Problemi di ottimizzazione non lineare

Locicero Giorgio

December 25, 2020

1 Introduzione

Questo tipo di problemi sono chiamati in questo modo perché la funzione obiettivo(con n variabili) o i vincoli(m vincoli) sono non lineari, più formalmente:

$$minf(x)$$

$$g(x) \leq 0$$
$$h(x) = 0$$
$$f : \mathbb{R}^n \implies \mathbb{R}$$
$$g : \mathbb{R}^n \implies \mathbb{R}^m$$
$$h : \mathbb{R}^n \implies \mathbb{R}^p$$

DEF: un punto si dice di minimo globale se $f(x\cdot) \leq f(x), \forall x \in X$

DEF: un punto si dice di minimo globale stretto se $f(x\cdot) < f(x), \forall x \in X, x\cdot \neq x$

DEF: un punto si dice di minimo locale se $\exists I(x\cdot, \delta), \delta > 0 | f(x\cdot) \leq f(x), \forall x \in X \cap I(x\cdot, \delta)$

DEF: un punto si dice di minimo locale stretto se $\exists I(x\cdot, \delta), \delta > 0 | f(x\cdot) \leq f(x), \forall x \in X \cap I(x\cdot, \delta), x \neq x\cdot$

1.1 vincolati e non vincolati

Si dicono vincolati problemi di ottimizzazione che hanno dei vincoli. Anche problemi con regione ammissibile aperta sono non vincolati.

DEF: (problema non vincolato)

$$\min f(x)$$
$$x \in \mathbb{R}^n$$
$$f : \mathbb{R}^n \implies \mathbb{R}$$

DEF: (problema vincolato)

$$\min f(x)$$
$$x \in X \cup \mathbb{R}^n$$

cioè:

$$minf(x)$$
$$g(x) \leq 0$$
$$h(x) = 0$$
$$f : \mathbb{R}^n \implies \mathbb{R}$$
$$g : \mathbb{R}^n \implies \mathbb{R}^m$$
$$h : \mathbb{R}^n \implies \mathbb{R}^p$$

con $p \leq n$ per evitare vincoli ridondanti.(?)

DEF: Sia dato un problema vincolato

$$minf(x)$$
$$g_1(x) \leq 0$$
$$\vdots$$
$$g_m(x) \leq 0$$

Un vincolo $g_i(x)$ si dice violato in x* se $g_i(x\cdot) > 0$

Un vincolo $g_i(x)$ si dice attivo in x* se $g_i(x\cdot) = 0$

1.2 Ottimizzazione non vincolata

Questa tipo di problemi si presenta nella seguente forma:

$$\min f(x)$$
$$x \in \mathbb{R}^n$$

Una soluzione lo cale deve soddisfare le condizioni di ottimalità.

DEF: (Condizioni necessaria di ottimalità).

Individua punti che potrebbero essere ottimi, con la contronominale si prova che se un punto non la rispetta, non potrà essere un punto ottimale.

DEF: (Condizioni necessaria di ottimalità).

Se un punto la soddisfa allora è un ottimo locale(ovviamente se rispetta pure la CN)

DEF: condizione necessaria di ottimalità è che sia un punto che potrebbe essere ottimo(bella questa definizione)

DEF: condizione sufficiente di ottimalità è che il punto ottimale è un ottimo locale

DEF: un vettore $d \in \mathbb{R}^n$ è una direzione di discesa per $\mathbb{R}^n \longrightarrow \mathbb{R}$

$$\exists \overline{\alpha} > 0 | f(x + \alpha d) < f(x), \forall \alpha \in (0, \overline{\alpha})$$

DEF: **CN** condizione necessaria affinché x* sia un minimo locale è che

$$\nabla f(x\cdot)^T d \geq 0 \in \mathbb{R}^n$$

Cioè il gradiente sia il vettore nullo nel punto da considerare. Poi punti stazionari con condizione necessaria gradiente =0 Inoltre si devono controllare l'hessiana e la convessità, in modo da ottenere una soluzione globale ottimale.

1.2.1 condizioni necessarie del 2° ordine

Affinché x* sia minimo locale il gradiente $\nabla f(x\cdot) = \vec{0}$ e la matrice hessiana sia semidefinita positiva

HESSIANA

$$\text{Hess } f_p(\mathbf{v}) = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Anche definita come la matrice jacobiana del gradiente di una funzione.

