

Metodi numerici per catene di Markov

Giulio Del Corso

15 agosto 2016

Indice

1	Introduzione alle Catene di Markov	5
1.1	Introduzione alle Catene di Markov	5
1.2	Classificazione per ricorrenza degli stati	8
1.3	Classi irriducibili	8
1.4	Processo Birth and Death	11
1.5	Teorema di Perron - Frobenius	12
1.6	Teorema Bound su $\rho(A)$	13
2	M-matrici	15
2.1	M-matrici	15
2.2	Applicazioni delle M-matrici	17
2.2.1	Primo metodo	17
2.2.2	Secondo metodo	18
2.2.3	Terzo metodo	18
2.2.4	Quarto metodo	19
2.2.5	Quinto metodo	19
2.3	Calcolo della fattorizzazione L, U	19
2.4	Splitting	21
2.4.1	Idea 1:	22
2.4.2	Idea 2: (Jacobi)	22
2.4.3	Idea 3: Gauss - Seidel	22
2.5	Modificare il partizionamento	24
2.6	Metodi non standard	26
2.7	M/G/1-type	27
2.8	Condizioni di positiva ricorrenza	27
3	Metodi di calcolo e iterazioni	29
3.1	Equazione Matriciale	29
3.1.1	Calcolo di Π_0	30
3.2	Metodi di iterazione funzionali	30
3.2.1	Metodo naturale N	30
3.2.2	Metodo tradizionale T	31
3.2.3	Metodo a U-based U	31

3.2.4	Confronto	31
3.3	Metodi a convergenza quadratica	32
3.3.1	Algoritmo di logarithmic reduction (LR)	32
3.4	Algoritmo di Cyclic - Reduction	33
3.4.1	Formule	34
3.4.2	Costo, stabilità e convergenza	34
3.5	Variante	34
3.6	Metodo di Newton	35

Capitolo 1

Introduzione alle Catene di Markov

1.1 Introduzione alle Catene di Markov

Definizione 1. Una **Catena di Markov** (informale) è un processo stocastico dipendente solo dai dati (stato) attuali e non dal passato.

Ci si interroga sull'esistenza di uno stato limite.

Esempio: Un esempio di catena di Markov sono le reti di comunicazione, esiste infatti un flusso di clienti (che accedono con una certa legge) che vanno distribuiti fra dei server affinché la fila sia il più corta possibile.

Definizione 2. Una **Catena di Markov** è un insieme $X = \{X_n\}$ con X_n variabili aleatorie, $n \in \mathbb{N} = T$ (Tempo). $X_n \in E$ (Eventi, E numerabile) tale che:

$$P\{X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} | X_n = i_n\}$$

Per ogni $n \in \mathbb{N}$ e per ogni $i_0, \dots, i_n \in E$.

Un'ipotesi aggiuntiva utilizzata è che le catene siano **omogenee rispetto al tempo**, ossia

$$P\{X_{n+1} = j | X_n = i\} = P_{i,j} \forall n \in \mathbb{N}, \forall i, j \in E.$$

Con $P_{i,j}$ la **probabilità di transizione**.

Esempio: Particella oscillante con probabilità p di passare alla posizione successiva ed $1 - p$ di passare alla precedente. Se è finita la

dimensione della barretta su cui può muoversi (fra 0 e k) allora bisogna imporre l'ipotesi aggiuntiva che in 0 si muova con probabilità 1 verso destra e che in k si muova con probabilità 1 verso sinistra. Se X_n è la posizione della particella, allora questa è una catena di Markov omogenea rispetto al tempo.

Definizione 3. Una **Matrice di transizione** (o Transition Matrix) è la matrice i cui coefficienti sono le probabilità di transizione:

$$\mathbb{P} = (P_{i,j})_{i,j \in E}$$

Una matrice si dice **stocastica** se:

$\mathbb{P} \geq 0$ (Non negative matrix)

$\dim \mathbb{R} = \text{Card}(E)$

$\sum_{j \in E} P_{i,j} = 1$

Esempio: Riprendendo l'esempio di prima:

$$\mathbb{P} = \begin{pmatrix} 0 & 1 & & & & & \\ 1-p & 0 & 1 & & & & \\ & 1-p & 0 & 1 & & & \\ & & \dots & \dots & \dots & & \\ & & & 1-p & 0 & 1 & \\ & & & & 1 & 0 & \end{pmatrix}.$$

Cerchiamo se esiste un valore per il limite:

$$\Pi_j := \lim_{n \rightarrow \infty} P\{X_n = j | X_0 = i\}$$

Osservazione 1. Non è detto che esista. L'esempio di prima, imponendo $k = 1$ da origine ad un ciclo fra i due punti 0 ed 1 che non si stabilizza.

Il limite è un "autovalore" (Lo è davvero se $\text{Card}(E) = 1$) della matrice di transizione per autovettore 1.

Esempio: Sappiamo che se $\text{Card}(E)$ = finita allora:

$$\mathbb{P} \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}$$

e $\rho(\mathbb{P}) = 1$ per Ghergorin, sono tutti cerchi con centro in $[0, 1]$ e raggio 1.

Inoltre $\|\mathbb{P}\|_\infty = 1$ e $\rho(\mathbb{P}) \leq \|\mathbb{P}\|_\infty = 1$

Quale è l'autovettore sinistro?

Se esiste è $\Pi = (\Pi_i)_{i \in E}$ (è già normalizzato) e vale:

$$\Pi^T \cdot \mathbb{P} = \Pi^T ; \sum_{i \in E} \Pi_i = 1.$$

È dunque diventato un problema di calcolo di autovettori.

Equivalente:

$$\Pi^T (Id - \mathbb{P}) = 0$$

Osservazione 2. Abbiamo già dei metodi per risolvere questo problema (metodo delle potenze, fattorizzazione LU, etc...). Vogliamo però sfruttare le proprietà di queste particolari matrici.

$$\text{Sia } e := \underline{1} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}.$$

Si osservi che una potenza di una matrice stocastica è ancora una matrice stocastica.

Teorema 1. $P\{X_n = j | X_0 = i\} = (P^n)_{i,j}$

Definizione 4. Definiamo il vettore Π .

$$\Pi^{(n)} = \left(\Pi_i^{(n)} \right)_{i \in E} ; \Pi_i^{(n)} = P\{X_n = 1 | X_0 = i\} \quad n \geq 1 ; \Pi_i^{(0)} = P\{X_0 = i\}.$$

Vale:

$$\Pi^{(n+1)T} = \Pi^{(n)T} P ; \quad n \geq 0$$

Si usano spesso i vettori riga, il simbolo di trasposizione non è dunque necessario.

Inoltre:

$$\Pi^{(n)T} = \Pi^{(0)T} P^n.$$

Siamo interessati a determinare se esiste (e nel caso a calcolare):

$$\Pi_i := \lim_{n \rightarrow \infty} \Pi_i^{(n)}$$

Se esiste, allora $\Pi = (\Pi_i)_{i \in E}$.

Osservazione 3. Dalla relazione fra i Π vale $\Pi^T = \Pi^T P$.

1.2 Classificazione per ricorrenza degli stati

Definiamo come segue il **tempo della prima visita allo stato** j :

$$j \in E ; T_j = \min n \geq 1 | X_n = j$$

e la probabilità di **ritornare allo stato** j :

$$f_j = P\{T_j < \infty | X_0 = j\}$$

Definizione 5. Lo stato j si dice **ricorrente** se $f_j = 1$.

Viceversa è detto **transiente** (transient) se $f_j < 1$.

Inoltre ci si interroga su “quanto ci metta lo stato a ripresentarsi.

Definizione 6. Se il tempo è finito si dice che uno stato ricorrente è **positive recurrent**.

$$\mathbb{E}\{T_j | X_0 = j\} < \infty$$

Se il tempo è infinito si dice che uno stato ricorrente è **null recurrent**.

$$\mathbb{E}\{T_j | X_0 = j\} = \infty$$

Esempio: $E = \{1, 2, 3\}$; $\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$.

Allora 1 è uno stato positive recurrent (absorbing), 2 e 3 sono invece transient.

Si osservi che questa matrice è riducibile.

1.3 Classi irriducibili

Si può associare ad una matrice \mathbb{P} un grafo seguendo le seguenti regole:

per ogni $i \in E$ si rappresenta un nodo,

se $P_{i,j} \neq 0$ si traccia una freccia fra il nodo i e il nodo j .

Definizione 7. Si dice che esiste un **cammino** (Path) fra due nodi i, j se esiste una successione di archi finita $(i, k_1) - (k_1, k_2) - \dots - (k_n, j)$.

Definizione 8. Gli stati i, j **comunicano** se esistono $\text{Path}(i, j)$ e $\text{Path}(j, i)$.

Per convenzione uno stato comunica con se stesso.

Questa relazione è di equivalenza e le classi sono dette **classi irriducibili**.

Definizione 9. Una catena di Markov è detta **irriducibile** se esiste un'unica classe di equivalenza. Equivalentemente se la matrice associata \mathbb{P} è irriducibile.

Definizione 10. Una classe si dice **finale** se $\forall i \in C$ non esiste $j \notin C$ tale che $i \rightarrow j$.

