PROBLEMI VINCOLATI

Consideriamo il solito problema di ottimizzazione P:

$$\begin{cases} \min f(x) \quad \forall x \in X \\ X = \{x \in \mathbb{R}^n : g_j(x) \leq 0,\ h_i(x) = 0,\ \forall i,j = 1 \dots \} \end{cases}$$

Con f, g e h continue e differenziabili.

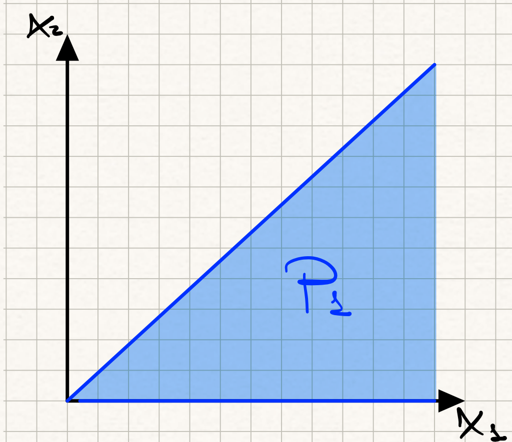
CONO TANGENTE

Definiamo il cono tangente in $x^* \in X$ come:

$$T_C(x^*) = \left\{ d \in \mathbb{R}^n : \exists \{z_k\} \subset X,\ \exists \{t_k\} > 0,\ z_k \to x^*,\ t_k \to 0,\ \lim_{k \to \infty} \frac{z_k - x^*}{t_k} = d \right\}$$

Esempio

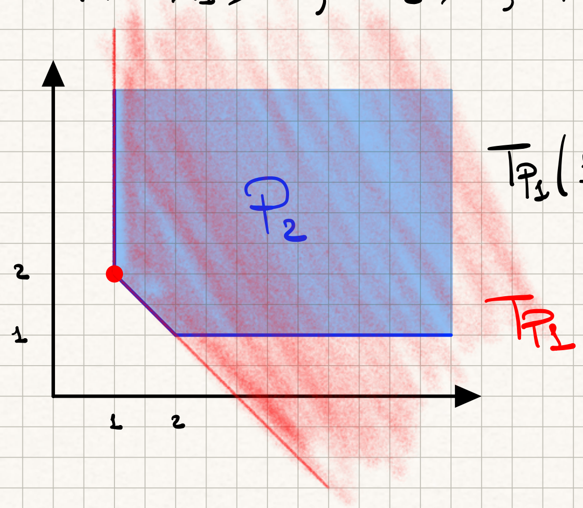
$$P_1 = \{ x \in \mathbb{R}^2 : x_2 \leq x_1,\ x_2 > 0 \}$$



$$T_{P_1}(0,0) = P_1$$

Esempio

$P_1 = \{x \in \mathbb{R}^2 : x_1 \geq 0, \; x_2 \geq 1, \; x_1 + x_2 \geq 3\}$



$T_{P_1}(1,2) = \{d \in \mathbb{R}^2 : d_1 \geq 0, \; d_1 + d_2 \geq 0\}$

SET DI DISEQUAZIONI ATTIVE

Si dice che un set A, formato da vincoli a modi disequazione, è attivo rispetto ad $x^* \in X$ se "tocca" la disequazione:

ATTIVO — NON-ATTIVO

Definiamo il set $A(x^*) = \{j : g_j(x^*) = 0\}$

FIRST ORDER FEASIBLE DIRECTION CONE

Definiamo il set $D(x^*)$:

$$D(x^*) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} d^T \nabla g_j(x^*) \leq 0 & \forall j \in \mathcal{A}(x^*) \\ d^T \nabla h_k(x^*) = 0 & \forall k = 1 \dots p \end{array} \right\}$$

È il cono di direzioni ammissibili del primo ordine.

DEFINIZIONE - Abadie constraint qualification (ACQ)

Diciamo che la qualificazione dei vincoli di Abadie è valida per un punto ammissibile $x^* \in X$ se:

$$T_X(x^*) = D(x^*)$$

Il nostro cono tangente $T_X(x^*)$ può essere separato con i gradienti dei vincoli $(\nabla g, \nabla h)$

TEOREMA

1) Vincoli affine

SE g_j e h_k sono funzioni affini $\forall j, k$, ALLORA la condizione ACQ è valida.

