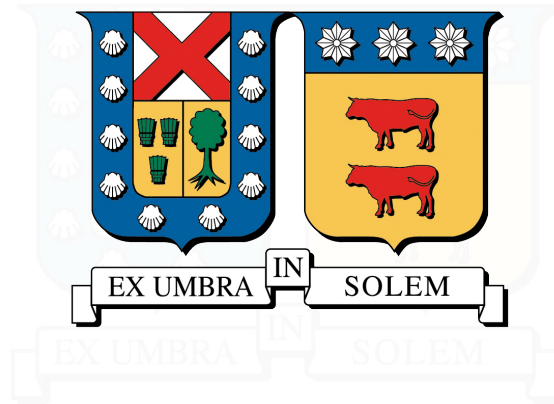


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS
VALPARAÍSO - CHILE



**ESTUDIO DE LOS PAGOS ADICIONALES EN MERCADOS ELÉCTRICOS A
TRAVÉS DE UNA FORMULACION BASADA EN REDES DEL PROBLEMA
CONVEX HULL PRICING**

CRISTIAN ANGELO ÁLVAREZ CÁRDENAS

TESIS PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
Y AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA INDUSTRIAL

PROFESOR GUÍA : SR. ALEJANDRO ANGULO C.
PROFESOR CORREFERENTE : SR. PABLO ESCALONA R.

JULIO 2019



A mis padres, Arturo y Marcia...

Agradecimientos

Agradezco a mis profesores guía Alejandro Angulo y Pablo Escalona, principalmente por la orientación en el desarrollo de este trabajo, pero también por la enseñanza que va más allá del trabajo de tesis, sin duda es una formación valiosa para mi futuro.

También quisiera agradecer a Conycit, que por medio del proyecto basal FB0008 “Advanced Center for Electrical and Electronic Engineering” y el proyecto Fondecyt N° 11170229, colaboró con el desarrollo de este trabajo.

No puedo dejar de mencionar en este agradecimiento a los amigos de la oficina por su constante disposición a ayudarme y a los profesores del área de operaciones por haberme dado la posibilidad de colaborar con ellos.

Agradezco a mis grandes amigos, mención especial a Pacman y Pedro porque estuvieron en todas, pero también a los muchachos del *ludus* y los muchachos de la *totita*, todos ellos un importante apoyo emocional en mi vida.

Finalmente, quisiera agradecer a toda mi familia y principalmente a mis padres, Arturo y Marcia. Les agradezco por el constante apoyo en mi educación, por la confianza y tantas otras cosas. Sin ellos, no podría haber alcanzado esta etapa de la vida.

Resumen ejecutivo

Este trabajo estudia los pagos adicionales en mercados mayoristas de electricidad, utilizando el método Convex Hull Pricing para determinar precios. El problema de precios es reformulado utilizando redes de flujo para representar la región de operación factible de las unidades, lo que implica que las restricciones asociadas definen un poliedro con la propiedad de integralidad. La formulación basada en redes entrega flexibilidad en el modelamiento de características de las unidades y permite obtener los precios resolviendo un problema en programación lineal. Para resolver el problema se utiliza un enfoque primal-dual basado en el algoritmo desarrollado por Bienstock y Zuckerberg. El algoritmo, junto a las técnicas de preproceso e inicialización implementadas, permiten alcanzar tiempos de solución menores a lo obtenido con algoritmos de optimizadores comerciales, tales como dual simplex y barrier. Por otro lado, los resultados económicos sugieren que la formulación propuesta obtiene el pago adicional mínimo incluso cuando las unidades tienen costos de partida dependientes del tiempo, lo cual hace que nuestro enfoque sea más robusto que la mejor formulación compacta encontrada en la literatura. Este trabajo también analiza el efecto de precios subóptimos sobre los pagos adicionales relajando el criterio de optimalidad del algoritmo implementado, observando el mayor impacto sobre el pago relacionado al costo de oportunidad perdido. Finalmente, la formulación propuesta puede ser clasificada como un problema de redes con restricciones adicionales y de esta forma abarcar otras aplicaciones, para lo cual el estudio computacional realizado entrega resultados de contraste detallados.

Palabras Clave. Mercado eléctrico, envoltura convexa, formulación de red, programación lineal, relajación Lagrangiana.

Abstract

This paper studies the convex hull pricing problem in electricity markets through a network-flow-based formulation. The network represents the feasible operating region of a generating unit, and the associated flow constraints define a polyhedron with an integrality property. These facts provide modeling flexibility with respect to the inclusion of unit features, and allow to obtain convex hull prices from a linear programming problem. The formulation is solved using a primal-dual approach based on the algorithm developed by Bienstock and Zuckerberg. The algorithm, together with the implemented pre-processing and initialization techniques, allows to achieve lower solution times than those obtained by state-of-the-art algorithms available in commercial solvers, e.g., barrier and dual simplex. Furthermore, results suggest that the proposed formulation obtains the minimum uplift payments even when time-dependent start-up costs are included, making the approach more robust than the best documented compact formulation. This paper also discusses the effect of sub-optimal prices on uplift payments by relaxing the optimality criterion of the algorithm, observing a significant impact on lost opportunity costs.

Keywords. Electricity market; convex hull; network-flow formulation; linear programming; Lagrangian relaxation.

Índice de Contenidos

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	3
1.3. Alcance	3
1.4. Estructura	4
2. Marco teórico	5
2.1. Optimización en redes	5
2.1.1. Problema de flujo en redes	5
2.1.1.1. Problema de ruta mínima	6
2.1.1.2. Restricciones adicionales	8
2.1.2. Problemas lineales de gran escala	9
2.1.2.1. Métodos de solución	9
2.2. Equilibrio competitivo	12
2.2.1. Mercados no-convexos	13
2.3. Mercados eléctricos	14
2.3.1. Entidades	15
2.3.2. Modelos de mercado	15
2.3.3. Tipos de producto	16
2.3.4. Mercado mayorista	16
2.3.5. Política de precios	17
2.3.5.1. Pagos adicionales	19
2.3.5.2. Precios marginales	19
2.4. Convex Hull Pricing	20
2.4.1. Modelo Unit Commitment	20
2.4.2. Cálculo de precios	22
2.4.3. Enfoque propuesto	24

3. Metodología	28
3.1. Modelo de red	28
3.1.1. Formulación del modelo	30
3.2. Algoritmo BZ	33
3.2.1. Preproceso de la red	35
3.2.2. Inicialización	36
3.2.3. Particiones	37
4. Resultados	40
4.1. Análisis de algoritmo BZ	40
4.1.1. Preproceso e inicialización	40
4.1.2. Iteraciones	42
4.2. Análisis comparativo	43
4.3. Estudio económico	45
4.3.1. Precios subóptimos	47
5. Conclusiones	49
5.1. Trabajos futuros	50
6. Anexo	51
Bibliografía	60

Índice de Tablas

2.1. Notación para formulación de UCP	21
2.2. Resultados de ejemplo por unidad	26
2.3. Resultados de ejemplo a nivel agregado	26
3.1. Parámetros de red para una unidad	29
3.2. Comparación de formulaciones en tiempo de solución	32
3.3. Efecto de preproceso e inicialización	37
4.1. Comparación de algoritmos según rendimiento computacional	45

Índice de Figuras

1.1. Costo marginal en barra Diego de Almagro	2
2.1. Envoltura convexa de la función de costos de ejemplo	25
3.1. Ejemplo de red para una unidad	28
3.2. Red preprocesada según estado inicial	36
3.3. Particiones en iteración k-ésima	38
3.4. Solución del problema relajado en una red	38
3.5. Particiones en iteración (k+1)-ésima	39
4.1. Ratios de tiempo de solución según técnica de preproceso	41
4.2. Tiempos de solución según tipo de inicialización	42
4.3. Número de iteraciones por instancia y horizonte de tiempo	42
4.4. Perfiles de rendimiento de los algoritmos	44
4.5. Pagos adicionales de las formulaciones	46
4.6. Diferencias de precio respecto a optimalidad	47
4.7. Ratio de pagos adicionales	48

1 | Introducción

1.1. Motivación

En la mayoría de los mercados eléctricos, rige un modelo de precios basado en el costo marginal del sistema. Esto implica que el precio de la energía en un instante de tiempo determinado, equivale al costo variable de aquella unidad de generación que puede aumentar o disminuir su producción para equilibrar oferta y demanda de energía. En resumen, el precio es el costo de suplir una unidad adicional de energía, o equivalentemente, costo marginal del sistema. Este precio resulta en ingresos adecuados si el costo marginal del sistema está determinado por la unidad más cara que se encuentra en funcionamiento, pero desafortunadamente lo anterior no siempre se puede garantizar debido a las restricciones operacionales de las unidades de generación que componen un sistema eléctrico. Existen situaciones en que algunas unidades no pueden ser apagadas y deben funcionar a mínimo técnico sin marcar el costo marginal del sistema, por lo tanto los ingresos obtenidos por la energía generada no son suficientes para cubrir los costos de operación [Vergara \(2016\)](#). Ante esta situación, la regulación de los mercados eléctricos establece un pago compensatorio o pago adicional que permite cubrir las pérdidas de aquellas unidades afectadas.

Esta política ha motivado la discusión del efecto de los pagos adicionales sobre los incentivos a la inversión y la competitividad en el mercado. Parte de la literatura señala que este tipo de pagos son una distorsión en el mercado, debido a la falta de transparencia y el carácter discriminatorio. Por otro lado, existe preocupación por el aumento observado en los pagos adicionales [Sauer \(2014\)](#), en parte debido a la inclusión de energía renovable no convencional (ERNC) en la matriz energética. Las fuentes de ERNC cuentan entre sus ventajas costos variables bajos o iguales a cero, y entre sus desventajas, la intermitencia e incertidumbre del suministro. Ejemplo de esto se refleja en la figura 1.1, donde se muestran los costos marginales horarios de una barra en el norte del sistema eléctrico chileno¹.

¹Información recuperada de https://cmg-sic.coordinador.cl/Modulos/CMg/CDEC_CMgBarras.aspx

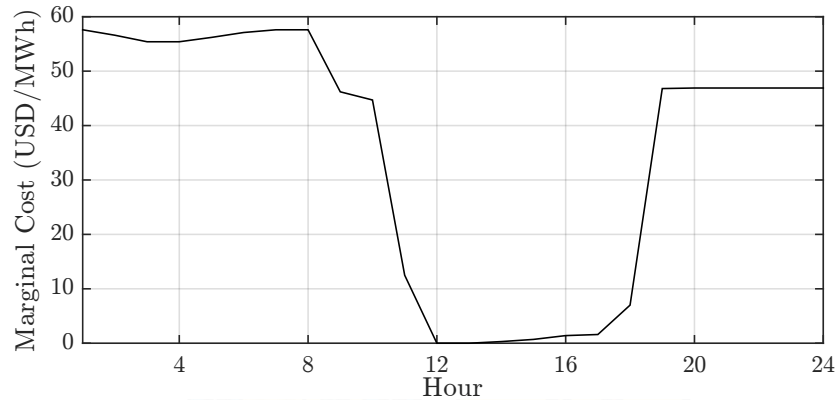


Figura 1.1: Costo marginal en barra Diego de Almagro, 26 de Abril de 2018.
(Fuente: elaboración propia.)

Existen al menos dos consecuencias que surgen a partir de situaciones como la mostrada en la figura 1.1. En primer lugar, la inclusión de ERNC ha demandado una mayor flexibilidad de las unidades convencionales, principalmente en decisiones de encendido y apagado que involucran costos fijos no considerados cuando la operación es en carga base, i.e., cuando el funcionamiento es continuo. En segundo lugar, y como se establece previamente, aquellas unidades que no pueden apagarse debido a restricciones técnicas, deben permanecer generando energía en horarios donde el precio va a estar por debajo de sus costos variables. El resultado final es que gran parte de la remuneración recibida por algunas unidades corresponde a un pago adicional. A medida que el pago por energía disminuye en la remuneración, la señal de precios pierde importancia como indicador para guiar las decisiones de participación en el mercado. Por lo tanto, una política de precios eficiente debería buscar reducir o eliminar los pagos adicionales.

Una de las alternativas propuestas para reducir un tipo de pago adicional es fijar precios con el método Convex Hull Pricing [Hogan y Ring \(2003\)](#). Para entender esta metodología, hay que saber que la curva de costo total mínimo de generación en función del nivel de demanda es no-convexa, y esto a su vez es la causa de los pagos adicionales. Para cierto valor de la demanda, el método de precios marginales equivale a la pendiente de la curva de costo en ese punto. Bajo el método Convex Hull Pricing, el precio es la pendiente de la envoltura convexa de la curva de costo en ese punto. Por lo tanto, se trata de un precio mayor, que incluye parte de los costos fijos de las unidades del sistema y que para ciertos valores de demanda es capaz de eliminar el pago adicional. Además, se obtienen precios no-decrecientes respecto al nivel de demanda, situación que no ocurre con el esquema marginalista. En la última década, el método Convex Hull Pricing ha adquirido amplia consideración en la literatura, pero no así a nivel práctico. Aún presenta un desafío desde el punto de vista de implementación: por una parte existen métodos computacionalmente complejos pero capaces de representar un buen modelo del sistema eléctrico, y por otro lado existen métodos sencillos en resolución pero con limitantes a nivel de modelamiento.

Este trabajo de tesis, busca contribuir a nivel de modelamiento y enfoque de solución del problema Convex Hull Pricing. La alternativa presentada muestra flexibilidad y rapidez, características que pueden considerarse importantes en el plano práctico. El trabajo en su totalidad es un aporte a la discusión del modelo de precios en el mercado eléctrico.

1.2. Objetivos

El objetivo principal es estudiar los resultados de precios y pagos adicionales en el mercado eléctrico con una nueva formulación del problema Convex Hull Pricing. La idea es mostrar que si bien esta formulación alternativa es más extensa que las formulaciones actuales, es capaz de obtener los pagos adicionales mínimos y además resulta flexible para la inclusión de características. Por otro lado, los objetivos específicos son los siguientes:

- Mostrar que la formulación propuesta puede ser resuelta en tiempos comparables a los enfoques actuales mediante el uso de un algoritmo especializado.
- Evaluar el desempeño del algoritmo especializado respecto a los algoritmos disponibles en un solver estándar.
- Comparar la formulación propuesta con la última formulación del problema Convex Hull Pricing, respecto a tiempos de solución, precios y pagos adicionales.
- Estudiar el efecto de precios subóptimos sobre los pagos adicionales obtenidos con la metodología Convex Hull Pricing.

1.3. Alcance

En primer lugar es necesario aclarar el contexto de aplicación de este trabajo de tesis. El escenario corresponde a un mercado mayorista eléctrico, donde el funcionamiento de las unidades de generación es programado por un operador independiente para cumplir al mínimo costo la demanda de energía pronosticada. Es importante señalar que sólo se va a considerar un nodo de demanda y por lo tanto se omite cualquier análisis relacionado a la transmisión de electricidad. El problema estudiado consiste en obtener precios de la energía, y con esto es posible evaluar las desviaciones entre la remuneración real basada en el funcionamiento ordenado por el operador y la remuneración “ideal” basada en un funcionamiento totalmente descentralizado. Este trabajo de tesis es principalmente exploratorio, debido a que desarrolla una alternativa de solución al problema de determinar precios en el mercado eléctrico usando el método Convex Hull Pricing. La contribución se apoya tanto en la formulación del problema como en la forma de resolverlo, proponiendo mejoras respecto a los enfoques actuales.

1.4. Estructura

Este trabajo de tesis se organiza en capítulos independientes, los cuales a su vez se organizan en secciones. El capítulo 2 presenta el marco teórico de la investigación y se divide en 4 secciones. La primera sección revisa el algoritmo de ruta mínima y resume distintos métodos genéricos para resolver formulaciones similares a la propuesta. La segunda sección explica el concepto de equilibrio competitivo en mercados no-convexos. La tercera sección revisa las principales componentes del mercado eléctrico para definir el escenario de estudio, mientras que la cuarta sección presenta el problema Convex Hull Pricing y un caso de ejemplo. El capítulo 3 muestra la metodología usada, la cual consiste en construir una formulación basada en redes e implementar un algoritmo especializado basado en el trabajo de Bienstock y Zuckerberg. El capítulo 4 presenta los resultados obtenidos y el capítulo 5 resume las conclusiones obtenidas, además de sugerir posibles direcciones para trabajos futuros. Finalmente, se adjunta como anexo el documento resumen de la tesis, el cual fue enviado para publicación a una revista de divulgación científica.

2 | Marco teórico

2.1. Optimización en redes

En este trabajo, el problema de interés se formula como un problema de red con restricciones adicionales, conocidas en la literatura como *side constraints*. Específicamente, se trata de múltiples problemas de ruta mínima acoplados por un conjunto de restricciones sistémicas. El objetivo es resolver la relajación continua, es decir, ignorar la integralidad de las variables en los problemas de ruta mínima acoplados, para obtener las variables duales asociadas con las restricciones adicionales, las cuales tienen la interpretación de precios en el contexto estudiado. Si bien se trata de un problema de programación lineal (LP, *linear programming*), el número de elementos considerados en el modelamiento deriva en un problema que puede considerarse de gran escala. En este capítulo se resumen algunas propiedades de los problemas de redes, se presenta la estructura del problema estudiado y se mencionan algunos métodos para resolver problemas LP de gran escala, además de introducir el método de solución implementado en este trabajo.

2.1.1. Problema de flujo en redes

Los problemas de flujo en redes son una clase de problemas de optimización ampliamente utilizada para el diseño de sistemas de gran escala, como en comunicaciones, transporte y manufactura. También, son una forma sencilla de representar algunos problemas combinatoriales. Un modelo basado en redes, utiliza nodos y arcos para representar la dinámica del sistema. El objetivo de modelar con redes es utilizar algoritmos especializados para resolver los cuales tienden a ser más eficiente que la resolución directa en un optimizador estándar Bertsekas (1998).

Dentro de los problemas de redes, se encuentran los problemas de flujo de costo mínimo, los cuales consisten en enviar una cantidad desde uno o más nodos de origen hacia uno o más nodos de destino a través de una red. En esta categoría entran los problemas de ruta mínima, flujo máximo, asignación y transporte, como también las variaciones de éstos incluyendo funciones de costos convexas, múltiples commodities en el transporte y restricciones adicionales. Otra clase de problemas corresponde a la optimización discreta en

redes, donde la cantidad transferida a través de la red debe tomar un único valor en un conjunto discreto de valores. Esta última clase de problemas es mucho más compleja que la anterior, y usualmente se resuelve de manera aproximada utilizando rutinas de los problemas de flujo de costo mínimo.

Asumir, sin pérdida de generalidad, que un problema de optimización en redes con función de costo lineal puede escribirse de la siguiente forma:

$$z = \text{mín} \{c^T x : x \in \mathbb{Z}^n, Ax \geq b, 0 \leq x \leq u\} \quad (2.1)$$

Donde c es el vector de costos, A es la matriz de coeficientes de las restricciones de flujo o restricciones de divergencia, el vector b es un vector de coeficientes enteros positivos que define los flujos en los extremos de la red, mientras que el vector u es un límite superior para la variable de flujo x . Se observa que la formulación dada es pura o exacta, pues no existen otras restricciones para los flujos además de las de divergencia. Una de las propiedades más interesantes de las formulaciones exactas de problemas de redes, es que la matriz A es totalmente unimodular (TU) y por lo tanto el problema z tiene una solución entera incluso si sus variables se consideran continuas. Es decir, aún si $x \in \mathbb{R}^n$, los vértices del poliedro $P = \{x : Ax \geq b, 0 \leq x \leq u\}$ son enteros. Lo anterior simplifica la resolución: permite que un problema de red pueda ser resuelto como un problema de programación lineal en vez de un problema entero o combinatorial. Sin embargo, como se detalla más adelante, es poco común que aplicaciones reales puedan ser modeladas como una red exacta.

2.1.1.1. Problema de ruta mínima

El problema estudiado en este trabajo, considera una modelación basada en redes acíclicas y dirigidas, sobre las que se resuelven problemas de ruta mínima. El problema de ruta mínima consiste en encontrar la ruta de mínimo costo o mínima distancia, dependiendo del parámetro asociado, entre un conjunto de rutas que comparten el mismo nodo de origen y el mismo nodo de destino.

Definiciones

- Grafo dirigido: una red dirigida o grafo dirigido es un conjunto de \mathcal{N} nodos y \mathcal{A} pares de nodos o arcos. Sean $i, j \in \mathcal{N}$, entonces el par $(i, j) \in \mathcal{A}$ representa un arco que ingresa en j y sale desde i , por lo tanto el orden del par (i, j) es relevante.
- Ruta y ciclo: una ruta es una secuencia de nodos (i_1, i_2, \dots, i_k) asociada a una secuencia de $k - 1$ arcos, considerando $k \geq 2$. Cada ruta posee un nodo de origen i_1 y un nodo de destino i_k . Un ciclo, es una ruta en la que el nodo de origen y nodo destino coinciden.
- Flujo: cantidad que circula por un arco de la red. Matemáticamente, el flujo se representa con una variable continua o binaria asociada a un arco específico, e.g. $x_{ij} \geq 0 \forall (i, j) \in \mathcal{A}$.

Considere la siguiente notación: sean \mathcal{N} y \mathcal{A} los conjuntos de nodos y arcos del grafo, respectivamente. Sean $\{s, t\} \in \mathcal{N}$ los nodos de origen y destino, respectivamente. Sean c_{ij} y x_{ij} el parámetro de costo o distancia y la variable de decisión asociada con el arco $(i, j) \in \mathcal{A}$, respectivamente. Entonces, el problema de ruta mínima se puede formular como:

$$\min_x \sum_{(i,j) \in \mathcal{A}} c_{ij} x_{ij} \quad (2.2)$$

$$\sum_{j:(i,j) \in \mathcal{A}} x_{ij} - \sum_{j:(j,i) \in \mathcal{A}} x_{ij} = b_i, \quad b_i = \begin{cases} 1, & \text{si } i = s \\ 0, & \text{si } i \in \mathcal{N} \setminus \{s, t\} \\ -1, & \text{si } i = t \end{cases} \quad (2.3)$$

$$0 \leq x_{ij} \leq 1 \quad \forall (i, j) \in \mathcal{A} \quad (2.4)$$

Notar que la variable x_{ij} tendría que ser binaria para tener la interpretación de activación del arco $(i, j) \in \mathcal{A}$. Sin embargo, debido a que las restricciones de flujo (2.3) definen un poliedro cuya matriz es TU, entonces el problema de ruta mínima definido con la variable continua tiene la propiedad de integridad, es decir, su solución óptima es entera [Nemhauser y Wolsey \(1988\)](#). Como se mencionaba anteriormente, la importancia de esta propiedad radica en que un problema combinatorial puede ser resuelto como un problema de programación lineal por algoritmos genéricos como simplex [Bertsimas y Tsitsiklis \(1997\)](#). En el caso particular del problema de ruta mínima, existen algoritmos especializados que resuelven con mejor desempeño que algoritmos basados en simplex.

Algoritmo de ruta mínima

En esta sección se resume el algoritmo genérico para resolver problemas de ruta mínima en redes dirigidas, pero las variantes de este método se encuentran detalladas en [Bertsekas \(1998\)](#). El algoritmo genérico está basado en comprobar iterativamente si se cumplen las condiciones de holgura complementaria del problema de ruta mínima, para lo cual se define un vector de etiquetas $d = (d_s, \dots, d_t)$ para todos los nodos que conforman la red.

La ruta $\{s, \dots, d\}$ es la ruta de mínimo costo entre s y d si cumple la ecuación de Bellman:

$$d_j = \begin{cases} \min_{(i,j) \in \mathcal{A}} \{d_i + c_{ij}\} & \text{si } j \neq s \\ 0 & \text{si } j = s \end{cases} \quad (2.5)$$

El algoritmo itera hasta encontrar etiquetas que cumplan la ecuación (2.5). Para actualizar las etiquetas se define un conjunto de nodos candidatos $V \subseteq \mathcal{N}$. El pseudocódigo del algoritmo genérico se describe a

continuación:

Algorithm 1 Algoritmo de ruta mínima.

Require: $k = 0$, $d^{(k)} = (0, \infty, \dots, \infty)$, $V^{(k)} = \{s\}$.

- 1: Remover un nodo $i \in V^{(k)}$.
 - 2: **if** $d_j^{(k)} > d_i^{(k)} + c_{ij}$ **then**
 - 3: $d_j^{(k+1)} = d_i^{(k)} + c_{ij}$
 - 4: $V^{(k+1)} \equiv V^{(k)} \cup \{j : (i, j) \in \mathcal{A}\}$
 - 5: **if** $V^{(k+1)} \equiv \emptyset$ **then return** $d^{(k+1)}$
 - 6: **else** $k \leftarrow k + 1$ y volver a 1:
-

Los distintos algoritmos para resolver el problema de ruta mínima difieren en la forma de remover el nodo del conjunto de candidatos. Una clasificación de algoritmos distingue los algoritmos *label setting* en los que el nodo que abandona la lista es aquel que posee la menor etiqueta (e.g. algoritmo de Dijkstra). Por otro lado están los algoritmos *label correcting* en que no existe un criterio para la selección de los nodos (e.g. algoritmo de Bellman-Ford). Por ejemplo, en el algoritmo de Dijkstra los nodos que salen del conjunto de candidatos no vuelven a ingresar, lo cual resulta en un número menor de iteraciones comparado con el algoritmo de Bellman-Ford.

