Seminario di Calcolo Scientifico

Un metodo per il precondizionamento di matrici Toeplitz per le iterazioni ai minimi quadrati

Mario Correddu

Dipartimento di Matematica
Università di Pisa

December 15, 2020

Introduzione

Definizione

Una matrice $m \times n$ T è detta Toeplitz se

$$t_{j,k} = t_{j-k}$$

$$j = 1 \dots m,\ i = 1 \dots n$$

ovvero se T è costante lungo le diagonali.

$$\begin{bmatrix} t_0 & t_1 & t_2 & \dots & t_n \\ t_{-1} & t_0 & t_1 & \ddots & \vdots \\ t_{-2} & t_{-1} & t_0 & \ddots & \vdots \\ t_{-3} & t_{-2} & t_{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{-m} & \dots & \dots & \dots & t_{n-m} \end{bmatrix}$$

Introduzione

Supponiamo che data T matrice Toeplitz, essa sia un elemento di una successione T_n ,dove T_n è il minore principale di ordine n di una matrice Toeplitz infinita T_∞,le cui entrate t_{i-j} di siano tali che

$$f(x) = \sum_{k=-\infty}^{\infty} t_k e^{-ikx}$$

per una certa f$\in C_{2\pi}$, lo spazio delle funzioni continue 2π-periodiche dotato della norma infinito, ovvero richiediamo che t_k siano i coefficienti di Fourier di f.

Introduzione

Ci occupiamo di risolvere un problema ai minimi quadrati ovvero:

$$\min_x \| Tx - b \|_2$$

dove T è una matrice Toeplitz $m \times n$ ottenuta a partire da una certa f.

Questo tipo di problemi può essere risolto tramite l'algoritmo del gradiente coniugato applicato al sistema di equazioni normali $T^*(Tx - b) = 0$ senza dover calcolare eplicitamente T^*T.
Per migliorare la velocità di convergenza si è soliti precondizionare il problema con una matrice invertibile C, ovvero risolvere il problema:

$$\min_y \left\| b - TC^{-1}y \right\|_2$$

E poi porre $x = C^{-1}y$.

Introduzione

Algoritmo del gradiente coniugato precondizionato per problemi ai minimi quadrati

Dato x_0 una prima approssimazione di x:

$r_0 = b - Tx_0$

$p_0 = s_0 = (C^{-1})^* T^* r_0$

$\gamma_0 = ||s_0||_2^2$

for $k = 0, 1, 2, ...$

$q_k = TC^{-1}p_k$

$\alpha_k = \frac{\gamma_k}{||q_k||_2^2}$

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

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

$s_{k+1} = (C^{-1})^* T^* r_{k+1}$

$\nu_{k+1} = ||s_{k+1}||_2^2$

$\beta_k = \frac{\nu_{k+1}}{\nu_k}$

$p_{k+1} = s_{k+1} + \beta_k p_k$

return x

Introduzione

Teorema

Sia A una matrice $\in \mathbb{C}^{n\times n}$ Hermitiana semidefinita positiva e b $\in \mathbb{C}^n$. Sia $\{x_k\}$ la successione delle approssimazioni della soluzione calcolate dal metodo del gradiente coniugato applicato al sistema $Ax = b$. Inoltre sia

$$Sp(A) \subseteq \bigcup_{i=1}^{p} \lambda_i \cup [a, b] \cup \bigcup_{j=1}^{q} \lambda_j^{'}$$

con $\lambda_i < a < b < \lambda_j^{'}$ per i=1...p, j=1...q. Allora, se $p = q$, fissato ϵ si ha che se

$$k \geq \left\lceil \left[ln\left(\frac{2}{\epsilon}\right) + \sum_{i=1}^{p} ln\left(\frac{\lambda_i^{'}}{4\lambda_i}\right)\left(1 - \frac{\lambda_i}{\lambda_i^{'}}\right)^2 \right] / ln(\sigma^{-1}) \right\rceil + 2p$$

dove $\sigma = \frac{1-\sqrt{\frac{a}{b}}}{1+\sqrt{\frac{a}{b}}}$. Allora esiste $\hat{x}$ una soluzione di $Ax = b$ tale che:

