# Dualidad
# Optimización II. Capítulo 12.

Oscar S. Dalmau Cedeño
dalmau@cimat.mx

13 de septiembre de 2018

1. Resumen
2. Sensibilidad
3. Dualidad

# Cono Crítico

## Cono Crítico

$$\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \quad \Leftrightarrow \quad \left\{ \begin{array}{ll} \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) = 0 & \text{si } i \in \mathcal{E} \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) = 0 & \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}, \lambda_i^* > 0 \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) \geq 0 & \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}, \lambda_i^* = 0 \end{array} \right.$$

# Condición necesaria de segundo orden

Sea $\boldsymbol{x}^*$ una solución del problema original y ademas satisface la condición LICQ. Si $\boldsymbol{\lambda}^*$ satisface las KKTs, entonces

$$\boldsymbol{w}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{w} \quad \geq \quad 0, \text{ para todo } \boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$$

# Condición suficiente de segundo orden

Suponga que $\boldsymbol{x}^*$ es un punto factible y que existe $\boldsymbol{\lambda}^*$ satisface las KKTs. Si

$$\boldsymbol{w}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{w} \quad > \quad 0, \text{ para todo } \boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$$

entonces $\boldsymbol{x}^*$ es un mínimo local del problema original.

# Resumen

## Pasos

1. Determinar Puntos Críticos usando las KKTs
2. Cacular $\nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$
3. Determinar el cono crítico $\mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$
4. Verificar si se cumple la condición:

$$\boldsymbol{w}^T \nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{w} \quad > \quad 0, \text{ para todo } \boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$$

Si se cumple la condición anterior, entonces la función tiene un mínimo local en $\boldsymbol{x}^*$

# Ejemplo 1

$$\text{min} \ -0{,}1(x_1-4)^2+x_2^2, \ \text{s.t. } x_1^2+x_2^2-1\geq 0$$

Lagrangiano

$$\mathcal{L}(x_1,x_2,\lambda) = -0{,}1(x_1-4)^2+x_2^2-\lambda(x_1^2+x_2^2-1)$$

KKTs

$$\begin{aligned}
\nabla_x\mathcal{L}(x_1,x_2,\lambda) &= 0\\
x_1^2+x_2^2-1 &\geq 0\\
\lambda &\geq 0\\
\lambda(x_1^2+x_2^2-1) &= 0
\end{aligned}$$

# Ejemplo 1

KKTs

$$
\begin{aligned}
-0{,}2(x_1 - 4) - 2\lambda x_1 &= 0 \\
2x_2 - 2\lambda x_2 &= 0 \\
x_1^2 + x_2^2 - 1 &\geq 0 \\
\lambda &\geq 0 \\
\lambda(x_1^2 + x_2^2 - 1) &= 0
\end{aligned}
$$

Solución $\boldsymbol{x}^* = [1, 0]^T$, $\lambda^* = 0{,}3$ y el conjunto activo es $\mathcal{A}(\boldsymbol{x}^*) = \{1\}$ ... es minimo?

**Nota:** Otros puntos críticos son $\boldsymbol{x}^* = [\frac{4}{11}, \pm\sqrt{1 - \frac{16}{121}}]$ y $\lambda^* = 1$.
$\boldsymbol{x}^* = [4, 0]^T$ y $\lambda^* = 0$

# Ejemplo 1

Cono critico: como $\lambda^* = 0{,}3 > 0$ y $\mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} = \{1\}$ entonces para $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$ se tiene que cumplir $\boldsymbol{w}^T \nabla c(\boldsymbol{x}^*) = 0$.

$$\nabla c(\boldsymbol{x}^*) \quad = \quad 2[\boldsymbol{x}_1^*, \boldsymbol{x}_2^*]^T = [2, 0]^T$$

Luego de $\boldsymbol{w}^T \nabla c(\boldsymbol{x}^*) = 0$ se tiene que $w_1 = 0$, por lo que

$$\mathcal{C}(\boldsymbol{x}^*, \lambda^*) \quad = \quad \{[0, w_2]^T |\, w_2 \in \mathbb{R}\}$$

# Ejemplo 1

Finalmente, para $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$ se tiene que

$$\boldsymbol{w}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \lambda^*) \boldsymbol{w} \quad = \quad 1{,}4 w_2^2 > 0,$$

por lo que $\boldsymbol{x}^* = [1, 0]^T$ es un mínimo!.