In pratica è una classe dalla quale non è più possibile uscire. Si osservi che esistono sempre delle classi finali. Una classe non finale è detta **classe di passaggio**.

Teorema 2. C_1, \dots, C_k classi irriducibili, allora esiste Π permutazione tale che:

$$\Pi^T P \Pi = \begin{pmatrix} P_{11} & & & \\ P_{21} & P_{22} & & \\ \dots & \dots & \dots & \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{pmatrix}$$

con P_{ii} classi finali.

Si osservi inoltre che se uno stato è di un determinato tipo, allora tutti gli elementi di quello stato sono del medesimo tipo.

Tutti gli stati di una classe di passaggio sono transienti.

Teorema 3. Se la dimensione è finita ogni classe è positiva ricorrente. Se la dimensione è infinita ogni possibilità è valida.

Osservazione 4. Si lavora con catene di Markov irriducibili.

Definizione 11. $\Pi = (\Pi_i)_{i \in E}$ è un **invariant vector** se $\Pi^T P = \Pi^T$.
É un **probability invariant vector** se $\Pi_i \geq 0 \forall i$ e $\sum_{i \in E} \Pi_i = 1$.

Teorema 4. Data X irriducibile.

Positive recurrent \iff esiste $\Pi > 0$ probability invariant vector.

Osservazione 5. Nel caso finito questo vettore è unico.

Nel caso positive recurrent $\mu_g = \mu_a = 1$.

Definizione 12. Uno stato $i \in E$ è detto **periodico** di periodo $\delta \geq 2, \delta \in \mathbb{N}$ se tutti i cammini da i in i sono formati da un numero di archi multiplo di δ .

Si osservi che è una proprietà di classe. Nell'esempio di prima gli stati sono periodici con periodo 2.

Teorema 5. X irriducibile, positiva ricorrente e A periodica. Allora:

$$\exists \lim_{n \rightarrow \infty} P\{X_n = j | X_0 = i\} = \Pi_j \quad \forall i, j \in E$$

Teorema 6. X irriducibile, null recurrent o transiente. Allora:

$$\exists \lim_{n \rightarrow \infty} P\{X_n = j | X_0 = i\} = 0 \quad \forall i, j \in E$$

Teorema 7. X irriducibile, positiva ricorrente e A periodica. Allora:

$$\exists \lim_{n \rightarrow \infty} (P^n)_{i,j} = \Pi_j \quad \forall i, j \in E$$

Esempio: Sia $|E| < \infty$; prendiamo ad esempio $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$.

Questa matrice è stocastica, irriducibile e aperiodica.

Si può cambiare di base ed ottenere $P = V \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{6} \end{pmatrix} V^{-1}$ da cui $\rho(P) = \{1, \frac{1}{6}\}$.

Si osservi che se ha elementi sulla diagonale allora è aperiodica (esistono i cicli liberi sull'elemento stesso).

Calcoliamone la potenza n -esima:

$P^n = V \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{6^n} \end{pmatrix} V^{-1}$ che, portata al limite, restituisce:

$$V \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} V^{-1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (\Pi_1 \quad \Pi_2).$$

Esempio (Teoria delle Code): Sia $E = \mathbb{N}$, prendiamo una coda di clienti (customers) con una determinata probabilità ed un server che ne soddisfa le richieste con una legge di servizio assegnata.

Sia $T = \mathbb{N}$. La legge di servizio in questo caso è di tipo **deterministico**, ossia 1 cliente viene servito in 1 unità di tempo.

Sia $\alpha_n =$ il numero di clienti che arrivano fra $[n-1, n]$ ed α_n variabili aleatorie indipendenti fra loro ed indipendenti da n .

Sia $q_i = P\{\alpha_n = i\}$.

La catena di Markov la descriviamo come:

$X_n := \#\{\text{Clienti in fila al tempo } n\}$. La matrice di transizione è dunque:

$$\mathbb{P} = \begin{pmatrix} q_0 + q_1 & q_2 & q_3 & \cdots & & & \\ q_0 & q_1 & q_2 & q_3 & \cdots & & \\ 0 & q_0 & q_1 & q_2 & q_3 & \cdots & \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \end{pmatrix}$$

Dove la diagonale rappresenta la probabilità di non cambiare stato, la sottodiagonale di aumentare di uno il numero di persone in fila, la sopradiagonale la probabilità di ridurre di uno il numero di persone in fila.

Si osservi che questa matrice è di Hessenberg (Ad ogni unità di tempo si può variare solo di un passo verso il basso) e di Toeplitz (la probabilità dipende solo dalla differenza degli indici).

1.4 Processo Birth and Death

Un processo è detto Birth and Death se i valori possono salire o scendere di livello (solo uno alla volta). Sia:

$$\mathbb{P} = \begin{pmatrix} b_0 & b_1 & 0 & \dots & & \\ a_{-1} & a_0 & a_1 & 0 & \dots & \\ 0 & a_{-1} & a_0 & a_1 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Osservazione 6. Nonostante il formalismo usato sfrutti le matrici, si osservi che se E non ha cardinalità finita, queste hanno un numero infinito di elementi; per questo non è possibile sfruttare i metodi già incontrati in analisi numerica.

Supponiamo la matrice irriducibile. A seconda dei valori assegnati agli a_i si ottengono casi distinti.

Innanzitutto, se $a_{-1} > a_1 \rightarrow$ lo stato limite è 0.

Se $a_{-1} < a_1 \rightarrow$ lo stato limite “scoppia”.

Cerchiamo $\Pi^T(Id - P) = 0$; $\sum \Pi_i < \infty \rightarrow$ caso positive recurrent, con $\Pi > 0$; $\Pi = (\Pi_1, \Pi_2, \dots)$.

Scriviamo esplicitamente le componenti:

$$\begin{aligned} \Pi_0(1 - b_0) - \Pi_1 a_{-1} &= 0 \\ -\Pi_0 b_1 + \Pi_1(1 - a_0) - \Pi_2 a_{-1} &= 0 \\ \dots & \end{aligned}$$

Consideriamo l'equazione a regime che è un'equazione alle differenze finite:

$$-\Pi_i a_1 + \Pi_{i+1}(1 - a_0) - \Pi_{i+2} a_{-1} = 0$$

Si associa il polinomio $P(\lambda) = -a_1 + \lambda(1 - a_0) - \lambda^2 a_{-1}$.

Siano $\lambda_1, \lambda_2 | P(\lambda_i) = 0$ e $\lambda_1 \neq \lambda_2$ da cui $\Pi_i = \alpha \lambda_1^{i-1} + \beta \lambda_2^{i-1}$.

Se $\lambda_1 = \lambda_2 \rightarrow \Pi_i = \alpha \lambda_1^{i-1} + \beta(i-1)\lambda_1^{i-1}$

Quali sono le radici del polinomio?

La somma dei coefficienti è 1 $\rightarrow \lambda_1 = 1, \lambda_2 = \frac{a_1}{a_{-1}}$.

Riscrivendolo:

$$\lambda_1 \neq \lambda_2 \rightarrow \Pi_i = \alpha + \beta \lambda_2^{i-1}$$

$$\lambda_1 = \lambda_2 \rightarrow \Pi_i = \alpha + \beta(i-1) \rightarrow \text{la soluzione esplode}$$

Poiché nel primo caso non esplode, allora deve valere $\alpha = 0$ e $\lambda_2 < 1$.

Quindi, se $a_1 = a_{-1} \rightarrow$ non è positive recurrent. È null recurrent.

Al contrario, se $a_1 \neq a_{-1}$ allora sviluppando il conto si ottiene $\alpha = 0$ da cui

$$\Pi_i = \beta \left(\frac{a_1}{a_{-1}} \right)^{i-1}; i \geq 1. \text{ È quindi positive recurrent se } a_1 < a_{-1}.$$

Casi:

$$a_1 < a_{-1} \rightarrow \text{PR.}$$

$$a_1 > a_{-1} \rightarrow \text{TR.}$$

$$a_1 = a_{-1} \rightarrow \text{NR.}$$

1.5 Teorema di Perron - Frobenius

Lemma 1. $A \geq 0$, irriducibile, $n \times n \implies (Id + A)^{n-1} > 0$.

Definizione 13. $z \in P$ tale che $Az \geq rz$ si dice **estremale**.

z estremale, $r_z = \min_{z_i \neq 0} \frac{\sum_j a_{ij} z_j}{z_i}$ dunque è $\geq \min \frac{r z_i}{z_i} = r$.

Quindi per ogni estremale vale $r_z \geq r \implies r_z = r$.

Gli estremali sono quindi quelli nei quali vale l'uguaglianza.

Lemma 2. Se $A \geq 0$ è irriducibile, allora:

1. $r > 0$.
2. Se $<$ è estremale $\implies z \geq 0$ e $Az = rz$ (Autovettore con autovalore r).

Lemma 3. $A \geq 0$, irriducibile, $B \in \mathbb{C}^{n \times n}; |B| \leq A$. Allora:

1. Se β è un autovalore per $B \implies |\beta| \leq r$.
2. Se $|\beta| = r \implies B = e^{i\theta} D A D^{-1}$, ossia $\beta = e^{i\theta} r$ con $D = \text{diag}(d_1, \dots, d_n)$, $|d_i| = 1$.

