Instituto Tecnológico Autónomo de México

Ecuaciones lineales diofantinas aplicadas a programas lineales enteros

Tesis

para obtener el título de

Licenciado en Matemáticas Aplicadas

presenta

Iñaki Sebastian Liendo Infante

Asesor

Dr. Andreas Wachtel

Ciudad de México 2025

"Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada "**Ecuaciones lineales diofantinas aplicadas a programas lineales enteros**", otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Bailléres Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación".

Iñaki Sebastian Liendo Infante

Fecha

Firma

Agradecimientos

> ¿Cómo iba yo a saber que la acumulación de esos "mañana" que ni siquiera distinguía, y que sin notarlo ya eran "hoy" y "ayer", harían de pasar no solo el tiempo, sino mi tiempo, el único mío?
>
> *Josefina Vicens, El libro vacío*

A mi hermano Cristóbal. A mis padres Javier y Yanitzin. A mis abuelos Isidoro y Gloria, Germán y Virginia. A mis tíos y a todos mis primos. Agradezco en lo más profundo de mi corazón tener el privilegio de llamarlos familia. Mi hogar es dondequiera que ustedes se encuentren.

A mis amigos María José Borges, Fernando Yedra, Hugo Nava, Luis Alfonso Maciel, Mauricio Pazos, Hazel Yáñez, Micho Altamirano, Santiago Prado, José Luis Bravo, Annia Pi-Suñer, Alejandro Naranjo, Pedro Olivares, Adrián González, Orlando Almazán, Ricardo Bravo, Irvin Ramírez, Adrián Berrón, Mauricio Díaz, Carlos Rivera, Rodrigo Arjona, Miguel Ángel Torres y Ariadna Irena. Búsquenme si quieren su agradecimiento personalizado, pues es buena excusa para salir y crear más memorias.

A mi asesor Andreas Wachtel por la inagotable paciencia con la que ha transformado una colección de símbolos inconexos en una tesis coherente. A mis sinodales Edgar Possani, Edith Vargas García, y Ezequiel Soto Sánchez por el tiempo que dedicaron a revisar este trabajo.

A todos los que he tenido la fortuna de conocer pero que nuestros caminos han sido interrumpidos. Las olas de nuestros recuerdos compartidos me atrapan en una marea de nostalgia y melancolía.

Nomenclatura

Abreviaciones

R&A Ramificación y Acotamiento

FPD Formulación de programación dinámica

CBC COIN-OR Branch and Cut

Símbolos reservados

$\boldsymbol{p}$ Vector esencialmente entero (ver definición 1.2.1).

$\boldsymbol{q}$ Vector coprimo asociado a $\boldsymbol{p}$ (ver definición 1.2.1).

$H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ k-ésima capa entera con parámetro k entero (ver definición 1.2.3).

η Entero que parametriza la primera capa entera en satisfacer la restricción presupuestaria.

$\boldsymbol{\nu}$ Vector entero (ver definición (1.37)).

M Matriz entera y rectangular (ver definición (1.38)).

σ Símplice de dimensión $m-1$ generado por un conjunto de m vectores linealmente independientes (ver definición 3.1.10).

$\hat{\boldsymbol{\sigma}}$ Baricentro del símplice σ (ver definición 3.1.12).

σ_j j-ésima faceta del símplice σ (ver definición 3.1.10).

$\hat{\boldsymbol{\sigma}}_j$ Baricentro de la j-ésima faceta σ_j del símplice σ (ver definición 3.1.12).

$\hat{\boldsymbol{\mu}}_j$ Vector unitario normal a σ_j que apunta al interior relativo de σ y que además es paralelo a σ.

S_i Región factible de un subproblema relajado de un programa lineal entero que se resolverá a través de R&A.

Convenciones

gen Espacio vectorial generado por una colección de vectores.

ker Espacio nulo de una transformación lineal o de su matriz asociada.

im Imagen de una transformación lineal o de su matriz asociada.

$\|\cdot\|$ Norma 2 de un vector o de un operador lineal acotado.

$\lfloor\cdot\rfloor$ Función piso.

$\lceil\cdot\rceil$ Función techo.

$:=$ Definición dentro de una expresión matemática.

Resumen

El algoritmo de *Ramificación y Acotamiento* (R&A) es uno de los más utilizados para resolver programas lineales enteros. Este método se basa en el famoso paradigma de "divide y vencerás", el cual combina las soluciones de subproblemas más pequeños a fin de obtener una solución del problema original. Estos problemas se estructuran en forma de árbol: el problema original genera una colección de subproblemas, y cada subproblema genera su propia colección de subsubproblemas, etcétera.

En esta tesis se muestra que existe una colección de programas lineales enteros con una sola restricción para los cuales el paradigma de "divide y vencerás" es inadecuado. En particular, las simetrías que exhiben estos problemas impiden que R&A los resuelva eficientemente. Además, se muestra la existencia de una subcolección de programas lineales enteros cuyos árboles asociados son tales que R&A no termina en tiempo finito. Se desarrollan dos nuevos algoritmos que resuelven de manera más eficiente este tipo de instancias problemáticas a partir de la búsqueda de soluciones de ecuaciones lineales diofantinas. De manera simultánea, se obtiene un método pseudo-polinomial para calcular cotas superiores del número de Frobenius en el *Problema Diofantino de Frobenius*. Finalmente, se generalizan estos algoritmos para encontrar soluciones de programas lineales enteros con más de una restricción.

Índice general

Introducción

Esta tesis analiza a profundidad el programa lineal entero

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{Z}^n} \quad & \boldsymbol{p}^T\boldsymbol{x}, && \text{(0.1a)} \\ \text{s.a.} \quad & \boldsymbol{p}^T\boldsymbol{x} \leq u, && \text{(0.1b)} \\ & \boldsymbol{x} \geq \mathbf{0}. \end{aligned}$$

donde $\boldsymbol{p} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ pertenece a una clase de vectores que definiremos en el primer capítulo, y $u \in \mathbb{R}$ es un escalar. Son dos las razones por las que nos dedicamos a estudiar casi exclusivamente este problema: en primer lugar, el hecho de que tenga una sola restricción facilita su análisis y su interpretación geométrica; y, en segundo lugar, las simetrías introducidas por el hecho de que el vector $\boldsymbol{p}$ define tanto la función objetivo (0.1a) como la restricción (0.1b) causan ineficiencias en el método de R&A al momento de resolver el problema. De esta manera, el programa lineal entero (0.1) es una especie de "ejemplo minimal" que motiva el diseño de algoritmos alternativos a Ramificación y Acotamiento.

En el capítulo 1 se presentan los prerrequisitos necesarios para obtener los resultados que se encuentran a lo largo de esta tesis. En particular, se define una clase de vectores a la cual supondremos que el vector objetivo $\boldsymbol{p}$ pertenece, y también se obtienen varias de sus propiedades. Esta clase de vectores contiene cualquier vector representable en aritmética finita por lo que, en la práctica, este supuesto es razonable. Entre los resultados origina-

les del autor que más destacan en esta parte de la tesis son el teorema 1.2.9, el cual separa en dos subclases las instancias de (0.1) y que son tratadas respectivamente en los capítulos 2 y 3; y el teorema 1.2.16, el cual da inicio a una clasificación de programas lineales enteros.

En el capítulo 2 se analiza el caso en el que dos entradas del vector $\boldsymbol{p}$ tienen signos distintos. Bajo esta hipótesis adicional, la solución del problema (0.1) se obtiene al resolver una sola ecuación lineal diofantina. Por un lado, mostramos que el valor objetivo del problema (0.1) se puede determinar de manera inmediata sin tener conocimiento de la solución óptima. Por el otro lado, presentamos un algoritmo que construye la solución óptima y cuya complejidad es polinomialmente acotada en la dimensión del vector $\boldsymbol{p}$. Finalmente, realizamos una serie de experimentos numéricos que permiten comparar los tiempos de terminación de este nuevo algoritmo con los de Ramificación y Acotamiento.

En el capítulo 3 se analiza el caso en el que todas las entradas del vector $\boldsymbol{p}$ tienen el mismo signo. Bajo esta hipótesis solamente podemos asegurar la finitud del número de ecuaciones lineales diofantinas que debemos resolver para encontrar la solución de (0.1). No obstante, mostramos que si el lado derecho de la restricción (0.1b) es suficientemente grande, entonces sí basta con resolver una sola ecuación lineal diofantina para obtener el óptimo. Además, presentamos un algoritmo que construye la solución óptima de (0.1). Es en este capítulo que, de manera simultánea, encontramos cotas superiores para el número de Frobenius mencionado en la motivación de esta tesis. Al igual que en el capítulo 2, se realizan experimentos numéricos que permiten comparar los tiempos de terminación de este nuevo algoritmo con los de Ramificación y Acotamiento.

En el capítulo 4 se introduce el caso de múltiples restricciones, y resulta ser que la división en casos del teorema 1.2.9 deja de ser vigente. Por lo tanto, se desarrolla un nuevo método que permite resolver este tipo de pro-

blemas más generales bajo la perspectiva de sistemas de ecuaciones lineales diofantinas. Puesto que un análisis detallado sería demasiado extenso para añadirlo a esta tesis de licenciatura, la discusión en este capítulo es más superficial, pero no por ello menos formal.

Finalmente, en el capítulo 5 se presenta una recopilación de los resultados originales y más destacables obtenidos a lo largo de esta tesis. Así también, se presenta una recopilación de problemas mencionados en esta tesis pero que no fueron tratados, bien porque eran tangenciales al objetivo central, bien porque sus respectivos análisis serían demasiado extensos para ser añadidos. Ciertamente, cada uno de estos problemas sirve como directriz inicial para la realización de futuras investigaciones.

Capítulo 1

Aspectos Teóricos

En este capítulo se presentan los prerrequisitos necesarios para obtener los resultados que se encuentran a lo largo de esta tesis. En primer lugar, en la sección 1.1 se recopilan resultados básicos de teoría de números y de programación lineal que forman parte de la literatura tradicional y constituyen nuestras bases a partir de las cuales derivaremos nuestros propios resultados. En segundo lugar, en la sección 1.2 se presentan enunciados y definiciones obtenidos de [BH09], los cuales utilizaremos para continuar con la construcción de nuestros resultados originales.

Como mencionamos en la introducción de esta tesis, nos concentraremos casi exclusivamente en problemas de programación lineal entera del tipo

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{Z}^n} \quad & \boldsymbol{p}^T\boldsymbol{x}, && \text{(1.1a)}\\ \text{s.a.} \quad & \boldsymbol{p}^T\boldsymbol{x} \leq u, && \text{(1.1b)}\\ & \boldsymbol{x} \geq \boldsymbol{0}, \end{aligned}$$

donde $\boldsymbol{p} \in \mathbb{R} \setminus \{\boldsymbol{0}\}$ pertenece a una clase que definiremos en la sección 1.2 y $u \in \mathbb{R}$ es un escalar. Comúnmente haremos referencia a la restricción (1.1b) como **restricción presupuestaria** debido a que cada unidad de x_i "cuesta" p_i unidades y queremos maximizar nuestro "gasto" (1.1a) sin exceder nuestro **presupuesto** u. Podemos suponer, sin pérdida de generalidad, que todas

las entradas de $\boldsymbol{p}$ son distintas de cero. En efecto, si alguna entrada p_i es nula, cualquier elección entera y no negativa de x_i es válida para este programa lineal entero. Esta suposición será implícita en lo que resta del capítulo, aunque se hará explícita al final cuando investiguemos la manera de clasificar instancias de este programa lineal entero.

1.1. Prerrequisitos

Se consideró pertinente no incluir demostraciones en esta sección, pues los enunciados son mostrados en los cursos de álgebra superior, programación lineal, o investigación de operaciones. Las referencias principales para las subsecciones de teoría de números y de programación lineal son [Lav14] y [Oli17], respectivamente.

1.1.1. Teoría de Números

Máximo común divisor y mínimo común múltiplo

Definición 1.1.1. Dados dos enteros a y b, decimos que a **divide a** b (y escribimos $a \mid b$) si existe un entero k tal que $a = k \cdot b$. También denotamos por $D(a)$ al **conjunto de divisores de** a, es decir, definimos

$$D(a) \coloneqq \{b \in \mathbb{Z} : b \mid a\}.$$

Observemos que si a es distinto de cero, entonces $D(a)$ es finito. En efecto, si $b \mid a$ y $a \neq 0$, es posible mostrar que $|b| \leq |a|$, lo cual implica que $|D(a)| \leq 2|a|$, donde $|D(a)|$ denota la cardinalidad del conjunto $D(a)$ y $|a|$ el valor absoluto de a. En caso de que a sea nulo, se sigue que $D(a) = \mathbb{Z}$.

Definición 1.1.2. Sean $a_1, \ldots, a_n$ enteros no todos iguales a cero, entonces definimos su **máximo común divisor** d como el máximo elemento del

conjunto $\bigcap_{i=1}^{n} D(a_i)$, y escribimos $d = \text{mcd}\{a_1, \ldots, a_n\}$. Si $d = 1$, entonces decimos que $a_1, \ldots, a_n$ son **coprimos**.

Puesto que alguna entrada a_i es distinta de cero en la definición anterior, encontramos que el conjunto $\bigcap_{i=1}^{n} D(a_i)$ es finito y también es no vacío, por lo que este conjunto tiene un máximo elemento. Es decir, el máximo común divisor está bien definido.

Además, el máximo común divisor siempre es positivo, pues se cumple que $1 \in D(a)$ para todo entero a, lo que implica por maximalidad que $1 \leq \text{mcd}\{a_1, \ldots, a_n\}$ para cualquier colección de enteros no todos nulos.

La definición más común del máximo común es dada de manera inductiva. Decimos que d es el máximo común divisor de dos enteros a_1, a_2, no ambos iguales a cero, si se satisface

1. $d \mid a_1$ y $d \mid a_2$, y también,

2. si $d' \mid a_1$ y $d' \mid a_2$, entonces $d' \mid d$.

Luego, para un conjunto de enteros $a_1, a_2 \ldots a_n$, no todos iguales a cero, definimos el máximo común divisor entre ellos de manera inductiva:

$$\text{mcd}\{a_1, a_2, \ldots, a_{n-1}, a_n\} \coloneqq \text{mcd}\{a_1, \text{mcd}\{a_2, \ldots, \text{mcd}\{a_{n-1}, a_n\} \cdots\}\}.$$

Sin embargo, debemos ser cuidadosos con esta manera de definir las cosas, pues puede ser el caso, por ejemplo, que a_{n-1} y a_n sean ambos cero y entonces $\text{mcd}\{a_{n-1}, a_n\}$ no está bien definido.

Para que esta manera de definir el máximo común divisor sea equivalente a la definición 1.1.2, deberemos presuponer o bien que $a_{n-1} \neq 0$ o bien que $a_n \neq 0$. A partir de este punto usaremos ambas definiciones de manera indistinta. Independientemente de la definición que utilicemos, es posible calcular el máximo común divisor a través del algoritmo de Euclides, el cual es introducido en la sección 4.5 de [Knu98].

Definición 1.1.3. Decimos que $c \in \mathbb{Z}$ es una **combinación lineal entera** de un conjunto de enteros $a_1, \ldots, a_n$ si existen enteros $x_1, \ldots, x_n$ tales que $c = a_1x_1 + \cdots + a_nx_n$. Si c es positivo, también decimos que esto último es una **combinación lineal entera positiva**.

Teorema 1.1.4. *Sea d un entero y sea $a_1, \ldots, a_n$ una colección de enteros no todos iguales a cero. Entonces $d = \text{mcd}\{a_1, \ldots, a_n\}$ si y solo si d es la mínima combinación lineal entera positiva de $a_1, \ldots, a_n$.*

Ejemplo 1.1.5. El máximo común divisor de los enteros 2, 3 y 5 es 1. Observemos que $-3 \cdot 2 - 1 \cdot 3 + 2 \cdot 5 = 1$.

Lema 1.1.6. *Si $d = \text{mcd}\{a_1, \ldots, a_n\}$, entonces* $\text{mcd}\left\{\frac{a_1}{d}, \ldots, \frac{a_n}{d}\right\} = 1$.

Además del máximo común divisor, haremos uso del mínimo común múltiplo, aunque será en menor medida.

Definición 1.1.7. Definimos el **conjunto de múltiplos** de un entero a como

$$M(a) \coloneqq \{x \in \mathbb{Z} : a \mid x\}.$$

También definimos el **mínimo común múltiplo** de un conjunto de enteros $a_1, \ldots, a_n$, no todos iguales a cero, como el mínimo elemento del conjunto

$$\mathbb{Z}_{\geq 0} \cap \bigcap_{i=1}^{n} M(a_i).$$

Escribimos $\text{mcm}\{a_1, \ldots, a_n\}$ para denotar a este mínimo común múltiplo.

Para mostrar que el mínimo común múltiplo está bien definido, basta observar que el producto $|a_1 \cdots a_n|$ es no negativo y también es un elemento de $M(a_i)$ para toda $i \in \{1, \ldots, n\}$.

Ecuaciones lineales diofantinas

Sea $c \in \mathbb{Z}$ y sea $a_1, \ldots, a_n$ una colección de enteros. Una ecuación lineal diofantina es una ecuación en la que deseamos determinar enteros $x_1, \ldots, x_n$ que satisfagan

$$a_1x_1 + \cdots + a_nx_n = c.$$

En esta sección nos enfocamos en el caso $n = 2$, y en la sección 1.2 generalizamos estos resultados para cualquier entero $n \geq 2$. Los siguientes enunciados abordan el problema de determinar la existencia y construcción de soluciones para este tipo de ecuaciones.

Teorema 1.1.8 (Existencia)**.** *Sean a y b enteros, no ambos iguales a cero. Entonces la ecuación lineal diofantina $ax + by = c$ tiene solución entera si y solo si* $\text{mcd}\{a, b\} \mid c$.

Para construir el conjunto de soluciones de una ecuación lineal diofantina, primero encontramos una solución particular.

Definición 1.1.9. Sea $d \coloneqq \text{mcd}\{a, b\}$ y sean x', y' números enteros tales que $ax' + by' = d$ (su existencia está garantizada por el teorema 1.1.4). Decimos entonces que x', y' son **coeficientes de Bézout** asociados a a y b, respectivamente.

Los coeficientes de Bézout se pueden calcular a través del algoritmo extendido de Euclides, el cual es presentado en la sección 4.5 de [Knu98]. Observemos que estos coeficientes no son únicos. En efecto, si x', y' son coeficientes de Bézout de a y b, entonces $x' + b$, $y' - a$ también lo son:

$$a(x' + b) + b(y' - a) = ax' + by' + ab - ab = ax' + by' = d.$$

Para fines de esta tesis basta la existencia de estos coeficientes, por lo que diremos de manera indistinta "los coeficientes de Bézout" y "una elección de coeficientes de Bézout".

Definamos $d \coloneqq \text{mcd}\{a, b\}$ y supongamos que la ecuación $ax + by = c$ tiene solución. Por el teorema 1.1.8, se sigue que $d \mid c$, y entonces existe $c' \in \mathbb{Z}$ tal que $c = c' \cdot d$. Sean x', y' los coeficientes de Bézout asociados a a, b respectivamente. Así,

$$a(c' \cdot x') + b(c' \cdot y') = c'(ax' + by') = c'd = c,$$

por lo que $(c' \cdot x', c' \cdot y')$ es una solución particular de la ecuación $ax+by = c$.

Teorema 1.1.10 (Construcción)**.** *Sea $ax + by = c$ una ecuación lineal diofantina con solución particular (x_0, y_0), entonces todas sus soluciones están dadas por*

$$\begin{cases} x = x_0 + \frac{b}{d}t, \\ y = y_0 - \frac{a}{d}t, \end{cases} \tag{1.2}$$

donde $d \coloneqq \text{mcd}\{a, b\}$ y $t \in \mathbb{Z}$ es una variable libre.

Ejemplo 1.1.11. Consideremos la ecuación lineal $2x + 3y = 5$. Los coeficientes de Bézout asociados a 2 y 3 son, respectivamente, -1 y 1. Luego, una solución particular para la ecuación es $(x_0, y_0) = (-5, 5)$. Por el teorema anterior encontramos que todas las soluciones están dadas por

$$\begin{cases} x = -5 + 3t, \\ y = 5 - 2t, \end{cases}$$

donde $t \in \mathbb{Z}$ es una variable libre. En efecto, sustituyendo obtenemos

$$2(-5 + 3t) + 3(5 - 2t) = -10 + 15 + 6t - 6t = 5.$$

1.1.2. Programación lineal

Definición 1.1.12. Sea $A \in \mathbb{R}^{m \times n}$ una matriz con renglones linealmente

independientes y $\boldsymbol{b} \in \mathbb{R}^m$ un vector. Entonces al conjunto definido por

$$P \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n \colon A\boldsymbol{x} \geq \boldsymbol{b}\} \tag{1.3}$$

lo llamamos **poliedro**. Si, además, P es acotado, entonces decimos que P es un **politopo**.

La programación lineal se encarga de resolver problemas de optimización de la forma

$$\max_{\boldsymbol{x} \in \mathbb{R}^n} \{\boldsymbol{c}^T\boldsymbol{x} \colon \boldsymbol{x} \in P\}, \tag{1.4}$$

donde $P \subseteq \mathbb{R}^n$ es un poliedro al cual llamamos **región factible**, $\boldsymbol{c} \in \mathbb{R}^n$ un vector, comúnmente conocido como el **vector objetivo**, y $\boldsymbol{x} \in \mathbb{R}^n$ un vector de variables a decidir.

Definición 1.1.13. Sea $\boldsymbol{a} \in \mathbb{R}^n$ un vector no nulo y sea $b \in \mathbb{R}$ un escalar. Llamamos **hiperplano afín** al conjunto de vectores $\boldsymbol{x} \in \mathbb{R}^n$ que satisfacen $\boldsymbol{a}^T\boldsymbol{x} = b$. Así también, llamamos **semi-espacios afines** a los conjuntos de vectores $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ que satisfacen $\boldsymbol{a}^T\boldsymbol{x} \geq b$ y $\boldsymbol{a}^T\boldsymbol{y} \leq b$.

Observemos que todo poliedro P es la intersección de m semi-espacios afines. Esto se debe a que, para una matriz $A \in \mathbb{R}^{m \times n}$ con renglones independientes y un vector $\boldsymbol{b} \in \mathbb{R}^m$, $A\boldsymbol{x} \geq \boldsymbol{b}$ si y solo si $\boldsymbol{a}_i^T\boldsymbol{x} \geq b_i$ para toda $1 \leq i \leq m$ y donde $\boldsymbol{a}_i^T$ denota el i-ésimo renglón de la matriz A. En la figura 1.1 se muestra la relación entre semi-espacios afines y poliedros.

Definición 1.1.14 ([Oli17])**.** Sea $P \subseteq \mathbb{R}^n$ un poliedro. Decimos que el vector $\boldsymbol{x} \in P$ es un **vértice** de P si existe $\boldsymbol{c} \in \mathbb{R}^n$ de manera que $\boldsymbol{c}^T\boldsymbol{x} < \boldsymbol{c}^T\boldsymbol{y}$ para todo $\boldsymbol{y} \in P \setminus \{\boldsymbol{x}\}$.

En términos gráficos, decimos que $\boldsymbol{x}$ es un vértice si se satisfacen dos condiciones: en primer lugar, existe un hiperplano afín que pasa por $\boldsymbol{x}$ y uno de sus semi-espacios contiene completamente al poliedro P; en segundo lugar, ningún otro punto de P se encuentra sobre este hiperplano.

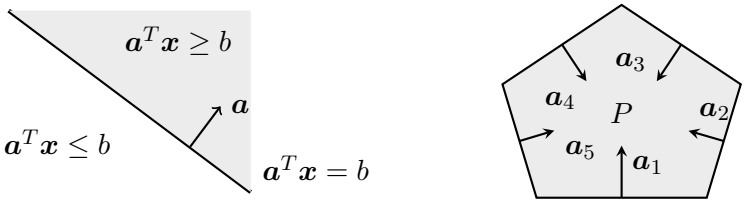

Figura 1.1: *Izquierda:* Un hiperplano afín $\{\boldsymbol{x}\colon \boldsymbol{a}^T\boldsymbol{x} = b\}$ junto con sus dos semi-espacios $\{\boldsymbol{x}\colon \boldsymbol{a}^T\boldsymbol{x} \geq \boldsymbol{b}\}$ y $\{\boldsymbol{x}\colon \boldsymbol{a}^T\boldsymbol{x} \leq \boldsymbol{b}\}$. *Derecha:* Un politopo P.

Definición 1.1.15. Sea P un poliedro y sea $\boldsymbol{c} \in \mathbb{R}^n$ un vector. Todo problema de optimización de la forma (1.4) entra en una de las siguientes tres categorías:

1. El valor óptimo no existe: ningún vector $\boldsymbol{x} \in \mathbb{R}^n$ satisface el sistema de desigualdades $A\boldsymbol{x} \geq \boldsymbol{b}$. Es decir, la región factible es vacía.

2. El valor óptimo existe y es infinito: el poliedro P no es acotado y existe una sucesión de vectores $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ en el poliedro P que satisface $\boldsymbol{c}^T\boldsymbol{x}_{k+1} > \boldsymbol{c}^T\boldsymbol{x}_k$ para todo $k \in \mathbb{N}$.

3. El valor óptimo existe y es finito: este caso es la negación de los dos puntos anteriores. Observemos que aquí cabe la posibilidad de que el poliedro P no sea acotado, pues puede que no exista la sucesión creciente de valores objetivo definida en el caso anterior.

En el primer caso decimos que **el problema es infactible**, mientras que en los últimos dos decimos que **el problema es factible**. También diremos comúnmente del segundo caso que **el problema es no acotado**.

En el capítulo 3 de [Oli17] se muestra que todo poliedro

$$P \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \geq \boldsymbol{b}\}$$

puede ser transformado a la forma estándar

$$\{\left(\boldsymbol{x}^+, \boldsymbol{x}^-, \boldsymbol{s}\right) \in \mathbb{R}^{n+n+m} : A(\boldsymbol{x}^+ - \boldsymbol{x}^-) - \boldsymbol{s} = \boldsymbol{b}, \left(\boldsymbol{x}^+, \boldsymbol{x}^-, \boldsymbol{s}\right) \geq \boldsymbol{0}\},$$

de manera que todo problema de optimización de la forma (1.4) puede ser escrito sin pérdida de generalidad como

$$\begin{aligned} \max_{\boldsymbol{x} \in \mathbb{R}^n} \quad & \boldsymbol{c}^T \boldsymbol{x}, && \text{(1.5a)} \\ \text{s.a.} \quad & A\boldsymbol{x} = \boldsymbol{b}, && \text{(1.5b)} \\ & \boldsymbol{x} \geq \boldsymbol{0}. \end{aligned}$$

De ahora en adelante nuestro análisis se concentrará exclusivamente en problemas lineales en forma estándar. A continuación enunciamos el teorema fundamental que nos permite resolver problemas lineales.

Teorema 1.1.16. *Sea P un poliedro que tiene al menos un vértice, consideremos el problema* (1.5)*, y supongamos que el valor óptimo z^* existe y es finito. Entonces el conjunto de soluciones óptimas contiene al menos un vértice de P.*

Este teorema constituye el primer paso para la construcción de varios algoritmos que encuentran soluciones del problema (1.5). El más famoso de todos es el algoritmo *simplex* (capítulo 13 de [NW06]), el cual "salta" de vértice en vértice hasta llegar a uno con valor óptimo. Otros, más modernos y conocidos como *métodos de puntos interiores* (capítulo 14 de [NW06]), comienzan en el interior del poliedro P y son "atraídos" como imanes a uno de los vértices con valor óptimo. No es el objetivo de esta tesis exponer la teoría detrás de estos algoritmos.

Ahora describimos brevemente los programas lineales enteros y explicamos el método de R&A. Supondremos en lo que resta de esta tesis que contamos con un algoritmo para resolver problemas lineales.

Definición 1.1.17. Sean $A \in \mathbb{R}^{m\times n}$ una matriz con renglones linealmente independientes, $\boldsymbol{c} \in \mathbb{R}^n$ un vector, y $\boldsymbol{b} \in \mathbb{R}^m$ otro vector. Al problema de optimización lineal (1.5) lo llamamos **problema relajado** del **programa lineal entero**

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{Z}^n} \quad & \boldsymbol{c}^T\boldsymbol{x}, && \text{(1.6a)} \\ \text{s.a.} \quad & A\boldsymbol{x} = \boldsymbol{b}, && \text{(1.6b)} \\ & \boldsymbol{x} \geq \boldsymbol{0}. \end{aligned}$$

Resalta el hecho de que la formulación de un programa lineal entero es idéntico a su formulación relajada, y solamente intercambiamos nuestro espacio de búsqueda $\mathbb{R}^n$ por $\mathbb{Z}^n$. Es decir, lo único que cambia es la región de factibilidad. De hecho, si definimos el poliedro

$$P \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\},$$

entonces tenemos que $P \cap \mathbb{Z}^n$ corresponde a la región factible de (1.6), mientras que P corresponde a la región factible de su problema relajado.

A partir de lo anterior, deducimos inmediatamente que el valor óptimo z^* de un problema relajado es una cota superior del valor óptimo z^*_{PE} de su programa lineal entero, pues ambos son problemas de maximización y es cierto que $P \cap \mathbb{Z}^n \subseteq P$. De aquí se sigue que si $z^*_{\text{PE}} = z^*$, entonces la solución óptima $\boldsymbol{x}^*$ del problema relajado también es la solución óptima del programa lineal entero si $\boldsymbol{x}^* \in \mathbb{Z}^n$.

Ramificación y Acotamiento encapsula una familia de métodos que comparten la misma estructura para encontrar soluciones de programas lineales enteros. Específicamente, R&A construye recursivamente un árbol de subproblemas lineales relajados a resolver. Si la solución de un subproblema no es entera, entonces su región factible es **ramificada** para generar otra colección de subproblemas y cuyas soluciones podrían ser enteras. Es posible

aplicar **criterios de poda** para eliminar porciones del árbol que no contienen soluciones óptimas del problema original. Ilustramos el procedimiento a través de un ejemplo concreto, en el cual enfatizamos estos componentes principales que caracterizan a la familia de métodos.

Ejemplo 1.1.18 ([Oli17])**.** Consideremos el programa lineal entero

$$\begin{aligned} \max_{(x_1,x_2)\in\mathbb{Z}^2} \quad & 4x_1 - x_2, \\ \text{s.a.} \quad & 7x_1 - 2x_2 \leq 14, \\ & 2x_1 - 2x_2 \leq 3, \\ & x_2 \leq 3, \\ & x_1, x_2 \geq 0. \end{aligned}$$

La región factible de este problema se muestra en la figura 1.2. La solución al problema relajado, cuya región factible denotamos por S, está dada por $\boldsymbol{x}^{(0)} \coloneqq (20/7, 3)$. Como $x_1^{(0)} = 20/7$ no es entero, **ramificamos binariamente** para generar los subproblemas con regiones factibles

$$\begin{aligned} S_0 &\coloneqq S \cap \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \leq \lfloor 20/7 \rfloor = 2\}, \\ S_1 &\coloneqq S \cap \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq \lceil 20/7 \rceil = 3\}. \end{aligned}$$

En este nivel del árbol así como en los siguientes, preferimos encontrar las soluciones de ambos subproblemas antes de ramificar cualesquiera de ellos. Es decir, realizamos una **búsqueda en anchura** (*breadth-first search*) para resolver el problema original.

Observamos en la figura 1.2 que S_1 es vacío, así que no será posible ramificarlo. Ahora bien, la solución del problema con región factible S_0 es $\boldsymbol{x}^{(1)} \coloneqq (2, 1/2)$. Puesto que $x_2^{(1)} = 1/2$ no es entero, ramificamos binaria-

mente para generar los nuevos subproblemas con regiones factibles

$$S_{00} \coloneqq S_0 \cap \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq \lfloor 1/2 \rfloor = 0\},$$
$$S_{01} \coloneqq S_0 \cap \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \geq \lceil 1/2 \rceil = 1\}.$$

La solución $\boldsymbol{x}^{(2)}$ del subproblema con región factible S_{01} es $(2, 1)$, y entonces es entera con valor objetivo $z_2^* \coloneqq 7$. Decimos que esta solución es la **solución incumbente**: no vale la pena ramificar otros subproblemas cuyo valor objetivo sea menor o igual que z_2^*. En estos casos diremos que podamos esos subproblemas por **cota**. No es difícil ver que cualquier descendiente de S_{01} será podado por cota. Además, observemos que la solución de S_{00} es $\boldsymbol{x}^{(3)} \coloneqq (3/2, 0)$, por lo que su valor objetivo es $6 < z_2*$, así que también podamos S_{00} por cota.

Como hemos agotado todos los subproblemas que podríamos explorar, concluimos que la solución óptima de este programa lineal entero es la única solución incumbente obtenida $\boldsymbol{x}^{(2)} = (2, 1)$ con valor objetivo 7.

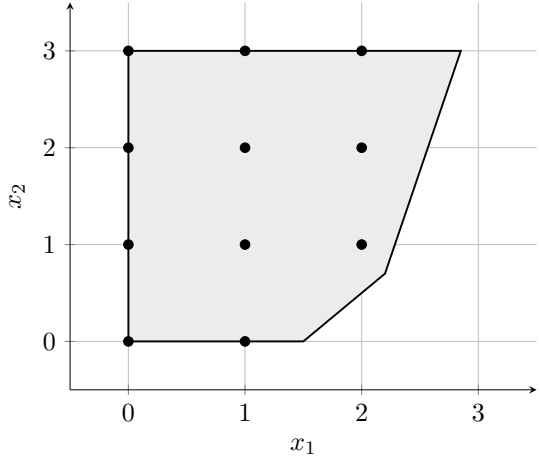

Figura 1.2: Los puntos negros forman la región factible del programa lineal entero del ejemplo 1.1.18, mientras que la región sombreada es la región factible de su problema relajado.

Como pudimos apreciar en este ejemplo, los componentes principales que caracterizan a Ramificación y Acotamiento fueron la estrategia de búsqueda, que indica el orden en el que resolvemos los subproblemas; la estrategia de ramificación, que indica cómo generar los descendientes de un nodo; y las políticas de poda, que son reglas para eliminar subárboles de nuestra búsqueda. Estas tres características están interrelacionadas: si hubiéramos resuelto el subproblema S_{00} antes que S_{01}, no habríamos tenido una solución incumbente, así que no lo habríamos podado y entonces lo habríamos ramificado en otros dos subproblemas. En [MJSS16] se recopilan avances en estos tres componentes que constituyen líneas de investigación abiertas.

El algoritmo 1 en la página 17 formaliza el procedimiento seguido en este ejemplo, a excepción de la búsqueda en anchura, pero que es posible implementar con una fila (*queue*). En cuanto a los resultados teóricos que obtendremos a lo largo de esta tesis, supondremos que nos referimos a este algoritmo en específico cuando hagamos mención genérica de R&A, y nos reservamos el calificativo de "implementación pura", pues es común que se enseñe esta versión en cursos introductorios de investigación de operaciones. En cuanto a los resultados numéricos, utilizaremos la implementación de código abierto *COIN-OR Branch and Cut* (CBC) de la fundación COIN-OR [Lou03] para encontrar soluciones a programas lineales enteros. Otras implementaciones comúnes de código abierto son HiGHS y SCIP, mientras que algunas implementaciones comerciales son Gurobi Optimizer, IBM ILOG CPLEX Optimizer, y Fico Xpress Solver.

Algoritmo 1: Ramificación y Acotamiento. Adaptado de [Oli17].

Datos: Problema lineal relajado S.

