Università di Catania

L.M. in Informatica

Ottimizzazione

Algoritmi per problemi di ottimizzazione

I problemi di ottimizzazione che si presentano nella pratica sono di solito molto complessi da risolvere (elevato numero di variabili e vincoli, funzioni non lineari) per cui non è possibile determinare una soluzione per via analitica. La soluzione analitica è possibile solo nel caso di poche variabili e di funzioni estremamente semplici. Nella realtà, per risolvere un problema di ottimizzazione occorre fare ricorso ad un algoritmo iterativo, cioè data una approssimazione corrente x_k della soluzione, si determina una nuova approssimazione x_{k+1}. A partire da una approssimazione iniziale x_0 si determina così una successione di punti.

Algoritmi per problemi di ONL non vincolata

Si consideri il problema di ottimizzazione non lineare:

$$\min_{x \in \mathbb{R}^n} f(x) = f(x^*).$$

I metodi iterativi di ricerca partono da un punto $x^0 \in \mathbb{R}^n$ e generano una successione infinita di punti $\{x^k\}_{k \in \mathbb{N}}$ che converge ad un punto appartenente ad un insieme che possiede una proprietà assegnata, detto insieme bersaglio. Tale insieme, nel caso di problemi non vincolati, coincide con l'insieme dei punti stazionari (insieme dei punti che soddisfano la condizione necessaria di ottimalità del primo ordine $\nabla f(x) = 0$).

Metodi basati sulle direzioni di ricerca

I metodi basati sulle direzioni di ricerca (line search) individuano, ad ogni iterazione, la lunghezza ottimale di spostamento lungo la direzione di ricerca scelta. Gli algoritmi di questo tipo si dividono in: algoritmi senza l'uso di derivate (di cui non ci occupiamo) e algoritmi con l'uso di derivate (di cui ci occupiamo).

Schema generale

- Si scelga una soluzione ammissibile iniziale x^0 e si ponga $k = 0$.
- Se la soluzione x^k verifica una condizione di arresto STOP. Altrimenti si scelga la direzione di ricerca $d^k \in \mathbb{R}^n$ e si determini il passo α^k minimizzando la funzione obiettivo lungo la direzione d^k, cioè risolvendo il problema

$$\min_{\alpha} f(x^k + \alpha d^k).$$

- La nuova approssimazione del punto di minimo è data da:

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

Il calcolo del passo può essere effettuato con ricerca esatta o inesatta. Se la funzione obiettivo è quadratica, il passo ottimale con ricerca esatta è dato da

$$\alpha_k = -\frac{\nabla f(x^k)^T d^k}{{d^k}^T H(x^k) d^k},$$

dove $H(x^k)$ è la matrice hessiana calcolata in x^k.
La condizione di arresto è in genere: $\|\nabla f(x^k)\| < \varepsilon$, $|f(x^{k+1}) - f(x^k)| < \varepsilon$, oppure $|x^{k+1} - x^k| < \varepsilon$, $\varepsilon > 0$.
I vari algoritmi di ricerca si differenziano per la scelta della direzione.

Metodo del gradiente (steepest descent)

Per trovare i punti stazionari di una funzione $f : \mathbb{R}^n \to \mathbb{R}$, con derivate parziali prime continue su $\mathbb{R}^n$, si costruiscono successioni minimizzanti $x^{k+1} = x^k + \alpha^k d^k$, dove la direzione di ricerca d^k è la direzione di massima discesa per la funzione obiettivo, ossia $d^k = -\nabla f(x^k)$ e il passo è scelto risolvendo il problema

$$\min_{\alpha>0} f\left(x^k - \alpha\nabla f(x^k)\right).$$

cioè effettuando una ricerca monodimensionale lungo la direzione di massima discesa $-\nabla f(x^k)$.

Passi dell'algoritmo

1. Si scelga un punto x^0 e si ponga $k = 0$;
2. Si scelga la direzione $d^k = -\nabla f(x^k)$, si calcoli α^k tale che risolva
$$\min_{\alpha > 0} f\left(x^k - \alpha \nabla f(x^k)\right),$$
e si ponga $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$.
3. Se è verificato un criterio di arresto, STOP. Altrimenti si ponga $k = k + 1$ e si torni al passo 2.

