COMPLEMENTI DI RICERCA OPERATIVA

Prof. Marco Trubian
6 CFU

Luca Cappelletti

Lecture Notes
Year 2017/2018

Magistrale Informatica
Università di Milano
Italy
25 aprile 2018

Indice

Parte I

Programmazione non lineare

1

Alcune definizioni base

1.1 La convessità

Definizione 1.1.1 (Insieme convesso). Un insieme $X \subset \mathbb{R}^n$ è convesso se comunque presi due punti $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in X$, allora $\lambda\underline{\boldsymbol{x}} + (1-\lambda)\underline{\boldsymbol{y}} \in X$, per ogni $\lambda \in [0,1]$.

La proprietà di convessità è invariante rispetto alle operazioni di moltiplicazione con uno scalare, unione e intersezione con un altro insieme convesso.

Definizione 1.1.2 (Funzione convessa). Una funzione $f : \mathbb{R}^n \to \mathbb{R}$ è convessa se il suo dominio è un insieme convesso $X \subseteq \mathbb{R}^n$ e comunque presi due punti $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in X$ vale la relazione:

$$f(\lambda\underline{\boldsymbol{x}} + (1-\lambda)\underline{\boldsymbol{y}}) \leqslant \lambda f(\underline{\boldsymbol{x}}) + (1-\lambda) f(\underline{\boldsymbol{y}}) \qquad \forall \lambda \in [0,1]$$

La proprietà di convessità è invariante rispetto a moltiplicazione con uno scalare e somma tra funzioni convesse.

Vale inoltre che la funzione max di una o più funzioni convesse e che il luogo dei punti $\underline{\boldsymbol{x}}$ per i quali vale che $f(\underline{\boldsymbol{x}}) \leqslant \alpha$ è convesso.

Definizione 1.1.3 (Problema convesso). Un problema di ottimizzazione con funzione obiettivo e regione ammissibile entrambe convesse viene detto problema convesso.

1.2 Massimi e minimi locali

Definizione 1.2.1 (Minimo globale). Un punto $\underline{\boldsymbol{x}}^* \in X$ è un punto di minimo globale di $f(\underline{\boldsymbol{x}})$ se:

$$f(\underline{\boldsymbol{x}}^*) \leqslant f(\underline{\boldsymbol{x}}) \quad \forall \underline{\boldsymbol{x}} \in X$$

Definizione 1.2.2 (Minimo locale). Un punto $\underline{\boldsymbol{x}}^* \in X$ è un punto di minimo locale di $f(\underline{\boldsymbol{x}})$ se esiste un intorno aperto $I(\underline{\boldsymbol{x}}^*, \epsilon)$ di $\underline{\boldsymbol{x}}^*$ avente raggio $\epsilon > 0$ tale che:

$$f(\underline{\boldsymbol{x}}^*) \leqslant f(\underline{\boldsymbol{x}}) \quad \forall \underline{\boldsymbol{x}} \in X \cap I(\underline{\boldsymbol{x}}^*, \epsilon)$$

2

Ottimizzazione non vincolata

Definizione 2.0.1 (Direzione di discesa). Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, un vettore $\underline{\boldsymbol{d}} \in \mathbb{R}^n$ si dice direzione di discesa per f in $\underline{\boldsymbol{x}}$ se:

$$\exists \lambda > 0 : f(\underline{\boldsymbol{x}} + \lambda \underline{\boldsymbol{d}}) < f(\underline{\boldsymbol{x}})$$

Definizione 2.0.2 (Derivata direzionale). Sia data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, un vettore $\underline{\boldsymbol{d}} \in \mathbb{R}^n$ e un punto dove f è definita. Se esiste il limite:

$$\lim_{0^+} \frac{f(\underline{\boldsymbol{x}} + \lambda \underline{\boldsymbol{d}}) - f(\underline{\boldsymbol{x}})}{\lambda}$$

allora tale limite prende il nome di derivata direzionale della funzione f nel punto $\underline{\boldsymbol{x}}$ lungo la direzione $\underline{\boldsymbol{d}}$

2.1 Condizioni necessarie di ottimalità del primo ordine

Teorema 2.1.1 (Condizioni necessarie di ottimalità del primo ordine). Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, derivabile in $\underline{\boldsymbol{x}}^* \in \mathbb{R}^n$, condizione necessaria affinchè il punto $\underline{\boldsymbol{x}}^*$ sia un minimo locale per f è che il gradiente della funzione calcolato in $\underline{\boldsymbol{x}}^*$ sia nullo.

2.2 Condizioni necessarie di ottimalità del secondo ordine

Teorema 2.2.1 (Condizioni necessarie di ottimalità del secondo ordine). Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$ di classe $\underline{\boldsymbol{C}}^2(\underline{\boldsymbol{x}}^*)$, condizione necessarie affinchè il punto $\underline{\boldsymbol{x}}^*$ sia un minimo locale per f è che il gradiente della funzione calcolato in $\underline{\boldsymbol{x}}^*$ sia nullo e che valga la relazione seguente:

$$\underline{\boldsymbol{d}}^T H(\underline{\boldsymbol{x}}^*) \underline{\boldsymbol{d}} \geqslant 0 \forall \underline{\boldsymbol{d}} \in \mathbb{R}^n$$

Cioè l'hessiana è definita come **semipositiva**.

