Programmazione non-lineare: ottimizzazione vincolata

Giovanni Righini

Ricerca Operativa

Università degli Studi di Milano

Ottimizzazione vincolata

Nell'ottimizzazione non-lineare vincolata, oltre alla funzione obiettivo

$$\textit{minimize} \quad f(x),$$

consideriamo anche l'effetto di

- vincoli di uguaglianza $h_i(x) = 0 \ \forall i \in \mathcal{E}$
- vincoli di disuguaglianza $g_i(x) \geq 0 \ \forall i \in \mathcal{I}$.

Un vincolo di disuguaglianza $j \in \mathcal{I}$ è attivo in una soluzione $\overline{x}$ se e solo se $g_j(\overline{x}) = 0$.

L'insieme attivo $A(\overline{x})$ è l'insieme dei vincoli attivi in $\overline{x}$.

In ogni punto $\overline{x}$ ammissibile, $A(\overline{x})$ comprende sempre tutti i vincoli di uguaglianza.

Vincoli di uguaglianza

Consideriamo un vincolo $c(x) = 0$ ed un punto $\overline{x}$ su di esso.

Indichiamo con $\nabla c(\overline{x})$ la direzione della normale al vincolo in $\overline{x}$.

Consideriamo un passo infinitesimo da $\overline{x}$ lungo una direzione d.

Per mantenere l'ammissibilità rispetto al vincolo, d deve essere tale che:

$$\nabla c(\overline{x})^T d = 0.$$

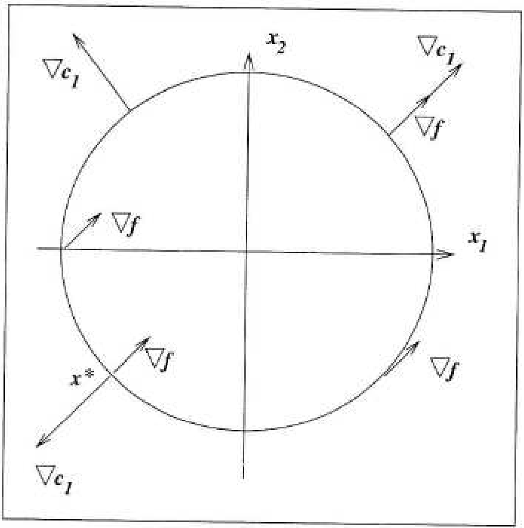
Il passo produce un miglioramento nel valore di $f(x)$ (da minimizzare) se e solo se

$$\nabla f(\overline{x})^T d < 0.$$

Quindi un passo migliorante da $\overline{x}$ *non* è possibile se e solo se

$$\exists \overline{\lambda} \neq 0 : \nabla c(\overline{x}) = \overline{\lambda} \nabla f(\overline{x})$$

Vincoli di uguaglianza



Vincoli di uguaglianza

Un modo alternativo di formulare la stessa condizione di ottimalità in un punto $\overline{x}$ consiste nell'introdurre la funzione Lagrangiana

$$\mathcal{L}(x, \lambda) = f(x) - \lambda c(x).$$

Si ha

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) - \lambda \nabla c(x).$$

Quindi la condizione di ottimalità in $\overline{x}$

$$\exists \overline{\lambda} \neq 0 : \nabla c(\overline{x}) = \overline{\lambda} \nabla f(\overline{x})$$

equivale alla condizione

$$\exists \overline{\lambda} \neq 0 : \nabla_x \mathcal{L}(\overline{x}, \overline{\lambda}) = 0.$$

Si tratta di una condizione necessaria del primo ordine, ma non sufficiente (proprio come nel caso non-vincolato).

Vincoli di disuguaglianza

Consideriamo un vincolo di disuguaglianza $g(x) \geq 0$ ed un punto $\overline{x}$ su di esso.

