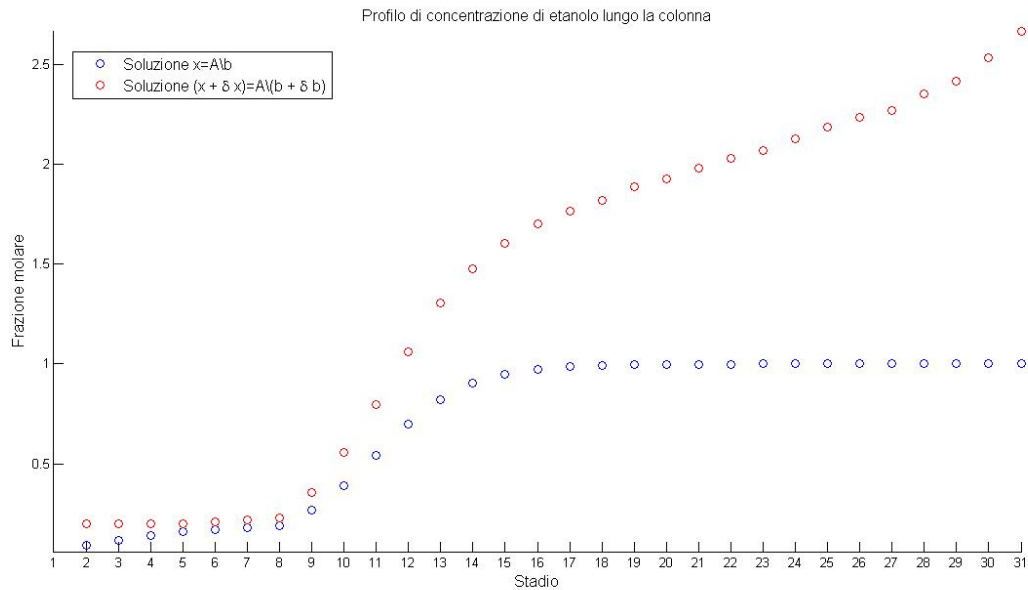


Figura 5: Confronto tra la soluzione perturbata e quella non perturbata

La soluzione perturbata è completamente inaccettabile, anche perché le concentrazioni possono assumere soltanto valori compresi tra 0 ed 1. Si deve infatti avere

$$x_i \in [0,1] \quad \text{per } i = 1, \dots, N$$

Una possibile soluzione potrebbe essere data dall'utilizzo della decomposizione a valori singolari troncata (TSVD): con questo metodo, la soluzione x del sistema è data dalla formula illustrata precedentemente.

In questo caso, è stato verificato che il miglior troncamento (quello che minimizza la norma dell'errore) si ottiene utilizzando un valore di k pari a 30. La soluzione ottenuta è illustrata in Figura 6, e risulta anch'essa inaccettabile (le concentrazioni non possono assumere un valore negativo).

Come secondo test è stato deciso di perturbare il solo termine $-Fz$ all'interno del vettore b , con il medesimo rumore gaussiano a media nulla e deviazione standard e_{Fz} . In Figura 7 sono illustrate la soluzione calcolata attraverso il metodo di Gauss e quella calcolata con la TSVD (che anche in

questo caso produce il miglior risultato con $k = 30$). È evidente come in questo caso la TSVD produca un risultato inaccettabile mentre la soluzione ricavata con Gauss è sicuramente migliore.

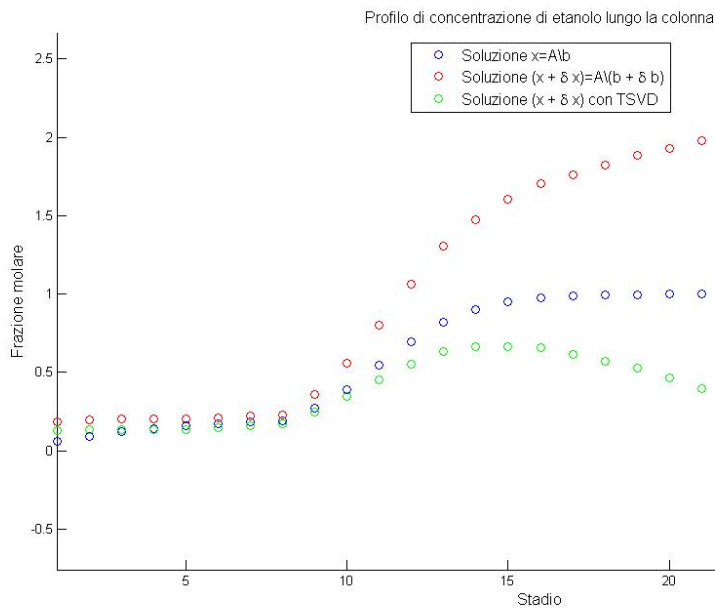


Figura 6: Confronto tra le due soluzioni perturbate (utilizzando Gauss e TSVD) e quella non perturbata nel caso con il vettore b completamente perturbato.

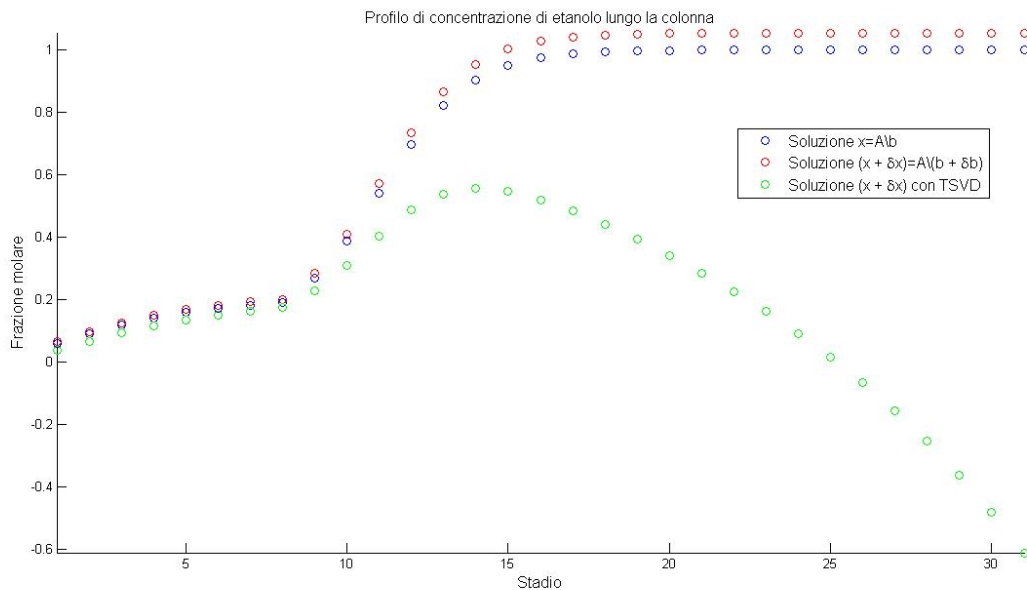


Figura 7: Confronto tra le due soluzioni perturbate (utilizzando Gauss e TSVD) e quella non perturbata nel caso con il vettore b non completamente perturbato.

