

Consiglio Nazionale delle Ricerche

**ISTITUTO DI ELABORAZIONE
DELLA INFORMAZIONE**

PISA

ANALISI DI ALCUNE CLASSI DI METODI
QUASI NEWTONIANI

M. Arioli, P. Favati

Nota interna B85-04

Luglio 1985

ANALISI DI ALCUNE CLASSI DI METODI QUASI NEWTONIANI

M. ARIOLI-P. FAVATI

Istituto di Elaborazione dell'Informazione - CNR,
via S. Maria 46, PISA

Introduzione

Una importante classe di algoritmi per trovare il minimo di funzioni non vincolate e' costituita dai metodi quasi Newton, chiamati cosi' perche' utilizzano lo schema del metodo di "Newton modificato" sostituendo ad ogni passo l'inversa della matrice hessiana con una certa matrice costruita in modo iterativo.

Nel seguito indicheremo con $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$ la funzione di cui si cerca il minimo e $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ il gradiente di f .

Lo schema iterativo comune a tutti i metodi e' il seguente:

fissati inizialmente $x_0 \in \mathbb{R}^n$ e H_0 matrice $n \times n$, si calcola

$$x_{i+1} = x_i - \alpha_i H_i^{-1} g_i$$

dove $g_i = g(x_i)$, $\alpha_i \in \mathbb{R}$ e' il passo calcolato secondo un criterio

fissato e la matrice H_{i+1} , posto $s_i = x_{i+1} - x_i$ e $y_i = g_{i+1} - g_i$,

viene costruita in modo che sia verificata la seguente condizione, detta "equazione quasi Newton":

$$H_{i+1} y_i = s_i$$

o quella piu' generale :

$$H_{i+1} y_i = \rho_i s_i, \quad \rho_i \in \mathbb{R}.$$

Osserviamo che, nel caso in cui la funzione f sia quadratica con matrice della forma quadratica G , l'equazione quasi Newton e' verificata ad ogni passo dall'inversa di G . Queste condizioni non determinano in modo univoco la matrice H_{i+1} ; generalmente H_{i+1} viene calcolata con una formula di aggiornamento del tipo:

$$H_{i+1} = H_i + a_1 b_1^T + a_2 b_2^T + \dots + a_r b_r^T$$

dove i vettori $a_1, \dots, a_r, b_1, \dots, b_r$ sono funzione di s_i, y_i, H_i, y_i e di eventuali parametri. Nel seguito saranno detti di rango t quei metodi in cui H_{i+1} si calcola con questa formula di aggiornamento e $r=t$.

Per la determinazione del passo esistono vari criteri: un primo criterio e' quello della ricerca del minimo unidimensionale

$$(0.1) \quad \alpha_i = \min f(x_i - \alpha H_i g_i), \quad \alpha \in \mathbb{R}.$$

Spesso approssimare nel miglior modo possibile il passo esatto è troppo costoso; si ricorre quindi a criteri di ricerca inesatta che garantiscano una soddisfacente decrescita della funzione f , [9], [7]; infine, in alcuni casi, una terza scelta possibile è quella di porre $\alpha_i = 1$ per ogni i .

Per quanto riguarda la convergenza esistono risultati specifici relativi alla scelta del passo e della formula di aggiornamento di H [18], [2], [21], [22], [23]; in generale comunque, se la matrice H è definita positiva, i vari metodi sono di discesa e, sotto opportune ipotesi sulla determinazione del passo e sulla funzione f , teoremi generali garantiscono la convergenza lineare [7]. Un'altra proprietà comune a molti (quasi tutti) metodi di tipo quasi Newton è la terminazione quadratica; cioè questi metodi determinano la soluzione in un numero finito di passi se applicati ad una funzione quadratica. Tale proprietà è interessante anche per lo studio della velocità di convergenza dei metodi applicati a funzioni non quadratiche [22], [23].

Piassumendo, nell'applicazione di un metodo quasi Newton i tre punti fondamentali sono: la scelta del punto iniziale, la scelta del passo e la scelta della direzione, che dipende dal modo di aggiornare la matrice H . Noi non ci occuperemo dei primi due aspetti, ma solo del terzo; pertanto nel seguito il passo corrisponderà al minimo unidimensionale (0.1).

Introdurremo la seguente relazione di equivalenza tra due metodi quasi Newton: considereremo "teoricamente" equivalenti due metodi se, pur essendo diversi dal punto di vista dell'algoritmo, generano le stesse direzioni. Indagare quali tra i metodi quasi Newton, noti in letteratura, determinano la stessa direzione risulta interessante perché i metodi tra loro equivalenti godono delle stesse proprietà teoriche (convergenza, terminazione quadratica, ecc...). Inoltre questo tipo di indagine prepara il terreno ad una seconda fase di ricerca che non è stata qui affrontata: la determinazione, nell'ambito di una classe di equivalenza, del metodo più efficiente dal punto di vista della complessità computazionale e della stabilità.

Nel primo capitolo descriveremo, enunciandone le proprietà, le più importanti famiglie di metodi quasi Newton; più precisamente ci occuperemo della famiglia di Broyden, di quella di

Huang a cinque parametri, dei metodi di Oren e Luenberger e di quelli di Greenstadt.

Nel secondo capitolo studieremo l'equivalenza teorica dei metodi presi in esame, raccogliendo i risultati noti in letteratura [4], [5], [20] e stabilendo un nuovo risultato che mette in relazione i metodi di Greenstadt con quelli di Broyden e con la famiglia di Huang simmetrica.

Cap.1 Descrizione dei metodi.