Il gradiente $\nabla g(x)$ è un vettore che punta verso l'interno della regione ammissibile, dato che il vincolo è nella forma $g(x) \geq 0$ (se $f(x)$ fosse da massimizzare porremmo i vincoli di disuguagalianza nella forma $g(x) \leq 0$).

Il punto $\overline{x}$ non è ottimo se esiste uno spostamento infinitesimo d tale da migliorare il valore dell'obiettivo e da mantenere l'ammissibilità, cioè tale che

$$\nabla f(\overline{x})d < 0$$

e

$$\nabla g(\overline{x})d \geq 0.$$

Tali condizioni non possono essere vere entrambe solo se

$$\exists \overline{\lambda} \geq 0 : \nabla f(\overline{x}) = \overline{\lambda} \nabla g(\overline{x}).$$

Vincoli di disuguaglianza

Quando invece il punto $\overline{x}$ non è sul vincolo, allora si può avere uno spostamento infinitesimo d ammissibile e migliorante quando

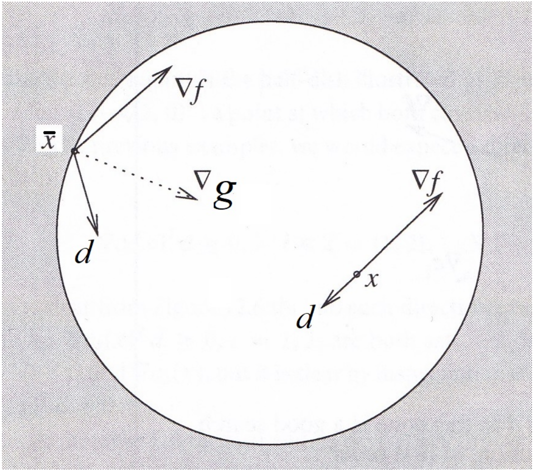
$$\nabla f(\overline{x})d < 0$$

e d è abbastanza piccolo da non superare lo slack del vincolo.

Quindi la condizione necessaria del primo ordine per l'ottimalità in $\overline{x}$ è la stessa del caso non-vincolato

$$\nabla f(\overline{x}) = 0.$$

Vincoli di disuguaglianza



Vincoli di disuguaglianza

Un modo alternativo di formulare la stessa condizione di ottimalità in un punto $\overline{x}$ consiste nell'introdurre la funzione Lagrangiana

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x).$$

Si ha

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) - \lambda \nabla g(x).$$

Quindi la condizione di ottimalità in $\overline{x}$

$$\begin{cases} \exists \overline{\lambda} \geq 0 : \nabla f(\overline{x}) = \overline{\lambda} \nabla g(\overline{x}) & \text{se } g(\overline{x}) = 0 \\ \nabla f(\overline{x}) = 0 & \text{se } g(\overline{x}) > 0 \end{cases}$$

equivale alle condizioni

$$\exists \overline{\lambda} \geq 0 : \nabla_x \mathcal{L}(\overline{x}, \overline{\lambda}) = 0$$

$$\overline{\lambda} g(\overline{x}) = 0$$

Due vincoli di disuguaglianza

Consideriamo due vincoli di disuguaglianza $g_1(x) \geq 0$, $g_2(x) \geq 0$ ed un punto $\overline{x}$, dove entrambi sono attivi.

Indichiamo con $\nabla g_1(\overline{x})$ e $\nabla g_2(\overline{x})$ la direzione della normale ai vincoli in $\overline{x}$.

Consideriamo un passo infinitesimo da $\overline{x}$ lungo una direzione d.

