

7 Approfondimenti di algebra lineare e non lineare numerica

aggiornamento: 14 aprile 2014

(!) indica un argomento fondamentale, (F) un argomento facoltativo, (*) un argomento o dimostrazione impegnativi, (NR) una dimostrazione non richiesta; per approfondimenti di analisi numerica, si vedano ad es. V. Comincioli: *Analisi Numerica* - McGraw-Hill cartaceo e Apogeo e-book, G. Rodriguez: *Algoritmi Numerici* - Pitagora, e le dispense online di A. Sommariva in <http://www.math.unipd.it/~alvise/didattica> con le referenze citate; per approfondimenti di analisi funzionale si veda ad es. A.N. Kolmogorov, S.V. Fomin: *Elementi di teoria delle funzioni e di analisi funzionale* - Mir; per le basi sulle differenze finite per equazioni differenziali si veda ad es. A. Quarteroni, F. Saleri: *Introduzione al calcolo scientifico* - Springer.

7.1 Algebra lineare numerica

1. (!) *teorema fondamentale di invertibilità*: data una norma matriciale in $\mathbb{C}^{m \times m}$ tale che $\|AB\| \leq \|A\| \|B\|$ (come sono tutte le norme indotte da norme vettoriali, ovvero $\|A\| = \sup_{x \neq 0} \|Ax\|/\|x\| = \sup_{\|x\|=1} \|Ax\|$), se $\|A\| < 1$ allora $I - A$ è invertibile e si ha

$$(I - A)^{-1} = \sum_{j=0}^{\infty} A^j, \quad \|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}$$

(sugg.: $\|\sum_{j=m}^n A^j\| \leq \sum_{j=m}^n \|A\|^j$ e la serie geometrica $\sum \|A\|^j$ è convergente e quindi di Cauchy, ... ; nel caso $\|A\|$ sia indotta da una norma vettoriale, la dimostrazione di invertibilità e la stima si possono fare in modo più diretto osservando che $\|(I - A)x\| \geq \|x\| - \|Ax\| > 0$, ..., e detta $S = (I - A)^{-1}$ si ha $1 = \|S(I - A)\| \geq \|S\| - \|AS\| \geq \dots$).

2. si verifichi che $\|A\| = \max_{i,j} |a_{ij}|$ è una norma matriciale ma non soddisfa la disuguaglianza $\|AB\| \leq \|A\| \|B\|$ per ogni coppia di matrici.
3. localizzazione rozza degli autovalori: data una norma matriciale come sopra, gli autovalori di $A \in \mathbb{C}^{m \times m}$ stanno in $\mathcal{C}[0, \|A\|]$ (il cerchio complesso chiuso di centro 0 e raggio $\|A\|$).
(sugg.: se $\lambda \in \mathbb{C}$, $|\lambda| > \|A\|$, scrivendo $(\lambda I - A) = \lambda(I - A/\lambda)$, ...).
4. (!) localizzazione fine degli autovalori (*teorema dei cerchi Gershgorin*): gli autovalori di $A \in \mathbb{C}^{m \times m}$ stanno in $\bigcup_{i=1}^m \mathcal{C}[a_{ii}, \sum_{j \neq i} |a_{ij}|]$.
(sugg.: se $\lambda \in \mathbb{C}$ e $\lambda \notin \bigcup_{i=1}^m \mathcal{C}[a_{ii}, \sum_{j \neq i} |a_{ij}|]$, scrivendo $A = D + E$, dove D è la matrice diagonale che coincide con la parte diagonale di A , si ha $\lambda I - A = (\lambda I - D) - E = (\lambda I - D)(I - (\lambda I - D)^{-1}E)$, dove $\|(\lambda I - D)^{-1}E\|_{\infty} < 1$, ...).

5. si deduca dal teorema di Gershgorin che una matrice quadrata *diagonalmente dominante in senso stretto*, ovvero tale che $|a_{ii}| > \sum_{j \neq i} |a_{ij}| \forall i$, è invertibile.
6. applicazione (stime di condizionamento): dato il sistema quadrato $Ax = b$ con $\det(A) \neq 0$ e il sistema perturbato $(A + \delta A)(x + \delta x) = b + \delta b$, se $\|k(A)\| \|\delta A\| < \|A\|$ (dove $k(A) = \|A\| \|A^{-1}\|$ è l'*indice di condizionamento* della matrice in una norma matriciale indotta da una norma vettoriale), vale la stima

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{k(A)}{1 - k(A)\|\delta A\|/\|A\|} \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right)$$

(sugg.: partendo da $(A + \delta A)\delta x = \delta b - \delta Ax$ e osservando che $(A + \delta A) = A(I + A^{-1}\delta A)$ con $\|A^{-1}\delta A\| \leq \|A^{-1}\| \|\delta A\| < 1$, ...).

