

UNIVERSITÀ DEGLI STUDI DI CAGLIARI
SCUOLA DI DOTTORATO IN INGEGNERIA INDUSTRIALE

Applicazione dei metodi del gradiente, del gradiente coniugato, GMRES e TSVD alla modellazione di una colonna di distillazione

Mariarosa Brundu
Anuta Chindris*
Andrea Frau*
Francesco Desogus**
Jesus Rodriguez Ruiz**

***Modello dinamico e applicazione della TSVD (precedentemente consegnata)**

****Applicazione di metodi iterativi (gradiente, gradiente coniugato e GMRES)**

**Versione del
22 Ottobre 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.

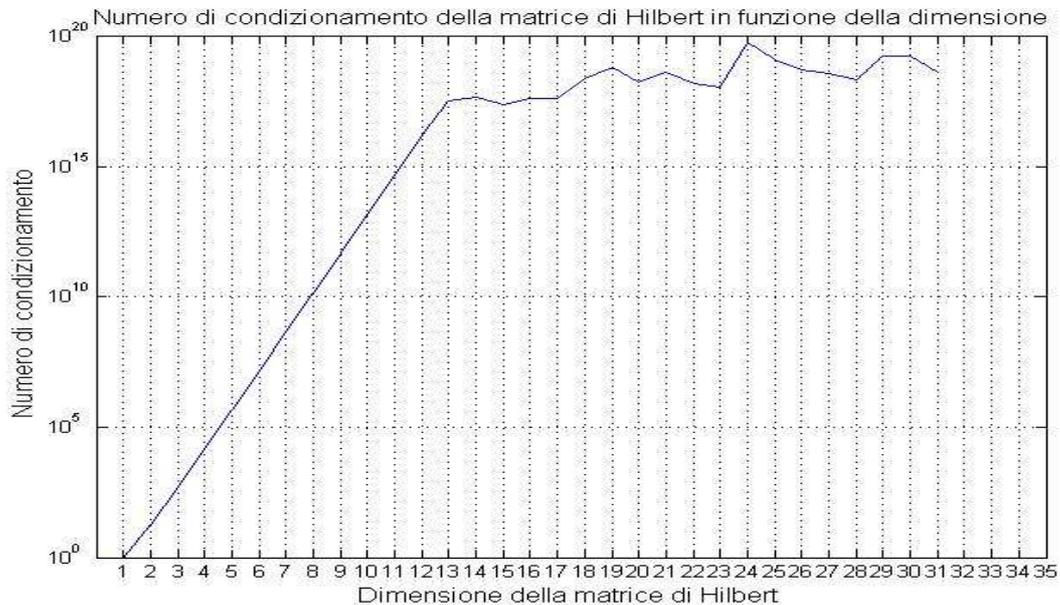


Figura 1: Condizionamento della matrice di Hilbert in funzione della dimensione

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 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.

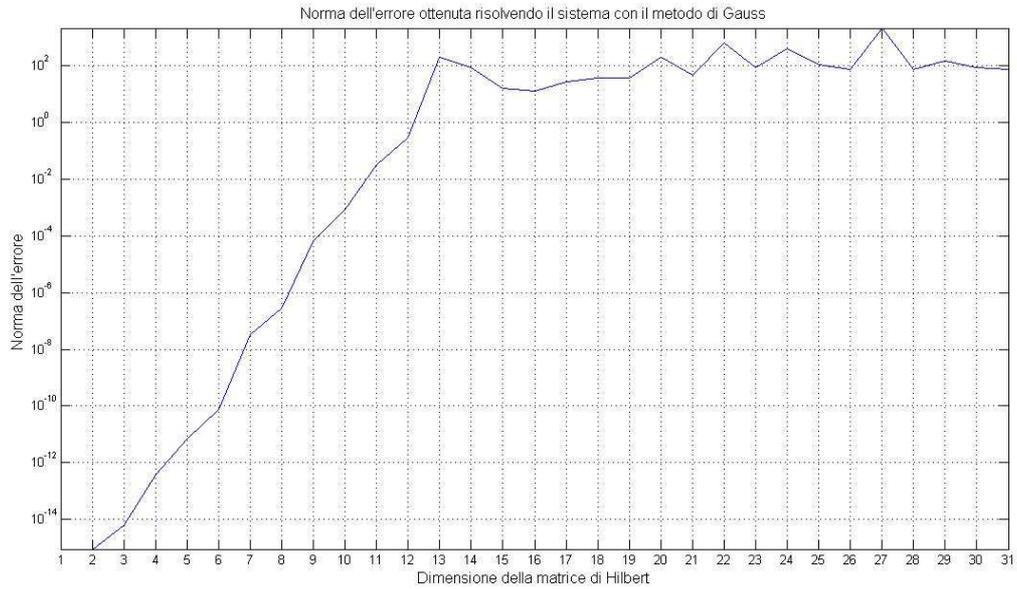


Figura 2: Norma dell'errore in funzione della dimensione della matrice di Hilbert