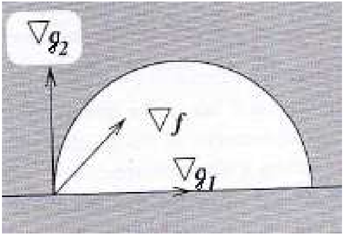
Per l'ammissibilità, d deve essere tale che:

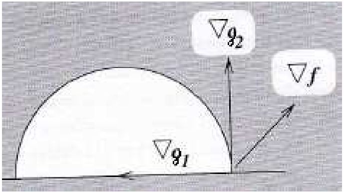
$$\nabla g_1(\overline{x})^T d \geq 0 \quad \text{e} \quad \nabla g_2(\overline{x})^T d \geq 0.$$

Il passo produce un miglioramento di $f(x)$ se e solo se

$$\nabla f(\overline{x})^T d < 0.$$

Vincoli di disuguaglianza





Direzioni ammissibili

Una direzione d è ammissibile in $\overline{x}$ se e solo se:

$$\nabla c_i(\overline{x})^T d = 0 \;\; \forall i \in \mathcal{E} \quad \text{e} \quad \nabla c_i(\overline{x})^T d \geq 0 \;\; \forall i \in \mathcal{A}(\overline{x}) \cap \mathcal{I},$$

dove $\mathcal{A}(\overline{x})$ indica l'insieme dei vincoli di disuguaglianza attivi in $\overline{x}$.

Proprietà *linear independence constraint qualification (LICQ)* in un punto $\overline{x}$: tutti i gradienti dei vincoli attivi in $A(\overline{x})$ sono linearmente indipendenti.

Condizioni di ottimalità del primo ordine

Definita la funzione Lagrangiana

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

si hanno le seguenti condizioni necessarie del primo ordine affinché un punto sia un minimo locale.

Condizioni di Karush-Kuhn-Tucker (KKT). Se

- x^* è un minimo locale di $f(x)$,
- $f(x)$ e $c_i(x)$ sono funzioni continue e differenziabili,
- la LICQ è soddisfatta in x^*,

allora esiste λ^* tale che

$$\begin{aligned}
&\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 && \\
&c_i(x^*) = 0 && \forall i \in \mathcal{E} \\
&c_i(x^*) \geq 0 && \forall i \in \mathcal{I} \\
&\lambda_i^* \geq 0 && \forall i \in \mathcal{I} \\
&\lambda_i^* c_i(x^*) = 0 && \forall i \in \mathcal{E} \cup \mathcal{I}
\end{aligned}$$

Complementarità

Le condizioni di complememtarità

$$\lambda_i^* c_i(x^*) = 0 \quad \forall i \in \mathcal{E} \cup \mathcal{I}$$

richiedono che

- o il vincolo $c_i(x)$ sia attivo,
- o il corrispondente moltiplicatore λ_i sia nullo,
- o entrambe le cose.

Poiché $\lambda_i^* = 0 \, \forall i \notin A(x^*)$, la condizione del primo ordine si può riscrivere come

$$\nabla f(x^*) - \sum_{i \in A(x^*)} \lambda_i^* \nabla c_i(x^*) = 0$$

Complementarità stretta

Si ha complementarità stretta quando solo una tra λ_i^* e $c_i(x^*)$ è nulla $\forall i \in A(x^*)$.

Per uno stesso punto x^* potrebbero esistere diversi λ_i^* che soddisfano le condizioni KKT.

Ma se valgono le condizioni LICQ, allora λ_i^* è unico.

Condizioni necessarie del secondo ordine

Assumiamo che $f(x)$ e $c_i(x)$ siano tutte continue e differenziabili fino al secondo ordine.

Siano

- $\mathcal{F}(x^*)$, l'insieme delle direzioni ammissibili in x^*;
- λ^* un vettore di moltiplicatori Lagrangiani che soddisfa le KKT in x^*.

