Algoritmi per problemi vincolati: il metodo di penalità

L'idea di base: "appesantire" la funzione obiettivo in corrispondenza dei punti non ammissibili e affrontare il problema come se fosse non vincolato.

Le condizioni sono molto generali: chiediamo solo che la funzione obiettivo sia continua. D'altra parte, supponiamo di disporre di un algoritmo capace di fornire un ottimo globale della funzione cui viene applicato.

$$\left\langle \begin{array}{l} \min f(x) \\ x \in S \subset \mathbb{R}^n \end{array} \right\rangle$$

Definiamo una funzione di penalità $P: \mathbb{R}^n \to \mathbb{R}$ con le proprietà

1) P è continua
2) $P(x) \geq 0 \quad \forall x \in \mathbb{R}^n$
3) $P(x) = 0 \iff x \in S$

Esempio $\quad S = \{x \mid x \in \mathbb{R}^n, \ g_i(x) \geq 0 \ \ i = 1, \dots, m\}$

Due possibili funzioni di Penalità:

a) $$P(x) = \frac{1}{2} \sum_{i=1}^{m} \left[\min(0, g_i(x))\right]^2$$

b) $$P(x) = \sum_{i=1}^{m} \left| \min(0, g_i(x)) \right|$$

Nota la funzione di penalità a) è differenziabile se lo sono le $g_i(x)$ mentre la b) non lo è anche quando le $g_i(x)$ sono di classe $C^{(1)}$.

Utilizziamo $P(x)$ per "penalizzare" la f (cioè aggiungendo ad essa qualcosa di positivo quando $x \notin S$ e lasciandola invariata quando $x \in S$).

Introduciamo un parametro di penalità $c > 0$ e definiamo

$$q_c(x) = f(x) + c P(x)$$

Supponiamo di disporre di un algoritmo capace di fornire x_c^*,

$$x_c^* = \arg\min_{x \in \mathbb{R}^n} \left[f(x) + c P(x)\right]$$

Problema di ottimo non vincolato

Si consideri una sequenza di valori positivi, crescenti e tendenti a $+\infty$ del parametro di penalità:

$$c_1 < c_2 < \dots, c_k < c_{k+1} < \dots$$

e definiamo

$$x^{(k)} = \arg\min_{x \in \mathbb{R}^n} q_{c_k}(x)$$

Vale il seguente teorema

Teorema

1) $q_{c_k}(x^{(k)}) \le q_{c_{k+1}}(x^{(k+1)})$ (La q "peggiora")

2) $P(x^{(k)}) \ge P(x^{(k+1)})$ (La P "migliora")

3) $f(x^{(k)}) \le f(x^{(k+1)})$ (La f "peggiora")

1. $$q_{c_{k+1}}(x^{(k+1)}) = f(x^{(k+1)}) + \underbrace{c_{k+1}}_{> c_k} P(x^{(k+1)}) \ge$$

$$\ge f(x^{(k+1)}) + c_k P(x^{(k+1)}) \ge f(x^{(k)}) + c_k P(x^{(k)}) = q_{c_k}(x^{(k)})$$

Perché $x^{(k)}$ è il punto di minimo di $f(x) + c_k P(x)$

2. $$f(x^{(k)}) + c_k P(x^{(k)}) \leq f(x^{(k+1)}) + c_k P(x^{(k+1)})$$

per l'ottimalità di $x^{(k)}$

$$f(x^{(k+1)}) + c_{k+1} P(x^{(k+1)}) \leq f(x^{(k)}) + c_{k+1} P(x^{(k)})$$

per l'ottimalità di $x^{(k+1)}$

Sommando primo e secondo membro si ottiene

$$(c_{k+1} - c_k) P(x^{(k+1)}) \leq (c_{k+1} - c_k) P(x^{(k)})$$

da cui, tenendo conto che $c_{k+1} - c_k > 0$, segue $P(x^{(k+1)}) \leq P(x^{(k)})$

3. $$f(x^{(k+1)}) + c_k P(x^{(k+1)}) \geq f(x^{(k)}) + c_k P(x^{(k)}) \geq f(x^{(k)}) + c_k P(x^{(k+1)})$$

per l'ottimalità di $x^{(k)}$ — per la 2.

$$\Rightarrow f(x^{(k+1)}) \geq f(x^{(k)})$$

Teorema Sia $x^* = \arg\min_{x \in S} f(x)$. Vale $f(x^*) \geq q_{c_k}(x^{(k)}) \geq f(x^{(k)})$, $\forall k$

