

1 Lezione del 16-05-25

1.1 Intervallo di convergenza

Vorremo estendere la definizione di convergenza data in 19.2 in un certo intervallo $[a, b]$.
Iniziamo con un esempio:

1.1.1 Esempio: convergenza del metodo di Newton

Prendiamo la funzione:

$$f(x) = 3x^2 e^{-x} - 1$$

e poniamo:

$$f(x) = 0$$

cioè equivalentemente uguagliamo le due funzioni:

$$3x^2 = e^x$$

che sono rispettivamente una parabola simmetrica rispetto all'asse y e un esponenziale.
Queste intersezioni saranno 3, che chiamiamo $\alpha_{1,2,3}$. Avremo che, scelti i punti $x_{0,1,2}$:

$$x_0 = -\frac{1}{2}, \quad x_1 = 1, \quad x_2 = 5$$

si avrà la convergenza nei rispettivi punti $\alpha_{1,2,3}$.

1.1.2 Convergenza in un intervallo

Una domanda che potremo farci, in relazione all'ultimo esempio, è quando possiamo assicurare che un metodo converge $\forall x_0 \in [a, b]$.

Esiste il seguente teorema:

Teorema 1.1: Teorema di convergenza in un intervallo

Se $\phi(x) \in C^1([a, b])$ e:

1. $|\phi'(x)| < 1, \quad \forall x \in [a, b];$
2. $\phi(x) \in [a, b], \quad \forall x \in [a, b];$

allora $\phi(x)$ converge su $[a, b]$.

Dove l'ipotesi in più rispetto al teorema di convergenza locale (teorema 19.2) è la 2.

1.1.3 Condizioni di convergenza globale per Newton

Nello specifico per Newton, la convergenza in un intervallo si assicura attraverso condizioni su concavità e convessità.

Esiste infatti il seguente teorema:

Teorema 1.2: Convergenza globale per Newton

Se $\alpha \in [a, b]$ è radice semplice di $f \in C^2([a, b])$, e f, f'' sono di segno costante su $[a, b]$, il metodo di Newton converge globalmente in maniera superlineare (≥ 2) $\forall x_0$ che verifica:

$$f(x_0) \cdot f''(x_0) > 0$$

Dove si mette enfasi su *globalmente*, in quanto il teorema 20.1 ha valenza solo locale. Sostanzialmente, questo teorema ci è utile per fare stime iniziali, probabilmente arbitrariamente distanti dall'origine, che convergono con velocità superlineare.

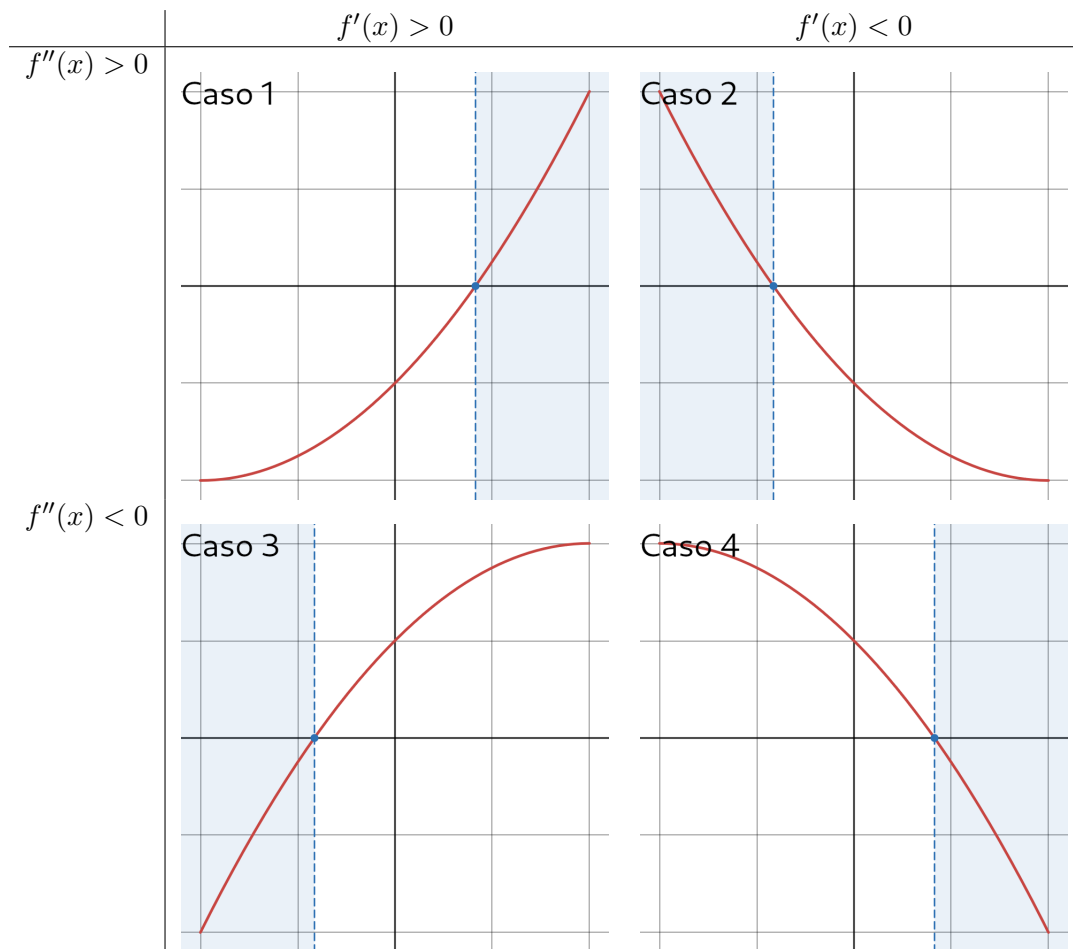
Osserviamo che:

