1 Lezione del 09-12-24

1.1 Problema di ottimizzazione quadratico

Vediamo un problema di esempio sull'ottimizzazione quadratica. Sia la funzione:

$$f(x_1, x_2) = 2x_1^2 + 4x_2^2 + 4x_1 - x_2$$

sul poliedro P di vertici:

$$V = \{(0,1), (4,1), (4,3), (-1,5)\}$$

Ci interessiamo a trovare massimi e minimi globali.

- **Massimi**
  La funzione è concava e definita su un poliedro, quindi i possibili massimi non possono che stare sui vertici. Si calcola allora per enumerazione diretta:

  $$\begin{cases} f(0,1) = 3 \\ f(4,1) = 51 \\ f(4,3) = 81 \\ f(-1,5) = 93 \end{cases}$$

  da cui risulta chiaro il massimo globale in $(-1, 5)$, che è anche unico.

- **Minimi**
  Per il calcolo dei minimi non possiamo sfruttare la concavità della funzione e la convessità del poliedro, in quanto potremmo avere minimi in **punti interni** o **punti sul bordo**.

  Il vertice del paraboloide si trova come $\left(\frac{c}{2a}, \frac{d}{2b}\right) = \left(-1, -\frac{1}{8}\right)$. che non ricade in P, ergo siamo nel secondo caso.

  Si presentano quindi due metodi di risoluzione: uno per via geometrica, e uno che sfrutta il sistema LKT.

  1. La f definisce un paraboloide concavo. In generale, il paraboloide avrà un punto di minimo globale libero $\bar{x}$, e il minimo vincolato sarà il punto $x \in P$ che minimizza la distanza $d(x, \bar{x})$, dove d è scalata per le dimensioni degli elissoidi dati dall'insieme di livello di $\text{lev}_{=k}(f)$. Avevamo adottato un approccio simile negli esempi precedenti, notando il fatto che parlavamo di circonferenze, e quindi d era la comune norma euclidea.

     Cerchiamo innanzitutto di trovare un equazione che espliciti i semiassi e il punto centrale dell'insieme di livello stesso. Per un paraboloide qualsiasi con semiassi paralleli agli assi cartesiani, cioè in forma:

     $$f(x_1, x_2) = ax_1^2 + bx_2^2 + cx_1 + dx_2 + e$$

     siamo interessati all'insieme:

     $$\text{lev}_{=k}(f) = \left\{(x_1, x_2) : ax_1^2 + bx_2^2 + cx_1 + dx_2 + e = k\right\}$$

Converrà riportare la f in una forma del tipo:

$$f(x_1, x_2) = \frac{(x_1 - p_1)^2}{\alpha^2} + \frac{(x_2 - p_2)^2}{\beta^2} - r^2$$

in modo da poter esprimere gli insiemi di livello:

$$\text{lev}_{=k}(f) = \left\{ (x_1, x_2) : \frac{(x_1 - p_1)^2}{\alpha^2} + \frac{(x_2 - p_2)^2}{\beta^2} = r^2 \right\}$$

che non è altro che il fascio di ellissi centrate in (p_1, p_2), con semiassi di dimensione αr e βr.
Troviamo quindi i parametri di tali ellissi, dato:

$$f(x_1, x_2) = \frac{(x_1 - p_1)^2}{\alpha^2} + \frac{(x_2 - p_2)^2}{\beta^2} - r^2$$
$$= \frac{1}{\alpha^2} x_1^2 + \frac{1}{\beta^2} x_2^2 + \left(-\frac{2p_1}{\alpha^2} \right) x_1 + \left(-\frac{2p_2}{\beta^2} \right) x_2 + \frac{p_1^2}{\alpha^2} + \frac{p_2^2}{\beta^2} - r^2$$

Questo si traduce nel sistema:

$$\begin{cases} a = \frac{1}{\alpha^2} \\ b = \frac{1}{\beta^2} \\ c = -\frac{2p_1}{\alpha^2} \\ d = -\frac{2p_2}{\beta^2} \\ e - k = \frac{p_1^2}{\alpha^2} + \frac{p_2^2}{\beta^2} - r^2 \end{cases}$$

da cui si ha:

$$\begin{cases} \alpha = \frac{1}{\sqrt{a}} \\ \beta = \frac{1}{\sqrt{b}} \\ p_1 = -\frac{c}{2a} \\ p_2 = -\frac{d}{2b} \\ r = \sqrt{\frac{c^2}{4a} + \frac{d^2}{4b} - g + k} \end{cases}$$

che sostituendo i valori dati dall'esempio diventa:

$$\begin{cases} \alpha = \frac{\sqrt{2}}{2} \\ \beta = \frac{1}{2} \\ p_1 = -1 \\ p_2 = \frac{1}{8} r = \sqrt{\frac{33}{16} + k} \end{cases}$$

