
BOLLETTINO

UNIONE MATEMATICA ITALIANA

Sezione A – La Matematica nella Società e nella Cultura

GERMANA LANDI

Metodi numerici per la minimizzazione di funzioni

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. 4-A—La Matematica nella Società e nella Cultura (2001), n.3 (Fascicolo Tesi di Dottorato), p. 471–474.

Unione Matematica Italiana

http://www.bdim.eu/item?id=BUMI_2001_8_4A_3_471_0

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

*Articolo digitalizzato nel quadro del programma
bdim (Biblioteca Digitale Italiana di Matematica)
SIMAI & UMI*

<http://www.bdim.eu/>

Metodi numerici per la minimizzazione di funzioni.

GERMANA LANDI

Il lavoro svolto si inserisce nell'ambito dello studio, teorico e numerico, dei metodi per la minimizzazione vincolata di funzioni non lineari. Sono stati presi in esame due metodi fra loro assai differenti: il metodo del lagrangiano aumentato ed un metodo delle direzioni ammissibili. Particolare attenzione è stata rivolta allo studio dei suddetti metodi in presenza di vincoli di disuguaglianza non lineari.

1. - Il metodo del lagrangiano aumentato.

Si consideri il problema

$$(1) \quad \begin{aligned} & \min f(x) \\ & \text{sotto i vincoli: } h_i(x) = 0, \quad i = 1, \dots, m \\ & \quad \quad \quad g_j(x) \leq 0, \quad j = 1, \dots, p \end{aligned}$$

con $f, h_i, g_j: R^n \rightarrow R, i = 1, \dots, m \leq n, j = 1, \dots, p$ funzioni assegnate.

Pierre e Lowe hanno introdotto [3] una funzione lagrangiana aumentata che permette di risolvere il problema (1) con il metodo del lagrangiano aumentato in maniera assai efficace.

Tale funzione lagrangiana aumentata è così definita:

$$(2) \quad \begin{aligned} L_a(x, \lambda, \mu, w) = & f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x) + \\ & + w_1 \sum_{i=1}^m (h_i(x))^2 + w_2 \sum_{j \in C_a} (g_j(x))^2 + w_3 \sum_{j \in C_b} (g_j(x))^2 \end{aligned}$$

con $\mu_j \geq 0, w_i > 0, C_a = \{j | \mu_j > 0\}$ e $C_b = \{j | g_j(x) \geq 0, \mu_j = 0\}$.

L'algoritmo proposto è il seguente: supposti dati i moltiplicatori λ^k e $\mu^k \geq 0$ ed un vettore dei pesi $w^k > 0$, si determina il punto x^k che minimizza su R^n la funzione $L_a(x, \lambda^k, \mu^k, w^k)$; in seguito si aggiornano i moltiplicatori λ^k e μ^k in modo tale che $\nabla f(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^{k+1} \nabla g_j(x^k) = \nabla L_a(x^k, \lambda^k, \mu^k, w^k)$ e $\mu^{k+1} \geq 0$.

Il teorema seguente mette in relazione, per un'ampia classe di problemi, i punti di minimo vincolato di $f(x)$ con quelli di minimo non vincolato di $L_a(x, \lambda, \mu, w)$.

TEOREMA 1. - *Siano dati $x^* \in R^n$ ammissibile per i vincoli, $\lambda^* \in R^m$ e $\mu^* \in R^p$ soddisfacenti con x^* le condizioni di Kuhn-Tucker del primo ordine per il problema (1); infine sia dato un vettore $w > 0$ sufficientemente grande ma finito.*

La funzione $L_a(x, \lambda^, \mu^*, w)$ soddisfa in x^* le condizioni sufficienti del secondo ordine per un minimo locale non vincolato se e solo se il problema (1) sod-*

disfa in x^* , λ^* , μ^* le condizioni sufficienti del secondo ordine per un minimo locale vincolato.

È stato effettuato uno studio comparativo tra $L_a(x, \lambda, \mu, w)$ e la funzione lagrangiana aumentata proposta da Rockafellar [4] per il problema (1) che ha l'espressione:

$$L_R(x, \lambda, \mu, c) = f(x) + \sum_{i=1}^n \lambda_i h_i(x) + \frac{c}{2} \sum_{i=1}^n h_i(x)^2 + \\ + \frac{1}{4c} \sum_{j=1}^n \{[\theta(\mu_j + 2cg_j(x))]^2 - (\mu_j)^2\}$$

con $\theta(t) = \max\{0, t\}$.

Sia $c = w_1 = w_2 = w_3 > 0$ e siano $\lambda \equiv \mu \geq 0$ fissati.