Da questo segue che il raggio spettale di A è un autovalore.

Lemma 4. $A \geq 0$, irriducibile, B sottomatrice principale di $A \implies \rho(B) < \rho(A)$.

Teorema 8 (Perron - Frobenius). $A \in \mathbb{R}^{n \times n}; A \geq 0; A$ irriducibile. Allora:

1. $\rho(A)$ è autovalore e $\rho(A) > 0$.
2. $\exists x \in \mathbb{R}^n, x > 0 | Ax = \rho(A)x$.
3. $B \geq A, B \neq A \rightarrow \rho(B) > \rho(A)$.
4. $\rho(A)$ è semplice.

L'importanza di questo teorema per quanto riguarda le catene di Markov è data dal fatto che se $A = \mathbb{P}$ è una transition matrix di una catena irriducibile,

per Perron - Frobenius esiste $P \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}$ (Noto) ma anche:

$\exists \Pi > 0; \Pi^T P = \Pi^T \implies$ che la catena è irriducibile con $|E| < \infty$ e positive recurrent.

Teorema 9 (Weak Perron - Frobenius). $A \geq 0$, allora:

1. $\rho(A)$ è un autovalore.
2. $\exists x \in \mathbb{R}^n, x \geq 0 | Ax = \rho(A)x$.
3. $B \geq A, B \neq A \implies \rho(B) \geq \rho(A)$.
4. NON vale più $\rho(A) > 0$.
5. NON è più detto che l'autovalore sia semplice.

1.6 Teorema Bound su $\rho(A)$

Teorema 10 (Bound su $\rho(A)$). $A \geq 0$ irriducibile, allora vale una e una sola fra:

1. $\sum_{j=1}^n a_{ij} = \sigma \forall i$ e $\sigma = \rho(A)$.
2. (Se la somma è non costante) $\min_i \sum_{j=1}^n a_{ij} < \rho(A) < \max_i \sum_{j=1}^n a_{ij}$

Si osservi che se $A \geq 0$ è irriducibile, $\mathbb{P} \sim \mathbb{F} \implies \rho(A) > 0$ e semplice.

$\exists \lambda$ autovalore di $A | |\lambda| = \rho(A), \lambda \neq \rho(A)$?

Si, ad esempio se $\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma(A) = \{1, -1\}$.

Definizione 14. $A \geq 0$ irriducibile, A è **primitiva** se $\rho(A)$ è l'unico autovalore di modulo massimo.

A è **ciclica di indice** $k > 1$ se $\exists k$ autovalore di modulo $\rho(A)$.

Teorema 11. $A \geq 0$ irriducibile, ciclica di indice k , allora:

$$\lambda_j = e^{i\theta_j} \rho(A), j = 0, \dots, k-1; \theta_j = \frac{2\pi j}{k}.$$

Sono gli autovalori di modulo massimo.

Teorema 12. *A ciclico di indice $k \implies \exists \Pi$ di permutazione tale che:*

$$\Pi^T A \Pi = \begin{pmatrix} 0 & A_{12} & 0 & \cdots & \\ 0 & 0 & A_{23} & 0 & \cdots \\ 0 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & A_{k-1,k} \\ A_n & \cdots & \cdots & \cdots & 0 \end{pmatrix}.$$

Con blocchi quadrati e diagonali.

Una domanda interessante può essere: P stocastica, irriducibile e ciclica di indice k , allora $\exists \lim_{r \rightarrow \infty} P^r$?

La risposta è no, come controesempio basta scegliere $\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathbb{P}^2 = Id$.

Osservazione 7. *Ciclica di indice $k \sim$ Catena di Markov periodica di indice k .*

Capitolo 2

M-matrici

2.1 M-matrici

Definizione 15. $A \in \mathbb{R}^{n \times n}$ è una **M-matrice** se:

$$A = sId - B; B \geq 0; s \in \mathbb{R}^+ \text{ e } \rho(B) \leq s$$

Si dice **singular** se $\rho(B) = s$.

Si dice **non singular** se $\rho(B) < s$.

Esempio: P stocastica $\implies Id - P$ è una M-matrice singolare.

Teorema 13. A M-matrice $\implies A = \begin{pmatrix} * & & & \\ & * & & \\ & & \dots & \\ & & & * \end{pmatrix}$.

Dove ogni spazio vuoto è ≤ 0 .

Definizione 16. A è una **Z-matrix** se $a_{ij} \leq 0$ per $i \neq j$.

Quando una Z-matrice è una M-matrice?

Esistono più di 50 condizioni equivalenti, sono qui riportate le più interessanti.

Teorema 14. Z-matrix $\implies A$ non singular M-matrix. È vera se è valida una fra:

1. $Re(\lambda) > 0 \forall \lambda$ autovalore di A .
2. $\det A \neq 0$ e $A^{-1} \geq 0$.
3. $x \in \mathbb{R}^n; Ax \geq 0 \implies x \geq 0$.

$$4. \exists D = \text{diag}(d_1, \dots, d_n) \quad d_i > 0 \quad i = 1, \dots, n \quad \text{e} \quad AD \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}.$$

$$5. a_{ii} > 0 \quad \forall i \quad \text{e} \quad \exists D = \text{diag}(d_1, \dots, d_n) \quad d_i > 0 \quad \forall i = 1, \dots, n \quad \text{e} \\ a_{ii}d_i > -\sum_{j=1; j \neq i}^n a_{ij}d_j.$$

Teorema 15. $A \in \mathbb{R}^{n \times n}$, A è non singular matrix \iff vale una fra le seguenti:

1. $(A + D)^{-1} \geq 0 \quad \forall D$ diagonale con $d_i > 0$.
2. $(A + \alpha Id)^{-1} \geq 0 \quad \forall \alpha \geq 0$.
3. Ogni sottomatrice principale di A ha inversa non negativa.

Teorema 16 (Singolari). A Z-matrix, allora:

A è una M-matrix $\iff A + \epsilon Id$ è una non singular M-matrix $\forall \epsilon > 0$.

Lemma 5. A non singular M-matrix \implies ogni sottomatrice principale è una non singular M-matrix.

Lemma 6. A singular irriducibile M-matrix \implies ogni sottomatrice principale non banale è una non singular M-matrix.

Teorema 17. A non singular M-matrix $\implies \exists! A = LU$ e L, U sono non singular M-matrix.

Teorema 18. A singular irriducibile M-matrix $\implies \exists!$ fattorizzazione $A = LU$ con L, U M-matrix con $\det U = 0$.

Si osservi che deve valere $\alpha = 0$.

Teorema 19 (Sintesi). A M-matrix ($n \times n$) singular irriducibile, allora:

1. $\text{Rank}(A) = n - 1$.
2. $\exists x \in \mathbb{R}^n, x > 0, Ax = 0$.
3. Ogni sottomatrice principale è una M-matrix non singolare.
4. $\exists A = LU$ con L, U M-matrix e $\det U = 0$.

2.2 Applicazioni delle M-matrici

Sia P una stochastic irriducibile matrix, $A = Id - P$ una M-matrix singular e irriducibile, ammette quindi una fattorizzazione LU .

É vero che $\Pi^T(Id - P) = 0$?

I teoremi studiati permettono di applicare l'eliminazione gaussiana:

$$\Pi^T(LU) = 0 \text{ se } \Pi^T \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix} = 1.$$

Vogliamo risolvere:

$$\begin{cases} y^T U = 0 \\ \Pi^T L = y^T \end{cases}$$

$$\text{Ma } U = \begin{pmatrix} * & * & \dots & \\ 0 & * & * & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & * \\ 0 & \dots & \dots & 0 \end{pmatrix}$$

Con elementi diagonali positivi e sopradiagonali ≤ 0 . Da cui $y^T = (0 \ \dots \ 0 \ 1)$.

Per rispondere alla domanda abbiamo diversi metodi possibili.

2.2.1 Primo metodo

Come accennato prima si risolve $\Pi^T LU = 0$ risolvendo il sistema:

$$\begin{cases} y^T U = 0 \\ \Pi^T L = y^T \end{cases}$$

Con $y^T = (0 \ \dots \ 0 \ 1)$. Dunque:

$\Pi^T L = (0 \ \dots \ 0 \ 1)$. Ossia:

$$(\Pi_1 \ \dots \ \Pi_n) \begin{pmatrix} 1 & 0 & & \\ \leq 0 & 1 & 0 & \\ \dots & \dots & \dots & \\ \leq 0 & \dots & & 1 \end{pmatrix} = (0 \ \dots \ 0 \ 1)$$

Per **sostituzione** si ottiene:

$$\begin{cases} \hat{\Pi}_n = 1 \\ \hat{\Pi}_j = -\sum_{i=j+1}^n \hat{\Pi}_i l_{ij} \end{cases}$$

Per $j = n - 1, \dots, 1$. E $\Pi_i = \frac{\hat{\Pi}_i}{\sum_{j=1}^n \hat{\Pi}_j}$.

