



INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Apunte IN3171: Modelamiento y Optimización

Macarena V. Osorio A.¹

Imágenes por Matías F. Muñoz F.²

Basado en los apuntes de Andreas Wiese , Fernando Ordóñez y Víctor Bucarey
Universidad de Chile

¹Cualquier error en el presente apunte o petición de material complementario escribir a macarena.osorio.88@gmail.com

²Cualquier error en el presente apunte escribir a matias.srs@gmail.com

Índice

1. Para comenzar...	4
2. Introducción	5
2.1. Problemas de programación lineal (PPLs)	5
2.1.1. ¿Qué se verá en este apunte?	6
3. Modelamiento	7
3.1. Organizar horarios de guardia	7
3.2. Invirtiendo en emprendimientos	7
3.3. Empacando tu maleta	8
3.4. Problema de dimensionamiento de lote	9
3.5. Locaciones para estaciones de bomberos	9
3.6. Problema de secuencia	10
3.7. Flujos	11
3.8. Problema de transporte	12
3.9. Problema de asignación	12
3.10. Camino más corto	12
3.11. Aplicación: Producción de acero	12
3.12. Flujo máximo	13
3.13. Aplicación: Programación con derecho preferente	13
3.14. Problema del vendedor viajero (TSP)	14
3.14.1. Formulación alternativa	15
3.15. Detección de arbitraje	15
3.16. Funciones de costo lineal por partes	15
3.17. Sudoku	17
3.18. Restricciones disyuntivas	17
4. Geometría	18
4.1. Formas especiales de los PPLs	19
4.2. Puntos factibles	19
4.3. Puntos extremos	25
4.4. Degenerancia	34
5. Método Simplex	35
5.1. PPLs en forma estándar	35
5.1.1. Puntos extremos en PPLs en forma estándar	37
5.2. Moverse a un nuevo vértice	40
5.2.1. Degenerancia y Forma estándar	41
5.2.2. Cambio de costos al moverse en dirección d^j	44
5.3. Problemas degenerados	49
5.4. Encontrar solución básica factible inicial	51
5.5. Simplex: Método Full Tableau	52
6. Dualidad	53
6.1. Duales para PPL's arbitrarios	55
6.2. Dualidad débil	56
6.3. Dualidad fuerte	56
6.4. Holgura complementaria	58
6.5. Lema de Farkas	58
6.6. Simplex Dual	59

6.7. Multiplicador de Lagrange	60
6.8. Precios sombra	61
6.9. Interpretación dual de costos reducidos	63
7. Análisis de sensibilidad	63
7.1. Cambio en el vector b	64
7.2. Cambios en el vector de costos	67
7.3. Agregar nueva variable	70
7.4. Agregar nueva restricción de igualdad	70
7.5. Cambios en entrada de A	70
7.6. Dependencia global en b	70
7.7. Dependencia global en c	71
8. Programación entera, Branch & Bound	72
8.1. Métodos de programación entera	72
8.1.1. Planos cortantes	72
8.1.2. Problema de la mochila	73
8.1.3. Conjunto independiente	74
8.1.4. PPE general en forma estándar	74
8.1.5. Cortes de Gomory	74
8.1.6. Branch and bound	76
8.1.7. Branch and cut	81
8.1.8. Árbol de mínima expansión o Minimum Spanning Tree (MST)	81
8.2. Algoritmo de Prim	85
8.3. Algoritmo de Kruskal	87
8.4. Matrices totalmente unimodulares	88
8.4.1. El camino más corto	89
9. Optimización no lineal	92
9.1. Dualidad para problemas no lineales	94
9.2. Dualidad fuerte implica...	95
10. Condiciones de Karush-Kuhn-Tucker (KKT)	95
11. Descenso más empinado	102
11.1. Método de Newton	105
11.2. Ratio de convergencia	105

1. Para comenzar...

El apunte que tienes en tus manos es el primer apunte en español de este curso, el cual fue realizado por Macarena V. Osorio A., profesora auxiliar del curso IN3701 desde 2016 Otoño y MA3701 desde 2019 Primavera y las imágenes por Matías F. Muñoz F., profesor auxiliar del curso desde 2019 Otoño. El apunte se realizó a partir de diversas clases y apuntes de profesores que han dictado el curso como Andreas Wiese, Fernando Ordóñez y Víctor Bucarey.

¡Este apunte fue hecho con mucho amor, sudor y lágrimas! ¡Úsalo sabiamente!

Llegará un punto en que te sentirás como Neo al ser capaz de ver la matrix cada vez que resuelvas un ppl, esperamos que este apunte te ayude a lograrlo.

Maca y Mati 2019

2. Introducción

Aquí aprenderás programación lineal y sus métodos para optimizar que te servirán en tu futuro profesional y vida laboral.

Para comenzar con un problema cotidiano, imagina que invitas algunos amigos y amigas a comer en tu casa y quieres cocinar una rica pizza y lasaña, cabe destacar que dado el presupuesto tus invitados e invitadas puedes escoger una opción de las anteriores. Entonces, cuál sería la cantidad de cada opción a cocinar?

	Pizza	Lasaña	Disponible
Tomates	2	3	18
Queso	4	3	24

Para tu invitación, quieres maximizar el total de amigos y amigas a invitar en donde x_1 es el número de pizzas y x_2 el número de lasañas, por ende, el objetivo es maximizar $x_1 + x_2$, luego para los tomates se tiene $2x_1 + 3x_2 \leq 18$ y para el queso $4x_1 + 3x_2 \leq 24$, en donde cada opción de cena de ser igual o superior a cero $x_1 \geq 0$, $x_2 \geq 0$.

$$\begin{aligned} \text{máx } & x_A + x_B \\ & 2x_A + 3x_B \leq 18 \\ & 4x_A + 3x_B \leq 23 \\ & x_A, x_B \geq 0 \end{aligned}$$

El problema anterior es un problema de optimización lineal que se puede encontrar en situaciones cotidianas y se puede resolver tanto a mano como con programas específicos. Ahora, si las restricciones se grafican en un plano de dos dimensiones, se verifica que la solución óptimo no puede estar al interior de esta región ni fuera de esta, sino que en los extremos, en este caso el óptimo se encuentra en $(3, 4)$, lo cual al reemplazar en la función objetivo da un valor de 7.