2.1.1.2. Restricciones adicionales

El objetivo de utilizar modelos de red para representar situaciones es poder emplear los algoritmos especializados para resolver el problema de optimización asociado. Esto genera ventajas computacionales considerables respecto a un algoritmo estándar. Sin embargo, en la práctica es poco común que las formulaciones de red resultantes sean exactas; lo normal es que existan otras restricciones que relacionan los arcos de la red en una manera distinta a la divergencia. Considerando el poliedro P del problema (2.1), un problema de red con restricciones adicionales se define como:

$$z_{sc} = \min \{c^T x : x \in \mathbb{Z}^n, x \in P, Dx = d\}, \quad (2.6)$$

donde las restricciones adicionales son definidas por la matriz D y el vector d . Por ejemplo en un problema de ruta mínima, no es raro que sea necesario que la ruta de mínimo costo que cumpla con un tiempo límite (RCSP, *Resource Constrained Shortest Path*, Garcia (2009)). Estas restricciones adicionales o *side constraints* tienen dos implicancias: (i) destruyen la estructura TU de la matriz de coeficientes de las restricciones y (ii) impiden el uso tradicional de los algoritmos especializados Mathies y Mevert (1998). En resumen, las restricciones adicionales complican la resolución del problema. En el caso de un problema de red con variables enteras, ya no es posible resolver la relajación continua del mismo para obtener la solución óptima.

En este trabajo, se requiere resolver la relajación continua de un problema de red con restricciones adicionales. Si bien se trata de un problema de programación lineal, los solucionadores comerciales resultan ineficientes en el contexto considerado donde se necesitan resultados en un plazo de segundos o minutos. Debido a la cantidad de variables y restricciones, la formulación propuesta en este trabajo puede considerarse un problema lineal de gran escala.

2.1.2. Problemas lineales de gran escala

Gran parte de los problemas de programación lineal de gran escala tienen una estructura especial, donde la matriz de coeficientes de las restricciones posee un gran número de valores nulos (*sparse matrix*). Este patrón de ceros y no-ceros puede ser aprovechado para obtener ventajas computacionales en la resolución [Dantzig y Thapa \(2006\)](#). El problema tratado en este trabajo, entra en la categoría de problemas lineales de gran escala con una estructura de bloques diagonal (*block-angular*), donde el número de restricciones que acopla los bloques es relativamente bajo respecto al total de restricciones del problema. Considere la siguiente formulación de un problema con estructura de bloques diagonal:

$$\min_x \quad c_1^T x_1 + c_2^T x_2 + \dots + c_n^T x_n \quad (2.7)$$

$$D_1 x_1 + D_2 x_2 + \dots + D_n x_n = b_0 \quad (2.8)$$

$$A_1 x_1 = b_1 \quad (2.9)$$

$$A_2 x_2 = b_2 \quad (2.10)$$

$$\vdots = \vdots \quad (2.11)$$

$$A_n x_n = b_n \quad (2.12)$$

En la formulación *block-angular* mostrada, la variable $x_i \in \mathbb{R}_+^{n_i} \forall i = 1, \dots, n$ y los parámetros $b_0 \in \mathbb{R}_+^{m_0} \forall i = 0, \dots, n$. Asumir que las demás matrices y vectores $A_i, D_i, c_i \forall i = 1, \dots, n$ tienen las dimensiones apropiadas. Se puede notar que en ausencia de las restricciones (2.8), el problema se descompone en subproblemas más pequeños.

2.1.2.1. Métodos de solución

A continuación se mencionan algunos métodos de solución para este tipo de problemas.

- Método estándar. En esta categoría se incluyen los algoritmos basados en simplex y los algoritmos de punto interior, ambos disponibles en la mayoría de los programas de optimización [Koberstein \(2008\)](#). Los algoritmos basados en simplex iteran moviéndose a través de los vértices de la región factible, mientras que los de punto interior se acercan a la solución óptima desde el interior de la región

factible. En el caso de problemas lineales de gran escala, utilizar un algoritmo simplex puede resultar ineficiente debido a la degeneración del problema. Dentro de los métodos de punto interior, uno de los más eficientes es el algoritmo primal-dual infactible Andersen et al. (1996). En la implementación que se puede encontrar en los optimizadores comerciales, la función objetivo se transforma en una función log-barrier penalizada, a partir de la cual se obtienen condiciones Karush-Kuhn-Tucker (KKT) perturbadas. El sistema de ecuaciones resultante, se resuelve con el método de Newton hasta alcanzar una tolerancia en la perturbación. Una de las críticas a este tipo de algoritmos, es que no necesariamente terminan con la solución óptima y tampoco con una solución básica. A pesar de lo anterior, se ha demostrado que los algoritmos de punto interior son herramientas muy útiles para resolver sistemas de gran escala Bixby et al. (1992).

- Dantzig-Wolfe: el algoritmo Dantzig-Wolfe es un método de descomposición adecuado para problemas con la estructura block-angular mostrada Ho y Louie (1981). El método consiste en reformular el problema en base a los puntos y rayos extremos de los poliedros que definen cada bloque, de tal forma que ahora las variables son los coeficientes asociados a dichos puntos y rayos extremos. El problema reformulado se denomina problema maestro, y se caracteriza por un número menor de restricciones pero un aumento considerable en el número de variables. Sabiendo que gran parte de las variables nunca formará parte de la base óptima, en vez de evaluar los costos reducidos de cada variable, el algoritmo resuelve subproblemas donde encuentra el costo reducido mínimo, lo que equivale a resolver un subproblema por cada bloque. Si se encuentra un costo reducido negativo, la solución del subproblema corresponde a un punto o rayo extremo, por lo tanto el coeficiente asociado debe entrar a la base en el problema maestro, el cual se debe resolver nuevamente. Si ningún costo reducido es negativo, entonces se ha alcanzado la solución óptima.
- Relajación Lagrangiana: en un problema con la estructura *block-angular*, es fácil notar que en ausencia de las restricciones que acoplan los bloques bastaría con resolver subproblemas por separado. La relajación lagrangiana en un problema de este tipo se basa en omitir las restricciones que acoplan el sistema pero incluirlas en la función objetivo con una penalización (multiplicadores lagrangianos), de manera que una solución relajada que es infactible en el problema original tiene un costo adicional. La utilidad de la relajar la restricción del sistema es separar el problema en subproblemas más sencillos. Considere la relajación lagrangiana del problema (2.7)-(2.12):

$$\max_{\lambda} \left\{ \min_x \sum_{i=1}^n c_i^T x_i + \lambda^T (b_0 - \sum_{i=1}^n D_i x_i) \right\} \quad (2.13)$$

$$A_1 x_i = b_i \quad \forall i = 1, \dots, n \quad (2.14)$$

El algoritmo itera actualizando los valores de los multiplicadores $\lambda \in \mathbb{R}^{m_0}$, usualmente utilizando el método del subgradiente. El procedimiento busca la penalización óptima, es decir, un valor de

los multiplicadores que logre que la restricción relajada se cumpla sin necesidad de incluirla como restricción.

- Algoritmo Bienstock-Zuckerberg (BZ): un método de descomposición más reciente para la resolución de problemas lineales con cierta estructura es el algoritmo de Bienstock y Zuckerberg desarrollado en [Bienstock y Zuckerberg \(2015\)](#). Originalmente, el algoritmo fue diseñado para el problema de minería en cielo abierto (*Open Pit Mining Problem*), donde se debe programar el orden de extracción de los bloques que reporten mayor beneficio cumpliendo con la precedencia inherente de las capas subterráneas. En aplicaciones reales, el problema tiene millones de variables binarias y además de las restricciones de precedencia se deben considerar restricciones de recursos como el tiempo para la extracción. Si las restricciones de recursos no estuvieran presentes, el problema sería sencillo de resolver mediante un algoritmo *minimum-cut*. Sin embargo, la inclusión de estas restricciones complica la formulación, haciendo que incluso la relajación continua del problema sea difícil de resolver para un optimizador estándar. La implementación desarrollada en el trabajo de Bienstock y Zuckerberg, muestra que el algoritmo resuelve la relajación Lagrangiana del problema, pero utiliza la solución relajada obtenida para generar restricciones en el problema principal. Estas restricciones disminuyen considerablemente el tamaño del problema principal. A su vez, la solución del problema principal restringido entrega nuevas variables duales que permiten actualizar la relajación. El algoritmo itera hasta que se alcanza la igualdad entre el problema restringido y el problema relajado. Un planteamiento mucho más general del algoritmo BZ se muestra en el trabajo de [Munoz et al. \(2016\)](#), donde se clasifica al algoritmo BZ como un método de generación de columnas que hace uso de la estructura combinatorial del problema permitiendo mejorar la convergencia. De acuerdo a lo señalado en [Munoz et al. \(2016\)](#) el algoritmo BZ difiere de un método general de generación de columnas en dos aspectos: (i) las columnas generadas son vectores ortogonales, por lo tanto la inclusión en el problema maestro equivale a una operación de contracción de variables y (ii) la operación de contracción nunca elimina una solución óptima previa del problema maestro.

En este trabajo, se implementa un método de solución basado en el algoritmo BZ que además incluye algunos de los elementos para mejorar la velocidad de convergencia discutidos en [Munoz et al. \(2016\)](#). Hasta donde sabemos, el algoritmo BZ no ha sido implementado en un problema de red con restricciones adicionales. En particular, en nuestra formulación los subproblemas corresponden a problemas de ruta mínima. La implementación del algoritmo BZ en nuestra formulación permite resolver una secuencia de problemas más pequeños en vez de un problema de gran escala, logrando un desempeño computacional superior a lo obtenido con un solver estándar.

2.2. Equilibrio competitivo

En economía, el equilibrio económico sucede cuando existe un balance entre la cantidad ofertada y demandada a nivel agregado, como resultado de la acción de múltiples participantes que buscan maximizar su beneficio a un determinado precio. Se establece que un equilibrio es competitivo o walrasiano, si ningún participante consume o produce una cantidad significativa como para alterar el precio de mercado, por lo tanto el precio es una variable exógena y los participantes son tomadores de precio [Varian \(1992\)](#). En teoría, el precio se ajusta hasta alcanzar la condición de equilibrio y este análisis ha sido útil para explicar el comportamiento de la mayoría de los mercados de *commodities*.

Para un mejor entendimiento, considere la siguiente notación que representa un mercado en general. Sea J el conjunto de productores, sea I el conjunto de consumidores. Sea x_j la cantidad producida por el productor j y sea y_i la cantidad comprada por el consumidor i . Además, sea $c_j(x_j)$ el costo para el productor j de producir la cantidad x_j , y sea $u_i(y_i)$ la utilidad de consumo del consumidor i al comprar una cantidad y_i . Sea p un precio determinado. Dado que ambos grupos de participantes son tomadores de precio, entonces el beneficio se calcula como:

- Productores: $S_j(x_j, p) = px_j - c_j(x_j) \quad \forall j \in J$
- Consumidores: $D_i(y_i, p) = u_i(y_i) - py_i \quad \forall i \in I$

El equilibrio competitivo se alcanza cuando dado un precio, cada participante produce o consume la cantidad que maximiza su beneficio individual y como consecuencia resulta un balance entre oferta y demanda, donde ningún participante del mercado tiene incentivo a ofertar o demandar una cantidad diferente. El precio que permite esta condición se conoce como precio de mercado o precio de equilibrio. Sea \bar{p} el precio de equilibrio, entonces en equilibrio competitivo se cumple:

- (i) Cada productor oferta la cantidad que maximiza su beneficio, es decir:

$$\bar{x}_j \in \arg \max_{x_j} S_j(x_j, \bar{p}) \quad \forall j \in J$$

- (ii) Cada consumidor compra la cantidad que maximiza su beneficio, es decir:

$$\bar{y}_i \in \arg \max_{y_i} D_i(y_i, \bar{p}) \quad \forall i \in I$$

- (iii) La cantidad total ofertada equivale a la cantidad total comprada.

$$\sum_{j \in J} \bar{x}_j = \sum_{i \in I} \bar{y}_i$$

Las condiciones anteriores muestran que los incentivos individuales de los participantes son compatibles con el equilibrio, y a la vez, con el beneficio social. Para ver esto último, considere el punto de vista de un planificador que busca maximizar el beneficio de todos los participantes cumpliendo el equilibrio de oferta y demanda. Este problema de optimización se conoce como *social planner's problem* (SPP):

$$\text{SPP: } \max_{x,y} \sum_{i \in I} u_i(y_i) - \sum_{j \in J} c_j(x_j) \quad (2.15)$$

$$\text{s.a. } \sum_{j \in J} x_j = \sum_{i \in I} y_i \quad (2.16)$$

$$x_j, y_i \geq 0 \quad \forall i \in I, j \in J \quad (2.17)$$

Es fácil notar que la función objetivo (2.15) corresponde a la suma de los beneficios de productores y consumidores. La solución a este problema se conoce como asignación eficiente de los recursos debido a que los bienes son producidos al mínimo costo y son consumidos por aquellos participantes que más lo valoran. A pesar de que el precio de mercado no se incluye explícitamente en la formulación del problema, si existe un precio que hace consistente el equilibrio obtenido de manera centralizada, al resolver el SPP, con la decisión descentralizada de los participantes, expresada en las condiciones (i) y (ii) antes mencionadas. Por lo tanto, la solución del SPP es un equilibrio competitivo.

En particular, el precio de equilibrio equivale al costo marginal de los productores o la utilidad marginal de los consumidores. Sea $c(x) = \sum_{j \in J} c_j(x_j)$ el costo de los productores para un nivel de producción $x = (x_1, x_2, \dots, x_{|J|})$ y sea $u(y) = \sum_{i \in I} u_i(y_i)$ la utilidad de los consumidores para un nivel de consumo $y = (y_1, y_2, \dots, y_{|I|})$. Sean (\bar{x}, \bar{y}) las cantidades producidas y consumidas que resuelven el problema de equilibrio, respectivamente. Entonces, el precio \bar{p} bajo el cual la asignación eficiente es un equilibrio competitivo equivale a:

$$\bar{p} = c'(\bar{x}) = u'(\bar{y}) \quad (2.18)$$

Es importante notar que este resultado se obtiene asumiendo que las funciones de utilidad y costo de los participantes son continuas y diferenciables en su dominio (*well-behaved*), usualmente se asume concavidad para la utilidad y convexidad para los costos. Bajo estos supuestos, el precio \bar{p} también equivale a la variable dual óptima asociada a la restricción de equilibrio (2.16) en el SPP.

2.2.1. Mercados no-convexos

En la práctica, muchos modelos de mercados involucran no-convexidades en la forma de economías de escala, costos fijos de uso de los factores productivos, costos de encendido o apagado de las máquinas, requerimientos de producción mínima, entre otros. Ha sido ampliamente analizado que en presencia de no-convexidades, puede que no exista un precio que sea consistente con el equilibrio competitivo [O'Neill et al.](#)

(2005). En otras palabras, si el planificador central decide las cantidades de asignación eficiente, puede que no exista un precio que haga compatible tales cantidades con la decisión individual de los participantes. Por ejemplo, un precio basado en costo o utilidad marginal, puede resultar en pérdidas para algunos participantes que producen o consumen en base a lo definido por el planificador. Desde un punto de vista matemático, lo anterior se resume en que las variables duales asociadas a la restricción de equilibrio no tienen la interpretación económica del caso convexo. Como señala Scarf (1990), cuando existen no-convexidades la prueba de precios del método simplex no sirve para decidir si una nueva actividad genera beneficio.

En las últimas décadas, el problema de encontrar precios en mercados no-convexos ha atraído mucho interés debido a la desregulación del sector eléctrico. A nivel internacional, los mercados eléctricos han migrado hacia una estructura tipo *pool* para las transacciones en el mercado mayorista diario, de manera de promover la competencia entre los participantes. En este tipo de modelos, los productores envían su información técnica y económica al operador del mercado y sistema, quien se encarga de planificar la operación menos costosa del sistema. Los costos de partida, apagado y las restricciones operativas generan una función de costos totales no-convexa, por lo tanto el costo marginal es menor al costo promedio Liberopoulos y Andrianesis (2016). Bajo un precio basado en costo marginal, puede que algunos productores no alcancen a recuperar los costos de operación cuando funcionan de acuerdo a la planificación del operador. Lo anterior es un problema característico de los modelos de mercado no-convexos pero es especialmente relevante en el contexto eléctrico. En el mercado eléctrico, los participantes son entidades privadas que buscan maximizar su beneficio individual pero están obligatoriamente subordinadas a la decisión de un operador. Por lo tanto, un participante que no está generando beneficios en principio no puede modificar la cantidad producida.

La práctica común para solucionar este problema ha sido proveer pagos adicionales a aquellos productores que no logran recuperar todos los costos de operación cuando son remunerados al precio de mercado. De esta manera, la decisión del operador se vuelve compatible con las decisiones de los productores. Si bien esto ayuda a mitigar el problema, últimamente se han planteado posibles consecuencias negativas en el largo plazo. Algunos autores señalan directamente que se tratan de una ineficiencia del mercado. Estos conceptos se abordan con más detalle en las siguientes secciones.

2.3. Mercados eléctricos

Los mercados eléctricos, sólo pueden funcionar si existe una red de soporte. A diferencia de otros commodities, la energía eléctrica necesita de un sistema para ser comprada o vendida Kirschen y Strbac (2018). La red eléctrica consiste básicamente de cuatro componentes: generación, transmisión, distribución y sistemas de carga. Antes de la desregulación del mercado eléctrico, era común que varios subsistemas sean propiedad de una misma empresa. Sin embargo, en los 80s, la industria eléctrica empezó a migrar desde la

integración vertical hacia una estructura más distribuida, la cual fortaleció la competencia principalmente a nivel de generación. Paralelo a la desregulación, se fortaleció el concepto de operador independiente del sistema (ISO, *independent system operator*). El ISO manipula la red eléctrica, coordinando la operación de los participantes en busca de dos objetivos: seguridad y eficiencia. La seguridad se sustenta en el balance en tiempo real entre la oferta y demanda de energía eléctrica y el cumplimiento de las restricciones de transmisión en el sistema. La eficiencia por su parte, se logra con la operación menos costosa del sistema y el libre acceso de los participantes a la red.

Esta sección resume la estructura básica de un mercado eléctrico, definiendo los tipos de mercado, algunas de las entidades que participan y las principales reglas para su funcionamiento, en base a las clasificaciones realizadas en [Shahidehpour et al. \(2003\)](#), [Kirschen y Strbac \(2018\)](#). Al final de la sección se establecen las características del modelo considerado para el problema tratado en este trabajo.

2.3.1. Entidades

- **ISO:** el operador independiente del sistema tiene como principal función garantizar la seguridad del sistema, manteniendo un balance de oferta y demanda en tiempo real y verificando el cumplimiento de los límites en la red de transmisión. Para llevar a cabo esta función, el ISO puede decidir el nivel de funcionamiento de todos o parte de los recursos del sistema.
- **Generadores:** conjunto de empresas propietarias de centrales eléctricas. Las centrales difieren principalmente en la tecnología y fuente de energía primaria utilizada. En las últimas décadas, ha existido un incremento de las centrales a base de energía renovable no convencional.
- **Transmisión:** conjunto de empresas encargadas del transporte de energía desde los centros de generación hacia los centros de demanda a través de una red de alta tensión.
- **Distribución:** conjunto de empresas que licita la concesión del servicio de distribución en un área geográfica determinada, para lo cual mantienen y operan las instalaciones de distribución.
- **Consumidores:** la principal distinción en este grupo es entre clientes libres y clientes regulados. Los primeros, tienen un consumo por sobre un mínimo y pueden negociar directamente precios y cantidades en el mercado mayorista. Los últimos son los consumidores finales de la energía, en su mayoría domicilios que consumen en base a una tarifa fija regulada.

2.3.2. Modelos de mercado

El modelo de mercado rige la forma de realizar las transacciones entre productores y consumidores. En general, se reconocen dos modelos de mercado eléctrico. Sin embargo, la mayoría de los países ha implementado estructuras mixtas que adquieren propiedades de ambos modelos.

- **Modelo *pool*:** modelo de mercado centralizado, donde los productores compiten por entregar energía al sistema en lugar de cumplir la demanda de consumidores específicos. Los productores informan la estructura de costos al ISO, el cual en base a la información de demanda de los consumidores, determina la operación de los productores de manera de cumplir la demanda al mínimo costo. Si bien los participantes deciden libremente entrar al mercado, una vez en él, la participación es obligatoria si el operador lo requiere. En estos modelos, las transacciones se rigen por el precio spot de la energía, el cual usualmente queda determinado por el costo marginal del productor más costoso despachado.
- **Modelo de contratos bilaterales:** modelo de mercado descentralizado, donde los participantes establecen libres acuerdos de precio y cantidad. En estos modelos, el ISO sólo verifica si las transacciones pueden llevarse a cabo de acuerdo a la capacidad de transmisión del sistema.

2.3.3. Tipos de producto

Una clasificación más específica de los mercados tiene relación con el producto de las transacciones realizadas. Para facilitar la explicación, esta sección se basa en un mercado *pool*.

- **Energía:** el mercado de energía es donde se desarrolla la competencia entre empresas generadoras. Los productores, envían información detallada de su estructura de costos y funcionamiento al ISO. Este último, construye una curva de oferta, la cual al ser intersectada con la curva de demanda agregada, determina la operación del sistema y el precio de mercado. Si bien este precio finalmente se ajusta mediante cargos por transmisión, es un indicador del comportamiento del mercado no sólo para los participantes sino también para posibles inversionistas.
- **Servicios complementarios:** los servicios complementarios son necesarios para la operación confiable del sistema. En general, los productores envían al ISO ofertas de capacidad disponible para ser utilizada en caso de una contingencia.

2.3.4. Mercado mayorista

En muchos países, el mercado eléctrico mayorista se liquida en base a una subasta centralizada. El ISO recibe ofertas de suministro y demanda de energía eléctrica, donde se especifican cantidades y precios. Las ofertas reunidas sirven como información para que el ISO determine la operación de los participantes en base a algún modelo implementado, como por ejemplo resolviendo el problema de programación y predespacho de unidades de generación (UCP, *unit commitment problem*).

El mercado eléctrico mayorista se divide entre un mercado *forward* y un mercado en tiempo real o mercado spot. En general, el mercado forward considera un horizonte diario basado en horas, por lo mismo es comúnmente conocido como mercado diario. Por su parte, en el mercado spot se realiza el balance de los

desajustes del mercado forward. En ambos mercados, el ISO se encarga de realizar el balance entre oferta y demanda para cumplir con la seguridad y eficiencia del sistema.

En el mercado diario, el proceso del ISO puede ser resumido en cuatro etapas: (1) recopilación de ofertas de despacho y consumo, (2) programación de unidades de generación, (3) determinación del despacho de las unidades programadas y (4) determinación de precios por período. Si bien existe consenso respecto a los primeros tres procesos, el último presenta variaciones entre sistemas. El esquema de precios es fundamental en el diseño de los mercados, debido a que orienta las decisiones de operación e inversión de los participantes. Actualmente, la mayoría de los mercados ha implementado lo que se conoce en la literatura como precios marginales locales (LMP, *local marginal prices*). Este modelo determina precios uniformes para cada hora basados en el costo marginal del sistema más un cargo local por pérdidas en líneas de transmisión. Al final de un horizonte de tiempo definido, usualmente un día, se determinan en caso de existir, las unidades con pérdidas económicas las cuales son compensadas mediante un pago adicional. El aumento de los pagos adicionales con la evolución del mercado eléctrico ha llevado a cuestionar el método de precios basado en costo marginal y a la vez desarrollar, al menos teóricamente, nuevas políticas de precios las cuales son discutidas en la siguiente sección.

2.3.5. Política de precios

El esquema de precios en un diseño de mercado es relevante debido a que los participantes realizan decisiones de operación e inversión basados en la señal de precios. En la literatura, se han propuesto tres indicadores para evaluar el precio: eficiencia, transparencia y simplicidad. Un precio es eficiente si es que maximiza el beneficio social e incentiva a los participantes a seguir esa solución de mercado. La transparencia tiene relación con que cada participante puede conocer los beneficios de los demás. La simplicidad evalúa si el esquema de precios es simple de implementar y fácil de interpretar por los participantes.

La determinación de precios en mercados eléctricos mayoristas es un problema que ha atraído interés en el último tiempo, principalmente debido a la dificultad de definir un esquema de precios que compatibilice los incentivos individuales con el objetivo social [O'Neill et al. \(2005\)](#), [Madani et al. \(2018\)](#). Como se mencionaba anteriormente, los mercados eléctricos están altamente acoplados con el sistema físico, por lo que la operación del sistema y regulación del mercado son en general funciones de una entidad independiente conocida como ISO. El ISO se encarga de programar la operación y despacho de las unidades de generación para cumplir con la demanda de energía en todo momento al mínimo costo posible y además, paralelo a la operación eficiente del sistema, el ISO determina el precio de la energía para regular las transacciones en el mercado mayorista. Idealmente, el precio de la energía debiera reflejar exactamente el costo de producirla.