Osservazione 8. Queste formule coinvolgono solo somme di numeri reali di segno concorde, dunque forniscono un algoritmo numericamente stabile.

Si può dimostrare anche una stabilità relativa all'indietro, è possibile approssimare numeri anche più piccoli della precisione di macchina.

2.2.2 Secondo metodo

P stocastica irriducibile, allora $(Id - P^T)\Pi = 0$.

Ma $Id - P^T$ è una M-matrix irriducibile e singolare, dunque:

$Id - P^T = LU$, M matrici con U sopradiagonale con ultimo elemento (n, n) nullo, elementi positivi sulla diagonale e ≤ 0 nella parte superiore.

L'equazione $LU\Pi = 0$ può essere scritta come un sistema:

$$\begin{cases} Ly = 0 \\ U\Pi = y \end{cases}$$

Essendo L non singolare, ha soluzione $y = 0$, dall'equazione otteniamo dunque $U\Pi = 0$. Da questo, mediante una **sostituzione all'indietro**:

$$\begin{cases} \hat{\Pi}_n = 1 \\ \hat{\Pi}_i = \frac{-\sum_{j=i+1}^n u_{ij}\hat{\Pi}_j}{u_{ii}}; i = n-1, \dots, 1 \end{cases}$$

Con tutti gli $u_{ij} \leq 0$ e $u_{ii} > 0$.

Osservazione 9. Come prima, la parte di somma di oggetti di segno concorde è numericamente stabile.

In questo caso si incorre però nel rischio di Overflow per elementi diagonali prossimi allo 0.

2.2.3 Terzo metodo

Si applica al Secondo metodo. Data $Id - P^T$ M-matrix singolare e irriducibile, eliminiamo l'ultima riga e l'ultima colonna ottenendo un sistema ridotto:

$$(Id - \hat{P}^T) \begin{pmatrix} \Pi_1 \\ \dots \\ \Pi_{n-1} \end{pmatrix} = \begin{pmatrix} p_{n,1} \\ p_{n,2} \\ \dots \\ p_{n,n-1} \end{pmatrix} := b$$

Sappiamo che le sottomatrici principali di $Id - P^T$ sono M-matrici non singolari. Sfruttando una fattorizzazione LU di $Id - \hat{P}^T$ otteniamo $LU\hat{\Pi} = b$ da cui:

$$\begin{cases} Ly = b \\ U\hat{\Pi} = y \end{cases}$$

Essendo tutti termini positivi è possibile risolverlo con un algoritmo numericamente stabile (sostituzione in avanti).

Osservazione 10. Il vantaggio (non molto importante) di questo metodo è dato dal fatto che la matrice ottenuta è non singolare.

Il grosso svantaggio è dato dal fatto che abbiamo due sistemi distinti da risolvere, costo doppio rispetto agli altri metodi.

2.2.4 Quarto metodo

Sviluppiamo $\Pi^T(Id - P) = 0$, con la matrice fra parentesi di rango $n - 1$. Se aggiungiamo la condizione $\Pi^T e = 1$ la soluzione è unica e possiamo direttamente inserirlo nel sistema.

Sostituiamo all'ultima colonna tutti 1, il sistema diventa:

$$(\Pi_1 \quad \cdots \quad \Pi_n) \begin{pmatrix} 1 - p_{1,1} & -p_{1,2} & \cdots & -p_{1,n-1} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & 1 - p_{n-1,n-1} & 1 \\ -p_{n,1} & \cdots & \cdots & \cdots & 1 \end{pmatrix} = (0 \quad \cdots \quad 0 \quad 1)$$

Ora abbiamo una matrice non singolare e il sistema ha un'unica soluzione.

Osservazione 11. Il condizionamento e i valori di Π al variare di $(Id - P)$ non è buono. Sostituendo l'ultima colonna si perde infatti la struttura di M-matrice \implies non è detto che L, U lo siano, quindi non abbiamo nessuna garanzia che nel sistema non sia presente una sottrazione \implies errori anche molto grandi.

2.2.5 Quinto metodo

Il quinto metodo sfrutta una tecnica di Pivoting sulla fattorizzazione LU . Data $(Id - P)$, permutando righe e colonne per sviluppare una strategia basata sui pivot.

In questo modo miglioriamo l'errore di costruzione della LU ma la permutata non è più una M-matrice, dunque L, U non è detto che lo siano. Per lo stesso motivo di prima non è dunque detto che convenga un metodo del genere.

2.3 Calcolo della fattorizzazione L, U

Abbiamo dimostrato che se $(Id - P) = LU \implies$ stabile, ma il calcolo di L ed U è anch'esso stabile?

Sia P una matrice stocastica irriducibile $n \times n$.

Siano I_1, I_2' degli stati che ricoprono l'insieme degli stati:

$$\begin{cases} I_1 \cap I_2 = \emptyset \\ I_1, I_2 \subset \{1, \dots, n\} \\ I_1 \cup I_2 = \{1, \dots, n\} \end{cases}$$

Vogliamo restringere la catena di Markov ad uno solo dei due stati (censurando gli altri, **censoring**). Sia:

$$P_{I_h, I_k} = (p_{ij})_{i \in I_h; j \in I_k} \text{ con } h, k \in \{1, 2\}.$$

Teorema 20. $P_1 = P_{I_1, I_1} + P_{I_1, I_2}(Id - P_{I_2, I_2})P_{I_2, I_1}^{-1}$ è stocastica (è la matrice associata alla catena ristretta).

Semplificando le notazioni:

$$I_1 = \{1, \dots, k\}$$

$$I_2 = \{k+1, \dots, n\}$$

$$P = \begin{pmatrix} P_{I_1, I_1} & P_{I_1, I_2} \\ P_{I_2, I_1} & P_{I_2, I_2} \end{pmatrix}.$$

Osserviamo che l'intera inversa è $(Id - P_{I_2, I_1})^{-1} \rightarrow \sum_{j=0}^{\infty} P_{I_1, I_2}^j$. La j rappresenta quello che rimane in I_2 .

Sfruttando il complemento di Schur:

$$Id - P = \begin{pmatrix} Id - P_{I_1, I_1} & -P_{I_1, I_2} \\ -P_{I_2, I_1} & Id - P_{I_2, I_2} \end{pmatrix}$$

e applicando l'eliminazione gaussiana si può costruire il seguente procedimento: sia $\Pi^T = (\Pi_{I_1}^T \quad \Pi_{I_2}^T) \rightarrow \Pi_{I_1}^T (Id - P_1) = 0$.

Consideriamo:

$$Se = (Id - P_1)e = 0$$

$$(Id - P)e = 0$$

$$\begin{pmatrix} Id - P_{11} & -P_{12} \\ -P_{21} & Id - P_{22} \end{pmatrix} \begin{pmatrix} e \\ e \end{pmatrix} = 0$$

$$LUe \begin{pmatrix} Id & 0 \\ -P_{21}(Id - P_{11})^{-1} & 0 \end{pmatrix} \begin{pmatrix} Id - P_{11} & -P_{12} \\ 0 & S \end{pmatrix} \begin{pmatrix} e \\ e \end{pmatrix} = 0$$

$$S = Id - P_{22} - P_{21}(Id - P_{11})^{-1}P_{12} \text{ e } Se = 0.$$

Essendo $\det L \neq 0$ questo è il complemento di Schur.

Osservazione 12. Abbiamo anche osservato che il complemento di Schur della matrice $Id - P$ è I -stocastica. Da questo è possibile costruire una buona fattorizzazione LU .

L'algoritmo è dunque:

$$(I - P) = \begin{pmatrix} 1 - P_{11} & -P_{12} & \cdots & -P_{1n} \\ -P_{21} & & & \\ \cdots & & A^{(0)} & \\ -P_{n1} & & & \end{pmatrix} = [a_{ij}]$$

$$B^{(1)} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & & & \\ \cdots & & A^{(1)} = a_{ij}^{(1)} & \\ 0 & & & \end{pmatrix}$$

Con $a_{ij}^{(1)} = a_{ij} - \frac{a_{1i}a_{1j}}{a_{11}}$. Di segno concorde se $i \neq j \implies$ stabile.
Se $i = j$? (Diagonale)

$$a_{ii} - \frac{a_{i1}a_{1i}}{a_{11}}$$

Dove entrambi i termini (senza la sottrazione) sono positivi, questo può dare grossi problemi di stabilità.

Il trucco è calcolare quegli elementi diagonali con una formula distinta che si origina dalle proprietà del completamento di Schur.

Osserviamo che $A^{(1)} = (a_{ij}^{(1)})$ può essere vista come un complemento di Schur:

$$A^{(1)} = (A^{(0)}) - \begin{pmatrix} P_{21} \\ \dots \\ P_{n1} \end{pmatrix} (1 - P_{11})^{-1} (P_{12} \dots P_{1n}).$$

Ma questa è una catena di Markov censurando il nodo 1, dunque:

$UI d - P^{(1)}$ con $P^{(1)}$ stocastica (insieme delle transizioni della catena senza lo stato 1).

Ma stocastica vuol dire che:

$$(Id - P^{(1)}) \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix} = 0 \implies \text{sono tutti di segno concorde, quindi il calcolo}$$

è stabile.