2.1. Problemas de programación lineal (PPLs)

Los problemas a estudiar consisten en:

1. Parámetros: Los "números" del problema, por ejemplo, cantidad de tomates y queso en el problema anterior.
2. Variables: Especifica cada posible solución, por ejemplo, x_1 y x_2 .
3. Restricciones: Condiciones dadas del problema, por ejemplo, la restricción de tomates $2x_1 + 3x_2 \leq 18$.
4. Función objetivo: Asigna valores a cada solución.

Como se especifica en el título, estos problemas son lineales, es por ello que a continuación se recuerdan algunas definiciones fundamentales.

Definición 1. Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es *lineal* si $f(ax + by) = a \cdot f(x) + b \cdot f(y)$ para cualquier $x, y \in \mathbb{R}^n$ y cualquier $a, b \in \mathbb{R}$.

En otras palabras f es lineal si y sólo si hay valores a_1, \dots, a_n tal que $f(x_1, \dots, x_n) = a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n = \sum_{i=1}^n a_i \cdot x_i = (a_1 \ a_2 \ \dots \ a_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = a^T x$

Luego, una restricción es llamada *lineal* si es de alguna de las formas siguientes:

- $a^T x \leq b$.
- $a^T x \geq b$.
- $a^T x = b$.

Lo anterior para un vector $a \in \mathbb{R}^n$ y un escalar $b \in \mathbb{R}$.

Entonces, un programa lineal tendría:

- Función lineal objetivo en donde se maximice o minimice $c^T x$.
- Un conjunto de restricciones lineales.
- Un conjunto de variables continuas, es decir, $x \in \mathbb{R}^n$.

Algunas consideraciones importantes son las siguientes:

- Los problemas de programación entera poseen variables entera, es decir, $x \in \mathbb{Z}^n$.
- Las variables binarias son las que $x \in \{0, 1\}^n$, por ejemplo, toma el valor de 1 si se abre la tienda de pizza, 0 si no se abre.

2.1.1. ¿Qué se verá en este apunte?

El modelamiento de problemas lineales posee mucho poder en el sentido de que puede modelar diversas situaciones de industrias, vida cotidiana, entre otros, pudiendo encontrar así la solución óptima y ahorrar dinero al mismo tiempo. Algunos ejemplos son: Asignar estudiantes a escuelas y colegios de Santiago, reducir el tiempo de espera de pasajeros en el metro de Berlín, entre otros.

Luego al usar geometría en este curso se puede comprender la importancia de la región factible, convexidad y concavidad, soluciones posibles en las esquinas de la región factible.

Al aprender algoritmo Simplex, se podrá calcular el óptimo de un PPL dado comenzando en una esquina para luego poder moverse a una esquina adyacente con mejor costo hasta encontrar un óptimo. Cabe destacar que en la práctica el algoritmo funciona bien y ha sido utilizado por millones de empresas.

Luego, en Dualidad se podrá probar que la solución encontrada es óptima y transformar el problema inicial a uno que usualmente se podrá resolver de manera más fácil, encontrando un estimado de la mejor solución posible y cambiar vector costos c , elemento del lado derecho de la ecuación b restricciones o variables para verificar cómo cambia la solución óptima del problema inicial, todo ello en el capítulo de Sensibilidad.

En la programación entera, se podrán discutir temas como los cortes en los planos y Branch and Bound en los diversos problemas que se encuentren para luego en optimización no lineal llegar a encontrar problemas del estilo:

$$\begin{aligned}
 & \min f_0(x) \\
 \text{s.a.} \quad & f_i(x) \leq 0 \quad \forall i = 1, \dots, m \\
 & h_i(x) = 0 \quad \forall i = 1, \dots, p \\
 & x \in \mathbb{R}^n
 \end{aligned}$$

En donde se pueda probar que la solución es óptima mediante condiciones de Karush-Kuhn-Tucker más conocidas como KKT.

Entonces se pueden verificar las diferencias entre este tipo de programas con los de programación lineal o PPLs, en donde se tiene una función objetivo lineal que puede maximizar o minimizar $c^T x$ con un conjunto de restricciones de la forma $a^T x \leq b$, $a^T x \geq b$ o $a^T x = b$ creadas con un conjunto de variables continuas de la forma $x \in \mathbb{R}^n$. Allí se pueden encontrar diversas variables, por ejemplo, para programación entera como la incorporación de variables enteras, es decir, $x \in \mathbb{Z}^n$. Un ejemplo de lo anterior es:

$$\begin{aligned}
\max \quad & R(x_1, x_2) = 8x_1 + 7x_2 \\
\text{s.a.} \quad & 2x_1 + 3x_2 \leq 18 \\
& 4x_1 + 3x_2 \leq 24 \\
& x_1 \geq 0, x_2 \geq 0
\end{aligned}$$

3. Modelamiento

En el presente capítulo se mostrarán diversos problemas de programación lineal y su modelo propuesto, los cuales se han utilizado de manera real en diferentes industrias o problemas de la vida cotidiana.