2.3 Condizioni necessarie di ottimalità in senso stretto del secondo ordine

Teorema 2.3.1 (Condizioni necessarie di ottimalità del secondo ordine). Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$ di classe $\underline{\boldsymbol{C}}^2(\underline{\boldsymbol{x}}^*)$, condizione necessarie affinchè il punto $\underline{\boldsymbol{x}}^*$ sia un minimo locale in **senso stretto** per f è che il gradiente della funzione calcolato in $\underline{\boldsymbol{x}}^*$ sia nullo e che valga la relazione seguente:

$$\underline{\boldsymbol{d}}^T H(\underline{\boldsymbol{x}}^*) \underline{\boldsymbol{d}} > 0 \forall \underline{\boldsymbol{d}} \in \mathbb{R}^n$$

Cioè l'hessiana è definita come **positiva**.

3
Programmazione quadratica

Nella programmazione quadratica si approssima f con il seguente modello quadratico:

$$\min f(\underline{\boldsymbol{x}}) = \frac{1}{2}\underline{\boldsymbol{x}}^T Q\underline{\boldsymbol{x}}\underline{\boldsymbol{b}}^T \underline{\boldsymbol{x}}$$

Esistono pochi casi possibili:

Q **non è semidefinita positiva** f non ha un punto di minimo.

Q **è definita positiva** $\underline{\boldsymbol{x}}^* = Q^{-1}\underline{\boldsymbol{b}}$ è l'unico minimo globale.

Q **è definita semi-positiva** Q **non è singolare** $\underline{\boldsymbol{x}}^* = Q^{-1}\underline{\boldsymbol{b}}$ è l'unico minimo globale.

Q **è singolare** non esiste una soluzione o esistono infinite soluzioni.

4
Convergenza

Ovviamente un algoritmo è buono se converge rapidamente.

4.1 Algoritmo convergente localmente e globalmente

4.1.1 Convergente localmente

Un algoritmo è convergente localmente quando converge solo se il punto da cui parte è in un intorno del punto ottimo.

4.1.2 Convergente globalmente

Un algoritmo è convergente globalmente quando converge partendo da qualsiasi punto del dominio.

4.2 Velocità di convergenza

4.2.1 Q-lineare

Quando il rapporto tra lo step k e lo step $k+1$ mantiene un valore costante.

4.2.2 Q-superlineare

Quando il rapporto tra lo step k e lo step $k+1$ all'infinito è nullo.

4.2.3 Q-quadratica

Quando il rapporto tra lo step k quadro e lo step $k+1$ mantiene un valore costante.

5
Metodi di ottimizzazione continua

Si tratta del problema di determinare a partire dallo steo k lo step $k+1$ in modo tale da avvicinarsi all'ottimo della funzione.

Esistono due strategie:

Line search Si determina una direzione e quindi si stabilisce una distanza con cui muoversi in questa direzione.

Trust region Si determina un raggio (una distanza) di confidenza e quindi si stabilisce una direzione nel limite di questa circonferenza verso cui muoversi.

Esiste chiaramente un tradeoff tra velocità e precisione nello stabilire la distanza con cui muoversi.

5.1 Condizioni di Wolfe

Per essere efficace, la line-search approssimata richiede che siano rispettate le condizioni di Wolfe. Esse sono due, una di **decremento sufficiente** ed una di **curvatura**, dove i coefficienti $\underline{\boldsymbol{c}}$ sono tali che $0 < c_1 < c_2 < 1$

$$f(\underline{\boldsymbol{x}} + \alpha\underline{\boldsymbol{d}}) \leqslant f(\underline{\boldsymbol{x}}) + c_1\alpha\nabla f(\underline{\boldsymbol{x}})^T\underline{\boldsymbol{d}}$$
$$\phi(\alpha) \leqslant \phi(0) + \alpha c_1\phi'(0)$$

(a) Condizione di decremento sufficiente

$$f(\underline{\boldsymbol{x}} + \alpha\underline{\boldsymbol{d}})^T\underline{\boldsymbol{d}} \geqslant c_2\alpha\nabla f(\underline{\boldsymbol{x}})^T\underline{\boldsymbol{d}}$$
$$\phi'(\alpha) \geqslant c_2\phi'(0)$$

(b) Condizione di curvatura

Figura 5.1: Condizioni di Wolfe

Le condizioni di Wolfe forti introducono un vincolo di segno sulla curvatura (valore assoluto).

5.2 Metodo di Armijo per stabilire la stepsize

Si itera riducendo gradualmente la distanza di un fattore σ, usualmente pari circa a 0.9 sino a che il valore dello step successivo è più vicino all'ottimo dello step corrente. Nei metodi Newton e quasi Newton il coefficiente della distanza è inizializzato usualmente a 1.

