

**UNIVERSIDAD NACIONAL DE INGENIERÍA  
FACULTAD DE CIENCIAS**



**TESIS**

**“DUALIDAD NO CONVEXA Y OPTIMIZACIÓN  
ENERGÉTICA”**

PARA OBTENER EL GRADO ACADÉMICO DE  
MAESTRO EN CIENCIAS EN MATEMÁTICA APLICADA

**ELABORADA POR:**

AMERICO ANDRES CHULLUNCUY REYNOSO

**ASESOR:**

Dr. ELADIO OCAÑA ANAYA (IMCA, Perú)

**ASESORA EXTERNA:**

Dra. CLAUDIA SAGASTIZÁBAL (UNICAMP, Brasil)

**LIMA-PERÚ**

2018

## **Dedicatoria**

El presente trabajo está dedicado a mis padres, pilares fundamentales en mi vida. Les estoy infinitamente agradecido por su gran apoyo incondicional.

A mis abuelos: Alejandro, Nestor y Lucina, por sus sabios consejos. Doy gracias a dios por haberme permitido conocerlos.

## **Agradecimientos**

Mis profundos y sinceros agradecimientos a la Dra Claudia Sagastizábal, por su constante apoyo, sus valiosos consejos y su gran paciencia. De igual forma al Dr. Eladio Ocaña por haberme guiado en el desarrollo del presente trabajo.

## Abstract

We develop the classical theory of duality through the classical Lagrangian as well as a more general theory (based on the Legendre-Fenchel conjugate) through the augmented Lagrangian, with the objective of eliminating the duality gap that arises when optimization problems are non-convex. In particular, we study the sharp augmented Lagrangian for a *DC* optimization problem, that is, a problem whose objective function or constraints are given as the difference of two convex functions. We also describe two nonsmooth optimization algorithms the subgradient and the cutting planes ones, for the augmented dual problem.

Finally we describe an application to an energy problem and we give the formulation via the sharp augmented Lagrangian for the problem, as well as some initial results to solve it.

## Resumen

En este trabajo desarrollamos la teoría clásica de dualidad por medio del lagrangiano clásico así como la teoría general (basada en la conjugada de Legendre-Fenchel) mediante el lagrangiano aumentado, con el objetivo de eliminar el salto de dualidad que surge cuando tratamos con problemas de optimización no convexos. En particular estudiaremos el lagrangiano aumentado sharp para un problema de optimización  $DC$ , es decir, un problema cuya función objetivo o restricciones es la diferencia de dos funciones convexas. Describimos además los algoritmos de subgradiente y de los planos cortantes para el problema dual aumentado.

Finalmente describimos una aplicación a un problema de energía y realizamos una reformulación de este, vía el lagrangiano aumentado sharp, además damos algunos resultados iniciales para resolverlo.

## Índice de contenidos

<b>Dedicatoria</b>	<b>ii</b>
<b>Agradecimientos</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Resumen</b>	<b>v</b>
<b>Lista de figuras</b>	<b>viii</b>
<b>Introducción</b>	<b>1</b>
<b>1. Definiciones y nociones básicas</b>	<b>3</b>
1.1. Análisis convexo y análisis variacional . . . . .	3
1.2. Optimización DC . . . . .	6
<b>2. Aspectos prácticos de dualidad</b>	<b>8</b>
2.1. Introducción a la dualidad en optimización matemática . . . . .	8
2.2. Teoría necesaria para resolver el problema primal . . . . .	10
2.2.1. El problema dual . . . . .	11
2.2.2. Caracterización de optimalidad primal-dual . . . . .	14
2.2.3. Existencia de solución dual . . . . .	16
2.2.4. Relación entre dualidad y función primal . . . . .	18
2.3. Lagrangiano aumentado . . . . .	22
<b>3. Dualidad no convexa</b>	<b>25</b>
3.1. Problemas duales en optimización . . . . .	30
3.2. Función Lagrangiano . . . . .	33
3.3. Lagrangiano aumentado generalizado y dualidad no convexa . . . . .	36
<b>4. Métodos de optimización no diferenciable</b>	<b>45</b>
4.1. Métodos duales clásicos . . . . .	45
4.1.1. Algoritmo del subgradiente . . . . .	46

4.1.2. Método de planos cortantes . . . . .	48
4.2. Aplicación al lagrangiano sharp . . . . .	52
<b>5. Aplicación a un problema de energía y perspectivas futuras</b>	<b>58</b>
5.1. Un problema de manejo de energía . . . . .	58
5.2. Formulación binivel . . . . .	59
5.3. Formulación dual generalizada del problema . . . . .	63
5.4. Aproximaciones en la resolución del problema . . . . .	64
<b>Bibliografía</b>	<b>66</b>

## Índice de figuras

2.1. Salto de dualidad . . . . .	19
2.2. Función primal y salto de dualidad . . . . .	20
3.1. Conjugada en espacios producto . . . . .	28
3.2. Salto de dualidad en problemas de minimización. . . . .	37
3.3. Salto de dualidad en un problema no convexo . . . . .	38



## Introducción

Dualidad es una herramienta muy utilizada en optimización matemática, ya que nos permite ver un problema de optimización desde dos perspectivas, el problema primal y el problema dual. La solución del problema dual nos da una cota inferior para la solución del primal (en el caso de un problema de minimización). Sin embargo en general los valores óptimos del problema primal y dual no necesariamente coinciden y su diferencia es llamada salto de dualidad.

La teoría de dualidad obtenida a través del lagrangiano clásico es bien conocida, sobre todo cuando es aplicada a problemas primales convexos. Sin embargo cuando el problema primal es no convexo y utilizamos el lagrangiano clásico puede existir un salto de dualidad. Esto justifica la búsqueda de otros tipos de lagrangianos. Trabajos recientes se han enfocado en problemas duales construidos utilizando funciones lagrangianas aumentadas. Un tipo particular de tales funciones es el lagrangiano aumentado sharp, que ha sido estudiado para resolver problemas no convexos y no diferenciables en [6], [8].

Lagrangianos clásicos y lagrangianos aumentados han sido utilizados en el diseño de métodos para resolver problemas de optimización con restricciones. Las principales dificultades en el desarrollo de tales métodos son garantizar el salto de dualidad nulo y probar la convergencia de los métodos a una solución del problema original.

Los métodos basados en lagrangianos clásicos convergen bajo hipótesis restrictivas, desde el punto de vista de las aplicaciones, tales como asumir convexidad del conjunto factible o de la función objetivo [10].

El uso de lagrangianos aumentados permite desarrollar algoritmos para una amplia clase de problemas de optimización restringidos no convexos donde el salto de dualidad nulo no puede ser alcanzado mediante el lagrangiano clásico.

El primer algoritmo genérico basado en el lagrangiano sharp, denominado Algoritmo de Subgradiente Modificado, fue introducido por Gasimov [5]. Este algoritmo no requiere de convexidad ni diferenciableidad de la función objetivo ni las restricciones. Se demostró que el algoritmo genera una sucesión creciente de valores duales que convergen al óptimo dual. El método necesita hallar soluciones globales de los problemas lagrangianos que hay que resolver para evaluar la función dual aumentada en cada operación. Este requerimiento es en general difícil de satisfacer en las aplicaciones. En una versión modificada, desarrollada por Kasimbeyli [9], se propusieron fórmulas especiales para actualizar las variables duales en el algoritmo y se probó la convergencia para el

caso donde no se requiere el cálculo de un mínimo global del lagrangiano aumentado en cada iteración. Por lo tanto en esta variante cualquier algoritmo de minimización local puede ser aplicado para la minimización no restringida del lagrangiano aumentado. Otras formas similares para actualizar las variables duales fueron consideradas en [6],[7].

En [8], se propone dos variantes del Algoritmo de Subgradiente Modificado que, a diferencia de las anteriores, poseen convergencia primal y dual, es decir, todo punto de acumulación de una sucesión primal (respectivamente dual) generada por el método es una solución primal (respectivamente dual) del problema original.

Este trabajo está organizado del modo siguiente. El capítulo 1 contiene algunas definiciones y resultados de análisis convexo y análisis variacional que serán utilizados en los siguientes capítulos. Asimismo introducimos algunos hechos básicos sobre funciones  $DC$  y presentamos diferentes tipos de condiciones de optimalidad utilizadas en optimización  $DC$ .

En el capítulo 2 se estudia la relajación lagrangiana, que consiste en aproximar un problema difícil de optimización con restricciones por un problema de menor dificultad. Para un problema (primal) dado formulamos su problema dual respectivo utilizando la relajación lagrangiana y estudiamos algunos resultados que garantizan la existencia de solución para dichos problemas.

En el capítulo 3 se estudia la transformada de Legendre-Fenchel, sus consecuencias y el rol que desempeña en la dualización de problemas de optimización. Definimos la función lagrangiana y el lagrangiano aumentado en un contexto general y abordamos el estudio del salto de dualidad en problemas no convexos.

En el capítulo 4 revisamos los algoritmos clásicos destinados a resolver el problema dual, en particular estudiamos el método de subgradiente y de planos cortantes, así como sus respectivos resultados de convergencia. Además aplicaremos estos algoritmos a un tipo particular de problema de optimización  $DC$  (con restricciones en forma de desigualdad) utilizando el lagrangiano sharp y realizamos una comparación con el trabajo de Gasimov, [5], que se enfoca en problemas primales con restricciones en forma de igualdad.

Finalmente, en el capítulo 5 se estudia un problema de manejo de energía, donde describimos la formulación binivel para dicho problema y una formulación utilizando el lagrangiano aumentado sharp así como algunas aproximaciones iniciales para la resolución del problema via esta última formulación.

## Definiciones y nociones básicas

Este capítulo contiene algunas definiciones y resultados de análisis convexo que serán utilizados en los siguientes capítulos. Siendo de nuestro interés los problemas de optimización *DC*, también introducimos algunos hechos básicos sobre funciones *DC*. Finalmente debido a su importancia para detectar soluciones candidatas, donde la ejecución de un algoritmo pueda parar, presentamos diferentes tipos de condiciones de optimalidad utilizadas en optimización *DC*.

### 1.1. Análisis convexo y análisis variacional

Sea  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función, el *epígrafo* de  $f$  es el conjunto

$$\text{epi}(f) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : \alpha \geq f(x)\},$$

la imagen de  $\text{epi } f$  bajo la proyección  $(x, \alpha) \mapsto x$  es el *dominio* de  $f$ , y será denotado por  $\text{dom } f$ .

Un conjunto  $C$  de  $\mathbb{R}^n$  es convexo si para todo  $x_0 \in C$  y  $x_1 \in C$  tenemos:

$$(1-t)x_0 + tx_1 \in C \quad \text{para todo } t \in (0, 1).$$

Una función  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  es convexa si su epígrafo es un conjunto convexo. Se dice que  $f$  es propia si  $f(x) > -\infty$  para todo  $x \in \text{dom } f \neq \emptyset$ .

**Definición** (Semicontinuidad). Una función  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  es *semicontinua inferior* (*sci*) en  $\bar{x}$  si

$$\liminf_{x \rightarrow \bar{x}} f(x) \geq f(\bar{x}).$$

Si esta propiedad se tiene para todo  $\bar{x} \in \mathbb{R}^n$ , se dice que  $f$  es *semicontinua inferior* en  $\mathbb{R}^n$ . La función  $f$  es *semicontinua superior* (*scs*) en  $\bar{x}$  si  $-f$  es *sci* en  $\bar{x}$ .

El siguiente resultado da una caracterización de la semicontinuidad inferior:

- (i)  $f$  es *sci* en  $\mathbb{R}^n$
- (ii) El epígrafo de  $f$  es cerrado en  $\mathbb{R}^n \times \mathbb{R}$ .

- (ii) Todos los conjuntos de subnivel  $[f \leq \alpha] := \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$  son cerrados en  $\mathbb{R}^n$

Basado en este resultado se define la regularización sci (clausura) de  $f$ , denotada por  $\text{cl } f$ , como la mayor función sci tal que  $\text{cl } f \leq f$ . Análogamente definimos la envoltura convexa de  $f$ , denotada por  $\text{co } f$ , como la mayor función convexa tal que  $\text{co } f \leq f$ .

**Teorema 1.1.** *Dada una función convexa  $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ , entonces  $\text{cl } f$  es convexa. Más aún  $\text{cl } f$  es propia si y sólo si  $f$  es propia.*

Se dice que una función afín  $l$  soporta a una función  $f$  en  $\bar{x}$  si  $l(x) \leq f(x)$  para todo  $x$  y  $l(\bar{x}) = f(\bar{x})$ . Uno de los principales resultados de análisis convexo afirma que: Una función propia, sci y convexa  $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  es el supremo puntual de sus soportes afines, es decir

$$f = \sup\{l : l(x) \leq f(x) \ \forall x, \ l \text{ afín}\} \quad (1.1)$$

**Definición** (Acotación de nivel). *Se dice que una función  $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  es de nivel acotado si para todo  $\alpha \in \mathbb{R}$  el conjunto de subnivel  $[f \leq \alpha]$  es acotado.*

El siguiente resultado da condiciones suficientes para que una función  $f$  alcance su mínimo sobre  $\mathbb{R}^n$

**Teorema 1.2** (Existencia de un mínimo). *Sea  $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  una función propia, sci y de nivel acotado, entonces el valor  $\inf f$  es finito y el conjunto  $\text{argmin } f$  es no vacío y compacto.*

La siguiente definición es una generalización del concepto de función de nivel acotado, y es útil en optimización paramétrica.

**Definición** (Acotación de nivel uniforme). *Una función  $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$  con valores  $f(x, u)$  es de nivel acotado en  $x$ , localmente uniforme en  $u$  si para cada  $\bar{u} \in \mathbb{R}^m$  y  $\alpha \in \mathbb{R}$  existe una vecindad  $V$  de  $\bar{u}$  tal que el conjunto  $\{(x, u) : u \in V, f(x, u) \leq \alpha\}$  es acotado en  $\mathbb{R}^n \times \mathbb{R}^m$ .*

**Teorema 1.3** (Optimización paramétrica). *Sea  $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$  una función propia y sci tal que  $f(x, u)$  es de nivel acotado en  $x$ , localmente uniforme en  $u$ . Consideremos*

$$p(u) := \inf_x f(x, u), \quad P(u) := \text{argmin}_x f(x, u),$$

*entonces, la función  $p$  es propia y sci en  $\mathbb{R}^m$ . Además, para cada  $u \in \text{dom } p$ , el conjunto  $P(u)$  es no vacío y compacto.*

Dado un conjunto  $C \subseteq \mathbb{R}^n$ , la cápsula afín de  $C$ , denotada por  $\text{aff } C$ , es la intersección de todos los subespacios afines que contienen a  $C$ .

**Definición** (Interior relativo). *El interior relativo de un conjunto convexo  $C \subseteq \mathbb{R}^n$ , denotado por  $\text{ri } C$ , es el interior de  $C$  en la topología relativa a la cápsula afín de  $C$ .*

El siguiente resultado muestra la importancia de esta definición

**Teorema 1.4.** *Sea  $C \subseteq \mathbb{R}^n$  un conjunto convexo no vacío, entonces  $\text{ri } C \neq \emptyset$ . Más aún existe un simplex  $\Delta := \{(t_0, \dots, t_n) \in \mathbb{R}^{n+1} : \sum_{i=0}^n t_i = 1, t_i \geq 0\}$  tal que  $\text{ri } \Delta \subseteq C$*

**Definición** (Subgradiente, subgradiente proximal). Sea  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función propia y convexa con  $\bar{x} \in \text{dom } f$ , el subgradiente de  $f$  en  $\bar{x}$  está definido por

$$\partial f(\bar{x}) := \left\{ v : f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle \text{ para todo } x \in \mathbb{R}^n \right\} \quad (1.2)$$

Un vector  $v$  es un subgradiente proximal de  $f$  en  $\bar{x}$  si existen  $\rho > 0$  y  $\delta > 0$  tales que

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle - \frac{1}{2}\rho|x - \bar{x}|^2 \text{ para } |x - \bar{x}| \leq \delta \quad (1.3)$$

El conjunto  $\partial f(\bar{x})$  es convexo; no vacío si  $\bar{x} \in \text{ri dom } f$  y se reduce al conjunto unitario  $\{\nabla f(x)\}$  si y sólo si  $f$  es diferenciable en  $\bar{x}$ . Una aplicación directa de la definición (1.2) da el siguiente resultado.

Para  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  propia y convexa, las siguientes afirmaciones son equivalentes

- (i)  $\bar{x}$  es un minimizador de  $f$ , es decir,  $f(x) \geq f(\bar{x})$  para todo  $x \in \mathbb{R}^n$ ,
- (ii)  $0 \in \partial f(\bar{x})$

La conjugada de Fenchel de una función  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  está definida por

$$f^*(v) := \sup_x \{ \langle v, x \rangle - f(x) \}.$$

Los dos teoremas siguientes contienen reglas de cálculo con subgradientes.

**Teorema 1.5** (Subgradiente del supremo de funciones convexas). Sea  $J$  un conjunto compacto de índices y  $\{f_j\}_{j \in J}$  una colección de funciones convexas de  $\mathbb{R}^n$  en  $\mathbb{R}$  tales que las funciones  $j \mapsto f_j(x)$  son scs para cada  $x \in \mathbb{R}^n$ . Si

$$f(x) := \sup \{ f_j(x) : j \in J \} < +\infty \text{ para todo } x \in \mathbb{R}^n,$$

donde  $J(x) := \{j \in J : f_j(x) = f(x)\}$  es el conjunto de índices activos, entonces

$$\partial f(x) = \text{co} \{ \cup \partial f_j(x) : j \in J(x) \},$$

**Teorema 1.6** (Pre-composición con una transformación afín). Sea  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  una función convexa y cerrada y  $A_0: \mathbb{R}^n \rightarrow \mathbb{R}^m$  una transformación lineal. Definamos  $A(x) := A_0(x) - q$  y supongamos que  $A(\mathbb{R}^n) \cap \text{dom } f \neq \emptyset$ . Entonces  $f \circ A$  es una función convexa y

$$(f \circ A)^*(\gamma) = \text{cl inf} \{ f^*(u) + \langle q, u \rangle : A_0^*u = \gamma \}$$

**Definición** (Coercividad de nivel). Una función  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  es de nivel coercivo si este es acotado inferiormente en conjuntos acotados y satisface

$$\liminf_{|x| \rightarrow \infty} \frac{f(x)}{|x|} > 0.$$

**Proposición 1.7** (Nivel acotado). Sea  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  una función convexa, sci y propia. Si algún conjunto subnivel  $[f \leq \alpha]$  es no vacío y acotado (por ejemplo el conjunto subnivel  $\text{argmin } f$ ), entonces  $f$  es de nivel acotado.

**Proposición 1.8** (Coercividad). Sea  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  una función sci y propia, entonces  $f$  es de nivel-coercivo si y sólo si existen  $\gamma \in (0, \infty)$  y  $\beta \in \mathbb{R}$  tales que  $f(x) \geq \gamma|x| + \beta$  para todo  $x$ .

**Proposición 1.9** (Coercividad y convexidad). *Sea  $f$  propia y  $sci$ . Si  $f$  es de nivel-coercivo entonces  $f$  es de nivel-acotado. Cuando  $f$  es además convexa, las dos propiedades son equivalentes.*

**Definición** (función “calm” inferiormente). *Una función  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  es “calm” inferiormente en  $\bar{x}$  con módulo  $\kappa \geq 0$  si  $f(\bar{x})$  es finito y en alguna vecindad  $V$  de  $\bar{x}$  se tiene*

$$f(x) \geq f(\bar{x}) - \kappa|x - \bar{x}| \quad \text{para todo } x \in V.$$

En particular, si  $f$  es convexa y lineal por partes,  $f(x) := \max_{i=1, \dots, m} \{a_i, x\} + b_i$ , entonces  $f$  es “calm” inferiormente. En general toda función convexa es “calm” inferiormente, esto es debido a que toda función convexa es localmente lipschitz en el interior de su dominio.

## 1.2. Optimización DC

Dado el problema de optimización

$$\begin{aligned} \min \quad & f_1(x) - f_2(x) \\ \text{s.a.} \quad & x \in X \\ & c_1(x) - c_2(x) \leq 0, \end{aligned} \tag{1.4}$$

donde  $X$  es un conjunto convexo y  $f_1, f_2, c_1, c_2: \mathbb{R}^n \rightarrow \mathbb{R}$  son funciones convexas. La función objetivo del problema (1.4) se denomina función *DC* por ser la diferencia de dos funciones convexas. El problema (1.4) es conocido como programa general *DC*. Si el conjunto factible no presenta la restricción de desigualdad, el problema correspondiente es llamado programa *DC* con restricción convexa.

Una función  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  es una función *DC* si esta puede ser representada como

$$f(x) = f_1(x) - f_2(x),$$

donde  $f_1$  y  $f_2$  son funciones convexas.

Las funciones *DC* son localmente lipschitzianas y pueden ser no diferenciables, además estas preservan su estructura *DC* bajo algunas operaciones (multiplicación por un escalar, suma, máximo, producto, valor absoluto) utilizadas con frecuencia en optimización.

Un problema de optimización *DC* consiste en minimizar una función objetivo *DC*. Luego el problema de minimización *DC* irrestricto es de la forma

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & x \in \mathbb{R}^n \end{aligned} \tag{1.5}$$

donde la función objetivo es una función *DC*. Por definición  $\bar{x}$  es un mínimo global de (1.5) si  $f(\bar{x}) \leq f(x)$  para todo  $x \in \mathbb{R}^n$ . Sin embargo debido a que una función *DC* es usualmente no convexa, el problema (1.5) puede tener muchos minimizadores locales diferentes del global.

Un punto  $\bar{x} \in \mathbb{R}^n$  es llamado minimizador local si existe  $\epsilon > 0$  tal que  $f(\bar{x}) \leq f(x)$  para todo  $x \in B(\bar{x}, \epsilon)$ . Aún cuando un minimizador global es siempre un minimizador local, puede ser difícil de distinguir de un minimizador global ya que los minimizadores

globales y locales usualmente satisfacen las mismas condiciones de optimalidad en los algoritmos. Por lo tanto estaremos satisfechos con métodos de optimización local que permitan encontrar soluciones locales.

El siguiente resultado contiene tres condiciones necesarias de optimalidad

**Teorema 1.10** (Condiciones necesarias de optimalidad-*DC*). *Dado las funciones convexas  $f_1, f_2: \mathbb{R}^n \rightarrow \mathbb{R}$ . Si  $\bar{x}$  es un minimizador local de una función *DC*  $f = f_1 - f_2$ , entonces se tiene las siguientes condiciones*

$$\partial f_2(\bar{x}) \subseteq \partial f_1(\bar{x}) \quad (1.6)$$

$$0 \in \partial f(\bar{x}) \quad (1.7)$$

$$\partial f_1(\bar{x}) \cap \partial f_2(\bar{x}) \neq \emptyset. \quad (1.8)$$

Los puntos que satisfacen la primera condición (1.6) son denominados puntos inf-estacionarios, esta condición es la más fuerte entre las presentadas pues implica las otras dos. Más aún, esta condición garantiza optimalidad local si  $f_2$  es una función poliedral de la forma  $f_2(x) = \max_{i=1, \dots, m} \{ \langle a_i, x \rangle + b_i \}$ . Por lo tanto podría resultar útil usar (1.6) como un criterio de parada en un algoritmo. Desafortunadamente esta condición es difícil de utilizar en la práctica puesto que necesitamos conocer todos los subdiferenciales de las componentes *DC*.

Los puntos satisfaciendo la condición (1.7) son llamados puntos estacionarios de Clarke. En el caso convexo esta condición también es suficiente, puesto que este garantiza un óptimo global. Aún cuando estas condiciones son utilizadas en optimización convexa y no convexa, ésta puede ser difícil de verificar para una función *DC* si realizamos los cálculos solamente para las componentes *DC*. Esto se sigue de la regla de cálculo para subdiferenciales

$$\partial f(x) \subseteq \partial f_1(x) - \partial f_2(x)$$

para un función *DC*,  $f = f_1 - f_2$ . Sin embargo para una descomposición *DC* arbitraria la estimación  $\partial f_1(x) - \partial f_2(x)$  puede ser muy mala.

La tercera condición (1.8) es una relajación de la condición inf-estacionario. Los puntos satisfaciendo esta condición relajada se denominan puntos críticos y son comúnmente utilizadas como criterio de parada en optimización *DC* ya que es más fácil de verificar a diferencia de las condiciones (1.6) y (1.7) y en general prevee buenas soluciones candidatas.

En el siguiente capítulo, estudiamos la dualidad lagrangiana (clásica) y algunos resultados para la existencia de solución del problema dual.

## Aspectos prácticos de dualidad

El tema central del presente capítulo es la relajación lagrangiana, que consiste en aproximar un problema difícil de optimización con restricciones por un problema aparentemente menos difícil. Para un problema (primal) dado se formulará su problema dual respectivo utilizando la relajación lagrangiana y estudiaremos algunos resultados que garantizan la existencia de solución para dichos problemas.

Entre las aplicaciones más relevantes de minimización convexa, se destacan: la *descomposición* en programación matemática y los algoritmos duales. Por un lado los algoritmos de minimización convexa tienen su campo de aplicación en temas de descomposición; por otro lado cuando descomponemos un problema de optimización (de gran escala) mediante un algoritmo, esencialmente estamos minimizando una función convexa, llamada la función dual asociada con el problema. Gran parte de este capítulo corresponden a [1, Cap XII].

### 2.1. Introducción a la dualidad en optimización matemática

Consideremos el siguiente problema de optimización,

$$\begin{aligned} \inf \quad & \varphi(x) \\ \text{s.a.} \quad & x \in U, \\ & c_j(x) = 0, \quad j = 1, \dots, m, \end{aligned} \tag{2.1}$$

al cual llamaremos *problema primal*, caracterizado por un conjunto no vacío  $U$  de variables de decisión factibles  $x$ , una función objetivo  $\varphi: U \rightarrow \mathbb{R}$  y funciones de restricción  $c_1, \dots, c_m: U \rightarrow \mathbb{R}$ . En particular, comparando (2.1) con (1.4),  $\varphi = f_1 - f_2$ ,  $U = X$  y cada  $c_j$  sería la diferencia de dos funciones convexas.

Denotando por  $c(x) := (c_1(x), \dots, c_m(x)) \in \mathbb{R}^m$  al vector de valores de las restricciones en  $x \in U$ , consideramos la función de Lagrange, definida por:

$$L(x, \lambda) := \varphi(x) + \langle \lambda, c(x) \rangle \quad \text{para todo } \lambda \in \mathbb{R}^m \text{ y } x \in U.$$

Para  $\lambda \in \mathbb{R}^m$  dado, esta función se asemeja a una perturbación de la función objetivo del problema primal (2.1) y nos dice que una violación de las restricciones es aceptada pero un precio  $\langle \lambda, c(x) \rangle$  debe ser pagado. Por ahora ninguna hipótesis es hecha en los



datos  $(U, \varphi, c)$ . Sin embargo, para propósitos prácticos supondremos lo siguiente:

**Hipótesis 2.1** (Práctica). *Asumiremos que el problema de optimización en  $x$*

$$\inf\{L(x, \lambda) : x \in U\}, \quad (2.2)$$

donde  $\lambda$  está fijo en  $\mathbb{R}^m$ , es considerablemente más simple que el problema primal (2.1).

El problema (2.2) será llamado *problema de Lagrange* asociado con  $\lambda$ .

En los siguientes ejemplos consideraremos problemas del tipo descomponible, es decir:

- $U = U^1 \times U^2 \times \cdots \times U^n$ ; las variables de optimización serán denotadas por  $x = (x^1, \cdots, x^n)$ , donde  $x^i \in U^i$  para  $i = 1, \cdots, n$ ;
- $\varphi$  es una suma de “funciones objetivo individuales”:  $x = (x^1, \cdots, x^n) \mapsto \varphi(x) = \sum_{i=1}^n \varphi^i(x^i)$ ;
- de igual manera, cada restricción es una suma:  $c_j(x) = \sum_{i=1}^n c_j^i(x^i)$ . Entonces el problema primal (2.1) adopta el formato

$$\begin{aligned} \inf \quad & \sum_{i=1}^n \varphi^i(x^i) \\ \text{s.a.} \quad & x^i \in U^i, \quad i = 1, \cdots, n \\ & \sum_{i=1}^n c_j^i(x^i) = 0, \quad j = 1, \cdots, m. \end{aligned} \quad (2.3)$$

**Observación 2.2.** *Si la restricción (acoplante)  $\sum_{i=1}^n c_j^i(x^i) = 0$ ,  $j = 1, \cdots, m$ , no estuviera presente, (2.3) se reduciría a  $n$  problemas, cada uno planteado en el conjunto  $U^i$ . Es esto precisamente lo que sucede con el problema de Lagrange, que se divide en*

$$\inf\{\varphi^i(x^i) + \langle \lambda, c^i(x^i) \rangle : x^i \in U^i\} \quad \text{para } i = 1, \cdots, n.$$

En particular, el problema de Lagrange resulta ser “considerablemente más simple” de resolver pues cada subproblema está definido en un conjunto más simple,  $U^i$ .

