

Clases 7-10: Dualidad

Juan Bazerque

23 de septiembre de 2019 *

Índice

1. Introducción	2
2. Procedimiento para obtener el problema dual	2
3. Ejemplos	3
4. Convexidad del problema dual	4
5. Dualidad débil	5
6. Dualidad fuerte	6
6.1. Problema convexo sin restricciones de igualdad	6
6.2. Problema convexo con restricciones de igualdad	6
6.3. Resolución de problemas a través de su dual	7
6.4. Óptimo como punto silla del Lagrangeano	8
6.5. Complementary slackness	9
6.6. Condiciones de Karush Kuhn Tucker	10
7. Multiplicadores interpretados como precios	11
8. Sensibilidad	12
8.0.1. Ejemplo: Flujo de carga óptimo	13
9. Algoritmo primal-dual	15
9.1. Subgradiente de la función dual	15
9.2. Ascenso por subgradiente	16
9.3. Condiciones del paso para asegurar convergencia	16
9.4. Algoritmo primal-dual	17
10. Ejemplo de clasificación	17

*Corregidas al 12/10/23.

1. Introducción

La siguiente teoría permite analizar un problema de optimización a partir de su problema dual. Este análisis permite ganar intuición y conectar distintos métodos de optimización. Entre otros motivos para estudiar dualidad encontramos que:

- El problema dual es siempre **convexo**, incluso si el primario no lo es.
- **Dualidad débil:** el problema dual da una cota inferior que es útil en el análisis del problema original, en particular cuando éste no es convexo.
- **Dualidad fuerte:** si se cumple dualidad fuerte, entonces de resolver el problema dual resultan los multiplicadores de Lagrange óptimos μ^* y λ^* .
- El dual puede ser **más sencillo**, como en el caso de que las variables del primario sean de alta dimensión, pero que haya pocas restricciones.
- **Sensibilidad:** μ y λ nos dan una idea de precio por imponer restricciones.
- Aparece una nueva clase de **algoritmos duales** que optimizan el problema primario subiendo por el dual.

A su vez hay líneas de investigación recientes en diversas áreas incluyendo ingeniería de procesos [2], sistemas de energía eléctrica [1], y comunicaciones [3], en que se demuestra dualidad fuerte en problemas no convexos particulares, o se intenta reducir el gap de dualidad.

En la siguiente sección empezaremos definiendo el problema dual y observando algunos ejemplos.

2. Procedimiento para obtener el problema dual

Sea una problema en su forma canónica (no necesariamente convexo) con M restricciones de desigualdad y P restricciones de igualdad.

$$\begin{aligned} \min_{x \in X} f(x) & \tag{P} \\ \text{s.a: } g_i(x) \leq 0, \forall i = 1, \dots, M \\ h_j(x) = 0, \forall j = 1, \dots, P. \end{aligned}$$

La condición $x \in X$ permite definir restricciones adicionales que no se “dualizarán”. Un ejemplo usual es $X = \{x \geq 0\}$.

paso 1: Escribir el Lagrangeano:

$$L(x, \mu, \lambda) := f(x) + \sum_{i=1}^M \mu_i g_i(x) + \sum_{j=1}^P \lambda_j h_j(x) = f(x) + \mu^T g(x) + \lambda^T h(x).$$

paso 2: Obtener la función dual:

$$d(\mu, \lambda) := \inf_{x \in X} L(x, \mu, \lambda) = \inf_{x \in X} \{f(x) + \mu^T g(x) + \lambda^T h(x)\}.$$

paso 3: Plantear el problema dual

$$\begin{aligned} & \sup_{\mu \in \mathbb{R}^M, \lambda \in \mathbb{R}^P} d(\mu, \lambda) & (D) \\ & \text{s. a: } \mu \geq 0; \end{aligned}$$

donde la restricción $\mu \geq 0$ significa $\mu_i \geq 0$, para todo $i = 1, \dots, M$.

En la mayoría de los casos ínf y sup podrán ser substituidos por mín y máx, pero hay algunos casos que veremos en los ejemplos a continuación en que los extremos no son alcanzados, por lo que se escribe ínf y sup para dar más generalidad y formalidad a la teoría.

3. Ejemplos

Ejemplo QP

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} \frac{1}{2} x' P_0 x + q_0' x + r_0, \quad P_0 > 0 \\ & \text{s.to : } x \geq 0 \\ & Ax = b \end{aligned}$$

Lagrangeano:

$$\begin{aligned} L(x, \mu, \lambda) &= \frac{1}{2} x' P_0 x + q_0' x - \mu' x + \lambda'(Ax - b) \\ &= \frac{1}{2} x' P_0 x + (q_0 + A' \lambda - \mu)' x - b' \lambda \end{aligned}$$

Función dual: Dado que $x^* = -P_0^{-1}(q_0 - \mu + A' \lambda)$, entonces

$$d(\lambda, \mu) = -\frac{1}{2} (q_0 - \mu + A' \lambda)' P_0^{-1} (q_0 - \mu + A' \lambda) + r_0 - b' \lambda$$

Problema dual:

$$\begin{aligned} & \max_{\lambda, \mu} -\frac{1}{2} (q_0 - \mu + A' \lambda)' P_0^{-1} (q_0 - \mu + A' \lambda) + r_0 - b' \lambda \\ & \text{s. to : } \mu \geq 0 \end{aligned}$$

Ejemplo LP

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} c^T x \\ & \text{s.a : } Ax \leq b \\ & Cx = d \end{aligned}$$

Lagrangeano

$$\begin{aligned} L(x, \mu, \lambda) &= c'x + \mu'(Ax - b) + \lambda'(Cx - d) \\ &= (c + A'\mu + C'\lambda)'x - b'\mu - d'\lambda \end{aligned}$$

Función dual

$$d(\mu, \lambda) = \begin{cases} -b'\mu - d'\lambda & \text{si } A'\mu + C'\lambda = -c \\ -\infty & \text{sino} \end{cases}$$

Problema dual

$$\begin{aligned} &\underset{\lambda, \mu}{\text{máx}} -b'\mu - d'\lambda \\ &\text{s. to : } \mu \geq 0 \\ &A'\mu + C'\lambda = -c \end{aligned}$$

Nota: No es conceptualmente correcto anular la derivada $\frac{dL}{dx} = 0$ para minimizar el Lagrangeano. Si bien se obtiene la misma condición $A'\mu + C'\lambda = -c$, el ínfimo del Lagrangeano existe incluso cuando $A'\mu + C'\lambda \neq -c$.

Ejemplo QCQP

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{mín}} x'P_0x + 2q_0'x + r_0, \quad P_0 > 0 \\ &\text{s.to : } x'P_i x + 2q_i'x + r_i \leq 0, \quad P_i \geq 0, \quad i = 1, \dots, M \end{aligned}$$

Lagrangeano

$$\begin{aligned} L(x, \mu) &= \underset{x \in \mathbb{R}^n}{\text{mín}} \frac{1}{2} x' \left(P_0 + \sum_{i=1}^M P_i \mu_i \right) x + \left(q_0 + \sum_{i=1}^M q_i \mu_i \right)' x + r_0 + \sum_{i=1}^M r_i \mu_i \\ &= x'P(\mu)x + q'(\mu)x + r(\mu) \end{aligned}$$

Dual

$$\underset{\mu \geq 0}{\text{máx}} -\frac{1}{2} q'(\mu) P^{-1}(\mu) q(\mu) + r(\mu)$$

Se prueba en el apéndice A que es equivalente a un SDP, por ende convexo.

4. Convexidad del problema dual

Para probar que la función dual es cóncava, recordamos el siguiente resultado: Sea una familia de funciones afines de variable z , y paramétricas en u : $\{l_u(z)\}_{u \in \mathcal{U}}$. Se cumple que el ínfimo en $u \in \mathcal{U}$ es una función cóncava. Es decir:

$$d(z) := \inf_{u \in \mathcal{U}} \{l_u(z)\},$$

es cóncava (la prueba de este resultado se encuentra en el Apéndice B).

Volviendo al problema dual, considérese un valor fijo para x , y considérese el Lagrangeano como función de λ y μ , paramétrico en x :

$$l_x(\lambda, \mu) = f(x) + \sum_{i=1}^M \mu_i g_i(x) + \sum_{j=1}^P \lambda_j h_j(x).$$

La notación $l_x(\lambda, \mu)$ para el Lagrangeano se eligió para emular la notación del resultado anterior, con $u = x$ y $z = (\lambda, \mu)$. Fijando x , la función $l_x(\lambda, \mu)$ es afín respecto a las variables λ y μ , y por ende $d(\lambda, \mu)$ es cóncava.

Hemos probado que la función dual $d(\lambda, \mu) := \inf_{x \in X} l_x(\lambda, \mu)$ es una función cóncava, luego el problema dual es convexo.

5. Dualidad débil

Probaremos que para todo punto factible x que cumpla $g_i(x) \leq 0$ y $h_j(x) = 0$, y para variables duales μ, λ con $\mu \geq 0$, se cumple:

$$f(x) \geq d(\mu, \lambda) \tag{1}$$

Prueba: De los signos de $g_i(x)$, μ_i y $h_j(x)$ se desprende que $\lambda_j h_j(x) = 0$ y $\mu_i g_i(x) \leq 0$, por lo que

$$\begin{aligned} f(x) &\geq f(x) + \sum_{i=1}^M \mu_i g_i(x) + \sum_{j=1}^P \lambda_j h_j(x) \\ &\geq \inf_{x \in X} \left\{ f(x) + \sum_{i=1}^M \mu_i^* g_i(x) + \sum_{j=1}^P \lambda_j h_j(x) \right\} = d(\mu, \lambda). \end{aligned}$$

Nótese que (1) no requiere convexidad del problema primario, sino que es una desigualdad válida para cualquier problema de optimización. Como se cumple para cualesquiera variables primarias y duales factibles, entonces se cumple para los extremos de las funciones $f(x)$ y $d(\mu, \lambda)$

$$f^* = \inf_{\substack{x \in X \\ g_i(x) \leq 0 \\ h_j(x) = 0}} f(x) \geq \sup_{\mu \geq 0} d(\mu, \lambda) = d^*. \tag{2}$$

A partir de la desigualdad en (2) se define el “gap” de dualidad

$$\text{gap} := f^* - d^* = \inf_{\substack{x \in X \\ g_i(x) \leq 0 \\ h_j(x) = 0}} f(x) - \sup_{\mu \geq 0} d(\mu, \lambda) \geq 0 \tag{3}$$

Si además el ínfimo y supremo en (1), (2), y (3) se alcanzan entonces las soluciones óptimas x^* y (λ^*, μ^*) para (P) y (D) cumplen

$$d^* = d(\mu^*, \lambda^*) \leq f(x^*) = f^*; \tag{4}$$

y el gap de dualidad se puede escribir como:

$$\text{gap} = f(x^*) - d(\mu^*, \lambda^*) = \min_{\substack{x \in X \\ g_i(x) \leq 0 \\ h_j(x) = 0}} f(x) - \max_{\mu \geq 0} d(\mu, \lambda) \geq 0. \quad (5)$$

6. Dualidad fuerte

Las desigualdades (2)-(5) son útiles en sí mismas para acotar el valor óptimo de (P). En algunos casos, estas desigualdades se convierten en igualdades; por lo que el gap de dualidad se anula, y los problemas se vuelven equivalentes, lo que permite resolver (P) a partir de la solución de (D). Por ser éste un resultado más fuerte, se conoce como dualidad fuerte.

