

Universidad Nacional del Litoral
Facultad de Ingeniería Química

Tesis presentada como parte de los requisitos para la obtención del
grado académico de:

Magister en Matemática

Título de la tesis

“Convergencia y optimización global en Programación no
lineal. Teoría y algoritmos”

Departamento de Matemática de la Facultad de Ingeniería Química
INGAR- Instituto de Investigación y Desarrollo

María Laura Taverna

Director: Dr. Pío A. Aguirre

Miembros del jurado:

Dr. Hugo Aimar

Dr. José Espinosa

Dr. Marcelo Montagna

Dr. Pedro Morin

Dr. Aldo Vechietti

Agradecimientos

A mi director Dr. Pío Aguirre por sus ideas claras que me guiaron y enseñaron a investigar y su calidad humana que primó en todo momento.

A mi familia por su incondicional paciencia y apoyo, a los que les dedico este trabajo.

A Marian y Lorena por darme el empujón inicial.

A Omar Chiotti por hacerme conocer y gustar del área de optimización.

A todos mis amigos que siento cerca constantemente y no dejaron de alentarme.

Indice

Capítulo 1	Página
Introducción.....	1
1.1 Formulación general de problemas de programación matemática	2
1.2 Optimización convexa	5
1.2.1 Condiciones necesarias y suficientes para un problema convexo	5
1.2.2 Algoritmos para NLP convexos	7
1.3 Optimización global	8
1.3.1 Clasificación de las técnicas.....	8
1.3.2 Desarrollo en el área	9
1.4 Objetivos de la tesis	10

Capítulo 2

Técnicas de optimización global: ramificar y acotar (Branch and Bound)	
2.1 Introducción	12
2.2 Ideas generales de los algoritmos de ramificar y acotar	12
2.3 Relajación de un problema no convexo	13
2.4. Ramificar y acotar en optimización global	15
2.5 Envoltentes convexas y Subestimaciones	16
2.5.1 Envoltentes convexas.....	16
2.5.2 Envoltentes convexas de una función cóncava.....	17
2.5.3 Envoltentes convexas del producto de dos funciones.....	19
2.5.4 Envoltente convexa de una función univariada arbitraria.....	20
2.5.5 Subestimaciones de una función bilineal.....	22
2.5.6 Subestimaciones de una función fraccionaria.....	24
2.5.6.1 Subestimaciones lineales según McCormick.....	24
2.5.6.2 Subestimaciones según Quesada y Grossmann	24
2.5.6.3 Comparación entre las distintas estimaciones propuestas para términos fraccionarios.....	26
2.5.6.4 Subestimaciones según Maranas y Floudas.....	31
2.5.6.5 Subestimaciones alternativas de Zamora y Grossmann.....	32

2.5.7 Subestimaciones para potencias.....	32
2.6 Reducción de cotas y de la región factible en el proceso de ramificación.....	32
2.6.1 Reducción de cotas usando optimización.....	35
2.6.2 Test de reducción de cotas propuesta por Ryoo y Sahinidis... ..	36
2.7 Ideas centrales de algunos algoritmos para problemas no convexos	41
2.7.1 Algoritmo de Ryoo y Sahinidis.....	41
2.7.2 Algoritmo de Byrne y Bogle	43
2.7.3 Algoritmo de Smith and Pantelides.....	49
2.7.4 Algoritmo α BB.....	53
2.7.5 Algoritmo de optimización global GOP (Global OPTimization).....	55
2.8 Cuadro resumen de algunos aspectos de los algoritmos comentados.....	74
2.9 Conclusiones.....	75

Capítulo 3

Método de optimización global: “Improve and Branch”

3.1 Introducción	76
3.2 Formulación del modelo.....	76
3.3 Problema Reformulado: RP.....	77
3.4 Relajación convexa.....	79
3.5 Problema principal.....	82
3.6 Ramificación.....	86
3.7 Pasos del Algoritmo.....	88
3.8 Convergencia.....	90
3.9 Reducción de cotas.....	90
3.10 Conclusiones.....	92

Capítulo 4

Método de optimización global: una variante de “Improve and Branch” con una forma diferente de partición

4.1 Introducción.....	93
4.2 Variantes del algoritmo.....	93

4.3 Estrategias de reducción de la región factible.....	97
4.4 Formulación de nuevos algoritmos de optimización global usando una forma diferente de partición	111
4.4.1 Introducción.....	111
4.4.2 Lineamientos generales del algoritmo.....	111
4.4.3 Procedimiento para la partición de regiones.....	112
4.4.4 Consideraciones adicionales acerca de la reformulación.....	121
4.4.5 Un resultado teórico.....	122
4.4.6 Reglas de selección para la partición.....	127
4.4.7 Convergencia.....	129
4.4.8 Algoritmo.....	131
4.4.9 Ejemplos.....	134
4.4.10 Conclusiones.....	153
5. Apéndice.....	155

Capítulo 1

Introducción

Una gran variedad de problemas de toma de decisión que surgen de áreas tan diversas como las ciencias, ingeniería, economía, administración entre otras, pueden ser modelados matemáticamente como problemas de optimización restringidos. La mejor decisión en un contexto real es aquella que además de cumplir con todas las condiciones impuestas y requeridas, optimiza un objetivo perseguido, es decir maximiza o minimiza el criterio o los criterios sobre los que se elige tomar la decisión.

Cuando queremos optimizar y modelar situaciones reales, surgen modelos matemáticos formulados mediante funciones de diversos tipos, lineales y no lineales. Bajo condiciones de convexidad para las funciones existe una teoría robusta que garantiza la solución global de los problemas. En el área de ingeniería, por ejemplo, muchos diseños de procesos, planificación de producción, tareas y problemas operacionales se encuentran con frecuencia funciones no convexas, dando lugar a problemas de programación no lineal NLP (Nonlinear Programming Problems) y problemas de programación no lineal entero mixto MINLP (Mixed Integer Nonlinear Programming Problems) dependiendo de la naturaleza de las variables. Las técnicas de optimización convexa aplicadas a estos problemas no convexas no garantizan el óptimo global del problema, por lo que son motivo de estudio, conocido como optimización global. El interés reciente en esta área si bien ha producido grandes progresos, no es suficiente y se considera la existencia de un importante campo de investigación. Grossmann y Biegler^(4,8) exponen que la investigación en optimización global comienza a crecer rápidamente a partir de 1991 cuando se publica el Journal of Global Optimization.

En esta tesis se estudian las características generales de los algoritmos determinísticos de Ramificación y Acotamiento (Branch and Bound) desarrollados para modelos no convexas, estudiando su problemática en algunas de las técnicas propuestas en el área de optimización global. Finalmente se realiza una reformulación del método "Improve and Branch" desarrollado por Marcovecchio, Bergamini y Aguirre⁽¹¹⁾, algoritmo determinístico de tipo ramificar y optimizar al que se le implementan reducción de cotas y restricciones de reducción de la región factible.

En la introducción se presenta una revisión del estado del arte del área de optimización global determinística y clasificación de los algoritmos existentes. En el

capítulo 2 se exponen aspectos generales de algoritmos determinísticos de tipo ramificar y acotar, las características que desarrollan en optimización global y lineamientos generales de algunos trabajos valiosos en la investigación del área, entre otros se estudia el algoritmo GOP, **G**lobal **O**ptimization desarrollado por Floudas y otros. Además se estudian las diferentes subestimaciones de funciones no convexas usadas con mayor frecuencia en los algoritmos. En el capítulo 3 se presenta el algoritmo “Improve and Branch” y en el capítulo 4 se proponen cambios al mismo en busca de otras nuevas propuestas para resolver problemas no convexas. Finalmente se encuentra un apéndice con las definiciones más importantes necesarias para la lectura de este trabajo.

1.1 **Formulación general de problemas de programación matemática**

La forma general de un problema de optimización se puede enunciar como sigue:

Dadas las funciones $f, g_i, h_j : \mathfrak{R}^n \rightarrow \mathfrak{R}$ para $i = 1, \dots, m$ $j = 1, \dots, p$.

Encontrar los valores de las variables x_1, \dots, x_n tales que dan solución al problema:

$$\begin{aligned}
 & \text{Minimizar } f(x) \\
 & \text{Sujeto a: } g_i(x) \leq 0 \quad i = 1, \dots, m \\
 & \quad \quad \quad h_j(x) = 0 \quad j = 1, \dots, p \\
 & \quad \quad \quad x \in X \subseteq \mathfrak{R}^n
 \end{aligned} \tag{1.1}$$

donde $x = (x_1, \dots, x_n)$ es el vector de variables de decisión definidas en $X \subseteq \mathfrak{R}^n$.

El modelo está definido por f que es la función objetivo a optimizar, tiene m restricciones de desigualdad g_i , y p restricciones de igualdad h_j .

Todo problema en el cual una o más de las funciones que intervienen en el modelo f , g_i y h_j son funciones no lineales, define un problema de programación no lineal NLP (Nonlinear Programming). Si además en el modelo se combinan variables continuas con algunas variables enteras o binarias, define un problema de programación no lineal entero mixto MINLP (Mixed Integer Nonlinear Programming) y su forma general es:

$$\begin{aligned}
& \text{Minimizar } f(x, y) \\
& \text{Sujeto a: } g_i(x, y) \leq 0 \quad i = 1, \dots, m \\
& \quad \quad h_j(x, y) = 0 \quad j = 1, \dots, p \\
& \quad \quad x \in X \subseteq \mathfrak{R}^l, y \in Z^{n-l}
\end{aligned} \tag{1.2}$$

donde $f, g_i, h_j : \mathfrak{R}^n \rightarrow \mathfrak{R}$ para $i = 1, \dots, m \quad j = 1, \dots, p$

Si todas las variables son enteras y/o binarias el modelo es entero puro.

La región factible de un modelo es un subconjunto de \mathfrak{R}^n formado por todos los vectores x que satisfacen simultáneamente todas las restricciones del problema, a estos puntos se los llama puntos o soluciones factibles. Si no existen puntos que cumplan con todas las restricciones a la vez el problema resulta infactible.

Encontrar la solución al modelo significa hallar el valor de las variables de decisión $x = (x_1, \dots, x_n)$ factible para el problema y que dan el menor valor a la función objetivo f .

Dentro de los NLP existe una clasificación en problemas diferenciables y no diferenciables, según la naturaleza de las funciones que lo definen. Buscar extremos de una función diferenciable definida en un intervalo implica hallar puntos donde el gradiente de la función es cero o en los extremos del intervalo de definición. No así en funciones que no son diferenciables en todo su dominio de definición, como por ejemplo la función valor absoluto, que tiene un mínimo en $x=0$ y sin embargo la función no es diferenciable (Ver Figura 1.1).

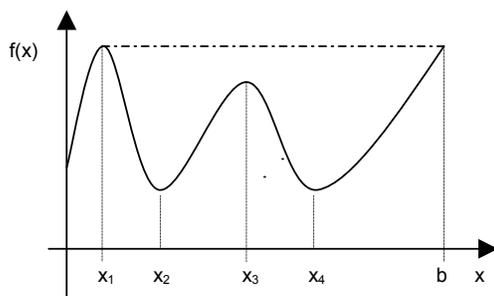


Figura 1.1 a)

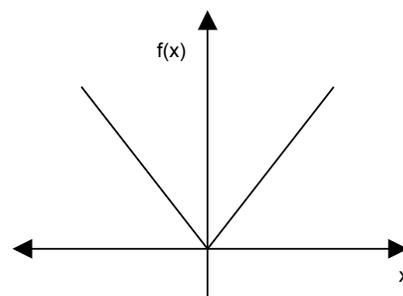


Figura 1.1 b)

Figura 1.1 a) Función diferenciable con $f'(x)=0$ en x_1, x_2, x_3 y x_4 , x_2 y x_4 mínimos globales; x_3 un máximo local y x_1 y la cota superior b máximos globales.

Figura 1.1 b) Función no diferenciable con $x=0$ mínimo global.

Otra clasificación basada en la naturaleza de las funciones surge de estudiar si son convexas o no. Los resultados teóricos de programación no lineal para NLP sólo garantizan hallar al mínimo global cuando el problema es convexo, es decir cuando todas las funciones que lo conforman son funciones convexas para restricciones de menor o igual y función objetivo, y lineales para igualdades (condiciones que luego se verán más relajadas). Como se muestra en la Figura 1.2 (b), para buscar el mínimo de una función convexa no restringida, la condición que el gradiente en un punto es cero es suficiente para asegurar que dicho punto es un óptimo global, no así en funciones que no son ni cóncavas ni convexas Figura 1.2 (a).

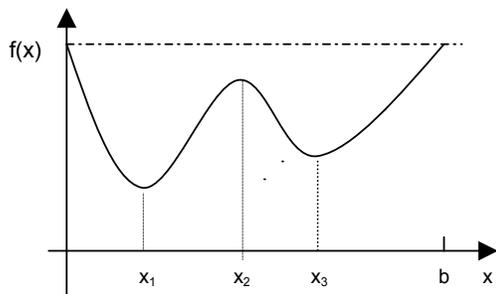


Figura 1.2 a) función no convexa

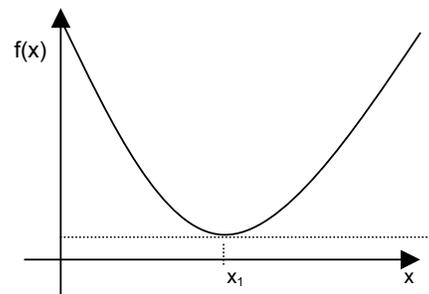


Figura 1.2 b) función convexa

Aplicar un método de optimización convexa a un problema no convexo, por sí solo no garantiza la llegada al óptimo y la solución local calculada depende del punto de comienzo requerido.

Si el modelo es no convexo, es decir al menos una función del modelo es no convexa, un algoritmo de optimización local permite hallar una solución que provee una cota superior del valor objetivo óptimo. Los métodos de optimización local se utilizan en aplicaciones donde se requiere una buena solución, no la mejor, por ejemplo en problemas de diseño en ingeniería donde se necesita mejorar una propuesta conocida o utilizada hasta el momento.

Si el problema tiene una función objetivo no convexa puede tener mínimos locales múltiples, y no necesariamente corresponden a extremos de la región factible. Si el problema tiene una región factible no convexa, estamos en la misma situación, encontrar una solución global o establecer si un punto hallado es una solución global

puede ser una tarea compleja. Es un área de investigación conocida como optimización global en busca de mejores propuestas aún.

En el campo de optimización global se tratan problemas con óptimos globales múltiples.

1.2 Optimización convexa

Cuando en un NLP de la forma (1.1) las funciones tienen una estructura particular, puede hallarse la solución del mismo usando técnicas de optimización convexa.

1.2.1 Condiciones necesarias y suficientes para un problema convexo

La teoría clásica de programación no lineal convexa establece condiciones necesarias y suficientes de óptimo global desarrolladas por Karush- Kuhn-Tucker⁽³⁾

Teorema 1.2.1.1:

Condición necesaria de óptimo local de Karush- Kuhn – Tucker (KKT)

Sean $f, g_i, h_j: \mathbb{R}^n \rightarrow \mathbb{R}$, $i=1, \dots, m$, $j=1, \dots, p$ que definen el problema (1.1). Sea x^* una solución factible del problema y a partir de la cual se puede construir un conjunto $I = \{i; g_i(x^*)=0\}$. Supongamos que $f, g_i(x^*)$ para $i \in I$ son funciones diferenciables en x^* , $g_i(x^*)$ para $i \notin I$ continuas en x^* y $h_j, j=1, \dots, p$ continuamente diferenciables en x^* . Si los vectores $\nabla g_i(x^*)$ para $i \in I$ y $\nabla h_j(x^*) \forall j$ son linealmente independientes y x^* es un mínimo local de (1.1) luego

$$\begin{aligned} & \exists \lambda_i^*, \nu_j^*, i=1, \dots, m \text{ y } j=1, \dots, p / \\ & \nabla f(x^*) + \sum_{i \in I} \lambda_i^* \cdot \nabla g_i(x^*) + \sum_{j=1}^p \nu_j^* \cdot \nabla h_j(x^*) = 0 \\ & \lambda_i^* \geq 0 \quad i \in I \end{aligned}$$

Si además para $g_i(x^*)$ para $i \notin I$ son funciones diferenciables, luego las condiciones pueden escribirse equivalentemente:

$$\begin{aligned} & \exists \lambda_i^*, v_j^*, i=1, \dots, m \text{ y } j=1, \dots, p / \\ & \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \cdot \nabla g_i(x^*) + \sum_{j=1}^p v_j^* \cdot \nabla h_j(x^*) = 0 \\ & \lambda_i^* \cdot g_i(x^*) = 0 \quad i=1, \dots, m \\ & \lambda_i^* \geq 0 \quad i=1, \dots, m \end{aligned}$$

De aquí se seleccionan todos los puntos factibles donde el gradiente de la función objetivo en el punto puede escribirse como combinación lineal de los gradientes de las restricciones que son activas en él. Esta condición necesaria provee un conjunto de puntos que contiene la solución del problema.

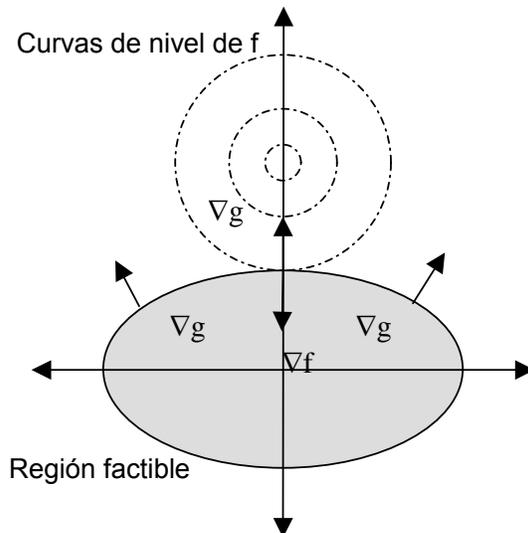


Figura 1.4: Ilustración de las condiciones de (KKT) para un NLP con una restricción de desigualdad g y dos variables.

Teorema 1.2.1.2:

Condición suficiente de óptimo global de Karush- Kuhn – Tucker para un problema (1.1).

Sea x^* una solución factible para un NLP de la forma **(1.1)** que cumple con la condición necesaria de Karush- Kuhn – Tucker, Sean los conjuntos $J = \{j : v_j^* > 0\}$ y $K = \{j : v_j^* < 0\}$. Si la función objetivo f es pseudoconvexa en x^* , las restricciones $g_i(x^*)$ para $i \in I$ cuasiconvexas en x^* , las restricciones de igualdad

h_j para $j \in J$ son cuasiconvexas en x^* y h_j para $j \in K$ son cuasicóncavas en x^* entonces x^* es una solución óptima global para (1.1).

(Ver definiciones en el apéndice)

Si bien para garantizar optimalidad global es posible adoptar condiciones sobre las funciones tales como f , g_i convexa y h_j lineales; este teorema trabaja más cuidadosamente relajando las mismas, logrando así una herramienta más poderosa que permite asegurar las condiciones suficientes para una mayor cantidad de problemas.

Las condiciones necesarias implican resolver un sistema de igualdades y desigualdades no lineales, tarea no muy sencilla y de la que se obtienen una cantidad de puntos que, si bien entre ellos está la solución del problema, los restantes son puntos calculados y posteriormente desechados. Luego las condiciones de *Karush-Kuhn – Tucker* proveen un marco teórico importante para el desarrollo de algoritmos para problemas convexos.

1.2.2 Algoritmos para NLP convexos

Los algoritmos de programación matemática para NLP convexos utilizan los resultados de la teoría clásica de la sección anterior para garantizar la llegada a la solución óptima con una precisión deseada.

Existe una gran variedad de métodos desarrollados, uno de los métodos más difundidos para resolver NLP convexos es el *Método del Gradiente Reducido Generalizado (GRG)*, que mediante expansiones de Taylor alrededor de un punto factible, realiza linealizaciones de las restricciones de igualdad no lineales. Estas linealizaciones forman un sistema de ecuaciones que, mediante la descomposición de variables en dependientes e independientes, permite trabajar el problema en un espacio de dimensión reducida. En el espacio reducido se calcula un vector llamado gradiente reducido. A partir del gradiente reducido se calcula una dirección de búsqueda que tiene la propiedad de ser una dirección de mejora pero que resulta no factible. El método en cada iteración se mueve a lo largo de la dirección construida y debe recuperar factibilidad usando el método de Newton. Existen diferentes implementaciones de GRG, entre ellas el sistema GAMS dispone de MINOS, CONOPT y CONOPT2 (Brooke, Kendrick, Meeraus, & Raman, 1998) ⁽⁵⁾

Otras técnicas de resolución se basan en Programación cuadrática sucesiva, (SQP), y su implementación en GAMS se llama SNOPT (Gill, Murray, Saunders y Drud, 2002). Las condiciones de *Karush- Kuhn – Tucker* se satisfacen sólo localmente en un problema donde la falta de convexidad no permite verificar la condición suficiente, luego es necesario buscar diferentes procedimientos para hallar la solución global de estos problemas.

1.3 Optimización global

1.3.1 Clasificación de las técnicas

Las técnicas de optimización global se pueden clasificar en estocásticas y determinísticas. Las técnicas estocásticas emplean algunos elementos de aleatoriedad en su investigación y consecuentemente se basan en argumentos estadísticos para asegurar convergencia al óptimo global, como las estrategias basadas en clustering o algoritmos de evolución genética.

Los métodos determinísticos trabajan el problema desde un punto de vista matemático, garantizando rigurosamente la llegada al óptimo global dentro de una tolerancia prefijada. Sin embargo la mayoría de estos métodos requieren de alguna suposición en la forma de la función objetivo y las restricciones del modelo.

Estos métodos incluyen:

- Métodos Lipschizianos
- Métodos de ramificar y acotar
- Métodos de plano de corte
- Métodos de diferencias de funciones DC
- Métodos de aproximación exterior
- Métodos primal dual
- Métodos de reformulación-linealización
- Métodos de intervalos

La optimización global rigurosa se puede aplicar cuando los problemas tienen una estructura especial, como por ejemplo términos continuos, con funciones bilineales, fraccionarias lineales y cóncavas separables⁽²¹⁾.

Las técnicas de optimización global, a partir de un problema no convexo usan subestimaciones o envolventes convexas para formular y resolver problemas MINLP convexos. En particular los algoritmos de ramificar y acotar espaciales dividen la región factible en subregiones que permiten la generación de cotas inferiores y superiores que se van aproximando (Floudas, 2000⁽⁷⁾; Quesada & Grossmann 1995⁽¹⁵⁾; Ryoo & Sahinidis 1995⁽¹⁷⁾)

1.3.2 Desarrollo en el área

Existen numerosos trabajos que plantean algoritmos determinísticos para resolver problemas no convexos que llegan a la solución para algunos problemas test en forma eficiente, pero en otros requieren de un importante esfuerzo computacional.

Dentro de los problemas no convexos se pueden clasificar en NLP con todas las variables continuas, y MINLP problemas con variables continuas y variables binarias.

Para problemas NLP Floudas y Visweswaran (1990)^(7,8) presentaron una descomposición primal-dual, basada en la descomposición generalizada de Benders, dando origen al método GOP, **G**lobal **O**ptimization.

Sahinidis⁽¹⁹⁾, Ryoo y Sahinidis (1995)⁽¹⁷⁾ usaron subestimaciones convexas y restricciones de reducción de intervalo en un gran número de problemas de ingeniería. El trabajo lo continúan Tawarmalani Sahinidis (2000)^(23,24) implementado en el algoritmo conocido bajo el nombre de BARON⁽²⁰⁾. Su implementación fue premiada en Agosto 2006 por Mathematical Programming Society con "The Beale-Orchard-Hays Prize" por la eficiencia y el desarrollo en nuevos métodos en programación matemática computacional

Adjiman, Androulakis, Maranas and Floudas (1996, 1998)⁽⁷⁾ presentaron un método para funciones dos veces diferenciables, el método alphaBB, basado en una estrategia de ramificar y acotar sobre variables continuas y discretas, usando subestimaciones convexas

Smith y Pantelides (1996, 1999)⁽²²⁾ presentaron un algoritmo para transformar conjuntos de restricciones no lineales en un conjunto de restricciones con términos simples no lineales manejados por un algoritmo de tipo ramificar y acotar especial para resolver problemas de programación no lineal enteros mixtos no convexos MINLP.

Quesada y Grossmann (1995)⁽⁹⁾ desarrollaron subestimaciones convexas para resolver una clase de problemas de transporte de masa bilineales y problemas de

síntesis de intercambiadores de calor y más recientemente trabajaron en problemas más generales.

Lee y Grossmann (2001)⁽⁹⁾ plantearon un método de ramificar y acotar en dos niveles para resolver problemas de programación disyuntiva no convexos.

Marcovecchio, Bergamini y Aguirre (2005)⁽¹¹⁾ han desarrollado una nueva propuesta para resolver problemas de programación no convexa NLP con una estructura de ramificación y reducción de rango.

La mayoría de los algoritmos de tipo ramificar y acotar existentes, tienen ejes de desarrollo comunes.

- Proponen técnicas de resolución de problemas de optimización con una estructura especial en las funciones que lo definen y generalmente presentan estrategias algebraicas útiles para transformar funciones generales en la forma requerida.
- Transforman el problema realizando subestimaciones y sobreestimaciones de la función objetivo y la región factible para trabajar sobre un problema de optimización convexo cuya región factible contiene a la región factible del problema inicial y a partir de su solución inferir resultados de la solución del mismo.
- Se realizan particiones de la región factible sobre las que se repite el proceso de estimación logrando que dichas estimaciones sean cada vez más ajustadas a las funciones iniciales y por lo tanto una mejor aproximación a la solución.
- Si se trabajan con particiones de la región factible, para garantizar la solución óptima global buscada es necesario una búsqueda exhaustiva en todas las subregiones obtenidas de particiones generadas por el método.
- Para acelerar el proceso del punto anterior algunos algoritmos implementan reglas de reducción de la región factible, tanto en reducción de rango de variables como ajuste en las restricciones.

1.4 Objetivos de la tesis

En esta tesis se presentan las ideas centrales de algunos algoritmos determinísticos para resolver problemas no convexos, estudiando las distintas estrategias seguidas en busca de dar solución a las dificultades propias de los algoritmos del tipo ramificar y acotar. Dentro del alcance de este trabajo se incluye un estudio de dos factores importantes y determinantes en la convergencia de estos algoritmos:

a) El nivel de falta de convexidad de un problema impacta directamente en la convergencia de los algoritmos, luego resulta importante un estudio de la calidad de las estimaciones convexas utilizadas para aproximar funciones no convexas.

b) Un inconveniente que presentan los algoritmos de tipo ramificar y acotar, es la cantidad de problemas que generan las particiones y que se deben resolver para garantizar que una solución hallada es una solución global del problema. Con frecuencia, el esfuerzo es mayor en garantizar la globalidad de la misma que en calcularla. Una forma de disminuir esta situación es reducir la región factible de cada subproblema planteado o detectar situaciones de infactibilidad.

Las sucesivas particiones de la región factible, producen subregiones donde en cada una de ellas, se particiona el rango de variación de una variable en particular. Si bien los rangos de variación iniciales de las otras variables son válidos, éstos no son los más ajustados a la nueva subregión y por lo tanto no son los más convenientes para utilizar en los cálculos siguientes. El estudio de estrategias para dar solución a estas cuestiones, es motivo de preocupación de los investigadores en el área. Un objetivo de este trabajo es analizar las propuestas existentes.

En particular se estudia un algoritmo desarrollado recientemente por Marcovecchio, Bergamini y Aguirre⁽¹¹⁾ que permite dar solución a problemas NLP no convexos con resultados satisfactorios en los problemas test presentados en el trabajo. Se prueban algunas variantes en el proceso de reducción de cotas en problemas test ya estudiados por otros autores.

Se presenta además una variante de este algoritmo donde se adoptan las reformulaciones del problema, implementando una forma diferente en el proceso de partición. Para evaluarlo se utilizan y presentan aquí algunos de los problemas test.

Capítulo 2

Técnicas de optimización global: ramificar y acotar (Branch and Bound)

2.1 Introducción

Dentro de las técnicas de optimización global determinísticas, se incluyen los algoritmos de ramificar y acotar. Estos métodos generan cotas superiores e inferiores del valor de la función objetivo, las cuales se van aproximando entre sí. Para lograr este proceso la región factible se va subdividiendo y se eliminan subregiones teniendo en cuenta la factibilidad y optimalidad del problema. En esta sección se dan ideas generales de estas técnicas y se estudian algunas implementaciones particulares.

Los métodos estudiados utilizan relajaciones de la región factible del problema que requieren del manejo de subestimaciones de las funciones intervinientes en el modelo. Se estudian diferentes propuestas presentadas en la bibliografía y se demuestran cuales subestimaciones son más ajustadas y por lo tanto más convenientes de usar en las reformulaciones. Se prueban además, condiciones que garantizan el óptimo global bajo estructuras particulares de algunos problemas de optimización no convexos.

2.2 Ideas generales de los algoritmos de ramificar y acotar

Los algoritmos del tipo ramificar y acotar se representan con una estructura de árbol. Se comienza con la resolución de un problema relajado que representa el nodo raíz. El problema relajado consiste en una sobreestimación de la región factible y una subestimación de la función objetivo en el caso de mínimo del problema original. Si la solución óptima del problema relajado es factible para el problema original, es la solución óptima buscada. De no ser así, el valor objetivo asociado es una cota inferior del valor objetivo óptimo de un problema de minimización y comienza el proceso de ramificación a partir de la definición de un criterio.

El proceso de ramificación sigue un criterio para particionar la región factible en subregiones. Para cada subregión se define un subproblema que debe resolverse y representa un nodo del árbol. La ramificación se trunca en un nodo, generalmente, por tres situaciones posibles:

- La solución del subproblema que lo representa es solución factible del problema inicial, y por lo tanto proporciona una cota superior para la solución óptima buscada.

- El subproblema es infactible.
- La solución del subproblema que lo representa tiene asociado un valor objetivo mayor (para un problema de minimización) que el de una cota superior obtenida hasta el momento.

Para cualquiera de las tres situaciones anteriores se dice que el nodo está agotado.

El primer punto concreta el paso de acotamiento del valor objetivo y el último punto implica una eliminación implícita de soluciones factibles con peores valores objetivos al logrado hasta el momento.

El desarrollo del árbol puede obtenerse a lo largo o a lo ancho según el orden en que se realizan las ramificaciones, por ejemplo, priorizando el orden de aparición de los nodos; o ramificando sobre el último nodo que se ramificó, conocido como LIFO (last in, first out). El proceso termina cuando todos los nodos están agotados y no es posible ramificar más, o cuando la diferencia entre la cota inferior y superior del valor objetivo es menor que una tolerancia prefijada.

2.3 Relajación de un problema no convexo

Dado un problema de la forma:

$$\text{Minimizan } f(x_1, \dots, x_n) \quad (2.1)$$

$$\text{Sujeto a: } g_i(x_1, \dots, x_n) \leq 0 \quad i = 1, \dots, m \quad x = (x_1, \dots, x_n) \in X \subset \mathbb{R}^n$$

Con frecuencia los algoritmos de resolución requieren al problema de optimización en diferentes reformulaciones. Las reformulaciones se generan a partir de alguna propiedad matemática estructural especial, por ejemplo la forma estándar para los problemas lineales.

Definición 2.3.1: Una reformulación es exacta si la solución global del problema original se puede inferir de su reformulación directamente o a través de un número finito de pasos. Una reformulación que no es exacta se llama relajación del problema.

Una relajación para un problema no convexo es:

$$\text{Minimizan } \bar{f}(x_1, \dots, x_n) \quad (2.2)$$

$$\text{Sujeto a: } \bar{g}_i(x_1, \dots, x_n) \leq 0 \quad i = 1, \dots, m \quad x \in X$$

Donde \bar{f}, \bar{g}_i son funciones convexas que aproximan las funciones del modelo original, tal que $RF(2.1) \subset RF(2.2)$ y $\bar{f}(x) \leq f(x) \quad \forall x \in X$.

La relajación de un problema no convexo se obtiene convexificando la región factible original, esto es, se reemplazan las funciones no convexas por subestimaciones convexas de las mismas. Este procedimiento construye una región convexa mayor que contiene a la región factible original. Además, en el caso de no ser convexa la función objetivo, ésta se subestima por una función convexa, obteniendo una función que es menor punto a punto que la función objetivo original. Por lo tanto la solución de este problema relajado subestima la solución global buscada. Más formalmente:

Afirmación: Dado un problema no convexo de la forma **(2.1)** y una relajación convexa **(2.2)** con x^* y \bar{x} las soluciones óptimas respectivamente. Luego se cumple $\bar{f}(\bar{x}) \leq f(x^*)$ proporcionando una cota inferior del valor objetivo óptimo global.

Demostración: $RF(2.1) \subset RF(2.2)$ entonces se cumple que $\text{Min}_{x \in RF(2.2)} f(x) \leq \text{Min}_{x \in RF(2.1)} f(x)$

Además como $\bar{f}(x) \leq f(x) \quad \forall x \in X$ luego $\text{Min}_{x \in RF(2.2)} \bar{f}(x) \leq \text{Min}_{x \in RF(2.2)} f(x)$

Uniendo ambas desigualdades $\bar{f}(\bar{x}) = \text{Min}_{x \in RF(2.2)} \bar{f}(x) \leq \text{Min}_{x \in RF(2.1)} f(x) = f(x^*)$

Para un NLP no convexo diferentes relajaciones convexas de las funciones que lo definen proveen diferentes problemas relajados. En la sección 2.7 se detallan algoritmos con diferentes relajaciones propuesta para un problema de programación global, usando reformulaciones convexas, lineales, funciones DC, entre otras y reformulaciones iniciales al problema que usan el concepto de separabilidad y requieren del agregado de variables y restricciones. El problema relajado puede resolverse por técnicas para problemas convexas y así su valor objetivo asociado a su solución permite generar una cota inferior del mínimo global buscado.

En un proceso de relajación se pueden agregar variables para redefinir términos que intervienen en las funciones, requiriendo además la incorporación de restricciones, luego el problema relajado puede resultar un problema de mayor dimensión que el original en cuanto a número de variables y número de restricciones. La región factible de este nuevo problema proyectada sobre las variables originales debe resultar una región que cumpla con las propiedades antes mencionadas, es decir, convexa y que contenga a la región original.

2.4. Ramificar y acotar en optimización global

Siguiendo un procedimiento particular para generar el problema relajado de la sección anterior, se obtiene un problema convexo al que, como ya se dijo, se le puede hallar su solución óptima global y cuyo valor objetivo es una cota inferior del mínimo global buscado. Si la solución óptima es una solución factible del problema original, y aún se está trabajando con la región sin particionar, entonces se tiene la solución óptima global. Soluciones del problema original obtenidas por alguna técnica local proporcionan una cota superior de la solución global del NLP no convexo (óptimo local). Cuando las cotas superior e inferior están lo suficientemente próximas, el algoritmo termina. Si no, se subdivide la región factible en dos o más partes y en cada una de ellas se plantean nuevos problemas relajados. Estos problemas permiten aproximaciones cada vez más precisas para las funciones que la definen.

La siguiente figura muestra una función objetivo no convexa y una subestimación convexa definida en la región inicial R_0 . El mínimo de la relajación, en este ejemplo, no coincide con el mínimo global buscado, pero proporciona una cota inferior, además de generar una partición de la región factible.

Una partición genera dos nuevas subregiones $R_1 \neq \emptyset$ y $R_2 \neq \emptyset$ tal que $R_0 = R_1 \cup R_2$ y cuyo único punto de intersección es el punto que la produce.

En cada subregión puede repetirse el proceso de subestimar a f , logrando subestimaciones de los nuevos problemas relajados más ajustadas a la función original (figura 2.1)

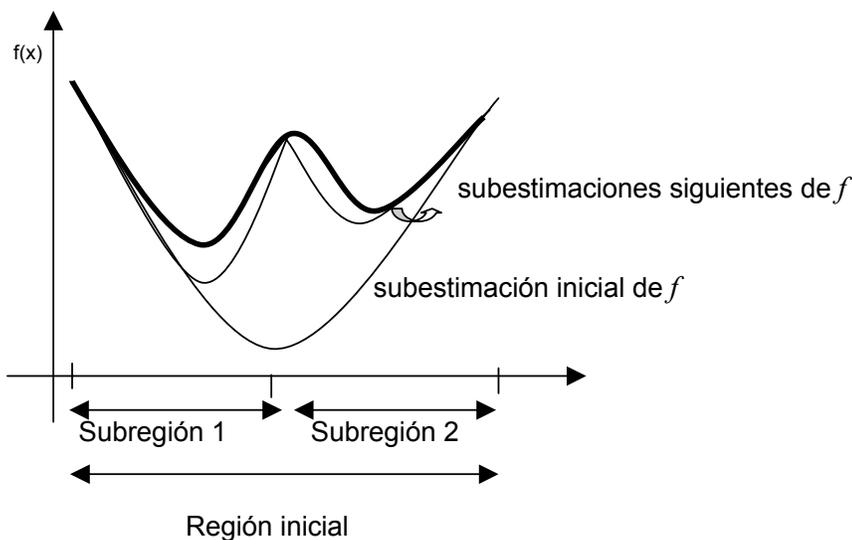


Figura 2.1: función no convexa y sus subestimaciones.

Este procedimiento genera un árbol de búsqueda cuyos nodos son los subproblemas relajados convexos. La ramificación se trunca cuando la solución de un subproblema tiene un valor objetivo mayor que la última cota superior o el subproblema resulta infactible. En todo otro caso, se debe estudiar los subproblemas generados por las particiones, resultando un proceso extenso aunque necesario para garantizar que el punto finalmente seleccionado es un óptimo global del problema. El agregado de reducción de las cotas de las variables y reducción de la región factible optimiza el procedimiento, reduciéndose la secuencia de partición y búsqueda.

2.5 Envoltentes convexas y Subestimaciones

Como ya se dijo, una estrategia utilizada para resolver problemas no convexos, es reemplazar las funciones no convexas por subestimaciones y sobreestimaciones convexas, obteniendo un problema relajado convexo asociado. Cuando las subestimaciones que se usan son más ajustadas, es decir que su diferencia punto a punto con la función original es más pequeña, el problema relajado estará más próximo al problema original y por lo tanto encontrar la solución deseada es una tarea menos costosa.

La idea de mejor subestimación o subestimación más ajustada se formaliza en la definición de envoltente convexa que se da en la siguiente sección.

2.5.1 Envoltentes convexas⁽⁷⁾

Definición 2.5.1: sea f una función semicontinua inferior definida sobre un conjunto convexo no vacío $S \subseteq \mathfrak{R}^n$. Luego la envoltente convexa de $f(x)$ sobre S es una función $\phi(x)$ que satisface:

- i) $\phi(x)$ es una función convexa sobre S .
- ii) $\phi(x) \leq f(x) \quad \forall x \in S$
- iii) Si $h(x)$ es una función convexa definida sobre S tal que $h(x) \leq f(x) \quad \forall x \in S$ luego $h(x) \leq \phi(x) \quad \forall x \in S$.

La envoltente convexa de una función representa la mejor subestimación convexa

Teorema 2.5.1:

Sea f una función semicontinua definida sobre un conjunto convexo compacto no vacío $S \subseteq \mathfrak{R}^n$ y sea $\phi(x)$ la envolvente convexa de $f(x)$ sobre S luego

$$\min_{x \in S} f(x) = \min_{x \in S} \phi(x)$$

Si bien el resultado anterior es de suma importancia, hallar la envolvente convexa de una función generalmente es una tarea compleja. Situaciones particulares sobre la región factible hacen posible el cálculo de dicha función convexa de manera más sencilla que se detallan a continuación.

2.5.2 Envolventes convexas de una función cóncava**Teorema 2.5.2:**

Sea $f(x)$ una función cóncava definida sobre un politopo P acotado, sean v_1, \dots, v_m los vértices del politopo. Luego la envolvente convexa $\phi(x)$ de la función cóncava $f(x)$ sobre el politopo P se puede obtener resolviendo el siguiente problema:

$$\begin{aligned} \phi(x) &= \min_{\lambda} \sum_{i=1}^m \lambda_i f(v_i) \\ \text{su. a: } &\sum_{i=1}^m \lambda_i v_i = x \\ &\sum_{i=1}^m \lambda_i = 1 \\ &\lambda = (\lambda_1, \dots, \lambda_m) \geq 0 \end{aligned}$$

Demostración:

Otra forma de caracterizar analíticamente la envolvente convexa es usando la noción de cápsula convexa y epígrafo de una función. Si f es una función arbitraria

$$\text{Envolvente convexa } (f(x)) = \inf \{ \mu / (x, \mu) \in \text{cápsula convexa}(\text{epi}(f)) \} \quad (2.3)$$

Sea P el politopo con vértices v_1, \dots, v_m , para cada

$$x \in P \text{ luego } \exists \lambda_i / x = \sum_{i=1}^m \lambda_i v_i \quad \sum_{i=1}^m \lambda_i = 1 \quad \lambda = (\lambda_1, \dots, \lambda_m) \geq 0$$

$(x, \mu) \in \text{cápsula convexa}(\text{epi}(f))$ sí y sólo si

$$(x, \mu) = \lambda_1(v_1, \mu_1) + \dots + \lambda_m(v_m, \mu_m) = \left(\sum_{i=1}^m \lambda_i v_i, \sum_{i=1}^m \lambda_i \mu_i \right) \text{ con } \sum_{i=1}^m \lambda_i = 1 \quad \lambda = (\lambda_1, \dots, \lambda_m) \geq 0 \text{ y}$$

$(v_i, \mu_i) \in \text{epi}(f) = \{(x, \mu) \in P \times \mathcal{R} \mid f(x) \leq \mu\}$ luego se verifica $f(v_i) \leq \mu_i$

$$\sum_{i=1}^m \lambda_i f(v_i) \leq \sum_{i=1}^m \lambda_i \mu_i$$

$\phi(x) = \inf \left\{ \sum_{i=1}^m \lambda_i f(v_i) \mid x = \sum_{i=1}^m \lambda_i v_i \right\}$ Como f es cóncava definida en un conjunto acotado

alcanza su mínimo \square

Teorema 2.5.3:

Sea $f(x)$ una función cóncava definida sobre un simplex S , y los vértices del simplex v_1, \dots, v_n . Luego la envolvente convexa $\phi(x)$ de la función cóncava $f(x)$ sobre el

simplex S es la función afín $\phi(x) = c^T x + b$ donde los vectores c, b están unívocamente determinados por el sistema de ecuaciones lineal

$$f(v_i) = c^T v_i + b \quad i = 1, \dots, n.$$

Demostración:

Sea S simplex cuyos vértices v_1, \dots, v_n son afinmente independientes. Por el Teorema 2.5.2 tomando $\lambda_i = 1$, $f(v_i) = \phi(v_i)$ $i = 1, \dots, n$

Para todo otro punto que no es vértice si $x \in S$ luego

$$\exists \lambda_i / x = \sum_{i=1}^n \lambda_i v_i \quad \sum_{i=1}^n \lambda_i = 1 \quad \lambda = (\lambda_1, \dots, \lambda_n) \geq 0$$

$$\phi(x) = \min_{\lambda} \sum_{i=1}^n \lambda_i (c^T v_i + b) = c^T \sum_{i=1}^n \lambda_i v_i + b = c^T x + b \quad \square$$

La envolvente convexa de una función cóncava univariada $f(x)$ sobre $[x^L, x^U]$ es la linealización definida a través de los puntos extremos. Aplicando el teorema anterior

los vectores c, b son solución del sistema:
$$\begin{cases} f(x^L) = c^T x^L + b \\ f(x^U) = c^T x^U + b \end{cases}$$

$$c^T = \frac{f(x^U) - f(x^L)}{x^U - x^L} \quad b = f(x^L) - \frac{f(x^U) - f(x^L)}{x^U - x^L} x^L$$

$$\text{resulta: } f(x^L) + \frac{f(x^U) - f(x^L)}{x^U - x^L} (x - x^L) \quad (2.4)$$

2.5.3 Envoltentes convexas del producto de dos funciones

Teorema 2.5.3.1:

Sean $f: [x^L, x^U] \rightarrow \mathbb{R}$ y $g: [y^L, y^U] \rightarrow \mathbb{R}$ funciones univariadas continuamente diferenciables dos veces y

$$l(x,y) = \max \left\{ \varphi(f^L g(y)) + \varphi(g^L f(x)) - f^L g^L, \varphi(f^U g(y)) + \varphi(g^U f(x)) - f^U g^U \right\}$$

Donde f^L es el ínfimo de la función f en su dominio de definición, f^U es el supremo de la función f en su dominio de definición, g^L es el ínfimo de la función g en su dominio de definición, y g^U es el supremo de la función g en su dominio de definición. $\varphi(f^L g(y))$, $\varphi(g^L f(x))$, $\varphi(f^U g(y))$, y $\varphi(g^U f(x))$ son las envoltentes convexas de las funciones univariadas $f^L g(y)$, $g^L f(x)$, $f^U g(y)$, y $g^U f(x)$ respectivamente. Luego se tiene:

- i) $l(x,y)$ es una función convexa, $\forall (x,y) \in [x^L, x^U] \times [y^L, y^U]$
- ii) $f(x)g(y) \geq l(x,y) \quad \forall (x,y) \in [x^L, x^U] \times [y^L, y^U]$
- iii) si $g^L f(x)$, $g^U f(x)$ y $f^L g(y)$, $f^U g(y)$ son funciones cóncavas en $[x^L, x^U]$ y $[y^L, y^U]$ respectivamente, f y g funciones monótonas. Luego $l(x,y) = \varphi(f(x)g(y)) \quad \forall (x,y) \in [x^L, x^U] \times [y^L, y^U]$ es la envolvente convexa de $f(x)g(y)$.

