
BOLLETTINO

UNIONE MATEMATICA ITALIANA

Sezione A – La Matematica nella Società e nella Cultura

GIAMPAOLO LIUZZI

Nuovi metodi per l'ottimizzazione vincolata non lineare

*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. 479–482.*

Unione Matematica Italiana

[<http://www.bdim.eu/item?id=BUMI_2001_8_4A_3_479_0>](http://www.bdim.eu/item?id=BUMI_2001_8_4A_3_479_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/>

Nuovi metodi per l'ottimizzazione vincolata non lineare.

GIAMPAOLO LIUZZI

1. - Introduzione.

L'argomento di questa tesi è quello di definire nuovi metodi per la soluzione di problemi di ottimizzazione vincolata non lineare a grandi dimensioni. Dato il problema di ottimizzazione vincolata (P)

$$\begin{aligned} \min f(x) \\ g(x) \leq 0 \end{aligned}$$

dove $f: R^n \mapsto R$ e $g: R^n \mapsto R^m$ sono funzioni due volte continuamente differenziabili, consideriamo il caso in cui il numero di variabili n e/o il numero dei vincoli m sono grandi. L'insieme dei punti a cui siamo interessati ovvero il nostro insieme delle soluzioni è l'insieme dei punti stazionari del problema (P). Sotto certe ipotesi «di regolarità», diciamo che un punto x^* è un punto stazionario di (P) quando è possibile trovare m numeri λ_i , $i = 1, \dots, m$ (moltiplicatori di Lagrange) tali che:

$$(1) \quad \nabla f(x^*) + \sum_{i=1}^m \nabla g_i(x^*) \lambda_i = 0, \quad g(x^*) \leq 0, \quad \lambda^\top g(x^*) = 0, \quad \lambda \geq 0.$$

Un generico algoritmo per la soluzione del problema (P) può essere schematizzato con l'iterazione

$$x^{k+1} := x^k + \alpha^k d^k$$

in cui: (a) x^k è il punto corrente ovvero l'approssimazione corrente del punto soluzione, (b) $d^k \in R^n$ è una direzione di spostamento (o di ricerca), (c) α^k è un passo effettuato lungo d^k ed infine (d) x^{k+1} è il nuovo punto determinato dall'algoritmo e che dovrebbe meglio approssimare la soluzione x^* .

Un buon algoritmo di soluzione per il problema (P) deve soddisfare almeno i seguenti due requisiti:

- essere «globalmente» convergente cioè avere la capacità di determinare un punto stazionario x^* qualunque sia il punto iniziale x^0 ;

- avere una rapidità di convergenza almeno «superlineare» cioè essere tale per cui la norma dell'errore $\|x^k - x^*\|$ tenda a zero in maniera sufficientemente veloce.

Molti algoritmi presenti in letteratura, pur avendo una buona rapidità di convergenza, sono solo localmente convergenti ovvero in grado di determinare un punto stazionario solo se il punto iniziale x^0 è sufficientemente vicino ad x^* . È

possibile, anche se non banale, a partire da questi, ottenere algoritmi che siano anche globalmente convergenti. Tuttavia, come dimostrano Marazzi e Nocedal [1], nel globalizzare un algoritmo bisogna sempre garantire un «sufficiente controllo dell'ammissibilità» ovvero prevedere un meccanismo che sia in grado di forzare l'ammissibilità dei punti prodotti dall'algoritmo stesso. La mancanza di un tale meccanismo porta alla definizione di algoritmi assolutamente inaffidabili anche quando applicati alla soluzione di problemi «semplici».

Un tale meccanismo di controllo può essere facilmente ottenuto mediante l'utilizzo di funzioni di merito esatte continuamente differenziabili dipendenti da un parametro di penalità (ϵ) con il quale è possibile pesare la violazione delle condizioni di stazionarietà del punto corrente. Tali funzioni di merito sono intrinsecamente capaci di combinare i due sottoproblemi che sono alla base del problema vincolato (P), ossia quello di minimizzare la funzione obiettivo e quello di garantire l'ammissibilità dei punti generati (almeno al limite).

Generalmente, nell'ambito delle funzioni di merito esatte, se il numero m dei vincoli di (P) è contenuto, è preferibile l'utilizzo di una funzione di penalità esatta. Se, invece, il numero dei vincoli è grande, allora è generalmente accettabile l'uso di una funzione Lagrangiana aumentata esatta che però necessita di essere minimizzata sullo spazio esteso delle variabili di decisione (primali) e dei moltiplicatori di Lagrange o variabili duali.