En particular, cuando (P) es convexo, **y bajo ciertas condiciones adicionales**, podemos garantizar que se cumple dualidad fuerte. También existen problemas que no son convexos, y que cumplen la dualidad fuerte. Por ejemplo, en [3] y [1] prueban que se cumple dualidad fuerte para algunos problemas concretos en matemática e ingeniería, aunque estos no sean convexos.

6.1. Problema convexo sin restricciones de igualdad

Para analizar el caso convexo, consideremos primero un problema sin restricciones de igualdad. En este caso se tiene el siguiente resultado.

Teorema 1. [4, p.589] *Sea X convexo, y $f(x)$ y $g_i(x)$ convexas, para todo $i = 1, \dots, M$. Consideremos el siguiente problema convexo:*

$$\begin{aligned} & \min_{x \in X} f(x) \\ & \text{s.a.} : g_i(x) \leq 0, \quad \forall i = 1, \dots, M; \end{aligned}$$

Si $\exists \bar{x} \in X$ tal que $g_i(\bar{x}) < 0, \forall i$, entonces:

$$d^* = \sup_{\mu \geq 0} d(\mu) \geq \inf_{\substack{x \in X \\ g_i(x) \leq 0}} f(x) = f^*.$$

Siempre se cumple que $d^* \leq f^*$ por dualidad débil, por lo que el teorema anterior implica la dualidad fuerte: $f^* = d^*$. La prueba, que se encuentra en el Apéndice C, es valiosa no solo por el resultado sino también por la intuición geométrica que contiene. La hipótesis $g_i(\bar{x}) < 0$ se conoce como condición de Slater. Significa que existe un punto en el interior del conjunto definido por las funciones g_i , y es una condición técnica para asegurar que los multiplicadores sean finitos.

6.2. Problema convexo con restricciones de igualdad

El resultado anterior se puede generalizar a un problema convexo con restricciones de igualdad y desigualdad. Para que las restricciones de igualdad sean convexas, estas deben ser restricciones lineales (afines), por lo que el problema a analizar es:

$$\begin{aligned} & \min_{x \in X} f(x) \\ & \text{s.a.} : g_i(x) \leq 0, \quad \forall i = 1, \dots, M \\ & \quad Ax - b = 0 \\ & \quad Cx - d \leq 0; \end{aligned}$$

con $f(x)$ y $g_i(x)$ convexas, y X convexo. Nuevamente, en este caso podemos garantizar dualidad fuerte si se cumple Slater. En este caso es suficiente con que se cumpla la condición de Slater solamente para las restricciones no lineales; mientras que las restricciones lineales admiten hipótesis más débiles, como se demuestra en [4, p.592]. Una prueba más directa puede encontrarse en [5, p.234] asumiendo que A es de rango completo (no hay restricciones redundantes), y sin distinguir las restricciones de desigualdad lineal.

6.3. Resolución de problemas a través de su dual

Supongamos que se cumple dualidad fuerte: $f^* = d^*$, y que (μ^*, λ^*) es una solución del problema dual. Entonces:

$$x^* \in \arg \min_{\substack{x \in X \\ g_i(x) \leq 0 \\ h_j(x) = 0}} f(x) \Rightarrow x^* \in \arg \min_{x \in X} L(x, \mu^*, \lambda^*). \quad (6)$$

Es decir: si se obtiene una solución (μ^*, λ^*) del problema dual, entonces cualquier solución x^* del problema primario, debe minimizar al Lagrangeano asociado a los multiplicadores (μ^*, λ^*) . Para probar (6), observe que si hay dualidad fuerte, entonces

$$\begin{aligned} f(x^*) &= d(\mu^*, \lambda^*) \\ &= \inf_{x \in X} L(x, \mu^*, \lambda^*) \\ &\leq L(x^*, \mu^*, \lambda^*) \\ &= f(x^*) + \sum_{i=1}^M \mu_i^* g_i(x^*) + \sum_{j=1}^P \lambda_j^* h_j(x^*) \\ &\leq f(x^*) \end{aligned}$$

donde se utilizó que $\mu_i^* \geq 0$, $g_i(x^*) \leq 0$ y $h_j(x^*) = 0$, y se concluye que todas las desigualdades se deben cumplir con igualdad. En particular, esto implica que el ínfimo del Lagrangeano se alcanza en x^* , como queríamos probar:

$$\inf_{x \in X} L(x, \mu^*, \lambda^*) = L(x^*, \mu^*, \lambda^*).$$

Observación 1. *El lagrangeano $L(x, \mu^*, \lambda^*)$, asociado a (μ^*, λ^*) , podría tener otras soluciones $\hat{x} \in X$ que no sean solución del problema primal.*

Ejemplo Considérese el siguiente problema convexo en la variable escalar x

$$\begin{aligned} &\min_{x \in \mathbb{R}} x^2 + 1 \\ &s.a : x \leq -1 \end{aligned}$$

El Lagrangeano correspondiente tiene la forma:

$$L(x, \mu) = x^2 + 1 + \mu(x + 1)$$

con mínimo $x(\mu)$ paramétrico en lambda:

$$x(\mu) = \arg \min_{x \in \mathbb{R}} x^2 + 1 + \mu(x + 1) = -\mu/2$$

Luego, insertando $x(\mu)$ en el Lagrangeano resulta la función dual:

$$d(\mu) = x^2(\mu) + 1 + \mu(x(\mu) + 1) = -\frac{1}{4}\mu^2 + \mu + 1.$$

y el problema dual resulta

$$\max_{\mu \geq 0} -\frac{1}{4}\mu^2 + \mu + 1$$

Este alcanza máximo en $\mu^* = 2$.

Como el problema es convexo, y además se cumple Slater, podemos afirmar que se cumple dualidad fuerte: $f^* = d^* = d(\mu^*) = 2$. Por otro lado, el lagrangeano asociado a la solución del dual $L(x, \mu^*)$, alcanza mínimo en un único punto: $x(\mu^*) = -1$. Por lo tanto, se deduce que el problema primal alcanza mínimo en $x^* = x(\mu^*) = -1$.

6.4. Óptimo como punto silla del Lagrangeano

Estos resultados pueden verse geoméricamente como sigue. En la Fig. 1(a) se ve el Lagrangeano como función de dos variables x y μ . El punto (x^*, μ^*) es un punto silla en que se minimiza el Lagrangeano y maximiza el dual. Se detalla en la Fig. 1(b) que el dual es cóncavo como función de μ . En la Fig. 1(c) se muestra que no hay duality gap, y que tanto el Lagrangeano como el primario se minimizan para x^* .

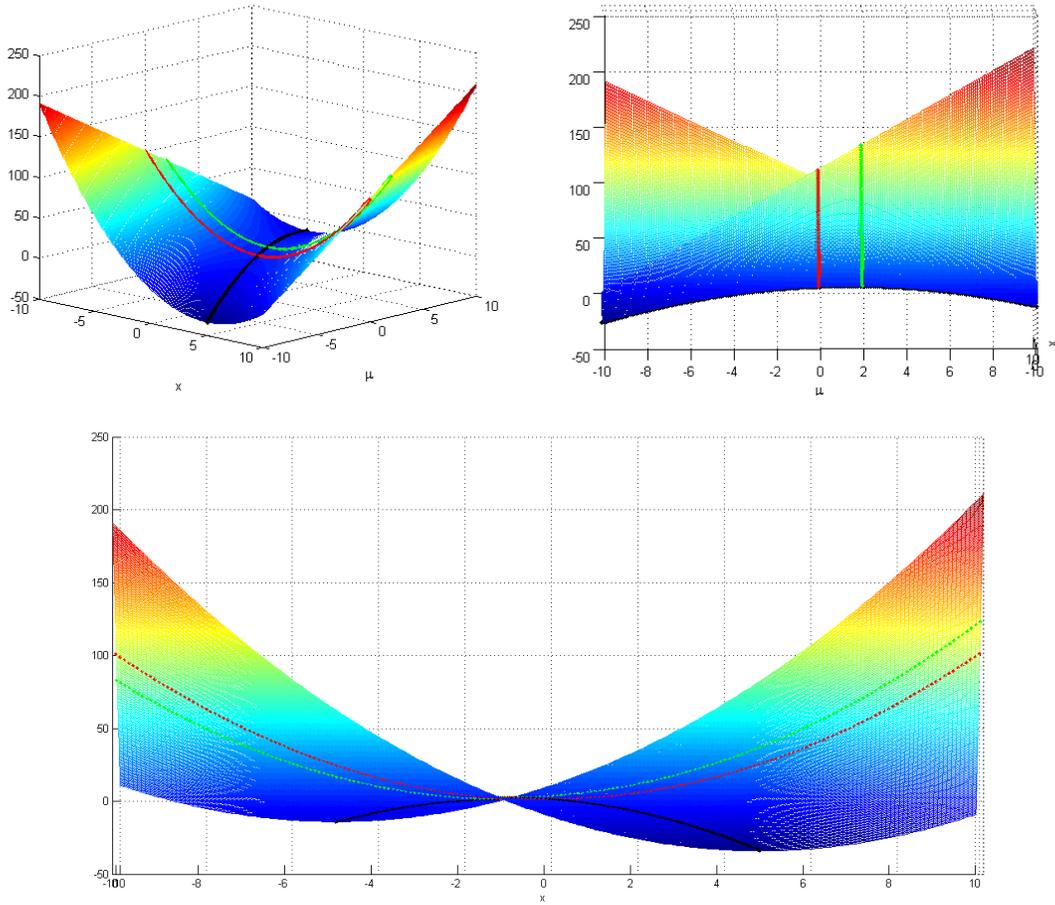


Figura 1: Óptimo (x^*, μ^*) como punto silla del Lagrangeano sin duality gap.

