DUALITA' DI WOLFE - PROBLEMI QUADRATICI

$$\min f(x) \quad g(x) \quad h(x)=0 \quad (P)$$

con $f, g_i \in C^1(\mathbb{R}^n)$ convesse, h_j lineari

$\not\Rightarrow$ (D) $\max \mathcal{L}(x,\lambda,\mu)$, $\nabla_x \mathcal{L}(x,\lambda,\mu)=0$, $\lambda \geq 0$

(D) → PROBLEMA DUALE DI WOLFE

x^* ottimo globale di (P) $\xLeftarrow{\text{sotto ipotesi di regolarità dei vincoli}}$ (x^*, λ^*, μ^*) PUNTO DI SELLA DELLA LAGRANGIANA $\Longleftrightarrow$ (x^*, λ^*, μ^*) PUNTO DI KKT $\Rightarrow$ (x^*, λ^*, μ^*) è soluzione ottima del duale di Wolfe $\mathcal{L}(x^*, \lambda^*, \mu^*) = f(x^*)$

↑ teorema di Wolfe di dualità forte

$\not\Leftarrow$

↓ quello che faremo oggi

TEOREMA

$f, g_i \in C^1(\mathbb{R}^n)$ convesse ; h_j lineari $i = 1, \dots, m$, $j = 1, \dots, p$

Sia (x^*, λ^*, μ^*) PUNTO KKT per il problema primale (P)

Se $(\hat{x}, \hat{\lambda}, \hat{\mu})$ è soluzione ottima del problema duale di Wolfe e la funzione Lagrangiana è STRETTAMENTE CONVESSA in $\hat{x}$

allora

$\hat{x} = x^*$ è un punto di minimo globale del problema primale (P)

e $f(\hat{x}) = \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\mu})$

DIM:

Supponiamo per assurdo che $\hat{x} \neq x^*$

(x^*, λ^*, μ^*) è KKT per (P) $\Rightarrow$ ⊙ x^* è soluzione ottima di (P)

per il teorema dualità forte

⊙ $f(x^*) = \mathcal{L}(x^*, \lambda^*, \mu^*)$

con (x^*, λ^*, μ^*) soluzione ottima di (D)

⇓

deve valere

$$\mathcal{L}(x^*, \lambda^*, \mu^*) = \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\mu})$$

$\mathcal{L}(x, \lambda, \mu)$ è strettamente convessa in $\hat{x}$

$$\mathcal{L}(x^*, \hat{\lambda}, \hat{\mu}) > \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\mu}) + \nabla \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\mu})^T (x^* - \hat{x})$$

Poiché $\nabla_x \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\mu}) = 0$ (essendo $(\hat{x}, \hat{\lambda}, \hat{\mu})$ ammissibile per il problema duale di Wolfe)

$$\mathcal{L}(x^*, \hat{\lambda}, \hat{\mu}) > \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\mu}) = \mathcal{L}(x^*, \lambda^*, \mu^*)$$

condizioni di complementarietà di KKT

$$\cancel{f(x^*)} + \hat{\lambda}^T g(x^*) + \underbrace{\hat{\mu} h(x^*)}_{=0} > \cancel{f(x^*)} + \underbrace{\lambda^{*T} g(x^*)}_{=0} + \underbrace{\mu^{*T} h(x^*)}_{=0}$$

(x^*, λ^*, μ^*) è KKT per (P)

Quindi arriviamo a dire che

$$\hat{\lambda}^T g(x^*) > 0 \qquad \text{ASSURDO!}$$

poiché:

- $\hat{\lambda} \geq 0 \quad \leftarrow \quad \hat{\lambda}$ è ammissibile per il problema duale di Wolfe
- $g(x^*) \leq 0 \quad \leftarrow \quad x^*$ è ammissibile per il problema primale

e quindi $\hat{\lambda}^T g(x^*) \leq 0$

$\Rightarrow$ Quindi deve essere $x^* = \hat{x}$

poiché $f(x^*) = \mathcal{L}(x^*, \lambda^*, \mu^*) = \mathcal{L}(\hat{x}, \hat{\lambda}, \hat{\mu}) = f(\hat{x})$ ■

PROBLEMI DI PROGRAMMAZIONE QUADRATICA

I problemi di programmazione quadratica hanno:

- funzione obiettivo quadratica

$$f(x) = \frac{1}{2} x^T Q x + c^T x + d$$

$$\begin{bmatrix} \min \ \frac{1}{2} x^T Q x + c^T x + d \\ Ax \leq b \end{bmatrix}$$

- vincoli lineari

$$a_j^T x \leq b_j \qquad j = 1, \dots, m$$

$\downarrow$

forma compatta

$$Ax \leq b \qquad \text{con } A \in \mathbb{R}^{m \times n} \ ; \ b \in \mathbb{R}^m$$

Supponiamo $Q \in \mathbb{R}^{n \times n}$ simmetrica, semidefinita positiva

$\Rightarrow$ $f(x) = \frac{1}{2} x^T Q x + c^T d + d$ è convessa

Quindi il problema di programmazione quadratica è un problema di programmazione convessa con f convessa

g_i lineari (e quindi convessa) $i = 1, \dots, m$

$\Rightarrow$ possiamo definire il problema duale di Wolfe.