**Ejemplo 2.3** (Problema de la mochila). *Tenemos una mochila y  $n$  objetos que deseamos poner en ella. Cada objeto tiene un precio  $p^i$  y un volumen  $v^i$ . Nuestro objetivo es obtener la mochila de mayor precio, sabiendo que la mochila tiene un volumen máximo  $v$ . Asignando a la variable de optimización  $x^i$  el valor 1 si el  $i$ -ésimo objeto es introducido en la mochila; 0 en caso contrario. El problema puede escribirse como:*

$$\max \sum_{i=1}^n p^i x^i \quad \text{sujeto a} \quad x^i \in \{0, 1\}, \quad \sum_{i=1}^n v^i x^i \leq v,$$

que expresado en el formato del problema primal (2.1) es

$$\begin{aligned} \min \quad & \sum_{i=1}^n -p^i x^i + 0x^0 \\ \text{s.a.} \quad & x^0 \geq 0, \quad x^i \in \{0, 1\}, \quad i = 1, \cdots, n \\ & c(x) := \sum_{i=1}^n v^i x^i + x^0 - v = 0, \end{aligned}$$

donde  $x^0$  es una variable de holgura añadida para obtener la restricción  $\sum_{i=1}^n v^i x^i - v \leq 0$  en forma de igualdad. Asociando con la restricción de igualdad un multiplicador  $\lambda \in \mathbb{R}$ , el problema de Lagrange correspondiente es

$$\inf\{\lambda x^0 - \lambda v + \sum_{i=1}^n (\lambda v^i - p^i)x^i : x^0 \geq 0, \quad x^i \in \{0, 1\}, \quad i = 1, \dots, n\}.$$

Si  $\lambda < 0$ , minimizando con respecto a  $x^0$ , obtenemos  $L(x_\lambda, \lambda) = -\infty$ . Para  $\lambda \geq 0$ , obtenemos

$$L(x_\lambda, \lambda) = \sum_{i \in I(\lambda)} (\lambda v^i - p^i) - \lambda v, \quad c(x_\lambda) = \sum_{i \in I(\lambda)} v^i - v,$$

donde

$$I(\lambda) = \{i : \lambda v^i - p^i < 0\}.$$

**Ejemplo 2.4** (Maximización de la entropía). Consideremos  $U = L_1(\Omega, \mathbb{R})$ , donde  $\Omega$  es un intervalo de  $\mathbb{R}$ . Queremos resolver el siguiente problema:

$$\begin{aligned} \inf \quad & \int_{\Omega} \psi(w, x(w)) dw \\ \text{s.a.} \quad & x \in L_1(\Omega, \mathbb{R}), \\ & \int_{\Omega} \gamma_j(w, x(w)) dw = 0, \quad j = 1, \dots, m, \end{aligned} \quad (2.4)$$

donde  $\psi, \gamma_1, \dots, \gamma_m : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$  son funciones dadas.

El lagrangiano para este problema es la integral sobre  $\Omega$  de la función

$$l(w, x, \lambda) = \psi(w, x(w)) + \sum_{j=1}^m \lambda_j \gamma_j(w, x(w)).$$

En particular si cada función  $\gamma_j(w, \cdot)$  es afín y la función  $\psi(w, \cdot)$  es estrictamente convexa, por ejemplo,

$$\gamma_j(w, x(w)) = \cos(2\pi j w)x(w) - b_j, \quad j = 1, \dots, m;$$

$$\psi(w, t) = \begin{cases} t \log t & \text{si } t > 0, \\ 0 & \text{si } t = 0, \\ +\infty & \text{si } t < 0, \end{cases}$$

entonces se cumple la Hipótesis 2.1: el problema de Lagrange tiene la solución explícita dada por

$$x_\lambda(w) = \exp\left(-1 - \sum_{j=1}^m \lambda_j \cos(2\pi j w)\right), \quad (2.5)$$

## 2.2. Teoría necesaria para resolver el problema primal

Con el objetivo de resolver el problema primal (2.1) formularemos un  $\lambda$ -problema explícito (llamado problema dual), estudiaremos la existencia de solución y algunas condiciones bajo las cuales el problema dual resuelve el problema primal.

Para resolver el problema (2.1) debemos encontrar al menos un punto factible. Esto resulta ser suficiente, gracias al siguiente resultado.

**Teorema 2.5** (Everett). *Sea  $\lambda \in \mathbb{R}^m$  fijo, supongamos que el problema de Lagrange (2.2) tiene una solución óptima  $x_\lambda \in U$ , entonces  $x_\lambda$  también es una solución óptima de*

$$\begin{aligned} \inf \quad & \varphi(x) \\ \text{s.a.} \quad & x \in U, \\ & c(x) = c(x_\lambda). \end{aligned} \tag{2.6}$$

**Prueba.** Sea  $x \in U$  arbitrario. Por definición de  $x_\lambda$ , tenemos

$$\varphi(x) + \langle \lambda, c(x) \rangle = L(x, \lambda) \geq L(x_\lambda, \lambda) = \varphi(x_\lambda) + \langle \lambda, c(x_\lambda) \rangle.$$

Si además,  $c(x) = c(x_\lambda)$ , entonces  $x$  es factible para (2.6) y  $\varphi(x) \geq \varphi(x_\lambda)$ .  $\square$

**Observación 2.6.** *El Teorema 2.5 nos dice que, una vez resuelto el problema de Lagrange (2.2) para algún  $\lambda$ , tenemos resuelto al mismo tiempo una perturbación de (2.1). En particular si  $x_\lambda$  es factible para el problema (2.1), entonces  $x_\lambda$  también es solución óptima para dicho problema.*

### 2.2.1. El problema dual

**Definición 2.7** (Función dual). *La función dual, denotada por  $\Theta$ , está definida por*

$$\Theta(\lambda) := \inf\{L(x, \lambda) : x \in U\}.$$

Se deduce inmediatamente de la definición que, para todo  $\lambda \in \mathbb{R}^m$  y todo  $x$  factible de (2.1) se cumple

$$\Theta(\lambda) \leq \varphi(x) \tag{2.7}$$

La desigualdad (2.7) es conocida con el nombre de *dualidad débil*. Así cada valor de la función dual es una cota inferior para el valor óptimo primal; luego es razonable encontrar la mayor de las cotas inferiores:

**Definición 2.8** (Problema dual). *El problema de optimización en  $\lambda$*

$$\sup\{\Theta(\lambda) : \lambda \in \mathbb{R}^m\} \tag{2.8}$$

*es llamado el problema dual asociado con  $(U, \varphi, c)$  del problema primal (2.1).*

El siguiente resultado muestra la importancia del problema dual

**Teorema 2.9** (Relación primal-dual). *Si  $\bar{\lambda}$  es una solución del problema*

$$\text{Encontrar } \lambda : \exists x \in \underset{x \in U}{\operatorname{argmin}} L(\cdot, \lambda) : c(x) = 0. \tag{2.9}$$

*Entonces  $\bar{\lambda}$  es también una solución de problema dual (2.8). Recíprocamente si  $\bar{\lambda}$  resuelve el problema dual, entonces  $\bar{\lambda}$  es solución de (2.9) si este admite solución.*

**Prueba.** Como  $\bar{\lambda}$  es solución de (2.9), entonces existe  $\bar{x} \in U$  con  $c(\bar{x}) = 0$ , tal que

$$\Theta(\bar{\lambda}) = \inf_{x \in U} L(x, \bar{\lambda}) = L(\bar{x}, \bar{\lambda}) = \varphi(\bar{x}).$$

Utilizando la desigualdad de dualidad débil (2.7) obtenemos

$$\varphi(\bar{x}) \geq \sup_{\lambda} \Theta(\lambda) \geq \Theta(\bar{\lambda}) = \varphi(\bar{x}),$$

de donde concluimos que  $\bar{\lambda}$  es solución del problema dual.

Recíprocamente si (2.9) tiene una solución  $\lambda^*$ , entonces existe algún  $x^* \in U$  tal que  $c(x^*) = 0$  y  $L(x^*, \lambda^*) = \Theta(\lambda^*) = \varphi(x^*)$ . Utilizando (2.7) tenemos que

$$\varphi(x^*) \geq \Theta(\bar{\lambda}) = \sup_{\lambda} \Theta(\lambda) \geq \Theta(\lambda^*) = \varphi(x^*),$$

por lo tanto el valor máximo de la función dual,  $\Theta(\bar{\lambda})$ , debe ser igual al valor  $\Theta(\lambda^*)$ , luego

$$L(x^*, \bar{\lambda}) = \varphi(x^*) + \langle \bar{\lambda}, c(x^*) \rangle = \varphi(x^*) = \Theta(\lambda^*) = \Theta(\bar{\lambda}) = \inf_{x \in U} L(x, \bar{\lambda}),$$

es decir,  $x^*$  factible minimiza  $L(\cdot, \bar{\lambda})$ .

□

El siguiente resultado nos dice que la existencia de una solución para (2.9) nos garantiza pleno éxito en el enfoque dual

**Corolario 2.10.** *Supongamos que (2.9) tiene solución. Entonces para cualquier solución  $\lambda$  del problema dual, el conjunto solución primal está formado por aquellos  $x$  que minimizan  $L(\cdot, \lambda)$  y que son factibles para el problema primal (2.1), es decir*

$$\operatorname{argmin}_{x \in U, c(x)=0} \varphi = \left\{ x \in \operatorname{argmin}_{x \in U} L(\cdot, \lambda) : c(x) = 0 \right\}.$$

**Prueba.** Cuando (2.9) tiene una solución, por el Teorema 2.9,  $\lambda$  maximiza  $\Theta$  si y sólo si  $\lambda$  resuelve (2.9) y existe algún  $x$  tal que  $\varphi(x) = L(x, \lambda) = \Theta(\lambda)$ .

Entonces, para cualquier  $\bar{x}$  resolviendo (2.1), tenemos  $c(\bar{x}) = 0$  y  $\varphi(\bar{x}) = \varphi(x)$ , por lo tanto  $L(\bar{x}, \lambda) = \Theta(\lambda)$ . □

**Proposición 2.11.** *La función dual  $\Theta$  es cóncava y scs en  $\mathbb{R}^m$ . Si  $\Theta$  no es idénticamente  $-\infty$ , entonces para cualquier solución óptima  $x_{\lambda}$  del problema de Lagrange (2.2), el  $-c(x_{\lambda})$  correspondiente es un subgradiente de  $-\Theta$  en  $\lambda$ .*

**Prueba.** Por definición de  $\Theta(\mu)$ . Para cada  $\lambda$  y  $\mu$  tenemos

$$\begin{aligned} \Theta(\mu) &\leq \varphi(x_{\lambda}) + \langle \mu, c(x_{\lambda}) \rangle \\ &= \varphi(x_{\lambda}) + \langle \lambda, c(x_{\lambda}) \rangle + \langle \mu - \lambda, c(x_{\lambda}) \rangle \\ &= \Theta(\lambda) + \langle \mu - \lambda, c(x_{\lambda}) \rangle, \end{aligned}$$

luego  $-c(x_{\lambda}) \in \partial(-\Theta)(\lambda)$ . Para probar la otra afirmación, por definición:

$$\mathbb{R}^m \ni \lambda \mapsto \Theta(\lambda) = \inf_{x \in U} [\varphi(x) + \langle \lambda, c(x) \rangle],$$

luego  $\Theta$  es el ínfimo de funciones afines indexadas por  $x$ , por lo tanto cóncava. □

Note que el caso  $\Theta \equiv -\infty$  es posible: basta considerar  $U = \mathbb{R}$ ,  $\varphi(x) = -x^2$ ,  $c(x) = x$ , entonces para cualquier  $\lambda$ ,  $\Theta(\lambda) = \inf_x \{-x^2 + \lambda x\} = -\infty$ . Note también que en general  $\Theta$  no es diferenciable a menos que el problema de Lagrange tenga solución única.

**Ejemplo 2.12.** Consideremos el Ejemplo 2.4. Reemplazando el valor de  $x_\lambda(w)$  obtenido en dicho ejemplo. Las funciones  $\Theta$  y  $c$  vienen dadas por

$$\Theta(\lambda) = -\langle b, \lambda \rangle - \int_{\Omega} \exp\left(-1 - \sum_{j=1}^m \lambda_j \cos(2\pi jw)\right) dw,$$

$$c_l(x_\lambda(w)) = \int_{\Omega} \cos(2\pi lw) \exp\left(-1 - \sum_{j=1}^m \lambda_j \cos(2\pi jw)\right) dw - b_l \quad \text{para } l = 1, \dots, m.$$

Luego, siendo  $\Theta$  diferenciable,  $-c(x_\lambda(w)) \in \partial(-\Theta)(\lambda) = \{\nabla - \Theta(\lambda)\}$ .

Volviendo al caso general, el problema dual posee propiedades importantes:

- Tiene sentido maximizar en  $\mathbb{R}^m$  una función cóncava scs como  $\Theta$ .
- El conjunto de maximizadores de una función cóncava, tal como  $\Theta$ , está bien definido: no existe ambigüedad entre máximo global y local. Más aún, la desigualdad de dualidad débil (2.7) nos dice que  $\Theta$  está acotado superiormente, a menos que no exista  $xd$  factible.

La aproximación dual es equivalente a explorar el siguiente conjunto primal parametrizado por  $\lambda \in \mathbb{R}^m$

$$\tilde{U} := \{x \in U : x \text{ resuelve el problema de Lagrange (2.2) para algún } \lambda\}. \quad (2.10)$$

Usualmente,  $\tilde{U}$  está contenido estrictamente en  $U$ : por ejemplo las funciones (2.5) solo describen una parte muy pequeña de  $U = L_1(\Omega, \mathbb{R})$  cuando  $\lambda$  describe  $\mathbb{R}^m$ . El interés de la dualidad reside precisamente aquí: la aproximación dual producirá una solución primal si (y sólo si)  $\tilde{U}$  contiene puntos factibles; de lo contrario esta fracasa.

**Ejemplo 2.13.** Consideremos un problema de la mochila de la forma

$$\min -x, \quad \text{sujeto a } 2x \leq 1, \quad x \in \{0, 1\}.$$

Con unos cálculos simples resulta que el óptimo dual tiene solución única  $\bar{\lambda} = 1/2$  en la cual el valor óptimo es  $\Theta(1/2) = -1/2$ , sin embargo el valor óptimo primal es 0. En  $\bar{\lambda} = 1/2$  el problema de Lagrange tiene dos soluciones (con holgura)

$$(x, x^0) = (0, 0) \quad \text{y} \quad (x, x^0) = (1, 0)$$

con valor de restricción  $-1$  y  $1$  luego ninguna de ellas es factible. Por otro lado la solución sin holgura  $x = 0$  es factible con respecto a la restricción con desigualdad.

Este ejemplo ilustra el siguiente concepto

**Definición 2.14** (Salto de dualidad). Se dice que en un problema de optimización, existe salto de dualidad si la diferencia entre los valores óptimos primal y dual,

$$\inf_{x \in U} \{\varphi(x) : c(x) = 0\} - \sup_{\lambda \in \mathbb{R}^m} \Theta(\lambda),$$

es diferente de cero.

Como consecuencias inmediatas de esta definición tenemos:

- Por la desigualdad de dualidad débil (2.7), la diferencia anterior es no negativa.
- Si existe salto de dualidad, entonces  $\tilde{U}$  no contiene ningún punto factible. En efecto si existiese un punto factible  $x_\lambda \in \tilde{U}$ , por dualidad débil obtendríamos que  $\Theta(\lambda) = \varphi(x_\lambda)$ .
- La ausencia de salto de dualidad no es suficiente para concluir que ambos problemas tienen solución: por ejemplo si el problema primal (2.1) es no acotado, entonces el dual correspondiente resulta infactible. No existe salto de dualidad pues ambos problemas tienen valor óptimo  $-\infty$  y ninguno de los problemas tiene solución.

### 2.2.2. Caracterización de optimalidad primal-dual

Supongamos que el problema dual (2.8) ha sido resuelto, con el objetivo de garantizar que existe al menos una solución para el problema primal (2.1) debemos de hacer algunas suposiciones.

Una solución dual  $\lambda$  es caracterizada por  $0 \in \partial(-\Theta)(\lambda)$ , pues  $\operatorname{argmax} \Theta = \operatorname{argmin} -\Theta$ . Consideremos el conjunto óptimo del problema de Lagrange (2.2):

$$U(\lambda) := \{x \in U : L(x, \lambda) = \Theta(\lambda)\}; \quad (2.11)$$

por la Proposición 2.11.

$$\partial(-\Theta)(\lambda) \supset \operatorname{co}\{-c(x) : x \in U(\lambda)\}.$$

Por lo tanto, la inclusión recíproca es crucial, pues nos da una descripción alternativa del conjunto  $\partial(-\Theta)$ . Consideremos la siguiente propiedad

$$\partial(-\Theta)(\lambda) = \operatorname{co}\{-c(x) : x \in U(\lambda)\}. \quad (2.12)$$

Observamos que el lado derecho de (2.12) es cerrado, pues el lado izquierdo lo es.

El siguiente resultado nos da condiciones suficientes para una descripción sencilla del conjunto  $\partial(-\Theta)$  en términos de puntos primales.

**Lema 2.15.** *Suponiendo que el conjunto  $U$  en (2.1) sea compacto, en el cual la función  $\varphi$  es semicontinua inferior y cada función  $c_j$  continua. Entonces se cumple la propiedad (2.12) para cada  $\lambda \in \mathbb{R}^m$*

**Prueba.** Por las hipótesis sobre  $\varphi$  y  $c_j$ , la función de Lagrange  $L(\cdot, \lambda) = \varphi(\cdot) + \langle \lambda, c(\cdot) \rangle$  resulta ser semicontinua inferior, luego tiene un mínimo para cada  $\lambda$  debido a que  $U$  es compacto. Así  $\operatorname{dom} \Theta = \mathbb{R}^m$ . El resultado se sigue de aplicar el Teorema 1.5 al supremo de funciones afines  $\lambda \mapsto -\varphi(x) - \langle \lambda, c(x) \rangle$ , indexadas por  $x \in U$ .

□

El principal resultado respecto a las relaciones primal-dual es el siguiente:

**Teorema 2.16.** *Supongamos que se cumple la propiedad (2.12) y sea*

$$C(\lambda) := \{c(x) \in \mathbb{R}^m : x \in U(\lambda)\},$$

la imagen por  $c$  del conjunto (2.11). Entonces  $\bar{\lambda}$  es un óptimo dual si y sólo si existen puntos  $x_1, \dots, x_k$  en  $U(\bar{\lambda})$  con  $k \leq m + 1$  y escalares no negativos  $\alpha_1, \dots, \alpha_k$ , con  $\sum_{i=1}^k \alpha_i = 1$ , tales que

$$\sum_{i=1}^k \alpha_i c(x_i) = 0.$$

En particular si  $C(\bar{\lambda})$  es convexo para algún  $\bar{\lambda}$  óptimo, entonces, para cualquier solución dual  $\lambda$ , los puntos factibles en  $U(\lambda)$  forman el conjunto solución del problema primal (2.1).

**Prueba.** Como  $\partial(-\Theta)(\bar{\lambda}) = \text{co}\{-c(x) : x \in U(\bar{\lambda})\}$ , la condición de minimalidad nos dice que  $0 \in \partial(-\Theta)(\bar{\lambda})$  si y sólo si existen  $\{x_i, \alpha_i\}$  con la propiedad requerida; es decir,  $0 \in \text{co} C(\bar{\lambda})$ . Si  $C(\bar{\lambda})$  es convexo,  $\text{co} C(\bar{\lambda}) = C(\bar{\lambda})$  y  $0$  es el valor de restricción de algún  $x$  minimizando  $L(\cdot, \bar{\lambda})$ . El resto del Teorema se sigue por el Corolario 2.10.  $\square$

**Ejemplo 2.17.** *Consideremos el problema de la mochila de la forma*

$$\min -x \quad \text{sujeto a} \quad 2x \leq 1, \quad x \in \{0, 1\}.$$

La solución óptima de este problema es  $0$ , la función de Lagrange viene dado por

$$L(x, \lambda) = -x + \langle \lambda, 2x - 1 \rangle,$$

para  $\lambda \geq 0$ ;  $-\infty$  en otro caso, luego la función dual  $\Theta$  viene dada por

$$\Theta(\lambda) = \begin{cases} -\infty & \text{si } \lambda < 0, \\ \lambda - 1 & \text{si } 0 \leq \lambda < 1/2, \\ -\lambda & \text{si } \lambda \geq 1/2; \end{cases}$$

en el óptimo dual  $\bar{\lambda} = 1/2$ ,  $C(\bar{\lambda}) = \{-1, +1\}$  no es convexo y existe salto de dualidad igual a  $1/2$ .

Los únicos casos en los cuales no existe salto de dualidad son aquellos descritos por el siguiente resultado, consecuencia del Teorema 2.16.

**Corolario 2.18.** *Supongamos que se cumple la propiedad (2.12). En cualquiera de las siguientes situaciones (i) ó (ii), no existe salto de dualidad; para cada solución dual  $\lambda^*$  (si existe), los puntos factibles en  $U(\lambda^*)$  forman el conjunto solución del problema primal (2.1).*

(i) *Para alguna solución dual  $\bar{\lambda}$ , la función de Lagrange asociada  $L(\cdot, \bar{\lambda})$  es minimizada en un único  $\bar{x}$ ; entonces  $\bar{x}$  es la única solución de (2.1).*

(ii) *En (2.1),  $U$  es convexo,  $\varphi$  convexa y  $c : U \rightarrow \mathbb{R}^m$  es afín.*

Para el Ejemplo 2.4, el problema de Lagrange tiene solución única dada por  $x_\lambda(y)$ , luego por el Corolario 2.18(i) no existe salto de dualidad. Por otro lado para el problema de la mochila del Ejemplo 2.17 existe un salto de dualidad igual a  $1/2$ , un resultado esperado pues el conjunto  $U$  no es convexo y la función dual  $\Theta$  no es diferenciable en el punto donde alcanza su máximo.

### 2.2.3. Existencia de solución dual

Para que el problema dual (2.8) tenga una solución  $\bar{\lambda} \in \mathbb{R}^m$ , la función dual  $\Theta$  debe ser acotada superiormente ; más aún, este debe alcanzar su valor supremo. El objeto de interés es la imagen del conjunto  $U$  bajo la función de restricción  $c : U \rightarrow \mathbb{R}^m$ ,

$$C(U) := \{\gamma \in \mathbb{R}^m : \gamma = c(x) \text{ para algún } x \in U\}. \quad (2.13)$$

Note que si  $0 \in C(U)$ , el problema primal resulta ser factible, en cuanto al problema dual los casos en el cual  $0 \notin C(U)$  también son de interés.

**Observación 2.19.** Denotando por  $\Gamma$  a la cápsula afín de  $C(U)$ :

$$\Gamma := \left\{ \gamma \in \mathbb{R}^m : \gamma = \sum_{i=1}^k \alpha_i c(x_i), x_i \in U, \sum_{i=1}^k \alpha_i = 1 \text{ } k = 1, 2, \dots \right\}$$

y por  $\Gamma_0$  al subespacio paralelo a  $\Gamma$ . Fijando  $x_0 \in U$ , tenemos que  $c(x) - c(x_0) \in \Gamma_0$  para todo  $x \in U$ , por lo tanto, para todo  $(\lambda, \mu) \in \mathbb{R}^m \times \Gamma_0^\perp$ ,

$$L(x, \lambda + \mu) = L(x, \lambda) + \langle \mu, c(x) \rangle = L(x, \lambda) + \langle \mu, c(x_0) \rangle,$$

tomando ínfimo sobre los  $x \in U$  en esta igualdad obtenemos

$$\Theta(\lambda + \mu) = \Theta(\lambda) + \langle \mu, c(x_0) \rangle \text{ para todo } \lambda \in \mathbb{R}^m, \mu \in \Gamma_0^\perp. \quad (2.14)$$

En otras palabras,  $\Theta$  es una función afín en el subespacio  $\Gamma_0^\perp$ .

Tenemos dos casos posibles a considerar:

- Si  $0 \notin \Gamma$  ( $0 \notin C(U)$ ), sea  $\mu_0 \neq 0$  la proyección del origen sobre  $\Gamma$ . Fijando  $\lambda$  en (2.14) y tomando  $\mu = t\mu_0$  con  $t \rightarrow +\infty$ . Debido a que  $c(x_0) \in \Gamma$  obtenemos

$$\Theta(\lambda + t\mu_0) = \Theta(\lambda) + t\langle \mu_0, c(x_0) \rangle = \Theta(\lambda) + t\|\mu_0\|^2 \rightarrow +\infty,$$

luego por la desigualdad de dualidad débil (2.7), el problema primal resulta ser infactible.

- Por lo tanto el único caso interesante es cuando  $0 \in \Gamma$ . Entonces el conjunto óptimo dual es  $\Lambda + \Gamma_0^\perp$ , donde  $\Lambda \subset \Gamma_0$  es el conjunto óptimo (posiblemente vacío) de

$$\sup\{\Theta(\lambda) : \lambda \in \Gamma_0 = \text{aff } C(U) = \text{lin } C(U)\}. \quad (2.15)$$

En cierta forma el problema dual relevante a considerar, suponiendo que  $\Gamma_0$  es conocido, es (2.15). En particular el problema dual permanecerá sin cambios si asumimos que  $\Gamma = \Gamma_0 = \mathbb{R}^m$ .

El siguiente resultado contiene las condiciones esenciales para la existencia de una solución óptima dual.

**Teorema 2.20.** Supongamos que  $\Theta$  no es idénticamente  $-\infty$ , sea  $C(U)$  definida en (2.13),

- (i) si  $0 \notin \text{clco } C(U)$ , entonces  $\sup_\lambda \Theta(\lambda) = +\infty$ ;



(ii) si  $0 \in \text{co } C(U)$ , entonces  $\sup_{\lambda} \Theta(\lambda) < +\infty$ ;

(iii) si  $0 \in \text{ri co } C(U)$ , entonces el problema dual tiene una solución.

**Prueba.**

(i) Por los teoremas de separación, el conjunto convexo cerrado  $\text{cl co } C(U)$  es separado de  $\{0\}$ , esto significa que para algún  $\mu_0 \neq 0$  y  $\delta > 0$ , se cumple

$$\langle \mu_0, c(x) \rangle \geq \delta > 0 \quad \text{para todo } x \in U.$$

Sea  $\lambda_0 < +\infty$ , escribiendo para todo  $x \in U$

$$\begin{aligned} L(x, \lambda_0 + t\mu_0) &= \varphi(x) + \langle \lambda_0, c(x) \rangle + t\langle \mu_0, c(x) \rangle \\ &\geq \Theta(\lambda_0) + t\langle \mu_0, c(x) \rangle \\ &\geq \Theta(\lambda_0) + t\delta. \end{aligned}$$

Se deduce que  $\Theta(\lambda_0 + t\mu_0) \geq \Theta(\lambda_0) - t\delta$  de donde  $\Theta(\lambda_0 + t\mu_0) \rightarrow \infty$  cuando  $t \rightarrow \infty$ .

(ii) Como  $0 \in \text{co } C(U)$ , existen un número finito de puntos  $x_1, \dots, x_p$  en  $U$  y escalares no negativos  $\alpha_1, \dots, \alpha_p$ , con  $\sum_{i=1}^p \alpha_i = 1$ , tales que  $0 = \sum_{i=1}^p \alpha_i c_i(x)$ . Entonces escribiendo para todo  $\lambda \in \mathbb{R}^m$ :

$$\Theta(\lambda) \leq \varphi(x_i) + \langle \lambda, c(x_i) \rangle \quad \text{para } i = 1, \dots, p,$$

luego

$$\alpha_i \Theta(\lambda) \leq \alpha_i \varphi(x_i) + \alpha_i \langle \lambda, c(x_i) \rangle \quad \text{para } i = 1, \dots, p,$$

de donde sumando y teniendo en cuenta que  $\sum_{i=1}^p \alpha_i = 1$ , obtenemos

$$\Theta(\lambda) \leq \sum_{i=1}^p \alpha_i \varphi(x_i) + \langle \lambda, \sum_{i=1}^p \alpha_i c(x_i) \rangle = \sum_{i=1}^p \alpha_i \varphi(x_i) \leq \max_{1 \leq i \leq p} \varphi(x_i).$$

(iii) Por el Teorema 1.4, si  $0 \in \text{ri co } C(U)$ , entonces  $0$  está en el interior relativo de algún simplex contenido en  $\text{co } C(U)$ , esto significa que existe un número finito de puntos  $x_1, \dots, x_p$  en  $U$  y  $\delta > 0$  tales que

$$B(0, \delta) \cap \Gamma_0 \subset \text{co}\{c(x_1), \dots, c(x_p)\},$$

donde  $\Gamma_0 := \text{aff } C(U)$ . Por definición de  $\Theta(\lambda)$ ,

$$\Theta(\lambda) \leq \varphi(x_i) + \langle \lambda, c(x_i) \rangle \quad \text{para } i = 1, \dots, p. \quad (2.16)$$

Si  $\Gamma_0 = \{0\}$ , es decir,  $C(U) = \{0\}$ ,  $L(x, \lambda) = \varphi(x)$ , así  $\Theta$  es una función constante diferente de  $-\infty$  por hipótesis. Si  $\Gamma_0 \neq \{0\}$ , tomando  $0 \neq \lambda \in \Gamma_0$  en (2.16) y

$$\gamma := -\delta \frac{\lambda}{\|\lambda\|} \in B[0, \delta] \cap \Gamma_0;$$

por lo tanto, este  $\gamma$  es una combinación convexa de los  $c(x_i)$ . Para cada  $i = 1, \dots, p$  multiplicando por  $\alpha_i$  la  $i$ -ésima desigualdad en (2.16), luego sumando obtenemos,

$$\Theta(\lambda) \leq \sum_{i=1}^p \alpha_i \varphi(x_i) + \langle \lambda, \sum_{i=1}^p \alpha_i c(x_i) \rangle = \sum_{i=1}^p \alpha_i \varphi(x_i) - \delta \|\lambda\| \leq \max_{1 \leq i \leq p} \varphi(x_i) - \delta \|\lambda\|.$$

De donde concluimos que  $\Theta(\lambda) \rightarrow -\infty$  si  $\lambda$  crece indefinidamente en  $\Gamma_0$ : Por la Proposición 2.11,  $\Theta$  es una función scs, luego tiene un máximo en  $\Gamma_0$  y por lo tanto en  $\mathbb{R}^m$  debido a (2.15). □

El siguiente ejemplo analiza el caso  $0 \in \text{fr } C(U)$  en el Teorema 2.20

**Ejemplo 2.21.** Consideremos un ejemplo de maximización de la entropía de la forma

$$\inf \varphi(x) \quad \text{s.a.} \quad x \in U \quad x - x_0 = 0,$$

Consideremos dos casos particulares

$$U_1 = [0, +\infty[ \quad \varphi_1 = x \log x \quad U_2 = [0, +\infty[ \quad \varphi_2 = -\log x$$

luego los conjuntos  $C(U_i)$  correspondientes son ambos convexos y vienen dados por

$$C(U_1) = [-x_0, +\infty[ \quad \text{y} \quad C(U_2) = [-x_0, +\infty[,$$

calculando las funciones duales correspondientes obtenemos

$$\Theta_1(\lambda) = -\exp^{-1-\lambda} - \lambda x_0 \quad \Theta_2(\lambda) = \log \lambda - \lambda x_0 \quad \text{para } \lambda > 0.$$

Note que los casos  $x_0 < 0$  y  $x_0 > 0$  son justificados por el Teorema 2.20. Sin embargo el para el caso  $x_0 = 0$  obtenemos que  $0 \in \text{fr } C(U_i)$  para  $i = 1, 2$  y  $\sup_\lambda \Theta_1(\lambda) = 0$ ,  $\sup_\lambda \Theta_2(\lambda) = +\infty$ .