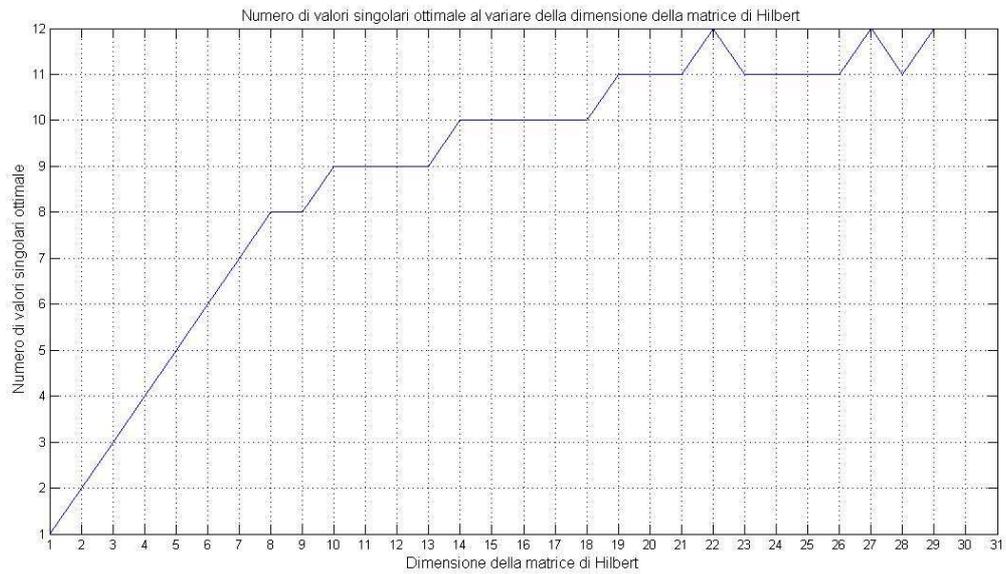


Figura 3: Troncamento ottimale al variare della dimensione della matrice di Hilbert

Metodo del gradiente

Il metodo del gradiente è valido per sistemi di equazioni del tipo

$$Ax = b$$

Dove la matrice A è simmetrica e definita positiva:

- $A^T = A$ (simmetrica)
- $x^T Ax > 0$ qualunque sia il vettore $x \neq 0$ (positiva)

sotto queste condizioni la soluzione del problema presentato, è equivalente alla minimizzazione della forma quadratica $\phi(x) = \frac{1}{2} x^T Ax - x^T b$ e la soluzione coincide con il punto in cui il gradiente diventa nullo.

Fissando una direzione di decrescita d si studia il comportamento della forma quadratica con passi di lunghezza **Errore. Non si possono creare oggetti dalla modifica di codici di campo.** in modo che ogni passo sia dato dalla:

Errore. Non si possono creare oggetti dalla modifica di codici di campo.

Il metodo consiste nella scelta di un primo punto a caso nello spazio multidimensionale e la valutazione della funzione e del suo gradiente nel punto stesso. Quest'ultimo, in particolare, indicherà la direzione più ripida secondo la quale la funzione tende ad un minimo. Il punto successivo viene dunque scelto nella direzione indicata dal gradiente. Se la funzione al secondo punto ha un valore inferiore al valore corrispondente al primo la discesa continua fino a trovare il punto che minimizza ϕ .

Quindi, data la matrice simmetrica A, la forma quadratica per x_{k+1} sarà:

$$\phi(x_{k+1}) = \phi(x_k) + \frac{1}{2} d_k^T A d_k \cdot \alpha^2 - d_k^T (Ax_k - b) \cdot \alpha$$

Uguagliando a zero la derivata:

$$\frac{d}{dt} \phi(x_{k+1}) = \frac{1}{2} d_k^T A d_k \cdot \alpha - d_k^T (Ax_k - b) = 0$$

Si troverà un minimo per questa funzione per:

$$\alpha_k = \frac{d_k^T (Ax_k - b)}{d_k^T A d_k}$$

Se prendiamo d_k pari al gradiente negativo di ϕ in x_k , ossia scegliendo la direzione opposta alla direzione del gradiente nel punto x_k , ciò equivarrà a prendere il residuo al passo k-simo

$$r_k = b - Ax_k,$$

si ottiene la direzione di discesa più ripida, il metodo viene infatti spesso chiamato “steepest descent”.