3.1. Organizar horarios de guardia

La planificación horaria de los guardias de Beaucheff implica tener d_j en el día j de la semana, por ejemplo, $j = 1$ si es Lunes, $j = 2$ si es Martes y así sucesivamente. Cada guardia trabaja 5 días seguidos, por ende, se puede decidir qué día de la semana comienzan a trabajar. Entonces el problema es encontrar el número mínimo de guardias a contratar, teniendo así un modelo en donde la variable x_j modele el número de guardias que comienzan su semana de trabajo en el día j .

$$\begin{aligned}
\min \quad & \sum_{j=1}^7 x_j \\
& x_1 + x_4 + x_5 + x_6 + x_7 \geq d_1 \\
& x_1 + x_2 + x_5 + x_6 + x_7 \geq d_2 \\
& x_1 + x_2 + x_3 + x_6 + x_7 \geq d_3 \\
& x_1 + x_2 + x_3 + x_4 + x_7 \geq d_4 \\
& x_1 + x_2 + x_3 + x_4 + x_5 \geq d_5 \\
& +x_2 + x_3 + x_4 + x_5 + x_6 \geq d_6 \\
& +x_3 + x_4 + x_5 + x_6 + x_7 \geq d_7 \\
& x_j \geq 0 \\
& x_j \text{ entero}
\end{aligned}$$

3.2. Invirtiendo en emprendimientos

Posees 5 amigos que comenzarán un emprendimiento prontamente. Ellos te ofrecen invertir en sus emprendimiento y para ello tienes 1.000.000 CLP disponibles. Los números describen el dinero a invertir/recibir al inicio del año, cabe destacar que el dinero que se posee al inicio del año k puede ser invertido en el mismo año.

Año	Beat	Spreadshirt	Cornershop	Uber	GoPlaceIt
2020	inviertes x CLP	0	inviertes x CLP	inviertes x CLP	0
2021	recibes $0,3x$ CLP	inviertes x CLP	recibes $1,1x$ CLP	0	0
2022	recibes x CLP	recibes $0,3x$ CLP	0	0	inviertes x CLP
2023	0	recibes x CLP	0	recibes $1,75x$ CLP	recibes $1,4x$ CLP
Límite	500.000 CLP	No existe	500.000 CLP	No existe	750.00 CLP

Cabe destacar que el dinero adicional disponible se transfiere automáticamente a cuentas de ahorro con un porcentaje de 4 por interés. Dado 1.000.000 CLP, debe decidir cuánto invertir en cada startup, intentando maximizar el dinero al comienzo del 2023, quedando el siguiente modelamiento:

$$\begin{array}{rcllcl}
\text{máx} & \mathbf{1,04} \cdot y_{2022} + x_B + 1,75 \cdot x_D + 1,4 \cdot x_E & & & & \\
& 0 & \leq & x_A & \leq & 500,000 \\
& 0 & \leq & x_B & & \\
& 0 & \leq & x_C & \leq & 500,000 \\
& 0 & \leq & x_D & & \\
& 0 & \leq & x_E & \leq & 750,000 \\
& 0 & \leq & y_{2020} & & \\
& 0 & \leq & y_{2021} & & \\
& 0 & \leq & y_{2022} & & \\
& x_A + x_C + x_D + y_{2020} & = & & = & 1,000,000 \\
& 0,3 \cdot x_A + 1,1 \cdot x_C + \mathbf{1,04} \cdot y_{2020} & = & x_B + y_{2021} & & \\
& x_A + 0,3x_B + \mathbf{1,04} \cdot y_{2021} & = & x_E + y_{2022} & &
\end{array}$$

3.3. Empacando tu maleta

En el 2023, gracias a las inversiones inteligentes que has realizado, eres increíblemente adinerada/o, ahora te vas de vacaciones, pero llega el momento de empacar la maleta de volumen C y posees una lista de elementos I que quieres llevar a tu viaje, cada elemento $i \in I$ posee un volumen v_i y una utilidad u_i , entonces el objetivo es seleccionar el subconjunto de elementos $I' \subseteq I$ en el cual el volumen máximo es de C .

El objetivo es maximizar la suma de utilidades de elementos en I' . Entonces, el modelo del problema con variables binarias queda:

$$x_i = \begin{cases} 1 & , \text{ si } i \in I' \\ 0 & , \text{ si } i \notin I' \end{cases} \quad \forall i \in I$$

En donde,

$$\begin{array}{l}
\text{máx} \sum_{i \in I} u_i \cdot x_i \\
\sum_{i \in I} v_i \cdot x_i \leq C \\
x_i \in \{0, 1\}
\end{array}$$

Si se llega al caso de *más maletas*, se tendrá el conjunto J de maletas, en donde cada maleta $j \in J$ posee capacidad C_j .

$$x_{ij} = \begin{cases} 1 & , \text{ si el objeto } i \text{ está en la mochila } j \\ 0 & , \text{ si no} \end{cases}$$

$$\begin{array}{l}
\text{máx} \sum_{j \in J} \sum_{i \in I} u_i \cdot x_{ij} \\
\sum_{i \in I} v_i \cdot x_{ij} \leq C_j \quad \forall j \\
\sum_{j \in J} x_{ij} \leq 1 \quad \forall i \\
x_{ij} \in \{0, 1\}
\end{array}$$

Si dos objetos i, i' no pueden estar en la misma mochila: $x_{ij} + x_{i'j} \leq 1 \quad \forall j \in J$.

Este problema se conoce como el problema de la mochila (se utilizan mochilas en vez de maletas), entonces para el conjunto $K \subseteq I$ de objetos que no pueden ir en la misma maleta, se tiene entonces que $\sum_{i \in K} x_{ij} \leq 1 \quad \forall j \in J$.

3.4. Problema de dimensionamiento de lote

Un familiar posee un emprendimiento en donde fabrica poleras únicas en donde el plan de producción se realiza para las siguientes T semanas para cada $t \in \{0, \dots, T-1\}$ con $d_t \in \mathbb{N}_0$ como la demanda al fin de la semana t con el costo de setup c_t , el costo de producir una unidad p_t y el costo unitario de almacenamiento durante la semana. Por ende, el objetivo del problema es producir el costo total mínimo tal que al final de cada semana existan d_t unidades de producto listo para entregar.