Ejemplo 2.5.3.1: sea $h(x,y) = f(x)g(y) = x \frac{1}{y}$, $(x,y) \in [x^L, x^U] \times [y^L, y^U]$ con $x^U \leq 0$, $y^U > 0$

f y g son funciones univariadas continuamente diferenciables dos veces.

f es creciente luego alcanza su mínimo en $f^L = x^L$, y su máximo en $f^U = x^U$.

g es decreciente luego alcanza su mínimo en $g^L = \frac{1}{y^U}$ y su máximo en $g^U = \frac{1}{y^L}$.

$g^L f(x) = \frac{x}{y^U}$ y $g^U f(x) = \frac{x}{y^L}$ son lineales y por lo tanto convexas.

Las envoltentes convexas de las funciones cóncavas univariadas

$f^L g(y) = \frac{x^L}{y}$ y $f^U g(y) = \frac{x^U}{y}$ por (2.4) resultan respectivamente:

$$\varphi(f^L g(y)) = -\frac{x^L}{y^L y^U} y + \frac{x^L}{y^L} + \frac{x^L}{y^U} \quad \text{y} \quad \varphi(g^U f(x)) = -\frac{x^U}{y^L y^U} y + \frac{x^U}{y^L} + \frac{x^U}{y^U}$$

$$l(x,y) = \max \left\{ -\frac{x^L}{y^L y^U} y + \frac{x^L}{y^L} + \frac{x}{y^U}, -\frac{x^U}{y^L y^U} y + \frac{x^U}{y^L} + \frac{x}{y^L} \right\}$$

este máximo de dos planos, según el *Teorema 2.5.3.1* es una subestimación convexa de h y (iii) asegura que es la subestimación convexa más ajustada de h .

La misma función en un intervalo de definición donde $x^L \geq 0$ e $y > 0$ permite construir una subestimación convexa pero no garantiza que sea la envolvente de la función ya que no cumple con la hipótesis que $f^L g(y), f^U g(y)$ sean funciones cóncavas.

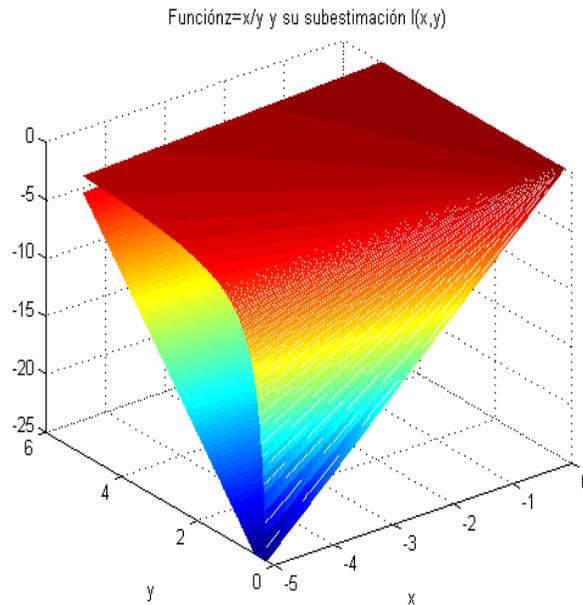


Figura 2.3: función $z=x/y$ definida en $(-5,0) \times (0,2,5)$ y su envolvente convexa.

2.5.4 Envolvente convexa de una función univariada arbitraria

Sea $f(x)$ una función dos veces diferenciable con partes cóncavas y partes convexas, su envolvente convexa se puede construir calculando los puntos críticos de la función, máximos, mínimos y puntos de inflexión.

Si la función definida en un intervalo cambia de cóncava a convexa, la envolvente convexa está formada por una recta que pasa por el extremo inferior del intervalo $x = x_0$ y un punto sobre la curva tal que la recta sea tangente a la misma:

$f'(x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$. El punto x_1 hallado es un punto donde f es convexa, la recta que

define está por debajo de la curva de f y no la corta en ningún punto. La envolvente convexa de f queda definida con la recta que conecta los puntos $(x_0, f(x_0))$ y $(x_1, f(x_1))$ y a partir de él, la misma f que es convexa. Esta aproximación se repite cada vez que se esté en la misma situación.

Si la función cambia de convexa a cóncava, análogamente la función que subestima la curva está formada por un punto x_1 donde la pendiente es igual a f' y el extremo superior $x = x_2$, quedando la envolvente convexa formada primero por la curva f donde es convexa hasta el punto x_1 hallado, y luego por la recta que una los puntos $(x_1, f(x_1))$ y $(x_2, f(x_2))$.

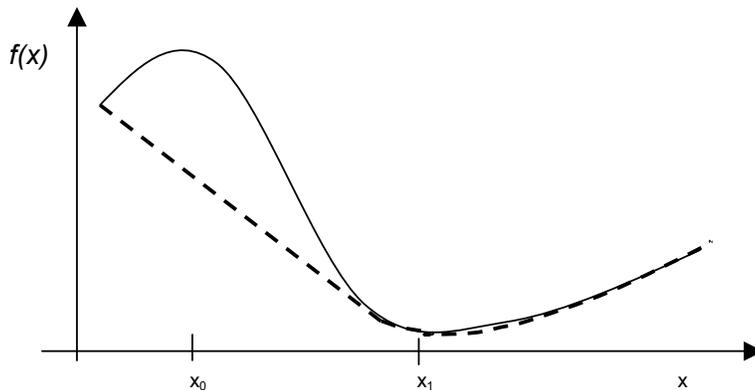


Figura 2.2: ejemplo de una función ni cóncava ni convexa con su envolvente convexa

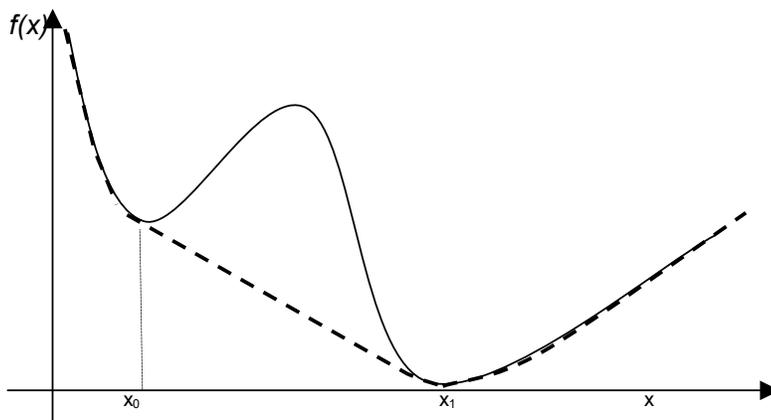


Figura 2.3: ejemplo de otra función arbitraria con su envolvente convexa.

En funciones donde la función tiene un tramo convexo, luego cóncavo y nuevamente convexo la envolvente convexa se logra con un tramo de la curva f convexo, un segmento de una recta soporte de f y nuevamente un tramo de curva de f convexo.

2.5.5 Subestimaciones de una función bilineal

Sea $f: [x^L, x^U] \times [y^L, y^U] \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ función bilineal $f(x,y) = xy$. McCormick (1976), Al Khayyal y Falk (1983), mostraron que la envolvente convexa sobre su dominio se obtiene agregando una nueva variable w que reemplaza a xy con la siguiente relación:

$$w = \max \{ x^L y + y^L x - x^L y^L, x^U y + y^U x - x^U y^U \} \quad (2.5)$$

Define el máximo de dos hiperplanos, el primero coincide con el valor de la función bilineal en los puntos (x^L, y^L) , (x^L, y^U) y (x^U, y^L) y el segundo en (x^U, y^U) , (x^U, y^L) y (x^L, y^U) . Los hiperplanos se intersecan por lo tanto en (x^U, y^L) , (x^L, y^U) y en todo un segmento de recta $y = \frac{x^L y^L - x^U y^U}{x^L - x^U} + \frac{y^U - y^L}{x^L - x^U} x$ comprendido en el hiperrectángulo $[x^L, x^U] \times [y^L, y^U]$ que corresponde a la diagonal de la misma.

La función máximo es una función lineal a trozos, no es diferenciable en todo punto, si es convexa por ser el máximo de funciones convexas.

Androulakis et al. (1995) determinaron que la máxima distancia de separación entre la función bilineal y su subestimación convexa se da en el punto medio del hiperrectángulo $[x^L, x^U] \times [y^L, y^U]$

Esta función máximo es no diferenciable en todo su dominio y con fines prácticos la condición de máximo se relaja mediante dos desigualdades:

$$\begin{aligned} w &\geq x^L y + y^L x - x^L y^L & (2.6) \\ w &\geq x^U y + y^U x - x^U y^U \end{aligned}$$

Estas desigualdades pueden derivarse teniendo en cuenta las restricciones de cotas

$$y^L \leq y \leq y^U \quad \text{y} \quad x^L \leq x \leq x^U$$

Como $x - x^L \geq 0$ e $y - y^L \geq 0$ se cumple que $(x - x^L)(y - y^L) \geq 0$, luego $xy - xy^L - x^L y + x^L y^L \geq 0$ despejando $w=xy$ se obtiene la primer desigualdad.

Como $x - x^U \leq 0$ e $y - y^U \leq 0$ se cumple que $(x - x^U)(y - y^U) \geq 0$, luego $xy - xy^U - x^U y + x^U y^U \geq 0$ despejando $w=xy$ se obtiene la segunda desigualdad.

Una cota superior se puede considerar sobre w para una mejor aproximación del problema original McCormick (1976), con el agregado de dos desigualdades más:

$$\begin{aligned} w &\leq x^U y + y^L x - x^U y^L & (2.7) \\ w &\leq x^L y + y^U x - x^L y^U \end{aligned}$$

Que análogamente a las subestimaciones, se derivan de las desigualdades $(x - x^U)(y - y^L) \leq 0$ y $(x - x^L)(y - y^U) \geq 0$ respectivamente.

Estas estimaciones si bien son más relajadas que su envolvente convexa, son muy utilizadas por los algoritmos de la bibliografía por su nobleza.

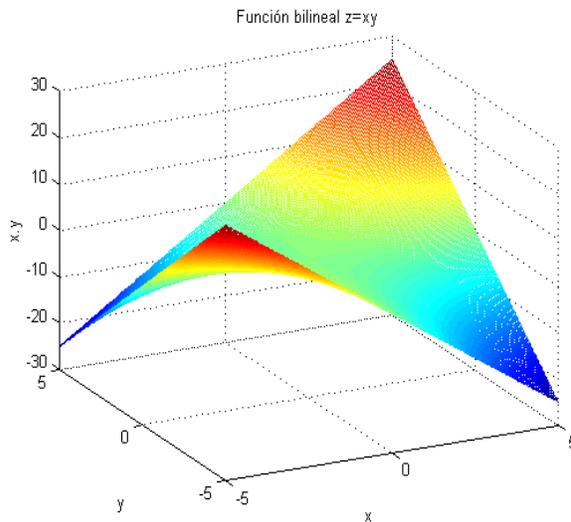


Figura 2.4 : función bilineal $f(x,y) = xy$ definida en $(x,y) \in (-5,5) \times (-5,5)$

Para términos bilineales, $w = xy$, Quesada y Grossmann (1995) propusieron estimadores que dependen de las cotas de las variables en una región o subregión.

$$\begin{aligned} w^L &\geq \min \{x^L y^L, x^L y^U, x^U y^L, x^U y^U\} \\ w^U &\leq \max \{x^L y^L, x^L y^U, x^U y^L, x^U y^U\} & (2.8) \end{aligned}$$

Estos estimadores son más sencillos de manejar, pero aproximan de forma mucho menos ajustada que lo propuesto por McCormick.

2.5.6 Subestimaciones de una función fraccionaria

2.5.6.1 Subestimaciones lineales según McCormick

Sea $f: [x^L, x^U] \times [y^L, y^U] \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ función fraccionaria $f(x,y) = x/y$ con $y \neq 0$, se pueden obtener estimaciones aplicando las subestimaciones y sobreestimaciones de McCormick para términos bilineales a $wy = x$, con $w = f(x,y)$ resultando:

$$\begin{aligned}
 x = w y &\geq w^L y + y^L w - w^L y^L \\
 x = w y &\geq w^U y + y^U w - w^U y^U & (2.9) \\
 x = w y &\leq w^U y + y^L w - w^U y^L \\
 x = w y &\leq w^L y + y^U w - w^L y^U
 \end{aligned}$$

Para utilizar estas estimaciones lineales, es necesario calcular las cotas de w , w^L y w^U que dependen de los signos de x e y según los intervalos de definición de la región o subregión en que se esté trabajando.

2.5.6.2 Subestimaciones según Quesada y Grossmann

Análogamente para términos fracciones $w = x / y$ sobre el dominio $[x^L, x^U] \times [y^L, y^U]$ con $y \neq 0$, Quesada y Grossmann (1995) propusieron estimadores no lineales que dependen de las cotas de las variables en una región o subregión.

$$\begin{aligned}
 w &\geq \frac{x^U}{y} + \frac{x}{y^L} - \frac{x^U}{y^L} & (A) \\
 w &\geq \frac{x^L}{y} + \frac{x}{y^U} - \frac{x^L}{y^U} & (B) & (2.10) \\
 w &\leq \frac{x^U}{y} + \frac{x}{y^U} - \frac{x^U}{y^U} & (C) \\
 w &\leq \frac{x^L}{y} + \frac{x}{y^L} - \frac{x^L}{y^L} & (D)
 \end{aligned}$$

Demostración:

Las desigualdades se derivan de las subestimaciones convexas propuestas por McCormick para términos bilineales:

$$\text{a) } xy \geq x^L y + y^L x - x^L y^L$$

$$\text{b) } xy \geq x^U y + y^U x - x^U y^U$$

$$\text{c) } xy \leq x^U y + y^L x - x^U y^L$$

$$\text{d) } xy \leq x^L y + y^U x - x^L y^U$$

Para que $w=x/y$ esté bien definido debe ser $y \neq 0$, luego $[y^L, y^U]$ es un intervalo enteramente positivo o enteramente negativo, resultando $yy^L \geq 0$ e $yy^U \geq 0$.

Si en (a) se divide ambos lados de la desigualdad por yy^L :

$$\frac{x}{y^L} \geq \frac{x^L}{y^L} + \frac{x}{y} - \frac{x^L}{y} \quad \text{reemplazando } x/y \text{ por } w \text{ y despejando } w$$

$$w \leq \frac{x^L}{y} + \frac{x}{y^L} - \frac{x^L}{y^L} \quad \text{que es la desigualdad (D)}$$

Si en (b) se divide ambos lados de la desigualdad por yy^U :

$$\frac{x}{y^U} \geq \frac{x^U}{y^U} + \frac{x}{y} - \frac{x^U}{y} \quad \text{reemplazando } x/y \text{ por } w \text{ y despejando } w$$

$$w \leq \frac{x^U}{y} + \frac{x}{y^U} - \frac{x^U}{y^U} \quad \text{que es la desigualdad (C)}$$

Análogamente si en (c) se divide ambos lados de la desigualdad por yy^L :

$$\frac{x}{y^L} \leq \frac{x^U}{y^L} + \frac{x}{y} - \frac{x^U}{y} \quad \text{reemplazando } x/y \text{ por } w \text{ y despejando } w$$

$$w \geq \frac{x^U}{y} + \frac{x}{y^L} - \frac{x^U}{y^L} \quad \text{que es la desigualdad (A)}$$

Si en (d) se divide ambos lados de la desigualdad por yy^U :

$$\frac{x}{y^U} \leq \frac{x^L}{y^U} + \frac{x}{y} - \frac{x^L}{y} \quad \text{reemplazando } x/y \text{ por } w \text{ y despejando } w$$

$$w \geq \frac{x^L}{y} + \frac{x}{y^U} - \frac{x^U}{y^U} \text{ que es la desigualdad (B)}$$

Todas estas estimaciones son válidas sin importar si $y^L > 0$ o $y^U < 0$ \square

Cada desigualdad tiene una parte lineal en x , una constante y otra hiperbólica en y . Para obtener una subestimación convexa, el término hiperbólico evaluado en y debe ser positivo, dependiendo de los signos de y y las cotas de x .

En el siguiente cuadro se presentan cuales subestimaciones son convexas y cuales no para las diferentes situaciones sobre el signo de las variables:

Cotas	$y > 0$		$y < 0$	
	convexo	No convexo	convexo	No convexo
$x^L \leq 0$	(D)	(B)	(B)	(D)
$x^L \geq 0$	(B)	(D)	(D)	(B)
$x^U \leq 0$	(C)	(A)	(A)	(C)
$x^U \geq 0$	(A)	(C)	(C)	(A)

La mayoría de los algoritmos propuestos en optimización global buscan subestimaciones convexas para poder aplicar los resultados que provee la teoría de programación convexa.

2.5.6.3 Comparación entre las distintas estimaciones propuestas para términos fraccionarios

Se pueden comparar las estimaciones para términos fraccionarios propuestas por Quesada y Grossmann y las obtenidas en (2.9)

Se va a demostrar que la estimación propuesta por Quesada y Grossmann aproxima de forma más ajustada.

En ambas estimaciones se trabaja con las cotas de w : $w^L = \frac{x^L}{y^L}$ y $w^U = \frac{x^U}{y^U}$

Primero se busca la forma de las subestimaciones y sobreestimaciones de Mc. Cornick a partir de las cotas de w . Para ello es necesario fijar una situación de signo para las variables, por ejemplo $0 < y^L \leq y^U$ y $x^L \leq x^U \leq 0$.

Si en las desigualdades (2.9) se despeja w resultan:

$$\begin{aligned}
 \text{a)} & -\frac{w^L}{y^L} y + \frac{x}{y^L} + w^L \geq w \\
 \text{b)} & -\frac{w^U}{y^U} y + \frac{x}{y^U} + w^U \geq w \\
 \text{c)} & -\frac{w^U}{y^L} y + \frac{x}{y^L} + w^U \leq w \\
 \text{d)} & -\frac{w^L}{y^U} y + \frac{x}{y^U} + w^L \leq w
 \end{aligned} \tag{2.11}$$

Luego reemplazando las cotas de w :

$$\begin{aligned}
 \text{a)} & -\frac{x^L}{(y^L)^2} y + \frac{x}{y^L} + \frac{x^L}{y^L} \geq w \\
 \text{b)} & -\frac{x^U}{(y^U)^2} y + \frac{x}{y^U} + \frac{x^U}{y^U} \geq w \\
 \text{c)} & -\frac{x^U}{y^U y^L} y + \frac{x}{y^L} + \frac{x^U}{y^U} \leq w \\
 \text{d)} & -\frac{x^L}{y^L y^U} y + \frac{x}{y^U} + \frac{x^L}{y^L} \leq w
 \end{aligned} \tag{2.12}$$

Considerando (a) y (b) cuando $w = -\frac{x^L}{(y^L)^2} y + \frac{x}{y^L} + \frac{x^L}{y^L}$ se cumple

$-\frac{x^L}{(y^L)^2} y + \frac{x}{y^L} + \frac{x^L}{y^L} \leq -\frac{x^U}{(y^U)^2} y + \frac{x}{y^U} + \frac{x^U}{y^U}$ trabajando la desigualdad

algebraicamente para despejar y , sabiendo que $\frac{x^U}{(y^U)^2} - \frac{x^L}{(y^L)^2} > 0$

$$y \leq \frac{y^L y^U (y^L - y^U)}{(y^L)^2 x^U - (y^U)^2 x^L} x + \frac{y^L x^U - y^U x^L}{(y^L)^2 x^U - (y^U)^2 x^L},$$

como $y^L y^U > 0$ se obtiene reordenando: $y \leq \frac{(y^L - y^U)}{y^L \frac{x^U}{y^U} - y^U \frac{x^L}{y^L}} x + \frac{\frac{x^U}{y^U} - \frac{x^L}{y^L}}{y^L \frac{x^U}{y^U} - y^U \frac{x^L}{y^L}}$

luego w resulta de las sobreestimaciones lineales de McCormick y para los signos propuestos en este caso.

$$w = \begin{cases} -\frac{x^L}{(y^L)^2}y + \frac{x}{y^L} + \frac{x^L}{y^L} & \text{si } y \leq \frac{(y^L - y^U)}{y^L \frac{x^U}{y^U} - y^U \frac{x^L}{y^L}} x + \frac{\frac{x^U}{y^U} - \frac{x^L}{y^L}}{y^L \frac{x^U}{y^U} - y^U \frac{x^L}{y^L}} \\ -\frac{x^U}{(y^U)^2}y + \frac{x}{y^U} + \frac{x^U}{y^U} & \text{en otro caso} \end{cases} \quad (2.13)$$

Por otro lado para los signos propuestos sobre x e y , la sobrestimación que resulta convexa de Quesada y Grossmann es (C): $w \leq \frac{x^U}{y} + \frac{x}{y^U} - \frac{x^U}{y^U}$

Si se compara esta desigualdad con (b) $-\frac{x^U}{(y^U)^2}y + \frac{x}{y^U} + \frac{x^U}{y^U} \geq w$

Se puede demostrar que (C) aproxima de forma más ajustada un término fraccionario que (b), chequeando la desigualdad:

$$\frac{x^U}{y} + \frac{x}{y^U} - \frac{x^U}{y^U} \leq -\frac{x^U}{(y^U)^2}y + \frac{x}{y^U} + \frac{x^U}{y^U}$$

Simplificando y acomodando la expresión: $\frac{x^U}{y} \leq 2\frac{x^U}{y^U} - \frac{x^U}{(y^U)^2}y$ o equivalentemente

$$-2\frac{x^U}{y^U} + \frac{x^U}{(y^U)^2}y + \frac{x^U}{y} \leq 0$$

Para $x^U \leq 0$ se debe verificar: $f(y) = -2\frac{1}{y^U} + \frac{1}{(y^U)^2}y + \frac{1}{y} \geq 0$

Como $0 \leq y \leq y^U$ se cumple que $\frac{1}{(y^U)^2} \leq \frac{1}{y^2}$ y por lo tanto $f'(y) = \frac{1}{(y^U)^2} - \frac{1}{y^2} \leq 0$

f resulta una función decreciente y como $f(y^U) = 0$ se verifica que

$$f(y) \geq 0 \quad \forall y \in [y^L, y^U]$$

Geoméricamente se puede ver comparando el comportamiento de la hipérbola que está a la izquierda de la primer desigualdad planteada con la recta del otro lado de la misma, ambas definidas para la variable y . Luego la estimación no lineal aproxima al término fraccionario de forma más exacta que la aproximación lineal para

$$y \geq \frac{(y^L - y^U)}{y^L \frac{x^U}{y^U} - y^U \frac{x^L}{y^L}} x + \frac{\frac{x^U}{y^U} - \frac{x^L}{y^L}}{y^L \frac{x^U}{y^U} - y^U \frac{x^L}{y^L}} .$$

Si se compara la estimación no lineal de Quesada y Grossmann dada por (D) y se compara con (a) dada por McCormick, para demostrar que la estimación no lineal es más ajustada que la lineal, se sigue la misma idea:

$$\frac{x^L}{y} + \frac{x}{y^L} - \frac{x^L}{y^L} \leq -\frac{x^L}{(y^L)^2} y + \frac{x}{y^L} + \frac{x^L}{y^L} \quad \text{simplificando} \quad \text{y} \quad \text{ordenando:}$$

$$\frac{x^L}{y} \leq -\frac{x^L}{(y^L)^2} y + 2\frac{x^L}{y^L}$$

Desigualdad que se puede demostrar trabajando y demostrando para la función

$$f(y) = \frac{1}{y} + \frac{1}{(y^L)^2} y - \frac{2}{y^L} \geq 0$$

Como $0 \leq y^L \leq y$, $\frac{1}{(y^L)^2} \geq \frac{1}{y^2}$ Resulta $f'(y) = \frac{1}{(y^L)^2} - \frac{1}{y^2} \geq 0$

f resulta así una función creciente y como $f(y^L) = 0$ se verifica que

$$f(y) \geq 0 \quad \forall y \in [y^L, y^U]$$

Nuevamente, mirando la primer desigualdad pedida su lado izquierdo es una hipérbola cuya gráfica está en el cuarto cuadrante y tiene por encima la gráfica de la recta del lado derecho y ambas coinciden en el punto $y = y^L$. Así para todo valor de y en su intervalo de definición, las sobreestimaciones propuestas por Quesada y Grossmann provocan menos errores al sobreestimar un término fraccionario.

Para las subestimaciones no convexas (A) $w \geq \frac{x^U}{y} + \frac{x}{y^L} - \frac{x^U}{y^L}$ y (B)

$w \geq \frac{x^L}{y} + \frac{x}{y^U} - \frac{x^L}{y^U}$, se pueden comparar con (c) $-\frac{x^U}{y^U y^L} y + \frac{x}{y^L} + \frac{x^U}{y^U} \leq w$ y (d)

$-\frac{x^L}{y^L y^U} y + \frac{x}{y^U} + \frac{x^L}{y^L} \leq w$ respectivamente.

Comparando primero (A) con (c) se debe ver que:

$-\frac{x^U}{y^U y^L} y + \frac{x}{y^L} + \frac{x^U}{y^U} \leq \frac{x^U}{y} + \frac{x}{y^L} - \frac{x^U}{y^L}$ reordenando es lo mismo probar que:

$$-\frac{x^U}{y^U y^L} y + \frac{x^U}{y^U} + \frac{x^U}{y^L} \leq \frac{x^U}{y}$$

Pero esta desigualdad es cierta ya que la hipérbola $\frac{x^U}{y}$ es una función cóncava

cuya envolvente convexa es el segmento de recta definido por los puntos

$(y^L, \frac{x^U}{y^L})$ y $(y^U, \frac{x^U}{y^U})$ cuya ecuación es: $-\frac{x^U y}{y^L y^U} + \frac{x^U}{y^L} + \frac{x^U}{y^U}$.

Aplicando esta subestimación al término en (A) y simplificando se obtiene:

$$w \geq \frac{x}{y^L} - \frac{x^U y}{y^L y^U} + \frac{x^U}{y^U} \quad \text{se obtiene (c)}$$

Por lo tanto la subestimación no lineal es mejor que la lineal.

Ahora comparando (B) con (d) se debe probar que:

$$-\frac{x^L}{y^L y^U} y + \frac{x}{y^U} + \frac{x^L}{y^L} \leq \frac{x^L}{y} + \frac{x}{y^U} - \frac{x^L}{y^U}$$

Simplificando y ordenando es lo mismo ver que: $-\frac{x^L}{y^L y^U} y + \frac{x^L}{y^L} + \frac{x^L}{y^U} \leq \frac{x^L}{y}$

El término $\frac{x^L}{y}$ resulta cóncavo, luego su envolvente convexa es el segmento de recta

definido por los puntos $(y^L, \frac{x^L}{y^L})$ y $(y^U, \frac{x^L}{y^U})$: $-\frac{x^L y}{y^L y^U} + \frac{x^L}{y^L} + \frac{x^L}{y^U}$.

Por ser una subestimación de la función resulta la desigualdad buscada y aplicando esta subestimación al término en (B): $w \geq \frac{x}{y^U} - \frac{x^L y}{y^L y^U} + \frac{x^L}{y^L}$ se obtiene (d)

Queda así demostrado para una situación particular de signos de x e y que las estimaciones de Quesada y Grossmann para términos fraccionarios son mejores que las reformuladas por McCormick. El ejemplo 2.5.1 muestra que ninguna de las dos subestimaciones es la envolvente convexa de la función. Razonamientos análogos pueden aplicarse para demostrar este resultado ante diferentes signos sobre las variables.

2.5.6.4 Subestimaciones según Maranas y Floudas

Maranas y Floudas (1995) proponen análogamente a lo realizado para términos bilineales, para términos fraccionarios x/y sobre el dominio $[x^L, x^U] \times [y^L, y^U]$ con $y \neq 0$, subestimarlos mediante el agregado de una nueva variable w y dos nuevas restricciones que dependen del signo de la cota sobre x .

$$\begin{aligned}
 \text{i) } w &\geq \frac{x^L}{y} + \frac{x}{y^U} - \frac{x^L}{y^U} && \text{si } x^L \geq 0 \\
 \text{ii) } w &\geq \frac{x}{y^U} - \frac{x^L y}{y^L y^U} + \frac{x^L}{y^L} && \text{si } x^L < 0 \\
 \text{iii) } w &\geq \frac{x^U}{y} + \frac{x}{y^L} - \frac{x^U}{y^L} && \text{si } x^U \geq 0 \\
 \text{iv) } w &\geq \frac{x}{y^L} - \frac{x^U y}{y^L y^U} + \frac{x^U}{y^U} && \text{si } x^U < 0
 \end{aligned} \tag{2.14}$$

(i) y (iii) son subestimaciones ya propuestas por Quesada y Grossmann en (B) y (A) que resultan convexas cuando $y \geq 0$ por ser suma de términos convexas.

(ii) y (iv) son las subestimaciones derivadas de aplicar McCormick cuando $y > 0$ y $x^L \leq x^U \leq 0$.

2.5.6.5 Subestimaciones alternativas para términos fraccionarios

Una subestimación alternativa propuesta por Quesada y Grossmann (1998)⁽¹⁵⁾ para términos fraccionarios es la siguiente:

$$\frac{x}{y} \geq \frac{1}{y} \left(\frac{x + \sqrt{x^L x^U}}{\sqrt{x^L} + \sqrt{x^U}} \right)^2 \quad (2.15)$$

2.5.7 Subestimaciones para potencias

En el caso de funciones univariadas que sean una potencia $w = x^n$ se pueden pensar diferentes situaciones:

- i) Si n es entero par, la función que define es convexa.
- ii) Si n es entero impar, la función que define es convexa para $x \geq 0$ y cóncava para $x \leq 0$. Se puede realizar una relajación convexa, lograda con una secante definida por la cota inferior de x y un punto $x > 0$ donde la secante es tangente a la curva en su parte convexa, y luego continuar con la curva donde ya es convexa.

Si $w = x^{1/n}$ análogamente las diferentes situaciones:

- i) Si n es entero par y $x \geq 0$, la función que define es cóncava y se puede subestimar con su envolvente convexa.
- ii) Si n es entero impar, la función que define es convexa para $x \leq 0$ y cóncava para $x \geq 0$. Se puede realizar una relajación convexa, lograda con la curva donde ya es convexa, unida con una secante definida por un punto $x < 0$ donde la recta es tangente a la curva en su parte convexa y la cota superior de x .

2.6 Reducciones de cotas y de la región factible en el proceso de ramificación

Como se expresaba en la introducción de este trabajo, los algoritmos de ramificar y acotar tienden a generar fácilmente una gran cantidad de subproblemas a resolver en

busca de la solución o para garantizar la globalidad de la misma cuando ésta se halla. Los algoritmos de la literatura proponen reglas que tienden a disminuir este proceso. Una propuesta que se explicará en detalle en la sección 2.7.5 es la implementada por Floudas y Visweswaran⁽⁷⁾ (1990,1993) en su algoritmo GOP. A partir de la forma en que se particiona la región factible en el algoritmo, se deducen propiedades de los problemas que dan condiciones para eliminar subregiones, reduciendo la búsqueda. Otra forma de reducir el trabajo en el proceso de ramificación, se logra trabajando con regiones factibles lo más exactas posible, de manera que, a partir de regiones más pequeñas el proceso de ramificación que le sigue llegará más rápidamente a nodos agotados. Esta propuesta la implementan algunos algoritmos usando diferentes metodologías que se estudian a continuación.

En el proceso de ramificación, se realiza una partición de una región a partir de una variable x_j , generalmente generando dos nuevos subproblemas con cambios en las cotas de la misma y permiten modificar las cotas de aquellas variables con las que se relaciona.

Más formalmente, sean (P) un problema de programación no lineal de la forma **(2.1)**, (R) su relajación de la forma **(2.2)** y x^* una solución óptima de (R) no factible para (P). Se propone una partición de la región a partir de una variable x_j^* interior del rango de la variable, que hace activa la relajación convexa (y por lo tanto no cumple con la restricción original) luego los nuevos subproblemas son:

$$\begin{aligned}
 & \text{Minimizar } f(x_1, \dots, x_n) \\
 \text{Sujeto a: } & \begin{aligned} & g_i(x_1, \dots, x_n) \leq 0 \quad i = 1, \dots, m \quad x \in X \\ & x^L \leq x_j \leq x_j^* \end{aligned} \tag{2.16}
 \end{aligned}$$

$$\begin{aligned}
 & \text{Minimizar } f(x_1, \dots, x_n) \\
 \text{Sujeto a: } & \begin{aligned} & g_i(x_1, \dots, x_n) \leq 0 \quad i = 1, \dots, m \quad x \in X \\ & x_j^* \leq x_j \leq x^U \end{aligned}
 \end{aligned}$$

La definición de estos dos nuevos problemas puede llevarse a cabo de dos maneras: incorporando la partición del rango de variación de la variable elegida en el problema original, como se hizo en esta propuesta; o incorporando la partición en el problema relajado y por lo tanto la formulación de los mismos debe hacerse con las funciones

$\overline{f}, \overline{g}_i$. Si se toman las funciones originales, cuando se busca la relajación en cada nuevo subproblema, las estimaciones hechas para términos no convexos donde interviene la variable x_j , serán más ajustadas que las relajaciones hechas antes de la partición. Esto se debe a la reducción de la longitud del intervalo de variación de la variable que provoca la partición. Esta forma cuenta con la desventaja que el punto solución x^* del problema relajado no es factible para las relajaciones de ninguno de los subproblemas, no pudiéndose tomar como punto inicial para sus resoluciones. Si se desea que el punto x^* solución del problema relajado se mantenga factible debe particionarse el problema relajado.

El cambio de cotas de la variable x_j provoca también una reducción en las cotas de las variables restantes relacionadas con ella. Si estas reducciones pueden cuantificarse en cada partición, las subregiones con las que se trabajan posteriormente serían menores.

Dos elementos motivan una reducción en la región factible, el proceso de partición realizado sobre las subregiones provocan reducciones de cotas, y las subestimaciones convexas de las funciones que se van mejorando a partir de las nuevas cotas de las variables.

El cálculo de nuevas cotas para las variables puede llevarse a cabo mediante un procedimiento algebraico, trabajando con las restricciones del problema, o mediante la resolución de problemas de optimización.

A partir de restricciones lineales, se pueden obtener cotas de las variables que la definen, es decir si:

$$\sum_j a_{ij}x_j \leq b_i, \text{ despejando } x_k: a_{ik}x_k \leq -\sum_{j \neq k} a_{ij}x_j + b_i$$

$$x_k^L = \max \left\{ x_k^L, \frac{1}{a_{ik}} \left(-\sum_{\substack{j \neq k \\ a_{ij} \geq 0}} a_{ij}x_j^L - \sum_{\substack{j \neq k \\ a_{ij} \leq 0}} a_{ij}x_j^U + b \right) \right\} \quad \text{si } a_{ik} \leq 0$$

se obtiene una cota inferior para la variable, y si es mayor que la cota inferior actual puede actualizarse.

$$x_k^U = \min \left\{ x_k^U, \frac{1}{a_{ik}} \left(-\sum_{\substack{j \neq k \\ a_{ij} \geq 0}} a_{ij}x_j^L - \sum_{\substack{j \neq k \\ a_{ij} \leq 0}} a_{ij}x_j^U + b \right) \right\} \quad \text{si } a_{ik} \geq 0 \quad (2.17)$$

se obtiene una cota superior para la variable, y si es menor que la cota superior actual, puede actualizarse. En ambos casos se logra formulación capaz de reducir el intervalo de variación de la variable.

Para restricciones no lineales sencillas como $x_1 x_2 \geq a$ con $x_1, x_2 > 0$, despejando una variable:

$x_1 \geq a/x_2$ se puede obtener una cota inferior para x_1 haciendo

$$x_1^L = \max\{x_1^L, a/x_2^U\}$$

El proceso de examinar restricciones particulares e inferir cotas lo realizó Hansen y otros (1991), determinando por ejemplo la monotonía de las funciones y haciendo análisis de intervalos.

2.6.1 Reducción de cotas usando optimización

Numerosos trabajos implementan en sus algoritmos una actualización de las cotas de las variables resolviendo dos problemas de optimización:

$$\begin{aligned} &\text{Minimizar } x_j \\ &\text{Sujeto a: } g_i(x_1, \dots, x_n) \leq 0 \quad i = 1, \dots, m \quad x \in R_k \\ & \\ &\text{Maximizar } x_j \\ &\text{Sujeto a: } g_i(x_1, \dots, x_n) \leq 0 \quad i = 1, \dots, m \quad x \in R_k \end{aligned} \tag{2.18}$$

El primer problema obtiene una actualización de la cota inferior de la variable y el segundo de la cota superior. Si el problema tiene n variables, y se desean actualizar sus cotas, por cada subregión, R_k se requiere resolver $2n$ problemas de optimización.

Como resulta una metodología muy costosa, distintos autores proponen distintas alternativas a partir de esta idea. Por ejemplo aplicar esta reducción sólo en el nodo raíz, siendo beneficioso para aquellos problemas donde se comienza con cotas más grandes que las definidas por la región factible formada por las restricciones restantes. Para resolver problemas convexos, las restricciones se reemplazan por sus

estimaciones convexas, o estimaciones lineales⁽⁶⁾ menos ajustadas pero planteando LPs más sencillos de resolver.

2.6.2 Test de reducción de cotas propuesta por Ryoo y Sahinidis

Ryoo y Sahinidis en su trabajo ^(17,18) presentan test de reducción de rangos basados en optimalidad y factibilidad. La reducción basada en optimalidad, utiliza la perturbación de un problema relajado y a partir de su resolución construye restricciones adicionales donde intervienen las variables duales asociadas.

Más formalmente, tomando el problema relajado presentado en **(2.2)**, se puede construir el problema perturbado:

$$\begin{aligned} \varphi(y) = \text{Minimizar } \bar{f}(x_1, \dots, x_n) & \quad (2.19) \\ \text{Sujeto a: } \bar{g}_i(x_1, \dots, x_n) \leq y \quad i = 1, \dots, m, \quad x \in X \end{aligned}$$

Teorema 2.6.2.1:

Supongamos que el problema **(2.2)** tiene una solución óptima finita, λ^* es el multiplicador sí y solo sí el hiperplano con ecuación $z = \varphi(0) - \lambda^* y$ es un hiperplano soporte en $y=0$ del gráfico de la función perturbación $\varphi(y)$.

Demostración: [ver en (17)]

Teorema 2.6.2.2:

Sea **(2.2)** problema de optimización convexo con valor de la función objetivo óptima igual a L . Sea U una cota superior conocida del problema original **(2.1)**. En el óptimo supongamos que la restricción $x_j - x_j^U \leq 0$ es activa con multiplicador $\lambda_j^* > 0$.

Luego el siguiente test es válido:

Test 1: $K_j = x_j^U - \frac{U-L}{\lambda_j^*}$. Si $x_j^L < K_j$ luego reemplazar $x_j^L \leftarrow K_j$

Demostración:

Sea **(2.20)** la perturbación del problema **(2.1)** :

$$\begin{aligned} \Phi(y) = \text{Minimizan } f(x_1, \dots, x_n) & \quad (2.20) \\ \text{Sujeto a: } g_i(x_1, \dots, x_n) \leq y \quad i = 1, \dots, m \quad x \in X \end{aligned}$$

(i) $\varphi(y) \leq \Phi(y)$ ya que (2.19) es una relajación convexa de (2.20) para todo y .

Sea (2.19) el problema convexo donde la única restricción perturbada es $x_j - x_j^U \leq y$

Por el Teorema 2.6.2.1 existe un hiperplano soporte del gráfico de la función $\varphi(y)$ en $y=0$: $\varphi(y) \geq \varphi(0) - \lambda_j^* y$ (ii) (ver Figura 2.6)

donde λ_j^* es el multiplicador asociado a una solución óptima finita de (2.2). Luego por (i), (ii) y teniendo en cuenta que U es una cota superior para la solución de (2.1) y L es la solución óptima de (2.2), se tiene $U \geq \varphi(y) \geq \varphi(0) - \lambda_j^* y = L - \lambda_j^* y$

Como por hipótesis la restricción $x_j - x_j^U \leq 0$ es activa en el óptimo de (2.2), para la restricción perturbada en (2.19) también es activa con $y=0$. Si se toma $y \leq 0$, es posible considerar $y = x_j - x_j^U$

Sustituyendo en $U \geq L - \lambda_j^* y = L - \lambda_j^* (x_j - x_j^U)$ y despejando teniendo en cuenta que $\lambda_j^* > 0$ resulta:

$$x_j > x_j^U - \frac{U-L}{\lambda_j^*}$$

Si esta nueva cota es mayor que la actual x_j^L , luego la cota inferior

de la variable puede actualizarse \square

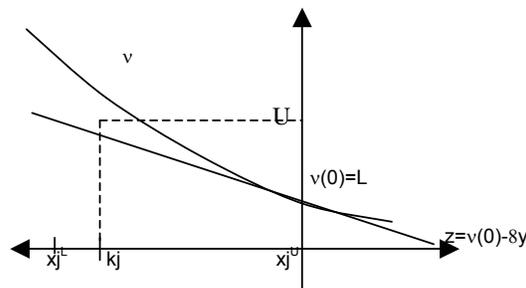


Figura 2.6: Representación del test 1

De manera similar si $x_j - x_j^L \leq 0$ es activa en la solución de (2.2), con multiplicador $\lambda_j^* > 0$.

Test 2: $\pi_j = x_j^L + \frac{U-L}{\lambda_j^*}$. Si $x_j^U > \pi_j$ luego reemplazar $x_j^U \leftarrow \pi_j$

Demostración:

Análogamente considerando que **(2.19)** el problema relajado convexo donde la única restricción perturbada es $x_j^L - x_j \leq y$.

Se tiene $U \geq \Phi(y) \geq \varphi(0) - \lambda_j^* y = L - \lambda_j^* y$

donde λ_j^* es el multiplicador asociado a $x_j^L - x_j \leq 0$ restricción activa en la solución óptima finita de **(2.2)**. Para la restricción perturbada en **(2.19)** también es activa con $y = 0$. Si se toma $y \leq 0$, es posible considerar $y = x_j^L - x_j$.

Sustituyendo $U \geq L - \lambda_j^* y = L - \lambda_j^* (x_j^L - x_j)$ y despejando teniendo en cuenta

que $\lambda_j^* > 0$ resulta: $x_j \leq x_j^L + \frac{U - L}{\lambda_j^*}$

Si esta nueva cota es menor que la actual x_j^U , luego la cota superior de la variable puede actualizarse \square

Para aquellas variables que en la solución óptima del problema relajado no están en su cota, no se pueden aplicar los test, en estos casos los autores proponen fijar dichas variables en sus cotas y resolver nuevamente el problema convexo relajado. Luego los test que se aplican son los siguientes:

Test 3: Resolver **(2.2)** tomando $x_j = x_j^U$. Si $\lambda^* > 0$, sea $\pi_j = x_j^U - \frac{L - U}{\lambda^*}$

Si $x_j^U > \pi_j$ luego reemplazar $x_j^U \leftarrow \pi_j$

Demostración:

Para lograr la integridad de la variable $x_j = x_j^U$, se incorpora al modelo la restricción $x_j^U \leq x_j$ en **(2.2)**. Si el problema aumentado tiene solución, sea el valor de la función objetivo óptima asociado igual a L .

Sea **(2.19)** el problema convexo donde la única restricción perturbada es $x_j^U - x_j \leq y$.

Ya que **(2.19)** es una relajación convexa de **(2.20)**, se cumple que $\varphi(y) \leq \Phi(y)$ para todo y .

Por el Teorema 2.6.2.1 existe un hiperplano soporte del gráfico de la función $\varphi(y)$ en $y=0$: $\varphi(y) \geq \varphi(0) - \lambda_j^* y$ donde λ_j^* es el multiplicador asociado a una solución óptima finita de **(2.2)**.

Una subestimación para el problema relajado también es una subestimación de **(2.20)**:
 $U \geq \Phi(y) \geq \varphi(0) - \lambda_j^* y = L - \lambda_j^* y$ donde U es la mejor cota superior obtenida hasta el momento.

Considerar $y \leq 0$. Como $x_j^U - x_j \leq y$, es activa con $y=0$, también será activa en la solución de **(2.19)** la restricción $y \leq 0$, luego es posible considerar $y = x_j^U - x_j$.

Sustituyendo $U \geq L - \lambda_j^* y = L - \lambda_j^* (x_j^U - x_j)$.

Si $\lambda_j^* > 0$ despejando resulta: $x_j \leq x_j^U - \frac{L-U}{\lambda_j^*}$

Si esta nueva cota es menor que la actual x_j^U , luego la cota superior de la variable puede actualizarse \square

Test 4: Resolver **(2.2)** tomando $x_j = x_j^L$. Si $\lambda_j^* > 0$, sea $K_j = x_j^L + \frac{L-U}{\lambda_j^*}$

Si $x_j^L > K_j$ luego reemplazar $x_j^L \leftarrow K_j$

Demostración:

Para lograr la integridad de la variable $x_j = x_j^L$, se incorpora al modelo la restricción $x_j \leq x_j^L$ en **(2.2)**. Si tiene solución, existe una solución óptima y sea el valor de la función objetivo óptima igual a L .

Sea **(2.19)** el problema convexo donde la única restricción perturbada es $x_j - x_j^L \leq y$.

Se cumple que $\varphi(y) \leq \Phi(y)$ para todo y , donde $\varphi(y)$ es la solución de **(2.19)** y $\Phi(y)$ de **(2.20)**.

Por el Teorema 2.6.1.1 existe un hiperplano soporte del gráfico de la función $\varphi(y)$ en $y=0$: $\varphi(y) \geq \varphi(0) - \lambda_j^* y$ donde λ_j^* es el multiplicador asociado a una solución óptima finita de **(2.2)**.

Como la solución de **(2.20)** debe ser mejor que la última cota superior obtenida, se tiene $U \geq \Phi(y) \geq \varphi(0) - \lambda_j^* y = L - \lambda_j^* y$

Considerar $y \leq 0$. Como $x_j - x_j^L \leq y$ es activa con $y=0$ también será activa en la solución de **(2.19)** la restricción $y \leq 0$ luego es posible considerar.