7. applicazione (cond. suff. per la convergenza delle approssimazioni successive): un sistema quadrato della forma $x = Bx + c$ con $\|B\| < 1$ (in una norma matriciale indotta da una norma vettoriale) ha soluzione unica che si può ottenere come limite della successione di approssimazioni successive $x_{n+1} = Bx_n + c$ $n \geq 0$, a partire da un qualsiasi vettore iniziale x_0 .
(sugg.: il sistema ha soluzione unica se e solo se $I - B$ è invertibile, ...).

8. (!) *cond. nec. e suff. per la convergenza delle approssimazioni successive*: il metodo delle approssimazioni successive $x_{n+1} = Bx_n + c$, $n > 0$, converge alla soluzione di $x = Bx + c$ per qualsiasi scelta dei vettori x_0 e c se e solo se $\rho(B) < 1$ (dove $\rho(B)$ è il *raggio spettrale* della matrice quadrata B , ovvero il max dei moduli degli autovalori).

(sugg.: supponendo per semplicità che B sia diagonalizzabile, $B = Q^{-1}\Lambda Q$ dove Λ è la matrice diagonale degli autovalori di A ((NR) il caso generale che si può trattare con la forma canonica di Jordan), si avrà $\sum B^j = Q^{-1}(\sum \Lambda^j)Q$, ...).

9. (!) dato uno splitting di una matrice quadrata, $A = P - N$, con $\det(P) \neq 0$, il sistema $Ax = b$ si può scrivere nella forma $x = Bx + c$ dove $B = P^{-1}N$ e $c = P^{-1}b$. Esempi di corrispondenti metodi delle approssimazioni successive nell'ipotesi $a_{ii} \neq 0 \forall i$ sono (posto $A = D - (E + F)$, dove D è la parte diagonale di A ed $-E, -F$ le parti triangolare inferiore e superiore di $A - D$)

- il metodo di Jacobi: $P = D, N = E + F$
- il metodo di Gauss-Seidel: $P = D - E, N = F$

Si scrivano per componenti tali metodi, e si dimostri che il metodo di Jacobi è convergente per matrici diagonalmente dominanti in senso stretto. Si può dimostrare (NR) che anche il metodo di Gauss-Seidel converge in tale ipotesi nonché per matrici simmetriche definite positive.

10. il metodo delle approssimazioni successive si può riscrivere come

$$x_{n+1} = (I - P^{-1}A)x_n + P^{-1}b = x_n + P^{-1}r(x_n)$$

(dove $r(x_n) = b - Ax_n$ è il vettore *residuo* al passo n -esimo). Il ruolo della matrice P^{-1} può essere visto come quello di *precondizionatore*: l'azione di P^{-1} è efficace quando $P^{-1} \approx A^{-1}$, nel senso che gli autovalori di $P^{-1}A$ si accumulano intorno ad 1 (e nel contempo dato un vettore v , il calcolo di $z = P^{-1}v$, ovvero la soluzione del sistema $Pz = v$, ha basso costo computazionale). Vari metodi introducono un parametro di rilassamento α ,

$$x_{n+1} = x_n + \alpha P^{-1}r(x_n)$$

che aumenti l'efficacia del preconditionatore (cercando di diminuire o addirittura minimizzare il raggio spettrale di $B(\alpha) = I - \alpha P^{-1}A$).