- Se la funzione è **concava**, chiediamo $f(x_0) < 0$;
- Se la funzione è **convessa**, chiediamo $f(x_0) > 0$.

In particolare distinguiamo quindi i 4 casi:

1. $f' > 0, f'' > 0$ convessa crescente, vorremo partire a *destra* di α ;
2. $f' < 0, f'' > 0$ convessa decrescente, vorremo partire a *sinistra* di α ;
3. $f' > 0, f'' < 0$ concava crescente, vorremo partire a *sinistra* di α ;
4. $f' < 0, f'' < 0$ concava decrescente, vorremo partire a *destra* di α ;

Cioè riassumendo:



1.1.4 Esempio: ricerca di un intervallo di convergenza

Prendiamo la funzione:

$$f(x) = x + \log(x)$$

e cerchiamo un intervallo dove questa ha una radice, in particolare fornendo un punto di partenza x_0 per cui Newton risulta convergente.

Chiaramente quello che cerchiamo sono le intersezioni dei grafici:

$$\log(x) = -x$$

che saranno una sola, compresa fra 0 e 1, perciò:

$$\alpha \in (0, 1)$$

Vediamo quindi se si applicano le condizioni di convergenza per Newton. Calcoliamo innanzitutto le derivate:

$$f'(x) = 1 + \frac{1}{x}, \quad > 0 \text{ su } (0, 1)$$

$$f''(x) = -\frac{1}{x^2}, \quad < 0 \text{ su } (0, 1)$$

Per 1 la condizione non è soddisfatta, in quanto si troverà a *destra* di α (e trovandoci nel caso 3, vorremo prendere punti a *sinistra*).

Dovremo quindi cercare un $x_0 \in (0, \alpha)$, ovvero $x_0 : f(x_0) < 0$. Proviamo $\frac{1}{e}$:

$$f\left(\frac{1}{e}\right) = \frac{1}{e} - 1 < 0$$

per cui possiamo scegliere $x_0 = \frac{1}{e}$ ed essere sicuri, per il teorema 21.2, di convergere ad α .

1.2 Equazioni non lineari in \mathbb{R}^m

Vediamo quindi di considerare sempre equazioni non lineari, ma in più variabili (in particolare m).