Prova Poiché $P(x) = 0 \ \forall x \in S$ e $x^* \in S$ possiamo scrivere:

$$f(x^*) = f(x^*) + c_k \underbrace{P(x^*)}_{=0} \geq f(x^{(k)}) + c_k P(x^{(k)}) = q_{c_k}(x^{(k)}) \geq f(x^{(k)})$$

per l'ottimalità di $x^{(k)}$

Perché $P(x^{(k)}) \geq 0$

Teorema Se la sequenza $\{x^{(k)}\}$ ha un punto di accumulazione $\bar{x}$, allora $\bar{x} = \arg\min_{x \in S} f(x)$

Prova Dire che $\{x^{(k)}\}$ ha un punto di accumulazione equivale a dire che <u>esiste</u> una <u>sottosequenza</u> $\{x^{(k)}\}$, $k \in K$, convergente

$$(*) \quad \{x^{(k)}\} \xrightarrow[k \in K]{} \bar{x}$$

Abbiamo dimostrato che $\forall k$, e in particolare per $k \in K$, è $q_{c_k}(x^{(k)}) \leq f(x^*)$ e (dalla 2.) che $q_{c_k}(x^{(k)})$ è una <u>sequenza di numeri reali</u> <u>monotona crescente</u>

$\Rightarrow$ La sequenza $\{q_{c_k}(x^{(k)})\}$, $k \in K$, ha un limite q^*

$$\lim_{k \in K} q_{c_k}(x^{(k)}) = q^* \leq f(x^*) \quad , \text{ cioè}$$

$$(**) \quad \lim_{k \in K} \left[f(x^{(k)}) + c_k P(x^{(k)}) \right] = q^* \leq f(x^*)$$

Dalla continuità di f e da $(*)$ segue che $\lim_{k \in K} f(x^{(k)}) = f(\bar{x})$

e quindi $\lim_{\substack{k \to \infty \\ k \in K}} \underset{\downarrow \text{ tende a } +\infty}{c_k} P(x^{(k)}) = \underbrace{q^* - f(\bar{x})}_{\text{quantità finita}}$

$\Rightarrow \lim_{\substack{k \to \infty \\ k \in K}} P(x^{(k)}) = 0 = P(\bar{x})$ $\Rightarrow$ $\bar{x}$ è ammissibile

continuità di $P(x)$

Poichè $\bar{x}$ è ammissibile, tenendo conto che x^* è un punto di minimo,

$$\underbrace{f(\bar{x}) \geq f(x^*)}\qquad (a)$$

D'altra parte dalla (**), tenendo conto che $P(\bar{x})=0$,

segue che $\lim_{\substack{k \in K \\ k \to \infty}} f(x^{(k)}) = \underbrace{f(\bar{x}) \leq f(x^*)}$ (b)

a) e b) $\Longrightarrow$ $f(\bar{x}) = f(x^*)$ e il teorema è dimostrato

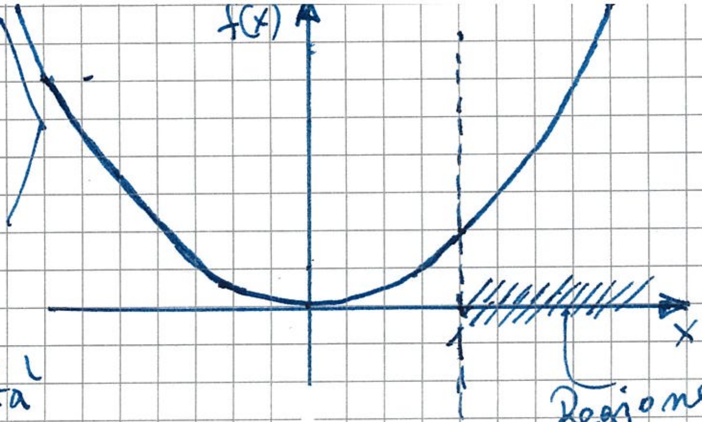
L'approccio della funzione di penalità consente di trasformare un problema di ottimo vincolato in una sequenza di problemi di ottimo non vincolato (Un'ottima cosa!!!)

Il problema è che, per ottenere una soluzione, può essere necessario "spingere" c_k verso valori molto grandi. Ciò può comportare difficoltà numeriche nella risoluzione del problema non vincolato (Mettere insieme quantità grandi e piccole è all'origine dei problemi di mal condizionamento (vedi ad esempio la lentezza della convergenza del metodo del gradiente quando $\lambda_{max} \gg \lambda_{min}$.