Sustituyendo.

Si $\lambda_j^* > 0$ despejando resulta: $x_j \geq x_j^L + \frac{L-U}{\lambda_j^*}$

Si esta nueva cota es mayor que la actual x_j^L luego la cota inferior de la variable puede actualizarse \square

Los test 3 y 4 si se aplicaran en cada subproblema relajado implicaría resolver $2n$ problemas cada vez, por tal motivo los autores proponen sólo aplicarlo en el nodo raíz. Con estos test se logra una nueva cota para las variables, que además permite actualizar las cotas de las variables relacionadas con ellas. Los autores proponen algunos lineamientos de un trabajo algebraico para lograr cotas de variables agregadas, por ejemplo, buscando una similitud con el trabajo de actualizar cotas en programación entera.

Otros resultados presentados se refieren al agregado de restricciones o modificación del lado derecho de restricciones existentes, con el fin de reducir la región factible.

Teorema 2.6.2.3:

Sea **(2.2)** un problema de optimización convexo con valor de la función objetivo óptima L , donde la restricción $\bar{g}_i(x) \leq 0$ es activa en el óptimo de **(2.2)** con valor del multiplicador asociado $\mu_i^* > 0$. Sea U una cota superior para el problema **(2.1)**. Luego

la restricción: $\bar{g}_i(x) \geq -\frac{U-L}{\mu_i^*}$ no excluye ninguna solución con valor de la función

objetivo mejor que U y puede agregarse a la formulación de **(2.1)**.

Demostración:

El razonamiento es similar al aplicado en el teorema anterior.

Sea **(2.2)** un problema con solución cuyo valor de la función objetivo óptima es L , y donde la restricción $\bar{g}_i(x) \leq 0$ es activa en el óptimo con valor del multiplicador asociado $\mu_i^* > 0$.

Sea **(2.3)** que se obtiene a partir de **(2.2)**, perturbando solo la restricción $\bar{g}_i(x) \leq 0$ es decir: $\bar{g}_i(x) \leq y$.

Por el Teorema 2.5.1: $\varphi(y) \geq \varphi(0) - \mu_i^* y$ donde $\varphi(y)$ es el valor objetivo óptimo de **(2.3)** para cada y definimos $L = \varphi(0)$.

Sea U una cota superior para el problema **(2.1)** que es cota válida también para las soluciones de **(2.20)**: $U \geq \Phi(y) \geq \varphi(y)$. Combinando las desigualdades resulta $U \geq L - \mu_i^* y$.

Considerar $y \leq 0$. Como $\bar{g}_i(x) \leq 0$ es activa en el óptimo de **(2.2)**, $y \leq 0$ es activa en **(2.19)**, luego tomando, $y = \bar{g}_i(x)$ se obtiene: $U \geq L - \mu_i^* \bar{g}_i(x)$

Por las condiciones de Kuhn-Tucker $\mu_i^* > 0$ luego $\bar{g}_i(x) \geq -\frac{U-L}{\mu_i^*}$ \square

En general estas desigualdades que generan el teorema son no convexas.

Otra propuesta para realizar reducción de cotas se explica en el capítulo 3 y fue implementada en el algoritmo desarrollado por Marcovecchio y otros.

2.7 Ideas centrales de algunos algoritmos para problemas no convexos

2.7.1 Algoritmo de Ryoo y Sahinidis

Ryoo y Sahinidis⁽¹⁷⁾ (1995) en su trabajo presentan un algoritmo para encontrar óptimos globales para problemas de programación no lineal NLP y problemas de programación no lineal entera mixta MINLP. Formulan un algoritmo que genera una secuencia de subproblemas relajados a partir de la subdivisión de la región investigada. Los subproblemas se construyen a partir de subestimaciones convexas de todas las funciones no convexas que forman el modelo. Para llevar a cabo este proceso, si es necesario se realiza una reformulación del problema en caso de ser separable, sumando subestimaciones de funciones de variable simple; si no es separable, se agregan nuevas variables y restricciones para hacerla separable (McCormick, 1972)⁽¹²⁾. Una vez que se subestiman las funciones del modelo se genera un nuevo problema relajado convexo que se resuelve por técnicas de optimización convexa. La solución de este problema relajado permite nuevas subdivisiones. Se elige entre las variables que definen términos no convexos aquella que tiene mayor diferencia entre la subestimación y el valor real de la función. Si x_j^* es la variable elegida, los nuevos subproblemas generados tendrán la variación de x_j como sigue: $[x_j^L, x_j^*]$ y $[x_j^*, x_j^U]$. En el caso que un mínimo local se conozca, éste puede usarse como elemento de partición. Esto hace las subestimaciones

exactas en la solución candidata (Swaney, 1990). Para elegir la variable que produce la partición, se busca una variable que cumple con la subestimación de un término no convexo pero no con la restricción a la que pertenece el término. Como el valor actual de la variable no cumple con la restricción no convexa, indica que es un punto interior del intervalo y garantiza que las particiones que definan al mismo darán intervalos estrictamente menores. De igual manera para particionar a partir de un mínimo local, éste deberá ser un punto interior del intervalo.

Transformación del modelo

Si el NLP tiene variables binarias, se modelan en un marco continuo de la forma:

$$0 \leq x \leq 1, \quad x(1-x) \leq 0$$

Las variables enteras pueden tratarse como en los algoritmos de ramificar y acotar estándar (Nemhauser and Wolsey, 1998)

Para los términos bilineales xy se agrega una nueva variable $u=xy$, se realiza una sustitución con la igualdad $2xy = u^2 - x^2 - y^2$ y los términos cóncavos se reemplazan por subestimaciones apropiadas.

Términos multilineales se tratan en forma recursiva aplicando la regla para términos bilineales, por ejemplo para xyz , se toma $w=yz$ y se reemplaza el triple producto por xw .

El elemento clave del algoritmo es la producción de test de reducción de rangos basados en optimalidad y factibilidad para acelerar la convergencia. La reducción basada en optimalidad, utiliza la perturbación del problema relajado y a partir de su resolución construye restricciones adicionales donde intervienen las variables duales asociadas. Cada vez que se resuelve el problema relajado, se pueden aplicar los test 1 y 2 (Ver sección 2.6) sobre las variables que en la solución están en alguna de sus cotas, permitiendo una reducción del intervalo de variación de la misma y en consecuencia de todas las variables que están relacionadas con ellas. Para aquellas variables que en la solución no están en alguna de sus cotas, mediante una integralidad de las mismas (se las fija temporalmente en su cota superior o inferior), se resuelve el problema resultante. Si este problema tiene solución provee de variables duales que permiten generar restricciones que excluyan soluciones con peor valor objetivo que la cota superior hasta el momento obtenida (test 3 y 4 de la sección 2.6).

Otra reducción de la región factible que utilizan está presentada en el Teorema 2.6.2.3 de la misma sección incorporando nuevas restricciones generalmente no convexas al problema. Las reducciones basadas en factibilidad, utilizan las restricciones del problema para inferir cotas más ajustadas de las variables, como la obtención de nuevas cotas a partir de la ecuación (2.17) presentada en la sección 2.6.

Para el trabajo de la ramificación se considera primero aquel problema con cota inferior menor.

Se proponen tres estrategias probadas en 20 problemas test. La primer estrategia considera un paso donde se busca, en lo posible, una solución del problema original para ajustar la cota superior en cada iteración. La segunda estrategia, a la primer estrategia se le agregan los test 1 y 2 y considera reducciones a partir de la reformulación del problema y de las restricciones del mismo. Finalmente la tercera, a partir de la segunda se incorporan los test 3 y 4. La evaluación se realiza considerando tiempo y numero de nodos que recorrió.

De los resultados se pueden observar las mejoras que se logra en las estrategias 2 y 3 cuando se incluye una estrategia de reducción de dominio, desde no converger a lograr convergencia y en general, en tiempo reducido. Para aquellos problemas donde el número de nodos fue grande, se implementó para términos bilineales las aproximaciones propuestas por McCormick, logrando mejores resultados.

Si bien las mejoras en la convergencia logradas en las dos últimas estrategias son significativas, como en ambas situaciones los test implementados se realizan en forma conjunta con un trabajo de reducción a partir de la factibilidad, no es posible ver cuan beneficiosos son estos test por sí solos.

2.7.2 Algoritmo de Byrne y Bogle

Byrne y Bogle⁽⁶⁾ (1999) en su trabajo plantean la estrategia de análisis de intervalo como un procedimiento de acotamiento, en dos algoritmos para resolver problemas no convexos usando reformulaciones lineales del modelo. Los algoritmos se llaman IGOR de "Interval global optimization relaxation" y RIGOR de "reduced Interval global optimization relaxation". Se manifiesta que los algoritmos anteriores a este trabajo no explotan las representaciones simbólicas obtenidas para los problemas y muestran como pueden ser utilizadas para mejorar la aplicación de los algoritmos basados en intervalos.

El trabajo propone reemplazar cada término no convexo por una nueva variable a la cual se le eligen restricciones de cotas de forma conveniente. Este procedimiento

llamado α - sustitución y β -sustitución, sugerido por McCormick (1976) y que también utilizaron Smith y Pantelides (1996) en su trabajo.

Definiciones

α - sustitución: se agrega una nueva variable en la función objetivo.

$$\begin{aligned} & \text{Minimizar } \alpha \\ & \text{Sujeto a: } f(x_1, \dots, x_n) \leq \alpha \\ & g_i(x_1, \dots, x_n) \leq 0 \quad i = 1, \dots, m \quad x \in X \end{aligned} \tag{2.21}$$

donde f es la función objetivo del modelo.

β -sustitución: se agrega una nueva variable en las restricciones y se separan en dos restricciones más simples.

$$g(u(x)) \leq 0 \quad g(\beta) \leq 0 \quad \beta = u(x) \tag{2.22}$$

Se utilizan dos tipos de subestimaciones lineales derivadas de una extensión natural y formas del valor medio, donde en la primer subestimación puede realizarse a cualquier término, mientras que en la segunda subestimación es aplicable a funciones diferenciables una vez.

Relajación usando análisis de intervalo

Sea $F : I^n \rightarrow I$, una inclusión de $f : R^n \rightarrow R$,

$$x \in X \Rightarrow f(x) \in F(X) \quad \text{donde } x \in R^n, X \in I^n \tag{2.23}$$

Las funciones inclusión pueden producirse de diferentes maneras como extensión natural, expansiones de Taylor y formas de valor medio con diferentes propiedades.

Relajación de la extensión natural (NE)

En este trabajo se plantea un análisis de intervalo para las variables. Para $\alpha = x_1 x_2$ en un intervalo, se plantea que su extensión natural es la propuesta por Quesada y Grossmann (1995) ((2.7) de la sección 2.5), pero la opinión de los autores es que si bien produce cotas constantes sobre α , son más conservativas que necesarias.

Es posible pensar una de las variables del término bilineal como variable continua y producir una subestimación y una sobreestimación para α :

$$\alpha \geq \min\{x_1 x_2^L, x_1 x_2^U\} \quad \text{y} \quad \alpha \leq \max\{x_1 x_2^L, x_1 x_2^U\} \quad (2.24)$$

Estas estimaciones definen una restricción que no es diferenciable en $x=0$. Para incorporar estas estimaciones y evitar la diferenciable, se hace una primer partición de la región tomando como punto de corte $x_1 = 0$.

Estimaciones análogas se obtienen tomando x_2 como variable continua y bajo las mismas consideraciones:

$$\alpha \geq \min\{x_2 x_1^L, x_2 x_1^U\} \quad \text{y} \quad \alpha \leq \max\{x_2 x_1^L, x_2 x_1^U\} \quad (2.25)$$

Relajación inclusión del valor medio (MV)

Esta relajación se puede aplicar a términos diferenciables. Considerar la inclusión valor medio de análisis de intervalo:

$$f(x) \in f(c) + (x-c)^t F'(X) \quad c \in X \quad F'(X) = [d^L, d^U]$$

$$f(x) \geq f(c) + \min\{(x-c)^t d^L, (x-c)^t d^U\}$$

es una subestimación lineal no diferenciable en c . Luego para incluir una relajación al modelo se puede restringir c a un intervalo de X que esté incluido, con signo constante para $x-c$ sobre X , luego se agrega al modelo en ese intervalo seleccionado dos subestimaciones:

$$f(x) \geq f(x^L) + (x-x^L)^t d^L \quad (2.26)$$

$$f(x) \geq f(x^U) + (x-x^U)^t d^U$$

y dos sobreestimaciones:

$$f(x) \leq f(x^L) + (x-x^L)^t d^U \quad (2.27)$$

$$f(x) \leq f(x^U) + (x-x^U)^t d^L$$

Se reformula el modelo incorporando estas relajaciones y se reemplaza cada término no convexo por una nueva variable a las que se les genera cotas eligiéndolas de una forma más exacta que tomar el mayor y menor valor que puede tomar el término en discusión.

Luego se construyen dos tipos de subestimaciones lineales derivadas de una extensión natural y formas del valor medio. La relajación del valor medio usa la expansión de Taylor de primer orden para dar cotas a las nuevas variables. La expansión se hace alrededor de las cotas inferior y superior de las variables intervinientes.

Ejemplo 2.7.2.1: Considerar la restricción de un NLP: $f(x) = x^2 + 10 \geq 0$, en $[-2, 3]$

Ya que $f'(x) = 2x$, la relajación MV (2.26) define las siguientes estimaciones:

$$x^2 + 10 \geq 9 + 10 + (x - 3)6 \quad (i)$$

$$x^2 + 10 \geq 4 + 10 + (x + 2)(-4) \quad (ii)$$

$$x^2 + 10 \leq 4 + 10 + (x + 2)6 \quad (iii)$$

$$x^2 + 10 \leq 9 + 10 + (x - 3)(-4) \quad (iv)$$

Observando las siguientes figuras se puede ver que para esta función las cotas superiores logradas por la inclusión natural: $10 \leq f(x) \leq 19$ están más ajustadas que las cotas logradas por la relajación MV. Lo mismo sucede para cualquier función cóncava.

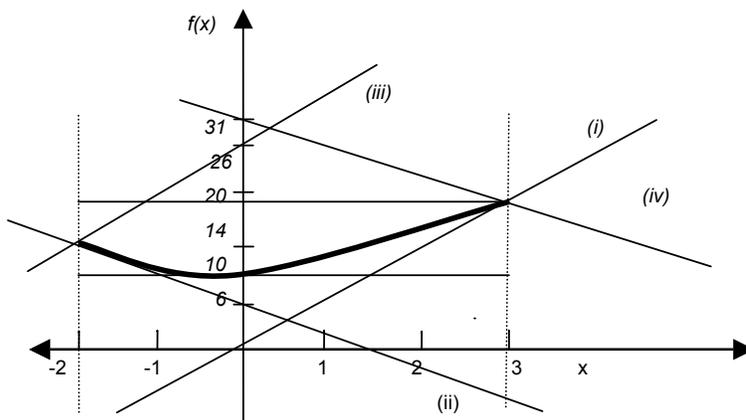


Figura 2.7.2.1 a) $f(x) = x^2 + 10$, su relajación MV e inclusión de intervalo

La relajación NE se puede aplicar realizando una partición del intervalo $[-2, 3]$

tomando el origen como punto de corte. El agregado de la relajación NE puede mejorar las estimaciones logradas hasta el momento (ver figura 2.7.2.1 a)), especialmente en problemas con rangos con gran longitud.

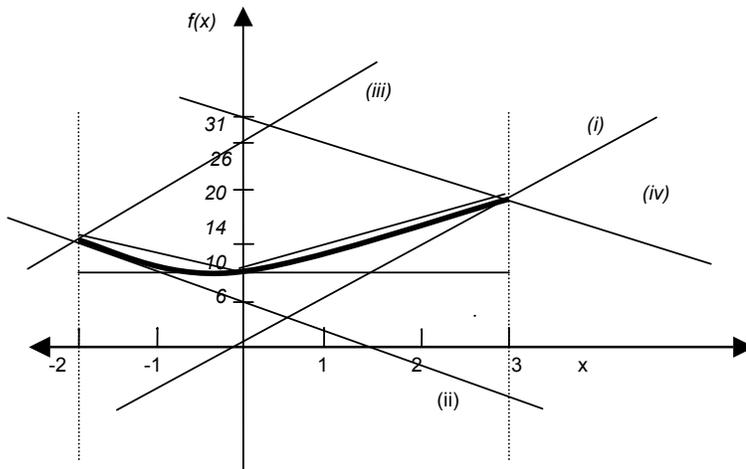


Figura 2.7.2.1 b) $f(x) = x^2 + 10$, y sus relajaciones MV y NE

Ejemplo 2.7.2.2: Un problema propuesto por Floudas, Aggarwal y Ciric (1989) es

$$\text{Min } f(x) = -x_1 + x_1 x_2 - x_2$$

$$\text{Suj. a: } -6x_1 + 8x_2 \leq 3$$

$$3x_1 - x_2 \leq 3$$

$$x_1 \geq 0 \quad x_2 \leq 5$$

El problema tiene dos mínimos en (0.916, 1.062) y (1.167, 0.5) con $f(x) = -1.0052$ y $f(x) = -1.0833$ respectivamente. La solución global está sobre la segunda restricción.

Este problema tiene una función objetivo con un término bilineal y el resto es lineal.

Usando α -sustitución para el término bilineal y la relajación de extensión natural (NE) el problema se redefine de la siguiente manera:

$$\text{Min } f(x) = -x_1 + \alpha - x_2$$

$$\text{Suj. a: } \alpha \geq \min \{x_1 x_2^L, x_1 x_2^U\}$$

$$\alpha \geq \min \{x_2 x_1^L, x_2 x_1^U\}$$

$$-6x_1 + 8x_2 \leq 3$$

$$3x_1 - x_2 \leq 3$$

$$x_1 \geq 0, \quad x_2 \leq 5, \quad \alpha \in X_1 X_2$$

donde $X_1 X_2$ es el intervalo de inclusión natural. Para que el problema recién formulado sea diferenciable se particiona en el origen.

Aplicando la relajación MV el problema resulta:

$$\text{Min } f(x) = -x_1 + \alpha - x_2$$

$$\text{Suj. a: } \alpha \geq x_1^L x_2^L + (x_1 - x_1^L) x_2^L + (x_2 - x_2^L) x_1^L$$

$$\alpha \geq x_1^U x_2^U + (x_1 - x_1^U) x_2^U + (x_2 - x_2^U) x_1^U$$

$$-6x_1 + 8x_2 \leq 3$$

$$3x_1 - x_2 \leq 3$$

$$x_1 \geq 0$$

$$x_2 \leq 5$$

$$\alpha \in X_1 X_2$$

donde las relajaciones que se aplican al término bilineal, resultan la cápsula convexa propuesta por Mc Cormick ((2.4) de la sección 2.5)

El algoritmo propuesto es un algoritmo de ramificación actuando sobre las variables que intervienen en términos no lineales. Por la sustitución que se hace, se disminuye la cantidad de variables sobre las que ramifica. De éstas, ramifica sobre las ramas con cota inferior menor.

Se generan cotas inferiores y superiores del valor objetivo que se van aproximando. Las cotas inferiores se obtienen resolviendo el problema relajado que resulta lineal ya que se relaja todo término no lineal, con relajaciones a partir de la extensión natural para términos no diferenciables y relajaciones con valor medio para términos diferenciables.

Las cotas superiores se obtienen del punto solución del LP acotado inferiormente si es punto factible, o resolviendo el problema no convexo a partir de ese punto. Los autores exponen que la solución se obtiene en las primeras iteraciones, por tal motivo el trabajo más extenso es la aproximación de las cotas inferiores a la cota superior. Por este motivo no es necesario buscar actualizar la cota superior con la misma frecuencia que la cota inferior.

Una variante de esta propuesta da lugar al segundo algoritmo que surge de agregar una etapa de reducción de intervalos para las variables que intervienen en términos no lineales del problema. Esta etapa de reducción se lleva a cabo resolviendo dos problemas lineales por cada variable, cuyos objetivos son maximizar y minimizar el

valor de la variable sujeta al conjunto de restricciones linealizadas del modelo y una restricción adicional formada por la parte lineal de la función objetivo acotada superiormente por la mejor cota superior \bar{y} hallada hasta el momento ((2.16) de la sección 2.6) Este ajuste de cotas es beneficioso cuando las cotas modeladas inicialmente son más grandes que la región factible definida por las restricciones restantes.

Si el problema reformulado es de la forma:

$$\begin{aligned} & \text{Min}_{x \in X^k} c^t x \\ & \text{su}j.a: Ax \leq b \end{aligned}$$

Para cada x^i que aparece en un término no lineal se resuelven los dos subproblemas:

$$\begin{array}{ll} \text{Min}_{x \in X^k} x^i & \text{Max}_{x \in X^k} x^i \\ \text{su}j.a: c^t x \leq \bar{y} & \text{su}j.a: c^t x \leq \bar{y} \\ Ax \leq b & Ax \leq b \end{array}$$

Si estos problemas son infactibles se eliminan las regiones asociadas.

Los problemas test resueltos mostraron que la relajación MV ajusta mejor que la subestimación NE; y que el algoritmo RIGOR es más eficiente que IGOR.

Está implementado con Matlab 4.2c y usa Toolbox de Optimización.

La propuesta del algoritmo es trabajar con subestimaciones lineales permitiendo una resolución fácil en cada iteración de las relajaciones por ser PLs, en contraposición a usar subestimaciones más relajadas se buscan subestimaciones lineales desde distintas estrategias para lograr un ajuste que lo compense.

2.7.3 Algoritmo de Smith and Pantelides

Smith and Pantelides⁽²²⁾ (1999) en su trabajo presentan un algoritmo para resolver problemas de programación no convexos enteros mixtos MINLP. El aporte que se considera más significativo de este trabajo es la reformulación simbólica general del problema que se plantea con poca intervención de usuario (mismos autores 1996). El problema reformulado luego se resuelve con un algoritmo de ramificación y acotación espacial el cual ramifica sobre variables enteras y continuas.

Se presenta una forma de implementación eficiente usando la idea de maestro y esclavo trabajando con equipos en red y en paralelo. El algoritmo se implementó en un entorno de modelado *gPROMS*

Reformulación del problema

A partir de un MPNL no convexo en las desigualdades se agregan variables de holgura para obtener un MPNL con restricciones de igualdad. Luego la reformulación del problema se obtiene identificando en las funciones intervinientes, términos algebraicos simples relacionados por cinco operaciones elementales: suma, resta, multiplicación, división y exponenciación y un número pequeño de funciones trascendentes que mediante composición de funciones definen funciones más complejas.

Ejemplo 2.7.3.1:

Dada la restricción de la forma $Fx - yw - 2Ae^{-E/RT} w^2 V = 0$

donde x , y , w y T son variables y el resto son constantes, se reemplaza por una expresión lineal de la forma: $Fx - w_1 - 2AVw_5 = 0$

junto a las relaciones no lineales:

$$\begin{aligned} w_1 &= yw \\ w_2 &= -\frac{E}{RT} \\ w_3 &= e^{w_2} \\ w_4 &= w^2 \\ w_5 &= w_3 w_4 \end{aligned}$$

w_j son variables auxiliares donde w_1 y w_5 son bilineales, w_2 es una fracción, w_3 una función exponencial, y w_4 potencia. Para cada término individual de una expresión, se reemplaza por una nueva variable y se hace una relajación convexa sobre ella. La idea seguida es la misma que implementaron Byrne y Bogle⁽⁶⁾ en su trabajo y llamaron β -sustitución (sección 2.7.2)

Todo problema puede reformularse por el procedimiento anterior y llevarse a la forma general:

$$\begin{aligned}
& \text{Min}_{w} \quad w_{\text{obj}} \\
& \text{Suj.a: } Aw = b \\
& \quad w^L \leq w \leq w^U \\
& \quad w_k \equiv w_i w_j \quad \forall (i,j,k) \in F_{bt} \\
& \quad w_k \equiv \frac{w_i}{w_j} \quad \forall (i,j,k) \in F_{ffi} \\
& \quad w_k \equiv w_i^n \quad \forall (i,k,n) \in F_{et} \\
& \quad w_k \equiv \ln(w_i) \quad \forall (i,k) \in F_{uft} \\
& \quad y \in [y^L, \dots, y^U]
\end{aligned} \tag{2.28}$$

donde el vector de variables w comprende las variables continuas x , las variables discretas y , las variables de holguras que son necesarias para convertir las desigualdades en igualdades y las variables continuas auxiliares que se introducen en la reformulación. A y b es una matriz y un vector respectivamente con coeficientes reales.

Los conjuntos F_{bt}, F_{ffi}, F_{et} y F_{uft} corresponden a términos bilineales, fracciones simples, potencias simples y términos univariados simples respectivamente. Las restricciones de igualdad lineales no se reformulan. El problema reformulado por su construcción es equivalente al problema original. Se plantea un algoritmo donde se automatiza este procedimiento usando una estructura de árbol (mismos autores 1996).

Reformulación de variables

Si el NLP tiene variables binarias, se adoptó, al igual que Ryoo y Sahinidis, la formulación continua $0 \leq x \leq 1, \quad x(1-x) \leq 0$

Al que se le aplican las subestimaciones de McCormick:

$$0 \leq x \leq 1, \quad x-w = 0, \quad w \geq 0, \quad w \geq 2x-1$$

Para llevar variables enteras y , a una expresión continua se propone: $\prod_{k=y^L}^{y^U} (y-k) = 0$

Donde k son todos los valores posibles que toma la variable entera y . Esta propuesta puede resultar compleja e ineficiente.

Relajaciones del problema no convexo

Se proponen dos formas de relajación del problema para el cálculo de cotas inferiores. La primera surge de usar relajaciones convexas lineales (McCormic), obteniendo primero un MILP problema de programación lineal entero mixto al que si se formulan todas las variables como continuas, la relajación define un LP.

La segunda opción utiliza relajaciones convexas no lineales (Grossmann) obteniéndose un MINLP problema de programación no lineal entero mixto y si se formulan todas las variables continuas, la cota inferior se obtiene resolviendo un NLP convexo.

De las opciones propuestas las más recomendadas son las que obtienen la cota inferior resolviendo NLP o LP con todas variables continuas por ser propuestas menos costosas.

Las estimaciones usadas para términos bilineales y fraccionarias son las propuestas por Quesada y Grossmann y la reformulación de Mc. Cornick (sección 2.5).

Funciones tales como logaritmos, raíces cuadradas, potencias, son cóncavas o convexas en todo su dominio se puede usar la función secante para la relajación convexa.

Para el cálculo de las cotas superiores se proponen tres esquemas: mediante puntos factibles; una segunda alternativa, plantea resolver el problema tratando todas las variables del mismo como variables continuas y luego agregar restricciones de la forma $x(1-x)=0$ para forzarlas a tomar valores enteros a aquellas que lo requieran. Sin embargo, estas restricciones hacen en general el problema fuertemente no convexo y puede hallar la convergencia en la búsqueda de un punto factible. Una tercer alternativa es fijar las variables discretas y luego resolver el NLP no convexo. Para obtener el valor de las variables discretas, se puede resolver el problema relajado convexo MILP o MINLP definido por subestimaciones y fijarlas en sus valores óptimos; o resolver el problema LP o NLP y redondear los valores óptimos.

El proceso de ramificación se realiza primero sobre las variables con mayor diferencia entre la relajación convexa (valor de la nueva variable) y el valor del término no lineal en valor absoluto. Se determinan reglas para evaluar el error cometido en la aproximación de cada término no lineal. Las variables asociadas a términos univariados no convexos son entonces las que propondrán la partición en subregiones y se priorizan las variables discretas sobre las continuas.

Ajuste de cotas

El algoritmo cuenta con un paso de ajuste de cotas que sigue lo propuesto por Quesada y Grossmann (1995). Se calculan cotas de las nuevas variables resolviendo dos problemas convexos cuyas restricciones son las funciones relajadas (Ver sección 2.6.1). Este procedimiento se realiza en una etapa inicial ya que para un problema con n variables requiere de la resolución de $2n$ problemas de optimización, resultando computacionalmente costoso. En las etapas siguientes se actualizan las cotas con un procedimiento menos exacto, por ejemplo basado en la factibilidad usando las cotas de las variables en una región o subregión.

Para las restricciones lineales $\sum_j a_{ij}x_j = b_i$ obtenidas de la reformulación:

$$\text{Si } a_{ik} \geq 0: \quad \begin{aligned} x_k^U &= \min \left\{ x_k^U, \frac{1}{a_{ik}} \left(- \sum_{j \neq k} \min(a_{ij}x_j^L, a_{ij}x_j^U) + b \right) \right\} \\ x_k^L &= \max \left\{ x_k^L, \frac{1}{a_{ik}} \left(- \sum_{j \neq k} \max(a_{ij}x_j^L, a_{ij}x_j^U) + b \right) \right\} \end{aligned}$$

$$\text{Si } a_{ik} \leq 0: \quad \begin{aligned} x_k^U &= \min \left\{ x_k^U, \frac{1}{a_{ik}} \left(- \sum_{j \neq k} \max(a_{ij}x_j^L, a_{ij}x_j^U) + b \right) \right\} \\ x_k^L &= \max \left\{ x_k^L, \frac{1}{a_{ik}} \left(- \sum_{j \neq k} \min(a_{ij}x_j^L, a_{ij}x_j^U) + b \right) \right\} \end{aligned}$$

Estas estimaciones son similares a las explicadas en la sección 2.6 para desigualdades lineales.

Para actualización de cotas de los términos bilineales, se usan las ecuaciones (2.8) desarrolladas por Quesada y Grossmann (sección 2.5) y análogamente pueden deducirse para fracciones.

2.7.4 Algoritmo α BB

Maranas, Androulakis, Adjiman y Floudas⁽⁷⁾ (1994, 1995, 1996) dan el marco teórico del algoritmo α BB para resolver problemas de optimización no lineal restringidos para funciones diferenciables dos veces. Las propiedades teóricas del algoritmo garantizan la llegada al óptimo global con una ε -convergencia finita.

En cada iteración se genera una cota inferior y una cota superior con un paso de ramificación y otro de acotación. En el paso de acotación la cota inferior se obtiene usando subestimaciones de las funciones del problema y resolviendo el NLP convexo. Para las subestimaciones se descompone cada función en suma de términos de la forma lineal, bilineal, trilineal, fraccionaria, trilineal fraccionaria, convexo, cóncavo univariado, o no convexo general. Los términos lineales y convexos no se transforman. Para subestimar términos bilineales se usan las subestimaciones de McCormick y Al.Khayyal y Falk (1983). Para subestimar términos trilineales, trilineales fraccionarios y términos fraccionarios se emplean las subestimaciones propuestas por Maranas y Floudas (1995), (sección 2.5). Los términos cóncavos se subestiman con su envolvente convexa. El principal aporte del trabajo es la presentación de una subestimación para los términos no convexos generales con una función $\mathfrak{I}(x)$ convexa propuesta por Maranas y Floudas (1994), definida como:

$$\mathfrak{I}(x) = f(x) + \sum_{i=1}^n \alpha_i (x_i^L - x_i)(x_i^U - x_i)$$

donde α_i son escalares positivos elegidos para que resulte convexa. Como a la función f se le suman todos términos negativos, la nueva función es una subestimación; y que además no agrega nuevas variables al problema.

El siguiente teorema da una herramienta más detallada de la forma de calcular los valores α_i .

Propiedad 2.7.4.1:

Sean las funciones $f(x) \in C^2$ y $\mathfrak{I}(x)$ definida como:

$$\mathfrak{I}(x) = f(x) + \alpha \sum_{i=1}^n (x_i^L - x_i)(x_i^U - x_i) \quad \text{donde } \alpha \geq \max \left\{ \max_{x_i^L \leq x_i \leq x_i^U} (0, (-\frac{1}{2} \lambda_i^f)) \right\} \quad (2.30)$$

λ_i^f es el autovalor de la matriz hessiana de la función f . Luego la función $\mathfrak{I}(x)$ es una función convexa.

En el libro de Floudas⁽⁷⁾ se detallan distintos métodos y sus órdenes de convergencia para obtener valores para este nuevo parámetro, siendo la principal preocupación.

Las cotas superiores se obtienen de soluciones locales del NLP no convexo.

El paso de ramificación en su primer versión, se realiza, subdividiendo la región a partir de la variable con mayor longitud en su intervalo de definición actual. Luego se proponen otras estrategias como por ejemplo buscar la máxima distancia entre el término y su subestimación; con el agregado de variantes como calcular una medida de relevancia de la variable en el problema, entre otras.

Los resultados computacionales se obtienen trabajando con MINOS 5.4 en una computadora HP9000/730 con una tolerancia para la convergencia de 0.001 generalmente, casos específicos se establecen otras tolerancias menores.

Otros algoritmos que se proponen son SMIN- α BB y GMIN- α BB para resolver problemas MINLP.

2.7.5 Algoritmo de optimización global GOP (Global OPTimization)

El algoritmo de optimización global GOP fue desarrollado por Floudas y Visweswaran⁽⁷⁾ (1990,1993, 1996) y es un algoritmo de descomposición primal dual relajado.

Formulación del modelo

La forma general del problema a resolver es:

$$\begin{aligned} & \underset{x,y}{\text{Min}} f(x,y) \\ & \text{Suj.a: } g(x,y) \leq 0 \\ & \quad h(x,y) = 0 \\ & \quad x \in X, y \in Y \end{aligned} \tag{2.31}$$

donde X, Y son conjuntos convexos compactos de \mathfrak{R}^n y \mathfrak{R}^m y g y h son vectores de restricciones de desigualdad e igualdad. Todas las funciones son diferenciables. Un problema satisface la condición (A) si satisface:

- a) $f(x,y)$ es convexa en x para todo y fijo, y convexa en y para todo x fijo.
- b) $g(x,y)$ es convexa en x para todo y fijo, y convexa en y para todo x fijo.
- c) $h(x,y)$ es afín en x para todo y fijo, y afín en y para todo x fijo.
- d) Un calificador de primer orden apropiado se satisface para todo y fijo.

Por ejemplo el calificador de Slater: sean X conjunto abierto, g vector de funciones pseudoconvexas en x' . Para toda restricción activa en x' , $(I = \{i : g_i(x') = 0\})$, y continua en toda restricción no activa y $\exists x \in X / g_i(x) < 0, \forall i \in I$.

Problema primal y problema dual relajado

Si se fija un valor de $y = y^k \in Y$ resulta un subproblema que por tener la condición A es un problema de programación convexa y la solución que se obtiene de él es global. Este problema se llama Problema Primal:

$$\begin{aligned} \underset{x \in X}{\text{Min}} f(x, y^k) \\ g(x, y^k) \leq 0 \\ h(x, y^k) = 0 \end{aligned} \quad (2.32)$$

El problema (2.31) puede describirse como:

$$\begin{aligned} \underset{y}{\text{Min}} v(y) \\ \text{Suj.a : } v(y) = \underset{x \in X}{\text{Min}} f(x, y) \\ \text{Suj.a : } h(x, y) = 0 \\ g(x, y) \leq 0 \\ y \in Y \cap V \\ V = \{y : h(x, y) = 0, g(x, y) \leq 0 \text{ para } x \in X\} \end{aligned} \quad (2.33)$$

El problema de minimización interno es paramétrico en y . Para un y^k fijo la solución del NLP interior es idéntica a la solución de su correspondiente problema dual por el Teorema de dualidad fuerte.

El problema dual es de la forma:

$$\left\{ \begin{array}{l} \underset{x \in X}{\text{Min}} f(x, y^k) \\ g(x, y^k) \leq 0 \\ h(x, y^k) = 0 \end{array} \right\} = \sup_{\mu \geq 0, \lambda} \inf_{x \in X} \{f(x, y^k) + \lambda h(x, y^k) + \mu g(x, y^k)\} \\ \forall y^k \in Y \cap V \quad (2.34)$$

donde λ y μ son los vectores multiplicadores de Lagrange de las restricciones del problema primal. Usando el teorema de dualidad fuerte $v(y)$ puede reformularse de la siguiente manera:

$$v(y) = \sup_{\mu \geq 0, \lambda} \inf_{x \in X} \{f(x, y) + \lambda h(x, y) + \mu g(x, y)\} \quad \forall y \in Y \cap V \quad (2.35)$$

Esta condición se puede relajar obteniendo una cota superior:

$$v(y) \geq \inf_{x \in X} \{f(x, y) + \lambda h(x, y) + \mu g(x, y)\} \quad \forall \mu \geq 0, \lambda \quad (2.36)$$

Asumiendo una solución factible para el problema interno, una formulación equivalente a (2.34) que surge usando el dual antes planteado es:

$$\begin{aligned} & \underset{y}{\text{Min}} v(y) \\ & \text{Suj. a: } v(y) \geq \underset{x \in X}{\text{Min}} \{f(x, y) + \lambda h(x, y) + \mu g(x, y)\} \\ & \text{Suj. a: } \begin{cases} y \in Y \cap V \\ V = \{y : h(x, y) = 0, g(x, y) \leq 0 \text{ para } x \in X\} \end{cases} \end{aligned} \quad (2.37)$$

Este problema es difícil de resolver por la representación implícita del dominio de la variable dada en las dos últimas restricciones. Esto se evita eliminando los dos últimos conjuntos de restricciones del problema trabajando luego con un *problema dual relajado* que se muestra a continuación. Cuando las restricciones del problema original (2.31) son convexas, se incorporan directamente al problema dual, evitando la relajación.

$$\begin{aligned} & \underset{y \in Y}{\text{Min}} \mu_B \\ & \text{Suj. a: } \mu_B \geq \underset{x \in X}{\text{Min}} L(x, y, \lambda^k, \mu^k) = \underset{x \in X}{\text{Min}} \{f(x, y) + \lambda^k h(x, y) + \mu^k g(x, y)\} \quad \forall \mu^k \geq 0, \forall \lambda^k \end{aligned}$$

λ^k, μ^k son los multiplicadores que se obtienen de resolver el problema primal para y^k y $L(x, y, \lambda^k, \mu^k) = f(x, y) + \lambda^k h(x, y) + \mu^k g(x, y)$

es la función de Lagrange de (2.34)

El problema de minimización interno se llama Problema Dual Relajado Interno.

Propiedades

La función de Lagrange es convexa en x para cada y y por ser una combinación lineal de funciones convexas en x para cada y , y convexa en y para cada x de igual forma.

Para las demostraciones de todos los resultados a continuación presentados (ver referencias⁽⁷⁾).

Propiedad 2.7.5.1:

$$\text{Min}_{x \in X} L(x, y^d, \lambda^k, \mu^k) \geq \text{Min}_{x \in X} L(x, y^d, \lambda^k, \mu^k) \Big|_{x^k}^{lin} \quad \forall y = y^d \geq 0$$

Donde el lado derecho es la linealización de la función de Lagrange alrededor del punto x^k y x^k la solución del subproblema primal k -ésimo.

Demostración:

Para cada $y = y^d$ la función de Lagrange es convexa en x , luego una linealización alrededor de algún x es una subestimación válida de la función para todo x y tomando el mínimo a ambos lados se obtiene el resultado \square

Observación: si la función de Lagrange es bilineal, ella coincide con su linealización y por lo tanto la desigualdad anterior se convierte en igualdad.

Definición 2.7.5.1: Sea $g_i^k(y) = \nabla_{x_i} L(x, y, \lambda^k, \mu^k) \Big|_{x^k}$ donde x_i es el i -ésimo elemento del vector x . La linealización de la función de Lagrange en x^k está dada por:

$$L(x, y^d, \lambda^k, \mu^k) \Big|_{x^k}^{lin} = L(x, y, \lambda^k, \mu^k) + \sum_{i=1}^n g_i^k(y) \cdot (x_i - x_i^k) \quad (2.38)$$

Las condiciones de optimalidad de Karush- Kuhn-Tucker para el problema primal requieren en parte que $g_i^k(y^k) = 0 \quad \forall i = 1, \dots, n$. Luego para y^k , para toda variable x_i para la cual $g_i^k(y^k)$ es una función de y se dice que es una variable conectada.

El conjunto de variables conectadas es $I_c^k = \{i : g_i^k(y) \text{ es una función de } y\}$.

Propiedad 2.7.5.2:

La linealización de la función de Lagrange depende del conjunto de variables y y el conjunto de variables conectadas solamente. La solución del problema dual relajado

interno_con la función de Lagrange remplazada por su linealización en x^k , depende sólo de aquellas x_i que cumplen que $g_i^k(y)$ es función de y (variables conectadas x_i).

Observación: El problema dual relajado interno se puede reemplazar por un problema que minimice la linealización de la función lagrangeana sobre las variables conectadas, reduciendo el esfuerzo de cálculo de óptimos globales en los problemas no convexos.

Propiedad 2.7.5.3:

Para todo y^k y para todo $i \in I_c^k$, sean x_i^L y x_i^U cotas inferior y superior de la variable x_i . Sean B_j el conjunto de cotas superior e inferior de las variables conectadas y CB el conjunto de todas sus combinaciones. Luego:

$$L^*(\bar{x}, y, \lambda^k, \mu^k) \geq \min_{B_j \in CB} \left\{ \begin{array}{l} \mu_B \geq L(x^{B_j}, y, \lambda^k, \mu^k) \Big|_{x^k}^{lin} \\ \text{con } g_i^k(y) \leq 0 \text{ si } x_i^{B_j} = x_i^U \\ g_i^k(y) \geq 0 \text{ si } x_i^{B_j} = x_i^L \end{array} \right\} \quad \forall y \quad (2.39)$$

Un nuevo conjunto de restricciones (las derivadas parciales de la función lagrangeana respecto a las variables conectadas) se introducen y se llaman calificadores de restricciones de la función de Lagrange $g_i^k(y) \leq 0$ o $g_i^k(y) \geq 0$.

Estos gradientes dividen el espacio Y y en cada región la función de Lagrange se subestima Como estos calificadores se introducen al problema dual relajado como restricciones, para garantizar la convexidad en dicho problema se requiere $g_i^k(y)$ sea lineal en y para todo i . Si no se tiene este caso, se debe linealizar la función lagrangeana, además, con respecto a y . Esta linealización es válida ya que se busca una cota inferior y la linealización es una subestimación de la función.

Los calificadores de restricciones lo que hacen es dividir el dominio en dos si y es un vector de una componente, en cuatro si y es de 2×1 . En cada subdivisión se va a subestimar la función $v(y)$ con linealizaciones de la función lagrangeana tomando una combinación de cotas apropiadas solo para el conjunto de variables conectadas (por propiedad 2.7.5.2). Estas subestimaciones se van mejorando con el refinamiento del dominio del vector y el cual se produce con el paso de las iteraciones del algoritmo.

Si ocurre que $g_i^k(y) = 0$, los calificadores se perturban con un número positivo muy pequeño de manera de no alterar la solución del problema.

Propiedad 2.7.5.4:

En la iteración k se define

i) $UL(k,K)$ el conjunto de funciones de Lagrange en la iteración k ($k < K$) cuyos calificadores se tienen en y^k , el valor fijo de las variables y para el problema primal en la iteración K.

ii) μ_B^K es el valor óptimo del K-ésimo problema dual relajado, es decir:

$$\mu_B^K = \left\{ \begin{array}{l} \text{Min}_{y, \mu_B} \mu_B \\ \text{Suj.a : } \mu_B \geq \text{Min}_{x \in X} L(x, y, \lambda^k, \mu^k) \quad k = 1, \dots, K-1 \\ \mu_B \geq \text{Min}_{x \in X} L(x, y, \lambda^K, \mu^K) \end{array} \right\}$$

luego

$$\mu_B^k \geq \min_{B_j \in CB} \left\{ \begin{array}{l} \text{Min}_{y, \mu_B} \mu_B \\ \mu_B \geq L(x^{B_j}, y, \lambda^k, \mu^k) \Big|_{x^k}^{lin} \\ \text{con } g_i^k(y) \leq 0 \text{ si } x_i^{B_j} = x_i^U \\ g_i^k(y) \geq 0 \text{ si } x_i^{B_j} = x_i^L \\ \mu_B \geq L(x^{B_i}, y, \lambda^K, \mu^K) \Big|_{x^k}^{lin} \\ \text{con } g_i^K(y) \leq 0 \text{ si } x_i^{B_i} = x_i^U \\ g_i^K(y) \geq 0 \text{ si } x_i^{B_i} = x_i^L \end{array} \right\} \quad (2.40)$$

Problemas primales infactibles

Puede ocurrir que para un y^k el subproblema primal no sea factible.

Situaciones donde se garantiza factibilidad:

i) si se comienza con un valor inicial y factible y en los subproblemas duales podamos incluir las restricciones, por ejemplo si son lineales o convexas, dejando de ser éste una relajación.

ii) Si el problema tiene solo una restricción conjunta en x e y y para cada valor de la variable x se convierte en restricción de cota para y .

En otro caso, si para un y^k el subproblema primal no tiene solución factible, para generar los multiplicadores de Lagrange apropiados, se puede plantear el siguiente problema primal relajado:

$$\begin{aligned} \underset{\substack{\alpha > 0 \\ x \in X}}{\text{Min}} \delta &= \sum_{i=1}^m \alpha_i & (2.41) \\ \text{Suj.a : } &g(x, y^k) - \alpha \leq 0 \\ &h(x, y^k) = 0 \end{aligned}$$

Es un problema con función objetivo lineal, restricciones de igualdad lineales y restricciones de desigualdad convexas y por el teorema de dualidad fuerte su valor objetivo óptimo coincide con el de su problema dual:

$$\bar{\delta} = \max_{\substack{\lambda_1^k \\ \mu_1^k \geq 0}} \min_{x \in X} \left\{ \delta + \lambda_1^{k^t} h(x, y^k) + \mu_1^{k^t} (g(x, y^k) - \alpha) \right\}$$

Como $\alpha \geq 0$, de las condiciones de Kuhn-Tucker surge la condición que $(1 - \mu_1^k)\alpha = 0$ lo que implica que

$$\bar{\delta} = \max_{\substack{\lambda_1^k \\ \mu_1^k \geq 0}} \min_{x \in X} \left\{ \lambda_1^{k^t} h(x, y^k) + \mu_1^{k^t} g(x, y^k) \right\}$$

Como se busca minimizar la infactibilidad, la idea es agregar al problema dual relajado una restricción que tenga como función cumplir con la condición:

$$\max_{\substack{\lambda_1^k \\ \mu_1^k \geq 0}} \min_{x \in X} \left\{ \lambda_1^{k^t} h(x, y) + \mu_1^{k^t} g(x, y) \right\} = 0$$

Luego es necesario resolver el problema primal relajado para el valor fijo $y = y^k$, y obtener los multiplicadores de Lagrange que surgen de este problema que se usan para armar restricciones definiendo la región factible de los subproblemas duales relajados futuros en términos de y , pero no para resolver subproblemas duales relajados de la iteración actual.