Cono critico.

$$\mathcal{C}(x^*, \lambda^*) = \{w \in \mathcal{F}(x^*) : \nabla c_i(x^*)^T w = 0 \ \forall i \in A(x^*) \cap \mathcal{I} : \lambda_i^* > 0\}.$$

Quindi:

$$w \in \mathcal{C}(x^*, \lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0 & \forall i \in \mathcal{E} \\ \nabla c_i(x^*)^T w = 0 & \forall i \in A(x^*) \cap \mathcal{I} : \lambda_i^* > 0 \\ \nabla c_i(x^*)^T w \geq 0 & \forall i \in A(x^*) \cap \mathcal{I} : \lambda_i^* = 0 \end{cases}$$

Cono critico

Dato che $\lambda_i^* = 0 \ \forall i \notin A(x^*)$,

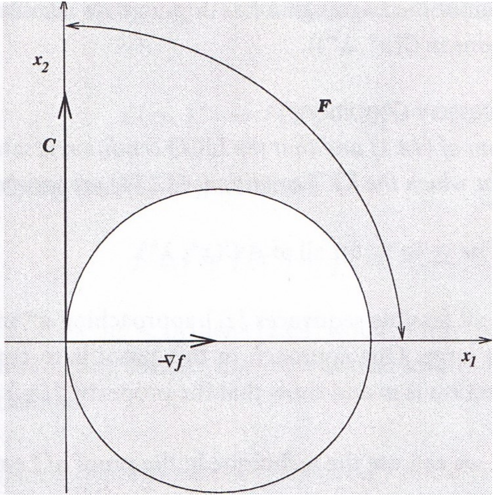
$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0 \ \ \forall i \in \mathcal{E} \cup \mathcal{I}.$$

Dalla definizione di $\mathcal{L}(x, \lambda)$ e dalle KKT

$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow w^T \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* w^T \nabla c_i(x^*) = 0.$$

Quindi il cono critico contiene quelle direzioni ammissibili per le quali le derivate prime non danno informazioni sufficienti.

Cono critico



Condizioni del secondo ordine

Condizioni necessarie del secondo ordine. Sia x^* un minimo locale di $f(x)$ in cui sono soddisfatte le LICQ. Sia λ^* un vettore di moltiplicatori Lagrangiani che soddisfa le KKT in x^*. Allora

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0 \;\; \forall w \in \mathcal{C}(x^*, \lambda^*).$$

Condizioni sufficienti del secondo ordine. Sia x^* una soluzione ammissibile e sia λ^* un vettore di moltiplicatori Lagrangiani che soddisfa le KKT in x^*. Se

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0 \;\; \forall w \in \mathcal{C}(x^*, \lambda^*), w \neq 0,$$

allora x^* è un minimo locale di $f(x)$.

Algoritmi

Per ogni dato sottoinsieme di vincoli attivi (il *working set*), è possibile risolvere un problema di PNL non vincolata.

Tuttavia questo metodo soffre per l'esplosione combinatoria nel numero di sottinsieme che è necessario considerare.

I metodi *active set* eseguono una ricerca "intelligente", scartando a priori alcuni sottinsiemi.

I metodi del punto interno o metodi a barriera invece producono sequenze di punti che non rendono attivo alcun vincolo di disuguaglianza, bensì si avvicinano asintoticamente al contorno della regione ammissibile.

Funzioni di merito e filtri

In generale gli algoritmi di PNL devono bilanciare due effetti di ogni passo:

- il miglioramento della funzione obiettivo
- il peggioramento nella violazione di alcuni vincoli

Una funzione di merito combina insieme i due effetti tramite un opportuno *penalty parameter* μ.

Una funzione di merito $\Phi(x, \mu)$ è esatta quando esiste un valore scalare positivo μ^* tale che per ogni valore $\mu > \mu^*$ ogni minimo locale del problema di PNL vincolata è un minimo locale di $\Phi(x, \mu)$.

Funzioni di merito

Un esempio è la l_1-*penalty function*:

$$\Phi_1(x, \mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} [c_i(x)]^-$$

dove $[k]^-$ indica $\max\{0, -k\}$.

La funzione $\Phi_1(x, \mu)$ non è differenziabile ovunque, ma è esatta.