$$\frac{\|\hat{x} - x_k\|}{\|\hat{x} - x_0\|} < \epsilon$$

Caso quadrato

Definizione

Una matrice $n \times n$ C è detta circolante se è Toeplitz e le diagonali hanno la proprietà $c_{n-j} = c_{-j}$ per $1 \leq j \leq n$

$$\begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_{n-2} & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \ddots & \ddots & c_{n-2} \\ c_{n-2} & c_n & c_0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ c_2 & \ddots & \ddots & \ddots & \ddots & c_{n-1} \\ c_1 & c_2 & \dots & \dots & c_{n-1} & c_0 \end{bmatrix}$$

Caso quadrato

Ogni matrice circolante può essere scritta come

$$C = \sum_{i=0}^{n} c_i \overline{C}^i$$

$$\overline{C} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ 1 & 0 & \dots & 0 \end{bmatrix}$$

per cui ogni matrice circolante è diagonalizzabile tramite la DFT:

$$C = FDF^*$$

con $F = \frac{1}{\sqrt{n}}\Omega_n^*$, dove Ω_n è la matrice di Fourier

Caso quadrato

Per costruire la matrice di precondizionamento che useremo partiamo dalla matrice di precondizionamento C definita per matrici quadrate T $n \times n$, come la matrice circolante che minimizza:

$$F(X) = \|T - X\|_F$$

Poichè stiamo lavorando con matrici Toeplitz la definizione coincide con:

$$c_k = \left\{ \begin{array}{ll} \frac{(n-k)a_k + k a_{k-n}}{n} & se \quad 0 \leq k < n \\ c_{n+k} & se \quad 0 < -k < n \end{array} \right\}$$

Caso quadrato

Lemma 1

Sia f$\in C_{2\pi}$, allora:

$$||T_n||_2 \leq 2||f||_\infty < \infty$$

inoltre se f non ha zeri, ovvero, $\min_{\theta\in[-\pi,\pi]} |f(\theta)| > 0$ allora esiste $c > 0$ t.c. per ogni n sufficientemente grande si ha

$$||T_n||_2 > c$$

Caso quadrato

Lemma 1

Sia f$\in C_{2\pi}$, allora:

$$||T_n||_2 \leq 2||f||_\infty < \infty$$

inoltre se f non ha zeri, ovvero, $\min_{\theta\in[-\pi,\pi]} |f(\theta)| > 0$ allora esiste $c > 0$ t.c. per ogni n sufficientemente grande si ha

$$||T_n||_2 > c$$

Lemma 2

Sia $f \in C_{2\pi}$ allora si ha:

$$||C||_2 \leq 2||f||_\infty < \infty$$

per n=1,2... se f non ha zeri allora per n sufficientemente grande:

$$||C^{-1}||_2 \leq 2\left\|\frac{1}{f}\right\|_\infty$$

Caso quadrato

Lemma 3

Sia f$\in C_{2\pi}$, T la corrispondente matrice Toeplitz $n \times n$ e C la corrispettiva matrice di precondizionamento di T allora per ogni $\epsilon > 0$ esistono $N, M \in \mathbb{N}$, $M, N > 0$ tali che per ogni $n > N$:

$$T - C = U + V$$

Dove U, V sono matrici $n \times n$ tali che: $rankU \leq M$ e $\|V\|_2 \leq \epsilon$

Costruzione della matrice

La nostra matrice di partenza però è $m \times n$ per cui ci poniamo in un caso più generale ovvero:

$$T = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ \vdots \\ T_k \end{bmatrix}$$

Dove T_i sono matrici Toeplitz $n \times n$ e supponiamo, a meno di estendere T in modo Toeplitz con degli zeri, che $m = nk$. L'analisi che svolgeremo considererà k costante indipendente da n.

Costruzione della matrice

Definiamo quindi la matrice di precondizionamento C di T come:

$$C^*C = \sum_{j=1}^{k} C_j^* C_j$$

Dove le matrici C_j sono le precondizionatrici quadrate definite prima per i blocchi T_j. Poichè le matrici C_j sono circolanti:

$$C = F(\sum_{j=1}^{k} \Lambda_j^* \Lambda_j)^{\frac{1}{2}} F^*$$

Dove F è la trasformata discreta di Fourier e Λ_j sono matrici diagonali date da $\Lambda_j = F^* C_j F$.