Resultado: Solución óptima entera $\boldsymbol{x}^*$ y valor óptimo z^*_{PE}.

inicio 1

- $\mathcal{L} \leftarrow \{S\}$ 2
- $\boldsymbol{x}^* \leftarrow -\boldsymbol{\infty}$ 3
- $z^*_{\text{PE}} \leftarrow -\infty$ 4
- **mientras** $\mathcal{L} \neq \emptyset$ **hacer** 5
  - elegir subproblema S_i de $\mathcal{L}$ `// estrategia de búsqueda` 6
  - obtener valor óptimo z^*_i y solución óptima $\boldsymbol{x}^{(i)}$ de S_i 7
  - $\mathcal{L} \leftarrow \mathcal{L} \setminus \{S_i\}$ 8
  - **si** $S_i = \emptyset$ *o* $z^*_i \leq z^*_{PE}$ **entonces** 9
    - `// podar por cota` 10
    - ir a la línea 5 11
  - **si** $\boldsymbol{x}^{(i)} \in \mathbb{Z}^n$ **entonces** 12
    - `// solución incumbente` 13
    - $\boldsymbol{x}^* \leftarrow \boldsymbol{x}^{(i)}$ 14
    - $z^*_{\text{PE}} \leftarrow z^*_i$ 15
    - ir a la línea 5 16
  - `// ramificación binaria` 17
  - elegir $x^{(i)}_j \notin \mathbb{Z}$ y definir 18

$$S_{i0} \leftarrow S_i \cap \{\boldsymbol{x} \in \mathbb{R}^n : x_j \leq \lfloor x^{(i)}_j \rfloor\},$$
$$S_{i1} \leftarrow S_i \cap \{\boldsymbol{x} \in \mathbb{R}^n : x_j \geq \lceil x^{(i)}_j \rceil\}$$

  - $\mathcal{L} \leftarrow \mathcal{L} \cup \{S_{i0}, S_{i1}\}$.
- **devolver** $(\boldsymbol{x}^*, z^*_{PE})$ 19

1.2. Fundamentos

En primer lugar, se dan a conocer las definiciones y encunciados originalmente dados en [BH09]. Es importante aclarar que se tradujeron los términos *projectively rational vectors* y *c-layers* como "vectores esencialmente enteros" y "capas enteras" en las definiciones 1.2.1 y 1.2.3, respectivamente, a falta de encontrar fuentes en español que hicieran uso de ellos.

En segundo lugar, se muestra una equivalencia entre resolver el programa lineal entero (1.1) y resolver ecuaciones lineales diofantinas. Este resultado se basa en los teoremas 1.1.8 y 1.1.10 para construir inductivamente el conjunto de soluciones de una ecuación lineal diofantina en n incógnitas. Estas soluciones serán definidas a partir de una relación de recurrencia y dependerán de una colección de $n-1$ variables libres.

En tercer lugar, se exhibe una transformación lineal entre el conjunto de soluciones de una ecuación lineal diofantina en n incógnitas y el conjunto de $n-1$ variables libres que determinan estas soluciones. Luego, se investigan las propiedades de esta transformación lineal que serán de gran utilidad teórica para los siguientes capítulos, en especial para el capítulo 4.

Finalmente, se muestra que el vector objetivo $\boldsymbol{p}$ del problema original (1.1) induce una descomposición de $\mathbb{Z}^n$ y se analiza cómo se relacionan las descomposiciones asociadas a distintos vectores $\boldsymbol{p}$. Esto permitirá mostrar equivalencias entre distintas instancias del problema (1.1).

1.2.1. Capas enteras

Definición 1.2.1. Decimos que un vector $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ es **esencialmente entero** si existen un vector $\boldsymbol{w} \in \mathbb{Z}^n$ y un escalar $m \in \mathbb{R} \setminus \{0\}$ tales que $\boldsymbol{v} = m\boldsymbol{w}$. Además, decimos que $\boldsymbol{w}$ es el **múltiplo coprimo** de $\boldsymbol{v}$ si sus entradas son coprimas y si su primera entrada no nula es positiva.

Ejemplo 1.2.2. El vector $\left(-\sqrt{2}, 1/\sqrt{2}\right) = 2\sqrt{2}(-2,1)$ es esencialmente

entero y $(2, -1)$ es su múltiplo coprimo. En contraste, el vector $\left(\sqrt{2}, \sqrt{3}\right)$ no es esencialmente entero. Para ver esto último, supongamos que existe un escalar $m \in \mathbb{R} \setminus \{0\}$ y enteros a, b tales que $m(a, b) = \left(\sqrt{2}, \sqrt{3}\right)$. Claramente, $a, b \neq 0$ y se debe cumplir $a/b = \sqrt{2/3}$, pero el lado izquierdo es un número racional mientras que el derecho es un número irracional, así que obtenemos una contradicción, por lo que el vector $\left(\sqrt{2}, \sqrt{3}\right)$ no es esencialmente entero.

Observación. Todo vector $\boldsymbol{v}$ esencialmente entero tiene a lo más dos vectores coprimos asociados. Sean $m \in \mathbb{R}$ y $\boldsymbol{w} \in \mathbb{Z}^n$ tales que $\boldsymbol{v} = m\boldsymbol{w}$. Entonces

$$\pm \frac{1}{\operatorname{mcd}\{w_1, \ldots, w_n\}} \boldsymbol{w}$$

son dos vectores cuyas entradas son coprimas, de acuerdo al lema 1.1.6. Como la primera entrada no nula w_i también debe ser positiva, se sigue que solo uno de estos dos vectores es el múltiplo coprimo de $\boldsymbol{v}$. Así, el múltiplo coprimo de un vector esencialmente entero es único.

Observación. Todo vector racional $\boldsymbol{v} \in \mathbb{Q}^n \setminus \{\boldsymbol{0}\}$ es esencialmente entero. En efecto, para cada $i \in \{1, \ldots, n\}$ existen $p_i, q_i \in \mathbb{Z}$ con $q_i \neq 0$ tales que $v_i = p_i/q_i$. Luego, si definimos $m \coloneqq \operatorname{mcm}\{q_1, \ldots, q_n\} \neq 0$ y también $\boldsymbol{w} \coloneqq m\boldsymbol{v} \in \mathbb{Z}^n$, encontramos que $\boldsymbol{v} = \frac{1}{m}\boldsymbol{w}$.

Definición 1.2.3. Sea $\boldsymbol{v} \in \mathbb{R}^n$ un vector esencialmente entero y sea $t \in \mathbb{R}$ un escalar. Decimos que su hiperplano afín asociado

$$H_{\boldsymbol{v},t} \coloneqq \ker\{\boldsymbol{x} \mapsto \boldsymbol{v}^T\boldsymbol{x}\} + t\boldsymbol{v} = \{\boldsymbol{v}^\perp + t\boldsymbol{v} : \boldsymbol{v}^T\boldsymbol{v}^\perp = 0\} \tag{1.7}$$

es una **capa entera** si contiene al menos un punto entero.

La ventaja principal de utilizar la clase de vectores esencialmente enteros en vez de los vectores sobre $\mathbb{R}^n$ o sobre $\mathbb{Q}^n$ se debe a que esta clase es la más grande, de acuerdo al teorema 9 de [BH09], en la que el número de capas enteras es finita entre cualesquiera dos puntos. Los algoritmos que

presentaremos en el capítulo 3 enumeran estas capas enteras, así que su finitud asegura la terminación en tiempo finito de estos algoritmos.

Lema 1.2.4. *Sean $\boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^n$ con $\boldsymbol{v}$ distinto de cero. Entonces $\boldsymbol{x} \in H_{\boldsymbol{v},t_{\boldsymbol{x}}}$, donde $t_{\boldsymbol{x}} \coloneqq \frac{\boldsymbol{v}^T\boldsymbol{x}}{\|\boldsymbol{v}\|^2}$.*

Las capas enteras son invariantes ante reescalamientos en el vector $\boldsymbol{v}$: si $r \neq 0$, entonces $H_{\boldsymbol{v},t} = H_{r\boldsymbol{v},t/r}$. Esta igualdad se prueba por contenciones. Sea $\boldsymbol{x} \in H_{\boldsymbol{v},t}$, entonces existe $\boldsymbol{v}^\perp$ ortogonal a $\boldsymbol{v}$ tal que

$$\boldsymbol{x} = \boldsymbol{v}^\perp + t\boldsymbol{v} = \boldsymbol{v}^\perp + \frac{t}{r}(r\boldsymbol{v}).$$

Pero si $\boldsymbol{v}^\perp$ es ortogonal a $\boldsymbol{v}$, entonces también es ortogonal a $r\boldsymbol{v}$. Así, encontramos que $\boldsymbol{x} \in H_{r\boldsymbol{v},t/r}$. La otra contención se muestra de manera similar.

En particular, si $\boldsymbol{w}$ es el múltiplo coprimo de $\boldsymbol{v}$, se cumple que

$$\{H_{\boldsymbol{v},t} : t \in \mathbb{R}\} = \{H_{\boldsymbol{v}/m,tm} : t \in \mathbb{R}\} = \{H_{\boldsymbol{w},t} : t \in \mathbb{R}\}, \tag{1.8}$$

donde $m \neq 0$ es el escalar que satisface $\boldsymbol{v} = m\boldsymbol{w}$. Así pues, para analizar las capas enteras, basta con fijarnos en los múltiplos coprimos que las definen en vez de sus vectores esencialmente enteros asociados.

Teorema 1.2.5. *Sea $\boldsymbol{v} \in \mathbb{R}^n$ un vector esencialmente entero y sea $\boldsymbol{w}$ su múltiplo coprimo. Entonces una cobertura de $\mathbb{Z}^n$ está dada por la familia de capas enteras $\{H_{\boldsymbol{w},k/\|\boldsymbol{w}\|^2} : k \in \mathbb{Z}\}$.*

Lema 1.2.6. *Sea $\boldsymbol{v} \in \mathbb{R}^n$ un vector esencialmente entero y sea $\boldsymbol{w}$ su múltiplo coprimo. Entonces $\boldsymbol{w}^T\boldsymbol{x} = k$ para todo $\boldsymbol{x} \in H_{\boldsymbol{w},k/\|\boldsymbol{w}\|^2}$.*

Demostración. Sea $\boldsymbol{x} \in H_{\boldsymbol{w},k/\|\boldsymbol{w}\|^2}$, por lo que existe un vector $\boldsymbol{w}^\perp$ ortogonal a $\boldsymbol{w}$ tal que

$$\boldsymbol{x} = \boldsymbol{w}^\perp + \frac{k}{\|\boldsymbol{w}\|^2}\boldsymbol{w}.$$

Luego,

$$\boldsymbol{w}^T\boldsymbol{x} = \boldsymbol{w}^T\boldsymbol{w}^\perp + \frac{k}{\|\boldsymbol{w}\|^2}\boldsymbol{w}^T\boldsymbol{w} = 0 + \frac{k}{\|\boldsymbol{w}\|^2}\|\boldsymbol{w}\|^2 = k.$$

que es lo que deseábamos obtener. □

Supongamos que el vector objetivo $\boldsymbol{p}$ del problema (1.1) es esencialmente entero y sea $\boldsymbol{q}$ su múltiplo coprimo. Puesto que el espacio de búsqueda de este problema es $\mathbb{Z}^n$, se sigue de (1.8) y del teorema 1.2.5 que podemos restringir la búsqueda a la familia de capas enteras $\{H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} : k \in \mathbb{Z}\}$.

Del lema 1.2.6 encontramos que los puntos de la **k-ésima capa entera** o bien respetan todos la restricción presupuestaria (1.1b), o bien ninguno lo hace. Sea $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ y observemos que

$$\boldsymbol{p}^T\boldsymbol{x} = m\boldsymbol{q}^T\boldsymbol{x} = mk \leq u \iff \begin{cases} k \geq u/m, & m < 0, \\ k \leq u/m, & m > 0, \end{cases} \tag{1.9}$$

donde $m \neq 0$ es el escalar que satisface $\boldsymbol{p} = m\boldsymbol{q}$. Obtenemos de manera inmediata el siguiente lema.

Lema 1.2.7. *Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y sea $\boldsymbol{q}$ su múltiplo coprimo, de manera que $\boldsymbol{p} = m\boldsymbol{q}$ para algún escalar $m \neq 0$. Entonces la primera capa entera $H_{\boldsymbol{q},\eta/\|\boldsymbol{q}\|^2}$ en satisfacer la restricción* (1.1b) *está parametrizada por*

$$\eta \coloneqq \begin{cases} \lceil u/m \rceil, & m < 0, \\ \lfloor u/m \rfloor, & m > 0. \end{cases} \tag{1.10}$$

Puesto que la gran mayoría de nuestros enunciados y algoritmos dependen del parámetro η, tendremos que separarlos al menos en dos casos. Creemos que es prudente considerar solamente el caso $m > 0$, ya que las demostraciones y los enunciados para el caso $m < 0$ son completamente análogos.

En resumen, si $m > 0$, encontramos que las capas enteras que satisfacen la restricción (1.1b) están parametrizadas por $k \in \{\eta, \eta - 1, \ldots\}$. Además,

por el lema 1.2.6, todo $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ satisface la ecuación $\boldsymbol{q}^T\boldsymbol{x} = k \leq \eta \leq u$, donde u es el lado derecho de la restricción (1.1b).

Teorema 1.2.8. *Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y sea $\boldsymbol{q}$ su múltiplo coprimo. Entonces el problema* (1.1) *es infactible si y solo si $\boldsymbol{q} \geq \boldsymbol{0}$ y el lado derecho u de la restricción* (1.1b) *es negativo.*

Demostración. ($\Rightarrow$) Supongamos que $\boldsymbol{q} \geq \boldsymbol{0}$ y $u < 0$. Si $\boldsymbol{x} \in \mathbb{Z}^n_{\geq \boldsymbol{0}}$ entonces $\boldsymbol{q}^T\boldsymbol{x} \geq 0 > u$ y por lo tanto $\boldsymbol{x}$ no es factible. Luego,

$$\mathbb{Z}^n_{\geq \boldsymbol{0}} \cap \{\boldsymbol{x} : \boldsymbol{q}^T\boldsymbol{x} \leq u\} = \emptyset,$$

y el problema no es factible.

($\Leftarrow$) Procedemos por contraposición. Si $u \geq 0$ observamos que $\boldsymbol{0} \in \mathbb{Z}^n$ es factible. Se debe cumplir $u < 0$. Similarmente, si $q_i < 0$ para algún $i \in \{1, \ldots, n\}$, encontramos que $\lceil u/q_i \rceil \boldsymbol{e}_i \in \mathbb{Z}^n$ es factible:

$$\boldsymbol{q}^T \left\lceil \frac{u}{q_i} \right\rceil \boldsymbol{e}_i = q_i \left\lceil \frac{u}{q_i} \right\rceil \leq q_i \frac{u}{q_i} = u,$$

además, como $u < 0$, concluimos que $\lceil u/q_i \rceil \boldsymbol{e}_i$ es no negativo. □

El siguiente teorema muestra que nuestro análisis para resolver los problemas factibles debe dividirse en dos casos.

Teorema 1.2.9. *Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y sea $\boldsymbol{q}$ su múltiplo coprimo, de manera que $\boldsymbol{p} = m\boldsymbol{q}$ para alguna $m > 0$. Supongamos que el problema* (1.1) *es factible y tomemos η del lema 1.2.7. Entonces se satisface lo siguiente:*

1. *Si $q_i < 0$ para algún $i \in \{1, \ldots, n\}$, entonces la η-ésima capa entera $H_{\boldsymbol{q},\eta/\|\boldsymbol{q}\|^2}$ contiene un número infinito de puntos factibles.*
2. *Si $\boldsymbol{q} > \boldsymbol{0}$ entonces, para todo $k \in \{\eta, \eta - 1, \ldots, 0\}$, la k-ésima capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ contiene un número finito de puntos factibles.*

Demostración.

1. En la subsección 1.2.2 mostraremos que, como $\boldsymbol{q}$ es un vector coprimo, entonces existe un punto entero $\boldsymbol{x}$ que satisface la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = \eta$. Por el momento confiemos que esto es verdadero. Luego,
$$\boldsymbol{p}^T\boldsymbol{x} = m\boldsymbol{q}^T\boldsymbol{x} = m\eta = m\left\lfloor \frac{u}{m} \right\rfloor \leq m\frac{u}{m} = u,$$
y se satisface la restricción (1.1b).

   Como no tenemos asegurada la no negatividad de $\boldsymbol{x}$, construiremos un vector entero $\boldsymbol{x}^+$ que sí satisface la restricción de no negatividad y también la restricción presupuestaria $\boldsymbol{q}^T\boldsymbol{x}^+ = \eta$, de manera que $\boldsymbol{x}^+$ sí será factible.

   Definamos los siguientes conjuntos de índices:
$$I^+ \coloneqq \{i : q_i > 0\}, \quad I^\circ \coloneqq \{\ell : q_\ell = 0\}. \quad I^- \coloneqq \{j : q_j < 0\}.$$

   Podemos suponer sin pérdida de generalidad que I° es vacío. En efecto, si $x_k < 0$ para alguna $k \in I^\circ$, esa entrada no sería factible, pero fácilmente podríamos definir $x_k^+ = 0$ para hacerla factible.

   Por hipótesis, sabemos que $\boldsymbol{q}$ tiene una entrada negativa y por lo tanto $I^- \neq \emptyset$. Además, por la definición 1.2.1, $\boldsymbol{q}$ tiene una entrada positiva y por lo tanto $I^+ \neq \emptyset$. Luego, ambos conjuntos I^+ e I^- forman una partición del conjunto $\{1, \ldots, n\}$. Podemos escoger enteros positivos $\{c_i\}_{i\in I^+}$ que satisfagan simultáneamente
$$x_k + \sum_{i\in I^+} q_i c_i \geq 0, \quad \forall k \in I^-, \tag{1.11}$$
$$x_k - \sum_{j\in I^-} q_j c_k \geq 0, \quad \forall k \in I^+. \tag{1.12}$$

Definamos el vector $\boldsymbol{x}^+ \in \mathbb{Z}^n$ de manera que

$$x_k^+ \coloneqq \begin{cases} x_k + \sum_{i\in I^+} q_i c_i, & k \in I^-, \\ x_k - \sum_{j\in I^-} q_j c_k, & k \in I^+. \end{cases}$$

Se verifica que $\boldsymbol{x}^+$ es no negativo y, además,

$$\begin{aligned} \boldsymbol{q}^T\boldsymbol{x}^+ &= \boldsymbol{q}^T\boldsymbol{x} + \sum_{k\in I^-}\sum_{i\in I^+} q_k q_i c_i - \sum_{k\in I^+}\sum_{j\in I^-} q_k q_j c_k \\ &= \eta + \sum_{j\in I^-}\sum_{i\in I^+} q_j q_i c_i - \sum_{i\in I^+}\sum_{j\in I^-} q_i q_j c_i \\ &= \eta. \end{aligned}$$

Así pues, tenemos existencia de un punto factible. Para concluir que hay un número infinito de puntos factibles, basta observar que si la elección de coeficientes $\{c_i\}_{i\in I^+}$ satisface ambas desigualdades (1.11) y (1.12), entonces cualquier múltiplo entero positivo de estos coeficientes también las satisface.

2. Se sigue del teorema anterior que $u \geq 0$. Definamos

$$P_k \coloneqq \left(H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \cap \mathbb{Z}^n_{\geq \mathbf{0}}\right) = \left\{\boldsymbol{x} \in \mathbb{Z}^n : \boldsymbol{q}^T\boldsymbol{x} = k, \boldsymbol{x} \geq \mathbf{0}\right\}, \qquad (1.13)$$

y observemos que $P_k = \emptyset$ para todo k negativo, pues $\boldsymbol{q} > \mathbf{0}$ y por lo tanto $\boldsymbol{q}^T\boldsymbol{x} \geq 0$ para cualquier $\boldsymbol{x} \in \mathbb{Z}^n_{\geq \mathbf{0}}$. Esto implica que ningún punto sobre capas enteras con parámetros negativos es factible.

Sea $k \in \{\eta, \eta-1, \ldots, 0\}$. La capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ interseca los ejes positivos en $\frac{k}{q_i}\boldsymbol{e}_i$, así que definamos $\ell_i \coloneqq \lceil k/q_i \rceil$. Luego,

$$H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \cap \mathbb{R}^n_{\geq \mathbf{0}} \subseteq \prod_{i=1}^n [0, \ell_i],$$

pues del caso contrario podemos escoger $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ con $\boldsymbol{x} \geq \mathbf{0}$ y

también $x_i > \ell_i$ para alguna $i \in \{1, \ldots, n\}$. Por el lema 1.2.6 tenemos

$$k = \boldsymbol{q}^T \boldsymbol{x} > \sum_{j \neq i} q_j x_j + q_i \ell_i = \sum_{j \neq i} q_j x_j + q_i \left\lceil \frac{k}{q_i} \right\rceil$$

y entonces

$$0 \geq k - q_i \left\lceil \frac{k}{q_i} \right\rceil > \sum_{j \neq i} q_j x_j \geq 0.$$

De donde la última desigualdad se obtiene del hecho que ambos $\boldsymbol{q}$ y $\boldsymbol{x}$ son no negativos. Así pues, obtenemos una contradicción y la contención anterior es cierta. De esta manera, obtenemos

$$\left(H_{\boldsymbol{q}, k/\|\boldsymbol{q}\|^2} \cap \mathbb{R}^n_{\geq \boldsymbol{0}} \right) \cap \mathbb{Z}^n = P_k \subseteq \left(\prod_{i=1}^{n} [0, \ell_i] \right) \cap \mathbb{Z}^n = \prod_{i=1}^{n} \left([0, \ell_i] \cap \mathbb{Z} \right).$$

Pero $|[0, \ell_i] \cap \mathbb{Z}| = \ell_i + 1$. Así,

$$|P_k| \leq \prod_{i=1}^{n} (\ell_i + 1) < \infty.$$

Entonces la k-ésima capa entera contiene un número finito de puntos factibles.

□

Ciertamente el primer caso del teorema 1.2.9 es el menos interesante, pues conocemos inmediatamente el valor óptimo de estas instancias. No obstante, existen muchos elementos en común que comparten ambos casos. También es cierto que esta división dejará de existir una vez que introduzcamos múltiples restricciones en el capítulo 4.

Además, antes de analizar los dos casos que el teorema anterior impone, primero debemos mostrar que la ecuación lineal diofantina $\boldsymbol{q}^T \boldsymbol{x} = k$ admite soluciones enteras para toda $k \in \mathbb{Z}$ siempre que las entradas de $\boldsymbol{q}$ sean coprimas. Habíamos supuesto esto en la demostración anterior.

Así también, la construcción de soluciones enteras de ecuaciones lineales diofantinas proveerá herramientas teóricas útiles para demostrar la gran mayoría de resultados que presentaremos. Por lo tanto, la siguiente subsección se encarga de construir solamente soluciones enteras de estas ecuaciones. Será cuestión de los capítulos 2 y 3 obtener soluciones sean no negativas.

1.2.2. Construcción de soluciones enteras

Debido al teorema 1.2.5, las soluciones del problema (1.1) se encuentran en una capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$. Luego, por el lema 1.2.6, los puntos $\boldsymbol{x} \in \mathbb{Z}^n$ que se encuentran sobre esa capa satisfacen la ecuación lineal diofantina

$$\boldsymbol{q}^T\boldsymbol{x} = q_1x_1 + q_2x_2 + \cdots + q_nx_n = k. \tag{1.14}$$

con $\boldsymbol{q} = (q_1, \ldots, q_n)$. Como hemos mencionado previamente, podemos suponer sin pérdida de generalidad que ninguna entrada de $\boldsymbol{q}$ es nula.

En la sección 1.1.1 mostramos condiciones de existencia de este tipo de ecuaciones, así como su construcción, cuando $n = 2$. Partimos de la observación que podemos resolver recursivamente esta ecuación. Definamos, por conveniencia, $g_1 \coloneqq \operatorname{mcd}\{q_1, \ldots, q_n\}$ y también $\omega_1 \coloneqq k$. Puesto que $\boldsymbol{q}$ es un vector con entradas coprimas, sabemos que $g_1 = 1$. También definamos

$$\omega_2 \coloneqq \frac{q_2}{g_2 \cdot g_1}x_2 + \cdots + \frac{q_n}{g_2 \cdot g_1}x_n, \tag{1.15}$$

donde $g_2 \coloneqq \operatorname{mcd}\{q_2/g_1, \ldots, q_n/g_1\}$. Como $q_n \neq 0$, tenemos que g_2 está bien definido y además es positivo. Así, la ecuación (1.14) es equivalente a

$$\frac{q_1}{g_1}x_1 + g_2\omega_2 = \omega_1. \tag{1.16}$$

Observemos que

$$\operatorname{mcd}\left\{\frac{q_1}{g_1}, g_2\right\} = \operatorname{mcd}\left\{\frac{q_1}{g_1}, \operatorname{mcd}\left\{\frac{q_2}{g_1}, \ldots, \frac{q_n}{g_1}\right\}\right\}$$

$$= \operatorname{mcd}\left\{ \frac{q_1}{g_1}, \frac{q_2}{g_1}, \ldots, \frac{q_n}{g_1} \right\} = 1.$$

Por el teorema 1.1.8, existen soluciones enteras para todo $\omega_1 \in \mathbb{Z}$. Como q_1/g_1 y g_2 son coprimos, encontramos que sus coeficientes de Bézout (ver definición 1.1.9) asociados x_1', ω_2' son soluciones particulares de la ecuación

$$\frac{q_1}{g_1} x_1' + g_2 \omega_2' = 1.$$

Deducimos del teorema 1.1.10 que las soluciones de la ecuación (1.16) están dadas por

$$\begin{cases} x_1 = \omega_1 x_1' + g_2 t_1, \\ \omega_2 = \omega_1 \omega_2' - \frac{q_1}{g_1} t_1, \end{cases} \tag{1.17}$$

donde $t_1 \in \mathbb{Z}$ es una variable libre.

Observación. Los coeficientes de Bézout x_1' y ω_2' dependen exclusivamente de $\boldsymbol{q}$ y no del punto $\boldsymbol{x}$. En efecto, x_1' está asociado a q_1/g_1 y ω_2' está asociado a g_2. Pero ambos g_1 y g_2 son el máximo común divisor de $q_1, \ldots q_n$ y de $q_1/g_1, \ldots, q_n/g_1$, respectivamente.

Para el siguiente paso de la recursión, escogemos cualquier $t_1 \in \mathbb{Z}$ para fijar ω_2. Tenemos de (1.15) que debemos resolvemos la ecuación

$$\frac{q_2}{g_2 \cdot g_1} x_2 + \frac{q_3}{g_2 \cdot g_1} x_3 + \cdots + \frac{q_n}{g_2 \cdot g_1} x_n = \omega_2. \tag{1.18}$$

Como $g_2 = \operatorname{mcd}\{q_2/g_1, \ldots, q_n/g_1\}$, sabemos del lema 1.1.6 que

$$\operatorname{mcd}\left\{ \frac{q_2}{g_2 \cdot g_1}, \ldots, \frac{q_n}{g_2 \cdot g_1} \right\} = 1.$$

En el mismo espíritu del primer paso de la recursión, definimos

$$\omega_3 \coloneqq \frac{q_3}{g_3 \cdot g_2 \cdot g_1} x_3 + \cdots + \frac{q_n}{g_3 \cdot g_2 \cdot g_1} x_n,$$

donde

$$g_3 \coloneqq \operatorname{mcd}\left\{ \frac{q_3}{g_2 \cdot g_1}, \ldots, \frac{q_n}{g_2 \cdot g_1} \right\}.$$

Nuevamente, como q_n es distinto de cero, g_3 está bien definido y además es positivo. Así pues, la ecuación (1.18) es equivalente a

$$\frac{q_2}{g_2 \cdot g_1} x_2 + g_3 \omega_3 = \omega_2. \tag{1.19}$$

También se cumple que

$$\operatorname{mcd}\left\{ \frac{q_2}{g_2 \cdot g_1}, g_3 \right\} = 1,$$

y entonces (1.19) tiene una infinidad de soluciones para todo $\omega_2 \in \mathbb{Z}$, las cuales están dadas por

$$\begin{cases} x_2 = \omega_2 x_2' + g_3 t_2, \\ \omega_3 = \omega_2 \omega_3' - \frac{q_2}{g_2 \cdot g_1} t_2, \end{cases}$$

donde $t_2 \in \mathbb{Z}$ es una variable libre, y x_2', ω_3' son los coeficientes de Bézout asociados a $\frac{q_2}{g_2 \cdot g_1}$ y g_3, respectivamente.

De manera general, para $i \in \{1, \ldots, n-2\}$, el i-ésimo paso de la recursión provee la ecuación

$$\frac{q_i}{\prod_{j=1}^{i} g_j} x_i + \frac{q_{i+1}}{\prod_{j=1}^{i} g_j} x_{i+1} + \cdots + \frac{q_n}{\prod_{j=1}^{i} g_j} x_n = \omega_i, \tag{1.20}$$

donde

$$g_i \coloneqq \operatorname{mcd}\left\{ \frac{q_i}{\prod_{j=1}^{i-1} g_j}, \ldots, \frac{q_n}{\prod_{j=1}^{i-1} g_j} \right\}, \tag{1.21}$$

por el lema 1.1.6 se sigue que

$$\operatorname{mcd}\left\{ \frac{q_i}{\prod_{j=1}^{i} g_j}, \ldots, \frac{q_n}{\prod_{j=1}^{i} g_j} \right\} = 1. \tag{1.22}$$

Ahora bien, definamos

$$g_{i+1} \coloneqq \operatorname{mcd}\left\{ \frac{q_{i+1}}{\prod_{j=1}^{i} g_j}, \ldots, \frac{q_n}{\prod_{j=1}^{i} g_j} \right\}. \tag{1.23}$$

Como q_n es distinto de cero, se sigue que g_{i+1} está bien definido y es positivo. Definamos

$$\omega_{i+1} \coloneqq \frac{q_{i+1}}{\prod_{j=1}^{i+1} g_j} x_{i+1} + \cdots + \frac{q_n}{\prod_{j=1}^{i+1} g_j} x_n,$$

de manera que la ecuación (1.20) es equivalente a

$$\frac{q_i}{\prod_{j=1}^{i} g_j} x_i + g_{i+1}\omega_{i+1} = \omega_i. \tag{1.24}$$

A partir de (1.22) y de (1.23), encontramos que

$$\operatorname{mcd}\left\{ \frac{q_i}{\prod_{j=1}^{i} g_j}, g_{i+1} \right\} = \operatorname{mcd}\left\{ \frac{q_i}{\prod_{j=1}^{i} g_j}, \frac{q_{i+1}}{\prod_{j=1}^{i} g_j}, \ldots, \frac{q_n}{\prod_{j=1}^{i} g_j} \right\} = 1,$$

y del teorema 1.1.8 se sigue que la ecuación (1.24) tiene soluciones enteras para todo $\omega_i \in \mathbb{Z}$. Por el teorema 1.1.10, las soluciones enteras de (1.24) están dadas por

$$\begin{cases} x_i = \omega_i x_i' + g_{i+1} t_i, \\ \omega_{i+1} = \omega_i \omega_{i+1}' - \frac{q_i}{\prod_{j=1}^{i} g_j} t_i, \end{cases} \tag{1.25}$$

donde $t_i \in \mathbb{Z}$ es la i-ésima variable libre. Es valioso mencionar, otra vez, que los coeficientes de Bézout x_i', ω_{i+1}' dependen exclusivamente de $\boldsymbol{q}$ a través de sus entradas q_i y de los máximos común divisores entre ellas. En efecto, por el teorema 1.1.4, estos coeficientes son soluciones particulares de la ecuación

$$\frac{q_i}{\prod_{j=1}^{i} g_j} x_i' + g_{i+1}\omega_{i+1}' = 1. \tag{1.26}$$

Finalmente, en el último paso de la recursión obtenemos la ecuación lineal

diofantina

$$\frac{q_{n-1}}{\prod_{j=1}^{n-1} g_j} x_{n-1} + \frac{q_n}{\prod_{j=1}^{n-1} g_j} x_n = \omega_{n-1}. \tag{1.27}$$

Por construcción, los coeficientes de x_{n-1} y x_n son coprimos. A causa del teorema 1.1.10 las soluciones enteras están dadas por

$$\begin{cases} x_{n-1} = \omega_{n-1} x'_{n-1} + \frac{q_n}{\prod_{j=1}^{n-1} g_j} t_{n-1}, \\ x_n = \omega_{n-1} x'_n - \frac{q_{n-1}}{\prod_{j=1}^{n-1} g_j} t_{n-1}, \end{cases} \tag{1.28}$$

donde x'_{n-1}, x'_n son los coeficientes de Bézout asociados a $\frac{q_n}{\prod_{j=1}^{n-1} g_j}$ y $\frac{q_{n-1}}{\prod_{j=1}^{n-1} g_j}$, respectivamente, por lo que satisfacen

$$\frac{q_{n-1}}{\prod_{j=1}^{n-1} g_j} x'_{n-1} + \frac{q_n}{\prod_{j=1}^{n-1} g_j} x'_n = 1. \tag{1.29}$$

Hemos demostrado, que la ecuación lineal diofantina (1.14) tiene al menos una solución, siempre que $\boldsymbol{q}$ sea un vector coprimo. Así pues, saldamos nuestra cuenta pendiente con respecto a una parte de la demostración 1.2.9. Además, construimos una infinidad de soluciones enteras, pues la elección de cada variable libre $t_i \in \mathbb{Z}$ provee una solución distinta. Aún más, por el teorema 1.1.10, sabemos que el conjunto de estas soluciones es exhaustiva. En la siguiente subsección estableceremos, a partir de la construcción de soluciones aquí obtenida, una relación lineal entre las soluciones enteras $\boldsymbol{x} = (x_1, \ldots, x_n)$ de la ecuación lineal diofantina $\boldsymbol{q}^T \boldsymbol{x} = k$ y el vector de variables libres $\boldsymbol{t} = (t_1, \ldots, t_{n-1})$.

Con respecto a la no negatividad de las soluciones mencionamos brevemente lo siguiente. Observamos de (1.25) que t_i debe satisfacer

$$t_i \geq \left\lceil -\frac{\omega_i x'_i}{g_{i+1}} \right\rceil, \tag{1.30}$$

para todo $i \in \{1, \ldots, n-2\}$. Ahora bien, para asegurar la no negatividad de x_{n-1} y x_n, observamos de (1.27) que dependemos de los signos de q_{n-1} y

de q_n. Debido al teorema 1.2.9, relegamos esta discusión para los siguientes dos capítulos.

1.2.3. Soluciones y variables libres

Hemos encontrado una relación entre un vector de variables libres $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ y un vector solución $\boldsymbol{x} \in \mathbb{Z}^n$ de la ecuación (1.14). Sabemos de (1.25) que la relación está dada de manera recursiva. Puesto que deseamos establecer una transformación lineal entre $\boldsymbol{t}$ y $\boldsymbol{x}$, resulta sumamente conveniente determinar una forma cerrada de esta relación. Recordemos que habíamos definido, por construcción, $\omega_1 \coloneqq k$. Combinando esto con la segunda igualdad de (1.25), obtenemos la relación de recurrencia

$$\begin{cases} \omega_1 = k, \\ \omega_{i+1} = \omega_i \cdot \omega'_{i+1} - \frac{q_i}{\prod_{\ell=1}^{i} g_\ell} \cdot t_i. \end{cases} \tag{1.31}$$

Lema 1.2.10. *La forma cerrada de la relación de recurrencia* (1.31) *está dada para toda* $i \in \mathbb{N}$ *por*

$$\omega_i = k \cdot \prod_{j=2}^{i} \omega'_j - \sum_{j=1}^{i-1} t_j \cdot \frac{q_j}{\prod_{\ell=1}^{j} g_\ell} \cdot \prod_{\ell=j+2}^{i} \omega'_\ell, \tag{1.32}$$

donde $q_1, q_2, \ldots$ *son enteros no nulos coprimos,* g_i *está definido recursivamente por* (1.23) *con* $g_1 = 1$, ω'_{i+1} *satisface* (1.26), *y* $k, t_1, t_2, \ldots$ *son enteros. Asignamos el valor de 0 a la suma vacía y el valor de 1 al producto vacío.*

Demostración. Utilizaremos inducción matemática sobre $i \in \mathbb{N}$. Observemos que para $i = 1$, se tiene que

$$\omega_1 = k \cdot \prod_{j=2}^{1} \omega'_j - \sum_{j=1}^{0} t_j \cdot \frac{q_j}{\prod_{\ell=1}^{j} g_\ell} \cdot \prod_{\ell=j+2}^{1} \omega'_\ell = k,$$

debido a que definimos el producto vacío como 1 y la suma vacía como 0.