El problema con la determinación de precios en el mercado eléctrico es que al tratarse de un modelo

de mercado no-convexo, contrario a la teoría económica clásica, puede que no exista un precio de equilibrio. En otras palabras, un determinado precio a la energía puede generar inconsistencia entre la cantidad dispuesta a ofertar por parte de las unidades de generación y la cantidad total demandada, la cual por cierto se asume inelástica. En un contexto distinto al eléctrico, si un participante observa un precio inferior a sus costos de operación, puede simplemente decidir no participar del mercado. Sin embargo, en principio la operación de las unidades obedece a la programación de despacho del ISO, por lo tanto se genera un problema de incentivos: puede que el precio y la cantidad determinada por el ISO resulte en pérdidas para algunas unidades. Por otro lado, un precio muy alto, si bien podría eliminar las pérdidas para las unidades despachadas, posiblemente no sería consistente con los incentivos de las unidades no despachadas.

Una forma de establecer un sistema de precios consistente con el equilibrio en el mercado eléctrico, es definir un pago adicional al pago por energía para asegurar que ninguna unidad despachada por el ISO tenga incentivos para despachar otra cantidad. De esta forma, el despacho de mínimo costo definido por el ISO para un determinado horizonte de planificación se alinea con la decisión descentralizada de las unidades.

Desde el punto de vista de eficiencia del mercado, lo ideal sería definir conjuntamente el precio de la energía y el pago adicional para cada unidad. Sin embargo, lo anterior resulta en un problema de optimización difícil de resolver, conocido como el problema de equilibrio en mercados no-convexos. En [Huppmann y Siddiqui \(2018\)](#) se formula un problema bi-nivel y multi-objetivo para obtener las cantidades, precios y pagos adicionales que logran el equilibrio de mercado y maximizan el bienestar social; sin embargo el procedimiento es complejo. Un enfoque más tratable para obtener precios de equilibrio en mercados no-convexos es el propuesto por [O'Neill et al. \(2005\)](#), donde se definen además de precios a la energía, precios a las decisiones discretas como el encendido y apagado. Los precios obtenidos están basados en la variable dual óptima de la relajación continua de la formulación de UCP cuando las variables binarias de programación han sido fijadas en su valor óptimo. Si bien los precios calculados permiten llegar de manera descentralizada al equilibrio de mercado, poseen algunas propiedades indeseables como la volatilidad, debido a que no siguen un patrón con respecto a la demanda, y el carácter discriminatorio, debido a que cada unidad es remunerada en base a precios diferentes. Por otro lado, los beneficios de las unidades son altamente sensibles a la optimalidad del UCP, lo cual es crítico si se considera que las dimensiones prácticas del problema implican recurrir a soluciones subóptimas [Johnson et al. \(1997\)](#).

En el trabajo de [Motto y Galiana \(2002\)](#), el UCP es reformulado para incluir funciones de desincentivo de tal forma de determinar conjuntamente precios a la energía y pagos adicionales. Los parámetros de la función de desincentivo son determinados resolviendo un problema cuadrático de tal forma que la suma total de los desincentivos sea cero. Por lo tanto, se obtiene equilibrio con la misma solución del UCP, equivalente al despacho de mínimo costo del ISO. Sin embargo, el alcance de la formulación es limitado porque sólo considera mínimos y máximos de generación. Además, es poco práctico porque el resultado de la función de desincentivo puede ser negativo, lo cual implicaría que algunas unidades serían las responsables de incurrir

en el pago adicional.

Alternativamente, una condición de cuasi-equilibrio de mercado se puede alcanzar con otros métodos. En [Ruiz et al. \(2012\)](#) se definen precios en conjunto con la solución de despacho, asegurando que ninguna unidad obtenga pérdidas. Sin embargo, la solución de despacho se desvía levemente de la solución de mínimo costo y por otra parte, como los precios deben ser altos para cubrir costos marginales y parte de los costos fijos, pueden no ser compatible con los incentivos de las unidades no despachadas. El método que actualmente se aplica en mercados eléctricos, es definir un precio y posterior a la realización del pago por energía, definir pagos adicionales a las unidades que están teniendo pérdidas económicas.

2.3.5.1. Pagos adicionales

Un método alternativo para alcanzar una condición de cuasi-equilibrio en el mercado consiste en establecer pagos compensatorios o pagos adicionales a las unidades que no logren maximizar beneficios cuando funcionan en base a las cantidades y precios establecidos por el operador del sistema. Este pago adicional (*side-payment* o *uplift* en la literatura) se calcula posterior a la realización del despacho de las unidades, en base a los ingresos obtenidos con el precio de la energía.

Los pagos adicionales se pueden clasificar en dos tipos: pago completo (MWP, *make-whole payment*) y pago al costo de oportunidad perdido (LOC, *lost opportunity cost*). MWP se realiza cuando el ingreso que recibe la unidad por despachar según la decisión del operador al precio del mercado, no es suficiente para cubrir sus costos. Un MWP lleva a la unidad a beneficio cero, es decir, la hace indiferente entre seguir la solución de mercado y no hacer nada. El pago LOC asegura que una unidad reciba el máximo beneficio posible al precio de mercado y se realiza cuando una unidad no está operando a máxima capacidad pero podría generar mayores ingresos si aumenta su nivel de generación.

Los pagos adicionales son indeseables en una economía. La falta de transparencia debido a que no es explícito el pago adicional recibido por cada unidad, y el carácter discriminatorio debido a que no es un pago uniforme para todas las unidades, distorsionan la señal de precios en el mercado. Incluso, algunos autores han señalado que la existencia de estos pagos pueden incentivar a que los participantes se comporten estratégicamente respecto a la información de sus costos y cantidades [Liberopoulos y Andrianesis \(2016\)](#), [Van Vyve et al. \(2011\)](#).

2.3.5.2. Precios marginales

En los mercados eléctricos, la práctica común es implementar un esquema de precios uniforme basado en el costo marginal del sistema. En un inicio esta práctica establecía precios por zonas geográficas, sin embargo a partir de los años 90 en muchos mercados empezó una transición hacia precios por nodo, denominados precios marginales locales (LMP). Este tipo de precios enfatiza las diferencias entre nodos

de demanda debido a las pérdidas y límites en la transmisión Litvinov (2010). La interpretación del LMP es el incremento total en el costo del sistema cuando se despacha una unidad adicional de energía en el nodo correspondiente. El LMP representa una buena señal de precios para un período de tiempo, pues puede capturar los costos marginales del despacho pero no los costos fijos asociados al encendido y apagado de las unidades. La naturaleza discreta de esas decisiones genera la no-convexidad en el modelo de mercado que impide la existencia de un precio de equilibrio, es decir, un precio que permita maximizar los beneficios de los participantes y a la vez equilibrar oferta con demanda. Una alternativa similar en funcionamiento a los LMP consiste en resolver el problema *Convex Hull Pricing* (CHP) propuesto en Hogan y Ring (2003), Gribik et al. (2007). Este método permite obtener precios no-decrecientes respecto a la demanda y además minimiza cierto tipo de pagos adicionales. Si bien son propiedades interesantes en este contexto, aún existen desafíos para la implementación del método, como se explica en la siguiente sección.

2.4. Convex Hull Pricing

Anteriormente se señalaba la problemática en la determinación de precios del mercado mayorista eléctrico. En resumen, las características del sistema eléctrico y las condiciones del mercado respectivo resultan en un modelo que necesita pagos adicionales. Una forma de minimizar cierto tipo de pagos adicionales, en particular LOC, es el método Convex Hull Pricing. Este método, determina los precios en base a la pendiente de la envoltura convexa de la función de costos del UCP parametrizada según la demanda. Para entender mejor el método CHP, al inicio de esta sección se presenta una formulación del modelo UCP. Posteriormente se explica el procedimiento de cálculo de los precios y finalmente se presenta un ejemplo para mostrar algunas diferencias con el método LMP, pues ambas políticas implican determinar un precio y posteriormente pagos adicionales.

2.4.1. Modelo Unit Commitment

Existe extensa literatura respecto al problema de programación y despacho de unidades de generación, más conocido como *unit commitment problem* (UCP). Este problema consiste en programar el funcionamiento de las unidades y a la vez decidir el nivel de despacho de cada una para cumplir la demanda pronosticada para un horizonte de tiempo. El modelo matemático de UCP corresponde a un problema de programación lineal entera mixta (MILP, *mixed integer linear programming*). De manera de formular el modelo UCP, considere la notación de la tabla 2.1:

Tabla 2.1: Notación para la formulación de UCP.

Conjuntos	
\mathcal{G}	: conjunto de unidades de generación.
\mathcal{T}	: conjunto de períodos de tiempo.
\mathcal{S}_g	: conjunto de tipo de partidas de la unidad $g \in \mathcal{G}$.
Parámetros	
CG_g	: costo unitario de generación del generador $g \in \mathcal{G}$.
CNL_g	: costo sin carga del generador $g \in \mathcal{G}$.
CSU_{gs}	: costo de partida tipo $s \in \mathcal{S}_g$ del generador $g \in \mathcal{G}$.
P_g^{\min}	: potencia mínima del generador $g \in \mathcal{G}$.
P_g^{\max}	: potencia máxima del generador $g \in \mathcal{G}$.
RU_g	: límite de aumento o toma de carga del generador $g \in \mathcal{G}$.
RD_g	: límite de disminución o desprendimiento de carga del generador $g \in \mathcal{G}$.
TU_g	: tiempo mínimo de encendido del generador $g \in \mathcal{G}$.
TD_{gs}	: tiempo mínimo de apagado para partida tipo $s \in \mathcal{S}_g$ del generador $g \in \mathcal{G}$.
D_t	: demanda de energía en el período $t \in \mathcal{T}$.
R_t	: requerimiento de reserva para el período $t \in \mathcal{T}$.
Variables	
u_{gt}	: estado on/off de la unidad $g \in \mathcal{G}$ en el período $t \in \mathcal{T}$.
v_{gt}	: decisión de encendido de la unidad $g \in \mathcal{G}$ en el período $t \in \mathcal{T}$.
z_{gts}	: decisión de encendido tipo $s \in \mathcal{S}_g$ de la unidad $g \in \mathcal{G}$ en el período $t \in \mathcal{T}$.
w_{gt}	: decisión de apagado de la unidad $g \in \mathcal{G}$ en el período $t \in \mathcal{T}$.
r_{gt}	: reserva disponible de la unidad $g \in \mathcal{G}$ en el período $t \in \mathcal{T}$.
p_{gt}	: potencia generada por la unidad $g \in \mathcal{G}$ en el período $t \in \mathcal{T}$.

La formulación presentada en esta sección está basada en el trabajo de [Morales-España et al. \(2013\)](#) y, exceptuando las restricciones de rampa, equivale a la misma formulación que utilizamos para contrastar nuestros resultados. Sólo se considera un nodo de demanda, por lo tanto no se modelan restricciones de transmisión. Por simplicidad, el requerimiento de reserva es omitido.

$$\min \sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}} \left(\sum_{s \in S_g} CSU_{gs} z_{gts} + CNL_g u_{gt} + CG_g p_{gt} \right) \quad (2.19)$$

$$\text{s.a.} \quad \sum_{g \in \mathcal{G}} p_{gt} = D_t \quad \forall t \in \mathcal{T} \quad (2.20)$$

$$P_g^{\min} u_{gt} \leq p_{gt} \leq P_g^{\max} u_{gt} \quad \forall t \in \mathcal{T}, \forall g \in \mathcal{G} \quad (2.21)$$

$$v_{gt} - w_{gt} = u_{gt} - u_{gt-1} \quad \forall t \in \mathcal{T}, \forall g \in \mathcal{G} \quad (2.22)$$

$$\sum_{i=t-TU_g+1}^t v_{gi} \leq u_{gt} \quad \forall t \in \mathcal{T}, \forall g \in \mathcal{G} \quad (2.23)$$

$$\sum_{i=t-TD_g+1}^t w_{gi} \leq 1 - u_{gt} \quad \forall t \in \mathcal{T}, \forall g \in \mathcal{G} \quad (2.24)$$

$$\sum_{s \in S_g} z_{gts} = v_{gt} \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \quad (2.25)$$

$$z_{gts} \leq \sum_{i=TD_{gs}}^{TD_{gs+1}-1} w_{gt-i} \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \quad (2.26)$$

$$p_{gt} - p_{gt-1} \leq RU_g u_{gt} \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \quad (2.27)$$

$$p_{gt-1} - p_{gt} \leq RD_g u_{gt} \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \quad (2.28)$$

$$u_{gt}, v_{gt}, w_{gt}, z_{gts} \in \{0, 1\}, p_{gt} \geq 0 \quad \forall s \in S_g, \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \quad (2.29)$$

En la formulación, la función objetivo a minimizar representa los costos totales en todo el horizonte de planificación, específicamente costos de generación, costos fijos de funcionamiento y costos de partida. La restricción (2.20) relaciona la potencia total generada con el cumplimiento de la demanda para cada período. La restricción (2.21) representa los límites de generación para cada unidad. La restricción (2.22) relaciona temporalmente las variables de encendido y apagado con la variable de unidad activa. La restricción (2.23) y (2.24) cumplen el tiempo mínimo de encendido y apagado para cada unidad, respectivamente. La restricción (2.25) y (2.26) definen el tipo de partida de la unidad de acuerdo al tiempo de apagado de la misma. Las restricciones (2.27) y (2.28) permiten modelar las rampas de toma y deprendimiento de carga de las unidades, i.e., los límites en la variación de potencia generada entre períodos consecutivos. Finalmente, la naturaleza de las variables está dada en (2.29).

2.4.2. Cálculo de precios

De acuerdo a la revisión bibliográfica, los métodos de resolución propuestos hasta ahora para el problema CHP se pueden clasificar en dos enfoques: (i) dual Lagrangiano y (ii) primal. El primero está basado en resolver el dual Lagrangiano de la formulación del UCP cuando la restricción de demanda es relajada

mediante el uso de multiplicadores. Según lo mostrado en [Gribik et al. \(2007\)](#), el vector de multiplicadores óptimos del dual Lagrangiano pertenece al subgradiente de la envoltura convexa de la función de costos parametrizada en la demanda y por lo tanto, corresponde a los precios del método CHP. Considere el dual Lagrangiano del UCP formulado anteriormente:

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^{|\mathcal{T}|}} \left\{ \min_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}} \left(\sum_{s \in S_g} CSU_{gs} z_{gts} + CNL_g u_{gt} + CG_g p_{gt} \right) - \sum_{t \in \mathcal{T}} \lambda_t (D_t - \sum_{g \in \mathcal{G}} p_{gt}) \right\} \\ \text{s.a. (2.21) - (2.29)} \end{aligned}$$

Se tiene un problema dual Lagrangiano que puede ser resuelto iterativamente mediante el método del subgradiente. Notar que dada una solución al problema de minimización interno, entonces los multiplicadores óptimos del problema externo corresponden a los precios que maximizan los beneficios de todas las unidades. De igual forma, dado un vector de multiplicadores, el problema interno consiste en definir individualmente la operación que maximiza el beneficio de cada unidad.

Si bien es posible aprovechar la estructura separable del problema, aún es necesario resolver una cantidad de problemas MILP equivalentes al número de unidades del problema, lo cual puede resultar computacionalmente costoso. Lo anterior puede ser simplificado si las restricciones (2.21)-(2.28) definen exactamente la envoltura convexa de su respectivo espacio factible. Sin embargo, incluso teniendo la envoltura convexa, existen oscilaciones en la convergencia cuando se utiliza el método del subgradiente tradicional, debido a la estructura de la función dual Lagrangiana [Wang et al. \(2013a\)](#). De manera de solucionar este problema, se han desarrollado variaciones al método del subgradiente. En [Wang et al. \(2013a\)](#) se utiliza el método de subgradiente en conjunto con un procedimiento de planos cortantes que va eliminando los multiplicadores no óptimos, lo cual logra reducir el tiempo computacional pero carece de robustez [Stevens et al. \(2015\)](#). En [Wang et al. \(2013b\)](#), en vez de utilizar el subgradiente tradicional se resuelve un problema cuadrático convexo para encontrar una dirección de mejora en el algoritmo. El procedimiento converge en un número finito de iteraciones pero involucra encontrar y almacenar todos los puntos extremos del problema de minimización. A pesar de que los solucionadores comerciales poseen herramientas para realizar esta tarea, el número de puntos extremos crece exponencialmente con el número de unidades, lo cual es relevante si se consideran problemas de tamaño real.

Un enfoque más reciente para el problema CHP es resolver el problema primal, el cual corresponde a la relajación continua de la formulación UCP. De esta forma, los precios se obtienen a partir de las variables duales de la restricción de demanda (2.20) si y solo si, las restricciones (2.21)-(2.28) definen exactamente la envoltura convexa del poliedro respectivo [Schiro et al. \(2016\)](#), [Hua y Baldick \(2016\)](#). Cuando las restricciones definen la envoltura convexa, la relajación Lagrangiana y la relajación continua del problema tienen el mismo valor de función objetivo y los multiplicadores Lagrangianos óptimos son equivalentes a las variables duales

óptimas de la restricción relajada [Hua y Baldick \(2016\)](#). Comparado con los métodos que resuelven el dual Lagrangiano, este enfoque es más sencillo debido a que requiere resolver directamente un problema de programación lineal (LP, *linear programming*), lo cual puede ser realizado por cualquier solucionador que utiliza métodos basados en simplex. Sin embargo, definir la envoltura convexa de un conjunto no-convexo puede ser difícil debido a que un cambio en el modelamiento requiere desarrollar restricciones o desigualdades válidas especiales. En [Rajan et al. \(2005\)](#) las restricciones definen la envoltura convexa de un conjunto que considera límites de generación, tiempo mínimo de encendido y apagado, costos de generación lineales y costos de partida independientes del tiempo de apagado. En [Morales-España et al. \(2013\)](#), se incluyen rampas de partida y apagado. En [Pan y Guan \(2016\)](#) se desarrollan una serie de desigualdades válidas para incluir restricciones de rampas entre períodos consecutivos para máximo tres períodos de tiempo. Es importante notar que en ausencia de las restricciones de rampas, la envoltura convexa se puede definir para la mayoría de las características de las unidades, independiente del número de períodos de tiempo. En concreto, las restricciones (2.21)-(2.25) definen una envoltura convexa si sólo se considera un tipo de partida.

2.4.3. Enfoque propuesto

Una descripción alternativa de la envoltura convexa puede ser construida a partir de la representación de red de las operaciones factibles de las unidades de generación. Tradicionalmente, la representación de red ha sido utilizada en programación dinámica para determinar la programación y despacho óptimos de una unidad dado un conjunto de multiplicadores en la relajación Lagrangiana del UCP, [Fan et al. \(2002\)](#); [Guan et al. \(1992\)](#). En este trabajo se utiliza la representación de red para construir una formulación monolítica del problema CHP y obtener los precios mediante la resolución del problema primal. La principal idea de utilizar la red es disponer de flexibilidad en el modelamiento. La contribución del enfoque propuesto es doble:

- Se desarrolla una nueva formulación del problema CHP basada en la representación de red mostrada en [Guan et al. \(1992\)](#). Dado que las restricciones de red definen un poliedro con una matriz de coeficientes totalmente unimodular [Nemhauser y Wolsey \(1988\)](#), la formulación desarrollada describe exactamente la envoltura convexa de las características representadas con la red y por lo tanto se pueden obtener los precios resolviendo un problema LP. Cabe señalar que a excepción de las rampas, la mayoría de las características de las unidades pueden ser incluidas en la representación de red, sin necesidad de alterar la formulación obtenida.
- Se utiliza un enfoque de solución primal-dual para el problema CHP basado en el algoritmo de Bienstock y Zuckerberg (BZ) desarrollado en [Bienstock y Zuckerberg \(2015\)](#). El método de solución implementado permite obtener tiempos comparables e incluso menores a lo obtenido con los enfoques más recientes para resolver el problema CHP. Por otra parte, el algoritmo permite realizar un estudio del comportamiento de los pagos adicionales respecto al parámetro de optimalidad utilizado para obtener los precios.

Ejemplo

De manera de mostrar algunas propiedades del método CHP y contrastar con el método basado en costo marginal (LMP), la formulación UCP se aplica a un caso de dos unidades y un período de tiempo. No es necesario considerar restricciones intertemporales, por lo tanto sólo se restringe la generación. Por simplicidad sólo se consideran costos variables de generación y costos fijos de funcionamiento en las unidades. La formulación resultante es la siguiente:

$$\begin{aligned} \underset{u,p}{\text{mín}} \quad & 10p_1 + 50p_2 + 100u_1 + 50u_2 \\ & p_1 + p_2 = D \\ & 10u_i \leq p_i \leq 50u_i \quad \forall i \in \{1, 2\} \\ & u_i \in \{0, 1\} \quad \forall i \in \{1, 2\} \end{aligned}$$

Recordar que las variables binarias $u_i \forall i \in \{1, 2\}$ representan el estado de la unidad, encendido o apagado, mientras que las variables continuas $p_i \forall i \in \{1, 2\}$ representan el nivel de generación. El objetivo consiste en cumplir la demanda D al costo mínimo. La función de costo total mínimo según nivel de demanda y la envoltura convexa de dicha función se grafican en la figura 2.1.

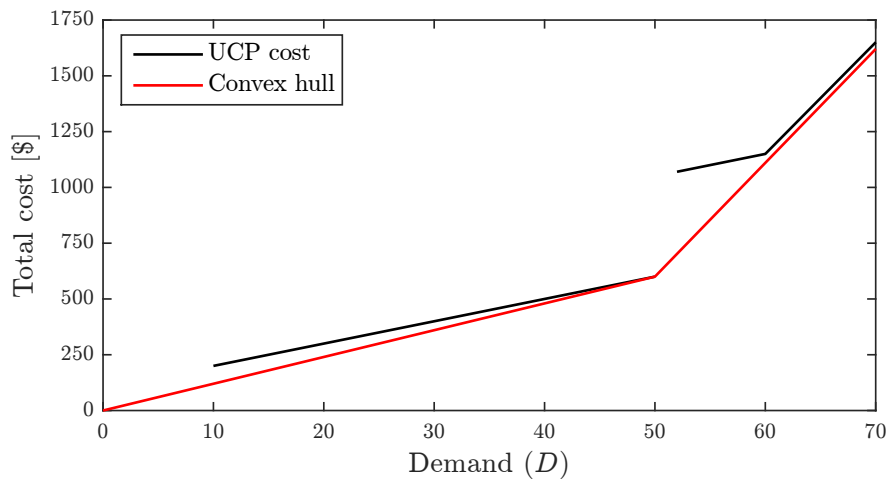


Figura 2.1: Función de costo mínimo según nivel de demanda y envoltura convexa de la función.
(Fuente: elaboración propia.)

Notar que la función de costo total parametrizada por la demanda es no convexa, presenta discontinuidades y saltos. Para resaltar las diferencias entre la función y su envoltura convexa se acota el dominio de la demanda a $[0, 70]$. Con el método de precios basado en costo marginal, el precio sería equivalente a la pendiente de esta curva, por lo que se identifican dos tramos: \$10 cuando $D \in [10, 60]$ y \$50 cuando $D \in (60, 100]$. Con el método CHP, el precio equivale a la pendiente de la envoltura convexa de la función de costos parametrizada. También se identifican dos tramos: \$12 cuando $D \in [10, 50]$ y \$51 cuando

$D \in (50, 100]$. Para cuantificar las diferencias respecto a pagos entre ambos métodos de precios, la tabla 2.2 muestra resultados para distintos niveles de demanda.

Tabla 2.2: Resultados del caso ejemplo por unidad y método de precio.

Demanda	Beneficio 1		MWP 1		LOC 1		Beneficio 2		MWP 2		LOC 2	
	CHP	LMP	CHP	LMP	CHP	LMP	CHP	LMP	CHP	LMP	CHP	LMP
18	-64	-100	64	100	64	100	0	0	0	0	0	0
52	1622	-100	0	100	328	100	-40	-450	40	450	40	450
59	1909	-100	0	100	41	100	-40	-450	40	450	40	450
65	1950	1900	0	0	0	0	-35	-50	35	50	35	50

La tabla 2.2 muestra los resultados de beneficio y pago adicional según el método CHP y el método LMP. El beneficio se calcula como la diferencia entre los ingresos que obtiene la unidad por pago de energía y los costos operacionales. Recordar que el MWP corresponde a la cantidad necesaria para alcanzar beneficio cero y sólo se realiza cuando el actual beneficio es negativo. Por otra parte, el LOC es la diferencia entre el máximo beneficio que podría obtener la unidad y el beneficio actual obtenido a un precio determinado. En este ejemplo, se da la situación de que en algunos niveles de demanda el LOC es igual al MWP. Estas situaciones ocurren cuando el beneficio actual es negativo y la mejor decisión que puede tomar una unidad a un determinado precio es permanecer sin funcionar obteniendo beneficio cero. Es importante aclarar que el LOC mínimo con el método CHP es un concepto global, pues en el caso resaltado en azul, el pago LOC es mayor para la unidad 1 pero es compensado con un LOC menor para la unidad 2. Lo anterior se observa en la tabla 2.3.

Tabla 2.3: Resultados del caso ejemplo a nivel agregado según método de precio.

Demanda	Costo	Pago por energía		MWP		LOC	
		CHP	LMP	CHP	LMP	CHP	LMP
18	280	216	180	64	100	64	100
52	1070	2652	520	40	550	368	550
59	1140	3009	590	40	550	81	550
65	1400	3315	3250	35	50	35	50

Puede resultar preocupante el aumento en el pago por energía que resulta con el método CHP. Por ejemplo, en azul se resalta un caso que ocurre en el tramo de demanda $(50, 60]$, donde el método CHP propone un precio de \$51 a diferencia del precio \$10 del método LMP. Claramente, en este caso el método CHP no resulta conveniente para el consumidor. Sin embargo, hay otras lecturas del mismo resultado: en el método LMP, el pago adicional corresponde a más del 90 % del pago por energía, mientras que en el método CHP este porcentaje está bajo el 3 %. Con los últimos resultados, se tiene un mercado más transparente y un precio que refleja mejor las transacciones del mercado. En el caso resaltado, el precio LMP sólo justifica el 51 % del pago total.