Supongamos que la restricción $i$ se activa en el óptimo $\boldsymbol{x}^*$, ie, $c_i(\boldsymbol{x}^*) = 0$, vamos a perturbar dicha restricción, de modo que el resto de las activas e inactivas no cambia, es decir tenemos la nueva restricción

$$\tilde{c}_i(\boldsymbol{x}) = c_i(\boldsymbol{x}) + \epsilon \|\nabla c_i(\boldsymbol{x}^*)\| \geq 0$$

Supongamos que la nueva solución perturbada $\boldsymbol{x}^*(\epsilon)$, activa la restricción anterior, ie, como $c_i(\boldsymbol{x}^*) = 0$

$$\begin{aligned} c_i(\boldsymbol{x}^*(\epsilon)) + \epsilon \|\nabla c_i(\boldsymbol{x}^*)\| &= 0 \\ c_i(\boldsymbol{x}^*(\epsilon)) - c_i(\boldsymbol{x}^*) &= -\epsilon \|\nabla c_i(\boldsymbol{x}^*)\| \\ (\boldsymbol{x}^*(\epsilon) - \boldsymbol{x}^*)^T \nabla c_i(\boldsymbol{x}^*) &\approx -\epsilon \|\nabla c_i(\boldsymbol{x}^*)\| \text{ por Taylor} \end{aligned}$$

El resto de las restricciones se mantienen activas $c_j(\boldsymbol{x}^*(\epsilon)) = c_j(\boldsymbol{x}^*) = 0$ usando taylor

$$0 = c_j(\boldsymbol{x}^*(\epsilon)) - c_j(\boldsymbol{x}^*) \quad \approx \quad (\boldsymbol{x}^*(\epsilon) - \boldsymbol{x}^*)^T \nabla c_j(\boldsymbol{x}^*)$$

Usando Taylor para la $f(.)$

$$\begin{aligned}
f(\boldsymbol{x}^*(\epsilon)) - f(\boldsymbol{x}^*) &\approx (\boldsymbol{x}^*(\epsilon) - \boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^*) \\
&= (\boldsymbol{x}^*(\epsilon) - \boldsymbol{x}^*)^T \sum_{k \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_k^* \nabla c_k(\boldsymbol{x}^*) \\
&= \sum_{k \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_k^* (\boldsymbol{x}^*(\epsilon) - \boldsymbol{x}^*)^T \nabla c_k(\boldsymbol{x}^*) \\
&= -\epsilon \lambda_i^* \|\nabla c_i(\boldsymbol{x}^*)\|
\end{aligned}$$

Supongamos que el resto de las restricciones se mantienen activas $c_j(\boldsymbol{x}^*(\epsilon)) = c_j(\boldsymbol{x}^*) = 0$ usando taylor

$$0 = c_j(\boldsymbol{x}^*(\epsilon)) - c_j(\boldsymbol{x}^*) \quad \approx \quad (\boldsymbol{x}^*(\epsilon) - \boldsymbol{x}^*)^T \nabla c_j(\boldsymbol{x}^*)$$

Usando Taylor para la $f(.)$

$$\begin{aligned} \frac{f(\boldsymbol{x}^*(\epsilon)) - f(\boldsymbol{x}^*)}{\epsilon} &\approx -\lambda_i^* \|\nabla c_i(\boldsymbol{x}^*)\| \\ \frac{df(\boldsymbol{x}^*(\epsilon))}{d\epsilon} &= -\lambda_i^* \|\nabla c_i(\boldsymbol{x}^*)\| \end{aligned}$$

- Si $\lambda_i^* \|\nabla c_i(\boldsymbol{x}^*)\|$ es muy grande entonces el valor óptimo es muy sensible
- Si $\lambda_i^* \|\nabla c_i(\boldsymbol{x}^*)\|$ es muy pequeño entonces el valor óptimo no es muy sensible
- Si $\lambda_i^* = 0$ entonces pequeñas perturbaciones en la restricción afectaria poco o no afecta al valor óptimo de la función objetivo.