Ahora hacemos uso de la hipótesis inductiva, así que supongamos que (1.32) se satisface para alguna $i \in \mathbb{N}$, entonces, debemos mostrar que también se satisface para $i+1$. Veamos que

$$
\begin{aligned}
& k \cdot \prod_{j=2}^{i+1} \omega'_j - \sum_{j=1}^{i} t_j \cdot \frac{q_j}{\prod_{\ell=1}^{j} g_\ell} \cdot \prod_{\ell=j+2}^{i+1} \omega'_\ell \\
&= k \cdot \prod_{j=2}^{i} \omega'_j \cdot \omega'_{i+1} - \sum_{j=1}^{i-1} t_j \cdot \frac{q_j}{\prod_{\ell=1}^{j} g_\ell} \cdot \prod_{\ell=j+2}^{i} \omega'_\ell \cdot \omega'_{i+1} - \frac{q_i}{\prod_{\ell=1}^{i} g_\ell} \cdot \prod_{\ell=i+2}^{i+1} \omega'_\ell \cdot t_i \\
&= \left(k \cdot \prod_{j=2}^{i} \omega'_j - \sum_{j=1}^{i-1} t_j \cdot \frac{q_j}{\prod_{\ell=1}^{j} g_\ell} \cdot \prod_{\ell=j+2}^{i} \omega'_\ell \right) \omega'_{i+1} - \frac{q_i}{\prod_{\ell=1}^{i} g_\ell} \cdot t_i \\
&= \omega_i \cdot \omega'_{i+1} - \frac{q_i}{\prod_{\ell=1}^{i} g_\ell} \cdot t_i \\
&= \omega_{i+1}.
\end{aligned}
$$

Observemos que el producto de la derecha en el segundo renglón es 1 por ser un producto vacío, y que del tercer al cuarto renglón hicimos uso de la hipótesis inductiva. Así pues, por el principio de inducción se sigue que (1.32) satisface (1.31) para todo $i \in \mathbb{N}$. Concluimos con que esta fórmula es la forma cerrada de la relación de recurrencia propuesta. □

Ahora que encontramos una forma cerrada a la relación de recurrencia (1.31), somos capaces de establecer una transformación lineal entre el vector de variables libres $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ y el vector solución $\boldsymbol{x} \in \mathbb{Z}^n$ de (1.14). Definamos, por conveniencia, los coeficientes $m_{ij} \in \mathbb{Z}$ con $i > j$ como

$$
m_{ij} \coloneqq \frac{q_j}{\prod_{\ell=1}^{j} g_\ell} \cdot \prod_{\ell=j+2}^{i} \omega'_\ell. \tag{1.33}
$$

Sustituyendo en la forma cerrada (1.32), obtenemos la fórmula simplificada

$$\omega_i = k \cdot \prod_{j=2}^{i} \omega'_j - \sum_{j=1}^{i-1} m_{ij} t_j, \tag{1.34}$$

Así pues, juntando esto último con (1.25), obtenemos para $i \in \{1, \dots, n-2\}$:

$$\begin{aligned} x_i &= \omega_i \cdot x'_i + g_{i+1} t_i \\ &= k \cdot \prod_{j=2}^{i} \omega'_j \cdot x'_i - \sum_{j=1}^{i-1} m_{ij} x'_i t_j + g_{i+1} t_i. \end{aligned} \tag{1.35}$$

Similarmente, usando (1.34) y sustituyendo en (1.28), llegamos a

$$x_{n-1} = k \cdot \prod_{j=2}^{n-1} \omega'_j \cdot x'_{n-1} - \sum_{j=1}^{n-2} m_{n-1,j} x'_{n-1} t_j + \frac{q_n}{\prod_{j=1}^{n-1} g_j} t_{n-1}, \tag{1.36a}$$

$$x_n = k \cdot \prod_{j=2}^{n-1} \omega'_j \cdot x'_n - \sum_{j=1}^{n-2} m_{n-1,j} x'_n t_j - \frac{q_{n-1}}{\prod_{j=1}^{n-1} g_j} t_{n-1}. \tag{1.36b}$$

Así pues, definimos el vector $\boldsymbol{\nu} \in \mathbb{Z}^n$ como

$$\nu_i \coloneqq x'_i \cdot \prod_{j=2}^{\min\{i, n-1\}} \omega'_j. \tag{1.37}$$

También definimos la matriz $M \in \mathbb{Z}^{n \times (n-1)}$ a través de

$$M_{ij} \coloneqq \begin{cases} -m_{ij} x'_i, & j < i, \\ g_{i+1}, & i = j < n-1, \\ \frac{q_n}{\prod_{k=1}^{n-1} g_k}, & i = j = n-1, \\ -\frac{q_{n-1}}{\prod_{k=1}^{n-1} g_k}, & i = n, j = n-1, \\ 0, & \text{e.o.c.} \end{cases} \tag{1.38}$$

Sustituyendo en (1.35) y (1.36) obtenemos la siguiente proposición original.

Proposición 1.2.11. *Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector con entradas coprimas y última entrada no nula. Entonces todas las soluciones enteras de la ecuación* (1.14) *son de la forma*

$$\boldsymbol{x} = k\boldsymbol{\nu} + M\boldsymbol{t}, \tag{1.39}$$

donde $\boldsymbol{t} \in \mathbb{Z}^{n-1}$, $\boldsymbol{\nu} \in \mathbb{Z}^n$ está definida en (1.37) *y $M \in \mathbb{Z}^{n\times(n-1)}$ en* (1.38).

Observación. Si $q_n = 0$, entonces puede ser que M no esté bien definida. Por ejemplo, si $q_n = q_{n-1} = 0$, encontramos que

$$g_{n-1} \coloneqq \operatorname{mcd}\left\{ \frac{q_{n-1}}{\prod_{j=1}^{n-2} g_j}, \frac{q_n}{\prod_{j=1}^{n-2} g_j} \right\} = \operatorname{mcd}\{0,0\}\,.$$

Pero el máximo común divisor de dos números no está bien definido si ambos son cero. Esto implica que la entrada $M_{n-2,n-2} \coloneqq g_{n-1}$ no está bien definida.

En la subsección 1.2.2 mencionamos a lo largo de la construcción de soluciones que los coeficientes de Bézout ω_i', x_i' están asociados a términos exclusivamente dependientes de $\boldsymbol{q}$, por lo que no dependen de la elección $\boldsymbol{x} \in \mathbb{Z}$. Se sigue de (1.37) y de (1.38) que $\boldsymbol{\nu}$ y M dependen exclusivamente de $\boldsymbol{q}$ y no de $\boldsymbol{x}$.

Lema 1.2.12. *Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector coprimo. Entonces el vector $\boldsymbol{\nu} \in \mathbb{Z}^n$ definido en* (1.37) *satisface $\boldsymbol{q}^T\boldsymbol{\nu} = 1$.*

Demostración. Primero mostramos por inducción hacia atrás que se cumple

$$\sum_{j=i}^{n} q_j\nu_j = \prod_{j=2}^{i} \omega_j' \cdot \prod_{j=1}^{i} g_j, \tag{1.40}$$

para todo $i \in \{1,\dots,n-1\}$. Empezamos con el caso base $i = n-1$. De

(1.37), encontramos que

$$q_{n-1}\nu_{n-1} + q_n\nu_n = q_{n-1}\left(x'_{n-1}\cdot\prod_{j=2}^{n-1}\omega'_j\right) + q_n\left(x'_n\cdot\prod_{j=2}^{n-1}\omega'_j\right)$$
$$= \prod_{j=2}^{n-1}\omega'_j\cdot\left(q_{n-1}x'_{n-1} + q_nx'_n\right). \tag{1.41}$$

Recordemos que x'_{n-1} y x'_n son coeficientes de Bézout asociados a los coeficientes del lado izquierdo de (1.27), los cuales son coprimos. Entonces se cumple, por el teorema 1.1.4, que

$$\frac{q_{n-1}}{\prod_{j=1}^{n-1}g_j}x'_{n-1} + \frac{q_n}{\prod_{j=1}^{n-1}g_j}x'_n = 1,$$

o, equivalentemente,

$$q_{n-1}x'_{n-1} + q_nx'_n = \prod_{j=1}^{n-1}g_j.$$

Sustituyendo en (1.41), obtenemos la base de la inducción:

$$q_{n-1}\nu_{n-1} + q_n\nu_n = \prod_{j=2}^{n-1}\omega'_j\cdot\prod_{j=1}^{n-1}g_j.$$

Supongamos que (1.40) se satisface para alguna $i \in \{2,\ldots,n-1\}$. Para ver qué pasa con $i-1$, ocupamos (1.37) y utilizamos la hipótesis inductiva:

$$\sum_{j=i-1}^{n} q_j\nu_j = q_{i-1}\nu_{i-1} + \sum_{j=i}^{n} q_j\nu_j$$
$$= \prod_{j=2}^{i-1}\omega'_j\cdot q_{i-1}x'_{i-1} + \prod_{j=2}^{i}\omega'_j\cdot\prod_{j=1}^{i}g_j$$
$$= \prod_{j=2}^{i-1}\omega'_j\cdot\left(q_{i-1}x'_{i-1} + \omega'_i\prod_{j=1}^{i}g_j\right).$$

Nuevamente, x'_{i-1} y ω'_i son coeficientes de Bézout asociados, respectivamente, a $\frac{q_{i-1}}{\prod_{j=1}^{i-1} g_j}$ y g_i, los cuales son coprimos. De esta manera satisfacen (1.26) pero sustituyendo i por $i-1$. Es decir, se satisface

$$\frac{q_{i-1}}{\prod_{j=1}^{i-1} g_j} x'_{i-1} + g_i \omega'_i = 1,$$

o, equivalentemente,

$$q_{i-1} x'_{i-1} + \omega'_i \prod_{j=1}^{i} g_j = \prod_{j=1}^{i-1} g_j.$$

Sustituyendo, obtenemos el resultado (1.40) para $i-1$. Así, por inducción hacía atrás, (1.40) se cumple para todo $i \in \{1, \ldots, n-1\}$. Finalmente, para demostrar este lema, observamos que

$$\boldsymbol{q}^T \boldsymbol{\nu} = \sum_{j=1}^{n} q_j \nu_j = \prod_{j=2}^{1} \omega'_j \cdot \prod_{j=1}^{1} g_j = g_1 = 1.$$

El primer producto es uno por ser el producto vacío. Recordemos también que g_1 es el máximo común divisor de $q_1, \ldots, q_n$, los cuales son coprimos por hipótesis, y entonces $g_1 = 1$. □

Lema 1.2.13. *Si $\boldsymbol{q} \in \mathbb{Z}^n$ es un vector coprimo con $q_n \neq 0$, entonces $\boldsymbol{q}$ genera el complemento ortogonal de la imagen de M, es decir, se satisface* $\operatorname{gen}\{\boldsymbol{q}\} = \operatorname{im}\{M\}^{\perp}$, *donde la matriz M está definida en* (1.38).

Demostración. Por el teorema de la dimensión sabemos que se satisface $\operatorname{im}\{M\}^{\perp} = \ker\{M^T\}$, así que basta mostrar que $\operatorname{gen}\{\boldsymbol{q}\} = \ker\{M^T\}$. La matriz M es triangular inferior y su diagonal principal es distinta de cero. En efecto, para todo $i \in \{1, \ldots, n-2\}$, tenemos

$$M_{ii} = g_{i+1} = \operatorname{mcd}\left\{ \frac{q_i}{\prod_{j=1}^{i} g_j}, \ldots, \frac{q_n}{\prod_{j=1}^{i} g_j} \right\}.$$

Pero el máximo común divisor entre cualesquiera enteros siempre es positivo. También tenemos por hipótesis que

$$M_{n-1,n-1} = \frac{q_n}{\prod_{j=1}^{n-1} g_j} \neq 0.$$

Se sigue que las $n-1$ columnas de M son linealmente independientes. Por lo tanto, M^T tiene $n-1$ renglones linealmente independientes y entonces $\dim \ker\{M^T\} = 1$, así que basta mostrar que $\boldsymbol{q} \in \ker\{M^T\}$.

Sea $\boldsymbol{x} \in \mathbb{Z}^n$. Por el teorema 1.2.5, existe una capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ que contiene a $\boldsymbol{x}$. Así, por el lema 1.2.6, $\boldsymbol{x}$ satisface la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$. Por la proposición 1.2.11 sabemos que existe $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ tal que $\boldsymbol{x} = k\boldsymbol{\nu} + M\boldsymbol{t}$. Luego, por el lema 1.2.12, tenemos

$$k = \boldsymbol{q}^T\boldsymbol{x} = k\boldsymbol{q}^T\boldsymbol{\nu} + \boldsymbol{q}^T M\boldsymbol{t} = k + (\boldsymbol{q}^T M)\boldsymbol{t}.$$

De donde obtenemos $(\boldsymbol{q}^T M)\boldsymbol{t} = 0$. Pero $\boldsymbol{x}$ fue arbitrario, así que también lo fue $\boldsymbol{t}$. Entonces $\boldsymbol{q}^T M = \boldsymbol{0}^T$, lo que implica $\boldsymbol{q} \in \ker\{M^T\}$. □

La gran mayoría de nuestra argumentación para demostrar los resultados ha sido fundamentada a través de las capas enteras $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$, al igual que por el teorema 1.2.5. Sin embargo, estas capas enteras contienen puntos que no son enteros y por lo tanto no son de nuestro interés, ya que nos gustaría concentrarnos exclusivamente en puntos enteros, al mismo tiempo que buscamos caracterizarlos por medio de $\boldsymbol{q}$.

Definición 1.2.14 ([Sch98])**.** Decimos que un subconjunto Λ del espacio vectorial $(\mathbb{R}^n, +, \cdot)$ es un **grupo aditivo** si

1. $\boldsymbol{0} \in \Lambda$, y
2. si $\boldsymbol{x}, \boldsymbol{y} \in \Lambda$, entonces $\boldsymbol{x} + \boldsymbol{y} \in \Lambda$, y también $-\boldsymbol{x} \in \Lambda$.

Además, decimos que Λ es una **red** si existen vectores $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ linealmente independientes tales que

$$\Lambda = \{\boldsymbol{x} : \boldsymbol{x} = \lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_n \boldsymbol{v}_n, \lambda_i \in \mathbb{Z}\}.$$

A los vectores $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ los llamamos la **base de la red** Λ.

Ejemplo 1.2.15. No es difícil ver que $\mathbb{Z}^n$ es un grupo aditivo. Si consideramos los vectores canónicos $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, entonces encontramos que son linealmente independientes, pero también se cumple

$$\mathbb{Z}^n = \{\boldsymbol{x} : \boldsymbol{x} = \lambda_1 \boldsymbol{e}_1 + \cdots + \lambda_n \boldsymbol{e}_n, \lambda_i \in \mathbb{Z}\}.$$

De esta manera, $\mathbb{Z}^n$ es una red que tiene como base canónica a los vectores $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

Teorema 1.2.16. *Sean $\boldsymbol{q} \in \mathbb{Z}^n$ un vector coprimo y supongamos que q_n es distinto de cero. Entonces $\boldsymbol{\nu}$ y las columnas de M (definidas en* (1.37) *y* (1.38)*, respectivamente) forman una base de la red $\mathbb{Z}^n$.*

Demostración. En el lema 1.2.13 mostramos que las columnas de M son linealmente independientes. Ahora mostramos por contradicción que $\boldsymbol{\nu}$ es linealmente independiente de las columnas de M, así que supongamos que no lo es, por lo que existen escalares $\lambda_1, \ldots, \lambda_{n-1}$, no todos cero, tales que

$$\boldsymbol{\nu} = \lambda_1 \boldsymbol{m}_1 + \cdots + \lambda_{n-1} \boldsymbol{m}_{n-1},$$

donde $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{n-1}$ son las columnas de M. De los lemas 1.2.12 y 1.2.13 obtenemos

$$1 = \boldsymbol{q}^T \boldsymbol{\nu} = \lambda_1 \boldsymbol{q}^T \boldsymbol{m}_1 + \cdots + \lambda_{n-1} \boldsymbol{q}^T \boldsymbol{m}_{n-1} = 0,$$

lo cual es una contradicción. Se sigue que $\{\boldsymbol{\nu}, \boldsymbol{m}_1, \ldots, \boldsymbol{m}_{n-1}\}$ es un conjunto de vectores linealmente independiente.

Ahora bien, sea $\boldsymbol{x} \in \mathbb{Z}^n$, por el teorema 1.2.5, sabemos que $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ para alguna $k \in \mathbb{Z}$. Por el lema 1.2.6, $\boldsymbol{x}$ satisface la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$. Por la proposición 1.2.11 sabemos que existe $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ que satisface

$$\boldsymbol{x} = k\boldsymbol{\nu} + M\boldsymbol{t} = k\boldsymbol{\nu} + t_1\boldsymbol{m}_1 + \cdots + t_{n-1}\boldsymbol{m}_{n-1}.$$

Pero $\boldsymbol{x}$ fue arbitrario, lo que implica que

$$\mathbb{Z}^n = \{k\boldsymbol{\nu} + t_1\boldsymbol{m}_1 + \cdots + t_{n-1}\boldsymbol{m}_{n-1} : k, t_1, \ldots, t_{n-1} \in \mathbb{Z}\}$$

De esta manera, se cumple que $\{\boldsymbol{\nu}, \boldsymbol{m}_1, \ldots, \boldsymbol{m}_{n-1}\}$ forma una base de la red $\mathbb{Z}^n$. □

El siguiente corolario es presentado sin demostración, pero cabe mencionar que es una consecuencia directa del teorema anterior junto con las equivalencias encontradas en el teorema 4.3 de [Sch98].

Corolario 1.2.17. *Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector coprimo y supongamos que q_n es distinto de cero. Consideremos $\boldsymbol{\nu} \in \mathbb{Z}^n$ y las columnas $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{n-1} \in \mathbb{Z}^n$ de la matriz M (definidas en* (1.37) *y* (1.38)*, respectivamente), entonces la matriz*

$$[\boldsymbol{\nu} \mid \boldsymbol{m}_1 \mid \cdots \mid \boldsymbol{m}_{n-1}] \in \mathbb{Z}^{n\times n}$$

es unimodular, es decir, su determinante es ± 1.

Geométricamente, tenemos que si $\boldsymbol{q} \in \mathbb{Z}^n$ es un vector coprimo tal que $q_n \neq 0$, entonces $\boldsymbol{q}$ induce una descomposición de $\mathbb{Z}^n$ como la suma directa de las subredes Λ_p y Λ_h, donde

$$\Lambda_p \coloneqq \{k\boldsymbol{\nu} : k \in \mathbb{Z}\}, \quad \Lambda_h \coloneqq \{M\boldsymbol{t} : \boldsymbol{t} \in \mathbb{Z}^{n-1}\}. \tag{1.42}$$

Para todo vector $\boldsymbol{x} \in \Lambda_p$ existe una $k \in \mathbb{Z}$ tal que $\boldsymbol{x} = k\boldsymbol{\nu}$. Luego, por el lema 1.2.12, $\boldsymbol{x}$ es una solución particular de la ecuación lineal diofantina

$\boldsymbol{q}^T\boldsymbol{x} = k$. Similarmente, por el lema 1.2.13, todo vector $\boldsymbol{x} \in \Lambda_h$ es una solución de la ecuación lineal diofantina homogénea $\boldsymbol{q}^T\boldsymbol{x} = 0$.

Por el teorema 1.2.16, encontramos que $\mathbb{Z}^n$ se puede descomponer como la suma directa de sus subredes Λ_p y Λ_h, es decir, $\mathbb{Z}^n = \Lambda_p \oplus \Lambda_h$. De esta manera, tenemos que todo vector $\boldsymbol{x} \in \mathbb{Z}^n$ se puede descomponer de la forma $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_h$, donde $\boldsymbol{x}_p \in \Lambda_p$ por lo que $\boldsymbol{q}^T\boldsymbol{x}_p = k$ para alguna k entera y, similarmente, $\boldsymbol{x}_h \in \Lambda_h$ lo que implica que $\boldsymbol{q}^T\boldsymbol{x}_h = 0$. Ahora bien, por el teorema 1.2.5 y el lema 1.2.6, $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ y entonces el componente $\boldsymbol{x}_p$ se encarga de que $\boldsymbol{x}$ se encuentre sobre la k-ésima capa entera, mientras que el componente $\boldsymbol{x}_h$ se desliza sobre esta capa. Esta idea de descomponer el espacio a partir de soluciones particulares y homogéneas de una ecuación no es novedosa, pues es sumamente frecuente hacerlo para espacios vectoriales.

Observemos que si $q_n = 0$, es válido permutar las entradas de $\boldsymbol{q}$ de manera que el vector permutado $\tilde{\boldsymbol{q}}$ cumpla el supuesto $\tilde{q}_n \neq 0$. Podemos preguntarnos cómo se relacionan las imágenes de las matrices M y $\tilde{M}$ definidas en (1.38) y utilizando $\boldsymbol{q}$ y $\tilde{\boldsymbol{q}}$ respectivamente para construirlas. Pero si $q_n = 0$, entonces puede ser que M no esté bien definida. Requerimos de un supuesto más fuerte para responder la pregunta.

Corolario 1.2.18. *Sea $\boldsymbol{q}$ un vector coprimo y sea $\tilde{\boldsymbol{q}}$ un vector con las entradas de $\boldsymbol{q}$ permutadas. Supongamos que $q_n, \tilde{q}_n \neq 0$. Sean M y $\tilde{M}$ sus respectivas matrices definidas en* (1.38)*, entonces se satisface*

$$\ker\{M\} \cong \ker\{\tilde{M}\}, \quad \operatorname{im}\{M\} \cong \operatorname{im}\{\tilde{M}\},$$
$$\ker\{M^T\} \cong \ker\{\tilde{M}^T\}, \quad \operatorname{im}\{M^T\} \cong \operatorname{im}\{\tilde{M}^T\}.$$

Demostración. Por el teorema de la dimensión basta mostrar uno de estos isomorfismos, así que mostramos que $\operatorname{im}\{M\} \cong \operatorname{im}\{\tilde{M}\}$. Por hipótesis, existe una matriz de permutación $P \in \mathbb{Z}^{n\times n}$ tal que $\tilde{\boldsymbol{q}} = P\boldsymbol{q}$. Puesto que P es

invertible, tenemos $\text{gen}\{\boldsymbol{q}\} \cong \text{gen}\{\tilde{\boldsymbol{q}}\}$. Usando el lema 1.2.13 obtenemos

$$\text{im}\{M\}^{\perp} = \text{gen}\{\boldsymbol{q}\} \cong \text{gen}\{\tilde{\boldsymbol{q}}\} = \text{im}\{\tilde{M}\}^{\perp}$$

y entonces $\text{im}\{M\} \cong \text{im}\{\tilde{M}\}$, que es lo que queríamos demostrar. □

Observación. Si $\tilde{\boldsymbol{q}} = P\boldsymbol{q}$ no es cierto que $\tilde{M} = PM$. Por ejemplo, consideremos el vector $\boldsymbol{q} \coloneqq (1, 1, -2)^T$ y la matriz de permutación

$$P \coloneqq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

de donde obtenemos $\tilde{\boldsymbol{q}} = (1, -2, 1)^T$. Calculamos de (1.38) que

$$M = \begin{pmatrix} 1 & 0 \\ 1 & -2 \\ 1 & -1 \end{pmatrix}, \quad \tilde{M} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad PM = \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 1 & -2 \end{pmatrix}.$$

Se verifica inmediatamente que $\tilde{M} \neq PM$.

En esta última parte del capítulo, exploramos las consecuencias del corolario 1.2.18. Esto nos llevará de regreso a justificar de forma algebraica o geométrica por qué es posible ignorar las entradas nulas del vector $\boldsymbol{q}$, o tan siquiera por qué podemos concentrarnos en este vector coprimo en vez del vector esencialmente entero $\boldsymbol{p}$ en el problema original (1.1).

Definición 1.2.19. Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector coprimo con entradas distintas de cero, entonces definimos su **órbita** como

$$\text{orb}(\boldsymbol{q}) \coloneqq \{P\boldsymbol{q} : P \in \mathbb{Z}^{n\times n} \text{ es matriz de permutación}\}.$$

Lema 1.2.20. *Sean $\boldsymbol{q}, \tilde{\boldsymbol{q}} \in \mathbb{Z}^n$ vectores con entradas coprimas distintas de cero. Entonces $\tilde{\boldsymbol{q}} \in \text{orb}(\boldsymbol{q})$ si y solo si $\text{orb}(\tilde{\boldsymbol{q}}) = \text{orb}(\boldsymbol{q})$.*

Demostración. ($\Rightarrow$) Supongamos que $\text{orb}(\tilde{\boldsymbol{q}}) = \text{orb}(\boldsymbol{q})$. Debemos mostrar que $\tilde{\boldsymbol{q}} \in \text{orb}(\tilde{\boldsymbol{q}})$, lo cual se cumple puesto que $\tilde{\boldsymbol{q}} = I_n\tilde{\boldsymbol{q}}$ y la matriz identidad $I_n \in \mathbb{Z}^{n\times n}$ es una matriz de permutación.

($\Leftarrow$) Supongamos que $\tilde{\boldsymbol{q}} \in \text{orb}(\boldsymbol{q})$, luego existe una matriz de permutación $P \in \mathbb{Z}^{n\times n}$ tal que $\tilde{\boldsymbol{q}} = P\boldsymbol{q}$. Ahora bien, sea $\hat{\boldsymbol{q}} \in \text{orb}(\tilde{\boldsymbol{q}})$, por lo que existe una matriz de permutación $P' \in \mathbb{Z}^{n\times n}$ tal que se satisface $\hat{\boldsymbol{q}} = P'\tilde{\boldsymbol{q}} = (P'P)\boldsymbol{q}$. Como el producto de matrices de permutación también es una matriz de permutación, se sigue que $\hat{\boldsymbol{q}} \in \text{orb}(\boldsymbol{q})$. Con esto mostramos la contención $\text{orb}(\tilde{\boldsymbol{q}}) \subseteq \text{orb}(\boldsymbol{q})$. La otra contención se muestra de manera análoga, pues $\boldsymbol{q} = P^{-1}\tilde{\boldsymbol{q}}$ y la inversa de una matriz de permutación también es una matriz de permutación. □

Lema 1.2.21. *Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector con entradas coprimas distintas de cero y sea $\tilde{\boldsymbol{q}} \in \text{orb}(\boldsymbol{q})$. Entonces las redes $\tilde{\Lambda}_h$ y Λ_h definidas en* (1.42) *son isomorfas. Similarmente, las redes $\tilde{\Lambda}_p$ y Λ_p son isomorfas.*

Demostración. Se sigue inmediatamente de la definición 1.2.19 junto con el corolario 1.2.18. □

El lema anterior nos permite identificar la descomposición de $\mathbb{Z}^n$ a partir de $\text{orb}(\boldsymbol{q})$ y no solamente de $\boldsymbol{q}$. Ahora bien, supongamos que $\boldsymbol{q} \in \mathbb{Z}^{n+\ell}$ es un vector coprimo con ℓ entradas nulas. Definamos el conjunto ordenado de índices no nulos de $\boldsymbol{q}$ como $\sigma = (i : q_i \neq 0)$, y también definamos la proyección $\pi^{(n+\ell)} \colon \mathbb{Z}^{n+\ell} \to \mathbb{Z}^n$ a partir de $\pi^{(n+\ell)}(\boldsymbol{q})_i \coloneqq q_{\sigma_i}$. Luego, $\pi^{(n+\ell)}(\boldsymbol{q})$ es un vector coprimo con entradas distintas de cero. Por el teorema 1.2.16 (y también por la discusión que le sucede), sabemos que $\pi^{(n+\ell)}(\boldsymbol{q})$ induce la descomposición

$$\mathbb{Z}^n \cong \Lambda_p \oplus \Lambda_h,$$

donde Λ_p y Λ_h están definidas en (1.42). Pero en el lema 1.2.21 vimos que todo vector $\tilde{\boldsymbol{q}} \in \text{orb}(\pi^{(n+\ell)}(\boldsymbol{q}))$ induce una descomposición isomorfa.

Esto justifica el hecho de que podamos ignorar las entradas nulas del vector coprimo $\boldsymbol{q}$, y también motiva el siguiente resultado.

Teorema 1.2.22. *Sean $\boldsymbol{q} \in \mathbb{Z}^n$ y $\tilde{\boldsymbol{q}} \in \mathbb{Z}^m$ vectores coprimos con $n < m$. Si $\operatorname{orb}(\pi^{(n)}(\boldsymbol{q})) = \operatorname{orb}(\pi^{(m)}(\tilde{\boldsymbol{q}}))$, entonces $\boldsymbol{q}$ y $\tilde{\boldsymbol{q}}$ inducen descomposiciones isomorfas de una subred de $\mathbb{Z}^n$.*

Demostración. Supongamos, sin pérdida de generalidad, que $\boldsymbol{q}$ no tiene entradas nulas. Luego, $\pi^{(n)}(\boldsymbol{q}) = \boldsymbol{q}$. Entonces $\tilde{\boldsymbol{q}}$ tiene $m-n$ entradas no nulas, pues de otra forma las dimensiones de $\pi^{(n)}(\boldsymbol{q})$ y $\pi^{(m)}(\tilde{\boldsymbol{q}})$ serían distintas y por lo tanto sus órbitas no serían iguales. Tenemos por hipótesis y del lema 1.2.20 que $\pi^{(m)}(\tilde{\boldsymbol{q}}) \in \operatorname{orb}(\boldsymbol{q})$. Finalmente, del lema 1.2.21 junto con el teorema 1.2.16 obtenemos lo que queríamos demostrar. □

Definición 1.2.23. Sean $\boldsymbol{p} \in \mathbb{R}^n$ y $\tilde{\boldsymbol{p}} \in \mathbb{R}^m$ vectores esencialmente enteros (ver definición 1.2.1). Entonces decimos que $\boldsymbol{p}$ y $\tilde{\boldsymbol{p}}$ son equivalentes si y solo si $\operatorname{orb}(\pi^{(n)}(\boldsymbol{q})) = \operatorname{orb}(\pi^{(m)}(\tilde{\boldsymbol{q}}))$, donde $\boldsymbol{q}$ y $\tilde{\boldsymbol{q}}$ son sus respectivos múltiplos coprimos. En este caso escribimos $\boldsymbol{p} \sim \tilde{\boldsymbol{p}}$.

Puesto que el múltiplo coprimo de un vector esencialmente entero es único (ver la discusión al inicio de la subsección 1.2.1), esta relación está bien definida. Luego, por las propiedades de la igualdad, tenemos que esta relación es una de equivalencia sobre el conjunto de vectores esencialmente enteros, independientemente de su dimensión.

Con esta definición y el teorema 1.2.22 encontramos que si los vectores esencialmente enteros $\boldsymbol{p} \in \mathbb{R}^n$ y $\tilde{\boldsymbol{p}} \in \mathbb{R}^m$ son equivalentes, entonces el programa lineal (1.1) es esencialmente el mismo. En conclusión, encontramos una especie de clasificación de este tipo de programas lineales.

En el siguiente capítulo observaremos que conviene imponer un orden sobre las entradas del vector coprimo $\boldsymbol{q} \in \mathbb{Z}^n$ de manera que se satisfaga $q_{n-1} < 0 < q_n$. Esto se reduce a escoger un elemento en $\operatorname{orb}(\boldsymbol{q})$ que cumpla

con esta característica. Equivalentemente, dado un vector esencialmente entero $\boldsymbol{p} \in \mathbb{R}^{n+\ell}$ con ℓ entradas nulas, veremos que conviene escoger un representante en su clase de equivalencia $[\boldsymbol{p}]$ tal que su múltiplo coprimo asociado $\boldsymbol{q} \in \mathbb{Z}^n$ cumpla $q_{n-1} < 0 < q_n$. Esto ejemplificará concretamente la conveniencia de definir una clase de equivalencia sobre el conjunto de vectores esencialmente enteros.

Capítulo 2

El caso infinito

Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y recordemos de la definición 1.2.1 que tiene un único múltiplo coprimo $\boldsymbol{q} \in \mathbb{Z}^n$. Es decir, existe un único escalar $m \in \mathbb{R}$ que satisface tres cosas: $\boldsymbol{p} = m\boldsymbol{q}$, las entradas $q_1, \ldots, q_n$ son coprimas, y la primera entrada no nula q_i es positiva. Supondremos, sin pérdida de generalidad, que m es positivo.

Retomemos el entero $\eta \in \mathbb{Z}$ del lema 1.2.7 que parametriza la primera capa entera que satisface el presupuesto (1.1b). A causa del teorema 1.2.9, sabemos que si $q_i < 0$ para alguna $i \in \{1, \ldots, n\}$, entonces la η-ésima capa entera contiene un número infinito de puntos factibles. No es difíci ver, a través del lema 1.2.6, que estos puntos factibles también son soluciones óptimas.

Corolario 2.0.1. *Supongamos que $q_i < 0$ para algún $i \in \{1, \ldots, n\}$. Entonces el valor óptimo del programa lineal entero (1.1) es $m\eta$. Además, si m es positivo, tenemos que η es el múltiplo de m más grande que satisface $m\eta \leq u$, donde u es el lado derecho de la restricción presupuestaria* (1.1b).

Demostración. Por hipótesis, una entrada de $\boldsymbol{q}$ es negativa. Del teorema 1.2.9 sabemos que existen una infinidad de soluciones en la η-ésima capa entera, así que sea $\boldsymbol{x}^*$ una de ellas. Por el lema 1.2.6 se sigue que $\boldsymbol{q}^T\boldsymbol{x}^* = \eta$,

pero $\boldsymbol{p} = m\boldsymbol{q}$ por la definición 1.2.1, por lo que obtenemos $\boldsymbol{p}^T\boldsymbol{x}^* = m\boldsymbol{q}^T\boldsymbol{x}^* = m\eta$.

Ahora bien, supongamos m es positivo pero que η no es el múltiplo más grande de m que satisface $m\eta \leq u$. Supongamos que $\xi \in \mathbb{Z}$ satisface $m\xi \leq u$ y también $\eta < \xi$. Por el lema 1.2.7 tenemos $\eta = \lfloor u/m \rfloor$. Luego,

$$m\eta = m\left\lfloor \frac{u}{m} \right\rfloor < m\xi \leq u \implies \left\lfloor \frac{u}{m} \right\rfloor < \xi \leq \frac{u}{m},$$

pero esto contradice las propiedades de la función piso. Finalmente, por hipótesis tenemos que m es positivo y por lo tanto η debe ser el múltiplo más grande de m que satisface $m\eta \leq u$. □

Observación. Para ilustrar la conveniencia de restringir m a que sea positivo, consideremos el caso cuando $m < 0$. Del lema 1.2.7 tenemos que $\eta \coloneqq \lceil u/m \rceil$ parametriza también la primera capa entera que satisface el presupuesto, pues ahora tenemos de la restricción (1.1b) que $\boldsymbol{p}^T\boldsymbol{x} \leq u$ si y solo si $\boldsymbol{q}^T\boldsymbol{x} \geq u/m$. Se sigue cumpliendo que el valor óptimo del problema (1.1) es $m\eta$. Sin embargo, η ahora es el múltiplo más pequeño de m que satisface $m\eta \geq u$.