Cercheremo quindi $f(x) = 0$ per $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$, cioè:

$$f = \begin{pmatrix} f_1(x_1, \dots, x_m) \\ \vdots \\ f_m(x_1, \dots, x_m) \end{pmatrix}$$

e vorremo trovare $\alpha \in \mathbb{R}^m$, cioè:

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}$$

tale che:

$$f(\alpha) = \begin{pmatrix} f_1(\alpha_1, \dots, \alpha_m) \\ \vdots \\ f_m(\alpha_1, \dots, \alpha_m) \end{pmatrix} = 0$$

Questo problema corrisponde effettivamente a cercare soluzioni a sistemi di m equazioni in m incognite, non lineari.

In particolare, ci può essere utile per trovare i punti stazionari di funzioni in più variabili, ovvero le soluzioni di $\nabla F(x) = 0$, con $F(x) : \mathbb{R}^m \rightarrow \mathbb{R}$ cioè:

$$\nabla F(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} F(x_1, \dots, x_m) \\ \vdots \\ \frac{\partial}{\partial x_m} F(x_1, \dots, x_m) \end{pmatrix}$$

Nel caso dei sistemi lineari, avevamo la condizione semplice:

$$f(x) = b - Ax, \quad x = A^{-1}b$$

mentre in questo caso non avremo a disposizione tali semplificazioni.

Ci potremo quindi chiedere come risolvere $f(x)$ nel caso non lineare. La soluzione sarà quella di generalizzare i metodi di punto fisso a \mathbb{R}^m , definendoli come:

$$x = \phi(x), \quad \phi : \mathbb{R}^m \rightarrow \mathbb{R}^m$$

e quindi usare la solita formula di aggiornamento:

$$x^{(n+1)} = \phi(x^{(n)})$$

Una possibile formula di aggiornamento può essere:

$$\phi(x) = x - G(x) \cdot f(x)$$

dove $G(x)$ è:

$$G(x) : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$$

cioè:

$$G(x) = \begin{pmatrix} G_{11}(x) & \dots & G(1m)(x) \\ \vdots & \ddots & \vdots \\ G_{m1}(x) & \dots & G(mm)(x) \end{pmatrix}$$

e si cerca $G(x)$ tale che $G(\alpha)$ è non singolare, ovvero

$$\det(G(\alpha)) \neq 0$$

Quindi si considera l'iterazione:

$$\begin{cases} x^{(0)} \in \mathbb{R}^m, & \text{dato} \\ x^{(n+1)} = x^{(n)} - G(x^{(n)})f(x^{(n)}) = \phi(x^{(n)}) \end{cases}$$

Ridiamo la definizione di convergenza in \mathbb{R}^n :

Definizione 1.1: Convergenza multivariabile

La successione $\{x^{(n)}\}_{n \in \mathbb{N}}$ converge ad $\alpha \in \mathbb{R}^m$ se:

$$\lim_{n \rightarrow +\infty} |x^{(n)} - \alpha|_2 = 0$$

In particolare, la convergenza ha **ordine** $p > 0$ se vale:

$$\lim_{n \rightarrow +\infty} \frac{|x^{(n+1)} - \alpha|_2}{|x^{(n)} - \alpha|_2^p} = c$$

con $c \neq 0$, $c < +\infty$, e in particolare nel caso $p = 1$ è $C \in (0, 1)$, cioè si ha la stessa situazione della definizione 19.3.

Nel caso esistano e siano continue le derivate delle componenti di $\phi(x)$ si introduce il **Jacobiano** di ϕ , come la funzione $J_\phi : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$ definita da:

$$J_\phi(x) = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(x) & \dots & \frac{\partial \phi_1}{\partial x_m}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_m}{\partial x_1}(x) & \dots & \frac{\partial \phi_m}{\partial x_m}(x) \end{pmatrix}$$

oppure per componenti:

$$[J_\phi(x)]_{ij} = \frac{\partial \phi_i}{\partial x_j}(x), \quad i, j = 1, \dots, m$$

Potremo quindi enunciare il seguente teorema:

Teorema 1.3: Convergenza locale multivariabile

Sia $\phi(x) \in C^1(\Omega)$, $\Omega \subseteq \mathbb{R}^m$ con $\alpha \in \Omega$ tale che $\phi(\alpha) = \alpha$. Se:

$$\rho(J_\phi(\alpha)) < 1$$

cioè il raggio spettrale del Jacobiano di ϕ in α è < 1 , allora $\exists \delta > 0$ tale che:

$$\forall x^{(0)} : |x^{(0)} - \alpha|_2 < \delta$$

in cui la successione:

$$x^{(n+1)} = \phi(x^{(n)})$$

converge ad α , ed α è l'unico punto fisso di ϕ in:

$$\{z \in \mathbb{R}^m : |z - \alpha|_2 < \delta\}$$

Osserviamo quindi che se per una qualsiasi norma si trova che:

$$|J_\phi(\alpha)| < 1$$

allora vale la tesi del teorema, ovvero la convergenza locale. Viceversa, se si trova che:

$$|J_\phi(\alpha)| \geq 1$$

non si può concludere nulla.