Evidentemente, lo ideal sería disminuir el pago adicional y el pago de energía simultáneamente, pero como se menciona anteriormente esto corresponde a un problema de optimización más complejo no abordado

en este trabajo. La contribución de este trabajo radica en un nuevo enfoque de modelamiento y solución para el método CHP.



3 | Metodología

El método desarrollado en este trabajo consta de una parte de modelamiento, donde se presenta el modelo de red utilizado y se desarrolla la formulación para abordar el problema de precios, y otra parte que explica el algoritmo implementado para la resolución del problema.

3.1. Modelo de red

A partir de red acíclica presentada en Guan et al. (1992) se desarrolla una formulación basada en redes con variables semicontinuas. La estructura de la red utilizada permite incluir la mayoría de las características de las unidades abordadas en la literatura. La red representa todas las operaciones de programación factible para las unidades: los nodos representan el número de períodos que la unidad ha estado apagada o encendida en cada período de tiempo, mientras que los arcos representan transiciones factibles entre períodos de tiempo consecutivos. En general, el tamaño de la red está determinado horizontalmente por el horizonte de tiempo del problema y verticalmente por la suma del tiempo mínimo de apagado de la partida más fría de la unidad y el tiempo mínimo de encendido. La figura 3.1 muestra un ejemplo de la representación de red, para una unidad con un tiempo mínimo de encendido de 3 períodos, un tiempo mínimo de apagado de 2 períodos y un horizonte de tiempo de 4 períodos.

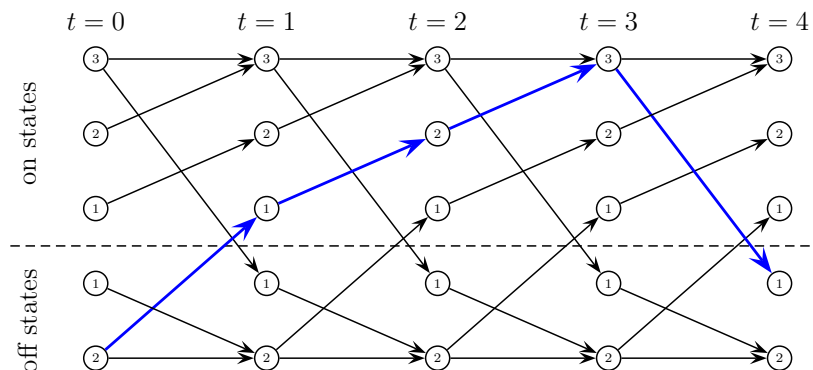


Figura 3.1: Ejemplo de red para representar el funcionamiento de la unidad.
(Fuente: elaboración propia.)

Los nodos o estados de la red se clasifican en encendido y apagado. A modo de ejemplo, los nodos incluyen números que representan el número de períodos que la unidad ha estado encendida o apagada. En particular, en los nodos extremos de cada período del horizonte de tiempo, el número indica que la unidad lleva al menos tal cantidad de períodos encendida o apagada. De esta forma, una ruta como la resaltada en la red representa una programación factible para la unidad, la cual satisface restricciones de mínimo tiempo de apagado y encendido y restricciones lógicas [Rajan et al. \(2005\)](#). Por ejemplo, la ruta resaltada en la figura 3.1 indica que la unidad lleva 2 o más períodos apagada al inicio del horizonte de planificación, se enciende en el primer período y permanece en ese estado hasta el tercer período, donde se apaga. Se observa en la ruta que el único modo de pasar al estado apagado es cumplir con con al menos 3 períodos de encendido, lo cual ocurre en el tercer período de planificación.

Notar que los arcos de la red mostrada pueden ser clasificados de acuerdo al tipo de estado de origen y destino, lo cual genera cuatro categorías: *on-on*, *on-off*, *off-on* y *off-off*. Considerando la notación dada en la tabla 2.1, los parámetros de una unidad se relacionan a la red de acuerdo a lo mostrado en la tabla 3.1.

Tabla 3.1: Parámetros de la red para una unidad g arbitraria.

Tipo de arco	Costo fijo	Costo variable	Potencia mínima	Potencia máxima
<i>on-on</i>	CNL_g	CG_g	P_g^{\min}	P_g^{\max}
<i>on-off</i>	CSD_g	0	0	0
<i>off-on</i>	$CNL_g + CSU_{g1}$	CG_g	P_g^{\min}	P_g^{\max}
<i>off-off</i>	0	0	0	0

Una ventaja de la representación en red recae en que la inclusión de características puede ser fácilmente realizada en una etapa de preproceso, modificando nodos y arcos de la red. En el ejemplo de red y los parámetros mostrados en la tabla 3.1 se trata de una unidad con costos de partida constantes y costos de generación lineales. En la práctica, los costos de generación son cuadráticos, y se suelen aproximar por funciones lineales por tramos. Para representar esto en la red, bastaría con generar arcos según el número de tramos de la función de costos entre los pares de nodos *on-on* y *on-off*, de modo que cada arco tiene los límites de generación y costos respectivos al tramo que representa. Si se quiere incluir distintos tipos de partida, se deben agregar más estados *off* y los arcos de la categoría *off-on* deben dividirse según el estado de origen, el cual representa el número de períodos que la unidad ha estado apagada, cf. [Guan et al. \(1992\)](#). El estado de origen determina el tipo de partida y por lo tanto el costo respectivo. Para las rampas de encendido, se pueden definir límites de generación diferentes para los arcos *off-on* que llegan al primer estado de encendido. El mismo procedimiento puede aplicarse para las rampas de apagado. La flexibilidad en el modelamiento otorgada por la red permite desarrollar una formulación más robusta en comparación a formulaciones más compactas encontradas en la literatura.

3.1.1. Formulación del modelo

De manera de construir una formulación a partir de la red mostrada, es necesario asociar no sólo parámetros a los arcos sino que también variables. Dada una red que representa a la unidad $g \in \mathcal{G}$, sean \mathcal{V}_g y \mathcal{E}_g el conjunto de nodos y arcos, respectivamente. Sea $\mathcal{E}_g(t) \subseteq \mathcal{E}_g$ el subconjunto de arcos que llega a un nodo o estado en el tiempo $t \in \mathcal{T}$. Sea x_e una variable binaria asociada con el uso del arco $e \in \mathcal{E}_g$ y sea z_e una variable continua no negativa relacionada con la potencia generada en el arco $e \in \mathcal{E}_g$. Sean \underline{P}_e y \overline{P}_e los parámetros de potencia mínima y máxima del arco $e \in \mathcal{E}_g$. Entonces, se debe cumplir lo siguiente:

$$\underline{P}_e x_e \leq z_e \leq \overline{P}_e x_e \quad \forall e \in \mathcal{E}_g \quad (3.1)$$

Es decir, la potencia de un arco debe estar dentro de los límites si el arco está siendo utilizado. A partir de (3.1) se puede derivar la siguiente relación:

$$\begin{aligned} 0 &\leq z_e - \underline{P}_e x_e \leq \overline{P}_e x_e - \underline{P}_e x_e & \forall e \in \mathcal{E}_g \\ 0 &\leq \frac{z_e - \underline{P}_e x_e}{\overline{P}_e - \underline{P}_e} \leq x_e & \forall e \in \mathcal{E}_g \\ 0 &\leq y_e \leq x_e & \forall e \in \mathcal{E}_g, \end{aligned} \quad (3.2)$$

donde se utiliza la variable semicontinua $y_e \forall e \in \mathcal{E}_g$, la cual tiene la interpretación de fracción de potencia generada sobre el mínimo y está acotada por la variable binaria $x_e \forall e \in \mathcal{E}_g$. Por lo tanto, la generación de potencia de una unidad por período de tiempo se expresa de la siguiente forma:

$$\sum_{e \in \mathcal{E}_g(t)} (q_e x_e + s_e y_e) = p_{gt} \quad \forall t \in \mathcal{T}, g \in \mathcal{G} \quad (3.3)$$

En (3.3), $q_e = \underline{P}_e$ y $s_e = \overline{P}_e - \underline{P}_e$ son parámetros preprocesados de la red para simplificar la notación. Por otra parte, los costos fijos como los costos de partida pueden ser incluidos en los costos fijos de la red, mientras que los costos de generación pueden ser considerados en los costos variables como muestra la tabla 3.1. Sean CF_e y CV_e los costos fijos y variables del arco $e \in \mathcal{E}_g$, respectivamente. Entonces, el costo de la red queda definido por la siguiente función:

$$\sum_{e \in \mathcal{E}_g} (c_e x_e + f_e y_e) \quad \forall g \in \mathcal{G} \quad (3.4)$$

En (3.4), $c_e = q_e CV_e + CF_e$ y $f_e = s_e CV_e$ son parámetros preprocesados de la red para simplificar la notación. Como se menciona anteriormente, una programación factible de la unidad corresponde a una ruta en la red. El estado de la unidad al inicio del horizonte de planificación es conocido, por lo tanto la ruta

queda definida por restricciones de flujo o divergencia. Sean $\delta^+(i)$ y $\delta^-(i)$ los conjuntos de arcos que salen e ingresan al nodo $i \in \mathcal{V}_g$. Sea b_i un parámetro asociado al nodo $i \in \mathcal{V}_g$, entonces la restricción de flujo en cada nodo es la siguiente:

$$\sum_{e \in \delta^+(i)} x_e - \sum_{e \in \delta^-(i)} x_e = b_i \quad \forall i \in \mathcal{V}_g, \forall g \in \mathcal{G} \quad (3.5)$$

En particular, el parámetro b_i sólo toma valor 1 en el estado inicial de la unidad, en el período $t = 0$. Dado que en todos los demás estados el parámetro b_i no tiene valor, la restricción de divergencia asegura que se genera una única ruta en la red. Considere ahora la formulación completa del modelo. Por simplicidad, se omiten las restricciones de requerimiento de reserva. Para facilitar la exposición de resultados, en lo sucesivo la formulación será denotada como NFU:

$$\text{NFU:} \quad \min_{x,y} \sum_{g \in \mathcal{G}} \sum_{e \in \mathcal{E}_g} (c_e x_e + f_e y_e) \quad (3.6)$$

$$\text{s.t.} \quad \sum_{g \in \mathcal{G}} \sum_{e \in \mathcal{E}_g(t)} (q_e x_e + s_e y_e) = D_t \quad \forall t \in \mathcal{T} \quad (3.7)$$

$$\sum_{e \in \delta^+(i)} x_e - \sum_{e \in \delta^-(i)} x_e = b_i \quad \forall i \in \mathcal{V}_g, \forall g \in \mathcal{G} \quad (3.8)$$

$$0 \leq y_e \leq x_e \quad \forall e \in \mathcal{E}_g, \forall g \in \mathcal{G} \quad (3.9)$$

$$x_e \in \{0, 1\} \quad \forall e \in \mathcal{E}_g, \forall g \in \mathcal{G} \quad (3.10)$$

La función objetivo (3.6) busca minimizar los costos fijos y variables de generación. La restricción (3.7) corresponde a la satisfacción de demanda en cada período del horizonte de planificación. La restricción (3.8) cumple la divergencia en la red. La restricción (3.9) permite acotar la variable de potencia generada en los arcos, para que se cumplan las ecuaciones derivadas previamente. Finalmente, la restricción (3.10) restringe la activación de los arcos a una decisión binaria. Notar que el modelo es un problema de flujo en redes con restricciones adicionales o *side constraints*. Específicamente, la demanda impone una restricción sistémica que acopla a todas las redes que representan a las unidades. Dado que las restricciones de flujo definen un poliedro cuya matriz de coeficientes totalmente unimodular [Nemhauser y Wolsey \(1988\)](#), las restricciones (3.8)-(3.10) definen la envoltura convexa del poliedro para las características de la unidad abordadas en la estructura de la red.

Teniendo la envoltura convexa del espacio de operación factible de las unidades, basta con resolver la relajación continua del problema para obtener los precios a partir de las variables duales óptimas de la restricción (3.7). En otras palabras, se debe resolver la relajación en programación lineal del problema NFU, lo que equivale a asumir como continuas las variables binarias. En adelante, esta relajación será denotada como NFU-LP.

Para tener una primera aproximación de los resultados, se realiza una prueba a la formulación NFU-LP, contrastando a esta última con la formulación compacta mostrada en [Hua y Baldick \(2016\)](#), la cual ha sido modificada para incluir costos de partida dependientes del tiempo de acuerdo a lo presentado en [Morales-España et al. \(2013\)](#). Esta formulación corresponde al modelo UCP de la sección anterior, considerando como continuas las variables binarias. En adelante, esta formulación relajada es denotada como COM-LP y de acuerdo a lo investigado corresponde a la mejor formulación compacta en la literatura. La primera prueba se lleva a cabo en AMPL usando CPLEX 12.6.1 sobre un conjunto de instancias de 48 y 120 horas (períodos de tiempo), considerando entre 50 y 500 unidades de generación. El tiempo de solución es obtenido a partir de la solución del nodo raíz cuando se resuelve la relajación continua del problema MILP. La media geométrica de los resultados se muestra en la tabla 3.2. Si bien es sabido que la formulación de red es más extensa, se observa que el tiempo de solución para instancias de gran tamaño se vuelve comparable debido a las técnicas de preproceso que incluyen los solvers para problemas MILP, reducen considerablemente el tamaño de la red.

Tabla 3.2: Comparación de formulaciones en tiempo de solución.

Horizonte de tiempo	Media geométrica del tiempo de solución [s]	
	NFU-LP	COM-LP
48 horas	30.3	7.9
120 horas	111.0	115.9

La formulación basada en redes es un enfoque interesante debido a la flexibilidad en la inclusión de características para el modelamiento. Exceptuando las restricciones de rampas, otras como rampas de partida, costos de generación lineales por tramos, costos de partida dependientes del tiempo de apagado, son fácilmente consideradas cambiando la estructura de nodos y arcos de la red en vez de la formulación, a diferencia de lo sucedido con los enfoques primales actuales que resuelven el problema CHP. La formulación propuesta es más robusta, pero implica un mayor desafío en la resolución debido al aumento en el número de variables.

Lo anterior motiva a la búsqueda de un algoritmo adecuado para esta clase de problemas lineales de gran escala, de manera de lograr resultados más eficientes a lo mostrado en la tabla 3.2. Se observa que la formulación es un problema de red con restricciones adicionales [Bertsekas \(1998\)](#). Por lo tanto, la relajación lagrangiana de las restricciones adicionales deriva en un problema con la propiedad de integralidad y además separable por unidad de generación. Considerando los problemas de convergencia relacionados al tradicional método del subgradiente, decidimos implementar un algoritmo basado en el trabajo [Bienstock y Zuckerberg \(2015\)](#). En adelante nos vamos a referir a este método como algoritmo BZ. El esquema de descomposición utilizado en el algoritmo BZ permite explotar la estructura de nuestra formulación logrando una convergencia más rápida a la solución óptima.

3.2. Algoritmo BZ

En este trabajo se resuelve el problema NFU-LP utilizando un método basado en el algoritmo BZ. El algoritmo BZ es un método de descomposición adecuado para resolver problemas LP de gran escala [Munoz et al. \(2016\)](#). A pesar de que el algoritmo es genérico, la implementación realizada está basada en el enfoque de relajación Lagrangiana mostrado en [Bienstock y Zuckerberg \(2015\)](#), donde las restricciones adicionales son relajadas y el problema dual Lagrangiano resultante es sencillo de resolver. La solución de la relajación Lagrangiana es usada para generar restricciones sobre el problema principal NFU-LP, con el objetivo de lograr la convergencia a la solución óptima resolviendo una secuencia de problemas pequeños en vez de un problema de gran escala. De manera de explicar el funcionamiento del algoritmo BZ, considere la siguiente formulación matricial del problema NFU-LP:

$$\mathbf{NFU-LP} : \quad \min_{x,y} \quad c^T x + f^T y \quad (3.11)$$

$$\text{s.t.} \quad Qx + Sy = d \quad (\lambda) \quad (3.12)$$

$$(x, y) \in \mathcal{X}, \quad (3.13)$$

donde $\mathcal{X} = \{(x, y) \in \mathbb{R}_+^n : Ax = b, y \leq x \leq 1\}$. Dado un vector de multiplicadores λ , la relajación Lagrangiana del problema NFU-LP o bien “problema relajado”, se formula de la siguiente manera:

$$\mathbf{NFU-LP}_\lambda : \quad \min_{x,y} \quad c^T x + f^T y + \lambda^T (d - Qx - Sy) \quad (3.14)$$

$$\text{s.t.} \quad (3.13)$$

La solución al problema relajado entrega una cota inferior para el problema NFU-LP. A pesar de las variables semicontinuas en el poliedro, $y \in \mathcal{X}$, se mantiene la propiedad de integridad en el problema relajado como se muestra en la proposición 1.

Proposición 1. *Si A es TU, entonces el problema de programación lineal $z_{LP} = \min\{c^T x + d^T y : (x, y) \in \mathcal{X}\}$ tiene una solución óptima entera.*

Demostración. Considere la siguiente reformulación de z_{LP} :

$$z_{LP} = \min_{0 \leq x \leq 1} \{c^T x : Ax = b\} + \min_{y \geq 0} \{d^T y : y \leq x\}$$

Notar que la segunda expresión en la suma tiene una solución trivial, donde $y_e = x_e$ si $d_e < 0$, mientras que $y_e = 0$ en otro caso. Por lo tanto, la solución de la variable y es una función de la variable x . Sea $\underline{d}_e = \min(0, d_e)$, entonces:

$$\begin{aligned}
z_{LP} &= \min_{0 \leq x \leq 1} \{c^T x : Ax = b\} + \min_{x \geq 0} \underline{d}^T x \\
&= \min_{0 \leq x \leq 1} \{(c + \underline{d})^T x : Ax = b\}
\end{aligned}$$

Si la matriz A es TU, es bien sabido que el problema z_{LP} tiene una solución óptima entera [Nemhauser y Wolsey \(1988\)](#). □

El problema relajado es separable por unidad de generación y el método de solución para cada problema por separado es un algoritmo de ruta mínima. En el algoritmo BZ implementado, la solución del problema relajado es utilizada para generar particiones de variables, las cuales imponen restricciones en el problema NFU-LP tratando de suponer el patrón de su solución óptima. En adelante, el problema NFU-LP con las restricciones impuestas por las particiones será denotado como “problema restringido”. Como se muestra en [Bienstock y Zuckerberg \(2015\)](#), dado que la relajación Lagrangiana del problema NFU-LP posee la propiedad de integralidad, el número de valores fraccionarios distintos en su solución óptima está en el orden del número de restricciones adicionales, mientras que las demás variables están en los límites 0 y 1. Si el patrón generado con las particiones no resulta ser el óptimo, entonces las particiones se actualizan a partir de las variables duales de las restricciones adicionales (3.12) en el problema restringido. Sea H la matriz de coeficientes que define el nuevo conjunto de restricciones, entonces el problema restringido se formula de la siguiente manera:

$$\begin{aligned}
\text{NFU-LP}_H : \quad & \min_{x,y} \quad c^T x + f^T y \\
& \text{s.t.} \quad (3.12), (3.13) \\
& \quad \quad Hx = 0
\end{aligned} \tag{3.15}$$

Las restricciones (3.15) están basadas en las particiones generadas. Sea N el conjunto de índices de variables. Sea R el número de particiones, sea I_r la partición r -ésima del conjunto N y sea \mathcal{N} el conjunto de particiones. Entonces, las restricciones (3.15) obligan a que las variables relacionadas a una misma partición I_r tengan el mismo valor al resolver el problema restringido. La solución del problema restringido entrega una cota superior para el problema NFU-LP. Al resolver el problema restringido se obtienen nuevas variables duales a partir de las restricciones adicionales (3.12), las cuales a su vez permiten formar un nuevo problema relajado. Este proceso se repite hasta alcanzar un gap relativo de optimalidad entre la cota superior y la cota inferior del problema principal NFU-LP. El pseudocódigo de nuestra implementación del algoritmo BZ se muestra en el algoritmo 2.

Algorithm 2 Algoritmo basado en BZ.

Require: $\epsilon, k = 0, \lambda^{(k)} = \lambda_0, \mathcal{N}^{(k)} \equiv \cup_{r=1}^{R^{(k)}} I_r$.

1: Resolver NFU-LP $_{\lambda^{(k)}}$. Sea \bar{x} su solución óptima, entonces:

$$N^1 = \{i \in N : \bar{x}_i = 1\}, N^0 = \{i \in N : \bar{x}_i = 0\}$$

2: Actualizar particiones: $\mathcal{N}^{(k)} \leftarrow \{I_r \cap N^1, I_r \cap N^0 : I_r \in \mathcal{N}^{(k)}\}$

3: Generar restricciones $H^{(k)}x = 0$ tal que:

$$x_i = x_j \text{ if } (i, j) \in I_r \forall r \in \{1, \dots, R^{(k)}\}$$

4: Resolver NFU-LP $_{H^{(k)}}$. Sea $\bar{\lambda}$ su variable dual óptima:

5: **if** *optimality gap* $< \epsilon$ **then return** $\bar{\lambda}$ y finalizar

6: **else** $k \leftarrow k + 1, \lambda^{(k)} = \bar{\lambda}$ y volver a 1:

Una descripción formal y detallada del algoritmo BZ genérico se encuentra en [Bienstock y Zuckerberg \(2015\)](#), [Munoz et al. \(2016\)](#). En base a las técnicas sugeridas para mejorar el rendimiento del algoritmo en [Munoz et al. \(2016\)](#), en este trabajo se desarrolla una inicialización de particiones que evita resolver el problema restringido desde “cero”, i.e. con $\mathcal{N}^{(k)} \equiv N$. Por otro lado, también se realiza un preproceso a la red para reducir el tamaño del problema, tratando de emular lo realizado por los optimizadores de problemas MILP. La base de la implementación se explica a través de las siguientes subsecciones.

3.2.1. Preproceso de la red

Notar que el tamaño de la red puede ser considerablemente reducido en base al estado inicial, el cual es siempre un parámetro del problema. Dependiendo del nodo de partida, existen nodos que nunca son alcanzados y por tanto, los arcos respectivos no son utilizados. Esto permite eliminar variables de la formulación en una etapa previa a la resolución. Para mostrar el efecto del preproceso, considere la red mostrada en la figura 3.2. Esta red, en donde los arcos han sido etiquetados con una variable, corresponde a la red de la figura 3.1 preprocesada considerando que el nodo inicial es el tercer estado de encendido. En otras palabras, esta unidad de ejemplo inicia el horizonte de planificación con tres o más períodos de encendido. El preproceso logra reducir de 35 a 23 arcos.

Otro tipo de preproceso que se puede realizar tiene que ver con los nodos que tienen un arco de entrada y más de dos arcos de salida. Este tipo de nodos *fork* puede ser eliminado y el arco de entrada unido a cada arco de salida. Tanto este como el anterior tipo de preproceso buscan reducir el número de variables del problema. En nuestra implementación utilizamos sólo el primer tipo, pero como ejemplo se prueban ambos en un conjunto de instancias para mostrar el efecto del preproceso en el tiempo de solución. La tabla 3.3 muestra la media geométrica de los resultados obtenidos considerando sólo preproceso, sin la técnica de inicialización que se discute más adelante. A pesar de que el preproceso 2 permite eliminar más variables que

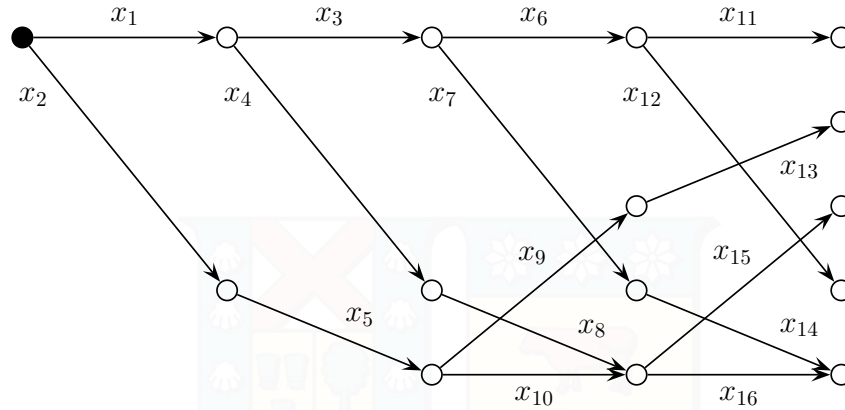


Figura 3.2: Red de la figura 3.1 preprocesada en base al estado inicial (nodo negro).
(Fuente: elaboración propia.)

el preproceso 1, la reducción en tiempo de solución puede ser menos en promedio debido a que se utiliza más tiempo en preprocesar. Como el segundo tipo de preproceso es una transformación no lineal de variables, podría ser más útil en un contexto de resolución de problemas MILP como la formulación NFU.