Il valore soglia è dato da

$$\mu^* = \max_{i \in \mathcal{E} \cup \mathcal{I}} \{|\lambda_i^*|\},$$

dove λ_i^* indica il vettore dei moltiplicatori duali corrispondenti ad una soluzione ottima x^*.

Dato che λ_i^* non è noto a priori, occorre iterativamente ri-calibrare il valore di μ.

Funzioni di merito

Un altro esempio è la l_2-*penalty function* che nel caso di vincoli di uguaglianza ha la forma

$$\Phi_2(x,\mu) = f(x) + \mu\|c_i(x)\|_2.$$

Anche questa non è differenziabile perché la derivata non è definita dove $c(x) = 0$.

Funzioni di merito

La funzione di merito *Fletcher's augmented Lagrangian* è sia differenziabile che esatta:

$$\Phi_F(x, \mu) = f(x) - \lambda(x)^T c(x) + \frac{1}{2}\mu \sum_{i \in \mathcal{E}} c_i(x)^2$$

dove

$$\lambda(x) = [A(x)A(x)^T]^{-1} A(x) \nabla f(x)$$

e $A(x)$ indica lo Jacobiano di $c(x)$.

Tuttavia è pesante a causa del calcolo di $\lambda(x)$.

Funzioni di merito

La funzione di merito Lagrangiana aumentata è

$$\mathcal{L}_A(x, \lambda, \mu) = f(x) - \lambda^T c(x) + \frac{1}{2}\mu\|c(x)\|_2^2.$$

Si accetta un punto prossimo (x^{k+1}, λ^{k+1}) se la $\mathcal{L}_A$ diminuisce rispetto al punto corrente (x, λ).

Gli algoritmi che usano questa funzione di merito includono criteri per modificare opportunamente i valori di λ e μ.

Derivate direzionali

Le funzioni non differenziabili hanno tuttavia derivate direzionali: data una funzione $f(x)$ ed una direzione p, la derivata direzionale di $f(x)$ nella direzione p è

$$D(f(x), p) = \lim_{\epsilon \to 0} \frac{f(x + \epsilon p) - f(x)}{\epsilon}.$$

Quando $f(x)$ è continua e differenziabile in un intorno di x, si ha

$$D(f(x), p) = \nabla f(x)^T p.$$

Derivate direzionali

In un metodo *line search* la condizione per accettare un passo α è che sia abbastanza piccolo affinché la disequazione

$$\Phi(x + \alpha p, \mu) \leq \Phi(\mu, x) + \eta\alpha D(\Phi(x, u), p)$$

sia soddisfatta per qualche $0 \leq \eta \leq 1$.

I metodi *trust region* usano tipicamente un modello quadratico q per stimare il valore di Φ dopo un passo p.

La condizione sufficiente per accettare un passo è

$$\Phi(x + p, \mu) \leq \Phi(x, \mu) - \eta(q(0) - q(p))$$

per qualche $0 \leq \eta \leq 1$.

Filtri

Negli algoritmi basati sui filtri l'ottimalità e l'ammissibilità vengono trattate come due obiettivi distinti, come nella PMO, e vengono accettate le soluzioni x non-dominate, cioè quelle per cui non è stata trovata in precedenza alcuna soluzione x' con $f(x') \leq f(x)$ e $h(x') \leq h(x)$, dove

$$h(x) = \sum_{i \in \mathcal{E}} |c_i(x)| + \sum_{i \in \mathcal{I}} [c_i(x)]^-$$

indica una misura della violazione dei vincoli.

Nei metodi *line search* una soluzione $x^{k+1} = x^k + \alpha_k p_k$ viene accettata se (f^{k+1}, h^{k+1}) è una coppia di valori non-dominata.

Nei metodi *trust region*, se una soluzione x^{k+1} non viene accettata, si riduce il raggio e si ripete l'iterazione.

In entrambi i casi vengono intercalate iterazioni di ripristino dell'ammissibilità (*feasibility restoration phases*), dove viene minimizzata solo $h(x)$.