In questo caso $a_{ii}^{(1)} = -\sum_{j=2; j \neq i}^n a_{ij}^{(1)}$, questo trucco è detto **GTH-trick**.

2.4 Splitting

Sia A una matrice $n \times n$, prendiamone uno splittin $A = M - N$, se $\det M \neq 0$ vale:

$$Ax = b \implies (M - N)x = b \implies (Id - M^{-1}N)x = M^{-1}b \implies x = M^{-1}Nx + M^{-1}b. \text{ Da cui l'iterazione:}$$

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b$$

Siamo interessati ad un **regular splitting**, ossia al caso nel quale $M^{-1} \geq 0; N \geq 0$.

Dai risultati noti di analisi numerica se $\exists \lim_{k \rightarrow \infty} x^{(k)} = x^* \implies Ax^* = b$.

Definizione 17. Un metodo si dice **convergente** se $\exists x^* \forall$ scelta del punto iniziale.

Una condizione necessaria e sufficiente è che $\rho(M^{-1}N) < 1$.

Osservazione 13. Se A è singolare ($\det A = 0$) e vogliamo determinare:

$$Ax = b \implies \exists \bar{x} | A\bar{x} = 0 \implies (M - N)x = 0; (I - M^{-1}N)x = 0 \implies M^{-1}Nx = x \implies \rho(M^{-1}N) \geq 1 \implies \text{non convergenza del metodo } (\forall x_0).$$

Quindi $\det A \neq 0$.

Sia $H = M^{-1}N = (A + N)^{-1}N = (A(Id + A^{-1}N))^{-1}$ che, posto $G = A^{-1}N$ è uguale a:

$$(Id + G)^{-1}G.$$

$$\text{Sia } Gx = \lambda x \implies Hx = (Id + G)^{-1}Gx = \frac{\lambda}{1+\lambda}x$$

$$\text{Sia } Hx = \mu x \implies (Id + G)^{-1}Gx = \mu x \implies Gx = (Id + G)\mu x \implies G(1 - \mu)x\mu x \implies Gx = \frac{\mu}{1-\mu}x.$$

Quindi λ è un autovalore con x autovettore di $G \iff \mu = \frac{\lambda}{1+\lambda}$ è autovalore di H con x autovettore corrispondente.

Teorema 21. *det $A \neq 0$ con A reale, $A^{-1} \geq 0$; $A = M - N$ regular splitting. Allora:*

$$\rho(H) = \frac{\rho(G)}{1 + \rho(G)} < 1$$

Corollario 1: Ogni regular splitting di matrici con inversa non negativa è convergente.

Sia A una M-matrix non singular, vogliamo risolvere $Ax = b$.

2.4.1 Idea 1:

Sia $A = sId - B \implies B \geq 0; \rho(B) \leq s, M = sId; N = B$.
La matrice di iterazione $M^{-1}N = \frac{B}{s}$ con $\rho(M^{-1}N) < 1$.

2.4.2 Idea 2: (Jacobi)

L'idea è lavorare isolando la diagonale.

$A = D - L - U$ con D diagonale, L triangolare inferiore con diagonale nulla ed U triangolare superiore con diagonale nulla.

$M_2 = D; N_2 = L + U \geq 0$ e $M_2^{-1}N_2 = J \geq 0 \implies$ regular splitting \implies metodo di Jacobi convergente.

2.4.3 Idea 3: Gauss - Seidel

$$M_3 = D - L, N_3 = U \geq 0, G_s = (D - L)^{-1}U.$$

$D - L$ è una matrice triangolare inferiore con elementi diagonali positivi e il resto negativi, questa matrice ha inversa non negativa.

Ma $D - L$ è una M-matrice in quanto L sottomatrice di B ($0 \leq L \leq B$) $\implies D - L = sId - \hat{B}$ con $0 \leq \hat{B} \leq B \implies \rho(\hat{B}) \leq \rho(B) < 1 \implies D - L$ non singular M-matrix \implies converge.

Osservazione 14. Le suddivisioni sono sempre fatte selezionando la diagonale e qualcosa di facilmente invertibile.

Osservazione 15. $\rho(H)$ dipende solo da $\rho(G)$ che dipende solo da $N \implies$ il metodo è più veloce se $\rho(N)$ è il minore possibile.

Corollario 2: Se $N_1 \leq N_2 \implies \rho(M_1^{-1}N_1) \leq \rho(M_2^{-1}N_2)$, $A^{-1}N_1 \leq A^{-1}N_2 \implies \rho(A^{-1}N_1) \leq \rho(A^{-1}N_2)$.

Questo risultato implica:

$$\rho(G_s) \leq \rho(J) \leq \rho(M_1^{-1}N_1)$$

Bisogna però mediare la velocità con il costo di inversione (o delle operazioni necessarie). Gauss - Seidel, Jacobi e l'ultimo costano circa uguali, se però le sottomatrici non sono triangolari bisogna stare attenti.

Studiamo il caso singolare.

Definizione 18. H si dice **semiconvergente** se $\exists \lim_{n \rightarrow \infty} H^n$.

Teorema 22. H è semiconvergente \iff valgono (tutte):

1. $\rho(H) \leq 1$.
2. L'unico autovalore di modulo 1 deve essere $\lambda = 1$.
3. Se $\rho(H) = 1 \implies 1$ deve essere semplice (semisingolo), $\mu_g(1) = \mu_a(1)$.

Possiamo determinare a cosa converge H^m ?

In forma di Jordan:

$$H = V \begin{pmatrix} Id_k & 0 \\ 0 & \hat{J} \end{pmatrix} V^{-1}$$

con $\rho(\hat{J}) < 1$. Dunque:

$V \begin{pmatrix} Id_k & 0 \\ 0 & 0 \end{pmatrix} V^{-1}$ è formata dagli autovettori relativi all'autovalore 1.

Data P transition matrix irriducibile, $\exists \lim_{n \rightarrow \infty} P^n$?

É semiconvergente?

Non è detto, come controesempio basta scegliere $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^2$, $\rho(P) = 1$; $\lambda = 1$

non è detto che sia 1.

Il punto 3 vale perché per Perron -Frobenius è semplice $\implies P^m =$

$V \begin{pmatrix} 1 & 0 \\ 0 & \hat{S}^m \end{pmatrix} V^{-1} = \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix} (\Pi_1 \quad \dots \quad \Pi_n) = \begin{pmatrix} \Pi^T \\ \dots \\ \Pi^T \end{pmatrix}$. Il caso 2, catena aperi-

odica (P primitiva).

Quindi questo è un metodo per il calcolo di Π^T se la catena è primitiva.

Per non calcolare P^m usiamo $x^{(k+1)T} = x^{(k)T} P$.

Vogliamo che converga a Π^T , come scegliere x_0^T , non vogliamo che tenda ad un autovettore di alto valore $\implies x^{(k)T} = x^{(0)T} P^k = x^{(0)T} \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix} \Pi^T$.

Vogliamo che il primo prodotto sia $\neq 0$.

Quindi $x^{(0)} \geq 0$ e $\|x^{(0)}\|_1 = 1$ la successione converge a Π^T già ortogonalizzata.

Si osservi che la **velocità di convergenza** è il $\rho(\hat{J})$, quindi è data dal valore del secondo autovalore.

Teorema 23. *A M-matrice, singolare e irriducibile. $A = M - N$ regolare splitting, $H = M^{-1}N$ allora:*

(Ricordiamo che vogliamo risolvere $Ax = 0$; $(Id - P^T)\Pi = 0$, $(M - N)\Pi = 0$; $\Pi^{(k+1)} = (M^{-1}N)x^{(k)}$, quindi converge?).

1. $\rho(H) = 1$.
2. $\lambda = 1$ è autovalore semplice di H .

Manca come condizione solo 1, l'unico di modulo 1.

2.5 Modificare il partizionamento

Per riepilogare quanto detto, A singular irriducibile M-matrix, $A = M - N$ regular splitting, $H = M^{-1}N$ valgono:

1. $\rho(H) = 1$
2. 1 è autovalore semplice di H

Osservazione 16. Si presti attenzione al fatto che possono esistere altri autovalori di modulo 1 (Se P è stocastica ed irriducibile, $Id - P^T$ potrebbe averne altri e quindi non essere semiconvergente).

Spesso si trova un partizionamento che converge più velocemente (ad esempio con il metodo delle potenze).

Esempio: $M = Id$; $N = P^T$; $H = M^{-1}N$ è semiconvergente se P è primitiva (Cioè la catena di Markov è non periodica).

L'idea è modificare il partizionamento per rendere semiconvergente i metodi come quello di Gauss - Seidel ($\epsilon = 0$).

Teorema 24. *A singular irreducible M-matrix $A = D - L - U; \epsilon > 0$. Allora: $M = D - L + \epsilon Id$, $N = U + \epsilon Id$ è un regular splitting semiconvergente. ($\epsilon = 0 \implies M^{-1}N = G$ Gauss - Seidel).*

Osservazione 17. Se ϵ è vicino a 0, allora $M_\epsilon^{-1}N_\epsilon$ hanno autovalori vicini ad 1 $\implies \epsilon$ troppo alto si perde consistenza con il problem, stessa cosa per α vicino ad 1 per il teorema successivo. Esistono però in questo caso degli α ottimali.