3.2.2. Inicialización

El procedimiento de inicialización busca generar particiones rápidamente a partir de valores duales aproximados. Para esto, se resuelve una nueva relajación del problema NFU-LP que resulta de eliminar las restricciones de flujo (3.13). Al hacer esto, el nuevo problema relajado puede ser separado por período de tiempo. El procedimiento de inicialización consiste en resolver una muestra de períodos de tiempo y obtener de los valores duales de la restricción (3.12) para los períodos resueltos. A partir de los pares demanda-dual de la muestra, se construye una función lineal por tramos de los valores duales respecto al nivel de demanda, lo que permite aproximar los demás valores duales de los períodos no considerados en la muestra. El problema relajado NFU-LP $_{\lambda}$ se resuelve con los valores duales aproximados para obtener particiones iniciales. Este procedimiento se repite un número determinado de veces para distintas muestras, permitiendo obtener diferentes aproximaciones de valores duales sin resolver el problema restringido. Por lo tanto, se obtienen múltiples particiones de manera rápida. Es importante notar que sin este procedimiento, cada red empezaría sólo con dos particiones. En resumen, la inicialización tiene dos parámetros: (i) el tamaño de la muestra que equivale a la cantidad de períodos de tiempo que se van a resolver por separado y (ii) el número de veces que se va a tomar una muestra, que equivale al número de aproximaciones construidas. Para decidir los parámetros para nuestros experimentos, se realizan pruebas con cuatro tipos de inicialización en el mismo conjunto de instancias de la subsección previa. La media geométrica de los tiempos de solución se muestra en la tabla 3.3. Se observa que una mala inicialización puede ser perjudicial para el algoritmo. Si bien el proceso para aproximar valores duales es sencillo y rápido, un tamaño de muestra muy grande puede

demandar tiempo considerable. En un buen caso de inicialización, se puede resolver incluso con un 30 % menos de tiempo en promedio.

Tabla 3.3: Efecto del preproceso y la inicialización sobre el tiempo de solución.

Inicialización	Media geométrica del tiempo [s]		
	Sin preproceso	Preproceso 1	Preproceso 2
Sin inicialización	76.8	74.9	75.3
1 muestra de 2 períodos	–	52.9	53.3
9 muestras de 5 períodos	–	76.0	76.1
10 muestras de 20 períodos	–	90.2	90.6

3.2.3. Particiones

En el problema restringido, el conjunto de variables es particionado. El procedimiento de generación y actualización de particiones es similar al utilizado en [Espinoza y Moreno \(2014\)](#). Recordar que la solución al problema relajado corresponde a rutas en las respectivas redes, en las cuales sólo una fracción de variables toma valor 1 mientras la mayoría se mantiene en 0. Siguiendo con la notación anterior donde N es el conjunto de índices de variables, entonces a partir de la solución del problema relajado el conjunto N puede ser dividido en dos subconjuntos: $N^0 = \{i \in N : x_i = 0\}$ y $N^1 = \{i \in N : x_i = 1\}$. Para generar las particiones, los conjuntos N^0 y N^1 de cada iteración son intersectados con las particiones de la iteración previa, i.e., $N^0 \cap I_r \forall r$ y $N^1 \cap I_r \forall r$. Como las rutas del problema relajado debieran ser diferentes debido a los sucesivos cambios en la función de costos penalizada (3.14), se van obteniendo distintos conjuntos N^0 y N^1 , por lo tanto las intersecciones aumentan el número de particiones R . Las particiones generadas equivalen a imponer el siguiente conjunto de restricciones de igualdad en el problema:

$$x_i - x_j = 0 \quad \forall i, j \in I_r, \forall r \in \{1, \dots, R\}$$

De esta forma, el problema restringido sólo soporta un número reducido de soluciones factibles dadas por las combinaciones de particiones. En cada iteración, se van agregando nuevas soluciones factibles cuando las particiones se vuelven más pequeñas. El método escogido para generar particiones tiene una propiedad fundamental: la actualización de restricciones nunca descarta una solución factible previamente generada. De esta forma, el problema restringido en una iteración es siempre una relajación del problema restringido previo, lo cual es una condición para que el algoritmo tenga una dirección de mejora estricta en la optimización del problema [Bienstock y Zuckerberg \(2015\)](#). El procedimiento aplicado a una red de ejemplo se muestra en las figuras 3.3, 3.4 y 3.5. Tomando en cuenta las variables que etiquetan al respectivo arco, se sabe que el conjunto $N = \{1, \dots, 16\}$. La figura 3.3 muestra las particiones en alguna iteración arbitraria (k), donde se tiene un total de $R = 3$ particiones del conjunto de variables, representadas con distintos tipos de línea. Siguiendo con la notación del algoritmo 2, el conjunto de particiones $\mathcal{N}^{(k)} = \{I_1, I_2, I_3\}$.

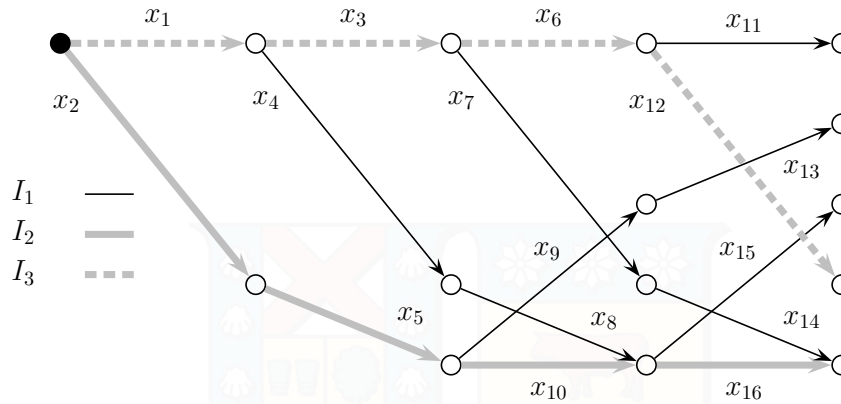


Figura 3.3: Particiones en iteración (k): $I_2 = \{2, 5, 10, 16\}$, $I_3 = \{1, 3, 6, 12\}$ y $I_1 = N \setminus \cup_{r=2}^3 I_r$.
(Fuente: elaboración propia.)

Por otro lado, la figura 3.4 muestra la solución a una red del problema relajado en la iteración k -ésima. La solución es una ruta a través de la red, en la cual sólo 4 arcos toman valor y los demás se mantienen nulos. Esta solución permite generar los ya mencionados conjuntos N^0 y N^1 .

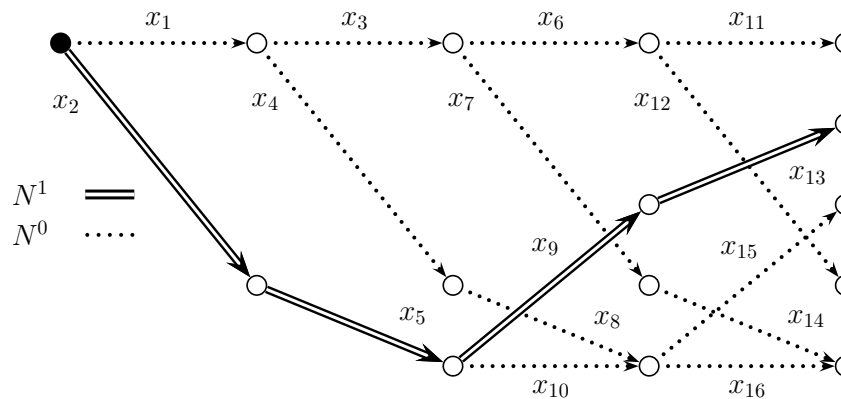


Figura 3.4: Solución al problema relajado en iteración (k): $N^1 = \{2, 5, 9, 13\}$ y $N^0 = N \setminus N^1$.
(Fuente: elaboración propia.)

El siguiente paso del algoritmo está representado en la figura 3.5, donde las particiones de la iteración (k) son actualizadas a partir de la solución al problema relajado. Las nuevas particiones surgen de la intersección de los conjuntos N^0 y N^1 con las particiones existentes: $N^1 \cap I_1 = \{9, 13\}$, $N^1 \cap I_2 = \{2, 5\}$, $N^0 \cap I_2 = \{10, 16\}$, $N^0 \cap I_3 = I_3$ y $N^0 \cap I_1 = I_1 \setminus \{9, 13\}$. Las particiones resultantes se representan con distintos tipos de línea. Siguiendo con la notación del algoritmo 2, ahora el total de particiones es $R = 5$ y el conjunto $\mathcal{N}^{(k+1)} = \{I_1, I_2, I_3, I_4, I_5\}$.

Notar que las particiones pueden mantener su tamaño o volverse más pequeñas. Las particiones determinan las restricciones en el problema restringido. Al tomar en cuenta la actualización de particiones, se puede notar que en la iteración (k) el problema restringido sólo admitía 2 soluciones equivalentes a las particiones I_2 y I_3 . En cambio, en la iteración ($k + 1$), el problema restringido admite 3 soluciones,

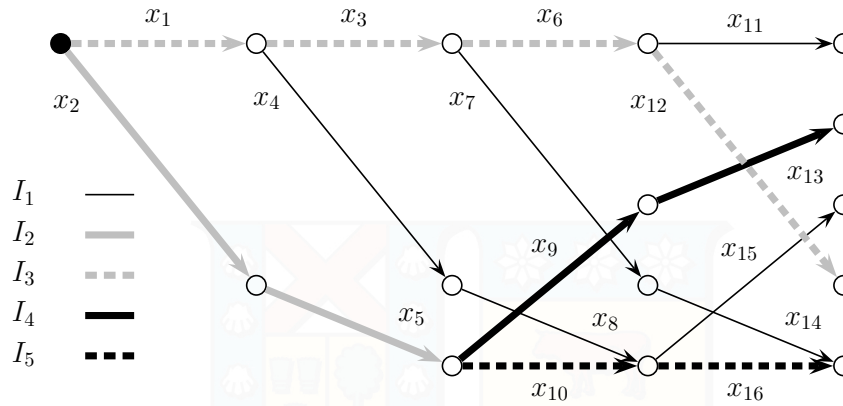


Figura 3.5: Particiones en iteración $(k + 1)$: $I_2 = \{2, 5\}$, $I_3 = \{1, 3, 6, 12\}$, $I_4 = \{9, 13\}$, $I_5 = \{10, 16\}$ y $I_1 = \mathbb{N} \setminus \cup_{r=2}^5 I_r$.
(Fuente: elaboración propia.)

conformadas por las uniones de particiones $I_2 \cup I_4$, $I_2 \cup I_5$ y la partición I_3 . Esto permite apreciar cómo el problema restringido se va relajando sucesivamente al admitir nuevas soluciones, producto de la intersección de particiones. Además, las soluciones del problema restringido en la iteración (k) aún son soluciones factibles en la iteración posterior.

4 | Resultados

Este capítulo se divide en tres secciones. La primera sección analiza con más detalle algunas características del algoritmo BZ como la inicialización, el preproceso y el número de iteraciones primal-dual. La segunda parte corresponde al estudio computacional donde se analizan los tiempos de resolución de la formulación propuesta, comparando el algoritmo implementado con los algoritmos disponibles en los solucionadores comerciales. La última sección analiza los resultados económicos al comparar la formulación propuesta con la mejor formulación compacta desde el punto de vista de los precios y pagos adicionales obtenidos.

4.1. Análisis de algoritmo BZ

4.1.1. Preproceso e inicialización

La implementación del algoritmo BZ realizada en este trabajo depende de algunos parámetros que pueden afectar considerablemente el desempeño en tiempo computacional. De manera de seleccionar parámetros adecuados, se estudia el tiempo de solución de un conjunto de 105 instancias. Las instancias comprenden entre 50 y 2000 unidades de generación para un horizonte de planificación de 48 horas. El requerimiento de reserva es omitido, es decir, la demanda de energía es la única restricción sistémica. En cuanto a las unidades, se modelan 2 tipos de costos de partida dependientes del tiempo de apagado y la función de costos de generación se aproxima mediante una función lineal por tramos de 3 intervalos. Por último, no se modelan restricciones de rampas para la generación.

Primero se analiza el efecto de los dos tipos de preproceso en el tiempo de solución. Recordar que el preproceso tipo 1 tiene relación con la eliminación de nodos y arcos en base al estado inicial de la red, mientras que el preproceso tipo 2 aprovecha la estructura de los nodos *fork* para acoplar el arco de entrada con cada arco de salida y eliminar arcos. La figura 4.1 muestra los ratios de tiempo computacional entre lo obtenido con preproceso y lo obtenido sin preproceso de cada instancia. Los resultados están en porcentajes y ordenados de menor a mayor en cada técnica de preproceso para facilitar la exposición. Cabe mencionar

que el promedio geométrico de los resultados fue mostrado en la tabla 3.3 del capítulo anterior. En el gráfico se observa que el preproceso 1 obtiene tiempos menores a la ejecución sin preproceso en todas las instancias, a diferencia del preproceso 2 que en aproximadamente el 9 % de las instancias excede el tiempo sin preproceso, utilizando un 15 % adicional en el peor caso. Para ambos tipos de preproceso, la reducción en tiempo en promedio está cerca del 3 % y si bien no existe mucha diferencia entre ambos, para los estudios computacionales posteriores se procede con el preproceso 1 debido a que presenta mayor estabilidad. Como se menciona en el capítulo anterior, el preproceso 2 puede resultar más adecuado en un contexto de solución de problemas de programación entera mixta.

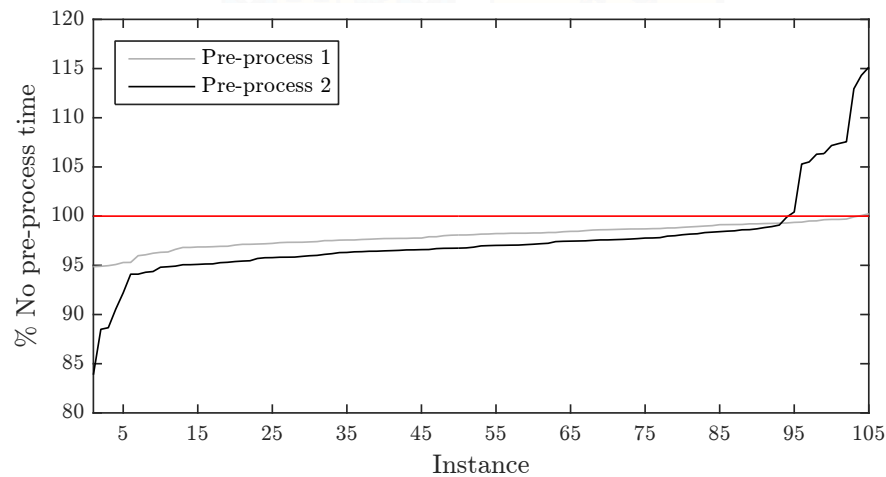


Figura 4.1: Gráfico de ratios de tiempo de solución según preproceso.
(Fuente: elaboración propia.)

Lo segundo que se analiza es el efecto de la inicialización. Recordar que la inicialización resuelve una muestra de problemas sencillos relacionados a distintos períodos de tiempo para obtener valores duales aproximados de todo el horizonte de tiempo. Los valores duales obtenidos son utilizados para generar las particiones y no empezar a resolver el problema principal desde cero. La figura 4.2 muestra los resultados de tiempo de solución de cada instancia, ordenados de menor a mayor, para 3 inicializaciones distintas: (i) 1 muestra de 2 períodos, (ii) 9 muestras de 5 períodos y (iii) 10 muestras de 20 períodos. Los resultados fueron resumidos en la tabla 3.3 del capítulo anterior. Se observa que la muestra más pequeña obtiene los mejores resultados. Posiblemente debido al tamaño del problema, una muestra muy grande o muchas repeticiones de la misma resultan contraproducentes. Por ejemplo, la inicialización (iii) repite 10 veces la resolución de 20 problemas y como resultado llega a tener una diferencia de alrededor de 100 segundos sobre el tiempo de solución para las instancias más complejas. Al momento de elegir los parámetros de la inicialización, parece relevante el concepto de muestra en estadística. La inicialización (iii) considera una selección de 20 períodos cuando el horizonte de planificación es de 48 períodos en todas las instancias. Caso contrario es la inicialización (i) que considera una muestra equivalente al 4 % del total de períodos de tiempo.

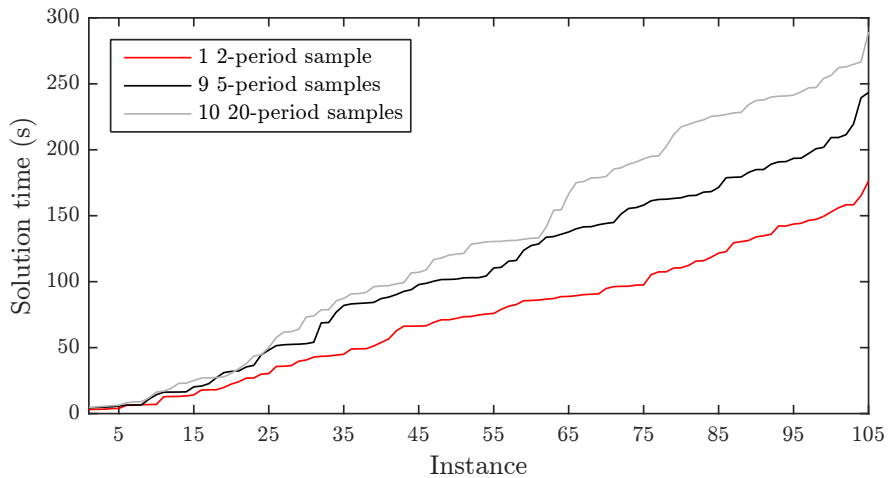


Figura 4.2: Gráfico de tiempo de solución según tipo inicialización.
(Fuente: elaboración propia.)

4.1.2. Iteraciones

El siguiente análisis considera adicionalmente un conjunto de 105 instancias de 24 horas y un conjunto de 105 instancias de 120 horas, con las mismas características de las instancias abordadas en el análisis de preproceso e inicialización. Los resultados corresponden al número de iteraciones primal-dual del algoritmo BZ, es decir, el número de veces que se resuelve el problema relajado, el problema restringido y se comprueba el criterio de optimalidad. La figura 4.3 muestra los resultados de iteraciones ordenados de menor a mayor y clasificados según el horizonte de tiempo de las instancias. El orden creciente de los resultados se elige para mostrarlos de manera más clara y también por la inexistencia de un patrón de la cantidad de iteraciones según el número de unidades de las instancias.

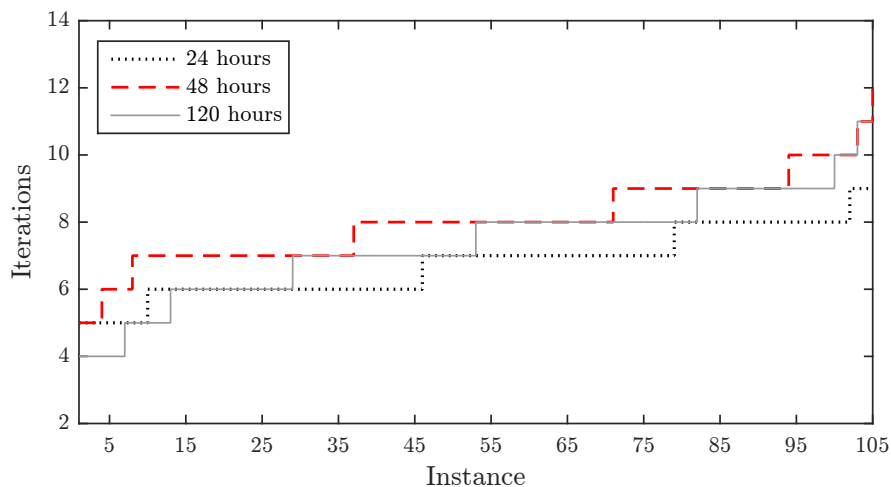


Figura 4.3: Gráfico de iteraciones por instancia y horizonte de tiempo.
(Fuente: elaboración propia.)

En términos globales, el máximo y el mínimo fue de 12 y 4 iteraciones, respectivamente. Se observa que existe poca variabilidad en el número de iteraciones respecto al horizonte de tiempo, puesto que las instancias con mayores diferencias no superan las 2 iteraciones. Lo anterior sugiere que el número de iteraciones no depende del número de períodos o el número de unidades y la complejidad de la instancia está principalmente determinada por las características de las unidades

4.2. Análisis comparativo

La escalabilidad de la formulación propuesta y el rendimiento del algoritmo implementado son puestos a prueba en instancias de gran escala generadas de manera aleatoria. El estudio evalúa 315 casos que consideran entre 50 y 2000 unidades para horizontes de tiempo de 24, 48 y 120 horas. Se modelan dos tipos de costos de partida dependientes del tiempo de apagado y la función de costos de generación es aproximada mediante una función lineal por tramos de 3 intervalos. No se modelan restricciones de rampas y el requerimiento de reserva se omite para los estudios.

El computador utilizado funciona con la arquitectura de 64-bit de Windows 10. Posee un procesador AMD de 8 núcleos y 12 GB de memoria RAM. El algoritmo implementado se compara con los algoritmos Dual simplex y Barrier que se encuentran disponibles en el solucionador comercial CPLEX 12.6.1. Por otro lado, también se prueban distintos parámetros para la tolerancia de optimalidad del algoritmo BZ. Cabe mencionar que el parámetro de tolerancia de 10^{-6} fue suficiente para alcanzar optimalidad en todas las instancias, es decir, con dicho parámetro las instancias alcanzaban un *optimality gap* igual a cero. Toda la implementación en esta sección se lleva a cabo utilizando lenguaje de programación C. El límite de tiempo para la resolución de instancias es de 1 hora y en todas se aplica el preproceso tipo 1 a las redes, independiente del algoritmo utilizado. En el algoritmo BZ, el problema restringido se resuelve con el algoritmo Barrier, mientras que el problema relajado se separa por unidad y se resuelve con un algoritmo de ruta mínima para redes acíclicas. Se utiliza una inicialización de 9 muestras de 5 períodos para las instancias de 120 horas, mientras que para las instancias de 24 y 48 horas se utiliza 1 muestra de 2 períodos.

Un análisis global de los resultados se obtiene utilizando los perfiles de rendimiento de [Dolan y Moré \(2002\)](#). Primero, se identifica el mejor tiempo de solución para cada instancia. Luego, para cada algoritmo se grafica la proporción de instancias resueltas dentro de un factor τ del mejor tiempo de solución. Si t_{ps} es el tiempo de solución del algoritmo s en la instancia p , y t_p^* es el mejor tiempo de solución obtenido en la instancia p , entonces el rendimiento del algoritmo queda caracterizado por la curva:

$$\rho_s(\tau) = \frac{\text{Número de instancias } p : t_{ps} \leq \tau t_p^*}{\text{Número de instancias}} \quad (4.1)$$

La figura 4.4 muestra los perfiles de rendimiento obtenidos para los algoritmos analizados. Se puede

notar que el algoritmo BZ, denotado como $BZ 10^{-6}$ pues es la tolerancia que resuelve a optimalidad, tiene una curva de rendimiento fija en el valor 1.0. Esto significa que el algoritmo fue el más rápido en cada una de las instancias, es decir, el total de las instancias son resueltas dentro del mejor tiempo de solución. No es el caso del algoritmo Dual simplex, el cual recién es capaz de resolver una proporción de 5 % de las instancias cuando emplea 5 veces el tiempo de solución del algoritmo BZ. El algoritmo Barrier es capaz de resolver todas las instancias, lo cual se concluye debido a que su curva de rendimiento alcanza la proporción total de 1.0 con el factor de 3.5 aproximadamente. Esto significa que en el peor tiempo de solución de Barrier se utilizó 3.5 veces el tiempo del algoritmo BZ. Además, en cerca del 50 % de las instancias, Barrier pudo resolver utilizando a lo más dos veces el mejor tiempo de solución.

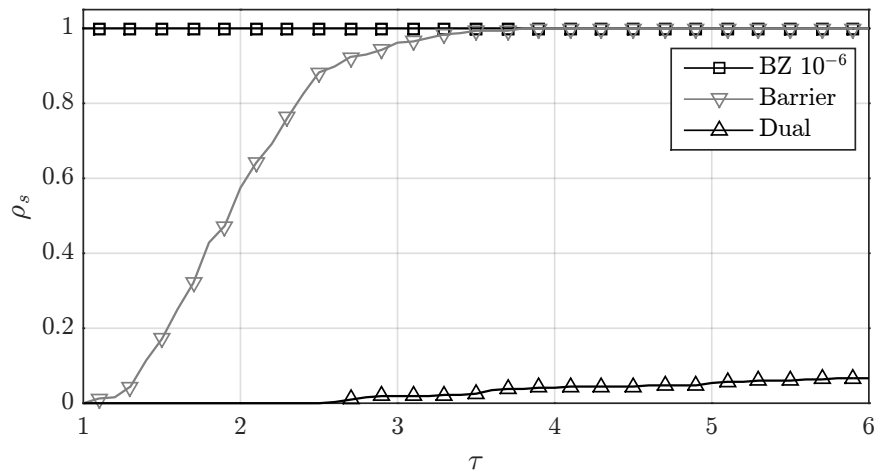


Figura 4.4: Perfiles de rendimiento de los algoritmos dentro de un factor de 6.
(Fuente: elaboración propia.)

Un análisis más detallado de los resultados se resume en la tabla 4.1. Las instancias han sido agrupadas en categorías denotadas $umin_umax_nper$, donde $umin$ y $umax$ se refieren al número mínimo y máximo de unidades de generación en la categoría, respectivamente, mientras que $nper$ corresponde al número de períodos del horizonte de planificación. En cada categoría se muestra la media geométrica de los tiempos de solución de cada algoritmo.