**Ejemplo 2.22.** Consideremos una variante del problema de la mochila dada por:

$$\inf -x, \quad 2x = 1, \quad x \in \{0, 1\}.$$

Es claro que este problema es infactible, luego su valor ínfimo es  $+\infty$ . Por otro lado como  $c(x) = 2x - 1$  entonces  $C(U) = \{-1, 1\}$  y  $0 \in \text{ri co } C(U) = [-1, 1]$ , por el Teorema 2.20 (iii) existe solución dual. En efecto podemos verificar que la función dual para este ejemplo viene dada por

$$\Theta(\lambda) = \begin{cases} \lambda - 1 & \text{si } \lambda < 1/2 \\ -1/2 & \text{si } \lambda = 1/2 \\ -\lambda & \text{si } \lambda > 1/2, \end{cases}$$

de donde  $\sup_{\lambda \in \mathbb{R}} \Theta(\lambda) = \Theta(1/2) = -1/2$ .

## 2.2.4. Relación entre dualidad y función primal

**Definición 2.23** (Función primal). La función primal está definida por

$$\mathbb{R}^m \ni \gamma \mapsto p(\gamma) := \inf\{\varphi(x) : x \in U \quad c(x) = \gamma\} \quad (2.17)$$

Es decir, el lado derecho de las restricciones del problema primal (2.1) es considerado como parámetro variable  $\gamma$ .

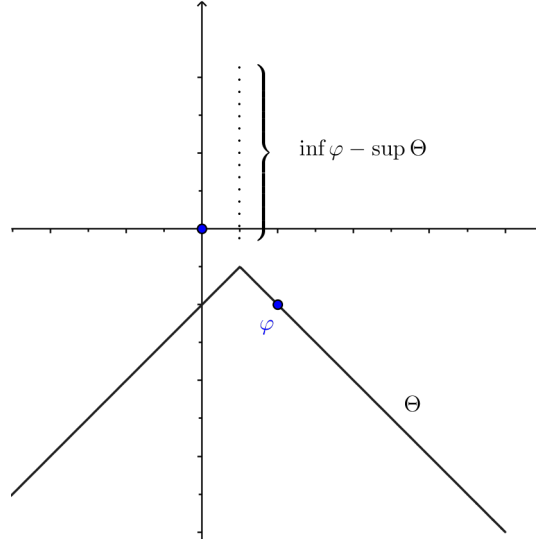


Figura 2.1: Salto de dualidad

**Teorema 2.24** (Relación entre las funciones primal y dual). *Supongamos que  $\text{dom } \Theta \neq \emptyset$ . Entonces*

$$p^*(\lambda) = -\Theta(-\lambda), \quad \text{para todo } \lambda \in \mathbb{R}^m.$$

**Prueba.** Por definición

$$\begin{aligned} p^*(\lambda) &= \sup_{\gamma} \{ \langle \lambda, \gamma \rangle - p(\gamma) \} \\ &= \sup_{\gamma} \{ \sup_{x \in U} [ \langle \lambda, \gamma \rangle - \varphi(x) : c(x) = \gamma ] \} \\ &= \sup_{\gamma} \{ \sup_{x \in U} [ \langle \lambda, c(x) \rangle - \varphi(x) ] \} \\ &= \sup_{x \in U} [ \langle \lambda, c(x) \rangle - \varphi(x) ] = -\Theta(-\lambda) \end{aligned}$$

□

Como consecuencia del Teorema 2.24 tenemos que:

$$p^{**}(\gamma) = \text{cl co } p(\gamma) = (-\Theta)^*(-\gamma) \quad \text{para todo } \gamma \in \mathbb{R}^m. \quad (2.18)$$

Esto nos da algunas conexiones entre el problema primal (2.1) y el problema dual (2.8)

(i) Cuando hay salto de dualidad, este es igual al número

$$p(0) - \sup_{\mathbb{R}^m} \Theta = p(0) - (-\Theta)^*(0) = p(0) - \text{cl co } p(0).$$

luego, no existe salto de dualidad si  $p$  es convexa y cerrada en 0.

(ii) El mínimo de la función convexa y cerrada,  $-\Theta$ , está caracterizado por

$$\lambda_0 \in \text{argmin}\{(-\Theta)(\lambda) : \lambda \in \mathbb{R}^m\} \iff 0 \in \partial(-\Theta)(\lambda_0) \iff \lambda_0 \in \partial(-\Theta)^*(0).$$

Luego

$$\text{argmax}\{\Theta(\lambda) : \lambda \in \mathbb{R}^m\} = \text{argmin}\{(-\Theta)(\lambda) : \lambda \in \mathbb{R}^m\} = \partial(-\Theta)^*(0) = -\partial(\text{cl co } p)(0).$$

Si  $0 \in \text{ri dom}(\text{cl co } p)$ , entonces  $\partial(\text{cl co } p)(0) \neq \emptyset$ , luego existe solución dual, esto explica el Teorema 2.20(iii).

(iii) El problema dual resuelve la “versión convexa cerrada” del problema primal. Por ejemplo, para simplificar, consideremos en particular los datos  $(U, \varphi, c)$ , donde  $U$  es un subconjunto acotado de  $\mathbb{R}^m$ ,  $\varphi(x) := \langle q, x \rangle$  lineal y  $c(x) := Ax - b$  afín, luego el problema primal correspondiente es

$$\begin{aligned} \inf \quad & \langle q, x \rangle \\ \text{s.a.} \quad & x \in U \\ & Ax - b = 0 \end{aligned} \tag{2.19}$$

**Ejemplo 2.25.** Consideremos un problema de la mochila de la forma

$$\min -x \quad \text{s.a.} \quad 2x \leq 1, \quad x \in \{0, 1\}$$

es claro que el valor óptimo de este problema es 0, mientras que el valor óptimo dual es  $-1/2$ , luego existe un salto de dualidad igual a  $1/2$ . Por otro lado, un cálculo de  $p$  y  $p^{**}$  en cero nuevamente nos da como resultado un salto de dualidad de  $1/2$  como lo muestra la Figura 2.2.

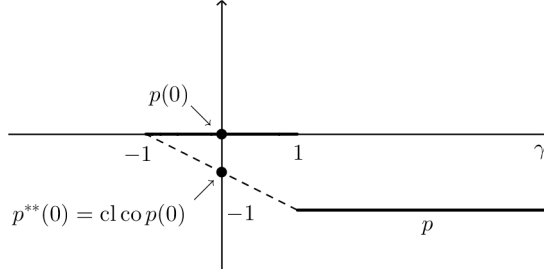


Figura 2.2: Función primal y salto de dualidad

**Proposición 2.26.** Consideremos el problema dual asociado con (2.19). Entonces su valor supremo coincide con el valor ínfimo

$$\inf \{ \langle q, x \rangle : x \in \text{cl co } U, Ax - b = 0 \}. \tag{2.20}$$

Más aún, asumiendo que el problema dual tenga alguna solución óptima  $\bar{\lambda}$ , entonces las soluciones de (2.20) son aquellas

$$x \in (\text{cl co } U)(\bar{\lambda}) = \{x \in \text{cl co } U : L(x, \bar{\lambda}) = \Theta(\bar{\lambda})\}$$

que satisfacen  $Ax = b$ .

**Prueba.** El problema primal (2.19) puede ser re-escrito utilizando la definición de función primal como

$$p(\gamma) = \inf \{ I_U(x) + \langle q, x \rangle : Ax = \gamma + b \} \quad \text{para todo } \gamma \in \mathbb{R}^m.$$

Calculando la conjugada de  $p$  obtenemos

$$\begin{aligned} p^*(\lambda) &= \sup_{\gamma} \{ \langle \lambda, \gamma \rangle - \inf_x \{ I_U(x) + \langle q, x \rangle : Ax = \gamma + b \} \} \\ &= \sup_u \{ \langle A^* \lambda - q, x \rangle - I_U(x) \} - \langle \lambda, b \rangle \\ &= I_U^*(A^* \lambda - q) - \langle \lambda, b \rangle = \sigma_U(A^* \lambda - q) - \langle \lambda, b \rangle. \end{aligned}$$

Como  $I_U^* = (\text{cl co } I_U)^* = I_{\text{cl co } U}^*$  entonces  $\sigma_U = \sigma_{\text{cl co } U}$ , luego tenemos

$$p^* + \langle \cdot, b \rangle = \sigma_{\text{cl co } U} \circ (A^*(\cdot) - q),$$

tomando conjugada en ambos miembros y aplicando el Teorema 1.6 al segundo miembro obtenemos

$$\begin{aligned} \text{cl co } p(\gamma - b) &= \text{cl inf} \{ I_{\text{cl co } U}(x) + \langle q, x \rangle : Ax = \gamma \} \\ &= \text{inf} \{ \langle q, x \rangle : x \in \text{cl co } U, Ax = \gamma \} \text{ para todo } \gamma \in \mathbb{R}^m. \end{aligned}$$

En particular para  $\gamma = b$  y teniendo en cuenta el comentario que sigue al Teorema 2.24 obtenemos  $0 = p(0) - \text{cl co } p(0) = p(0) - \sup_{\mathbb{R}^m} \Theta$ , es decir, no existe salto de dualidad para los problemas primal y dual considerados.

Para concluir la prueba, notemos que el problema primal (2.20) satisface las hipótesis del Lema 2.15 y el Corolario 2.18. □

Un ejemplo de (2.19) es la programación lineal entera. Considerando el problema primal en  $\mathbb{R}^n$  (un tipo de problema de la mochila generalizado).

$$\begin{aligned} \text{inf } & \langle c, x \rangle \\ \text{s.a. } & Ax = a \in \mathbb{R}^m, \quad Bx \leq b \in \mathbb{R}^p \\ & x^i \in \mathbb{N}, \quad x^i \leq \bar{x}, \quad i = 1, \dots, n, \end{aligned} \tag{2.21}$$

donde  $\bar{x}$  es algún entero positivo. El resultado siguiente utiliza el esquema de dualidad.

$$\mathbb{N}^n \times \mathbb{R}^m \times (\mathbb{R}^+)^p \ni (x, \lambda, \mu) \mapsto L(x, \lambda, \mu) = \langle (c + A^\top \lambda + B^\top \mu), x \rangle - \langle \lambda, a \rangle - \langle \mu, b \rangle$$

**Corolario 2.27.** *El valor óptimo dual asociado con (2.21) es el valor óptimo de*

$$\begin{aligned} \text{inf } & \langle c, x \rangle \\ \text{s.a. } & Ax = \alpha, \quad Bx \leq b \\ & 0 \leq x^i \leq \bar{x}, \quad i = 1, \dots, n. \end{aligned} \tag{2.22}$$

**Prueba.** Introducimos variables de holgura en (2.21) para escribir las restricciones de desigualdad como  $Bx + z = b$ , con  $z \geq 0$ . Como el conjunto admisible de los  $x$  es acotada entonces  $z$  también lo es, poniendo  $z^j \leq \bar{z}$  para  $j = 1, \dots, p$ , estamos en el caso (2.19) con

$$U = \{ (x, z) \in \mathbb{N}^n \times [0, \bar{z}]^p \mid z = b - b_x, \quad x^i \leq \bar{x}, \quad i = 1, \dots, n, \}$$

cuya cápsula convexa cerrada es  $[0, \bar{x}]^n \times [0, \bar{z}]^p$ , luego aplicando la Proposición 2.26 y eliminando la variable de holgura  $z$  obtenemos (2.22). □

**Observación 2.28.** *Es claro que (2.22) es un programa lineal estandar; el cual podemos dualizar mediante la relajación lagrangiana. El problema (2.22) también es llamado relajación convexa de (2.21), las restricciones enteras son relajadas para formar un problema de minimización convexa. El corolario anterior nos dice que ambas técnicas son equivalentes para programación lineal entera.*

## 2.3. Lagrangiano aumentado

Para  $r \geq 0$ , consideremos el problema de optimización

$$\begin{aligned} \inf \quad & \varphi(x) + \frac{1}{2}r\|c(x)\|^2 \\ \text{s.a.} \quad & x \in U, \\ & c(x) = 0 \in \mathbb{R}^m. \end{aligned} \tag{2.23}$$

Es claro que (2.23) es equivalente al problema primal (2.1), pues tienen el mismo conjunto factible y la misma función objetivo, sin embargo en el espacio dual la equivalencia ya no se cumple. La función de Lagrange asociada con (2.23) es

$$L_r(x, \lambda) := \varphi(x) + \frac{1}{2}r\|c(x)\|^2 + \langle \lambda, c(x) \rangle = L(x, \lambda) + \frac{1}{2}r\|c(x)\|^2,$$

llamado *Lagrangiano aumentado* asociado con (2.1). La función *dual aumentada* correspondiente es

$$\Theta_r(\lambda) := \inf_{x \in U} L_r(x, \lambda). \tag{2.24}$$

**Observación 2.29.** *Aplicando la desigualdad de dualidad débil (2.7) al par primal-dual (2.23) y (2.24) obtenemos*

$$\Theta_r(\lambda) \leq \varphi(x) + \langle \lambda, c(x) \rangle + \frac{1}{2}r\|c(x)\|^2 = \varphi(x).$$

para todo  $r \geq 0$ ,  $\lambda \in \mathbb{R}^m$  y  $x$  factible en (2.23). Por otro lado  $L_r \geq L$ , implica que  $\Theta_r \geq \Theta$ , obteniendo así

$$\inf_{x \in U, c(x)=0} \varphi(x) \geq \Theta_r \geq \Theta$$

esto nos dice que el lagrangiano aumentado no empeora el salto de dualidad. En efecto, esta aproximación resulta ser eficiente cuando la falta de convexidad de la función primal puede ser corregida por un término cuadrático:

**Teorema 2.30** (Salto de dualidad nulo). *Sea  $p$  la función primal definida en (2.17), supongamos que existen  $r_0 \geq 0$  y  $\bar{\lambda} \in \mathbb{R}^m$  tales que*

$$p(\gamma) \geq p(0) - \langle \bar{\lambda}, \gamma \rangle - \frac{1}{2}r_0\|\gamma\|^2 \quad \text{para todo } \gamma \in \mathbb{R}^m. \tag{2.25}$$

Entonces, para todo  $r \geq r_0$  el salto de dualidad asociado con lagrangiano aumentado  $L_r$  es cero y

$$p(0) = \Theta_r(\bar{\lambda}) \geq \Theta_r(\lambda) \quad \text{para todo } \lambda \in \mathbb{R}^m$$

**Prueba.** La desigualdad (2.25) también se cumple con  $r_0$  reemplazado por cualquier  $r \geq r_0$  luego para cualquier  $\gamma$  tenemos:

$$\begin{aligned} -p(0) & \geq -p(\gamma) + \langle \bar{\lambda}, \gamma \rangle + \frac{1}{2}r\|\gamma\|^2 \\ & = -\inf_{x \in U} \{ \varphi(x) + \langle \bar{\lambda}, \gamma \rangle + \frac{1}{2}r\|\gamma\|^2 : c(x) = \gamma \} \\ & = -\inf_{x \in U} \{ L_r(x, \bar{\lambda}) : c(x) = \gamma \}. \end{aligned}$$

Como  $\gamma$  era arbitrario, tenemos que

$$-p(0) \geq -\inf_{x, \gamma} \{ L_r(x, \bar{\lambda}) : c(x) = \gamma \} = -\Theta_r(\bar{\lambda}),$$

es decir  $p(0) \leq \Theta_r(\bar{\lambda})$ . Utilizando la dualidad débil (2.7), también se cumple que  $\Theta_r(\bar{\lambda}) \leq p(0)$ , concluimos así que no existe salto de dualidad y  $\bar{\lambda}$  maximiza  $\Theta_r$ .  $\square$

**Observación 2.31.** Por el Teorema 2.24,  $p^*(\lambda) = -\Theta(-\lambda)$  para todo  $\lambda \in \mathbb{R}^m$ , luego

$$p^{**}(\gamma) = (-\Theta)^*(-\gamma) := \sup_{\lambda} \{ \langle -\lambda, x \rangle - (-\Theta)(-\lambda) \}$$

por lo tanto,  $p^{**}(0) = \sup_{\lambda} \Theta(-\lambda) = \sup_{\lambda} \Theta(\lambda)$ . Por otro lado, veremos en el capítulo 3 que la desigualdad (2.25) nos da una condición necesaria y suficiente para la “exactitud del lagrangiano aumentado”.

**Ejemplo 2.32.** Consideremos el ejemplo siguiente

$$\min -x \quad \text{sujeto a } 2x \leq 1 \quad x \in \{0, 1\}$$

cuyo valor óptimo es 0 y por el Ejemplo 2.17, tenemos que el valor óptimo dual es  $-1/2$ . Considerando el lagrangiano aumentado obtenemos

$$L_r(x, v, \lambda) = -x + \lambda(2x - 1 + v) + \frac{1}{2}r(2x - 1 + v)^2,$$

minimizando respecto de  $v$  obtenemos  $v = \max\{0, -(2x - 1) - \frac{\lambda}{r}\}$ , luego

$$L_r(x, \lambda) = \begin{cases} -x + \lambda(2x - 1) + \frac{1}{2}r(2x - 1)^2 & \text{si } 2x - 1 \geq -\frac{\lambda}{r} \\ -x - \frac{\lambda^2}{2r} & \text{si } 2x - 1 \leq -\frac{\lambda}{r} \end{cases}$$

Considerando  $|\lambda| \leq r$  obtenemos

$$\Theta_r(\lambda) = \min\{-1 + \frac{1}{2}r + \lambda, -\frac{1}{2r}\lambda^2\}.$$

Si  $r \geq 2$  y  $\lambda$  próximo a 0, entonces

$$-1 + \frac{1}{2}r + \lambda \geq \lambda > \lambda^2 \geq -\lambda^2 \geq -\frac{\lambda^2}{2r},$$

por lo tanto  $\Theta_r(\lambda) = -\frac{1}{2r}\lambda^2$  cuyo valor máximo es 0, Así tenemos salto de dualidad nulo.

**Observación 2.33.** La función primal asociado con (2.23) es

$$p_r(\gamma) = \inf_{x \in U} \{ \varphi(x) + \frac{1}{2}r\|\gamma\|^2 : c(x) = \gamma \} = p(\gamma) + \frac{1}{2}r\|\gamma\|^2.$$

Esta función primal perturbada establece una conexión entre el lagrangiano aumentado y la regularización de Moreau-Yoshida. Dados  $c > 0$  y una función  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  convexa y cerrada, la regularización Moreau-Yosida de  $f$  está definida por

$$(f \uplus \frac{1}{2}c\|\cdot\|^2)(x) := \min\{f(y) + \frac{1}{2}c\|x - y\|^2 : y \in \mathbb{R}^m\}.$$

**Proposición 2.34.** Supongamos que  $p$  es una función convexa y cerrada en  $\mathbb{R}^m$ ; entonces para todo  $r > 0$

$$-\Theta_r(\lambda) = (-\Theta \uplus \frac{1}{2r}\|\cdot\|^2)(\lambda) = \min\{-\Theta(\mu) + \frac{1}{2r}\|\mu - \lambda\|^2 : \mu \in \mathbb{R}^m\}$$

**Prueba.** Aplicando el Teorema 2.24 a la función primal perturbada  $p_r$  obtenemos

$$-\Theta_r(-\lambda) = p_r^*(\lambda) = (p + \frac{1}{2}r\|\cdot\|^2)^*(\lambda),$$

siendo  $p$  y  $\|\cdot\|$  funciones convexas y cerradas

$$-\Theta_r(\lambda) = (p^* \uplus \frac{1}{2}r\|\cdot\|^2)(-\lambda) = \min\{-\Theta(-\mu) + \frac{1}{2}r\|\nu\|^2 : \mu + \nu = -\lambda\}.$$

□

Supongamos que nuestro problema inicial (2.1) sea tal que  $p + \frac{1}{2}r_0\|\cdot\|^2$  es convexa y cerrada para algún  $r_0 \geq 0$  entonces la aproximación por medio del Lagrangiano aumentado resulta ser eficiente, pues:

- Elimina el salto de dualidad, debido al Teorema 2.30
- Suaviza la función dual, pues por la Proposición 2.34 para  $r > r_0$

$$-\Theta_r = -\Theta_{r_0} \downarrow \frac{1}{2(r-r_0)} \|\cdot\|^2$$

es una regularización de Moreau-Yoshida

Sin embargo la técnica del lagrangiano aumentado termina con la Hipótesis 2.1 como lo muestra el siguiente ejemplo.

**Ejemplo 2.35.** *Consideremos el problema*

$$\min -x_1 - x_2 \quad \text{s.a.} \quad 2x_1 + x_2 = 1, \quad x_i \in \{0, 1\}, i = 1, 2.$$

la función dual aumentada viene dada por

$$\Theta_r(\lambda) = \inf\{-x_1 - x_2 + \lambda(2x_1 + x_2 - 1) + \frac{1}{2}r(2x_1 + x_2 - 1)^2 : x_i \in \{0, 1\}\}$$

en contraste la función dual (sin aumento) es

$$\begin{aligned} \Theta(\lambda) &= \inf_{x_1, x_2} \{-x_1 - x_2 + \lambda(2x_1 + x_2 - 1) + \frac{1}{2}r(2x_1 + x_2 - 1)^2 : x_i \in \{0, 1\}\} \\ &= \inf_{x_1} \{x_1(2\lambda - 1)\} + \inf_{x_2} \{x_2(\lambda - 1)\} - \lambda = \begin{cases} 2\lambda - 2 & \text{si } \lambda < 1/2 \\ -1 & \text{si } 1/2 \leq \lambda < 1 \\ -\lambda & \text{si } \lambda \geq 1 \end{cases} \end{aligned}$$

Nos centraremos ahora en estudiar la teoría general de dualidad y veremos su capacidad para eliminar el salto de dualidad en ciertos problemas no convexos.



## Dualidad no convexa

El capítulo anterior, dedicado a la dualidad lagrangiana, resulta en problemas duales convexos. De este modo, el salto de dualidad con problemas primales no convexos es inevitable. En este capítulo estudiaremos un enfoque más general, que incluye las lagrangianas estudiadas, pero además permite considerar otras funciones, cuyos problemas duales (no convexos) eliminan el salto de dualidad. Iniciamos estudiando la transformada de Legendre-Fenchel, sus consecuencias y el rol que desempeña en la dualización de problemas de optimización, en particular se estudiará la dualización en problemas convexos. Definiremos la función lagrangiana y el lagrangiano aumentado en un contexto general y abordaremos el estudio del salto de dualidad en problemas no convexos. El marco general para la teoría de dualidad explicada en este capítulo corresponde a [2, Cap XI].

Para cualquier función  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , la función  $f^*: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  definida por

$$f^*(v) := \sup_x \{ \langle v, x \rangle - f(x) \}$$

es la *conjugada* de  $f$ , mientras que la función  $f^{**} = (f^*)^*$  definida por

$$f^{**}(x) := \sup_v \{ \langle v, x \rangle - f^*(v) \}$$

es la *biconjugada* de  $f$ .

La interpretación geométrica de la función conjugada, utilizando la noción de epígrafo de una función, es la siguiente: por definición de  $f^*$ ,

$$\begin{aligned} (v, \beta) \in \text{epi } f^* &\iff \beta \geq \langle v, x \rangle - \alpha \text{ para todo } (x, \alpha) \in \text{epi } f \\ &\iff l_{v, \beta} \leq f, \text{ donde } l_{v, \beta}(x) := \langle v, x \rangle - \beta, \end{aligned}$$

de la primera equivalencia resulta

$$f^*(v) = \sup \{ \langle v, x \rangle - \alpha : (x, \alpha) \in \text{epi } f \},$$

esto nos dice que  $f^*$  es el supremo de la familia de todas las funciones afines  $\langle \cdot, x \rangle - \alpha$  para  $(x, \alpha) \in \text{epi } f$ . De manera análoga la definición de  $f^{**}$  nos dice que

$$\begin{aligned} (x, \alpha) \in \text{epi } f^{**} &\iff \alpha \geq \langle x, v \rangle - \beta \text{ para todo } (v, \beta) \in \text{epi } f^* \\ &\iff \alpha \geq \langle x, v \rangle - \beta \text{ para todo } l_{v, \beta} \leq f, \end{aligned}$$

de donde obtenemos

$$f^{**} := \sup\{\langle x, v \rangle - \beta : l_{v,\beta} \leq f\}, \quad (3.1)$$

esto significa que  $f^{**}$  es el supremo de todas las funciones afines  $l_{v,\beta}$  mayoradas por  $f$ .

**Teorema 3.1** (Legendre-Fenchel). *Sea  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función tal que  $\text{co } f$  es propia, entonces  $f^*$  y  $f^{**}$  son propias, sci y convexas, además*

$$f^{**} = \text{cl co } f.$$

*Si  $f$  es propia, sci y convexa, tenemos que  $f = f^{**}$ . Independientemente de las hipótesis, siempre se cumple que:*

$$f^* = (\text{co } f)^* = (\text{cl } f)^* = (\text{cl co } f)^*.$$

**Prueba.** Siendo  $f^*$  y  $f^{**}$  supremo de funciones afines, estas resultan ser convexas y sci. más aún  $\text{co } f$  propia implica que  $f^*$  sea propia. Por otro lado, como  $\text{co } f$  es propia entonces por el Teorema 1.1  $\text{cl co } f$  resulta ser propia (sci y convexa); luego por el Teorema 1.1 obtenemos que  $f^{**} = \text{cl co } f$ .