La traccia della matrice è il Laplaciano. Il determinante della matrice è l'Hessiano.

Vedere pure il concetto di Bordered Hessian che ingloba i vincoli nel caso della lagrangiana.

il segno di una matrice si definisce dal segno dei suoi autovalori:

- autovalori solo positivi $\implies$ matrice positiva
- autovalori solo negativi $\implies$ matrice negativa
- autovalori positivi e negativi $\implies$ matrice indefinita
- almeno un autovalore nullo $\implies$ matrice semidefinita(con aggiunta dei casi precedenti)

1.2.2 condizioni sufficienti del 2° ordine

Sia x* un punto stazionario $\nabla f(x\cdot) = 0$ interno alla regione ammissibile.

- Se la matrice Hessiana è definita positiva => x* minimo locale
- Se la matrice Hessiana è definita negativa(lei intende tutti autovalori tutti negativi forse) => massimo locale
- Se la matrice è seminegativa o semipositiva=> punto di sella?
- Se la matrice Hessiana è indefinita non si può dire niente.

1.3 Ottimizzazione non lineare vincolata

1.3.1 Ottimizzazione con sole uguaglianze

$$min(f(x))$$

$$h(x) = 0$$

$$f : X- > \mathbb{R}$$

$$h : X- > \mathbb{R}^p$$

DEF: Regolarità dei vincoli i vincoli sono regolari se i gradienti dei singoli vincoli sono linearmente indipendenti, si nota come irregolarità presuma una sorta di ridondanza tra i vincoli che rende complesso lo studio del problema, da ora in poi si considerano solo problemi con vincoli regolari

DEF: Condizione di ottimalità del 1° ordine(molto simile a Lagrange, infatti è un introduzione al concetto) Sia x* un punto di minimo ed i vincoli siano regolari, allora si ha

$$\nabla f(x\cdot) + \lambda \nabla h(x\cdot) = 0$$

$$h(x\cdot) = 0$$

DEF: forma equivalente della CN del 1° ordine(con lagrangiana) La lagrangiana è la funzione:

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x)$$

Dove $\lambda_1, ..., \lambda_p$ sono i moltiplicatori di Lagrange associati ad ogni vincolo di eguaglianza. rivedendo la condizione sopracitata, si riscrive:

$$\nabla f(x\cdot) + \lambda^T \nabla h(x\cdot) = \nabla_x \mathcal{L}(x\cdot, \lambda)$$

$$\nabla_\lambda \mathcal{L}(x, \lambda) = 0$$

dove a pedice vi sono le derivate in base alle variabili di derivazione parziale(prima gradiente sulle x, poi gradiente su lambda, che lascia solo il vincolo). I sistemi da risolvere sono sempre gli stessi.

1.3.2 Ottimizzazione non lineare vincolata con diseguaglianze

$$min(f(x))$$

$$g(x) \leq 0$$

$$h(x) = 0$$

$$f : X- > \mathbb{R}$$

$$h : X- > \mathbb{R}^p$$

$$g : X- > \mathbb{R}^m$$

stesso concetto dei vincoli attivi con attività nel caso di eguaglianze per i vincoli di diseguaglianze.

DEF: regolarità vincoli

Nei vincoli attivi, questi (insieme a quelli h(x)) devono essere regolari, cioè i gradienti devono essere linearmente indipendenti, tutti tra tutti(come se fossero tutte eguaglianze, dato che stiamo considerando dove i vincoli sono attivi).

DEF: Condizione di ottimalità del 1° ordine(condizioni di Karush-Kuhn-Tucker)

Sia x* punto di minimo locale regolare dei vincoli. Allora $\exists \lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^m$ tali che

$$\nabla f(x\cdot) + \lambda^T \nabla h(x\cdot) + \mu^T \nabla g(x\cdot) = 0$$

$$\mu^T g(x\cdot) = 0$$

quella sopra è la condizione di complementarietà.