Notar que el algoritmo Dual simplex no sólo tiene el peor rendimiento, sino que también es incapaz de resolver algunas de las categorías dentro del tiempo de solución de 1 hora. Mientras el algoritmo BZ tiene mejor rendimiento que el algoritmo Barrier en todas las categorías, las diferencias se enfatizan en las instancias grandes, especialmente en aquellas con un alto número de períodos de tiempo. De esto se puede concluir que el incremento en períodos de tiempo es más crítico en Barrier. Como el horizonte de planificación define el tamaño de las redes, el incremento es más fácil de manejar con la descomposición realizada con el algoritmo BZ. Los números resaltados en la tabla 4.1 indican las categorías donde el *speedup* promedio del algoritmo BZ es más prominente. En estas categorías el algoritmo implementado puede llegar a ser en promedio hasta 2.5 veces más rápido que el algoritmo Barrier.

Tabla 4.1: Comparación de algoritmos según rendimiento computacional.

Categorías	Media geométrica del tiempo de solución [s]				
	Dual	Barrier	BZ 10^{-6}	BZ 10^{-4}	BZ 10^{-2}
50_500_24	46.7	7.7	5.2	4.7	4.3
50_500_48	148.5	21.1	13.5	12.3	11.0
600_1000_24	715.3	34.6	20.9	17.8	15.0
50_500_120	687.8	77.7	31.1	26.2	21.4
1100_1500_24	2003.2	65.0	36.2	30.4	25.6
1600_2000_24	3019.8	106.2	50.3	44.1	39.5
600_1000_48	1917.7	90.5	60.6	51.7	45.9
1100_1500_48	3521.5	175.7	92.2	77.1	66.3
600_1000_120	–	318.9	134.8	115.8	95.0
1600_2000_48	–	290.9	138.1	125.6	103.2
1100_1500_120	–	600.2	264.4	215.4	167.5
1600_2000_120	–	978.5	454.9	403.9	340.3

Las últimas columnas de la tabla 4.1 muestran el desempeño del algoritmo BZ con distintos parámetros para la tolerancia de optimalidad. Por ejemplo, BZ 10^{-4} finaliza su ejecución cuando el gap de optimalidad es menor o igual a 10^{-4} . El criterio de BZ 10^{-2} es aún más relajado. Cabe mencionar que en estos últimos dos algoritmos, se obtienen resultados subóptimos. Esta característica del algoritmo permite hacer los estudios de pagos adicionales con precios subóptimos en la próxima sección. La tabla 4.1 muestra que en promedio, el tiempo de solución se reduce un 15 % cuando el parámetro de tolerancia se incrementa 100 veces, lo cual puede ser especialmente útil en las instancias más grandes.

4.3. Estudio económico

En esta sección, se compara la efectividad económica de la formulación propuesta respecto a la formulación compacta mostrada en [Hua y Baldick \(2016\)](#), la cual ha sido modificada para incluir costos de partida dependientes del tiempo en base a [Morales-España et al. \(2013\)](#) (COM-LP). El estudio evalúa 50 instancias generadas de manera aleatoria que van desde 8 a 30 unidades de generación en un horizonte de tiempo de 24 horas. El motivo de estudiar instancias más pequeñas en comparación al análisis computacional anterior, radica en poder resolver a optimalidad la formulación de UCP, para estudiar resultados óptimos de programación y despacho, en base a los cuales se realizarán los pagos a las unidades. Las características modeladas son equivalentes a las de la sección anterior pero es importante aclarar los problemas que son resueltos en esta sección: (i) NFU-LP y COM-LP para la obtención de precios a la energía, (ii) la formulación del UCP para obtener el despacho óptimo del ISO y (iii) los problemas de maximización de beneficios que enfrenta cada unidad de manera descentralizada dado el precio calculado, donde el beneficio es la diferencia entre los ingresos por energía y los costos de operación. A excepción del modelo NFU-LP, todos los demás modelos son programados en AMPL y resueltos con CPLEX 12.6.1.

Cuando los costos de partida no dependen del tiempo, entonces la formulación NFU-LP obtiene

exactamente los mismos precios que la formulación de [Hua y Baldick \(2016\)](#). Sin embargo, cuando la formulación se perturba para incluir costos de partida que varían con el tiempo de apagado, entonces nuestra formulación obtiene precios distintos que conducen a un pago adicional menor o igual. En la figura 4.5, los resultados de pago adicional obtenidos con la formulación NFU-LP se muestran ordenados y como un porcentaje de los resultados obtenidos con la formulación COM-LP. Se puede apreciar que en el costo de oportunidad perdido (LOC), el pago adicional es siempre menor o igual con nuestra formulación. Por ejemplo, en el 60 % de los casos el pago adicional obtenido con la formulación NFU-LP es estrictamente inferior, y en promedio corresponde a aproximadamente un 20 % menos de pago adicional.

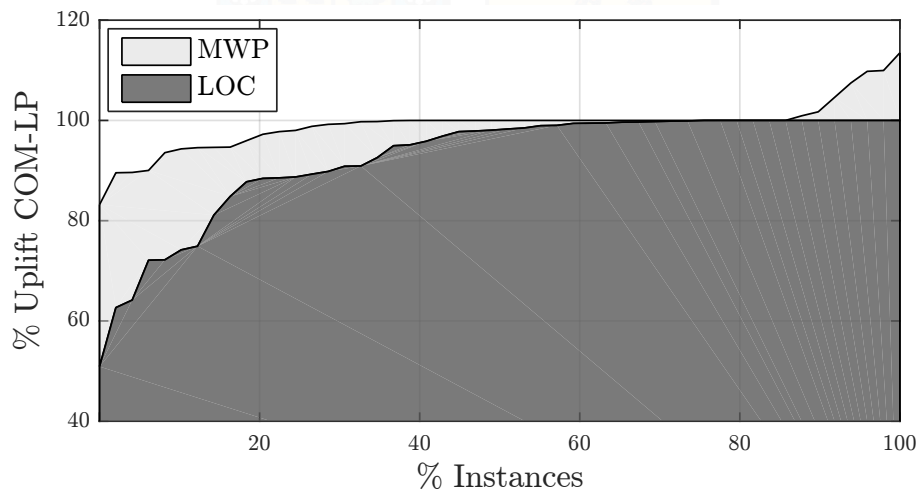


Figura 4.5: Pago adicional total NFU-LP como porcentaje del pago adicional total COM-LP.
(Fuente: elaboración propia.)

Esto se explica debido a que la inclusión de costos de partida dependientes del tiempo no afecta a la formulación propuesta, simplemente equivale a un cambio en la estructura de las redes. De hecho, se agregan más estados y arcos a las redes, pero la formulación permanece intacta y por lo tanto no se perturba la envoltura convexa del poliedro que define las operaciones factibles de la unidad. Para que esta condición se cumpla con la formulación compacta, sería necesario desarrollar desigualdades válidas especiales que modelen distintos tipos de partidas, complejizando la formulación. Claramente, la modificación realizada a partir de la mejor formulación compacta de la literatura [Morales-España et al. \(2013\)](#), perturba la envoltura convexa existente cuando los costos de partida son invariantes.

Respecto al pago *make-whole* (MWP), es sabido que definir la envoltura convexa no implica un pago adicional mínimo. Por lo tanto, no hay una relación de orden entre los resultados de MWP obtenidos con nuestra formulación y los obtenidos con la formulación compacta, pero se observa que el comportamiento promedio indica que lo obtenido con NFU-LP es menor. Por otro lado, estas diferencias están en promedio en torno al $\pm 10\%$. De lo anterior, se puede concluir que el pago adicional MWP es más robusto que el pago LOC, pues las variaciones son menores respecto a la formulación utilizada.

4.3.1. Precios subóptimos

Este apartado analiza el efecto de relajar el criterio de optimalidad de los precios sobre los pagos adicionales. Se obtienen precios subóptimos del método CHP con un parametro de optimalidad preestablecido. Primero se comparan los precios óptimos obtenidos con una tolerancia $\epsilon = 10^{-6}$ con los precios subóptimos obtenidos con una tolerancia $\epsilon = 10^{-2}$, y posteriormente se analizan los pagos adicionales derivados de ambos resultados. Los gráficos de la figura 4.6 son una muestra de los casos probados, donde se aprecia que las diferencias en precios son leves e incluso inexistentes en algunas horas. En base al casi nulo efecto sobre los precios, se analiza el efecto en los pagos adicionales.

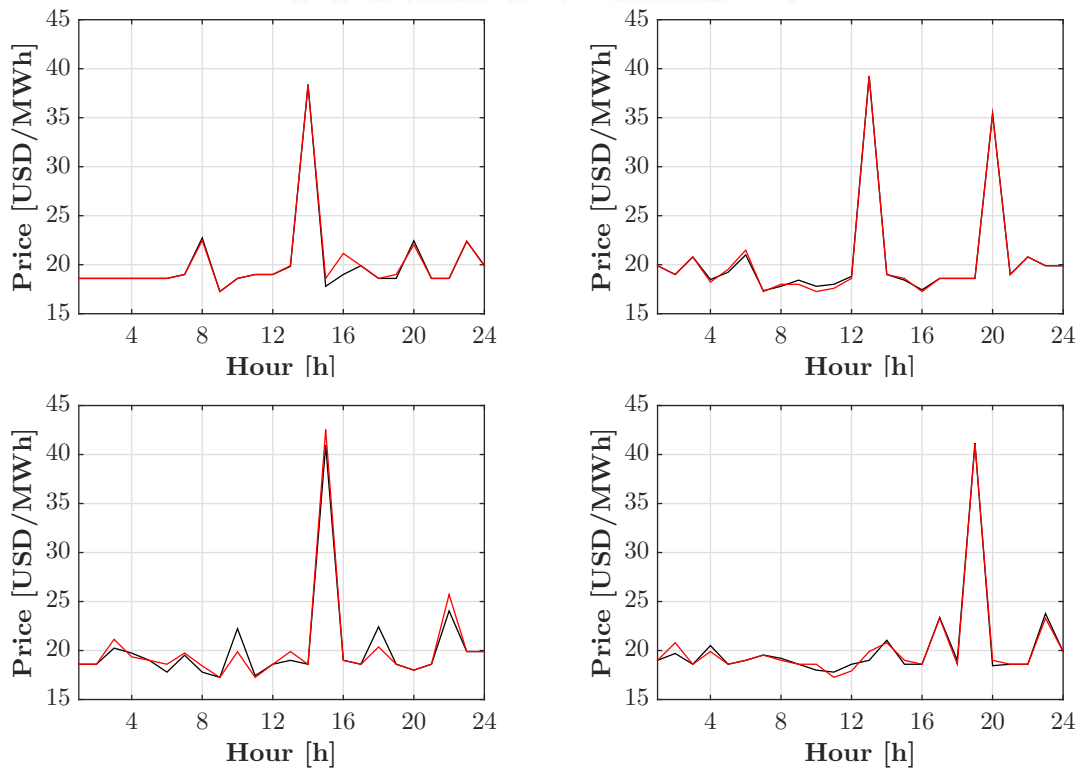


Figura 4.6: Diferencias entre precios óptimos (rojo) y subóptimos (negro) en 24 horas.

(Fuente: elaboración propia.)

En la figura 4.7 se grafica el ratio entre los pagos adicionales subóptimos y los óptimos, calculado tanto para MWP como para LOC, ambos ordenados en orden creciente. Se observa cierta estabilidad en el comportamiento de MWP, y el efecto en promedio tiende a ser un pago adicional menor con precios subóptimos, situación que ocurre en aproximadamente el 45 % de las instancias. En promedio, las diferencias están en un 10 % y en muy pocos casos MWP subóptimo fue superior al óptimo, lo cual puede sugerir que en este tipo de pago adicional la optimalidad en los precios del método CHP no es crítica. Respecto al LOC, si bien es sabido que precios subóptimos llevan a pagos adicionales mayores, los resultados muestran que las variaciones son más significativas. En nuestros experimentos, el pago subóptimo puede ser hasta más de 3

veces superior al óptimo y en cerca del 40 % de los casos el incremento es mayor al 50 %. Especialmente este efecto es el que enfatiza la importancia de resultados óptimos cuando se utiliza el método CHP.

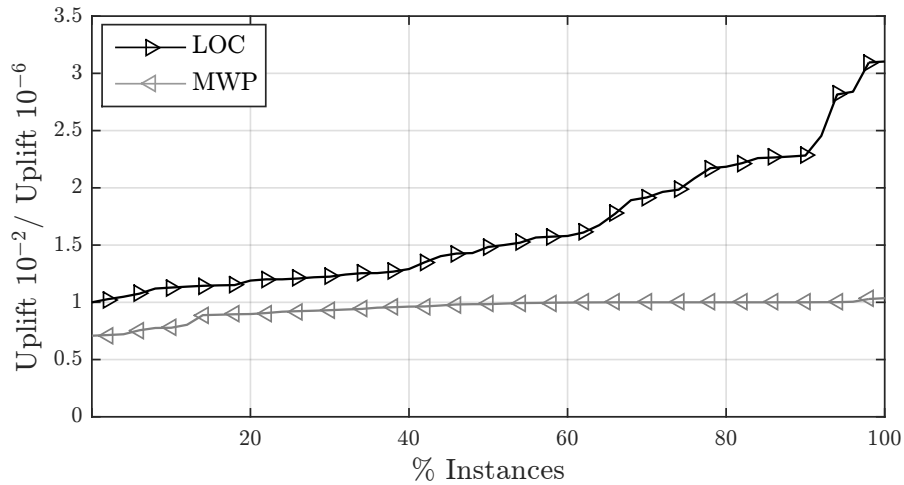


Figura 4.7: Ratio entre pago adicional subóptimo y pago adicional óptimo.
(Fuente: elaboración propia.)

A pesar de que el uso del método CHP con una política de pagos adicionales MWP podría permitir relajar la condición de optimalidad en la obtención de precios, i.e. utilizar un ϵ mayor, es recomendable considerar conjuntamente el efecto sobre LOC. Este último, es una medida más representativa del equilibrio pues cuantifica la desviación entre la decisión centralizada del ISO y la decisión individual de la unidad. Si se observa sólo el MWP, un precio alto puede llevar a que ninguna unidad despachada obtenga pérdidas anulando completamente el pago adicional. Sin embargo, es posible que existan unidades no despachadas capaces de obtener beneficios a dicho precio, lo cual se traduce en un incentivo a generar que no es cuantificado por el MWP. En resumen, incluso si el LOC no se implementa como un pago adicional, ya que la política generalizada es el MWP, es relevante en un esquema de precios al menos cuantificarlo. Para una unidad, un aumento sostenido del LOC en el largo plazo puede definir su salida del mercado y esto a su vez desincentivar la entrada de nuevos participantes.

5 | Conclusiones

Este trabajo presenta un nuevo enfoque de solución para el problema de precios de envoltura convexa conocido como *Convex Hull Pricing*. Nuestro enfoque propone una nueva formulación del problema basada en la representación en red de las operaciones de las unidades. Dado que las restricciones de red definen exactamente la envoltura convexa, los precios del método CHP pueden ser obtenidos resolviendo un problema de programación lineal. Sin embargo, considerando el tamaño de nuestra formulación, se implementa un algoritmo tipo BZ para resolver el problema de manera más eficiente que los solucionadores comerciales, logrando resultados comparables a lo obtenido con las formulaciones compactas del problema.

La formulación obtenida resulta ser robusta, debido a que la red puede representar la mayoría de las características de las unidades documentadas en la literatura, con excepción de las rampas de potencia entre períodos consecutivos arbitrarios. Como las características son abordadas con la estructura de nodos y arcos la red, la envoltura convexa que define nuestra formulación se ve inalterada respecto a cambios de modelamiento, a diferencia de lo que ocurre con formulaciones más compactas donde un cambio en el modelamiento obliga al desarrollo de desigualdades válidas especiales para redefinir la envoltura convexa. Por otro lado, la formulación basada en redes puede tener ventajas para representar los modos de operación de unidades de ciclo combinado, como se muestra en [Sepúlveda y Palma \(2011\)](#).

La implementación del algoritmo de Bienstock y Zuckerberg realizada en este trabajo permite tener un desempeño computacional superior a lo obtenido con solucionadores estándar. En promedio, las instancias se resuelven 2 veces más rápido que un algoritmo de punto interior disponible en un optimizador comercial y es posible obtener precios diarios para casos de 1000 unidades en menos de un minuto. El estudio computacional no sólo muestra que es posible obtener precios en tiempos acordes a lo que demanda el contexto; también contribuye de manera más general como referencia del desempeño del algoritmo BZ en problemas de redes con restricciones adicionales. Como se mencionaba, nuestra formulación representa problemas de ruta mínima acoplados por restricciones de sistema y posiblemente puede ser extrapolada a otros modelos.

El análisis de optimalidad realizado muestra claramente la importancia de precios óptimos con el método CHP, a diferencia de otros contextos donde es usual preferir resultados subóptimos en beneficio del tiempo computacional. En el problema estudiado, las variaciones en precios respecto al nivel de optimali-

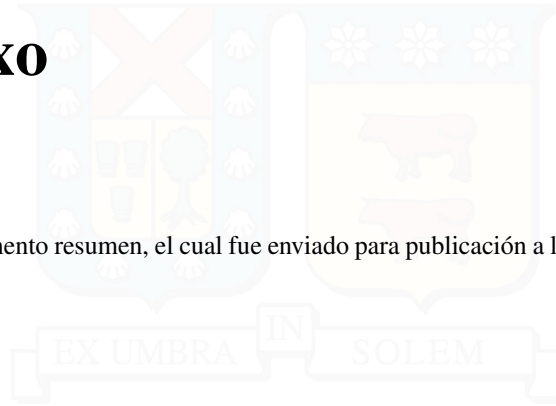
dad pueden ser leves pero con un significativo impacto sobre los pagos adicionales calculados como *lost opportunity cost*. Este resultado implica que el método CHP no es compatible con un enfoque de solución aproximado o heurístico. Lo anterior también implica exactitud en el modelamiento, mientras la formulación de la envoltura convexa pueda capturar un modelamiento más ajustado al funcionamiento real del sistema, es más seguro que los precios obtenidos repercuten en un pago adicional mínimo. Por otro lado, a pesar de que el esquema de pagos más implementado en los mercados se asimila al pago *make-whole*, sobre el cual el efecto de la suboptimalidad es poco influyente, es necesario considerar conjuntamente el pago LOC. Este último, representa una mejor medida de la desviación a la condición de equilibrio en el mercado.

5.1. Trabajos futuros

Una posible dirección de trabajo es la inclusión de rampas de toma y desprendimiento de carga en el modelo. Si bien esto no puede ser realizado en la estructura de la red, a menos que se asuma como variable discreta la potencia generada, se puede incluir en la formulación como restricciones adicionales. Esto perturba la envoltura convexa definida pero es compatible con la aplicación del algoritmo, por lo que es posible estudiar el efecto de inclusión de las rampas y compararlos con los resultados de la formulación compacta. Por otro lado, dado que la formulación es un problema de flujo en redes con restricciones adicionales, tiene aplicaciones en otros contextos donde existen otros métodos de solución. El contraste del desempeño de estos métodos respecto lo obtenido con la implementación del algoritmo BZ, puede ser la base para abrir otras líneas de investigación.

6 | Anexo

Se adjunta el documento resumen, el cual fue enviado para publicación a la revista “IEEE Transactions on Power Systems”.



A Bienstock-Zuckerberg-based Algorithm for Solving a Network-Flow Formulation of the Convex Hull Pricing Problem

C. Álvarez, F. Mancilla–David, *Member, IEEE*, P. Escalona, and A. Angulo, *Member, IEEE*

Abstract—This paper studies the convex hull pricing problem in electricity markets through a network-flow-based formulation. The network represents the feasible operating region of a generating unit, and the associated flow constraints define a polyhedron with an integrality property. These facts provide modeling flexibility with respect to the inclusion of unit features and allow to obtain convex hull prices from a linear programming problem. The formulation is solved using a primal-dual approach based on the algorithm developed by Bienstock and Zuckerberg. The algorithm, together with the implemented pre-processing and initialization techniques, allows achieving lower solution times than those obtained by state-of-the-art algorithms available in commercial solvers, e.g., barrier and dual simplex. Furthermore, results suggest that the proposed formulation obtains the minimum uplift payments even when time-dependent start-up costs are included, making the approach more robust than the best documented compact formulation. The paper also discusses the effect of sub-optimal prices on uplift payments by relaxing the optimality criterion of the algorithm, observing a significant impact on lost opportunity costs.

Index Terms—Convex hull pricing; Lagrangian relaxation; network-flow; unit commitment; BZ algorithm.

I. INTRODUCTION

IN regulated electricity markets, the dispatch of generating units over a time horizon is performed by an independent system operator (ISO), which solves the well-known unit commitment problem (UCP) in order to meet demand at minimum cost [1]. Besides that, as the market operator, the ISO must determine the energy price in the spot market for every time period. However, in general it is not possible to set prices that support market equilibrium [2]. Since generating units seek to maximize their profit, a given price may incentivize some of them to generate power levels different from those ordered by the ISO. In general, there is no price under which the units maximum profit decision in a decentralized way can be consistent with the ISO solution [3]. The aforementioned issue stems from the fact that the operational characteristics

of an electric system, such as units startup cost and minimum power output, results in a nonconvex mathematical model that prevents reaching the classical competitive equilibrium [4].

As a way to achieve market equilibrium, uplift payments can be settled in addition to the energy price, so that each unit obtains nonnegative profit and the ISO optimal dispatch is compatible with the incentives of the units. However, reaching market equilibrium can be hard for practical size instances. One of the first attempts to study equilibrium prices in nonconvex markets proposes prices for startup and capacity limits in addition to energy price, based on the dual variables of the continuous relaxation of the UCP when the binary variables has been fixed at the optimum [2]. Besides being necessary to solve the UCP in advance, the prices obtained have undesirable properties because they are volatile and discriminatory. Moreover, the computational study of [5] shows that slight differences in the optimality of the UCP solution have significant consequences in the profits of the units. In [6], the UCP is reformulated including disincentive functions, where the uplift for each unit can be positive or negative and is the result of its disincentive function. The parameters of the disincentive functions are determined by solving a quadratic problem that cancels the sum of the disincentives. Therefore, the method obtains equilibrium with the same minimum cost solution of the UCP. However, the scope is limited because the formulation considers one time period and only generation limit constraints. A broader formulation is solved in [7] using a multiobjective and bilevel approach, but the procedure is NP-hard.

Alternatively, a near-equilibrium condition can be achieved with other methods. In [8], the prices ensure that the dispatched units do not incur in losses. Since the price tends to be higher enough to cover fixed costs, it may be not compatible with incentives of nondispatched units. Another method relies in setting nonnegative uplift payments after market-clearance, i.e., once the energy price is already defined. To this end, two uplift approaches stand out: (i) make-whole payment (MWP) and (ii) lost opportunity cost (LOC). MWP covers the losses of those units which fail to recover their operating costs when they perform according to ISO dispatch [9]. LOC payment ensures that each unit receives the maximum profit at the given price. In the latter, even units which are not operating at loss can obtain an uplift payment if they still have not reached their maximum profit. The common practice is to set uniform prices according the variable cost of the marginal unit (marginal pricing) and to make MWP to units. A review of different

This work was partially supported by Fondecyt Chile under Grant 11170229 and by the Basal Project FB0008 “Advanced Center for Electrical and Electronic Engineering”.

C. Álvarez and P. Escalona are with the Department of Industrial Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile. (e-mail: cristian.alvarez@alumnos.usm.cl; pablo.escalona@usm.cl).

F. Mancilla–David is with the Department of Electrical Engineering, University of Colorado Denver, 1200 Larimer St., Denver, CO 80217, USA. (e-mail: Fernando.Mancilla-David@ucdenver.edu)

A. Angulo is with the Department of Electrical Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile. (e-mail: alejandro.angulo@usm.cl)

pricing schemes and uplift payments can be found in [10].

Regardless of the approach, the uplift payments are undesirable in an economy due to their lack of transparency, which in turn distorts the competitiveness [11]. Even it has been pointed that uplifts disincentivizes units to bid their true cost, allowing the use of game strategies [10], [12]. Furthermore, it remains unclear whether those payments must be charged to consumers or producers [8]. Another concern about uplift payments is that they may increase due to penetration of nonconventional renewable energy-based units under the traditional marginal price scheme [13]. In hours of high renewable generation, when the generation excess can not be distributed due to transmission constraints, some thermal units that usually operate in base-load must reduce their generation to the minimum [14]. In these hours, the marginal cost is determined by the renewable energy-based units. For example, Fig. 1 shows the hourly marginal cost registered in a bus in northern Chile during a day of high renewable generation, where zero marginal costs are observed. This situation presses downward the spot price, causing the revenues of some thermal units to be based almost entirely on uplifts. And if the energy payments shrink, the price loses its role as a good indicator to guide long term investment decisions. Although the setting of the price affects the energy payment and the uplift inversely, the latter is considered a market inefficiency [15]. Therefore, an efficient pricing mechanism should reduce or eliminate the uplift payments.