1.1 Metodi di Broyden

La famiglia di rango 2 di Broyden [1] e' costruita secondo il seguente schema:

posto
$$H_{i+1} = H_i + R_i,$$

affinche' sia verificata l'equazione quasi Newton deve essere

$$B_i y_i = s_i - H_i y_i ;$$

si sceglie allora $R_i = s_i q_i^T - H_i y_i z_i^T$, dove q_i e z_i verificano

$$q_i^T y_i = z_i^T y_i = 1$$

In [1] vengono proposte le seguenti due scelte di q e z :

(1.1)
$$q_i^T = z_i^T = (y_i^T H_i y_i - s_i^T y_i)^{-1} (y_i^T H_i - s_i^T) ;$$

(1.2a)
$$q_i^T = \gamma_i s_i^T - \sigma_i y_i^T H_i / s_i^T y_i$$

(1.2b)
$$z_i^T = \delta_i y_i^T H_i + \theta_i s_i^T / s_i^T y_i, \text{ dove}$$

(1.2c)
$$\gamma_i = (s_i^T y_i + \sigma_i y_i^T H_i y_i) / (s_i^T y_i)^2$$

(1.2d)
$$\delta_i = (1 - \sigma_i) / y_i^T H_i y_i ;$$

la prima scelta conduce ad un metodo di rango 1 che ha grossi svantaggi in quanto il denominatore che compare nella (1.1) puo' annullarsi ed il metodo stesso risultare non definito; la seconda fa dipendere q_i e z_i da un parametro σ_i e conduce alla famiglia ben nota in letteratura di rango 2, la cui formula di aggiornamento di H e':

$$H_{i+1} = H_i + \frac{s_i s_i^T}{s_i^T y_i} - \frac{H_i y_i y_i^T H_i}{y_i^T H_i y_i} + (y_i^T H_i y_i) \sigma_i v_i v_i^T, \text{ dove}$$

$$(1.3) \quad v_i = \frac{s_i}{s_i^T y_i} - \frac{H_i y_i}{y_i^T H_i y_i}$$

Da questa formula per $\sigma_i = 0$ si ottiene il metodo di Davidon Fletcher Powell (DFP) che cronologicamente e' il primo metodo di

tipo quasi Newton proposto [3], [6]. Un altro metodo ottenuto indipendentemente da vari autori e' quello corrispondente a $\alpha=1$ detto metodo BFGS (Broyden, Fletcher, Goldfarb, Shanno) che sperimentalmente mostra buone proprieta' nel caso di determinazione del passo con tecniche di ricerca inesatta; la formula che aggiorna la matrice H puo' essere scritta nella forma seguente (nel seguito per semplificare le notazioni indicheremo con H^* la matrice aggiornata e tralascieremo gli indici delle matrici, dei vettori e dei parametri relativi al passo corrente):

$$H_{BFGS}^* = H + \left(1 + \frac{y^T H y}{s^T y} \right) \frac{ss^T}{s^T y} - \left(\frac{sy^T H + Hys^T}{s^T y} \right)$$

Infine la famiglia di Broyden include la formula di Hoshino per

$$\bar{\theta} = \left(1 + \frac{y^T H y}{s^T y} \right)^{-1}$$

Anche questo metodo da' buoni risultati, se utilizzato in congiunzione con criteri di ricerca inesatta del passo [7].

Le principali proprieta' verificate dalla famiglia di Broyden sono le seguenti:

- i) Se H_0 e' simmetrica ogni H_i risulta simmetrica.
- ii) Se il metodo e' applicato a problema quadratico termina in $m \leq n$ iterazioni e se $m=n$ $H_{m+1} = G^{-1}$ (dove G indica la matrice della forma quadratica).
- iii) Se H_0 e' definita positiva e ad ogni passo risulta $s_i^T y_i > 0$ e $\bar{\theta}_i > \bar{\theta}_{i+1}$ dove

$$(1.4) \quad \bar{\theta} = \left(1 - \frac{y^T H y}{s^T y} \frac{s^T H s}{2} \right)^{-1}$$

allora H_{i+1} risulta definita positiva. Osserviamo inoltre che se H_i e' definita positiva allora $\bar{\theta}_i < 0$.

- iv) Se $H_0 = I$ ed il metodo e' applicato a funzioni quadratiche genera le stesse direzioni del gradiente coniugato di Fletcher Reeves.

1.2 Metodi di Huang

Huang [11] determina una classe di metodi di tipo quasi Newton dipendente da 5 parametri. Essendo interessato alla terminazione quadratica del metodo, considera il caso in cui

$$f(x) = \frac{1}{2} x^T G x + b^T x + a$$

con G definita positiva e imposta un metodo iterativo del tipo:

$$x_{i+1} = x_i - \alpha_i p_i, \quad \text{dove} \quad \alpha_i = \frac{g_i^T p_i}{p_i^T G p_i} \quad \text{realizza il}$$

minimo unidimensionale indicato in (0.1)

Suppone che p_i sia del tipo $p_i = H_i^T g_i$ con H_i matrice $n \times n$ ed impone le seguenti condizioni:

- a) $g_i^T p_i \neq 0$ "condizione di nonortogonalita"
- b) $p_i^T G p_j = 0$ per $i \neq j$ "condizione di coniugio"

che garantiscono le proprietà di discesa e di terminazione quadratica.