Da qui potremmo già applicare un approccio grafico variando k e trovando la prima intersezione con il poliedro P, sia questa sul bordo o su un vertice.
Un metodo più informato prevede di definire una funzione distanza $d(a, b)$, scalata sulle dimensioni degli elissoidi di $\text{lev}_{=k}(f)$. A questo punto basterà minimizzare $d(p_m, p_\Omega) \in \mathbb{R}$, con $p_m = (p_1, p_2)$ e p_Ω un punto qualsiasi del bordo di P. Definiamo quindi d come:

$$d(a, b) = \sqrt{\frac{(a_x - b_x)^2}{\alpha^2} + \frac{(a_y - b_y)^2}{\beta^2}}$$

Dove gli α e β sono gli stessi semiassi $/r$ di prima.
Questa definizione deriva dalla *distanza di Mahalanobis* su matrici di covarianza, presa nel caso di matrici diagonali (quindi semiassi paralleli agli assi cartesiani). Si potrebbe fare una trattazione più approfondita del caso non strettamente parallelo, ma sono le 2 di notte e domani c'è Reti Logiche.
Prendiamo quindi il punto p_Ω come uno qualsiasi fra i:

$$\phi_{ij}(t) = v_i(1-t) + v_j t$$

segmenti del bordo di P. Consideremo solo i segmenti $(1,2)$ e $(1,4)$, cioè $(0,1) - (4,1)$ e $(0,1) - (-1,5)$, per esemplificare i calcoli. Una discussione completa richiederebbe la verifica di tutti i segmenti. Si hanno quindi le parametrizzazioni:

$$\phi_{12}(t) = v_1(1-t) + v_2 t$$
$$\phi_{14}(t) = v_1(1-t) + v_4 t$$

E vogliamo calcolare:

$$\min d(p_m, \phi_{12}(t))$$
$$\min d(p_m, \phi_{14}(t))$$

Queste sono funzioni reali di variabile reale t, e hanno minimo in 0 di valore $\frac{9}{4}$. Fossero state in disaccordo sul minimo, si sarebbe preso il minimo minore. Il fatto che qui i minimi sono identici e sullo stesso punto significa che il minimo è su un vertice. In particolare, secondo le parametrizzazioni riportate sopra, il vertice è v_1, quindi $(0,1)$.
Notiamo come ad $r = \frac{9}{4}$ si ha $k = 3$, e l'insieme $\text{lev}_{=3}(f)$ sia proprio quello che intercetta per primo il vertice $(0,1)$ del poliedro. Questa è una conseguenza di aver scelto una norma d che è correttamente scalata sui semiassi delle ellissi.

2. Lagrange

1.2 Algoritmo del gradiente proiettato

Sia dato un problema di ottimizzazione non lineare simile al precedente, di massimizzazione, cioè:

$$\begin{cases} \max f(x) \\ x \in P \end{cases}$$

dove P è un poliedro limitato. Di questo poniamo di conoscere il gradiente $\nabla f(x_k)$ ad un certo punto x_k sul bordo di P.

L'idea è quella di **proiettare** il gradiente $\nabla f(x_k)$ sul poliedro P, o meglio sulla retta data dal vincolo attivo di P nel punto x_k. Se la proiezione è non nulla, l'angolo θ_k che il gradiente proiettato ha con il gradiente proprio è $< \frac{\pi}{2}$, e quindi:

$$\nabla f(x_k) \cdot d_k > 0$$

posta d_k come la proiezione ottenuta, che è esattamente l'ipotesi di una direzione di crescita per un algoritmo di PNL.

Resta quindi il problema di calcolare tale proiezione. Avevamo quindi che il poliedro P è definito come:

$$Ax \leq b$$

e gli indici attivi erano definiti come:

$$A(x_k) = \{i \in I : A_i x_k = b\}$$

Definiamo quindi M come la sottomatrice di A ottenuta selezionando le righe di indice $i \in A(x_k)$. Si introduce allora la **matrice di proiezione**:

Definizione 1.1: Matrice di proiezione

Definiamo la matrice di proiezione in un punto x_k di un poliedro P di un problema di PNL come:

$$H = \begin{cases} I, & A(x_k) = \emptyset \\ I - M^T(MM^T)^{-1}M, & \text{altrimenti} \end{cases}$$

data M matrice dei vincoli attivi.

Si può dimostrare il seguente teorema:

Teorema 1.1: Teorema della proiezione

Data una matrice di proiezione H, Hy è la proiezione di y sulla varietà geometrica M.

Nel nostro caso, potremo sfruttare il teorema per calcolare la proiezione del gradiente su M, cioè:

$$H \cdot \nabla f(x_k) = d_k$$

Posto $x_{k+1} = x_k + td_k$, vorremo seguire tale direzione d_k fino al punto di uscita dal poliedro (diciamo $\hat{t}_k$). Questo punto sarà dato dal problema LP:

$$\hat{t}_k = \begin{cases} \max t \\ A \cdot (x_k + td_k) \leq b \end{cases}$$

su m disequazioni per m vincoli. Massimizzeremo poi la funzione sulla direzione d_k fino a un massimo di $\hat{t}_k$ cioè:

$$t_k = \text{argmax}_{[0,\hat{t}_k]} f(x_k + td_k) = \text{argmax}_{[0,\hat{t}_k]} f(x_k + t \cdot H \cdot \nabla f(x_k))$$