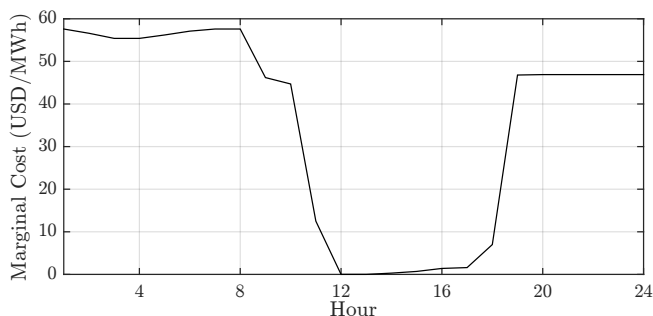


Fig. 1. Marginal cost at the “Diego de Almagro” bus on April 26th 2018, Chilean Electric National System.

A. Convex hull pricing

One way to minimize uplift payments when they are defined as LOCs is by setting prices from the solution of the convex hull pricing problem (CHPP) [16]. This procedure obtains a price vector based on the subgradient of the convex envelope of the UCP optimal cost function parametrized by demand. As shown in [16], the convex hull prices can be obtained from the Lagrangian relaxation (LR) of the UCP. The results are prices nondecreasing in demand, which implicitly include a fraction of the fixed costs. In order to explain the CHPP, consider the following simplified formulation of the UCP:

$$\text{SUC} : \quad \min_{u,p} \quad \sum_{t \in \mathcal{T}} \sum_{g \in \mathcal{G}} C_{tg}(p_{tg}, u_{tg}) \quad (1)$$

$$\text{s.t.} \quad \sum_{g \in \mathcal{G}} p_{tg} = d_t, \quad \forall t \in \mathcal{T} \quad (2)$$

$$(\mathbf{p}_g, \mathbf{u}_g) \in \Pi_g, \quad \forall g \in \mathcal{G}. \quad (3)$$

Without loss of generality, the reserve requirement and other system constraints have been omitted. The sets \mathcal{T} and \mathcal{G} represent the time periods and generating units, respectively. The nonnegative continuous variable p_{tg} is the power generated by unit g during period t . The binary variable u_{gt} is the on/off state of unit g in t . Vectors \mathbf{p}_g and \mathbf{u}_g represent power outputs and states of unit g over the entire time horizon, respectively. The costs of each unit g in period t are represented by the function $C_{tg}(\cdot)$ and the demand in each period t is d_t . The feasible operating region for unit g in the entire time horizon is represented by the set Π_g , which includes the usual constraints: minimum up/down time, minimum/maximum output power, ramping limits and startup-type costs [17], [18], [19]. By dualizing constraint (2), the LR of problem SUC is stated as follows:

LRU:

$$\max_{\lambda} \left[\begin{array}{l} \min_{u,p} \sum_{t \in \mathcal{T}} \sum_{g \in \mathcal{G}} C_{tg}(p_{tg}, u_{tg}) + \sum_{t \in \mathcal{T}} \lambda_t (d_t - \sum_{g \in \mathcal{G}} p_{tg}) \\ \text{s.t.} \quad (\mathbf{p}_g, \mathbf{u}_g) \in \Pi_g, \quad \forall g \in \mathcal{G} \end{array} \right] \quad (4)$$

For a given vector of demand, the optimal vector of multipliers in the problem LRU belongs to the subgradient of the optimal objective function convex hull (1). Thus, the convex hull prices are equivalent to the optimal multipliers of the dualized demand constraint of SUC problem [16], [20].

While it is possible to exploit the decomposition by unit in LRU with an LR-based algorithm, every iteration must solve as many mixed-integer programming (MIP) subproblems as units in the problem. Even if the convex hull of Π_g is available, achieving optimality in multipliers is computationally expensive due to the concave piecewise-linear structure of the dual Lagrangian function, which in turn produces oscillation in the convergence when traditional subgradient methods are employed [21]. In order to reduce computational burden, in [21] they use simplex and subgradient techniques in a cutting plane method to eliminate nonoptimal multipliers. The method is able to decrease computational time but lacks robustness [22]. In [20], instead of using an arbitrary subgradient, a convex quadratic problem is solved to find a steepest ascent direction. The procedure converges in a finite number of iterations but requires to find and storage all the extreme points in the unit maximum profit problems in each iteration. Despite the existence of tools in off-the-shelf solvers to do so, the number of extreme points grows exponentially with the number of generators.

Another approach is to define the convex hull of the set Π_g and to solve the linear programming (LP) relaxation of problem SUC, in order to obtain the optimal dual variables of the demand constraint [23], [24]. If the constraints of the set Π_g define exactly its convex hull then the optimal dual variables are equivalent to the optimal multipliers of the dual Lagrangian function [23]. Compared to LR-based algorithms, this approach is more straightforward as involves solving directly a LP problem. However, defining the convex hull of a nonconvex set may be extremely difficult, because usually a change in the modeling requires special-purpose inequalities. In [25] the defined convex hull includes generation limits,

time-invariant startup costs, minimum on and minimum off time. In [17] startup and shutdown ramp rates are added. In [19] ramp constraints are included for up to three time periods. Except for ramp constraints, most of the characteristics of units can be included in an exact description of the convex hull of Π_g for an arbitrary number of time periods.

An alternative description of the convex hull can be defined from the network representation of the set Π_g . Traditionally, the network representation has been used in dynamic programming to determine the optimal scheduling and generation level for a unit given a set of multipliers in the problem LRU [26], [27]. In this work, we use the network-flow approach to build a monolithic formulation of the CHPP, and we obtain the prices solving the primal problem.

B. Contribution

The contribution of this work is twofold. First, we build a monolithic formulation of the CHPP based on the network representation proposed in [26]. Since network-flow constraints define a polyhedron whose matrix is totally unimodular [28], the formulation describes the exact convex hull of Π_g . Therefore, it is possible to obtain the convex hull prices by solving an LP problem. With the exception of ramp constraints, the network can handle almost all the other characteristics of the units through changes in its structure, without altering the proposed formulation. Second, we propose a primal-dual solution approach for the CHPP based on the Bienstock-Zuckerberg (BZ) algorithm [29]. The approach allows to exploit the structure of the formulation, reaching improvements in solution times when compared against commercial available solvers. On the other hand, by modifying the optimality tolerance criterion of the algorithm, we study the effect of prices suboptimality on uplifts payments.

The remainder of the paper is organized as follows. Section II states the network-flow formulation, followed by the implementation of the BZ-based algorithm in Section III. In section IV we present numerical results and in Section V we present concluding remarks and provide directions for further research.

II. NETWORK-FLOW FORMULATION

We build a network-flow formulation with semi-continuous variables using the acyclic network presented in [26]. The structure of this network can include most of the characteristics of the units addressed in the literature. In the network, nodes represent periods in on/off state of a unit in every time period, while the edges model feasible transitions between consecutive time periods. In general, the network size is determined by the time horizon of the problem, the minimum up time and the minimum down time of the coldest startup of the unit. Fig. 2 shows an example of the network for a unit with a minimum up and down time of three and two periods, respectively. The number in the node indicates the periods that the unit has been in the respective on/off state. A path in the network is a feasible scheduling of the unit, which satisfy minimum up/down time constraints and logical constraints [17]. In order to define a formulation from the network, the edges are related

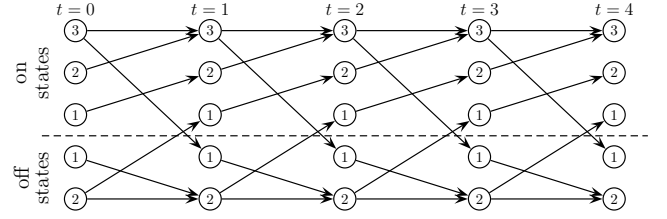


Fig. 2. Network example of a unit with minimum up time of 3 periods and minimum down time of 2 periods, for a time horizon of 4 periods.

to variables and parameters. Given a network for a unit $g \in \mathcal{G}$, let \mathcal{V}_g and \mathcal{E}_g be the set of nodes and edges, respectively. Let $\mathcal{E}_g(t) \subseteq \mathcal{E}_g$ be the subset of edges that reach a state in time period $t \in \mathcal{T}$. Let x_e be the binary variable related to the use of the edge $e \in \mathcal{E}_g$ and let y_e be a continuous variable bounded by x_e in the edge $e \in \mathcal{E}_g$ and related to power output. Let \underline{P}_e and \overline{P}_e be the minimum and maximum power output of the edge $e \in \mathcal{E}_g$, respectively. Then, the unit output power per period is expressed as follows:

$$\sum_{e \in \mathcal{E}_g(t)} [q_e x_e + s_e y_e] = p_{tg} \quad \forall t \in \mathcal{T}, \forall g \in \mathcal{G}. \quad (5)$$

In (5), $q_e = \underline{P}_e$ and $s_e = \overline{P}_e - \underline{P}_e$. It is noted that no-load costs and startup costs of the unit can be charged to fixed costs of the network, while power generation costs can be charged to the variable costs. See appendix A for details on network parameters. Let CF_e and CV_e be the fixed and variable cost of the edge $e \in \mathcal{E}_g$, respectively. Then, the cost of the network is defined by:

$$\sum_{e \in \mathcal{E}_g} [c_e x_e + f_e y_e] \quad \forall g \in \mathcal{G}. \quad (6)$$

In (6), $c_e = q_e CV_e + CF_e$ and $f_e = s_e CV_e$. Consider now the full formulation. Let $\delta^+(i)$ and $\delta^-(i)$ be the subsets of edges that come out and come in from the node $i \in \mathcal{V}_g$, respectively. For simplicity, reserve requirement and other system constraints have been omitted. Ramping constraints are not addressed with the network, and therefore, are not defined with network-flow constraints. Although they can be formulated as side constraints, this may perturb the convex hull description and is not considered in this work. The network-flow formulation may be stated as,

NFU :

$$\min_{\mathbf{x}, \mathbf{y}} \sum_{g \in \mathcal{G}} \sum_{e \in \mathcal{E}_g} (c_e x_e + f_e y_e) \quad \text{s.t.} \quad (7)$$

$$\sum_{g \in \mathcal{G}} \sum_{e \in \mathcal{E}_g(t)} (q_e x_e + s_e y_e) = d_t \quad \forall t \in \mathcal{T} \quad (8)$$

$$\sum_{e \in \delta^+(i)} x_e - \sum_{e \in \delta^-(i)} x_e = b_i \quad \forall i \in \mathcal{V}_g, \forall g \in \mathcal{G} \quad (9)$$

$$0 \leq y_e \leq x_e \quad \forall e \in \mathcal{E}_g, \forall g \in \mathcal{G} \quad (10)$$

$$x_e \in \{0, 1\} \quad \forall e \in \mathcal{E}_g, \forall g \in \mathcal{G}. \quad (11)$$

In problem NFU, constraint (8) is the demand constraint that links all networks. Constraint (9) is the flow constraint in

which the parameter \mathbf{b} only takes the value one at the initial state. Finally, constraint (10) bounds the continuous variable to their binary values. Note that **NFU** is a network-flow problem with side constraints and is equivalent to **SUC**. Since network-flow constraints define a polyhedron whose matrix is totally unimodular (TU) [28], constraints (9), (10), and (11) define the convex hull of the set Π_g for the constraints and characteristics addressed in the network.

As we are interested in the dual variables of the demand constraint, it is necessary to solve the LP relaxation of the **NFU** problem. We will refer to this relaxation as **NFU-LP** hereinafter, and considering proper size matrices and vectors the formulation can be stated as follows:

$$\text{NFU-LP} : \min_{\mathbf{x}, \mathbf{y}} \mathbf{c}^T \mathbf{x} + \mathbf{f}^T \mathbf{y} \quad (12)$$

$$\text{s.t. } \mathbf{Q}\mathbf{x} + \mathbf{S}\mathbf{y} = \mathbf{d} \quad (13)$$

$$(\mathbf{x}, \mathbf{y}) \in \mathcal{X}, \quad (14)$$

where $\mathcal{X} = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{y} \leq \mathbf{x} \leq \mathbf{1}\}$. A first computational experiment is performed in AMPL using CPLEX 12.6.1 over a small set of 48- and 120-hour cases, considering between 50 and 500 units. The proposed formulation is compared to the compact formulation of [23], which has been modified to include time-dependent startup costs according to [17], and will be referred to as **COM-LP** hereafter. The solution time is obtained from the root node relaxation in the MIP solver. Results are summarized in Table I. Although it is known that the network-flow formulation is computationally expensive, it is observed that the solution times in the larger instances are comparable because the pre-processing techniques available in the solver reduce the size of the network considerably.

TABLE I
NETWORK-FLOW AND COMPACT FORMULATIONS COMPARISON.

Periods	Solution time geometric mean (s)	
	NFU-LP	COM-LP
48 hours	30.3	7.9
120 hours	111.0	115.9

The network approach is interesting because of the flexibility in the inclusion of modeling features. With the exception of ramp constraints, others like startup/shutdown ramps, piecewise-linear generation costs and time-dependent startup costs can be handled only by modifying the nodes and edges of the network. Therefore, a network-based model provides a more robust formulation. This motivates us to look for a suitable algorithm for this kind of large-scale LP problem in order to have a more efficient solution approach. We observe that the formulation is a network-flow problem with side constraints [30]. Therefore, a LR of the side constraints leaves a problem with integrality property and separable by unit. Considering the convergence issues related to the subgradient method, we propose a solution approach based on the Bienstock-Zuckerberg (BZ) algorithm [29]. A key reason for this choice is that the decomposition scheme of the BZ algorithm allows to exploit the structure of the proposed formulation, which in turn leads to faster converge rates.

III. BZ-BASED ALGORITHM FOR CONVEX HULL PRICING

The BZ algorithm is a decomposition scheme useful for solving large-scale LP relaxations [31]. Although the BZ algorithm is generic, the proposed implementation is based on the decomposition within the LR framework discussed in [29]. There, the side constraints are relaxed and the LR solution is used to successively generate constraints in the problem to converge to the optimal solution. Using this approach in our network-flow formulation leads to solving a sequence of small problems instead of one large-scale problem. In order to explain the implementation of the BZ-based algorithm, consider a vector of multipliers λ associated to constraint (13). Then, the LR of the **NFU-LP** problem may be formulated as,

$$\text{NFU-LP}_\lambda : \min_{\mathbf{x}, \mathbf{y}} \mathbf{c}^T \mathbf{x} + \mathbf{f}^T \mathbf{y} + \lambda^T (\mathbf{d} - \mathbf{Q}\mathbf{x} - \mathbf{S}\mathbf{y}) \quad (15)$$

$$\text{s.t. } (14).$$

NFU-LP $_\lambda$ is also referred to as “relaxed” problem throughout the rest of the paper. The relaxed problem **NFU-LP** $_\lambda$ provides a lower bound for **NFU-LP**, and despite the variables bounded by \mathbf{x} in the polyhedron \mathcal{X} , the problem preserves its integrality property.

Proposition 1. *If A is TU, then the LP problem $z_{LP} = \min\{\mathbf{c}^T \mathbf{x} + \mathbf{d}^T \mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \mathcal{X}\}$ has an integral optimal solution.*

Proof. Consider the following reformulation of z_{LP} :

$$z_{LP} = \min_{\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}} \{\mathbf{c}^T \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\} + \min_{\mathbf{y} \geq \mathbf{0}} \{\mathbf{d}^T \mathbf{y} : \mathbf{y} \leq \mathbf{x}\}.$$

The second term in the sum above has a trivial solution, where $y_e = x_e$ if $d_e < 0$ and $y_e = 0$ otherwise. Therefore, it is a function of \mathbf{x} . Let $\underline{\mathbf{d}}_e = \min(0, d_e)$, then:

$$\begin{aligned} z_{LP} &= \min_{\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}} \{\mathbf{c}^T \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\} + \min_{\mathbf{x} \geq \mathbf{0}} \underline{\mathbf{d}}^T \mathbf{x} \\ &= \min_{\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}} \{(\mathbf{c} + \underline{\mathbf{d}})^T \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}. \end{aligned}$$

If matrix A is TU, it is well-known that problem z_{LP} has an integral optimal solution [28]. \square

In addition, the relaxed problem is separable by unit and can be solved straightforwardly by means of a shortest path algorithm. The solution of the relaxed problem is used to generate partitions of the whole set of variables (that constrains the problem **NFU-LP**) attempting to guess the structure of the optimal solution. As shown in [29], given that the LR features the integrality property, the number of distinct fractional values is in the order of the number of side constraints, while the remainder of variables are at the bounds 0 or 1. If the guessed structure is not optimal, then the partitions can be updated from the dual information of the constrained problem. Let H be the matrix that defines the partitions, then the constrained problem is stated as follows:

$$\begin{aligned} \text{NFU-LP}_H : \min_{\mathbf{x}, \mathbf{y}} \mathbf{c}^T \mathbf{x} + \mathbf{f}^T \mathbf{y} \\ \text{s.t. } (13), (14) \\ H\mathbf{x} = \mathbf{0}. \end{aligned} \quad (16)$$

Constraints (16) are based on partitions. Let N be the set of indexes of variables. Let R be the number of partitions, let I_r be the r th partition of the set N and let \mathcal{N} be the set of partitions. Then, by imposing constraint (16), the variables related to the same partition are forced to have the same value in the constrained problem. The constrained problem thus provides an upper bound for **NFU-LP**. Solving the constrained problem allows to update the former dual values, which lead to a new relaxed problem. This process is repeated until achieving a relative optimality gap between the upper and lower bound. The pseudocode of the BZ-based algorithm used in this work is presented in Algorithm 1. A formal description and details of the general BZ algorithm can be found in [29], [31]. Based on the speedup items discussed in [31], we implemented an initialization that avoids solving the restricted problem from scratch. The core of our implementation is explained through the next subsections.

Algorithm 1: BZ-based algorithm.

- 1 Let $k = 0$, $\lambda^{(k)} = \lambda_0$, $\mathcal{N}^{(k)} \equiv \cup_{r=1}^{R^{(k)}} I_r$.
 - 2 Solve **NFU-LP** $_{\lambda^{(k)}}$. Let $\bar{\mathbf{x}}$ be its optimal solution, then:
 $N^1 = \{i \in N : \bar{x}_i = 1\}$, $N^0 = \{i \in N : \bar{x}_i = 0\}$
 - 3 Update partitions:
 $\mathcal{N}^{(k)} \leftarrow \{I_r \cap N^1, I_r \cap N^0 : I_r \in \mathcal{N}^{(k)}\}$
 Generate $H^{(k)} \mathbf{x} = \mathbf{0}$ such that:
 $x_i = x_j$ if $(i, j) \in I_r \forall r \in \{1, \dots, R^{(k)}\}$
 - 4 Solve **NFU-LP** $_{H^{(k)}}$. Let $\bar{\lambda}$ be its optimal dual variable:
 If *optimality gap* $< \epsilon$, then STOP.
 Else, $k \leftarrow k + 1$, $\lambda^{(k)} = \bar{\lambda}$, return to step 2.
-

A. Pre-processing

Note that the size of a network can be reduced based on the initial state, which always is a parameter of the problem. Depending on the start node, there are nodes that are never reached, thus the related edges are not used. Consider the network in Fig. 3 which represents the pre-processed network from Fig. 2 given that the black node is the initial state. A

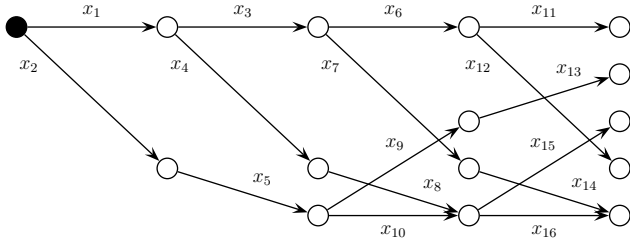


Fig. 3. Network of Fig. 2 with a pre-process based on the initial state and edges labeled by its binary variables.

second pre-processing approach is related to nodes with one input edge and two or more output edges. This kind of fork node can be eliminated and the input edge joined to the output edges. Both initial state- and fork node-based pre-processing aim at reducing the number of variables in the problem and, therefore, the solution time. In our implementation we use the former, but as an example we test both of them in a subset of the instances. Experimental results are summarized on the

first row of Table II. Although fork node-based pre-processing allows the elimination of more variables as compared to initial state pre-processing, the reduction in solution time is less on average due to the time spent in pre-processing. Since fork node pre-processing is a nonlinear transformation of variables, we believe it may be more useful in a MIP context, e.g., the **NFU** problem.

B. Initialization

We approximate initial dual values by solving another relaxation of the problem **NFU-LP**. By dropping network-flow constraints of (14), the relaxed problem can be decoupled in time. We solve a small sample of time periods to obtain the related dual values. Given the demand-dual variable pairs of the sample, we build a piecewise linear approximation of the dual values as a function of the demand. The relaxed problem **NFU-LP** $_{\lambda}$ is solved with the approximated dual values in order to obtain partitions. This procedure is repeated a given number of times for different samples, leading to different approximations of dual values without solving the constrained problem. Therefore, the partitions are quickly updated. In summary, this process has two parameters: (i) the length of a sample and (ii) the number of times the sample is taken. We test three types of initialization in the same set of instances from the previous subsection. The geometric mean of solution times is shown in the second to fourth rows of Table II. While a bad initialization can be worse than the lack of it, a suitable one can save close to 30% of time on average.

TABLE II
PRE-PROCESSING AND INITIALIZATION EFFECTS.

Initialization	Solution time geometric mean (s)		
	No pre-process	Initial state pre-process	Fork node pre-process
No initialization	76.8	74.9	75.3
1 2-period sample	–	52.9	53.3
9 5-period samples	–	76.0	76.1
10 20-period samples	–	90.2	90.6

C. Partitions

In the constrained problem the set of variables is partitioned. The procedure of generation and update of partitions is similar to that used in [32]. Given that each network solution is the minimum cost path in which only a fraction of variables take the value 1 while the others remain at 0, the set N can be split into two sets: $N^0 = \{i \in N : x_i = 0\}$ and $N^1 = \{i \in N : x_i = 1\}$. To generate partitions, at every iteration, the current sets N^0 and N^1 are intersected with the partitions of the previous iteration, i.e. $N^0 \cap I_r \forall r$ and $N^1 \cap I_r \forall r$. As the shortest path should be different due to changes in the network cost, the intersection lead to new partitions, increasing the number of partitions R . The partitions are equivalent to the following set of equality constraints (H matrix):

$$x_i - x_j = 0 \quad \forall i, j \in I_r, \forall r \in \{1, \dots, R\}.$$

In this way, the constrained problem only supports a reduced number of feasible solutions given by the combination of partitions. In every iteration, new feasible solutions are

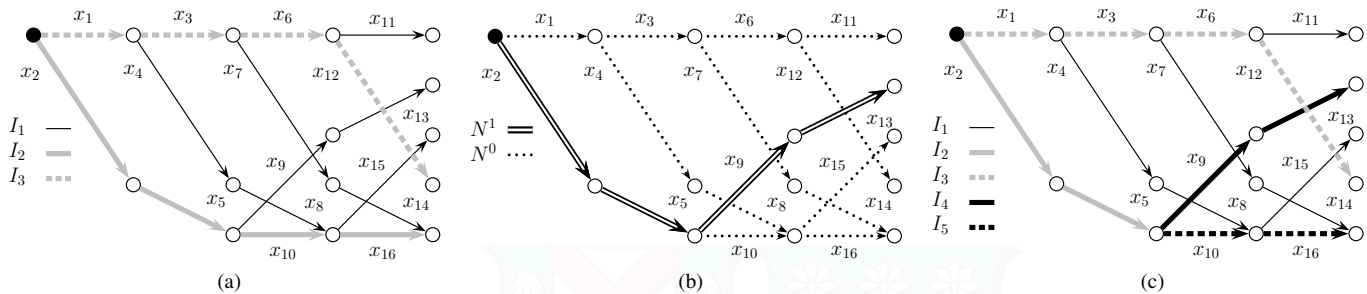


Fig. 4. Partition update procedure: (a) shows partitions at iteration k , where $I_2 = \{2, 5, 10, 16\}$, $I_3 = \{1, 3, 6, 12\}$ and $I_1 = N \setminus \cup_{r=2}^3 I_r$; (b) shows subsets generated by the relaxed problem solution at iteration k , where $N^1 = \{2, 5, 9, 13\}$ and $N^0 = N \setminus N^1$; and (c) shows updated partitions at iteration $k + 1$, where $I_2 = \{2, 5\}$, $I_3 = \{1, 3, 6, 12\}$, $I_4 = \{9, 13\}$, $I_5 = \{10, 16\}$ and $I_1 = N \setminus \cup_{r=2}^5 I_r$.

being allowed. The method to generate partitions has a key property: the update of the constraints matrix never cuts off a previously generated feasible solution. Thus, a constrained problem is always a relaxation of the previous one, which is a condition for a strictly improving direction of the problem optimization in the algorithm [29]. The procedure for one network is illustrated in the following example. Consider the edge variables in Fig 3. Fig. 4a shows three partitions at some iteration, k . Fig. 4b shows the solution to a relaxed problem at the same iteration, leading to sets N^0 and N^1 . Finally, Fig. 4c shows the intersection of the sets N^0 and N^1 with the existing partitions at iteration k . As a result, there are five partitions at iteration $k + 1$. Note that the partitions always keep its size or become smaller. Furthermore, the constrained problem at iteration $k + 1$ has only three feasible paths for this network: I_3 , $I_2 \cup I_4$ and $I_2 \cup I_5$.