In particolare, vale la relazione:

$$L_a(x, \lambda, \mu, c) - L_R(x, \lambda, \mu, c) = \sum_{\substack{\mu_j > 0 \\ g_j(x) \leq -\frac{\mu_j}{2c}}} \left(\mu_j g_j(x) + cg_j(x)^2 + \frac{\mu_j^2}{4c} \right) \geq 0.$$

Pertanto, se x_R è un punto di minimo locale di $L_R(x, \lambda, \mu, c)$ e se l'insieme $A = \left\{ j \mid \mu_j > 0, g_j(x_R) \leq -\frac{\mu_j}{2c} \right\}$ è vuoto allora, in un intorno di x_R , si ha:

$$L_a(x_R, \lambda, \mu, c) = L_R(x_R, \lambda, \mu, c) \leq L_R(x, \lambda, \mu, c) \leq L_a(x, \lambda, \mu, c)$$

e dunque x_R è minimo locale di $L_a(x, \lambda, \mu, c)$.

Per condurre la sperimentazione numerica è stato utilizzato il pacchetto software reso disponibile da Pierre e Lowe. Allo scopo di analizzare il comportamento dell'algoritmo in presenza di vincoli di disuguaglianza, tutti i problemi test, tratti da [2] e [3], sono stati considerati sia nella forma originale che in quella equivalente con soli vincoli di uguaglianza, ottenuta introducendo delle variabili ausiliarie. I risultati numerici mostrano che il programma di Pierre e Lowe è robusto ed efficiente e, quando lo si utilizza, non è conveniente introdurre le variabili ausiliarie perchè, aumentando le dimensioni del problema, si ha un aumento del costo (inteso come numero di iterazioni compiute dal programma e numero di valutazioni della funzione obiettivo) necessario per ottenere la convergenza.

2. - Un metodo delle direzioni ammissibili.

Si consideri ora il problema (1) senza vincoli di uguaglianza.

È stato analizzato, sia teoricamente che numericamente, un metodo delle direzioni ammissibili proposto da Herskovits. Questo metodo dipende da vari parametri per i quali sono stati forniti nuovi validi criteri di scelta.

Pertanto, con questi criteri l'algoritmo proposto assume la forma seguente.

Si supponga che i punti di $\Omega_\alpha = \{x \mid g_i(x) \leq 0, f(x) \leq \alpha\}$ siano punti regolari, che i punti di Ω_α^0 siano ammissibili e che $f, g \in C^2(\Omega_\alpha)$. Detta $l(x, \lambda) = f(x) +$

$\lambda^t g(x)$ la funzione lagrangiana, siano $H(x, \lambda)$ la sua matrice hessiana e $G(x)$ e A le matrici diagonali tali che $G_{ii}(x) = g_i(x)$ e $A_{ii} = \lambda_i$, per $i = 1, \dots, p$.

Algoritmo

Parametri: $\xi \in (0, 1)$; $\eta \in (0, 1)$; $\varphi > 0$; $\varepsilon \in (0, 1)$; $\bar{g} \in (0, 1)$; $\lambda^I, \omega_1 \in (0, 1)$; $\lambda^S, \omega_2 > 0$ sufficientemente grandi.

Dati Iniziali: $x^0 \in \Omega_a^0, \omega^0 = (\omega, \dots, \omega)^t \in R^p$ con $\omega = \sigma \frac{\sum_{j=1}^p |g_j(x^0)|}{\sigma \in (0, 1]}$ e $B^0 \in R^{n \times n}$ approssimazione simmetrica e definita positiva di $H(x^0, \lambda^0)$.

Passo 1) (Determinazione della direzione di ricerca)

Determinare (d_a^k, λ_a^k) e (d_b^k, λ_b^k) soluzioni rispettivamente di:

$$B^k d_a^k + \nabla g(x^k) \lambda_a^k = -\nabla f(x^k)$$

$$(3) \quad A^k \nabla g(x^k)^t d_a^k + G(x^k) \lambda_a^k = 0$$

e

$$B^k d_b^k + \nabla g(x^k) \lambda_b^k = 0$$

$$(4) \quad A^k \nabla g(x^k)^t d_b^k + G(x^k) \lambda_b^k = A^k \omega^k$$

Porre: $d^k = d_a^k + \varrho^k d_b^k$ e $\mu^k = \lambda_a^k + \varrho^k \lambda_b^k$ con $\varrho^k = \min \left[\varphi \|d_a^k\|_2^2; (\xi - 1) \frac{(d_a^k)^t \nabla f(x^k)}{(d_b^k)^t \nabla f(x^k)} \right]$ se $(d_b^k)^t \nabla f(x^k) > 0$ oppure $\varrho^k = \varphi \|d_a^k\|_2^2$.