Graficamente questo metodo si può rappresentare come l'avanzamento in una iperelissoide il cui centro geometrico costituisce la soluzione x del sistema lineare.

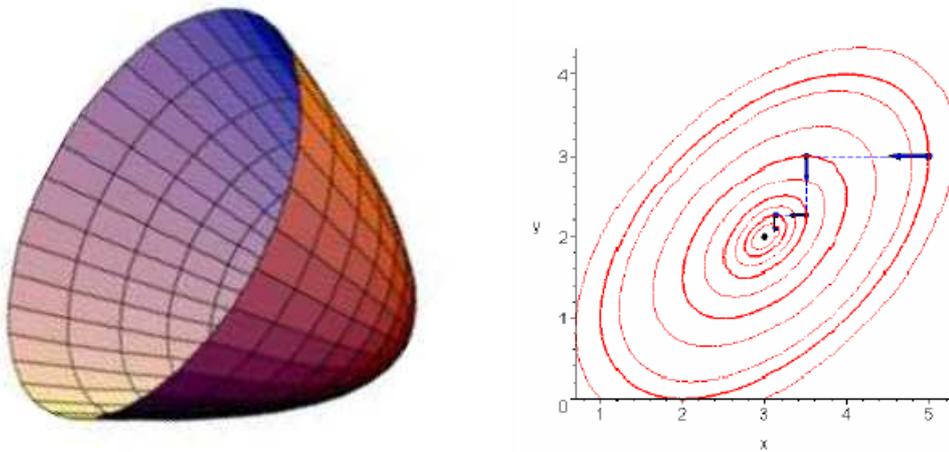


Figura 4: Rappresentazione geometrica del metodo del gradiente

Metodo del gradiente coniugato

Il metodo precedente, è in generale molto lento se le curve della forma quadratica sono molto vicine. Attraverso l'utilizzo di condizioni di ortogonalità si può modificare il metodo in modo da ottenere una convergenza veloce. In questo caso si parla del metodo del gradiente coniugato.

L'idea principale di questo metodo consiste nel costruire una base di vettori ortogonali e utilizzarla per realizzare una ricerca della soluzione in maniera più efficiente. Questo metodo fornisce una direzione di discesa p_k basata sull'ortogonalizzazione del residuo r_k rispetto alle direzioni di discesa precedenti p_i , $i = 1, \dots, k - 1$.

Le direzioni ottenute soddisfano $\forall i \neq k \langle Ap_k, p^f \rangle$ e vengono chiamate direzioni coniugate o A-ortogonali.

Considerando quindi iterazioni del tipo:

$$x_{k+1} = x_k + \alpha_k p_k$$

Si impone ora che la direzione di discesa al passo k-simo sia presa nel piano generato dai due vettori p_k e r_{k+1} , che sappiamo essere tra loro ortogonali. Sia dunque:

$$p_{k+1} = r_{k+1} + \beta_{k+1} p_k$$

Per la condizione di ortogonalità tra la direzione di discesa p_k ed il suo residuo relativo al punto x_{k+1} si ha che:

$$p_k^T r_{k+1} = 0$$

Fissando la direzione iniziale $p_0 = r_0$, si osserva che le due direzioni p_k e p_{k+1} sono ortogonali nel prodotto scalare, cioè A-coniugate ossia

$$p_k^T A p_{k+1} = p_k^T A (r_{k+1} + \beta_{k+1} p_k) = 0$$

Il che implica che $\beta_k = -\frac{p_k^T A r_{k+1}}{p_k^T A p_k}$

In questo modo dato x_0

$$p_0 = r_0 = b - A x_0$$

per $k = 0, 1, \dots$

$$\alpha = \frac{p_k^T r_k}{p_k^T A p_k}$$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$r_{k+1} = r_k - \alpha_k A p_k$$

$$\beta_k = -\frac{p_k^T A r_{k+1}}{p_k^T A p_k}$$

$$p_{k+1} = r_{k+1} + \beta_{k+1} p_k$$

Si dimostra che la direzione p_{k+1} è A-coniugata non solo a p_k ma a tutte le precedenti direzioni di discesa e che il residuo r_k è ortogonale a tutti i precedenti residui.

$$p_i^T A p_{k+1} = 0 \quad r_i^T r_{k+1} = 0 \quad i = 0, \dots, k-1$$

Da quest'ultima considerazione possibile evincere che il metodo del gradiente coniugato converge in un numero finito di passi non superiore ad n .