Entonces las variables serían:

- $z_t \in \{0, 1\}$: Si se produce o no en semana t .
- $x_t \in \mathbb{N}_0$: Cantidad producida durante semana t .
- $y_t \in \mathbb{N}_0$: Cantidad que se almacena durante la semana t .

En donde la función objetivo minimiza $\sum_{t=0}^{T-1} (z_t \cdot c_t + x_t \cdot p_t + y_t \cdot h_t)$ con las restricciones siguientes:

- $y_0 = 0$.
- $x_t + y_t \geq d_t \quad \forall t$.
- $y_t = y_{t-1} + x_{t-1} - d_{t-1} \quad \forall t \geq 1$.
- $x_t \leq z_t \cdot \sum_{t=0}^{T-1} d_t \quad \forall t$ implica que si $z_t = 0$ entonces $x_t = 0$ y si $z_t = 1$ entonces x_t puede ser arbitrariamente grande.
- $x_t \in \mathbb{N}_0$ (Pequeño para $x_t \in \mathbb{Z}$ y $x_t \geq 0$).
- $y_t \in \mathbb{N}_0$.
- $z_t \in \{0, 1\}$.

Ahora, se supone que se permite decrecer la demanda a un costo de b_t por unidad perdida. Entonces la nueva variable sería w_t por cantidad de disminución de la demanda al final de la semana t , entonces la función objetivo sería minimizar $\sum_{t=0}^{T-1} (z_t \cdot c_t + x_t \cdot p_t + y_t \cdot h_t + b_t \cdot w_t)$, en donde $x_t + y_t \geq d_t - w_t \quad \forall t$.

Luego se supone que la producción sólo se permite en a lo más cinco semanas, lo cual se modela de la siguiente forma $\sum_{t=0}^{T-1} z_t \leq 5$, pero si se supone que ninguna de las dos puede ser consecutiva, entonces se tiene que $z_k + z_{k+1} \leq 1 \quad \forall k = 1, \dots, T-1$.

3.5. Locaciones para estaciones de bomberos

Sea el conjunto de distritos I y J el conjunto de posibles locaciones para estaciones de bomberos, se tiene que para cada $j \in J$ hay un costo c_j por construir una estación de bomberos en j . La distancia $d_{ij} \geq 0$ entre cada distrito i y locación j .

Para este problema se puede construir a lo más una estación de bomberos en cada locación en J , cada distrito debe ser asignado a una estación de bomberos, una estación de bomberos puede responsabilizarse de más de un distrito.

Luego, las variables son:

- $x_j := \begin{cases} 1 & \text{Si en la locación } j \text{ se instala una estación de bomberos} \\ 0 & \text{si no} \end{cases}$

- $y_{ij} := \begin{cases} 1 & \text{Si el distrito } i \text{ se asigna a una estación de bomberos en la locación } j \\ 0 & \text{si no} \end{cases}$

Las restricciones son:

$$\begin{aligned} x_j &\geq y_{ij} \quad \forall i, j && \text{Abrir estación de bomberos en } j \text{ cuando se asigna el distrito } i \\ \sum_j y_{ij} &= 1 \quad \forall i && \text{Cada distrito tiene una estación de bomberos asignada a él.} \\ x_j &\in \{0, 1\}, y_{ij} \in \{0, 1\} && \text{Naturaleza de variables.} \end{aligned}$$

Los posibles objetivos para el problema son:

- Cada distrito es servido por una estación de bomberos a una distancia $\leq D$, entonces se minimiza el costo se tiene que: $\text{Min } \sum_j x_j \cdot c_j$ con $y_{ij} = 0$ si $d_{ij} > D$.
- Se fija el presupuesto B , entonces se minimiza el promedio de la distancia del distrito a su estación de bomberos, entonces se tiene que: $\text{Min } \frac{1}{|I|} \sum_i \sum_j d_{ij} \cdot y_{ij}$ con la restricción del presupuesto como $\sum_j x_j \cdot c_j \leq B$.
- Minimizar la máxima distancia con un presupuesto fijo B : Primero se crea la variable $z :=$ Máxima distancia, con la función objetivo mín z , teniendo como restricción de presupuesto $\sum_j x_j \cdot c_j \leq B$. Al mismo tiempo se quiere tener la restricción en donde si $y_{ij} = 1$ implique que $z \geq d_{ij}$ y si $y_{ij} = 0$ implique que $z \geq 0$, es decir, la restricción se desactive, entonces $z \geq d_{ij} \cdot y_{ij} \quad \forall i, j$, lo que implica que $z \geq d_{ij}$ para todos los i, j s.a. $y_{ij} = 1$, es decir, en la solución óptima $z = \max_{i,j:y_{ij}=1} d_{ij}$, ya que el objetivo es minimizar z .

Ahora, hay relaciones usando variables binarias x_1, \dots, x_n , las cuales se pueden denotar como:

- $y_{ij} \leq x_j \Leftrightarrow y_{ij}$ implica x_j
- $\sum_j y_{ij} = 1 \Leftrightarrow$ Exactamente una de las variables es 1.
- $\sum_{j=1}^n x_j \leq 1 \Leftrightarrow$ A lo más una de las variables es 1.
- $\sum_{j=1}^n x_j \geq 1 \Leftrightarrow$ Al menos una de las variables es 1.
- $x_1 = 1 - x_2 \Leftrightarrow x_1$ Si sólo si x_2 no sucede.
- $x_i - x_j = 0 \Leftrightarrow$ Cualquiera de los dos elementos suceden de forma simultánea o ninguno de los dos sucede.
- $x_1 \leq x_2, x_1 \leq x_3, x_1 + 1 \geq x_2 + x_3 \Leftrightarrow x_1$ Sucede si y sólo si x_2 y x_3 suceden.
- $x_1 \geq x_2, x_1 \geq x_3, x_1 \leq x_2 + x_3 \Leftrightarrow x_1$ Sucede si y sólo si x_2 or x_3 suceden.