Puesto que somos capaces de decidir si un escalar u^* es el valor óptimo del problema (1.1), nos preguntamos ahora cómo obtener la solución óptima.

Por hipótesis de este capítulo, el vector coprimo $\boldsymbol{q}$ tiene al menos una entrada negativa. Del teorema 1.2.9, sabemos que las infinitas soluciones de (1.1) se encuentran en la η-ésima capa entera. Luego, del lema 1.2.6 tenemos que estas soluciones satisfacen la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = \eta$.

Supongamos, por el momento, que $\boldsymbol{q}$ no tiene entradas nulas. Podemos permutar las entradas de $\boldsymbol{q}$ sin afectar la el problema (1.1). En efecto, toda solución $\boldsymbol{x} \in \mathbb{Z}^n$ de (1.1) es no negativa y satisface $\boldsymbol{q}^T\boldsymbol{x} = \eta$. Sea $P \in \mathbb{Z}^{n\times n}$ una matriz de permutación. Observemos que $\boldsymbol{x}$ es no negativo si y solo si

$P\boldsymbol{x}$ es no negativo. Además,

$$\eta = \boldsymbol{q}^T\boldsymbol{x} = \boldsymbol{q}^T P^T P\boldsymbol{x} = (P\boldsymbol{q})^T(P\boldsymbol{x}).$$

Por lo que podemos encontrar una solución entera no negativa $\boldsymbol{x}$ de la ecuación lineal diofantina $(P\boldsymbol{q})^T\boldsymbol{x} = \eta$ y entonces $P\boldsymbol{x}$ es solución de (1.1).

En particular podemos suponer, sin pérdida de generalidad, que las entradas de $\boldsymbol{q}$ satisfacen $q_{n-1} < 0 < q_n$. Como $\boldsymbol{q}$ no tiene entradas nulas, de la definición 1.2.1 sabemos que q_1 es positivo y, por suposición de este capítulo, alguna entrada q_j es negativa, así que podemos permutar estas entradas con las de q_n y q_{n-1}, respectivamente.

De la proposición 1.2.11 sabemos que el conjunto de soluciones enteras de la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = \eta$ es $\{\eta\boldsymbol{\nu} + M\boldsymbol{t} : \boldsymbol{t} \in \mathbb{Z}^{n-1}\}$. Así pues, basta encontrar condiciones suficientes en $\boldsymbol{t}$ para asegurar la no negatividad de la solución $\boldsymbol{x} \coloneqq \eta\boldsymbol{\nu} + M\boldsymbol{t}$.

Para que las primeras $n-2$ entradas de $\boldsymbol{x}$ sean no negativas, debe ser el caso que $t_i \in \mathbb{Z}$ satisfaga (1.30) para todo $1 \leq i \leq n-2$. Recuperamos de (1.28) que las últimas dos entradas de $\boldsymbol{x}$ son

$$\begin{cases} x_{n-1} = \omega_{n-1}x'_{n-1} + \frac{q_n}{\prod_{j=1}^{n-1} g_j}t_{n-1}, \\ x_n = \omega_{n-1}x'_n - \frac{q_{n-1}}{\prod_{j=1}^{n-1} g_j}t_{n-1}, \end{cases}$$

donde los enteros g_i están definidos por (1.23) con $g_1 = 1$, ω_{n-1} está definida a través de la relación de recurrencia (1.31) con condición inicial $\omega_1 = \eta$, y x'_{n-1}, x'_n son coeficientes de Bézout que satisfacen (1.29). Definamos, por conveniencia,

$$b_1 \coloneqq -\frac{\omega_{n-1}x'_{n-1}}{q_n} \cdot \prod_{j=1}^{n-1} g_j, \quad b_2 \coloneqq \frac{\omega_{n-1}x'_n}{q_{n-1}} \cdot \prod_{j=1}^{n-1} g_j. \tag{2.1}$$

Puesto que $q_{n-1} < 0 < q_n$, podemos despejar t_{n-1} de las dos últimas

soluciones de (1.28) y reproducidas aquí por conveniencia,

$$\begin{cases} x_{n-1} = \omega_{n-1} x'_{n-1} + \frac{q_n}{\prod_{j=1}^{n-1} g_j} t_{n-1}, \\ x_n = \omega_{n-1} x'_n - \frac{q_{n-1}}{\prod_{j=1}^{n-1} g_j} t_{n-1}, \end{cases}$$

encontramos que las últimas dos entradas de $\boldsymbol{x}$ son no negativas si y solo si t_{n-1} satisface

$$t_{n-1} \geq \lceil \max\{b_1, b_2\} \rceil . \tag{2.2}$$

El siguiente lema fortalece nuestros resultados al generalizarlo para todo lado derecho de (1.14).

Lema 2.0.2. *Sea $\boldsymbol{p} \in \mathbb{R}$ un vector esencialmente entero, y supongamos que su múltiplo coprimo $\boldsymbol{q} \in \mathbb{Z}^n$ tiene entradas no nulas y al menos una de ellas es negativa. Entonces la ecuación lineal diofantina* (1.14) *tiene una infinidad de soluciones enteras no negativas.*

Demostración. Supongamos, sin pérdida de generalidad, que $q_{n-1} < 0 < q_n$. Sea $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ tal que

$$t_i \geq \begin{cases} \left\lceil -\frac{\omega_i x'_i}{g_{i+1}} \right\rceil , & i < n-1, \\ \lceil \max\{b_1, b_2\} \rceil , & i = n-1, \end{cases}$$

y sea

$$\boldsymbol{x} \coloneqq k\boldsymbol{\nu} + M\boldsymbol{t},$$

donde recuperamos $\boldsymbol{\nu}$ y M de (1.37) y (1.38), respectivamente. De la proposición 1.2.11, así como de (1.30) y (2.2), se sigue que $\boldsymbol{x}$ es entero y no negativo. Finalmente, de los lemas 1.2.12 y 1.2.13 encontramos que

$$\boldsymbol{q}^T \boldsymbol{x} = k\boldsymbol{q}^T \boldsymbol{\nu} + \boldsymbol{q}^T M = k,$$

por lo que $\boldsymbol{x}$ es solución de (1.14). □

En la práctica es mejor usar la relación de recurrencia (1.25) y "construir" las entradas x_i al mismo tiempo que definimos t_i de manera que satisfaga (1.30) y (2.2). Si procedemos de esta forma no tenemos que encontrar primero $\boldsymbol{\nu}$ y M, determinar $\boldsymbol{t}$ y luego recuperar $\boldsymbol{x}$. El algoritmo 2 en la página 51 muestra este procedimiento constructivo. Para ello, suponemos la existencia de una subrutina `Bezout` que, como su nombre lo indica, calcula los coeficientes de Bézout entre dos enteros. Reiteramos, así como lo hicimos en la sección 1.1.1, que estos coeficientes se pueden calcular por medio del algoritmo extendido de Euclides.

Lema 2.0.3. *El algoritmo 2 en la página 51 es correcto.*

Demostración. Basta observar que el algoritmo sigue la construcción recursiva de la sección 1.2.2, donde escogemos las variables libres t_i como lo indican (1.30) y (2.2) para asegurar que $\boldsymbol{x} \in \mathbb{Z}^n$ sea no negativo. □

Ahora bien, supongamos que $\boldsymbol{q}$ tiene entradas nulas. Sea

$$I^\circ \coloneqq \{i : q_i = 0\},$$

y también definamos el vector $\tilde{\boldsymbol{q}}$ cuyas entradas son las entradas no nulas de $\boldsymbol{q}$. Supongamos que la penúltima entrada de $\tilde{\boldsymbol{q}}$ es negativa y que su última entrada es positiva. A causa del lema anterior, el algoritmo 2 encuentra un vector $\tilde{\boldsymbol{x}}$ entero no negativo que satisface $\tilde{\boldsymbol{q}}^T\tilde{\boldsymbol{x}} = \eta$. Luego, encontramos que el vector $\boldsymbol{x}$ dado por

$$x_i \coloneqq \begin{cases} \tilde{x}_i, & i \notin I^\circ, \\ 0, & i \in I^\circ, \end{cases}$$

es entero, no negativo, y también satisface $\boldsymbol{q}^T\boldsymbol{x} = \eta$.

El algoritmo 3 en la página 52 extiende el algoritmo 2 en el sentido que incorpora esta forma de manejar las entradas nulas de $\boldsymbol{q}$. Además, modifica el vector $\tilde{\boldsymbol{q}}$ de manera que aseguramos que su penúltima entrada es negativa

y su última entrada es positiva, por lo que podemos deshacernos de este supuesto. Suponemos la existencia de las subrutinas `length` y `swap` (véase capítulos 1 y 2 de [Knu97]), las cuales determinan la dimensión de un vector $\boldsymbol{q}$ y permutan dos de sus entradas, respectivamente.

Teorema 2.0.4. *El algoritmo 3 en la página 52 es correcto.*

Demostración. Primero mostramos que el vector $\tilde{\boldsymbol{q}}$ satisface las hipótesis del algoritmo 2. Por definición, en la línea 3, tenemos que ninguna entrada de $\tilde{\boldsymbol{q}}$ es nula.

Recordemos de la definición 1.2.1 que, como $\boldsymbol{q}$ es el múltiplo coprimo de un vector esencialmente entero $\boldsymbol{p}$, su primera entrada no nula es positiva. Así, es cierto que $\tilde{q}_1 > 0$. A partir de la transposición en la línea 5 encontramos que $\tilde{q}_m > 0$.

Del ciclo que comienza en la línea 6 recuperamos el primer índice j tal que $\tilde{q}_j < 0$ y lo transponemos con la $(m-1)$-ésima entrada de $\tilde{\boldsymbol{q}}$ en la línea 10, de manera que obtenemos $\tilde{q}_{m-1} < 0$.

Con los tres puntos anteriores, encontramos que el vector $\tilde{\boldsymbol{q}}$ satisface las hipótesis del algoritmo 2 y por lo tanto el vector $\tilde{\boldsymbol{x}}$ es no negativo y satisface la ecuación lineal diofantina $\tilde{\boldsymbol{q}}^T\tilde{\boldsymbol{x}} = \eta$.

Las siguientes dos líneas se encargan de invertir las transposiciones hechas previamente. Finalmente, en el ciclo que comienza en la línea 14 insertamos en $\boldsymbol{x}$ los valores de $\tilde{\boldsymbol{x}}$ donde las entradas de $\boldsymbol{q}$ son no nulas. En otro caso, donde las entradas de $\boldsymbol{q}$ son nulas, podemos escoger cualquier no negativo para las entradas correspondientes de $\boldsymbol{x}$, entonces dejamos que estas entradas sean cero. Así pues, el vector $\boldsymbol{x} \in \mathbb{Z}^n$ es no negativo y también satisface

$$\boldsymbol{q}^T\boldsymbol{x} = \sum_{i=1}^{n} q_i x_i = \sum_{i=1}^{m} \tilde{q}_i \tilde{x}_i = \eta,$$

por lo que concluimos que el algoritmo 3 es correcto. □

Algoritmo 2: `NonNegativeIntSolInf`

Datos:

$\boldsymbol{q} \in \mathbb{Z}^n$ con entradas coprimas no nulas y tal que $q_{n-1} < 0 < q_n$.

$\eta \in \mathbb{Z}_{\geq 0}$.

Resultado:

$\boldsymbol{x} \in \mathbb{Z}^n_{\geq \mathbf{0}}$ tal que $\boldsymbol{q}^T\boldsymbol{x} = \eta$.

inicio 1

- $\boldsymbol{x} \leftarrow \mathbf{0}$ 2
- $\omega_1 \leftarrow \eta$ 3
- $p \leftarrow 1$ `// producto acumulado` $g_1 g_2 \cdots g_{i+1}$ 4
- **para** $i \leftarrow 1$ **a** $n-2$ **hacer** 5
  - $g_{i+1} \leftarrow \mathrm{mcd}\{q_{i+1}/p, \ldots, q_n/p\}$ `// ecuación` (1.23) 6
  - $x'_i, \omega'_{i+1} \leftarrow$ `Bezout(`q_i/p, g_{i+1}`)` 7
  - $t_i \leftarrow \lceil -\omega_i x'_i / g_{i+1} \rceil$ `// ecuación` (1.30) 8
  - `// ecuación` (1.25) 9
  - $x_i \leftarrow \omega_i x'_i + g_{i+1} t_i$ 10
  - $\omega_{i+1} \leftarrow \omega_i \omega'_{i+1} - q_i t_i / p$ 11
  - `// actualizar producto` 12
  - $p \leftarrow p g_{i+1}$ 13
- $x'_{n-1}, x'_n \leftarrow$ `Bezout(`q_{n-1}/p, q_n/p`)` 14
- `// ecuación` (2.1) 15
- $b_1 \leftarrow -\omega_{n-1} x'_{n-1} \cdot p / q_n$ 16
- $b_2 \leftarrow \omega_{n-1} x'_n \cdot p / q_{n-1}$ 17
- $t_{n-1} \leftarrow \lceil \mathrm{máx}\{b_1, b_2\} \rceil$ `// ecuación` (2.2) 18
- `// ecuación` (1.28) 19
- $x_{n-1} \leftarrow \omega_{n-1} x'_{n-1} + q_n t_{n-1} / p$ 20
- $x_n \leftarrow \omega_{n-1} x'_n - q_{n-1} t_{n-1} / p$ 21
- **devolver** $\boldsymbol{x}$ 22

Algoritmo 3: `Dioph`

Datos:

$\boldsymbol{q} \in \mathbb{Z}^n$ coprimo con al menos una entrada negativa.

$\eta \in \mathbb{Z}_{\geq 0}$.

Resultado:

$\boldsymbol{x} \in \mathbb{Z}^n_{\geq \mathbf{0}}$ tal que $\boldsymbol{q}^T\boldsymbol{x} = \eta$.

$\boldsymbol{x} \leftarrow \mathbf{0}$ 1

$\boldsymbol{\sigma} \leftarrow (i \colon q_i \neq 0)$ 2

$\tilde{\boldsymbol{q}} \leftarrow (q_i \colon q_i \neq 0)$ 3

$m \leftarrow$ `length(`$\tilde{\boldsymbol{q}}$`)` 4

`swap(`$\tilde{\boldsymbol{q}}$`,` 1`,` m`)` `// ` $\tilde{\boldsymbol{q}}_m > 0$ 5

para $i \leftarrow 1$ **a** $m-1$ **hacer** 6

  si $\tilde{q}_i < 0$ **entonces** 7

    $j \leftarrow i$ 8

    ir a la línea 10 9

`swap(`$\tilde{\boldsymbol{q}}$`,` j`,` $m-1$`)` `// ` $\tilde{\boldsymbol{q}}_{m-1} < 0$ 10

$\tilde{\boldsymbol{x}} \leftarrow$ `NonNegativeIntSolInf(`$\tilde{\boldsymbol{q}}$`,` η`)` `// algoritmo 2` 11

`swap(`$\tilde{\boldsymbol{x}}$`,` j`,` $m-1$`)` 12

`swap(`$\tilde{\boldsymbol{x}}$`,` 1`,` m`)` 13

para $i \leftarrow 1$ **a** m **hacer** 14

  $x_{\sigma_i} \leftarrow \tilde{x}_i$ 15

devolver $\boldsymbol{x}$ 16

La correctud de los dos algoritmos anteriores nos permite mostrar el siguiente teorema, el cual postula que existe una "reducción" del problema (1.1) al problema de resolver una sola ecuación lineal diofantina. Cuando determinemos en la siguiente sección la complejidad de ambos algoritmos, observaremos que el problema de resolver esta única ecuación lineal diofantina se puede resolver en tiempo polinomial para un vector coprimo $\boldsymbol{q}$ fijo. En caso de que esta reducción sea también polinomial, el problema (1.1) se puede resolver en tiempo polinomial con respecto a la dimensión del vector objetivo. A fin de evitar introducir formalmente estos conceptos y arriesgar desviarnos por una tangente, no mostramos que esta reducción es, en efecto, polinomial. Todo esto se cumple bajo el supuesto adicional de que el vector coprimo $\boldsymbol{q}$ contenga al menos una entrada negativa.

Teorema 2.0.5. *Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero tal que su múltiplo coprimo $\boldsymbol{q} \in \mathbb{Z}^n$ tiene al menos una entrada negativa. Entonces el problema* (1.1) *se puede resolver a través de encontrar la solución de una ecuación lineal diofantina en n incógnitas.*

Demostración. Como $\boldsymbol{q}$ es el múltiplo coprimo de $\boldsymbol{p}$, existe un escalar m tal que $\boldsymbol{p} = m\boldsymbol{q}$. Supongamos, sin pérdida de generalidad, que m es positivo. Recuperemos η del lema 1.2.7. Por hipótesis, una entrada de $\boldsymbol{q}$ es negativa, y entonces este vector satisface las condiciones del algoritmo 3. Por el teorema 2.0.4 podemos encontrar, a partir de resolver solo una ecuación lineal diofantina, un vector entero no negativo $\boldsymbol{x}$ que satisface $\boldsymbol{q}^T\boldsymbol{x} = \eta$. Observemos que

$$\boldsymbol{p}^T\boldsymbol{x} = m\boldsymbol{q}^T\boldsymbol{x} = m\eta.$$

Por el corolario 2.0.1 concluimos que $\boldsymbol{x}$ no solo es factible para el problema (1.1), sino que también es un punto óptimo. □

2.1. Experimentos numéricos

En primer lugar, determinamos la complejidad algorítmica del algoritmo 3 y lo usamos como base de nuestras comparaciones. En segundo lugar, mostramos que la complejidad del algoritmo 1 en la página 17 depende del número de decimales utilizados para especificar las entradas del vector objetivo, así como de la magnitud en el lado derecho de las restricciones. En tercer lugar, diseñamos un experimento numérico que ilustre esta dependencia.

2.1.1. Análisis de la complejidad de nuestros algoritmos

En ambos algoritmos 2 y 3 dados en las páginas 51 y 52, respectivamente, suponemos que el costo de realizar operaciones aritméticas es constante.

Primero analizamos el algoritmo 2. En el ciclo que comienza en la línea 5 realizamos $\mathcal{O}(n)$ multiplicaciones y $\mathcal{O}(n^2)$ divisiones. Luego, para $1 \leq i \leq n-2$, se sigue de la línea 6 que calculamos una vez el máximo común divisor de $n-(i+1)+1 = n-i$ números. Vimos en la subsección 1.1.1 que el máximo común divisor puede ser definido de manera inductiva. Entonces calcular el máximo común divisor entre $n-i$ números es equivalente a calcular $n-i$ veces el máximo común divisor entre dos números. Por lo tanto, calculamos

$$\sum_{i=1}^{n-2} n - i = n(n-2) - \frac{(n-2)(n-1)}{2} = \frac{(n-2)(n+1)}{2} \tag{2.3}$$

veces el máximo común divisor entre dos números.

Además, tenemos de las líneas 7 y 14 que realizamos $n-1$ llamadas a la subrutina `Bezout`. Pero habíamos mencionado que los coeficientes de Bézout se pueden calcular usando el algoritmo extendido de Euclides. Sin embargo, la complejidad de este es un múltiplo constante de la complejidad del algoritmo de Euclides, el cual calcula el máximo común divisor entre dos números. Así pues, comparando estas aproximadamente n llamadas a

`Bezout` con el cálculo de aproximadamente n^2 máximos común divisores entre dos números, podemos ignorar las llamadas a la subrutina `Bezout`.

En la sección 4.5 de [Knu98] se muestra que la complejidad de calcular el máximo común divisor entre dos enteros $a \leq b$ es $\mathcal{O}(\log_2 b)$. De esta manera, tenemos de las $\mathcal{O}(n)$ multiplicaciones y $\mathcal{O}(n^2)$ divisiones, así como de (2.3) que la complejidad del algoritmo 2 es

$$\mathcal{O}(n^2 \log_2 \|\boldsymbol{q}\|_\infty) + \mathcal{O}(n) + \mathcal{O}(n^2) = \mathcal{O}(n^2(1 + \log_2 \|\boldsymbol{q}\|_\infty)).$$

Del algoritmo 3 tenemos que las subrutinas `length` y `swap` tienen complejidad constante. Además, recorremos el vector $\boldsymbol{q}$ en el ciclo 6 una sola vez así como el vector $\tilde{\boldsymbol{x}}$ en el ciclo 14. Entonces, debido a la única llamada del algoritmo 2, este algoritmo también tiene complejidad cuadrática.

Finalmente, mencionamos que el número de operaciones de estos dos algoritmos dependen exclusivamente de la dimensión n y no del lado derecho η de la ecuación $\boldsymbol{q}^T\boldsymbol{x} = \eta$. Es decir, la complejidad de estos algoritmos con respecto a η es constante. Los resultados numéricos en la subsección 2.1.3 confirman esto.

2.1.2. Árbol de Ramificación y Acotamiento con profundidad dependiente del número de decimales

Primero ilustramos a través de un ejemplo que el árbol generado por una implementación pura de R&A correspondiente a instancias de (1.1) tiene profundidad infinita en algunos casos. Luego mostramos que, aún cuando esta profundidad sea finita, el número de decimales utilizados para expresar el vector objetivo de (1.1) afecta exponencialmente en los tiempos de terminación de este método.

Ejemplo 2.1.1. Consideremos el programa lineal entero

$$\begin{aligned} \max_{(x,y)\in\mathbb{Z}^2} \quad & x - y, \\ \text{s.a.} \quad & x - y \leq 0.3, \\ & x, y \geq 0, \end{aligned}$$

Adoptamos la siguiente convención de notación con respecto a los subproblemas: agregamos un 0 como sufijo al subíndice si la última restricción añadida fue del tipo $\leq$ y, similarmente, agregamos un 1 si la última restricción añadida fue del tipo $\geq$.

En el lado izquierdo de la figura 2.1 (página 57) se muestra la región factible S del problema relajado asociado. Encontramos que su solución es $(0.3, 0)$. Puesto que 0.3 no es entero, ramificamos sobre esta variable y generamos los subproblemas con regiones factibles

$$\begin{aligned} S_0 &:= S \cap \{(x, y) \in \mathbb{R}^2 : x \leq 0\}, \\ S_1 &:= S \cap \{(x, y) \in \mathbb{R}^2 : x \geq 1\}. \end{aligned}$$

La solución del subproblema con región factible S_0 es $(0, 0)$, así que obtenemos una solución incumbente y podamos por cota. Ahora bien, la solución del subproblema con región factible S_1 es $(1, 0.7)$ y como 0.7 no es entero, generamos los subproblemas con regiones factibles

$$\begin{aligned} S_{10} &:= S_1 \cap \{(x, y) \in \mathbb{R}^2 : y \leq 0\}, \\ S_{11} &:= S_1 \cap \{(x, y) \in \mathbb{R}^2 : y \geq 1\}. \end{aligned}$$

Encontramos que el poliedro S_{10} es vacío, así que no ramificamos.

En el lado derecho de la figura 2.1 se muestra el poliedro S_{11} es una traslación de S. En otras palabras, S_{11} es un caso particular de una homotecia de S. Consecuentemente, hay un comportamiento periódico en cuanto a la

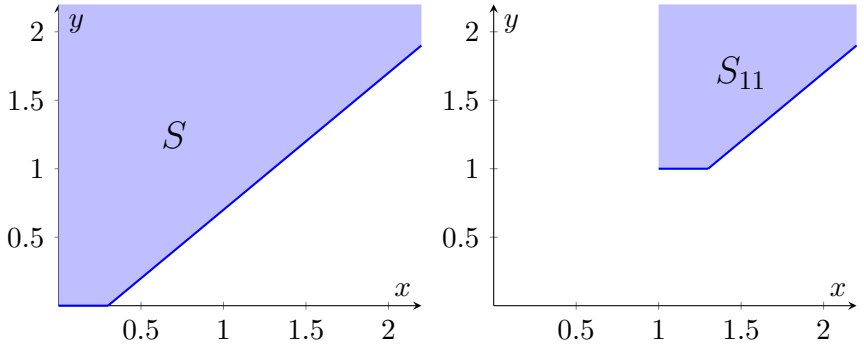

Figura 2.1: *Izquierda*: Poliedro asociado al problema relajado del ejemplo 2.1.1. *Derecha*: Poliedro asociado a un subproblema de S.

ramificación y solución de estos problemas: resolver el subproblema con región factible S_{11} se reduce a resolver el que tiene como región factible a S, pero el primero es un subproblema del segundo. Siguiendo este razonamiento, encontramos que R&A genera una cadena de subproblemas con regiones factibles autosimilares

$$S, S_{11}, S_{1111}, S_{111111}, \ldots$$

y el algoritmo 1 jamás terminará con una solución.

Observemos que si el lado derecho hubiera sido 0 en lugar de 0.3, una implementación pura de R&A habría terminado inmediatamente con la solución $(0, 0)$ al resolver el problema relajado. Es decir, un cambio de 0.3 unidades en el lado derecho de la única desigualdad provocó que el árbol generado por R&A tuviera profundidad unitaria a profundida infinita. Así pues, este ejemplo también muestra que implementaciones puras de R&A dependen de la magnitud del lado derecho de las restricciones.

Sospechamos que el fenómeno presentado en el ejemplo anterior ocurre en todas las instancias de (1.1) cuando $\boldsymbol{p}$ es esencialmente entero y el lado derecho u de (1.1b) es tal que $u/m \notin \mathbb{Z}$, donde m es el escalar que satisface

$\boldsymbol{p} = m\boldsymbol{q}$, y $\boldsymbol{q}$ es el múltiplo coprimo de $\boldsymbol{p}$. Es decir, conjeturamos que, para problemas que cumplen con estas condiciones, existe una colección de subproblemas homotéticos entre sí. Además, si el múltiplo coprimo $\boldsymbol{q}$ de $\boldsymbol{p}$ tiene una entrada negativa, entonces esta colección es infinita. Por cuestiones de longitud de esta tesis de licenciatura y por el hecho de que es un tema de carácter abierto, no investigamos más en este respecto.

Para que una implementación pura de R&A termine en tiempo finito al resolver una de estas instancias, debe ser el caso que el valor óptimo del problema relajado de (1.1) sea el mismo que el de su programa entero. Es decir, tenemos terminación finita si y solo si el hiperplano definido por $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{p}^T\boldsymbol{x} = u\}$ contiene un punto entero, donde u es el lado derecho de (1.1b). Por el teorema 1.2.5, hay terminación finita si y solo si este hiperplano en realidad es una capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ con parámetro $k \in \mathbb{Z}$.

Supongamos que tenemos terminación finita y que estamos utilizando una implementación pura de R&A para resolver una instancia. Por lo discutido anteriormente, todas las soluciones de los subproblemas se encuentran en la k-ésima capa entera. El algoritmo de R&A genera cortes del estilo

$$x_i \leq \lfloor x_i^* \rfloor, \quad x_i \geq \lceil x_i^* \rceil, \tag{2.4}$$

donde $\boldsymbol{x}^*$ es la solución de un subproblema relajado con $x_i^* \notin \mathbb{Z}$. Sin pérdida de generalidad, supongamos que añadimos a este subproblema relajado un corte como el de la derecha de (2.4) y que se mantiene la factibilidad. El punto de intersección entre esta restricción junto con la restricción presupuestaria (1.1b) introduce un único vértice nuevo para el siguiente subproblema relajado. Tenemos entonces del teorema 1.1.16 que este único vértice es una solución. Los algoritmos más comúnes para resolver problemas relajados siempre encuentran como solución un vértice del poliedro, por lo que podemos suponer razonablemente que R&A encontrará este vértice como solución del siguiente subproblema relajado. Es decir, si $\boldsymbol{x}^*$ es solución de

un subproblema relajado, entonces $\lceil x_i^* \rceil$ constituye la i-ésima entrada de la solución del siguiente subproblema relajado obtenido al añadir el corte $x_i \geq \lceil x_i^* \rceil$. En particular, encontramos que

$$0 < \left| x_i^* - \lceil x_i^* \rceil \right| \leq 1. \tag{2.5}$$

Por el teorema 1.2.16, existe $(k^*, \boldsymbol{t}^*) \in \mathbb{R}^n$ tal que $\boldsymbol{x}^* = k^*\boldsymbol{\nu} + M\boldsymbol{t}^*$, donde la matriz M y el vector $\boldsymbol{\nu}$ están definidos, respectivamente, en (1.37) y (1.38). Habíamos supuesto que el algoritmo de R&A tiene terminación finita, y entonces $\boldsymbol{x}^*$ se encuentra en la k-ésima capa entera, con parámetro $k \in \mathbb{Z}$. Se sigue de los lemas 1.2.6, 1.2.12 y 1.2.13 que

$$k = \boldsymbol{q}^T\boldsymbol{x}^* = k^*\boldsymbol{q}^T\boldsymbol{\nu} + \left(\boldsymbol{q}^T M\right)\boldsymbol{t}^* = k^*,$$

y entonces $\boldsymbol{x}^* = k\boldsymbol{\nu} + M\boldsymbol{t}^*$. Sea $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ el vector resultante de redondear las entradas de $\boldsymbol{t}^*$ a sus enteros más cercanos. No es difícil ver que $\boldsymbol{x} \coloneqq k\boldsymbol{\nu} + M\boldsymbol{t}$ es un punto entero sobre la k-ésima capa entera más cercano a $\boldsymbol{x}^*$. También es cierto que $\boldsymbol{t} = \boldsymbol{t}^* + \boldsymbol{\delta}$ para alguna $\boldsymbol{\delta} \in \mathbb{R}^{n-1}$ que satisface $0 < \|\boldsymbol{\delta}\|_\infty \leq 1/2$. Luego,

$$\left| x_i^* - x_i \right| = \left| \left(\boldsymbol{x}^* - \boldsymbol{x}\right)_i \right| = \left| \boldsymbol{e}_i^T M \left(\boldsymbol{t}^* - \boldsymbol{t}\right) \right| = \left| \boldsymbol{e}_i^T M\boldsymbol{\delta} \right|. \tag{2.6}$$

Dividiendo (2.6) entre (2.5) obtenemos

$$\left| \frac{x_i^* - x_i}{x_i^* - \lceil x_i^* \rceil} \right| \geq \left| \boldsymbol{e}_i^T M\boldsymbol{\delta} \right|. \tag{2.7}$$

Entonces el número de decimales utilizados para expresar las entradas del vector objetivo $\boldsymbol{p}$ del problema (1.1) afecta la eficiencia de estos cortes. En efecto, si usamos un decimal más para especificar las entradas de $\boldsymbol{p}$, su vector coprimo $\boldsymbol{q}$ se multiplica aproximadamente por 10. De (1.38) tenemos que las entradas de M se multiplican aproximadamente por 10 y, por lo tanto, el lado derecho de (2.7) es 10 veces mayor. Es decir, los cortes de

R&A se vuelven aproximadamente 10 más ineficientes al utilizar un decimal adicional. En conclusión, la complejidad de Ramificación y Acotamiento es exponencial en el número de decimales utilizados para expresar $\boldsymbol{p}$, pues R&A requiere 10 veces más cortes por decimal adicional. No obstante, es cierto que los cortes se vuelven más efectivos una vez que $\boldsymbol{x}^*$ se encuentra cerca de una solución entera, pues $\|\boldsymbol{\delta}\|_\infty \approx 0$.

Como mencionamos en la subsección 1.1.2, R&A encapsula una familia de métodos para encontrar soluciones de programas lineales enteros. Los resultados hasta ahora obtenidos solamente aplican para el algoritmo 1, donde contamos con una política de poda y ramificamos binariamente. Si en el ejemplo 2.1.1 hubiéramos deducido una cota superior más justa que $x - y \leq 0.3$, R&A habría encontrado inmediatamente la solución. En efecto, como $(x, y) \in \mathbb{Z}^2$, se sigue que $x - y \in \mathbb{Z}$ y puesto que $x - y \leq 0.3$, obtenemos la cota superior $x - y \leq 0$. De esta manera, la implementación pura de R&A falla en este ejemplo particular porque no genera cortes de Gomory. Redirigimos al lector interesado en esta y otras políticas de poda a la sección 5 del artículo [MJSS16].

A pesar de que implementaciones comerciales y de código abierto utilicen otro orden de búsqueda, ramifiquen de distinta manera, extiendan las políticas de poda, introduzcan heurísticas, preprocesen el problema, etcétera, veremos en la siguiente subsección que la obtención de soluciones por parte del *solver* CBC ([Lou03]) todavía es lenta si comparamos sus tiempos de terminación con nuestro algoritmo 3, o incluso con la cota inferior (2.7).

2.1.3. Análisis de resultados

Primero detallamos la manera en la que medimos los resultados de nuestro experimento numérico. Luego explicamos y justificamos el diseño del experimento. Finalmente mostramos y analizamos los resultados obtenidos.

Para realizar nuestro experimento numérico con R&A, usamos el *solver*

CBC a través de la interfaz de PuLP versión 3.1 implementada en Python versión 3.12. Nuestra configuración permite métodos de preprocesamiento, cortes de mochila y de Gomory-Chvátal. Prohibimos el uso de múltiples hilos (*threads*) a fin de garantizar una comparación justa con nuestro método.

Cada observación de tiempo representa el promedio de 20 corridas. Para cada observación realizamos 2 corridas preliminares a fin de evitar sesgos en el tiempo por cuestión de temperatura en la computadora, o por cargar cosas a la memoria, o por compilación al momento de crear archivos `.pyc`, etcétera. En total, cada observación es el resultado de haber corrido el mismo experimento 22 veces.

Por cuestiones de tiempo, cada corrida fue automáticamente terminada si sobrepasaba los 300 segundos. Decimos que el método no encontró una solución si fue automáticamente terminado en más de 10 de las 20 veces que medimos el tiempo de terminación. Para verificar la correctud en la implementación de nuestro algoritmo, comparamos los valores objetivo de nuestra implementación con los de CBC y en ningún momento encontramos que fueran distintos. Nuestra implementación se encuentra libre en GitHub.[1]

Finalmente, los experimentos fueron realizados en una computadora portátil Dell XPS 15 equipada con un procesador Intel Core i7-8750H (6 núcleos físicos y 12 hilos, frecuencia base de 2.20 GHz y frecuencia máxima de 4.10 GHz). El sistema cuenta con 12 CPU lógicos disponibles y una memoria RAM de 32 GiB. Todos los cálculos fueron ejecutados bajo la arquitectura x86-64. El sistema operativo usado fue Fedora Linux 42 (Server Edition).

Inspirados por el hecho de que la implementación pura de R&A es sensible ante la precisión numérica, generamos instancias de (1.1) con dimensión $n = 4$. Este experimento tiene por objetivo mostrar que los tiempos de terminación de CBC son lentos y aumentan significativamente cuando utilizamos más cifras decimales para especificar el vector objetivo $\boldsymbol{p}$ de (1.1).

[1] Véase `https://github.com/tempdata73/thesis`

Así pues, generamos un vector $\boldsymbol{p}_3$ tomado de una distribución uniforme discreta de enteros sobre el intervalo $[9{,}000, 10{,}000)^n$. A este vector inicial lo dividimos por 1,000, de manera que sea un vector racional con tres cifras decimales y entradas en $[9, 10)$. Escogimos aleatoriamente 2 de estas $n = 4$ entradas y las multiplicamos por -1, con lo que aseguramos que el vector coprimo $\boldsymbol{q}_3$ asociado a $\boldsymbol{p}_3$ tenga al menos una entrada negativa.

Luego, muestreamos $n = 4$ observaciones de una distribución uniforme discreta de enteros sobre $[0, 10)$, las multiplicamos por 10^{-4} y sumamos cada observación a una entrada de $\boldsymbol{p}_3$, de manera que obtenemos un vector racional $\boldsymbol{p}_4$ con cuatro cifras decimales. Repetimos una vez más el procedimiento anterior pero usando $\boldsymbol{p}_4$ para generar el vector racional $\boldsymbol{p}_5$ con cinco cifras decimales.