PROP:

Sia (P) un problema di programmazione quadratica con Q semidefinita positiva.
Il suo problema duale di Wolfe è dato da

$$\min \frac{1}{2} x^T Q x + b^T \lambda$$
$$Qx + A^T \lambda + c = 0$$
$$\lambda \geq 0 \qquad \text{con} \quad \lambda \in \mathbb{R}^m$$

DIM:

Per definizione il duale di Wolfe è dato da

$$\max \mathcal{L}(x, \lambda, \mu)$$
$$\nabla_x \mathcal{L}(x, \lambda, \mu) = 0$$
$$\lambda \geq 0$$

Per il problema (P) di programmazione quadratica otteniamo:

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T g(x) = \frac{1}{2} x^T Q x + c^T x + d + \lambda^T (Ax - b)$$

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) + \nabla g(x) \lambda$$

$$\text{con} \quad \nabla f(x) = Qx + c$$

Calcoliamo il gradiente* di

$$\lambda^T (Ax - b) = \lambda^T A x - \lambda^T b$$

* rispetto a x

$$\lambda^T A x = \lambda^T \begin{pmatrix} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n \\ \vdots \\ a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n \end{pmatrix} = \begin{matrix} \lambda_1 a_{11} x_1 + \lambda_1 a_{12} x_2 + \dots + \lambda_1 a_{1n} x_n + \\ + \lambda_2 a_{21} x_1 + \lambda_2 a_{22} x_2 + \dots + \lambda_2 a_{2n} x_n + \\ + \dots \qquad\qquad + \\ + \lambda_m a_{m1} x_1 + \lambda_m a_{m2} x_2 + \dots + \lambda_m a_{mn} x_n \end{matrix}$$

$$w^T v = v^T w$$

$$\lambda^T A x = x^T A^T \lambda = (Ax)^T \lambda$$

Calcolando le derivate parziali $\lambda^T A x$ rispetto a $x_1, x_2, \dots, x_n$ otteniamo

$$\frac{\partial(\lambda^T A x)}{\partial x_1} = \lambda_1 a_{11} + \lambda_2 a_{21} + \dots + \lambda_m a_{m1} = \sum_{i=1}^{m} \lambda_i a_{i1} \Big) \, (A^T\lambda)_1$$

$$\frac{\partial(\lambda^T A x)}{\partial x_1} = \lambda_2 a_{12} + \lambda_2 a_{22} + \dots + \lambda_m a_{m2} = \sum_{i=1}^{m} \lambda_i a_{i2} \Big) \text{ seconda riga di } A^T\lambda$$

$$\vdots$$

$$\frac{\partial(\lambda^T A x)}{\partial x_2} = \lambda_1 a_{12} + \lambda_2 a_{22} + \dots + \lambda_m a_{m2} = \sum_{i=1}^{m} \lambda_i a_{i2} \Big) \text{ n-esima riga di } A^T\lambda$$

$$\nabla_x \left(\lambda^T (Ax - b) \right) = A^T\lambda$$

Quindi

$$\nabla_x \mathcal{L}(x,\lambda) = \underbrace{Qx + c}_{\nabla f(x)} + \overbrace{A^T\lambda}^{= \nabla g(x)\lambda}$$

e il duale di Wolfe di (P) è dato da

$$\max \ \frac{1}{2} x^T Q x + c^T x + d + \lambda^T (Ax - b)$$
$$Qx + c + A^T\lambda = 0$$
$$\lambda \geq 0$$

Scriviamo $\frac{1}{2} x^T Q x =$ manca un pezzo

...

$$\frac{1}{2} x^T Q x + c^T x + d + \lambda^T (Ax - b) =$$

$$= -\frac{1}{2} x^T Q x + x^T Q x + c^T x + d + \lambda^T A x - \lambda^T b$$

$$= -\frac{1}{2} x^T Q x + x^T Q x + x^T c + d + x^T A^T \lambda - b^T \lambda$$

$$= -\frac{1}{2} x^T Q x + x^T \overbrace{\left(Qx + c + A^T\lambda \right)}^{= \nabla_x \mathcal{L}(x,\lambda)} + d - b^T\lambda$$