Cuando  $f$  es propia convexa y sci, entonces  $\text{co } f = f$  es propia y  $\text{cl co } f = f$  luego  $f^{**} = f$ . La última parte del teorema se sigue de la interpretación de  $f^*$  y de las siguientes equivalencias

$$\begin{aligned} f^*(v) \leq \beta &\iff \langle v, x \rangle - \alpha \leq \beta \quad \text{para todo } (x, \alpha) \in \text{epi } f \\ &\iff \langle v, x \rangle - \alpha \leq \beta \quad \text{para todo } (x, \alpha) \in \text{epi co } f \\ &\iff \langle v, x \rangle - \alpha \leq \beta \quad \text{para todo } (x, \alpha) \in \text{epi cl } f \\ &\iff \langle v, x \rangle - \alpha \leq \beta \quad \text{para todo } (x, \alpha) \in \text{epi cl co } f \end{aligned}$$

□

**Ejemplo 3.2.** *La condición  $\text{co } f$  propia no es superflua, por ejemplo para la función  $f: \mathbb{R} \rightarrow \overline{\mathbb{R}}$  definida por*

$$f(x) = \begin{cases} \log x & \text{si } x \in (0, +\infty) \\ 0 & \text{si } x = 0 \\ +\infty & \text{en otro caso.} \end{cases}$$

*tenemos que  $\text{co } f(x) = -\infty$  si  $x \in [0, +\infty)$ ;  $+\infty$  en otro caso, así  $f^* \equiv +\infty$  y  $f^{**} \equiv -\infty$ ; mientras que  $\text{cl co } f(x) = -\infty$  en el caso que  $x \in [0, \infty)$  y  $+\infty$  en otro caso. Por lo tanto  $f^{**} \neq \text{cl co } f$ .*

Por definición de la conjugada de una función, si  $f_1 \leq f_2$  entonces  $f_1^* \geq f_2^*$ . El hecho que  $f = f^{**}$  cuando  $f$  es propia, lsc y convexa significa que: si  $f^*$  es la conjugada de  $f$  entonces  $f$  es la conjugada de  $f^*$ , lo que denotaremos por  $f \xleftrightarrow{*} f^*$ ; es decir

$$\begin{aligned} f^*(v) &= \sup_x \{\langle v, x \rangle - f(x)\}, \\ f(x) &= \sup_v \{\langle v, x \rangle - f^*(v)\}. \end{aligned}$$

Las siguientes relaciones se siguen inmediatamente por definición de conjugación. En cada caso la expresión del lado izquierdo es una función de  $x$  mientras la del lado derecho es función de  $v$ . Bajo la hipótesis que  $f$  sea propia, convexa y sci.

$$\begin{aligned} f(x) - \langle a, x \rangle &\xleftrightarrow{*} f^*(v + a), \\ f(x) + c &\xleftrightarrow{*} f^*(v) - c, \\ \alpha f(x) &\xleftrightarrow{*} \alpha f^*(\alpha^{-1}v) \text{ para } \alpha > 0. \end{aligned}$$

**Proposición 3.3** (Relación de subgradientes). *Sea  $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función propia, sci y convexa, entonces*

$$\bar{v} \in \partial f(\bar{x}) \iff \bar{x} \in \partial f^*(\bar{v}) \iff f(\bar{x}) + f^*(\bar{v}) = \langle \bar{x}, \bar{v} \rangle$$

además

$$\partial f(\bar{x}) = \operatorname{argmax}_v \{ \langle v, \bar{x} \rangle - f^*(v) \}, \quad \partial f^*(\bar{v}) = \operatorname{argmax}_x \{ \langle \bar{v}, x \rangle - f(x) \}.$$

**Prueba.** El resultado se sigue utilizando el hecho que una función afín  $l_{\bar{v}, \bar{\beta}}(\cdot) := \langle \bar{v}, \cdot \rangle - \bar{\beta}$  soporta  $f$  en  $\bar{x}$ , con  $\bar{\alpha} = f(\bar{x})$  si y sólo si la función afín  $l_{\bar{x}, \bar{\alpha}}(\cdot) := \langle \bar{x}, \cdot \rangle - \bar{\alpha}$  soporta  $f^*$  en  $\bar{v}$ , con  $\bar{\beta} = f^*(\bar{v})$ .  $\square$

**Ejemplo 3.4** (Funciones soporte y cono polar).

Para cualquier  $C \subseteq \mathbb{R}^n$ , la conjugada de la función indicatriz  $I_C$  es la función soporte  $\sigma_C$ , En efecto

$$I_C^*(v) = \sup_x \{ \langle v, x \rangle - I_C(x) \} = \sup_{x \in C} \langle v, x \rangle = \sigma(v),$$

Si  $K \subseteq \mathbb{R}^n$  es un cono  $I_K^* = I_{K^*}$ , donde  $K^*$  es un cono polar de  $K$  definido por:

$$K^* = \{ v : \langle v, x \rangle \leq 0 \quad \forall x \in K \}$$

**Proposición 3.5** (Conjugación en espacios producto). *Para funciones  $f_i$  definidas en  $\mathbb{R}^{n_i}$ , la función conjugada de  $f(x_1, \dots, x_m) = f_1(x_1) + \dots + f_m(x_m)$  es*

$$f^*(v_1, \dots, v_m) = f_1^*(v_1) + \dots + f_m^*(v_m).$$

**Prueba.**

$$f^*(v_1, \dots, v_m) = \sup_x \{ \langle v, x \rangle - f(x) \} = \sum_{i=1}^m \sup_{x_i} \{ \langle v_i, x_i \rangle - f_i(x_i) \} = \sum_{i=1}^m f_i^*(v_i)$$

$\square$

**Ejemplo 3.6.** *En el problema de la mochila, consideremos una función objetivo de la forma  $\varphi(x_1, x_2) = x_1 + 2x_2$ , con  $U_1 = U_2 = \{0, 1\}$ . Entonces por la Proposición 3.5 la conjugada de  $\varphi$  viene dada por*

$$\begin{aligned} \varphi^*(v_1, v_2) &= \sup_{x_1 \in U_1} \{ (v_1 - 1)x_1 \} + \sup_{x_2 \in U_2} \{ (v_2 - 2)x_2 \} \\ &= \begin{cases} v_1 - 1 & \text{si } v_1 > 1 \\ 0 & \text{si } v_1 \leq 1 \end{cases} + \begin{cases} v_2 - 2 & \text{si } v_2 > 2 \\ 0 & \text{si } v_2 \leq 2 \end{cases} \\ &= \begin{cases} v_1 + v_2 - 3 & \text{si } v_1 > 1, v_2 > 2 \\ v_2 - 2 & \text{si } v_1 \leq 1, v_2 > 2 \\ v_1 - 1 & \text{si } v_1 > 1, v_2 \leq 2 \\ 0 & \text{si } v_1 \leq 1, v_2 \leq 2 \end{cases} \end{aligned}$$

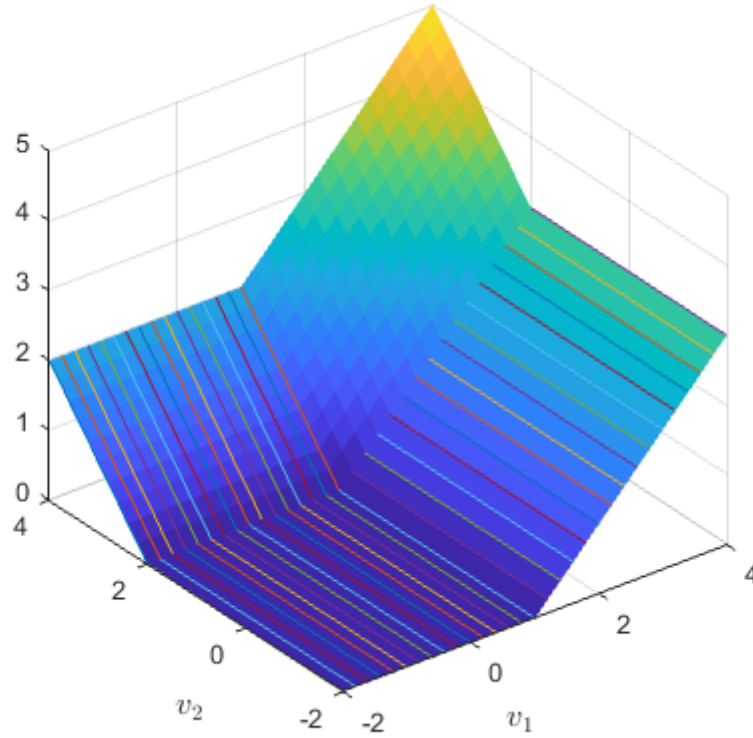


Figura 3.1: Conjugada en espacios producto

El siguiente resultado muestra la dualidad que existe entre  $f$  y  $f^*$  cuando  $f$  representa un problema de optimización.

**Teorema 3.7** (Propiedades duales en minimización).

- (a)  $\inf f = -f^*(0)$  y  $\operatorname{argmin} f = \partial f^*(0)$ .
- (b)  $\operatorname{argmin} f = \{\bar{x}\}$  si y sólo si  $f^*$  es diferenciable en 0 con  $\nabla f^*(0) = \bar{x}$ .
- (c)  $f$  es nivel-coerciva (ó nivel-acotada) si y sólo si  $0 \in \operatorname{int}(\operatorname{dom} f^*)$

**Prueba.** La primera propiedad en (a) se sigue por definición, pues

$$f^*(0) = \sup_x \{\langle 0, x \rangle - f(x)\} = -\inf f,$$

mientras que la segunda es consecuencia de la Proposición 3.3 y el hecho que  $\operatorname{argmin} f$  consiste de los puntos  $x$  tal que  $0 \in \partial f(x)$ .

Para la parte restante del teorema ver [2, Teorema 11.8] □

**Teorema 3.8** (Operaciones duales). *Se cumplen las siguientes propiedades*

- (a) **(Adición / epi-adición)** Sean  $f_i$  funciones propias, si  $f = f_1 \downarrow f_2$ , entonces  $f^* = f_1^* + f_2^*$ . Dualmente si,  $f = f_1 + f_2$  siendo  $f_1, f_2$  propias, sci y convexas tales que  $\operatorname{dom} f_1$  intersecta al  $\operatorname{dom} f_2$ , entonces  $f^* = \operatorname{cl}(f_1^* \downarrow f_2^*)$ . La operación clausura puede omitirse cuando  $0 \in \operatorname{int}(\operatorname{dom} f_1 - \operatorname{dom} f_2)$ .

- (b) **(Composición / epi-composición)**. Si  $g = Af$  para  $f: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  y  $A \in \mathbb{R}^{m \times n}$ , donde  $(Af)(u) := \inf\{f(x) : Ax = u\}$ , entonces  $g^* = f^*A^*$ , con  $(f^*A^*)(y) := f^*(A^*y)$  y  $(A^* = A^t)$ . Dualmente si  $f = gA$  para una función  $g: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  propia, sci y convexa tal que el subespacio  $\text{ran } A$  intersecta al  $\text{dom } g$ , entonces  $f^* = \text{cl}(A^*g^*)$ . La operación clausura puede omitirse cuando  $0 \in \text{int}(\text{dom } g - \text{ran } A)$ .
- (c) **(Restricción / proyección infimal)**. Si  $p(u) = \inf_x f(x, u)$ , donde  $f: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  es una función arbitraria, entonces  $p^*(\lambda) = f^*(0, \lambda)$ . Dualmente, si  $f$  es propia, sci y convexa, entonces la función  $\varphi$  definida por  $\varphi(x) = f(x, \bar{u})$  para algún  $\bar{u} \in U := \{u : \exists x, f(x, u) < \infty\}$  cumple  $\varphi^* = \text{cl } q$ , donde  $q(v) = \inf_\lambda \{f^*(v, \lambda) - \langle x, \bar{u} \rangle\}$ . Aquí la operación de clausura es superflua si  $\bar{u} \in \text{int } U$ .
- (d) **(supremo / ínfimo puntual)**. Para una familia de funciones  $\{f_i\}_{i \in I}$ , si  $f = \inf_{i \in I} f_i$ , entonces  $f^* = \sup_{i \in I} f_i^*$ . Dualmente si  $f = \sup_{i \in I} f_i$ , con  $f_i$  propia, sci y convexa, si además  $f$  es propia, entonces  $f^* = \text{cl co } \inf_{i \in I} f_i^*$ .

**Prueba.** La primera relación en (a) es consecuencia de la definición de  $f^*$  y  $f_1 \downarrow f_2$ :

$$\begin{aligned} f^*(v) &= \sup_x \{\langle v, x \rangle - f(x)\} = \sup_x \sup_y \{\langle v, x \rangle - f_1(y) + f_2(x - y)\} \\ &= \sup_z \sup_y \{\langle v, y \rangle - f_1(y) + \langle v, y \rangle - f_2(z)\} = f_1^*(v) + f_2^*(v). \end{aligned}$$

Considerando  $f_i^*$  en lugar de  $f_i$ , la relación anterior implica que  $(f_1^* \downarrow f_2^*)^* = f_1^{**} + f_2^{**} = f_1 + f_2$ ; la última igualdad se debe a que  $f_1, f_2$  son propias sci y convexas. Si  $\text{dom } f_1 \cap \text{dom } f_2 \neq \emptyset$ , entonces  $f = f_1 + f_2$  es propia, luego por el Teorema 3.1 concluimos que  $f^* = (f_1^* \downarrow f_2^*)^{**} = \text{cl co}(f_1^* \downarrow f_2^*) = \text{cl}(f_1^* \downarrow f_2^*)$ , la última igualdad se debe a que la convexidad de los  $f_i^*$  implica la convexidad de  $f_1^* \downarrow f_2^*$ .

Para (b), por definición de función conjugada obtenemos

$$\begin{aligned} g^*(v) &= \sup_u \{\langle v, u \rangle - \inf\{f(x) : Ax = u\}\} \\ &= \sup_u \sup_x \{\langle v, u \rangle - f(x) : Ax = u\} \\ &= \sup_x \{\langle A^*v, x \rangle - f(x)\} = (f^*A^*)(v), \end{aligned}$$

de esto, considerando  $A^*$  en lugar de  $A$  y  $g^*$  en lugar de  $f$ , obtenemos  $(A^*g^*)^* = g^{**}A^{**} = gA$  cuando  $g$  es propia, sci y convexa. Si  $\text{ran } A \cap \text{dom } g \neq \emptyset$ , entonces  $gA$  es propia, luego por el Teorema 3.1 obtenemos que  $(gA)^* = (A^*g^*)^{**} = \text{cl co } A^*g^* = \text{cl } A^*g^*$ . La última igualdad se debe a que la convexidad de  $g^*$  implica la convexidad de  $A^*g^*$ .

Análogamente en (c) las definiciones de  $p$  y  $p^*$  nos da

$$\begin{aligned} p^*(\lambda) &= \sup_u \{\langle \lambda, u \rangle - \inf_x f(x, u)\} \\ &= \sup_{x, u} \{\langle (0, \lambda), (x, u) \rangle - f(x, u)\} = f^*(0, \lambda). \end{aligned}$$

Mediante un cálculo similar obtenemos que  $q^*(x) = f^{**}(x, \bar{u})$ . Cuando  $f$  es propia, sci y convexa obtenemos que  $f^{**} = f$ , así  $q^* = \varphi$ . Por otro lado, la convexidad de  $f^*$  implica la convexidad de  $q$ . Si  $\bar{u} \in U$  entonces  $\varphi$  es propia. Por el Teorema 3.1 concluimos que  $\varphi^* = q^{**} = \text{cl co } q = \text{cl } q$ .

La primera relación es consecuencia inmediata de la definición, de donde, considerando  $f_i^*$  en lugar de  $f_i$  obtenemos  $(\inf_{i \in I} f_i^*)^* = \sup_{i \in I} f_i^{**}$ ; cuando  $f = \sup_{i \in I} f_i$  con  $f_i$  propia, sci y convexa, entonces por el Teorema 3.1 tenemos  $f_i = f_i^{**}$ . Si además  $f$  es propia y convexa, obtenemos que  $f^* = (\inf_{i \in I} f_i^*)^{**} = \text{cl co}(\inf_{i \in I} f_i^*)$ .

Para la demostración de las condiciones suficientes para omitir la operación de clausura, ver [2, Teorema 11.23].  $\square$

### 3.1. Problemas duales en optimización

En esta sección estudiaremos la dualidad en el sentido de Fenchel para problemas de optimización. El Teorema 3.8 (c), nos da una forma alternativa de calcular la conjugada de una función dada en un punto dado por minimización. Esta idea es capturada en el siguiente resultado.

**Lema 3.9** (Cálculo dual en optimización paramétrica). *Sea  $f: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  una función, definamos  $p(u) := \inf_x f(x, u)$ ,  $\varphi(x) := f(x, 0)$  y  $\psi(\lambda) := -f^*(0, \lambda)$ , entonces*

$$p(0) = \inf_x \varphi(x) \quad \text{y} \quad p^{**}(0) = \sup_\lambda \psi(\lambda).$$

**Prueba.** Por definición,  $p(0) = \inf_x f(x, 0) = \inf_x \varphi(x)$ . Por otro lado,

$$p^*(\lambda) = \sup_u \{ \langle \lambda, u \rangle - p(u) \} = \sup_{x,u} \{ \langle (0, \lambda), (x, u) \rangle - f(x, u) \} = f^*(0, \lambda),$$

de donde  $p^{**}(0) = \sup_\lambda \{ \langle 0, \lambda \rangle - p^*(\lambda) \} = \sup_\lambda \{ -f^*(0, \lambda) \} = \sup_\lambda \psi(\lambda)$ . □

**Observación 3.10.** *Las condiciones bajo las cuales  $p(0) = p^{**}(0)$  y por lo tanto  $\inf \varphi = \sup \psi$  son de gran interés en el sentido que dichas condiciones nos garantizarán salto de dualidad nulo y en general están gobernadas por el Teorema 3.1.*

El siguiente resultado, consecuencia del Lema 3.9, distingue las principales características de los problemas de optimización de tipo convexo.

**Teorema 3.11** (Problemas duales en optimización convexa). *Sea  $f: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  propia, sci y convexa. Consideremos el problema primal*

$$\min \varphi(x) \quad \text{en} \quad \mathbb{R}^m, \quad \varphi(x) := f(x, 0),$$

*junto con el problema dual*

$$\max \psi(\lambda) \quad \text{en} \quad \mathbb{R}^n, \quad \psi(\lambda) := -f^*(0, \lambda).$$

*Sean  $p(u) = \inf_x f(x, u)$  con  $U = \text{dom } p$  y  $q(v) = \inf_\lambda f^*(v, \lambda)$  con  $V = \text{dom } q$ , entonces estos conjuntos y funciones son convexas. Además,*

- (a) *Siempre se verifica que  $\inf_x \varphi(x) \geq \sup_\lambda \psi(\lambda)$ . Por otro lado  $\inf_x \varphi(x) < \infty$  si y sólo si  $0 \in U$ , mientras que  $\sup_\lambda \psi(\lambda) > -\infty$  si y sólo si  $0 \in V$ . Además si  $0 \in \text{int } U$  ó  $0 \in \text{int } V$ , entonces*

$$\inf_x \varphi(x) = \sup_\lambda \psi(\lambda).$$

- (b) *El conjunto  $\text{argmax}_\lambda \psi(\lambda)$  es no vacío y acotado si y sólo si  $0 \in \text{int } U$  y el valor  $\inf_x \varphi(x) = p(0)$  es finito, en cuyo caso  $\text{argmax}_\lambda \psi(\lambda) = \partial p(0)$*
- (c) *El conjunto  $\text{argmin}_x \varphi(x)$  es no vacío y acotado si y sólo si  $0 \in \text{int } V$  y el valor  $\sup_\lambda \psi(\lambda) = -q(0)$  es finito, en cuyo caso  $\text{argmin}_x \varphi(x) = \partial q(0)$ .*
- (d) *Las soluciones óptimas de los problemas primal y dual están caracterizadas por:*

$$\left. \begin{array}{l} \bar{x} \in \text{argmin}_x \varphi(x) \\ \bar{\lambda} \in \text{argmax}_\lambda \psi(\lambda) \\ \inf_x \varphi(x) = \sup_\lambda \psi(\lambda) \end{array} \right\} \iff (0, \bar{\lambda}) \in \partial f(\bar{x}, 0) \iff (\bar{x}, 0) \in \partial f^*(0, \bar{\lambda})$$

**Prueba.** La convexidad de los conjuntos y las funciones en el preámbulo es consecuencia directa de la convexidad de  $f$  y  $f^*$ . Para probar (a), por definición de conjugada  $f^*(0, \lambda) = \sup_x \{ \langle (x, 0), (0, \lambda) \rangle - f(x, 0) \} \geq -f(x, 0)$  para todo  $x \in \mathbb{R}^m$ , de donde obtenemos que  $\inf_x \varphi(x) \geq \sup_\lambda \psi(\lambda)$ . Por otro lado tenemos que

$$p(0) = \inf_x \varphi(x) < \infty \iff 0 \in \text{dom } p = U.$$

$$q(0) = \inf_y f^*(0, \lambda) > -\infty \iff -q(0) = \sup_\lambda \psi(\lambda) < +\infty \iff 0 \in \text{dom } -q = V.$$

Finalmente si  $0 \in \text{int } U$ , entonces por el Teorema 3.8(c), con  $\varphi(x) = f(x, 0)$ , tenemos que  $\varphi^*(v) = \inf_y \{ f^*(v, y) \}$ , de donde, considerando  $v = 0$ , obtenemos que  $\inf_x \varphi(x) = \sup_\lambda \psi(\lambda)$ .

Análogamente, si  $v \in \text{int } V$ , por el Teorema 3.8 con  $-\psi(y) = f^*(0, y)$  y  $p(u) = \inf_x f(x, u)$ , tenemos que  $(-\psi)^* = p$ , de donde, considerando  $u = 0$  obtenemos  $\sup_\lambda \psi(\lambda) = \inf_x \varphi(x)$ .

Para la parte (b), Por el Teorema 3.7(c) tenemos que  $0 \in \text{int } U$  y  $p(0) := \inf_x \varphi(x) > -\infty$  si y sólo si  $-\psi$  es propia y de nivel acotado pues

$$p^*(\lambda) = \sup_u \{ \langle \lambda, u \rangle - p(u) \} = \sup_{x,u} \{ \langle (0, \lambda), (x, u) \rangle - f(x, u) \} = f^*(0, \lambda) = -\psi(\lambda).$$

Pero por el Teorema 1.2, la función propia, convexa y sci  $-\psi$  es de nivel acotado si y sólo si su conjunto  $\text{argmin } -\psi$  es no vacío y acotado, entonces también, por el Teorema 3.7(a),  $\partial p(0) = \text{argmin } p^* = \text{argmax } \psi$ .

La prueba de (c) es análoga a (b), reemplazando  $f$  por  $f^*$  y usando la relación  $f = f^{**}$ . Luego  $-\sup \psi = q(0) \geq q^{**} = -\inf \varphi$  con  $\varphi = q^*$ , en particular  $0 \in \text{int } V$  implica que  $q(0) = q^{**}(0)$ . Así  $0 \in \text{int } V$  y  $q(0) > -\infty$  corresponde, por argumentos paralelos a (c), a que  $\text{argmin } \varphi$  sea no vacío y acotado. Esto también nos da  $\partial q(0) = \text{argmin } q^* = \text{argmin } \varphi$ .

Finalmente en (d), por el Teorema 3.3,  $(0, \bar{\lambda}) \in \partial f(\bar{x}, 0)$  y  $(\bar{x}, 0) \in \partial f^*(0, \bar{\lambda})$  si y sólo si  $(\bar{x}, 0) \in \partial f^*(0, \bar{\lambda})$  si y sólo si  $\varphi(\bar{x}) = \psi(\bar{\lambda})$ . Por (a),  $\inf \varphi \geq \sup \psi$ , luego también es equivalente a tener  $\varphi(\bar{x}) = \inf \varphi = \sup \psi = \psi(\bar{\lambda})$ .  $\square$

**Corolario 3.12.** *En el contexto del Teorema 3.11, las siguientes condiciones son equivalentes y sirven para garantizar que  $-\infty < \min \varphi = \max \psi < \infty$ .*

- (a)  $0 \in \text{int } U$  y  $0 \in \text{int } V$ ;
- (b)  $0 \in \text{int } U$ ,  $\text{argmin } \varphi$  es no vacío y convexo;
- (c)  $0 \in \text{int } V$ ,  $\text{argmax } \psi$  es no vacío y acotado;
- (d)  $\text{argmin } \varphi$  y  $\text{argmax } \psi$  es no vacío y acotado

**Ejemplo 3.13.** *Consideremos la siguiente relajación convexa del problema de la mochila:*

$$\min -x \quad \text{s.a. } x \in [0, 1], \quad 2x \leq 1$$

este problema es equivalente a  $\min_{x \in \mathbb{R}} \varphi(x)$ , con  $\varphi(x) = f(x, 0)$  y  $f$  dado por

$$f(x, u) = \begin{cases} -x + I_{[0,1]}(x) & \text{si } 2x - u \leq 1 \\ +\infty & \text{en otro caso.} \end{cases}$$

Luego el problema dual de Fenchel es  $\max_{\lambda \in \mathbb{R}} \psi(\lambda)$ , donde

$$\begin{aligned} \psi(\lambda) &= -f^*(0, \lambda) = -\sup_{x,u} \{ \langle (0, \lambda), (x, u) \rangle - f(x, u) \} \\ &= -\sup_{x,u} \{ \langle \lambda, u \rangle + x - I_{[0,1]}(x) : 2x + u + v = 1, v \geq 0 \} \\ &= \inf_{x \in [0,1]} \{ -x + \langle \lambda, 2x - 1 + v \rangle : v \geq 0 \} \\ &= \begin{cases} -\infty & \text{si } \lambda < 0 \\ \lambda - 1 & \text{si } 0 \leq \lambda < 1/2 \\ -\lambda & \text{si } 1/2 \leq \lambda. \end{cases} \end{aligned}$$

Observe que el dual de Fenchel coincide con el dual lagrangiano estudiado en el Capítulo 2 (ver Ejemplo 2.17). Por otro lado, calculando la función  $p$  definida en el Teorema 3.11 obtenemos

$$\begin{aligned} p(u) &= \inf_x f(x, u) = \inf \{ -x : x \in [0, 1], 2x - 1 + v = u, v \geq 0 \} \\ &= \begin{cases} +\infty & \text{si } u < -1, \\ -(u+1)/2 & \text{si } -1 \leq u < 1, \\ -1 & \text{si } 1 < u. \end{cases} \end{aligned}$$

Note que el valor óptimo primal y dual coinciden y es igual a  $-1/2$ , luego no existe salto de dualidad. Este resultado es justificado por el Teorema 3.11(a), pues  $0 \in \text{int dom } p$

Note también que  $\partial p(0) = \{1/2\} = \text{argmax}_{\lambda} \psi(\lambda)$ , resultado que es justificado por el Teorema 3.11(b), pues  $p(0) = 0$  es finito.

El hecho que los valores óptimos primal y dual sean finitos es justificado por el Corolario 3.12(c), pues  $0 \in \text{int dom } p$  y  $\text{argmax } \psi = \{1/2\}$ .

**Ejemplo 3.14** (Esquema de dualidad tipo Fenchel). *Para el siguiente problema de minimización:*

$$\min \varphi \text{ en } \mathbb{R}^n, \quad \varphi(x) := \langle c, x \rangle + k(x) + h(b - Ax),$$

donde  $k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  y  $h : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  son propias, sci y convexas;  $A \in \mathbb{R}^{m \times n}$  y  $c \in \mathbb{R}^n$ .

Este problema encaja en el formato del Teorema 3.11 definiendo  $f$  como:

$$f(x, u) := \langle c, x \rangle + k(x) + h(b - Ax + u),$$

con lo cual obtenemos el problema dual

$$\max \psi \text{ en } \mathbb{R}^m, \quad \psi(\lambda) := \langle b, \lambda \rangle - h^*(\lambda) - k(A^* \lambda - c).$$

En efecto,

$$\begin{aligned} \psi(\lambda) &= -f^*(0, \lambda) = -\sup_{x,u} \{ \langle \lambda, u \rangle - \langle c, x \rangle - k(x) - h(b - Ax + u) \} \\ &= -\sup_{x,w} \{ \langle \lambda, w - b + Ax \rangle - \langle c, x \rangle - k(x) - h(w) \} \\ &= -\sup_x \{ \langle A^* \lambda - c, x \rangle - k(x) \} + \sup_w \{ \langle \lambda, w \rangle - h(w) \} - \langle y, b \rangle \\ &= \langle b, \lambda \rangle - h^*(\lambda) - k^*(A^* \lambda - c). \end{aligned}$$



Por otro lado, considerando  $p$  y  $q$  definidos como en el Teorema 3.11, obtenemos

$$\begin{aligned}
u \in \text{dom } p &\iff \inf_x f(x, u) < \infty \\
&\iff \exists x_0 : f(x_0, u) < \infty \\
&\iff \exists x_0 : \langle c, x_0 \rangle + k(x_0) + h(b - Ax_0 + u) \\
&\iff x_0 \in \text{dom } k, b - Ax_0 + u \in \text{dom } h \\
&\iff u \in \text{dom } h - b + A(\text{dom } k),
\end{aligned}$$

luego,  $0 \in \text{int dom } p$  si y sólo si  $b \in \text{int}(A(\text{dom } k) + \text{dom } h)$ . De manera análoga obtenemos que  $0 \in \text{int dom } q$  si y sólo si  $c \in \text{int}(A^* \text{dom } h^* - \text{dom } k^*)$ . Más aún, las soluciones óptimas están caracterizadas por

$$\left. \begin{array}{l} \bar{x} \in \text{argmin } \varphi \\ \bar{\lambda} \in \text{argmax } \psi \\ \inf \varphi = \sup \psi \end{array} \right\} \iff \begin{array}{l} A^* \bar{\lambda} - c \in \partial k(\bar{x}), \\ \bar{\lambda} \in \partial h(b - A\bar{x}) \end{array} \iff \begin{array}{l} \bar{x} \in \partial k^*(A^* \bar{\lambda} - c), \\ b - A\bar{x} \in \partial h^*(\bar{\lambda}). \end{array}$$

En efecto, escribiendo  $f = g \circ F$ , con

$$g(x, w) = \langle c, x \rangle + k(x) + h(b + w) \quad y \quad F(x, u) = (x, -Ax + u),$$

note que  $F$  es lineal y no singular. Como  $\partial g(x, w) = \{(c + v, \lambda) : v \in \partial k(x), \lambda \in \partial h(b + w)\}$ , aplicando la regla de la cadena para subdiferenciales obtenemos

$$\begin{aligned}
\partial f(x, u) &= \partial(g \circ F)(x, u) = \nabla F(x, u)^* \partial g(F(x, u)) \\
&= \{(c + v - A^* \lambda, \lambda) : v \in \partial k(x), \lambda \in \partial h(b - Ax + u)\}.
\end{aligned}$$

La condición  $(0, \bar{\lambda}) \in \partial f(\bar{x}, 0)$  en el Teorema 3.11(d) se reduce a tener  $A^* \bar{\lambda} - c \in \partial k(\bar{x})$  para  $\bar{\lambda} \in \partial h(b - A\bar{x})$ . La expresión de optimalidad en términos de  $\partial k^*$  y  $\partial h^*$  es consecuencia inmediata de la Proposición 3.3.

## 3.2. Función Lagrangiano

La dualidad en el contexto del Teorema 3.11 caracteriza solamente problemas de optimización del tipo convexo. En general, para un problema de optimización representado como minimizar  $\varphi(\cdot) = f(\cdot, 0)$  sobre  $\mathbb{R}^n$  para una función  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  independiente de la convexidad, decimos que el problema de maximizar  $\psi(\cdot) = -f^*(0, \cdot)$  sobre  $\mathbb{R}^m$  es el problema dual asociado. Nuestro objetivo principal es garantizar que no existe salto de dualidad para dichos problemas, es decir,  $\inf_x \varphi(x) = \sup_\lambda \psi(\lambda)$ . La siguiente definición será el primer paso hacia dicho objetivo.

**Definición 3.15** (Lagrangiano y parametrización dualizante). *Para un problema de minimizar  $\varphi(x)$  en  $\mathbb{R}^n$ , una parametrización dualizante es una representación  $\varphi(x) = f(x, 0)$ , donde  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  es una función propia tal que  $f(x, u)$  es sci y convexa en  $u$ . El lagrangiano asociado  $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ , está definido por*

$$L(x, \lambda) = \inf_u \{f(x, u) - \langle \lambda, u \rangle\}$$

**Observación 3.16.** Para cada  $x \in \mathbb{R}^n$  tenemos,

$$f^*(x, \lambda) = \sup_u \{ \langle \lambda, u \rangle - f(x, u) \} = -\inf_u \{ f(x, u) - \langle \lambda, u \rangle \} = -L(x, \lambda), \quad (3.2)$$

por lo tanto, la función  $-L(x, \cdot)$  es conjugada a  $f(x, \cdot)$  en  $\mathbb{R}^m$ . Por otro lado utilizando las hipótesis sobre  $f$ , para cada  $x \in \mathbb{R}^n$  obtenemos

$$f(x, u) = f^{**}(x, u) = \sup_\lambda \{ \langle \lambda, u \rangle - f^*(x, \lambda) \} = \sup_\lambda \{ \langle \lambda, u \rangle - (-L(x, \lambda)) \}, \quad (3.3)$$

es decir la función  $f(x, \cdot)$  es conjugada a  $-L(x, \cdot)$ .