GMRES

Il metodo GMRES (Generalized Minimum Residual), sviluppato nel 1986 da Yousef Saad e Martin H. Schultz rientra nella classe dei metodi iterativi per la soluzione di sistemi lineari in sottospazi di Krylov. Il metodo approssima la soluzione con un vettore dello spazio di Krylov che minimizza il residuo, ricavato tramite il metodo di Arnoldi.

Dato un sistema lineare di equazioni:

$$Ax = b$$

Dove A è una matrice invertibile di dimensioni $m \times n$ e sia b normalizzato rispetto alla sua norma il sottospazio di Krylov per il problema è definito come:

$$K_n = \text{span}\{b, Ab, A^2b, \dots, A^{n-1}b\}$$

Allora il metodo approssima la soluzione del sistema con un vettore $x_n \in K_n$ che minimizzi la norma del residuo $Ax_n - b$.

Essendo i vettori $b, Ab, A^2b, \dots, A^{n-1}b$ quasi linearmente dipendenti, per trovare una base di vettori ortonormale, al posto di essi viene usata l'iterazione di Arnoldi.

In questo modo il vettore $x_n \in K_n$ può essere scritto come

$$x_n = Q_n y_n$$

dove $y_n \in \mathbf{R}^n$, e Q_n è la matrice di dimensioni $m \times n$ formata dai vettori q_1, \dots, q_n ricavati attraverso Arnoldi.

Attraverso il processo di Arnoldi viene inoltre prodotta una matrice di Hessenberg \tilde{H}_n di dimensioni $(n+1) \times n$ tale che

$$AQ_n = Q_{n+1} \tilde{H}_n$$

Essendo Q_n ortogonale si verifica che:

$$\|Ax_n - b\| = \|\tilde{H}_n y_n - \beta e_1\|$$

dove

$$e_1 = (1, 0, 0, \dots, 0)$$

è il primo vettore della base standard di \mathbf{R}^{n+1} e $\beta = \|b - Ax_0\|$.

Chiamato x_0 il vettore di primo tentativo, la soluzione del sistema x_n potrà essere trovata attraverso una minimizzazione del residuo così definito:

$$r_n = \tilde{H}_n y_n - \beta e_1$$

I passi fondamentali per ogni iterazione dell'algoritmo sono quindi:

Applicazione del metodo di Arnoldi;

Trovare y_n che minimizza il residuo $\|r_n\|$;

calcolare $x_n = Q_n y_n$;

ripetere se il residuo così ottenuto non è sufficientemente basso.

L'iterazione n -sima minimizza il residuo nel sottospazio di Krylov K_n ma poiché ogni sottospazio è contenuto in quello successivo il residuo diminuisce in maniera monotona e dopo m iterazioni, dove m è la dimensione della matrice A lo spazio di Krylov K_m corrisponde ad R^m perciò il metodo è arrivato alla soluzione. Per sistemi molto grandi non è tuttavia necessario arrivare ad m iterazioni e già un numero inferiore potrebbe essere considerato una buona approssimazione. Ciò non è sempre verificato, secondo il teorema di Greenbaum, Pták e Strakoš, infatti, è possibile trovare una matrice per la quale il residuo resti costante $m - 1$ iterazioni e poi cada a zero nell'ultima.

GMRES viene talvolta combinato con un preconditionatore che ne accelera la convergenza, inoltre poiché il numero di calcoli aumenta con le iterazioni, per velocizzare il calcolo nel caso di sistemi di grosse dimensioni il metodo viene fatto ripartire dopo un numero di iterazioni k usando il vettore x_k come primo tentativo.

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 1 e Tabella 2.

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 1: Parametri

	A	B	C
Etanolo	18.9119	3803.98	-41.68
Acqua	18.3036	3816.44	-46.13

Tabella 2: 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 11.

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 5.

Come illustrato in Figura 5, 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.