6.5. Complementary slackness

De la prueba anterior, podemos ver que si se cumple dualidad fuerte, se tiene:

$$\sum_{i=1}^M \mu_i^* g_i(x^*) + \sum_{j=1}^P \lambda_j^* h_j(x^*) = 0.$$

Como además $h_j(x^*) = 0$, y $\mu_i^* g_i(x^*) \leq 0$, esto implica:

$$\mu_i^* g_i(x^*) = 0, \quad \forall i.$$

Estas últimas igualdades se conocen como condiciones de *Complementary Slackness*. Bajo estas condiciones, si x^* es solución del primal, y μ^* es solución del dual, se cumple:

1. si el multiplicador $\mu_i^* > 0$, entonces la restricción debe estar activa: $g_i(x^*) = 0$, y
2. si la restricción no está activa $g_i(x^*) < 0$, entonces el multiplicador debe anularse: $\mu_i^* = 0$.

6.6. Condiciones de Karush Kuhn Tucker

Bajo la hipótesis de dualidad fuerte, hemos establecido que para que una terna (x^*, μ^*, λ^*) sea óptima (del dual y del primal), es necesario que x^* minimice el Lagrangeano asociado a (μ^*, λ^*) , y que se cumpla complementary slackness.

Esta condición es también suficiente. Es decir que si por algún medio se obtiene una terna candidata $(x^\dagger, \mu^\dagger, \lambda^\dagger)$, tal que x^\dagger minimiza el Lagrangeano $L(x, \mu^\dagger, \lambda^\dagger)$, y además cumple:

$$g(x^\dagger) \leq 0, \quad h(x^\dagger) = 0, \quad \mu^\dagger \geq 0, \quad \mu_i^\dagger g_i(x^\dagger) = 0, \quad \forall i = 1, \dots, M,$$

entonces se cumple dualidad fuerte, y x^\dagger es solución del problema primario. En efecto, bajo las hipótesis anteriores, se tiene:

$$\begin{aligned} f(x^\dagger) &\geq f^* && \text{definición de ínfimo, con } x^\dagger \in X \\ &\geq d^* && \text{dualidad débil} \\ &\geq d(\mu^\dagger, \lambda^\dagger) && \text{definición de supremo, con } \mu^\dagger \geq 0 \\ &= \inf_{x \in X} L(x, \mu^\dagger, \lambda^\dagger) && \text{definición de función dual} \\ &= L(x^\dagger, \mu^\dagger, \lambda^\dagger) && x^\dagger \text{ minimiza el Lagrangeano} \\ &= f(x^\dagger) + \sum_{i=1}^M \mu_i^\dagger g_i(x^\dagger) + \sum_{j=1}^P \lambda_j^\dagger h_j(x^\dagger) && \text{definición de Lagrangeano} \\ &= f(x^\dagger) && h(x^\dagger) = 0 \text{ y complementary slackness} \end{aligned}$$

Por lo tanto, las desigualdades deben ser igualdades. En particular: $f(x^\dagger) = f^* = d^* = d(\mu^\dagger)$.

Observación 2. Si el problema es convexo y el conjunto X representa todo el espacio, entonces el Lagrangeano es convexo (para $\mu \geq 0$), y su mínimo se obtiene anulando el gradiente.

Ejemplo: Resolver KKT directamente sin hallar el problema dual.

Volvamos al ejemplo anterior: $\min_{x \in \mathbb{R}} x^2 + 1$, sujeto a: $x \leq -1$. Su Lagrangeano es:

$$L(x, \mu) = x^2 + 1 + \mu(x + 1).$$

Como el problema es convexo, y se cumple la condición de Slater, sabemos que cualquier par óptimo (x^*, μ^*) debe cumplir complementary slackness. Es decir, se debe dar uno de los tres siguientes casos:

- $x + 1 < 0 \rightarrow \mu = 0$
- $\mu > 0 \rightarrow x + 1 = 0$
- $x + 1 = 0$ y $\mu = 0$

En el primer caso, la derivada del Lagrangeano igualada a cero nos da la ecuación $2x = 0$, que corresponde al caso en que la restricción no está activa. En el segundo y tercer caso, la restricción está activa y se convierte en una ecuación extra $x + 1 = 0$, que en este caso resuelve el problema.

Si se tiene más de una restricción, entonces se deben considerar estos tres casos para cada una de ellas (ver ejemplo en la próxima sección). Esto genera un número combinatorio de casos a considerar, lo que se puede hacer en problemas con pocas restricciones, o en aquellos en que el conocimiento del fenómeno modelado nos ayude a eliminar casos. Sino este abordaje no es práctico y se opta por resoluciones numéricas, por ejemplo, con algoritmos de punto interior.

7. Multiplicadores interpretados como precios

En esta sección estudiaremos el óptimo social de un problema de teoría microeconómica, lo cual nos llevará a pensar los multiplicadores como precios. Esta intuición se consolida más adelante con el teorema de sensibilidad. Los dos ejemplos de esta sección fueron presentados en el curso *Optimización convexa aplicada al flujo de carga óptimo* dictado en UTE en 2016 por el Prof. Fernando Paganini, a quién agradecemos su autorización para incluir estos ejemplos aquí.

Ejemplo: mínimo de un problema económico

Considérese un conjunto de N firmas que producen un bien a consumir por un conjunto de M consumidores. Cada firma genera una cantidad x_i del bien incurriendo en un costo $c_i(x_i)$, $i = 1, \dots, N$. A su vez cada consumidor $j = 1, \dots, M$ obtendrá una cantidad y_j que les generará una utilidad $u_j(y_j)$. Se supone que los costos son estrictamente convexos y las utilidades estrictamente cóncavas, y se adopta un óptimo social para decidir las cantidades x_i y y_j , esto es

$$\min_{x \in \mathbb{R}^N, y \in \mathbb{R}^M} \sum_{i=1}^N c_i(x_i) - \sum_{j=1}^M u_j(y_j) \quad (7)$$

$$s.to : \sum_{j=1}^M y_j - \sum_{i=1}^N x_i = 0 \quad (8)$$

donde la restricción implica una balance en el mercado.

El Lagrangeano para este problema es

$$L(x, y, \lambda) = \sum_{i=1}^N c_i(x_i) - \sum_{j=1}^M u_j(y_j) + \lambda \left(\sum_{i=1}^N x_i - \sum_{j=1}^M y_j \right) \quad (9)$$

$$= \sum_{i=1}^N (c_i(x_i) - \lambda x_i) - \sum_{j=1}^M (u_j(y_j) - \lambda y_j). \quad (10)$$

Es fundamental observar que al escribir el Lagrangeano el problema desacopla por firma y consumidor, es decir que para un valor fijo de $\lambda \in \mathbb{R}$, las firmas y consumidores pueden hallar los valores x_i e y_j que minimizan el Lagrangeano separadamente según

$$x_i(\lambda) = \arg \max_{x_i} \lambda x_i - c_i(x_i) \quad (11)$$

$$y_j(\lambda) = \arg \max_{y_j} u_j(y_j) - \lambda y_j \quad (12)$$

donde se maximizó en vez de minimizar cambiando a su vez el signo.

Para un λ dado $x_i(\lambda)$ e $y_j(\lambda)$ son los que anulan la derivada, satisfaciendo las ecuaciones

$$\frac{d}{dx_i} c_i(x_i(\lambda)) = \lambda \quad (13)$$

$$\frac{d}{dy_j} u_j(y_j(\lambda)) = \lambda \quad (14)$$

El multiplicador se interpreta entonces como el precio del bien. Esto es, si el bien se intercambiara en el mercado a precio λ , entonces las firmas decidirían su nivel de producción igualando su costo

marginal al precio. Como se observa en la figura 2 producir una unidad infinitesimal extra del bien le costaría a la firma lo mismo que obtendría por venderlo en el mercado, lo que fija el punto de equilibrio donde parar la producción (a precio λ). En forma similar el consumidor ajusta su consumo igualando su utilidad marginal al precio. Interpretando λ como un precio (11) y (12) significan que las firmas y consumidores maximizan sus excedentes, como también se nota en la figura 2 donde $s_i(\lambda) = \max_{x_i} \lambda x_i - c_i(x_i)$ y $s_j(\lambda) = \max_{y_j} u_j(y_j) - \lambda y_j$.

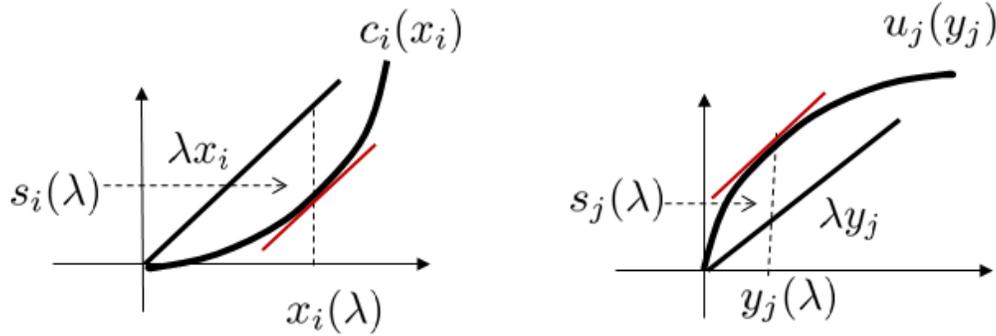


Figura 2: Costos y utilidades marginales se igualan al precio obteniéndose los máximos excedentes.

Existe un precio óptimo λ^* que balancea el mercado haciendo que se cumpla la restricción con igualdad. Se prueba en el apéndice D que este precio maximiza el dual. El efecto alrededor del óptimo λ^* puede intuirse observando la figura 2. Para $\lambda < \lambda^*$ disminuye la pendiente de las rectas λx_i y λy_j y entonces las firmas y consumidores bajan su producción y suben su consumo para adaptar sus costos y utilidades marginales al nuevo precio. Esto lleva a que $\sum_{j=1}^M y_j(\lambda) > \sum_{i=1}^N x_i(\lambda)$ y el mercado no cierra. Por el contrario si $\lambda > \lambda^*$ aumenta la producción y baja el consumo resultando luego $\sum_{j=1}^M y_j(\lambda) < \sum_{i=1}^N x_i(\lambda)$. Si bien en este caso se satisface la demanda (podría haberse admitido una desigualdad en la restricción al plantear (8)), estas cantidades no resultan óptimas pues puede bajarse el costo social manteniendo el consumo y bajando la producción.