11. (!) metodi di discesa: risolvere un sistema lineare $Ax = b$ con A *simmetrica e definita positiva* è equivalente a risolvere il problema di minimo globale

$$\min_{x \in \mathbb{R}^m} F(x), \quad F(x) = \frac{1}{2} x^t A x - x^t b$$

(sugg.: $\nabla x^t A x = Ax + A^t x, \dots$). I metodi di discesa corrispondono a costruire un'iterazione del tipo

$$x_{n+1} = x_n + \alpha_n d_n, \quad n \geq 0$$

per diverse scelte della direzione di discesa d_n , dove il parametro α_n viene scelto in modo di minimizzare $F(x_{n+1})$. Si mostri che

$$\alpha_n = \frac{d_n^t r(x_n)}{d_n^t A d_n}$$

Il *metodo del gradiente* corrisponde ad usare la direzione di discesa localmente più ripida, $d_n = -\nabla F(x_n) = r(x_n)$. Si può dimostrare (NR) che per il metodo del gradiente vale la stima dell'errore $\|x - x_n\|_2 = \mathcal{O}(\theta^n)$, dove $\theta = (k_2(A) - 1)/(k_2(A) + 1) < 1$; per matrici mal condizionate $\theta \approx 1$, diventa quindi importante utilizzare un buon preconditionatore, che è equivalente (NR) nel caso P^{-1} sia simmetrica e definita positiva ad applicare il metodo al sistema $P^{-1}Ax = P^{-1}b$ con $k_2(P^{-1}A) \ll k_2(A)$ (pur non essendo in generale la matrice $P^{-1}A$ simmetrica e definita positiva).

12. test di arresto dello step: l'errore del metodo delle approssimazioni successive con $\rho(B) < 1$, supposta B diagonalizzabile (ovvero $B = Q^{-1}\Lambda Q$ con Λ matrice diagonale degli autovalori di B) si può stimare come

$$\|x - x_n\| \leq \frac{k(Q)}{1 - \rho(B)} \|x_{n+1} - x_n\|$$

purché la norma matriciale indotta dalla norma vettoriale soddisfi $\|D\| = \max\{|d_{ii}|\}$ per qualsiasi matrice diagonale. Si faccia un esempio in cui ha senso arrestare le iterazioni quando lo step $\|x_{n+1} - x_n\| \leq \varepsilon$, dove ε è una tolleranza prefissata (sugg.: se B è simmetrica, ...).

13. (!) test di arresto del residuo: dato un qualsiasi metodo iterativo *convergente* per la soluzione di un sistema lineare non singolare $Ax = b$ con $b \neq 0$ (approssimazioni successive, gradiente, ...), si mostri che vale la seguente stima dell'errore relativo

$$\frac{\|x - x_n\|}{\|x\|} \leq k(A) \frac{\|r(x_n)\|}{\|b\|}$$

14. *teorema di Bauer-Fike* (sul condizionamento del problema degli autovalori): data una matrice complessa diagonalizzabile $A \in \mathbb{C}^{m \times m}$, $A = Q^{-1}\Lambda Q$ con Λ matrice diagonale degli autovalori di A , e una perturbazione matriciale E , detto μ un autovalore fissato di $A + E$, si ha la stima (NR)

$$\min_{\lambda \in \sigma(A)} |\lambda - \mu| \leq k_2(Q) \|E\|_2$$

dove $\sigma(A)$ è lo spettro di A e $k_2(Q) = \|Q\|_2 \|Q^{-1}\|_2$ (da cui si vede che il problema degli autovalori per una matrice hermitiana è ottimamente condizionato). (sugg.: se μ autovalore di $A + E$ non è autovalore di A , $A - \mu I$ è non singolare e $A + E - \mu I$ è invece singolare: allora esiste un vettore $z \neq 0$ tale che $Q(A + E - \mu I)Q^{-1}z = (\Lambda - \mu I + QEQ^{-1})z = 0$ e raccogliendo $\Lambda - \mu I$ si ottiene $z = -(\Lambda - \mu I)^{-1}QEQ^{-1}z$, da cui $\|z\|_2 \leq \dots$).

15. (!) *metodo delle potenze* per il calcolo degli autovalori estremali: data una matrice complessa diagonalizzabile $A \in \mathbb{C}^{m \times m}$, con un unico autovalore di modulo massimo (di qualsiasi molteplicità), e la successione di vettori $x_{n+1} = Ax_n$, $n \geq 0$ dove x_0 abbia componente non nulla nell'autospazio corrispondente, i *quozienti di Rayleigh* $R(x_n) = (Ax_n, x_n)/(x_n, x_n)$ (dove (x, y) è il prodotto scalare euclideo di $x, y \in \mathbb{C}^m$) convergono a tale autovalore e $x_n/\|x_n\|_2$ tende ad un autovettore (normalizzato) corrispondente; come si può modificare il metodo per calcolare l'autovalore di modulo minimo quando A è non singolare? (sugg.: si scriva x_0 nella base di autovettori, ...; per l'autovalore di modulo minimo, si osservi che gli autovalori di A^{-1} sono ...).