DEF: funzione Lagrangiana

dato il problema

$$min(f(x))$$
$$g(x) \leq 0$$
$$h(x) = 0$$

La funzione Lagrangiana:

$$\mathcal{L}(x.\lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x) = 0$$

$$\forall j, \lambda_j \xrightarrow{moltdelvincolo} h_j(x) = 0$$

$$\forall i, \mu_i \xrightarrow{moltdelvincolo} g_i(x) = 0$$

con h_j e g_i moltiplicatore dei vincoli

Le condizioni KKT si scrivono:

$$\nabla_x \mathcal{L}(x\cdot, \lambda, \mu) = \nabla f(x\cdot) + \lambda^T \nabla h(x\cdot) + \mu^T \nabla g(x\cdot) = 0$$

$$\mu^T g(x\cdot) = 0$$

OSS: i moltiplicatori λ possono assumere valori positivi e negativi,i moltiplicatori μ(dei vincoli di g(x)) sono sempre ≥0

Se il problema è convesso le condizioni KKT si possono invertire e cono così condizioni necessarie e sufficienti per l'ottimalità(minimo-massimo globale).

problema	C.N.1° ordine
minf(x) $x \in \mathbb{R}$	$\nabla f(x\cdot) = 0$ condizione di Fermat(punto stazionario)
minf(x) h(x)=0	$\nabla f(x\cdot) + \lambda^T \nabla h(x\cdot) = 0$ $h(x\cdot) = 0$ metodo moltiplicatori di Lagrange
min$f(x)$ $g(x) \leq 0$ $h(x) = 0$	$\nabla f(x\cdot) + \lambda^T \nabla h(x\cdot) + \mu^T \nabla g(x) = 0$ $\mu^T g(x\cdot) = 0$ condizioni KKT

1.4 convessi e concavi

Un problema si chiama convesso o concavo in base alle caratteristiche dei problemi di ottimizzazione convessi:

$min(f(x))$ con X insieme convesso $\quad$ $x \in X$ con f funzione convessa

$max(f(x))$ con X insieme convesso $\quad$ $x \in X$ con f funzione concava

Coincidenza tra minimi locali e minimi globali.

Problema di ottimizzazione concavi:

$min(f(x))$ con X insieme convesso $\quad$ $x \in X$ con f funzione concava

$max(f(x))$ con X insieme convesso $\qquad x \in X$ con f funzione convessa

Non c'è coincidenza tra minimi globali e minimi globali. La soluzione, se esiste, si trova sulla frontiera della regione ammissibile

Si può verificare che una funzione è convessa provando che la matrice hessiana è definita positiva. I punti stazionari di minimo locali trovati in una funzione convessa sono minimi globali

1.5 Esempi reali

1.5.1 Problema non lineare che si trasforma in lineare

Problemi con valore assoluto:

$\min \sum_j c_j|x_j| + \sum_k d_k y_k \qquad (x, y) \in C$

Con C un poliedro, $c_j \geq 0, \forall j$

Risulta $|x_j| = max x_j, -x_j$

Il problema diventa:

$min(\sum_j c_j max x_j, -x_j + \sum_k d_k y_k)$

Si utilizza la variabile ausiliaria $z_j = max x_j, -x_j$ e si aggiunge il vincolo $-z_j \leq x_j \leq z_j$

Alla fine il problema diventa

$min(\sum_j c_j z_j + \sum_k d_k y_k) \qquad -z_j \leq x_j \leq z_j$

1.5.2 Support Vector Machine

massimizzazione del margine tra le classi, m come $\frac{2}{||w||}$, quindi cercare il massimo del margine significa minimizzare la norma del vettore dei pesi. si risolve il problema

$$min\frac{1}{2}||w||^2$$

$$y_i(w^T x_i + w_0) \geq 1, \forall i$$

Con $y_i \in \{-1, 1\}$ indicatore di classe per il singolo record, w vettore dei pesi.