2) Slater condition

SE g_j è convessa $\forall j = 1 \dots m$, h_k è affine $\forall k = 1 \dots p$ e $\exists \bar{x} \in X$: $g(\bar{x}) < 0$ e $h(\bar{x}) = 0$ ALLORA la condizione ACQ è valida $\forall x \in X$.

3) Dipendenza lineare dei gradienti dei vincoli attivi.

SE $x^* \in X$ e i vettori:

$$\nabla g_j(x^*) \quad \forall j \in A(x^*)$$

$$\nabla h_k(x^*) \quad \forall k = 1 \dots p$$

sono linearmente indipendenti ALLORA la condizione ACQ è valida in x^*.

TEOREMA - KKT

SE x^* è un minimo locale e la condizione ACQ è verificata ALLORA $\exists \lambda^* \in \mathbb{R}^m$, $\exists \mu^* \in \mathbb{R}^m$: (λ^*, μ^*, x^*) soddisfano il seguente sistema

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) = \emptyset \\ \lambda_i^* g_i(x^*) = \emptyset \qquad \forall i = 1 \dots m \\ \lambda^* \geq \vec{\emptyset} \\ g(x^*) \leq 0 \\ h(x^*) = \emptyset \end{cases}$$

LAGRANGIANA

Definiamo la funzione lagrangiana $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$:

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Considerando la Lagrangiana il sistema KKT diventa

$$\begin{cases} \nabla_x L(x, \lambda, \mu) = 0 \\ \lambda_i g_i(x) = 0 \quad \forall i = 1 \dots m \\ \lambda \geq 0 \\ h(x) = 0 \\ g(x) \leq 0 \end{cases}$$

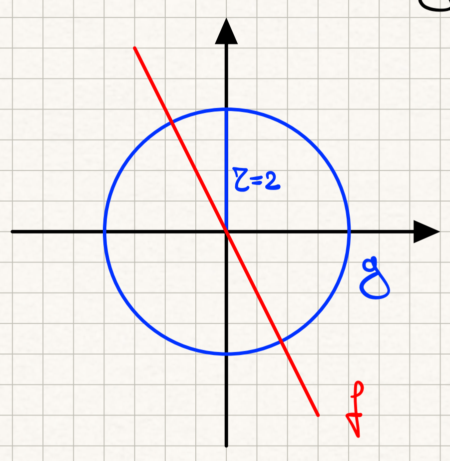
Nota: la condizione $\lambda_i g_i(x) = 0$ può essere riscritta in forma vettoriale:

$$\lambda^T g(x) = 0 \quad \text{o} \quad \langle \lambda, g(x) \rangle = 0$$

Esempio

$$\begin{cases} \min f(x) = 2x_1 + x_2 \\ x \in X = \{ x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 4 \} \end{cases}$$

La regione X è chiusa e limitata, $f(x)$ è continua in $\mathbb{R}^2$ quindi anche in $X \subset \mathbb{R}^2$, da qui vale il teorema di Weierstrass: $\exists x^* \in X$ minimo globale di $f(x)$ in X.



La Lagrangiana è:

$$L(x, \lambda, \mu) = 2x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 4)$$

Da cui il KKT:

$$\begin{cases} \begin{pmatrix} 2 + \lambda 2x_1 \\ 1 + \lambda 2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \lambda(x_1^2 + x_2^2 - 4) = 0 \\ \lambda > 0 \\ x_1^2 + x_2^2 - 4 \leq 0 \end{cases} \Longrightarrow \begin{cases} 2 + \lambda 2x_1 = 0 \\ 1 + \lambda 2x_2 = 0 \\ \lambda(x_1^2 + x_2^2 - 4) = 0 \\ \lambda > 0 \\ x_1^2 + x_2^2 - 4 \leq 0 \end{cases}$$

Nota che per $\lambda = 0$ il sistema è impossibile perché la prima equazione verrebbe $2 = 0$.