Utilizzando solo la funzione e la sua derivata, le informazioni al passo corrente e a quello precedente ottiene la seguente formula iterativa, piuttosto generale:

$$H^* = H + \rho \frac{(c_1 s + c_2^T y) s^T - (k_1 s - k_2^T y) y^T H}{(c_1 s + c_2^T y) y^T - (k_1 s + k_2^T y) y^T}$$

(ricordiamo che i cinque parametri ρ, c_1, c_2, k_1, k_2 possono variare ad ogni passo)

Affinche' sia verificata la condizione di nonortogonalita' si richiede che la parte simmetrica della matrice iniziale, cioe'

la matrice $(H + H^T)/2$, sia definita positiva o negativa.

Questa famiglia di metodi verifica l'equazione:

$$H_{i+1} y_i = \rho_i s_i \quad \text{e se } \rho_i = \rho \text{ per ogni } i \text{ risulta } H_{n+1} = \rho G^{-1}$$

Si ottiene una classe molto generale in cui si ritrovano per opportune scelte dei parametri i metodi di Broyden, il metodo del gradiente coniugato ed alcuni interessanti metodi di rango 1; si perdono pero' in generale le proprietà di simmetria e di definita positività di H_i , pur partendo con H_0 simmetrica e definita positiva.

Un sottoinsieme speciale della famiglia di Huang e'

costituito dalla cosiddetta famiglia simmetrica di Huang (Hs) che dipende da due successioni di parametri $\{\rho_i\}$ e $\{\sigma_i\}$, mantiene la simmetria delle H_i ed ha la seguente formula:

$$(1.5) \quad H_{Hs}^* = H + \rho \frac{SS^T}{S^T Y} - \frac{H Y Y^T H}{Y^T H Y} + \sigma (Y^T H Y) v v^T, \text{ con } v \text{ come in (1.3)}$$

1.3 Metodi di Oren e Luenberger

Oren e Luenberger [13] introducono un metodo quasi Newton dipendente da due successioni di parametri $\{\gamma_i\}$, $\{\sigma_i\}$, la formula di aggiornamento della matrice risulta:

$$(1.6) \quad H_{O-L}^* = \gamma \left(H - \frac{H Y Y^T H}{Y^T H Y} + \sigma (Y^T H Y) v v^T \right) + \frac{S S^T}{S^T Y},$$

con $\gamma > 0$, $\sigma > 0$, v definito in (1.3).

Il metodo di O-L ad ogni passo verifica l'equazione quasi Newton per qualunque scelta dei parametri. Se H_0 e' definita positiva, $\gamma > 0$, $\sigma > 0$ e $S^T Y > 0$ le matrici H_i risultano definite positive

E' stato dimostrato [13] che ogni algoritmo di tipo quasi Newton con matrice H_i definita positiva applicato a funzione quadratica definita positiva ha tasso di convergenza ad ogni passo

$$\left(\frac{\kappa(R_i) - 1}{\kappa(R_i) + 1} \right)^2$$

(con $R_i = G_i^{\frac{1}{2}} H_i G_i^{\frac{1}{2}}$, G matrice della forma quadratica e

$\kappa(A) = \frac{\|A\|}{\|A^{-1}\|}$ il condizionamento della matrice $n \times n$ A invertibile).

Questo fatto implica che la convergenza e' tanto migliore quanto piu' e' piccolo $\kappa(R_i)$.

Tenendo conto di questo risultato, Oren e Luenberger specializzano la scelta dei parametri $\{\gamma_i\}$, $\{\sigma_i\}$ in modo tale che risulti $\kappa(R_{i+1}) < \kappa(R_i)$. Precisamente ricavano le seguenti formule per i parametri:

$$(1.7) \quad \gamma = \phi \frac{S^T H^{-1} S}{S^T Y} + (1-\phi) \frac{S^T Y}{Y^T H Y}, \quad \phi \in [0, 1]$$

$$(1.8) \quad \sigma \in [0, 1]$$

Il metodo ottenuto considerando la formula (1.6) insieme alle (1.7) e (1.8) viene chiamato Self Scaling Variable Metric

Algorithm (SSVM). Tale metodo (molto studiato anche in congiunzione con criteri di ricerca inesatta per il passo) gode delle seguenti proprietà:

a) Se H_0 è definita positiva allora, purché $s^T y > 0$, ogni H_i risulta definita positiva.

b) Le direzioni di ricerca sono invarianti per scalatura della funzione obiettivo e della matrice iniziale H_0 .

Se la funzione è quadratica definita positiva, con matrice della forma quadratica G , l'algoritmo ha anche le seguenti proprietà:

c) Se il passo è calcolato con criterio di ricerca esatta il metodo ha terminazione quadratica

d) Al crescere dell'indice i il numero di condizionamento della matrice $G^k H_i G^k$ decresce.

e) Se $\alpha_i = 1$ per ogni i allora:

$$\lim_{i \rightarrow \infty} \frac{\|x_{i+1}^* - x^*\|}{\|x_i^* - x^*\|} = 0 \quad \text{dove } x^* \text{ è la soluzione.}$$

In [17] viene proposta la seguente formula per $\{\theta_i\}$, (omettiamo anche qui gli indici per semplicità):

$$(1.9) \quad \theta = \frac{s^T y (s^T H^{-1} s - \gamma s^T y)}{(s^T H^{-1} s + \gamma H y - (s^T y)^2)}$$

(se γ è calcolato secondo (1.7), θ scelto come in (1.9) verifica la (1.8))