5.3 Convergenza dei metodi di ricerca lineare approssimata

Se definiamo θ_k come l'angolo tra $\underline{\boldsymbol{d}}_k$ e $-\nabla f_k$, allora possiamo determinare il coseno dell'angolo come:

$$\cos\theta_k = \frac{-\nabla f(\underline{\boldsymbol{x}})_k^T\underline{\boldsymbol{d}}_k}{\|\nabla f(\underline{\boldsymbol{x}})_k\| \cdot \|\underline{\boldsymbol{d}}_k\|}$$

Teorema 5.3.1. Sia $\underline{\boldsymbol{d}}_k$ una direzione di discesa e sia α_k una distanza che rispetti le condizioni di Wolfe. Sia inoltre f una funzione limitata inferiormente su $\mathbb{R}^n$, **differenziabile continuamente** sull'insieme M che contiene $L_f = \{\underline{\boldsymbol{x}} : f(\underline{\boldsymbol{x}}) \leqslant f(\underline{\boldsymbol{x}}_0)\}$, dove $\underline{\boldsymbol{x}}_0$ è il punto di inizio. Assumiamo inoltre che ∇f sia **lipschitziana** sull'insieme M. Allora:

$$\sum_{j=0}^{\infty} \cos^2 \theta_j \left\| \nabla f(\underline{\boldsymbol{x}}_j) \right\|^2 \leqslant \infty$$

Dimostrazione. Essendo valida la **condizione di curvatura**, allora vale la disequazione:

$$\nabla f_{k+1}^T \underline{\boldsymbol{d}}_k \geqslant c_2 \nabla f_k^T \underline{\boldsymbol{d}}_k$$

Sottraggo ad ambo i lati $\nabla f_k \underline{\boldsymbol{d}}_k$ ed ottengo:

$$(\nabla f_{k+1}^T - \nabla f_k)\underline{\boldsymbol{d}}_k \geqslant (c_2 - 1)\nabla f_k^T \underline{\boldsymbol{d}}_k$$

Siccome il gradiente della funzione è **lipschitziano** vale la disequazione:

$$(\nabla f_{k+1}^T - \nabla f_k)\underline{\boldsymbol{d}}_k \leqslant \|\nabla f_{k+1}^T - \nabla f_k\| \, \|\underline{\boldsymbol{d}}_k\| \leqslant L \|\underline{\boldsymbol{x}}_{k+1} - \underline{\boldsymbol{x}}_k\| \, \|\underline{\boldsymbol{d}}_k\| = \alpha_k L \|\underline{\boldsymbol{d}}_k\|^2$$

Da questa disequazione ricaviamo il valore di α_k:

$$\alpha_k \geqslant \frac{c_2 - 1}{L} \frac{\nabla f_k^T \underline{\boldsymbol{d}}_k}{\|\underline{\boldsymbol{d}}_k\|^2}$$

Essendo quindi valida la **condizione di decremento sufficiente** vale la disequazione:

$$f_{k+1} \leqslant f_k + c_1 \alpha_k \underline{\boldsymbol{d}}_k^T \nabla f_k = f_k - c_1 \frac{1 - c_2}{L} \frac{(\nabla f_k^T \underline{\boldsymbol{d}}_k)^2}{\|\underline{\boldsymbol{d}}_k\|^2}$$

Pongo $c = c_1 \frac{1-c_2}{L}$ ed ottengo:

$$f_{k+1} \leqslant f_k + c_1 \alpha_k \underline{\boldsymbol{d}}_k^T \nabla f_k = f_k - c \frac{(\nabla f_k^T \underline{\boldsymbol{d}}_k)^2}{\|\underline{\boldsymbol{d}}_k\|^2}$$

Sostituisco $\cos\theta_k = \frac{-\nabla f(\underline{\boldsymbol{x}})_k^T \underline{\boldsymbol{d}}_k}{\|\nabla f(\underline{\boldsymbol{x}})_k\| \cdot \|\underline{\boldsymbol{d}}_k\|}$ ed ottengo:

$$f_{k+1} \leqslant f_k + c_1 \alpha_k \underline{\boldsymbol{d}}_k^T \nabla f_k = f_k - c \cos^2 \theta_k \|\nabla f_k\|^2$$

Per ricorsione ottengo la sommatoria:

$$f_{k+1} \leqslant f_k + c_1 \alpha_k \underline{\boldsymbol{d}}_k^T \nabla f_k = f_0 - c \sum_{j=0}^{k} \cos^2 \theta_j \|\nabla f_j\|^2$$

Siccome la funzione f è limitata inferiormente si ottiene la **condizione di Zoutendijk**:

$$c \sum_{j=0}^{k} \cos^2 \theta_j \|\nabla f_j\|^2 \leqslant f_0 - f_{k+1} < \infty$$

Questo implica che:

$$\cos^2 \theta_j \|\nabla f_j\|^2 \to 0$$

Quindi se l'algoritmo soddisfa anche la **condizione angolare** (cioè sceglie una direzione di discesa che la rispetta) $\cos\theta_k \geqslant \epsilon > 0$ allora **converge**:

$$\lim_{k \to \infty} \|\nabla f(\underline{\boldsymbol{x}}_k)\| = 0$$

□

Parte II

Programmazione lineare intera

6
Programmazione lineare intera