IV. COMPUTATIONAL EXPERIMENTS

A. Algorithm benchmark

The scalability of our network-flow formulation and the performance of the implemented algorithm are tested in large-scale instances randomly generated. The study evaluates 315 cases which range between 50 and 2000 generating units for a time horizon of 24, 48 and 120 hours. We consider two time-dependent startup costs and the generation costs function is approximated by a piecewise-linear function with three steps. Ramping constraints are not considered and the reserve requirements have been omitted. Experiments are conducted in a PC running Windows 10, 64-bit architecture, one AMD 8-Core processor and 12 GB of RAM. The BZ-based algorithm is compared to dual simplex [33] and barrier [34] algorithms available in the commercial solver CPLEX 12.6.1. Also, we test different relative optimality tolerances for the BZ algorithm. It is noted that the relative tolerance parameter of 10^{-6} was enough to achieve optimality in all the instances. The implementation of all three algorithms is carried out using C programming language. Also, the time limit is 1 hour and initial state pre-processing is applied. In the BZ-based algorithm, the constrained problem is solved using the barrier algorithm, while the relaxed problem separated by unit is solved using a shortest path algorithm for acyclic networks. We use a initialization of 9 samples of 5 periods for instances

of 120 hours and 1 sample of 2 periods for instances of 24 and 48 hours.

A global analysis of the results is made using the performance profiles of [35]. For each solver we plot the proportion of instances solved within a factor τ of the best solution time. If t_{ps} is the solution time in instance p by solver s and t_p^* is the best solver time in such instance, then the algorithm performance is represented by the function:

$$\rho_s(\tau) = \frac{\text{Number of instances } p : t_{ps} \leq \tau t_p^*}{\text{Number of instances}}$$

Fig. 5 shows the performance profiles obtained. It can be noted that the proposed BZ-based algorithm is the fastest in all the instances. Dual simplex algorithm has the worst performance, it can solve near 5% of the instances using 5 times the solution time of the BZ algorithm. Barrier algorithm can solve all the instances, and in about the 50% of them the solution time required is at most twice the BZ time. However, in some instances barrier algorithm needs more than a factor of 3.5 to solve, which can be noted when its function reaches a value of 1.

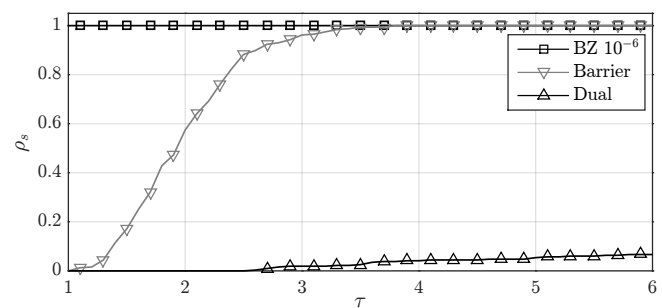


Fig. 5. Performance profiles of the algorithms for $\tau \leq 6$.

A more detailed analysis is summarized in the first three columns of Table III. In order to analyze complexity within the results, the instances have been grouped into categories named $umin_umax_nper$, where $umin$ and $umax$ stand for the minimum and maximum units number, while $nper$ is the time horizon considered. In each category average times are computed using geometric mean. Note that the dual simplex algorithm not only has the worst performance, but it is also unable to solve some of the categories within the time limit. Although BZ algorithm has always a better performance than

barrier algorithm, the differences are emphasized in the largest instances, specially in those with a high number of time periods. We can state that the increase in time periods is more critical for the barrier algorithm. Since the number of periods defines the size of the networks, the increase is easier to be handled with the decomposition scheme of the BZ-based algorithm. The bold numbers in the table show the categories in which the average speedup of the BZ algorithm is more prominent. In those categories, the implemented algorithm can be up to 2.5 times in average faster than barrier algorithm.

A feature of the BZ algorithm is the possibility to finish within a preset relative optimality tolerance. Results for different tolerances are shown in the last three columns in Table III. We note an average saving of 15% of solution time when the tolerance is one hundred times increased. This may be particularly useful for larger instances.

TABLE III
COMPARATIVE ANALYSIS OF COMPUTATIONAL PERFORMANCE.

Instances	Dual	Solution time geometric mean (s)			
		Barrier	BZ 10^{-6}	BZ 10^{-4}	BZ 10^{-2}
50_500_24	46.7	7.7	5.2	4.7	4.3
50_500_48	148.5	21.1	13.5	12.3	11.0
600_1000_24	715.3	34.6	20.9	17.8	15.0
50_500_120	687.8	77.7	31.1	26.2	21.4
1100_1500_24	2003.2	65.0	36.2	30.4	25.6
1600_2000_24	3019.8	106.2	50.3	44.1	39.5
600_1000_48	1917.7	90.5	60.6	51.7	45.9
1100_1500_48	3521.5	175.7	92.2	77.1	66.3
600_1000_120	–	318.9	134.8	115.8	95.0
1600_2000_48	–	290.9	138.1	125.6	103.2
1100_1500_120	–	600.2	264.4	215.4	167.5
1600_2000_120	–	978.5	454.9	403.9	340.3

B. Convex hull prices

We compare the economic effectiveness of our network-flow approach against the compact formulation **COM-LP** [23]. The study evaluates 50 randomly generated instances ranging from 8 to 30 units in a time horizon of 24 hours. The modeling features are the same as those from the computational study of the previous section. All the problems were programmed in AMPL using CPLEX 12.6.1. When startup costs are time-invariant we obtain exactly the same convex hull prices as those of the formulation of [23]. However, if we modify the compact formulation to include time-dependent startup costs using the constraints shown in [17], then results differ. The uplifts obtained with **NFU-LP** as a percentage of the uplifts obtained with **COM-LP** are shown in Fig. 6, sorted in increasing order. Note that LOC obtained with **NFU-LP** is always lower or equal than the results of **COM-LP**. This is explained because time-dependent startup costs do not change the convex hull description of Π_g in the network-flow formulation, unlike its compact counterpart which needs special constraints. Therefore, our formulation hold the minimum uplift. While the minimum uplift property does not hold with MWP, the uplifts of **NFU-LP** are lower on average. Also, MWP payments prove to be a more robust measure because the dispersion of the differences is around 10% on average.

Another interesting result concerns to sub-optimal convex hull prices obtained when we relax the optimality tolerance.

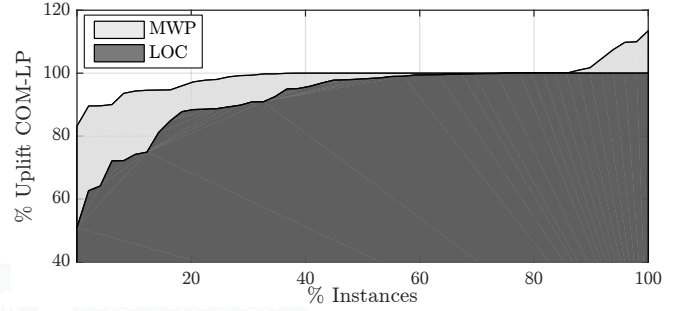


Fig. 6. Total uplift costs over a 24-hour horizon: **NFU-LP** uplift results are shown as a percentage of **COM-LP** uplift results.

We compare the optimal prices obtained with a tolerance of 10^{-6} against the results with a tolerance of 10^{-2} . Since the differences in prices are small, we plot the ratio between the sub-optimal and optimal uplift payments. See Fig. 7. We observe some stability of the MWP regarding the optimality of the prices. Although the use of a MWP policy with convex hull prices may allow a relaxation of the optimality criterion, the behavior of LOCs should be considered jointly. While it is known that sub-optimal prices lead to higher LOCs, we observe that the increasing is on average larger than the variation of MWP. In our tests, it is observed that LOCs increases threefold or more. Furthermore, in near 40% of the cases LOCs increases at least a 50%. This effect emphasizes the importance of price optimality. LOC is a more representative measure of the incentives of units, so even if not implemented, its behavior in the long run can determine whether the participants stay or leave the market.

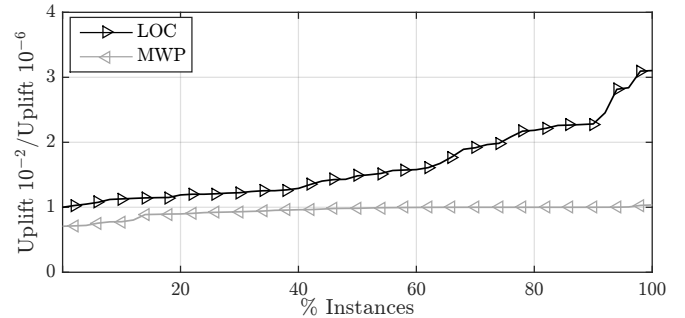


Fig. 7. Ratios between sub-optimal and optimal total uplift over a 24-hour horizon, shown in increasing order.

V. CONCLUSIONS AND FUTURE WORK

This work presented a new solution approach for the convex hull pricing problem. The approach uses a network-flow formulation of the units feasible operation, and a BZ-based algorithm to efficiently solve the mathematical programming problem. The proposed formulation turns out to be flexible as it allows to include different units characteristics and leads to the exact convex hull prices. The implementation of the BZ algorithm allows improvements in solution times when compared against other solvers, proving to solve large instances twice as fast on average. The optimality analysis performed

shows the importance of exactness in convex hull prices. Since lost opportunity costs are a more representative measure of market equilibrium, slight differences in prices can be considerable over long-term incentives of the units. A direction of further research is the inclusion of ramping constraints as side constraints in the unit network-flow formulation, based on the last works on the single unit commitment problem [19].

REFERENCES

[1] D. S. Kirschen and G. Strbac, *Fundamentals of power system economics*. John Wiley & Sons, 2018.

[2] R. P. O'Neill, P. M. Sotkiewicz, B. F. Hobbs, M. H. Rothkopf, and W. R. Stewart, "Efficient market-clearing prices in markets with non-convexities," *European journal of operational research*, vol. 164, no. 1, pp. 269–285, 2005.

[3] W. W. Hogan and B. J. Ring, "On minimum-uplift pricing for electricity markets," *Electricity Policy Group*, 2003.

[4] H. E. Scarf, "Mathematical programming and economic theory," *Operations Research*, vol. 38, no. 3, pp. 377–385, 1990.

[5] R. B. Johnson, S. S. Oren, A. J. Svoboda *et al.*, "Equity and efficiency of unit commitment in competitive electricity markets," *Utilities Policy*, vol. 6, pp. 9–20, 1997.

[6] A. L. Motto and F. D. Galiana, "Equilibrium of auction markets with unit commitment: The need for augmented pricing," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 798–805, 2002.

[7] D. Huppmann and S. Siddiqui, "An exact solution method for binary equilibrium problems with compensation and the power market uplift problem," *European Journal of Operational Research*, vol. 266, no. 2, pp. 622–638, 2018.

[8] C. Ruiz, A. J. Conejo, and S. A. Gabriel, "Pricing non-convexities in an electricity pool," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1334–1342, 2012.

[9] R. Sioshansi, R. O'Neill, and S. S. Oren, "Economic consequences of alternative solution methods for centralized unit commitment in day-ahead electricity markets," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 344–352, 2008.

[10] G. Liberopoulos and P. Andrianesis, "Critical review of pricing schemes in markets with non-convex costs," *Operations Research*, vol. 64, no. 1, pp. 17–31, 2016.

[11] B. Zhang, P. B. Luh, E. Litvinov, T. Zheng, and F. Zhao, "On reducing uplift payment in electricity markets," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*. IEEE, 2009, pp. 1–7.

[12] M. Van Vyve *et al.*, "Linear prices for non-convex electricity markets: models and algorithms," *CORE Discussion Paper 2011/50*, 2011.

[13] N. Li and K. W. Hedman, "Economic assessment of energy storage in systems with high levels of renewable resources," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 1103–1111, 2015.

[14] C. Vergara, "Efecto del mínimo técnico de unidades generadoras en el cálculo del costo marginal de corto plazo sobre la señal de inversión de largo plazo," 2016, undergraduate thesis.

[15] A. J. Cavicchi, "The polar vortex: Implications for improving the efficiency of wholesale electricity spot market pricing," 2014.

[16] P. R. Gribik, W. W. Hogan, and S. L. Pope, "Market-clearing electricity prices and energy uplift," *Cambridge, MA*, 2007.

[17] G. Morales-España, J. M. Latorre, and A. Ramos, "Tight and compact milp formulation for the thermal unit commitment problem," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4897–4908, 2013.

[18] G. Morales-España, C. Gentile, and A. Ramos, "Tight mip formulations of the power-based unit commitment problem," *OR spectrum*, vol. 37, no. 4, pp. 929–950, 2015.

[19] K. Pan and Y. Guan, "A polyhedral study of the integrated minimum-up/down time and ramping polytope," *arXiv preprint arXiv:1604.02184*, 2016.

[20] G. Wang, U. V. Shanbhag, T. Zheng, E. Litvinov, and S. Meyn, "An extreme-point subdifferential method for convex hull pricing in energy and reserve markets—part i: Algorithm structure," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2111–2120, 2013.

[21] C. Wang, T. Peng, P. B. Luh, P. Gribik, and L. Zhang, "The subgradient simplex cutting plane method for extended locational marginal prices," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2758–2767, 2013.

[22] N. Stevens, Y. Nesterov, and Y. Smeers, "Models and Algorithms for Pricing Electricity in Unit Commitment," Master's thesis, Université Catholique de Louvain, 2016.

[23] B. Hua and R. Baldick, "A convex primal formulation for convex hull pricing," *IEEE Transactions on Power Systems*, 2016.

[24] D. A. Schiro, T. Zheng, F. Zhao, and E. Litvinov, "Convex hull pricing in electricity markets: Formulation, analysis, and implementation challenges," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 4068–4075, 2016.

[25] D. Rajan, S. Takriti *et al.*, "Minimum up/down polytopes of the unit commitment problem with start-up costs," *IBM Res. Rep.*, no. RC23628, pp. W0506–050, 2005.

[26] X. Guan, P. B. Luh, H. Yan, and J. Amalfi, "An optimization-based method for unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 14, no. 1, pp. 9–17, 1992.

[27] W. Fan, X. Guan, and Q. Zhai, "A new method for unit commitment with ramping constraints," *Electric Power Systems Research*, vol. 62, no. 3, pp. 215–224, 2002.

[28] G. Nemhauser and L. Wolsey, *Integer and Combinatorial Optimization*. Wiley, 1988.

[29] D. Bienstock and M. Zuckerberg, "A new lp algorithm for precedence constrained production scheduling," *Optimization Online*, pp. 1–33, 2015.

[30] D. P. Bertsekas, *Network optimization: continuous and discrete models*. Citeseer, 1998.

[31] G. Muñoz, D. Espinoza, M. Goycoolea, E. Moreno, M. Queyranne, and O. R. Letelier, "A study of the bienstock–zuckerberg algorithm: applications in mining and resource constrained project scheduling," *Computational Optimization and Applications*, vol. 69, no. 2, pp. 501–534, 2018.

[32] D. Espinoza and E. Moreno, "A primal-dual aggregation algorithm for minimizing conditional value-at-risk in linear programs," *Computational Optimization and Applications*, vol. 59, no. 3, pp. 617–638, 2014.

[33] A. Koberstein, "Progress in the dual simplex algorithm for solving large scale lp problems: techniques for a fast and stable implementation," *Computational Optimization and Applications*, vol. 41, no. 2, pp. 185–204, 2008.

[34] E. D. Andersen, J. Gondzio, C. Mészáros, X. Xu *et al.*, *Implementation of interior point methods for large scale linear programming*. HEC/Université de Geneve, 1996.

[35] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical programming*, vol. 91, no. 2, pp. 201–213, 2002.

APPENDIX

This section aims to show the equivalence of the network parameters and the parameters of a compact formulation like [17]. In addition to the notation given in the paper, let P_{\min} , P_{\max} be the minimum and maximum power output of unit $g \in \mathcal{G}$, respectively. Let CS_g , CN_g and CG_g be the startup, no-load and generation cost of unit $g \in \mathcal{G}$, respectively. Note that the edges shown in Fig. 2 can be classified by the states of origin and destination nodes in four types: *on-on*, *on-off*, *off-on* and *off-off*. Therefore, the network parameters can be evaluated from the mentioned parameters according to Table IV:

TABLE IV
NETWORK PARAMETERS

$e \in \mathcal{E}_g$	CF_e	CV_e	\underline{P}_e	\overline{P}_e
<i>on-on</i>	CN_g	CG_g	P_{\min}	P_{\max}
<i>on-off</i>	CD_g	0	0	0
<i>off-on</i>	$CS_g + CN_g$	CG_g	P_{\min}	P_{\max}
<i>off-off</i>	0	0	0	0

One advantage of the network representation lies in the fact that the inclusion of additional features for a unit does not change the formulation structure, only implies manageable modification on nodes and edges. For instance, for the inclusion of piecewise-linear generation cost, the edges can be replicated according to the number of steps of that function. For the addition of start-up ramps, edges reaching the first state in the *on* side must have the minimum power output equal to the maximum. In the case we want to incorporate shutdown ramps, a similar procedure can be applied.

Bibliografía

- Andersen, Erling D; Gondzio, Jacek; Mészáros, Csaba; Xu, Xiaojie; et al. (1996). *Implementation of interior point methods for large scale linear programming*. HEC/Université de Geneve. 2.1.2.1
- Bertsekas, Dimitri P (1998). *Network optimization: continuous and discrete models*. Citeseer. 2.1.1, 2.1.1.1, 3.1.1
- Bertsimas, Dimitris y Tsitsiklis, J (1997). Introduction to linear programming. *Athena Scientific*, 1, 997. 2.1.1.1
- Bienstock, Daniel y Zuckerberg, Mark (2015). A new lp algorithm for precedence constrained production scheduling. *Optimization Online*, (pp. 1–33). 2.1.2.1, 2.4.3, 3.1.1, 3.2, 3.2, 3.2, 3.2.3
- Bixby, Robert E; Gregory, John W; Lustig, Irvin J; Marsten, Roy E; y Shanno, David F (1992). Very large-scale linear programming: A case study in combining interior point and simplex methods. *Operations Research*, 40(5), 885–897. 2.1.2.1
- Dantzig, George B y Thapa, Mukund N (2006). *Linear programming 2: theory and extensions*. Springer Science & Business Media. 2.1.2
- Dolan, Elizabeth D y Moré, Jorge J (2002). Benchmarking optimization software with performance profiles. *Mathematical programming*, 91(2), 201–213. 4.2
- Espinoza, Daniel y Moreno, Eduardo (2014). A primal-dual aggregation algorithm for minimizing conditional value-at-risk in linear programs. *Computational Optimization and Applications*, 59(3), 617–638. 3.2.3
- Fan, Wei; Guan, Xiaohong; y Zhai, Qiaozhu (2002). A new method for unit commitment with ramping constraints. *Electric Power Systems Research*, 62(3), 215–224. 2.4.3
- Garcia, Renan (2009). *Resource constrained shortest paths and extensions*. PhD thesis, Georgia Institute of Technology. 2.1.1.2
- Gribik, Paul R; Hogan, William W; y Pope, Susan L (2007). Market-clearing electricity prices and energy uplift. *Cambridge, MA*. 2.3.5.2, 2.4.2

- Guan, Xiaohong; Luh, Peter B; Yan, Houzhong; y Amalfi, JA (1992). An optimization-based method for unit commitment. *International Journal of Electrical Power & Energy Systems*, 14(1), 9–17. 2.4.3, 3.1, 3.1
- Ho, James K y Loute, Etienne (1981). An advanced implementation of the dantzig—wolfe decomposition algorithm for linear programming. *Mathematical Programming*, 20(1), 303–326. 2.1.2.1
- Hogan, William W y Ring, Brendan J (2003). On minimum-uplift pricing for electricity markets. *Electricity Policy Group*. 1.1, 2.3.5.2
- Hua, Bowen y Baldick, Ross (2016). A convex primal formulation for convex hull pricing. *IEEE Transactions on Power Systems*. 2.4.2, 3.1.1, 4.3
- Huppmann, Daniel y Siddiqui, Sauleh (2018). An exact solution method for binary equilibrium problems with compensation and the power market uplift problem. *European Journal of Operational Research*, 266(2), 622–638. 2.3.5
- Johnson, Raymond B; Oren, Shmuel S; Svoboda, Alva J; et al. (1997). Equity and efficiency of unit commitment in competitive electricity markets. *Utilities Policy*, 6, 9–20. 2.3.5
- Kirschen, Daniel S y Strbac, Goran (2018). *Fundamentals of power system economics*. John Wiley & Sons. 2.3
- Koberstein, Achim (2008). Progress in the dual simplex algorithm for solving large scale lp problems: techniques for a fast and stable implementation. *Computational Optimization and Applications*, 41(2), 185–204. 2.1.2.1
- Liberopoulos, George y Andrianesis, Panagiotis (2016). Critical review of pricing schemes in markets with non-convex costs. *Operations Research*, 64(1), 17–31. 2.2.1, 2.3.5.1
- Litvinov, Eugene (2010). Design and operation of the locational marginal prices-based electricity markets. *IET generation, transmission & distribution*, 4(2), 315–323. 2.3.5.2
- Madani, Mehdi; Ruiz, Carlos; Siddiqui, Sauleh; y Van Vyve, Mathieu (2018). Convex hull, ip and european electricity pricing in a european power exchanges setting with efficient computation of convex hull prices. *arXiv preprint arXiv:1804.00048*. 2.3.5
- Mathies, Susanne y Mevert, Peter (1998). A hybrid algorithm for solving network flow problems with side constraints. *Computers & operations research*, 25(9), 745–756. 2.1.1.2
- Morales-España, Germán; Latorre, Jesus M; y Ramos, Andres (2013). Tight and compact milp formulation for the thermal unit commitment problem. *IEEE Transactions on Power Systems*, 28(4), 4897–4908. 2.4.1, 2.4.2, 3.1.1, 4.3, 4.3

- Motto, Alexis L y Galiana, Francisco D (2002). Equilibrium of auction markets with unit commitment: The need for augmented pricing. *IEEE Transactions on Power Systems*, 17(3), 798–805. 2.3.5
- Munoz, Gonzalo; Espinoza, Daniel; Goycoolea, Marcos; Moreno, Eduardo; Queyranne, Maurice; y Rivera, Orlando (2016). A study of the bienstock-zuckerberg algorithm. *arXiv preprint arXiv:1607.01104*. 2.1.2.1, 3.2, 3.2
- Nemhauser, G. y Wolsey, L. (1988). *Integer and Combinatorial Optimization*. Wiley. 2.1.1.1, 2.4.3, 3.1.1, 1
- O’Neill, Richard P; Sotkiewicz, Paul M; Hobbs, Benjamin F; Rothkopf, Michael H; y Stewart, William R (2005). Efficient market-clearing prices in markets with nonconvexities. *European journal of operational research*, 164(1), 269–285. 2.2.1, 2.3.5
- Pan, Kai y Guan, Yongpei (2016). A polyhedral study of the integrated minimum-up/-down time and ramping polytope. *arXiv preprint arXiv:1604.02184*. 2.4.2
- Rajan, Deepak; Takriti, Samer; et al. (2005). Minimum up/down polytopes of the unit commitment problem with start-up costs. *IBM Res. Rep.*, (RC23628), W0506–050. 2.4.2, 3.1
- Ruiz, Carlos; Conejo, Antonio J; y Gabriel, Steven A (2012). Pricing non-convexities in an electricity pool. *IEEE Transactions on Power Systems*, 27(3), 1334–1342. 2.3.5
- Sauer, William (2014). Uplift in rto and iso markets. *Federal Energy Regulatory Commission, Tech. Rep.* 1.1
- Scarf, Herbert E (1990). Mathematical programming and economic theory. *Operations Research*, 38(3), 377–385. 2.2.1
- Schiro, Dane A; Zheng, Tongxin; Zhao, Feng; y Litvinov, Eugene (2016). Convex hull pricing in electricity markets: Formulation, analysis, and implementation challenges. *IEEE Transactions on Power Systems*, 31(5), 4068–4075. 2.4.2
- Sepúlveda, Rodrigo y Palma, Rodrigo (2011). Modelación de centrales térmicas de ciclo combinado y su aplicación en el problema de pre despacho de unidades. 5
- Shahidehpour, Mohammad; Yamin, Hatim; y Li, Zuyi (2003). *Market operations in electric power systems: forecasting, scheduling, and risk management*. John Wiley & Sons. 2.3
- Stevens, Nicolas; Nesterov, Yurii; y Smeers, Yves (2015). Models and algorithms for pricing electricity in unit commitment. 2.4.2
- Van Vyve, Mathieu et al. (2011). Linear prices for non-convex electricity markets: models and algorithms. *CORE Discussion Paper 2011/50*. 2.3.5.1
- Varian, Hal R (1992). *Microeconomic analysis*. 2.2

- Vergara, Carlos (2016). Efecto del mínimo técnico de unidades generadoras en el cálculo del costo marginal de corto plazo sobre la señal de inversión de largo plazo. Undergraduate thesis. 1.1
- Wang, Congcong; Peng, Tengshun; Luh, Peter B; Gribik, Paul; y Zhang, Li (2013a). The subgradient simplex cutting plane method for extended locational marginal prices. *IEEE Transactions on Power Systems*, 28(3), 2758–2767. 2.4.2
- Wang, Gui; Shanbhag, Uday V; Zheng, Tongxin; Litvinov, Eugene; y Meyn, Sean (2013b). An extreme-point subdifferential method for convex hull pricing in energy and reserve markets—part i: Algorithm structure. *IEEE Transactions on Power Systems*, 28(3), 2111–2120. 2.4.2