Il problema è convesso con funzione obiettivo strettamente convessa $\implies$ ha una sola soluzione

1.5.3 SVM con KKT

$$min\frac{1}{2}||w||^2$$

$$y_i(w^T x_i + w_0) \geq 1, \forall i$$

che diventa

$$min\frac{1}{2}||w||^2$$

$$-y_i(w^T x_i + w_0) + 1 \leq 0, \forall i$$

L'ottimo globale si trova come soluzione del sistema KKT:

$$\mathcal{L}(w, \lambda) = \frac{1}{2}||w||^2 + \sum_{i=1}^{n} \lambda_i(-y_i(w^T x_i + w_0) + 1) = 0$$

Condizioni KKT risultanti:

$$\mathcal{L}(w, \lambda) = \nabla\tfrac{1}{2}||w||^2 + \textstyle\sum_{i=1}^{n} \lambda_i \nabla(-y_i(w^T x_i + w_0) + 1) = 0 \qquad \text{(Lagrangiana)}$$
$$\lambda_i(-y_i(w^T x_i + w_0) + 1) = 0, \forall i \qquad \text{(Condizione di complementarietà)}$$
$$-y_i(w^T x_i + w_0) + 1 \leq 0, \forall i \qquad \text{(Ammissibilità)}$$
$$\lambda_i \geq 0 \qquad \text{(Non negatività dei moltiplicatori)}$$

Soluzione $w = \sum_i \lambda_i y_i x_i$ e $w = \sum_i \lambda_i y_i = 0$

1.6 algoritmi di ottimizzazione

algoritmi iterativi che cercano punti che appartengono ad un insieme ω di punti desiderati. ω è detto bersaglio. ω è l'insieme dei punti che soddisfano le condizioni necessarie di ottimalità.

1.6.1 ottimizzazione non vincolata

consideriamo $\quad \min f(x)$

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$

$x \in \mathbb{R}^n$

$f \in C^1$ cioè funzione continua e derivate parziali continue, quindi differenziabilità di primo livello, vedere smoothness ed appartenenza di una funzione a C^n.

Supponiamo che il problema ammetta soluzioni.

Se il problema ammette soluzione, l'insieme bersaglio Ω (dei punti stazionari) è il seguente

$$\Omega = \{x | \nabla f(x) = 0\}$$

Se non ci sono punti stazionari penso si proceda all'analisi della frontiera.

Lo schema generale:

1. Si fissa $x^0 \in \mathbb{R}^n$, k=0 ;
2. Se $x^k \in \Omega \implies$ STOP ;
3. Si calcola una direzione di ricerca $d^k \in \mathbb{R}^n$;
4. Si calcola il passo α^k lungo la direzione d^k ;
5. $x^{k+1} = x^k + \alpha^k d^k$ e si pone k=k+1 ;
6. Si ritorna al passo 2 .

Per il punto iniziale non esistono criteri di scelta di x^0, si utilizza una strategia di scelta di punti casuale appartenenti all'insieme ammissibile prendendo l'ottimo migliore tra questi.

Consideriamo principalmente algoritmi di discesa che diminuiscono il valore della funzione alla ricerca di minimi.

Per il criterio di arresto, a $||\nabla f(x^k)|| \leq \varepsilon_1$ si aggiungono anche

$||x^{k+1} - x^k|| \leq \varepsilon_2$ e $||f(x^{k+1}) - f(x^k)|| \leq \varepsilon_3$

Scelta della direzione, I criteri si dividono nei vari metodi:

- senza uso di derivate
- uso di derivate prime
- uso di derivate prime e seconde

Scelta del passo, α^k può essere scelto in maniera esatta o inesatta. Si generano una successione di spostamenti definiti lungo la direzione (ottenuta con i metodi precedentemente accennati), e si considera quando la funzione diminuisce di valore, cioè $f(x^k + \alpha^k d^k) \leq f(x^k)$

Sia $\phi(\alpha) = f(x^k + \alpha d^k)$, il passo ottimo α si trova cercando il minimo della funzione $\phi(\alpha)$ (ricerca unidimensionale, ottimizzazione nella variabile α) Problema molto facile da risolvere perché è un problema di minimo in una sola variabile.

Se f è C^1(differenziabile) $\implies$ ϕ derivabile rispetto ad α.

$$\phi'(\alpha) = \nabla f(x^k + \alpha d^k)^T d^k$$

Quindi si pone = 0 per la ricerca di punti stazionari.

Geometricamente la condizione di ottimalità di ϕ vuole che il gradiente di f in $(x^k + \alpha d^k)$ sia ortogonale a d^k dato che il prodotto scalare deve essere =0 per i punti stazionari.