Con questa scelta del parametro viene minimizzata la seguente limitazione superiore del condizionamento di H^* , valida se H^* e γ sono definite positive:

$$\kappa(H^*) \leq \kappa(H) \frac{\max(\lambda + (\lambda^2 - \mu)^{1/2}, \gamma)}{\min(\lambda - (\lambda^2 - \mu)^{1/2}, \gamma)}$$

dove $\lambda = (s^T H^{-1} s + \gamma H y) / 2 s^T y$,

$$\mu = \gamma ((s^T y)^2 + \theta (s^T H^{-1} s + \gamma H y - (s^T y)^2) / \gamma H y s^T y)$$

1.4 Metodo di Greenstadt

I metodi fin qui esposti sono stati ottenuti seguendo un approccio di tipo quadratico, cioè imponendo che il metodo verificasse determinate proprietà se applicato a funzioni quadratiche. Greenstadt [10] costruisce una famiglia di metodi seguendo un approccio completamente diverso, usa infatti una

tecnica di tipo variazionale. Posto

$$H_{i+1} = H_i + E_i$$

la correzione E_i viene determinata minimizzando la norma di Frobenius della matrice $A E_i A$ soggetta ai vincoli di simmetria e di quasi Newton, con A matrice fissata. Osservando che:

$$\text{Tr}(A E_i A^T) = \text{Tr}(A^T A E_i) \quad \text{e ponendo} \quad A A^T = A^T A = W$$

si conclude che E_i può essere determinato risolvendo il seguente problema:

$$(1.10) \quad \begin{cases} \min \text{Tr} (W E_i W E_i^T) \\ E_i^T - E_i = 0 & \text{simmetria} \\ E_i y_i = s_i - H_i y_i & \text{eq. quasi Newton} \end{cases}$$

dove la matrice W , scelta come peso, è simmetrica e definita positiva. Risolvendo questo problema e ponendo $W^{-1} = M$ si ottiene la seguente formula:

$$H_G^* = H + \frac{1}{y^T M y} (s y^T M + M y s^T - H y y^T M - M y y^T H - \frac{1}{y^T M y} (y^T s - y^T H y) M y y^T M)$$

Greenstadt analizza le formule corrispondenti alle seguenti scelte della matrice M : $M=H$, $M=I$. Nessuna delle due scelte garantisce che sia mantenuta la definita positività della matrice H^* , per $M=H$ è dimostrata la terminazione quadratica del metodo [10].

Golfarb [8] propone di considerare $M=H^*$, con questa scelta si ottiene il metodo BFGS le cui proprietà sono state già elencate. Golfarb osserva che facendo una combinazione convessa delle correzioni E_H e E_{H^*} relative alle scelte $M=H$ e $M=H^*$ si riottengono i metodi di Broyden; più precisamente la correzione

$$E = \alpha E_H + (1-\alpha) E_{H^*} \quad \text{con} \quad \alpha = \frac{(1-\sigma) y^T H y}{y^T H y + y^T s}$$

coincide con quella del metodo di Broyden di parametro σ . Golfarb, inoltre esibisce diverse matrici peso che determinano una correzione uguale a quella del metodo DFP, tali matrici non sono però in generale definite positive e le corrispondenti formule non possono essere considerate appartenenti alla famiglia di Greenstadt.

C'e' da osservare comunque che anche le scelte $M=H$ e $M=H^*$ sono incoerenti con l'impostazione del problema (1.10) e le ipotesi fatte sulla matrice dei pesi. Infatti ponendo $M=H$ non e' garantita la definita positività di H^* e quindi al passo successivo non puo' essere ripetuta la medesima scelta ; inoltre porre $M=H^* = H+E$ significa far dipendere M da E , mentre nel problema (1.10) M e' costante rispetto alla matrice incognita E .

Cap. 2 Analisi comparativa di alcuni metodi.

Molti dei metodi descritti nel capitolo precedente risultano, per opportune scelte dei parametri, equivalenti nel senso esposto nell'introduzione.

Dixon [4] dimostra che le sequenze di punti generate applicando le formule appartenenti alla famiglia di Huang a funzioni qualunque, con passo determinato come in (0.1), sono identiche se e solo se il valore di ρ e' lo stesso. La famiglia di Broyden al variare di σ costituisce il sottoinsieme di quella di Huang corrispondente alla seguente scelta dei parametri

$$\rho=1, k_1=c_2=1, k_2=\frac{1-\sigma}{\sigma} \frac{s^T y}{y^T H y} \text{ e } c_1=-\frac{1}{\sigma} - \frac{y^T H y}{s^T y} \text{ per } \sigma \neq 0$$

e $\rho=1, k_1=c_2=0, k_2=c_1=1$ per $\sigma=0$

Dal teorema di Dixon segue quindi che tutti i metodi di Broyden, al variare di σ , ed i metodi di Huang con $\rho=1$ sono equivalenti.

Spedicato [20] mette in relazione la famiglia di Huang simmetrico con quella di Oren Luenberger; piu' precisamente dimostra che se i parametri $\{\rho_i\}$ e $\{\gamma_i\}$ (che compaiono in (1.5) e (1.6) rispettivamente) soddisfano la relazione :

$$1/\rho_i = \prod_{j=0}^i \gamma_j$$

allora i metodi di Oren Luenberger e di Huang simmetrico generano le stesse direzioni.

Nel paragrafo successivo daremo condizioni sui parametri affinche' un metodo di Greenstadt ed uno di Broyden siano equivalenti.