Teorema 25. *A M-matrix, singular e irriducibile. $A = M - n$ regular splitting, $H = M^{-1}N$, $\alpha \in \mathbb{R}; 0 < \alpha < 1$. $H_\alpha := (1 - \alpha)Id + \alpha H$. Allora H_α è semiconvergente.*

Osservazione 18. La soluzione H_α e H è la stessa.

$$Hx = x \iff H_\alpha x = (1 - \alpha)x + \alpha x = x$$

Un approccio distinto è il seguente:

$$(Id - P^T)\Pi = 0 \implies (Id - \hat{P}^T)\hat{\Pi} = \begin{pmatrix} 0 \\ \dots \\ 0 \\ 1 \end{pmatrix}. \text{ Ottenuta cancellando come}$$

prima l'ultima riga e l'ultima colonna, non è più una matrice singolare.

Osservazione 19. $(Id - P^T)\Pi = 0$; $(M - N)\Pi = 0$ regular splitting $M^{-1}N\Pi = \Pi$ ipotesi di semiconvergenza.

Quindi l'iterazione è:

$$\Pi^{(k+1)} = M^{-1}N\Pi^{(k)}; \Pi^{(k)} = (M^{-1}N)^k \Pi^{(0)}$$

Che converge alla matrice di rango 1 fatta da autovettori.

Scelta di $\Pi^{(0)}$? Vogliamo che converga alla soluzione non banale.

$$(M^{-1}N)^k \rightarrow |\Pi| \cdot (\geq 0 \dots \geq 0)\Pi_0.$$

Non vogliamo che il prodotto sia 0, basta dunque prendere $\Pi_0 > 0$. Dovremmo normalizzare ad ogni passo per evitare overflow.

$$\Pi^{(k+1)} = \frac{\Pi^{(k+1)}}{\|\Pi^{(k+1)}\|_1}$$

Chi è l'autovettore sinistro?

Sappiamo che $(1 \dots 1)(Id - P^T) = 0$ dunque:

$$1^T(M - N) = 0 \implies 1^T M (Id - M^{-1}N) = 0 \implies 1^T M \text{ è l'autovettore sinistro.}$$

2.6 Metodi non standard

Sia P stocastica, irriducibile e partizionabile a blocchi.

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1n} & \\ \cdots & & \cdots & \cdots \\ P_{m1} & & \cdots & P_{nn} \end{pmatrix}$$

con blocchi diagonali (imp).

Un metodo è l'IAD (**I**terative **A**ggregation/**D**isaggregation).

Sia $\Pi^T = (\Pi_1^T \ \cdots \ \Pi_n^T)$ con $\Pi^T P = \Pi^T$, $\Pi^T \mathbf{1} = \mathbf{1}$.

Teorema 26. $A = (a_{ij})$, $n \times n$, $a_{ij} = \phi_i^T P_{ij} \mathbf{1}$ con $\phi_i = \frac{\Pi_i}{\|\Pi_i\|_1}$.
Allora A è stocastica e irriducibile.

Osservazione 20. Ora come ora sembrerebbe che si debba conoscere Π , ossia che il problema debba essere già stato risolto.

Teorema 27. $\xi^T = (\|\Pi_1\|_1 \ \cdots \ \|\Pi_n\|_1)$

Teorema 28. $\Pi^T = (\xi_1 \phi_1^T \ \cdots \ \xi_n \phi_n^T)$

Con $\phi_i^T S_i = \phi_i^T$; $\phi_i^T \mathbf{1} = 1$.

Con S_i matrice di transizione ottenuta con censoring ristretto agli stati con indice i a blocchi.

Sia P_i la matrice ottenuta da P cancellando l' i -esima riga/colonna a blocchi.

Sia $P_{i*} = (P_{i,1} \ \cdots \ P_{i,i-1} \ P_{i,i+1} \ \cdots \ P_{i,N})$. Ottenuta eliminando l' i -esima colonna. P_{*i} idem sulla riga. Vale allora:

$$S_i = P_{i,i} + P_{i*} + (Id - P_i)^{-1} P_{*i}$$

È stato già visto che questa matrice è stocastica ed è il complemento di Schur generalizzato.

Quindi da questo caratterizziamo i $\|\Pi_i\|_1$ di prima.

In sintesi:

1. Si calcolano S_i e ϕ_i tali che rispettino $\phi_i^T S_i = \phi_i^T$ e $\phi_i^T \mathbf{1} = 1$.
2. Si costruisce A , $a_{i,j} = \phi_i^T P_{i,j} \mathbf{1}$.
3. Si calcolano $\xi^T A = \xi^T$, $\xi^T \mathbf{1} = 1$.
4. $\Pi^T = (\xi_1 \phi_1^T \ \cdots \ \xi_N \phi_N^T)$.

Si presti attenzione al fatto che per il calcolo di S_i servono n inversioni che costano $(N-1)^3 \implies \sim N^4$.

Senza metodi efficienti per il calcolo di S_i si preferisce un metodo iterativo.

2.7 M/G/1-type

La prima lettera rappresenta la legge di arrivo degli elementi della coda, M sta per Markovian.

La seconda il servizio reso, in questo caso G sta per General.

1-type, infine, indica che è presente una sola coda.

Sia $E = \mathbb{N} \times S$, $S = \{1, 2, \dots, m\}$. $m \geq 1, m \in \mathbb{N}$.

$(i, j) \in \mathbb{N} \times S$ si dicono rispettivamente livello e fase.

La matrice di transizione risulta $\mathbb{P} = \begin{pmatrix} B_0 & B_1 & B_2 & \cdots & \\ A_{-1} & A_0 & A_1 & \cdots & \\ 0 & A_{-1} & A_0 & A_1 & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots \end{pmatrix}$.

Dove le colonne sono dette **livelli**.

Con B_i, A_{i-1} matrici $m \times m$, $\sum_{i \geq 0} B_i; \sum_{i \geq -1} A_i$ stocastiche.

Questa matrice è di Hessenberg e Toeplitz a blocchi (Conta solo la differenza fra gli indici dei blocchi).

Osservazione 21. Una MG1 è una QBD (quasi Birth and Death) se la matrice è tridiagonale a blocchi ($B_i = A_i = 0 \forall i \geq 2$).

2.8 Condizioni di positiva ricorrenza

In questa sezione si assumono sempre valide le ipotesi:

1. $A = \sum_{i=-1}^{\infty} A_i$ irriducibile.
2. P irriducibile.

Teorema 29. Sia $\sum_{i=0}^{\infty} (i+1)A_i$ convergente. Sia $\underline{a} = \sum_{i=-1}^{\infty} iA_i e$ ($e = \underline{1}$). Sia $\underline{\alpha}$ (straight vector) tale che $\underline{\alpha}^T A = \underline{\alpha}^T$, $\underline{\alpha}^T e = 1$, $\mu = \underline{\alpha}^T \underline{a}$. Allora: La catena di Markov è **ricorrente** se $\mu \leq 0$, **transiente** se $\mu > 0$, **positive recurrent** se $\mu < 0$ e $\sum_{i \geq 0} iB_i$ è convergente.

Osservazione 22. M è detto **drift** (tendenza). Per Perron - Frobenius $\underline{\alpha} > 0$ quindi $\mu \equiv \underline{\alpha}^T \underline{a}$ e $\underline{a} = -A_{-1} + (A_1 + 2A_2 + 3A_3 + \dots)$. Siccome tutti i termini tranne il primo sono positivo, perché si allunghi la coda A_{-1} deve dominare.

In questo caso $P = \begin{pmatrix} b_0 & b_1 & 0 & & \\ a_{-1} & a_0 & a_1 & 0 & \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$.

Con $\underline{a} = -a_{-1} + a_1$, $\underline{\alpha} = 1$, $\underline{\mu} = a_1 - a_{-1} < 0$.

Sviluppiamo i conti per calcolare la soluzione:

$X = A_{-1} + A_0 X + A_1 X^2 + \dots$, $X = \sum_{i=-1}^{\infty} A_i X^{i+1}$ X matrice $m \times m$. Va risolto in quanto sono matrici e non commutano.

Teorema 30. $\exists!$ soluzione $G \geq 0$ tale che $G \leq T \forall Y$ matrice non negativa soluzione.

Vale $Ge \leq e$ e inoltre $Ge = e \iff$ la catena di Markov è ricorrente.

Sia $G_{i,j} = P\{\text{passare dal livello } i \text{ e fase } j \text{ al livello } 0 \text{ e fase } j \text{ in tempo finito}\}$.

Avendo così definito G allora se ricorrente si arriva in ogni fase in tempo finito \implies stocastica.

Osservazione 23. $(\lambda Id - \sum_{i=-1}^{\infty} \lambda^{i+1} A_i)v = 0$.

Definiamo $A(z) := zId - \sum_{i=-1}^{\infty} z^{i+1} A_i$, $a(z) = \det A(z)$.