**Ejemplo 3.17** (Lagrangianos en el esquema de Fenchel). Para el problema del Ejemplo 3.14, el lagrangiano correspondiente es

$$\begin{aligned} L(x, \lambda) &= \inf_u \{ \langle c, x \rangle + k(x) + h(b - Ax + u) - \langle \lambda, u \rangle \} \\ &= \langle c, x \rangle + k(x) + \langle b, \lambda \rangle - \langle \lambda, Ax \rangle - h^*(\lambda). \end{aligned}$$

Por otro lado,  $\partial_x L(\bar{x}, \bar{\lambda}) = c + \partial k(\bar{x}) - A^* \bar{\lambda}$  y  $\partial_\lambda [-L](\bar{x}, \bar{\lambda}) = -b + \partial h^*(\bar{\lambda}) + A\bar{x}$ , luego las condiciones de optimalidad en el Ejemplo 3.14 pueden ser reescritas de manera equivalente como

$$0 \in \partial_x L(\bar{x}, \bar{\lambda}), \quad 0 \in \partial_\lambda [-L](\bar{x}, \bar{\lambda}). \quad (3.4)$$

Note además que el lagrangiano  $L(x, \lambda)$  es convexo en  $x$  y cóncavo en  $\lambda$ .

**Proposición 3.18** (Convexidad del lagrangiano). Para cualquier parametrización dualizante  $\varphi = f(x, 0)$ , el lagrangiano asociado  $L(x, \lambda)$  es scs y cóncavo en  $\lambda$ . Este es además convexo en  $x$  si y sólo si  $f(x, u)$  es convexa en  $(x, u)$ . En este caso tenemos

$$(v, \lambda) \in \partial f(x, u) \iff v \in \partial_x L(x, \lambda), \quad u \in \partial_\lambda [-L](x, \lambda),$$

además el valor de  $L(x, \lambda)$  es finito e igual a  $f(x, u) - \langle \lambda, u \rangle$ .

**Prueba.** Dado  $x$ , por la Observación 3.16  $-L(x, \lambda) = f^*(x, \lambda)$ , luego  $L(x, \cdot)$  es cóncava, scs en  $\lambda$ . Si  $f(x, \lambda)$  es convexa en  $(x, \lambda)$ , entonces para todo  $\lambda \in \mathbb{R}^m$  la función  $f_\lambda(x, u) := f(x, u) - \langle \lambda, u \rangle$  es convexa, luego  $\inf_u f_\lambda(x, u) = L(x, \lambda)$  es convexa en  $x$ .

Recíprocamente, si  $L(x, \lambda)$  es convexa en  $x$ , la función  $L_\lambda(x, \lambda) := L(x, \lambda) + \langle \lambda, u \rangle$  es convexa en  $(x, u)$  y por la Observación 3.16  $f(x, u) = \sup_\lambda \{ L_\lambda(x, u) \}$ , luego  $f$  es convexa en  $(x, u)$ , pues es el supremo de la colección de la familia de funciones  $L_\lambda$  cuando  $\lambda$  varía sobre  $\mathbb{R}^m$ .

Para probar la equivalencia en la relación de subgradientes, sean  $x_0, u_0, v_0, \lambda_0$ , por el Teorema 3.3 y el hecho que  $f^*(x_0, \cdot) = -L(x_0, \cdot)$ , obtenemos

$$\begin{aligned} \lambda_0 \in \partial_u f(x_0, u_0) &\iff u_0 \in \partial_\lambda f^*(x_0, \lambda_0) = \partial_\lambda [-L](x_0, \lambda_0) \\ &\iff f(x_0, u_0) + f^*(x_0, \lambda_0) = \langle u_0, \lambda_0 \rangle \\ &\iff L(x_0, \lambda_0) = f(x_0, u_0) - \langle \lambda_0, u_0 \rangle. \end{aligned}$$

Si  $f$  es convexa, por la Definición 1.2, la condición  $(v_0, \lambda_0) \in \partial f(x_0, u_0)$  es equivalente a

$$f(x, u) \geq f(x_0, u_0) + \langle v_0, x - x_0 \rangle + \langle \lambda_0, u - u_0 \rangle \quad \text{para todo } x, u, \quad (3.5)$$

Para  $x = x_0$  tenemos que  $\lambda_0 \in \partial_u f(x_0, u_0)$  si y sólo  $u_0 \in \partial_\lambda[-L](x_0, \lambda_0)$ , además de la desigualdad (3.5) obtenemos que

$$\inf_u \{f(x, u) - \langle \lambda_0, u \rangle\} \geq f(x_0, u_0) - \langle \lambda_0, u_0 \rangle + \langle v_0, x - x_0 \rangle \quad \text{para todo } x,$$

esto es equivalente a  $L(x, \lambda_0) \geq L(x_0, \lambda_0) + \langle v_0, x - x_0 \rangle$  para todo  $x$ , por la Definición 1.2 y la convexidad de  $L(x, \lambda_0)$  en  $x$ , obtenemos que  $v_0 \in \partial_x L(x_0, \lambda_0)$ . Así  $(v_0, \lambda_0) \in \partial f(x_0, u_0)$  si y sólo si  $v_0 \in \partial L(x_0, \lambda_0)$  y  $u_0 \in \partial_\lambda[-L](x_0, \lambda_0)$  y  $u_0 \in \partial_\lambda[-L](x_0, \lambda_0)$   $\square$

**Ejemplo 3.19.** Consideremos un problema de la mochila de la forma

$$\min -x \quad \text{sujeto a } 2x \leq 1, \quad x \in U = \{0, 1\}.$$

este problema es equivalente a  $\min_{x \in \mathbb{R}} \varphi(x)$ , donde

$$\varphi(x) = f(x, 0), \quad f(x, u) = \begin{cases} -x + I_U(x) & \text{si } 2x + u \leq 1 \\ +\infty & \text{en otro caso.} \end{cases}$$

El Lagrangiano para este problema viene dado por

$$\begin{aligned} L(x, \lambda) &= \inf_u \{f(x, u) - \langle \lambda, u \rangle\} = \inf_u \{-x + I_U(x) - \langle \lambda, u \rangle : 2x + u + v = 1, v \geq 0\} \\ &= \inf \{-x + I_U(x) + \langle \lambda, 2x - 1 + v \rangle : v \geq 0\} \\ &= \begin{cases} -x + I_U(x) + \langle \lambda, 2x - 1 \rangle & \text{si } \lambda \geq 0 \\ -\infty & \text{si } \lambda < 0, \end{cases} \end{aligned}$$

expresión equivalente al lagrangiano (clásico) del Capítulo 2, ver el Ejemplo 2.17. Por otro lado vemos que la función  $L(x, \cdot)$  es cóncava, esto es justificado por la Proposición 3.18. Note también que  $L(\cdot, \lambda)$  no es convexo, pues la función  $f$  no es convexa.

Siempre que  $L(x, \lambda)$  sea convexo en  $x$  y cóncavo en  $\lambda$ , la condición de optimalidad en términos del lagrangiano (3.4) puede ser interpretado a través del siguiente concepto

**Definición 3.20** (Puntos de silla). *Un par de vectores  $(\bar{x}, \bar{\lambda})$  es punto silla de la función  $L$  definida en  $\mathbb{R}^n \times \mathbb{R}^m$  si y sólo si  $\inf L(x, \bar{\lambda}) = L(\bar{x}, \bar{\lambda}) = \sup_\lambda L(\bar{x}, \lambda)$ , es decir,*

$$L(x, \bar{\lambda}) \geq L(\bar{x}, \bar{\lambda}) \geq L(\bar{x}, \lambda) \quad \text{para todo } x, \lambda.$$

El conjunto de todos los puntos silla  $(\bar{x}, \bar{\lambda})$  será denotado por  $\text{argminimax}_{x, \lambda} L(x, \lambda)$ .

**Teorema 3.21** (Relaciones minimax). *Dado un problema de  $\min \varphi$  en  $\mathbb{R}^n$  con parametrización dualizante  $\varphi = f(x, 0)$ ,  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$ , sea  $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$  el lagrangiano asociado y  $\psi = -f^*(0, \cdot)$  en  $\mathbb{R}^m$ , entonces*

$$\varphi(x) = \sup_\lambda L(x, \lambda), \quad \psi(\lambda) = \inf_x L(x, \lambda), \quad (3.6)$$

$$\inf_x \varphi(x) = \inf_x [\sup_\lambda L(x, \lambda)] \geq \sup[\inf_x L(x, \lambda)] = \sup_\lambda \varphi(\lambda), \quad (3.7)$$

además

$$\begin{aligned} \bar{x} \in \text{argmin}_x \varphi(x) \\ \bar{\lambda} \in \text{argmax}_\lambda \psi(\lambda) &\iff (\bar{x}, \bar{\lambda}) \in \text{argminimax}_{x, \lambda} L(x, \lambda) \\ \inf_x \varphi(x) = \sup_\lambda \psi(\lambda) & \\ &\iff \varphi(\bar{x}) = \psi(\bar{\lambda}) = L(\bar{x}, \bar{\lambda}). \end{aligned} \quad (3.8)$$

La condición de punto silla  $(\bar{x}, \bar{\lambda}) \in \operatorname{argminimax}_{x,\lambda} L(x, \lambda)$  implica la condición de subgradiente

$$0 \in \partial_x L(\bar{x}, \bar{\lambda}), \quad 0 \in \partial_\lambda [-L](\bar{x}, \bar{\lambda}) \quad (3.9)$$

y son equivalentes siempre que  $L(x, \lambda)$  sea convexo en  $x$  y cóncavo en  $\lambda$ , en cuyo caso también es lo mismo que tener  $(0, \bar{\lambda}) \in \partial f(\bar{x}, 0)$ .

**Prueba.** La expresión para  $\varphi$  en (3.6) es consecuencia de (3.3) con  $u = 0$ , mientras que para  $\psi$  se sigue considerando  $v = 0$  en

$$-f^*(v, \lambda) = \inf_{x,u} \{f(x, u) - \langle v, x \rangle - \langle \lambda, u \rangle\} = \inf_x \{L(x, \lambda) - \langle v, x \rangle\}.$$

Las relaciones en (3.7) se sigue inmediatamente de (3.6). Como  $\inf_x \varphi(x) = p(0)$  y  $\sup_\lambda \psi(\lambda) = p^{**}(0)$  para  $p(u) := \inf_x f(x, u)$ , con  $u = 0$  obtenemos (3.8). Para la última parte del Teorema, por definición es claro que  $(\bar{x}, \bar{\lambda}) \in \operatorname{argminimax}_{x,\lambda} L(x, \lambda)$  implica la condición en (3.8), mientras que el recíproco es consecuencia de la Proposición 3.18.  $\square$

### 3.3. Lagrangiano aumentado generalizado y dualidad no convexa

Estamos interesados en conocer bajo que condiciones se cumple la relación de dualidad  $\inf_x \varphi(x) = \sup_\lambda \psi(\lambda)$  para problemas no convexos, es decir, no existe salto de dualidad. No obstante anular el salto de dualidad no quiere decir que al resolver el dual se encuentra una solución del problema primal (ver el comentario después de la Definición 2.14), simplemente se encuentra el valor óptimo del problema primal. Encontrar soluciones del problema primal via el dual depende de otras condiciones, llamadas en programación no lineal “exactitud del lagrangiano aumentado”, ver el Ejemplo 3.33.

Una interpretación de las relaciones dadas por el Lema 3.9

$$\begin{aligned} p(0) &= \inf_x \varphi(x) \\ p^{**}(0) &= \sup_\lambda \psi(\lambda) \end{aligned} \quad \text{para} \quad \begin{cases} p(u) = \inf_x f(x, u) \\ \varphi(x) = f(x, 0) \\ \psi(\lambda) = -f^*(0, \lambda) = -p^*(\lambda). \end{cases} \quad (3.10)$$

será nuestra principal herramienta en el objetivo de eliminar el salto de dualidad en problemas no convexos.

El caso general se muestra en la Figura 3.2, por definición de  $p^{**}$ , el valor  $\sup_\lambda \psi(\lambda) = p^{**}(0)$  es el supremo de las funciones afines mayoradas por  $p$  que interceptan al eje vertical; los vectores  $\bar{\lambda} \in \operatorname{argmax}_\lambda \psi(\lambda)$ , si existe, corresponden a dichas funciones afines. Un salto de dualidad,  $\inf_x \varphi(x) > \sup_\lambda \psi(\lambda)$ , ocurre cuando los interceptos no alcanzan el valor  $\inf_x \varphi(x) = p(0)$ . En particular la falta de convexidad (de la función  $p$ ) puede ser la fuente de tal salto.

**Lema 3.22.** *Con las notaciones en (3.10) tenemos que*

$$\left. \begin{aligned} \inf_x \varphi(x) &= \sup_\lambda \psi(\lambda) \\ \bar{\lambda} &\in \operatorname{argmax}_\lambda \psi(\lambda) \end{aligned} \right\} \iff \begin{cases} p(u) \geq p(0) + \langle \bar{\lambda}, u \rangle & \text{para todo } u \\ \text{con } p(0) \neq -\infty \end{cases}$$

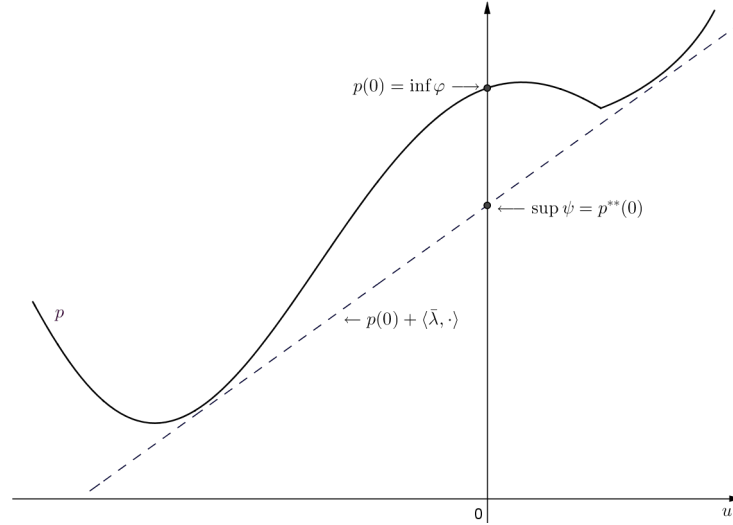


Figura 3.2: Salto de dualidad en problemas de minimización.

**Prueba.** Basta observar que

$$\begin{aligned} p(0) &= p^{**}(0) = -f^*(0, \bar{\lambda}) = -\sup_{x,u} \{ \langle (0, \bar{\lambda}), (x, u) \rangle - f(x, u) \} \\ &\leq f(x, u) - \langle \bar{\lambda}, u \rangle. \end{aligned}$$

□

**Ejemplo 3.23.** Volviendo a considerar el ejemplo de la mochila dado por

$$\min -x \quad \text{sujeto a } 2x \leq 1, \quad x \in U = \{0, 1\}.$$

que es equivalente al problema  $\min_{x \in \mathbb{R}} \varphi(x)$ , donde  $\varphi(x) = f(x, 0)$ , siendo  $f$  una parametrización dualizante dada por

$$f(x, u) = \begin{cases} -x + I_U(x) & \text{si } 2x - u \leq 1 \\ +\infty & \text{en otro caso.} \end{cases}$$

Un cálculo de la función  $p$  nos da

$$\begin{aligned} p(u) &= \inf_x f(x, u) = \{-x : x \in \{0, 1\}, 2x - 1 + v = u, v \geq 0\} \\ &= \begin{cases} +\infty & \text{si } u < -1 \\ 0 & \text{si } -1 \leq u < 1 \\ -1 & \text{si } 1 \leq u \end{cases} \end{aligned}$$

Note que la función  $p$  no es otra cosa que la función primal estudiada en 2.2.4. Por otro lado observe que  $p$  no es convexa y que existe salto de dualidad  $p(0) - p^{**}(0) = 1/2$  como lo muestra la Figura 3.3

Este resultado sugiere que el salto de dualidad puede evitarse si el problema dual se define de manera conveniente, lo cual motiva la siguiente definición.

**Definición 3.24** (Función de aumento). Una función  $\sigma: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  se denomina función de aumento si  $\sigma$  es propia, sci, convexa y verifica:

$$\min \sigma = 0, \quad \operatorname{argmin} \sigma = \{0\}.$$

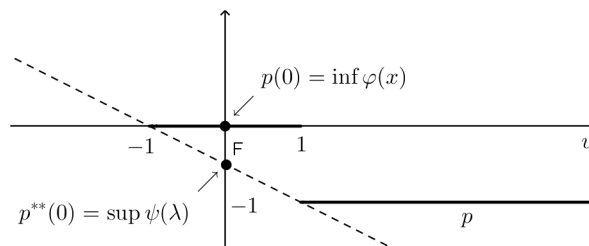


Figura 3.3: Salto de dualidad en un problema no convexo

Como ejemplos podemos citar la función norma y la función valor absoluto.

**Definición 3.25** (Lagrangiano aumentado). *Para un problema primal de minimizar  $\varphi(x)$  sobre  $x \in \mathbb{R}^n$  y cualquier parametrización dualizante  $\varphi = f(\cdot, 0)$  con  $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ , el lagrangiano aumentado con parámetro  $r > 0$  correspondiente a una función de aumento  $\sigma: \mathbb{R}^m \rightarrow \mathbb{R}$  es la función  $\ell: \mathbb{R}^n \times \mathbb{R}^m \times (0, \infty) \rightarrow \mathbb{R}$  definida por*

$$\ell(x, \lambda, r) := \inf_u \left\{ f(x, u) + r\sigma(u) - \langle \lambda, u \rangle \right\}.$$

El problema dual correspondiente está dado por

$$\sup_{(\lambda, r) \in \mathbb{R}^m \times (0, \infty)} \bar{\psi}(\lambda, r),$$

donde  $\bar{\psi}(\lambda, r) := \inf_x \ell(x, \lambda, r)$ .

**Observación 3.26.** *Note que, como con el lagrangiano clásico la función dual  $\bar{\psi}$  es el ínfimo del lagrangiano aumentado en la variable primal  $x$ .*

*La idea del lagrangiano aumentado es reemplazar la desigualdad  $p(u) \geq p(0) + \langle \bar{\lambda}, u \rangle$  para todo  $u$  por una de la forma*

$$p(u) \geq p(0) + \langle \bar{\lambda}, u \rangle - r\sigma(u) \quad \text{para todo } u, \quad (3.11)$$

*para alguna función de aumento  $\sigma$  y un valor del parámetro  $r$  suficientemente grande. Esta aproximación del problema dual elimina el salto de dualidad debido a que la desigualdad (3.11) es idéntica a*

$$p_{r\sigma}(u) \geq p_{r\sigma}(0) + \langle \bar{\lambda}, u \rangle \quad \text{para todo } u$$

*donde  $p_{r\sigma}(u) = \inf_x f_{r\sigma}(x, u)$  para la función  $f_{r\sigma}(x, u) = f(x, u) + r\sigma(u)$ . En efecto  $p_{r\sigma}(u) = p(u) + r\sigma(u)$  y  $p_{r\sigma}(0) = p(0)$ , debido a que  $\sigma(0) = 0$ .*

*Por otro lado tenemos que,  $\varphi(\cdot) = f_{r\sigma}(\cdot, 0)$ . Más aún como  $\sigma$  es propia, *sci* y convexa, la representación de  $\varphi$  en términos de  $f_{r\sigma}$  (como en términos de  $f$ ) es una parametrización dualizante. El lagrangiano asociado con  $f_{r\sigma}$  es  $L_{r\sigma} = \ell(x, \lambda, r)$ . El problema dual resultante, que consiste en maximizar  $\psi_{r\sigma}(\cdot) = -f_{r\sigma}^*(0, \cdot)$  sobre  $\lambda \in \mathbb{R}^m$ , verifica  $\psi_{r\sigma}(\lambda) = \bar{\psi}(\lambda, r)$ .*

El siguiente resultado contiene algunas consecuencias importantes de la Definición 3.25.

**Proposición 3.27** (Propiedades del lagrangiano aumentado). *Para una parametrización dualizante  $f$  y una función de aumento  $\sigma$ , el lagrangiano aumentado  $\ell(x, \lambda, r)$  es cóncavo y scs en  $(\lambda, r)$ , más aún éste es no decreciente en  $r$ . Si  $f(x, u)$  es convexa en  $(x, u)$ , entonces  $\ell(x, \lambda, r)$  es convexo en  $x$ . Si  $\sigma$  es finita en todas partes, el lagrangiano aumentado es dado en términos del lagrangiano ordinario  $L(x, \lambda)$  por*

$$\ell(x, \lambda, r) = \sup_z \{L(x, \lambda - z) - r\sigma^*(r^{-1}z)\} \quad (3.12)$$

Análogamente, la expresión dual aumentada  $\bar{\psi}(\lambda, r)$  es cóncava y scs en  $(\lambda, r)$ , más aún, esta es no decreciente en  $r$ . Si  $f(x, u)$  es convexa en  $(x, u)$  y  $\sigma$  finita en todas partes tenemos que

$$\bar{\psi}(\lambda, r) = \sup_z \{\psi(\lambda - z) - r\sigma^*(r^{-1}z)\}. \quad (3.13)$$

Luego  $(\bar{\lambda}, \bar{r})$  maximiza  $\bar{\psi}$  si y sólo si  $\bar{\lambda}$  maximiza  $\psi$ ; el valor de  $\bar{r} > 0$  puede ser escogido arbitrariamente.

**Prueba.** Las propiedades de  $\ell$  es consecuencia directa de la Definición 3.25, mientras que (3.12) se debe a que  $-\ell(x, \lambda, r) = \sup_u \{\langle \lambda, u \rangle - (f(x, u) + r\sigma(u))\} = [f(x, \cdot) + r\sigma(\cdot)]^*(\lambda)$  y el Teorema 3.8(a), pues

$$\begin{aligned} \ell(x, \lambda, r) &= -\sup_u \{\langle \lambda, u \rangle - (f(x, u) + r\sigma(u))\} = -[f(x, \cdot) + r\sigma(\cdot)]^*(\lambda) \\ &= -\inf_z \{f^*(x, \lambda - z) + r\sigma^*(r^{-1}z)\} = \sup_z \{L(x, \lambda - z) - r\sigma^*(r^{-1}z)\}. \end{aligned}$$

La última igualdad es debido a (3.2). De manera análoga se deducen las propiedades de  $\bar{\psi}$ .  $\square$

**Ejemplo 3.28** (Lagrangiano proximal). *Un lagrangiano aumentado generado con la función de aumento  $\sigma(u) = \frac{1}{2}|u|^2$  ( $|\cdot|$  es la norma 2 definida por  $|u| = (\sum_{i=1}^m u_i^2)^{1/2}$ ) se denomina lagrangiano proximal. Entonces*

$$\begin{aligned} \ell(x, \lambda, r) &= \inf_u \left\{ f(x, u) + \frac{r}{2}|u|^2 - \langle \lambda, u \rangle \right\} \\ &= \sup_z \left\{ L(x, \lambda - z) - \frac{1}{2r}|z|^2 \right\} = \sup_z \left\{ L(x, z) - \frac{1}{2r}|z - \lambda|^2 \right\}, \end{aligned}$$

donde la segunda igualdad se debe a (3.12). En particular para el problema de la mochila de la forma

$$\min -x \quad \text{sujeto a } 2x \leq 1 \quad x \in U = \{0, 1\},$$

el cual es equivalente a  $\min_{x \in \mathbb{R}} \varphi(x)$ , con  $\varphi(x) = f(x, 0)$  y  $f(x, u) := -x + I_U(x)$  si  $2x + u \leq 1$ ;  $+\infty$  en caso contrario. Por el Ejemplo 3.19, el lagrangiano para este problema viene dado por

$$L(x, \lambda) = \begin{cases} -x + I_U(x) + \langle \lambda, 2x - 1 \rangle & \text{si } \lambda \geq 0 \\ -\infty & \text{si } \lambda < 0, \end{cases}$$

luego el lagrangiano proximal es

$$\begin{aligned} \ell(x, \lambda, r) &= \sup_z \left\{ L(x, z) - \frac{1}{2r}|z - \lambda|^2 \right\} \\ &= \sup_z \left\{ -x + I_U(x) + \langle z, 2x - 1 \rangle - \frac{1}{2r}|z - \lambda|^2 \right\} \\ &= \begin{cases} -x + I_U(x) + \lambda(2x - 1) + \frac{1}{2}r(2x - 1)^2 & \text{si } 2x - 1 \geq -\frac{\lambda}{r} \\ -x + I_U(x) - \frac{\lambda^2}{2r} & \text{si } 2x - 1 \leq -\frac{\lambda}{r}. \end{cases} \end{aligned}$$

Observe que esta expresión es igual (excepto por el término  $I_U(x)$ ) al lagrangiano aumentado estudiado en el Capítulo 2, por lo tanto se ha eliminado el salto de dualidad, ver el Ejemplo 2.32. La diferencia está en que en el capítulo anterior  $r$  era un parámetro externo y aquí es una componente más de la variable dual.

**Ejemplo 3.29** (Lagrangiano sharp). *Es un lagrangiano aumentado generado con la función de aumento  $\sigma(u) = \|u\|$  (cualquier norma  $\|\cdot\|$ ). Entonces*

$$\begin{aligned}\ell(x, \lambda, r) &= \inf_u \left\{ f(x, u) + r\|u\| - \langle \lambda, u \rangle \right\} \\ &= \sup_z \left\{ L(x, \lambda - z) - I_{rB^\circ}(z) \right\} = \sup_{\|z-\lambda\|^\circ \leq r} L(x, z)\end{aligned}$$

donde la segunda igualdad es debida a (3.12). En particular, si consideramos la norma 1, entonces la norma dual es la norma del máximo, luego para el Ejemplo de la mochila de la forma

$$\min -x \quad \text{sujeto a } 2x \leq 1 \quad x \in U = \{0, 1\},$$

el lagrangiano sharp es

$$\begin{aligned}\ell(x, \lambda, r) &= \sup_{\|z-\lambda\|^\circ \leq r} L(x, z) \\ &= \sup_{-r+\lambda \leq z \leq r+\lambda} -x + I_U(x) + z(2x - 1) \\ &= \begin{cases} -x + I_U(x) + \lambda(2x - 1) + r(2x - 1) & \text{si } 2x - 1 > 0 \\ -x + I_U(x) + \lambda(2x - 1) - r(2x - 1) - [r - \lambda]^+(2x - 1)^- & \text{si } 2x - 1 \leq 0, \end{cases}\end{aligned}$$

en cualquier caso tenemos que

$$\ell(x, \lambda, r) = -x + I_U(x) + \lambda(2x - 1) + r|2x - 1| - [r - \lambda]^+(2x - 1)^-.$$

Teniendo en cuenta que  $U = \{0, 1\}$  la función dual aumentada viene dada por

$$\begin{aligned}\bar{\psi}(\lambda, r) &= \min\{\ell(0, \lambda, r), \ell(1, \lambda, r)\} \\ &= \min\{-\lambda + r - [r - \lambda]^+, -1 + \lambda + r\} \\ &= \min\{[r - \lambda]^+ - [r - \lambda]^- - [r - \lambda]^+, -1 + \lambda + r\} \\ &= \min\{r - \lambda, 0, -1 + \lambda + r\}\end{aligned}$$

El valor óptimo dual es

$$\sup_{\lambda, r} \bar{\psi}(\lambda, r) = 0,$$

por lo tanto se ha eliminado el salto de dualidad.

Aquí hemos utilizado la parte positiva y negativa de un número real  $a$ , definida respectivamente por  $a^+ := \max\{0, a\}$  y  $a^- := \max\{0, -a\}$ , de donde obtenemos que  $a = a^+ - a^-$  y  $|a| = a^+ + a^-$ .

El siguiente resultado nos dice que la teoría de dualidad desarrollada en el contexto del lagrangiano aumentado es capaz de eliminar el salto de dualidad incluso en ciertos problemas no convexos.