2.1 Metodi di Greenstadt e di Broyden

Supponendo di partire con la stessa matrice H_0 i metodi di Greenstadt e di Broyden costruiscono la stessa matrice H_{i+1} se vale:

$$(2.1) \frac{1}{y^T M y} (s y^T M + M y s^T - H y y^T M - M y y^T H - \frac{1}{y^T H y} (y^T s - y^T H y) M y y^T M) =$$

$$= \frac{ss^T}{s^T y} - \frac{Hy^T H}{y^T Hy} + \theta vv^T (y^T Hy) \quad \text{con } v \text{ come in (1.3)}$$

Se si pone $w = \frac{s}{s^T y} - \frac{My}{y^T My}$, $v = -\frac{s}{s^T y}$

$z = \frac{Hy}{y^T Hy} - \frac{My}{y^T My}$, $\mu = y^T Hy$

la condizione (2.1) diventa:

$$(2.2) \quad v w^T + \mu z z^T = \mu \theta v v^T$$

Se la (2.2) e' verificata, deve valere per ogni vettore p

$$(2.3) \quad v (w^T p) w + \mu (z^T p) z - \mu \theta (v^T p) v = 0$$

Poiche' $v = w - z$, se w e z fossero linearmente indipendenti dovrebbe risultare:

$$v (w^T p) = \mu \theta (v^T p) \quad \text{e} \quad \mu (z^T p) = -\mu \theta (v^T p) \quad \text{per ogni } p,$$

$$\text{ossia} \quad v (w^T p) = -\mu (z^T p) \quad \text{per ogni } p.$$

Scegliendo p ortogonale a w e appartenente al piano generato da w e da z si arriva ad una contraddizione, quindi perche' sia verificata la (2.1) w e z devono essere paralleli, cioe:

$$(2.4) \quad w = \omega z, \quad \omega \in \mathbb{R}.$$

Con semplici passaggi dalle (2.4), (2.2) tenendo presente la definizione di w e z si ottengono le seguenti condizioni per M e

$$(2.5) \quad \frac{My}{y^T My} = \left(\frac{s}{s^T y} - \omega \frac{Hy}{y^T Hy} \right) \frac{1}{1-\omega}, \quad \omega \neq 1$$

$$(2.6) \quad \theta = \varphi(\omega) = \frac{(\nu/\mu) \omega^2 + 1}{(\omega - 1)^2}$$

Osserviamo che posto :

$$\left(\frac{s}{s^T y} - \omega \frac{Hy}{y^T Hy} \right) \frac{1}{1-\omega} = h, \quad \text{la (2.5) diventa}$$

$$(2.7) \quad (I - b y^T) M y = 0 \quad \text{da cui, essendo } b^T y = 1, \text{ segue}$$

$$(2.8) \quad M y = \beta b, \quad \beta \in \mathbb{R}$$

Dalla formula (2.1) segue che assumere $\beta = 1$ non lede la generalita'.

Se fissato un valore di ω , esiste una matrice M simmetrica

e definita positiva che risolve (2.8) il metodo di Greenstadt corrispondente coincide con il metodo di Broyden di parametro dato dalla (2.6). Per verificare se una matrice che gode di queste proprietà esiste impostiamo il seguente problema di minimo vincolato:

$$(2.9) \begin{cases} \min \|M - X\|_F^2 \\ My = b \\ x - \bar{x} = 0 \end{cases}$$

con X matrice $n \times n$.

La funzione obiettivo di questo problema è convessa; scrivendo la condizione di Lagrange per il problema (2.9) si determina la seguente soluzione ottima:

$$(2.10) \quad M = P X P + \frac{1}{Y^T Y} \left((I+P) b y^T \right)^T S$$

dove $A = 1/2 (A + A^T)$ e $P = I - \frac{Y Y^T}{Y^T Y}$ è il proiettore ortogonale

sulla varietà $\{x: x^T y = 0\}$

Se indichiamo con s_0 il più piccolo autovalore di X^S , si può dimostrare che condizione sufficiente affinché M data in (2.10) sia definita positiva è che risulti:

$$(2.11) \quad s > b^T b$$

(Per la dimostrazione di questa proprietà e i passaggi con cui si arriva alla formula (2.10) vedere appendice 1) Fissata quindi una matrice X che verifica (2.11), risolvendo (2.9) si determina M simmetrica e definita positiva con $My=b$; questo in particolare risponde alla questione che ci eravamo posti. Segue quindi che ogni metodo di Greenstadt con M simmetrica e definita positiva tale che:

$$My = \left(\frac{S}{Y^T Y} - \omega \frac{Hy}{Y^T Hy} \right) \frac{1}{1 - \omega} \quad \text{con } \omega \neq 1$$

è un metodo di Broyden per opportuni $\{\omega\}$

Studiamo ora la funzione $\varphi(\omega)$ espressa in (2.6); se poniamo

$$-\frac{2}{\mu} = a \quad \text{risulta } \varphi(\omega) = \frac{-a\omega^2 + 1}{(\omega - 1)^2}$$

Nelle figure 1 e 2 sono riportati i grafici di $\varphi(\omega)$

rispettivamente per $a < 1$ e $a > 1$.

Come si vede in figura 1, per $a < 1$ risulta

$$\text{Im}(\varphi) = \left[\frac{a}{a-1}, \infty \right) = \left[\frac{\frac{\text{S}^T \text{Y}}{\text{S}^T \text{Y} - \text{Y}^T \text{H} \text{Y}}}{\text{S}^T \text{Y} - \text{Y}^T \text{H} \text{Y}}, \infty \right)$$

questo significa che, se al passo i -esimo $\text{S}^T \text{Y} < \text{Y}^T \text{H} \text{Y}$, ogni metodo

di Broyden con parametro $\sigma > \frac{\text{S}^T \text{Y}}{\text{S}^T \text{Y} - \text{Y}^T \text{H} \text{Y}}$ e' un metodo di

Greenstadt corrispondente a $M(\omega)$ definita come in (2.10), con ω tale che $\sigma = \varphi(\omega)$. La funzione φ non e' invertibile, quindi la scelta di ω non e' unica.