Il metodo modificato $z_{n+1} = Ay_n$, $y_{n+1} = z_{n+1}/\|z_{n+1}\|_2$, $n \geq 0$ a partire da $y_0 = x_0$ evita overflow e underflow in aritmetica di macchina quando l'autovalore di modulo massimo è molto grande o molto piccolo (si mostri che $y_n = x_n/\|x_n\|_2$). Perché facendo una scelta random di x_0 ci si aspetta comunque convergenza in aritmetica di macchina?

(F) cosa succede se l'autovalore di modulo massimo non è unico? come si può modificare il metodo per calcolare l'autovalore più vicino ad un valore prefissato?

16. *metodo QR* per il calcolo dell'intero spettro: se gli autovalori di A sono tutti distinti in modulo, si può dimostrare (NR) che la successione di matrici $\{A_n\}$ definita da

$$A_n = Q_n R_n, \quad A_{n+1} = R_n Q_n, \quad n \geq 0; \quad A_0 = A$$

dove Q_n è ortogonale (unitaria nel caso complesso) ed R_n triangolare superiore, converge ad una matrice triangolare T ; perchè T ha gli stessi autovalori di A ?

(e quindi la diagonale di A_n converge agli autovalori di A). Si osservi che se A è simmetrica tali sono le A_n da cui si vede che T è una matrice diagonale. Il metodo può essere adattato al caso in cui ci siano autovalori con lo stesso modulo (NR).

17. dato un polinomio $p(\lambda) = a_0 + a_1\lambda + \dots + a_m\lambda^m$, $a_m \neq 0$, si vede facilmente per induzione che la matrice $(m+1) \times (m+1)$

$$C(p) = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0/a_m \\ 1 & 0 & \dots & 0 & -a_1/a_m \\ 0 & 1 & \dots & 0 & -a_2/a_m \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{m-1}/a_m \end{pmatrix}$$

detta “*matrice companion*” di p , ha polinomio caratteristico $\det(\lambda I - C(p)) = p(\lambda)/a_m$. È quindi possibile calcolare tutti gli zeri di un polinomio applicando il metodo QR (modificato per moduli non distinti) alla matrice companion.

7.2 Algebra non lineare numerica

1. sia $\phi : (K \subseteq \mathbb{R}^m) \rightarrow K$, dove K è un sottoinsieme chiuso (anche non limitato), una mappa tale che $\|\phi(x) - \phi(y)\| \leq \theta\|x - y\|$, $0 < \theta < 1$, per una qualche norma in \mathbb{R}^m (*contrazione*): allora per il *metodo delle approssimazioni successive* (iterazione di punto fisso)

$$x_{n+1} = \phi(x_n), \quad x_0 \in K$$

vale la disuguaglianza fondamentale

$$\|x_m - x_n\| \leq (1 + \theta + \dots + \theta^{m-n-1}) \|x_{n+1} - x_n\|, \quad \forall m > n$$

(sugg.: $x_m - x_n = x_m - x_{m-1} + x_{m-1} - x_{m-2} + \dots + x_{n+2} - x_{n+1} + x_{n+1} - x_n = \phi(x_{m-1}) - \phi(x_{m-2}) + \dots + \phi(x_{n+1}) - \phi(x_n) + x_{n+1} - x_n, \dots$).

2. (!) *teorema delle contrazioni*: dato il sistema di punto fisso $x = \phi(x)$ con ϕ contrazione di K in K , il metodo delle approssimazioni successive converge, per qualsiasi scelta di $x_0 \in C$, all’unico $\xi \in K$ tale che $\xi = \phi(\xi)$.
(sugg.: dalla disuguaglianza fondamentale e da $\|x_{n+1} - x_n\| \leq \theta^n \|x_1 - x_0\|$, si ricava che la successione $\{x_n\}$ è di Cauchy, ...; si osservi che l’enunciato è valido in qualsiasi spazio *normato completo*).

3. una condizione sufficiente affinché ϕ sia una contrazione in $\|\cdot\|_\infty$ è che sia di classe $C^1(K)$, con K chiusura di un aperto convesso e $\sup_{x \in K} \|J\phi(x)\|_\infty < 1$, dove $J\phi$ è la matrice jacobiana di ϕ (si utilizzi per componenti il teorema del valor medio in più variabili).