Un esempio in R

$$\begin{cases} \min \frac{1}{2}x^2 \\ \text{s.v.}\ x \geq 1 \end{cases}$$

$g(x) = x - 1$



KKT

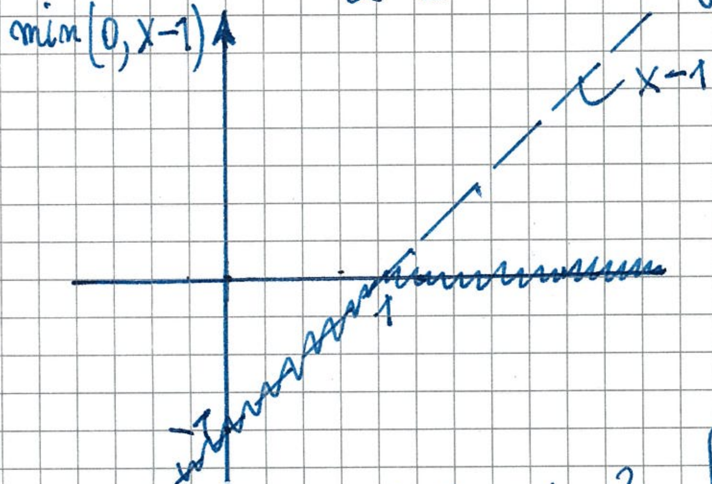
$f'(x) = \lambda g'(x) \Rightarrow$

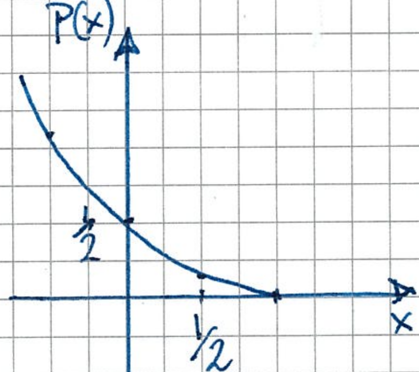
$$\begin{cases} x = \lambda \\ g(x) = x - 1 = 0 \end{cases}$$

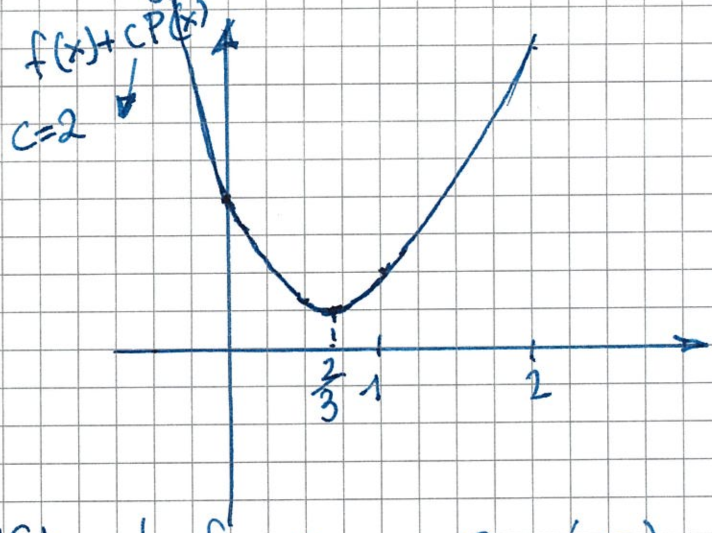
$\Rightarrow x^* = 1, \lambda = 1$

Utilizziamo la funzione di penalità

$$P(x) = \frac{1}{2}\left[\min(0, g(x))\right]^2$$







$$q_{c_k}(x) = \frac{1}{2}x^2 + c_k \begin{cases} 0 & \text{se } x \geq 1 \\ \frac{1}{2}(x-1)^2 & \text{se } x < 1 \end{cases}$$

Nota La funzione $q_{c_k}(x)$ è differenziabile ovunque (basta osservare che in $x = 1$ derivata sinistra e destra sono uguali)

Poiché la derivata non si annulla mai per $x > 1$, per ottenere il minimo basta derivare il termine

$$\frac{1}{2}x^2 + c_k \frac{1}{2}(x-1)^2 \Rightarrow x + c_k(x-1) = 0 \Big|_{x = x_k} \Rightarrow x_k = \frac{c_k}{c_k + 1}$$

Quindi $x_k \longrightarrow x^*$ per $c_k \to \infty$

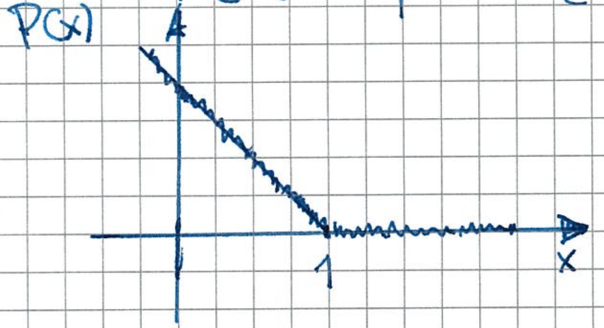
Nota Nell'esempio precedente $x_k \to x^*$ ma è sempre $x_k \neq x^*$!!