Se la funzione obiettivo è quadratica, il passo ottimale con ricerca esatta è dato da

$$\alpha_k = \frac{\|\nabla f(x^k)\|^2}{\nabla f(x^k)^T H(x^k) \nabla f(x^k)},$$

dove $H(x^k)$ è la matrice hessiana calcolata in x^k.

Il metodo è molto semplice e converge sotto ipotesi molto deboli. È purtroppo molto lento, a causa della scarsa precisione con cui si dirige verso l’ottimo, il metodo può infatti avere un comportamento a zig-zag. Si può ovviare a questo inconveniente considerando una direzione perturbata.

Metodo di Newton

La funzione f da minimizzare viene approssimata localmente da una funzione quadratica. La funzione f è approssimata dalla funzione $\phi(x)$ che si ottiene dallo sviluppo in serie di Taylor di $f(x)$ arrestato ai termini del secondo ordine. Si minimizza quindi la funzione:

$$\phi(x) = f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T H(x^k)(x - x^k),$$

dove H è l'hessiana di f in x^k. Il punto successivo si ottiene minimizzando ϕ ed è dato da:

$$x^{k+1} = x^k - [H(x^k)]^{-1}\nabla f(x^k),$$

dove $H(x^k)]^{-1}$ è la matrice inversa della matrice hessiana.

Passi dell'algoritmo

1. Si scelga un punto x^0 e si ponga $k = 0$;
2. Si calcoli
$$x^{k+1} = x^k - [H(x^k)]^{-1} \nabla f(x^k).$$
3. Se è verificato un criterio di arresto, STOP. Altrimenti si ponga $k = k + 1$ e si torni al passo 2.

Convergenza

Converge rapidamente all'ottimo purchè si parta da un punto sufficientemente prossimo alla soluzione ottima, altrimenti può non convergere. Il metodo è computazionalmente costoso, essendo necessario calcolare ad ogni passo il gradiente della funzione e la matrice inversa della matrice hessiana. Questi inconvenienti possono essere superati utilizzando i metodi Quasi-Newtoniani che si basano sull'utilizzo di un'approssimazione della matrice inversa della matrice hessiana.

Algoritmi per problemi di ONL vincolata

Consideriamo il problema:

$$\begin{aligned}
&\min \; f(x) \\
&h(x) = 0 \\
&g(x) \leq 0 \\
&x \in S \subseteq \mathbb{R}^n,
\end{aligned}$$

dove $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$.

I metodi risolutivi si dividono in due categorie. La prima comprende i metodi basati sulla trasformazione del problema vincolato in un problema non vincolato o in una successione di problemi non vincolati. La seconda categoria comprende i metodi basati sulla trasformazione del problema vincolato in una successione di problemi di programmazione quadratica (di cui non ci occupiamo).

Metodo delle funzioni di penalità

Il metodo delle funzioni di penalità trasforma un problema vincolato in un problema non vincolato o in una successione di problemi non vincolati. I vincoli sono inseriti nella funzione obiettivo mediante un parametro di penalità, che penalizza la loro violazione. A seconda della funzione penalità scelta si possono individuare funzioni di penalità esterne o interne.

Funzioni di penalità sequenziali esterne

Con il metodo di penalità con funzioni sequenziali esterne, si considera una successione monotona divergente $\{c_i\}$, con $c_i \geq 0$ e $c_{i+1} > c_i$, $\forall i$, e si risolve il problema non vincolato per ogni i:

$$\min_{x \in \mathbb{R}^n} (f(x) + c_i P(x)).$$

La funzione di penalità $P(x)$ è tale che

$$P(x) \begin{cases} = 0 & \text{se } x \text{ è ammissibile} \\ > 0 & \text{altrimenti.} \end{cases}$$

Si risolve quindi una successione di problemi non vincolati con un metodo di discesa.