Risolvendo il sistema, abbiamo che:

$$\begin{cases} x_1 = -1/\lambda \\ x_2 = -1/2\lambda \\ x_1^2 + x_2^2 = 4 \\ \lambda \geq 0 \end{cases} \Longrightarrow \frac{1}{\lambda^2} + \frac{1}{4\lambda^2} = 4 \Longrightarrow \lambda = \frac{\pm\sqrt{5}}{4}$$

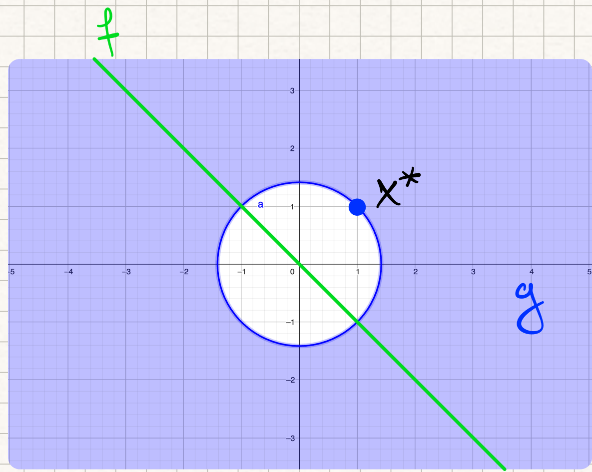
Da cui troviamo $x_1 = -\dfrac{4\sqrt{5}}{5}$, $x_2 = -\dfrac{2\sqrt{5}}{5}$

Per cui l'ottimo è $x^* = \left(-\dfrac{4\sqrt{5}}{5}; -\dfrac{2\sqrt{5}}{5}\right)$

Attenzione: Il KKT ci da una condizione NECESSARIA ma non SUFFICIENTE, rispetto all'ottimalità

Esempio

$$\begin{cases} \min \ x_1 + x_2 \\ -x_1^2 - x_2^2 + 2 \leq 0 \end{cases}$$



Attraverso la risoluzione del KKT arriviamo a ottenere

$$x^* = (1, 1) \ ; \quad \lambda^* = 1/2$$

Però x^* è un minimo locale. Infatti non esiste una sfera $B(x^*, \varepsilon)$ tale che $f(x^*)$ sia il minimo di tutti i punti $x \in B$.

TEOREMA - KKT PER PROBLEMI CONVESSI

SE il problema di ottimizzazione P è convesso e (x^*, λ^*, μ^*) risolve il KKT ALLORA x^* è un ottimo globale.

Dimostrazione

Assumiamo che (x^*, λ^*, μ^*) risolve il KKT, allora:

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

Allora abbiamo trovato il punto x^* minimo globale per L, dato che il nostro problema è convesso, da cui:

$$L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*) \quad \forall x \in \mathbb{R}^n$$

Riscrivendo L:

$$f(x^*) + (\lambda^*)^T g(x^*) + (\mu^*)^T h(x^*) \leq f(x) + (\lambda^*)^T g(x) + (\mu^*)^T h(x)$$

RILASSAMENTO LAGRANGIANO

Chiamiamo $v(P)$ l'ottimo valore per P.
Dato $\lambda \geq 0$ e $\mu \in \mathbb{R}^p$, il problema

$$\begin{cases} \inf L(x, \lambda, \mu) \\ x \in \mathbb{R}^n \end{cases}$$

È detto Lagrangian Relaxation di P.

Invece:

$$\varphi(\lambda, \mu) = \inf L(x, \lambda, \mu)$$

È detto Lagrangian dual function.
Quest'ultima funzione φ è:

1) Concava, perché è l'estremo inferiore di un set affine
2) Potrebbe succedere che $\varphi(\lambda, \mu) = -\infty$ per qualche punto
3) Potrebbe non essere differenziabile in qualche punto

TEOREMA

$\forall \lambda \geq 0$ e $\mu \in \mathbb{R}^p \Rightarrow \varphi(\lambda, \mu) \leq v(P)$

Dimostrazione

Se prendiamo $x \in X$, per definizione abbiamo che:

$$g(x) \leq 0 \qquad h(x) = 0$$

Allora:

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \leq f(x)$$

Ma dato che per definizione $\varphi(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$

$$\inf_{x \in \mathbb{R}^n} \left(L(x, \lambda, \mu) \right) \leq \inf_{x \in X} \left(L(x, \lambda, \mu) \right) \leq \inf_{x \in X} \left(f(x) \right) = v(P)$$

Assumendo che P abbia un valore ottimo finito.

PROBLEMA DUALE LAGRANGIANO

Definiamo il problema

$$\begin{cases} \max \varphi(\lambda, \mu) \\ \lambda \geq 0 \end{cases} = D$$

Lagrangian Dual Problem

In altre parole il duale D consiste nel trovare il miglior limite inferiore di $v(P)$.