# Dualidad

Sea el siguiente problema con restricciones de desigualdad

$$\begin{aligned} \text{min } & f(\boldsymbol{x}) \\ & s.t. \quad c(\boldsymbol{x}) \succeq \mathbf{0} \end{aligned}$$

donde $c(\boldsymbol{x}) = [c_1(\boldsymbol{x}), c_2(\boldsymbol{x}), \cdots, c_m(\boldsymbol{x})]^T$

# Función dual

Lagrangiano

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \quad = \quad f(\boldsymbol{x}) - \boldsymbol{\lambda}^T c(\boldsymbol{x})$$

## Función dual

$$\begin{aligned} q(\boldsymbol{\lambda}) &\overset{def}{=} \inf_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \\ \mathcal{D} &= \{\boldsymbol{\lambda} |\; q(\boldsymbol{\lambda}) > -\infty\} \end{aligned}$$

donde $\mathcal{D}$ es el dominio de $q()$.

# Función dual

- Encontrar el óptimo global $\mathcal{L}(\cdot, \boldsymbol{\lambda})$ puede ser muy dificil
- Sin embargo, si $f()$ y las restricciones $-c_i()$ son funciones convexas (lo que ocurre en varias aplicaciones) entonces $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ es convexa, y sus minimos locales son minimos globales.
- Determinar $q(\boldsymbol{\lambda})$ en el caso anterior puede ser mas practico

# Problema dual

## Problema (Lagrangiano) dual

$$\text{max}\ _{\boldsymbol{\lambda}} q(\boldsymbol{\lambda}) \qquad s.t.\ \ \boldsymbol{\lambda} \succeq \mathbf{0}$$

# Ejemplo

$$\text{min } 0{,}5(x_1^2 + x_2^2),\ s.t:\ x_1 - 1 \geq 0$$

Lagrangiano

$$\mathcal{L}(x_1, x_2, \lambda) = 0{,}5(x_1^2 + x_2^2) - \lambda(x_1 - 1)$$

Para $\lambda$ fijo, busquemos el minimo!

$$\begin{aligned} \nabla_{x_1}\mathcal{L} &= x_1 - \lambda = 0 \\ \nabla_{x_2}\mathcal{L} &= x_2 = 0 \end{aligned}$$

Luego $[x_1, x_2] = [\lambda, 0]$

# Ejemplo: Problema dual

$$\text{max } q(\lambda) = -0{,}5\lambda^2 + \lambda,\ s.t:\ \lambda \geq 0$$

y la solución es $\lambda = 1$.
Luego $[x_1, x_2] = [\lambda, 0] = [1, 0]$

# Problema dual

**Teorema: Concavidad**

La función dual (definida arriba) es concava y su dominio els convexo

La demostración es directa y se basa en la siguiente propiedad

**Propiedad del ínfimo**

$$\inf f(\boldsymbol{x}) + \inf g(\boldsymbol{x}) \quad \leq \quad \inf f(\boldsymbol{x}) + g(\boldsymbol{x})$$

Nota: Ver en Mathematica PropiedadDeInfimos.nb

# Problema dual

- Sean $\boldsymbol{\lambda_0}$ y $\boldsymbol{\lambda_1}$ en el dominio $\mathcal{D}$ de $q(\cdot)$, ie, $q(\boldsymbol{\lambda_0}) > -\infty$ y $q(\boldsymbol{\lambda_1}) > -\infty$.
- Sea $\alpha \in [0, 1]$ entonces, por la linealidad de $\mathcal{L}(., \boldsymbol{\lambda})$

$$\mathcal{L}(\boldsymbol{x}, \alpha\boldsymbol{\lambda_0} + (1-\alpha)\boldsymbol{\lambda_1}) \quad = \quad \alpha\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda_0}) + (1-\alpha)\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda_1})$$

- Luego

$$\begin{aligned} \text{ínf}\, \mathcal{L}(\boldsymbol{x}, \alpha\boldsymbol{\lambda_0} + (1-\alpha)\boldsymbol{\lambda_1}) &\geq \alpha\, \text{ínf}\, \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda_0}) + (1-\alpha)\, \text{ínf}\, \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda_1}) \\ q(\alpha\boldsymbol{\lambda_0} + (1-\alpha)\boldsymbol{\lambda_1}) &\geq \alpha q(\boldsymbol{\lambda_0}) + (1-\alpha)q(\boldsymbol{\lambda_1}) \end{aligned}$$

Y se concluye que $q()$ es concava.