3.6. Problema de secuencia

Sean m máquinas y n trabajos, entonces cada trabajo debe ser procesado en cada máquina. Por cada trabajo j el orden está predeterminado: $j(1), j(2), \dots, j(m)$, luego cada trabajo j posee $p_{j,j(k)}$ unidades de tiempo en la máquina $j(k)$. Cada uno de los trabajos anteriores no puede ser interrumpido hasta su término.

El objetivo es minimizar el promedio de tiempos de finalización de trabajos. Entonces el modelo quedaría como sigue:

- La variable $t_{j,j(k)}$ denota el tiempo de finalización del trabajo j en la máquina $j(k)$.
- La función objetivo minimiza $\frac{1}{n} \sum_{j=1}^n t_{j,j(m)}$.
- Se tiene también que $t_{j,j(1)} \geq p_{j,j(1)}$ para cada j y también $t_{j,j(k)} \geq t_{j,j(k-1)} + p_{j,j(k)}$ para cada j y cada $k = 2, \dots, m$.

- Por cada máquina i y cada par de trabajos j, j' se introduce una variable binaria $x_{j,j',i}$ s.a. $x_{j,j',i} = 1$ si y sólo si en la máquina i , el trabajo j es procesado antes que j' , no necesariamente justo antes, entonces $x_{j,j',i} = 1 - x_{j',j,i}$ para cada j, j', i . Ahora si $x_{j,j',i} = 1$ se requiere que $t_{j',i} \geq t_{j,i} + p_{j',i}$.
- Sea $P := \sum_j \sum_i p_{j,i}$, implica que $t_{j',i} \geq t_{j,i} + p_{j',i} - (1 - x_{j,j',i})(P + p_{j',i})$ para cada i, j, j' , entonces si $x_{j,j',i} = 1$ la restricción sería $t_{j',i} \geq t_{j,i} + p_{j',i}$, en cambio si $x_{j,j',i} = 0$ la restricción quedaría $t_{j',i} \geq t_{j,i} - P$. Cabe destacar que se satisface para cualquier solución óptima, ya que en cualquier punto el tiempo del trabajo es procesado y así $t_{j,i} - P \geq 0$.
- Luego $t_{j,j(k)} \geq 0$ para cada trabajo j y máquina $j(k)$.
- Entonces, $x_{j,j',i} \in \{0, 1\}$ para todos los trabajos j, j' y cada máquina i .

El modelo posee variables enteras y continuas, por ende, es un programa lineal mixto entero.

3.7. Flujos

En el presente problema se quiere distribuir agua en donde cada locación se conecta con otra mediante tuberías. Cabe destacar que las locaciones poseen oferta y demanda. El objetivo es encontrar la forma de mover agua para balancear la oferta y demanda.

Definición 2. Un grafo dirigido $G = (V, A)$ consiste en un conjunto de nodos o vértices V y un conjunto de arcos $A \subseteq V \times V$. Cabe destacar que un arco $a = (i, j) \neq (j, i)$ (Si $i \neq j$).

La función objetivo se remite a minimizar los costos, en donde cada arco $a \in A$ posee un costo c_a , el cual envía x_a unidades de flujo de costo $x_a c_a$. Entonces si se tiene un grafo dirigido G con la demanda/oferta de la siguiente forma $b : V \rightarrow \mathbb{R}$, en donde:

- Si $b(i) > 0$ es un nodo de oferta.
- Si $b(i) < 0$ es un nodo de demanda.
- Si $b(i) = 0$ es un nodo de paso o transporte.

Luego, los costos se denotan como $c : A \rightarrow \mathbb{R}$ y así la capacidad máxima viene dada por $u : A \rightarrow \mathbb{R}_0^+ \cup \{\infty\}$. Finalmente, se define la variable $x_{ij} :=$ Cantidad de flujo desde el nodo i hasta el nodo j , la avariables existe si y sólo si $(i, j) \in A$.

$$\begin{aligned} & \text{mín} \sum_{(i,j) \in A} c_{ij} x_{ij} \\ & \sum_{\{j:(i,j) \in A\}} x_{ij} - \sum_{\{j:(j,i) \in A\}} x_{ij} = b(i) \quad \forall i \in V \\ & x_{ij} \leq u_{ij} \quad \forall (i,j) \in A \\ & x_{ij} \geq 0 \quad \forall (i,j) \in A \end{aligned}$$

Cabe destacar que la instancia del problema es factible si y sólo si:

$$\sum_{i \in N} b_i = 0 \quad \Rightarrow \quad \sum_{i \in N_{oferta}} b_i = - \sum_{i \in N_{demanda}} b_i$$

onde $N_{oferta} = \{i \in N : b(i) > 0\}$ y $N_{demanda} = \{i \in N : b(i) < 0\}$

3.8. Problema de transporte

Este modelo consiste en seguir exactamente la ruta del agua. Supone que sólo se interesa la cantidad de agua que cada nodo de abastecimiento mueve hacia cada nodo de demanda.

Para el caso especial de costo mínimo de flujo:

- No existen nodos intermedios, es decir, no hay un nodo i con $b(i) = 0$.
- Cada arco conecta un nodo de abastecimiento con uno de demanda: Los nodos se particionan entre $N_1 = \{i \in N : b(i) > 0\}$ y $N_2 = \{i \in N : b(i) < 0\} \forall (i, j) \in A : i \in N_1$ y $j \in N_2$.