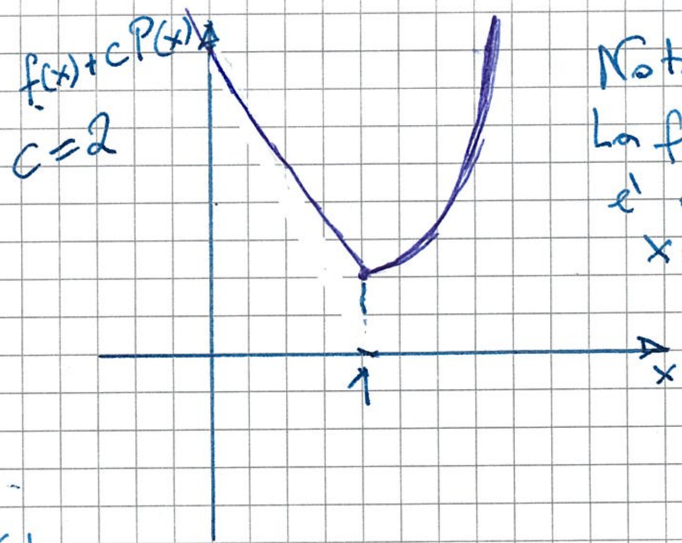
E' possibile costruire una funzione di penalità esatta, tale che per qualche valore finito di c_k sia $x^{(k)} = x^*$?

Consideriamo, per l'esempio precedente, la funzione di penalità b)

$P(x) = \sum_{i=1}^{m} \left| \min(0, g_i(x)) \right|$, che nel nostro caso diventa

$P(x) = \left| \min(0, x-1) \right|$





Nota.
La funzione non è derivabile in $x=1$

$$q_{c_k}(x) = \frac{1}{2}x^2 + c_k \begin{cases} 0 & x \geq 1 \\ |x-1| & x < 1 \end{cases}$$

Si può dimostrare che per $c_k \geq 1$ $x^{(k)} = x^*$

Funzione di penalità esatta ma non differenziabile.

Il metodo del piano di taglio (funzioni convesse)

Sia $f \in C^{(1)}$ e sia Q un insieme compatto in $\mathbb{R}^n$

f è convessa. Vogliamo risolvere

$$\left\langle \begin{array}{l} \min f(x) \\ x \in Q \end{array} \right\rangle$$

f convessa e differenziabile $\Longrightarrow$

$$\Longrightarrow \quad f(x) \geq f(y) + \nabla f(y)^T (x-y) \quad \forall x, \forall y \qquad (*)$$

"Scegliamo" ora un certo numero k di punti $x^{(1)}, x^{(2)}, \dots, x^{(k)}$

Calcoliamo in essi $f(x^{(i)})$ e $\nabla f(x^{(i)})$, $i=1,\dots,k$ e costruiamo le k linearizzazioni, ciascuna "radicata" nello specifico $x^{(i)}$.