# Problema dual

- Sean $\boldsymbol{\lambda_0}$ y $\boldsymbol{\lambda_1}$ en el dominio $\mathcal{D}$ de $q(\cdot)$, ie, $q(\boldsymbol{\lambda_0}) > -\infty$ y $q(\boldsymbol{\lambda_1}) > -\infty$.
- Sea $\alpha \in [0, 1]$ entonces, por la concavidad de $q()$

$$q(\alpha\boldsymbol{\lambda_0} + (1-\alpha)\boldsymbol{\lambda_1}) \quad \geq \quad \alpha q(\boldsymbol{\lambda_0}) + (1-\alpha)q(\boldsymbol{\lambda_1}) > -\infty$$

- Luego $\alpha\boldsymbol{\lambda_0} + (1-\alpha)\boldsymbol{\lambda_1} \in \mathcal{D}$ es decir, $\mathcal{D}$ es convexo

# Problema dual

**Teorema: Dualidad débil**

Para cualquier punto factible $\bar{\boldsymbol{x}}$ del problema original (el problema con restricciones de desigualdad) y para cualquier $\bar{\boldsymbol{\lambda}} \succeq \mathbf{0}$ se cumple $q(\bar{\boldsymbol{\lambda}}) \leq f(\bar{\boldsymbol{x}})$

$$q(\bar{\boldsymbol{\lambda}}) \quad = \quad \text{inf}_{\boldsymbol{x}} f(\boldsymbol{x}) - \bar{\boldsymbol{\lambda}}^T c(\boldsymbol{x}) \leq f(\bar{\boldsymbol{x}}) - \bar{\boldsymbol{\lambda}}^T c(\bar{\boldsymbol{x}}) \leq f(\bar{\boldsymbol{x}})$$

pues $c(\bar{\boldsymbol{x}}), \bar{\boldsymbol{\lambda}} \succeq \mathbf{0}$.

# Problema dual

**Teorema: Dualidad debil**

Para cualquier punto factible $\bar{\boldsymbol{x}}$ del problema original (el problema con restricciones de desigualdad) y para cualquier $\bar{\boldsymbol{\lambda}} \succeq \mathbf{0}$ se cumple $q(\bar{\boldsymbol{\lambda}}) \leq f(\bar{\boldsymbol{x}})$

Luego, el óptimo del problema dual, nos da una cota inferior del valor de la función objetivo del problema primal u original.

# Margen dual (duality gap)

- El *margen o diferencia de dualidad* es la diferencia entre las soluciones primal y dual.
- Si $d^*$ es el valor dual óptimo de la función objetivo y $p^*$ es el valor primal óptimo entonces la *diferencia de dualidad* es igual a $p^* - d^*$.
- Este valor siempre es mayor o igual a 0 (para problemas de minimizacion).

# Margen dual (duality gap)

- La dualidad fuerte es un concepto de optimización tal que las soluciones primal y dual son equivalentes.
- La diferencia de la dualidad es cero si y solo si se cumple la dualidad fuerte.
- De lo contrario, la diferencia es estrictamente positiva y se cumple la dualidad débil

# Dual de Wolfe

$$\begin{aligned} \max_{\boldsymbol{x},\boldsymbol{\lambda}} \quad & \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) \\ s.t.: \quad & \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \mathbf{0},\ \boldsymbol{\lambda} \succeq 0 \end{aligned}$$

Desde el punto de vista de calculo, esta formulación es mas conveniente

# Dual de Wolfe

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\
s.t.: \quad & A\boldsymbol{x} = \boldsymbol{b}, \\
& \boldsymbol{x} \succeq 0
\end{aligned}
$$

$$
\begin{aligned}
\max_{\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\lambda}} \quad & \mathcal{L}(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\lambda}) = \boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{\mu}^T(A\boldsymbol{x}-\boldsymbol{b}) - \boldsymbol{\lambda}^T\boldsymbol{x} \\
s.t.: \quad & \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \boldsymbol{0}, \Rightarrow A^T\boldsymbol{\mu} + \boldsymbol{\lambda} = \boldsymbol{c} \\
& \boldsymbol{\lambda} \succeq \boldsymbol{0}
\end{aligned}
$$