4. un sistema lineare quadrato della forma $x = Bx + c$ con $\|B\| < 1$ (in una norma matriciale indotta da una norma vettoriale) è un caso particolare di sistema di punto fisso con una contrazione $\phi(x) = Bx + c$ definita su $K = \mathbb{R}^m$.
5. convergenza locale: se ϕ è di classe C^1 in un intorno di ξ punto fisso e $\|J\phi(\xi)\|_\infty < 1$, allora il metodo delle approssimazioni successive converge localmente a ξ (sugg.: prendendo come K una opportuna palla chiusa centrata in ξ dove $\|J\phi(x)\|_\infty < 1$, ...).
6. un altro risultato di convergenza locale: se $\phi \in C^1(B_\infty[x_0, r])$ (la palla chiusa di centro x_0 e raggio r in $\|\cdot\|_\infty$) e $\theta = \max_{x \in B_\infty[x_0, r]} \|J\phi(x)\|_\infty < 1$, allora il metodo delle approssimazioni successive converge quando $\|x_1 - x_0\|_\infty \leq r(1 - \theta)$. (sugg.: prendendo $K = B_\infty[x_0, r]$, si verifichi che $\phi(K) \subseteq K$, ...).
7. (!) valgono le seguenti *stime dell'errore*:

- *a priori*

$$\|\xi - x_n\| \leq \frac{\theta^n}{1 - \theta} \|x_1 - x_0\|$$

$$\|\xi - x_n\| \leq \theta^n \|\xi - x_0\|$$

- *a posteriori*

$$\|\xi - x_n\| \leq \frac{1}{1 - \theta} \|x_{n+1} - x_n\|$$

8. velocità di convergenza del metodo delle approssimazioni successive nelle ipotesi del teorema delle contrazioni: la convergenza è comunque *almeno lineare* visto che $\|\xi - x_{n+1}\| \leq \theta \|\xi - x_n\|$; se ϕ è C^2 in un intorno del punto fisso ξ e $J\phi(\xi) = 0$ la convergenza diventa localmente *almeno quadratica*, ovvero $\|\xi - x_{n+1}\| \leq c\|\xi - x_n\|^2$ con una opportuna costante c per n abbastanza grande. (sugg.: detta $B_2[\xi, r]$ una palla euclidea centrata in ξ tale che $\phi \in C^2(B_2[\xi, r])$, utilizzando al formula di Taylor centrata in ξ arrestata al secondo ordine si ha $x_{n+1} - \xi = \phi(x_n) - \phi(\xi) = J\phi(\xi)(x_n - \xi) + \varepsilon_n$ con $(\varepsilon_n)_i = \frac{1}{2}(x_n - \xi)^t H\phi_i(z_{n,i})(x_n - \xi)$, dove $H\phi_i$ è la matrice Hessiana della componente ϕ_i e $z_{n,i}$ sta nel segmento di estremi x_n e ξ , da cui $|(\varepsilon_n)_i| \leq \frac{1}{2} \max_{1 \leq i \leq m} \max_{x \in B_2[\xi, r]} \|H\phi_i(x)\|_2 \|x_n - \xi\|_2^2$ e quindi $\|\varepsilon_n\|_2 \leq \dots$).
9. (!) stabilità del metodo delle approssimazioni successive: dato il seguente modello di metodo perturbato

$$\tilde{x}_{n+1} = \phi(\tilde{x}_n) + \varepsilon_{n+1}, \quad n \geq 0$$

dove ϕ verifica le ipotesi del teorema delle contrazioni, vale la seguente stima per ogni $N > 0$

$$\max_{1 \leq n \leq N} \|\tilde{x}_n - x_n\| \leq \frac{1}{1 - \theta} \max_{1 \leq n \leq N} \|\varepsilon_n\|$$