Otteniamo le funzioni affini $l_i(x)$, $i=1,\dots,k$

$$l_i(x) = f(x^{(i)}) + \nabla f(x^{(i)})^T (x - x^{(i)}) \quad i=1,\dots,k$$

La $(*)$ ci assicura che $\forall x$ è

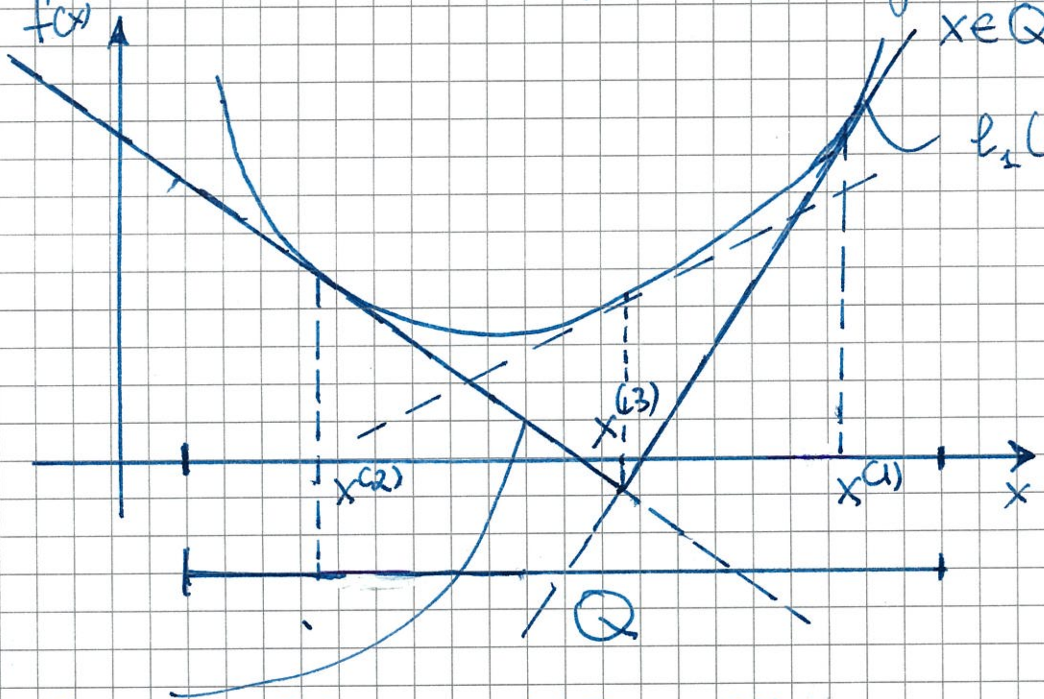
$$f(x) \geq l_i(x), \quad i=1,\dots,k$$

$$\Downarrow$$

$$f(x) \geq \max_{i=1,\dots,k} l_i(x) = f^{(k)}(x)$$

Il metodo del piano di taglio ("cutting plane") consiste nello schema iterativo:

$$x^{(k+1)} = \arg\min_{x\in Q} f^{(k)}(x) = \arg\min_{x\in Q} \max_{i=1,\dots,k} l_i(x)$$



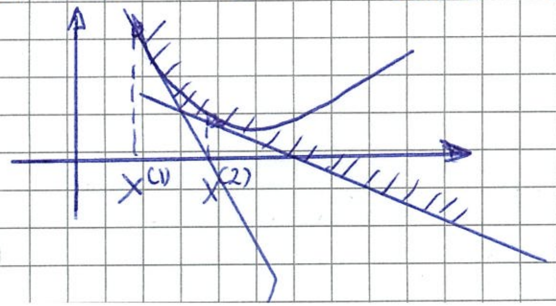
$$l_1(x) = f(x^{(1)}) + \nabla f(x^{(1)})^T (x - x^{(1)})$$

$$x^{(3)} = \arg\min f^{(2)}(x)$$

Dopo aver calcolato $x^{(3)}$, viene costruita la linearizzazione $l_3(x)$ che taglia un pezzo dell'epigrafo di $l_2(x)$ e migliora l'approssimazione di $f(x)$.

$$l_2(x) = f(x^{(2)}) + \nabla f(x^{(2)})^T (x - x^{(2)})$$

Nota. L'algoritmo può essere applicato anche al caso nonvincolato ($Q = \mathbb{R}^n$), ma possono esserci delle difficoltà. Se Q è compatto il punto di minimo $x^{(k+1)}$ esiste sempre ($f^{(k)}(x)$ è continua). Ciò non è più vero se $Q = \mathbb{R}^n$



Come si calcola $x^{(k+1)} = \arg\min_{x\in Q} \underbrace{\max_{i=1,\dots,k} \left[f(x^{(i)}) + \nabla f(x^{(i)})^T (x - x^{(i)}) \right]}_{v}$

$$\left\langle \begin{array}{l} \min\limits_{x\in\mathbb{R}^n,\, v} \; v \\ v \geq f(x^{(1)}) + \nabla f(x^{(1)})^T (x - x^{(1)}) \\ \vdots \\ v \geq f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)}) \\ x \in Q \end{array} \right\rangle$$

Se Q è rappresentato da vincoli lineari

(Es. $Q = \{ x \mid x \in \mathbb{R}^n, \; a_i \leq x_i \leq b_i, \; i = 1,\dots,n \}$ Vincoli "box")

il problema è del tipo Programmazione lineare.

E' possibile provare che $f^{(k)}(x^{(k+1)}) \xrightarrow[k\in K]{} f^*$ (Esiste una sottosuccessione convergente)

Nota La caratteristica più importante del metodo è che esso è il "progenitore" dei più efficienti metodi esistenti per il trattamento delle funzioni nondifferenziabili ("Bundle" methods)