En último lugar, para obtener el lado derecho de (1.1b), calculamos 128 puntos $u \in \mathbb{Z}_{\geq 0}$ espaciados aproximadamente logarítmicamente en el intervalo $[10^3, 10^7)$. En la figura 2.2 se muestran los tiempos de terminación promedio de las 20 corridas. Las rupturas de continuidad en las líneas son causadas porque CBC no encontró una solución antes de los 300 segundos.

Por los resultados obtenidos en las dos subsecciones anteriores, no sorprende que el algoritmo 2 tenga mejores tiempos de terminación que CBC, y que no se vea afectado por la precisión numérica del vector objetivo $\boldsymbol{p}$. Esto solo muestra que los tiempos de terminación de CBC son lentos para instancias de (1.1).

En realidad nos interesa mostrar la sensibilidad en los tiempos de terminación de CBC a medida que la precisión del vector objetivo $\boldsymbol{p}$ aumenta. Es decir, si $T(\boldsymbol{p}_i)$ es el tiempo de terminación de CBC con una instancia usando el vector esencialmente entero $\boldsymbol{p}_i$ que tiene i cifras decimales, entonces queremos medir la proporción $T(\boldsymbol{p}_{i+1})/T(\boldsymbol{p}_i)$. Llamamos **multiplicadores** a estas proporciones, pues representan el factor de aumento en los tiempos de terminación de CBC cuando utilizamos una cifra decimal adicional. Puesto

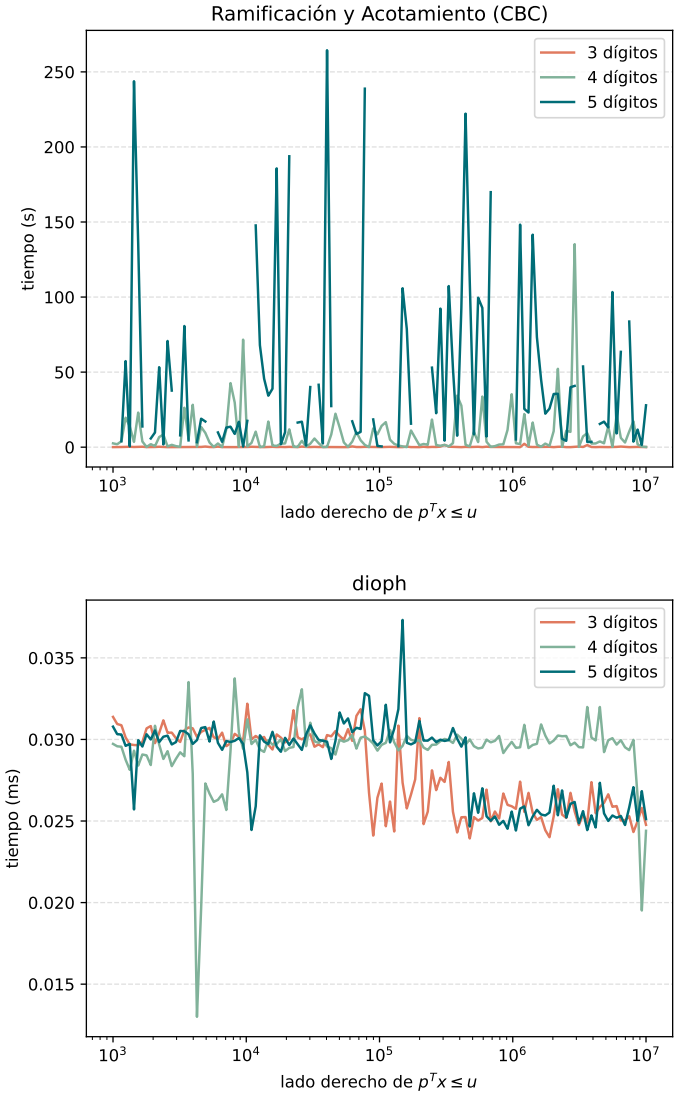

Figura 2.2: Tiempos de terminación promedio cuando varía el presupuesto y el número de digitos. *Arriba*: Resultados en segundos de R&A implementado en CBC. *Abajo*: Resultados en milisegundos de nuestro método `dioph`.

que cada instancia depende del presupuesto u de la restricción (1.1b) y tomamos 128 de estos presupuestos por cada nivel de precisión i, obtenemos a lo más $128 \times (3 - 1) = 256$ multiplicadores registrados, pues no podemos medir aquellos en donde $T(\boldsymbol{p}_i)$ es mayor a 300 segundos. La figura 2.3 en la página 65 muestra la distribución acumulada de estos multiplicadores. Por cuestiones visuales, juntamos todos los multiplicadores que fueron mayores a 100 en la última columna.

De acuerdo a la subsección anterior, deberíamos esperar que la gran mayoría de los multiplicadores se concentren alrededor de 10. Esto último ocurrió en el 40 % de los casos. Después de realizar un conteo en los archivos de registro generados por CBC, descubrimos que en todos los casos que se encontró una solución fue a causa de una heurística de redondeo y no por el método de ramificación. Debido la dificultad del árbol que genera el problema (1.1), de acuerdo a la subsección anterior, esto parece ser razonable.

Aún cuando las soluciones fueron obtenidas exclusivamente a través de heurísticas, en el 60 % de los casos el multiplicador es mayor a 10. Es decir, una cifra decimal adicional en $\boldsymbol{p}$ provoca, en la mayoría de los casos, que el tiempo de terminación de CBC sea al menos diez veces mayor. Así pues, los cortes generados por R&A son más ineficientes de lo que habíamos predicho.

Esperamos que lo realizado en esta sección muestre categóricamente que implementaciones puras (algoritmo 1) o prácticas (CBC) de Ramificación y Acotamiento no dependen solo del número de variables o de desigualdades utilizadas, sino que también de la precisión numérica con la que se especifican los datos. Con esto damos por concluido este capítulo y continuamos con nuestro desarrollo del segundo caso del teorema 1.2.9.

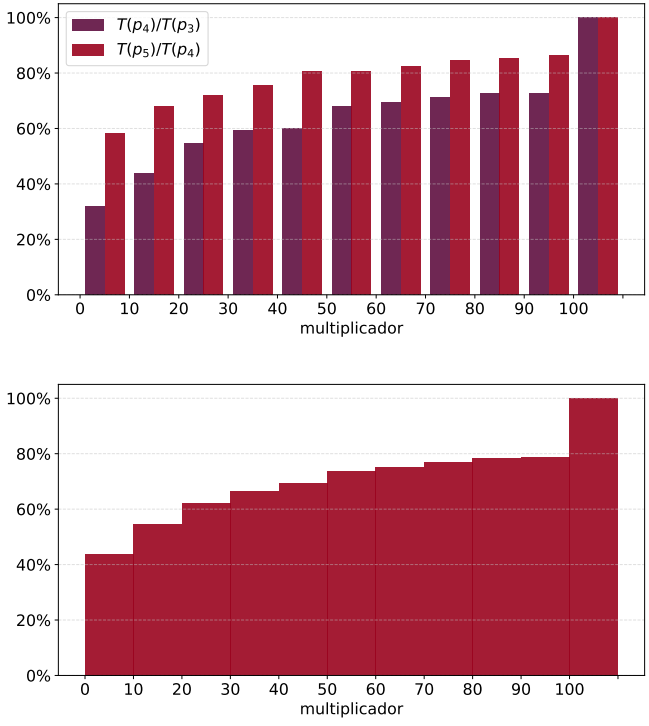

Figura 2.3: Distribución acumulada de los multiplicadores de tiempo. *Arriba:* Separados por el número de cifras decimales utilizadas. *Abajo:* Distribución conjunta sin tomar en cuenta el número de cifras decimales utilizadas.

Capítulo 3

El caso finito

Inspirados por el teorema 1.2.9, en este capítulo analizamos el caso en el que el vector coprimo $\boldsymbol{q}$ tiene entradas estrictamente positivas. Bajo este supuesto adicional, el problema (1.1) es una instancia particular del famoso *Problema de la Mochila*:

$$\max_{\boldsymbol{x}\in\mathbb{Z}^n} \quad \boldsymbol{s}^T\boldsymbol{x}, \tag{3.1a}$$

$$\text{s.a.} \quad \boldsymbol{w}^T\boldsymbol{x} \leq u, \tag{3.1b}$$

$$\boldsymbol{x} \geq \boldsymbol{0}, \tag{3.1c}$$

donde los vectores positivos $\boldsymbol{s}, \boldsymbol{w} \in \mathbb{Z}^n$ son conocidos como vector de útiles y vector de pesos, respectivamente. Puesto que no acotamos $\boldsymbol{x}$, el problema recibe el nombre de *Problema de la Mochila no Acotado*. Analizaremos el caso $\boldsymbol{p} = \boldsymbol{s} = \boldsymbol{w}$, el cual es es considerado como un *Problema de la Suma de Conjuntos no Acotado*. Referimos a los capítulos 3 y 4 de [MT90] para una descripción más detallada de estos problemas.

En la primera sección realizamos un análisis de capas enteras a fin de obtener un resultado análogo al teorema 2.0.5. En concreto, el teorema 3.1.19 enuncia que, para un presupuesto u suficientemente grande en el problema (1.1), la búsqueda de una solución se reduce a resolver solamente una ecuación lineal diofantina.

El resultado anterior, si bien es interesante, solo es de existencia y no muestra cómo obtener las soluciones enteras no negativas de ecuaciones lineales diofantinas. De manera similar a como lo hicimos en el capítulo anterior, la segunda sección se encarga de presentar la construcción de soluciones a partir de los algoritmos 4 y 5, presentados en las páginas 93 y 94, respectivamente.

Finalmente, en la tercera y última sección de este capítulo, realizamos algunos experimentos numéricos que comparan la eficacia de nuestros algoritmos recién desarrollados con la de Ramificación y Acotamiento, así como de una formulación de programación dinámica a fin de contar con otro punto de comparación y tener conclusiones más robustas.

3.1. Análisis de capas enteras

De acuerdo al segundo caso del teorema 1.2.9, el número de puntos enteros no negativos sobre la k-ésima capa entera es finito y, por lo tanto, puede ser cero. Sea $k \in \{\eta, \eta - 1, \ldots, 0\}$. Sabemos de la sección 1.2.2 que deseamos resolver la ecuación lineal diofantina (1.14), por lo que implementamos la misma estrategia para plantear una formulación recursiva.

Debido al supuesto $\boldsymbol{q} > \mathbf{0}$, observemos de (1.20):

$$\frac{q_i}{\prod_{j=1}^{i} g_j} x_i + \frac{q_{i+1}}{\prod_{j=1}^{i} g_j} x_{i+1} + \cdots + \frac{q_n}{\prod_{j=1}^{i} g_j} x_n = \omega_i,$$

que podemos agregar la condición $\omega_i \geq 0$, en efecto, buscamos que $\boldsymbol{x}$ sea no negativo y recordemos que g_i es un máximo común divisor (ver (1.21)), por lo que es estrictamente positivo. Juntando esto con el supuesto $\boldsymbol{q} > \mathbf{0}$, encontramos que ω_i es no negativo para toda $i \in \{1, \ldots, n-1\}$. Así pues,

despejando t_i de (1.25) obtenemos los intervalos de factibilidad

$$\left\lceil -\frac{\omega_i x'_i}{g_{i+1}} \right\rceil \leq t_i \leq \left\lfloor \frac{\omega_i \omega'_{i+1}}{q_i} \prod_{j=1}^{i} g_j \right\rfloor,$$

para todo $i \in \{1, \ldots, n-2\}$. Luego, como $0 < q_{n-1}, q_n$, se sigue de (1.28) que

$$\left\lceil -\frac{\omega_{n-1} x'_{n-1}}{q_n} \cdot \prod_{j=1}^{n-1} g_j \right\rceil \leq t_{n-1} \leq \left\lfloor \frac{\omega_{n-1} x'_n}{q_{n-1}} \cdot \prod_{j=1}^{n-1} g_j \right\rfloor.$$

Consecuentemente, el número de elecciones que podemos realizar para el vector de variables libres $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ es finito, como lo confirma el teorema 1.2.9. Si determinamos que no existe tal punto $\boldsymbol{t}$ en la k-ésima capa entera, descendemos a la $(k-1)$-ésima capa entera y continuamos con nuestra búsqueda.

Ahora bien, en esta primera parte de la sección nos encargamos de calcular una cota superior para el número de capas enteras que debemos analizar de manera que garanticemos la existencia de un punto entero no negativo sobre una de estas capas enteras.

Lema 3.1.1. *Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y sea $\boldsymbol{q} \in \mathbb{Z}^n$ su múltiplo coprimo, por lo que existe $m \in \mathbb{R}$ tal que $\boldsymbol{p} = m\boldsymbol{q}$. Supongamos que $m > 0$ y que $\boldsymbol{q} > \boldsymbol{0}$. Sea $q^* \coloneqq \max\{q_1, \ldots, q_n\}$, y sea*

$$\tau \coloneqq \left\lfloor \left\lfloor \frac{u}{q^*} \right\rfloor \frac{q^*}{m} \right\rfloor, \tag{3.2}$$

donde u es el lado derecho de (1.1b)*. Entonces la solución del problema* (1.1)*, de ser factible, se encuentra en una capa entera parametrizada por $k \in \{\eta, \eta - 1, \ldots, \tau\}$, donde recuperamos η del lema 1.2.7.*

Demostración. Sea $i^* \coloneqq \arg\max\{q_1, \ldots, q_n\}$ y consideremos el vector

$$\boldsymbol{v} \coloneqq \left\lfloor \frac{u}{q^*} \right\rfloor \boldsymbol{e}_{i^*}.$$

Por hipótesis tenemos $q^* > 0$ y, además, como el problema (1.1) es factible, se sigue del teorema 1.2.9 que el presupuesto u es no negativo. De esto obtenemos que $\boldsymbol{v} \geq \boldsymbol{0}$. Así también,

$$\boldsymbol{q}^T\boldsymbol{v} = \left\lfloor \frac{u}{q^*} \right\rfloor q^* \leq \frac{u}{q^*} q^* = u,$$

y entonces $\boldsymbol{v}$ es factible. De aquí se sigue que este vector provee una cota inferior para valor óptimo del problema (1.1). Así pues, todo vector $\boldsymbol{x}$ candidato a ser el óptimo del problema satisface

$$\boldsymbol{q}^T\boldsymbol{x} = \frac{\boldsymbol{p}^T\boldsymbol{x}}{m} \geq \frac{\boldsymbol{q}^T\boldsymbol{v}}{m} = \left\lfloor \frac{u}{q^*} \right\rfloor \frac{q^*}{m}.$$

Nos interesa determinar el entero τ más pequeño tal que todo punto sobre la capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ con $k \in \{\tau, \tau+1, \ldots\}$ satisfaga esta desigualdad. Del lema 1.2.4, encontramos que k debe satisfacer

$$\frac{k}{\|\boldsymbol{q}\|^2} = \frac{\boldsymbol{q}^T\boldsymbol{x}}{\|\boldsymbol{q}\|^2} \geq \left\lfloor \frac{u}{q^*} \right\rfloor \frac{q^*}{m} \frac{1}{\|\boldsymbol{q}\|^2},$$

equivalentemente,

$$k \geq \left\lfloor \frac{u}{q^*} \right\rfloor \frac{q^*}{m}.$$

Consecuentemente,

$$\tau = \left\lfloor \left\lfloor \frac{u}{q^*} \right\rfloor \frac{q^*}{m} \right\rfloor.$$

Finalmente, recordemos del lema 1.2.7 que η es la primera capa en satisfacer la restricción presupuestaria. Por lo tanto, el óptimo del problema (1.1) se encuentra en una capa entera parametrizada por $\tau \leq k \leq \eta$. □

Observación. Siempre se cumple que $\tau \leq \eta$. En efecto,

$$\left\lfloor \frac{u}{q^*} \right\rfloor q^* \leq \frac{u}{q^*} q^* = u,$$

como $m > 0$, tenemos

$$\left\lfloor \frac{u}{q^*} \right\rfloor \frac{q^*}{m} \leq \frac{u}{m}.$$

Aplicando la función piso a ambos lados de la desigualdad y comparando con (3.2) y el lema 1.2.7 encontramos que $\tau \leq \eta$.

Vimos en el primer capítulo que, si $m < 0$, entonces existe un parámetro η' análogo a η. En este caso, también aquí existe $\tau' \geq \eta'$ tal que la solución del problema (1.1) se encuentra en una capa parametrizada por $\eta' \leq k \leq \tau'$.

Lema 3.1.2. *Sean q y m enteros distintos de cero. Entonces la función $\Delta\colon \mathbb{R} \to \mathbb{R}$ dada por*

$$\Delta(x) \coloneqq \left\lfloor \frac{x}{m} \right\rfloor - \left\lfloor \left\lfloor \frac{x}{q} \right\rfloor \frac{q}{m} \right\rfloor,$$

es periódica con periodo $\operatorname{mcm}\{q, m\}$.

Demostración. Tenemos

$$\Delta(x + \operatorname{mcm}\{q,m\}) = \left\lfloor \frac{x}{m} + \frac{\operatorname{mcm}\{q,m\}}{m} \right\rfloor - \left\lfloor \left\lfloor \frac{x}{q} + \frac{\operatorname{mcm}\{q,m\}}{q} \right\rfloor \frac{q}{m} \right\rfloor,$$

pero $q, m \mid \operatorname{mcm}\{q,m\}$, y por las propiedades de la función piso obtenemos:

$$\begin{aligned}
\Delta(x + \operatorname{mcm}\{q,m\}) &= \left\lfloor \frac{x}{m} \right\rfloor + \frac{\operatorname{mcm}\{q,m\}}{m} - \left\lfloor \left\lfloor \frac{x}{q} \right\rfloor \frac{q}{m} + \frac{\operatorname{mcm}\{q,m\}}{q} \cdot \frac{q}{m} \right\rfloor \\
&= \left\lfloor \frac{x}{m} \right\rfloor + \frac{\operatorname{mcm}\{q,m\}}{m} - \left\lfloor \left\lfloor \frac{x}{q} \right\rfloor \frac{q}{m} \right\rfloor - \frac{\operatorname{mcm}\{q,m\}}{m} \\
&= \left\lfloor \frac{x}{m} \right\rfloor - \left\lfloor \left\lfloor \frac{x}{q} \right\rfloor \frac{q}{m} \right\rfloor = \Delta(x).
\end{aligned}$$

□

Definición 3.1.3. Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y sea $\boldsymbol{q} \in \mathbb{Z}^n$ su múltiplo coprimo. Consideremos los parámetros η y τ definidos, respectivamente, en los lemas lemas 1.2.7 y 3.1.1, como funciones del presupuesto u. Entonces decimos que la función $\Delta^* \colon \mathbb{R} \to \mathbb{R}$ dada por

$$\Delta^*(u) \coloneqq \eta(u) - \tau(u) \tag{3.3}$$

denota el **número de capas enteras a revisar** dado el presupuesto u.

Si queremos aplicar el lema 3.1.2 a la función de la definición anterior, debemos reducir nuestra atención a vectores $\boldsymbol{p}$ enteros. Esto se debe a que debemos asegurar que el múltiplo m sea entero. Aunque no lo demostraremos, es nuestra creencia que este resultado puede ser generalizado para m racional. Independientemente del comportamiento periódico de Δ^*, tenemos que esta función varía significativamente ante cambios en m. Esto último implica que el número de capas enteras a revisar depende del número de cifras decimales utilizadas para especificar $\boldsymbol{p}$.

Ejemplo 3.1.4. Si tenemos $\boldsymbol{p} \coloneqq (9.6, 7.2, 5.6)^T$, entonces $m = 0.8$ y por lo tanto el número de capas a revisar dado $u \coloneqq 119$ es $\Delta^*(u) = 13$. En cambio, si tenemos $\boldsymbol{p} \coloneqq (9.60, 7.28, 5.68)^T$, obtenemos $m = 0.08$, por lo que el número de capas a revisar dado u es $\Delta^*(u) = 1487$. Eso ilustra la sensitividad de Δ^* con respecto a m.

En esta segunda parte de la sección, demostraremos que para un presupuesto u suficientemente grande, la solución del problema (1.1) se encuentra en la η-ésima capa entera donde, como siempre, η es recuperada del lema 1.2.7. Este resultado será análogo al teorema 2.0.5. No obstante, para lograr aquello, necesitamos de un par de definiciones y lemas preliminares.

Para motivar al lector, primero mostramos que existe una vecindad fija de todo punto en $\mathbb{R}^n$ de manera que esa vecindad contiene al menos un punto entero. Esto será especificado en el teorema 3.1.6.

Luego, observamos que el "trozo" no negativo de una capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ crece a medida que k aumenta. Recordemos que $k \leq \eta$ y, por el lema 1.2.7, η depende del lado derecho u de (1.1b). Así pues, si el presupuesto u lo permite, habrá un punto sobre ese trozo no negativo cuya vecindad también se encuentra contenida en ese trozo y, por lo tanto, habrá un punto entero no negativo sobre ese trozo. Esto será especificado en el teorema 3.1.17.

Finalmente, relacionamos el punto entero que se encuentra sobre el pedazo no negativo de $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ con el problema (1.1). Así pues, concluiremos esta sección con los teoremas 3.1.18 y 3.1.19.

Definición 3.1.5. Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector coprimo y sea k un entero. Entonces definimos la **bola cerrada** sobre la k-ésima capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ con radio $r > 0$ y centro $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ como

$$B_r^{(k)}(\boldsymbol{x}) \coloneqq \{\boldsymbol{y} \in \mathbb{R}^n \colon \|\boldsymbol{y} - \boldsymbol{x}\| \leq r\} \cap H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}. \tag{3.4}$$

Teorema 3.1.6. *Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector coprimo y supongamos que $q_n \neq 0$. Sea k un entero. Entonces existe $r > 0$ tal que la familia de bolas*

$$\left\{B_r^{(k)}(\boldsymbol{x}) \colon \boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \cap \mathbb{Z}^n\right\}$$

es una cubierta de la k-ésima capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$.

Demostración. Como $q_n \neq 0$, recordemos del teorema (1.2.16) que

$$\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \cap \mathbb{Z}^n \iff \boldsymbol{x} = k\boldsymbol{\nu} + M\boldsymbol{t}$$

para algún vector $\boldsymbol{t} \in \mathbb{Z}^{n-1}$, donde recuperamos $\boldsymbol{\nu}$ y M de (1.37) y (1.38), respectivamente. Así, tenemos

$$\left\{B_r^{(k)}(\boldsymbol{x}) \colon \boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \cap \mathbb{Z}^n\right\} = \left\{B_r^{(k)}(k\boldsymbol{\nu} + M\boldsymbol{t}) \colon \boldsymbol{t} \in \mathbb{Z}^{n-1}\right\}.$$

Por la definición 3.1.5 sabemos que $B_r^{(k)}(k\boldsymbol{\nu} + M\boldsymbol{t}) \subseteq H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ para todo

$r > 0$ y para todo $\boldsymbol{t} \in \mathbb{Z}^{n-1}$. Luego, para cualquier $r > 0$ tenemos

$$\bigcup_{\boldsymbol{t}\in\mathbb{Z}^{n-1}} B_r^{(k)}(k\boldsymbol{\nu} + M\boldsymbol{t}) \subseteq H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}. \tag{3.5}$$

Ahora bien, sea $\boldsymbol{y}$ un punto sobre la k-ésima capa entera. Por el teorema 1.2.16 sabemos que las columnas de M son linealmente independientes, y entonces existe $\tilde{\boldsymbol{t}} \in \mathbb{R}^{n-1}$ tal que

$$\boldsymbol{y} = k\boldsymbol{\nu} + M\tilde{\boldsymbol{t}}.$$

Sea $\boldsymbol{t} \in \mathbb{Z}^{n-1}$ el vector resultante de redondear cada entrada de $\boldsymbol{t}$ al entero más cercano. Luego, $\tilde{\boldsymbol{t}} = \boldsymbol{t} + \boldsymbol{\delta}$, para alguna $\boldsymbol{\delta} \in \mathbb{R}^{n-1}$ que satisface $\|\boldsymbol{\delta}\|_\infty \leq 1/2$. Definamos

$$\boldsymbol{x} \coloneqq k\boldsymbol{\nu} + M\boldsymbol{t} \in \mathbb{Z}^{n-1},$$

de donde se sigue que

$$\|\boldsymbol{y} - \boldsymbol{x}\| = \|M\boldsymbol{\delta}\| \leq \|M\| \, \|\boldsymbol{\delta}\| \leq \sqrt{n-1} \, \|M\| \, \|\boldsymbol{\delta}\|_\infty \leq \frac{\sqrt{n-1}}{2} \, \|M\| \, .$$

Por lo tanto, si definimos

$$r \coloneqq \frac{\sqrt{n-1}}{2} \, \|M\| \, , \tag{3.6}$$

encontramos que $\boldsymbol{y} \in B_r^{(k)}(\boldsymbol{x})$. Luego, como $\boldsymbol{y} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ fue genérico, tenemos que si r está definido por (3.6), entonces

$$H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \subseteq \bigcup_{\boldsymbol{t}\in\mathbb{Z}^{n-1}} B_r^{(k)}(k\boldsymbol{\nu} + M\boldsymbol{t}). \tag{3.7}$$

Juntando esto con (3.5) obtenemos lo que queríamos demostrar. □

Lo que se encuentra a continuación se encarga de formalizar y caracterizar lo que nos referíamos anteriormente como "trozo no negativo" de la k-ésima capa entera. Todas las definiciones fueron tomadas del capítulo 2 de [BV04]

a excepción del baricentro, mientras que las demostraciones de los lemas y teoremas fueron realizadas de manera independiente.

Definición 3.1.7. Sean $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathbb{R}^n$ una colección de vectores, entonces definimos su **combinación afín** a partir de

$$\operatorname{aff}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \coloneqq \{\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{v} = \sum_{i=1}^{m} \theta_i \boldsymbol{v}_i, \sum_{i=1}^{m} \theta_i = 1\}$$

Lema 3.1.8. *Sean $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathbb{R}^n$ una colección de vectores. Entonces para todo $1 \leq j \leq m$ se tiene que*

$$\operatorname{aff}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} = \boldsymbol{v}_j + \operatorname{gen}\{\boldsymbol{v}_i - \boldsymbol{v}_j\}_{i=1}^{m} = \boldsymbol{v}_j + \operatorname{gen}\{\boldsymbol{v}_i - \boldsymbol{v}_j\}_{i \neq j}.$$

Demostración. Puesto que $\boldsymbol{v}_j - \boldsymbol{v}_j = \boldsymbol{0}$ es linealmente dependiente de los vectores $\{\boldsymbol{v}_i - \boldsymbol{v}_j\}_{i \neq j}$, se sigue que

$$\boldsymbol{v}_j + \operatorname{gen}\{\boldsymbol{v}_i - \boldsymbol{v}_j\}_{i=1}^{m} = \boldsymbol{v}_j + \operatorname{gen}\{\boldsymbol{v}_i - \boldsymbol{v}_j\}_{i \neq j}.$$

Sean $\theta_1, \ldots, \theta_m$ escalares tales que $\theta_1 + \cdots + \theta_m = 1$. Por un lado, tenemos

$$\begin{aligned} \sum_{i=1}^{m} \theta_i \boldsymbol{v}_i &= \sum_{i=1}^{m} \theta_i \boldsymbol{v}_j + \sum_{i=1}^{m} \theta_i (\boldsymbol{v}_i - \boldsymbol{v}_j) \\ &= \boldsymbol{v}_j + \sum_{i \neq j} \theta_i (\boldsymbol{v}_i - \boldsymbol{v}_j). \end{aligned}$$

De donde se sigue que $\operatorname{aff}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \subseteq \boldsymbol{v}_j + \operatorname{gen}\{\boldsymbol{v}_i - \boldsymbol{v}_j\}_{i \neq j}$.

Ahora bien, sea $\{\lambda_i\}_{i \neq j}$ un conjunto de $m - 1$ escalares y definamos

$$\lambda_j = 1 - \sum_{i \neq j} \lambda_i.$$

Observemos que $\lambda_1 + \cdots + \lambda_m = 1$ y, además,

$$\begin{aligned} \boldsymbol{v}_j + \sum_{i \neq j} \lambda_i (\boldsymbol{v}_i - \boldsymbol{v}_j) &= \left(1 - \sum_{i \neq j} \lambda_i\right) \boldsymbol{v}_j + \sum_{i \neq j} \lambda_i \boldsymbol{v}_i \\ &= \lambda_j \boldsymbol{v}_j + \sum_{i \neq j} \lambda_i \boldsymbol{v}_i \\ &= \sum_{i=1}^{m} \lambda_i \boldsymbol{v}_i. \end{aligned}$$

De donde se sigue que $\boldsymbol{v}_j + \operatorname{gen}\{\boldsymbol{v}_i - \boldsymbol{v}_j\}_{i \neq j} \subseteq \operatorname{aff}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$. Puesto que hemos mostrado ambas contenciones, obtenemos lo que queríamos demostrar. □

Ejemplo 3.1.9. Si $\boldsymbol{q} > \boldsymbol{0}$ es un vector coprimo y k un entero positivo, entonces la k-ésima capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ es la combinación afín de un conjunto de vectores. En efecto, recordemos de la definición 1.2.3 que esta capa entera es simplemente un hiperplano afín. Como $\boldsymbol{q}$ es el vector normal a este hiperplano, se sigue que puede ser escrito como $\boldsymbol{v} + \ker\{\boldsymbol{q} \mapsto \boldsymbol{q}^T\boldsymbol{x}\}$ para alguna $\boldsymbol{v} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$.

Sean $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ las intersecciones de la k-ésima capa entera con cada uno de los ejes. Es decir, sean, para cada $i \in \{1, \ldots, n\}$,

$$\boldsymbol{u}_i \coloneqq \frac{k}{q_i} \boldsymbol{e}_i. \tag{3.8}$$

Como cada $\boldsymbol{u}_i$ está en la k-ésima capa entera, se verifica que $\boldsymbol{q}^T\boldsymbol{u}_i = k$ y por lo tanto $\boldsymbol{u}_i - \boldsymbol{u}_j \in \ker\{\boldsymbol{q} \mapsto \boldsymbol{q}^T\boldsymbol{x}\}$. No es difícil ver entonces que el conjunto de vectores $\{\boldsymbol{u}_i - \boldsymbol{u}_j\}_{i \neq j}$ forma una base del espacio nulo de la transformación lineal $\boldsymbol{q} \mapsto \boldsymbol{q}^T\boldsymbol{x}$, por lo que obtenemos

$$H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} = \boldsymbol{u}_j + \operatorname{gen}\{\boldsymbol{u}_i - \boldsymbol{u}_j\}_{i \neq j}.$$

Comparando con la definición 1.2.3, obtenemos una construcción explícita

de $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$. Por el lema 3.1.8 concluimos que la k-ésima capa entera es la combinación afín de los vectores $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$.

Definición 3.1.10. Sean $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathbb{R}^n$ vectores linealmente independientes. Entonces definimos el **símplice** σ de dimensión $m-1$ como la combinación convexa de estos vectores:

$$\sigma = \operatorname{conv}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \coloneqq \{\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{v} = \sum_{i=1}^{m} \theta_i \boldsymbol{v}_i, \sum_{i=1}^{m} \theta_i = 1, \theta_i \geq 0\}$$

Decimos entonces que σ es generado por $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$. También definimos la **j-ésima faceta** σ_j de σ como el símplice generado por los vectores $\{\boldsymbol{v}_i\}_{i \neq j}$.

Observación. Comparando con la definición 3.1.7, encontramos que todo símplice σ generado por $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ está contenido en la combinación afín de estos vectores. Es decir,

$$\operatorname{conv}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \subseteq \operatorname{aff}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}. \tag{3.9}$$

Observación. Si σ es un símplice generado por m vectores, entonces tiene $\binom{m}{1} = m$ facetas. Tomaremos por hecho, puesto que de otra manera arriesgamos desviarnos por una tangente, que estas facetas constituyen la frontera relativa del símplice dentro de la k-ésima capa entera. Es decir, tomaremos por hecho que las facetas constituyen las "caras" de σ.

Lema 3.1.11. *Sea $\boldsymbol{q} > \boldsymbol{0}$ un vector coprimo y sea $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ la k-ésima capa entera, con parámetro k positivo. Consideremos el símplice σ generado por los vectores definidos en* (3.8)*, entonces*

$$\sigma = H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \cap \mathbb{R}^n_{\geq \boldsymbol{0}}.$$

Demostración. En el ejemplo 3.1.9 mostramos que

$$H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} = \operatorname{aff}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}.$$

($\subseteq$) Sea $\boldsymbol{x} \in \sigma$, de la definición 3.1.10 y de (3.9) encontramos que $\boldsymbol{x}$ se encuentra en la k-ésima capa entera. Además, existen escalares $\theta_1, \ldots, \theta_n$ no negativos tales que

$$\boldsymbol{x} = \theta_1 \boldsymbol{u}_1 + \cdots + \theta_n \boldsymbol{u}_n = k \begin{pmatrix} \theta_1/q_1 \\ \vdots \\ \theta_n/q_n \end{pmatrix}.$$

Como $\boldsymbol{q} > \boldsymbol{0}$ y $k > 0$ por hipótesis, tenemos que $\boldsymbol{x} \geq \boldsymbol{0}$, lo que implica que $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} \cap \mathbb{R}^n_{\geq \boldsymbol{0}}$.

($\supseteq$) El otro lado de la contención se muestra de análogamente. □

En el contexto del problema (1.1), sabemos que si σ es generado por los vectores en (3.8) entonces, por el lema anterior, todo punto entero sobre σ es un punto factible siempre que $0 < k \leq \eta$, donde recuperamos η del lema 1.2.7. Nos gustaría entonces garantizar la existencia de tal punto entero.

Adoptamos la siguiente estrategia: nos concentramos en un punto $\boldsymbol{x} \in \sigma$ y abrimos una bola (ver definición 3.1.5) con radio dado por (3.6). Si esa bola está contenida en el símplice σ, entonces el teorema 3.1.6 garantiza la existencia de un punto entero sobre σ. Por el lema anterior, garantizaríamos la existencia de un punto entero no negativo sobre la k-ésima capa entera.

Lo que se encuentra a continuación es un análisis para determinar qué tan grande debe ser k para asegurar que la bola de radio (3.6) esté contenida en el símplice σ, dado que la bola está centrada en un punto particular, a saber, en el baricentro del símplice.

Definición 3.1.12. Sea σ un símplice generado por $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathbb{R}^n$, definimos su **baricentro** $\hat{\boldsymbol{\sigma}}$ como

$$\hat{\boldsymbol{\sigma}} \coloneqq \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{v}_i.$$

Observación. El baricentro $\hat{\boldsymbol{\sigma}}$ es un elemento de σ. Esto se debe a que $\hat{\boldsymbol{\sigma}}$ es la combinación convexa de $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$, donde $\theta_1 = \cdots = \theta_m = \frac{1}{m}$.

Definición 3.1.13. Sea σ un símplice y sea $\hat{\boldsymbol{\sigma}}$ su baricentro. Entonces definimos el **radio de la bola inscrita** en σ con centro $\hat{\boldsymbol{\sigma}}$ como

$$r_\sigma \coloneqq \max\{r > 0 \colon B_r^{(k)}(\hat{\boldsymbol{\sigma}}) \subseteq \sigma\}, \tag{3.10}$$

donde $B_r^{(k)}(\hat{\boldsymbol{\sigma}})$ está dada en la definición 3.1.5.

Encontraremos que el radio de la bola inscrita está dado por el mínimo de las distancias entre el baricentro del símplice con cada una de sus facetas. Puesto que $\hat{\boldsymbol{\sigma}}_j \in \sigma_j$, sabemos bien por álgebra lineal, bien por optimización, que la distancia entre el baricentro $\hat{\boldsymbol{\sigma}}$ del símplice σ y su j-ésima faceta σ_j es

$$d(\hat{\boldsymbol{\sigma}}, \sigma_j) = |\hat{\boldsymbol{\mu}}_j^T(\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j)|, \tag{3.11}$$

donde $\hat{\boldsymbol{\mu}}_j$ es un vector unitario, normal a la j-ésima faceta, y que es paralelo a $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$. El siguiente lema caracteriza este vector.