La restricción apropiada que se debe agregar al problema dual relajado para aquellas iteraciones donde el problema primal es infactible es:

$$\text{Min}_{x \in X} L(x, y, \lambda_1^k, \mu_1^k) \leq 0 \quad (2.42)$$

$$L(x, y, \lambda_1^k, \mu_1^k) = \lambda_1^k h(x, y) + \mu_1^k g(x, y)$$

Actúa como corte de factibilidad, eliminando la posibilidad de que todo problema dual relajado retorne al mismo valor y^k .

Selección de una cota inferior

Luego de resolver el problema dual relajado para toda combinación de cotas de las variables conectadas, se debe elegir una nueva cota inferior del problema original, y un valor nuevo de y para comenzar otra iteración. Esto se realiza tomando el mínimo de todos los valores almacenados μ_B como cota inferior y su correspondiente solución $y^{k+1} = y$. Una vez elegido se elimina del conjunto para no ser considerado nuevamente, garantizando que el problema dual relajado no vuelva el mismo valor de y en iteraciones sucesivas.

Si la diferencia entre cota inferior y superior está dentro de una tolerancia prefijada, el algoritmo termina.

Ideas del algoritmo

Para aplicar el algoritmo a un NLP se debe elegir una partición de variables x e y de tal forma que cumpla con la condición A y un punto de partida $y^{(1)}$.

El cálculo del subproblema primal en cada iteración da una cota superior a la solución. El signo de las derivadas parciales de la función lagrangeana definen un conjunto de ecuaciones llamado calificadores de restricciones que particionan el espacio y . En cada parte se resuelve un problema dual relajado cuyas restricciones son linealizaciones de la función lagrangeana. Estos problemas dependen de y y las variables conectadas que, en cada iteración se reemplazan por sus cotas superior o inferior, resultando así un problema con variables y . En cada iteración además, se van incorporando un conjunto de restricciones y corresponden a linealizaciones hechas en los subdominios. Estas incorporaciones permiten un mejor ajuste a la función $v(y)$ y por lo tanto una mejor cota inferior del valor objetivo óptimo. Hablamos de linealizaciones ya que si bien pueden ser hasta aproximaciones no convexas (por ser g_i cualquier tipo de función), se pide cuando sea necesario una linealización en y . Por esta exigencia y por la condición A inicialmente pedida, los problemas duales relajados son problemas lineales y cada subproblema primal es convexo.

La solución de los problemas duales relajados demanda la mayor parte del tiempo. Como el número de problemas a resolverse es $2^{NI_c^k}$, donde NI_c^k es el número de variables conectadas, se ha propuesto una enumeración implícita de los mismos mediante la solución de un problema MILP.

De cada problema dual si se obtiene un valor μ_B que es inferior a la mejor cota superior obtenida del problema primal, luego se agrega al conjunto de cotas inferiores candidatas; en caso contrario no se considera más esa región para posteriores subdivisiones.

Cuando la función de Lagrange es bilineal como en el ejemplo, luego la linealización con respecto a x es la misma función.

Convergencia ε -finita y optimalidad ε -global

Lema 2.7.5.1:

Si la sucesión de soluciones $\langle y^k \rangle$ del problema dual relajado converge a \bar{y} , los correspondientes multiplicadores de Lagrange para el problema primal (2.34) convergen a $(\bar{\lambda}, \bar{\mu})$ luego $L(x^{B_j}, \bar{y}, \bar{\lambda}, \bar{\mu}) \Big|_{x^k}^{lin} = P(\bar{y}) \quad \forall x^{B_j}$

Donde $P(\bar{y})$ representa la solución óptima del problema primal en \bar{y} .

Lema 2.7.5.2:

Para un $y = y^k$ fijo del problema primal k -ésimo, sea $U(y^k)$ el conjunto de multiplicadores óptimos. Si $y^k \rightarrow \bar{y}$ y $\lambda^k \rightarrow \bar{\lambda}$, $\mu^k \rightarrow \bar{\mu}$ luego $(\bar{\lambda}, \bar{\mu}) \in U(\bar{y})$.

Lema 2.7.5.3:

Sea $P(y)$ el conjunto de soluciones óptimas del problema primal para valores fijos de y , luego si $y^k \rightarrow \bar{y}$ $P(y)$ es semicontinua inferior en \bar{y} .

Teorema 2.7.5.1: (Convergencia ε -finita)

Si se tienen las siguientes condiciones:

a) X e Y son conjuntos convexos compactos, e $Y \subset V$

$$V = \{y : h(x, y) = 0, g(x, y) \leq 0 \quad \text{para } x \in X\}$$

b) Condición A

c) $f(x, y), g(x, y)$ y $h(x, y)$ son continuas en $X \times Y$

d) El conjunto $U(y)$ de multiplicadores para el problema primal es no vacío $\forall y \in Y$ y uniformemente acotado en un entorno de todos esos puntos luego para todo $\varepsilon > 0$ dado, el algoritmo GOP termina en un número finito de pasos.

Observación: Si Y no es subconjunto de V se logra solo una convergencia asintótica.

Teorema 2.7.5.2: (Optimalidad global)

Si las condiciones del teorema 2.7.5.1 se tienen luego

- i) La solución del problema dual relajado en el paso 3 del algoritmo GOP siempre será una subestimación de la solución del problema (2.31).
- ii) El algoritmo GOP termina en el óptimo global de (2.31).

Propiedades adicionales

Al implementar el algoritmo GOP se puede ver que se trabaja con una gran cantidad de problemas duales relajados con igual solución o infactibles y que pueden ser evitados mirando la estructura de la función lagrangeana (Visweswaran and Floudas, 1993). Se introduce para tal fin una condición A':

En cada iteración la función de Lagrange que se usa en el problema dual relajado es lineal en x para y fijo y lineal en y para x fijo.

La condición A' garantiza que

$$L(x, y, \lambda^k, \mu^k) \Big|_{x^k}^{lin} = L(x, y, \lambda^k, \mu^k) \quad y$$

$$g_i^k(y) = \nabla_{x_i} L(x, y^d, \lambda^k, \mu^k) \Big|_{x^k}^{lin} = \nabla_{x_i} L(x, y, \lambda^k, \mu^k) \quad \forall i$$

Por lo tanto $g_i^k(y)$ son funciones lineales en y , por lo que es fácil determinar cotas inferiores y superiores para esta función sobre Y , garantizando el siguiente resultado:

Propiedad 2.7.5.5:

En cada iteración del algoritmo GOP, si $g_i^k(y) \geq 0$ para todo y (respectivamente $g_i^k(y) \leq 0$ para todo y) no es necesario resolver aquellos subproblemas duales relajados que tienen $g_i^k(y) \leq 0$ como un calificador de restricción (respectivamente $g_i^k(y) \geq 0$).

Propiedad 2.7.5.6:

En cada iteración del algoritmo GOP, si $g_i^k(y) = 0$ para todo i y luego es suficiente resolver sólo aquellas combinaciones de cotas donde x_j están en su cota inferior o superior.

Proceso de ramificación y acotamiento

Se ramifica siguiendo la partición que se produce en el espacio Y a partir de los calificadores de restricciones por iteración. Se comienza con el espacio completo Y del problema original. Sean I_c^1 el conjunto de variables conectadas en la primer iteración, NI_c^k el número de variables conectadas, la región original se subdivide en $2^{NI_c^1}$ subregiones.

Cada subregión implica la resolución de un subproblema dual relajado y corresponde a un nodo. Si un subproblema tiene un valor de μ_B mayor que la mejor cota superior, se trunca eliminándolo de posteriores refinamientos, si es menor se agrega al conjunto de cotas inferiores candidatas

Resumen del algoritmo GOP

Etapa inicial. Inicialización de parámetros

Definir $\mu_B^{stor}(K^{\max}, B_j), y^{stor}(K^{\max}, B_j), y^k(K^{\max}, B_j)$ sobre el conjunto de cotas CB (conjunto de todas las combinaciones entre x^L y x^U para las variables conectadas)

Sea el máximo número esperado de iteraciones K^{\max} . P^{UBD} y R^{LBD} conjuntos que almacenan valores objetivos y corresponden a cotas superior e inferior obtenidas de los problemas primal y dual relajado respectivamente.

$$\mu_B^{stor}(K^{\max}, B_j) = U, \quad P^{UBD} = U, \quad R^{LBD} = L$$

donde U es un número positivo muy grande y L un número negativo muy chico.

Seleccionar un conjunto inicial de valores y^1 y una tolerancia de convergencia ϵ .

Sean K^{feas}, K^{infeas} conjuntos vacíos y tomar $K=1$.

Proceso iterativo

Paso 1 (problema primal)

Almacenar el valor de y^K .

Resolver el problema primal

$$\underset{x}{\text{Min}} f(x, y^k)$$

$$\text{Suj.a: } g(x, y^k) \leq 0$$

$$h(x, y^k) = 0$$

Si este problema es factible, actualizar el conjunto K^{feas} y almacenar los multiplicadores de Lagrange asociados. Actualizar la cota superior usando la expresión $P^{UBD} = \min(P^{UBD}, P^K(y^K))$ donde $P^K(y^K)$ es la solución del problema primal actual.

Si el problema primal es infactible, actualizar el conjunto K^{infeas} y resolver el siguiente problema primal relajado para obtener los multiplicadores λ_1^K y μ_1^K

$$\underset{\substack{\alpha > 0 \\ x \in X}}{\text{Min}} \delta = \sum_{i=1}^m \alpha_i$$

$$\text{Suj.a: } g(x, y^k) - \alpha \leq 0$$

$$h(x, y^k) = 0$$

(la infactibilidad se da en las restricciones no lineales)

Paso 2 (Selección de funciones de Lagrange de iteraciones previas)

Para $k=1, \dots, K-1$ evaluar todos los calificadores de restricciones de toda función de Lagrange (es decir correspondiente a cada conjunto de cotas de x) en y^K . Si todo calificador de restricciones de la función de Lagrange se satisface en y^K , seleccionar y colocar la función de Lagrange en el conjunto $UL(k, K)$, es decir a estar en los problemas duales relajados actuales con sus calificadores de restricciones.

Paso 3 (Problema dual relajado)

Formular la función de Lagrange correspondiente al problema primal actual. Adicionar esto como una restricción al problema dual relajado. Luego:

- Determinar el conjunto de variables conectadas x , I_c^K
- Seleccionar una combinación de cotas de las variables conectadas en x , $B_l = B_1 \in CB$
- Resolver el siguiente subproblema dual relajado:

$$\left. \begin{array}{l} \text{Min } \mu_B \\ y \in Y, \mu_B \\ \text{Suj.a : } \mu_B \geq L(x^{B_j}, y, \lambda^k, \mu^k) \Big|_{x^k}^{lin} \\ g_i^k(y) \leq 0 \text{ si } x_i^{B_j} = x_i^U \\ g_i^k(y) \geq 0 \text{ si } x_i^{B_j} = x_i^L \end{array} \right\} \forall j \in UL(k, K), k = 1, 2, \dots, K-1, k \in K^{feas}$$

$$\left. \begin{array}{l} 0 \geq L_1(x^{B_j}, y, \lambda_1^k, \mu_1^k) \Big|_{x^k}^{lin} \\ g_i^k(y) \leq 0 \text{ si } x_i^{B_j} = x_i^U \\ g_i^k(y) \geq 0 \text{ si } x_i^{B_j} = x_i^L \end{array} \right\} \forall j \in UL(k, K), k = 1, 2, \dots, K-1, k \in K^{infeas}$$

$$\left. \begin{array}{l} \mu_B \geq L(x^{B_j}, y, \lambda^k, \mu^k) \Big|_{x^k}^{lin} \\ g_i^k(y) \leq 0 \text{ si } x_i^{B_j} = x_i^U \\ g_i^k(y) \geq 0 \text{ si } x_i^{B_j} = x_i^L \end{array} \right\} \text{Si } K \in K^{feas}$$

$$\left. \begin{array}{l} 0 \geq L_1(x^{B_j}, y, \lambda_1^k, \mu_1^k) \Big|_{x^k}^{lin} \\ g_i^k(y) \leq 0 \text{ si } x_i^{B_j} = x_i^U \\ g_i^k(y) \geq 0 \text{ si } x_i^{B_j} = x_i^L \end{array} \right\} \text{Si } K \in K^{infeas}$$

donde $L_1(x, y, \lambda_1^k, \mu_1^k) = \lambda_1^{k^t} h(x, y) + \mu_1^{k^t} g(x, y)$

Almacenar la solución si es factible, en $\mu_B^{stor}(K, B_1)$ e $y^{stor}(K, B_1)$

d) Seleccionar una nueva combinación de cotas para x , es decir $B_1 = B_2$.

e) Repetir (b)-(d) hasta que el problema dual relajado sea resuelto en cada conjunto de cotas de las variables conectadas en x , esto es para todo $B_1 \in CB$.

Paso 4 (Selección de una nueva cota inferior y y^{K+1})

Del conjunto μ_B^{stor} , seleccionar el mínimo μ_B^{\min} (incluyendo las soluciones de la iteración actual). Seleccionar el valor almacenado actual $y^{stor}(k, B_j)$ como y^{\min} . Sean

$R^{LBD} = \mu_B^{\min}$ y $y^{K+1} = y^{\min}$. Borrar μ_B^{\min} y y^{\min} del conjunto almacenado.

Paso 5 (Verificación de convergencia)

Chequear si $R^{LBD} > P^{UBD} - \varepsilon$, parar. Si no $K=K+1$ y volver al paso 1.

Ejemplo 2.7.5: Problema de “pooling” o fusión de Haverly y Lasdon (1979)

Haverly definió un problema de fusión con tres sustancias A , B y C , (ver Figura 2.7.5.1), con diferentes contenidos de sulfuro que son combinados para formar dos productos, x e y con diferentes contenidos de sulfuro máximos. Lasdon y otros (1979) agregó la restricción de calidad en la reformulación. El modelo resulta:

$$\text{Min } 6A + 16B + 10(C_x + C_y) - 9x - 15y$$

$$\text{Suj. a: } P_x + P_y - A - B = 0 \quad \text{Ecuación de balance}$$

$$\left. \begin{array}{l} x - P_x - C_x = 0 \\ y - P_y - C_y = 0 \end{array} \right\} \quad \text{Balance de componente}$$

$$p(P_x + P_y) - 3A - B = 0 \quad \text{Calidad de pool}$$

$$\left. \begin{array}{l} pP_x + 2C_x - 2.5x \leq 0 \\ pP_y + 2C_y - 1.5y \leq 0 \end{array} \right\} \quad \text{Restricciones de calidad de producto}$$

$$0 \leq x \leq x^U \quad \text{Cotas superiores de producto}$$

$$0 \leq y \leq y^U$$

$$P_x, P_y, C_x, C_y \geq 0$$

$$1 \leq p \leq 3$$

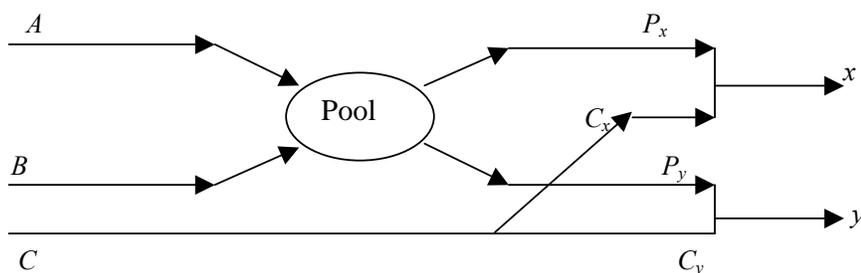


Figura 2.7.5.1: Problema de fusión de Haverly

Por la naturaleza del problema todas las variables son no negativas, $1 \leq p \leq 3$ y el problema se resuelve tomando $x^U = 100$ e $y^U = 200$. El modelo tiene tres restricciones de calidad, una de igualdad y dos de desigualdad, con términos bilineales; el resto, incluyendo la función objetivo son lineales.

Resolución usando GOP

La partición de variables se define de la siguiente forma: $y = p$ y las variables restantes forman el vector x .

Se debe fijar primero y , se toma $p^{(1)} = 3$

1º iteración

Problema primal

$$\text{Min } Z = 6A + 16B + 10(C_x + C_y) - 9x - 15y$$

$$\text{Suj. a: } P_x + P_y - A - B = 0$$

$$x - P_x - C_x = 0$$

$$y - P_y - C_y = 0$$

$$3(P_x + P_y) - 3A - B = 0$$

$$3P_x + 2C_x - 2.5x \leq 0$$

$$3P_y + 2C_y - 1.5y = 0$$

$$0 \leq x \leq 100$$

$$0 \leq y \leq 200$$

$$A, B, P_x, P_y, C_x, C_y \geq 0$$

Solución óptima del problema: $A^* = 50$, $B^* = 0$, $C_x^* = 50$, $C_y^* = 0$, $x^* = 100$, $y^* = 0$,
 $P_x^* = 50$, $P_y^* = 0$ $Z^* = -100$

Multiplicadores de Lagrange asociados: $\lambda^{(1)} : \lambda_1 = 21$, $\lambda_2 = 18$, $\lambda_3 = 30$, $\lambda_4 = -5$

$$\mu^{(1)} : \mu_1 = 4, \mu_2 = 10, \mu_3 = 1, \mu_4 = 0$$

Función de Lagrange

$$L(x, y, \lambda^{(1)}, \mu^{(1)}) = 6A + 16B + 10(C_x + C_y) - 9x - 15y + 21(P_x + P_y - A - B) + 18(x - P_x - C_x) + 30(y - P_y - C_y) - 5[p(P_x + P_y) - 3A - B] + 4(pP_x + 2C_x - 2.5x) + 10(pP_y + 2C_y - 1.5y) + x - 100$$

Variables conectadas: P_x , P_y

Calificadores de restricciones

$$\frac{\partial L}{\partial P_x} = 21 - 18 - 5p + 4p \geq 0 \text{ si } p \leq 3$$

$$\frac{\partial L}{\partial P_x} \leq 0 \text{ si } p \geq 3$$

$$\frac{\partial L}{\partial P_y} = 21 - 30 - 5p + 10p \geq 0 \text{ si } p \geq 9/5$$

$$\frac{\partial L}{\partial P_y} \leq 0 \text{ si } p \leq 9/5$$

3) Problemas duales relajados

i) Si

$$P_x^L = 0, P_y^L = 0 \quad \left(\frac{\partial L}{\partial P_x} \geq 0, \frac{\partial L}{\partial P_y} \geq 0 \right)$$

$$p \leq 3 \quad y \quad p \geq 9/5$$

min μ_B

$$Suj.a : \mu_B \geq L(x^{(1)}, y, \lambda^{(1)}, \mu^{(1)}) \Big|_{p=3} = L_1^{(1)} = 50 - 50p - (3-p)50 = -100$$

$$9/5 \leq p \leq 3$$

Solución óptima: $\mu_B^* = -100$, $9/5 \leq p^* \leq 3$

ii) Si

$$P_x^L = 0, P_y^U = 200 \quad \left(\frac{\partial L}{\partial P_x} \geq 0, \frac{\partial L}{\partial P_y} \leq 0 \right)$$

$$p \leq 3 \quad y \quad p \leq 9/5$$

min μ_B

$$Suj.a : \mu_B \geq L(x^{(1)}, y, \lambda^{(1)}, \mu^{(1)}) \Big|_{p=3} = L_2^{(1)} = 50 - 50p - (3-p)50 + (5p-9)200 = -1900 + 1000p$$

$$1 \leq p \leq 9/5$$

Solución óptima: $\mu_B^* = -900$, $p^* = 1$

iii) Si

$$P_x^U = 100, P_y^L = 0 \quad \left(\frac{\partial L}{\partial P_x} \leq 0, \frac{\partial L}{\partial P_y} \geq 0 \right)$$

$$p \geq 3 \quad y \quad p \geq 9/5$$

min μ_B

$$Suj.a : \mu_B \geq L(x^{(1)}, y, \lambda^{(1)}, \mu^{(1)}) \Big|_{p=3} = L_3^{(1)} = 50 - 50p + (3-p)(100 - 50) = 200 - 100p$$

$$3 \leq p \leq 3$$

Solución óptima: $\mu_B^* = -100$, $p^* = 3$

iv) Si

$$P_x^U = 100, P_y^U = 200 \quad \left(\frac{\partial L}{\partial P_x} \leq 0, \frac{\partial L}{\partial P_y} \leq 0 \right) \quad \text{problema no factible.}$$

$$p \geq 3 \quad y \quad p \leq 9/5$$

4) Cota inferior: $p(2) = p_{\min} = 1$, $\mu_B^{\min} = -900$ $-900 \leq Z^* \leq -100$

μ_{stor}	-100	-100
y_{stor}	[9/5,3]	3

2º iteración: $p(2) = 1$

1) Problema primal

$$\text{Min } Z = 6A + 16B + 10(C_x + C_y) - 9x - 15y$$

$$\text{Suj. a: } P_x + P_y - A - B = 0$$

$$x - P_x - C_x = 0$$

$$y - P_y - C_y = 0$$

$$P_x + P_y - 3A - B = 0$$

$$P_x + 2C_x - 2.5x \leq 0$$

$$P_y + 2C_y - 1.5y = 0$$

$$0 \leq x \leq 100$$

$$0 \leq y \leq 200$$

$$A, B, P_x, P_y, C_x, C_y \geq 0$$

Solución óptima: $A^* = 0$, $B^* = 100$, $C_x^* = 0$, $C_y^* = 100$, $x^* = 0$, $y^* = 200$, $P_x^* = 0$

$P_y^* = 100$ $Z^* = -400$. Nueva cota superior

Multiplicadores de Lagrange asociados: $\lambda^{(2)} : \lambda_1 = 21$, $\lambda_2 = 10$, $\lambda_3 = 22$, $\lambda_4 = -5$

$\mu^{(2)} : \mu_1 = 0$, $\mu_2 = 6$, $\mu_3 = 0$, $\mu_4 = 2$

Función de Lagrange

$$L(x, y, \lambda^{(2)}, \mu^{(2)}) = 6A + 16B + 10(C_x + C_y) - 9x - 15y + 21(P_x + P_y - A - B) + 10(x - P_x - C_x) + 22(y - P_y - C_y) - 5[p(P_x + P_y) - 3A - B] + 6(pP_y + 2C_y - 1.5y) + 2(y - 200)$$

Variables conectadas: P_x , P_y

Calificadores de restricciones

$$\frac{\partial L}{\partial P_x} = 21 - 10 - 5p \geq 0 \text{ si } p \leq 11/5$$

$$\frac{\partial L}{\partial P_x} \leq 0 \text{ si } p \geq 11/5$$

$$\frac{\partial L}{\partial P_x} = 21 - 22 - 5p + 6p \geq 0 \quad \text{si } p \geq 1$$

$$\frac{\partial L}{\partial P_y} \leq 0 \quad \text{si } p \leq 1$$

2) $p^{(2)} = 1$ cumple con los calificadores de restricciones asociados a
 $L_2^{(1)} = 1000p - 1900 \quad 1 \leq p \leq 9/5$

3) Problemas duales relajados

i) Si

$$P_x^L = 0, P_y^L = 0 \quad \left(\frac{\partial L}{\partial P_x} \geq 0, \frac{\partial L}{\partial P_y} \geq 0 \right)$$

$$p \leq 11/5 \quad \text{y} \quad p \geq 1$$

min μ_B

$$\text{Suj.a: } \mu_B \geq L_2^{(1)} = 1000p - 1900$$

$$1 \leq p \leq 9/5$$

$$\mu_B \geq L(x^{(2)}, y, \lambda^{(2)}, \mu^{(2)}) \Big|_{p=2} = L_1^{(2)} = 100p - 500 + (p-1)(-100) = -400$$

$$1 \leq p \leq 11/5$$

Solución óptima: $\mu_B^* = -400, \quad p^* = 1$

ii) Si

$$P_x^L = 0, P_y^U = 200 \quad \left(\frac{\partial L}{\partial P_x} \geq 0, \frac{\partial L}{\partial P_y} \leq 0 \right)$$

$$p \leq 11/5 \quad \text{y} \quad p \leq 1$$

min μ_B

$$\text{Suj.a: } \mu_B \geq L_2^{(1)} = 1000p - 1900$$

$$1 \leq p \leq 9/5$$

$$\mu_B \geq L(x^{(2)}, y, \lambda^{(2)}, \mu^{(2)}) \Big|_{p=2} = L_2^{(2)} = 100p - 500 + (p-1)100 = 200p - 600$$

$$1 \leq p \leq 1$$

Solución óptima: $\mu_B^* = -400, \quad p^* = 1$

iii) Si

$$P_x^U = 100, P_y^L = 0 \quad \left(\frac{\partial L}{\partial P_x} \leq 0, \frac{\partial L}{\partial P_y} \geq 0 \right)$$

$$p \geq 11/5 \quad \text{y} \quad p \geq 1$$

Pero considerando los calificadores de restricciones de la iteración anterior: $1 \leq p \leq 9/5$ da un problema no factible.

iv) Si $P_x^U = 100, P_y^U = 200$ $(\frac{\partial L}{\partial P_x} \leq 0, \frac{\partial L}{\partial P_y} \leq 0)$ problema no factible.

4)_Cota inferior: observando μ_{stor}, y_{stor} y las soluciones de los problemas duales de la iteración actual.

$$Z^* = -400, p = 1 \quad -400 \leq Z^* \leq -400$$

Por lo que se ha llegado a la solución óptima.

Solución óptima: $A^* = 0, B^* = 100, C_x^* = 0, C_y^* = 100, x^* = 0, y^* = 200, P_x^* = 0$
 $P_y^* = 100, p^* = 1, Z^* = -400$

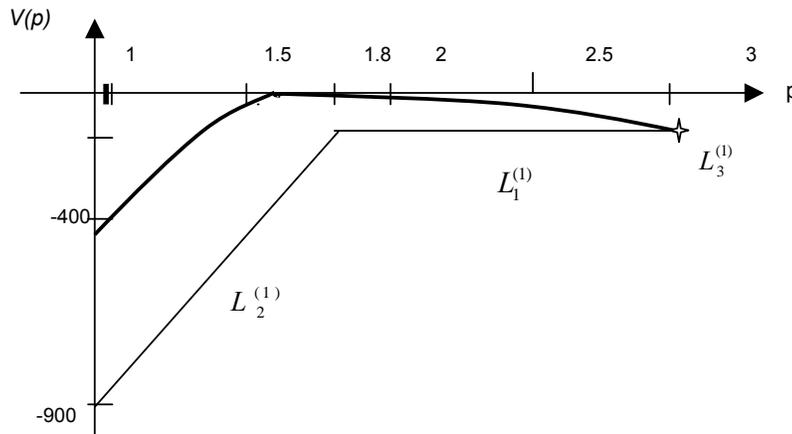


Figura 2.7.5.2: Soluciones del problema para distintos valores de p y aproximaciones lineales de la iteración 1

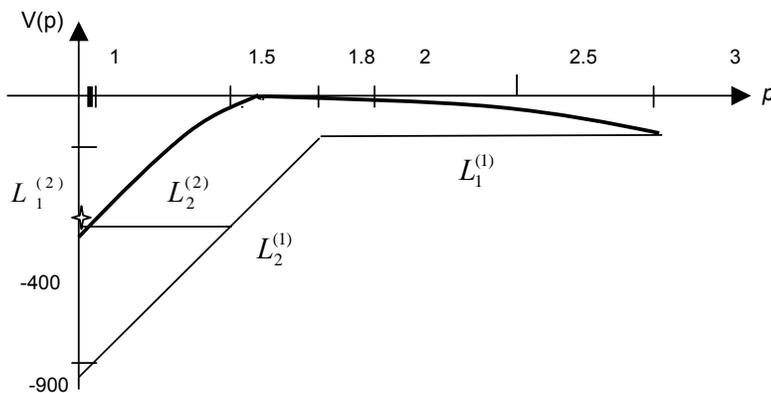


Figura 2.7.5.3: Soluciones del problema para distintos valores de p y aproximaciones lineales de la iteración 2

Conclusiones del algoritmo

GOP es un algoritmo para resolver problemas no convexos que requiere una forma general del modelo menos restrictiva a la que proponen otros algoritmos. La condición A permite trabajar con producto de funciones convexas tales como las funciones bilineales por ejemplo. Esto implica resolver los mismos problemas no convexos resueltos por otros algoritmos pero con un modelo original afectado a menos transformaciones y con lo que ello implica en cuanto a menos agregados de nuevas variables, restricciones y/o subestimaciones.

El signo de las funciones $g_i^k(y)$ definen un conjunto de ecuaciones llamado calificador de restricciones de la función lagrangeana en la iteración k, la cual particiona el espacio y en $2^{NI_c^k}$ subregiones. En cada una de estas subregiones, la función lagrangeana se construye y tal que subestima la solución global en la subregión; además al buscar el mínimo provee una cota inferior de la solución global en tal región. Luego de este proceso surge que el número de problemas duales relajados a resolver por iteración es $2^{NI_c^k}$ donde NI_c^k es el número de variables conectadas en la k-ésima iteración. Si bien los autores presentan resultados que permiten eliminar algunos problemas duales disminuyendo el trabajo del algoritmo, la solución de los problemas duales relajados demandan la mayor parte del tiempo.

2.8 Cuadro resumen de algunos aspectos de los algoritmos comentados

Autores	Nombre Implementación/ lenguaje	Tipo de problema que resuelve	Equipo	Criterio de parada. Precisión ϵ	Software/Algoritmo que usa	Usa paso de reducción
Ryoo-Sahinidis	BARON	MINLP	Sun SPARC station 2	10^{-6}	GMAS MINOS 5.4	Si
Byrne-Bogle	IGOR, RIGOR	MINLP	IBM RS6000	desconocida	Matlab Optimization toolbox 4.2c	RIGOR si
Smith-Pantelides	Lenguaje gPROMS	MINLP	Sun SPARC Cstation 10/51	Variable según el ejemplo	GAMS CONOP	Si
Maranas, Androulakis, Adjiman-Floudas	α BB	NLP	HP9000/730	Generalmente 0.001	MINOS 5.4	si

2.9 Conclusiones

Es difícil realizar comparaciones de los resultados obtenidos al resolver los mismos problemas test con diferentes propuestas de los autores debido a que las condiciones en cada trabajo en cuanto a equipo y precisión por ejemplo, son diferentes o desconocidas en algunos casos. Factores como el uso de escalamientos en el modelo y precisiones requeridas en procesos de cálculo internos de los algoritmos empleados, influyen sustancialmente en la ejecución de los mismos, elementos que los distintos referentes no detallan en sus trabajos.

Se ha comprobado que los procesos de reducción de intervalos son un elemento clave para acelerar la convergencia de los algoritmos. El trabajo de Ryoo y Sahinidis aporta test de reducción de rango y región factible distintos a las usadas hasta su publicación. Estas estrategias fueron probadas en la primer reformulación del algoritmo presentado en esta tesis.

En el trabajo de Smith y Pantelides se presenta una reformulación del problema mediante una descomposición de las funciones que lo forman, que posteriormente se adopta en el algoritmo desarrollado por Marcovecchio, Bergamini y Aguirre⁽¹¹⁾ y que será explicado en el capítulo 3.

En algoritmos de tipo ramificar y acotar se pueden extraer algunos criterios comunes que resultan ventajosos para un mejor funcionamiento de las alternativas propuestas. Los algoritmos expuestos en este capítulo usan diferentes alternativas de reformulación para un problema no convexo. En busca de problemas relajados convexos y a veces lineales, se utilizan y combinan diferentes estimaciones estudiadas en su gran parte en la sección 2.5.

En todos los trabajos se obtienen mejores resultados, en término de aceleración a la convergencia de un algoritmo, mediante la incorporación de un paso de reducción de cotas para las variables; es un factor determinante para lograr convergencia dentro de un límite preestablecido a no lograrla; luego se considera indispensable adoptar procesos de reducción de la región factible apropiados a un proceso iterativo determinado. Otro elemento que favorece la evolución de un algoritmo de este tipo es el agregado de subestimadores como los propuesto por McCormick, explícitamente enunciado por Ryoo y Sahinidis; y claro está que cuando se trabajan con subestimaciones más ajustados a la función en estudio, menores esfuerzos se requerirán para llegar a la solución del problema.

Capítulo 3

Método de optimización global: “Improve and Branch”

3.1 Introducción

En el presente capítulo se analiza en detalle el algoritmo “Improve and Branch” presentado por Marcovecchio, Bergamini y Aguirre⁽¹¹⁾ al que en el siguiente capítulo se proponen y realizan algunas variantes.

El algoritmo es de ramificación y mejora del objetivo y permite resolver problemas de programación no lineal no convexos.

En el desarrollo del algoritmo se resuelven dos problemas no convexos, un problema reformulado RP y un problema principal MP. RP es una reformulación equivalente del problema original y MP es una subestimación no convexa cuyo objetivo es encontrar, si existen, soluciones con valor objetivo mejor que las calculadas hasta el momento y factibles para el problema original. A partir de la solución de este problema, el problema RP busca mejores soluciones que serán cotas superiores del problema. De no encontrar una solución factible para RP, se implementan criterios para realizar ramificaciones. Se formulan distintas reglas de ramificación que particionan la región factible, generando subregiones que proporcionan subestimaciones de las funciones cada vez más precisas.

Para acelerar la convergencia, se propone una técnica de reducción de cotas implementada para cada ejemplo particular.

3.2 Formulación del modelo

La notación general de un NLP tomada en este trabajo es:

$$\begin{aligned} & \text{Min } f(\bar{x}) \\ & \text{suj.a: } \bar{h}(\bar{x}) = \bar{0} \\ & \bar{g}(\bar{x}) \leq \bar{0} \\ & \bar{x}^{\text{LO}} \leq \bar{x} \leq \bar{x}^{\text{UP}} \end{aligned} \quad \mathbf{P}$$

donde $\bar{x} \in \mathfrak{R}^n$ es un vector de variables de decisión continuas acotadas, el conjunto de restricciones está formado por las funciones vectoriales $\bar{g} : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ que son desigualdades, $\bar{h} : \mathfrak{R}^n \rightarrow \mathfrak{R}^q$ igualdades y $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ es la función objetivo a minimizar.

3.3 Problema Reformulado: RP

Para reformular el problema no convexo, se adoptó el trabajo algebraico propuesto por Smith y Pantelides⁽²²⁾, el cual fue descrito detalladamente en la sección 2.7.3. Como ya se explicó, se descomponen las funciones en términos algebraicos simples. Se sustituyen los términos no convexos por una nueva variable la que será trabajada algebraicamente si es necesario, o subestimada por una función convexa. Así con el agregado de variables y restricciones, las funciones no lineales se transforman en funciones simples y funciones univariadas trascendentes (logaritmos, exponenciales, trigonométricas). Luego todo problema que tiene expresiones solo aritméticas se puede transformar en un problema equivalente donde la no linealidad corresponde a términos bilineales, fracciones simples y funciones univariadas, donde la relajación convexa del problema reformulado también es una relajación válida para el problema original.

Una vez realizada la desagregación de las expresiones del modelo, se obtienen términos simples que serán tratados según su forma:

Para los términos fraccionales, como están definidos en intervalos para x_j enteramente

positivos o negativos, se realiza un cambio de variable $z_{ij} = \frac{x_i}{x_j}$ que reemplaza a cada

ocurrencia del término fraccional, y esta definición se agrega al modelo como una restricción bilineal $x_j z_{ij} = x_i$. Se calculan cotas de la nueva variable a partir de las cotas de x_j y x_i .

Para los términos bilineales se requieren tres nuevas variables z, α y β definidas de la forma: $\alpha = x_1 + x_2$

$$\beta = x_1 - x_2$$

$$\text{luego } x_1 x_2 = \frac{\alpha^2 - \beta^2}{4} = z$$

En el modelo, cada término bilineal se reemplaza por z y se le deben agregar las restricciones: $\alpha = x_1 + x_2$

$$\beta = x_1 - x_2$$

$$-4z + \alpha^2 - \beta^2 \leq 0$$

$$4z + \beta^2 - \alpha^2 \leq 0$$

Además se deben agregar restricciones de cotas para las nuevas variables que se derivan de las variables originales. En estas desigualdades hay dos términos cóncavos: $-\beta^2$ y $-\alpha^2$ que se respetarán inicialmente en la reformulación del modelo.

Completa la reformulación el nuevo problema es equivalente al original y tiene tres tipos de restricciones: de igualdad lineal, restricciones de desigualdad convexas y restricciones de desigualdad no convexas por tener términos univariados cóncavos.

Ejemplo 3.3.1: El siguiente ejemplo muestra una reformulación de un problema no lineal:

$$\begin{aligned} \text{Min } & x_1^2 - x_2^2 + 4x_3^4 - \sqrt{x_1} \\ \text{suj.a: } & 3x_1 + x_2 e^{x_3} \leq 20 \end{aligned}$$

Para este ejemplo, tanto la función objetivo como la primer restricción son funciones no convexas. Por tal motivo el problema debe reformularse con el agregado de nuevas variables como sigue:

$$\begin{aligned} \text{Min } & x_1^2 - z_1 + 4x_3^4 - \sqrt{x_1} \\ \text{suj. a: } & \text{eq1 } 3x_1 + z_2 \leq 20 \\ & \left. \begin{array}{l} \text{eq2 } x_2^2 \leq z_1 \\ \text{eq3 } -x_2^2 \leq -z_1 \end{array} \right\} z_1 = x_2^2 \\ & \left. \begin{array}{l} \text{eq4 } \alpha = x_2 + z_3 \\ \text{eq5 } \beta = x_2 - z_3 \\ \text{eq6 } -\beta^2 \leq 4z_2 - \alpha^2 \\ \text{eq7 } -\alpha^2 \leq -4z_2 - \beta^2 \end{array} \right\} z_2 = x_2 z_3 \\ & \left. \begin{array}{l} \text{eq8 } e^{x_3} \leq z_3 \\ \text{eq9 } -e^{x_3} \leq -z_3 \end{array} \right\} z_3 = e^{x_3} \end{aligned}$$

Se utilizan sustituciones válidas que llevan a una reformulación equivalente del problema original. El modelo tiene restricciones de igualdad lineales (eq4, eq5), restricciones de desigualdad convexas (eq1, eq2, eq8) y restricciones de desigualdad no convexas (eq3(x_2), eq6(β), eq7(α), eq9(x_3))

Si el problema P no está formado por funciones tales como las trigonométricas, la forma general del problema reformulado RP es:

$$\begin{aligned}
 & \text{Min } f^c(\bar{x}, \bar{z}) \\
 & \text{suj.a. } \bar{h}^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) = \bar{0} \\
 & \quad \bar{g}^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \leq \bar{0} \\
 & \quad \bar{t}cu + \bar{k}^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \leq \bar{0} \\
 & \quad \bar{x}^{LO} \leq \bar{x} \leq \bar{x}^{UP} \quad \bar{z}^{LO} \leq \bar{z} \leq \bar{z}^{UP} \\
 & \quad \bar{\alpha}^{LO} \leq \bar{\alpha} \leq \bar{\alpha}^{UP} \quad \bar{\beta}^{LO} \leq \bar{\beta} \leq \bar{\beta}^{UP}
 \end{aligned}
 \tag{RP}$$

donde \bar{h}^c son funciones lineales, f^c , \bar{g}^c y \bar{k}^c son funciones convexas, los términos cóncavos univariados están simbolizados tcu y las variables agregadas z están definidas a partir de x , $z = z(x)$. Si la reformulación se hace cuidadosamente, P y RP resultan completamente equivalentes, en el sentido que hay una correspondencia punto a punto entre sus regiones factibles, y la solución óptima de RP proporciona la solución óptima de P.

3.4 Relajación convexa

Es posible obtener una formulación convexa a partir de RP, subestimando los términos cóncavos $tcu(w)$, donde w es alguna de las variables de los vectores $(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta})$. La subestimación más apropiada es la envolvente convexa del término cóncavo, y es un segmento de recta definido por los puntos $(w^{LO}, tcu(w^{LO}))$ y $(w^{UP}, tcu(w^{UP}))$, donde w^{LO} y w^{UP} son las cotas de w (Ver capítulo 2).

Resulta:

$$\begin{aligned}
 & \lambda tcu(w^{LO}) + (1 - \lambda)tcu(w^{UP}) \leq -k^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \\
 & w = \lambda w^{LO} + (1 - \lambda)w^{UP} \\
 & 0 \leq \lambda \leq 1 \\
 & w^{LO} \leq w \leq w^{UP}
 \end{aligned}
 \tag{R'}$$

Observaciones: 1) si se reemplazan las cotas w^{LO} y w^{UP} por nuevas variables que cumplen $w_U \geq w_L$, sigue siendo una subestimación válida, no convexa y a medida que el intervalo $[w_L, w_U]$ es menor, la subestimación está más próxima a la función $tcu(w)$.

Para este caso R' queda de la forma:

$$\begin{aligned}
 \lambda tcu(w_L) + (1 - \lambda) tcu(w_U) &\leq -k^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \\
 w &= \lambda w_L + (1 - \lambda) w_U \\
 w_L &\leq w_U \\
 0 &\leq \lambda \leq 1 \\
 w^{LO} &\leq w_L \leq w^{UP} \\
 w^{LO} &\leq w_U \leq w^{UP}
 \end{aligned}
 \tag{R}$$

2) Si bien el procedimiento propuesto permite obtener regiones convexas que proyectadas al espacio original contienen a la región factible, no son las regiones más ajustadas. Desde el punto de vista matemático sería más exquisito trabajar con la cápsula convexa de la región factible, pero también más trabajoso en muchos casos. La reformulación propuesta implica un proceso simple y sin dificultades para implementar.

Se propone un ejemplo donde se puede observar las diferencias entre el conjunto convexo construido y la cápsula convexa para un conjunto definido por restricciones particulares.

Ejemplo 3.4.1: dado el conjunto de restricciones de un NLP:

$$(x - 3)^2 + (y - 3)^2 \geq 3, \quad 1 \leq x, y \leq 2$$

Según la construcción sugerida, la restricción debe reemplazarse por:

$$\begin{aligned}
 z_1 + z_2 &\leq -3 \\
 z_1 &\geq -(x^{(1)} - 3)^2 \lambda_1 - (x^{(2)} - 3)^2 (1 - \lambda_1) \\
 z_2 &\geq -(y^{(1)} - 3)^2 \lambda_2 - (y^{(2)} - 3)^2 (1 - \lambda_2) \\
 x &= x^{(1)} \lambda_1 + x^{(2)} (1 - \lambda_1) \\
 y &= y^{(1)} \lambda_2 + y^{(2)} (1 - \lambda_2) \\
 0 &\leq \lambda_1, \lambda_2 \leq 1
 \end{aligned}$$

Reemplazando en las desigualdades por $x^{(1)}=x_{lo}$ $x^{(2)}=x_{up}$ $y^{(1)}=y_{lo}$ $y^{(2)}=y_{up}$ se

obtiene:
$$\begin{aligned} z_1 &\geq -1 - 3\lambda_1 & x &= 2 - \lambda_1 \\ z_2 &\geq -1 - 3\lambda_2 & y &= 2 - \lambda_2 \end{aligned}$$

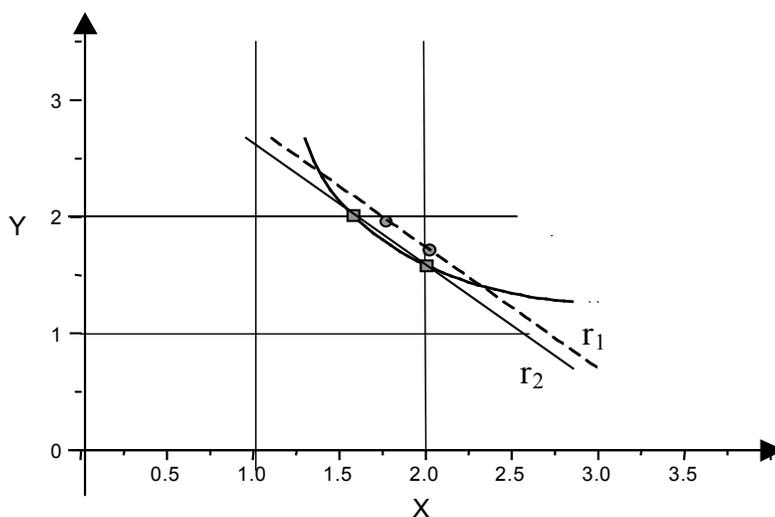
despejando λ_j $j=1,2$ en función de x e y , y reemplazando en las desigualdades de z_j :

$$3x + 3y \leq 11$$

es un semiespacio cuya frontera es una recta r_1 definida por la ecuación: $y + x = 11/3$ y representa la proyección en el espacio xy de una subestimación realizada en un espacio de mayor dimensión.

Por otro lado, es posible construir una región convexa definida por las cotas de las variables y el semiespacio cuya frontera es la recta r_2 : $y + x = 5 - \sqrt{2}$ definida por los puntos $(2, 3 - \sqrt{2})$ y $(3 - \sqrt{2}, 2)$, puntos de intersección entre la restricción no lineal y las cotas superior de x y superior de y respectivamente. Su ecuación es $x + y \leq 5 - \sqrt{2}$.