8. Sensibilidad

Los multiplicadores dan una noción de cuánto cuesta imponer una restricción, en el sentido de cuánto lograríamos bajar el costo $f(x^*)$ si nos permitimos relajar la restricción asociada. Más formalmente, definamos el problema [perturbado](#) y su costo mínimo

$$\begin{aligned}
 p^*(u, v) &:= \min_{x \in X} f(x) \\
 \text{s.a.} &: g_i(x) \leq u_i, \quad i = 1, \dots, M \\
 &: h_j(x) = v_j, \quad j = 1, \dots, P
 \end{aligned}$$

donde las restricciones $g_i(x) \leq 0$ y $h_j(x) = 0$ fueron substituidas por $g_i(x) \leq u_i$ y $h_j(x) = v_j$ respectivamente. Observe que cuando $(u, v) = (0, 0)$ se recupera el problema (P) original. Nos interesa conocer cómo varía $p^*(u, v)$ respecto a u y v , para lo que se tiene el siguiente resultado.

Teorema 2. *Supongamos que el problema sin perturbar (P) cumple dualidad fuerte, y que su problema dual alcanza máximo en (μ^*, λ^*) . El valor mínimo del problema primal perturbado es*

una función diferenciable de las perturbaciones u y v , y sus derivadas parciales vienen dadas por:

$$\frac{d}{du_i} p^*(0, 0) = -\mu_i^*, \quad \frac{d}{dv_j} p^*(0, 0) = -\lambda_j^*.$$

La prueba será parcial, saltando la demostración de que $p^*(u, v)$ es diferenciable y poniendo el foco en hallar los valores de las derivadas. Considérese la función dual para el problema [perturbado](#):

$$d_{u,v}(\mu, \lambda) := \inf_{x \in X} f(x) + \mu'(g(x) - u) + \lambda'(h(x) - v) = d(\mu, \lambda) - \mu'u - \lambda'v$$

Luego:

$$\begin{aligned} p^*(u, v) &\geq d_{u,v}(\mu_{u,v}^*, \lambda_{u,v}^*) && \text{dualidad débil} \\ &\geq d_{u,v}(\mu^*, \lambda^*) && \text{definición de supremo, con } \mu^* \geq 0 \\ &= d(\mu^*, \lambda^*) - (\mu^*)'u - (\lambda^*)'v && \text{definición de } d_{u,v} \\ &= p^*(0, 0) - (\mu^*)'u - (\lambda^*)'v && \text{dualidad fuerte} \end{aligned}$$

Entonces para $u = (0, \dots, u_i, \dots, 0) = u_i e_i$ con $u_i > 0$, se tiene:

$$\frac{p^*(u, 0) - p^*(0, 0)}{u_i} \geq -\mu_i^*$$

y para $u_i < 0$:

$$\frac{p^*(u, 0) - p^*(0, 0)}{u_i} \leq -\mu_i^*.$$

Luego, si existe la derivada, entonces los límites por izquierda y derecha deben coincidir, por lo que se deduce que:

$$\frac{d}{du_i} p^*(0, 0) = -\mu_i^*.$$

Análogamente se prueba que:

$$\frac{d}{dv_j} p^*(0, 0) = -\lambda_j^*.$$

8.0.1. Ejemplo: Flujo de carga óptimo

Este ejemplo analiza el precio de agregar restricciones. Considere el ejemplo de una red de tres barras de la Figura 3, con dos generadores y dos cargas.

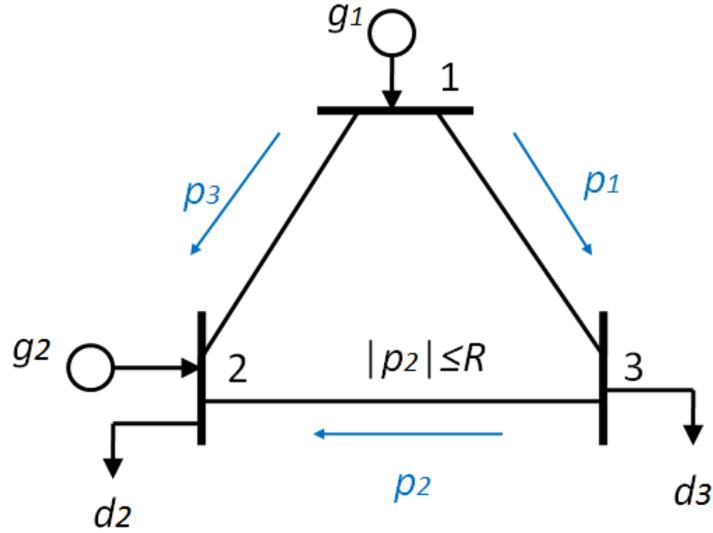


Figura 3: Red eléctrica de tres barras con dos generadores y dos cargas

Se asume que los generadores tienen costos lineales de generación $c_1(g_1) = \alpha_1 g_1$ y $c_2(g_2) = \alpha_2 g_2$ con $2\alpha_1 < \alpha_2$, donde α_1 y α_2 se miden en $\$/MWh$ (puede resultar más natural expresar los precios por energía (MWh) normalizando a un intervalo de tiempo arbitrario)

El problema de optimización que modela esta red es:

$$\min_{(p_1, p_2, p_3, g_1, g_2)} \alpha_1 g_1 + \alpha_2 g_2 \quad (15)$$

$$\text{sujeto a: } p_1 + p_3 = g_1 \quad (16)$$

$$g_2 + p_3 + p_2 = d_2 \quad (17)$$

$$p_1 - p_2 = d_3 \quad (18)$$

$$p_3 - p_1 - p_2 = 0 \quad (19)$$

$$|p_2| \leq R \quad (20)$$

$$g_1 \geq 0 \quad (21)$$

$$g_2 \geq 0 \quad (22)$$

Las ecuaciones (16), (17) y (18) corresponden a los balances en barras 1, 2, y 3 respectivamente. Por su lado (19) modela las potencias como diferencias de potencial según es inherente a las redes eléctricas. Finalmente (20) modela la restricción de máxima potencia, y se agregan restricciones de no negatividad para las potencias generadas.

Eliminando la variable p_3 y la restricción (19) el problema anterior a

$$\min_{(p_1, p_2, g_1, g_2)} \alpha_1 g_1 + \alpha_2 g_2 \quad (23)$$

$$\text{sujeto a: } g_1 - 2p_1 - p_2 = 0 \quad (\lambda_1) \quad (24)$$

$$g_2 + p_1 + 2p_2 - d_2 = 0 \quad (\lambda_2) \quad (25)$$

$$p_1 - p_2 - d_3 = 0 \quad (\lambda_3) \quad (26)$$

$$g_1 \geq 0 \quad (\mu_1) \quad (27)$$

$$g_2 \geq 0 \quad (\mu_2) \quad (28)$$

$$p_2 \leq R \quad (\mu_3) \quad (29)$$

donde (20) se convirtió en (29) eliminando la restricción $-R \leq p_2$ por asumirse con argumentos físicos que ésta no estará activa en el óptimo. Podría conservarse $-R \leq p_2$ y se comprobaría luego que ésta no se activa.

El óptimo de (23) se obtiene en el apéndice E para el caso $d_2 - d_3 > 3R$ y viene dado por $p_1^* = d_3 + R$, $p_2^* = R$, $g_1^* = 2d_3 + R$ y $g_2^* = d_2 - d_3 - 3R$. Los multiplicadores óptimos para las restricciones (24), (25), (26), y (29) son $\lambda_1^* = -\alpha_1$, $\lambda_2^* = -\alpha_2$, $\lambda_3^* = \alpha_2 - 2\alpha_1$ y $\mu_3 = 3(\alpha_2 - \alpha_1)$.

En particular se observa que el multiplicador $\lambda_3 = \alpha_2 - 2\alpha_1$ asociado al balance de energía en la barra 3 es positivo. Esto significa, siguiendo el teorema de sensibilidad, que la derivada del costo respecto a una relajación de la restricción es negativa. Es decir que si en vez de $p_1 - p_2 = d_3$ se admite $p_1 - p_2 = d_3 + u$ entonces el costo baja, lo que equivale a una mayor demanda en la barra 3. Es decir que si aumenta la demanda en la barra 3 el costo global disminuye. La intuición de este fenómeno es que aumentando la demanda en la barra 3 se puede transmitir más energía desde el generador en la barra 1 sin violar la restricción $p_2 \leq R$. Se concluye que la energía entregada en la barra 3 tiene un precio negativo, es decir que deberíamos incentivar a que se consuma más energía en esa barra, incluso regalarla o hasta pagar al cliente para que aumente su consumo. Hay varios reportes de casos reales en la literatura de flujo de potencia óptima, e.g., [6], en que se da este fenómeno de precios negativos debido a limitaciones de las redes de transmisión.

9. Algoritmo primal-dual

Considérese el siguiente problema:

$$\begin{aligned} & \min_{x \in X} f(x) & (P) \\ & \text{s.a. : } g_i(x) \leq 0, \quad i = 1, \dots, M \end{aligned}$$

Notar que esta formulación permite incluir restricciones de igualdad: $h(x) = 0$, pues estas se pueden expresar de forma equivalente como dos restricciones de desigualdad: $h(x) \leq 0$ y $h(x) \geq 0$.

9.1. Subgradiente de la función dual

Podemos calcular un subgradiente de la función dual, a partir de un mínimo del lagrangeano.

Teorema 3. *Dado μ cualquiera, el Lagrangeano que vamos a utilizar es:*

$$L(x, \mu) := f(x) + \mu^T g(x).$$

Sea x_μ un punto donde se minimiza este Lagrangeano:

$$x_\mu \in \arg \min_{x \in X} L(x, \mu).$$

Entonces $g(x_\mu)$ es subgradiente de la función dual $d(\mu)$. Es decir:

$$d(\bar{\mu}) \leq d(\mu) + (\bar{\mu} - \mu)^T g(x_\mu), \quad \forall \bar{\mu}.$$

Denotamos esto como: $\partial_\mu d = g(x_\mu)$.