**Teorema 3.30** (Dualidad sin convexidad). *Para un problema de minimizar  $\varphi$  en  $\mathbb{R}^n$ , consideremos el lagrangiano aumentado  $\ell(x, \lambda, r)$  asociado con una parametrización dualizante  $\varphi(\cdot) = f(\cdot, 0)$ ,  $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  y alguna función de aumento  $\sigma: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ . Suponiendo que  $f(x, u)$  es nivel acotado en  $x$  localmente uniforme en  $u$ , sea  $p(u) := \inf_x f(x, u)$ . Suponiendo además que existe  $(\lambda, r) \in \mathbb{R}^m \times (0, \infty)$  tal que  $\inf_x \ell(x, \lambda, r) > -\infty$ , entonces*

$$\varphi(x) = \sup_{\lambda, r} \ell(x, \lambda, r), \quad \overline{\psi}(\lambda, r) = \inf_x \ell(x, \lambda, r)$$

y

$$\inf_x \varphi(x) = \inf_x [\sup_{\lambda, r} \ell(x, \lambda, r)] = \sup_{\lambda, r} [\inf_x \ell(x, \lambda, r)] = \sup_{\lambda, r} \overline{\psi}(\lambda, r). \quad (3.14)$$

Más aún, las soluciones óptimas de los problemas primal y dual aumentado están caracterizadas como puntos de silla del lagrangiano aumentado:

$$\begin{aligned} \bar{x} \in \operatorname{argmin}_x \varphi(x) \quad y \quad (\bar{\lambda}, \bar{r}) \in \operatorname{argmax}_{\lambda, r} \overline{\psi}(\lambda, r) \\ \iff \inf_x \ell(x, \bar{\lambda}, \bar{r}) = \ell(\bar{x}, \bar{\lambda}, \bar{r}) = \sup_{\lambda, r} \ell(\bar{x}, \lambda, r), \end{aligned} \quad (3.15)$$

los elementos de  $\operatorname{argmax}_{\lambda, r} \overline{\psi}(\lambda, r)$  son los pares  $(\bar{\lambda}, \bar{r})$  satisfaciendo la propiedad

$$p(u) \geq p(0) + \langle \bar{\lambda}, u \rangle - \bar{r}\sigma(u) \quad \text{para todo } u. \quad (3.16)$$

**Prueba.** La primera igualdad es consecuencia de la definición de parametrización dualizante y conjugación, en efecto:

$$\begin{aligned} \varphi(x) &= f(x, 0) = f^{**}(x, 0) = \sup_u \{ \langle (x, 0), (0, \lambda) \rangle - f^*(0, \lambda) \} \\ &= \sup_{\lambda} -f^*(0, \lambda) = \sup_{\lambda} - \sup_u \{ \langle (0, \lambda), (x, u) \rangle - f(x, u) \} \\ &= \sup_{\lambda, r} - \sup_u \{ \langle \lambda, u \rangle - (f(x, u) + r\sigma(u)) \} \\ &= \sup_{\lambda, r} \inf_u \{ f(x, u) + r\sigma(u) - \langle \lambda, u \rangle \} = \sup_{\lambda, r} \ell(x, \lambda, r). \end{aligned}$$

Mientras que la segunda recae en la definición de función dual aumentada y lagrangiano aumentado, en efecto:

$$\begin{aligned} \overline{\psi}(\lambda, r) &= \inf_{x, u} \{ f(x, u) + r\sigma(u) - \langle \lambda, u \rangle \} \\ &= \inf_x \inf_u \{ f(x, u) + r\sigma(u) - \langle \lambda, u \rangle \} = \inf_x \ell(x, \lambda, r). \end{aligned}$$

Mostraremos ahora la segunda igualdad en (3.14). Por definición de supremo obtenemos

$$\inf_x [\sup_{\lambda, r} \ell(x, \lambda, r)] \geq \sup_{\lambda, r} [\inf_x \ell(x, \lambda, r)].$$

Como  $\overline{\psi}(\lambda, r) = \inf_x \ell(x, \lambda, r) > -\infty$ , entonces existe al menos un par  $(\tilde{\lambda}, \tilde{r})$  tal que  $\overline{\psi}(\tilde{\lambda}, \tilde{r})$  es finito. Para conseguir la igualdad requerida será suficiente mostrar que  $\overline{\psi}(\tilde{\lambda}, \tilde{r})$  converge a  $p(0) := \inf_x \varphi(x)$  cuando  $r \rightarrow \infty$ , pues la función  $\overline{\psi}(\tilde{\lambda}, r)$  es creciente en  $r$  (Proposición 3.27). En efecto, por definición de  $\overline{\psi}$  obtenemos, para todo  $r > 0$

$$\overline{\psi}(\tilde{\lambda}, r) = \inf_u \{ p(u) + r\sigma(u) - \langle \tilde{\lambda}, u \rangle \}.$$

Sea  $\tilde{p}(u) := p(u) + \tilde{r}\sigma(u) - \langle \tilde{\lambda}, u \rangle$ , notemos que  $\tilde{p}(0) = p(0)$  pues  $\sigma(0) = 0$ , luego  $\tilde{p}$  es una función acotada inferiormente por  $\overline{\psi}(\tilde{\lambda}, \tilde{r})$ , además  $\tilde{p}$  es propia y sci, pues  $p$  y  $\sigma$  lo son. Escribiendo

$$\overline{\psi}(\tilde{\lambda}, \tilde{r} + s) = \inf_u \{ \tilde{p} + s\sigma(u) \} \quad \text{para } s > 0,$$

probaremos que el límite de  $\bar{\psi}(\bar{\lambda}, \bar{r} + s)$  cuando  $s \rightarrow \infty$  es  $\tilde{p}(0)$ . Como  $\sigma$  es propia, sci, y  $\operatorname{argmin} \sigma = \{0\}$ , entonces por la Proposición 1.7, la función  $\sigma$  es nivel acotado. La convexidad de  $\sigma$  y la Proposición 1.9 nos dice que  $\sigma$  es nivel-coercivo, luego, debido a que  $\tilde{p}$  es una función acotada inferiormente, la función  $\tilde{p} + \sigma$  también es nivel-coercivo. Como  $\sigma \geq 0$  entonces  $\tilde{p} + \sigma$  crece puntualmente cuando  $s \rightarrow \infty$  a la función  $I_{\{0\}} + \tilde{p}(0)$ , por lo tanto  $\inf(\tilde{p} + \sigma) \rightarrow \inf(I_{\{0\}} + \tilde{p}(0)) = \tilde{p}(0)$   $\square$

La importancia de las soluciones del problema dual aumentado se encuentra en la siguiente definición.

**Definición 3.31** (Representación exacta de penalidad). *Un vector  $\bar{\lambda}$  soporta una representación exacta de penalidad para el problema de minimizar  $\varphi$  en  $\mathbb{R}^n$  si para todo  $r > 0$  suficientemente grande, este problema es equivalente a minimizar  $\ell(\cdot, \bar{\lambda}, r)$  en  $\mathbb{R}^n$  en el sentido que*

$$\inf_x \varphi(x) = \inf_x \ell(x, \bar{\lambda}, r) \quad y \quad \operatorname{argmin}_x \varphi(x) = \operatorname{argmin}_x \ell(x, \bar{\lambda}, r).$$

*Se dice que un valor específico  $\bar{r} > 0$  sirve como un umbral de penalización adecuado si la propiedad es válida para todo  $r \in (\bar{r}, \infty)$ .*

**Teorema 3.32** (Criterio para una representación exacta de penalidad). *Con las notaciones e hipótesis del Teorema 3.30, un vector  $\bar{\lambda}$  soporta una representación exacta de penalidad para el problema primal si y sólo si existen una vecindad  $W$  de 0 y  $\hat{r} > 0$  tales que*

$$p(u) \geq p(0) + \langle \bar{\lambda}, u \rangle - \hat{r}\sigma(u) \quad \text{para todo } u \in W \quad (3.17)$$

*Este criterio a su vez es equivalente a la existencia de un  $\bar{r} > 0$  con  $(\bar{\lambda}, \bar{r}) \in \operatorname{argmax}_{\lambda, r} \bar{\psi}(\lambda, r)$ . Más aún  $\bar{r}$  es un umbral de penalidad para la propiedad de penalidad exacta respecto de  $\bar{\lambda}$ .*

**Prueba.** Como  $f$  es propia, sci y de nivel acotado en  $x$  localmente uniforme en  $u$  (hipótesis del Teorema 3.30), entonces, por el Teorema 1.3, la función  $p$  es propia y sci en  $\mathbb{R}^m$ , luego por la desigualdad (3.15)  $p(0) = \inf_x \varphi(x)$  es finito.

Mostraremos primero que la condición  $(\bar{\lambda}, \bar{r}) \in \operatorname{argmax}_{\lambda, r} \bar{\psi}(\lambda, r)$  es necesaria y suficiente para que  $\bar{\lambda}$  soporte una representación exacta de penalidad con  $\bar{r}$  umbral de penalización. Para la necesidad, note que la última condición implica, por Definición 3.31, que  $\inf_x \varphi(x) = \inf_x \ell(x, \bar{\lambda}, r) = \bar{\psi}(\bar{\lambda}, r)$  para todo  $r \in (\bar{r}, \infty)$ , por lo tanto, debido a que  $\bar{\psi}$  es sci (por la Proposición 3.27) obtenemos que  $\bar{\psi}(\bar{\lambda}, \bar{r}) \geq \inf_x \varphi(x)$ . Como  $\sup_{\lambda, r} \bar{\psi}(\lambda, r) = \inf_x \varphi(x)$  por (3.14), se sigue que  $(\bar{\lambda}, \bar{r})$  maximiza  $\bar{\psi}$ .

Para la suficiencia, note que por (3.14)  $\bar{\psi}(\bar{\lambda}, \bar{r}) = \inf_x \ell(x, \bar{\lambda}, \bar{r}) = \inf_x \varphi(x)$ . Por otro lado, por el Teorema 3.30, la condición  $(\bar{\lambda}, \bar{r}) \in \operatorname{argmax}_{\lambda, r} \bar{\psi}(\lambda, r)$  corresponde a tener la desigualdad (3.16). Cuando el valor de  $\bar{r}$  es reemplazado por cualquier valor  $r > \bar{r}$ , esta desigualdad es estricta para  $u \neq 0$ , pues  $\sigma(0) = 0$  y  $\sigma(u) > 0$  para  $u \neq 0$ , entonces

$$\operatorname{argmin}_u \{p(u) + r\sigma(u) - \langle \bar{\lambda}, u \rangle\} = \{0\}$$

Fijando tal  $r > \bar{r}$ , consideremos la función  $g(x, u) := f(x, u) + r\sigma(u) - \langle \bar{\lambda}, u \rangle$  y sus inf-proyecciones asociadas  $h(u) := \inf_x g(x, u)$  y  $k(x) := \inf_u g(x, u)$ , notando que  $h(u) = p(u) + r\sigma(u) - \langle \bar{\lambda}, u \rangle$  mientras que  $k(x) = \ell(x, \bar{\lambda}, r)$ , tenemos que

$$\left. \begin{array}{l} \bar{u} \in \operatorname{argmin}_u h(u) \\ \bar{x} \in \operatorname{argmin}_x g(x, \bar{u}) \end{array} \right\} \iff \left\{ \begin{array}{l} \bar{x} \in \operatorname{argmin}_x k(x) \\ \bar{u} \in \operatorname{argmin}_u g(\bar{x}, u) \end{array} \right.$$

Con esto, vamos a mostrar que  $\operatorname{argmin}_x \varphi(x) = \operatorname{argmin}_x \ell(x, \bar{\lambda}, r)$ . Vimos que  $\operatorname{argmin}_u h(u) = \{0\}$ , luego si  $\bar{x} \in \operatorname{argmin}_x \varphi(x) = \operatorname{argmin}_x g(x, 0)$ , entonces  $\bar{x} \in \operatorname{argmin}_x k(x) = \operatorname{argmin}_x \ell(x, \bar{\lambda}, r)$ . Recíprocamente, notando que  $\operatorname{argmin}_u g(\bar{x}, u) = \{0\}$ , si  $\bar{x} \in \operatorname{argmin}_x \ell(x, \bar{\lambda}, r) = \operatorname{argmin}_x k(x)$ , entonces  $\bar{x} \in \operatorname{argmin}_x g(x, 0) = \operatorname{argmin}_x \varphi(x)$ .

Para completar la prueba del teorema basta mostrar que si se tiene (3.17), entonces debe existir  $\bar{r} \in (\hat{r}, \infty)$  tal que la condición más fuerte (3.16) se tenga. Tomando  $W = \epsilon B$  en (3.17) y teniendo en cuenta que la desigualdad (3.17) continua valiendo para  $\bar{r} > \hat{r}$ . Como  $\inf_x \ell(x, \lambda, r) > -\infty$  (hipótesis del Teorema 3.30) existe  $(\tilde{\lambda}, \tilde{r}) \in \mathbb{R}^m \times (0, \infty)$  tal que  $\bar{\psi}(\tilde{\lambda}, \tilde{r})$  es finito; entonces para  $\alpha = \bar{\psi}(\tilde{\lambda}, \tilde{r})$  tenemos

$$p(u) \geq \alpha + \langle \tilde{\lambda}, u \rangle - \tilde{r}\sigma(u) \quad \text{para todo } u$$

Será suficiente mostrar que, cuando  $\bar{r}$  es suficientemente grande, se tendrá

$$\alpha - \langle \tilde{\lambda}, u \rangle + \tilde{r}\sigma(u) > p(0) + \langle \tilde{\lambda}, u \rangle - \bar{r}\sigma(u) \quad \text{para } |u| > \epsilon.$$

Esto equivale a mostrar que para valores grandes de  $\bar{r}$  tendremos

$$\{u : (\bar{r} - \tilde{r})\sigma(u) \leq \langle \bar{\lambda} - \tilde{\lambda}, u \rangle + p(0) - \alpha\} \subseteq B(0, \epsilon).$$

denotando por  $s = \bar{r} - \tilde{r} > 0$ ,  $y = |\bar{\lambda} - \tilde{\lambda}|$  y  $\mu := p(0) - \alpha$ , solamente necesitamos probar que

$$C(s) := \{u : s\sigma(u) \leq y|u| + \mu\} \subseteq B(0, \epsilon)$$

cuando  $s$  es elegido suficientemente grande.

En efecto sabemos que  $\sigma$  es de nivel coercivo, pues  $\operatorname{argmin} \sigma = \{0\}$  (Proposición 1.7, 1.9), luego existe  $\gamma > 0$  y  $\beta$  tal que  $\sigma(u) \geq \gamma|u| + \beta$  para todo  $u$  (Proposición 1.8). Sea  $s_0 > 0$  suficientemente grande tal que  $s_0\gamma - y > 0$ . Para  $u \in C(s_0)$  tenemos que  $s_0(\gamma|u| + \beta) \leq y|u| + \mu$ , por lo tanto  $|u| \leq \rho := (\mu - s_0\beta)/(s_0\gamma - y)$ . Como  $C(s) \subseteq C(s_0)$  cuando  $s > s_0$ , puesto que  $\sigma(u) \geq 0$  para todo  $u$ . Por lo tanto para  $s > s_0$  tenemos

$$C(s) \subseteq \{u : \sigma(u) \leq (y\rho + \mu)/s\}.$$

Por otro lado, como  $\sigma(u) = 0$  sólo para  $u = 0$ , el conjunto nivel  $\{u : \sigma(u) \leq \delta\}$  debe estar incluido en  $B(0, \epsilon)$  para  $\delta$  pequeño, tomando tal  $\delta$  y notando que  $(y\rho + \mu)/s \leq \delta$  cuando  $s$  excede cierto  $s_1$ , concluimos que  $C(s) \subseteq B(0, \epsilon)$  cuando  $s > s_1$ .  $\square$

**Ejemplo 3.33.** Consideremos el lagrangiano sharp del Ejemplo 3.29 bajo la hipótesis que  $\inf_x \ell(x, 0, r) > -\infty$  para algún  $r$ . Entonces el vector  $\bar{\lambda} = 0$  soporta una representación exacta de penalidad si y sólo si la función  $p(u) = \inf_x f(x, u)$  es “calm” inferiormente en  $u = 0$ .

En efecto, por el Teorema 3.32,  $\bar{\lambda} = 0$  soporta una representación exacta de penalidad si y sólo si existe una vecindad  $W$  de 0 y  $\hat{r} > 0$ :  $p(u) \geq p(0) + \langle 0, u \rangle - \hat{r}\|u\|$  para todo  $u \in W$  si y sólo si  $p(u) \geq p(0) - \hat{r}\|u - 0\|$  para todo  $u \in W$ , es decir  $p$  es “calm” inferiormente en  $u = 0$ .

En particular para el problema de la mochila dado por

$$\min -x \quad \text{sujeto a } 2x \leq 1 \quad x \in U = \{0, 1\},$$

el cual es equivalente a  $\min_{x \in \mathbb{R}} \varphi(x)$ , con  $\varphi(x) = f(x, 0)$  y  $f$  definido por

$$f(x, u) := \begin{cases} -x + I_U(x) & \text{si } 2x + 1 - u \leq 1 \\ +\infty & \text{en caso contrario.} \end{cases}$$

Como  $p(0) = 0$ , pues  $p(u)$  está dado en el Ejemplo 3.13,  $\bar{\lambda} = 0$  soporta una representación exacta de penalidad si y sólo si existen  $W = (-1, 1)$ ;  $\hat{r} = 1$  tales que  $p(u) \geq p(0) - \tilde{r}|u|$ .

Por otro lado, considerando  $\bar{\lambda} = 0$ , una aplicación directa de la Definición 3.31 nos da

$$0 = \inf_{x \in \mathbb{R}} \varphi(x) = \inf_x \ell(x, 0, r) = \inf_{x \in U} -x + r(2x - 1)^+ = 0$$

para  $r > 1$ , es decir,  $\bar{\lambda} = 0$  soporta una representación exacta de penalidad. Además esto nos dice que resolviendo el problema dual aumentado se resuelve el problema primal.

Veremos ahora los métodos duales del subgradiente y de los planos cortantes, cuya finalidad es resolver el problema dual.

## Métodos de optimización no diferenciable

En este Capítulo describiremos algunos métodos clásicos destinados a resolver el problema dual, en particular estudiaremos el método del subgradiente y de planos cortantes, así como sus respectivos resultados de convergencia, además aplicaremos estos métodos a un tipo particular de problema de optimización *DC* (no convexo), utilizando el lagrangiano sharp.

### 4.1. Métodos duales clásicos

Los métodos clásicos de optimización tienen por objetivo resolver el problema dual, por lo tanto nuestro problema es el siguiente:

- Debemos maximizar una función cóncava, la función dual  $\Theta$ ;
- La única información disponible que tenemos es la “black box” que resuelve el problema de Lagrange (2.2), para cada  $\lambda \in \mathbb{R}^m$ ;
- Hipótesis: Para todo  $\lambda \in \mathbb{R}^m$  existe al menos una solución para el problema de Lagrange (2.2). Por lo tanto la “black box” calcula el número  $\Theta(\lambda)$  y algún vector  $s(\lambda) := -c(u_\lambda) \in \partial(-\Theta)(\lambda)$ .

Esta Hipótesis se cumple para el Ejemplo 2.3 (de la mochila), vemos que siempre existe una solución óptima para el problema de Lagrange y esta es única excepto cuando  $\lambda$  es igual a uno de los  $n$  valores  $p_i/v_i$ .

Luego la aproximación general para resolver nuestro problema será la siguiente: en la  $k$ -ésima iteración llamamos a la “black box”; es decir se resuelve el problema de Lagrange en  $\lambda_k$ , para obtener  $\Theta(\lambda_k)$  y  $s(\lambda_k)$ ; luego en la  $k + 1$ -ésima iteración se calcula  $\lambda_{k+1}$ . Así un algoritmo dual está caracterizado por un conjunto de reglas que permiten calcular  $\lambda_{k+1}$ .

### 4.1.1. Algoritmo del subgradiente

---

**Algoritmo 1:** Subgradiente

---

**Datos:** Dado una sucesión  $\{t_k\}$ , con  $t_k > 0$  para  $k = 1, 2, \dots$

- 1 Elegir  $\lambda_1 \in \mathbb{R}^m$  y obtener  $s_1 \in \partial(-\Theta)(\lambda_1)$ . Poner  $k = 1$
- 2 **if**  $s_k = 0$  **then**
- 3     | parar;
- 4 **else**
- 5     | poner  $\lambda_{k+1} = \lambda_k - t_k \frac{s_k}{\|s_k\|}$
- 6 **end**
- 7 Obtener  $s_{k+1} \in \partial(-\Theta)(\lambda_{k+1})$ . Reemplazar  $k$  por  $k + 1$  y retornar al paso 2.

---

**Observación 4.1.** *Note que el valor  $\Theta(\lambda_k)$  no es usado por el Algoritmo 1, esto se debe a que: si  $\mu$  es mejor que  $\lambda_k$ , por la desigualdad del subgradiente de  $-\Theta$  en  $\lambda_k$  obtenemos*

$$\langle s_k, \mu - \lambda_k \rangle \leq \Theta(\lambda_k) - \Theta(\mu) < 0,$$

*esto significa que el ángulo entre la dirección en la que nos movemos  $-s_k$  y la dirección  $\mu - \lambda_k$  es agudo. Si nuestro movimiento a lo largo de esta dirección ( $-s_k$ ) es suficientemente pequeño, lograremos acercarnos a  $\mu$ .*

Por la Observación 4.1 el tamaño del paso  $t_k$  debe ser pequeño, por lo que consideraremos

$$t_k \rightarrow 0 \text{ cuando } k \rightarrow +\infty. \quad (4.1)$$

Por otro lado tenemos que  $\|\lambda_{k+1} - \lambda_k\| = \|-t_k \frac{s_k}{\|s_k\|}\| = t_k$  y la desigualdad triangular implica que todas las iteradas están contenidas en alguna bola:

$$\lambda_k \in B(\lambda_1, T), \text{ donde } T := \sum_{k=1}^{\infty} t_k.$$

Suponiendo que el conjunto de minimizadores de  $-\Theta$ ;  $\operatorname{argmin} -\Theta$ , es no vacío, entonces, si  $B(\lambda_1, T)$  no intersecta al conjunto  $\operatorname{argmin} -\Theta$ , el Algoritmo 1 no producirá una sucesión minimizante (es decir una sucesión  $\{\lambda_k\}$  tal que  $\lambda_k \rightarrow \inf -\Theta$  cuando  $k \rightarrow \infty$ ), por lo tanto los tamaños de paso no deben de ser muy pequeños y se requiere que

$$\sum_{k=1}^{\infty} t_k = +\infty. \quad (4.2)$$

Los dos Lemas siguientes son la base para probar la convergencia del Algoritmo del subgradiente 1.

**Lema 4.2.** *Sea  $\{t_k\}$  una sucesión de números positivos satisfaciendo (4.1) y (4.2), denotando*

$$\tau_n := \sum_{k=1}^n t_k, \quad \rho_n := \sum_{k=1}^n t_k^2.$$

*Entonces  $\frac{\rho_n}{\tau_n} \rightarrow 0$  cuando  $n \rightarrow +\infty$ .*

**Prueba.** Como  $t_k \rightarrow 0$ , dado  $\delta > 0$ ; existe algún  $n(\delta)$  tal que  $t_k \leq \delta$  para todo  $n > n(\delta)$ , entonces

$$\rho_n = \rho_{n(\delta)-1} + \sum_{k=n(\delta)}^n t_k^2 \leq \rho_{n(\delta)-1} + \delta \sum_{k=n(\delta)}^n t_k \leq \rho_{n(\delta)-1} + \delta \tau_n$$

luego

$$\frac{\rho_n}{\tau_n} \leq \frac{\rho_{n(\delta)-1}}{\tau_n} + \delta \text{ para todo } n > n(\delta);$$

así,  $\limsup \frac{\rho_n}{\tau_n} \leq \delta$ . Como  $\delta$  era arbitrario, obtenemos el resultado.  $\square$

**Lema 4.3.** Sea  $\Theta: \mathbb{R}^m \rightarrow \mathbb{R}$  una función cóncava y  $\bar{\lambda} \in \mathbb{R}^m$  fijo. Para todo  $\lambda \in \mathbb{R}^m$  tal que  $\Theta(\lambda) < \Theta(\bar{\lambda})$  y para todo  $s \in \partial(-\Theta)(\lambda)$ , tenemos que

$$\bar{d}(\lambda) := \frac{\langle s, \lambda - \bar{\lambda} \rangle}{\|s\|} > 0. \quad (4.3)$$

Además, dado  $M > 0$  existe  $L > 0$  tal que

$$\bar{d}(\lambda) \leq M \Rightarrow \Theta(\bar{\lambda}) - \Theta(\lambda) \leq L\bar{d}(\lambda).$$

**Prueba.** Para  $s \in \partial(-\Theta)(\lambda)$  tenemos

$$-\Theta(\gamma) \geq -\Theta(\lambda) + \langle s, \gamma - \lambda \rangle \text{ para todo } \gamma \in \mathbb{R}^m$$

en particular para  $\gamma = \bar{\lambda}$  obtenemos que  $\bar{d}(\lambda) > 0$ . Sea  $\mu(\lambda)$  la proyección de  $\bar{\lambda}$  sobre el hiperplano

$$\{\mu \in \mathbb{R}^m : \langle s, \mu - \lambda \rangle = 0\}.$$

entonces  $\|\mu(\lambda) - \bar{\lambda}\| = \bar{d}(\lambda)$  y  $\Theta(\mu(\lambda)) \leq \Theta(\lambda)$ . Si  $\mu(\lambda) \in B[\bar{\lambda}, M]$  entonces siendo  $-\Theta$  convexa resulta ser localmente lipschitz, por lo tanto existe  $L$  tal que

$$-\Theta(\lambda) \leq -\Theta(\mu(\lambda)) \leq -\Theta(\bar{\lambda}) + L\|\mu(\lambda) - \bar{\lambda}\|$$

$\square$

Definimos la sucesión de mejores valores generados por el Algoritmo 1:

$$\bar{\Theta}_k := \min\{-\Theta(\lambda_i) : i = 1, \dots, k\},$$

luego necesitamos conocer cuándo esta función de mejores valores converge al ínfimo de  $-\Theta$  en  $\mathbb{R}^m$ .

**Teorema 4.4** (Convergencia del método del subgradiente). *Aplicando el Algoritmo 1 a la función convexa  $-\Theta: \mathbb{R}^m \rightarrow \mathbb{R}$  con los tamaños de pasos  $t_k$  satisfaciendo (4.1) y (4.2). Entonces*

$$\bar{\Theta}_k \rightarrow \inf_{\lambda \in \mathbb{R}^m} -\Theta(\lambda) \text{ cuando } k \rightarrow +\infty.$$

**Prueba.** Procedemos por contradicción, si existiera  $\bar{\lambda} \in \mathbb{R}^m$  y  $\eta > 0$  tal que

$$-\Theta(\lambda_k) \geq \bar{\Theta}_k \geq -\Theta(\bar{\lambda}) + \eta \text{ para todo } k; \quad (4.4)$$

entonces desarrollando

$$\begin{aligned}\|\bar{\lambda} - \lambda_{k+1}\|^2 &= \|\bar{\lambda} - \lambda_k\|^2 + 2\langle \bar{\lambda} - \lambda_k, \lambda_k - \lambda_{k+1} \rangle + \|\lambda_k - \lambda_{k+1}\|^2 \\ &= \|\bar{\lambda} - \lambda_k\|^2 - 2t_k \bar{d}(\lambda_k) + t_k^2,\end{aligned}$$

donde  $\bar{d}(\lambda_k) = \frac{\langle s_k, \lambda_k - \bar{\lambda} \rangle}{\|s_k\|}$  es mayor que cero, por el Lema 4.3. Para  $n \geq k$ , sea  $\delta_n := \min_{k=1, \dots, n} \bar{d}(\lambda_k)$  tenemos que

$$\|\bar{\lambda} - \lambda_{k+1}\|^2 \leq \|\bar{\lambda} - \lambda_k\|^2 - 2t_k \delta_n + t_k^2, \quad \text{para } k = 1, \dots, n;$$

sumando de 1 a  $n$ , obtenemos

$$2\delta_n \sum_{k=1}^n t_k \leq \|\bar{\lambda} - \lambda_1\|^2 + \sum_{k=1}^n t_k^2 \quad \text{para todo } n.$$

Por el Lema 4.2, se sigue que  $\delta_n \rightarrow 0$  cuando  $n \rightarrow +\infty$ .

Así tenemos un subconjunto infinito de enteros  $K$  tal que  $\lim_{k \in K} \bar{d}(\lambda_k) = 0$ . Por el Lema 4.3  $\{\bar{d}(\lambda_k)\}_{k \in K}$  es acotado y

$$\lim[\Theta(\bar{\lambda}) - \Theta(\lambda_k)] = 0$$

el cual es una contradicción con (4.4). □

Note que en el Algoritmo 1 no existe un criterio de parada adecuado, en particular  $s_k$  no necesariamente tiende a 0. El algoritmo debe ser detenido “manualmente”, cuando  $t_k$  es pequeño comparado con la escala del problema.

### 4.1.2. Método de planos cortantes

Se asume conocido un conjunto compacto y convexo  $C \subseteq \mathbb{R}^m$  que contiene una solución dual. Nuestro problema dual

$$\sup\{\Theta(\lambda) : \lambda \in \mathbb{R}^m\}, \tag{4.5}$$

$\Theta(\lambda) := \inf_x L(x, \lambda)$ , donde  $L(x, \lambda) = \varphi(x) + \langle \lambda, c(x) \rangle$  puede ser reformulado como

$$\begin{aligned}\max \quad & \theta \\ \text{s.a.} \quad & \theta \in \mathbb{R}, \lambda \in C, \\ & \theta \leq L(x, \lambda) \quad \text{para todo } x \in U\end{aligned} \tag{4.6}$$

La idea básica del algoritmo de planos cortantes es acumular las restricciones una tras otra en (4.6). Además, aprovechar el hecho de que el índice de restricción  $x$  puede



restringirse al conjunto más pequeño  $\tilde{U}$  de (2.10).

---

**Algoritmo 2:** Planos cortantes

---

**Datos:** Dado un conjunto compacto, convexo  $C$  y una tolerancia de parada  $\epsilon \geq 0$ .

- 1 [Inicialización]. Escoger  $\lambda_0 \in C$  y resolver el problema de Lagrange en  $\lambda_0$  para obtener  $x_0 := x_{\lambda_0}$ . Poner  $k = 0$ .
- 2 [Problema maestro]. Resolver la siguiente relajación del problema (4.6):

$$\begin{aligned} \max \quad & \theta \\ \text{s.a.} \quad & \theta \in \mathbb{R}, \quad \lambda \in C, \\ & \theta \leq L(x_i, \lambda) \quad \text{para } i = 0, \dots, k-1 \end{aligned} \tag{4.7}$$

para obtener una solución  $(\theta_k, \lambda_k)$ .

- 3 [Problema local]. Resolver el problema de Lagrange en  $\lambda_k$ :

$$\inf_x \{ \varphi(x) + \langle \lambda_k, c(x) \rangle : x \in U \},$$

para obtener el próximo punto primal  $x_k := x_{\lambda_k}$ .

- 4 [Criterio de parada] **if**

$$\theta_k \leq \Theta(\lambda_k) + \epsilon, \tag{4.8}$$

**then**

- 5 | parar
  - 6 **else**
  - 7 | reemplazar  $k$  por  $k + 1$  y retornar al paso 1
  - 8 **end**
- 

Típicamente el conjunto  $C$  está caracterizado por cotas conocidas en las componentes de  $\lambda$ ; entonces (4.7) es un programa lineal con  $2m + k$  restricciones. Es importante señalar que la descripción de (4.7) puede ser hecha sin hacer referencia a los puntos primales  $u$ : siendo la función de Lagrange afín en  $\lambda$ , tenemos

$$L(x_i, \lambda) = L(x_i, \lambda_i) - \langle -c(x_i), \lambda - \lambda_i \rangle \tag{4.9}$$

valido para todo  $\lambda$  y  $i = 0, 1, \dots$ . Así, (4.7) puede ser reformulado como

$$\begin{aligned} \max \quad & \theta \\ \text{s.a.} \quad & \theta \in \mathbb{R}, \quad \lambda \in C, \\ & \theta \leq \Theta(\lambda_i) - \langle s(\lambda_i), \lambda - \lambda_i \rangle \quad i = 0, \dots, k-1 \end{aligned} \tag{4.10}$$

el cual solamente involucra objetos duales, esto es, los valores de  $\Theta$  y  $s$ .

El siguiente resultado establece la convergencia del Algoritmo 2.