Il costo computazionale associato ad entrambe le precedenti funzioni di merito esatte, unitamente al loro possibile mal condizionamento ha limitato notevolmente, in letteratura, il loro utilizzo nella definizione di algoritmi di soluzione per il problema (P) nel caso di grandi dimensioni. Il lavoro svolto nella tesi trae spunto dalle precedenti considerazioni e si concreta nella definizione di nuovi metodi globalmente e superlinearmente convergenti, particolarmente adatti a trattare problemi di grandi dimensioni, unendo alle potenzialità teoriche delle predette funzioni di merito esatte, l'utilizzo di algoritmi locali estremamente efficienti. Più specificamente, questi algoritmi locali possono essere visti come l'applicazione del metodo di Newton alla soluzione di sistemi di equazioni, che approssimano le condizioni di stazionarietà (1), anche note come condizioni di ottimalità di Karush-Kuhn-Tucker (KKT), per fornire delle buone direzioni di ricerca. La chiave di volta che permette di raccordare questi metodi locali con l'uso di funzioni di merito esatte e quindi ottenere algoritmi globalmente e superlinearmente convergenti, è il fatto, provato nella tesi, che le direzioni generate dall'algoritmo locale sono in effetti direzioni di tipo Newton per la funzione di merito.

2. – Approccio di tipo Penalità esatta.

Supponiamo che il numero dei vincoli m sia piccolo. Come detto, in questo caso, è consigliabile l'uso di una funzione di penalità esatta continuamente differenziabile. La funzione di penalità adottata è la stessa introdotta e studiata da Con-

taldi, Di Pillo e Lucidi in [2]. L'innovazione dell'algoritmo proposto consiste nel calcolo della direzione di ricerca ovvero nella fase locale del metodo. La direzione d^k è calcolata usando un metodo di Newton troncato particolarmente adatto alla soluzione di sistemi di grandi dimensioni.

Il calcolo della direzione d^k si basa su una stima $A(x^k)$ dell'insieme degli indici dei vincoli attivi nel punto corrente, ovvero dei vincoli soddisfatti all'uguaglianza in x^k . La direzione d^k è definita come soluzione troncata del seguente sistema di equazioni

$$\begin{bmatrix} \nabla_x^2 L(x^k, \lambda(x^k)) & \nabla g_{A^k}(x^k) \\ \nabla g_{A^k}(x^k)^\top & 0 \end{bmatrix} \begin{bmatrix} d^k \\ z^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) \\ g_{A^k}(x^k) \end{bmatrix}.$$

Tuttavia, piuttosto che usare il metodo di Newton per risolvere l'intero sistema, si è preferito separare la direzione d^k in due componenti (vedi p.es. [3]) $d^k = d_v^k + d_o^k$, ovvero esprimere d^k come somma di un passo verticale ed uno orizzontale ortogonali tra loro. Questo procedimento ci ha consentito di isolare la parte difficile nel calcolo della direzione e cioè la determinazione del passo orizzontale. Al contrario, il passo verticale può essere facilmente ottenuto risolvendo il sistema di equazioni lineari $\nabla g_{A^k}(x^k)^\top d_v^k = -g_{A^k}(x^k)$. Una volta noto il passo verticale della direzione, quello orizzontale è ottenibile applicando uno schema di Newton troncato, basato su un gradiente coniugato, alla soluzione del sistema di equazioni

$$\mathcal{P}^k(\nabla_x^2 L(x^k, \lambda(x^k)) d_o^k) = - \mathcal{P}^k(\nabla_x^2 L(x^k, \lambda(x^k)) d_v^k + \nabla f(x^k)),$$

dove con \mathcal{P}^k si intende l'operatore di proiezione nel nullo dello Jacobiano dei vincoli stimati attivi nel punto x^k . Il vantaggio di questo approccio, oltre alla evidente predisposizione per le grandi dimensioni consiste anche nel fatto che tutti i passi necessari al calcolo della direzione sono eseguibili fattorizzando una volta per tutte una matrice quadrata, simmetrica e sparsa.

3. - Approccio di tipo Lagrangiano aumentato esatto.