Demostración.

$$\begin{aligned} d(\bar{\mu}) &= \inf_{x \in X} \{f(x) + \bar{\mu}^T g(x)\} && \text{definición de función dual} \\ &\leq f(x_\mu) + \bar{\mu}^T g(x_\mu) && \text{definición de ínfimo, con } x_\mu \in X \\ &= f(x_\mu) + \mu^T g(x_\mu) + (\bar{\mu} - \mu)^T g(x_\mu) && \text{suma y resta } \mu^T g(x_\mu) \\ &= L(x_\mu, \mu) + (\bar{\mu} - \mu)^T g(x_\mu) && \text{definición de Lagrangeano} \\ &= (\inf_{x \in X} L(x, \mu)) + (\bar{\mu} - \mu)^T g(x_\mu) && \text{definición de } x_\mu \\ &= d(\mu) + (\bar{\mu} - \mu)^T g(x_\mu), \quad \forall \bar{\mu} && \text{definición de función dual.} \end{aligned}$$

□

9.2. Ascenso por subgradiente

Utilizaremos el algoritmo de ascenso por el subgradiente con proyección (PGD) para resolver el siguiente problema dual:

$$\begin{aligned} &\underset{\mu}{\text{máx}} \quad d(\mu) && \text{(D)} \\ &s.a : \mu \geq 0 \end{aligned}$$

En este caso el algoritmo PGD toma la forma:

$$\begin{aligned} \mu_{k+1} &= \text{proy}_{\mu \geq 0}(\mu_k + \alpha_k \partial d(\mu_k)) \\ &= [\mu_k + \alpha_k \partial d(\mu_k)]_+ \\ &= [\mu_k + \alpha_k g(x_\mu)]_+ \end{aligned} \tag{30}$$

9.3. Condiciones del paso para asegurar convergencia

Teorema 4. *Supongamos que elegimos α_k tal que:*

$$0 < \alpha_k < 2 \left(\frac{d(\mu^*) - d(\mu_k)}{\|g(x_{\mu_k})\|} \right).$$

Entonces, la iteración (30) converge al óptimo del dual:

$$\lim_{k \rightarrow \infty} \|\mu_k - \mu^*\| = 0, \quad y \quad \lim_{k \rightarrow \infty} d(\mu_k) = d(\mu^*).$$

Esta forma de elegir el paso requiere conocer el valor óptimo $d(\mu^*)$, lo cual no es posible en general. Por lo tanto, es de interés la siguiente modificación del resultado anterior.

Teorema 5. [4, p.732] Si $\alpha_k \rightarrow 0$, con $\sum_{k=0}^{\infty} \alpha_k = \infty$, la iteración

$$\mu_{k+1} = \left[\mu_k + \alpha_k \frac{g(x_{\mu_k})}{\|g(x_{\mu_k})\|} \right]^+$$

converge al óptimo del dual:

$$\lim_{k \rightarrow \infty} \|\mu_k - \mu^*\| = 0, \quad \lim_{k \rightarrow \infty} d(\mu_k) = d(\mu^*).$$

Este algoritmo de ascenso por subgradiente requiere minimizar el Lagrangeano en cada paso para obtener x_{μ} . Si se necesita iterar para lograr esto, entonces se forma un algoritmo con dos ciclos anidados. Sin embargo, puede resultar ineficiente exigir la convergencia del ciclo interior, ya que el punto x_{μ} no es el mínimo buscado, sino solo un iterado intermedio en el ciclo exterior.

Esto sugiere un algoritmo más simple, en que en cada paso del PGD no se minimice el Lagrangeano, sino que solo se haga un paso de descenso por gradiente en la variable primaria x .

9.4. Algoritmo primal-dual

Sea el problema con restricciones de igualdad

$$\begin{aligned} \min_{x \in X} f(x) & \tag{P} \\ \text{s.a. : } g_i(x) \leq 0, \forall i = 1, \dots, M \end{aligned}$$

El algoritmo primal-dual asociado, es:

$$x_{k+1} = x_k - \alpha \nabla_x L(x_k, \mu_k) \tag{31}$$

$$\mu_{k+1} = \mu_k + \alpha h(x_k) \tag{32}$$

El siguiente resultado garantiza convergencia del algoritmo primal dual, bajo ciertas hipótesis de regularidad y diferenciabilidad.

Teorema 6. [4][pp. 530] Sean x^* y λ^* solución del primal y del dual, respectivamente. Supongamos que $f, h \in C^2$, $\nabla_{xx}^2 L(x^*, \lambda^*) > 0$, y x^* es regular. Entonces existen $\bar{\alpha} > 0$ y un conjunto S , tal que: si $(x_0, \lambda_0) \in S$, entonces, para todo $\alpha \in (0, \bar{\alpha}]$, la iteración (31)-(32) converge a (x^*, λ^*) .

10. Ejemplo de clasificación

Se desea ajustar la recta que discrimine linealmente un conjunto de puntos etiquetados. Esto se realizará según los métodos de Support vector Machines (SVMs). Véase [5, pp. 422].

Apéndice A: Dual de QCQP y SDP equivalente

Caso 1: QCQP convexo con $P_0 > 0$

Sea el problema convexo

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & x'P_0x + 2q'_0x + r_0, \quad P_0 > 0 \\ \text{s.to} \quad & x'P_ix + 2q'_ix + r_i \leq 0, \quad P_i \geq 0, \quad i = 1 \dots, m \end{aligned} \tag{33}$$

Su dual tiene la forma

$$\max_{\mu \geq 0} -q'(\mu)P^{-1}(\mu)q(\mu) + r(\mu)$$

Donde $P(\mu) > 0$ y por tanto invertible ya que $\mu \geq 0$. Agregando la variable mayorante t

$$\begin{aligned} \max_{\mu \geq 0, P, q, r, t} \quad & t \\ \text{s.to} \quad & q'P^{-1}q + r \geq t \\ & P = P_0 + \sum_{i=1}^m \mu_i P_i \\ & q = q_0 + \sum_{i=1}^m \mu_i q_i \\ & r = r_0 + \sum_{i=1}^m \mu_i r_i \end{aligned}$$

El costo y todas las restricciones son lineales salvo $-q'P^{-1}q + r - t \geq 0$. Para simplificar esta restricción se recurre al complemento de Schur. Sea una matriz definida en bloques

$$X = \begin{bmatrix} A & B \\ B' & C \end{bmatrix}$$

se cumple que si $A > 0$ (definida positiva), entonces $X \geq 0$ si y solo si $X|A \geq 0$ donde $X|A = C - B'A^{-1}B$ es el complemento de Schur de A en X [7, p.34]. Aplicado a $A = P$, $B = q$ y $C = r - t$ se cumple que $X \geq 0$ si y solo si $r - t - q'P^{-1}q \geq 0$. Esta última desigualdad es escalar y coincide con la que queríamos simplificar. Podemos entonces substituir $r - t - q'P^{-1}q \geq 0$ por $X \geq 0$ (positiva semidefinida) con

$$X = \begin{bmatrix} P & q \\ q' & r - t \end{bmatrix}$$

obteniendo el siguiente SDP convexo

$$\begin{aligned}
& \underset{\mu \geq 0, P, q, r, t, X}{\text{máx}} && t && (34) \\
& \text{s.to : } && X = \begin{bmatrix} P & q \\ q' & r - t \end{bmatrix} \\
& && X \geq 0 \\
& && P = P_0 + \sum_{i=1}^m \mu_i P_i \\
& && q = q_0 + \sum_{i=1}^m \mu_i q_i \\
& && r = r_0 + \sum_{i=1}^m \mu_i r_i
\end{aligned}$$

Caso 2: QCQP no convexo

Alcanza que una sola matriz P_i , $i = 0, \dots, m$ tenga un valor propio negativo, o que haya una restricción de igualdad para que el QCQP no sea convexo. Consideremos el siguiente caso con $P_i \not\geq 0$

$$\begin{aligned}
& \underset{x \in \mathbb{R}^n}{\text{mín}} && x' P_0 x + 2q_0' x + r_0 && (35) \\
& \text{s.to : } && x \in \mathbb{R}^n && x' P_i x + 2q_i' x + r_i = 0 \quad i = 1 \dots, m
\end{aligned}$$

Sin embargo el dual deberá permanecer siendo convexo, según se confirma a continuación. El Lagrangeano toma la forma

$$L(x, \lambda) = x' P(\lambda) x + 2q'(\lambda) x + r(\lambda)$$

con $P(\lambda)$, $q(\lambda)$ y $r(\lambda)$ definidas como en el caso anterior. A diferencia del caso anterior el vector λ puede tomar en principio cualquier valor en \mathbb{R}^m . Algunos de estos valores harán que $P(\lambda) \not\geq 0$ incluso si $P_i > 0$, $\forall i$. Para estos valores de λ el ínfimo del Lagrangeano es $-\infty$ lo que se obtiene haciendo x tender a infinito por la dirección del vector propio de $P(\lambda)$ con valor propio negativo.

Continuamos con $P = P(\lambda) \geq 0$. En el caso de que haya algún valor propio nulo, se debe separar en dos casos. En el caso en que $q = q(\lambda) \notin \mathcal{R}(P)$ (rango de $P = P(\lambda)$) entonces existen dos vectores u y v tales que $q = Pu + v$ y $P'v = 0$. Estos vectores pueden obtenerse resolviendo la proyección $u = \arg \min_v \|P\bar{v} - q\|^2$ y tomando $v = q - Pu$ que será no nulo dado que $q(\lambda) \notin \mathcal{R}(P(\lambda))$ y satisface $P'v = P'(q - Pu) = 0$ igualando a cero el gradiente en el problema de proyección. Luego tomando $x = -tv$ y usando que $v'P = 0$ se tiene

$$L(x, \lambda) = t^2(v'P)v - 2tv'q + r = -2tv'(Pu + v) + r = -2t\|v\|^2 + r \rightarrow -\infty$$

donde la dependencia respecto a λ está presente en P, q, r así como en u y v .

En el caso en que $q \in \mathcal{R}(P)$, entonces existe $u \in \mathbb{R}^m$ tal que $q = Pu$. Luego sumando $u'Pu -$

$u'Pu = 0$ se obtiene

$$\begin{aligned} L(x, \lambda) &= xPx - 2x'Pu + u'Pu - u'Pu + r \\ &= (x + u)'P(x + u) - u'Pu + r \geq -u'Pu + r > -\infty. \end{aligned}$$

por lo que en este caso el dual tiene un valor finito.

Si bien $P(\lambda) \geq 0$ no es invertible cuando tiene valores propios nulos, se puede definir la pseudo inversa $P^\dagger(\lambda)$ a partir de la descomposición en valores propios y considerando solo los valores propios positivos [7, p.41]. Luego el mínimo de $L(x, \lambda)$ (para $P(\lambda) \geq 0$ y $q \in \mathcal{R}(P)$) viene dado por

$$d(\lambda) = -q'(\lambda)P^\dagger(\lambda)q(\lambda) + r(\lambda) \quad (36)$$

Finalmente, si $P(\lambda) > 0$ entonces sigue valiendo (36), y se reduce al caso anterior ya que $P^\dagger(\lambda) = P^{-1}(\lambda)$ cuando $P(\lambda) > 0$.