**Teorema 4.5.** *Sea  $C \subset \mathbb{R}^m$  convexo compacto y  $\Theta: \mathbb{R}^m \rightarrow \mathbb{R}$  una función cóncava, consideremos el valor óptimo*

$$\bar{\Theta}_C := \max\{\Theta(\lambda), \lambda \in C\}.$$

*Entonces en el Algoritmo 2,  $\bar{\Theta}_C \leq \theta_k$  para todo  $k$  y se tiene la siguiente propiedad de convergencia:*

(i) Si  $\epsilon > 0$ , la parada ocurre en alguna iteración  $k_\epsilon$  con  $\lambda_{k_\epsilon}$  satisfaciendo

$$\bar{\Theta}_C \leq \Theta(\lambda_{k_\epsilon}) + \epsilon.$$

(ii) Si  $\epsilon = 0$ , las sucesiones  $\{\theta_k\}$  y  $\{\Theta(\lambda_k)\}$  tienden a  $\bar{\Theta}_C$  cuando  $k \rightarrow +\infty$ .

**Prueba.** Como el conjunto factible de (4.6) está incluido en el conjunto factible de (4.7) entonces  $\bar{\Theta}_C \leq \theta_k$ . Si se cumple el criterio de parada (4.8), entonces  $\bar{\Theta}_C \leq \theta_{k_\epsilon} \leq \Theta(\lambda_{k_\epsilon}) + \epsilon$

Por otro lado, supongamos por contradicción que, para  $k = 1, 2, \dots$

$$\Theta(\lambda_k) + \epsilon < \bar{\Theta}_C \leq \theta_k \leq \Theta(\lambda_i) - \langle s(\lambda_i), \lambda_k - \lambda_i \rangle \quad \text{para } i < k$$

por lo tanto

$$\epsilon \leq \Theta(\lambda_i) - \Theta(\lambda_k) + \|s(\lambda_i)\| \|\lambda_k - \lambda_i\|.$$

Como  $-\Theta$  es convexa, luego es lipschitziana en el conjunto acotado  $C$ , existe  $L$  tal que

$$\epsilon < 2L \|\lambda_k - \lambda_i\| \quad \text{para todo } k = 1, 2, \dots$$

esto contradice la acotación de  $\{\lambda_k\} \subseteq C$ . Como  $\epsilon > 0$  fue arbitrario y debido a que  $\theta_k \geq \bar{\Theta}_C \geq \Theta(\lambda_k)$ , hemos probado que  $\Theta(\lambda_k) - \theta_k \rightarrow 0$ , si  $k \rightarrow +\infty$  (esto sucede sólo si  $\epsilon = 0$ ); entonces  $\bar{\Theta}_C$  es el límite común entre  $\{\theta_k\}$  y  $\{\Theta(\lambda_k)\}$

□

El método del plano cortante tiene buenas propiedades de convergencia para funciones “sharp”; ver [?]. En particular para funciones poliedrales el método tiene convergencia finita: esto es una consecuencia del Teorema 4.5(ii).

Si  $\epsilon = 0$ , el Algoritmo 2 normalmente no se detiene. Un caso de particular interés es cuando  $U$  es un conjunto finito, pues el algoritmo se detiene siempre, incluso si  $\epsilon = 0$ . Esto es consecuencia del siguiente resultado

**Proposición 4.6.** *Sin ninguna hipótesis sobre el conjunto  $C$  en el Algoritmo 2. El  $x_k$  generado en el paso 2 es diferente de todos los puntos  $x_0, \dots, x_{k-1}$  a no ser que se accione el criterio de parada en el paso 4.*

**Prueba.** Si  $x_k = x_i$  para algún  $i \leq k - 1$ , entonces, por la reformulación (4.7), obtenemos

$$\Theta(\lambda_k) = L(x_i, \lambda_k) \geq \theta_k$$

□

La presencia del conjunto compacto  $C$  en el Algoritmo 2 no solamente está motivada por el Teorema 4.5. Puede suceder que (4.10) no tenga soluciones si  $C$  fuese no acotado: por ejemplo cuando  $C = \mathbb{R}^m$ , en la primera iteración teniendo en  $\theta$  y  $s(\lambda_0) \neq 0$ . Esto es una desventaja del Algoritmo, el cual corresponde a una inherente inestabilidad: la sucesión  $\{\lambda_k\}$  debe ser artificialmente estabilizado.

En algunos casos favorables  $C$  puede ser eventualmente eliminado. Esto usualmente sucede cuando

(i) La función  $\Theta$  tiene un conjunto acotado de maximizadores y

(ii) Se han realizado suficientes iteraciones duales, entonces el “conjunto-subnivel linealizado”

$$\{\lambda \in \mathbb{R}^m : -\Theta(\lambda_i) + \langle s(\lambda_i), \lambda - \lambda_i \rangle \leq -\theta \text{ para } i = 0, \dots, k-1\}$$

es acotado (para algún  $\theta$ , ver ...); este incluso puede estar incluido en  $C$ . Bajo estas condiciones, es claro que (4.10) puede ser reemplazado por

$$\max\{\theta : \theta \leq \Theta(\lambda_i) - \langle s(\lambda_i), \lambda - \lambda_i \rangle \text{ para } i = 0, \dots, k-1\} \quad (4.11)$$

Es interesante dualizar el problema(4.11). Denotando por  $\alpha \in \mathbb{R}^k$  a la variable dual, el lagrangiano respectivo viene dado por

$$L(\theta, \lambda, \alpha) = \theta(1 - \sum_{i=0}^{k-1} \alpha_i) + \sum_{i=0}^{k-1} \alpha_i[\Theta(\lambda_i) + \langle s(\lambda_i), \lambda_i \rangle] - \sum_{i=0}^{k-1} \alpha_i s(\lambda_i) \lambda.$$

el problema dual resultante es

$$\begin{aligned} \min & \sum_{i=0}^{k-1} \alpha_i[\Theta(\lambda_i) + \langle s(\lambda_i), \lambda_i \rangle] \\ & \sum_{i=0}^{k-1} \alpha_i s(\lambda_i) = 0 \\ & \sum_{i=0}^{k-1} \alpha_i = 1, \alpha_i \geq 1, i = 0, \dots, k-1 \end{aligned}$$

Este problema puede ser escrito en notación primal recordando que  $s(\lambda_i) = -c(x_i)$  y recordando que  $\Delta_k := \{\alpha \in \mathbb{R}^k : \sum_{i=0}^{k-1} \alpha_i = 1, \alpha_i \geq 1, i = 0, \dots, k-1\}$  obtenemos

$$\min_{\alpha \in \Delta_k} \left\{ \sum_{i=0}^{k-1} \alpha_i \varphi(x_i) : \sum_{i=0}^{k-1} \alpha_i c(x_i) = 0 \right\} \quad (4.12)$$

Cuando utilizamos (4.12) en lugar de (4.11) el algoritmo de planos cortantes (un mecanismo de generación por columnas) recibe el nombre de algoritmo de Dantzing-Wolfe (mecanismo de generacion por columnas). En términos del problema primal (2.1);  $\varphi(\cdot)$  y  $c(\cdot)$  son reemplazados en (4.12) por combinaciones convexas apropiadas. Por lo tanto, el punto

$$\Delta_k \ni \alpha \mapsto x(\alpha) := \sum_{i=0}^{k-1} \alpha_i x_i$$

puede ser visto como como una solución primal aproximada como lo menciona el siguiente resultado

**Teorema 4.7.** *En el problema primal (2.1) consideremos  $U$  convexo,  $\varphi$  cóncavo y  $c$  afín. Supongamos además que el Algoritmo 2 aplicado a la función cóncava  $\Theta: \mathbb{R}^m \rightarrow \mathbb{R}$ , puede ser usado con (4.11) en lugar de (4.10). Cuando la parada ocurre, denotando por  $\bar{\alpha} \in \Delta_k$  una solución óptima de (4.12) entonces,  $x(\bar{\alpha})$  es un  $\epsilon$ -solución del problema primal (2.1).*

**Prueba.** Note que por construcción  $x(\bar{\alpha})$  es factible, además no existe salto de dualidad entre los programas lineales (4.11) y (4.12), utilizando la convexidad de  $\varphi$  resulta

$$\theta_k = \sum_{i=0}^{k-1} \bar{\alpha}_i \varphi(x_i) \geq \varphi(x(\bar{\alpha})).$$

Por el criterio de parada (4.8) y la desigualdad de dualidad débil (2.7) obtenemos

$$\varphi(x(\bar{\alpha})) \leq \theta_k \leq \Theta_k + \epsilon \leq \varphi(x) + \epsilon,$$

para todo  $x$  factible para el problema primal (2.1).  $\square$

## 4.2. Aplicación al lagrangiano sharp

Dado el problema de optimización

$$\begin{cases} \min_x & f_1(x) \\ \text{s.a} & x \in X \\ & \langle \mathcal{C}, x \rangle - c_2(x) \leq 0, \end{cases} \quad (4.13)$$

donde  $f_1$ ,  $c_2$  son funciones convexas y lineales por partes y  $X$  es un conjunto poliedral, la principal dificultad en el problema (4.13) está en la restricción  $DC$  (diferencia de una función lineal y otra convexa). Estamos interesados en calcular el lagrangiano sharp de (4.13) considerando la norma 1. Asociando un multiplicador escalar  $\lambda > 0$  a la restricción  $DC$ , el lagrangiano (clásico) viene dado por

$$L(x, \lambda) = \begin{cases} f_1(x) + I_X(x) + \lambda(\langle \mathcal{C}, x \rangle - c_2(x)) & \text{si } \lambda \geq 0 \\ -\infty & \text{si } \lambda < 0 \end{cases}$$

La expresión del Ejemplo 3.29 nos dice que

$$\ell(x, \lambda, r) = \sup_{\|z-\lambda\|^0 \leq r} L(x, z)$$

como la norma dual de la norma 1 es la norma infinita, entonces

$$\ell(x, \lambda, r) = \begin{cases} \sup & L(x, z) \\ \text{s.a} & -r + \lambda \leq z \leq r + \lambda. \end{cases}$$

Calculando este supremo con  $z \in [\max\{0, -r + \lambda\}, r + \lambda] = [\lambda - r + \max\{0, r - \lambda\}, r + \lambda]$ , obtenemos

$$\ell(x, \lambda, r) = \begin{cases} L(x, \lambda) + r(\langle \mathcal{C}, x \rangle - c_2(x)) & \text{si } \langle \mathcal{C}, x \rangle - c_2(x) > 0 \\ L(x, \lambda) - r(\langle \mathcal{C}, x \rangle - c_2(x)) - [r - \lambda]^+ [\langle \mathcal{C}, x \rangle - c_2(x)]^- & \text{si } \langle \mathcal{C}, x \rangle - c_2(x) \leq 0. \end{cases}$$

Luego, en cualquier caso el lagrangiano sharp viene dado por

$$\ell(x, \lambda, r) = L(x, \lambda) + r|\langle \mathcal{C}, x \rangle - c_2(x)| - [r - \lambda]^+ [\langle \mathcal{C}, x \rangle - c_2(x)]^-.$$

El problema dual aumentado, consiste en maximizar sobre todo  $(\lambda, r) \in \mathbb{R} \times (0, \infty)$  la función

$$\bar{\psi}(\lambda, r) = \inf_x \ell(x, \lambda, r) \quad (4.14)$$

**Proposición 4.8.** Si  $\bar{\psi}$  no es idénticamente  $-\infty$ , entonces para cualquier solución óptima  $x_{\lambda,r} \in X$  del problema (4.14) y para

$$c(x_{\lambda,r}) := \langle \mathcal{C}, x_{\lambda,r} \rangle - c_2(x_{\lambda,r})$$

se tiene que

$$s(\lambda, r) = - \left( \begin{array}{c} c(x_{\lambda,r}) + \alpha c(x_{\lambda,r})^- \\ |c(x_{\lambda,r})| - \alpha c(x_{\lambda,r})^- \end{array} \right) \in \partial(-\bar{\psi})(\lambda, r),$$

con  $\alpha = 0$  si  $\lambda - r < 0$ ;  $\alpha \in [0, 1]$  si  $\lambda = r$  y  $\alpha = 1$  si  $\lambda - r > 0$ .

**Prueba.** Para cada  $\lambda_1$  y  $r_1$ , por definición de  $\bar{\psi}(\lambda_1, r_1)$ , tenemos que

$$\begin{aligned} \bar{\psi}(\lambda_1, r_1) &\leq f_1(x_{\lambda,r}) + \langle \lambda_1, c(x_{\lambda,r}) \rangle + r_1 |c(x_{\lambda,r})| - [r_1 - \lambda_1]^+ c(x_{\lambda,r})^- \\ &= \bar{\psi}(\lambda, r) + \langle c(x_{\lambda,r}), \lambda_1 - \lambda \rangle + |c(x_{\lambda,r})| (r_1 - r) + c(x_{\lambda,r})^- ([r - \lambda]^+ - [r_1 - \lambda_1]^+) \\ &= \bar{\psi}(\lambda, r) + \left\langle \left( \begin{array}{c} c(x_{\lambda,r}) \\ |c(x_{\lambda,r})| \end{array} \right), \left( \begin{array}{c} \lambda_1 - \lambda \\ r_1 - r \end{array} \right) \right\rangle + c(x_{\lambda,r})^- ([r - \lambda]^+ - [r_1 - \lambda_1]^+). \end{aligned}$$

Para limitar superiormente el último término a la derecha  $T := c(x_{\lambda,r})^- ([r - \lambda]^+ - [r_1 - \lambda_1]^+)$ , consideremos la función convexa  $h(z) := [z]^+$  definida para  $z \in \mathbb{R}$ . Es fácil ver que para todo  $z, z_1 \in \mathbb{R}$  vale la desigualdad

$$h(z) \leq h(z_1) + \alpha(z - z_1) \quad \text{con} \quad \alpha \begin{cases} = 0 & z < 0 \\ \in [0, 1] & z = 0 \\ = 1 & z > 0. \end{cases}$$

Como  $T = c(x_{\lambda,r})^- (h(r - \lambda) - h(r_1 - \lambda_1))$ , aplicando la regla de la cadena para  $\alpha$  como en la proposición vale que

$$T \leq \alpha c(x_{\lambda,r})^- \left\langle \left( \begin{array}{c} 1 \\ -1 \end{array} \right), \left( \begin{array}{c} \lambda_1 - \lambda \\ r_1 - r \end{array} \right) \right\rangle,$$

y, en consecuencia

$$s(\lambda, r) = - \left( \begin{array}{c} c(x_{\lambda,r}) + \alpha c(x_{\lambda,r})^- \\ |c(x_{\lambda,r})| - \alpha c(x_{\lambda,r})^- \end{array} \right),$$

satisface la desigualdad del subgradiente, como se quería mostrar.  $\square$

## Algoritmo de subgradiente para el problema dual aumentado

---

**Algoritmo 3:** Subgradiente modificado

---

**Datos:** Dado una sucesión  $\{t_k\}$ , con  $t_k > 0$  para  $k = 1, 2, \dots$

1 Elegir  $\lambda_1, r_1$  y obtener  $s_1 = s(\lambda_1, r_1) \in \partial(-\bar{\psi})(\lambda_1, r_1)$ . Poner  $k = 1$

2 **if**  $s_k = (0, 0)$  **then**

3 | parar;

4 **else**

5 | poner  $(\lambda_{k+1}, r_{k+1}) = (\lambda_k, r_k) - t_k \frac{s_k}{\|s_k\|}$

6 **end**

7 Obtener  $s_{k+1} \in \partial(-\bar{\psi})(\lambda_{k+1}, r_{k+1})$ . Reemplazar  $k$  por  $k + 1$  y retornar al paso 2.

---

## Algoritmo de planos cortantes para el problema dual aumentado

Recordando que la función dual aumentada está definida por

$$\bar{\psi}(\lambda, r) = \inf_{x \in X} \ell(x, \lambda, r),$$

haciendo el cambio de variable  $\theta = \bar{\psi}(\lambda, r)$ , la desigualdad  $\theta \leq \ell(x, \lambda, r)$  debe verificarse para todo  $x \in X$ , por lo tanto el problema dual de maximizar  $\bar{\psi}(\lambda, r)$  sobre  $(\lambda, r) \in \mathbb{R} \times (0, \infty)$  es equivalente al siguiente problema

$$\begin{aligned} \max \quad & \theta \\ \text{s.a.} \quad & \theta \in \mathbb{R}, \quad (\lambda, r) \in C, \\ & \theta \leq \ell(x, \lambda, r), \quad x \in X. \end{aligned} \tag{4.15}$$

Note que el problema (4.15) es un programa lineal en las variables  $x, \lambda$  y  $r$ . sin embargo las restricciones son infinitas y no se conocen explícitamente. Supongamos que tenemos los puntos  $x_0, \dots, x_{k-1}$  en  $X$  y consideremos el siguiente relajación del problema (4.15)

$$\begin{aligned} \max \quad & \theta \\ \text{s.a.} \quad & \theta \in \mathbb{R}, \quad (\lambda, r) \in C, \\ & \theta \leq \ell(x_i, \lambda, r) \quad \text{para } i = 0, \dots, k-1 \end{aligned} \tag{4.16}$$

Este problema es un programa lineal con un número finito de restricciones y puede ser resuelto. Sea  $(\theta_k, \lambda_k, r_k)$  una solución óptima. Si esta solución satisface (4.15), entonces es una solución para el problema dual. Para verificar cuándo se satisface (4.15), se considera el siguiente subproblema

$$\min \ell(x, \lambda_k, r_k) \quad \text{s.a. } x \in X$$

Sea  $x_k$  una solución para este problema, entonces  $\bar{\psi}(\lambda_k, r_k) = \ell(x_k, \lambda_k, r_k)$ . Si

$$\theta_k \leq \bar{\psi}(\lambda_k, r_k), \tag{4.17}$$

entonces  $(\lambda_k, r_k)$  es una solución para el problema dual. En otro caso, para  $(\lambda, r) = (\lambda_k, r_k)$  la desigualdad en (4.16) se satisface para  $x = x_k$ . Así, adicionamos la restricción  $\theta \leq \ell(x_k, \lambda, r)$  a las restricciones en (4.16) y resolvemos el programa lineal. Es claro que el punto óptimo actual  $(\theta_k, \lambda_k, r_k)$  contradice esta restricción añadida. Así, este punto se corta y por lo tanto el nombre, algoritmo de planos cortantes.

---

**Algoritmo 4:** Planos cortantes modificado

---

**Datos:** Dado un conjunto compacto, convexo  $C$  y una tolerancia de parada  $\epsilon \geq 0$ .

- 1 [Inicialización]. Escoger  $(\lambda_0, r_0) \in C$  y resolver el problema  $\inf_x \ell(x, \lambda_0, r_0)$  para obtener  $x_0 := x_{\lambda_0, r_0}$ . Poner  $k = 0$ .
- 2 [Problema maestro]. Resolver la relajación (4.16), para obtener una solución  $(\theta_k, \lambda_k, r_k)$ .
- 3 [Problema local]. Resolver el problema en  $(\lambda_k, r_k)$ :

$$\inf_x \{ \ell(x, \lambda_k, r_k) : x \in X \},$$

para obtener el próximo punto primal  $x_k := x_{\lambda_k, r_k}$ .

- 4 [Criterio de parada] **if**

$$\theta_k \leq \bar{\psi}(\lambda_k, r_k) + \epsilon, \tag{4.18}$$

**then**

- 5 | parar

**else**

- 7 | reemplazar  $k$  por  $k + 1$  y retornar al paso 2

**end**

---

En este algoritmo nada nos garantiza que en cada iteración, el valor de la función dual crecerá.

## Relación con el trabajo de Gasimov

En [5], Gasimov estudió el lagrangiano aumentado sharp para problemas de minimización con restricciones en forma de igualdad. En esta sección realizamos una descripción en el contexto de nuestro problema (4.13) considerando la restricción en forma de igualdad, es decir, consideramos el siguiente problema

$$\begin{cases} \min_x & f_1(x) \\ \text{s.a} & x \in X \\ & \langle \mathcal{C}, x \rangle - c_2(x) = 0, \end{cases} \tag{4.19}$$

donde  $f_1, c_2$  son funciones convexas y lineal por partes. En este caso los cálculos se simplifican considerablemente. Dada la parametrización dualizante

$$f(x, u) = \begin{cases} f_1(x) & \text{si } x \in X \text{ y } \langle \mathcal{C}, x \rangle - c_2(x) = -u \\ +\infty & \text{en caso contrario} \end{cases},$$

por definición, el lagrangiano sharp viene dado por

$$\begin{aligned} \ell(x, \lambda, r) &= \inf_u \{ f(x, u) + r|u| - \langle \lambda, u \rangle \} = \inf_{-u = \langle \mathcal{C}, x \rangle - c_2(x)} \{ f(x, u) + r|u| - \langle \lambda, u \rangle \} \\ &= f_1(x) + r|\langle \mathcal{C}, x \rangle - c_2(x)| + \langle \lambda, \langle \mathcal{C}, x \rangle - c_2(x) \rangle \end{aligned}$$

Note también que

$$\sup_{(\lambda, r) \in \mathbb{R} \times (0, \infty)} \ell(x, \lambda, r) = \begin{cases} f_1(x) & \text{si } \langle \mathcal{C}, x \rangle - c_2(x) = 0 \\ +\infty & \text{en caso contrario} \end{cases}$$

de esto se sigue que el problema primal (4.19) puede expresarse de manera equivalente como

$$\inf_{x \in X} \sup_{(\lambda, r) \in \mathbb{R} \times (0, \infty)} \ell(x, \lambda, r),$$

tal y como lo afirma la primera igualdad en (3.14) del Teorema 3.30.

En el presente contexto, el siguiente resultado, ver [5, Teorema 5], puede ser utilizado como criterio de parada en el algoritmo del subgradiente modificado.

**Teorema 4.9.** *Si no existe salto de dualidad entre el problema (4.19) y su problema dual aumentado correspondiente considerando el lagrangiano sharp y para algún  $(\bar{\lambda}, \bar{r}) \in \mathbb{R} \times (0, \infty)$  y  $\bar{x} \in X$ ,*

$$\min_{x \in X} \ell(x, \bar{\lambda}, \bar{r}) = f_1(\bar{x}) + \bar{r} |\langle \mathcal{C}, \bar{x} \rangle - c_2(\bar{x})| + \langle \bar{\lambda}, \langle \mathcal{C}, \bar{x} \rangle - c_2(\bar{x}) \rangle, \quad (4.20)$$

entonces  $\bar{x}$  es una solución para (4.19) y  $(\bar{\lambda}, \bar{r})$  es una solución para su problema dual correspondiente si y sólo si

$$\langle \mathcal{C}, \bar{x} \rangle - c_2(\bar{x}) = 0. \quad (4.21)$$

**Prueba.** Si se cumple (4.20) y se satisface (4.21), entonces  $\bar{x}$  es factible para el problema (4.19), por lo tanto,  $\langle \mathcal{C}, \bar{x} \rangle - c_2(\bar{x}) = 0$ .

Recíprocamente si se cumplen (4.20) y (4.21) pero  $\bar{x}$  y  $(\bar{\lambda}, \bar{r})$  no son soluciones, entonces existe  $\tilde{x}$  factible para el problema primal tal que  $f_1(\tilde{x}) < f_1(\bar{x})$ . Por lo tanto

$$\begin{aligned} f_1(\tilde{x}) < f_1(\bar{x}) &= f_1(\bar{x}) + \bar{r} |\langle \mathcal{C}, \bar{x} \rangle - c_2(\bar{x})| + \langle \bar{\lambda}, \langle \mathcal{C}, \bar{x} \rangle - c_2(\bar{x}) \rangle \\ &= \min_{x \in X} \ell(x, \bar{\lambda}, \bar{r}) \leq \sup_{(\lambda, r) \in \mathbb{R} \times (0, \infty)} \min_{x \in X} \ell(x, \lambda, r) \\ &= \inf_{x \in X} \sup_{(\lambda, r) \in \mathbb{R} \times (0, \infty)} \ell(x, \lambda, r) \leq f_1(\tilde{x}), \end{aligned}$$

esto es una contradicción y concluye la prueba del teorema.  $\square$

Recordando que el problema dual viene dado por

$$\sup_{(\lambda, r) \in \mathbb{R} \times (0, \infty)} \bar{\psi}(\lambda, r),$$

donde  $\bar{\psi}(\lambda, r) := \inf_{x \in X} \ell(x, \lambda, r)$ , la Proposición 4.8 en este contexto adopta un formato simplificado

**Proposición 4.10.** *Si  $\bar{\psi}$  no es idénticamente  $-\infty$ , entonces para cualquier solución óptima  $x_{\lambda, r} \in X$  del problema (4.19) y para*

$$c(x_{\lambda, r}) := \langle \mathcal{C}, x_{\lambda, r} \rangle - c_2(x_{\lambda, r})$$

se tiene que

$$s(\lambda, r) = - \left( c(x_{\lambda, r}), |c(x_{\lambda, r})| \right) \in \partial(-\bar{\psi})(\lambda, r),$$

**Prueba.** Para cada  $\lambda_1$  y  $r_1$ , por definición de  $\bar{\psi}(\lambda_1, r_1)$ , tenemos que

$$\begin{aligned} \bar{\psi}(\lambda_1, r_1) &\leq f_1(x_{\lambda, r}) + \langle \lambda_1, c(x_{\lambda, r}) \rangle + r_1 |c(x_{\lambda, r})| \\ &= \bar{\psi}(\lambda, r) + \langle c(x_{\lambda, r}), \lambda_1 - \lambda \rangle + |c(x_{\lambda, r})| (r_1 - r) \\ &= \bar{\psi}(\lambda, r) + \left\langle \left( \begin{array}{c} c(x_{\lambda, r}) \\ |c(x_{\lambda, r})| \end{array} \right), \left( \begin{array}{c} \lambda_1 - \lambda \\ r_1 - r \end{array} \right) \right\rangle \end{aligned}$$



luego,  $-(c(x_{\lambda,r}), |c(x_{\lambda,r})|) \in \partial(-\bar{\psi})(\lambda, r)$ , como se queria mostrar.  $\square$

En el contexto del problema (4.19) y con estos resultados, el Algoritmo 3 puede ser mejorado del siguiente modo

- En el paso 2 ahora tenemos un criterio de parada, si  $c(x_{\lambda,r}) = 0$  parar, por el Teorema 4.9,  $(\lambda, r)$  es una solución para el problema dual y  $x_{\lambda,r}$  es una solución para el problema primal.
- En el paso 5 modificamos la actualización del siguiente modo

$$\lambda_{k+1} = \lambda_k + t_k c(x_k), \quad r_{k+1} = r_k + (t_k + \epsilon_k) |c(x_k)|, \quad (4.22)$$

donde  $t_k, \epsilon_k$  son escalares positivos que representan los tamaños de pasos.

El siguiente resultado, ver [5, Teorema 7], muestra que a diferencia del método del subgradiente desarrollado para problemas duales formulados utilizando el lagrangiano clásico la nueva iterada mejora estrictamente para todos los valores de los tamaños de paso  $t_k$  y  $\epsilon_k$

**Teorema 4.11.** *Supongamos que el par  $(\lambda_k, r_k) \in R \times (0, +\infty)$  no es una solución para el problema dual y  $x_k \in \operatorname{argmin} \ell(x, \lambda_k, r_k)$ , entonces para la nueva iterada  $(\lambda_{k+1}, r_{k+1})$  calculada por (4.22) para todo tamaño de paso  $t_k$  y  $\epsilon_k$  tenemos*

$$0 < \bar{\psi}(\lambda_{k+1}, r_{k+1}) - \bar{\psi}(\lambda_k, r_k) \geq (2t_k + \epsilon_k) |c(x_k)|^2.$$

El siguiente resultado nos dice que para ciertos valores de los tamaños de pasos  $t_k$  y  $\epsilon_k$ , la distancia entre los puntos generado por el algoritmo y la solución del problema dual decrece en cada iteración (ver [10, Proposición 6.3.1])

**Teorema 4.12.** *Sea  $(\lambda_k, r_k)$  cualquier iteración, que no sea solución para el problema dual, luego  $c(x_k) \neq 0$ , entonces para cualquier solución dual  $(\bar{\lambda}, \bar{r})$  tenemos*

$$|(\bar{\lambda}, \bar{r}) - (\lambda_{k+1}, r_{k+1})| < |(\bar{\lambda}, \bar{r}) - (\lambda_k, r_k)|$$

para todos los tamaños de paso  $t_k$  tal que

$$0 < t_k < \frac{2(\bar{\psi}(\bar{\lambda}, \bar{r}) - \bar{\psi}(\lambda_k, r_k))}{5|c(x_k)|^2} \quad y \quad 0 < \epsilon_k < t_k.$$

Análogamente al caso del algoritmo de subgradiente, en [5], considerando (4.19) como problema primal, se hace una modificación del Algoritmo 4 y se presenta un algoritmo combinado que contiene iteraciones del método de planos cortantes usual (calculado para el problema dual construido respecto al lagrangiano sharp) e iteraciones del método de subgradiente modificado descrito en la sección anterior.

Esta modificación consiste en tomar en cada iteración el valor más reciente de la función dual e incrementando este al aplicar las iteradas del método de subgradiente modificado descrito. Asimismo, en este caso, se tiene un nuevo criterio de parada, puesto que (4.17) es solamente una condición suficiente mas no necesaria para la optimalidad, puede suceder que el procedimiento se convierta en la solución óptima  $(\lambda_k, r_k)$ , pero (4.17) aún no se satisfaga y por lo tanto el algoritmo continua realizando nuevas iteraciones. Esta situación logra solucionarse utilizando la condición necesaria y suficiente para la optimalidad presentado en el Teorema 4.9.

A continuación realizamos la formulación dual generalizada de un problema de manejo de energía.

## Aplicación a un problema de energía y perspectivas futuras

En este trabajo se ha estudiado la teoría de dualidad clásica y la dualidad en un marco general utilizando la transformada de Fenchel, resaltando la importancia de esta última debido a su capacidad para eliminar el salto de dualidad aún en problemas no convexos. En particular se ha puesto atención especial a problemas de optimización *DC* debido a su importancia en aplicaciones prácticas.

Un trabajo futuro pendiente es la implementación orientada a problemas de despacho energético utilizando una aproximación numérica via el lagrangiano generalizado.