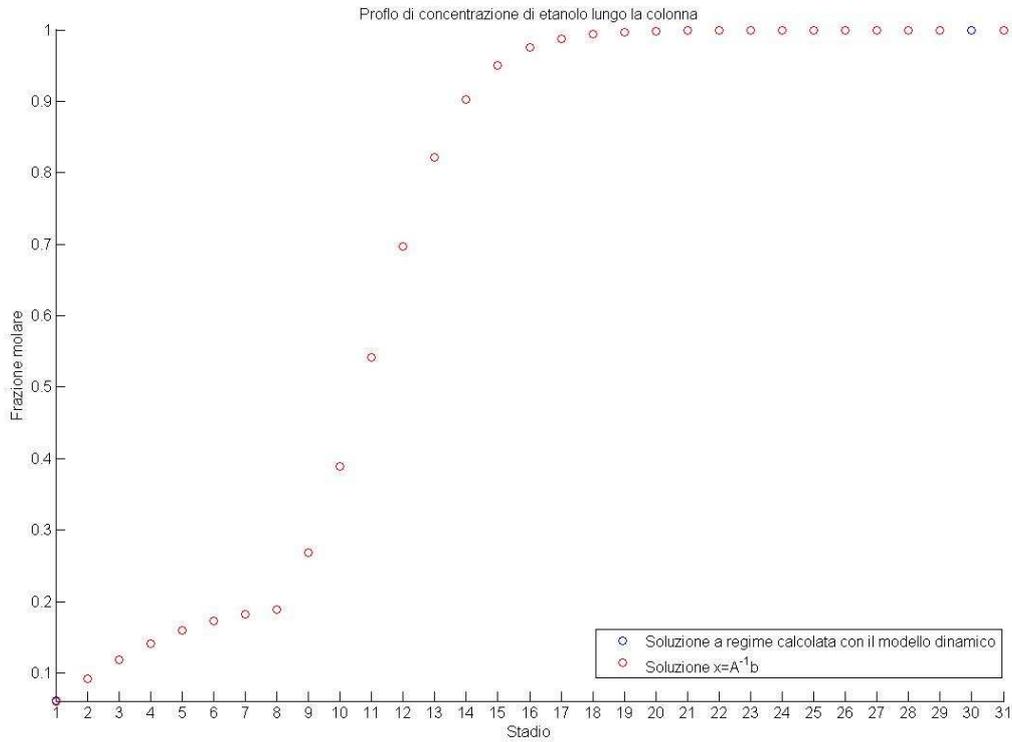


Figura 5: Profilo di concentrazione di etanolo lungo la colonna

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 accuratèzze 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 6

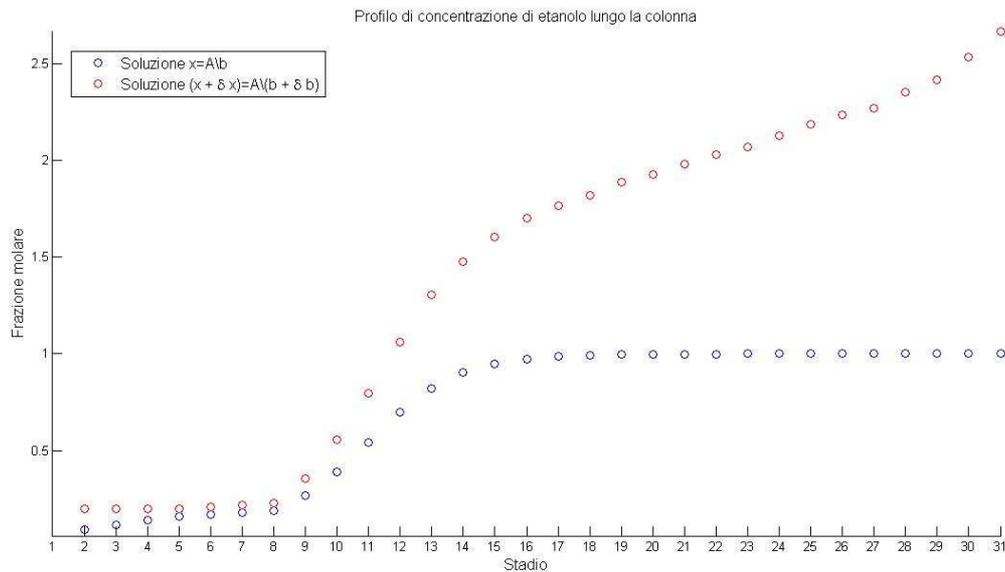


Figura 6: Confronto tra la soluzione perturbata e quella non perturbata

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 7, 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 8 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.