Nota: D è sempre un problema convesso, anche quando P non lo è. È un problema dove bisogna massimizzare una funzione concava in un set convesso

PROGRAMMAZIONE LINEARE

Abbiamo il problema primale P:

$$\begin{cases} \min \ c^T x \\ Ax \geq b \end{cases}$$

La Lagrangiana

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) = \lambda^T b + (c^T - \lambda^T A)\, x$$

La funzione duale

$$\varphi(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda) = \begin{cases} -\infty & c^T - \lambda^T A \neq \emptyset \\ \lambda^T b & c^T - \lambda^T A = \emptyset \end{cases}$$

Problema duale D:

$$\begin{cases} \max \ \varphi(\lambda) \\ \lambda \geq \emptyset \end{cases} \Longrightarrow \begin{cases} \max \ \lambda^T b \\ \lambda^T A = c^T \\ \lambda \geq \emptyset \end{cases}$$

Quest'ultimo è un problema di programmazione lineare

LEAST NORM SOLUTION OF LINEAR EQUATION

Definiamo il problema primale P

1) $\begin{cases} \min \frac{1}{2}x^Tx \\ Ax=b \end{cases}$

2) $L(x,\mu) = \frac{1}{2}x^Tx + \mu(Ax-b)$

3) $\varphi(\mu) = \inf_{x\in\mathbb{R}^n} L(x,\mu) \underset{\downarrow}{=} \min_{x\in\mathbb{R}^n} L(x,\mu) = A^T\mu$

L Quadratica e strongly convex

$$\varphi(\mu) = -\frac{1}{2}\mu^T A A^T \mu + b^T\mu$$

4) Duale

$$\begin{cases} \max -\frac{1}{2}\mu^T A A^T \mu + b^T\mu \\ \mu\in\mathbb{R}^p \end{cases}$$

TEOREMA

Per qualsiasi problema di ottimizzazione P abbiamo che

$$v(D) \leq v(P)$$

Da cui definiamo la strong duality se $v(D)=v(P)$

Attenzione: la dualità forte non è verificata in generale

TEOREMA 5 - Condizione sufficiente per la dualità

Supponiamo f, h e g continue e differenziabili, formiamo il primale P:

$$\begin{cases} \min f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

Sia P convesso, esiste un ottimo globale x^* e l'ACQ è verificata ALLORA:

1) $v(D) = v(P)$ e D amette l'ottimo

2) (λ^*, μ^*) è l'ottimo per D SSE (λ^*, μ^*) è un vettore dei moltiplicatori associati a x^*

Dimostrazioni

$L(x, \lambda, \mu)$ è convessa perchè abbiamo supposto P convesso.
Sia (λ^*, μ^*) un vettore dei moltiplicatori per x^* per il sistema KKT, ALLORA

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \quad \lambda^* \geq 0, \quad (\lambda^*)^T g(x^*) = 0$$

$$v(D) \geq \varphi(\lambda^*, \mu^*) = \inf_x L(x, \lambda^*, \mu^*) \underset{L \text{ convessa}}{=} \inf_x L(x^*, \lambda^*, \mu^*) =$$

$$= f(x^*) + \underbrace{(\lambda^*)^T g(x^*)}_{=0} + \underbrace{(\mu^*)^T h(x^*)}_{=0} = f(x^*) = v(P) \geq v(D)$$

Ma dato che siamo partiti da $v(D)$ allora vale l'uguaglianza $v(P) = v(D)$ e (λ^*, μ^*) è l'ottimo per D.

SE invece (λ^*, μ^*) è una qualsiasi soluzione ottima per D **ALLORA**

$$f(x^*) = v(P) = v(D) = \varphi(\lambda^*, \mu^*) = \inf_x L(x, \lambda^*, \mu^*) \leq$$
$$\leq L(x^*, \lambda^*, \mu^*) = f(x^*) + (\lambda^*)^T g(x^*) + \underbrace{(\mu^*)^T h(x^*)}_{=0} =$$
$$= f(x^*) + (\lambda^*)^T g(x^*) \leq f(x^*)$$

Ma $(\lambda^*)^T g(x^*) = 0$ e $\inf_x L(x, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*)$, da cui viene da se che:

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

Ovvero (λ^*, μ^*) sono una coppia di vettori moltiplicatori KKT associati a x^*.

TEOREMA - Caratterizzazione della dualità forte

(x^*, λ^*, μ^*) è un punto di sella per L, ovvero:

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*)$$

$\forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^m_+, \quad \forall \mu \in \mathbb{R}^p$

SSE x^* è ottimo per P, (λ^*, μ^*) è ottimo per D e $v(P) = v(D)$.