Ambas rectas son paralelas y sus ordenadas cumplen con la desigualdad $5 - \sqrt{2} < 11/3$ indicando que el semiespacio cuya frontera es r_1 contiene al semiespacio cuya frontera es r_2 y en consecuencia la región convexa relajada es mayor. Como se puede ver en la figura 3.4.1 la región más ajustada resultante es la cápsula convexa de la región factible original.



.Figura 3.4.1: Región factible no convexa y dos sobreestimaciones convexas

3.5 Problema principal

El problema principal es no convexo y se formula de la siguiente manera:

$$\begin{aligned}
 & \text{Min} \quad \sum_{i=1}^J (w_{U_i} - w_{L_i}) \\
 & \text{suj. a:} \quad f^c(\bar{x}, \bar{z}) \leq f^{UP} - \varepsilon \\
 & \quad \quad \quad \bar{h}^1(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) = \bar{0} \\
 & \quad \quad \quad \bar{g}^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \leq \bar{0} \\
 & \left. \begin{aligned}
 & \lambda_i tcu_i(w_{L_i}) + (1 - \lambda_i) tcu_i(w_{U_i}) \leq -k_i^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \\
 & w_i = \lambda_i w_{L_i} + (1 - \lambda_i) w_{U_i} \\
 & w_{L_i} \leq w_{U_i}
 \end{aligned} \right\} \quad i = 1, \dots, J \quad \text{MP} \\
 & \bar{x}^{LO} \leq \bar{x} \leq \bar{x}^{UP} \quad \bar{z}^{LO} \leq \bar{z} \leq \bar{z}^{UP} \\
 & \bar{\alpha}^{LO} \leq \bar{\alpha} \leq \bar{\alpha}^{UP} \quad \bar{\beta}^{LO} \leq \bar{\beta} \leq \bar{\beta}^{UP} \\
 & w_i^{LO} \leq w_{L_i} \leq w_i^{UP} \quad w_i^{LO} \leq w_{U_i} \leq w_i^{UP} \\
 & 0 \leq \lambda_i \leq 1 \quad i = 1, \dots, J
 \end{aligned}$$

donde f^{UP} es la mejor cota superior de la función objetivo f , lograda hasta el momento.

La función objetivo de MP minimiza la diferencia entre las cotas de las variables definidas en w . El valor de la nueva función objetivo siempre es no negativo y busca que para cada variable w_i coincida su cota superior con su cota inferior.

Cuando se resuelve MP y su solución resulta una solución factible del problema original, la restricción $\lambda_i tcu_i(w_{L_i}) + (1 - \lambda_i) tcu_i(w_{U_i}) \leq -k_i^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta})$ para algún $i = 1, \dots, J$ es activa en dicho punto.

En MP se incorporaron subestimaciones para los términos cóncavos $tcu(w)$ que resultan exactas, coincidiendo con el valor de la función, en los extremos w^{LO} y w^{UP} .

Si la solución w está en alguno de estos extremos luego la solución de MP es factible también para RP, de lo contrario en RP no cumple con la restricción

$$tcu_i(w) \leq -k_i^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) .$$

Si la solución de MP es factible para el problema RP, se toma como punto de partida para resolverlo en busca de mejores soluciones además de mejorar la cota superior obtenida hasta el momento.

En MP se agrega una restricción que condiciona la función objetivo a ser menor que la

última cota superior hallada, con el fin de obtener una solución factible del problema original y con mejor valor objetivo aún. Si el problema no tiene solución factible no existe solución con mejor valor objetivo f^{UP} y la región se puede eliminar del estudio. Este procedimiento realiza entonces una eliminación implícita de soluciones con peor o igual valor objetivo y por lo tanto también de subregiones.

Si w_{L_i} y w_{U_i} se fijan, el problema queda convexo, luego otro problema que se define es CP cuya región factible contiene a la región factible del problema original, resultando una relajación convexa de MP y por lo tanto su solución sirve como cota inferior del valor objetivo óptimo.

En el trabajo, los autores exponen una serie de resultados que fundamentan y proporcionan un buen funcionamiento del algoritmo.

$$\text{Min } \sum_{i=1}^J (w_{U_i} - w_{L_i})$$

$$\text{su. a: } f^c(\bar{x}, \bar{z}) \leq f^{UP} - \varepsilon$$

$$\bar{h}^1(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) = \bar{0}$$

$$\bar{g}^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \leq \bar{0}$$

$$\left. \begin{aligned} \lambda_i \text{tcu}_i(w_i^{LO}) + (1 - \lambda_i) \text{tcu}_i(w_i^{UP}) &\leq -k_i^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \\ w_i &= \lambda_i w_i^{LO} + (1 - \lambda_i) w_i^{UP} \end{aligned} \right\} \quad i = 1, \dots, J$$

CP

$$\bar{x}^{LO} \leq \bar{x} \leq \bar{x}^{UP} \quad \bar{z}^{LO} \leq \bar{z} \leq \bar{z}^{UP}$$

$$\bar{\alpha}^{LO} \leq \bar{\alpha} \leq \bar{\alpha}^{UP} \quad \bar{\beta}^{LO} \leq \bar{\beta} \leq \bar{\beta}^{UP}$$

$$0 \leq \lambda_i \leq 1 \quad i = 1, \dots, J$$

Teorema 3.5.1:

a) Si un punto factible en MP: $\bar{p}^* = (\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}_L^*, \bar{w}_U^*, \bar{\lambda}^*)$ tiene valor objetivo igual a 0, es la solución óptima global de MP y representa una solución factible de P con valor objetivo mejor que la solución conocida.

b) Si un punto en MP: $\bar{p}^* = (\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}_L^*, \bar{w}_U^*, \bar{\lambda}^*)$ tiene valor objetivo mayor que 0, pero $\lambda_i^* = 0$ o 1, para tales i tal que $0 < w_{U_i}^* - w_{L_i}^*$, corresponde a un punto factible de P con valor objetivo mejor que la solución conocida.

c) Si el óptimo global de MP: $\bar{p}^* = (\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}_L^*, \bar{w}_U^*, \bar{\lambda}^*)$ tiene un valor objetivo mayor que 0 luego no existe ninguna solución de P con valor objetivo menor que $(f^{UP} - \varepsilon)$.

d) Si MP es infactible, no hay ninguna solución de P con valor objetivo menor que $(f^{UP} - \varepsilon)$.

Demostración:

Sea $\bar{p} = (\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}, \bar{w}_L, \bar{w}_U, \bar{\lambda})$ un punto factible en MP.

Como es factible satisface la restricción $w_{U_i} \geq w_{L_i}$, luego $\sum_{i=1}^J (w_{U_i} - w_{L_i}) \geq 0$.

Sea $\bar{p}^* = (\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}_L^*, \bar{w}_U^*, \bar{\lambda}^*)$ la solución óptima de MP.

a) Sea \bar{p}^* tal que $\sum_{i=1}^J (w_{U_i}^* - w_{L_i}^*) = 0$ luego es solución óptima global del problema MP y

además como es suma de términos positivos debe cumplir que $\bar{w}_{U_i}^* - \bar{w}_{L_i}^* = 0 \quad \forall i$.

Por la restricción en MP: $w_i^* = \lambda_i w_{L_i}^* + (1 - \lambda_i) w_{U_i}^*$, $i = 1, \dots, J$ se cumple $\bar{w}_{L_i}^* = \bar{w}_{U_i}^* = \bar{w}_i^*$ para todo i .

Para ver que \bar{p}^* es un punto factible del problema P o equivalentemente RP, basta ver las restricciones no convexas de MP:

$$tcu_i(w_i^*) = \lambda_i^* tcu_i(w_{L_i}^*) + (1 - \lambda_i^*) tcu_i(w_{U_i}^*) \leq -k_i^c(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*) \quad \forall i = 1, \dots, J$$

luego $(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*)$ es un punto factible de RP, y \bar{x}^* punto factible de P que proporciona una nueva cota superior del valor objetivo óptimo buscado.

$f^c(\bar{x}^*, \bar{z}^*) \leq f^{UP} - \varepsilon$, por ser una restricción de MP y f^c es una subestimación de la función objetivo original, cumple que en este punto $f^c(\bar{x}^*, \bar{z}^*) = f(\bar{x}^*)$

luego $f(\bar{x}^*) \leq f^{UP} - \varepsilon$, dando una cota superior mejor que la actual.

b) Sea \bar{p}^* tal que $\sum_{i=1}^J (w_{U_i}^* - w_{L_i}^*) > 0$.

Para todo i tal que $w_{L_i}^* < w_{U_i}^*$: $\bar{w}_i^* = \bar{w}_{L_i}^*$ si $\lambda_i^* = 1$

$$\bar{w}_i^* = \bar{w}_{U_i}^* \quad \text{si} \quad \lambda_i^* = 0$$

estando en la misma situación que en (a) respecto de la factibilidad de RP y P.

c) Supongamos que existe un punto factible de P con $f(\bar{x}^*) \leq f^{UP} - \varepsilon$. Luego como P y RP son equivalentes, existe un punto factible en RP: $\bar{p} = (\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0, \bar{w}_0)$ que satisface todas las restricciones del problema, en particular $\overline{tcu}(\bar{w}_0) + \bar{k}^c(\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0) \leq \bar{0}$. Tomando $\bar{w}_{L_0} = \bar{w}_0$, $\bar{w}_{U_0} = \bar{w}_0$ y cualquier $\bar{\lambda}_0$ se tiene un punto $\bar{p} = (\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0, \bar{w}_{U_0}, \bar{w}_{L_0}, \bar{\lambda}_0)$ factible para MP con $\sum_{i=1}^J (w_{U_i}^* - w_{L_i}^*) = 0$.

Pero la solución óptima de MP tiene valor objetivo positivo. Absurdo que provino de suponer que existía una solución de P con mejor valor objetivo que hasta el momento hallada, luego no existe ninguna solución de P con valor objetivo menor que $(f^{UP} - \varepsilon)$.

d) Como $RF_{RP} \subseteq RF_{MP}$ si el problema MP no tiene solución factible luego tampoco la tiene RP y P luego no hay solución factible para P con valor objetivo menor que $(f^{UP} - \varepsilon)$ □

Teorema 3.5.2:

- a) CP es factible sí y solo si MP es factible.
- b) Si CP es infactible, no existe ninguna solución de P con valor objetivo menor que $f^{UP} - \varepsilon$.

Demostración:

a) Sea $\bar{p} = (\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0, \bar{w}_0)$ un punto factible del problema CP. Las cotas de w están fijas, $\bar{w}_{L_0} = \bar{w}^{LO}$ y $\bar{w}_{U_0} = \bar{w}^{UP}$.

$\bar{w}_{L_0} \leq \bar{w}_{U_0}$ y las restricciones restantes de MP se cumplen por ser las mismas que en CP. Luego el punto $(\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0, \bar{w}_{L_0}, \bar{w}_{U_0}, \bar{\lambda}_0)$ es factible en MP.

Ahora supongamos que $(\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0, \bar{w}_{L_0}, \bar{w}_{U_0}, \bar{\lambda}_0)$ es factible para MP. Luego se cumplen entre otras las restricciones $w_i = \lambda_{0i} w_{L_{0i}} + (1 - \lambda_{0i}) w_{U_{0i}}$ y

$$\lambda_{0i} tcu_i(w_{L_{0i}}) + (1 - \lambda_{0i}) tcu_i(w_{U_{0i}}) \leq -k_i^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \quad i = 1, \dots, J$$

Para que sea factible en CP debe hallarse $\bar{\lambda}$ tal que: $w_i = \lambda_i w_i^{LO} + (1 - \lambda_i) w_i^{UP}$ y $\lambda_i tcu_i(w_i^{LO}) + (1 - \lambda_i) tcu_i(w_i^{UP}) \leq -k_i^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}) \quad i = 1, \dots, J$.

Como $w_i^{LO} \leq w_{oi} \leq w_i^{UP}$ $i = 1, \dots, J$

$\exists \lambda_i$ de la forma $\lambda_i = (w_{oi} - w_i^{UP}) / (w_i^{LO} - w_i^{UP}) /$ cumple con las primeras restricciones.

Como $w_i^{LO} \leq w_{Loi} \leq w_{Uoi} \leq w_i^{UP}$ $i = 1, \dots, J$ y tcu es cóncava se verifica que

$$tcu_i(w_{Loi}) \geq \lambda_{Loi} tcu_i(w_i^{LO}) + (1 - \lambda_{Li}) tcu_i(w_i^{UP})$$

$$tcu_i(w_{Uoi}) \geq \lambda_{Uoi} tcu_i(w_i^{LO}) + (1 - \lambda_{Ui}) tcu_i(w_i^{UP})$$

Multiplicando cada desigualdad por $\lambda_{oi} \geq 0$ y $(1 - \lambda_{oi}) \geq 0$ y sumando:

$$\begin{aligned} -k_i^c(\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0) \geq tcu_i(w_{oi}) &\geq \lambda_{oi} tcu_i(w_{Loi}) + (1 - \lambda_{oi}) tcu_i(w_{Uoi}) \quad \text{y} \\ \lambda_{oi} tcu_i(w_{Loi}) + (1 - \lambda_{oi}) tcu_i(w_{Uoi}) &\geq \lambda_{oi} (\lambda_{Li} tcu_i(w_i^{LO}) + (1 - \lambda_{Li}) tcu_i(w_i^{UP})) + \\ (1 - \lambda_{oi}) (\lambda_{Ui} tcu_i(w_i^{LO}) + (1 - \lambda_{Ui}) tcu_i(w_i^{UP})) &\geq (\lambda_{oi} \lambda_{Li} + (1 - \lambda_{oi}) \lambda_{Ui}) tcu_i(w_i^{LO}) + \\ ((1 - \lambda_{oi}) \lambda_{Li} + (1 - \lambda_{oi}) (1 - \lambda_{Ui})) tcu_i(w_i^{UP}) & \end{aligned}$$

luego $\bar{\lambda}_i tcu_i(w_i^{LO}) + (1 - \bar{\lambda}_i) tcu_i(w_i^{UP}) \leq -k_i^c(\bar{x}_0, \bar{z}_0, \bar{\alpha}_0, \bar{\beta}_0)$ con

$$\bar{\lambda}_i = \lambda_{oi} \lambda_{Li} + (1 - \lambda_{oi}) \lambda_{Ui}$$

verificando las restricciones en CP.

b) Por a) si CP es infactible MP también lo es y por el teorema 3.5.2 d) se tiene el resultado \square

3.6 Ramificación

Sea $\bar{p} = (\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta}, \bar{w}_L, \bar{w}_U, \bar{\lambda})$ una solución óptima local de MP, con $0 < \lambda_i < 1$ para algún i , el punto hallado no es solución para el problema original y por lo tanto no es posible realizar ninguna inferencia. Para eliminar esta solución y poder continuar la búsqueda de la solución global de P, se realiza una partición de la región factible a partir de \bar{p}^* .

Teorema 3.6.1:

Sea $\bar{p}^* = (\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}_L^*, \bar{w}_U^*, \bar{\lambda}^*)$ una solución óptima para MP, con

$\sum_{i=1}^J (w_{Ui}^* - w_{Li}^*) > 0$ y $G^0 = \{i : w_{Li}^* < w_{Ui}^* \wedge 0 < \lambda_i^* < 1, i = 1, \dots, J\}$. Luego la restricción

$\lambda_i tcu_i(w_{Li}) + (1 - \lambda_i)tcu_i(w_{Ui}) \leq -k_i^c(\bar{x}, \bar{z}, \bar{\alpha}, \bar{\beta})$ es activa en \bar{p}^* para $i \in G^0$.

Demostración:

De las condiciones necesarias de KKT que se cumplen en \bar{p}^* , las condiciones donde intervienen las variables λ_i , w_{Li} y w_{Ui} son:

$$\mu_1(tcu(w_{Li}^*) - tcu(w_{Ui}^*)) + \gamma(w_{Li}^* - w_{Ui}^*) - \mu_\lambda^1 + \mu_\lambda^2 = 0 \quad (1)$$

$$-1 + \mu_1(\lambda_i^* tcu'(w_{Li}^*)) + \gamma(\lambda_i^*) + \mu_2 - \mu_{w_L}^1 + \mu_{w_L}^2 = 0 \quad (2)$$

$$1 + \mu_1((1 - \lambda_i^*)tcu'(w_{Ui}^*)) + \gamma(1 - \lambda_i^*) - \mu_2 - \mu_{w_U}^1 + \mu_{w_U}^2 = 0 \quad (3)$$

$$\mu_1(\lambda_i^* tcu_i(w_{Li}^*) + (1 - \lambda_i^*)tcu_i(w_{Ui}^*) + k_i^c(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*)) = 0$$

$$\mu_2(w_{Li}^* - w_{Ui}^*) = 0$$

$$\mu_\lambda^1(0 - \lambda_i^*) = 0 \quad \mu_\lambda^2(\lambda_i^* - 1) = 0$$

$$\mu_{w_L}^1(w_i^{LO} - w_{Li}^*) = 0 \quad \mu_{w_L}^2(w_{Li}^* - w_i^{UP}) = 0$$

$$\mu_{w_U}^1(w_i^{LO} - w_{Ui}^*) = 0 \quad \mu_{w_U}^2(w_{Ui}^* - w_i^{UP}) = 0$$

$$\mu_1, \mu_2, \mu_\lambda^1, \mu_\lambda^2, \mu_{w_L}^1, \mu_{w_L}^2, \mu_{w_U}^1, \mu_{w_U}^2 \geq 0$$

Como por hipótesis $0 < \lambda_i^* < 1$ para $i \in G^0$, $\mu_\lambda^1 = \mu_\lambda^2 = 0$

Además $w_i^{LO} \leq w_{Li}^* < w_{Ui}^* \leq w_i^{UP}$ para $i \in G^0$, $\mu_2 = 0$ y $\mu_{w_L}^2 = \mu_{w_U}^1 = 0$ Quedando:

$$\mu_1(tcu(w_{Li}^*) - tcu(w_{Ui}^*)) + \gamma(w_{Li}^* - w_{Ui}^*) = 0 \quad (1)$$

$$-1 + \mu_1(\lambda_i^* tcu'(w_{Li}^*)) + \gamma(\lambda_i^*) = 0 \quad (2)$$

$$1 + \mu_1((1 - \lambda_i^*)tcu'(w_{Ui}^*)) + \gamma(1 - \lambda_i^*) = 0 \quad (3)$$

$$\mu_1(\lambda_i^* tcu_i(w_{Li}^*) + (1 - \lambda_i^*)tcu_i(w_{Ui}^*) + k_i^c(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*)) = 0$$

Si $\mu_1 = 0$ de (1), como $w_{Li}^* < w_{Ui}^*$, $\gamma = 0$. Pero (2) queda inconsistente. Luego debe ser $\mu_1 \neq 0$ y por (3) la restricción debe ser activa \square .

Cuando se obtiene una solución óptima de MP no factible para P, se desea eliminar, para ello se recurre a particionar la región factible, particionando el rango para la variable w_{i_0} .

Cuando existe más de un w_{i_0} para $i_0 \in G^0$, se necesita un criterio para seleccionar la variable que realiza la partición. Los autores proponen distintas reglas para lograr la partición sobre un w_{i_0} tal que,

Regla 1: $i_0 = \arg \max_{i \in G^0} \{w_{U_i}^* - w_{L_i}^*\}.$

Regla 2: $i_0 = \arg \max_{i \in G^0} \{(w_{U_i}^* - w_{L_i}^*) / (w_i^{LO} - w_i^{UP})\}.$

Regla 3: $i_0 = \arg \max_{i \in G^0} \{tcu_i(w_i^*) - (\lambda_i^* tcu_i(w_{L_i}^*) + (1 - \lambda_i^*) tcu_i(w_{U_i}^*))\}.$

Regla 4: $i_0 = \arg \max_{i \in G^0} \{\min\{w_i^* - w_{L_i}^*, w_{U_i}^* - w_i^*\}\}$

Regla 5: versión ponderada de la regla 4.

Regla 6: elegir la variable siguiendo un orden de prioridades.

3.7 Pasos del Algoritmo

Dado un problema en la forma P.

Paso 1: Inicialización:

Seleccionar un valor para el parámetro ε . Reformular el problema en la forma RP y luego definir MP. Sea $P = \{R_0\}$ el conjunto de regiones, donde R_0 es el hiperrectángulo definido por las cotas en RP. Resolver RP con un algoritmo local. Si el problema tiene solución factible, su solución óptima es $(\bar{x}^{OPT}, \bar{z}^{OPT}, \bar{\alpha}^{OPT}, \bar{\beta}^{OPT})$ y f^{UP} su valor objetivo. Si no, tomar para f^{UP} un valor suficientemente grande. Sea $iter=1$ y $R_i = R_0$.

Paso 2: resolver el problema MP en la región R_i .

Si MP es infactible en la región R_i , $P=P \setminus \{R_i\}$. Ir al paso 5.

Si no sea $(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}_L^*, \bar{w}_U^*, \bar{\lambda}^*)$ el óptimo.

Definir $G^0 = \{i : w_{L_i}^* < w_{U_i}^* \wedge 0 < \lambda_i^* < 1, i = 1, \dots, J\}$.

Si $G^0 = \emptyset$, (la solución actual es factible para RP) ir al paso 3, sino ir al paso 4.

Paso 3: Resolver RP en la región R_i a partir del punto $(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}^*)$ agregando la restricción $f^c(\bar{x}, \bar{z}) \leq f^{UP} - \varepsilon$. Sea $(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*)$ la solución local y f^* el valor objetivo. Actualizar $(\bar{x}^{OPT}, \bar{z}^{OPT}, \bar{\alpha}^{OPT}, \bar{\beta}^{OPT}) = (\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*)$, $f^{UP} = f^*$. Ir al paso 5.

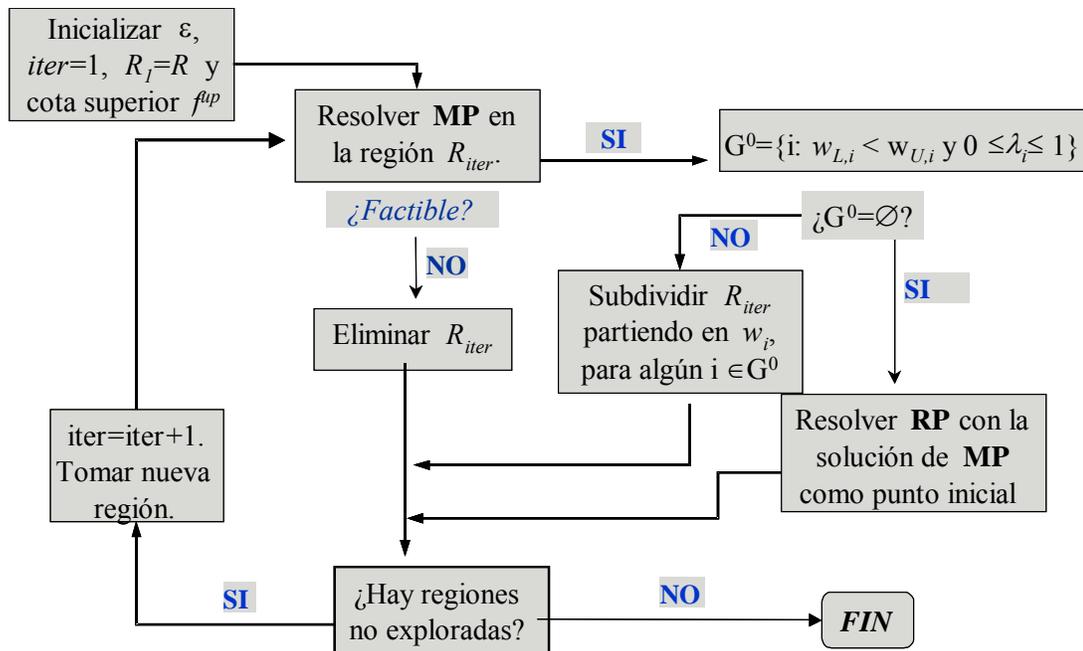
Paso 4: particionar R_i en una o más variables w_{i_0} para $i_0 \in G^0$, generando las subregiones $\{R_{i_1}, R_{i_2}, \dots, R_{i_q}\}$. Sea $P = P \setminus \{R_i\}$. Ir al paso 5.

Paso 5: Criterio de parada:

Si $P = \emptyset$, parar: la solución actual es el óptimo global. Si no hacer $iter = iter + 1$ y elegir R_i de P . Retornar al paso 2.

Cuando se implementa el algoritmo, los problemas RP y MP no convexos se resuelven con un algoritmo para problemas convexos, garantizando óptimos locales de estos problemas. Cuando se llega al paso de resolver RP siempre se cuenta con un punto factible proporcionado por MP previamente resuelto. En cambio, MP puede resultar infactible sin serlo. Para salvar este inconveniente se resuelve el problema convexo CP, si tiene solución óptima factible, ésta se toma como punto inicial para resolver MP y si es infactible, el teorema 3.5.2 permite concluir la infactibilidad de MP.

Diagrama del algoritmo



3.8 Convergencia

El algoritmo genera un árbol a partir de las particiones de la región factible, y la acotación de la función objetivo se incorpora en la formulación del modelo. Las subregiones se eliminan cuando el problema MP resulta infactible requiriendo que el valor de la función objetivo es menor que la mejor cota superior lograda hasta el momento. Luego en el proceso todas las subregiones deben estudiarse. Con una precisión ε , el proceso de particionar, chequear y eliminar regiones es finito, así las características de convergencia de los algoritmos de ramificación y acotación son válidas para el algoritmo propuesto. El algoritmo finaliza cuando todas las subregiones se descartan porque ha proporcionado una cota superior o por infactibilidad. Se propone una reducción de cotas, alternativa a las planteadas en el capítulo 2 por Ryoo y Sahinidis, (1995), Zamora y Grossmann, (1999).

3.9 Reducción de cotas

Se propone una reducción de rango basada en factibilidad. El ajuste de cotas propuesto en el trabajo, es similar al test de monotonía y acotamiento propuesto por Hansen y otros (1991).

Dado el conjunto de restricciones de igualdad originales:

$$\begin{aligned} h_1(x_1, x_2, \dots, x_n) &= 0 \\ h_2(x_1, x_2, \dots, x_n) &= 0 \\ &\vdots \\ h_p(x_1, x_2, \dots, x_n) &= 0 \end{aligned}$$

Si se supone que de cada ecuación se puede despejar una variable, de la forma,

$$x_k = g_{lk}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_p), \quad l \in \{1, \dots, p\} \text{ tal que para todo } j \neq k, \text{ el signo de } \frac{\partial g_{lk}}{\partial x_j},$$

$s_{ij} = \text{sgn}\left(\frac{\partial g_{lk}}{\partial x_j}\right)$, es constante en la caja R_i de variación de las variables que se esté trabajando.

Si la ramificación se realiza sobre una de las variables x_1, \dots, x_p el rango de x_k puede cambiar. Para actualizar las cotas, se debe calcular

$$u_{lm} = \begin{cases} x_m^{LO} & \text{if } s_{lm} = -1 \\ x_m^{UP} & \text{if } s_{lm} = 1 \end{cases}$$

Luego una cota superior para x_k es $x_k \leq \min \{x_k^{UP}, \min_{l=1, \dots, p} \{g_{lk}(u_{l0}, \dots, u_{lk-1}, u_{lk+1}, \dots, u_{lp})\}\}$.

Esta condición puede relajarse usando $x_k \leq g_{lk}^U(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_p)$

Una cota inferior de la variable es $x_k \geq \max \{x_k^{LO}, \max_{l=1, \dots, p} \{g_{lk}(l_{l0}, \dots, l_{lk-1}, l_{lk+1}, \dots, l_{lp})\}\}$ con

$$l_{lm} = \begin{cases} x_m^{UP} & \text{if } s_{lm} = -1 \\ x_m^{LO} & \text{if } s_{lm} = 1 \end{cases},$$

Análogamente puede relajarse a los fines prácticos usando $x_k \geq g_{lk}^L(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_p)$, donde g_{lk} , g_{lk}^U y g_{lk}^L se calculan con anterioridad. El resultado del cálculo de las cotas depende del orden en que se trabajan las variables. Es posible determinar un orden apropiado construyendo una lista de implicación para cada variable. El procedimiento se puede aplicar en forma recursiva hasta que la reducción se torne no significativa según una tolerancia prefijada.

Este procedimiento no ha sido implementado en forma automática, sino que se construye a partir de las expresiones presentadas para cada ejemplo.

Para aplicar el algoritmo a problemas MINLP, se adopta lo planteado en la sección 2.7.1 (Ryoo y Sahinidis, 1995) para trabajar las variables en forma continua:

$$0 \leq x \leq 1, \quad x(1-x) \leq 0$$

En MP se puede agregar la restricción no convexa:

$$x - \lambda(x_L)^2 - (1-\lambda)(x_U)^2 \leq 0$$

y $x_U - x_L$ forman parte de la función objetivo.

El algoritmo se implementó en GAMS (Brooke y otros, 1997) y se probó en 6 problemas test. Se usaron los resolvidores GAMS/CONOPT para resolver NLP en una PC Pentium 4 1.5 GHz.

3.10 Conclusiones

El trabajo plantea un algoritmo de optimización para resolver problemas no convexos con algunas características a tener en cuenta. Antes de realizar estimaciones convexas para las funciones no convexas, se reformula el problema siguiendo el procedimiento explicado e implementado por Pantelides y otros en su trabajo ⁽⁷⁾ donde se realizan sustituciones que si bien aumentan la dimensión del problema, desagregan funciones altamente no convexas en expresiones simples, que tienen una no convexidad conocida y de más fácil manejo.

El algoritmo utiliza un procedimiento de ramificación mediante particiones de la región factible, y definición y posterior resolución de subproblemas a partir de las subregiones. Seguido de un procedimiento de acotación donde se realiza una eliminación implícita de soluciones gracias al agregado de una restricción en un problema de optimización a resolver donde se fuerza a buscar soluciones con valor objetivo mejor en un factor ε al obtenido actualmente.

Si bien el proceso iterativo se planteó a partir de la resolución alternada de dos problemas no convexos generados a partir del problema inicial, en la evolución del algoritmo se observa la necesidad de agregar la resolución de un problema convexo. Esta situación será aprovechada en el capítulo siguiente para implementar una reducción de la región factible. Observando los problemas test resueltos, los resultados que se obtienen son altamente competitivos frente a otras propuestas, cuando se utiliza un proceso de reducción de rango que es implementado para cada ejemplo en particular.

Los algoritmos explicados en el capítulo anterior indican la obtención de cotas superiores de la solución óptima global del problema mediante puntos factibles del mismo, algunas veces sin explicar una estrategia para hallarlos. En este trabajo se puede ver como la reformulación y las relajaciones del problema permiten una metodología para llegar a un punto factible del mismo resolviendo un problema con más garantías para lograrlo como es un problema convexo.

Capítulo 4

Método de optimización global: una variante de “Improve and Branch” con una forma diferente de partición

4.1 Introducción

En el presente capítulo primeramente se proponen algunas variantes de implementación del algoritmo, y se estudian e implementan diferentes alternativas para la etapa de reducción de la región factible y rango de las variables.

Seguidamente, en busca de nuevas alternativas, se presenta un nuevo algoritmo de ramificación y mejora del objetivo que adopta las formas de RP y MP con reglas de reducción de rango y utiliza una forma diferente de realizar el proceso de partición.

A partir de una solución factible de RP, se construyen restricciones que particionarán la región factible del problema relajado en tres subregiones que subestiman de manera más ajustada la región factible original. Estas nuevas restricciones permiten además una estrategia de reducción de la región factible eficiente.

4.2 Variantes del algoritmo

Debido a que se va a trabajar con las mismas reformulaciones para el problema explicadas en el algoritmo del capítulo anterior, es necesario que las funciones del modelo sean separables o resulten de composiciones de funciones definidas por términos algebraicos simples relacionados por cinco operaciones elementales: suma, resta, multiplicación, división y exponenciación y un número pequeño de funciones trascendentes (ver sección 2.7.3)

El algoritmo presentado en el capítulo anterior, propone como ya se dijo, en cada iteración la resolución de dos problemas no convexos en forma secuencial primero el definido en MP y seguidamente RP.

RP es una reformulación equivalente que reemplaza al problema original P debido a que a partir de la solución de uno es posible obtener la solución del otro. Esta reformulación es de mayor dimensión en término de número de variables y número de restricciones, y se obtiene utilizando redefiniciones de variables y/o términos que forman las expresiones

algebraicas simples de cada función del modelo matemático para lograr un problema con función objetivo convexa, igualdades lineales, restricciones completamente convexas o restricciones que combinen términos univariados cóncavos y convexas. Las funciones trigonométricas que no se contemplan en esta reformulación podrían tratarse particionando la región en zonas de concavidad y convexidad de la misma.

En este capítulo se simplificará la notación tomando en xc todas las variables del problema que no intervienen en términos cóncavos, y las que sí lo hacen en w .

$$\text{Min } f^c(xc, w)$$

$$\text{suje.a: } \bar{h}^c(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

$$\overline{tcu}(w) + \bar{k}^c(xc, w) \leq \bar{0}$$

RP

$$xc^{LO} \leq xc \leq xc^{UP}$$

$$w^{LO} \leq w \leq w^{UP}$$

Se subestiman los términos cóncavos $tcu(w)$ por su envolvente convexa, se reemplazan las cotas w^{LO} y w^{UP} por nuevas variables que cumplen $w_U \geq w_L$, el problema resultante MP sigue siendo una subestimación válida, no convexa.

$$\text{Min } \sum_{i=1}^J (w_{Ui} - w_{Li})$$

$$\text{suje.a: } f^c(xc, w) \leq f^{UP} - \varepsilon$$

$$\bar{h}^l(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

$$\left. \begin{aligned} \lambda_i tcu_i(w_{Li}) + (1 - \lambda_i) tcu_i(w_{Ui}) &\leq -k_i^c(xc, w) \\ w_i &= \lambda_i w_{Li} + (1 - \lambda_i) w_{Ui} \\ w_{Li} &\leq w_{Ui} \\ 0 &\leq \lambda_i \leq 1 \end{aligned} \right\} i = 1, \dots, J$$

MP

$$xc^{LO} \leq xc \leq xc^{UP}$$

$$w_i^{LO} \leq w_{Li} \leq w_i^{UP} \quad w_i^{LO} \leq w_{Ui} \leq w_i^{UP}$$

Si bien se usan las envolventes convexas de las funciones cóncavas por término simple, para la restricción en su conjunto no siempre es la expresión más ajustada (ver sección 3.4). Además se cambia la función objetivo por una función cuyo fin es lograr que la solución óptima de este problema sea también factible para RP. De la forma que han sido expresadas y explicadas en el capítulo anterior, estas subestimaciones reemplazan los términos cóncavos por términos no convexos (bilineales) cuando w_{L_i} y w_{U_i} se las considera variables en el problema y convexa cuando se fijan.

Al resolver el primer problema no convexo MP en el proceso iterativo, no se cuenta con una solución factible inicial, luego se puede reportar una solución infactible sin serlo el problema. Para salvar este inconveniente cada vez que la resolución de MP resulte infactible, se planteó resolver el problema CP, que surge de fijar w_{L_i} y w_{U_i} , condición suficiente para que el problema se torne convexo. Si CP tiene solución óptima factible, ésta se toma como punto inicial para resolver MP y si es infactible, el teorema 3.5.2 permite concluir la infactibilidad del problema MP.

Se puede observar de la evolución del algoritmo que a medida que el proceso de partición avanza se genera una gran cantidad de subregiones infactibles que sólo se pueden eliminar luego de resolver dos problemas, la infactibilidad del problema convexo CP asociado permite la eliminación de la subregión. Basados en esta situación, se propone invertir el orden de resolución de los problemas resolviendo primero el problema CP y se modifican algunos aspectos de CP. De esta forma, se eliminan todos los subproblemas que se generan y son infactibles resolviendo solamente un problema convexo. Para aquellos subproblemas factibles, al resolver CP se genera una solución factible para MP con mejor valor objetivo que la proporcionada por la última cota superior por tener entre sus restricciones la restricción $f^c(xc, w) \leq f^{UP} - \varepsilon$.

Resolver el problema MP con funciones objetivos diferentes a la utilizada por los autores, tal como
$$\text{Min} \sum_i \left(\lambda_i tcu_i(w_{L_i}) + (1 - \lambda_i) tcu_i(w_{U_i}) - tcu_i(w_i) \right)$$
, no mejoran el trabajo de obtener un punto factible para RP.

La función objetivo de CP podría elegirse arbitrariamente y una opción conveniente vuelve a ser minimizar la función objetivo convexificada.

Cada vez que se obtiene una solución factible para la reformulación del modelo original, se actualiza la cota superior del valor objetivo, f^{UP} ; y mientras el algoritmo no realice

particiones de la región factible, el valor objetivo óptimo del problema convexo CP permite actualizar f^{LO} , la cota inferior del valor objetivo buscado en forma directa.

$$\text{Min } f^c(xc, w)$$

$$\text{suj.a: } f^c(xc, w) \leq f^{UP} - \varepsilon$$

$$\bar{h}^1(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

$$\left. \begin{aligned} \lambda_i tcu_i(w_i^{LO}) + (1 - \lambda_i) tcu_i(w_i^{UP}) &\leq -k_i^c(xc, w) \\ w_i &= \lambda_i w_i^{LO} + (1 - \lambda_i) w_i^{UP} \end{aligned} \right\} \quad i = 1, \dots, J \quad \text{CP}$$

$$xc^{LO} \leq xc \leq xc^{UP}$$

$$0 \leq \lambda_i \leq 1 \quad i = 1, \dots, J$$

El criterio utilizado para realizar una ramificación responde a observar la variable con mayor diferencia entre el valor de la función y su subestimación:

$$\text{Se particiona } w_{i_0} \text{ tal que } i_0 = \max_{i \in G^0} \{tcu_i(w_i^*) - (\lambda_i^* tcu_i(w_{Li}^*) + (1 - \lambda_i^*) tcu_i(w_{Ui}^*))\}$$

$$\text{donde } G^0 = \{i : w_{Li}^* < w_{Ui}^* \wedge 0 < \lambda_i^* < 1, i = 1, \dots, J\}$$

Con esta regla solo se realizarán particiones sobre las variables que participan en términos no convexos y tiene como objetivo lograr mejores aproximaciones al problema original.

Una condición que puede tenerse en cuenta es $z = f^c(xc, w) \geq f^{LO} + \varepsilon$, pero como resulta una representación no convexa, se agrega al problema mediante una cota inferior de la variable $z^{LO} = f^{LO} + \varepsilon$. Esta cota no se agrega en el problema MP para que el punto óptimo y factible de CP sea factible para MP y pueda usarse como punto inicial para su resolución. Esta cota puede ser útil para eliminar subproblemas que están al mismo nivel en la ramificación.

4.3 Estrategias de reducción de la región factible

En las conclusiones del capítulo 2 se destacó la importancia de desarrollar estrategias de reducción de las regiones que se vayan generando en el proceso de partición para acelerar la convergencia de cualquier algoritmo de partición y ramificación. En esta sección se buscan y se demuestran nuevos elementos que den soporte a esta situación.

Alternativa 1

Otro elemento motivador del ya mencionado es invertir el orden de resolución de los problemas CP y MP es poder implementar los mecanismos de reducción de rangos y región factible presentados por Ryoo y Sahinidis⁽¹⁷⁾ en su trabajo.

Al resolver el problema CP, si se obtiene una solución infactible se concluye que el problema MP es infactible y puede desecharse la subregión que lo define. De lo contrario si se obtiene una solución factible, se deben tener en cuenta dos situaciones posibles:

- La solución es factible también para el problema original y por lo tanto es la solución óptima global para el problema original en la subregión considerada.
- La solución es factible para MP no para el problema original.

Esta última situación es la que se da con más frecuencia. Los multiplicadores asociados a la solución del problema convexo pueden utilizarse para implementar el procedimiento de reducción de región factible propuesto por Ryoo y Sahinidis permitiendo reducir rangos de variación de algunas variables y construir restricciones para reducir la región factible. El procedimiento elimina partes de la región factible donde no existen soluciones con mejor valor objetivo hasta el momento hallado (Ver sección 2.6.2) y como la reducción mantiene la solución hallada de CP se propone resolver el problema MP en una región factible menor o igual en el peor caso, en busca de una solución factible para el problema original. Como en el problema convexo se toman las variables w_{Li} y w_{Ui} iguales a su cota inferior y superior respectivamente, se podría pensar en aplicar los test 3 y 4. Para obtener el problema convexo es necesario fijar los valores de las variables y no es posible incorporar las restricciones $w_{Li} \leq w_i^{LO}$ y $w_{Ui} \geq w_i^{UP}$. La siguiente afirmación demuestra que no es de utilidad la implementación de los test 3 y 4 al algoritmo que aquí se trabaja.

Afirmación: sean P un NLP no convexo, RP una reformulación equivalente de P y una subestimación no convexa de RP incorporando las nuevas variables w_{L_j} y w_{U_j} de la forma MP . Si se toman w_{L_j} y w_{U_j} iguales a su cota inferior y superior respectivamente resulta el problema convexo CP, luego los valores que se calculan para K_j y π_j usando los test 3 y 4, no reducen en ningún caso las cotas actuales de la variable.

Demostración

Para definir el problema convexo se fijan w_{L_i} y w_{U_i} en su cota inferior y superior respectivamente. Luego es posible aplicar el test 3, si $\lambda_j^* > 0$ sea $\pi_j = w_j^{UP} - \frac{L-U}{\lambda^*}$

En el algoritmo, $L = f^c(xc^*, w^*)$ es la mejor cota inferior de la subregión en que se está trabajando y $U = f^{UP}$ la mejor cota superior del problema hallada hasta el momento respectivamente.

Por otro lado el problema CP tiene incorporado la restricción $f^c(xc, w) \leq f^{UP} - \varepsilon$ luego se cumple $L-U \leq -\varepsilon$ para $\varepsilon > 0$ pequeño,

$$\frac{L-U}{\lambda_j^*} < 0 \text{ resultando siempre } w_j^{UP} < \pi_j$$

Análogamente, el test 4 se puede aplicar, si $\lambda_j^* > 0$ sea $K_j = w_j^{LO} + \frac{L-U}{\lambda_j^*}$

Luego siempre se verifica $w_j^{LO} > K_j$.

Por lo tanto los valores que se calculan para K_j y π_j no reducen en ningún caso las cotas actuales de la variable \square

El siguiente teorema asegura que la reducción de la región factible con los test 1 y 2 no eliminan la solución óptima del problema CP.

Teorema 4.3.1:

Sean P un NLP no convexo, RP una reformulación equivalente de P y una subestimación no convexa de RP incorporando las nuevas variables w_{L_j} y w_{U_j} de la forma MP . Si se

toman w_{L_j} y w_{U_j} iguales a su cota inferior y superior respectivamente resulta el problema convexo CP. La solución óptima del problema convexo CP es una solución factible para el problema RP con la región factible reducida por los test 1, 2 y el resultado del Teorema 2.6.2.3

Demostración:

Sea (xc^*, w^{LO}, w^{UP}) solución del problema CP.

Si $x_j^* = x_j^U$ para algún j con multiplicador $\lambda_j^* > 0$ asociado a la restricción $x_j - x_j^U \leq 0$ luego es posible aplicar el test 1:

Si $x_j^L < K_j$ se actualiza la cota inferior $x_j^L = K_j$, caso contrario no se modifica.

Donde $K_j = x_j^U - \frac{U-L}{\lambda_j^*} < x_j^U$ ya que $\frac{U-L}{\lambda_j^*} > 0$

Luego $x_j^* \in [K_j, x_j^U]$ perteneciendo al nuevo intervalo reducido.

Análogamente si $x_j - x_j^L \leq 0$ es activa en la solución óptima de CP con multiplicador $\lambda_j^* > 0$, se aplica el test 2.

Se calcula $\pi_j = x_j^L + \frac{U-L}{\lambda_j^*}$.

Si $x_j^U > \pi_j$ se redefine $[x_j^L, x_j^U] = [x_j^L, \pi_j]$ Luego $x_j^* \in [x_j^L, \pi_j]$ pertenece al nuevo intervalo reducido.

El resultado del Teorema 2.6.2.3 se aplica a restricciones activas en el óptimo, generando una cota inferior para las mismas, luego $g_i^c(xc^*, w^*) = 0 \geq g_i^c(xc, w) \geq -\frac{U-L}{\mu_i^*}$ esto implica

que el punto (xc^*, w^{LO}, w^{UP}) es factible para las nuevas restricciones.

Como (xc^*, w^{LO}, w^{UP}) es factible para CP y cumple con todas las restricciones adicionales de la reducción de la región factible e intervalo, es factible para el problema RP con la subregión reducida \square

Las restricciones generadas por el Teorema 2.6.2.3 que se agregan al problema MP no serán incorporadas en el problema CP ya que generalmente son no convexas. Las restricciones que se generan no son acumulativas en las iteraciones sucesivas, sino que

luego de resolver cada problema CP es posible actualizar la cota inferior asociada a cada restricción de desigualdad que se le ha aplicado el resultado del teorema.

Por otro lado, la reducción lograda de esta forma para una subregión en particular, será tomada en cuenta para las subregiones que son particiones de dicha subregión en nodos sucesivos.

Otra reducción de rangos de variables surge a partir de las restricciones

$$\bar{g}_i^c(xc,w) \geq -\frac{U-L}{\mu_i^*}$$

implementada para variables que intervienen en las mismas en forma lineal, llegando a una cota inferior de la variable si la derivada parcial de la función respecto a la variable es positiva, o una cota superior si es negativa.

Se implementaron las reducciones propuestas por los test 1 y 2 y el teorema 3.6.2.3 y se pudo comprobar que en ninguno de los ejemplos realizaron mejoras significativas en la cantidad de iteraciones que el algoritmo requirió para llegar al óptimo.