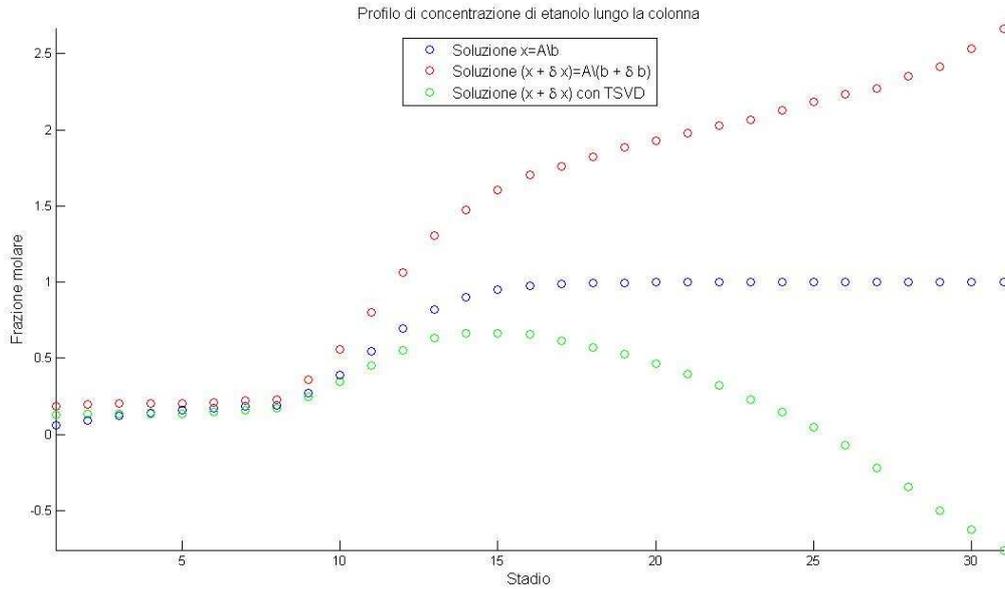


Figura 7: Confronto tra le due soluzioni perturbate (utilizzando Gauss e TSVD) e quella non perturbata nel caso con il vettore b completamente perturbato.

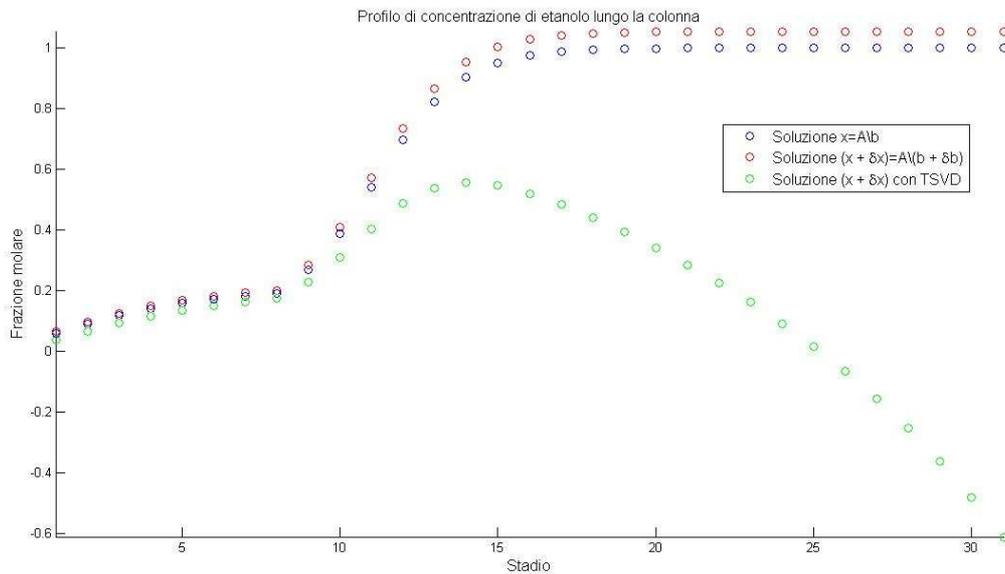


Figura 8: Confronto tra le due soluzioni perturbate (utilizzando Gauss e TSVD) e quella non perturbata nel caso con il vettore b non completamente perturbato.

Il sistema è stato inoltre risolto attraverso l'utilizzo dei metodi iterativi presentati.