1.2.1 Esempio: convergenza multivariabile

Prendiamo la funzione, con $m = 2$:

$$f(x_1, x_2) = \begin{pmatrix} x_1 - \frac{1}{4}(x_1^2 + x_2^2) \\ x_2 + x_1 - 2 \end{pmatrix}$$

e cerchiamone i punti $f(\alpha) = 0$, cioè che soddisfano:

$$\begin{cases} x_1 - \frac{1}{4}(x_1^2 + x_2^2) = 0 \\ x_2 + x_1 - 2 = 0 \end{cases}$$

Dalla f , posto $f(x) = 0$ si derivano le identità:

$$\begin{cases} x_1 = \frac{1}{4}(x_1^2 + x_2^2) \\ x_2 = 2 - x_1 \end{cases}$$

per cui si può considerare $\phi(x)$:

$$\phi(x) = \begin{pmatrix} \frac{1}{4}(x_1^2 + x_2^2) \\ 2 - x_1 \end{pmatrix}$$

e prendere l'aggiornamento dal metodo stazionario ad un punto:

$$x^{(n+1)} = \phi(x^{(n)})$$

Prima di valutare la convergenza, cerchiamo di trovare i punti α in maniera analitica. Per la prima equazione, avremo che:

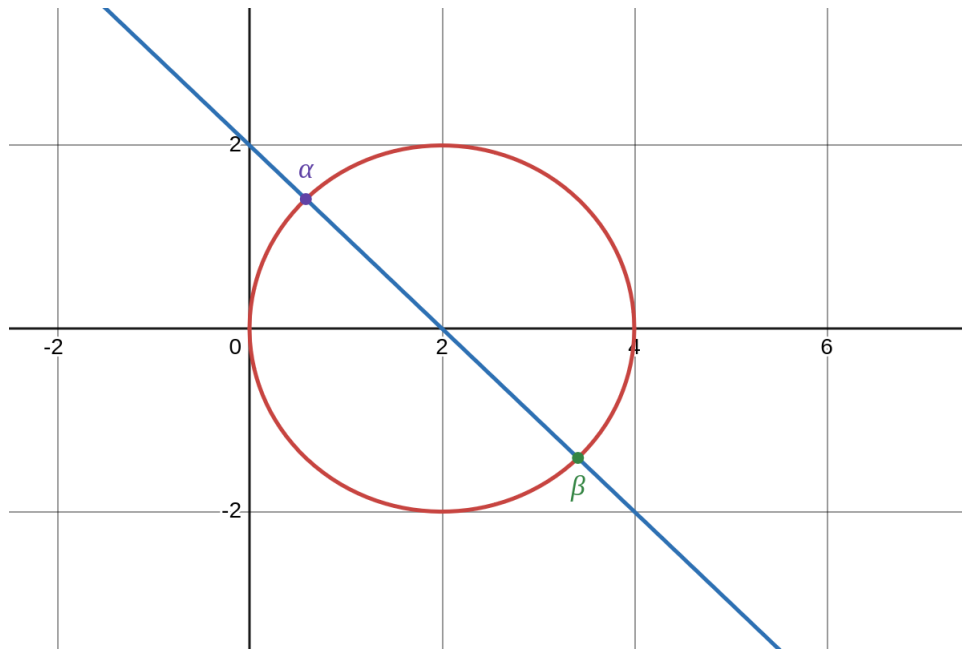
$$x - \frac{1}{4}(x^2 + y^2) = 0 \Rightarrow -\frac{(x-2)^2}{4} - \frac{y^2}{4} + 1 = 0$$

cioè si trova la circonferenza di centro $(2, 0)$ e raggio 2.