10. si studi l'applicabilità del metodo delle approssimazione successive al sistema $x_1 = \arctan(x_1 + x_2) \sin(x_2)/10$, $x_2 = \cos(x_1/4) + \sin(x_2/4)$ e al sistema $2x_1^2 + x_2^2 = 5$, $x_1 + 2x_2 = 3$ (nel secondo caso si consideri la soluzione nel semipiano destro isolandola in un rettangolo opportuno tramite un'interpretazione grafica del sistema).
11. (!) dato il sistema non lineare $f(x) = 0$, dove $f : (\Omega \subseteq \mathbb{R}^m) \rightarrow \mathbb{R}^m$ è un campo vettoriale differenziabile definito su un aperto Ω contenente ξ tale che $f(\xi) = 0$, il *metodo di Newton* corrisponde alla *linearizzazione* iterativa

$$f(x_n) + J_n(x - x_n) = 0, \quad n \geq 0$$

a partire da un opportuno vettore iniziale x_0 , dove $J_n = Jf(x_n)$ è la matrice Jacobiana (purché $x_n \in \Omega$ e J_n sia invertibile ad ogni iterazione), ovvero

$$x_{n+1} = x_n - J_n^{-1}f(x_n), \quad n \geq 0$$

12. (!) *velocità di convergenza* del metodo di Newton: se $f \in C^2(K)$ dove K è la chiusura di un aperto convesso e limitato contenente ξ , in cui la Jacobiana di f è invertibile, e supposto che le iterazioni x_n siano tutte in K , posto $e_n = \|\xi - x_n\|_2$ vale la seguente stima (convergenza *almeno quadratica*)

$$e_{n+1} \leq ce_n^2, \quad n \geq 0, \quad c = \frac{\sqrt{m}}{2} \max_{x \in K} \|(Jf(x))^{-1}\|_2 \max_{1 \leq i \leq m} \max_{x \in K} \|Hf_i(x)\|_2$$

dove Hf_i è la matrice Hessiana della componente f_i .

(sugg.: dalla formula di Taylor centrata in x_n arrestata al secondo ordine, $0 = f(\xi) = f(x_n) + J_n(\xi - x_n) + \varepsilon_n$, e dalla definizione del metodo, si arriva a $\xi - x_{n+1} = -J_n^{-1}\varepsilon_n$, dove $(\varepsilon_n)_i = \frac{1}{2}(\xi - x_n)^t Hf_i(z_{n,i})(\xi - x_n)$, con $z_{n,i}$ che sta nel segmento di estremi x_n e ξ , ...).

13. (!) *convergenza locale* del metodo di Newton: se $f \in C^2(K)$ e $Jf(x)$ è invertibile in $K = B_2[\xi, r]$ (dove ξ è soluzione del sistema, $f(\xi) = 0$), detta c la costante dell'esercizio precedente, scelto x_0 tale che $e_0 < \min\{1/c, r\}$, il metodo di Newton è convergente e vale la seguente stima dell'errore

$$ce_n \leq (ce_0)^{2^n}, \quad n \geq 0$$

(sugg.: per induzione $e_{n+1} \leq (ce_n)e_n < e_n$ e quindi $x_{n+1} \in B_2[\xi, r]$, ...).

14. nelle ipotesi di convergenza locale la stima *a posteriori* dell'errore con lo step $\|x_{n+1} - x_n\|$ è una buona stima (almeno per n abbastanza grande)

$$e_n = \|\xi - x_n\| \approx \|x_{n+1} - x_n\|$$

(sugg.: si osservi che f è localmente invertibile e che $Jf^{-1}(f(x)) = (Jf(x))^{-1}$, quindi $\xi - x_n = f^{-1}(f(\xi)) - f^{-1}(f(x_n)) \approx Jf^{-1}(f(x_n))(f(\xi) - f(x_n)) = \dots$).

15. il metodo di Newton corrisponde ad un'iterazione di punto fisso con $\phi(x) = x - (Jf(x))^{-1}f(x)$, da cui si deduce che se f è C^2 in un intorno di ξ la convergenza è localmente almeno quadratica perché $J\phi(\xi) = 0$ (sugg.: posto $(Jf(x))^{-1} = \{b_{ij}(x)\}$, si ha $\frac{\partial \phi_i}{\partial x_k}(x) = \frac{\partial}{\partial x_k}(x_i - \sum_j b_{ij}(x)f_j(x)) = \delta_{ik} - \sum_j \frac{\partial b_{ij}}{\partial x_k}(x)f_j(x) - \sum_j b_{ij}(x)\frac{\partial f_j}{\partial x_k}(x) = -\sum_j \frac{\partial b_{ij}}{\partial x_k}(x)f_j(x), \dots$).