Luego el conjunto de valores en que el ínfimo del Lagrangeano no es $-\infty$ es $\Lambda = \{\lambda : P(\lambda) \geq 0, q(\lambda) \in \mathcal{R}(P(\lambda))\}$, y el problema dual toma la forma

$$\begin{aligned} &\text{máx}_{\lambda \in \Lambda, P, q, r} t \\ &\text{s. to: } q'P^\dagger q + r - t \geq 0 \end{aligned} \quad (37)$$

Puede generalizarse el complemento de Schur en función de la pseudo inversa para matrices $A \geq 0$ con valores propios nulos. El complemento de Schur generalizado de X en A es $X|A = C - B'A^\dagger B$ y cumple que si $A \geq 0$ entonces $X \geq 0$ si y solo si $X|A = C - B'A^\dagger B$ y $B \in \mathcal{R}(A)$ [7, p.44]. Este resultado es exactamente lo que se necesita para reformular (37) como el siguiente SDP convexo

$$\begin{aligned} &\text{máx}_{\lambda, P, q, r, t, X} t \\ &\text{s.to: } X = \begin{bmatrix} P & q \\ q' & r - t \end{bmatrix} \geq 0 \\ &P = P_0 + \sum_{i=1}^m \lambda_i P_i, \quad q = q_0 + \sum_{i=1}^m \lambda_i q_i, \quad r = r_0 + \sum_{i=1}^m \lambda_i r_i \end{aligned} \quad (38)$$

Observe la sencillez del resultado, en que la matriz pseudo inversa desaparece y el SDP coincide con el obtenido para el caso convexo salvo por la eliminación de la restricción $\mu \geq 0$.

Apéndice B: Ínfimo de funciones **cóncavas**

Se quiere probar que si $l_u(z)$ es afín en $z \in Z$, para todo $u \in U$, entonces $d(z) = \inf_{u \in U} \{l_u(z)\}$ es **cóncava** en $z \in Z$.

Prueba:

$$\begin{aligned} d(\theta z_1 + (1 - \theta)z_2) &= \inf_{u \in U} \{l_u(\theta z_1 + (1 - \theta)z_2)\} \\ &= \inf_{u \in U} \{\theta l_u(z_1) + (1 - \theta)l_u(z_2)\} \\ &\geq \theta \inf_{u \in U} \{l_u(z_1)\} + (1 - \theta) \inf_{u \in U} \{l_u(z_2)\} \\ &= \theta d(z_1) + (1 - \theta)d(z_2) \end{aligned}$$

Apéndice C: Prueba del Teorema 1 (dualidad fuerte)

Se presentará primero la intuición geométrica tras la prueba de dualidad fuerte para luego pasar a una demostración formal. Como sugiere las pruebas de dualidad débil y de convexidad del dual conviene ver $\{g_i(x)\}_{i=1}^M$ como un punto que se mueve con x . Esto se logra definiendo el mapa $x \rightarrow (t, z) : t = g(x), z = f(x)$. Correspondientemente se aplica el mapa al conjunto factible X en el espacio $\mathcal{S} = \{(t, z) : \exists x \in X : t = g(x), z = f(x)\}$

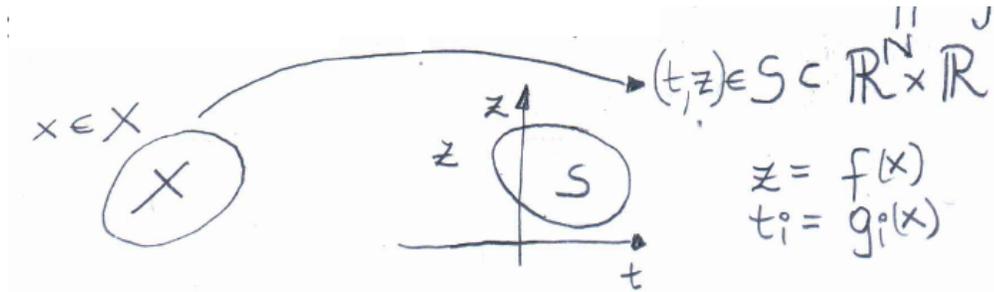


Figura 4: Mapeo a utilizar en la prueba de Strong duality.

Las ilustraciones de aquí en más se piensan con $M = 1$ pero las ecuaciones que se asocian generalizan dimensiones arbitrarias.

Clave de la prueba: Dualidad fuerte es fácil de probar una vez que se entiende la relación entre la función dual y el conjunto \mathcal{S} .

Dado $\mu \geq 0$ considérense las rectas de pendiente $-\mu$ que cortan el eje \bar{z} en $(0, k)$ con k arbitrario, lo que lleva a las siguientes tres ideas geométricas:

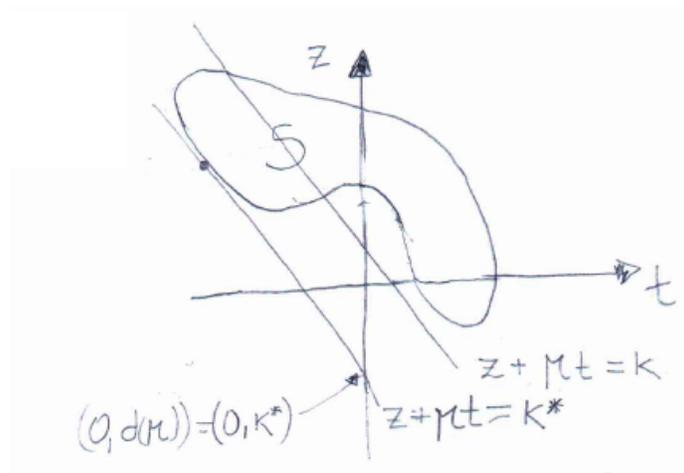


Figura 5: Rectas en el plano (z, t) .

Idea 1: $d(\mu)$ corresponde al corte del eje \bar{z} y la tangente a \mathcal{S} de pendiente $-\mu$.

Prueba: $d(\mu) = \inf_{x \in X} \{f(x) + \mu g(x)\} = \inf_{(z, t) \in \mathcal{S}} \{z + \mu t\} = k^*$

Idea 2: El óptimo del primario corresponde al punto más bajo de \mathcal{S} en el semiplano izquierdo $\{t \leq 0\}$. Prueba: $f^* = \min_{(t, z) \in \mathcal{S}} z : t \leq 0$.

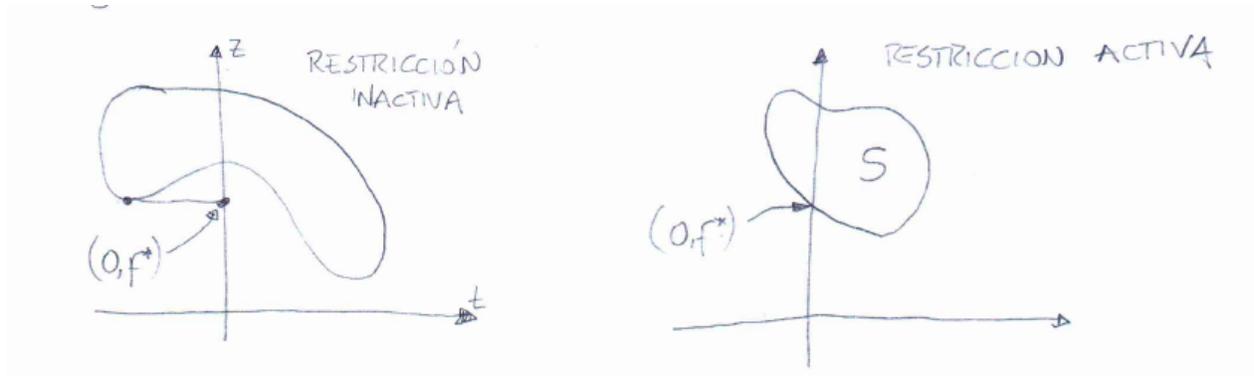


Figura 6: Óptimo del primario en el plano (z, t) .

Idea 3: Si S fuera convexo, el corte más alto de las tangentes y el piso de S en $\{t \leq 0\}$ coincidirían. Luego $d^* = f^*$

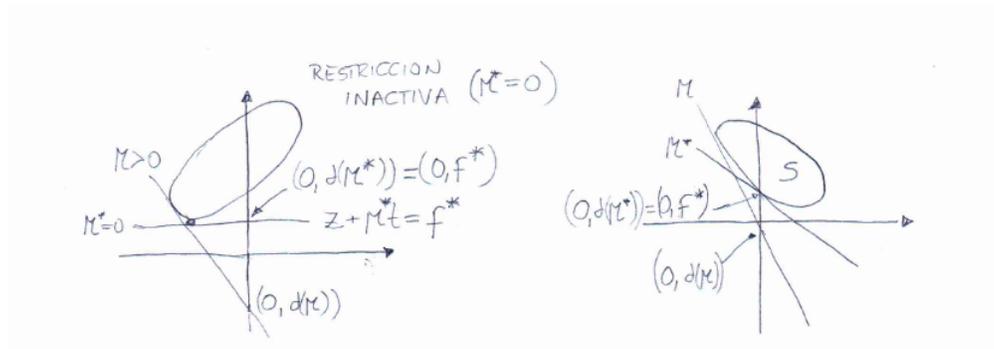


Figura 7: dualidad fuerte para S convexo.

La idea 3 debe ser más elaborada pues el conjunto S no es generalmente convexo aún en las hipótesis de convexidad de las funciones f y g_i como lo muestra el siguiente contra-ejemplo y la figura 8.

$$\min_{x \in X} (x + 1)^2 \quad \text{sujeto a : } x \leq 0.$$

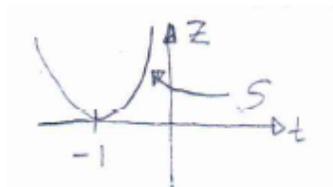


Figura 8: Contraejemplo para S convexo.