Per la seconda, si ha invece semplicemente la retta:

$$y = 2 - x$$

Sul grafico, questo ha l'aspetto:



Le soluzioni di $f(x) = 0$ saranno quindi i punti di intersezione fra la circonferenza e la retta, che vediamo essere 2, uno sopra l'asse x e l'altro sotto. Consideriamone solo il primo, che chiamiamo (α_1, α_2) .

Calcoliamo quindi il Jacobiano:

$$J_\phi(x) = \begin{pmatrix} \frac{x_1}{2} & \frac{x_2}{2} \\ -1 & 0 \end{pmatrix}$$

Vediamo che questa ha norma ≥ 1 , per cui bisogna calcolare gli autovalori. Guardiamo quindi al polinomio caratteristico:

$$\left(\frac{x_1}{2} - \lambda\right)(-\lambda) + \frac{x_2}{2} = \lambda^2 - \frac{x_1}{2}\lambda + \frac{x_2}{2} \rightarrow \lambda_{1,2} = \frac{\frac{x_1}{2} \pm \sqrt{\frac{x_1^2}{4} - 2x_2}}{2}$$

cioè:

$$\lambda_{1,2} = \frac{\frac{\alpha_1}{2} \pm \sqrt{\frac{\alpha_1^2}{4} - 2\alpha_2}}{2}$$

Valutando la posizione generica di α , si ha che:

$$\alpha_1 < 1, \quad \alpha_2 \in (1, 2)$$

che formalmente si ottiene notando che $\alpha_1 \geq 1$ porta ad un assurdo, e $\alpha_2 = 2 - \alpha_1$.

Da questo deriverà che il discriminante:

$$\Delta = \frac{\alpha_1^2}{4} - 2\alpha_2 < 0$$

e cioè $\lambda_{1,2}^\alpha$ sono necessariamente numeri complessi in forma:

$$\lambda_{1,2}^\alpha = \frac{\alpha_1}{4} \pm i\sqrt{\frac{\alpha_2}{2} - \frac{\alpha_1^2}{16}}$$

Il modulo di uno di questi (prendiamo il primo) sarà:

$$|\lambda_1^\alpha|^2 = \frac{\alpha_1^2}{16} + \frac{\alpha_1}{2} - \frac{\alpha_1^2}{16} = \frac{\alpha_1}{2}$$

per cui il raggio spettrale sarà $\frac{\alpha_1}{2} < 1$ (dal fatto che $\alpha_1 < 1$), e quindi il metodo convergente.

1.3 Metodo di Newton-Raphson

Vediamo quindi la versione multivariabile del metodo di Newton-Raphson.

Quindi, se abbiamo $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$ il metodo di Newton si generalizza nel seguente modo:

$$x^{(n+1)} = x^{(n)} - J_f(x^{(n)})^{-1} f(x^{(n)})$$

Osserviamo che ad ogni passo si valuta $J_f(x^{(n)})$ e $f(x^{(n)})$, si risolve il sistema lineare:

$$J_f(x^{(n)})y = f(x^{(n)})$$

ed infine si chiama $x^{(n)} - y$.

L'aggiornamento complessivo sarà quindi:

$$\begin{cases} x^{(0)}, & \text{dato} \\ x^{(n+1)} = x^{(n)} - J_f(x^{(n)})^{-1} f(x^{(n)}) \end{cases}$$

Per quanto riguarda la convergenza si ha il seguente risultato:

Teorema 1.4: Convergenza del metodo di Newton-Rapson

Sia $f \in C^2(\Omega)$, $\Omega \subseteq \mathbb{R}^m$, $\alpha \in \Omega$ tale che $f(\alpha) = 0$. Se $J_f(\alpha)$ è non singolare, allora $\exists S \subseteq \Omega$ intorno di α tale che $\forall x^{(0)} \in S$ la successione generata dal metodo di Newton-Raphson a partire da $x^{(0)}$ converge e vale:

$$|x^{(n+1)} - \alpha|_2 \leq \beta |x^{(n)} - \alpha|_2^2$$

ovvero la convergenza è almeno quadratica (quindi superlineare).