Se invece al passo i -esimo risulta $\text{S}^T \text{Y} > \text{Y}^T \text{H} \text{Y}$, la funzione $\varphi(\omega)$ e'

limitata superiormente da $\frac{a}{a-1} = \frac{\text{S}^T \text{Y}}{\text{S}^T \text{Y} - \text{Y}^T \text{H} \text{Y}}$, cosi' che solo i

metodi di Broyden con $\sigma < \frac{\text{S}^T \text{Y}}{\text{S}^T \text{Y} - \text{Y}^T \text{H} \text{Y}}$ possono essere interpretati

come metodi di Greenstadt. Si osserva infine che qualunque sia il valore di a , facendo variare opportunamente ω tra $-1/\sqrt{a}$ e 0 , ogni metodo della classe di Broyden ristretta $(0,1)$ e' un particolare metodo di Greenstadt; quindi i metodi della famiglia di Broyden ristretta godono dell'ulteriore proprieta' di minimizzare la norma pesata della correzione (rispetto alla matrice dei pesi $M^{-1}(\omega)$, con $M(\omega)$ definita in (2.10)).

2.2 Metodi di Greenstadt e di Huang simmetrico

Se conduciamo un'analisi analoga per la famiglia a due parametri di Huang, possiamo dedurre che l'intersezione tra la famiglia di Greenstadt e di Huang simmetrico coincide con l'intersezione gia' studiata della famiglia di Greenstadt con quella di Broyden.

Ponendo $u = \frac{\text{S}}{\text{S}^T \text{Y}}$, deve valere la seguente uguaglianza tra

matrici (in analogia con la (2.2)):

$$(2.17) \quad \nu w w^T + \mu z z^T = \mu \sigma v v^T + (1 - \rho) \nu u u^T$$

Ricordando che $v = w - z$, per ogni vettore p deve valere:

$$[\nu (w^T p) - \mu (v^T p)] w + [\mu (z^T p) + \mu \sigma (v^T p)] z + (\rho - 1) \nu (u^T p) u = 0$$

Poiche' u non puo' essere scritto come combinazione lineare di w e z , il numero $\nu (\rho - 1) (u^T p)$ deve annullarsi per ogni p ; questo e' possibile solo se $\rho = 1$.

APPENDICE 1

Sia X una matrice $n \times n$, consideriamo il seguente problema:

$$(a.1) \begin{cases} \min ||M - X||_F^2 \\ M\gamma = b \\ M^T - M = 0 \end{cases}$$

Proposizione 1

Il problema (a.1) ammette la seguente soluzione:

$$(a.2) \quad M = PX^S P + 1/\gamma \gamma^T ((I+P) b \gamma^T)^S, \text{ dove } A^S = 1/2 (A+A^T)$$

e $P = I - \gamma \gamma^T / \gamma^T \gamma$ e' la proiezione ortogonale su $\{ x: x \gamma^T = 0 \}$.

Dim.

La funzione lagrangiana associata al problema risulta:

$$(a.3) \quad L(M) = \text{Tr}(\gamma \gamma^T + M M^T - X M^T - M X^T) + \lambda (M \gamma - b) + \text{Tr}(\Gamma (M^T - M))$$

dove λ e Γ sono il vettore e la matrice dei parametri lagrangiani. Ponendo a zero la derivata di $L(M)$ si ottiene:

$$(a.4) \quad 2M - 2X + \lambda \gamma + \Gamma - \Gamma^T = 0$$

Risolviamo quindi il sistema costituito dalla (a.4) e dai vincoli del problema. Da (a.4) ricaviamo

$$(a.5) \quad M = 1/2 (\Gamma^T - \Gamma - \lambda \gamma^T) + X$$

$$(a.6) \quad M^T = 1/2 (\Gamma - \Gamma^T - \gamma \lambda^T) + X^T$$

Ponendo la differenza tra la (a.5) e la (a.6) uguale a zero si ricava:

$$(a.7) \quad \Gamma - \Gamma^T = X - X^T + 1/2 (\gamma \lambda^T - \lambda \gamma^T) \text{ da cui, sostituendo in (a.5):}$$

$$(a.8) \quad M = X^S - 1/4 (\lambda \gamma^T + \gamma \lambda^T)$$

Deve inoltre valere:

$$(a.9) \quad X^S \gamma - 1/4 (\lambda \gamma^T \gamma + \gamma \lambda^T \gamma) = b$$

Moltiplicando a sinistra per γ^T e svolgendo i calcoli si ottiene:

$$(a.10) \quad \lambda^T \gamma = 2/\gamma^T \gamma (\gamma^T X^S \gamma - \gamma^T b), \text{ da cui sostituendo in (a.9) si ha:}$$

$$(a.11) \quad \lambda = 2/\gamma^T \gamma ((I+P) (X^S \gamma - b)), \text{ dove } P = I - \gamma \gamma^T / \gamma^T \gamma$$

Sostituendo il valore di λ in (a.8) con semplici calcoli si ottiene la soluzione:

$$(a.12) \quad M = PX^S P + 1/\gamma \gamma^T ((I+P) b \gamma^T)^S, \text{ dove } A^S = 1/2 (A+A^T)$$

Proposizione 2

Se il più piccolo autovalore della matrice X^s , s_0 , verifica la seguente relazione:

$$(a.13) \quad s_0 > b^T b$$

la matrice M data in (a.2) risulta definita positiva.