# Dual de Wolfe

$$\begin{aligned} \max_{\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\lambda}} \quad \mathcal{L}(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\lambda}) &= (\boldsymbol{c} - A^T\boldsymbol{\mu} - \boldsymbol{\lambda})^T\boldsymbol{x} + \boldsymbol{\mu}^T\boldsymbol{b} \\ s.t.: \quad \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\lambda) &= 0, \Rightarrow A^T\boldsymbol{\mu} + \boldsymbol{\lambda} = \boldsymbol{c} \\ \boldsymbol{\lambda} \succeq 0 & \end{aligned}$$

$$\begin{aligned} \max_{\boldsymbol{\mu},\boldsymbol{\lambda}} \quad & \boldsymbol{\mu}^T\boldsymbol{b} \\ s.t.: \quad A^T\boldsymbol{\mu} + \boldsymbol{\lambda} &= \boldsymbol{c} \\ \boldsymbol{\lambda} \succeq 0 & \end{aligned}$$

$$\begin{aligned} \max_{\boldsymbol{\mu}} \quad & \boldsymbol{\mu}^T\boldsymbol{b} \\ s.t.: \quad A^T\boldsymbol{\mu} &\preceq \boldsymbol{c} \end{aligned}$$

# Dual de Wolfe

**Teorema**

Suponga que $f$ y $-c_i$ son funciones convexas y continuamente diferenciables. Si el punto $\boldsymbol{x}^*, \lambda^*$ es solución del problema original y se cumple la condición LICQ en dicho punto , entonces $\boldsymbol{x}^*, \lambda^*$ resuelve el *Dual de Wolfe*.

# Ejemplo: Dual Programación Lineal

$$\min_{\boldsymbol{x}} \ \boldsymbol{c}^T \boldsymbol{x}$$
$$A\boldsymbol{x} - \boldsymbol{b} \succeq \mathbf{0}$$

# Ejemplo: Dual Programación Lineal

Lagrangiano

$$\begin{aligned}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) &= \boldsymbol{c}^T\boldsymbol{x} - \lambda^T(A\boldsymbol{x}-\boldsymbol{b}) \\ &= (\boldsymbol{c}-A^T\lambda)^T\boldsymbol{x} + \boldsymbol{\lambda}^T\boldsymbol{b}\end{aligned}$$

Si $\boldsymbol{c} - A^T\boldsymbol{\lambda} \neq \mathbf{0}$ el infimo es $-\infty$, por ejemplo, $\boldsymbol{x} = -\alpha(\boldsymbol{c} - A^T\boldsymbol{\lambda})$ con $\alpha \to \infty$. Teniendo en cuenta que el dominio de la función dual no considera este caso, entonces se tiene $\boldsymbol{c} - A^T\boldsymbol{\lambda} = \mathbf{0}$, luego el problema dual es

$$\max \quad \boldsymbol{\lambda}^T\boldsymbol{b}, s.t.:\ \boldsymbol{\lambda} \succeq \mathbf{0},\ \boldsymbol{c} - A^T\boldsymbol{\lambda} = \mathbf{0}$$

donde ademas, se incluyó la restricción $\boldsymbol{c} - A^T\boldsymbol{\lambda} = \mathbf{0}$

# Ejemplo: Dual Programación Lineal

El dual de Wolfe es

$$\begin{aligned} & \text{max} \quad \boldsymbol{\lambda}^T \boldsymbol{b} + (\boldsymbol{c} - A^T \boldsymbol{\lambda})^T \boldsymbol{x}, \\ s.t.: \quad & \boldsymbol{c} - A^T \boldsymbol{\lambda} = \boldsymbol{0}; \\ & \boldsymbol{\lambda} \succeq \boldsymbol{0} \end{aligned}$$

En algunos problemas el dual (ie, el Lagrangiano dual), dependiendo de la matriz $A$, podria ser mas facil de resolver que el problema original.
Ver Ejemplos en : pagina 96, capitulo 4, Karush - Kuhn-Tucker conditions and duality, en el libro Peter Zornig, Peter Zeornig-Nonlinear Programming An Introduction-Walter de Gruyter (2014).pdf