Come precedentemente descritto il metodo del gradiente converge solo per matrici simmetriche definite positive. Nella rappresentazione grafica della matrice A , presente in Figura 11, è possibile osservare che questa non è simmetrica, perciò ci si aspetta che in questo caso il metodo non converga. Sorprendentemente, la sua applicazione al caso in esame risulta convergente per diversi valori del vettore di primo tentativo. In Figura 9 è riportato il confronto tra la soluzione ottenuta a partire da un vettore di primo tentativo costituito da tutti zeri e la soluzione ottenuta col modello dinamico. Si è inoltre notato che al variare del vettore di primo tentativo il metodo va sempre a convergenza, mentre il numero di iterazioni necessario varia. In Tabella 3 è mostrato il numero di iterazioni corrispondente al vettore di primo tentativo utilizzato.

A differenza di ciò che è stato osservato per il metodo del gradiente, nel caso il metodo del gradiente coniugato i calcoli non sono andati a convergenza in accordo con la teoria.

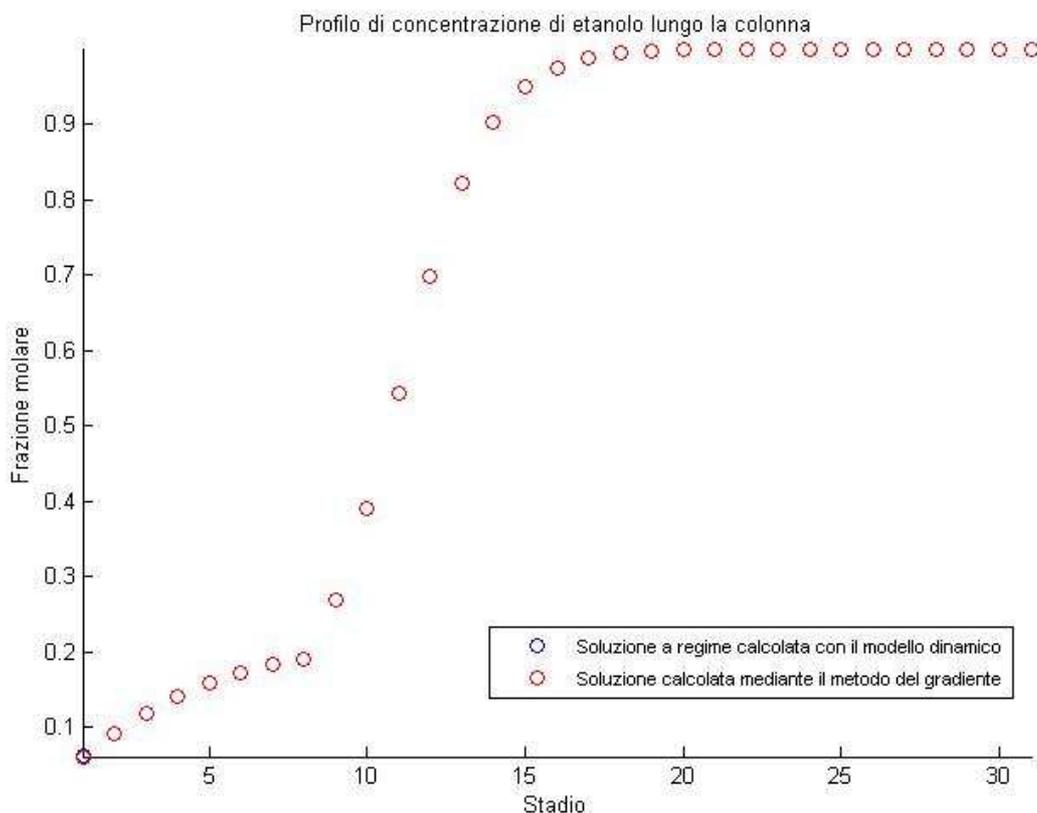


Figura 9: confronto tra la soluzione calcolata tramite il modello dinamico e quella calcolata mediante il metodo del gradiente.

Tabella 3: variazione del numero di iterazioni e del residuo con il vettore di primo tentativo per il metodo del gradiente.

Vettore di primo tentativo	Iterazioni	Residuo
(0,0,0, ..., 0)	8855	$9.98 \cdot 10^{-7}$
(1.5,1.5,1.5, ..., 1.5)	8723	$9.98 \cdot 10^{-7}$
(1,1,1, ..., 1)	7827	$9.98 \cdot 10^{-7}$
(0.15,0.15,0.15, ..., 0.15)	8540	$9.98 \cdot 10^{-7}$

Un metodo ampiamente utilizzato per la soluzione di sistemi con matrici non simmetriche è il metodo GMRES precedentemente descritto.