Osserviamo che $(TC^{-1})^* TC^{-1}$ è simile a $(C^*C)^{-1}(T^*T)$ per cui possiamo studiare equivalentemente la distribuzione degli autovalori di quest'ultima per stimare la velocità di convergenza.

Proprietà della matrice

Lemma 4

Sia $f_j \in C_{2\pi}$ per ogni j=1,2...k. Allora

$$||C||_2^2 \leq 4 \sum_{j=1}^{k} ||f_j||_\infty^2 \leq \infty \quad n = 1, 2, ...$$

inoltre se uno degli f_j no ha zeri, supponiamo f_l, allora per n sufficientemente grande si ha

$$||(C^* C)^{-1}||_2 \leq 4 \left\| \frac{1}{f_l} \right\|_\infty^2$$

Dimostrazione: il primo punto segue dalla definzione di C e dal lemma 1, per il secondo in aggiunta basta ricordare che C^*C sono matrici Hermitiane semidefinite positive e quindi $\lambda_{min}(C^*C) \geq \lambda_{min}(C_l^* C_l)$ e il Lemma 2.

Proprietà della matrice

Lemma 5

Data T_j per $1 \leq j \leq k$, allora per ogni $\epsilon > 0$ esistono N_j e $M_j > 0$ tali che per ogni $n > N_j$,

$$T_j^* T_j - C_j^* C_j = U_j + V_j$$

dove U_j e V_j sono matrici **Hermitiane** tali che $rank U_j \leq M_j$ e $\|V_j\|_2 \leq \epsilon$

Proprietà della matrice

Lemma 5

Data T_j per $1 \leq j \leq k$, allora per ogni $\epsilon > 0$ esistono N_j e $M_j > 0$ tali che per ogni $n > N_j$,

$$T_j^* T_j - C_j^* C_j = U_j + V_j$$

dove U_j e V_j sono matrici **Hermitiane** tali che $rank U_j \leq M_j$ e $\|V_j\|_2 \leq \epsilon$

$$T_j^*(\tilde{U}_j + \tilde{V}_j) + (\tilde{U}_j + \tilde{V}_j)^*(\tilde{U}_j + \tilde{V}_j) + (\tilde{U}_j + \tilde{V}_j)^* T_j$$

con $\tilde{U}_j$ e $\tilde{V}_j$ dati dal Lemma 3 applicato a $T_j - C_j$

Proprietà della matrice

Lemma 6

Data T_j per $1 \leq j \leq k$, allora per ogni $\epsilon > 0$ esistono N e $M > 0$ tali che per ogni $n > N$,

$$T^*T - C^*C = U + V$$

dove U e V sono matrici **Hermitiane** tali che $rankU \leq M$ e $\|V\|_2 \leq \epsilon$

La dimostrazione segue direttamente da

$$T^*T - C^*C = \sum_{j=1}^{k} T_j^*T_j - C_j^*C_j$$

unito al Lemma 5 e al fatto che k non dipende da n.

Proprietà della matrice

Teorema 1

Sia $f_j \in C_{2\pi}$ per $j = 1, ..., k$, se esiste l tale per cui f_l non si annulla in nessun punto, allora per ogni $\epsilon > 0$ esistono N e M > 0 tali che per ogni $n > N$, al più M autovalori di $(C^*C)^{-1}(T^*T) - I$ hanno valore assoluto $> \epsilon$.