1.6.2 metodo del gradiente

Anche detto della discesa più ripida(steepest descent), la direzione è quella dell'antigradiente.

quindi $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$

La generica iterazione $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$

Il metodo è globalmente convergente

1.6.3 metodo di Newton

Si utilizzano le derivate seconde, quindi o si prendono punti in cui la funzione sia differenziabile due volte oppure si prende una $f \in C^2$. Si costruisce una successione di punti minimizzando una approssimazione quadratica della funzione(tramite approssimazione di secondo ordine di Taylor). Sia x^k un punto assegnato

$$q^k(s) = f(x^k + s) = f(x^k) + \nabla f(x^k)^T s + \frac{1}{2} s^T \nabla^2 f(x^k) s$$

si determina il punto $x^{k+1} = x^k + s^k$ dove s^k è scelto in modo da minimizzare $q^k(s)$

Si pone $\nabla q^k(s) = 0 \implies s^k = -\nabla f(x^k)[\nabla^2 f(x^k)]^{-1}$

L'iterazione generica $x^{k+1} = x^k - \nabla f(x^k)[\nabla^2 f(x^k)]^{-1}$ dove $\nabla^2 f(x^k)$ è la matrice hessiana(in realtà dovrebbe essere l'operatore laplaciano aka la traccia della hessiana ma nella community la ambiguità c'è, dato che può essere visto come prodotto esterno o interno).

Il metodo converge localmente.

1.7 algoritmi ottimizzazione vincolata

Come avviene molto spesso in matematica e logica, un problema complesso viene scomposto in problemi più semplici, in questo caso un problema vincolato viene scomposto in sottoproblemi non vincolati con struttura più semplice. Risolvere problemi vincolati significa

- cercare l'ottimo della funzione obiettivo
- rimanere nella regione ammissibile

I metodi possono:a)trasformare il problema vincolato in una successione di problemi non vincolati, oppure in un solo problema non vincolato. b)trasformare il problema in problemi vincolati ma più semplici.

1.7.1 metodi di penalità

Problema vincolato classico:

$$min(f(x))$$

$$g(x) \leq 0$$

$$h(x) = 0$$

$$f : X- > \mathbb{R}$$

$$h : X- > \mathbb{R}^p$$

$$g : X- > \mathbb{R}^m$$

La regione ammissibile viene definita come :

$$\mathcal{I} = \{x \in X | g(x) \leq 0, h(x) = 0\}$$

Trasformiamo il problema vincolato in un problema non vincolato di minimizzazione di una funzione detta di penalità. La funzione di penalità è costruita a partire da f aggiungendo un termine di penalità che deve penalizzare la violazione del vincolo.

-FUNZIONI DI PENALITA' SEQUENZIALI Consideriamo la funzione di penalità:

$$P(x, \mu) = f(x) + \mu\phi(x)$$

Dove P è la funzione di penalità, μ è il parametro di penalità e phi è il termine di penalità.

$\phi(x)$ misura la violazione del vincolo. Risulta:

- $\phi(x)$ è differenziabile con continuità(ridondante?)
- $\phi(x) = 0$ se $x \in \mathcal{I}$ (non c'è penalità se x è ammissibile cioè soddisfa i vincoli)

- $\phi > 0$ se non è ammissibile

Come funzione ϕ si sceglie

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

La funzione ϕ non è esatta, cioè non esiste un valore di μ per cui il problema vincolato iniziale sia equivalente al problema $minP(x, \mu)$. Risulta però che la soluzione che ha soluzione x_μ del problema non vincolato tende alla soluzione del problema vincolato per $\mu \longrightarrow \infty$.

-FUNZIONI DI PENALITA' ESATTE Si pone

$$P(x, \mu) = f(x) + \mu\phi(x)$$

- $\phi(x)$ non è differenziabile
- $\phi(x) = 0$ se $x \in \mathcal{I}$ (non c'è penalità se x è ammissibile cioè soddisfa i vincoli)
- $\phi > 0$ se non è ammissibile

Come funzione ϕ si sceglie

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

Queste funzioni sono esatte, quindi per valori molto grandi di μ i valori ottimi del problema vincolato è un valore ottimo del problema non vincolato(minimo o massimo). Per valori grandi di μ, le condizioni KKT del problema vincolato coincidono con i punti stazionari di $P(x, \mu)$.

2 esercizi

min(-12x)

$h(x) = x - x^2$

$P(x, \mu) = -12x + \mu(x - x^2)^2$