Si bien S no es convexo, puede rellenarse para conseguir de un conjunto convexo \mathcal{A} con el que

se continuará la prueba. Se define entonces \mathcal{A} como

$$\begin{aligned}\mathcal{A} &= \{(t_A, z_A) : \exists (t_S, z_S) \in \mathcal{S} : t_S \leq t_A, z_S \leq z_A\} \\ &= \{(t, z) : \exists x \in X : g_i(x) \leq t_i, i = 1, \dots, M, f(x) \leq z\}\end{aligned}$$

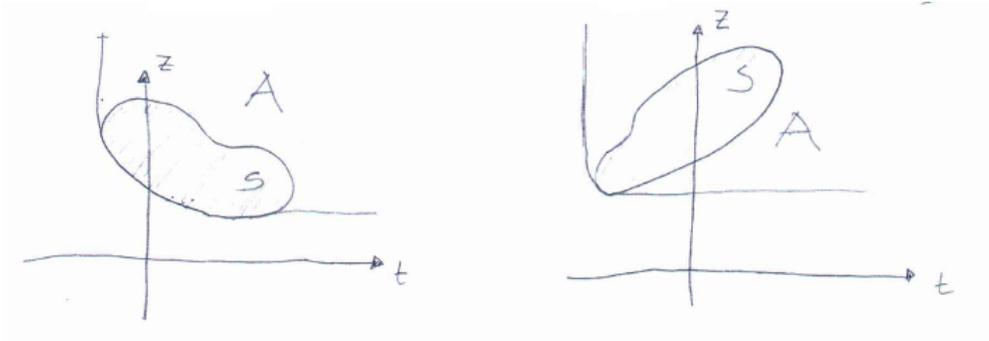


Figura 9: definición del conjunto \mathcal{A} convexo.

Se probará que

- 1) X, f, g convexos $\Rightarrow \mathcal{A}$ convexo.
- 2) \mathcal{A} convexo y hay un punto interior de Slater \Rightarrow existe un hiperplano que pasa por $(0, f^*)$ y soporta a \mathcal{A} de la forma $H = \{(t, z) : z + \mu_H t = f^*\}$.
- 3) Probaremos que tiene pendiente negativa, lo que corresponde a $\mu_H \geq 0$.
- 4) Luego μ_H es candidato a multiplicador óptimo, y probaremos $d(\mu_H) \geq f^*$.

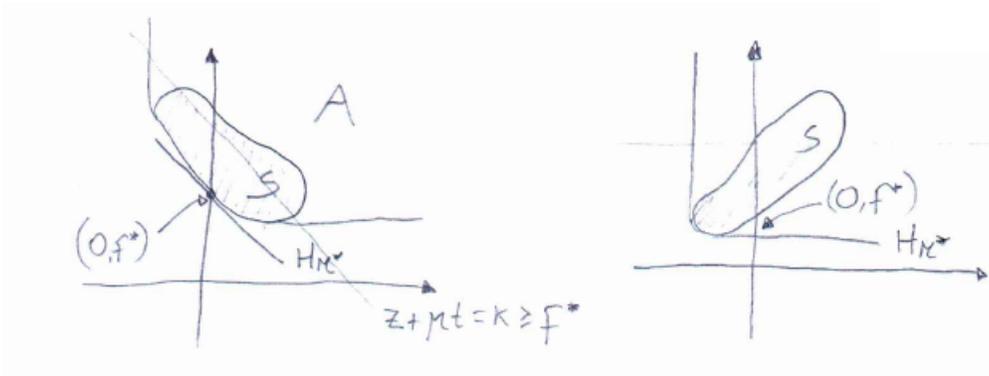


Figura 10: Existe hiper-plano H_{μ^*} tangente a \mathcal{A} por $(0, f^*)$ con pendiente $-\mu^* \Rightarrow d^* = f^*$.

En lo que sigue se prueba dualidad fuerte formalmente partiendo de las hipótesis del teorema 1 para llegar a $d^* \geq f^*$. Se necesita el siguiente lema que se presenta sin demostración

Lema 1. [4, p.794] Si $\mathcal{A} \in \mathbb{R}^M \times \mathbb{R}$ convexo y (t^*, z^*) no pertenece al interior de \mathcal{A} , entonces existe hiperplano H por (t^*, z^*) que soporta a \mathcal{A} , esto es $\exists (\beta, \mu) \neq (0, 0) : \beta z + \mu t \geq \beta z^* + \mu t^*, \forall (z, t) \in \mathcal{A}$

Éste es un resultado intuitivo de geometría con conjuntos convexos según se muestra en la figura 11

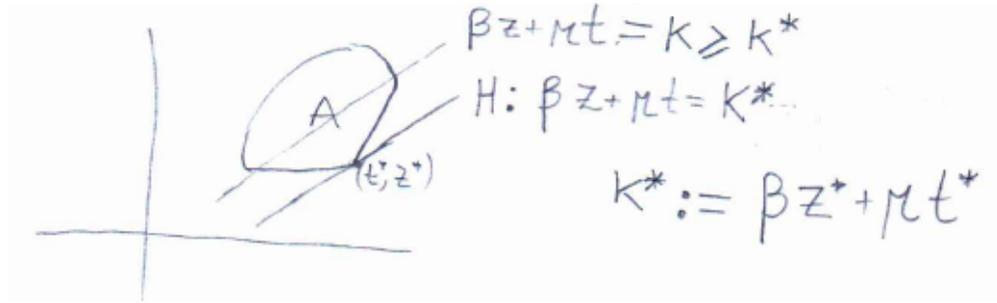


Figura 11: Dado un conjunto convexo \mathcal{A} y un punto en su frontera existe un hiperplano que soporta a \mathcal{A} .

Obsérvese que este lema no asume que $f(x)$ o $g(x)$ sean diferenciables ya que el hiperplano que soporta a un conjunto no necesita ser tangente a él.

Se pasa ahora a probar formalmente el Teorema 1.

Demostración. Como antes se define

$$\mathcal{A} = \{(t, z) : \exists x \in X : g_i(x) \leq t_i, f(x) \leq z\}$$

Se deja para probar al lector que \mathcal{A} es convexo, comprobando que si $(t, z) \in \mathcal{A}$ y $(\bar{t}, \bar{z}) \in \mathcal{A}$ entonces $\theta(t, z) + (1 - \theta)(\bar{t}, \bar{z}) \in \mathcal{A}$ con $\theta \in (0, 1)$ siempre que $X, f, y \{g_i\}$ sean convexos.

El punto $(0, f^*)$ no es interior a \mathcal{A} pues si lo fuera entonces $\exists \epsilon > 0$ tal que $(0, f^* - \epsilon) \in \mathcal{A}$ y por ende existiría $x \in X$ tal que $g_i(x) \leq 0$ y $f(x) \leq f^* - \epsilon$. Pero esto contradice que f^* sea óptimo del primario.

Usando el lema con $(t^*, z^*) = (0, f^*)$ se tiene que

$$\exists(\beta, \mu) \neq (0, 0) : \beta z + \mu t \geq \beta z^* + \mu t^* = \beta f^*, \forall (z, t) \in \mathcal{A} \quad (39)$$

En lo que resta hay que probar que $\beta > 0$ y $\mu \geq 0$. En particular $\beta \neq 0$ permitirá normalizar la desigualdad en (39) para convertirla en $z + \mu/\beta t \geq f^*$ y tomar $\mu_H := \mu/\beta$ como nuestro candidato a multiplicador óptimo. Primero se argumenta que $\beta \geq 0$ y $\mu \geq 0$ como sigue. Si $(t, z) \in \mathcal{A}$ entonces por construcción de \mathcal{A} se tiene que para todos $u > 0$ y $v > 0$ (escalares) $(t + e_i u, z) \in \mathcal{A}$ y $(t, z + v) \in \mathcal{A}$. Luego

$$\begin{aligned} \beta z + \beta v + \mu' t &\geq f^*, \forall v > 0 \Rightarrow \beta \geq 0 \\ \beta z + \mu' t + \mu_i u &\geq f^*, \forall u > 0 \Rightarrow \mu_i \geq 0 \end{aligned}$$

Esto significa que la pendiente de la tangente no puede ser positiva lo cual es razonable ya que por construcción \mathcal{A} “no levanta” hacia la derecha.

Probemos ahora que $\beta > 0$ estrictamente. Si $\mu = 0$ entonces $\beta \neq 0$ pues el lema asegura que $(\beta, \mu) \neq (0, 0)$, y como ya tenemos $\beta \geq 0$ entonces $\beta > 0$. Probemos ahora que $\beta > 0$ cuando $\mu \neq 0$, lo que corresponde a decir que la pendiente no es vertical. Para ello se utiliza la condición de Slater de punto interior que coloca un punto en el semiplano izquierdo. Tomemos el punto $\bar{x} \in X$ tal que $g_i(\bar{x}) < 0$ para todo $i = 1, \dots, M$ en la condición de Slater. Definiendo $\bar{t} = g(\bar{x}) < 0$ y $\bar{z} = f(\bar{x})$ se tiene que $(\bar{t}, \bar{z}) \in \mathcal{S} \subset \mathcal{A}$. Entonces puede usarse (39) para obtener $\beta(\bar{z} - f^*) \geq -\mu' \bar{t} \geq 0$. Como $\mu \neq 0$ entonces existe i tal que $\mu_i > 0$. Entonces $\beta(\bar{z} - f^*) \geq -\mu_i \bar{t}_i > 0$ lo que implica $\beta > 0$.