Dim.

Consideriamo un generico vettore di dimensione n non nullo, z . Tale vettore può essere decomposto nella seguente maniera:

$$z = \alpha y + r \quad \text{con } \alpha \in \mathbb{R} \text{ e } r^T y = 0$$

Con semplici passaggi si ottiene:

$$z^T M z = r^T X r + \alpha^2 + 2\alpha r^T b = \|r\|^2 s_0 + \alpha^2 - 2\alpha \|r\| \|b\|$$

Dall'ipotesi (a.11) segue $z^T M z > (\|r\| \|b\| - \alpha)^2 \geq 0$ e questo conclude la dimostrazione.

FIG. 1

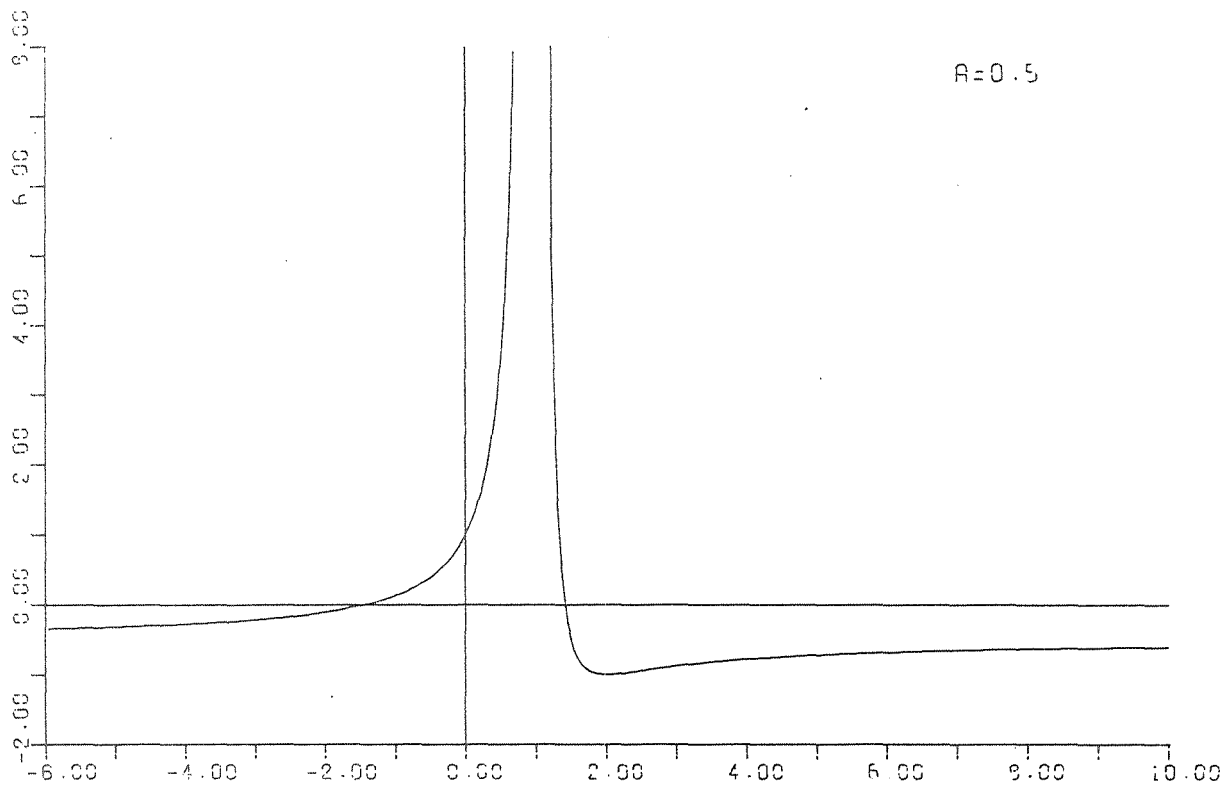
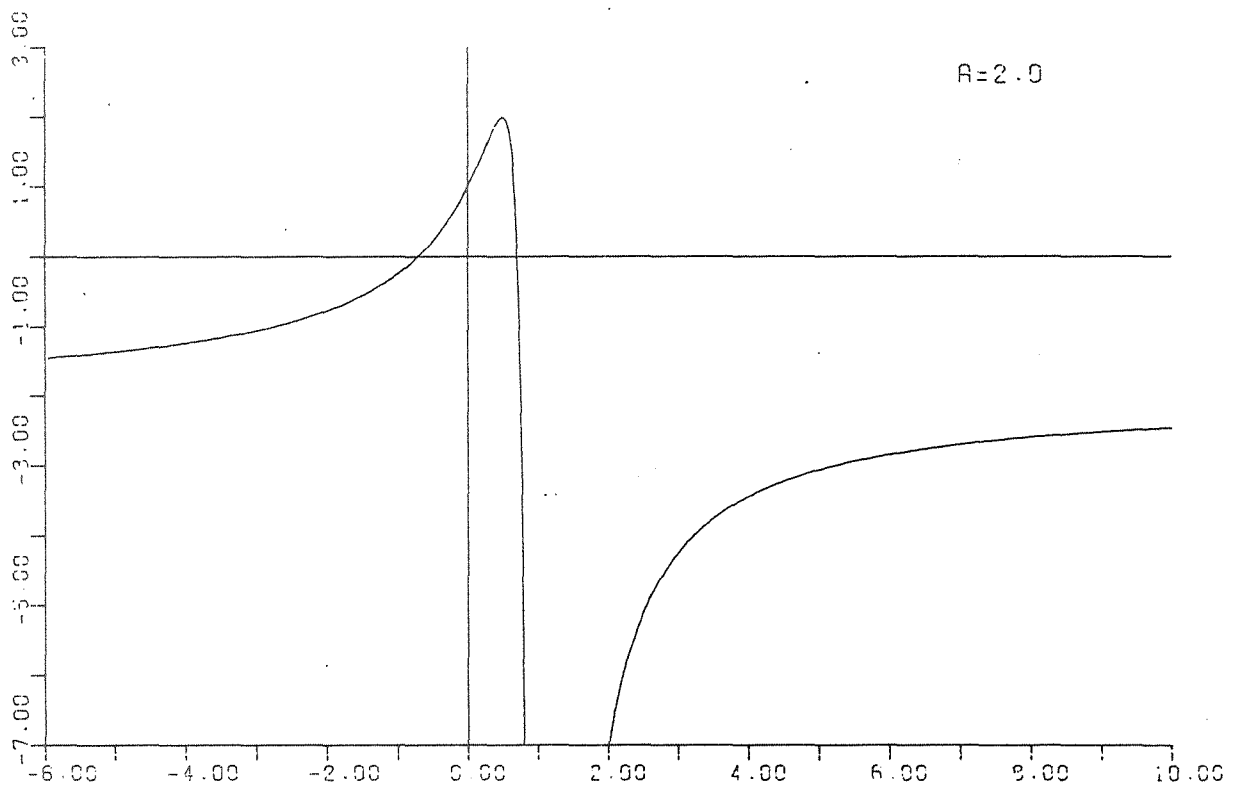