Por ejemplo en el problema propuesto por Hartman las reducciones de la alternativa 1 son más relajadas que las que se pueden proponer a partir de los cambios de variables propuestos por la reformulación del problema.

Un trabajo presentado posteriormente por Sheckman y Sahinidis⁽²¹⁾, demuestra que la reducción que se obtienen con estos test es menor a la lograda por otras estrategias presentadas en el mismo trabajo basadas en programación lineal paramétrica.

Alternativa 2

En el proceso de reformulación del modelo inicial, se trabajan matemáticamente las expresiones no convexas para lograr un modelo de la forma RP equivalente. En este proceso, se definen nuevas variables mediante una relación matemática con variables originales del modelo, luego esta dependencia puede utilizarse para el cálculo de las cotas de estas nuevas variables. Una alternativa válida para reducir rango de variación de las variables, es realizar el cálculo de las cotas de las nuevas variables usando expresiones que las actualicen ante cualquier cambio de cotas de las variables originales intervinientes.

Por ejemplo, si en el modelo se tiene un término bilineal, debe reemplazarse por nuevas variables z , α y β definidas de la forma:

$$\alpha = x_1 + x_2$$

$$\beta = x_1 - x_2$$

$$z = x_1 x_2 = \frac{\alpha^2 - \beta^2}{4}$$

En el modelo reformulado, cada término bilineal se reemplaza por z y se le deben agregar las restricciones:

$$\alpha = x_1 + x_2$$

$$\beta = x_1 - x_2$$

$$-4z + \alpha^2 - \beta^2 \leq 0$$

$$4z + \beta^2 - \alpha^2 \leq 0$$

Además deben incorporarse cotas para las nuevas variables, que se derivan de las variables originales. En vez de calcular externamente un valor inicial para estas cotas, se busca la manera general de calcularlas y se incorporan al algoritmo, en este caso:

$$x_1^L + x_2^L \leq \alpha \leq x_1^U + x_2^U$$

$$x_1^L - x_2^U \leq \beta \leq x_1^U - x_2^L$$

$$\min(x_1^L x_2^L, x_1^U x_2^U, x_1^L x_2^U, x_1^U x_2^L) \leq z \leq \max(x_1^L x_2^L, x_1^U x_2^U, x_1^L x_2^U, x_1^U x_2^L)$$

De esta forma por ejemplo para la variable z las cotas se pueden actualizar de acuerdo a:

$$z^L = \max\{z^L, \min(x_1^L x_2^L, x_1^U x_2^U, x_1^L x_2^U, x_1^U x_2^L)\}$$

$$z^U = \min\{z^U, \max(x_1^L x_2^L, x_1^U x_2^U, x_1^L x_2^U, x_1^U x_2^L)\}$$

Si las variables intervinientes son no negativas, cualquier partición del rango de variación de las mismas también lo serán, por lo tanto las expresiones se simplifican de la siguiente manera:

$$z^L = \max\{z^L, x_1^L x_2^L\}$$

$$z^U = \min\{z^U, x_1^U x_2^U\}$$

Este último intervalo, fue denominado intervalo de inclusión natural por Byrne y Bogle en su trabajo.

Cuando en la expresión $z=f(x)$ resulta sencillo hallar sus valores máximos y mínimos, las cotas de la variable z serán ajustadas, de lo contrario se buscan cotas a estos extremos que estimen de manera menos precisa.

En los ejemplos que se estudian en este capítulo, se expondrán cuando sea necesario, las nuevas relaciones que no hallan sido ya explicadas para llevar a cabo este procedimiento.

Alternativa 3

Para formalizar esta alternativa se usa la notación $f(x)$ y $f^c(x)$ para la función objetivo y su subestimación ya que lo aquí presentado sirve no sólo para la reformulación con la que trabaja el algoritmo, sino también para otras situaciones análogas.

Se propone una reducción de cotas para las variables intervinientes en la función objetivo. Las restricciones que permiten llevarla a cabo son:

$$f^{LO} \leq f^c(x) \leq f^{UP}$$

Generalmente la función objetivo convexificada resulta lineal, de no ser así, para implementar esta alternativa de manera más sencilla, se busca su reformulación lineal.

Las reducciones se realizan cada vez que se actualiza la cota superior o inferior del valor objetivo.

El algoritmo actualiza la cota superior del valor objetivo buscado cada vez que se obtiene una solución factible para el problema reformulado, y por lo tanto también para el problema original. Generalmente el algoritmo obtiene estas soluciones, incluyendo la solución óptima, sólo en las primeras iteraciones, luego las reducciones a partir de

$$f^c(x) \leq f^{UP}$$

son útiles en las primeras iteraciones.

Sea \bar{x}_j una variable que define la función objetivo reformulada, interviniendo de manera lineal, es decir $a_j \bar{x}_j$ es un término de $f^c(x)$

- Si $a_j = \frac{\partial f^c(x)}{\partial x_j} \geq 0$ se obtiene una cota superior para la variable.

Despejando la variable \bar{x}_j de (4.1) y evaluando las variables restantes $x_k, \forall k \neq j$ en su

cota superior si $\frac{\partial f^c(x)}{\partial x_k} < 0$ y en su cota inferior si $\frac{\partial f^c(x)}{\partial x_k} > 0$ resulta:

$$x_j \leq \frac{1}{a_j} f^{UP} + G(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_l) \Big|_{\text{cotas}}$$

- Si $a_j = \frac{\partial f^c(x)}{\partial x_j} < 0$ se obtiene una cota inferior de la variable.

Evaluando las variables restantes $x_k, \forall k \neq j$ en su cota inferior si $\frac{\partial f^c(x)}{\partial x_k} < 0$ y en su cota

superior si $\frac{\partial f^c(x)}{\partial x_k} > 0$ resulta: $x_j \geq \frac{1}{a_j} f^{UP} + G(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_l) \Big|_{\text{cotas}}$

Análogamente puede se puede trabajar para las variables en el vector \bar{z}_j .

Soluciones del problema convexo CP resuelto sobre una relajación convexa de la región factible original, resultan una cota inferior del valor objetivo óptimo buscado.

Se pueden obtener actualizaciones de la cota inferior de f^{LO} luego de comenzar el proceso de partición de la región factible. Formalmente se presentan dos resultados que permiten implementar esta actualización en el algoritmo.

Se necesita definir niveles en el proceso de ramificación del árbol. Se considera un nivel inicial con un primer nodo con problemas definidos en la región factible original sin particiones. A partir de él se realiza una partición de la región factible generalmente en dos regiones, definiendo otros dos nodos pero a un mismo nivel 1 El nivel siguiente comprende todos los nodos que provienen de ramificaciones de nodos del nivel anterior.

En la figura 4.2.1 se grafican los tres primeros niveles de un proceso de ramificación donde cada partición redefinen dos subregiones.

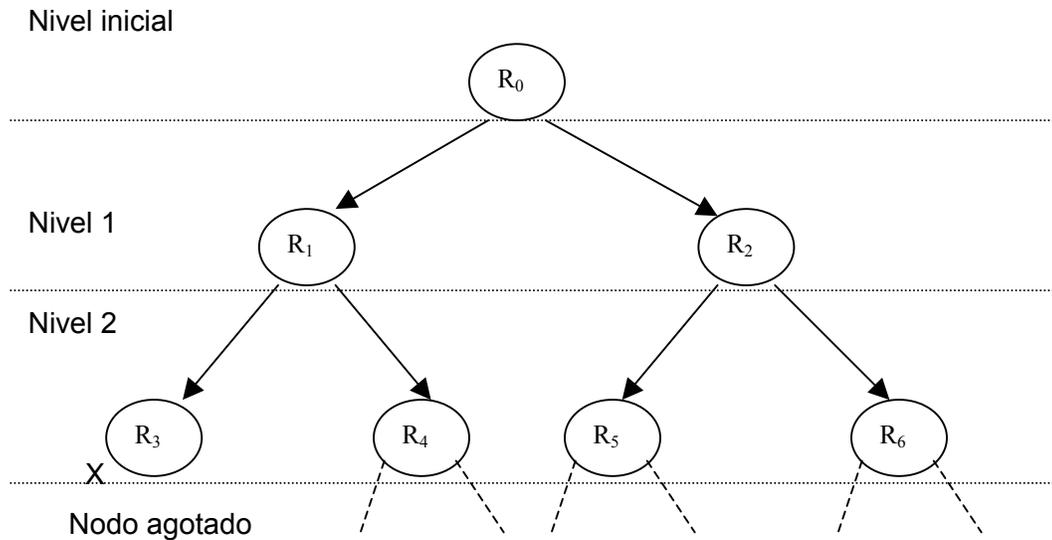


Figura 4.2.1: niveles en un proceso de partición-ramificación

Propiedad 4.2.1:

Dado un NLP no convexo cuya región factible es R_0 . Si se resuelve con un algoritmo del tipo ramificar y acotar sin reducción de cotas ni región factible, del proceso de partición y ramificación, se cumple para cada nivel $k \geq 0$: $R_0 = \prod_{\forall i \in \text{nivel}(k)} R_i$.

Demostración:

Sea R_0 la región factible de un NLP no convexo y sea \bar{R}_0 su relajación convexa.

A partir de una variable x_j se propone una primer partición de la región

$$R_1 = R_0 \cap [x^L, x_j^*]$$

$$R_2 = R_0 \cap [x_j^*, x^U]$$

$$R_1 \cap R_2 = (R_0 \cap [x_j^L, x_j^*]) \cap (R_0 \cap [x_j^*, x^U]) = R_0 \cap ([x_j^L, x_j^*] \cap [x_j^*, x^U]) = R_0$$

Supongo cierto para un nivel n-1: $R_0 = \prod_{\forall i \in \text{nivel}(n-1)} R_i$

Se demuestra para un nivel n: como $R_i \in \text{nivel}(n-1)$, si $R_i \neq \emptyset$ se realiza una partición

$R_i = R_{i1} \cup R_{i2} \quad \forall i \in \text{nivel}(n-1)$, y considerando una enumeración de las subregiones en

orden, se pueden tomar de a pares sucesivos, y nombrarlos R_{i1}, R_{i2} que provienen de una partición de una subregión i del nivel $n-1$. Luego

$$\prod_{\forall i \in \text{nivel}(n)} R_i = \prod_{\forall ij \in \text{nivel}(n)} R_{ij} = \prod_{\forall i \in \text{nivel}(n-1)} (R_{i1} \cup R_{i2}) = \prod_{\forall i \in \text{nivel}(n-1)} R_i = R_0 \quad \square$$

Propiedad 4.3.2:

Dado un NLP no convexo cuya región factible es R_0 y su relajación convexa \bar{R}_0 . Si se resuelve con el algoritmo del tipo ramificar y acotar sin reducción de cotas ni región factible, del proceso de partición y ramificación para cada nivel (k) se obtiene una cota inferior del valor objetivo óptimo de la forma:

$$\min_{i \in \text{nivel}(k)} \left\{ \min_{x \in R_i} f^c(x) \right\} \tag{4.1}$$

Demostración:

Sea nivel(k) un nivel del árbol generado por el proceso de resolución del algoritmo. Por simplicidad de notación llamemos R_1, R_2, \dots, R_m las regiones que definen los m nodos del nivel(k).

Sea \bar{R}_i la relajación convexa del conjunto $R_i, \forall i \geq 0$.

$$R_0 = \prod_{\forall i \in \text{nivel}(k)} R_i \subseteq \prod_{\forall i \in \text{nivel}(k)} \bar{R}_i \subseteq \bar{R}_0$$

Si el NLP no convexo es factible, sea x^* su solución óptima global con valor objetivo $f^c(x^*)$

$$f^c(x^*) = \min_{x \in R_0} f^c(x)$$

Como $x^* \in R_0$ y $R_0 = \prod_{\forall i \in \text{nivel}(k)} R_i$, $\exists i \in \text{nivel}(k) / x^* \in R_i \subseteq \bar{R}_i$ y $f^c(x^*) = \min_{x \in R_i} f^c(x)$

Por otro lado $\min_{i \in \text{nivel}(k)} \left\{ \min_{x \in R_i} f^c(x) \right\} \leq \min_{x \in R_i} f^c(x)$ luego $\min_{i \in \text{nivel}(k)} \left\{ \min_{x \in R_i} f^c(x) \right\} \leq f^c(x^*) \quad \square$

Corolario 4.3.1:

Dado un NLP no convexo cuya región factible es R_0 y su relajación convexa \bar{R}_0 . Si se resuelve con el algoritmo del tipo ramificar y acotar con reducción de cotas o de región factible, del proceso de partición y ramificación para cada nivel (k) se obtiene una cota inferior del valor objetivo óptimo de la forma:

$$\min_{i \in \text{nivel}(k)} \left\{ \min_{x \in R_i} f^c(x) \right\}$$

Demostración:

Sea nivel(k) un nivel del árbol generado por el proceso de resolución del algoritmo.

Si se aplica una estrategia de reducción, ésta debe trabajar de forma tal de eliminar regiones donde se obtienen soluciones con mayor valor objetivo (para un problema de mínimo) o regiones infactibles, luego

$$\min_{x \in R_1} f^c(x) = \min_{x \in R_1, \text{reducida}} f^c(x) \geq \min_{x \in \bar{R}_1, \text{reducida}} f^c(x)$$

Además esto asegura que para $x^* \in R_0$, solución óptima del problema que

$\exists i \in \text{nivel}(k) / x^* \in R_{i, \text{reducida}} \subseteq \bar{R}_{i, \text{reducida}}$ y análogamente a la propiedad anterior:

$$\min_{i \in \text{nivel}(k)} \left\{ \min_{x \in R_i, \text{reducida}} f^c(x) \right\} \leq \min_{x \in \bar{R}_i, \text{reducida}} f^c(x)$$

$$f^c(x^*) = \min_{x \in \bar{R}_i, \text{reducida}} f^c(x)$$

$$\text{luego } \min_{i \in \text{nivel}(k)} \left\{ \min_{x \in R_i, \text{reducida}} f^c(x) \right\} \leq f^c(x^*) \quad \square$$

Estos resultados dan una forma de obtener actualizaciones para la cota inferior de f^{LO} una vez comenzada la partición, luego de resolver todos los problemas convexificados de un mismo nivel, el mínimo de los valores objetivos asociados a las soluciones resultará una cota inferior mejor que la obtenida hasta el momento.

Si bien las restricciones de cotas para la función objetivo permiten realizar reducciones en las cotas de las variables que la definen, permiten realizar también reducciones de cotas sobre variables relacionadas.

En esta alternativa se combinan las reducciones que se obtienen de considerar tanto la cota inferior como la cota superior, donde su efectividad está sujeta a las funciones matemáticas que definen el problema.

Alternativa 4

Otro aspecto que puede estudiarse es la construcción de una restricción que acote el valor objetivo de la solución utilizando dualidad reproduciendo la idea planteada en algoritmo GOP (ver sección 2.7.5)

La condición A pedida en GOP impone condiciones sobre la naturaleza de la función lagrangeana a partir de una partición de las variables.

Si se definen los dos conjuntos de variables x e y considerando en y las variables w_i^{LO} y w_i^{UP} y en x el resto, al igual que el algoritmo GOP se comienza fijando el vector y , para luego resolver el problema resultante que en este caso es CP. Integrado un conjunto de variables el problema resultante debe ser convexo. No es posible implementar el procedimiento creado por Floudas de manera directa debido a que no se tiene en su totalidad la condición A. Fijando w_i^{LO} y w_i^{UP} el problema resultante es convexo, pero luego fijando las variables restantes y liberando w_i^{LO} y w_i^{UP} , el problema ahora definido tiene restricciones que no son convexas por tener términos cóncavos.

Luego es necesario realizar primero una subestimación para cada término cóncavo tcu en la reformulación. Nuevamente se recurre a la envolvente convexa.

$$\begin{aligned}
 w_L &= \lambda_L w^{LO} + (1 - \lambda_L) w^{UP} \\
 \lambda_L tcu(w^{LO}) + (1 - \lambda_L) tcu(w^{UP}) &\leq tcu(w_L) \\
 w_U &= \lambda_U w^{LO} + (1 - \lambda_U) w^{UP} \\
 \lambda_U tcu(w^{LO}) + (1 - \lambda_U) tcu(w^{UP}) &\leq tcu(w_U)
 \end{aligned}$$

Dada la función Lagrangeana:

$$L(xc, \lambda, xc, w_U, \delta_1, \mu_1, \delta_2, \mu_2, \mu_3) = f^c(xc, w) + \delta_1 h^c(xc, w) + \mu_1 g^c(xc, w) + \delta_2 (w - \lambda w_L - (1 - \lambda) w_U) + \mu_2 (\lambda w_L + (1 - \lambda) tcu(w_U)) + k^c(xc) + \mu_3 (w_L - w_U)$$

Se cumple que

$$L(xc, \lambda, w, w_L, w_U, \delta_1, \mu_1, \delta_2, \mu_2, \mu_3) \geq L(xc, \lambda, w, \lambda_L, \lambda_U, \delta_1, \mu_1, \delta_2, \mu_2, \mu_3)$$

donde

$$L(xc, \lambda(xc, \lambda_U, \delta_1, \mu_1, \delta_2, \mu_2, \mu_3)) = f^c(xc, w) + \delta_1 h^c(xc, w) + \mu_1 g^c(xc, w) + \delta_2 (w - \lambda (\lambda_L w^{LO} + (1 - \lambda_L) w^{UP}) - (1 - \lambda) (\lambda_U w^{LO} + (1 - \lambda_U) w^{UP})) + \mu_2 (\lambda (\lambda_L tcu(w^{LO}) + (1 - \lambda_L) tcu(w^{UP})) + (1 - \lambda) (\lambda_U tcu(w^{LO}) + (1 - \lambda_U) tcu(w^{UP}))) + k^c(xc) + \mu_3 (\lambda_L w^{LO} + (1 - \lambda_L) w^{UP} - (\lambda_U w^{LO} + (1 - \lambda_U) w^{UP}))$$

$$\text{Para todo } (xc, \lambda) \in X = \{xc, \lambda : xc^{LO} \leq xc \leq xc^{UP}, 0 \leq \lambda \leq 1\}$$

$$(w_L, w_U) \in W = \{w_L, w_U : w^{LO} \leq w_L \leq w^{UP}, w^{LO} \leq w_U \leq w^{UP}\}$$

Esta subestimación se mantiene convexa cuando se fijan xc y λ en función de las variables restantes y luego cuando se fijan w_L y w_U , se fijan λ_L y λ_U resultando también una función convexa en xc y λ , cumpliendo con la condición A.

El algoritmo aquí implementado, en cada iteración k fija las variables w_i^{LO} y w_i^{UP} definiendo el problema CP que al resolverse si tiene solución, proporciona valores para los multiplicadores de Lagrange δ^k y μ^k y su solución un valor particular para el vector xc^k .

Al igual que en el algoritmo GOP, si se aplica la Propiedad 2.7.5.1 para la función de Lagrange subestimada y una linealización de la función alrededor del punto xc^k se obtiene:

$$\text{Min}_{xc, \lambda \in X} L(xc, \lambda, w, w^{LO}, w^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \geq \text{Min}_{xc, \lambda \in X} L(xc, \lambda, w, \lambda^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \geq$$

$$\text{Min}_{xc, \lambda \in X} L(xc, \lambda, w, \lambda^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k}$$

para w^{LO} y w^{UP} fijos.

$$L(xc, \lambda, w, \lambda^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k} = L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) + \nabla_{\lambda} L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) (\lambda - \lambda^k)$$

ya que las únicas variables conectadas a w_L y w_U son λ_i , por las condiciones de Kuhn-

$$\text{Tucker } \nabla_{xc} L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) = 0$$

$$\nabla_{\lambda} L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) = (\lambda_U - \lambda_L) (\delta_2^k (w^{LO} - w^{UP}) + \mu_2^k (tcu(w^{UP}) - tcu(w^{LO})))$$

El problema a resolver es:

$$\begin{aligned} \underset{xc, \lambda \in X}{\text{Min}} L(xc, \lambda, w, \lambda^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k}^{lin} &= L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) + \\ \underset{xc, \lambda \in X}{\text{Min}} \nabla_{\lambda} L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) (\lambda - \lambda^k) \end{aligned}$$

cuya solución se obtiene fácilmente.

Siempre se verifica que $\lambda_L \geq \lambda_U$.

Si $(\delta_2^k (w^{LO} - w^{UP}) + \mu_2^k (tcu(w^{UP}) - tcu(w^{LO}))) \leq 0$ el mínimo del problema anterior se da en $\lambda_i = 0$.

Si $(\delta_2^k (w^{LO} - w^{UP}) + \mu_2^k (tcu(w^{UP}) - tcu(w^{LO}))) > 0$ el mínimo de la función lagrangeana resulta en $\lambda_i = 1$.

Se obtiene una subestimación para la solución del problema original, en términos de λ_L y λ_U que posteriormente puede llevarse en términos de w_L y w_U .

Para lograr una expresión más acabada de la linealización de la función lagrangeana se opera algebraicamente y se simplifican:

$$\begin{aligned} L(xc, \lambda(xc, \lambda^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k}^{lin} &= f^c(xc^k, w^k) + \\ \delta_2^k (w^k - \lambda^k (\lambda_L w^{LO} + (1 - \lambda_L) w^{UP})) - (1 - \lambda^k) (\lambda_U w^{LO} + (1 - \lambda_U) w^{UP})) &+ \\ \mu_2^k (\lambda^k \lambda_L tcu(w^{LO}) + (1 - \lambda_L) \lambda_L tcu(w^{UP}) + (1 - \lambda^k) \lambda_U tcu(w^{LO}) + (1 - \lambda^k) (1 - \lambda_U) tcu(w^{UP}) + k^c(xc^k)) &+ \\ + [\delta_2^k (-\lambda_L w^{LO} - (1 - \lambda_L) w^{UP}) + \lambda_U w^{LO} + (1 - \lambda_U) w^{UP}) + & \\ \mu_2^k (\lambda_L tcu(w^{LO}) + (1 - \lambda_L) tcu(w^{UP}) - \lambda_U tcu(w^{LO}) - (1 - \lambda_U) tcu(w^{UP}))] (\lambda - \lambda^k) \end{aligned}$$

reemplazando la igualdad

$$\begin{aligned} k^c(xc^k, w^k) &= -\lambda^k (tcu(w^{LO}) \lambda_L + (1 - \lambda_L) tcu(w^{UP})) - (1 - \lambda^k) (tcu(w^{LO}) \lambda_U + (1 - \lambda_U) tcu(w^{UP})) \\ \text{si } \mu_2^k &> 0 \end{aligned}$$

y simplificando resulta:

$$\begin{aligned} L(xc, \lambda(xc, \lambda^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k}^{lin} &= f^c(xc^k, w^k) + \\ \delta_2^k [w^k - w^{LO} (\lambda \lambda_L + (1 - \lambda) \lambda_U) - w^{UP} (1 - \lambda \lambda_L - (1 - \lambda) \lambda_U)] + \mu_2^k [(tcu(w^{UP}) - tcu(w^{LO})) (\lambda_U - \lambda_L) (\lambda - \lambda^k)] \end{aligned}$$

Para expresar la función en términos de w_L y w_U se usan las igualdades:

$$\lambda_L = \frac{(w_L - w^{UP})}{(w^{LO} - w^{UP})} \quad \text{y} \quad \lambda_U = \frac{(w_U - w^{UP})}{(w^{LO} - w^{UP})}$$

$$L(xc, \lambda(xc, w^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k}^{lin} = f^c(xc^k, w^k) +$$

$$\delta_2^k \left[w^k - w^{LO} \frac{\lambda w_L + (1-\lambda)w_U - w^{UP}}{w^{LO} - w^{UP}} - w^{UP} \left(1 - \frac{\lambda w_L + (1-\lambda)w_U - w^{UP}}{w^{LO} - w^{UP}} \right) \right]$$

$$+ \mu_2^k (tcu(w^{UP}) - tcu(w^{LO})) \frac{w_U - w_L}{w^{LO} - w^{UP}} (\lambda - \lambda^k)$$

Volviendo al problema a resolver:

$$\text{Min}_{xc, \lambda \in X} L(xc, \lambda, w, \lambda^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k}^{lin} = L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) +$$

$$\text{Min}_{xc, \lambda \in X} \nabla_{\lambda} L(xc^k, \lambda^k, w^k, \lambda_L, \lambda_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) (\lambda - \lambda^k)$$

Reemplazando por $\lambda = \frac{(w - w^{UP})}{(w^{LO} - w^{UP})}$ resulta:

$$L(xc, \lambda(xc, w^{LO}, \lambda^{UP}, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k) \Big|_{xc^k, \lambda^k}^{lin} = f^c(xc^k, w^k) -$$

$$\frac{(w - w^k)}{w^{LO} - w^{UP}} \left(\frac{\mu_2^k}{w^{LO} - w^{UP}} (w_L - w_U) (tcu(w^{UP}) - tcu(w^{LO})) + \delta_2^k \right)$$

Hasta ahora se consideraron las restricciones agrupadas en forma vectorial, para buscar el mínimo debe estudiarse cada una de ellas.

Como $w_{U_i} - w_{L_i} > 0$, si

$$\Delta_i = \delta_2^k (w_i^{LO} - w_i^{UP}) + \mu_2^k \frac{(tcu_i(w^{UP}) - tcu_i(w^{LO}))}{w_i^{LO} - w_i^{UP}} (w_{L_i} - w_{U_i}) \geq 0, \text{ para lograr el mínimo de la}$$

función se debe tomar $w = w^{UP}$ (igual condición ya obtenida anteriormente mirando el gradiente de la función de Lagrange respecto a las variables conectadas, $\lambda_i = 0$)

Como $\nabla_{\lambda} L(xc^k, \lambda^k, w^k, w_L, w_U, \delta_1^k, \mu_1^k, \delta_2^k, \mu_2^k, \mu_3^k)$ no es función de w_L, w_U , los calificadores de restricciones no producen particiones de la región factible, proporcionando una subestimación única en todos lados.

La solución de este problema es:

$$\mu_B = f^c(xc^k, w^k) + \sum_{i/\Delta_i \geq 0} \frac{(w_i^k - w_i^{UP})}{(w_i^{LO} - w_i^{UP})} \Delta_i + \sum_{i/\Delta_i < 0} \frac{(w_i^k - w_i^{LO})}{(w_i^{LO} - w_i^{UP})} \Delta_i$$

Se obtiene así en μ_B una cota inferior al valor objetivo óptimo de un problema que es una relajación de MP, luego es una cota inferior del valor objetivo óptimo de RP. Además se puede agregar a la resolución del problema MP la subestimación:

$z \geq \mu_B$ donde z es el valor objetivo del problema RP.

Los valores objetivos de f y f^c coinciden para cada solución del problema convexo CP.

4.4 Formulación de nuevos algoritmos de optimización global usando una forma diferente de partición

4.4.1 Introducción

El algoritmo estudiado en el capítulo 3 realiza las particiones de la región factible a partir del comportamiento de las variables que intervienen en restricciones con términos no convexo, por ejemplo teniendo en cuenta primero aquellas variables que provocan mayor error de aproximación en dichos términos.

Un proceso de ramificación diferente se tiene en cuenta aquí, donde se propone particionar la región factible a partir de la construcción de nuevas restricciones que reemplazan los términos cóncavos de la reformulación del problema, por funciones lineales.

Los cambios realizados en la forma de trabajar del algoritmo junto con una forma diferente de obtener particiones de la región factible, conducen a un mecanismo alternativo de búsqueda de la solución del problema, dando lugar a un nuevo algoritmo para problemas no convexos que obtiene la solución global en un número finito de pasos con una precisión prefijada.

Se evalúa su performance en algunos problemas test ya utilizados en los trabajos presentados en los capítulos 2 y 3.

4.4.2 Lineamientos generales del algoritmo

El algoritmo, en una etapa inicial resuelve una relajación convexa del problema reformulado CP y obtiene una cota inferior a partir del valor objetivo de su solución. Por otro lado dicha solución es un punto de partida conveniente para resolver MP ya que es un punto factible para el problema.

Si la solución de MP tiene valor objetivo cero, el punto es factible para RP y se toma como punto de partida para resolverlo. Como RP es un problema no convexo provee un óptimo

local y es una cota superior del valor objetivo de la solución buscada. El cálculo de la cota superior implica realizar particiones, de ser necesario, siguiendo el procedimiento de partición del algoritmo del capítulo 3, hasta llegar a una solución factible del problema original o equivalentemente del problema RP.

Una vez que el algoritmo logra una solución factible para el problema original, solución que dicho algoritmo obtiene generalmente en las primeras iteraciones, se comienza a implementar esta nueva forma de particionar la región factible. La partición se realiza definiendo una nueva restricción a partir de las restricciones no convexas activas en la solución factible del problema reformulado.

Luego en la etapa principal se realiza el proceso de ramificación y acotamiento ajustando la cota superior cada vez que se obtiene una nueva solución factible para el problema original, y la cota inferior luego de completar cada nivel en el árbol de ramificación.

Se requiere fijar una tolerancia que será utilizada para finalizar el algoritmo en el caso que se logre la desigualdad $f^U - f^L < \text{tolerancia}$ antes de haber testeado todas las subregiones generadas por el proceso de partición.

4.4.3 Procedimiento para la partición de regiones

Dada una solución factible del problema original, las particiones se realizan a partir de la construcción de una restricción que se incorpora al problema y se define de la forma que se detalla a continuación.

Sea (\bar{x}_c, \bar{w}) una solución del problema RP y sea $tcu_i(\bar{w}_i) + k^c_i(x_c, w) \leq 0$ una restricción activa en dicha solución tal que $w_{Li} < \bar{w}_i < w_{Ui}$.

Se define la recta $r_i(w_i)$ con pendiente $\frac{tcu_i(w_{Li}) - tcu_i(w_{Ui})}{w_{Li} - w_{Ui}}$, igual a la subestimación

$\lambda_i tcu_i(w_{Li}) + (1 - \lambda_i)tcu_i(w_{Ui})$, y que pasa por el punto \bar{w}_i .

Su ecuación es:

$$r_i : (tcu_i(w_{Li}) - tcu_i(w_{Ui}))\lambda_i + tcu_i(\bar{w}_i) - (tcu_i(w_{Li}) - tcu_i(w_{Ui}))\bar{\lambda}_i \quad (4.2)$$

con
$$\bar{w}_i = \bar{\lambda}_i w_i^{LO} + (1 - \bar{\lambda}_i)w_i^{UP}$$

$$0 \leq \bar{\lambda}_i \leq 1$$

o en términos de w

$$r_i(w_i) = \frac{tcu_i(w_{Li}) - tcu_i(w_{Ui})}{w_{Li} - w_{Ui}} w_i + tcu_i(\bar{w}_i) - \frac{tcu_i(w_{Li}) - tcu_i(w_{Ui})}{w_{Li} - w_{Ui}} \bar{w}_i \quad (4.3)$$

A partir de esta ecuación se proponen dos formas de particionar la región factible, generando dos algoritmos.

Algoritmo 1

Es posible generar dos nuevos subproblemas agregando a la región factible del problema convexo en uno la restricción convexa: $r_i(w_i) + k^c_i(xc, w) \leq 0$

y al otro la restricción: $r_i(w_i) + k^c_i(xc, w) \geq 0$ que resulta convexa sólo si $k^c_i(xc, w)$ es lineal.

Para obtener en cada nueva partición regiones que definan dos subproblemas convexos, una opción es trabajar con una reformulación del problema que tenga las restricciones que tienen términos cóncavos combinados con términos lineales. Bajo estas condiciones, a partir de una nueva restricción se particiona la región factible \bar{R}_0 en dos nuevas subregiones R_1, R_2 con la propiedad $R_1 \cup R_2 = \bar{R}_0$ y $R_1 \cap R_2 = \{r_i(w_i) + k^c_i(xc, w) = 0\}$

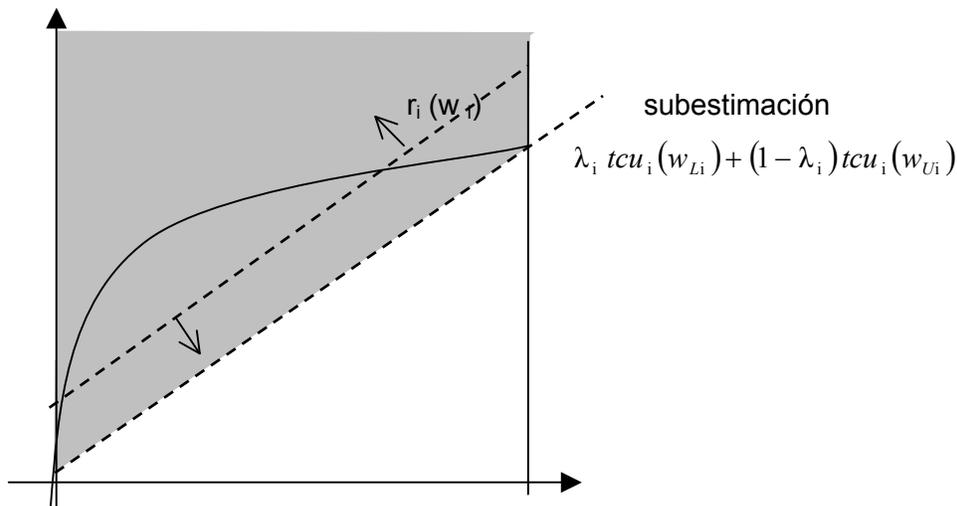


Figura 4.4.1: Término cóncavo con su envolvente convexa y la recta definida paralela

 Parte de la región factible definida a partir de la subestimación convexa

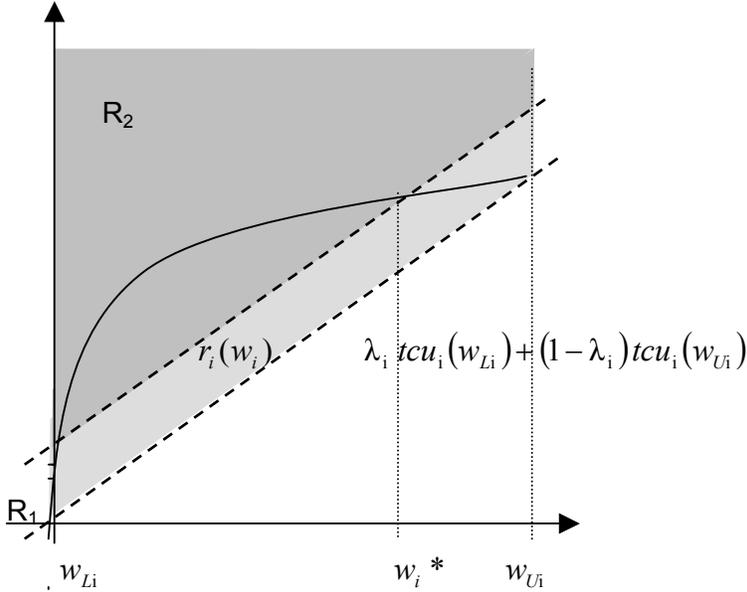


Figura 4.4.2: Partición de la región en dos subregiones a partir de la recta $r_i(w_i)$

Referencias: Región factible R_1 Región factible R_2

Elegida la restricción no convexa k , se definen los nuevos siguientes subproblemas:

$$\text{Min } f^c(xc, w)$$

$$\text{suj.a: } f^c(xc, w) \leq f^{\text{UP}} - \varepsilon$$

$$\bar{h}^1(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

$$r_k(w_k) \leq -k_k^c(xc, w)$$

$$w_k = \lambda_k w_k^{\text{LO}} + (1 - \lambda_k) w_k^{\text{UP}}$$

$$\left. \begin{array}{l} \lambda_i tcu_i(w_i^{\text{LO}}) + (1 - \lambda_i) tcu_i(w_i^{\text{UP}}) \leq -k_i^c(xc, w) \\ w_i = \lambda_i w_i^{\text{LO}} + (1 - \lambda_i) w_i^{\text{UP}} \end{array} \right\} \quad i = 1, \dots, k-1, k+1, \dots, J$$

(P1)

$$xc^{\text{LO}} \leq xc \leq xc^{\text{UP}}$$

$$0 \leq \lambda_i \leq 1 \quad i = 1, \dots, J$$

$$\text{Min } f^c(xc, w)$$

$$\text{suj.a: } f^c(xc, w) \leq f^{\text{UP}} - \varepsilon$$

$$\bar{h}^1(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

$$r_k(w_k) \geq -k_k^c(xc, w)$$

$$\left. \begin{aligned} \lambda_i \text{tcu}_i(w_i^{\text{LO}}) + (1 - \lambda_i) \text{tcu}_i(w_i^{\text{UP}}) &\leq -k_i^c(xc, w) \\ w_i &= \lambda_i w_i^{\text{LO}} + (1 - \lambda_i) w_i^{\text{UP}} \\ 0 &\leq \lambda_i \leq 1 \end{aligned} \right\} \quad i = 1, \dots, k-1, k, \dots, J \quad (\text{P2})$$

$$xc^{\text{LO}} \leq xc \leq xc^{\text{UP}}$$

En el proceso de ramificación el desarrollo por diferentes ramas pueden generar deferentes rectas para un mismo $\overline{w_k}$ debido a que proceso de reducción de cotas modifica las cotas de la variable y la pendiente de la recta está calculada en función de ellas.

El segundo subproblema tiene una región que en términos de la variable $\overline{w_k}$ resulta generalmente no factible. A los efectos de eliminar franjas infactibles del problema original, un mejor ajuste se realiza considerando 3 subproblemas en vez de dos en la partición, donde el segundo se reemplaza por dos nuevos subproblemas a partir de los rangos de variación de la variable $\overline{w_k}$ de la forma:

$$\left. \begin{aligned} [\overline{w_k^{\text{LO}}}, \overline{w_k}] &= [\overline{w_k^{\text{LO}}}, \overline{w_k}] \\ [\overline{w_k}, \overline{w_k^{\text{UP}}}] &= [\overline{w_k^{\text{calculado}}}, \overline{w_k^{\text{UP}}}] \end{aligned} \right\} \quad \text{si } \overline{w_k} \leq \overline{w_k^{\text{calculado}}} \quad (\text{4.4})$$

o

$$\left. \begin{aligned} [\overline{w_k^{\text{LO}}}, \overline{w_k}] &= [\overline{w_k^{\text{LO}}}, \overline{w_k^{\text{calculado}}}] \\ [\overline{w_k}, \overline{w_k^{\text{UP}}}] &= [\overline{w_k}, \overline{w_k^{\text{UP}}}] \end{aligned} \right\} \quad \text{si } \overline{w_k} \geq \overline{w_k^{\text{calculado}}}$$

Se reemplaza P2 por: P2 más la restricción de cota $w_k \in [\overline{w_k^{\text{LO}}}, \overline{w_k}]$.

Se define P3 como: P2 más la restricción de cota $w_k \in [\overline{w_k}, \overline{w_k^{\text{UP}}}]$.

En la figura 4.4.3 se muestran las tres subregiones que define la partición.

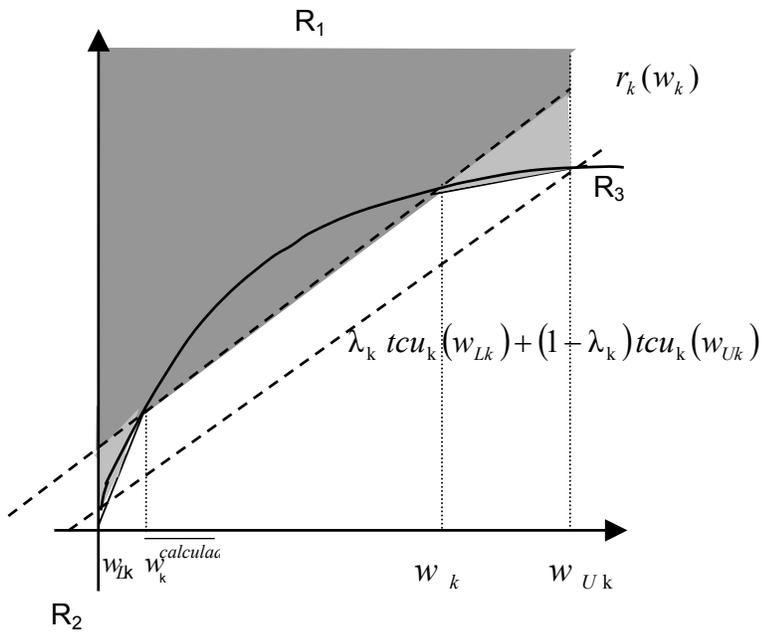


Figura 4.4.3: Gráfico de las tres subregiones definidas a partir de la partición

Referencias Región R₁ Regiones R₂ y R₃

El punto $\overline{w_k^{calculado}}$ en aquellos casos donde la función es compleja se calcula mediante un problema de optimización convexo definido como sigue:

$$\text{Min } w_k$$

$$\text{suj. a: } r_k(w_k) \leq tcu_k(w_k)$$

$$w_k = \lambda_k w_k^{LO} + (1 - \lambda_k) w_k^{UP}$$

$$0 \leq \lambda_k \leq 1$$

$$\text{si } \overline{w_k} \geq \overline{w_k^{calculado}}$$

(4.5)

$$\text{Max } w_k$$

$$\text{suj. a: } r_k(w_k) \leq tcu_k(w_k)$$

$$w_k = \lambda_k w_k^{LO} + (1 - \lambda_k) w_k^{UP}$$

$$0 \leq \lambda_k \leq 1$$

$$\text{si } \overline{w_k} \leq \overline{w_k^{calculado}}$$

Algoritmo 2

A partir de la recta (4.3) es posible definir otra partición diferente a la implementada en el algoritmo 1. En cada partición de la región factible se generan tres subregiones y sus respectivos subproblemas de la siguiente forma:

A partir de (4.4) se definen tres intervalos de variación de la variable que define el término no convexo. Para el primer subproblema resulta:

$$\left[\begin{array}{c} \bar{w}_k^L \\ \bar{w}_k^U \end{array} \right] = \left[\bar{w}_k, \bar{w}_k^{\text{calculado}} \right] \quad \text{si} \quad \bar{w}_k \leq \bar{w}_k^{\text{calculado}}$$

o

$$\left[\begin{array}{c} \bar{w}_k^L \\ \bar{w}_k^U \end{array} \right] = \left[\bar{w}_k^{\text{calculado}}, \bar{w}_k \right] \quad \text{si} \quad \bar{w}_k \geq \bar{w}_k^{\text{calculado}}$$

El subproblema resuelto en este intervalo además cuenta con el agregado de la restricción convexa $r_k(w_k) + k_k^c \leq 0$.

Luego:

$$\text{Min } f^c(xc, w)$$

$$\text{su.j.a: } f^c(xc, w) \leq f^{\text{UP}} - \varepsilon$$

$$\bar{h}^1(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

$$r_k(w_k) \leq -k_k^c(xc, w)$$

$$w_k = \lambda_k \bar{w}_k^L + (1 - \lambda_k) \bar{w}_k^U$$

(P1)

$$\left. \begin{array}{l} \lambda_i t c u_i(w_i^{\text{LO}}) + (1 - \lambda_i) t c u_i(w_i^{\text{UP}}) \leq -k_i^c(xc, w) \\ w_i = \lambda_i w_i^{\text{LO}} + (1 - \lambda_i) w_i^{\text{UP}} \end{array} \right\} \quad i = 1, \dots, k-1, k+1, \dots, J$$

$$xc^{\text{LO}} \leq xc \leq xc^{\text{UP}}$$

$$0 \leq \lambda_i \leq 1 \quad i = 1, \dots, J$$

Los otros subproblemas:

$$\text{Min } f^c(xc, w)$$

$$\text{su.j.a: } f^c(xc, w) \leq f^{\text{UP}} - \varepsilon$$

$$\bar{h}^1(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

(P2')

$$\left. \begin{array}{l} \lambda_i \text{tcu}_i(w_i^{\text{LO}}) + (1 - \lambda_i) \text{tcu}_i(w_k^{\text{-L}}) \leq -k_i^c(xc, w) \\ w_i = \lambda_i w_i^{\text{LO}} + (1 - \lambda_i) w_i^{\text{UP}} \end{array} \right\} \quad i = 1, \dots, J$$

$$xc^{\text{LO}} \leq xc \leq xc^{\text{UP}} \quad w_k^{\text{LO}} \leq w_k \leq w_k^{\text{-L}}$$

$$0 \leq \lambda_i \leq 1 \quad i = 1, \dots, J$$

$$\text{Min } f^c(xc, w)$$

$$\text{su.j.a: } f^c(xc, w) \leq f^{\text{UP}} - \varepsilon$$

$$\bar{h}^1(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

(P3)

$$\left. \begin{array}{l} \lambda_i \text{tcu}_i(w_k^{\text{-U}}) + (1 - \lambda_i) \text{tcu}_i(w_i^{\text{UP}}) \leq -k_i^c(xc, w) \\ w_i = \lambda_i w_i^{\text{LO}} + (1 - \lambda_i) w_i^{\text{UP}} \end{array} \right\} \quad i = 1, \dots, J$$

$$xc^{\text{LO}} \leq xc \leq xc^{\text{UP}} \quad w_k^{\text{-U}} \leq w_k \leq w_k^{\text{UP}}$$

$$0 \leq \lambda_i \leq 1 \quad i = 1, \dots, J$$

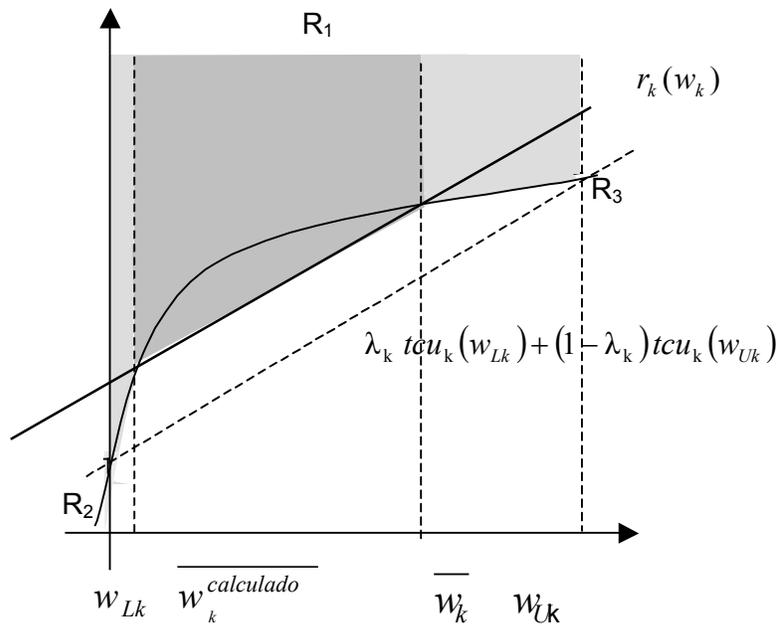


Figura 4.4.5: Gráfico de las tres subregiones para el algoritmo 2

Referencias Región R₁ Regiones R₂ y R₃

A diferencia del algoritmo 1, no es necesario ninguna reformulación adicional del problema ya que la única restricción que se incorpora es convexa. Otro aspecto ventajoso es que los tres subproblemas que se definen a partir de la partición tienen asociado una reducción del rango de variación de la variable elegida w y por lo tanto las reglas de reducción que acompañan la implementación del algoritmo actuarán en forma positiva en la reducción de la región factible (Figura 4.4.5)

En ambas propuestas se observa de cada partición una mejora en la subestimación de la restricción no convexa elegida para particionar.