Quindi il problema di Wolfe si può riscrivere come

$$\max\ -\frac{1}{2}x^TQx + x^T(Qx + c + A^T\lambda) - \lambda^Tb + d$$
$$Qx + c + A^T\lambda = 0$$
$$\lambda \geq 0$$

che equivale a

costante

$$\max\ -\frac{1}{2}x^TQx - \lambda^Tb + d \quad \Longleftrightarrow \quad -\min\ \frac{1}{2}x^TQx + \lambda^Tb - d$$
$$Qx + c + A^T\lambda = 0 \qquad\qquad Qx + C + A^T\lambda = 0$$
$$\lambda \geq 0 \qquad\qquad \lambda \geq 0$$

cambiando segno alla f.o

Supponiamo ora che sia Q definita positiva

$$(x^TQx > 0 \quad \forall x \in \mathbb{R}^n,\ x \neq 0)$$

abbiamo che

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^TQx + c^Tx + d + \lambda^T(Ax - b)$$

è ancora una f. quadratica rispetto a x

ed è strettamente convessa !

$$\mathcal{L}(x,\lambda) = \underbrace{\frac{1}{2}x^TQx}_{\text{parte quadratica rispetto } x} + \underbrace{x^T(c + A^T\lambda)}_{\text{parte lineare rispetto } x} \underbrace{- b^T\lambda + d}_{\text{costante rispetto a } x}$$

Siamo nelle ipotesi per cui

(x^*, λ^*) soluzione del problema duale di Wolfe $\Longrightarrow$ x^* soluzione ottima del problema primale (P)

la funzione $\mathcal{L}$ è strett. convessa $\forall x$ (poiché Q def positiva)

PROP:

Sia (P) un problema di programmazione quadratica con Q definita positiva

La soluzione ottima di (P) è data da

$$\boxed{x^* = -Q^{-1}(A^T\lambda^* + c)}$$

dove

$\lambda^* \in \mathbb{R}^m$ è la soluzione ottima di $\boxed{\begin{array}{l} \min \ \frac{1}{2}(\lambda^T A)Q^{-1}(A^T\lambda) + (AQ^{-1}c + b)^T\lambda \\ \lambda \geq 0 \end{array}}$

DIM:

Q definita positiva $\Rightarrow$ la funzione obiettivo del problema primale (è coerciva)

è strettamente convessa

$\Rightarrow \exists! \ x^*$ punto di minimo globale di (P)

· Q è invertibile e Q^{-1} è simmetrica

manca un pezzo

le soluzioni ammissibili del problema di Wolfe soddisfano

$$\nabla_x \mathcal{L}(x, \lambda) = 0$$
$$\Updownarrow$$
$$Qx + A^T\lambda + c = 0$$
$$\Updownarrow$$
$$Qx = -(A^T\lambda + c)$$
$$\Updownarrow$$

moltiplico a sx e dx per Q^{-1}

$$\underbrace{Q^{-1}Q}_{= I}\, x = -Q^{-1}(A^T\lambda + c)$$

I matrice identità

Quindi il duale di Wolfe diventa

$$\begin{array}{l} \min \ \frac{1}{2}x^T Q x + \lambda^T b + d \\ x = -Q^{-1}(A\lambda + c) \\ \lambda \geq 0 \end{array}$$

possiamo sostituire l'espressione di x fornita dal vincolo $\nabla_x \mathcal{L}(x,\lambda) = 0$ nella funzione obiettivo

$$\mathcal{L}(x,\lambda) = \mathcal{L}((-Q^{-1}(A^T\lambda + c)), \lambda) =$$

$$= \frac{1}{2}\left[-Q^{-1}(A^T\lambda + c)\right]^T Q \left[-Q^{-1}(A^T\lambda + c)\right] + \lambda^T b + d$$

$$= \frac{1}{2}\left[\lambda^T A \underbrace{Q^{-1} Q\, Q}_{I}{}^{-1}(A^T\lambda) + 2c^T \underbrace{Q^{-1} Q\, Q}_{I}{}^{-1}(A^T\lambda) + c^T Q c\right] + b^T\lambda + d$$

matrice identità

$$= \frac{1}{2}\left[\lambda^T A Q^{-1} A^T \lambda + 2c^T Q^{-1} A\lambda\right] + \frac{1}{2} c^T Q c + b^T \lambda + d$$

$$= \underbrace{\frac{1}{2}\lambda^T (A Q^{-1} A^T \lambda)}_{\text{parte quadratica in } \lambda} + \underbrace{(A Q^{-1} c + b)^T \lambda}_{\text{parte lineare in } \lambda} + \underbrace{\frac{1}{2} c^T Q c + d}_{\text{parte costante}}$$

Quindi il problema duale di Wolfe si riduce ad un problema nella sola variabile λ:

N.B la dimensione nel problema duale e' quella di λ

$$\begin{cases} \min \frac{1}{2}\lambda^T (A Q^{-1} A^T)\lambda + (A Q^{-1} c + b)^T \lambda + c^T Q c + d \\ \lambda \geq 0 \end{cases}$$

conveniente quando (P) ha pochi vincoli lineari

risolvendo il duale di Wolfe ottengo λ^* e inserendo λ^* nell'espressione di x^* ottenuta da $\nabla_x \mathcal{L}(x,\lambda) = 0$ si ottiene il punto di minimo globale di (P)

$$x^* = -Q^{-1}(A^T\lambda^* + c)$$