Supponiamo ora che il numero dei vincoli m sia anch'esso grande. In questo caso, è necessario l'utilizzo di una funzione Lagrangiana aumentata esatta. Tuttavia, considerato il fatto che, sempre più frequentemente, si riscontrano problemi con la seguente struttura (P_b)

$$\begin{aligned} \min f(x) \\ l(x)g \leq u \end{aligned}$$

con $l, u \in R^m$, lo sforzo maggiore di questa seconda parte della tesi è stato quello di definire una nuova funzione Lagrangiana aumentata in grado di sfruttare la particolare struttura del problema (P_b). Normalmente, quando il problema presenta vincoli come questi, essi vengono separati nella coppia di vincoli $-g(x) \leq -l$ e $g(x) \leq u$ seguendo quindi l'approccio classico con conseguente introduzione di

$2m$ moltiplicatori di Lagrange. Visto però che, per il problema (P_b) , è possibile dare condizioni di ottimalità usando solo m moltiplicatori, la precedente non è l'unica alternativa disponibile e certamente non è la più economica.

Allo scopo di trattare ogni doppio vincolo come un unico vincolo, definiamo, per ogni ε positivo, le seguenti funzioni

$$\gamma_i(x, \lambda; \varepsilon) = \begin{cases} g_i(x) - u_i & \text{se } g_i(x) - u_i \geq -\varepsilon p(x, \lambda) \lambda_i \\ g_i(x) - l_i & \text{se } g_i(x) - l_i \leq -\varepsilon p(x, \lambda) \lambda_i \\ -\varepsilon p(x, \lambda) \lambda_i & \text{altrimenti,} \end{cases} \quad i = 1, \dots, m$$

dove $p(x, \lambda)$ è una funzione di *barriera*, tale che cioè (a) $p(x, \lambda) > 0 \forall (x, \lambda) \in S \times \mathbb{R}^m$, essendo S un rilassamento aperto dell'insieme ammissibile del problema (P_b) (b) $\lim_{x \rightarrow \partial S} p(x, \lambda) = 0$ e (c) $\lim_{\|\lambda\| \rightarrow \infty} p(x, \lambda) = 0$.

Con queste premesse è stato possibile definire la seguente funzione Lagrangiana aumentata continuamente differenziabile

$$L_a(x, \lambda; \varepsilon) = f(x) + \lambda^\top \gamma(x, \lambda; \varepsilon) + \frac{1}{2\varepsilon p(x, \lambda)} \|\gamma(x, \lambda; \varepsilon)\|^2 +$$

$$\|\nabla g(x)^\top \nabla_x L(x, \lambda) + (G(x) - U)^2 (G(x) - L)^2 \lambda\|^2$$

dove $L(x, \lambda) = f(x) + \lambda^\top g(x)$ è la funzione Lagrangiana associata al problema (P_b) e dove si sono indicate con $G(x)$, U ed L le matrici diagonali costituite dagli elementi dei vettori $g(x)$, u ed l rispettivamente.

Seguendo Di Pillo e Lucidi in [4], per la funzione $L_a(x, \lambda; \varepsilon)$ sono state dimostrate le proprietà che la configurano come funzione di merito esatta. Oltre a questo, è stata condotta una esauriente analisi delle proprietà del secondo ordine.

BIBLIOGRAFIA

- [1] MARAZZI M. e NOCEDAL J., *Feasibility control in nonlinear optimization*, Tech. Rep. OTC 2000/04, Optimization Technology Center, Evanston, IL, USA, (2000).
- [2] CONTALDI G., DI PILLO G. e LUCIDI S., *A continuously differentiable exact penalty function for nonlinear programming problems with unbounded feasible set*, Operations Research Letters, **14** (1993), 153-161.
- [3] BYRD R. H., HRIBAR M. E. e NOCEDAL J., *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optimization, **9** (1999), 877-900.
- [4] DI PILLO G. e LUCIDI S., *An augmented lagrangian function with improved exactness properties*, To be published on SIAM J. Optimization (2001).

Dipartimento di Informatica e Sistemistica «Antonio Ruberti»,
Università di Roma «La Sapienza»
e-mail: liuzzi@dis.uniroma1.it

Dottorato in Ricerca Operativa (sede amministrativa: Roma «La Sapienza») - Ciclo XIII
Direttore di ricerca: Prof. Stefano Lucidi, Università di Roma «La Sapienza»