FIG. 2



Bibliografia

- [1] C.G. BROYDEN. Quasi Newton Methods and their Application to function minimization. Math. Comp. Vol.21 (1967) 368-381
- [2] C.G. BROYDEN, J. E. DENNIS JR and J.J. MORE'. On the local and superlinear convergence of quasi-Newton methods. J. Inst. Math Appics Vol.12 (1973) 223-245
- [3] W.C. DAVIDON. Variable metric method for minimization. A.E.C. Research and development report. ANL-5990 (1959)
- [4] L.C.W. DIXON. Variable metric algorithms: necessary and sufficient conditions for identical behavior of nonquadratic function. J. Optim. Theory Appl. Vol.10 (1972) 34-40
- [5] L.C.W. DIXON. Quasi-Newton algorithms generate identical point Math. Programming Vol.2 (1972) 383-387
- [6] R. FLETCHER and M.J.D. POWELL. A rapidly convergent descent method for minimization. Comput. J. Vol.6 (1963) 163-168
- [7] R. FLETCHER. Pratical methods of optimization Vol.1 (1980).
- [8] D. GOLDFARB. A family of variable-metric methods derived by variational means. Math. Comp. Vol.24 (1970) 23-26
- [9] A.A. GOLDSTEIN. On steepest descent. SIAM J. Control Optim. Vol.3 (1963) 147-151
- [10] J. GREENSTADT. Variations on variable-metric methods. Math. Comp. Vol. 24 (1970) 1-22
- [11] H.Y. HUANG. Unified approach to quadratically convergent algorithms for function minimization. J. Optim. Theory Appl. Vol.5 (1970) 405-423
- [12] H.Y. HUANG and A.V. LEVY. Numerical experiments on

- quadratically convergent algorithms for function minimization. J. Optim. Appl. vol.6 (1970) 269-282
- [13] S.S. OREN and D.G. LUENBERGER. Self scaling variable metric (SSVM) algorithms I: criteria and sufficient conditions for scaling a class of algorithms. Management Sci. Vol.20 (1974) 845-862
- [14] S.S. OREN. Self scaling variable metric (SSVM) algorithms II: Implementation and experiments. Management Sci. Vol.20 (1974) 863-874
- [15] S.S. OREN. Self scaling variable metric algorithms without line-search for unconstrained minimization. Math. Comp. Vol.27 (1973) 873-885
- [16] S.S. OREN. On the selection of parameters in self scaling variable metric algorithms. Math. Programming Vol.7 (1974) 351-367
- [17] S.S. OREN and E. SPEDICATO. Optimal conditioning of self scaling variable metric algorithms. Math. Programming Vol. 10 (1976) 70-90
- [18] M.J.D. POWELL. On the convergence of the variable metric algorithm. J. Inst. Maths. Applis Vol.7 (1971) 21-36
scaling variable metric algorithms. Math. Programming Vol.10 (1976) 70-90
- [19] C.R.RAO and S.K. MITRA. Generalized inverse of matrices and its applications. (1971) [20] E. SPEDICATO. On a conjecture of Dixon and other topics in variable metric methods. Math. Programming Vol. 15 (1978) 123-129
- [21] J. STOER. On the convergence rate of imperfect minimization algorithms in Broyden's β -class. Math. Programming Vol.9 (1975) 313-335
- [22] J. STOER. On the relation between quadratic termination and convergence properties of minimization algorithms. Part I.

Theory. Numer. Math. Vol.28 (1977) 343-366

[23] J. STOER. On the relation between quadratic termination and convergence properties of minimization algorithms. Part II. Applications. Numer. Math. Vol.28 (1977) 367-391

[24] J. STOER. The convergence of matrices generated by rank-2 methods from the restricted β -class of Broyden. Numer. Math. Vol. 44 (1984) 53-60