Usualmente, uno puede resolver instancias de problemas de transporte más rápido que el general de las instancias de costo mínimo de flujo.

3.9. Problema de asignación

En una Universidad existe un conjunto de clases y un conjunto de asignaturas. Ambos conjuntos anteriores son del mismo tamaño.

Algunas clases están calificadas para algunas asignaturas y se quiere asignar exactamente una clase calificada a cada clase.

Este ejemplo es un caso particular de transporte, el cual posee N_1 como conjunto de clases y N_2 de asignaturas, luego como se había comentado: $|N_1| = |N_2|$.

Luego se define $b(i) = 1 \forall i \in N_1$ y $b(i) = -1 \forall i \in N_2$, en donde cada clase es calificada para algunas asignaturas. Entonces se verifica el arco (i, j) si la clase $i \in N_1$ es calificada por la asignatura $j \in N_2$. Luego de ello se requiere que $x_{ij} \in \{0, 1\}$ por cada arco $a = (i, j)$.

La solución es llamada matching: Subconjunto de arcos A' tal que cada vértice es incidente a lo más un arco de A' .

Luego para minimizar el costo total, se definen costos c_{ij} por cada arco (i, j) .

Este modelamiento también es válido para asignar trabajadores a trabajos, personas a horarios, entre otros problemas.

3.10. Camino más corto

Una empresa quiere programar un GPS y encontrar el camino más corto desde su ubicación actual hasta el aeropuerto, por ende, se verifica que las calles de Santiago pueden ser modeladas por un grafo dirigido G , con $c_{ij} > 0$ por cada arco $(i, j) \in A$ y dos nodos s, t . Se quiere encontrar el camino desde s hasta t , es decir, conjunto de nodos $s = v_1, v_2, \dots, v_k = t$ sujeto a $(v_i, v_{i+1}) \in A$ por cada i .

El objetivo es minimizar el costo $\sum_{\ell=1}^{k-1} c_{\ell, \ell+1}$, el cual es un caso especial para el flujo de costo mínimo: Enviar una unidad de flujo desde s a t . Finalmente, define $b(s) = 1$, $b(t) = -1$ y $b(i) = 0 \forall i \neq s, t$.

$$\begin{aligned} & \text{mín} \sum_{(i,j) \in A} c_{ij} x_{ij} \\ & \sum_{\{j:(i,j) \in A\}} x_{ij} - \sum_{\{j:(j,i) \in A\}} x_{ji} = \begin{cases} 1 & \text{if } i = s \\ 0 & \text{if } i \in N \setminus \{s, t\} \\ -1 & \text{if } i = t \end{cases} \\ & x_{ij} \in \{0, 1\} \\ & (x_{ij} \geq 0 \text{ es suficiente}) \end{aligned}$$

3.11. Aplicación: Producción de acero

Una compañía produce piezas de acero con n largos diferentes $L_1 < L_2 < \dots < L_n$, en donde la demanda es D_i para cada largo L_i .

Para producir x unidades de largo L_i la compañía tiene que pagar $\underbrace{K_i}_{\text{costos de setup}} + \underbrace{c_i}_{\text{costos unitarios}} \cdot x$. El costo unitario cumple $c_1 \leq c_2 \leq \dots \leq c_n$. Luego, para satisfacer demanda para largo L_i , la compañía debe producir D_i piezas de largo L_j con $j \geq i$.

El modelo de camino mínimo posee un nodo i por cada largo L_i , un nodo binario 0, para cada tamaño i se definen arcos (i, j) para cada $j = i + 1, i + 2, \dots, n$. El arco (i, j) modela la estrategia para producir el largo L_j para satisfacer toda la demanda para largos $L_{i+1}, L_{i+2}, \dots, L_j$.

Los costos se definen como sigue $c_{ij} := K_j + c_j \cdot \sum_{k=i+1}^j D_k$.

Luego, el camino desde el nodo 0 al nodo n define la estrategia de la siguiente forma: Si la estrategia óptima produce acero de largos: $L_{a_1}, L_{a_2}, \dots, L_{a_k} = L_n$, entonces el largo del camino $0, a_1, a_2, \dots, a_k = n$ es igual al costo de producción para cualquier camino $0, a_i, a_2, \dots, a_k = n$.

3.12. Flujo máximo

Se tiene un grafo dirigido G en donde la máxima capacidad viene dada por: $u : A \rightarrow \mathbb{R}_0^+ \cup \{\infty\}$ y los nodos existentes son s, t . El objetivo es encontrar el valor más grande de F tal que se puedan enviar F unidades de flujo desde s hacia t .

$$\begin{aligned} \text{máx } & F \\ \sum_{\{j:(i,j) \in A\}} x_{ij} - \sum_{\{j:(j,i) \in A\}} x_{ij} &= \begin{cases} F & \text{if } i = s \\ 0 & \text{if } i \in N \setminus \{s, t\} \\ -F & \text{if } i = t \end{cases} \quad \forall i \in N \setminus \{s, t\} \\ 0 \leq x_{ij} &\leq u_{ij} \\ F &\geq 0 \end{aligned}$$

Se puede modelar como un problema de flujo de costo mínimo. Luego, si se agrega un arco desde t hacia s , se tiene:

$$\begin{aligned} \text{mín } & -x_{ts} \\ \sum_{\{j:(i,j) \in A\}} x_{ij} - \sum_{\{j:(j,i) \in A\}} x_{ij} &= 0 \quad \forall i \in N \\ 0 \leq x_{ij} &\leq u_{ij} \quad \forall (i, j) \neq (t, s) \\ x_{ts} &\geq 0 \end{aligned}$$