Allora $Gv = \lambda v \implies a(\lambda) = 0$ e $A(\lambda)v = 0$.

Osservazione 24. La serie è analitica per $|z| \leq 1$ e converge per $z = 1$ e siccome i termini sono positivi, converge se $|z| \leq 1$.

Se $\rho(G) \leq 1 \implies a(z)$ ha almeno m radici con $|\lambda| \leq 1$. Ce ne sono altre?

Teorema 31. 1. Se $\mu < 0 \implies a(z)$ ha $m - k$ radici di modulo < 1 e k radici semplici fra le radici k -esime di 1.

2. Se $\mu > 0 \implies a(z)$ ha m radici di modulo < 1 e k radici uguali alle radici k -esime di 1.

3. Se $\mu = 0 \implies a(z)$ ha $m - k$ radici di modulo < 1 e k radici doppie alle radici k -esime di 1 (Ogni radice k -esima di 1 è doppia).

Corollario 3: Gli autovalori di G sono le m radici di modulo minimo della funzione $a(z)$.

Capitolo 3

Metodi di calcolo e iterazioni

3.1 Equazione Matriciale

Sviluppiamo i conti sugli oggetti incontrati prima.

Sia $G = \sum_{i=-1}^{\infty} A_i G^{i+1}$, $G \geq 0$, $Ge \leq e$.

Siamo nel caso Positive Recurrent ($\mu < 0$, $\sum_{i \geq 0} iB_i < \infty$).

$$(\Pi_0^T \quad \Pi_1^T \quad \dots) \begin{pmatrix} Id - B_0 & -B_1 & -B_2 & \dots \\ -A_{-1} & Id - A_0 & \dots & \dots \\ 0 & \dots & \dots & \dots \end{pmatrix} = 0$$

Supponiamo che Π_0 sia noto, vogliamo determinare il resto del vettore. Allora:

$$(\Pi_1^T \quad \Pi_2^T \quad \dots) \begin{pmatrix} Id - A_0 & -A_1 & -A_2 & \dots \\ -A_{-1} & Id - A_0 & \dots & \dots \\ 0 & \dots & \dots & \dots \end{pmatrix} = \Pi_0^T (B_1 \quad B_2 \quad \dots)$$

Rispetto al sistema iniziale questo è non singolare e non abbiamo condizioni al contorno (Invece di Hessenberg è di Toeplitz a blocchi).

Scriviamo la serie di Laurent della matrice:

$$S(z) = Id - \sum_{i=-1}^{\infty} z^i A_i = (Id - \sum_{i=0}^{\infty} A_i^*) (Id - z^{-1} G)$$

Con $A_i^* = \sum_{j=i}^{\infty} A_j$.

Quindi otteniamo una fattorizzazione LU :

$$(\Pi_1^T \quad \Pi_2^T \quad \dots) \begin{pmatrix} Id - A_0^* & -A_1^* & -A_2^* & \dots \\ 0 & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} Id & 0 \\ -G & Id \\ 0 & \dots & 0 \\ 0 & \dots & -G & Id \end{pmatrix}$$

Osservazione 25. L è invertibile.

Allora:

$$\Pi_0^T (B_1 \ B_2 \ \cdots) \begin{pmatrix} Id & & & \\ G & Id & & \\ G^2 & G & Id & \\ G^3 & \cdots & \cdots & \cdots \end{pmatrix} = \Pi_0 (B_1^* \ B_2^* \ \cdots)$$

Con $B_i^* = \sum_{j=i}^{\infty} B_j G^{i-j}$.

Abbiamo quindi un termine noto calcolabile, i termini Π_i^T si ricavano per sostituzione, essendo la matrice triangolare. L'aspetto interessante è che possiamo continuare a calcolare i termini ma non sono mai approssimazioni.

La formula ricavata è stabile e vale la seguente regola ricorsiva $A_{i-1}^* = A_i^* G + A_{i-1}$.

Si presti attenzione al fatto che tutto questo è stato fatto sotto le ipotesi che conosciamo Π_0 .

3.1.1 Calcolo di Π_0

Si ricava come lo Straight Vector della matrice Q ($\Pi_0^T (Id - Q) = 0$) con:

$$Q = B_0 + (B_1 \ B_2 \ \cdots) \begin{pmatrix} Id - A_0 & -A_1 & & \\ -A_{-1} & Id - A_0 & \cdots & \\ 0 & \cdots & \cdots & \cdots \end{pmatrix}^{-1} \begin{pmatrix} A_{-1} \\ 0 \\ \cdots \\ 0 \end{pmatrix}$$

Questa catena è finita e ricaviamo Π_0 con la fattorizzazione UL vista prima da cui otteniamo:

$$Q = B_0 + B_1^* (Id - A_0^*)^{-1} A_{-1}$$

Un problema è dato dal fatto che questo è vero a condizione di normalizzare il tutto. Non potendo effettuare una normalizzazione (matrice infinita) servono delle proprietà teoriche che lo garantiscano.

Senza addentrarci nel meccanismo (molto tecnico) si ottengono delle condizioni per le quali Π è normalizzato.

3.2 Metodi di iterazione funzionali

3.2.1 Metodo naturale N

$$\begin{cases} X_{k+1}^{(N)} = \sum_{i=-1}^{\infty} A_i X_k^{(i+1)(N)} \\ X_0 = 0 \end{cases}$$

3.2.2 Metodo tradizionale T

$$\begin{cases} X_{k+1}^{(T)} = (Id - A_0)^{-1}(A_{-1} + \sum_{i=1}^{\infty} A_i X_k^{(i+1)(T)}) \\ X_0 = 0 \end{cases}$$

3.2.3 Metodo a U-based U

Posto $U := Id - \sum_{i=0}^{\infty} A_i G^i$

$$\begin{cases} X_{k+1}^{(U)} = (Id - \sum_{i=0}^{\infty} A_i X_k^i)^{(U)} A_{-1} \\ X_0 = 0 \end{cases}$$

3.2.4 Confronto

Tutte e tre convergono in modo monotono a G . E $X_{k+1}^{(*)} \geq X_k^{(*)} \geq 0$.

Perché non usare allora solo la Natural?

La riduzione asintotica del passo è molto maggiore in U e T .

Teorema 32 (Qualitativo). $\forall k X_k^{(N)} \leq X_k^{(T)} \leq X_k^{(U)}$

Siccome la convergenza è monotona crescente, maggiore significa anche più accurato.

Un risultato quantitativo invece è il seguente:

Sia $E_n = G - X_n^{(*)} \geq 0$ successioni monotone.

Stimiamo $E_{k+1}^{(N)} = G - X_{k+1}^{(N)} = \sum_{i=0}^{\infty} A_i (G^{i+1} - X_k^{i+1})$. Quindi l'errore al $k+1$ esimo passo è legato a quello del k -esimo. Vale:

$$\epsilon_{k+1} = R_k R_{k-1} \cdots R_0 \epsilon_0; R = \lim_{k \rightarrow \infty} R_k.$$

Si può dimostrare l'uguaglianza:

$$\lim_{k \rightarrow \infty} \sqrt[k]{\|E_k\|} = \rho(R)$$

Come è fatta R ?

N) $\sum_{i=0}^{\infty} A_i^*$; $M = Id$.

T) $(Id - A_0)^{-1}(\sum_{i=0}^{\infty} A_i^* - A_0)$; $M = Id - A_0$.

U) $(Id - A_0^*)^{-1} \sum_{i=1}^{\infty} A_i^*$; $M = Id - A_0^*$.

Osservazione 26. $R^{(*)} = (M^{(*)})^{-1} N^{(*)}$

$$M^{(*)} - N^{(*)} = Id - \sum_{i \geq 0} A_i^*$$

Sono tutte regular splitting della stessa matrice. Allora, per il teorema del confronto (regular splitting) più piccola è la R migliore è la convergenza (ρ).

Cosa cambia se viene variato X_0 ?

3.3 Metodi a convergenza quadratica

Quelli noti sono varianti del metodo di Newton. Si cerca di costruirne qualcuno da un'idea di base un po' diversa.

Lavoriamo nel caso quasi Birth and Death (QBD).

$$G = A_{-1} + A_0G + A_1G^2.$$

Definiamo R come la minima soluzione ≥ 0 di $X = X^2A_{-1} + XA_0 + A_1$ (In questa maniera è al contrario e anche in una matrice ∞ possiamo effettuare al sostituzione).

Vale:

1. $\rho(R) < 1 \iff \mu < 0$
2. $\rho(R) = 1 \iff \mu \leq 0$

$$Id - \sum_{i=-1}^1 z^i A_i = (Id - A_1^*(Id - A_0^*)^{-1}z)(Id - A_0^*)(Id - Z^{-1}G) = (Id - Rz)K(Id - Z^{-1}G).$$

Le matrici * le sappiamo calcolare da G , quindi R si ottiene da G e viceversa.