Dato il problema:

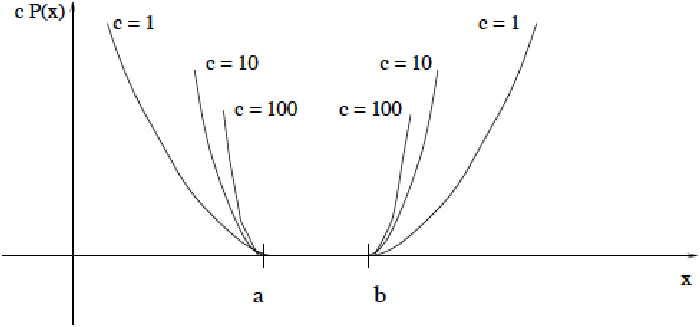
$$\begin{aligned}
&\min \ f(x)\\
&h(x) = 0\\
&g(x) \leq 0\\
&x \in S \subseteq \mathbb{R}^n,
\end{aligned}$$

la funzione di penalità $P(x)$ è in genere costruita come:

$$P(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\}^2 + \sum_{j=1}^{p} h_j^2(x).$$

Con questa scelta, le funzioni di penalità sono differenziabili.

La funzione di penalizzazione è esterna in quanto la successione di minimizzatori x_i tende al punto di minimo dall'esterno della regione ammissibile.



Le funzioni di penalità sequenziali esterne sono facili da implementare, tuttavia non sono esatte. Infatti i punti stazionari del problema penalizzato non coincidono con i ounti KKT del problema iniziale. Sotto opportune ipotesi l'algoritmo converge al minimo del problema originario quando c_i tende all'infinito.

Funzioni di penalità esatte non differenziabili

Il metodo di penalità con funzioni non differenziabili considera per ogni i il problema non vincolato

$$\min_{x \in \mathbb{R}^n} (f(x) + c_i P(x))$$

dove la funzione di penalità $P(x)$ è ancora tale che

$$P(x) \begin{cases} = 0 & \text{se } x \text{ è ammissibile} \\ > 0 & \text{altrimenti,} \end{cases}$$

ma non è differenziabile. In genere si pone:

$$P(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\} + \sum_{j=1}^{p} h_j(x).$$

Queste funzioni sono sempre esterne e sono esatte. Infatti, sotto ipotesi opportune, i punti di minimo della funzione $f(x)$ e di $f(x) + c_i P(x)$ coincidono (per c_i sufficientemente grande). Inoltre, i punti che soddisfano le condizioni KKT della funzione $f(x)$ sono punti stazionari per la funzione $f(x) + c_i P(x)$.

Metodo delle funzioni barriera

Le funzioni barriera vengono utilizzate per trasformare un problema vincolato in un problema non vincolato oppure in una successione di problemi non vincolati. Le funzioni barriera hanno lo scopo di impedire che la soluzione del problema non vincolato si trovi al di fuori della regione ammissibile. Per tale ragione si applicano a problemi con interno della regione ammissibile non vuoto, cioè del tipo

$$\begin{aligned}
&\min \; f(x) \\
&g_i(x) \leq 0, \quad i = 1, \dots, m \\
&x \in S \subseteq \mathbb{R}^n, \\
&\text{con } X = \{x : g_i(x) < 0, i = 1, \dots, m\} \neq \emptyset.
\end{aligned}$$

Le funzioni barriera più utilizzate sono la barriera inversa $B_1(x)$ e la barriera logaritmica $B_2(x)$:

$$B_1(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)}, \quad B_2(x) = -\sum_{i=1}^{m} \ln(-g_i(x)).$$

In generale, per ogni i, si considera il problema:

$$\min(f(x) + \frac{1}{c_i} B(x))$$

con x interno alla regione ammissibile e c_i successione monotona divergente con $c_i \geq 0$ per ogni i e $c_{i+1} > c_i$. $B(x)$ è una funzione continua non negativa e tende all'infinito se x tende alla frontiera.

La funzione barriera tende al punto di minimo dall'interno.