Dimostrazione

Avendo

$$(C^*C)^{-1}(T^*T) - I = (C^*C)^{-1}(U + V)$$

lavoriamo con

$$(C^*C)^{-1/2}(U + V)(C^*C)^{-1/2}$$

Sapendo che $rank(C^*C)^{-1/2}U(C^*C)^{-1/2} \leq M$ e che

$$\left\|(C^*C)^{-1/2}V(C^*C)^{-1/2}\right\|_2 \leq 4\hat{\epsilon}\left\|\frac{1}{f_l}\right\|_\infty^2$$

Proprietà della matrice

La tesi segue da:

Teorema

Siano A, B, C $\in \mathbb{C}^{n\times n}$ hermitiane, tali che C = A + B. Per gli autovalori

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \text{ di A,}$$
$$\mu_1 \geq \mu_2 \geq ... \geq \mu_n \text{ di B,}$$
$$\nu_1 \geq \nu_2 \geq ... \geq \nu_n \text{ di C,}$$

vale la relazione

$$\lambda_k + \mu_n \leq \nu_k \leq \lambda_k + \mu_1$$

per k = 1, . . . , n.

Quindi ponendo $A = (C^*C)^{-1/2}U(C^*C)^{-1/2}$ e $B = (C^*C)^{-1/2}V(C^*C)^{-1/2}$ si otteniene la tesi.

Test numerici

Test 1

Consideriamo T, una matrice $m \times n$ Toeplitz, definita tramite le sue righe e colonne come:

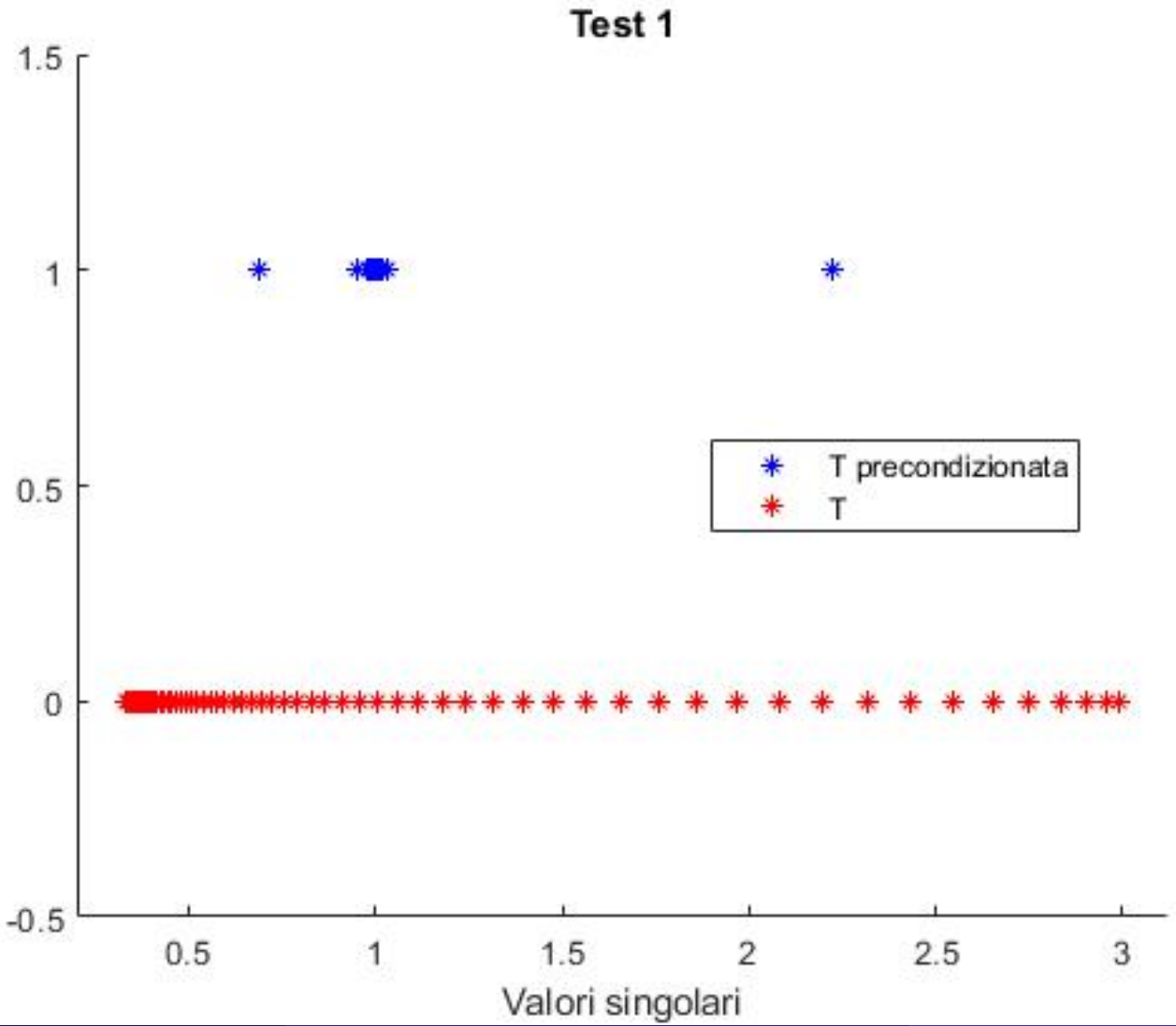
$$c(i) = \frac{1}{2^i} \text{ per i=1,...,m}$$
$$r(j) = \frac{1}{2^j} \text{ per j=1,...n}$$
$$m = 3n$$