3.3.1 Algoritmo di logarithmic reduction (LR)

Sempre nel caso QBD.

$$G = A_{-1} + A_0G + A_1G^2$$

$$(Id - A_0)G = A_{-1} + A_1G$$

$$G = B_{-1} + B_1G^2 \text{ con } B_{-1} = (Id - A_0)^{-1}A_{-1} \text{ e } B_1 = (Id - A_0)^{-1}A_1.$$

Da cui la formula:

$$G^{2^k} = B_{-1}^{(k)} + B_1^{(k)}G^{2^{k+1}}$$

Teorema 33 (Convergenza). $\mu_0, \|\dots\|$ norma matriciale indotta, $\epsilon > 0$ tale che $\epsilon + \eta < 1$ con $\eta = \max\{z \in \mathbb{C}; |z| < 1; \det Id - (A_{-1} + zA_0 - z^2A_1) = 0\}$ è il secondo autovalore di modulo massimo di G . Allora:

1. $\exists \gamma > 0; \|B_1^{(k)}\| \leq \gamma \rho(R)^{2^k}$ (Doppiamente esponenziale).
2. $\|B_{-1}^{(k)} - eg^T\| \leq \gamma(\rho(R)^{2^k} + (\eta + \epsilon)^{2^k})$ con $g^T G = g^T; g^T e = 1$.

Costo e stabilità?

É stabile ma il costo è circa quattro volte quello delle iterazioni, talvolta però è molto veloce.

3.4 Algoritmo di Cyclic - Reduction

Rispetto all'algoritmo precedente la permutazione è differente.

$$A_{-1} + (Id - A - 0)G - A_1G^2 = 0$$

$$\begin{pmatrix} Id - A_0 & -A_1 & 0 & & \\ -A_{-1} & Id - A_0 & -A_1 & 0 & \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix} \begin{pmatrix} G \\ G^2 \\ G^3 \\ \dots \end{pmatrix} = \begin{pmatrix} A_{-1} \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

Stavolta la permutazione è dispari/pari:

$$\begin{pmatrix} C & D \\ E & F \end{pmatrix} \begin{pmatrix} G^2 \\ G^4 \\ \dots \\ \dots \\ G \\ G^3 \\ \dots \\ \dots \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ A_{-1} \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

Con $C, F = \text{diag}(Id - A_0)$, $D = \begin{pmatrix} -A_1 & -A_1 & 0 \\ 0 & -A_{-1} & -A_1 & 0 \\ 0 & 0 & \dots & \dots \end{pmatrix}$,

$$E = \begin{pmatrix} -A_1 & 0 \\ -A_{-1} & -A_1 & 0 \\ 0 & \dots & \dots & 0 \end{pmatrix}$$

Applichiamo come prima l'eliminazione gaussiana.

$$(F - EC^{-1}D) = \begin{pmatrix} G \\ G^3 \\ G^5 \\ \dots \end{pmatrix} = \begin{pmatrix} A_{-1} \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

Come cambia la matrice del complemento di Schur?

Stavolta non sarà Toeplitz a causa dell'elemento di posto (1, 1).

Se come prima scriviamo:

$$\begin{pmatrix} Id - \hat{A}_0^{(k)} & -A_1^{(k)} & 0 & & \\ -A_{-1}^{(k)} & Id - A_0^{(k)} & -A_1^{(k)} & \dots & \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}$$

Quindi questa volta $A_1^{(k)} \rightarrow 0$ quadraticamente.

Invertendo la matrice si ricava direttamente $\begin{pmatrix} A_{-1} \\ 0 \\ \dots \\ 0 \end{pmatrix}$.

Inoltre:

$$G^{(k)} = (Id - \hat{A}_0^{(k)})^{-1} A_{-1}$$

Come prima $G^{(k)} \rightarrow G$ come $\rho(R)^{2^k}$.

3.4.1 Formule

$$\begin{aligned} \hat{A}_0^{(k+1)} &= \hat{A}_0^{(k)} + A_1^{(k)}(Id - A_0^{(k)})^{-1}A_{-1}^{(k)} \\ A_{-1}^{(k+1)} &= A_{-1}^{(k)}(Id - A_0^{(k)})^{-1}A_{-1}^{(k)} \\ A_1^{(k+1)} &= A_1^{(k)}(Id - A_0^{(k)})^{-1}A_1^{(k)} \\ A_0^{(k+1)} &= A_0^{(k)} + A_1^{(k)}(Id - A_0^{(k)})^{-1}A_{-1}^{(k)} + A_{-1}^{(k)}(Id - A_0^{(k)})^{-1}A_1^{(k)} \end{aligned}$$

3.4.2 Costo, stabilità e convergenza

1 inversione e 6 moltiplicazioni.

Analoga all'algoritmo precedente la stabilità e la convergenza, sebbene non sia stato dimostrato, nella pratica è un po' più stabile di quello precedente. A differenza dell'algoritmo *LR* si può estendere ad equazioni di grado arbitrario.

3.5 Variante

Sia $G = \sum_{i=-1}^{\infty} A_i G^{i+1}$; $(G - \sum_{i \geq -1} A_i G^i)G^k = 0$.

$$\begin{pmatrix} Id - A_0 & -A_1 & -A_2 & \dots \\ -A_{-1} & Id - A_0 & -A_1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} G \\ G^2 \\ G^3 \\ \dots \end{pmatrix} = \begin{pmatrix} A_{-1} \\ 0 \\ \dots \end{pmatrix}$$

Scrivendolo permutando:

$$\begin{pmatrix} C & D \\ E & F \end{pmatrix} \begin{pmatrix} G^2 \\ G^4 \\ \dots \\ G \\ G^3 \\ \dots \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \\ 0 \\ A_{-1} \\ 0 \\ \dots \end{pmatrix}$$

Con E triangolare superiore di diagonali $(Id - A_0), -A_2, -A_4, \dots$.

D triangolare superiore di diagonali $(-A_{-1}, -A_1, -A_3, \dots)$.

E triangolare superiore di diagonali $(-A_1, -A_3, -A_5, \dots)$ e sottodiagonale $-A_{-1}$. F triangolare superiore di diagonali $(Id - A_0, -A_2, -A_4, \dots)$.

Il cui complemento di Schur è dato da:

$$(F - EC^{-1}D) \begin{pmatrix} G \\ G^3 \\ G^5 \\ \dots \end{pmatrix} = \begin{pmatrix} A_{-1} \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

Si può scrivere una iterazione stabile per il calcolo degli A_i (Incompleto)

3.6 Metodo di Newton

Vogliamo estendere il metodo di Newton al nostro problema: ricerca di uno 0 di $X = \sum_{i=1}^{\infty} A_i X^{i+1}$.

Il primo passo consiste nel troncamento della serie con un polinomio $f(x) = \sum_{i=0}^p B_i X^i$ con $B_0 = -A_{-1}$; $B_1 = Id - A_0$, $B_i = -A_{i-1}$.

Il secondo passo è la ricerca di G soluzione minima (≥ 0) di $f(x) = 0$.

Nel caso di Newton (scalare) valeva la relazione:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Introduciamo la definizione di **derivata di Gateux o Frechet**.

Definizione 19. $g(x) = AX^r$, A, X matrici $m \times m$, la derivata di Frechet di $g(x)$ lungo la direzione $H \in \mathbb{R}^{m \times m}$ è

$$g'_H(x) = \lim_{t \rightarrow 0} \frac{g(x + tH) - g(x)}{t}$$

Quanto vale la derivata nel nostro caso (prestando attenzione al fatto che non commutano)?

$$\lim_{t \rightarrow 0} \frac{A(X + tH)^r - AX^r}{t}$$

Otteniamo quindi una funzione (lineare rispetto ad H):

$$g'_H(x) = A \sum_{j=0}^{r-1} X^j H X^{r-1-j}$$

A questo punto, detto (per linearità della derivata) $f'_H(x) = \sum_{i=1}^p B_i \sum_{j=0}^{i-1} X^j H X^{i-1-j}$, dobbiamo “invertirla” per scrivere il metodo di Newton.

La relazione è dunque:

$$X_{k+1} = X_k - F'_*(X_k)^{-1} F(X_k)$$

Lo * rappresenta la direzione, non ancora nota. Equivalente:

$$X_{k+1} - X_k = -F'_0(X_k)^{-1} F(X_k)$$

equivalente

$$F'(X_k)(X_{k+1} - X_k) = -F(X_k)$$

Poniamo $H_k = X_{k+1} - X_k$, quindi $X_{k+1} = X_k + H_k$ (Incremento di Newton) dove H_k è tale che $F'_{H_k}(X_k) = -F(X_k)$.

Osservazione 27. H_k risolve questa equazione (non banale):

$$\sum_{i=1}^p B_i \sum_{j=0}^{i-1} X_k^j Y X_k^{i-1-j} = -F(X_k)$$

Con Y l'incognita.

O vantaggi di questo metodo riguardano la convergenza molto più forte che locale. Lo svantaggio riguarda il dover risolvere quella equazione.

Teorema 34 (Convergenza). *Se $0 \leq X_0 \leq G \implies X_k \leq X_{k+1} \leq G$. Dunque la successione converge crescendo a G .*

Se $\mu \neq 0 \implies$ convergenza di ordine 2 (quadratica).

Per risolvere l'equazione la rappresentiamo come $\sum_{i=1}^p U_i Y V_i = -F(X_k)$ ed esistono diversi metodi per risolverla.