Lema 3.1.14. *Sean $\boldsymbol{q} > \boldsymbol{0}$, $k > 0$ y retomemos el símplice σ convexamente generado por los vectores $\{\boldsymbol{u}_i\}_{i=1}^n$ definidos en* (3.8)*. Definamos, para cada $j \in \{1, \ldots, n\}$,*

$$\boldsymbol{\mu}_j \coloneqq \boldsymbol{u}_j - \frac{\boldsymbol{q}^T\boldsymbol{u}_j}{\boldsymbol{q}^T\boldsymbol{q}}\boldsymbol{q} = \boldsymbol{u}_j - \frac{k}{\|\boldsymbol{q}\|^2}\boldsymbol{q}. \tag{3.12}$$

Entonces $\boldsymbol{\mu}_j$ es paralelo a $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ y también es normal a la j-ésima faceta σ_j del símplice σ.

Demostración. Observemos que $\boldsymbol{\mu}_j \in \ker\{\boldsymbol{q} \mapsto \boldsymbol{q}^T\boldsymbol{x}\}$, en efecto,

$$\boldsymbol{q}^T\boldsymbol{\mu}_j = \boldsymbol{q}^T\boldsymbol{u}_j - \frac{k}{\|\boldsymbol{q}\|^2}\boldsymbol{q}^T\boldsymbol{q} = k - k = 0,$$

Juntando esto con los resultados del ejemplo 3.1.9 encontramos que

$$\boldsymbol{u}_j + \boldsymbol{\mu}_j \in \boldsymbol{u}_j + \ker\{\boldsymbol{q} \mapsto \boldsymbol{q}^T\boldsymbol{x}\} = \boldsymbol{u}_j + \operatorname{gen}\{\boldsymbol{u}_i - \boldsymbol{u}_j\}_{i\neq j} = H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2},$$

por lo que $\boldsymbol{\mu}_j$ es paralelo a $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$.

También debemos mostrar que si $\boldsymbol{x} \in \sigma_j$, entonces $\boldsymbol{\mu}_j^T\boldsymbol{y} = 0$ para todo $\boldsymbol{y} \in \sigma_j - \boldsymbol{x}$. Por la definición 3.1.10, tenemos que la combinación convexa de los vectores $\{\boldsymbol{u}_i\}_{i\neq j}$ genera la j-ésima faceta σ_j, y entonces basta mostrar que $\boldsymbol{\mu}_j^T\boldsymbol{y} = 0$ para todo $\boldsymbol{y} \in \sigma_j - \boldsymbol{u}_m$ con $m \neq j$.

Sea, pues, $m \in \{1, \ldots, n\} \setminus \{j\}$. Tenemos de las definiciones 3.1.7 y 3.1.10, así como del lema 3.1.8 que

$$\sigma_j = \operatorname{conv}\{\boldsymbol{u}_i\}_{i\neq j} \subseteq \operatorname{aff}\{\boldsymbol{u}_i\}_{i\neq j} = \boldsymbol{u}_m + \operatorname{gen}\{\boldsymbol{u}_i - \boldsymbol{u}_m\}_{i\neq j}.$$

De donde obtenemos

$$\sigma_j - \boldsymbol{u}_m \subseteq \operatorname{gen}\{\boldsymbol{u}_i - \boldsymbol{u}_m\}_{i\neq j},$$

así que basta mostrar que $\boldsymbol{\mu}_j^T(\boldsymbol{u}_i - \boldsymbol{u}_m) = 0$ para todo $i \neq j$ Cabe mencionar que los vectores $\{\boldsymbol{u}_i\}_{i=1}^n$ son ortogonales entre sí (ver (3.8)). Sustituyendo con la definición de $\boldsymbol{\mu}_j$ en la hipótesis, obtenemos

$$\begin{aligned}
\boldsymbol{\mu}_j^T(\boldsymbol{u}_i - \boldsymbol{u}_m) &= \boldsymbol{u}_j^T\boldsymbol{u}_i - \boldsymbol{u}_j^T\boldsymbol{u}_m - \frac{k}{\|\boldsymbol{q}\|^2}(\boldsymbol{q}^T\boldsymbol{u}_i - \boldsymbol{q}^T\boldsymbol{u}_m) \\
&= 0 - 0 - \frac{k}{\|\boldsymbol{q}\|^2}(k - k) \\
&= 0.
\end{aligned}$$

De esta manera, concluimos que $\boldsymbol{\mu}_j$ es un vector normal a σ_j. □

Ahora que encontramos vectores normales $\boldsymbol{\mu}_j$ para cada faceta σ_j, podemos simplificar un poco más (3.11). Aprovechando el hecho de que los vectores $\{\boldsymbol{u}_i\}_{i=1}^n$ son todos ortogonales entre sí, a partir de la definición

3.1.12 y de la ecuación (3.8) obtenemos los siguientes cálculos:

$$\begin{aligned}\boldsymbol{\mu}_j^T \hat{\boldsymbol{\sigma}} &= \left(\boldsymbol{u}_j - \frac{k}{\|\boldsymbol{q}\|^2}\boldsymbol{q}\right)^T \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{u}_i \\ &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{u}_j^T \boldsymbol{u}_i - \frac{k}{n\|\boldsymbol{q}\|^2}\sum_{i=1}^{n} \boldsymbol{q}^T \boldsymbol{u}_i \\ &= \frac{1}{n}\|\boldsymbol{u}_j\|^2 - \frac{k}{n\|\boldsymbol{q}\|^2}\sum_{i=1}^{n} k \\ &= \frac{k^2}{nq_j^2} - \frac{k^2}{\|\boldsymbol{q}\|^2}.\end{aligned}$$

A través de un procedimiento similar, encontramos para el baricentro $\hat{\boldsymbol{\sigma}}_j$ de la j-ésima faceta σ_j que

$$\boldsymbol{\mu}_j^T \hat{\boldsymbol{\sigma}}_j = -\frac{k^2}{\|\boldsymbol{q}\|^2}, \tag{3.13}$$

y por lo tanto

$$\boldsymbol{\mu}_j^T (\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j) = \frac{k^2}{nq_j^2}. \tag{3.14}$$

Más adelante normalizaremos $\boldsymbol{\mu}_j$ de manera que este vector sea unitario. Cabe resaltar el hecho de que el lado derecho (3.14) es positivo. Geométricamente, lo anterior implica que los vectores normales $\boldsymbol{\mu}_j$ de cada faceta σ_j apuntan hacia el interior relativo del símplice σ. Esto sugiere una caracterización alternativa de σ que nos permite interpretarlo como un poliedro y que es importante para demostrar el teorema 3.1.16.

Lema 3.1.15. *Sea $\boldsymbol{q} > \boldsymbol{0}$ un vector coprimo y sea σ el símplice generado por los vectores definidos en* (3.8)*, con $k > 0$. Entonces*

$$\sigma = \bigcap_{j=1}^{n} \{\boldsymbol{x} \in \mathbb{R}^n \colon \hat{\boldsymbol{\mu}}_j^T (\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) \geq 0\} \cap H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}, \tag{3.15}$$

donde $\hat{\boldsymbol{\mu}}_j$ es el vector $\boldsymbol{\mu}_j$ definido en (3.12) *normalizado.*

Demostración. Denotemos por $\{\boldsymbol{u}_i\}_{i=1}^n$ los vectores ortogonales definidos en (3.8). Como $\hat{\boldsymbol{\mu}}_j$ es el vector $\boldsymbol{\mu}_j$ normalizado, se sigue que

$$\{\boldsymbol{x} \in \mathbb{R}^n \colon \hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) \geq 0\} = \{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{\mu}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) \geq 0\},$$

y entonces podemos trabajar con $\boldsymbol{\mu}_j$ sin normalizarlo.

($\subseteq$) Sea $\boldsymbol{x} \in \sigma$. Por el lema 3.1.11 sabemos que $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$. También sabemos, por el ejemplo 3.1.9, que existen escalares no negativos $\theta_1, \ldots, \theta_n$ que suman 1 y que satisfacen $\boldsymbol{x} = \theta_1\boldsymbol{u}_1 + \cdots + \theta_n\boldsymbol{u}_n$. Tenemos entonces

$$\begin{aligned}
\boldsymbol{\mu}_j^T\boldsymbol{x} &= \left(\boldsymbol{u}_j - \frac{k}{\|\boldsymbol{q}\|^2}\boldsymbol{q}\right)^T \sum_{i=1}^n \theta_i\boldsymbol{u}_i \\
&= \boldsymbol{u}_j^T\left(\theta_j\boldsymbol{u}_j + \sum_{i\neq j}\theta_i\boldsymbol{u}_i\right) - \frac{k}{\|\boldsymbol{q}\|^2}\sum_{i=1}^n \theta_i\boldsymbol{q}^T\boldsymbol{u}_i \\
&= \theta_j\|\boldsymbol{u}_j\|^2 - \frac{k}{\|\boldsymbol{q}\|^2}\sum_{i=1}^n \theta_i k \\
&= \theta_j\|\boldsymbol{u}_j\|^2 - \frac{k^2}{\|\boldsymbol{q}\|^2}.
\end{aligned}$$

Retomamos de (3.13) el valor de $\boldsymbol{\mu}_j^T\hat{\boldsymbol{\sigma}}_j$, así que obtenemos

$$\boldsymbol{\mu}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) = \boldsymbol{\mu}_j^T\boldsymbol{x} - \boldsymbol{\mu}_j^T\hat{\boldsymbol{\sigma}}_j = \theta_j\|\boldsymbol{u}_j\|^2 \geq 0,$$

para todo $1 \leq j \leq n$.

($\supseteq$) Mostramos por contradicción que si $\boldsymbol{x}$ no se encuentra en el lado izquierdo de (3.15), entonces tampoco se encuentra en el lado derecho. Así pues, supongamos que $\boldsymbol{x} \notin \sigma$. Por el lema 3.1.11 se sigue o bien que $\boldsymbol{x} \notin H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ o bien que $\boldsymbol{x} \notin \mathbb{R}^n_{\geq \boldsymbol{0}}$. En el primer caso obtenemos inmediatamente que $\boldsymbol{x}$ no se encuentra en el lado derecho de (3.15). Luego, como $\boldsymbol{x} \notin \sigma$, es válido suponer que $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ pero $\boldsymbol{x} \notin \mathbb{R}_{\geq \boldsymbol{0}}$. Como $\{\boldsymbol{u}_i\}_{i=1}^n$ es base de

$\mathbb{R}^n$, existen escalares $\{\lambda_i\}_{i=1}^n$ tales que

$$\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i.$$

Como las entradas de $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ son todas no negativas y $x_j < 0$ para alguna $1 \leq j \leq n$, se sigue que $\lambda_j < 0$. Observemos que

$$\begin{aligned}
\boldsymbol{\mu}_j^T \boldsymbol{x} &= \sum_{i=1}^{n} \lambda_i \boldsymbol{\mu}_j^T \boldsymbol{u}_i \\
&= \sum_{i=1}^{n} \lambda_i \left(\boldsymbol{u}_j - \frac{k}{\|\boldsymbol{q}\|^2} \boldsymbol{q} \right)^T \boldsymbol{u}_i \\
&= \sum_{i=1}^{n} \lambda_i \left(\boldsymbol{u}_j^T \boldsymbol{u}_i - \frac{k}{\|\boldsymbol{q}\|^2} \boldsymbol{q}^T \boldsymbol{u}_i \right) \\
&= \lambda_j \|\boldsymbol{u}_j\|^2 - \frac{k^2}{\|\boldsymbol{q}\|} \sum_{i=1}^{n} \lambda_i.
\end{aligned}$$

Pero $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} = \operatorname{aff}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ (véase el ejemplo 3.1.9) y entonces los escalares $\lambda_1, \ldots, \lambda_n$ suman a 1. Sustituyendo,

$$\boldsymbol{\mu}_j^T \boldsymbol{x} = \lambda_j \|\boldsymbol{u}_j\|^2 - \frac{k^2}{\|\boldsymbol{q}\|^2},$$

retomando el valor de $\boldsymbol{\mu}_j^T \hat{\boldsymbol{\sigma}}_j$ en (3.13), encontramos que

$$\boldsymbol{\mu}_j^T (\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) = \lambda_j \|\boldsymbol{u}_j\|^2 = \lambda_j \frac{k^2}{q_j^2} < 0$$

y entonces $\boldsymbol{x}$ no es elemento del semi-espacio afín $\{\boldsymbol{x} : \boldsymbol{\mu}_j^T (\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) \geq 0\}$, por lo que tampoco es elemento del lado derecho de (3.15). □

Hemos caracterizado completamente, salvo la constante de normalización de $\boldsymbol{\mu}$, las distancias entre el baricentro $\hat{\boldsymbol{\sigma}}$ del símplice σ y cada una de sus facetas σ_j. El siguiente teorema relaciona estas distancias con el radio de la

bola inscrita en el símplice σ (ver definición 3.1.13).

Teorema 3.1.16. *Sea $\boldsymbol{q} > \mathbf{0}$ un vector coprimo y sea σ el símplice generado por los vectores $\{\boldsymbol{u}_i\}_{i=1}^n$ definidos en* (3.8)*. Entonces el radio r_σ de la bola inscrita en σ con centro $\hat{\boldsymbol{\sigma}}$ está dado por*

$$r_\sigma = \min_{1\leq j\leq n} d(\hat{\boldsymbol{\sigma}}, \sigma_j) = \min_{1\leq j\leq n} \hat{\boldsymbol{\mu}}_j^T(\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j),$$

donde $\hat{\boldsymbol{\mu}}_j$ es el vector $\boldsymbol{\mu}_j$ definido en (3.12) *normalizado, y σ_j es la j-ésima faceta del símplice σ.*

Demostración. Como $\hat{\boldsymbol{\sigma}} \in \sigma$, tenemos del lema 3.1.15 que $\boldsymbol{\mu}_j^T(\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j) \geq 0$ y, por lo tanto, deducimos de (3.11) que la distancia entre $\hat{\boldsymbol{\sigma}}$ y la j-ésima faceta σ_j es

$$d(\hat{\boldsymbol{\sigma}}, \sigma_j) = \hat{\boldsymbol{\mu}}_j^T(\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j). \tag{3.16}$$

Supongamos que $r \leq d(\hat{\boldsymbol{\sigma}}, \sigma_j)$ para todo $j \in \{1, \ldots, n\}$ y sea $\boldsymbol{x} \in B_r^{(k)}(\hat{\boldsymbol{\sigma}})$. Observemos que

$$\begin{aligned}\hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) &= \hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}) + \hat{\boldsymbol{\mu}}_j^T(\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j) \\ &= \hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}) + d(\hat{\boldsymbol{\sigma}}, \sigma_j).\end{aligned}$$

Por la desigualdad de Cauchy-Schwarz, tenemos

$$\hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}) \geq -\left\|\hat{\boldsymbol{\mu}}_j\right\| \|\boldsymbol{x} - \hat{\boldsymbol{\sigma}}\| \geq -r,$$

pues $\hat{\boldsymbol{\mu}}_j$ es unitario y $\boldsymbol{x} \in B_r^{(k)}(\hat{\boldsymbol{\sigma}})$. Así pues, tenemos

$$\hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) \geq -r + d(\hat{\boldsymbol{\sigma}}, \sigma_j) \geq 0,$$

daod que supusimos $r \leq d(\hat{\boldsymbol{\sigma}}, \sigma_j)$ para todo $1 \leq j \leq n$. Además, como $\boldsymbol{x} \in B_r^{(k)}(\hat{\boldsymbol{\sigma}})$, tenemos por la definición 3.1.5 que $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$. De lo

anterior obtenemos

$$\boldsymbol{x} \in \bigcap_{j=1}^{n} \{\boldsymbol{x} \in \mathbb{R}^n \colon \hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) \geq 0\} \cap H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2} = \sigma,$$

donde la última igualdad se sigue del lema 3.1.15. Así pues, $B_r^{(k)}(\hat{\boldsymbol{\sigma}}) \subseteq \sigma$ si $r \leq d(\hat{\boldsymbol{\sigma}}, \sigma_j)$ para toda $j \in \{1, \ldots, n\}$. De la definición 3.1.13 encontramos entonces que el radio r_σ de la bola inscrita satisface

$$r_\sigma \geq \min_{1 \leq j \leq n} d(\hat{\boldsymbol{\sigma}}, \sigma_j). \tag{3.17}$$

Ahora bien, supongamos que $r > d(\hat{\boldsymbol{\sigma}}, \sigma_j)$ para alguna $j \in \{1, \ldots, n\}$. Como σ_j es cerrado, existe un punto $\boldsymbol{x} \in \sigma_j$ que satisface $d(\hat{\boldsymbol{\sigma}}, \sigma_j) = d(\hat{\boldsymbol{\sigma}}, \boldsymbol{x})$. Luego, $\|\boldsymbol{x} - \hat{\boldsymbol{\sigma}}\| < r$. Entonces existe $\varepsilon > 0$ tal que

$$\left\|(\boldsymbol{x} - \varepsilon\hat{\boldsymbol{\mu}}_j) - \hat{\boldsymbol{\sigma}}\right\| \leq r,$$

lo que implica que $\boldsymbol{x} - \varepsilon\hat{\boldsymbol{\mu}}_j \in B_r^{(k)}(\hat{\boldsymbol{\sigma}})$. Observemos que

$$\hat{\boldsymbol{\mu}}_j^T((\boldsymbol{x} - \varepsilon\hat{\boldsymbol{\mu}}_j) - \hat{\boldsymbol{\sigma}}_j) = \hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) - \varepsilon.$$

Pero $\boldsymbol{x}, \hat{\boldsymbol{\sigma}}_j \in \sigma_j$, así que $\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j \in \sigma_j - \hat{\boldsymbol{\sigma}}_j$. Del lema 3.1.14 encontramos que

$$\hat{\boldsymbol{\mu}}_j^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) = 0,$$

de donde obtenemos

$$\hat{\boldsymbol{\mu}}_j^T((\boldsymbol{x} - \varepsilon\hat{\boldsymbol{\mu}}_i) - \hat{\boldsymbol{\sigma}}_j) = -\varepsilon < 0,$$

lo cual implica que $\boldsymbol{x} - \varepsilon\hat{\boldsymbol{\mu}}_j$ no se encuentra en el semi-espacio afín definido por $\{\boldsymbol{x} : \hat{\boldsymbol{\mu}}^T(\boldsymbol{x} - \hat{\boldsymbol{\sigma}}_j) \geq 0\}$. Así pues, por el lema 3.1.15, encontramos que $\boldsymbol{x} - \varepsilon\hat{\boldsymbol{\mu}}_j \notin \sigma$. Pero $\boldsymbol{x} - \varepsilon\hat{\boldsymbol{\mu}}_j \in B_r^{(k)}(\hat{\boldsymbol{\sigma}})$. De aquí se desprende que $B_r^{(k)}(\hat{\boldsymbol{\sigma}}) \not\subseteq \sigma$ si $r > d(\hat{\boldsymbol{\sigma}}, \sigma_j)$ para alguna $j \in \{1, \ldots, n\}$. De la definición 3.1.13 obtenemos

entonces

$$r_\sigma \leq \min_{1\leq j\leq n} d(\hat{\boldsymbol{\sigma}}, \sigma_j). \tag{3.18}$$

De (3.17) y de (3.18) concluimos entonces con lo que queríamos demostrar. □

Continuamos con el cálculo de la distancia entre el baricentro $\hat{\boldsymbol{\sigma}}$ del símplice σ y cada una de sus facetas σ_j. Aplicando el teorema anterior, podemos obtener una expresión explícita del radio de la bola inscrita en σ. De (3.11) tenemos

$$d(\hat{\boldsymbol{\sigma}}, \sigma_j) = \hat{\boldsymbol{\mu}}_j^T(\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j) = \frac{\boldsymbol{\mu}_j^T(\hat{\boldsymbol{\sigma}} - \hat{\boldsymbol{\sigma}}_j)}{\|\boldsymbol{\mu}_j\|}. \tag{3.19}$$

Recordemos de (3.14) que ya contamos con el numerador, así que ahora debemos calcular la norma de $\boldsymbol{\mu}_j$. Tenemos

$$\begin{aligned}
\|\boldsymbol{\mu}_j\|^2 &= \boldsymbol{\mu}_j^T \boldsymbol{\mu}_j \\
&= \left(\boldsymbol{u}_j - \frac{k}{\|\boldsymbol{q}\|^2}\boldsymbol{q}\right)^T \left(\boldsymbol{u}_j - \frac{k}{\|\boldsymbol{q}\|^2}\boldsymbol{q}\right) \\
&= \|\boldsymbol{u}_j\|^2 - 2\frac{k}{\|\boldsymbol{q}\|^2}\boldsymbol{q}^T\boldsymbol{u}_j + \frac{k^2}{\|\boldsymbol{q}\|^4}\boldsymbol{q}^T\boldsymbol{q} \\
&= \frac{k^2}{q_j^2} - 2\frac{k^2}{\|\boldsymbol{q}\|^2} + \frac{k^2}{\|\boldsymbol{q}\|^2} \\
&= \frac{k^2}{q_j^2} - \frac{k^2}{\|\boldsymbol{q}\|^2}.
\end{aligned}$$

De donde obtenemos

$$\|\boldsymbol{\mu}_j\| = k\sqrt{\frac{1}{q_j^2} - \frac{1}{\|\boldsymbol{q}\|^2}}. \tag{3.20}$$

Usando (3.14) y (3.20) para sustituir en (3.19), obtenemos

$$d(\hat{\boldsymbol{\sigma}}, \sigma_j) = \frac{k}{n} \cdot \frac{1}{q_j^2\sqrt{q_j^{-2} - \|\boldsymbol{q}\|^{-2}}}$$

Finalmente, del teorema 3.1.16 encontramos que el radio r_σ de la bola inscrita en el símplice σ con baricentro $\hat{\boldsymbol{\sigma}}$ está dado por

$$r_\sigma = \min_{1\leq j\leq n}\{d(\hat{\boldsymbol{\sigma}}, \sigma_j)\} = \frac{k}{n}\cdot\frac{1}{\max_{1\leq j\leq n}\{q_j^2\sqrt{q_j^{-2}+\|\boldsymbol{q}\|^{-2}}\}} \tag{3.21}$$

Recordemos que la motivación detrás de todos estos cálculos se fundamenta en determinar la existencia de un punto entero sobre el "trozo" no negativo de la k-ésima capa entera. Esto permitirá determinar la existencia de soluciones enteras no negativas de la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$ y por lo tanto de obtener un resultado análogo al teorema 2.0.5.

Para empezar, en el siguiente teorema se proporciona una cota inferior de manera que podamos asegurar la existencia de puntos enteros en una vecindad del baricentro $\hat{\boldsymbol{\sigma}}$. Este punto no es especial, pues en realidad podemos realizar el mismo procedimiento enfocándonos en otros puntos del símplice σ para asegurar soluciones en sus respectivas vecindades. Entonces, dependiendo del punto, podemos obtener mejores o peores cotas para k. El punto más interesante es aquel que provee la cota inferior más pequeña. Aunque no lo demostraremos, creemos que el baricentro $\hat{\boldsymbol{\sigma}}$ provee la mejor cota.

Teorema 3.1.17. *Sea $\boldsymbol{q} > \boldsymbol{0}$ un vector coprimo y sea k un entero positivo suficientemente grande. Entonces existe un punto entero sobre el símplice σ generado por los vectores en* (3.8).

Demostración. Sea r el radio definido en (3.6) y sea r_σ el radio definido en (3.21). Por el teorema 3.1.6 sabemos que existe un punto entero $\boldsymbol{x}$ en $B_r^{(k)}(\hat{\boldsymbol{\sigma}})$, y por el teorema 3.1.16 sabemos que la bola $B_{r_\sigma}^{(k)}(\hat{\boldsymbol{\sigma}})$ está contenida en σ. Entonces basta mostrar que existe k suficientemente grande tal que $r \leq r_\sigma$, pues esto implica la contención de en medio en la cadena

$$\boldsymbol{x} \in B_r^{(k)}(\hat{\boldsymbol{\sigma}}) \subseteq B_{r_\sigma}^{(k)}(\hat{\boldsymbol{\sigma}}) \subseteq \sigma.$$

De (3.6) y de (3.21) obtenemos que $r \leq r_\sigma$ si y solo si

$$k \geq \frac{n\sqrt{n-1}}{2} \|M\| \max_{1 \leq j \leq n} \{q_j^2 \sqrt{q_j^{-2} + \|\boldsymbol{q}\|^{-2}}\}, \tag{3.22}$$

que es lo que queríamos demostrar. □

De manera inmediata obtenemos también los siguientes teoremas. Cabe mencionar que estos resultados solamente muestran la existencia de una solución entera $\boldsymbol{x}$ no negativa para la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$. Será en la sección 3.2 que discutiremos cómo encontrar una solución.

Teorema 3.1.18. *Sea $\boldsymbol{q} > \boldsymbol{0}$ un vector coprimo. Si $k \in \mathbb{Z}_{\geq 0}$ satisface* (3.22)*, entonces la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$ tiene soluciones enteras no negativas.*

Demostración. Consideremos el símplice σ generado por los vectores en (3.8). Por el teorema 3.1.17 existe un punto entero $\boldsymbol{x} \in \sigma$, lo cual implica que $\boldsymbol{x} \geq \boldsymbol{0}$ y además $\boldsymbol{x} \in H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ por el lema 3.1.11. Luego, por el lema 1.2.6, $\boldsymbol{x}$ satisface la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$. □

Teorema 3.1.19. *Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y supongamos que su múltiplo coprimo $\boldsymbol{q}$ tiene entradas estrictamente positivas. Entonces el problema* (1.1) *se puede resolver a través de encontrar la solución de una sola ecuación lineal en n incógnitas para un presupuesto u suficientemente grande.*

Demostración. Por la definición 1.2.1 sabemos que existe un escalar m tal que $\boldsymbol{p} = m\boldsymbol{q}$. Supongamos, sin pérdida de generalidad, que m es positivo. Del lema 1.2.7 tenemos que el entero η parametriza la primera capa entera en satisfacer el presupuesto y que $\eta = \lfloor u/m \rfloor$. Por el teorema 3.1.18 sabemos que si η es suficientemente grande, entonces la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = \eta$ tiene al menos una solución entera no negativa $\boldsymbol{x}$. Luego, $\boldsymbol{x}$ es factible para el problema (1.1), pero por la maximalidad de η encontramos que

$\boldsymbol{x}$ también es un punto óptimo. Así, solo es necesario resolver una ecuación lineal diofantina para determinar la solución del problema (1.1). □

El teorema 3.1.18 provee nuevas cotas superiores para el número de Frobenius en el Problema Diofantino de Frobenius [RA05]. De manera resumida, dada una colección de enteros $q_1, \ldots, q_n$ coprimos, el número de Frobenius es el entero F más grande tal que F no pueda ser expresado como una combinación lineal entera no negativa de $q_1, \ldots, q_n$. En efecto, o bien F es mayor o igual que el lado derecho de (3.22) o bien es estrictamente menor. El primer caso no puede ocurrir porque el teorema 3.1.18 asegura que la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = F$ tiene una solución no negativa, es decir, que F puede ser escrito como una combinación lineal entera no negativa de $q_1, \ldots, q_n$, pero entonces F no puede ser el número de Frobenius de estos enteros. Así pues, F debe ser estrictamente menor que el lado derecho de (3.22). Un estudio sobre cómo se compara esta colección de cotas con respecto a las presentadas en el capítulo 3 de [RA05], por ejemplo, queda fuera del propósito de esta tesis.

En último lugar, mencionamos que eventualmente es suficiente con revisar la primera capa entera. Sin embargo, no hemos demostrado que el número de capas enteras a revisar eventualmente decrece en cuanto el presupuesto u aumenta. Demostrar que este comportamiento siempre se cumple es mucho más difícil y queda fuera del propósito de esta tesis.

3.2. Construcción de soluciones

Sea $\boldsymbol{p} \in \mathbb{R}^n$ un vector esencialmente entero y supongamos que las entradas de su múltiplo coprimo $\boldsymbol{q}$ son todas estrictamente positivas. Supongamos, sin pérdida de generalidad, que el escalar m que satisface $\boldsymbol{p} = m\boldsymbol{q}$ es también positivo. Ya hemos visto cómo la solución del problema (1.1) se traduce a la

búsqueda de una solución entera no negativa de la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$ para alguna $k \leq \eta$, donde η es tomada del lema 1.2.7.

En esta sección presentamos los algoritmos 4 y 5 en las páginas 93 y 94, que se encargan de obtener soluciones enteras no negativas de la ecuación $\boldsymbol{q}^T\boldsymbol{x} = k$. De manera resumida, el algoritmo 4 utiliza la estrategia de "vuelta atrás" (*backtracking*) para buscar una solución entera no negativa de la ecuación anterior. En caso de que este algoritmo no encuentre ninguna solución, el algoritmo 5 desciende a la $(k-1)$-ésima capa entera para volver a realizar la búsqueda.

Teorema 3.2.1. *El algoritmo 4 en la página 93 es correcto.*

Demostración. Hacemos la demostración por inducción en la dimensión n del vector $\boldsymbol{q}$. Supongamos, para el caso base, que $n = 2$. Luego, queremos encontrar soluciones enteras no negativas de la ecuación

$$q_1x_1 + q_2x_2 = k. \tag{3.23}$$

Por hipótesis sabemos que q_1 y q_2 son coprimos. Luego, del teorema 1.1.10 encontramos que las soluciones enteras de esta ecuación están dadas por

$$\begin{cases} x_1 = kx_1' + q_2t, \\ x_2 = kx_2' - q_1t, \end{cases} \tag{3.24}$$

donde $t \in \mathbb{Z}$ es una variable libre, y x_1', x_2' son los coeficientes de Bézout (véase definición 1.1.9) de q_1 y q_2, respectivamente. Por claridad, escribimos x_1' y x_2' como x_{n-1}' y x_n' en la línea 4. Despejando de estas soluciones, encontramos que existen soluciones no negativas si y solo si existe $t \in \mathbb{Z}$ que satisfaga

$$b_1 \coloneqq \left\lceil -\frac{kx_1'}{q_2} \right\rceil \leq t \leq \left\lfloor \frac{kx_2'}{q_1} \right\rfloor =: b_2.$$

Los enteros b_1 y b_2 en las líneas 5 y 6 representan el lado izquierdo y dere-

cho de estas desigualdades, respectivamente. De esta manera, el algoritmo devuelve NIL si y solo si este intervalo no está bien definido, es decir, si y solo si no existen soluciones enteras no negativas. Supongamos, pues, que este intervalo sí está bien definido. Entonces, podemos escoger que la variable libre t sea b_1. Sustituyendo en (3.24) obtenemos una solución entera no negativa de la ecuación (3.23) (ver las líneas 9 y 10) y entonces el algoritmo es correcto para $n = 2$.

Haciendo uso de la hipótesis inductiva, queremos mostrar que el algoritmo también es correcto para $n \geq 3$ si lo es para $n - 1 \geq 2$. Entonces deseamos encontrar soluciones enteras no negativas de la ecuación (1.14) Haciendo la misma sustitución que en (1.16), recordando que $q_1, \ldots, q_n$ son coprimos por hipótesis, que definimos $\omega_1 \coloneqq k$, y renombrando las variables (x en vez de x_1, g en vez de g_2 y ω en vez de ω_2), obtenemos la ecuación

$$q_1 x + g\omega = k. \tag{3.25}$$

Observemos que, como $g_1 = 1$, el entero $g = \operatorname{mcd}\{q_2/g_1, \ldots, q_n/g_1\}$, es equivalente a lo que se encuentra en la línea 12. Por definición de g, tenemos que q_1 y g son coprimos (ver el lema 1.1.6), así que por el teorema 1.1.10 tenemos que las soluciones enteras de esta ecuación están dadas por

$$\begin{cases} x = kx' + gt, \\ \omega = k\omega' - q_1 t, \end{cases} \tag{3.26}$$

donde $t \in \mathbb{Z}$ es una variable libre, y x', ω' son los coeficientes de Bézout de q_1, g. Recordemos de (1.16) que

$$\omega = \frac{q_2}{g} x_2 + \cdots + \frac{q_n}{g} x_n. \tag{3.27}$$

Como $\boldsymbol{q} > \mathbf{0}$ por hipótesis, $g > 0$ porque el máximo común divisor siempre es positivo, y exigimos que $x_2, \ldots, x_n$ sean no negativos, debe ser el caso que

ω también sea no negativo. Luego, despejando de (3.26), existen soluciones no negativas de la ecuación (3.25) si y solo si existe $t \in \mathbb{Z}$ que satisfaga

$$\left\lceil -\frac{kx'}{g} \right\rceil \leq t \leq \left\lfloor \frac{k\omega'}{q_1} \right\rfloor . \tag{3.28}$$

Los enteros b_1 y b_2 en las líneas 14 y 15 representan el lado izquierdo y derecho de estas desigualdades, respectivamente. Si no existe tal variable libre $t \in \mathbb{Z}$ es porque el intervalo $[b_1, b_2]$ no está bien definido y por lo tanto $b_2 < b_1$. El algoritmo entonces salta a la línea 27 y devuelve NIL.

Si el intervalo $[b_1, b_2]$ está bien definido, entonces podemos asegurar la no negatividad de x y de ω en (3.26) para cualquier elección de t en $[b_1, b_2]$ a causa de (3.28) y en la línea 21 nos encargamos entonces de encontrar soluciones enteras no negativas de la ecuación (3.27). Se verifica automáticamente que los coeficientes del lado derecho de esta ecuación son coprimos y constituyen justamente las entradas del vector $\boldsymbol{q}^{\texttt{tail}}$, como se observa en la línea 17. Como $g > 0$ se sigue que $\boldsymbol{q}^{\texttt{tail}} > \mathbf{0}$. Luego, $\boldsymbol{q}^{\texttt{tail}}$ y ω satisfacen las hipótesis del algoritmo.

Por hipótesis inductiva, en la línea 21 tenemos o bien que $\boldsymbol{x}^{\texttt{tail}}$ es entero no negativo y solución de (3.27), o bien es NIL. En el primer caso y definiendo $\boldsymbol{x}$ como el vector de la línea 25 encontramos que

$$\boldsymbol{q}^T \boldsymbol{x} = q_1 x + g \left(\boldsymbol{q}^{\texttt{tail}}\right)^T \boldsymbol{x}^{\texttt{tail}} = q_1 x + g\omega = k.$$

Pero $x \geq 0$ por construcción y $\boldsymbol{x}^{\texttt{tail}} \geq \mathbf{0}$ por este caso de la hipótesis inductiva. Así, $\boldsymbol{x}$ también es no negativo.

Finalmente, en caso de que $\boldsymbol{x}^{\texttt{tail}}$ sea NIL, iteramos sobre otra elección de la variable libre t y regresamos al caso pasado. En caso de que este vector sea NIL para todas las elecciones posibles de t en el intervalo de factiblidad $[b_1, b_2]$, se sigue por hipótesis inductiva que la ecuación (3.27) no tiene solución entera no negativa y por lo tanto tampoco la tiene la ecuación

(1.14). Una vez agotadas estas elecciones finitas, devolvemos NIL en la línea 27.

En conclusión, si el algoritmo es correcto para vectores $\boldsymbol{q}$ con dimensión $n-1 \geq 2$, entonces también lo es para $\boldsymbol{q}$ con dimensión $n \geq 3$. Juntando esto con el caso base, se sigue por inducción que el algoritmo es correcto para toda $n \geq 2$, lo cual termina la demostración. □

Si recordamos el método de Ramificación y Acotamiento documentado en el algoritmo 1, podemos observar que la elección del parámetro libre t en el intervalo de factiblidad I definido en la línea 16 del algoritmo 4 es similar a la elección del subproblema S_i definido en la línea 5 del algoritmo 1. La diferencia radica en que, como todos los puntos enteros sobre la k-ésima capa entera tienen el mismo nivel de utilidad k, no es necesario desarrollar políticas de poda así como lo hicimos en el ejemplo 1.1.18. De cierta manera, la única política de poda posible es la de infactibilidad por no respetar la no negatividad de un punto entero.