Observando los ejemplos resueltos se puede ver que la solución óptima del problema se obtiene en las primeras iteraciones, esto permite implementar una regla empírica que consiste en resolver solamente el problema convexo en las regiones R_2 y R_3 sin resolver seguidamente MP. Esta regla evita una resolución de una gran cantidad de problemas no convexos MP.

En la implementación además se resuelve primeramente el problema convexo (P2), si es infactible no es necesario resolver los dos subproblemas en las regiones R_2 y R_3 .

A los fines de comparar esta forma de particionar con la partición propuesta en el capítulo 3 que se realiza a partir de los intervalos $[w_k^L, w_k^*]$ y $[w_k^*, w_k^U]$ se muestra en la figura 4.4.6 las dos subregiones que se obtienen en la misma situación. Si bien las subestimaciones de la función no convexa a partir de las nuevas cotas reducen parte de la región no factible, la nueva partición subestima mejor la región en términos de w_k eliminando más rápidamente la región infactible agregada por la subestimación

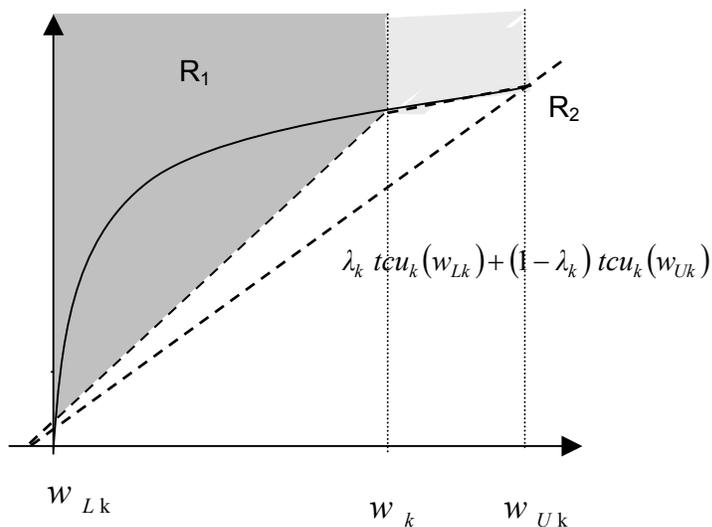


Figura 4.4.6: Subregiones obtenidas a partir de la partición convencional.

Referencia: Región R_1 Región R_2

Una vez realizada la partición se resuelve el problema convexo (P1) cuya solución es una solución factible para MP asociado en la subregión R_1 . Si dicho problema arroja una solución factible para RP se procede a optimizar RP mejorando la solución última y actualizando la última cota superior obtenida.

Nueva reducción de cotas

Una nueva reducción de cotas surge a partir de las nuevas restricciones incorporadas en el problema para particionar la región factible original.

Tanto $r_k(w_k) \leq -k_k^c(xc, w)$ para las dos alternativas como $r_k(w_k) \geq -k_k^c(xc, w)$ en el primer algoritmo permiten generar fácilmente cotas para la variable w_k y toda aquella variable que defina el término $k_k^c(xc, w)$.

Esta reducción se implementa en los algoritmos y se referencia en los ejemplos como alternativa 5 de reducción de cotas. Además de esta reducción se implementan las alternativas 2 y 3, donde en la 2 se tiene en cuenta la definición de las variables que se incorporan en la reformulación, y 3 las cotas de la función objetivo.

4.4.4 Consideraciones adicionales acerca de la reformulación

1) En la reformulación se consideran restricciones generales no convexas de la forma $\overline{tcu}(w) + \bar{k}^c(xc, w) \leq \bar{0}$. Existen situaciones donde se pueden resolver con expresiones matemáticas más convenientes que otras. Por ejemplo si un problema cuenta con las siguientes restricciones: $y \leq (x-1)^2$, $x \in [0,4]$ una restricción no convexa y restricciones de cotas para una variable.

En la reformulación se puede tratar de dos maneras:

a) Reemplazar $x' = x - 1$ resultando $y - x'^2 \leq 0$, $x' \in [-1,3]$. Si la variable interviene en otras restricciones el cambio de variable no genera problemas.

Cuando se plantea el problema convexo CP, la subestimación para el término cóncavo es: $-x'^2 \geq \lambda (-1)^2 + (1-\lambda) (-9)$ $x = \lambda (-1) + (1-\lambda) 3$ con $0 \leq \lambda \leq 1$

La restricción relajada: $y + \lambda (-1)^2 + (1-\lambda) (-9) \leq 0$

o en término de x' : $y - 2x' \leq 3$ junto con $x' \in [-1,3]$ definen un conjunto convexo que es la cápsula convexa de la región original.

b) $-x^2 + y + 2x - 1 \leq 0$, $x \in [0,4]$

donde $tcu(x) = -x^2$ y $k^c(x, y) = y + 2x - 1$

La subestimación para el término cóncavo es: $-x^2 \geq \lambda (0) + (1-\lambda) (-16)$

$x = \lambda (0) + (1-\lambda) 4$ con $0 \leq \lambda \leq 1$

La restricción relajada: $y + 2x - 1 + (1-\lambda) (-16) \leq 0$ o en términos de x

$$y - 2x \leq 13 \text{ junto con } x \in [0,4]$$

Si bien k^c es lineal, depende de la variable x ; cuando se subestima el término cóncavo, la subestimación de la restricción resulta mucho más relajada y por lo tanto menos conveniente.

En general cuando se tiene una restricción de la forma $y - x^2 + ax + b \leq 0$ es conveniente completar cuadrados sobre x logrando la expresión $y - (x - c)^2 + d \leq 0$ para hacer el cambio de variable $x' = x - c$. Esto garantiza una subestimación más ajustada.

2) Otra condición a tener en cuenta es definir un modelo escalado, es decir los rangos de las variables deben ser similares en magnitud, sino utilizar cambios de variables para lograr tal fin.

3) Si un término bilineal interviene en una desigualdad, éste puede reemplazarse por

$$x_1 x_2 = \frac{1}{2}(x_1^2 + x_2^2 - \alpha^2) \text{ con el agregado de } \alpha = x_1 - x_2 \quad x_1^L - x_2^U \leq \alpha \leq x_1^U - x_2^L$$

4.4.5 Un resultado teórico

A partir de la construcción de la recta $r_k(w_k)$ se define una franja F definida entre las rectas $\lambda_k tcu_k(w_{Lk}) + (1 - \lambda_k) tcu_k(w_{Uk})$, subestimación de tcu , y $r_k(w_k)$ que está formada por puntos factibles del problema original y puntos que no lo son y fueron agregados con la subestimación de tcu . Estos puntos pueden distinguirse por la siguiente propiedad:

$w_k \in F$ y es factible para el problema original si $\exists r(w_k)$ recta con igual pendiente a $r_k(w_k)$ que verifica $r_k(w_k) \leq r(w_k) \leq \lambda_k tcu_k(w_k^{LO}) + (1 - \lambda_k) tcu_k(w_k^{UP}) \quad / \quad tcu(w_k) - r(w_k) \leq 0$

De lo contrario para $w_k \in F$ y no factible para el problema original: $tcu(w_k) - r(w_k) > 0$

$\forall r(w_k)$ con igual pendiente a $r_k(w_k)$ que verifique

$$r_k(w_k) \leq r(w_k) \leq \lambda_k tcu_k(w_k^{LO}) + (1 - \lambda_k) tcu_k(w_k^{UP})$$

Luego es posible definir el siguiente NLP:

$$\text{Min } tcu(w_k) - r(w_k)$$

$$\text{suj.a: } f^c(xc, w) \leq f^{UP} - \varepsilon$$

$$\bar{h}^l(xc, w) = \bar{0}$$

$$\bar{g}^c(xc, w) \leq \bar{0}$$

$$r_k(w_k) \leq r(w_k) \leq \lambda_k tcu_k(w_k^{LO}) + (1 - \lambda_k) tcu_k(w_k^{UP})$$

$$\left. \begin{aligned} \lambda_i tcu_i(w_i^{LO}) + (1 - \lambda_i) tcu_i(w_i^{UP}) &\leq -k_i^c(xc, w) \\ w_i &= \lambda_i w_i^{LO} + (1 - \lambda_i) w_i^{UP} \\ 0 &\leq \lambda_i \leq 1 \end{aligned} \right\} \quad i = 1, \dots, J$$

$$xc^{LO} \leq xc \leq xc^{UP}$$

Si la solución óptima de este problema tiene valor objetivo asociado positivo está indicando que no hay soluciones con mejor valor objetivo que el hallado hasta el momento en la región estudiada, pudiendo desecharla.

El inconveniente de este planteo es que si bien es un problema con región factible convexa consiste en minimizar una función cóncava haciéndolo un problema no convexo.

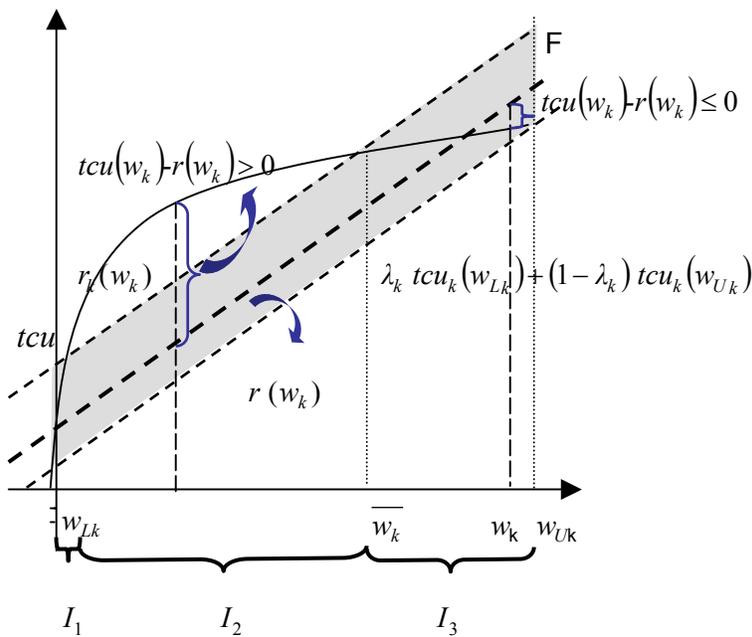


Figura 4.4.7: Identificación analítica de puntos factibles e infactibles en una franja F

Observando la figura 4.4.7 los puntos que están en la franja F y son factibles para el problema original son $w_k \in I_1 \cup I_3$ con la propiedad: $\exists r(w_k)$ recta con igual pendiente que $\lambda_k tcu(w_{Lk}) + (1 - \lambda_k) tcu(w_{Uk}) \leq r(w_k) \leq r_k(w_k)$ tal que $tcu(w_k) - r(w_k) \leq 0$.

Los puntos que están en la franja F y no son factibles para el problema original son $w_k \in I_2 : tcu(w_k) - r(w_k) > 0 \quad \forall r(w_k) / \lambda_k tcu(w_{Lk}) + (1 - \lambda_k) tcu(w_{Uk}) \leq r(w_k) \leq r_k(w_k)$

Ejemplo 4.4.1

a) Min $z = (x_1 - 1)^2 + (x_2 - 8)^2$

suja: $-x_1^2 + x_2 \leq 4$

$-1 \leq x_1 \leq 3$

$0 \leq x_2 \leq 8$

Mediante este ejemplo se muestra como trabaja esta nueva forma de partición y los subproblemas que se definen a partir de ella.

El problema está en la forma pedida por la reformulación RP. La reformulación CP resulta:

Min $z = (x_1 - 1)^2 + (x_2 - 8)^2$

suja: $-\lambda (x_1^L)^2 - (1-\lambda) (x_1^U)^2 + x_2 \leq 4$

$x_1 = \lambda x_1^L + (1-\lambda) x_1^U$

$-1 \leq x_1^L \leq x_1^U \leq 3$

$0 \leq \lambda \leq 1$

$-1 \leq x_1 \leq 3$

$0 \leq x_2 \leq 8$

CP

Se fija $x_1^L = -1, x_1^U = 3$ para resolver un problema convexo y se obtiene la solución que resulta la solución global: $x_1^* = 1,939 \quad x_2^* = 7,758 \quad z^* = 0,9395$.

La única restricción no convexa es activa en el punto y el valor de x_1^* no es ninguna de las cotas de la variable, luego es posible realizar una partición de la región factible a partir de la recta $r(x_1) = -2 x_1 + 0,1183$ definiendo dos subespacios opuestos a partir de la ecuación: $-2 x_1 + 0,1183 + x_2 = 4$

$$\text{Min } z = (x_1 - 1)^2 + (x_2 - 8)^2$$

$$\text{suj.a: } (x_1 - 1)^2 + (x_2 - 8)^2 \leq 0.939 - \varepsilon$$

$$-\lambda (x_1^L)^2 - (1-\lambda) (x_1^U)^2 + x_2 \leq 4$$

$$x_1 = \lambda x_1^L + (1-\lambda) x_1^U$$

$$-1 \leq x_1^L \leq x_1^U \leq 3$$

$$-2x_1 + x_2 \leq 3.8817$$

$$0 \leq \lambda \leq 1$$

$$-1 \leq x_1 \leq 3$$

$$0 \leq x_2 \leq 8$$

(P1)

$$\text{Min } z = (x_1 - 1)^2 + (x_2 - 8)^2$$

$$\text{suj.a: } (x_1 - 1)^2 + (x_2 - 8)^2 \leq 0.939 - \varepsilon$$

$$-\lambda (x_1^L)^2 - (1-\lambda) (x_1^U)^2 + x_2 \leq 4$$

$$x_1 = \lambda x_1^L + (1-\lambda) x_1^U$$

$$-1 \leq x_1^L \leq x_1^U \leq 3$$

$$-2x_1 + x_2 \geq 3.8817$$

$$0 \leq \lambda \leq 1$$

$$-1 \leq x_1 \leq 3$$

$$0 \leq x_2 \leq 8$$

(P2')

Este último problema se subdivide en dos nuevos subproblemas a partir del cálculo del otro punto que interseca la recta a la función no convexa $-x_1^2$ siguiente:

$$x_1 \text{ calculado} = x_1^L + x_1^U - x_1^* = 0.061$$

$$\text{Min } z = (x_1 - 1)^2 + (x_2 - 8)^2$$

$$\text{suj.a: } (x_1 - 1)^2 + (x_2 - 8)^2 \leq 0.939 - \varepsilon$$

$$-\lambda (x_1^L)^2 - (1-\lambda) (x_1^U)^2 + x_2 \leq 4$$

$$x_1 = \lambda x_1^L + (1-\lambda) x_1^U$$

$$-1 \leq x_1^L \leq x_1^U \leq 0.061$$

$$-2x_1 + x_2 \geq 3.8817$$

$$0 \leq \lambda \leq 1$$

$$0 \leq x_2 \leq 8$$

(P2)

$$\text{Min } z = (x_1 - 1)^2 + (x_2 - 8)^2$$

$$\text{suj.a: } (x_1 - 1)^2 + (x_2 - 8)^2 \leq 0.939 - \varepsilon$$

$$-\lambda (x_1^L)^2 - (1-\lambda) (x_1^U)^2 + x_2 \leq 4$$

$$x_1 = \lambda x_1^L + (1-\lambda) x_1^U$$

$$1.939 \leq x_1^L \leq x_1^U \leq 3$$

$$-2x_1 + x_2 \geq 3.8817$$

$$0 \leq \lambda \leq 1$$

$$0 \leq x_2 \leq 8$$

(P3)

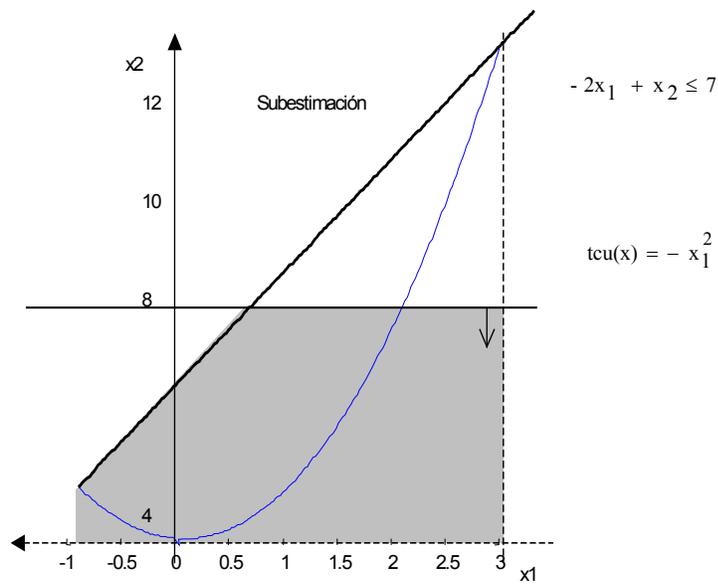


Figura 4.4.8: Envoltura convexa para un término cóncavo y región convexa que define

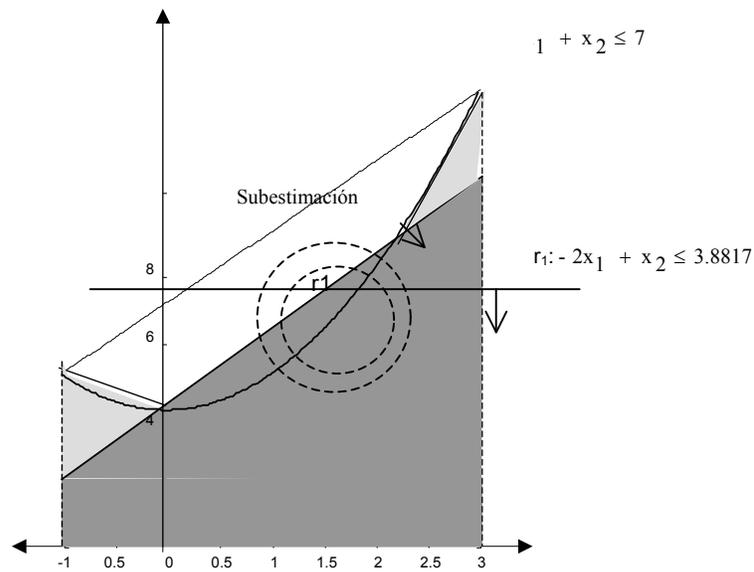


Figura 4.4.9: Regiones definidas a partir de una partición

Referencia: Regiones R_2 y R_3 Región R_1

b) Si se considera el mismo problema pero tomando los siguientes rangos de variación para las variables: $-1 \leq x_1 \leq 8$ $0 \leq x_2 \leq 16$

La subestimación resulta: $-7 x_1 + x_2 \leq 12$

Nuevas restricciones: $r(x_1) = -7 x_1 + x_2 \leq 2.1867$ y $-7 x_1 + x_2 \geq 2.1867$

$$x_1_{\text{calculado}} = x_1^L + x_1^U - x_1^* = 5.061$$

La partición de la región factible genera tres subregiones donde la región 3 con $-7 x_1 + x_2 \geq 2.1867$ y $5.061 \leq x_1 \leq 8$ resulta vacía.

4.4.6 Reglas de selección para la partición

Es posible implementar diferentes reglas para definir una partición de la región factible. Se definen aquí tres reglas alternativas.

Es necesario definir el conjunto de restricciones no convexas activas en la solución actual factible del problema reformulado:

$$H = \{i : tcu_i(\bar{w}_i) + k^c_i(\bar{x}^c) = 0 \text{ y } w_{Li} < \bar{w}_i < w_{Ui} \}$$

Regla 1: $i_0 = \arg \max_{i \in H} \{tcu_i(\bar{w}_i) - (\bar{\lambda}_i tcu_i(\bar{w}_{Li}) + (1 - \bar{\lambda}_i)tcu_i(\bar{w}_{Ui}))\}$

O en término de la recta

$$i_0 = \arg \max_{i \in H} \left\{ r_i(\bar{w}_i) - (\bar{\lambda}_i tcu_i(\bar{w}_{Li}) + (1 - \bar{\lambda}_i)tcu_i(\bar{w}_{Ui})) \right\}$$

Regla 2: En caso de tener las funciones cóncavas univariadas monótonas:

$$i_0 = \arg \max_{i \in H} \left\{ \frac{(tcu_i(\bar{w}_i) - (\bar{\lambda}_i tcu_i(\bar{w}_{Li}) + (1 - \bar{\lambda}_i)tcu_i(\bar{w}_{Ui})))}{abs(tcu_i(\bar{w}_{Li}) - tcu_i(\bar{w}_{Ui}))} \right\}.$$

$$\text{Regla 3: } i_0 = \arg \max_{i \in H} \left\{ \frac{(tcu_i(\bar{w}_i) - (\bar{\lambda}_i tcu_i(\bar{w}_{Li}) + (1 - \bar{\lambda}_i)tcu_i(\bar{w}_{Ui})))}{abs\left(tcu_i\left(\frac{w_{Li} + w_{Ui}}{2}\right) - (\bar{\lambda}_i tcu_i(\bar{w}_{Li}) + (1 - \bar{\lambda}_i)tcu_i(\bar{w}_{Ui}))\right)} \right\}.$$

4.4.7 Convergencia⁽¹⁹⁾

Definiciones: Sea $\Omega \subseteq \mathbb{R}^n$. S es una red para Ω si es una familia finita de conjuntos disjuntos de a pares tal que cubren todo Ω :

$$\forall s_i, s_j \in S \quad s_i \cap s_j = \emptyset \text{ con } i \neq j \quad \text{se cumple} \quad \Omega \subseteq \prod_{s_i \in S} s_i.$$

Una red S' es un refinamiento de S si está formada por pares disjuntos $s'_i \cap s'_j = \emptyset$ para todo $i \neq j$.

Para algún $s \in S$ se obtienen elementos $s'_i \in S' / s = \prod_i s'_i$ y $s \notin S'$

Es decir existe una partición de s que lo reemplazan, definiendo una nueva red S'.

Sea S_n una sucesión infinita de redes para $\Omega / \forall i \in \mathbb{N} \quad S_i$ es un refinamiento S_{i-1} .

Sea M_n una sucesión infinita de subconjuntos de $\Omega / M_i \in S_i$. M_n es un filtro para S_n si

$$\forall i \in \mathbb{N} \text{ se cumple } M_i \subseteq M_{i-1} \text{ y } M_\infty = \bigcap_{i \in \mathbb{N}} M_i \text{ es el límite del filtro.}$$

Lineamientos generales de algoritmos de partición y selección que resuelven problemas generales de la forma $\text{Min}\{f(x) : x \in \Omega\}$

Sean $\gamma \in \mathbb{R}$ y una red S para $\Omega \cap \{x / f(x) < \gamma\}$, la regla de selección que determina:

- 1) Un punto distinguido $w(M) \forall M \in S$.
- 2) Una subfamilia R de miembros calificados de S.
- 3) Un miembro distinguido $M^*(S)$ de R tal que $\forall x \in \prod_{s \in S \setminus R} s$ se tiene que $f(x) \geq \gamma$.

La regla rechaza regiones que no contiene el óptimo global eligiendo miembros calificados, para cada uno de esos miembros calcula un punto distinguido, por ejemplo utilizando un algoritmo de optimización local en la región prefijada, y un miembro distinguido para el refinamiento de la red posterior.

Una regla para particionar la región general se puede reformular como sigue:

- 1) Comenzar con una red S_1 para Ω_1 , $x_0 = \emptyset$, γ_0 es una cota superior para $f(\Omega)$.
- Sean $P_1 = S_1$, $k=1$, $\sigma_0 = \emptyset$.
- 2) Sea $\sigma_k = \{w(M) / M \in P_k\}$ conjunto de todos los puntos distinguidos.
 - 3) Sea x_k el punto en $\{x_{k-1}\} \cup \sigma_k / \gamma_k = f(x_k)$ es menor.

- 4) Determinar la familia R_k de miembros calificados de S_k (rechazar los no calificados, es decir aquellos que se pueden demostrar que no contienen mejor solución que γ_k)
- 5) Si $R_k = \phi$ terminar. El problema es infactible si $\gamma_k \geq \gamma_0$, de otra manera x_k es la solución óptima global.
- 6) Seleccionar el miembro distinguido $M_k = M^*(S_k) \in R_k$ y particionar M_k de acuerdo a la regla de brancheo prefijada. Sea P_{k+1} partición de M_k . En R_k reemplazar M_k por P_{k+1} obteniendo un nuevo refinamiento S_{k+1} .
- 7) $k \leftarrow k+1$ e ir al paso 2.

Un algoritmo de selección y partición es convergente si $\gamma^* = \inf f(\Omega) = \lim_{k \rightarrow \infty} \gamma_k$.

Una regla de selección es exacta si

- a) El valor de la función objetivo ínfimo de toda región que se mantiene calificada durante el proceso de solución completo es más grande que el valor objetivo óptimo global, es decir

$$\forall M \in \bigcap_{k=1}^{\infty} R_k \quad (\inf f(\Omega \cap M) \geq \gamma^*)$$

- b) El límite M_{∞} de todo filtro M_k es tal que $\inf f(\Omega \cap M_{\infty}) \geq \gamma^*$.

Teorema 4.4.4

Un algoritmo de selección y partición usando una regla de selección exacta es convergente

Demostración:

Por reducción al absurdo, suponga que el algoritmo finaliza en un punto $x \in \Omega$ con $f(x) < \gamma^*$. Sea $x \in M$ con $M \in R_n$ para algún $n \in \mathbb{N}$. Regiones no calificadas no pueden por hipótesis incluir puntos con mejores valores objetivos que el actual γ_k . Debe ser región calificada y como no puede mantenerse como tal, luego M debe ser particionada en alguna iteración $n' > n$. Así x pertenece a todo M_n en algún filtro $\{M_n\}$, así $x \in \Omega \cap M_{\infty}$. Por la condición (b) de regla exacta $f(x) \geq \inf f(\Omega \cap M_{\infty}) \geq \gamma^*$, absurdo que provino de suponer que el algoritmo finaliza en un punto con $f(x) < \gamma^*$ \square

Aquí no se habla de una convergencia finita, luego para las implementaciones se usa la idea de ϵ -optimalidad:

$x^* \in \Omega$ es ε - óptimo global si existen cotas m, M con $m \leq f(x^*) \leq M$ tal que $m \leq f(\bar{x}) \leq M$ y $M - m < \varepsilon$. Luego el algoritmo garantiza convergencia ε -global.

Este teorema garantiza la convergencia del algoritmo implementado en esta tesis por usar una regla de partición exacta.

4.4.8 Algoritmo

Paso de inicialización: Dado un problema en la forma P. Reformular el problema en la forma (RP) y luego definir MP y CP. Seleccionar un valor para el parámetro de mejora ε_1 y ε_2 , precisión con la que se obtiene la solución. Sea $R_{0,0}$ es el hiperrectángulo definido por las cotas en RP. Sean $iter=-1$, $niv=0$, $R = \{R_{0,0}\}$, $f^{\min} = \inf$.

Paso 1: Si $R = \Phi$, parar la solución $(\bar{x}^{OPT}, \bar{z}^{OPT}, \bar{\alpha}^{OPT}, \bar{\beta}^{OPT})$ es la solución óptima global del problema.

Si no, hacer $iter = iter+1$, $R = R \setminus \{R_{iter,niv}\}$.

Paso 2: Resolver el problema convexo CP en $R_{iter,niv}$. si no tiene solución y $R_{iter,niv} = R_{0,0}$ parar, el problema P es infactible.

Si CP no tiene solución y $R_{iter,niv} \neq R_{0,0}$ ir al paso 1.

Si no, CP tiene solución, hacer $f^{\min} = \min\{f^*, f^{\min}\}$. Si $R_{*,niv} \notin R$, actualizar la cota inferior $f^{LO} = f^{\min}$, y reasignar $f^{\min} = \inf$, y $niv = niv+1$.

Si $f^{UP} - f^{LO} < \varepsilon_2$ parar la solución $(\bar{x}^{OPT}, \bar{z}^{OPT}, \bar{\alpha}^{OPT}, \bar{\beta}^{OPT})$ es la solución óptima global del problema.

Paso 3: Tomar la solución óptima de CP como punto inicial para resolver MP en la región $R_{iter,niv}$ con un algoritmo local.

Si $(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}^*)$ es su solución con $f^* = \sum_{i=1}^J (w^*_{U_i} - w^*_{L_i}) = 0$, la solución hallada es factible para el problema (RP), ir al paso 4, sino ir al paso 7.

Paso 4: Resolver RP en la región $R_{iter,niv}$ a partir del punto $(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*, \bar{w}^*)$ agregando

la restricción $f^c(\bar{x}, \bar{z}) \leq f^{UP} - \varepsilon_1$. Sea $(\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*)$ la solución local y f^* el valor objetivo. Actualizar $(\bar{x}^{OPT}, \bar{z}^{OPT}, \bar{\alpha}^{OPT}, \bar{\beta}^{OPT}) = (\bar{x}^*, \bar{z}^*, \bar{\alpha}^*, \bar{\beta}^*)$, y la cota superior del valor objetivo $f^{UP} = f^*$. Definir $H = \{i : tcu_i(w_i^{OPT}) + k^c_i(xc^{OPT}) = 0 \text{ y } w_{Li} < w_i^{OPT} < w_{Ui}\}$

Ir al paso 5.

Paso 5: Si $H = \emptyset$, y $R = \emptyset$, la solución $(\bar{x}^{OPT}, \bar{z}^{OPT}, \bar{\alpha}^{OPT}, \bar{\beta}^{OPT})$ es la solución óptima global buscada del problema, parar.

Si $H = \emptyset$, y $R \neq \emptyset$ ir al paso 7. Si no ir al paso 6.

Paso 6: Aplicar regla de ramificación para elegir $k \in H$. Hacer $H = H \setminus \{k\}$. Particionar $R_{iter, niv}$ generando nuevas subregiones. Sea $R = R \cup \{R_{iter+1, niv}, R_{iter+2, niv}\}$ ir al paso 1.

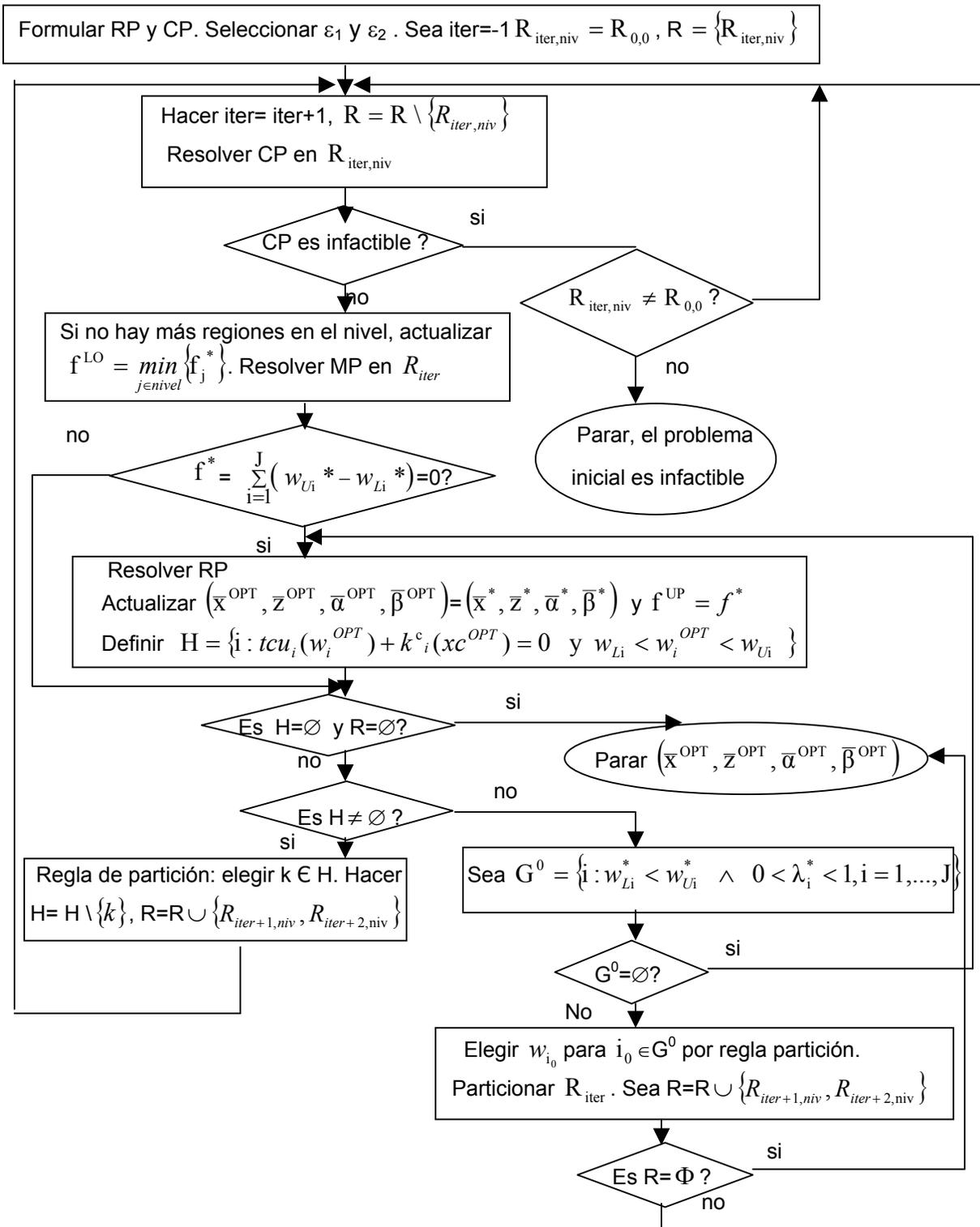
Paso 7: Definir $G^0 = \{i : w_{Li}^* < w_{Ui}^* \wedge 0 < \lambda_i^* < 1, i = 1, \dots, J\}$.

Si $G^0 = \emptyset$, (la solución actual es factible para RP) ir al paso 4, sino ir al paso 8.

Paso 8: Aplicar regla de ramificación eligiendo una variable w_{i_0} para $i_0 \in G^0$.

Particionar $R_{iter, niv}$ generando nuevas subregiones. Sea $R = R \cup \{R_{iter+1, niv}, R_{iter+2, niv}\}$ ir al paso 1.

Diagrama del algoritmo



4.4.9 Ejemplos

Ejemplo 4.4.2

Considerar el siguiente ejemplo presentado por Sahinidis y Grossmann (1991) y posteriormente Ryoo y Sahinidis en su trabajo ⁽¹³⁾.

$$\text{Min } f(\bar{x}) = -x_1 - x_2$$

$$\text{suje. a: } x_1 x_2 \leq 4$$

$$0 \leq x_1 \leq 6 \quad 0 \leq x_2 \leq 4$$

PNL está definido en dos variables con función objetivo lineal y una restricción bilineal.

El problema reformulado resulta:

$$\text{Min } f(\bar{x}) = -x_1 - x_2$$

$$\text{suje. a: } \alpha_1 = x_1 + x_2$$

$$\beta_1 = x_1 - x_2$$

$$\alpha_1^2 - \beta_1^2 \leq 16$$

$$0 \leq x_1 \leq 6$$

$$0 \leq x_2 \leq 4$$

$$0 \leq \alpha_1 \leq 10$$

$$-4 \leq \beta_1 \leq 6$$

(RP) es de 4 variables dos restricciones de igualdad lineales y una restricción no convexa, además de las restricciones de cotas.

Reemplazando cada término cóncavo por su envolvente convexa se obtiene el problema MP y fijando las variables β_{L1}, β_{U1} , el problema convexificado:

$$\text{Min } f(\bar{x}) = -x_1 - x_2$$

$$\text{suj.a : } \alpha_1 = x_1 + x_2$$

$$\beta_1 = x_1 - x_2$$

$$\alpha_1^2 - \lambda_1 \beta_{L1}^2 - (1 - \lambda_1) \beta_{U1}^2 \leq 16$$

$$\beta_1 = \lambda_1 \beta_{L1} + (1 - \lambda_1) \beta_{U1}$$

$$-4 \leq \beta_{L1} \leq \beta_{U1} \leq 6$$

$$0 \leq x_1 \leq 6 \quad 0 \leq x_2 \leq 4$$

$$0 \leq \alpha_1 \leq 10$$

$$-4 \leq \beta_1 \leq 6$$

$$0 \leq \lambda_1 \leq 1$$

El algoritmo resuelve primero el problema convexificado, fijando $\beta_{L1} = -4$, $\beta_{U1} = 6$ se obtiene como resultado:

$$x_1 = 6, \quad x_2 = 1.062, \quad f^c(\bar{x}) = -7.062 \quad \beta_1 = 4.938, \quad \alpha_1 = 7.062 \quad \lambda_1 = 0.106$$

Esto proporciona una cota inferior para el valor objetivo buscado: $f^{LO}(\bar{x}) = -7.062$

Como este punto no resulta (RP) factible, se busca un punto factible para el problema original resolviendo el problema MP que resulta de considerar β_{L1} , β_{U1} como variables y minimizar la función objetivo $\text{dif} = \sum_i \beta_{Ui} - \beta_{Li}$.

La solución resultante es:

$$x_1 = 6, \quad x_2 = 0.6667 \quad f(\bar{x}) = -6.6667 \quad \beta_1 = 5.333, \quad \alpha_1 = 6.6667 \quad \lambda_1 = 0, \quad \text{dif} = 0.$$

Luego es un punto factible para (RP) (y por lo tanto para el problema original) que el

algoritmo tomará como punto inicial para resolver (RP), resultando:

$$x_1 = 6, \quad x_2 = 0.6667 \quad f(\bar{x}) = -6.6667 \quad \beta_1 = 5.333, \quad \alpha_1 = 6.6667$$

Esto proporciona una cota superior para el valor objetivo buscado: $f^{UP} = -6.6667$.

Como $f^{UP} - f^{LO}(\bar{x}) = -6.6667 + 7.062 > \varepsilon$, se prosigue con el algoritmo.

Sea

$$H = \{i: tcu_i(w_i^{OPT}) + k^c_i(xc^{OPT}) = 0 \text{ y } w_{Li} < w_i^{OPT} < w_{Ui}\} = \{3: \alpha_1^{*2} - \beta_1^{*2} = 16, -4 < \beta_1^* \leq 6\}$$

La primer partición de la región factible se realiza a partir de la recta:

$$r: -2(\beta_1 - 6) - 29.778 = -2\beta_1 - 17.775$$

con igual pendiente a la envolvente convexa del término $-\beta_1^2$:
 $-\lambda_1\beta_{L1}^2 - (1-\lambda_1)\beta_{U1}^2 = \lambda_1(\beta_{U1}^2 - \beta_{L1}^2) - \beta_{U1}^2 = 20(\beta_1 - 6)/(-10) - 36 = -2\beta_1 - 24$ que pasa por el punto $(\beta_1^*, -\beta_1^{*2})$.

Si se implementa el algoritmo 1, las restricciones que se generan son:

$$\alpha_1^2 - 2\beta_1 \leq 33.775 \quad \text{restricción convexa}$$

$$\alpha_1^2 - 2\beta_1 \geq 33.775 \quad \text{restricción no convexa}$$

Como los problemas a resolver con las nuevas restricciones se desean que sean convexos, debe trabajarse de forma diferente la restricción no convexa. Anteriormente se planteó un procedimiento que consiste en reformular el problema (RP) con el cuidado de linealizar cada término no convexo que intervenga en las restricciones no convexas, luego las restricciones que se generan resultan lineales y por lo tanto ambas convexas. Una forma alternativa más eficiente de trabajar esta restricción se implementa, reemplazando

por una subestimación sólo cuando es necesario. Es decir, la primer restricción no se modifica ya que resulta convexa y en la segunda restricción en el momento de incorporarla, se subestiman los términos que sean necesarios para convertirla en una restricción lineal. En el ejemplo:

$$\lambda_2 \alpha_{L1}^2 + (1 - \lambda_2) \alpha_{U1}^2 - 2\beta_1 \geq 33.775$$

$$\alpha_1 = \lambda_2 \alpha_{L1} + (1 - \lambda_2) \alpha_{U1}$$

$$\alpha_{L1} = 0, \quad \alpha_{U1} = 10$$

resultando en término de α : $10\alpha_1 - 2\beta_1 \geq 33.775$ (Figura 4.4.10)

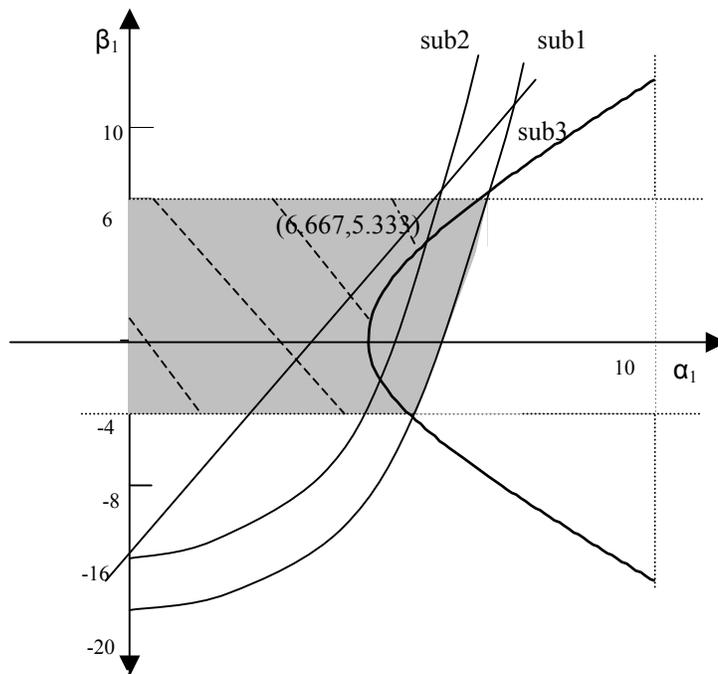


Figura 4.4.10: Región no convexa definida por $\alpha_1^2 - \beta_1^2 \leq 16$ y sus subestimaciones

Referencias: sub1: $\alpha_1^2 - 2\beta_1 \leq 40$ sub2: $\alpha_1^2 - 2\beta_1 \leq 33.775$ sub3: $10\alpha_1 - 2\beta_1 \geq 33.775$



Región original



Región convexificada

Los dos nuevos subproblemas a resolver son:

Subproblema 1:

$$\text{Min } f(\bar{x}) = -x_1 - x_2$$

$$\text{suj.a : } \alpha_1 = x_1 + x_2$$

$$\beta_1 = x_1 - x_2$$

$$f(x) \leq f^{UP} - \varepsilon_1 = -6.6667 - \varepsilon_1$$

$$\alpha_1^2 - 2\beta_1 \leq 33.775$$

$$0 \leq x_1 \leq 6$$

$$0 \leq x_2 \leq 4$$

$$0 \leq \alpha_1 \leq 10$$

$$-4 \leq \beta_1 \leq 6$$

$$0 \leq \lambda_1 \leq 1$$

Subproblema 2:

$$\text{Min } f(\bar{x}) = -x_1 - x_2$$

$$\text{suj.a : } \alpha_1 = x_1 + x_2$$

$$\beta_1 = x_1 - x_2$$

$$f(x) \leq f^{UP} - \varepsilon_1 = -6.6667 - \varepsilon_1$$

$$10\alpha_1 - 2\beta_1 \geq 33.775$$

$$0 \leq x_1 \leq 6$$

$$0 \leq x_2 \leq 4$$

$$0 \leq \alpha_1 \leq 10$$

$$-4 \leq \beta_1 \leq 6$$

$$0 \leq \lambda_1 \leq 1$$

Las técnicas de reducción de cotas implementadas modifican las siguientes cotas:

La cota superior de $f^{UP} = -6.6667$ permite actualizar $x_1^{LO} = 0.666$ y $x_2^{LO} = 0.666$

Luego la reducción propuesta a partir de la definición de las nuevas variables:

$$\alpha_1^{LO} = 1.332, \beta_1^{UP} = 5.333, \beta_1^{LO} = -3.334$$

Al resolver los subproblemas con estas reducciones de cotas resultan ambos infactibles por la restricción $f(x) \leq f^{UP} - \varepsilon_1$.

Para implementar el algoritmo se utilizó GAMS y se usaron el algoritmo GAMS/CONOP2 para resolver PNL en una PC Pentium 4 3.06 RAM DDR2 1G.

Las reducciones de cotas se incorporan al algoritmo mediante el uso de archivos de tipo "include" de manera de modificarse según el ejemplo.

El algoritmo se prueba en los siguientes ejemplos especificando tipo de reducciones de región factible en cada caso y en todos los casos se considera la regla 1 de selección para la partición que resultó la más efectiva.