A continuación se explica el modelo binivel considerado para manejar de modo eficiente una “smart grid”.

### 5.1. Un problema de manejo de energía

En optimización energética, el problema llamado *despacho de corto plazo* permite producir energía eléctrica durante un período de 24 ó 48 horas, coordinando la producción de un conjunto de generadores eléctricos de modo a satisfacer la demanda del sistema con costo mínimo<sup>1</sup>.

En contraste, los problemas llamados de “unit commitment” en la literatura consideran para cada período de tiempo y para cada usina, también cuales generadores deben ser puestos en marcha y cuales apagados. Modelar este tipo de acción requiere variables 0-1 para representar los estados “prendido” y “apagado”. En este trabajo no son consideradas variables binarias, habiéndose preferido representar incertezas en la generación renovable (por fuentes solares o eólicas intermitentes). En el contexto de optimización binivel presentado a continuación, el modelo estocástico en variables continuas ya complica suficientemente el problema de optimización que debe ser resuelto. La inclusión de variables 0-1 puede ser una extensión interesante de la investigación iniciada con este trabajo.

---

<sup>1</sup>o maximizando los ingresos de los generadores de energía

## 5.2. Formulación binivel

Nos concentraremos en estudiar un modelo de interacción entre un actor global y un actor local representando una “smart grid”. El actor local dispondrá de diversas tecnologías de generación de energía, tales como paneles solares o generación eólica, e intentará satisfacer la demanda local (de los componentes de la “smart grid”). Por su parte, el actor global representa un gran productor de energía, que dispone de varias usinas eléctricas y es responsable de mantener el balance entre generación y demanda en la red de transmisión de todo el sistema eléctrico considerado.

La interacción entre los dos tipos de agentes se da a través de la comercialización de energía eléctrica, ya sea porque el agente local no tuvo suficiente producción propia y debe comprar energía de la red, o porque la “smart grid” tuvo una generación mayor a su demanda, y le vende el excedente al actor global, para uso en la red “nacional”.

La energía intermitente tiene efectos indeseables en la calidad de la electricidad, pues causa variaciones bruscas de voltaje, que no fueron previstas al diseñar la red de transmisión. Por ejemplo, ráfagas intensas de viento aumentan subrepticia y desproporcionadamente el aporte de energía eólica. De modo similar, cuando el sol se pone, quienes durante el día eran productores de energía solar y no usaban luz del sistema, se tornan consumidores, todos al mismo tiempo, y en un horario de punta. Para atenuar estos vaivenes, el actor global aplica un dispositivo llamado “peak-shaving”, que le permite desplazar parte de la demanda del actor local que llega en horarios de punta para otro momento, menos crítico, porque la red estará menos congestionada. En compensación por la potencial pérdida parcial del servicio, el actor local recibe tarifas preferenciales para la compra de energía en algunas horas.

En líneas generales, el actor global optimiza la gestión global de energía del sistema, de modo a mantener la red eléctrica en equilibrio, satisfaciendo la carga/demanda en las barras de la red, con mínimo costo. Como el sistema contiene fuentes de energía intermitente, la demanda en cada paso de tiempo no se conoce de antemano, y está representada en el modelo por un conjunto de escenarios,  $\{d_s^g : s \in S\}$ , con probabilidad asociada  $\pi_s > 0$  satisfaciendo  $\sum_{s \in S} \pi_s = 1$ .

Por su parte, en función del precio fijado por el actor global para la compra y venta de energía, el agente local optimiza el funcionamiento de su “smart grid”, para un valor de demanda estimada en función de su conocimiento de los clientes locales, denotada por  $d^l$ , un valor determinístico.

Dependiendo de cuanto viento sople en un determinado momento, deberán efectuarse correcciones para ajustar la generación (y no producir energía de fuentes no eólicas, más caras, si hay suficiente viento). Para representar este fenómeno, algunas componentes del vector de decisión son variables aleatorias. El modelo considerado para tratar la incerteza funciona en dos etapas, con recurso (“two-stage with recourse” en inglés, [11]) Más precisamente, independientemente de la realización de la incerteza, tanto el agente global como el local deciden los valores de la generación e intercambios para un primer conjunto de variables, denominadas “first stage”, o “here-and-now”. En la segunda etapa, una vez que la generación incierta se torna conocida, *solamente* el actor global corrige sus decisiones de primera etapa, para ajustarlas a la realización de la incerteza, generando por ejemplo más electricidad en algunas usinas. Este segundo grupo de variables son llamadas de recurso, o de tipo “wait-and-see”.

Para un horizonte temporal con  $T$  periodos, se adopta la notación siguiente:

- $d \in \mathbb{R}^T$  representa la demanda total en cada paso de tiempo. Por simplicidad, se supone que no hay déficit de energía: la carga de todo el sistema es la suma algebraica de la carga del actor global y el actor local, es decir,  $d = d^g + d^l$ , con  $d^g \in \{d_s^g : s \in S\}$ .
- La generación de energía proviene de las  $m$  fuentes disponibles en el sistema, de las cuales las indexadas  $\{1, \dots, m_g\}$  son propiedad del actor global, en tanto que los índices  $\{m_g + 1, \dots, m\}$  corresponden al actor local.
- los vectores  $x_1, \dots, x_m$  representan decisiones relacionadas con la generación de la  $i$ -ésima planta o tecnología, para  $i = 1, \dots, m$  (nuclear, de carbón, hidroeléctrica, solar, eólica),
- Los conjuntos  $\mathcal{X}^1, \dots, \mathcal{X}^m$  contienen las restricciones tecnológicas asociadas con estas decisiones.
- $x_0$  es el vector de gestión de demanda, con costo  $c_0^x$  y conjunto de restricciones  $\mathcal{X}^0$ .
- Las funciones de costo  $c_0^x, \dots, c_m^x$  son lineales.
- $A^0, \dots, A^m$  son matrices tales que  $A^i x_i$  representa el efecto de una decisión dada sobre la cantidad de energía generada. De modo general, la ecuación de satisfacción de la demanda se expresa como una restricción afín de igualdad de la forma

$$\sum_{j=0}^m A^j x_j = d.$$

- Los intercambios de energía adoptan dos formas:
  - cuando el actor local tiene un excedente  $y^e$ , éste es comprado a precio  $c^e \in \mathbb{R}^T$  por el actor global, hasta una cota  $\bar{y}^e$ .
  - Si, en cambio, el actor local tiene escasez de energía generada con fuentes propias (no puede satisfacer la demanda en su smart grid), compra la electricidad faltante  $y^l$  del actor global, a precio  $c^l \in \mathbb{R}^T$ , hasta una cota  $\bar{y}^l$ .

En la formulación binivel el actor local debe resolver el siguiente problema de optimización:

$$\left\{ \begin{array}{l} \min_{x_{m_g+1}, \dots, x_m, y^e, y^l} \quad \sum_{i=m_g+1}^m \langle c_i^x, x_i \rangle - \langle c^e, y^e \rangle + \langle c^l, y^l \rangle \\ \text{s.a.} \quad x_j \in \mathcal{X}^j, \quad j = m_g + 1, \dots, m \\ y^e \in [0, \bar{y}^e] \subset \mathbb{R}^T, \quad y^l \in [0, \bar{y}^l] \subset \mathbb{R}^T \\ y^l - y^e + \sum_{j=m_g+1}^m A^j x_j = d^l - \sum_{j \in L} A^j x_j - A^0 x_0. \end{array} \right.$$

En la última ecuación, de demanda local, el conjunto  $L \subset \{1, \dots, m_g\}$  representa las usinas del actor global que pueden generar electricidad en la “smart grid”. En

consecuencia, como las variables  $x_0$  y  $\{x_j : j \in L\}$  son decididas en el nivel superior, por el actor global, el problema del agente local tiene la forma abstracta

$$c_2(\mathbf{u}^1) := \begin{cases} \min_{\mathbf{v}} & \langle \mathcal{C}, \mathbf{v} \rangle \\ \text{s.a.} & \mathbf{v} \in \mathbf{V} \\ & M\mathbf{v} = d^l - N\mathbf{u}^1. \end{cases} \quad (5.1)$$

donde

- $\mathbf{u}^1 := (x_0, \{x_j\}_{j \in L})$ ,  $\mathbf{u}^1 := \mathcal{X}_0 \times \prod_{j \in L} \mathcal{X}_j$ ,
- $\mathbf{v} := (\{x_j\}_{j=m_g+1}^m, y^e, y^l)$ ,  $\mathbf{V} := \prod_{j=m_g+1}^m \mathcal{X}_j \times \mathbb{R}^T \times \mathbb{R}^T$ ,
- $\mathcal{C} := (\{c_i^x\}_{i=m_g+1}^m, -c^e, c^l)$ ,
- $M := (\{A^j\}_{j=m_g+1}^m, -I_{T \times T}, I_{T \times T})$ ,
- $N := (A^0, \{A^j\}_{j \in L})$ .

Supondremos que el problema del agente local (5.1) tiene recurso completo. Para nuestra aplicación esto significa que la capacidad de generación del actor global y el actor local siempre conseguirán satisfacer la demanda, para cualquier escenario considerado. Esta hipótesis nos permite asegurar que el problema primal es factible y que hay un dual.

**Lema 5.1** (Propiedades del problema de nivel inferior). *El problema del agente local (5.1), considerando su parametrización en las variables “here-and-now” de primer nivel,  $\mathbf{u}^1$ , define una función convexa, lineal por partes y con subdiferencial dado por*

$$\partial c_2(\mathbf{u}_0^1) = -N \operatorname{argmax}\{\langle d^l - N\mathbf{u}_0^1, \pi \rangle : M^\top \pi \leq \mathcal{C}\}$$

**Prueba.** El problema dual de (5.1) viene dado por

$$\begin{aligned} \max_{\pi} & \langle d^l - N\mathbf{u}^1, \pi \rangle \\ \text{s.a.} & M^\top \pi \leq \mathcal{C}. \end{aligned}$$

definiendo

$$\nu(z) := \inf\{\langle \mathcal{C}, \mathbf{v} \rangle : M^\top \mathbf{v} = z, \text{ para } \mathbf{v} \in \mathbf{V}\}$$

entonces  $c_2(\mathbf{u}^1) = \nu(d^l - N\mathbf{u}^1)$ . Si el conjunto primal factible es diferente del vacío y el dual también es factible entonces  $\nu(z)$  es finito y

$$\nu(z) = \begin{cases} \min_{\mathbf{v}} & \langle \mathcal{C}, \mathbf{v} \rangle \\ \text{s.a.} & \mathbf{v} \in \mathbf{V} \\ & M\mathbf{v} = z. \end{cases} = \begin{cases} \max_{\pi} & \langle z, \pi \rangle \\ \text{s.a.} & M^\top \pi \leq \mathcal{C}, \end{cases}$$

de esta expresión, siendo  $\nu$  el máximo de funciones afines, concluimos que la función  $c_2$  es convexa, lineal por partes. Por otro lado, como  $\Pi = \Pi(\mathcal{C}) := \{\pi : \pi \in M^\top \pi \leq \mathcal{C}\}$

es un poliedro entonces  $\Pi = \text{co}\{w^1, \dots, w^L\} + \Pi_\infty$ , donde  $\Pi_\infty$  es un cono de recesión, luego

$$\begin{aligned} \nu(z) &= \max_{\pi} \langle z, \pi \rangle \\ &\text{s.a. } \pi = \sum_{l=1}^L \alpha_l w^l + d. \quad = \max_{\alpha} \langle z, \sum_{l=1}^L \alpha_l w^l \rangle \\ &\quad \text{s.a. } \alpha \in \{\alpha_l \geq 0, \sum_{l=1}^L \alpha_l = 1\} + \max_{\text{s.a.}} \langle z, d \rangle \\ &\quad \text{s.a. } M^\top d \leq 0. \\ &= \max_{\text{s.a.}} \langle z, w^l \rangle \quad + \begin{cases} 0 & \text{si } \langle d, z \rangle \leq 0 \\ +\infty & \text{si } \langle d, z \rangle > 0, \end{cases} \end{aligned}$$

luego  $\partial\nu(z_0) = \text{co}\{w^l : \langle z_0, w^l \rangle = \nu(z_0)\} = \text{argmax}\{\langle z_0, \pi \rangle : \pi \in \Pi\}$ , de donde aplicando la regla de la cadena obtenemos

$$\partial c_2(u_0^1) = \partial\nu(d^l - Nu_0^1) = -N \text{argmax}\{\langle d^l - Nu_0^1, \pi \rangle : \pi \in \Pi\}.$$

□

Recordando que el agente global usa variables de recurso para corregir las decisiones “here-and-now”, el problema de optimización del nivel superior se escribe como sigue:

$$\begin{cases} \min_{x,y} \langle c_0^x, x_0 \rangle + \langle c^e, y^e \rangle - \langle c^l, y^l \rangle + \sum_{s \in S} \pi_s Q^s(y) \\ \text{s.a. } x_j \in \mathcal{X}^j, j = 0, \dots, m_g \\ (x_{m_g+1}, \dots, x_m, y^e, y^l) \in \text{argmin} \end{cases} \quad (5.1)$$

con

$$Q^s(y) := \begin{cases} \min_{x_1^s, \dots, x_{m_g}^s} \sum_{j=1}^{m_g} \langle c_j^x, x_j^s \rangle \\ \text{s.a. } x_j^s \in \mathcal{X}^j, j = 1, \dots, m_g \\ \sum_{j=1}^{m_g} A^j x_j^s = d_s^g - y^e + y^l. \end{cases}$$

En optimización estocástica la función  $Q^s(y)$ , que representa el impacto de las correcciones realizadas para ajustarse a la realización  $s$ -ésima de la demanda incierta, se llama *función de recurso*. Por simplicidad, empleando nuevamente una formulación abstracta, resulta el problema del agente global

$$\begin{cases} \min_{u^1, v} \langle C^1, u^1 \rangle + \sum_{s \in S} \pi_s Q^s(v) \\ \text{s.a. } u^1 \in U^1 \\ v \in \text{argmin} \end{cases} \quad (5.2)$$

En formato abstracto, la función de recurso es

$$Q^s(v) := \begin{cases} \min_{u^2} \langle C^2, u^2 \rangle \\ \text{s.a. } u^2 \in U^2 \\ Wu^2 = d_s^g - Tv, \end{cases} \quad (5.3)$$

donde definimos  $U^2 := \prod_{j=1}^{m_g} \mathcal{X}^j$ , y las matrices  $W := \{A^j\}_{j=1}^{m_g}$  y  $T := [-I, I]$

**Lema 5.2** (Propiedades de la función de recurso). *Para cada escenario de demanda  $d_s^g$  dado y para cada vector de segundo nivel  $v$ , la función de recurso definida en (5.3) es convexa, lineal por partes y con subdiferencial dado por*

$$\partial Q^s(v_0) = -T \text{argmax}\{\langle d_s^g - Tv_0, \pi \rangle : W^\top \pi \leq C^2\}.$$

**Prueba.** La prueba es idéntica a la del Lema 5.1

□

Siguiendo la transformación de la función de valor óptimo debida a [4], reescribimos en el problema (5.1)-(5.2) la restricción “abstracta” como sigue

$$\mathbf{v} \in \operatorname{argmin}(5.1) \iff \langle \mathcal{C}, \mathbf{v} \rangle \leq c_2(\mathbf{u}^1) \quad \text{para todo } \mathbf{v} \in \mathbf{V} \text{ tal que } M\mathbf{u}^1 + N\mathbf{v} = d^l.$$

Resulta el problema

$$\begin{cases} \min_{\mathbf{u}^1, \mathbf{v}} & \langle \mathcal{C}^1, \mathbf{u}^1 \rangle + \sum_{s \in S} \pi_s \mathcal{Q}^s(\mathbf{v}) \\ \text{s.a} & \mathbf{u}^1 \in \mathbf{U}^1, \mathbf{v} \in \mathbf{V} \\ & M\mathbf{u}^1 + N\mathbf{v} = d^l \\ & \langle \mathcal{C}, \mathbf{v} \rangle \leq c_2(\mathbf{u}^1). \end{cases} \quad (5.4)$$

La función objetivo es convexa y lineal por partes, por el Lema 5.2, en tanto que la última restricción es de tipo DC (diferencia de una función lineal y otra convexa lineal por partes), por el Lema 5.1.

### 5.3. Formulación dual generalizada del problema

Para poder aplicar las técnicas de dualidad no convexa a la restricción DC, aligeramos la notación una vez más, definiendo

$$x := (\mathbf{u}^1, \mathbf{v}), \quad X := \mathbf{U}^1 \times \mathbf{V} \cap \{(\mathbf{u}^1, \mathbf{v}) : M\mathbf{u}^1 + N\mathbf{v} = d^l\}$$

la función objetivo

$$f_1(x) := \langle \mathcal{C}^1, \mathbf{u}^1 \rangle + \sum_{s \in S} \pi_s \mathcal{Q}^s(\mathbf{v})$$

y escribimos la restricción en variable  $x$ , para trabajar con el problema

$$\begin{cases} \min_x & f_1(x) \\ \text{s.a} & x \in X \\ & \langle \mathcal{C}, x \rangle - c_2(x) \leq 0. \end{cases} \quad (5.5)$$

Asociando un multiplicador escalar  $\lambda \geq 0$  a esta restricción resulta un lagrangiano

$$L(x, \lambda) = f_1(x) + I_X(x) + \lambda \left( \langle \mathcal{C}, x \rangle - c_2(x) \right).$$

Por el Ejemplo 3.29 el lagrangiano Sharp tiene la forma

$$\ell(x, \lambda, r) = L(x, \lambda) + r |\langle \mathcal{C}, x \rangle - c_2(x)| - [r - y]^+ [\langle \mathcal{C}, x \rangle - c_2(x)]^-$$

En efecto considerando la norma dual de la norma 1 (en este caso el valor absoluto) es la norma infinito, la expresión en el Ejemplo 3.29

$$\ell(x, \lambda, r) = \sup_{\|z-y\|^\circ \leq r} L(x, z)$$

resulta

$$\begin{aligned} \ell(x, \lambda, r) &= \begin{cases} \sup & L(x, z) \\ \text{s.a} & -r + \lambda \leq z \leq r + \lambda. \end{cases} \\ &= L(x, \lambda) + r |\langle \mathcal{C}, x \rangle - c_2(x)| - [r - \lambda]^+ [\langle \mathcal{C}, x \rangle - c_2(x)]^- . \end{aligned}$$

Como no tenemos disponible explícitamente ni  $f_1$  ni  $c_2$ , en la  $k$ -ésima iteración tendremos las aproximaciones de planos cortantes

$$\check{f}_1^k(x) = \max\{f_1(x^i) + \langle g^i, x - x^i \rangle : i \in I_f^k\} \quad \text{y} \quad \check{c}_2^k(x) = \max\{c_2(x^i) + \langle \gamma^i, x - x^i \rangle : i \in I_c^k\},$$

para conjuntos de índices  $I_f^k, I_c^k \subset \{1, \dots, k-1\}$ . Como tanto  $f_1$  y  $c_2$  son convexas,  $\check{f}_1^k \leq f_1$  y  $\check{c}_2^k \leq c_2$ , más aún por ser lineales por partes, las aproximaciones coinciden con las funciones a partir de algún  $k$ .

Para las lagrangianas resultan las aproximaciones

$$L^k(x, \lambda) = \check{f}_1^k(x) + I_X(x) + \lambda(\langle \mathcal{C}, x \rangle - \check{c}_2^k(x))$$

y

$$\ell^k(x, \lambda, r) = L^k(x, \lambda) + r|\langle \mathcal{C}, x \rangle - \check{c}_2^k(x)| + [r - \lambda]^- [\langle \mathcal{C}, x \rangle - \check{c}_2^k(x)]^-,$$

luego, la aproximación de la función dual generalizada en la  $k$ -ésima iteración viene dada por:

$$\Theta^k(\lambda, r) := \min\{\ell^k(x) : x \in X\}$$

Denotando por  $w = \langle \mathcal{C}, x \rangle - \check{c}_2^k(x)$ , y teniendo en cuenta que  $w = w^+ - w^-$  y que  $|w| = w^+ + w^-$  la función dual generalizada adopta la forma

$$\min_{x \in X} \{\check{f}_1^k(x) + (r + \lambda)[\langle \mathcal{C}, x \rangle - \check{c}_2^k(x)]^+ - (r - \lambda)^- [\langle \mathcal{C}, x \rangle - \check{c}_2^k(x)]^-\} \quad (5.6)$$

## 5.4. Aproximaciones en la resolución del problema

Considerando el cambio de variable  $y = \check{f}_1^k(x)$ ,  $z = \mathcal{C}(x) - \check{c}_2^k(x)$  en (5.6), obtenemos el problema lineal

$$\begin{aligned} \min_{x, y, z^+, z^-} \quad & 0x + y + (r + \lambda)z^+ - (r - \lambda)^- z^- \\ \text{s.a.} \quad & y \geq f_1^i + \langle g^i, x - x^i \rangle, \quad i = 1, \dots, k \\ & -z \geq -\langle \mathcal{C}, x \rangle + c_2^i + \langle \gamma^i, x - x^i \rangle, \quad i = 1, \dots, k \\ & Bx \geq d \\ & (x, y, z^+, z^-) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \end{aligned} \quad (5.7)$$

que escribiendo matricialmente resulta

$$\min \langle b, w \rangle \quad \text{s.a.} \quad Cw \geq h, \quad w = (x, y, z^+, z^-) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+,$$

con

$$b := (0, 1, (r + \lambda), -(r - \lambda)^-), \quad h := (f_1^k, c_2^k, d), \quad C = \begin{pmatrix} -G^k & 1^k & 0_{k \times 1} & 0_{k \times 1} \\ 1^k c_1^T - \Gamma^k & 0_{k \times 1} & -1^k & 1^k \\ B_{m \times n} & 0_{m \times 1} & 0_{m \times 1} & 0_{m \times 1} \end{pmatrix},$$

donde  $G^k$  y  $\Gamma^k$  son matrices de orden  $k \times n$  con los  $k$  gradientes  $g^i$  y  $\gamma^i$  en sus filas respectivas;  $1^k$  es un vector columna de orden  $k \times 1$  cuyas entradas son unos;  $f_1^k$  y  $c_2^k$  son vectores en  $\mathbb{R}^k$  con componentes respectivos definidos por

$$\begin{aligned} (f_1^k)_i &:= f_1(x_i) - \langle g^i, x^i \rangle \\ (c_2^k)_i &:= c_2(x^i) - \langle \gamma^i, x^i \rangle \end{aligned}, \quad i = 1, \dots, k$$



El problema dual de (5.7) viene dado por:

$$\begin{aligned}
& \max \quad \langle h, \alpha \rangle \\
& \text{s.a.} \quad \alpha_1(-G^k) + \alpha_2(1^k c_1^\top - \Gamma^k) + \alpha_3 B = 0 \\
& \quad \alpha_1 1^k = 1 \\
& \quad \alpha_2(-1^k) \leq (r + \lambda) \\
& \quad \alpha_2 1^k \leq -(r - \lambda)^- \\
& \quad \alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}_+^k \times \mathbb{R}_+^k \times \mathbb{R}_+^m.
\end{aligned} \tag{5.8}$$

**Observación 5.3.** *A partir del problema dual (5.8), podemos obtener información relevante de nuestro problema primal (5.7). Tenemos dos casos:*

- *Si  $-(r - \lambda)^- < 0$  ( $\Leftrightarrow \lambda > r$ ), entonces  $\alpha_2 1^k < 0$ ; dando como resultado un problema dual infactible, pues  $p \geq 0$  implica que  $\alpha_2 1^k \geq 0$ , luego el problema primal (5.7) será no acotado o no factible. Este caso es de poco interés.*
- *Si  $-(r - \lambda)^- = 0$  ( $\Leftrightarrow \lambda \leq r$ ), en este caso obtenemos  $\alpha_2 = 0$ , pues  $\alpha_2 \geq 0$  y la suma de sus componentes es menor o igual que cero.*

**Lema 5.4.** *Dados dos números reales  $z$  y  $b$ , denotando por  $z^+, b^+$  y  $z^-, b^-$  la parte positiva y negativa de  $z$  y  $b$  respectivamente, entonces  $z \leq b$  si y sólo si  $z^+ \leq b^+$  y  $z^- \geq b^-$*

**Prueba.** Si  $z \leq b$  tenemos que  $z^+ - z^- \leq b^+ - b^-$ , consideremos los tres casos posibles: si  $b = 0$  se cumple la conclusión. Si  $b > 0$  entonces  $z^+ \geq b^+ = 0$  y en cualquier de los subcasos  $|z| \leq b$  ó  $z \leq -b$  obtenemos  $z^+ \leq b^+$ . Finalmente si  $b < 0$  entonces  $|z| \geq |b|$  de donde obtenemos  $-z^+ - z^- \leq -b^+ - b^-$  y sumando con  $z^+ - z^- \leq b^+ - b^-$  implica que  $z^- \geq b^-$ , por otro lado dado que  $b < 0$  también se cumple que  $z^+ \geq b^+ = 0$ . Por lo tanto en cualquier caso se cumple  $z^+ \leq b^+$  y  $z^- \geq b^-$ . La otra implicación se deduce directamente por definición de parte positiva y parte negativa de un número real.  $\square$

**Observación 5.5.** *Aplicando el Lema 5.4 a las restricciones  $-z \geq -\langle c_1, x \rangle + c_2^i + \langle \gamma^i, x - x^i \rangle, i = 1, \dots, k$  del problema lineal (5.7) obtenemos.*

$$(\lambda + r)^+ z^+ \leq (\lambda + r)^+ [\langle \mathcal{C}, x \rangle - c_2^k(x)]^+ \quad y \quad -(\lambda - r)^- z^- \leq -(\lambda - r)^- [\langle \mathcal{C}, x \rangle - c_2^k(x)]^-.$$

**Proposición 5.6.** *Si  $(x_0, y_0, z_0^+, z_0^-)$  es solución óptima del problema lineal (5.7), entonces su valor óptimo es una cota inferior del problema (5.6). Más aún  $y_0 = \check{f}_1^k(x_0)$*

**Prueba.** Como  $(x_0, y_0, z_0^+, z_0^-)$  es solución óptima de (5.7), entonces

$$s_0 := y_0 + (r + \lambda)z_0^+ - (r - \lambda)^- z_0^- \leq v + (r + \lambda)z^+ - (r - \lambda)^- z^-,$$

para todo  $(x, y, z^+, z^-)$  factible para (5.7). Más aún  $y_0 = \check{f}_1^k(x_0)$  (en caso contrario, obtenemos una contradicción con la minimalidad de  $s_0$ ).

Por otro lado, si existiera  $x_1 \in X$ :

$$\check{f}_1^k(x_1) + (r + \lambda)[\langle \mathcal{C}, x_1 \rangle - \check{c}_2^k(x_1)]^+ - (r - \lambda)^- [\langle \mathcal{C}, x_1 \rangle - \check{c}_2^k(x_1)]^- < s_0$$

obtendríamos una contradicción, pues  $(x_1, y_1 := \check{f}_1^k(x_1), z_1^+ := [\langle \mathcal{C}, x_1 \rangle - \check{c}_2^k(x_1)]^+, z_1^- := [\langle \mathcal{C}, x_1 \rangle - \check{c}_2^k(x_1)]^-)$  es factible para (5.7) y  $s_0$  es el valor mínimo, en consecuencia para todo  $x \in X$  se cumple

$$\check{f}_1^k(x) + (r + y)(\langle \mathcal{C}, x \rangle - \check{c}_2^k(x))^+ - (r - y)^-(\langle \mathcal{C}, x \rangle - \check{c}_2^k(x))^- \geq \check{f}_1^k(x_0) + (r + y)z_0^+ - (r - y)^- z_0^-$$

$\square$

La próxima etapa del trabajo será la implementación a un problema real como el empleado en [3].

## Bibliografía

- [1] Jean-Baptiste Hiriart-Urruty; Claude Lemaréchal. *Convex Analysis and Minimization Algorithms II*. Springer-Verlag Berlin Heidelberg. 1993.
- [2] R. Tyrrell Rockafellar; Roger J-B Wets. *Variational Analysis*. Springer-Verlag Berlin Heidelberg. 1998.
- [3] W. van Ackooij and W. de Oliveira, *DC programming techniques with inexact subproblem' solution for general DC programs*. 2017
- [4] J. V. Outrata, *On the numerical solutions of a class of stackelberg games*. Z. Operations Research 34(1990), pp. 255-277.
- [5] R.N. Gasimov, *Augmented Lagrangian Duality and nondifferentiable optimization methods in nonconvex programming*. J. Global. Optim. 24(2002), pp. 187-203.
- [6] R.S. Burachik, R.N. Gasimov, N. A. Ismayilova, and C. Kaya, *On a modified subgradient algorithm for dual problems via sharp augmented lagrangian*, J. Global Optim. 34(2006), pp. 55-78
- [7] R.S. Burachik, C. Kaya, and M. Mammadov, *An inexact modified subgradient algorithm for nonconvex optimization*, Comput. Optm. Appl. 45 (2010), pp. 1-24.
- [8] R. Burachik, R.S., Iusen, A. N., Melo, G. J., *A primal dual modified subgradient algorithm with sharp lagrangian*, J. Global Optim. 46 (2010), pp. 347-361.
- [9] R. Kasimbeyli, O. Ostun, and A. M. Rubinov, *The modified subgradient algorithm based on feasible values*, Optimization 58(2009), pp. 535-560
- [10] D. Bertsekas, *Nonlinear programming* 2nd ed. Athena Scientific, Nashua, 1999.
- [11] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming. Modeling and Theory*. Philadelphia 2009.
- [12] J. F. Bonnans, J.C. Gilbert, C. Lemaréchal, C. Sagastizábal, *Numerical Optimization theoretical and practical aspects*. Springer-Verlag Berlin Heidelberg 2006.

- [13] J. P. Crouzeix, E. Ocaña, *Métodos algorítmicos Haces-Proximal-Lagrangiano aumentado para problemas de Optimización Matemática*. Monografía del IMCA 2011.
- [14] J. E. Martínez Legaz, *Generalized Convex Duality and its Economic Applications*. Monografía del IMCA 2002.