1.3.1 Esempio: metodo di Newton-Raphson

Vediamo quindi come valutare la convergenza del metodo di Newton-Raphson per il sistema:

$$f(x_1, x_2) = \begin{pmatrix} \frac{1}{3}(x_1 - x_2) + x_1^2 \\ \frac{1}{3}(-x_1 + x_2) + x_1x_2 \end{pmatrix}$$

Vogliamo quindi determinare le radici di f , che si otterranno come:

$$\begin{cases} \frac{x_1 - x_2}{3} + x_1^2 = 0 \\ \frac{x_2 - x_1}{3} + x_1x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_2 = 3x_1^2 + x_1 \\ \left(\frac{-x_1 + x_1 + 3x_1^2}{3}\right) + x_1(3x_1^2 + x_1) = 0 \end{cases} \Rightarrow \begin{cases} x_2 = 3x_1^2 + x_1 \\ 2x_1^2 + 3x_1^3 = 0 \end{cases}$$

da cui si ottengono i due punti:

$$\alpha = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$$

Vorremo quindi valutare il Jacobiano:

$$J_f(x_1, x_2) = \begin{pmatrix} \frac{1}{3} - 2x_1 & -\frac{1}{3} \\ -\frac{1}{3} - x_2 & \frac{1}{3} + x_1 \end{pmatrix}$$

Che nei due punti α e β risulta:

α :

$$J_f(\alpha) = \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

da cui $\det = 0$, e la convergenza non è superlineare;

β :

$$J_f(\beta) = \begin{pmatrix} -1 & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}$$

da cui $\det \neq 0$, e la convergenza è superlineare;

1.3.2 Varianti del metodo di Newton-Raphson

Abbiamo che ad ogni passo del metodo di Newton-Raphson dobbiamo:

1. Valutare $J_f(x^n)$; ovvero m^2 funzioni non lineari;
2. Risolvere $J_f(x^n)y = f(x^n)$, sistema lineare $m \times m$ (complessità $O(m^3)$).

Quando m è particolarmente grande siamo disposti a fare qualcosa di più economico al prezzo di perdere l'ordine di convergenza molto favorevole di Newton-Raphson.

1. La prima strategia è quella di non aggiornare mai il Jacobiano (**Newton semplificato**). Il Jacobiano viene quindi calcolato solo in x^0 punto iniziale, e poi si usa sempre lo stesso nelle iterazioni successive, cioè l'aggiornamento diventa:

$$x^{n+1} = x^n - J_f(x^0)^{-1} f(x^n)$$

Per la risoluzione del sistema, si può quindi sfruttare la fattorizzazione LU del Jacobiano appena nominato, e quindi svolgere complessivamente due operazioni $O(m^2)$.

2. La seconda strategia viene detta **metodo di Jacobi nonlineare**. Ad ogni passo si sostituisce $J_f(x^n)$ con la matrice diagonale:

$$\text{diag} \left(\frac{\partial f}{\partial x_1}(x^n), \dots, \frac{\partial f}{\partial x_m}(x^n) \right)$$

così che:

- (a) Bisogna fare m valutazioni di funzioni non lineari;
- (b) Risolvere il sistema lineare costa $O(m)$.

L'aggiornamento diventa quindi qualcosa del tipo:

$$x_i^{n+1} = x_i^n - \frac{f_i(x^n)}{\frac{\partial f}{\partial x_i}(x_1^n, \dots, x_m^n)}$$

3. La terza strategia viene detta **metodo di Gauss-Seidel nonlineare**.

In questo caso l'aggiornamento è sostanzialmente analogo al metodo di Jacobi, con la differenza:

$$x_i^{n+1} = x_i^n - \frac{f_i(x_1^{n+1}, \dots, x_i^n, \dots, x_m^n)}{\frac{\partial f}{\partial x_i}(x_1^n, \dots, x_m^n)}$$

dove le entrate x^{n+1} vengono calcolate una per volta attraverso le stesse considerazioni fatte per il metodo di Gauss-Seidel lineare.

Altre varianti sono date dai metodi **quasi-Newton**, dove si sostituisce $J_f(x)$ con qualcosa con cui è più semplice (meno costoso) risolvere sistemi lineari e che richiedono meno valutazioni di funzioni non lineari.