Table: Numero di iterazioni necessarie per soddisfare il criterio di arresto nel caso precondizionato e non precondizionato, per $m = 3n$

n	40	50	60	70	80	100	120
C-prec	7	7	7	7	7	7	7
non-prec	31	36	39	41	42	45	48

Test numerici

Figure: Valori singolari di T e TC^{-1} per m=210 e n=70



Test 2

Consideriamo tre matrici, T_1, T_2, T_3 definite, per m=2n:

T_1, ben condizionata:

$$c(k) = k^{-1.1} + i(k^{-1.1}) \text{ per k=1,...,m}$$
$$r(j) = j^{-1.1} + i(j^{-1.1}) \text{ per } j = 1, ...n$$

T_2, mal condizionata:

$$r(1) = c(1) = 0$$
$$c(k) = k^{-1.1} + i(k^{-1.1}) \text{ per k=2,...,m}$$
$$r(j) = j^{-1.1} + i(j^{-1.1}) \text{ per j=2,...n}$$

e T_3, sparsa, costruita a partire da T_2 ponendo a zero alcuni elementi della prima riga e colonna.

Sia M la matrice matrice tridiagonale ottenuta dalle diagonali di T^*T

Test numerici

Figure: residui a ogni passo per n=250, per matrice ben condizionata

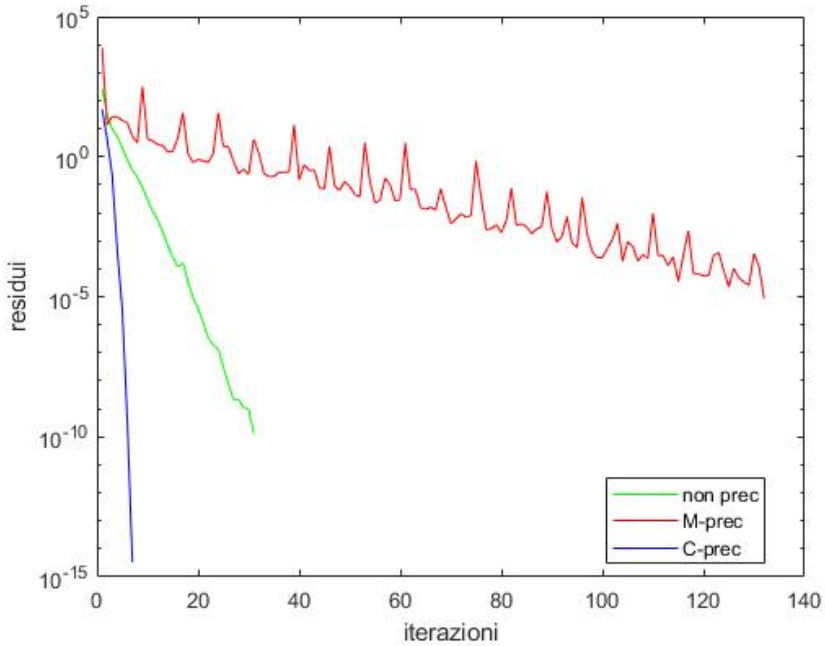
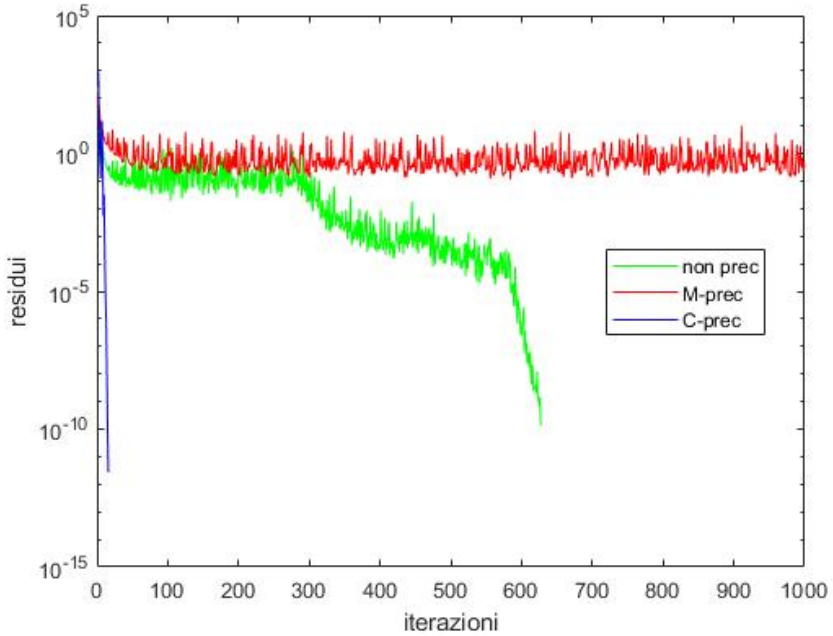