Teorema 3.2.2. *El algoritmo 5 en la página 94 es correcto.*

Demostración. A causa del teorema 3.2.1 basta verificar que el algoritmo termina y no devuelve NIL. Además, obtenemos la maximalidad de k debido a la manera en la que iniciamos el ciclo en la línea 2. Tenemos $0 \leq \eta$ por hipótesis y observemos que $\mathbf{0}$ es la única solución entera no negativa de la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = 0$. De esta manera, si la ecuación $\boldsymbol{q}^T\boldsymbol{x} = k$ no tiene solución para $0 < k \leq \eta$, entonces el algoritmo devuelve $\mathbf{0}$ debido al teorema 3.2.1. □

En realidad, sabemos por el lema 3.1.1 que el parámetro k definido en la línea 2 descenderá hasta 0 si y solo si el parámetro τ definido en (3.2) es nulo. No obstante, la demostración del teorema 3.2.2 se vuelve más simple

Algoritmo 4: `NonNegativeIntSolFin`

Datos:
$(\boldsymbol{q}, k)$, donde $\boldsymbol{q} \in \mathbb{Z}^n_{>\mathbf{0}}$ es coprimo con $n \geq 2$ y $k \geq 0$.
Resultado:
$\boldsymbol{x} \in \mathbb{Z}^n_{\geq \mathbf{0}}$ tal que $\boldsymbol{q}^T\boldsymbol{x} = k$ o NIL.

inicio 1
  $n \leftarrow$ `length(`$\boldsymbol{q}$`)`; 2
  si $n = 2$ **entonces** 3
    $x'_{n-1}, x'_n \leftarrow$ `Bezout(`q_1, q_2`)`; 4
    $b_1 \leftarrow \lceil -kx'_{n-1}/q_2 \rceil$; 5
    $b_2 \leftarrow \lfloor kx'_n/q_1 \rfloor$; 6
    si $b_2 < b_1$ **entonces** 7
      devolver NIL; 8
    $x_{n-1} \leftarrow kx'_{n-1} + b_1 q_2$; 9
    $x_n \leftarrow kx'_n - b_1 q_1$; 10
    devolver (x_{n-1}, x_n); 11
  $g \leftarrow \mathrm{mcd}\{q_2, \ldots, q_n\}$; 12
  $x', \omega' \leftarrow$ `Bezout(`q_1, g`)`; 13
  $b_1 \leftarrow \lceil -kx'/g \rceil$; 14
  $b_2 \leftarrow \lfloor k\omega'/q_1 \rfloor$; 15
  $I \leftarrow \{b_1, b_1 + 1, \ldots, b_2\}$; 16
  $\boldsymbol{q}^{\mathtt{tail}} \leftarrow (q_{i+1}/g : 1 \leq i \leq n-1)$; 17
  mientras $I \neq \emptyset$ **hacer** 18
    elegir $t \in I$; 19
    $\omega \leftarrow k\omega' - tq_1$; 20
    $\boldsymbol{x}^{\mathtt{tail}} \leftarrow$ `NonNegativeIntSolFin(`$\boldsymbol{q}^{tail}$, ω`)`; 21
    si $\boldsymbol{x}^{tail} \neq$ NIL **entonces** 22
      $r \leftarrow$ `length(`$\boldsymbol{x}^{tail}$`)`; 23
      $x \leftarrow kx' + tg$; 24
      devolver $(x, x_1^{tail}, \ldots, x_r^{tail})$; 25
    $I \leftarrow I \setminus \{t\}$; 26
  devolver NIL; 27

cuando en el algoritmo 5 dejamos que k se encuentre en $[0, \eta]$ en vez de $[\tau, \eta]$. Esta modificación no afecta en lo más mínimo la correctud del algoritmo.

Vale la pena mencionar lo siguiente con respecto al algoritmo 4. Varios lenguajes de programación, tales como Python, cuentan con un límite en las llamadas de recursión que el usuario puede realizar. Si bien este límite puede modificarse, aumenta la posibilidad de encontranos con un desbordamiento de pila, pues el algoritmo 4 no está expresado en forma de recursión terminal.

Además, este algoritmo no minimiza, por ejemplo, el número de llamadas para calcular el máximo común divisor en la línea 12 o los coeficientes de Bézout en la línea 13. En efecto, supongamos que un intervalo de factibilidad finito I definido en la línea 16 causa que $\boldsymbol{x}^{\texttt{tail}}$ sea NIL para todo $t \in I$. Entonces estaríamos haciendo $|I|$ llamadas recursivas a `NonNegativeIntSolFin` en la línea 21 con el mismo vector $\boldsymbol{q}^{\texttt{tail}}$ y, por lo tanto, estaríamos calculando $|I|$ veces la misma g y los mismos x', ω'.

A pesar de los puntos anteriores, decidimos escribir el algoritmo 4 de esa manera debido a que se simplificaba de manera significativa la demostración del teorema 3.2.1. Realizamos una implementación equivalente más eficiente a través de ciclos para obtener los resultados de la siguiente sección.

Algoritmo 5: `Dioph`

Datos:
$\boldsymbol{q} \in \mathbb{Z}_{>\mathbf{0}}$ coprimo tal que `length(`$\boldsymbol{q}$`)` ≥ 2.
$\eta \geq 0$.
Resultado:
$\boldsymbol{x} \in \mathbb{Z}^n_{\geq \mathbf{0}}$ tal que $\boldsymbol{q}^T\boldsymbol{x} = k$ con $0 \leq k \leq \eta$ maximal.

```
inicio                                                          1
    para k ← η a 0 hacer                                        2
        x ← NonNegativeIntSolFin(q, k);                         3
        si x ≠ NIL entonces                                     4
            devolver x;                                         5
```

3.3. Experimentos numéricos

En esta sección medimos la eficiencia y estabilidad del algoritmo 5. Discutimos detalles de implementación del algoritmo 4. Introducimos una formulación de programación dinámica (FPD) capaz de resolver instancias de (1.1) cuando $\boldsymbol{q} > \mathbf{0}$, de manera que obtenemos otro punto de comparación para nuestro algoritmo además de Ramificación y Acotamiento (R&A) implementado en el *solver* COIN-OR Branch and Cut (CBC) ([Lou03]). Finalmente, explicamos y realizamos nuestros experimentos numéricos.

Utilizamos el mismo equipo así como la misma configuración de CBC que se detalla en la página 60 para la realización de estos experimentos.

3.3.1. Detalles de implementación

Mencionamos en la sección pasada que el algoritmo 4 fue escrito de esa manera para demostrar la correctud de nuestro método. En esta sección preferimos realizar una implementación equivalente en ciclos debido al límite suave en las llamadas de recursión que permite Python. Este límite es de 3,000 llamadas recursivas en la máquina donde se realizaron los experimentos. Esto significa que, de manera predeterminada, solamente podríamos resolver problemas de dimensión $n \leq 3{,}000$. No obstante, en la subsección 3.3.3 realizamos experimentos con problemas de dimensión $n \leq 150{,}000$.

Hemos visto que algoritmo 4 calcula repetidamente los mismos máximos común divisores g y los mismos coeficientes de Bézout x', ω'. Puesto que estos números dependen exclusivamente del vector $\boldsymbol{q}$, decidimos calcularlos en una fase de preprocesamiento antes de llamar `NonNegativeIntSolFin`.

La implementación por ciclos utiliza una pila (*stack*) de estados (i, b_1, b_2, ω), donde i indica el nivel o la variable t_i que debemos escoger; el resto de los parámetros están definidos en el algoritmo 4. La elección de t_i ciertamente es la más importante, pues determina el intervalo de factibilidad $I = [b_1, b_2]$

en el siguiente nivel $i + 1$. Existen, *a priori*, dos estrategias que podemos adoptar para este tipo de elecciones.

La primera estrategia escoge $t_i \leftarrow b_1$ o $t_i \leftarrow b_2$ para todo nivel i. Si, en tal nivel i, el intervalo de factibilidad es vacío, entonces retrocedemos en la pila y hacemos la sustitución $t_i \leftarrow t_i + 1$ o $t_i \leftarrow t_i - 1$ siempre que t_i se encuentre dentro del intervalo de factibilidad $[b_1, b_2]$. Así también, actualizamos nuestra variable x como lo hacemos en la línea 24 del algoritmo 4. Repetimos el proceso hasta obtener una solución $\boldsymbol{x}$ entera no negativa o hasta agotar nuestro intervalo de factibilidad y devolvemos NIL.

La segunda estrategia consiste en construir un árbol de la siguiente manera: en el nivel i del árbol escogemos el punto medio $t_i \leftarrow \lfloor (b_1 + b_2)/2 \rfloor$ (el cual representa una elección de t en la línea 19 del algoritmo 4 en el nivel de recursión i) y empujamos a la pila los estados $(i, b_1, t_i - 1, \omega)$ en caso de que $b_1 < t_i$ y también empujamos $(i, t_i + 1, b_2, \omega)$ en caso de que $t_i < b_2$. Estos estados representan subproblemas que se deben resolver si nuestra elección de t_i causa que el siguiente intervalo de factibilidad sea vacío. En caso de que este intervalo no sea vacío, también empujamos a la pila el estado del siguiente nivel con parámetros b_1, b_2 y ω calculados como en el algoritmo 4. Así pues, en el nivel $i < n - 1$ del árbol intentamos encontrar la i-ésima entrada del vector solución $\boldsymbol{x} \geq \mathbf{0}$ de manera que $x_i \geq 0$ pero también que x_{i+1} pueda ser no negativo. Puesto que empujamos a la pila un estado del siguiente nivel $i + 1$ antes de agotar el intervalo de factiblidad en el nivel i, realizamos una búsqueda en profundidad del árbol (*depth-first search*).

Encontramos en los experimentos preliminares que la segunda estrategia es mucho más eficaz. En los experimentos de la subsección 3.3.3, la primera estrategia tenía tiempos de terminación aproximadamente equivalentes a los de CBC. Para los experimentos de la subsección 3.3.4, ambas estrategias tenían tiempos de terminación significativamente menores que CBC. Por ello, decidimos omitir los resultados de la primera estrategia.

Observemos que el orden en el que consideramos los estados $(i, b_1, t_i - 1, \omega)$ y $(i, t_i + 1, b_2, \omega)$ determina si visitamos primero el lado izquierdo o el lado derecho del árbol. Realizamos los experimentos con ambas posibilidades. Así pues, llamaremos `dioph_left` a la implementación que recorre primero el subintervalo izquierdo y definimos análogamente `dioph_right`.

3.3.2. Una formulación de programación dinámica

El libro [MT90] está dedicado a realizar una exposición de los métodos más conocidos para resolver el Problema de la Mochila, del cual el problema (1.1) es un caso particular si suponemos que el vector objetivo $\boldsymbol{p}$ tiene entradas con el mismo signo. A fin de contar con otro algoritmo además de Ramificación y Acotamiento para comparar la eficiencia del algoritmo 5, decidimos adaptar una formulación de programación dinámica (FPD) expuesta en las páginas 95 a 98 de [MT90] que permite resolver instancias de (1.1) bajo algunas suposiciones adicionales. A saber, supondremos en lo que resta de esta sección que el vector objetivo $\boldsymbol{p}$ de (1.1) es entero y tiene entradas estrictamente positivas. Así también, supondremos que el lado derecho u de (1.1b) es entero. Como queremos que nuestro problema sea factible, supondremos que $u \geq 0$ a causa del teorema 1.2.8.

Sea $f_m(\hat{u})$ el valor óptimo de (1.1) cuando el lado derecho de (1.1b) es $\hat{u} \leq u$ con $\hat{u}$ entero y nos restringimos a las primeras m entradas de $\boldsymbol{p}$, con lo que obtenemos un problema de dimensión $1 \leq m \leq n$. Proponemos la formulación de programación dinámica

$$f_1(\hat{u}) \coloneqq \lfloor \hat{u}/p_1 \rfloor \, p_1, \tag{3.29a}$$

$$f_m(\hat{u}) \coloneqq \begin{cases} f_{m-1}(\hat{u}), & \hat{u} < p_m, \\ \max\{f_{m-1}(\hat{u}), f_m(\hat{u} - p_m) + p_m\}, & p_m \leq \hat{u} \leq u. \end{cases} \tag{3.29b}$$

Teorema 3.3.1. *Supongamos que el vector objetivo* $\boldsymbol{p}$ *del problema* (1.1) *es*

entero y tiene entradas estrictamente positivas, y también supongamos que el lado derecho u de (1.1b) *es entero y no negativo. Entonces $f_n(u)$ dado recursivamente por* (3.29) *es el valor óptimo de este problema.*

Demostración. Realizamos una inducción lexicográfica sobre el conjunto de tuplas $\{(m, \hat{u}) : 1 \leq m \leq n, 0 \leq \hat{u} \leq u\}$. Si $m = 1$, se sigue inmediatamente que (3.29a) es el valor óptimo para todo presupuesto $0 \leq \hat{u} \leq u$. Ahora bien, supongamos inductivamente que $f_{m-1}(\hat{u})$ es el valor óptimo para alguna $m > 1$ y para todo $0 \leq \hat{u} \leq u$. De (3.29b) junto con la hipótesis inductiva sobre m se sigue que, si $\hat{u} < p_m$, entonces $f_m(\hat{u}) = f_{m-1}(\hat{u})$ es óptimo. Ahora utilizamos un argumento fuertemente inductivo sobre $\hat{u}$ para mostrar que $f_m(\hat{u})$ es el valor óptimo para todo $0 \leq \hat{u} \leq u$, donde el caso base es la optimalidad para $\hat{u} < p_m$. Supongamos inductivamente que $f_m(\hat{u})$ es el valor óptimo para todo $0 \leq \hat{u} < \tilde{u} \leq u$, por lo que debemos mostrar que $f_m(\hat{u}+1)$ también es el valor óptimo. Por el caso base podemos suponer que $\hat{u} \geq p_m$. Observemos que $\hat{u}+1-p_m \leq \hat{u}$ pues $p_m > 0$, de donde se sigue que $f_m(\hat{u} + 1 - p_m) + p_m$ es óptimo por hipótesis fuertemente inductiva sobre $\hat{u}$, así también, $f_{m-1}(\hat{u} + 1)$ es óptimo por hipótesis inductiva sobre m. De (3.29b) y del hecho que $\hat{u} \geq p_m$, se sigue que $f_m(\hat{u} + 1)$ es el valor óptimo. Utilizando el caso base sobre $\hat{u}$ encontramos que $f_m(\hat{u})$ es el valor óptimo para todo $0 \leq \hat{u} \leq u$. Luego, del caso base sobre m, concluimos que $f_m(\hat{u})$ es el valor óptimo para todo $1 \leq m \leq n$ y para todo $0 \leq \hat{u} \leq u$. □

Observemos que para calcular $f_n(u)$, debemos calcular, en el peor de los casos, $f_m(\hat{u})$ para todo $1 \leq m \leq n$ y para todo $0 \leq \hat{u} \leq u$. Se desprende que el costo de calcular $f_n(u)$ es $\mathcal{O}(nu)$. No obstante, ambos R&A y los métodos diofantinos (`dioph_left` y `dioph_right`) terminan con la solución óptima y no solo con el valor óptimo. Para obtener comparaciones justas, nuestra implementación de esta FPD incluye una manera de construir la solución, a pesar del costo adicional en eficiencia, el cual es $\mathcal{O}(u)$.

3.3.3. Experimentos en la dimensión

Para obtener resultados informativos en cuanto varía la dimensión del problema (1.1), debemos ser cuidadosos para evitar tener soluciones triviales.

En primer lugar, observemos de (1.14) que si alguna entrada del vector coprimo $\boldsymbol{q}$ es tal que $q_j = 1$, entonces obtenemos la solución trivial

$$x_i^* \coloneqq \begin{cases} u & i = j, \\ 0 & i \neq j. \end{cases}$$

Esto se vuelve aún más trivial para nuestro método cuando ordenamos las entradas de $\boldsymbol{q}$ de manera ascendente. En términos del vector objetivo $\boldsymbol{p}$, esto se traduce a que no debe existir una entrada p_i tal que todas las entradas que le sigan sean múltiplos de p_i. De caso contrario, tendremos $g_{i+1} = p_i$ y por lo tanto $q_i = 1$.

En segundo lugar, es posible mostrar que todo problema (1.1) tiene una reducción al problema binario con dimensión $n_b \in \mathbb{N}$:

$$\max_{\boldsymbol{x} \in \{0,1\}^{n_b}} \{\hat{\boldsymbol{p}}^T \boldsymbol{x} : \hat{\boldsymbol{p}}^T \boldsymbol{x} \leq u\},$$

y para alguna $\hat{\boldsymbol{p}} \in \mathbb{Z}^{n_b}$ que puede ser obtenida de $\boldsymbol{p}$ (ve página 82 de [MT90]). Observemos que si $u \geq \sum_{i=1}^{n_b} \hat{p}_i$, entonces obtenemos la solución trivial $x_i = 1$ para todo $1 \leq i \leq n_b$. Puesto que existen *solvers* que implícitamente reducen problemas como (1.1) a su forma binaria (el ejemplo más famoso es `KnapsackSolver` de Google OR-Tools), debemos ser cuidadosos con no introducir este tipo de trivialidades. Una forma simple de evitar esta situación es exigir que el lado derecho de (1.1b) satisfaga $u < \sum_{i=1}^{n} p_i$.

En tercer lugar, tenemos que si el vector $\boldsymbol{p}$ contiene entradas repetidas, entonces podemos reducir trivialmente la dimensión del problema (1.1). En

efecto, si $p_j = p_\ell$, encontramos que

$$\sum_{i=1}^{n} p_i x_i = \sum_{\substack{i=1 \\ i \neq \ell}}^{n} p_i z_i,$$

donde $\boldsymbol{z} \in \mathbb{Z}^{n-1}$ está definida como

$$z_i \coloneqq \begin{cases} x_j + x_\ell, & i \in \{j, \ell\}, \\ x_i, & \text{e.o.c.}, \end{cases}$$

y el problema (1.1) es equivalente a

$$\max_{\boldsymbol{z} \in \mathbb{Z}^{n-1}} \{\hat{\boldsymbol{p}}^T \boldsymbol{z} : \hat{\boldsymbol{p}}^T \boldsymbol{z} \leq u, \boldsymbol{z} \geq \boldsymbol{0}\},$$

donde $\hat{\boldsymbol{p}} \in \mathbb{R}^{n-1}$ resulta de remover del vector $\boldsymbol{p}$ su ℓ-ésima entrada.

Así pues, sea $n > 2$ la dimensión del problema. Dejamos que $\boldsymbol{p} \in \mathbb{Z}^n$ sea un vector aleatorio tomado de una distribución uniforme discreta sobre $[10, 5n)^n$. El hecho de que el soporte de la distribución dependa de la dimensión reduce la probabilidad de que $\boldsymbol{p}$ tenga entradas repetidas. Al calcular el vector coprimo $\boldsymbol{q}$ checamos que $q_i \neq 1$ para todo $1 \leq i \leq n$. De caso contrario, calculamos otro vector aleatorio $\boldsymbol{p}$. Finalmente, escogemos el lado derecho de la restricción (1.1b) de manera que

$$u \coloneqq \begin{cases} 0.5 \sum_{i=1}^{n} p_i & n \leq 20{,}000, \\ 0.1 \sum_{i=1}^{n} p_i & n > 20{,}000, \end{cases}$$

para evitar obtener un problema trivial de acuerdo a la reducción binaria. Decidimos cambiar el coeficiente de 0.5 a 0.1 para disminuir la probabilidad de que existan múltiples soluciones óptimas en problemas de tamaño grande.

La tabla 3.1 en la página 102 muestra los tiempos promedios de terminación así como las desviaciones estándar de los métodos utilizados para realizar cada experimento. La tabla 3.2 en la página 103 muestra los coefi-

cientes de variación de estos tiempos. Un coeficiente de variación mide, en proporción, qué tan concentrados están los tiempos alrededor de su media, y se calcula como σ/μ, donde μ y σ son la media y desviación estándar de los tiempos medidos, respectivamente.

De manera más inmediata, tenemos que los métodos diofantinos son significativamente más rápidos. Además, los tiempos de terminación de los métodos diofantinos así como de la FPD están altamente concentrados alrededor de la media, por lo que son estables. En efecto, los coeficientes de variación de los métodos diofantinos estuvieron, en la gran mayoría de los casos, por debajo de 0.5 %, mientras que los de la FPD estuvieron por debajo de 1 %. En contraste, solo la mitad de los coeficientes de variación de CBC estuvo por debajo del 1 %.

A pesar de que la FPD sea, en términos prácticos, igual de estable que los métodos diofantinos, no debemos olvidar que la FPD no encontró una solución en el 61 % de los experimentos realizados.

Una de las hipótesis por las que creemos que `dioph_left` presenta resultados aún más rápidos que su contraparte `dioph_right` es la que sigue. Puesto que estamos generando vectores aleatorios con dimensiones grandes, la probabilidad de que las últimas $n - i$ entradas tengan factores en común es pequeña. Por lo tanto, obtenemos $g_{i+1} = 1$. Recordando que los coeficientes de Bézout x'_i, ω'_{i+1} satisfacen (1.26), encontramos que $(x'_i, \omega'_{i+1}) = (0, 1)$ y también debe ser el caso que $\frac{q_i}{\prod_{j=1}^{i} g_j} = 1$. Sustituyendo en (1.25) tenemos

$$\begin{cases} x_i = t_i, \\ \omega_{i+1} = \omega_i - t_i. \end{cases}$$

De esta manera, nuestro método `dioph_left` recorre primero soluciones pequeñas en magnitud comparadas a las obtenidas por `dioph_right`. Es decir, el primero es menos voraz que el segundo.

(a) Tamaños pequeños (milisegundos). Valores: media ± desviación estándar.

n	FPD	CBC	`dioph_left`	`dioph_right`
50	0.16 (±0.01)	23.29 (±17.05)	0.31 (±0.01)	0.31 (±0.01)
100	1.14 (±0.02)	9.57 (±1.01)	0.64 (±0.01)	0.66 (±0.01)
200	8.98 (±0.04)	23.01 (±4.06)	1.42 (±0.01)	1.44 (±0.01)
500	142.68 (±0.46)	126.82 (±19.39)	4.53 (±0.01)	4.67 (±0.02)

(b) Tamaños grandes (segundos). Valores: media ± desviación estándar.

n	FPD	CBC	`dioph_left`	`dioph_right`
1,000	1.27 (±0.01)	0.11 (±0.02)	0.01 (±0.00)	0.01 (±0.00)
2,000	10.37 (±0.02)	0.14 (±0.02)	0.04 (±0.00)	0.04 (±0.00)
5,000	160.56 (±0.39)	1.67 (±0.20)	0.25 (±0.00)	0.27 (±0.00)
10,000		2.12 (±0.33)	1.06 (±0.00)	1.16 (±0.00)
20,000		7.68 (±0.34)	4.60 (±0.01)	4.99 (±0.01)
30,000		19.19 (±1.18)	10.74 (±0.02)	11.55 (±0.01)
40,000		24.33 (±0.19)	19.47 (±0.02)	20.99 (±0.08)
50,000		38.94 (±0.18)	30.89 (±0.03)	33.45 (±0.04)
60,000		51.96 (±0.53)	45.07 (±0.05)	48.55 (±0.05)
70,000		81.03 (±1.26)	61.89 (±0.04)	66.35 (±0.10)
80,000		116.62 (±0.93)	81.42 (±0.06)	87.73 (±0.12)
90,000		141.68 (±2.10)	103.96 (±0.04)	111.55 (±0.12)
100,000		170.40 (±1.75)	129.44 (±0.05)	139.65 (±0.16)
150,000			299.15 (±0.13)	

Tabla 3.1: Tiempos de terminación promedio de los distintos métodos cuando varía la dimensión. Las casillas vacías indican que el método no encontró una solución para el problema correspondiente en menos de los 300 segundos de tolerancia.

n	FPD	CBC	`dioph_left`	`dioph_right`
50	0.091	0.741	0.019	0.029
100	0.020	0.107	0.008	0.009
200	0.005	0.179	0.004	0.005
500	0.003	0.155	0.003	0.004
1,000	0.005	0.145	0.001	0.002
2,000	0.002	0.164	0.001	0.001
5,000	0.002	0.119	0.004	0.003
10,000		0.159	0.003	0.002
20,000		0.045	0.001	0.002
30,000		0.062	0.002	0.001
40,000		0.008	0.001	0.004
50,000		0.005	0.001	0.001
60,000		0.010	0.001	0.001
70,000		0.016	0.001	0.002
80,000		0.008	0.001	0.001
90,000		0.015	0.000	0.001
100,000		0.010	0.000	0.001
150,000			0.000	

Tabla 3.2: Coeficientes de variación en los tiempos de terminación de acuerdo a la tabla 3.1. Las casillas vacías indican que el método no encontró una solución para el problema correspondiente en menos de los 300 segundos de tolerancia.

3.3.4. Experimentos en el presupuesto

Tomamos las mismas precauciones que en la subsección pasada para evitar soluciones triviales.

En este caso, fijamos la dimensión en $n = 1{,}000$ y dejamos que el vector esencialmente entero $\boldsymbol{p} \in \mathbb{Z}^n$ sea un vector aleatorio tomado de una distribución uniforme discreta y entera sobre el intervalo $[10, 10{,}000)^n$. Luego, tomamos 128 puntos enteros aproximadamente equidistantes sobre el intervalo $[.01S, S)$, donde $S \coloneqq \sum_{i=1}^{n} p_i$ y cada punto representa el lado derecho de la restricción presupuestaria (1.1b). Medimos los tiempos de terminación promedio sobre las 20 corridas. La figura 3.1 muestra estos tiempos de terminación, mientras que la figura 3.2 muestra la distribución de sus coeficientes de variación.

Como mostramos en la subsección 3.3.2, se ilustra gráficamente que los tiempos de terminación de la FPD crecen linealmente con respecto al presupuesto. Aún así, destaca el hecho de que la FPD resultó ser más consistente en sus tiempos de terminación que los otros tres métodos. Ambos FPD y `dioph_right` tuvieron coeficientes de variación menor que 2.5 %, al igual que `dioph_left` con excepción de dos experimentos anómalos.

En ambos CBC y los métodos diofantinos hay presencia de un comportamiento oscilatorio. Se confirmó para los métodos diofantinos que esto coincide en la mayoría de los casos cuando el presupuesto u es múltiplo de una de las entradas de $\boldsymbol{q}$. En estos casos el problema es trivial, pues si $q_j \mid u$, entonces uno de `dioph_left` o `dioph_right` encuentra inmediatamente la solución $\boldsymbol{x} = \lfloor u/q_j \rfloor \boldsymbol{e}_j$. En estos casos donde una entrada de $\boldsymbol{q}$ divide a u, se observó que un repunte en `dioph_left` coincide con una caída en `dioph_right` y viceversa. Las excepciones más notables a este comportamiento fueron las dos observaciones anómalas en `dioph_left` (ver figuras 3.1 y 3.2). El comportamiento oscilatorio en CBC no coincide con los casos

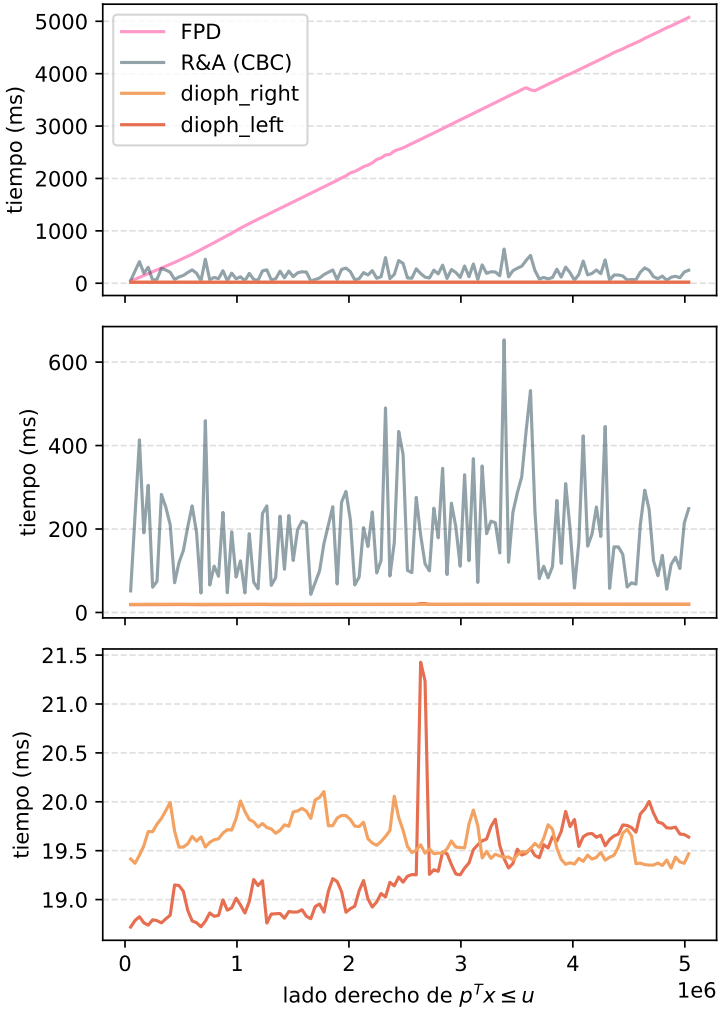

Figura 3.1: Tiempos de terminación promedio en milisegundos de los distintos métodos cuando varía el presupuesto. Para observar mejor los tiempos, eliminamos el método más lento y realizamos *zoom* a la imagen.

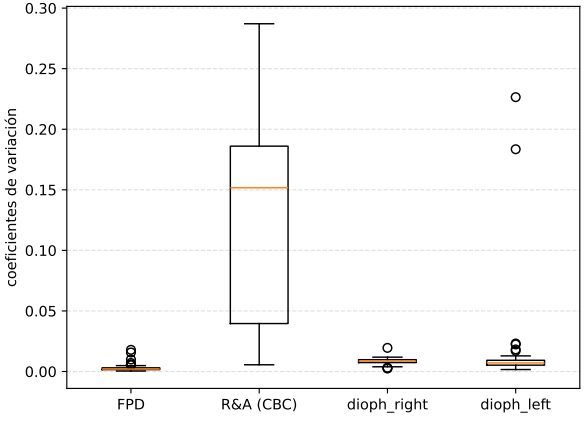

Figura 3.2: Coeficientes de variación en los tiempos de terminación a medida que el presupuesto varía.

en los que el presupuesto u es múltiplo de alguna entrada de $\boldsymbol{p}$.

Observamos en la figura 3.2 que la implementación de R&A en CBC es sumamente inestable. Se hipotetizó que esto se debe a la constante sobrecarga computacional de transferir datos entre los módulos del problema relajado y la generación de cortes dentro de CBC. Por ello, corrimos el mismo experimento tomando en cuenta una configuración de R&A que prohíba la generación de cortes. Debido a la dimensión $n = 1{,}000$ del problema y al fenómeno expuesto en la sección 2.1, no prohibimos el uso de heurísticas.

Se encontró que los tiempos de terminación son similares y por lo tanto no tiene sentido mostrarlos, pero sí hay un cambio significativo en la estabilidad de esta nueva configuración. La comparación de coeficientes de variación se muestra en la figura 3.3, en la cual se observa que la mediana de los coeficientes de variación es aproximadamente un tercio en la configuración sin cortes. Es decir, en el 50 % de los 128 experimentos que realizamos, las distribuciones de los tiempos de terminación están tres veces más con-

centradas alrededor de su media en la configuración sin cortes que en la configuración original. Aún así, el rango de ambas distribuciones es la misma, lo que sugiere que en ningún momento la generación de cortes ayudó a encontrar más rápidamente las soluciones óptimas. Al contrario, la generación de cortes provocó que CBC tuviera una mayor varianza relativa en los tiempos de terminación.

En conclusión, consideramos que el argumento a favor del uso de métodos diofantinos es fuerte. Mostramos en ambos experimentos de esta sección que los métodos diofantinos son más rápidos, más estables y significativamente más robustos ante cambios en la dimensión o en el presupuesto del problema (1.1) que sus contrapartes FPD o R&A implementado en CBC. El siguiente capítulo propone una extensión de estos métodos diofantinos para encontrar soluciones de programas lineales enteros generales.

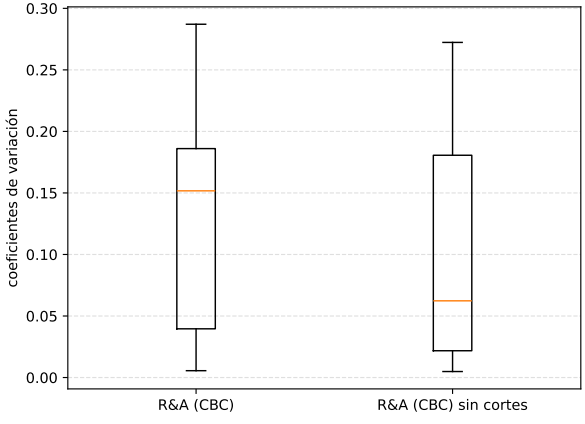

Figura 3.3: Comparación en la estabilidad de Ramificación y Acotamiento al prohibir la generación de cualquier tipo de corte.

Capítulo 4

Múltiples restricciones

En este último capítulo construimos un método que permite resolver programas lineales enteros generales. Mostramos que la complejidad exponencial de este tipo de programas se reduce a resolver sistemas de desigualdades lineales en los enteros. Además, encontramos una formulación alternativa a la manera tradicional de introducir programas lineales enteros que simplifica el árbol de subproblemas generado por el algoritmo de Ramificación y Acotamiento, lo cual podría resultar en mejores tiempos de terminación.

En la exposición de este capítulo dependemos en gran medida de las formas normales de Hermite y de Smith, las cuales son tratadas extensivamente en el capítulo 4 de [Sch98] y capítulo 2 de [New72], respectivamente.

Sea $\boldsymbol{p} \in \mathbb{R}^n$ esencialmente entero y sea $\boldsymbol{q} \in \mathbb{Z}^n$, de manera que $\boldsymbol{p} = m\boldsymbol{q}$ para algún escalar $m > 0$. Vimos en la subsección 1.2.2 que el problema (1.1) es equivalente a

$$\max_{\boldsymbol{x}\in\mathbb{Z}^n}\{\boldsymbol{q}^T\boldsymbol{x} \colon \boldsymbol{q}^T\boldsymbol{x} \in \{0,\dots,\eta\}, \boldsymbol{x} \geq \boldsymbol{0}\},$$

donde $\eta \in \mathbb{Z}$ está definida en el lema 1.2.7. Por ello, introducimos el problema con múltiples restricciones (4.1) en términos del vector $\boldsymbol{q}$ y el parámetro η en vez del vector $\boldsymbol{p}$ y el lado derecho u de (1.1b). Así pues, sea $A \in \mathbb{Q}^{m\times n}$ una matriz racional con renglones linealmente independientes y sea $\boldsymbol{b} \in \mathbb{Q}^m$

un vector. Consideremos el problema

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{Z}^n} \quad & \boldsymbol{q}^T\boldsymbol{x}, && \text{(4.1a)}\\ \text{s.a.} \quad & \boldsymbol{q}^T\boldsymbol{x} \leq \eta, && \text{(4.1b)}\\ & A\boldsymbol{x} = \boldsymbol{b}, && \text{(4.1c)}\\ & \boldsymbol{x} \geq \boldsymbol{0}. \end{aligned}$$

Supongamos que el problema (4.1) es factible. Debido a la restricción presupuestaria (4.1b), sabemos que la solución se encuentra en alguna capa entera $H_{\boldsymbol{q},k/\|\boldsymbol{q}\|^2}$ con parámetro entero $k \leq \eta$.