L'applicazione del metodo al nostro sistema ha portato a migliorie nella convergenza rispetto alla risoluzione attraverso il metodo del gradiente. In particolare è diminuito il tempo di calcolo e le iterazioni necessarie.

In un primo momento è stato utilizzato il metodo semplice, si è poi passati al metodo con restart alla 10 iterazione ed infine è stato introdotto un preconditionatore M al fine di velocizzare la convergenza. Dove M è una matrice diagonale i cui termini sono costituiti dalla diagonale principale della matrice A . Com'è riportato in figura la soluzione ottenuta attraverso il modello dinamico e quella ottenuta attraverso il metodo GMRES sono quasi coincidenti.

L'aggiunta del preconditionatore si è dimostrata efficace infatti il metodo GMRES con restart richiede un numero di iterazioni pari a mentre in questo caso sono sufficienti. Come è riportato in figura

Tabella 4: del numero di iterazioni e del residuo con per i metodi GMRES con restart e aggiunta di un preconditionatore.

Metodo	Iterazioni	Residuo
GMRES con restart	56	$9.9 \cdot 10^{-7}$
GMRES con restart e preconditionamento	17	$9.3 \cdot 10^{-7}$

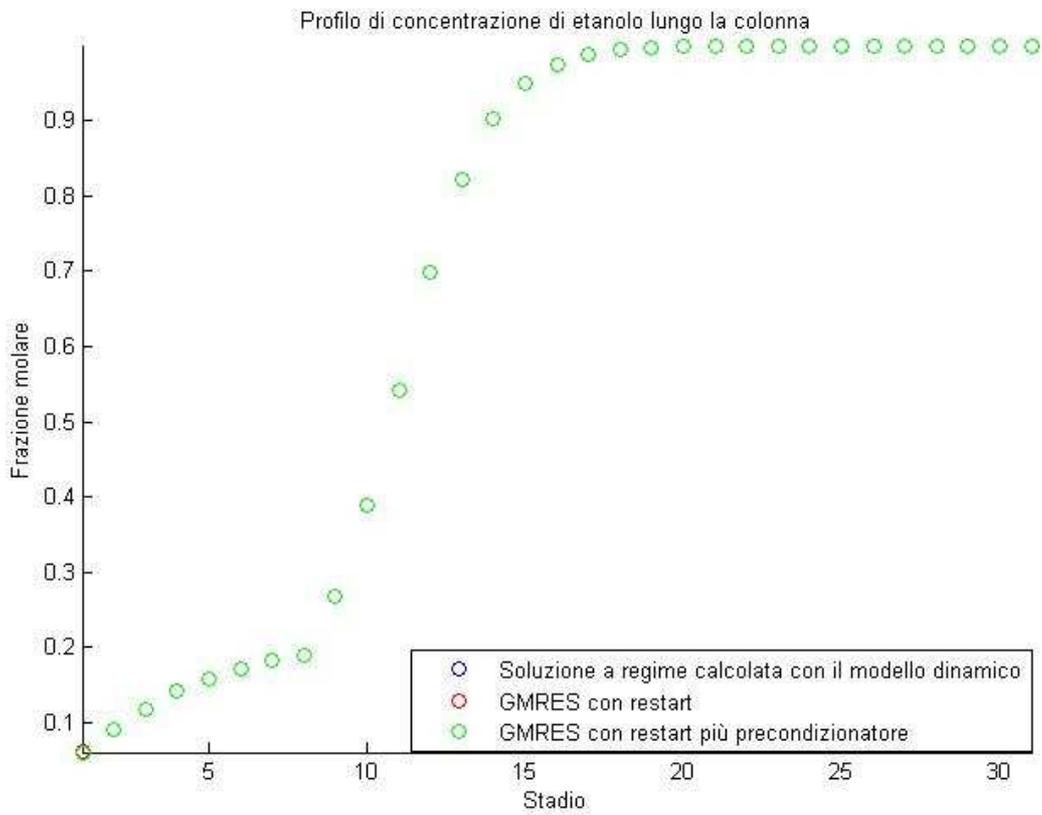


Figura 10: confronto tra la soluzione calcolata tramite il modello dinamico e quella calcolata mediante il metodo GMRES con restart e l'aggiunta del preconditionatore.