Maximizar el flujo en el arco (t, s) es igual a maximizar la cantidad de flujo que va desde s hacia t en el resto de la red, entonces $\text{máx } x_{ts} = \text{mín } -x_{ts}$ es una instancia del problema de flujo de costo mínimo con:

$$\begin{aligned} c_{ij} &= 0 \quad \forall (i, j) \neq (t, s) \\ c_{ts} &= -1 \\ b_i &= 0 \quad \forall i \\ u_{ts} &= +\infty \end{aligned}$$

3.13. Aplicación: Programación con derecho preferente

Se tienen m máquinas idénticas, un conjunto de trabajos J para ser procesados. Cada uno de ellos poseen tiempo de procesamiento $p_j \geq 0$, tiempo de entrega $r_j \geq 0$ y vencimiento $d_j \geq 0$, entonces el objetivo es encontrar un programa para trabajos en máquinas en donde se deba trabajar en cada trabajo j por

exactamente p_j unidades de tiempo durante $[r_j, d_j)$ con la consideración de que los trabajos pueden ser preferentes o prioritarios en relación a otros y, por ende, se deberán terminar de después posiblemente en otra máquina. Es importante mencionar que un trabajo no puede ser ejecutado en dos máquinas al mismo tiempo, lo que implica que $r_j + p_j \leq d_j$ para cada trabajo j .

Entonces si se modela como un problema de flujo, se tendría entonces que el orden $\{r_j, d_j | j \in J\}$ no decreciente, luego el tiempo se divide $[0, \max_{j \in J} d_j)$ en intervalos T_1, \dots, T_k . Así, en la instancia de máximo flujo se tiene el nodo de inicio s y el de fin t , como también el nodo j por cada trabajo $j \in J$, en donde se tiene el arco (s, j) con capacidad p_j para cada trabajo j , luego se tiene el nodo T_ℓ para cada intervalo de tiempo T_ℓ con el arco (j, T_ℓ) que posee capacidad $|T_\ell|$ para cada trabajo j y en cada intervalo T_ℓ s.a., $T_\ell \subseteq [r_j, d_j)$ (Es importante notar que $T_\ell \subseteq [r_j, d_j)$ o $T_\ell \cap [r_j, d_j) = \emptyset$). Luego, el arco (T_ℓ, t) con capacidad $m \cdot |T_\ell|$ por cada intervalo T_ℓ .

Luego, se busca que la programación factible S exista si y sólo si la instancia del flujo máximo posee tamaño de flujo como $\sum_j p_j$, por ende, se quiere demostrar.

Entonces por cada trabajo j se envían p_j unidades de flujo en el arco (s, j) , luego sea $y_{j\ell}$ la cantidad de trabajos j procesados durante el intervalo T_ℓ en S . Luego, $y_{j\ell} > 0$ si y sólo si $T_\ell \subseteq [r_j, d_j)$, luego $y_{j\ell} \leq |T_\ell|$ desde que j comienza, se puede trabajar sólo en una máquina en el tiempo específico de S . Entonces se envían $y_{j\ell}$ unidades de flujo en (j, T_ℓ) por cada intervalo T_ℓ , es decir, mediante el arco (T_ℓ, t) , acotado por $|T_\ell \cdot m|$, entonces la conservación de flujo se mantiene lo que implica que la cantidad total de flujo $\sum_{j \in J} p_j$, lo anterior es una implicancia de que se supone que existe un flujo que envíe $\sum_j p_j$ unidades.

Luego, se considera el intervalo T_ℓ y para cada trabajo j sea $x_{j\ell} \geq 0$ el flujo del arco (j, T_ℓ) , entonces $x_{j\ell} > 0$ sólo si $T_\ell \subseteq [r_j, d_j)$, por ende, se construye una programación para el intervalo T_ℓ via regla envolvente. Sea $\sum_j x_{j\ell} \leq m \cdot |T_\ell|$ implica que todos los trabajos calzan, entonces $x_{j\ell} \leq |T_\ell|$ para cada trabajo j lo que implica que ningún trabajo es paralizado.

3.14. Problema del vendedor viajero (TSP)

Sea un conjunto de n ciudades y distancias entre ellas, se modela como un grafo dirigido $G = (V, A)$, en donde por cada arco $(i, j) \in A$ se tiene c_{ij} como su largo.

Entonces se debe empezar en una ciudad arbitraria para luego visitar cada ciudad exactamente una vez y retornar a la ciudad de inicio.

Cabe destacar que la ciudad de inicio es arbitraria, ya que se busca un ciclo, es decir, un conjunto de enlaces que cumplan las condiciones anteriormente mencionadas, por ello no importa qué ciudad sea.

Entonces se modelan las siguientes variables:

$$x_{ij} = \begin{cases} 1 & , \text{ si se selecciona el arco } (i, j) \\ 0 & , \text{ si no} \end{cases}$$

Con la función objetivo como sigue:

$$\text{mín} \sum_{(i,j) \in A} c_{ij} \cdot x_{ij}$$

Y las restricciones:

$$\sum_{j:(i,j) \in A} x_{ij} = 1 \quad \forall i \in V \quad (1)$$

$$\sum_{j:(j,i) \in A} x_{ji} = 1 \quad \forall i \in V \quad (2)$$

Entonces el número de variables se calcula como: $|V|(|V| - 1)$ y el número de restricciones sería mayor que $2^{|V|} - 2$, entonces cuando hay mayor cantidad de ciudades, la cantidad de restricciones aumenta, por ende, el problema es más complicado de resolver.

3.14.1. Formulación alternativa

Dadas un conjunto de n ciudades y distancias entre ciudades modeladas en un grafo dirigido $G = (V, A)$, entonces por cada arco $(i, j) \in A$ el largo del mismo se denota como c_{ij} .