Teorema 4.0.1. *Sea $M \in \mathbb{Z}^{n\times(n-1)}$ la matriz definida en* (1.38) *y $\boldsymbol{\nu} \in \mathbb{Z}^n$ el vector definido en* (1.37)*. Entonces el problema* (4.1) *es equivalente al problema*

$$\begin{aligned} \max_{k\in\mathbb{Z},\boldsymbol{t}\in\mathbb{Z}^{n-1}} \quad & k, && \text{(4.2a)}\\ s.a. \quad & k \leq \eta, && \text{(4.2b)}\\ & AM\boldsymbol{t} = \boldsymbol{b} - kA\boldsymbol{\nu}, && \text{(4.2c)}\\ & M\boldsymbol{t} \geq -k\boldsymbol{\nu}. && \text{(4.2d)} \end{aligned}$$

Demostración. Por el teorema 1.2.16 sabemos que la transformación lineal

$$(k, \boldsymbol{t}) \mapsto \boldsymbol{x} \coloneqq k\boldsymbol{\nu} + M\boldsymbol{t}$$

es una descomposición de $\mathbb{Z}^n$ entre las redes $\Lambda_p \oplus \Lambda_h$ definidas en (1.42). Por consiguiente,

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} &\iff AM\boldsymbol{t} = \boldsymbol{b} - kA\boldsymbol{\nu},\\ \boldsymbol{x} \geq \boldsymbol{0} &\iff M\boldsymbol{t} \geq -k\boldsymbol{\nu}, \end{aligned}$$

y por lo tanto basta mostrar que si un vector es factible para un proble-

ma, entonces satisface la correspondiente restricción presupuestaria (4.1b) o (4.2b) del otro problema.

Sea $\boldsymbol{x} \in \mathbb{Z}^n$ un vector factible de (4.1), entonces existe $(k, \boldsymbol{t}) \in \mathbb{Z}^n$ que satisface $\boldsymbol{x} = k\boldsymbol{\nu} + M\boldsymbol{t}$. Por los lemas 1.2.12 y 1.2.13 encontramos que

$$k = \boldsymbol{q}^T \boldsymbol{x} \leq \eta,$$

y entonces $(k, \boldsymbol{t})$ es factible. Como $\boldsymbol{x}$ fue arbitrario, se sigue que la solución del problema (4.1) es una cota inferior del problema (4.2). La demostración de que la solución de (4.2) es una cota inferior de (4.1) es análoga.

Ahora mostramos por contradicción que podemos recuperar la solución de un problema a partir de la solución del otro. Supongamos que $(k, \boldsymbol{t}) \in \mathbb{Z}^n$ es solución de (4.2). Si existe $\tilde{\boldsymbol{x}}$ factible para (4.1) con utilidad $\boldsymbol{q}^T\tilde{\boldsymbol{x}} = \tilde{k}$ estrictamente mayor, entonces consideramos $(\tilde{k}, \tilde{\boldsymbol{t}})$ tal que $\tilde{\boldsymbol{x}} = \tilde{k}\boldsymbol{\nu} + M\tilde{\boldsymbol{t}}$. Este vector también es factible con utilidad $k < \tilde{k} \leq \eta$, y entonces $(k, \boldsymbol{t})$ no era la solución de (4.2). Obtenemos una contradicción. □

Este problema equivalente induce a que las políticas de poda en Ramificación y Acotamiento sean más eficientes. En efecto, la consecuencia principal del siguiente teorema es que siempre es mejor priorizar ramificaciones en k puesto que eliminamos de manera inmediata subproblemas infactibles.

Teorema 4.0.2. *Sea $(k^*_{PR}, \boldsymbol{t}^*_{PR})$ el óptimo del problema relajado de (4.2) y supongamos que k^*_{PR} no es entero. Entonces el subproblema generado al añadir la restricción $k \geq \lceil k^*_{PR} \rceil$ es infactible.*

Demostración. Supongamos que el subproblema es factible. Puesto que k^*_{PR} no es entero, existe $\tau \in \mathbb{Z}$ tal que $\tau - 1 < k^*_{\mathrm{PR}} < \tau$. Al añadir la restricción $k \geq \lceil k^*_{\mathrm{PR}} \rceil = \tau$ al problema (4.2), encontramos que el valor óptimo de este subproblema es estrictamente mayor que k^*_{PR}. Pero esto es una contradicción ya que en problemas de maximización el valor óptimo de un problema es una cota superior del valor óptimo de cualesquiera de sus subproblemas. □

A continuación desacoplamos el problema (4.2) en un subproblema de maximización y en otro de factibilidad. Supongamos, sin pérdida de generalidad, que las entradas de A y $\boldsymbol{b}$ son enteras. En el capítulo 2 de [New72] es introducida la forma normal de Hermite de la matriz A, la cual afirma que existe una matriz unimodular $U \in \mathbb{Z}^{n\times n}$ que satisface $AU = [H \mid \mathbf{0}]$, donde $H \in \mathbb{Z}^{m\times m}$ es triangular inferior y no singular.

Con esto en mente, introducimos el subproblema de (4.2) como

$$\max_{k\in\mathbb{Z},\tilde{\boldsymbol{y}}\in\mathbb{Z}^n} k, \tag{4.3a}$$

$$\text{s.a.} \quad k \leq \eta, \tag{4.3b}$$

$$A\tilde{\boldsymbol{y}} = \boldsymbol{b} - kA\boldsymbol{\nu}, \tag{4.3c}$$

donde

$$\tilde{\boldsymbol{y}} \coloneqq U \begin{pmatrix} \tilde{\boldsymbol{y}}_m \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = U_m \tilde{\boldsymbol{y}}_m + U_{n-m}\tilde{\boldsymbol{y}}_{n-m} \in \mathbb{Z}^n, \tag{4.4}$$

con $\tilde{\boldsymbol{y}}_m \in \mathbb{Z}^m$ y $\tilde{\boldsymbol{y}}_{n-m} \in \mathbb{Z}^{n-m}$. Denotamos por U_m y U_{n-m} las primeras m columnas y últimas $n-m$ columnas de U, respectivamente. Observemos que para toda $k \in \mathbb{Z}$ se cumple

$$AU \begin{pmatrix} H^{-1}(\boldsymbol{b} - kA\boldsymbol{\nu}) \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = [H, \mathbf{0}] \begin{pmatrix} H^{-1}(\boldsymbol{b} - kA\boldsymbol{\nu}) \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = \boldsymbol{b} - kA\boldsymbol{\nu}, \tag{4.5}$$

comparando con la restricción (4.3c) y la definición de $\tilde{\boldsymbol{y}}$ en (4.4), esto sugiere definir

$$\tilde{\boldsymbol{y}}_m \coloneqq H^{-1}(\boldsymbol{b} - kA\boldsymbol{\nu}). \tag{4.6}$$

No obstante, debemos asegurarnos que este vector sea entero. Observemos que $\tilde{\boldsymbol{y}}_{n-m}$ es un vector libre, así que en realidad este subproblema tiene dimensión $m+1$. Definimos el conjunto de factibilidad

$$\mathcal{F} \coloneqq \{k \in \mathbb{Z} \colon H^{-1}(\boldsymbol{b} - kA\boldsymbol{\nu}) \in \mathbb{Z}^m, k \leq \eta\} \tag{4.7}$$

Puesto que $H \in \mathbb{Z}^{m\times m}$ es no singular, para cada $k \in \mathbb{Z}$, existe una única solución $\tilde{\boldsymbol{y}}_m \in \mathbb{R}^m$ del sistema de ecuaciones $H\tilde{\boldsymbol{y}}_m = \boldsymbol{b} - kA\boldsymbol{\nu}$. Como, además, H es triangular inferior, podemos resolver rápidamente este sistema de ecuaciones y verificar si, para cada $k \leq \eta$ entera, la correspondiente solución $\tilde{\boldsymbol{y}}_m$ es entera o no.

Si $\mathcal{F}$ es vacío, deducimos que el subproblema (4.3) es infactible y por lo tanto (4.2) también lo es. Supongamos, pues, que $\mathcal{F} \neq \emptyset$. No es difícil observar que $\mathcal{F}$ tiene un elemento maximal k^* y que este elemento es la solución al subproblema (4.3). Luego, dada esta solución $k^* \in \mathbb{Z}$, buscamos resolver el subproblema de (4.2)

$$M\boldsymbol{t} = \tilde{\boldsymbol{y}}, \tag{4.8a}$$

$$M\boldsymbol{t} \geq -k^*\boldsymbol{\nu}. \tag{4.8b}$$

Tenemos un sistema de n ecuaciones lineales con $2n - m - 1$ incógnitas, por lo que tendremos que lidiar con $n - m - 1$ variables libres. En efecto, sustituyendo (4.4) en (4.8a), obtenemos

$$\begin{aligned} M\boldsymbol{t} = \tilde{\boldsymbol{y}} &= U_m\tilde{\boldsymbol{y}}_m + U_{n-m}\tilde{\boldsymbol{y}}_{n-m} \\ &\iff [M \mid -U_{n-m}] \begin{pmatrix} \boldsymbol{t} \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = U_m\tilde{\boldsymbol{y}}_m. \end{aligned} \tag{4.9}$$

En el capítulo 2 de [New72] también se introduce la forma normal de Smith, de la cual obtenemos dos matrices unimodulares $S \in \mathbb{Z}^{n\times n}$ y $T \in \mathbb{Z}^{(2n-m-1)\times(2n-m-1)}$ que satisfacen

$$S[M \mid -U_{n-m}]T = D \in \mathbb{Z}^{n\times(2n-m-1)},$$

donde D es una matriz diagonal cuyas n primeras entradas son distintas de cero y las restantes $n - m - 1$ son cero. Si multiplicamos S por la izquierda

en ambos lados de la ecuación (4.9), tenemos

$$DT^{-1}\begin{pmatrix} \boldsymbol{t} \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = SU_m\tilde{\boldsymbol{y}}_m =: \boldsymbol{z}. \tag{4.10}$$

Si D_{ii} no divide a z_i para alguna $i \in \{1, \ldots, n\}$, encontramos que la primera ecuación del subproblema (4.8) no tiene solución en los enteros, lo que implica que la elección de k^* fue la incorrecta para asegurar soluciones enteras a este subproblema. De ser este el caso, redefinimos nuestro conjunto de factibilidad $\mathcal{F}$ (ver (4.7)) como $\mathcal{F} \leftarrow \mathcal{F} \setminus \{k^*\}$. Si $\mathcal{F}$ ahora es vacío, entonces (4.2) es infactible, y en caso contrario escogemos el nuevo elemento maximal de $\mathcal{F}$ y repetimos el proceso.

Supongamos, pues, que $D_{ii} \mid z_i$ para todo $i \in \{1, \ldots, n\}$, por lo que obtenemos un vector $\boldsymbol{r} \in \mathbb{Z}^n$ de soluciones enteras con $r_i \coloneqq z_i/D_{ii}$ y otro vector $\boldsymbol{s} \in \mathbb{Z}^{n-m-1}$ de variables libres:

$$T^{-1}\begin{pmatrix} \boldsymbol{t} \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{s} \end{pmatrix}.$$

Por lo tanto, nuestro vector $\boldsymbol{t}$ es una función afín de $\boldsymbol{s}$, es decir, $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{s})$. En términos del problema original (4.1), hemos encontrado, hasta este punto, los vectores $\boldsymbol{x}(\boldsymbol{s}) \coloneqq k^*\boldsymbol{\nu} + M\boldsymbol{t}(\boldsymbol{s})$ que maximizan la utilidad y que satisfacen todas las restricciones excepto, posiblemente, las de no negatividad.

Consideremos el conjunto de vectores $\boldsymbol{s} \in \mathbb{Z}^{n-m-1}$ que obligan a que $\boldsymbol{t}(\boldsymbol{s})$ satisfaga (4.8b):

$$\mathcal{S} \coloneqq \{\boldsymbol{s} \in \mathbb{Z}^{n-m-1} : M\boldsymbol{t}(\boldsymbol{s}) \geq -k^*\boldsymbol{\nu}\}$$

Por un lado, es sabido que los programas enteros tales como (4.1) o (4.2) son problemas difíciles de resolver, a excepción de cuando la matriz de restricciones $A \in \mathbb{Z}^{m\times n}$ es totalmente unimodular. De manera superficial, decimos que un problema es difícil de resolver si no es conocida la existencia

de un algoritmo con complejidad polinomial que lo pueda resolver.

Por el otro lado, a lo largo de este capítulo hemos resuelto todos los problemas en tiempo polinomial. En efecto, obtener M y $\boldsymbol{\nu}$ de (1.38) y (1.37) se reduce a multiplicar números y calcular coeficientes de Bézout, al igual que máximos común divisores. En [Knu97], [Sch98] y [New72] se muestra que realizar este tipo de cálculos, así como de obtener las formas normales de Hermite y de Smith, son problemas acotados en tiempo polinomial.

Entonces, la única deducción posible es que el problema de determinar si el conjunto $\mathcal{S}$ es vacío, o cuántos elementos tiene, o cuáles son los elementos que contiene, son todos problemas difíciles de resolver. Esta complejidad se reduce drásticamente en dos casos especiales.

En primer lugar, si $m = n-1$, entonces no hay parámetros libres. De manera gráfica, el poliedro factible resultante es un semirrayo o un segmento de línea. Al momento de escoger la k^*-ésima capa entera, estamos agregando la ecuación $k = k^*$, con lo que obtenemos un sistema lineal entero de n ecuaciones con n incógnitas, y entonces la solución es única. Resta verificar que esta solución es entera y satisface (4.8b). Este caso se ilustra en el ejemplo 4.0.3. En segundo lugar, si $m = n - 2$, obtenemos un solo parámetro libre $s \in \mathbb{Z}$, con lo que podemos determinar rápidamente la existencia o inexistencia de un conjunto de factibilidad en s que obliga a que $\boldsymbol{t}(s)$ satisfaga (4.8b). Este caso se ilustra en el ejemplo 4.0.4.

A modo de resumen, mostramos en el pseudocódigo 6 en la página 115 la forma de resolver problemas del tipo (4.2). Por el teorema 4.0.1, este método también resuelve problemas del tipo (4.1). Después de presentar los ejemplos 4.0.3 y 4.0.4, mostramos una manera con la cual podemos deshacernos del ciclo infinito en la línea 7 de este pseudocódigo.

Pseudocódigo 6:

Datos: Vector coprimo $\boldsymbol{q} \in \mathbb{Z}^n$, $\eta \in \mathbb{Z}$, $A \in \mathbb{Z}^{m\times n}$ y $\boldsymbol{b} \in \mathbb{Z}^m$.

Resultado: Solución óptima $\boldsymbol{x}^*$ de (4.1).

inicio 1

- Calcular M y $\boldsymbol{\nu}$ de (1.38) y (1.37); 2
- Obtener U y H de la forma normal de Hermite de A; 3
- Particionar U en U_m y U_{n-m} tal que $[U_m \mid U_{n-m}] = U$; 4
- Obtener S y T de la forma normal de Smith de $[M \mid -U_{n-m}]$; 5
- $k \leftarrow \eta$; 6
- **mientras** $1 + 1 = 2$ **hacer** 7
  - Obtener $\tilde{\boldsymbol{y}}_m$ de $H\tilde{\boldsymbol{y}}_m = \boldsymbol{b} - kA\boldsymbol{\nu}$; 8
  - **si** $\tilde{\boldsymbol{y}}_m \in \mathbb{Z}^m$ **entonces** 9
    - ir a la línea 12; 10
  - $k \leftarrow k - 1$; 11
- $\boldsymbol{z} \leftarrow SU_m\tilde{\boldsymbol{y}}_m$; 12
- $\boldsymbol{r} \leftarrow \boldsymbol{0}_n$; 13
- **para** $i \leftarrow 1$ **a** n **hacer** 14
  - **si** D_{ii} no divide a z_i **entonces** 15
    - $k \leftarrow k - 1$; 16
    - ir a la línea 7; 17
  - $r_i \leftarrow z_i/D_{ii}$; 18
- $(\boldsymbol{t}(\boldsymbol{s}), \tilde{\boldsymbol{y}}_{n-m}(\boldsymbol{s})) \leftarrow T(\boldsymbol{r}, \boldsymbol{s})^T$; 19
- **si** existe $\boldsymbol{s}$ tal que $M\boldsymbol{t}(\boldsymbol{s}) \geq -k\boldsymbol{\nu}$ **entonces** 20
  - $\boldsymbol{x}^* \leftarrow k\boldsymbol{\nu} + M\boldsymbol{t}(\boldsymbol{s})$; 21
  - **devolver** $\boldsymbol{x}^*$; 22
- $k \leftarrow k - 1$; 23
- ir a la línea 7; 24

Para calcular las formas normales de Hermite y de Smith de la matriz de restricciones $A \in \mathbb{Z}^{m\times n}$ de los siguiente ejemplos, utilizamos la librería `hsnf` de Python[1].

Ejemplo 4.0.3. Consideremos el problema con $n = 2$ variables y $m = 1$ restricciones de igualdad

$$\begin{aligned} \text{máx } & x - y, \\ \text{s.a. } & x - y \leq 12, \\ & 3x + 5y = 25, \\ & x, y \geq 0. \end{aligned}$$

En este caso tenemos $A = (3, 5), \boldsymbol{b} = 25$, y también $\boldsymbol{q} = (1, -1)^T$, al igual que $\eta = 12$. De (1.37) y (1.38) calculamos

$$\boldsymbol{\nu} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad M = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

De la forma normal de Hermite de A tenemos

$$H = 1, \quad U = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix},$$

y de la forma normal de Smith de $[M \mid -U_{n-m}]$,

$$S = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 8 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 5 \\ 0 & 1 \end{pmatrix}.$$

Como $H = 1$, se sigue que $H^{-1}(\boldsymbol{b} - kA\boldsymbol{\nu}) = 25 - 3k$ es entero para todo $k \in \mathbb{Z}$. Así, el conjunto factible $\mathcal{F}$ definido en 4.7 está dado por

$$\mathcal{F} = \{k \in \mathbb{Z} : k \leq \eta = 12\}.$$

[1] Véase `https://hsnf.readthedocs.io/en/latest/index.html`.

Entonces escogemos $k^* = 12$ por ser el elemento maximal de $\mathcal{F}$. Luego,

$$\boldsymbol{z} \coloneqq SU_m\tilde{\boldsymbol{y}}_m = SU_m\left(H^{-1}(\boldsymbol{b} - k^*A\boldsymbol{\nu})\right) = \begin{pmatrix} 22 \\ -33 \end{pmatrix}.$$

Observemos que D_{22} no divide a z_2, y entonces el subproblema (4.8) no es factible para la elección de $k^* = 12$. Escogemos el segundo elemento de $\mathcal{F}$ más grande, con lo que tenemos $k^* = 11$. Siguiendo con el mismo procedimiento, encontramos ahora que $\boldsymbol{z} = (16, -24)$. En este caso, la diagonal de D sí divide, elemento a elemento, las entradas de $\boldsymbol{z}$, y entonces $\boldsymbol{r} = (16, -3)$. Puesto que $n - m - 1 = 0$, no hay variables libres. Tenemos de (4.10):

$$\begin{pmatrix} \boldsymbol{t} \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = T\begin{pmatrix} 16 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix},$$

y verificamos que se satisfaga (4.8b):

$$M\boldsymbol{t} + k^*\boldsymbol{\nu} = 1\begin{pmatrix} -1 \\ -1 \end{pmatrix} + 11\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 10 \\ -1 \end{pmatrix} \ngeq \mathbf{0}.$$

Ahora la elección de $k^* = 11$ dio un punto entero pero con una entrada negativa. Es decir, el subproblema (4.8) es infactible dada esta elección.

Repetimos este procedimiento hasta llegar a $k^* = 3$. En este caso encontramos que $(\boldsymbol{t}, \tilde{\boldsymbol{y}}_{n-m}) = (-2, 6)^T$. Por lo tanto,

$$M\boldsymbol{t} + k^*\boldsymbol{\nu} = -2\begin{pmatrix} -1 \\ -1 \end{pmatrix} + 3\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix} \geq \mathbf{0}.$$

Luego, $(k^*, \boldsymbol{t}) \coloneqq (3, -2)$ es el óptimo del programa (4.2). Por el teorema 4.0.1, concluimos que $(x^*, y^*) = (5, 2)$ es el óptimo de (4.1).

Ejemplo 4.0.4. Ahora consideremos el problema con $n = 3$ variables y $m = 1$ restricciones de igualdad

$$\begin{aligned} \text{máx } & x - y + 2z, \\ \text{s.a.} \quad & x - y + 2z \leq 10 \\ & 3x + 4y - z = 15 \\ & x, y, z \geq 0. \end{aligned}$$

En este caso tenemos $A = (3, 4, -1), \boldsymbol{b} = 15$, y también $\boldsymbol{q} = (1, -1, 2)^T$, al igual que $\eta = 10$. De (1.37) y (1.38) calculamos

$$\boldsymbol{\nu} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad M = \begin{pmatrix} 1 & 0 \\ -1 & 2 \\ -1 & 1 \end{pmatrix}.$$

De la forma normal de Hermite de A tenemos

$$H = 1, \quad U = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 4 & 3 \end{pmatrix},$$

y de la forma normal de Smith de $[M \mid -U_{n-m}]$,

$$S = \begin{pmatrix} 1 & 0 & 0 \\ -1 & -1 & 0 \\ 3 & 4 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 7 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Puesto que $H = 1$, tenemos de (4.7) que el conjunto factible es

$$\mathcal{F} = \{k \in \mathbb{Z} : k \leq \eta = 10\}.$$

Ahora bien, seguimos exactamente el mismo procedimiento que en el

ejemplo 4.0.3 hasta llegar a $k^* = 5$. Llegando a la línea 19 del pseudocódigo 6, encontramos que

$$T^{-1}\begin{pmatrix} \boldsymbol{t} \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ s \end{pmatrix} \implies \begin{pmatrix} \boldsymbol{t} \\ \tilde{\boldsymbol{y}}_{n-m} \end{pmatrix} = s\begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix},$$

donde $s \in \mathbb{Z}$ es la única variable libre. En este caso podemos determinar rápidamente un intervalo de existencia: tenemos $M\boldsymbol{t}(s) \geq -k^*\boldsymbol{\nu}$ si y solo si

$$s\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \geq \begin{pmatrix} -5 \\ 0 \\ 0 \end{pmatrix},$$

de donde se sigue inmediatamente que $-5 \leq s \leq 0$. Sustituyendo cada valor entero de s en $\boldsymbol{t}(s)$ y transformando a $\boldsymbol{x}^*(s) = k^*\boldsymbol{\nu} + M\boldsymbol{t}(s)$, tenemos que

$$\left\{ \begin{pmatrix} 0 \\ 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix} \right\}$$

son las seis soluciones del problema (4.1). Ciertamente, todas alcanzan un nivel de utilidad $k^* = 5$.

Si el problema (4.1) es factible, por la equivalencia del teorema 4.0.1, existe $k^* \in \mathbb{Z}$ que es el valor óptimo del problema (4.2), así que eventualmente saldremos del ciclo infito de la línea 7. En caso de que el problema (4.1) sea infactible, nada asegura, por el momento, que salgamos de este ciclo infinito. A continuación veremos cómo arreglar este problema, y en el proceso seremos capaces de eliminar la restricción presupuestaria (4.1b). Por lo tanto, en esta última parte, podremos encontrar soluciones a programas lineales enteros generales.

Sea $A \in \mathbb{Z}^{m\times n}$ una matriz con renglones linealmente independientes y sea $\boldsymbol{b} \in \mathbb{Z}^m$ un vector. Definamos el poliedro

$$P \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}.$$

Sea $\boldsymbol{q} \in \mathbb{Z}^n$ un vector coprimo y consideremos ambos problemas de maximización y minimización sobre este poliedro

$$\ell^* \coloneqq \min_{\boldsymbol{x}\in P}\{\boldsymbol{q}^T\boldsymbol{x}\}, \quad u^* \coloneqq \max_{\boldsymbol{x}\in P}\{\boldsymbol{q}^T\boldsymbol{x}\}, \tag{4.11}$$

y definamos

$$\tau \coloneqq \lceil \ell^* \rceil, \quad \eta \coloneqq \lfloor u^* \rfloor. \tag{4.12}$$

Observemos de (4.11) que siempre se cumple que $\tau \leq \eta$. Ciertamente, las restricción $\tau \leq \boldsymbol{q}^T\boldsymbol{x} \leq \eta$ es válida para el programa lineal entero $\max_{P\cap\mathbb{Z}^n}\{\boldsymbol{q}^T\boldsymbol{x}\}$ y, por lo tanto, este problema es equivalente a

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{Z}^n} \quad & \boldsymbol{q}^T\boldsymbol{x}, && \text{(4.13a)}\\ \text{s.a.} \quad & \tau \leq \boldsymbol{q}^T\boldsymbol{x} \leq \eta, && \text{(4.13b)}\\ & A\boldsymbol{x} = \boldsymbol{b}, && \text{(4.13c)}\\ & \boldsymbol{x} \geq \boldsymbol{0}, \end{aligned}$$

En primer lugar, si $\eta = \infty$, entonces el problema relajado es no acotado. En el teorema 1 de [BGH87] se demuestra que si esto ocurre y además las entradas de A, de $\boldsymbol{b}$ y de $\boldsymbol{q}$ son todas racionales, entonces el programa lineal entero o es infactible o es no acotado. Como tenemos integralidad de A y de $\boldsymbol{b}$ por suposición, así como integralidad de $\boldsymbol{q}$ por la definición 1.2.1, no tiene sentido usar el pseudocódigo 6 para buscar una solución de este problema, independientemente de que el problema sea infactible o no acotado.

En segundo lugar, si $\tau = -\infty$ y $\eta < \infty$, entonces el problema (4.13) es factible y representa exactamente el mismo problema que (4.1). En este

caso, como lo hemos discutido, siempre saldremos del ciclo infinito en la línea 7, por lo que el método delineado por el pseudocódigo 6 eventualmente terminará con una solución de este problema.

Finalmente, si $-\infty < \tau \leq \eta < \infty$ nos encontramos en la situación ideal. Esto se debe a que podemos reemplazar el ciclo en la línea 7 por algo del estilo "`para` $k \leftarrow \eta$ `a` τ `hacer`...". Es decir, sabemos exactamente cuántas capas enteras debemos recorrer para que el método delineado por el pseudocódigo 6 termine. Observemos que, en este caso, existe la posibilidad de que $P \neq \emptyset$ pero $P \cap \mathbb{Z}^n = \emptyset$. Sin realizar modificaciones grandes al pseudocódigo 6, encontramos que, o bien termina con una solución $\boldsymbol{x}^*$ del problema (4.13), o bien certifica en $\eta - \tau + 1$ pasos que este problema es infactible.

Cabe mencionar que en la subsección (1.1.2) indicamos que existen diversos algoritmos capaces de resolver rápidamente problemas lineales del estilo (4.11). Así pues, la parte de calcular los valores τ y η puede ser considerada como una parte de preprocesamiento. Recordemos que el método de Ramificación y Acotamiento, en el peor de los casos, necesita resolver un número exponencial de problemas relajados de (4.13). Nuestro método, en cambio, solo necesita resolver dos problemas relajados.

A modo de conclusión de este capítulo, mencionamos que futuras líneas de investigación podrían estar concentradas en resolver el problema de la línea 20 del pseudocódigo 6. Esto se reduce a investigar sistemas de desigualdades lineales en los enteros. Existen tres posibilidades para estas investigaciones con respecto al vector de variables libres $\boldsymbol{s} \in \mathbb{Z}^{n-m-1}$:

1. Decidir la existencia de este vector: si bien no podríamos obtener la solución entera $\boldsymbol{x}^*$, sí podríamos concluir que k^* es el valor óptimo de (4.2) y, por el lema 1.2.12 así como del teorema 4.0.1, también es el valor óptimo de (4.1).

2. En caso de tener existencia, determinar el número de estos vectores:

además de saber que k^* es el valor óptimo de (4.1), también conoceríamos el número de soluciones que tiene este problema.

3. En caso de tener existencia, calcular todos estos vectores: además de saber que k^* es el óptimo de (4.1) y de conocer cuántas soluciones tiene este problema, conoceríamos también cuáles son esas soluciones.

Otra posible futura línea de investigación, más aplicada pero no por ello menos interesante, es desarrollar las consecuencias del teorema 4.0.2. Es nuestra creencia que los tiempos de terminación de Ramificación y Acotamiento usando la formulación equivalente (4.2) serán menores que usando la formulación tradicional (4.1). Para lograr esto, necesitaremos calcular rápida y eficientemente la matriz $M \in \mathbb{Z}^{n\times(n-1)}$ y el vector $\boldsymbol{\nu} \in \mathbb{Z}^n$ definidos en (1.38) y (1.37), respectivamente.

Capítulo 5

Conclusiones

Durante el trabajo de tesis se obtuvieron los siguientes resultados teóricos, los cuales son de carácter original y fueron desarrollados por el autor.

En los teoremas 1.2.8 y 1.2.9 se simplifica y estructura el análisis del problema (1.1). A partir de ellos podemos, en primer lugar, deshacernos automáticamente de instancias infactibles y, en segundo lugar, de separar en casos las instancias factibles.

En la proposición 1.2.11 se muestra una relación lineal entre el vector solución $\boldsymbol{x} \in \mathbb{Z}^n$ de la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$ con un vector de variables libres $\boldsymbol{t} \in \mathbb{Z}^{n-1}$. Esta proposición da entrada para analizar propiedades de la matriz $M \in \mathbb{Z}^{n\times(n-1)}$ y del vector $\boldsymbol{\nu} \in \mathbb{Z}^n$ definidos en (1.37) y (1.38), respectivamente. En el teorema 1.2.16 aprovechamos estas propiedades para descomponer la red $\mathbb{Z}^n$ como la suma directa de dos subredes Λ_p y Λ_h (ver (1.42)) que contienen, respectivamente, soluciones particulares y soluciones homogéneas de la ecuación lineal diofantina $\boldsymbol{q}^T\boldsymbol{x} = k$. En el teorema 1.2.22 se sugiere que esta descomposición no es exclusivamente generada por $\boldsymbol{q}$, sino que lo es por su órbita. Esto último permite que consideremos una clasificación de programas lineales enteros a partir de los vectores coprimos asociados a un vector esencialmente entero $\boldsymbol{p}$.

En los teoremas 2.0.4 y 3.2.2, así como sus respectivos algoritmos 3 y

5 mostramos cómo las ecuaciones lineales diofantinas son esenciales para resolver instancias de (1.1). Además, en el caso infinito, tenemos que la complejidad para resolver este tipo de instancias es polinomialmente acotada. La dificultad radica en instancias cuando todas las entradas del vector coprimo $\boldsymbol{q}$ son estrictamente positivas. A pesar de ello, en el teorema 3.1.19 se indica que esta complejidad no depende del número de ecuaciones lineales diofantinas a resolver (pues eventualmente es suficiente con resolver una), sino que radica en cómo se resuelven estas ecuaciones.

En el teorema 3.1.18 se muestra que el lado derecho de (3.22) es una cota superior para el número de Frobenius F. En realidad, lo que obtenemos es una familia de cotas superiores, pues F está en función del número de enteros coprimos $q_1, \ldots, q_n$, así como de sus valores respectivos, es decir, $F = F(q_1, \ldots, q_n; n)$, donde $n \in \mathbb{Z}_{>0}$. Puesto que podemos calcular la matriz M definida en (1.38) en tiempo polinomial, entonces podemos calcular en tiempo polinomial la cota superior (3.22) de $F(q_1, \ldots, q_n; n)$ para $n \in \mathbb{Z}_{>0}$ fija y para cualesquiera $q_1, \ldots, q_n$.

Finalmente, en el teorema 4.0.1 se demuestra que la formulación (4.2) es equivalente al problema (4.1). En el teorema 4.0.2 se muestra que el algoritmo de Ramificación y Acotamiento podría beneficiarse al utilizar esta formulación equivalente, pues podemos priorizar ramificaciones en k para deshacernos rápidamente de subproblemas infactibles.

Ahora presentamos problemas abiertos que fueron descubiertos a lo largo de esta tesis y que podrían ser de interés para futuras líneas de investigación:

1. En el ejemplo 2.1.1 mostramos para una instancia particular que Ramificación y Acotamiento genera una sucesión de subproblemas trasladados. Mostrar o refutar que existe una clase instancias de (1.1) que contienen subproblemas homotéticos. En caso afirmativo, mostrar o refutar que este conjunto de subproblemas es infinito cuando el vector coprimo $\boldsymbol{q}$ tiene una entrada negativa.

2. Generalizar el lema 3.1.2 para m racional.

3. Para encontrar la cota inferior en el lado derecho de (3.22) tuvimos que calcular el radio de la bola inscrita en el símplice σ y centrada en el baricentro $\hat{\boldsymbol{\sigma}}$. Es cierto que podemos obtener distintas cotas inferiores si centramos la bola inscrita en σ en distintos puntos de este símplice. Mostrar o refutar que el baricentro $\hat{\boldsymbol{\sigma}}$ genera la menor de estas cotas.

4. Realizar un análisis detallado de la cota superior dada en el lado derecho de (3.22) para el número de Frobenius $F(q_1, \ldots, q_n; n)$ y compararla con las cotas establecidas en el capítulo 3 de [RA05].

5. Construir un algoritmo que resuelva uno de los tres problemas descritos en la conclusión del capítulo 4.

6. Realizar experimentos numéricos que comparen los tiempos de terminación de Ramificación y Acotamiento al utilizar la formulación (4.1) contra su forma equivalente (4.2).

> Would it save you a lot of time if I just gave up and went mad now?
>
> *Douglas Adams, The Hitchhiker's Guide to the Galaxy*

Bibliografía

[BGH87] R. H. Byrd, A. J. Goldman, and Miriam Heller, *Recognizing unbounded integer programs*, Operations Research **35** (1987), no. 1, 140–142.

[BH09] Robert F. Bodi and Katrin Herr, *Symmetries in integer programs*, arXiv: Combinatorics (2009).

[BV04] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[Knu97] Donald E Knuth, *The art of computer programming: Fundamental algorithms, volume 1*, Addison-Wesley Professional, 1997.

[Knu98] ———, *The art of computer programming: Seminumerical algorithms, volume 2*, Addison-Wesley Professional, 1998.

[Lav14] Carmen Gómez Laveaga, *Álgebra superior: Curso completo*, primera edición ed., Programa Universitario del Libro de Texto, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, México, 2014 (spanish), Primera reimpresión: julio de 2015.

[Lou03] Robin Lougee, *The common optimization interface for operations research: Promoting open-source software in the operations research community*, IBM Journal of Research and Development **47** (2003), 57 – 66.

[MJSS16] David Morrison, Sheldon Jacobson, Jason Sauppe, and Edward Sewell, *Branch-and-bound algorithms: Recent advances in searching, branching, and pruning*, Discrete Optimization **19** (2016), 79–102.

[MT90] Silvano Martello and Paolo Toth, *Knapsack problems: algorithms and computer implementations*, John Wiley & Sons, Inc., USA, 1990.

[New72] Morris Newman, *Integral matrices*, Pure and Applied Mathematics, vol. 45, Academic Press, New York, 1972.

[NW06] Jorge Nocedal and Stephen J. Wright, *Numerical Optimization*, 2 ed., Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.

[Oli17] Fabricio Oliveira, *Linear optimisation notes*, `https://github.com/gamma-opt/linopt-notes`, 2017.

[RA05] Jorge L. Ramírez Alfonsín, *The diophantine frobenius problem*, Oxford University Press, 12 2005.

[Sch98] Alexander Schrijver, *Theory of linear and integer programming*, John Wiley & Sons, Chichester, UK, 1998.