Figure: residui a ogni passo per n=250, per matrice mal condizionata



Test numerici

Table: Numero di iterazioni necessarie per soddisfare il criterio di arresto nel caso non precondizionato, precondizionato da M e da C, per $m = 2n$

	Matrice ben condizionata			Matrice mal condizionata			Matrice sparsa		
n	no-pre	M-pre	C-pre	no-pre	M-pre	C-pre	no-pre	M-pre	C-pre
100	22	65	7	168	493	15	39	37	14
150	25	84	7	297	>1000	15	42	41	14
250	31	132	7	627	>1000	16	45	45	15
500	36	246	7	>1000	>1000	17	49	48	15

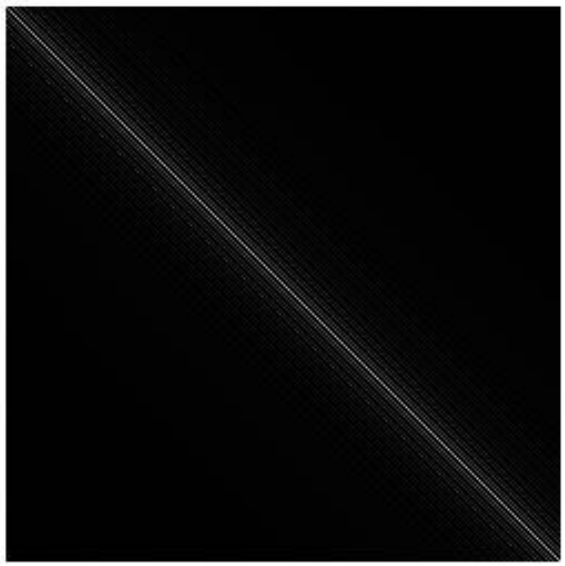


Figure: $T_3^* T_3$ per $m = 500$ e $n = 250$

Test 3

Consideriamo T definita come

$$c(k) = i(k^{-1.1}) \text{ per k=1,...,m}$$
$$r(j) = j^{-1.1} \text{ per j=1,...n}$$
$$m = 4n$$

Table: Numero di iterazioni e tempi necessari per soddisfare il criterio di arresto nel caso precondizionato da C per m=4n

n	iterazioni	tempo(s)
31250	30	3.97
62500	29	6.43
125000	33	36.30
250000	34	129.58