Con $\beta > 0$ normalizamos $\mu_H = \mu/\beta$ y (39) se convierte en

$$(t_A, z_A) \in \mathcal{A} \Rightarrow z_A + \mu_H t_A \geq z_H + \mu_H t_H = f^*$$

A partir de ello, la siguiente cadena de implicancias muestra que μ_H y H pueden tomarse como multiplicador óptimo y Lagrangeano, respectivamente, para concluir que $f^* \geq d^*$ concluyendo la prueba.

$$\forall (t_A, z_A) \in \mathcal{A} \Rightarrow z_A + \mu_H t_A \geq f^* \quad (40)$$

$$\Rightarrow \forall x \in X \ (g(x), f(x)) \in \mathcal{S} \subset \mathcal{A} \Rightarrow f(x) + \mu'_H g(x) \geq f^* \quad (41)$$

$$\Rightarrow d(\mu_H) = \inf_{x \in X} \{f(x) + \mu'_H g(x)\} \geq f^* \quad (42)$$

$$\Rightarrow d^* = \sup_{\mu} d(\mu) \geq d(\mu_H) \geq f^* \quad (43)$$

□

Apéndice D: Dual de un problema económico

Se quiere obtener el precio λ^* que maximiza el dual. Substituyendo $x_i(\lambda)$ y $y_j(\lambda)$ en (10) se obtiene

$$\begin{aligned} d(\lambda) &= \sum_{i=1}^N (c_i(x_i(\lambda)) - \lambda x_i(\lambda)) - \sum_{j=1}^M (u_j(y_j(\lambda)) - \lambda y_j(\lambda)) \\ &= - \sum_{i=1}^N s_i(\lambda) - \sum_{j=1}^M s_j(\lambda) \end{aligned} \quad (44)$$

donde $s_i(\lambda) := \lambda x_i(\lambda) - c_i(x_i(\lambda))$ y $s_j(\lambda) := u_j(y_j(\lambda)) - \lambda y_j(\lambda)$ representan los excedentes de las firmas y los consumidores. Las derivadas de estos excedentes vienen dadas por

$$\dot{s}_i(\lambda) = x_i(\lambda) + \dot{x}_i(\lambda) \left(\lambda - \frac{d}{d\lambda} c_i(x_i(\lambda)) \right) = x_i(\lambda) \quad (45)$$

$$\dot{s}_j(\lambda) = \dot{x}_i(\lambda) \left(\frac{d}{d\lambda} u_j(y_j(\lambda)) - \lambda \right) - y_j(\lambda) = -y_j(\lambda) \quad (46)$$

donde se uso (13) y (14). Por ende

$$\frac{d}{d\lambda} d(\lambda) = - \sum_{i=1}^N \dot{s}_i(\lambda) - \sum_{j=1}^M \dot{s}_j(\lambda) = \sum_{j=1}^M y_j(\lambda) - \sum_{i=1}^N x_i(\lambda) \quad (47)$$

Luego se deduce de igualar a cero (47) que el precio óptimo λ^* balancea el mercado asegurando que $\sum_{j=1}^M y_j(\lambda^*) = \sum_{i=1}^N x_i(\lambda^*)$. El resultado (47) dice que el gradiente del dual esta dado por la restricción. Este es un resultado general que se probó para el subgradiente al construir el algoritmo primal-dual.

Apéndice E: Flujo de potencia óptimo

Partiendo del siguiente problema con $\alpha_2 > 2\alpha_1$ y $R > d_2 - d_3$

$$\min_{(p_1, p_2, g_1, g_2)} \alpha_1 g_1 + \alpha_2 g_2 \quad (48)$$

$$\text{sujeto a: } g_1 - 2p_1 - p_2 = 0 \quad (\lambda_1) \quad (49)$$

$$g_2 + p_1 + 2p_2 - d_2 = 0 \quad (\lambda_2) \quad (50)$$

$$p_1 - p_2 - d_3 = 0 \quad (\lambda_3) \quad (51)$$

$$g_1 \geq 0 \quad (\mu_1) \quad (52)$$

$$g_2 \geq 0 \quad (\mu_2) \quad (53)$$

$$p_2 \leq R \quad (\mu_3) \quad (54)$$

se tiene

$$\begin{aligned} L(g, p, \lambda, \mu) &= \alpha_1 g_1 + \alpha_2 g_2 + \lambda_1(g_1 - 2p_1 - p_2) + \lambda_2(g_2 + p_1 + 2p_2 - d_2) \\ &\quad + \lambda_3(p_1 - p_2 - d_3) - \mu_1 g_1 - \mu_2 g_2 + \mu_3(p_2 - R) \\ &= (\alpha_1 + \lambda_1 - \mu_1)g_1 + (\alpha_2 + \lambda_2 - \mu_2)g_2 \\ &\quad + (\lambda_3 + \lambda_2 - 2\lambda_1)p_1 + (2\lambda_2 - \lambda_3 - \lambda_1 + \mu_3)p_2 + c \end{aligned}$$

donde c es el remanente constante que no depende de g o p .

Para que $\inf_{g,p} L(g, p, \lambda, \mu) > -\infty$, las cantidades en paréntesis deben ser nulas, lo cual arroja las siguientes ecuaciones

$$\begin{aligned} \alpha_1 + \lambda_1 - \mu_1 &= 0 \\ \alpha_2 + \lambda_2 - \mu_2 &= 0 \\ \lambda_3 + \lambda_2 - 2\lambda_1 &= 0 \\ 2\lambda_2 - \lambda_3 - \lambda_1 + \mu_3 &= 0 \end{aligned}$$

Agregando las restricciones de igualdad

$$\begin{aligned} g_1 - 2p_1 - p_2 &= 0 \\ g_2 + p_1 + 2p_2 - d_2 &= 0 \\ p_1 - p_2 - d_3 &= 0 \end{aligned}$$

se tienen 7 ecuaciones para 10 incógnitas. Las restantes 3 ecuaciones se obtienen a partir de complementary slackness para las restricciones de desigualdad y sus multiplicadores asociados. Debe cumplirse que

$$\mu_1 = 0 \text{ o } g_1 = 0 \quad (55)$$

$$\mu_2 = 0 \text{ o } g_2 = 0 \quad (56)$$

$$\mu_3 = 0 \text{ o } p_2 = R \quad (57)$$

donde los “o” son no excluyentes, lo que significa que por ejemplo puede darse que $\mu_1 = 0$ y $g_1 = 0$, pero si $\mu_1 > 0$ entonces $g_1 = 0$, y si $g_1 > 0$ entonces $\mu_1 = 0$. Luego se obtendrán a partir de (55)-(57) ocho posibles tripletas de ecuaciones que completen las 10 necesarias. Deben estudiarse todas las ocho combinaciones en (55)-(57), pero en este caso la física del problema original nos permite elegir

las combinaciones más plausibles. Cabe recordar que por ser las condiciones de KKT suficientes para el caso convexo, nos alcanza con encontrar una combinación compatible. Si las consideraciones físicas resultan poco sistemáticas para el lector, se invita a éste a resolver el problema sin usarlas y chequeando las 8 combinaciones posibles.

Consideremos entonces que $g_1 > 0$ lo cual implica usar energía más barata mientras las restricciones de red lo permitan. Se tiene entonces a partir de (55) que $\mu_1 = 0$.

Conjeturemos también la restricción de potencia está activa ($p_2 = R$) con lo que la transferencia de energía desde la barra 1 estará acotada y se requerirá abastecer las cargas con $g_2 > 0$. De estas consideraciones físicas se obtiene $p_2 = R$ y $\mu_2 = 0$. El sistema de ecuaciones resultante según esta conjetura tiene las siguientes 10 ecuaciones para 10 incógnitas

$$\begin{aligned}
 \alpha_1 + \lambda_1 - \mu_1 &= 0 \\
 \alpha_2 + \lambda_2 - \mu_2 &= 0 \\
 \lambda_3 + \lambda_2 - 2\lambda_1 &= 0 \\
 2\lambda_2 - \lambda_3 - \lambda_1 + \mu_3 &= 0 \\
 g_1 - 2p_1 - p_2 &= 0 \\
 g_2 + p_1 + 2p_2 - d_2 &= 0 \\
 p_1 - p_2 - d_3 &= 0 \\
 \mu_1 &= 0 \\
 \mu_2 &= 0 \\
 p_2 &= R
 \end{aligned}$$

y su solución es

$$\begin{aligned}
 p_1 &= d_3 + R \\
 p_2 &= R \\
 g_1 &= 2d_3 + 3R \\
 g_2 &= d_2 - d_3 - 3R \\
 \lambda_1 &= -\alpha_1 \\
 \lambda_2 &= -\alpha_2 \\
 \lambda_3 &= \alpha_2 - 2\alpha_1 \\
 \mu_1 &= 0 \\
 \mu_2 &= 0 \\
 \mu_3 &= 3(\alpha_2 - \alpha_1)
 \end{aligned}$$

donde la cuarta restricción es compatible con $g_2 > 0$ al haber planteado el problema con demandas tales que $d_2 - d_3 \geq 3R$.

Si las constantes del problema son tales que $d_2 - d_3 < 3R$ la conjetura nos lleva a un sistema de ecuaciones incompatibles. En este caso la combinación en (55)-(57) que nos lleva a un sistema compatible es que la restricción de línea no esté activa y por ende no sea necesario usar el generador más costoso en la barra 2. Entonces las ecuaciones $p_2 = R$ y $\mu_2 = 0$ se descartan, pero se substituyen por $\mu_3 = 0$ y $g_2 = 0$ recuperando 10 ecuaciones ¹

¹Cuando $g_2 = 0$ entonces tanto $\mu_2 = 0$ como $\mu_2 > 0$ son compatibles con (56), por lo que $\mu_2 = 0$ aún es posible, pero no puede forzarse como ecuación.

$$\begin{aligned}
\alpha_1 + \lambda_1 - \mu_1 &= 0 \\
\alpha_2 + \lambda_2 - \mu_2 &= 0 \\
\lambda_3 + \lambda_2 - 2\lambda_1 &= 0 \\
2\lambda_2 - \lambda_3 - \lambda_1 + \mu_3 &= 0 \\
g_1 - 2p_1 - p_2 &= 0 \\
g_2 + p_1 + 2p_2 - d_2 &= 0 \\
p_1 - p_2 - d_3 &= 0 \\
\mu_1 &= 0 \\
g_2 &= 0 \\
\mu_3 &= 0
\end{aligned}$$

y su solución es $p_1 = \frac{1}{3}(2d_3 + d_2)$, $p_2 = \frac{1}{3}(d_2 - d_3)$, $g_1 = d_2 + d_3$, $g_2 = 0$ y $\lambda_1 = -\alpha_1$, $\lambda_2 = -\alpha_1$, $\lambda_3 = -\alpha_1$, $\mu_1 = 0$, $\mu_2 = \alpha_2 - \alpha_1$, $\mu_3 = 0$.

Referencias

- [1] Lavaei, Javad, and Steven H. Low. "Zero duality gap in optimal power flow problem." *IEEE Transactions on Power Systems* 27.1 (2012): 92-107.
- [2] V. Gupta, I. E. Grossmann "A new decomposition algorithm for multistage stochastic programs with endogenous uncertainties" *Computers and Chemical Engineering*, vol. 62, pp. 62-79, 2013.
- [3] A. Ribeiro and G. B. Giannakis. "Separation principles in wireless networking," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4488-4505, 2010.
- [4] D. P. Bertsekas *Nonlinear programming 3rd Ed.* Athena Scientific, 2017
- [5] S Boyd and L. Vandenberghe *Convex optimization* Cambridge University Press, 2006.
- [6] F. Genoese, M. Genoese, and M. Wietschel, "Occurrence of negative prices on the German spot market for electricity and their influence on balancing power markets," *IEEE Conf. European Electricity Market*, pp. 1-6, 2010.
- [7] F. Zhang, editor *The Schur complement and its applications*, vol. 4. Springer Science and Business Media, 2006.