Debe notarse que en cada ejemplo se mencionarán resultados obtenidos con otros algoritmos como el explicado en el capítulo anterior, considerando además del tiempo de ejecución, el número de iteraciones, a pesar que este último valor no resulta comparable con los obtenidos en este capítulo debido a que la forma de contar las iteraciones de los algoritmos es diferente. En los algoritmos aquí explicados una iteración resulta luego de resolver todos los problemas en un nivel de ramificación.

Ejemplo 4.4.3

El primer ejemplo es el problema propuesto por Himmelblau (1972):

$$\begin{aligned} \text{Max } f(\bar{x}) &= (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2 \\ &-4 \leq x_1 \leq 4 \quad -4 \leq x_2 \leq 4 \end{aligned}$$

Este problema es un problema con función objetivo no cóncava y restricciones de cotas para las variables. Tiene 4 mínimos globales, 1 máximo global y 4 puntos de silla.

La solución global es el punto: $\bar{x}^* = (0.3124484, -4)$ con $f^* = -308.8025$ y corresponde a un punto frontera.

Con el agregado de 4 variables nuevas, este problema puede reformularse como un problema restringido como sigue:

$$\begin{aligned}
 \text{(RP)} \quad & \text{Min } f = -c_1 - c_2 \\
 & \text{Suj a: } -c_1 + w_1^2 \leq 0 \\
 & \quad -c_2 + w_2^2 \leq 0 \\
 & \quad -w_1 + x_1^2 + x_2 - 11 \leq 0 \\
 & \quad -w_2 + x_2^2 + x_1 - 7 \leq 0 \\
 & \quad c_1 - w_1^2 \leq 0 \\
 & \quad c_2 - w_2^2 \leq 0 \\
 & \quad w_1 - x_1^2 - x_2 + 11 \leq 0 \\
 & \quad w_2 - x_2^2 - x_1 + 7 \leq 0 \\
 & \quad 0 \leq c_1 \leq 225 \quad 0 \leq c_2 \leq 169 \\
 & \quad -15 \leq w_1 \leq 9 \quad -11 \leq w_2 \leq 13 \\
 & \quad -4 \leq x_1 \leq 4 \quad -4 \leq x_2 \leq 4
 \end{aligned}$$

La reformulación cuenta con 4 restricciones de desigualdad convexas y 4 no convexas, además de las restricciones de cotas para las variables y una función objetivo lineal. Las restricciones no convexas están acompañadas de término lineales, condición suficiente para que en el algoritmo 1 se trabaje con el agregado de restricciones convexas en todos los casos.

Una reformulación subestimando los términos cóncavos de las 4 últimas restricciones se obtiene:

$$\text{Min } f = -c_1 - c_2$$

$$\text{Suj a: } -c_1 + w_1^2 \leq 0$$

$$-c_2 + w_2^2 \leq 0$$

$$-w_1 + x_1^2 + x_2 - 11 \leq 0$$

$$-w_2 + x_2^2 + x_1 - 7 \leq 0$$

$$c_1 - \lambda_1 w_1^{L^2} - (1 - \lambda_1) w_1^{U^2} \leq 0$$

$$c_2 - \lambda_1 w_2^{L^2} - (1 - \lambda_1) w_2^{U^2} \leq 0$$

$$w_1 - \lambda_1 x_1^{L^2} - (1 - \lambda_1) x_1^{U^2} - x_2 + 11 \leq 0$$

$$w_2 - \lambda_1 x_2^{L^2} - (1 - \lambda_1) x_2^{U^2} - x_1 + 7 \leq 0$$

$$0 \leq c_1 \leq 225 \quad 0 \leq c_2 \leq 169$$

$$-15 \leq w_1 \leq 9 \quad -11 \leq w_2 \leq 13$$

$$-4 \leq x_1 \leq 4 \quad -4 \leq x_2 \leq 4$$

$$0 \leq \lambda_1 \leq 1 \quad 0 \leq \lambda_2 \leq 1$$

Cuando están fijas $w_i^L, w_i^U, x_i^L, x_i^U$ el problema es convexo, de lo contrario definen un problema no convexo al que se le debe agregar las restricciones:

$$w_i^L \leq w_i^U, \quad x_i^L \leq x_i^U \quad \text{para } i=1,2$$

$$-15 \leq w_1^L, \quad w_1^U \leq 9$$

$$-11 \leq w_2^L, \quad w_2^U \leq 13$$

$$-4 \leq x_i^L, \quad x_i^U \leq 4 \quad \text{para } i=1,2$$

A partir de la definición de nuevas variables se pueden implementar reducciones de cotas como sigue (alternativa 2):

$$w_1 = x_1^2 + x_2 - 11$$

$$w_2 = x_2^2 + x_1 - 7$$

$$w_1^L = \max\{w_1^L, (\max(0, x_1^L, -x_1^U))^2 + x_2^L - 11\}$$

$$w_2^L = \max\{w_2^L, (\max(0, x_2^L, -x_2^U))^2 + x_1^L - 7\}$$

$$w_1^U = \min\{w_1^U, \max((x_1^L)^2, (x_1^U)^2) + x_2^U - 11\}$$

$$w_2^U = \min\{w_2^U, \max((x_2^L)^2, (x_2^U)^2) + x_1^U - 7\}$$

$$x_1^L = \max\{x_1^L, 7 - \max((x_2^L)^2, (x_2^U)^2) + w_2^L\}$$

$$x_1^U = \min\{x_1^U, 7 - (\max(0, x_2^L, -x_2^U))^2 + w_2^U\}$$

$$x_2^L = \max\{x_2^L, 11 - \max((x_1^L)^2, (x_1^U)^2) + w_1^L\}$$

$$x_2^U = \min\{x_2^U, 11 - (\max(0, x_1^L, -x_1^U))^2 + w_1^U\}$$

A partir de $c_j = w_j^2$ $j=1,2$:

$$c_j^L = \max\{c_j^L, (\max(0, w_j^L, -w_j^U))^2\} \quad j=1,2$$

$$c_j^U = \min\{c_j^U, \max((w_j^L)^2, (w_j^U)^2)\}$$

A partir de las cotas superior e inferior que se actualizan en el proceso iterativo se obtienen formulaciones para reducir cotas de c_1, c_2 (alternativa 3):

$$c_1^L = \max\{c_1^L, -f^U - c_2^U\}$$

$$c_2^L = \max\{c_2^L, -f^U - c_1^U\}$$

$$c_1^U = \min\{c_1^U, -f^L - c_2^L\}$$

$$c_2^U = \min\{c_2^U, -f^L - c_1^L\}$$

En la tabla 4.4.1 se muestra tiempos e iteraciones que le llevan a ambos algoritmos implementados con la regla 1 para el proceso de partición y con las reglas de reducción de rango con diferentes precisiones ϵ_1 , para garantizar la solución global del problema. La solución óptima se obtiene en la primer iteración lo que implica que el proceso de partición sigue inicialmente la nueva propuesta.

Por ejemplo la primer restricción que ambos algoritmos incorporan para realizar la partición es:

$$c_2 + 48 \lambda_2 \leq 94.096$$

con $\lambda_2 = (13 - w_2)/24$ resulta $c_2 - 2w_2 \leq 68.096$

particionando el intervalo de variación de la variable w_2 con la reducción de cotas.

ϵ_1	Algoritmo 1			Algoritmo 2		
	Iteraciones	Tiempo de CPU (segundos)	Nº regiones examinadas ^(*)	Iteraciones	Tiempo de CPU (segundos)	Nº regiones examinadas ^(*)
10^{-3}	6	1.6	12	5	0.6	5
10^{-4}	13	3.7	30	7	1.2	9
10^{-6}	14	7.6	35	6	1	6

Tabla 4.4.1: resultados del ejemplo 4.4.3 con $\epsilon_2 = 10^{-4}$

^(*)Nº de regiones donde se resuelve el problema convexo respectivo, sin contar las regiones eliminadas por reducción de cotas.

El algoritmo 2 sin reducción de cotas y con las precisiones $\epsilon_1 = 10^{-4}$ y $\epsilon_2 = 10^{-4}$ en las tolerancias sin reducción de cotas, requirió de 7 iteraciones estudiando 47 subregiones en un tiempo de CPU de 4.5 segundos.

Usando los valores $\epsilon_1 = 0.001$ y $\epsilon_2 = 10^{-4}$ en las tolerancias, se probó el rendimiento del algoritmo 1 con la alternativa 3 de reducción de cotas, el algoritmo requirió de 7 iteraciones estudiando 50 subregiones en un tiempo de CPU de 3.9 segundos.

Este ejemplo resuelto con el algoritmo del capítulo anterior, implementado con

reducciones de cotas para $\varepsilon_1=0.001$ llevó 31 iteraciones y 4.4 seg de tiempo de CPU.

Ejemplo 4.4.4

El siguiente problema es minimizar la función de Hartman (Dixon y Szegö, 1978).

$$f(x) = -\sum_{i=1}^m c_i \exp\left(-\sum_{j=1}^n a_{ij}(x_j - p_{ij})^2\right) \quad \text{para } 0 \leq x \leq 1$$

La función a optimizar corresponde a una familia de funciones definida en n variables, llamadas funciones de Hartman.

Para la reformulación del problema se realizan las siguientes sustituciones:

$$z_i = a_{ij}(x_j - p_{ij})^2$$

$$y_i = c_i \exp(-z_i) \quad i = 1, \dots, m$$

$$w_j = x_j^2 \quad j = 1, \dots, n$$

A partir de estas igualdades la reformulación del problema (RP) es:

$$\min f(x) = -\sum_{i=1}^m y_i$$

$$\text{suj. a: } y_i - c_i \exp(-z_i) \leq 0$$

$$-y_i + c_i \exp(-z_i) \leq 0 \quad i = 1, \dots, m$$

$$z_i = \sum_{j=1}^n a_{ij} w_j - 2a_{ij} x_j p_{ij} + a_{ij} p_{ij}^2$$

$$w_j - x_j^2 \leq 0 \quad j = 1, \dots, n$$

$$-w_j + x_j^2 \leq 0$$

$$x_j \leq 1 \quad j = 1, \dots, n$$

$$z_i, y_i, w_j, x_j \geq 0 \quad i = 1, \dots, m \quad j = 1, \dots, n$$

El problema reformulado tiene $m+n$ variables en términos no convexos y $2(n+m)$ variables en total. Cuenta con m restricciones lineales, $n+m$ desigualdades no convexas y $n+m$ convexas. Si se observan las restricciones no convexas de la reformulación, pueden observarse dos aspectos, el primero es que nuevamente los términos que acompañan al término no convexo en cada una de ellas, resultan lineales en todos los casos. El segundo aspecto se refiere a las restricciones no lineales donde interviene la función exponencial, para poder obtener el punto de intersección entre la restricción y la linealización que ambos algoritmos van incorporando para el proceso de ramificación se recurrirá a la resolución de un problema de optimización convexo. Los parámetros que definen el problema están dados en la tabla 4.4.2

i	A_{ij}			c_i	p_{ij}		
1	3.0	10	30	1.0	0.3689	0.1170	0.2673
2	0.1	10	35	1.2	0.4699	0.4387	0.7470
3	3.0	10	30	3	0.1091	0.8732	0.5547
4	0.1	10	35	3.2	0.03815	0.5743	0.8828

Tabla 4.4.2: Datos del ejemplo con $m=4$ y $n=3$.

Esta función test definida por 3 variables, $n=3$, se conoce como Hartman3 y cuenta con 3 mínimos. La solución óptima es $x^* = (0.11464, 0.55565, 0.85255)$ con un valor objetivo $f^* = -3.8627821478$

En la tabla 4.4.3 se muestran los resultados obtenidos para garantizar la solución global del algoritmo en sus dos variantes implementados con la regla 1 para el proceso de partición y alternativas de reducción 2, 3 y 4. La solución óptima se obtiene en la primer iteración y sólo se realizan particiones usando las restricciones antes explicadas.

ϵ_1	Algoritmo 1			Algoritmo 2		
	Iteraciones	Tiempo de CPU (segundos)	Nº regiones examinadas (*)	Iteraciones	Tiempo de CPU (segundos)	Nº regiones examinadas (*)
10^{-3}	4	1.56	16	4	1	9
10^{-4}	4	1.68	16	4	0.93	9
10^{-5}	4	1.71	16	4	1.02	9
10^{-6}	4	1.51	16	4	0.96	9

Tabla 4.4.3: resultados del ejemplo 4.4.4 con $\epsilon_2 = 10^{-4}$

(*)Nº de regiones donde se resuelve el problema convexo respectivo, sin contar las regiones eliminadas por reducción de cotas.

El algoritmo 2 sin reducciones de cotas requirió para este ejemplo con tolerancias $\epsilon_1 = 10^{-4}$ y $\epsilon_2 = 10^{-4}$, 8 iteraciones, 14.2 segundos de tiempo de CPU y examinó 225 regiones.

Incorporando reducciones a partir de las nuevas restricciones que se generan y mismas tolerancias, 7 iteraciones, 14.1 segundos de tiempo de CPU y generó 148 regiones.

Con el agregado de las alternativas 2 y 3 en reducción de cotas y mismas tolerancias, 4 iteraciones, 1.09 segundos de tiempo de CPU y resolvió 9 de las 15 regiones generadas,

De estos resultados se concluye que en este ejemplo la reducción que se realiza a partir de las restricciones agregadas no producen mejoras significativas.

El algoritmo 1 sin reducciones de cotas con las mismas tolerancias requirió de 7 iteraciones, 14.29 segundos de tiempo de CPU y examinó 177 regiones. En este caso también sólo se utilizan las restricciones para definir las particiones.

Este ejemplo resuelto con el algoritmo del capítulo anterior, implementado con reducciones de cotas para $\epsilon_1 = 0.001$ llevó 38 iteraciones y 8.15 seg de tiempo de CPU.

Ejemplo 4.4.5

Este ejemplo es un problema de diseño de un reactor simple (Figure 4.4.11).

$$\text{Min } f = -x_4$$

$$\text{Suj a: } x_1 - 1 + k_1 x_1 V_1 = 0$$

$$x_2 - x_1 + k_2 x_2 V_2 = 0$$

$$x_3 + x_1 - 1 + k_3 x_3 V_1 = 0$$

$$x_4 - x_3 + x_2 - x_1 + k_4 x_4 V_2 = 0 \quad (\text{P})$$

$$V_1^{0.5} + V_2^{0.5} \leq 4$$

$$0 \leq x_i \leq 1 \quad i = 1, \dots, 4$$

$$0 \leq V_i \leq 16 \quad i = 1, 2$$

con $k_1=0.09755988$, $k_2=0.99k_1$, $k_3=0.0391908$, $k_4=0.9k_3$.

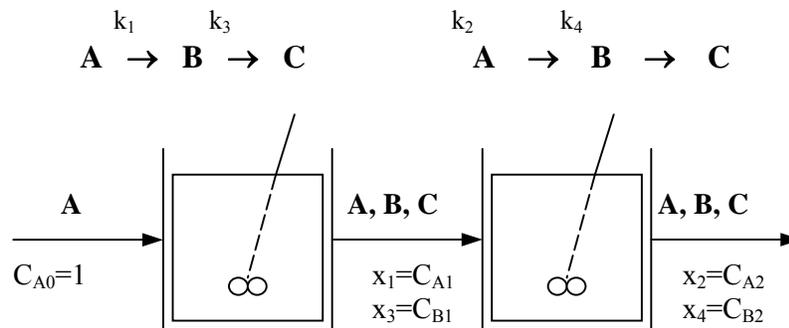


Figura 4.4.11: Diseño de una red de reactores

El problema está definido por 6 variables, una función objetivo lineal y 5 restricciones no convexas, 4 de igualdad con términos bilineales y una de desigualdad.

Se realizan las siguientes sustituciones:

$$\bar{V}_i = \frac{V_i}{16} \quad i = 1, 2 \quad \text{proceso de escalado}$$

$$x_9 = \bar{V}_1^{0.5} \quad x_{10} = \bar{V}_2^{0.5}$$

Los términos bilineales se reemplazan por :

$$\alpha_1 = x_1 + \bar{V}_1 \quad \beta_1 = x_1 - \bar{V}_1 \quad \alpha_2 = x_2 + \bar{V}_2 \quad \beta_2 = x_2 - \bar{V}_2$$

$$\alpha_3 = x_3 + \bar{V}_1 \quad \beta_3 = x_3 - \bar{V}_1 \quad \alpha_4 = x_4 + \bar{V}_2 \quad \beta_4 = x_4 - \bar{V}_2$$

$$x_5 = \frac{1}{4}(\alpha_1^2 - \beta_1^2) \quad x_6 = \frac{1}{4}(\alpha_2^2 - \beta_2^2) \quad x_7 = \frac{1}{4}(\alpha_3^2 - \beta_3^2) \quad x_8 = \frac{1}{4}(\alpha_4^2 - \beta_4^2)$$

Reformulación del problema:

$$\text{Min } f = -x_4$$

$$\text{Suj a: } x_1 - 1 + k_1 16x_5 = 0$$

$$\alpha_1 = x_1 + \bar{V}_1$$

$$\beta_1 = x_1 - \bar{V}_1$$

$$x_5 - \frac{1}{4}(\alpha_1^2 - \beta_1^2) \leq 0$$

$$\frac{1}{4}(\alpha_1^2 - \beta_1^2) - x_5 \leq 0$$

$$x_2 - x_1 + k_2 16x_6 = 0$$

$$\alpha_2 = x_2 + \bar{V}_2$$

$$\beta_2 = x_2 - \bar{V}_2$$

$$x_6 - \frac{1}{4}(\alpha_2^2 - \beta_2^2) \leq 0$$

$$\frac{1}{4}(\alpha_2^2 - \beta_2^2) - x_6 \leq 0$$

$$x_3 + x_1 - 1 + k_3 16x_7 = 0$$

$$\alpha_3 = x_3 + \bar{V}_1$$

$$\beta_3 = x_3 - \bar{V}_1$$

$$x_7 - \frac{1}{4}(\alpha_3^2 - \beta_3^2) \leq 0.$$

$$\frac{1}{4}(\alpha_3^2 - \beta_3^2) - x_7 \leq 0$$

$$x_4 - x_3 + x_2 - x_1 + k_4 16x_8 = 0$$

$$\alpha_4 = x_4 + \bar{V}_2$$

$$\beta_4 = x_4 - \bar{V}_2$$

$$x_8 - \frac{1}{4}(\alpha_4^2 - \beta_4^2) \leq 0$$

$$\frac{1}{4}(\alpha_4^2 - \beta_4^2) - x_8 \leq 0$$

$$x_9 - \bar{V}_1^{0.5} \leq 0$$

$$\bar{V}_1^{0.5} - x_9 \leq 0$$

$$x_{10} - \bar{V}_2^{0.5} \leq 0$$

$$\bar{V}_2^{0.5} - x_{10} \leq 0$$

$$x_9 + x_{10} \leq 1$$

$$0 \leq x_i \leq 1 \quad i = 1, \dots, 10$$

$$0 \leq V_i \leq 1 \quad i = 1, 2$$

$$0 \leq \alpha_i \leq 2 \quad i = 1, \dots, 4$$

$$-1 \leq \beta_i \leq 1 \quad i = 1, \dots, 4$$

El problema reformulado (RP) cuenta con 20 variables y 35 restricciones más las de cotas, 12 restricciones lineales, 10 restricciones de desigualdad no convexas y 3 restricciones de desigualdad convexas. 10 variables intervienen en restricciones no convexas.

Se implementaron reducciones de cotas para las nuevas variables a partir de sustituciones realizadas (alternativa 2), reducciones a partir de la actualización de las cotas del valor objetivo (alternativa 3), también se usaron las restricciones lineales del modelo y reducciones a partir de las nuevas restricciones que se agregan (alternativa 4). Se implementaron las subestimaciones de Mc. Cormick para los 4 términos bilineales del modelo solamente en el algoritmo 1. Se consideró la regla 1 como regla de selección para realizar la partición con la variante de priorizar primero la partición sobre las restricciones no convexas de la forma $\overline{tcu}(w) + \bar{k}^c(xc, w) \leq \bar{0}$ donde $\bar{k}^c(xc, w)$ resulta lineal, la cual mejoró sustancialmente los tiempos.

La solución óptima del problema es el vector $\mathbf{x}^*=(0.7715149717, 0.5169987806, 0.2041928334, 0.3888083092, 3.03558314, 5.09705646)$ con valor objetivo óptimo igual a $f^*=-0.3888083092$. Cuenta con dos mínimos locales: $\mathbf{x}_1=(0.390, 0.390, 0.375, 0.375, 16, 0)$ y $\mathbf{x}_2=(1, 0.393, 0, 0.388, 0, 16)$ cuyos valores objetivos están muy cerca del de la solución global, para el último punto la $f(\mathbf{x}^*)-f(\mathbf{x}_2) < 0.0001$ resultando un problema test de complejidad.

En la tabla 4.4.4 se dan los resultados obtenidos a partir de diferentes precisiones. Una primer solución factible se obtiene en la etapa inicial, por lo tanto el algoritmo comienza con el proceso de partición propuesto en la sección anterior y no necesita utilizar la partición convencional. La solución óptima se obtiene en la primer iteración.

ϵ_1	Algoritmo 1			Algoritmo 2		
	Iteraciones	Tiempo de CPU (segundos)	Nº regiones examinadas ^(*)	Iteraciones	Tiempo de CPU (segundos)	Nº regiones examinadas ^(*)
10^{-3}	17	11.78	99	6	1.11	16
10^{-4}	18	11.3	96	6	1.29	16
10^{-5}	15	10.06	84	6	1.25	16
10^{-6}	12	8.55	64	6	1.34	16

Tabla 4.4.4: resultados del ejemplo 4.4.5 con $\epsilon_2= 10^{-4}$.

^(*)Nº de regiones donde se resuelve el problema convexo respectivo, sin contar las regiones eliminadas por reducción de cotas.

El algoritmo 1 implementado sin reducciones, con las subestimaciones y sobreestimaciones de Mc Cormick, con tolerancias de $\epsilon_1=10^{-4}$ y $\epsilon_2= 10^{-4}$, finalizó luego de 7 iteraciones 60 segundos de tiempo de CPU inspeccionando 360 subregiones.

Considerando las reducciones a partir de las nuevas restricciones, con tolerancias de $\epsilon_1=10^{-4}$ y $\epsilon_2= 10^{-4}$, realizó 8 iteraciones 49.6 segundos de tiempo de CPU definiendo 362 subregiones pero 80 fueron eliminadas sin resolver un problema convexo.

Este ejemplo resuelto con el algoritmo del capítulo anterior, implementado con reducciones de cotas para $\epsilon_1=10^{-4}$ fue resuelto con diferentes reglas para la selección de la variable a particionar y la más eficiente finalizó en 117 iteraciones y 13.77 seg de tiempo de CPU.

Manousiouthakis y Sourlas (1992) le demandó 7950 nodos, posteriormente a Manaras y Floudas 299 iteraciones en 20 seg. CPU trabajando en Workstation HP-730; Byne y Bogle 26 iteraciones en 163 seg. CPU trabajando con Matlab en IBM RS6000. Ryoo y Sahinidis en su trabajo para $\epsilon_1=10^{-6}$ inspeccionan 413 nodos y requieren de 205 seg de tiempo de CPU de una Sun SPARC station 2. Se estudian algunas variantes donde implementan subestimaciones como la propuesta por Mc. Cormick para términos bilineales logrando reducir significativamente a 91 nodos y requieren de 30 seg de tiempo de CPU

Ejemplo	Con reducción de cotas $\epsilon_1=0.0001$		Sin reducción de cotas $\epsilon_1=0.0001$	
	Tiempo CPU	Nº subregiones examinadas	Tiempo CPU	Nº subregiones
4.4.3	1.2	9	4.5	47
4.4.4	0.93	9	14.2	225
4.4.5	1.29	16	60	360

Tabla 4.4.5: Tabla comparativa del algoritmo implementado con y sin reducción de cotas.

En la tabla 4.4.5 se muestran resultados del nuevo algoritmo para los diferentes ejemplos implementado con y sin reducción en las cotas. Se puede observar la ventaja de incluir

reglas de reducción de cotas, reduciendo número de subregiones y tiempo de ejecución.

En la tabla 4.4.6 se muestran las dimensiones de los problemas originales y sus reformulaciones equivalentes. Si bien el problema reformulado es de mayor dimensión, el problema original generalmente tiene términos altamente no convexos que son desagregados y simplificados en la reformulación.

Cantidad	Ejemplo 4.3.3		Ejemplo 4.3.4		Ejemplo 4.3.5	
	PI	PR	PI	PR	PI	PR
VARIABLES CONTINUAS	2	6	3	14	6	20
VARIABLES BINARIAS	0	0	0	0	0	0
VARIABLES EN TÉRMINOS NO CONVEXOS	2	4	7		6	10
Función objetivo	No convexa	Lineal	No convexa	Lineal	Lineal	Lineal
Restricciones de igualdad lineales	0	0	0	4	0	12
Restricciones de igualdad no lineales	0	0	0	–	4	–
Restricciones de desigualdad convexas (sin considerar rest. de cotas)	0	4	0	7	0	3
Restricciones de desigualdad no convexas	0	4	0	7	1	12

Tabla 4.4.6: Dimensiones del problema original y su reformulación

Referencias: PI: problema inicial

PR: problema reformulado

4.4.7 Conclusiones

El procedimiento explicado en este capítulo es un algoritmo en sus dos variantes, del tipo ramificar y reducir para resolver problemas no convexos con una convergencia ε -finita. Se adoptan las reformulaciones del problema no convexo seguidas en el capítulo anterior trabajando con los problemas: RP, MP no convexos y CP convexo. RP es una reformulación no convexa equivalente al original desagregando las funciones del modelo en suma de términos simples hasta obtener un problema con restricciones lineales y restricciones con a lo sumo un término cóncavo tcu . MP es la reformulación no convexa cuyo objetivo es obtener un punto factible del problema original, reemplazando los términos cóncavos por la reformulación $\lambda tcu(w_L) + (1-\lambda) tcu(w_U)$ con w_L y w_U nuevas variables y minimizando la función objetivo $\sum (w_{Uj} - w_{Lj})$.

CP es una reformulación convexa lograda a partir de (RP) fijando w_L y w_U .

Se implementa una nueva forma de partición de la región factible generando una nueva restricción a partir de las restricciones no convexas del problema reformulado. En cada partición se definen tres nuevas subregiones y sus subproblemas asociados, con la característica que el problema con región mayor con la ayuda de la restricción

$f(x) \leq f^{UP} - \varepsilon_1$ se torna más rápidamente infactible y mayores porciones de regiones son eliminadas en el proceso. Se pudo observar que para los ejemplos resueltos en ningún caso fue necesario recurrir a otro proceso de partición que no sea el nuevo aquí propuesto.

De sus dos variantes resultó más eficiente la segunda propuesta la ventaja que tiene esta propuesta frente a la primera es que las tres restricciones que se generan en cada partición permiten hacer mayores reducciones en las subregiones asociadas.

El proceso de reducción de cotas y de región factible es fundamental para que el algoritmo arribe a la solución con una precisión prefijada anticipadamente, de forma más eficiente en tiempo y número de iteraciones en todos los casos. El proceso de reducción permite eliminar subregiones sin llegar a resolver el problema de optimización asociado.

La incorporación de la restricción $f(x) \leq f^{UP} - \varepsilon_1$ a la reformulación es de gran utilidad para eliminar subregiones.

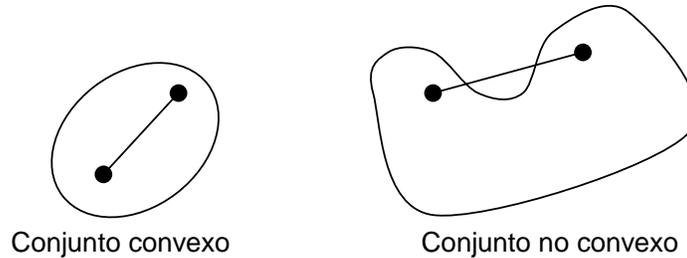
Se pueden destacar dos factores que deben ser motivo de estudios futuros para los algoritmos determinísticos del tipo partición, ramificación y reducción como el aquí

presentado: primero lograr reformulaciones del problema que no permitan un aumento en gran medida el tamaño del problema original. El segundo factor, ya dicho, es que el mayor esfuerzo sigue siendo el trabajo que el algoritmo debe realizar para garantizar que la solución a la que arriba es la óptima global del problema.

5. Apéndice

Convexidad de conjuntos

Definición 1: Un conjunto de puntos S se dice que es *convexo* si el segmento de línea que une dos puntos cualesquiera de S también pertenece al conjunto. Esto es, si $x, y \in S$, entonces, $\alpha \cdot x + (1 - \alpha) \cdot y \in S$, para todo $0 \leq \alpha \leq 1$.



Propiedad 1.1: La intersección de conjuntos convexos, es otro conjunto convexo.

Conjuntos distinguidos

Definición 2: La intersección de un número finito de semiespacios cerrados en \mathfrak{R}^n de la forma $\{x : c^t x \leq b, x \in \mathfrak{R}^n\}$ se llama politopo. Un politopo acotado es un poliedro.

Ambos conjuntos por propiedad 1.1 son convexos.

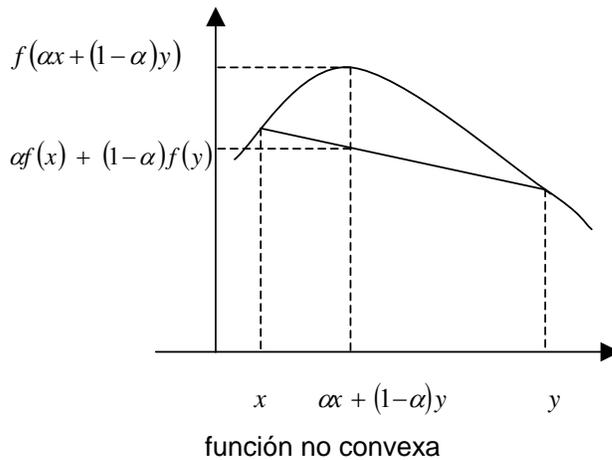
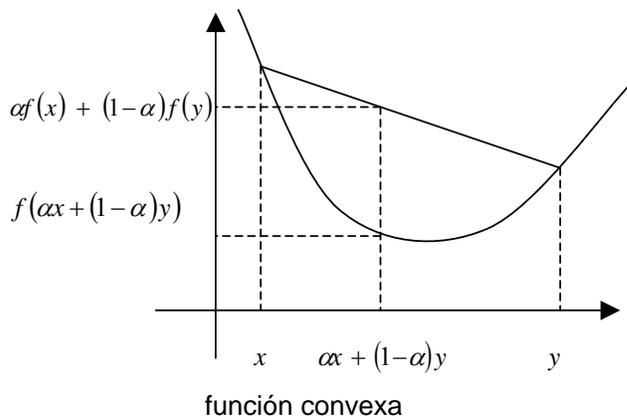
Definición 3: Un k -simplex en V es la cápsula convexa de $k+1$ puntos afínmente independientes de V .

Concavidad y Convexidad de funciones

Definición 4: Sea $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$

$$(1) \quad f \text{ es convexa si } f(\alpha \cdot x + (1 - \alpha) \cdot y) \leq \alpha \cdot f(x) + (1 - \alpha) \cdot f(y) \\ \forall x, y \in \mathfrak{R}^n, \quad 0 \leq \alpha \leq 1.$$

(2) f es cóncava si $f(\alpha \cdot x + (1-\alpha) \cdot y) \geq \alpha \cdot f(x) + (1-\alpha) \cdot f(y)$
 $\forall x, y \in \mathfrak{R}^n, 0 \leq \alpha \leq 1$



Si la desigualdad anterior es estricta, se dice que la función es *estrictamente convexa*, para $0 < \alpha < 1$.

Toda función que no cumple con la condición (1) o (2) no es *ni cóncava ni convexa*.

Propiedad 2.1: si f es cóncava, $-f$ es convexa.

Propiedad 2.2: una función lineal es tanto cóncava como convexa.

Propiedad 2.3: la suma de funciones convexas es una función convexa.

Lema 1: Si f es convexa, entonces el conjunto $S_K = \{x / f(x) \leq K\}$ es convexo, $\forall K \in \mathfrak{R}$

Similarmente, si f es cóncava, entonces el conjunto $S_K = \{x / f(x) \geq K\}$ es convexo, $\forall K \in \mathfrak{R}$

Me falta una definición :encuentro la definición de semicontinua inferior no semicontinua

Definición 5: Sean $f : X \rightarrow \mathfrak{R}$, y $a \in X$, f es semicontinua inferior en a si para cada $K < f(a)$ existe un entorno U de a tal que $K < f(U)$.

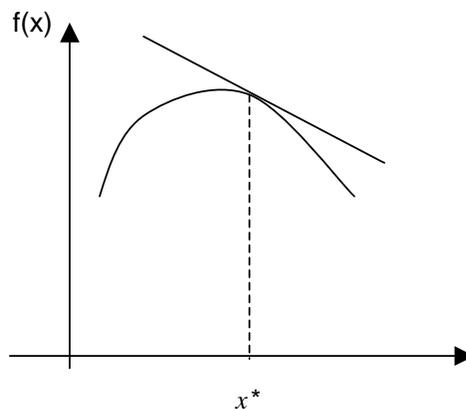
$$f(x) = \liminf_{y \rightarrow x} f(y)$$

La semicontinuidad superior es similar, con inf es reemplazado por sup y si ambas semicontinuidades se tienen en un punto, implica continuidad ordinaria en dicho punto.

Otra caracterización de las funciones cóncavas y convexas

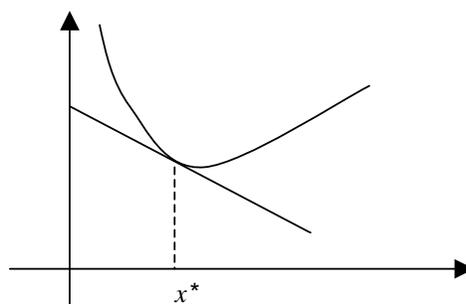
Sea $f(x) : S$, continua y diferenciable en todo punto $x \in S$, luego $f(x)$ es cóncava si y sólo si para cualquier par de puntos \underline{x}^* y $\underline{x} \in S$ se cumple:

$$f(\underline{x}) \leq f(\underline{x}^*) + \underline{\nabla}^T f(\underline{x}^*) \cdot (\underline{x} - \underline{x}^*)$$



y es convexa en S si:

$$f(\underline{x}) \geq f(\underline{x}^*) + \underline{\nabla}^T f(\underline{x}^*) \cdot (\underline{x} - \underline{x}^*)$$



Definición 6: Sea $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$, la matriz Hessiana asociada definida por:

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- es definida positiva en \bar{x} si $y' \cdot H(\bar{x}) \cdot y > 0$, $\forall y \neq 0$ en \mathfrak{R}^n , y
- es definida negativa en \bar{x} si $y' \cdot H(\bar{x}) \cdot y < 0$, $\forall y \neq 0$ en \mathfrak{R}^n .

$H(x)$ es semidefinida positiva (negativa) en \bar{x} si:

$$y' \cdot H(\bar{x}) \cdot y \geq (\leq) 0 \quad \forall y \in \mathfrak{R}^n$$

Una matriz D es **positiva definida** si y sólo si las variables pueden ser ordenadas de manera tal que $d_{11} > 0$ y además se verifica que:

$$\begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} > 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} > 0 \quad \mathbf{K} |D| > 0$$

Una matriz D es **positiva semidefinida** si y sólo si las variables pueden ser ordenadas de manera tal que $d_{11} > 0$ y además se verifica que:

$$\begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \geq 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} \geq 0 \quad \mathbf{K} |D| \geq 0$$

Una matriz D es **negativa definida** si y sólo si las variables pueden ser ordenadas de manera tal que $d_{11} < 0$ y además se verifica que:

$$\begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} > 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} < 0 \quad \mathbf{K} (-1)^n |D| > 0$$

Una matriz D es **negativa semidefinida** si y sólo si las variables pueden ser ordenadas de manera tal que $d_{11} < 0$ y además se verifica que:

$$\begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \geq 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} \leq 0 \quad \mathbb{K} (-1)^n |D| \geq 0$$

Herramientas útiles para determinar la convexidad de una función

Teorema 1: Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$ una función dos veces diferenciable en todo punto $x \in S$, siendo S un conjunto convexo, luego $f(x)$ es convexa (cóncava) si y sólo si su matriz Hessiana es positiva (negativa) semidefinida para todo $x \in S$.

Definición 7: Sea A un conjunto de \mathbb{R}^n , la cápsula convexa $co(A)$ es el menor subconjunto convexo de \mathbb{R}^n que contiene a A .

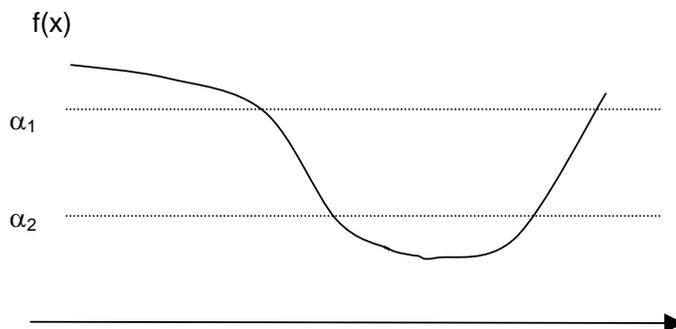
$$co(A) = \{x : x = \lambda x_1 + (1-\lambda)x_2; 0 \leq \lambda \leq 1, x_1, x_2 \in A\}$$

Definición 8: Sean X un conjunto y una función $f: X \rightarrow \mathbb{R}$, el epígrafo de f , $epi(f)$ es el conjunto $\{(x, \lambda) \in X \times \mathbb{R} / f(x) \leq \lambda\}$

Definición 9: Una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$, es cuasiconvexa si todos los conjuntos $S_\alpha = \{x \in \text{dom } f / f(x) \leq \alpha\} \quad \forall \alpha \in \mathbb{R}$ son convexos.

Una función es cuasicóncava si $-f$ es cuasiconvexa.

Ejemplo 1:



$f(x)$ es una función cuasiconvexa

Ejemplo 2: $f(x,y)=xy$, $x \geq 0, y \geq 0$, es una función cuasicóncava ya que los conjuntos

$$S_\alpha = \{(x, y) \in \mathbb{R}_+^2 / xy \geq \alpha\} \quad \forall \alpha \in \mathbb{R} \text{ son convexos.}$$

Ejemplo 3: $f(x)=x^3$ es cuasiconvexa.

Teorema 2:

Sea X un conjunto convexo, una función $f : X \rightarrow \mathbb{R}$, es cuasiconvexa sí y sólo sí

$$f(\alpha x + (1-\alpha)y) \leq \max\{f(x), f(y)\} \quad \forall x, y \in X \subseteq \mathbb{R}^n, \quad 0 \leq \alpha \leq 1.$$

Definición 10: f es estrictamente cuasiconvexa si $f(\alpha x + (1-\alpha)y) < \max\{f(x), f(y)\}$

$$\forall x, y \in X \subseteq \mathbb{R}^n, \quad x \neq y, \quad 0 \leq \alpha \leq 1.$$

Definición 11: Sea $f : X \rightarrow \mathbb{R}$, una función diferenciable sobre su dominio de definición, f es

pseudoconvexa si para cada $x, y \in X$, con $\nabla f(x) \cdot (y-x) \geq 0$ se tiene $f(y) \geq f(x)$, o en forma

equivalente si $f(y) < f(x)$, luego $\nabla f(x) \cdot (y-x) < 0$.

Una función es pseudocóncava si $-f$ es pseudoconvexa.

Teorema 3:

Sean X un conjunto convexo abierto, $f : X \rightarrow \mathbb{R}$, una función diferenciable y pseudoconvexa

en X , luego f es estrictamente cuasiconvexa y cuasiconvexa.

Definición 12: Una función f es separable si puede escribirse de la forma: $f = \sum_{i=1}^n f_i(x_i)$ donde

las funciones $f_i(x_i)$ son funciones univariadas en x_i .

Definición 13: Sean f, p, q funciones definidas sobre un subconjunto convexo $X \subset \mathbb{R}^n$ y p y q

son funciones convexas. Si la función f se puede expresar: $f(x)=p(x)-q(x) \quad \forall x \in X$ luego f se

dice que es una función D.C. sobre su dominio.

El conjunto de funciones DC es denso en el espacio de las funciones continuas. Una función continuamente diferenciable puede transformarse en una función DC.

Definición 14: (Mc. Cormick, 1976) ⁽¹³⁾ Una función es factorizable o separable si es una composición recursiva de sumas y productos de funciones de variables simples.

Ejemplo 4: $f(x,y,z,w) = \sqrt{\exp(xy + z \ln w)z^3}$ es factorizable ya que puede descomponerse de la

siguiente forma:

$$x_1 = xy$$

$$x_2 = \ln w$$

$$x_3 = zx_2$$

$$x_4 = x_1 + x_3$$

$$x_5 = \exp(x_4)$$

$$x_6 = z^3$$

$$x_7 = x_5 x_6$$

$$f = (x_7)^{0.5}$$

Definición 15: Dada $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$, \bar{x} es un mínimo (máximo) *local* de f , si existe un $\varepsilon > 0$ tal que $f(\bar{x}) \leq (\geq) f(x), \forall x \in N_\varepsilon(\bar{x})$ (entorno de \bar{x} de radio ε).

Definición 16: Sea $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$, \bar{x} es un mínimo (máximo) global de f si $f(\bar{x}) \leq (\geq) f(x), \forall x \in \mathfrak{R}^n$.

Referencias

- (1) Adjiman, C.S., Androulakis, I.P. and Floudas, C.A. "Global optimization of mixed-integer nonlinear problems" , *AIChE Journal* 46 p. 1769–1797, (2000)
- (2) Androulakis, I., Maranas, C.D. and Floudas, C.A. "αBB: A global optimization method for general constrained nonconvex problems", *Journal of Global Optimization* , Vol. 7, Number 4 p.337-363 (1995)
- (3) Bazaraa, M., Shetty, C. "Nonlinear Programming. Theory and Algorithms". Wiley, Ed. , second edition, (1993)
- (4) Biegler, Lorenz T., Grossmann Ignacio E. "Retrospective on optimization". *Computers and Chemical Engineering* 28 p.1169–1192, (2004)
- (5) Brooke, Anthony, David A. Kendrick, Alexander Meeraus and Ramesh Raman (1998), *GAMS: A Users Guide*, disponible en GAMS Development Corporation web site <http://www.gams.com>
- (6) Byrne 1 R.P., Bogle I.D.L., "Global optimisation of constrained non-convex programs using reformulation and interval analysis". *Computers and Chemical Engineering* 23 p.1341–1350, (1999)
- (7) Floudas, C. "Deterministic Global Optimization. Theory, methods and Applicatios", Kluwer Academic Publishers (2000)
- (8) Grossmann,I., Lorenz t, Biegler. "Part II: Future perspective on optimization" *Computers and Chemical Engineering* 28 p.1193–1218, (2004)
- (9) Lee, S., Grossmann I.E. "A global optimization algorithm for nonconvex generalized disjunctive programming and applications to process systems". *Computers and Chemical Engineering* 25 p.1675–1697, (2001)
- (10) Liberti, S.L., Tesis Doctoral: "Reformulation and Convex Relaxation Techniques for Global Optimization Department of Chemical Engineering and Chemical Technology". Imperial College London South Kensington Campus London SW7 2AZ (2004)
- (11) Marcovecchio, M., Bergamini, L. Aguirre, P. "Improve-and-Branch Algorithm for the Global Optimization of Nonconvex NLP Problems". *Journal of Global Optimization* , Vol. 34, Number 3 p.339-368 (2006)

- (12) McCormick, G.P. "Converting general nonlinear programming problems to separable nonlinear programming problems", Technical Report T-267, The George Washington University, Washington D.C., (1972)
- (13) . McCormick, G.P. "Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems", *Mathematical Programming* 10, 147–175. (1976)
- (14) Neumaier, Arnold "Complete Search in Continuous Global Optimization and Constraint Satisfaction". A chapter for: *Acta Numerica 2004* (A. Iserles, ed.), Cambridge University Press (2004)
- (15) Quesada, Ignacio , Grossmann, Ignacio. "Global optimization Algorithm for heat Exchanger Networks" *Ind. Eng. Chem. Res.* 32, p. 487–99 (1995)
- (16) Rockafellar R.T., "Convex Analysis". Princeton University Press (1970)
- (17) Ryoo, Sahinidis N., "Global optimization of nonconvex NLPs y MINLPs with applications in process design". *Computers chemical Engng.* Vol.19. Number 5, p.551-566, (1995).
- (18) Sahinidis N., "Global Optimization and Constraint Satisfaction: The Branch and Reduce Approach". LNCS 2861, p.1-16 Springer-Verlag Berlin Heidelberg (2003)
- (19) Sahinidis, N.V. "BARON: A general purpose global optimization software package", *Journal Global Optimization* Vol. 8, Number 2 p.201-205 (1996)
- (20) Sahinidis, N.V. BARON. Branch And Reduce Optimization Navigator. User's Manual, WWW-document. <http://archimedes.scs.uiuc.edu/baron/baron.html>
- (21) Shectman, P., Sahinidis N., "A finite Algorithm for Global Minimization of Separable Concave Programs". *Journal of Global Optimization* Vol. 12, Number 1: p.1-36, (1998).
- (22) Smith, E.M.B. Pantelides C.C. "A symbolic reformulation:spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs". *Computers and Chemical Engineering* Vol. 23, p. 457–478. (1999).
- (23) Tawarmalani, M. and N. V. Sahinidis, "Convex extensions and envelopes of lower semi-continuous functions", *Mathematical Programming*, Vol. 93 Number 2, p. 247-263, (2002)
- (24) Tawarmalani M., Sahinidis NV, <http://archimedes.scs.uiuc.edu/baron/baron.html>