Passo 2) (Ricerca del passo) Determinare $t^k \in (0, 1]$ tale che:

$$f(x^k + t^k d^k) \leq f(x^k) + \eta t^k \nabla f(x^k)^t d^k,$$

$$g_i(x^k + t^k d^k) < 0 \text{ se } \mu_i^k \geq 0, \quad i = 1, \dots, p,$$

$$g_i(x^k + t^k d^k) \leq g_i(x^k) \text{ se } \mu_i^k < 0, \quad i = 1, \dots, p,$$

Passo 3) (Aggiornamenti) Porre: $x^{k+1} = x^k + t^k d^k$ e $k = k + 1$. Risolvere:

$$[G(x^k)^2 + \nabla g(x^k)^t \nabla g(x^k)] v = -\nabla g(x^k)^t \nabla f(x^k)$$

e porre:

$$\lambda_i = \begin{cases} \lambda^S & \text{se } v_i > \lambda^S, \\ \lambda^I & \text{se } v_i < \varepsilon \text{ oppure } v_i < \lambda^I \text{ e } g_i(x^k) \geq -\bar{g}, \\ v_i & \text{altrimenti.} \end{cases}$$

Aggiornare ω^k in modo che $\omega_1 \leq \omega_i^k \leq \omega_2, i = 1, \dots, p$.

Porre:

$$B^k = \begin{cases} H(x^k, \lambda^k) & \text{se definita positiva,} \\ B^{k-1} & \text{altrimenti.} \end{cases}$$

Il metodo proposto, a partire da un punto iniziale strettamente ammissibile, genera una successione di iterati x^k strettamente ammissibili nei quali il valore

della funzione obiettivo è ridotto. La direzione d^k , ammissibile per i vincoli e di discesa, è determinata risolvendo i sistemi lineari (3) e (4) che derivano dalle equazioni delle condizioni necessarie del primo ordine di Kuhn-Tucker. L'algoritmo è parallelizzabile: i due sistemi hanno la stessa matrice dei coefficienti e possono essere risolti contemporaneamente. Il metodo risulta pertanto particolarmente adatto per risolvere problemi di grandi dimensioni. Se si assume che i punti stazionari del problema siano punti isolati è possibile provare i seguenti risultati di convergenza.

TEOREMA 2. - *Nelle ipotesi fatte vale: $\{x^k\}_{k \in N} \xrightarrow[k \rightarrow \infty]{} x^*$ e $\{d^k\}_{k \in N} \xrightarrow[k \rightarrow \infty]{} d^* = 0$ dove x^* è un punto di Kuhn-Tucker con moltiplicatore $\lambda_a^* = \lambda_a(x^*)$.*

TEOREMA 3. - *Se $V^* = [G(x^*)^2 + \nabla g(x^*)^t \nabla g(x^*)]$ è non singolare allora:*

$$\{\lambda_a^k\}_{k \in N}, \{\mu^k\}_{k \in N} \xrightarrow[k \rightarrow \infty]{} \lambda_a^*.$$

La sperimentazione numerica è stata condotta su un insieme di problemi test tratti da [2]. Tali problemi, seppur di dimensioni ridotte, hanno permesso di mettere in luce il ruolo svolto dai numerosi parametri del metodo. Dai risultati ottenuti è emerso che essi influenzano notevolmente l'algoritmo. Per chiarire tale affermazione sono riportati i risultati relativi al problema numero 15 di [2]. Per $\xi = 0.7$, $\varphi = 1$ e $\mu = 0.3$ sono necessarie 33 iterazioni per la convergenza: il punto iniziale è vicino al bordo della regione ammissibile, le direzioni d^k hanno componenti grandi in modulo e quindi i passi t^k risultano assai piccoli. È possibile variare i valori dei parametri in modo da «diminuire» il valore delle componenti di d^k «troppo grandi» e rendere possibile la scelta di un passo t^k più vicino ad uno. L'assegnazione: $\varphi = 0.0001$ e $\xi = 0.99$ comporta una riduzione del valore di q^k ; in questo caso la convergenza è raggiunta in 8 iterazioni. In entrambi i casi l'errore compiuto sul valore minimo di $f(x)$ è dell'ordine di 10^{-2} . È stato inoltre rilevato che il metodo, quando gli iterati x^k sono molto vicini al bordo della regione ammissibile, può avere una convergenza molto lenta o addirittura può divergere.

BIBLIOGRAFIA

- [1] HERSKOVITS J., *Feasible direction interior-point technique for nonlinear optimization*, J.O.T.A., **99** (1998), 121-146.
- [2] HOCK W. and SCHITTKOWSKI K., *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, Springer Verlag, **187** (1981).
- [3] PIERRE D. A. and LOWE M. J., *Mathematical Programming Via Augmented Lagrangians*, Addison-Wesley (1975).
- [4] ROCKAFELLAR R. T., *A dual approach to solving nonlinear programming problems by unconstrained optimization*, M.P., **5** (1973), 354-373.

Dipartimento di Matematica, Università di Bologna
e-mail: landig@dm.unibo.it

Dottorato in Matematica Computazionale (sede amministrativa: Padova) - Ciclo XIII
Direttore di Ricerca: Prof. I. Galligani, Università di Bologna