Entonces si se calcula el viaje, teniendo en cuenta que visita sólo una vez cada ciudad, se presenta la formulación alternativa:

$$\begin{array}{ll}
 \text{mín} & \sum_{(i,j) \in A} c_{ij} \cdot x_{ij} \\
 u_i - u_j + n \cdot x_{ij} & \leq n - 1 \quad \forall (i, j) \in A, \text{ with } j \neq 1 \\
 \sum_{j:(i,j) \in A} x_{ij} & = 1 \quad \forall i \in V \quad (*) \\
 \sum_{j:(j,i) \in A} x_{ji} & = 1 \quad \forall i \in V \quad (**) \\
 x_{ij} & \in \{0, 1\} \quad \forall (i, j) \in A \\
 u_i & \geq 0 \quad \forall i \in V
 \end{array}$$

Si el viaje comienza en la ciudad 1 se denota de la siguiente forma $u_1 = 1$, luego si i es la k -ésima ciudad visitada, entonces $u_i = k$. Pero ¿Por qué es una formulación válida? Para cada viaje factible existe una solución correspondiente en donde $x_{ij} := 1$ si el arco (i, j) es parte del viaje y $x_{ij} := 0$ si no. Entonces se asume que el viaje comienza en la ciudad 1, por ende, se define $u_1 := 1$. Si la ciudad i es la k -ésima ciudad visitada entonces se define $u_i := k$. Luego, si $x_{ij} = 1$ entonces la restricción para (i, j) significa que $u_i - u_j + n \leq n - 1$, es decir, $u_i + 1 \leq u_j$, se satisface (Con $u_j = u_i + 1$), luego si $x_{ij} = 0$ la restricción para (i, j) significa que $u_i - u_j \leq n - 1$ se satisface con $u_i \leq n$ y $u_j \geq 1$.

3.15. Detección de arbitraje

Se asume que hay n tipos de moneda y los bancos ofrecen tasas de cambio, en donde hay un conjunto de moneda de origen i y moneda de destino j . Entonces la tasa de cambio α_{ij} significa que por una unidad de moneda i se obtiene α_{ij} unidades de moneda j .

El modelo es representado en un grafo dirigido $G = (V, A)$. Entonces el objetivo es definir un problema de programación entera para encontrar arbitraje si es que hay un ciclo $i_1, i_2, \dots, i_{k-1}, i_k, i_{k+1} = i_1$ tal que $\prod_{i=1}^k \alpha_{i, i+1} > 1$, es decir, $\log(\prod_{i=1}^k \alpha_{i, i+1}) > \log 1$ como también, $\sum_{i=1}^k \log \alpha_{i, i+1} > 0$. En donde se le asigna un costo $c_{ij} := \log \alpha_{i, i+1}$ para cada arco (i, j) , por ende, el objetivo es encontrar un ciclo positivo.

Finalmente, el modelo de programación entera queda como la maximización de $\sum_{(i,j) \in A} x_{ij} c_{ij}$ sujeto a $\sum_{j:(i,j) \in A} x_{ij} \leq 1$ para todo i para que a lo más un arco entre a cada vértice i , luego $\sum_{j:(j,i) \in A} x_{ij} \leq 1$ para todo i para que a lo más un arco de salida exista para cada vértice i , $\sum_{j:(i,j) \in A} x_{ij} = \sum_{j:(j,i) \in A} x_{ij}$ para todo i para que todo lo que entre como flujo salga y la naturaleza de variables $x_{ij} \in \{0, 1\}$ para cada arco $(i, j) \in A$.

Entonces se usa la programación entera para resolver el problema, en donde si el arbitraje es posible, entonces la solución óptima es positiva. Primero, se seleccionan todos los enlaces del ciclo de arbitraje $i_1, i_2, \dots, i_{k-1}, i_k, i_{k+1} = i_1$, que satisface todas las restricciones $0 < \sum_{i=1}^k \log \alpha_{i, i+1} = \sum_{i=1}^k c_{i, i+1}$. Entonces cada arco tiene 0 ó 1 arco entrante, pero si tiene 1 arco entrante, debe tener uno de salida, lo que implica que cualquier solución es una colección de ciclos $\sum_{(i,j) \in A} x_{ij} c_{ij} > 0$, es decir, un ciclo $i_1, i_2, \dots, i_{k-1}, i_k, i_{k+1} = i_1$ debe satisfacer $0 < \sum_{j=1}^k c_{i_j, i_{j+1}} = \sum_{i=1}^k \log \alpha_{i, i+1}$, entonces el ciclo produce arbitraje.

3.16. Funciones de costo lineal por partes

Como ejemplo de este modelamiento, se puede ver lo siguiente:

$$\begin{array}{ll} \text{mín} & |x_1| \\ & x_1 \geq -4 \\ & x_1 \leq 5 \end{array}$$

Luego, se introduce la variable z_1 , entonces por intuición se tiene que: $z_1 = |x_1|$, teniendo así:

$$\begin{array}{ll} \text{mín} & x_1 \\ & x_1 \geq -4 \\ & x_1 \leq 5 \\ & x_1 \leq z_1 \\ & -x_1 \leq z_1 \end{array}$$

Entonces en cualquier solución factible $z_1 \geq |x_1|$, luego si en la solución se tiene que $z_1 > |x_1|$ entonces la solución no es óptima, lo que implica que en el óptimo $z_1 = |x_1|$.

Luego, generalizando se obtiene:

$$\begin{array}{ll} \text{mín} & \sum_{i=1}^n c_i \cdot |x_i| \\ & a_j^T x \geq b_j \quad \forall \quad j \\ & c_i \geq 0 \quad \text{Para cada } i \end{array}$$

Se reescribe a:

$$\begin{array}{ll} \text{mín} & \sum_{i=1}^n c_i \cdot z_i \\ & a_j^T x \geq b_j \quad \forall \quad j \\ & x_i \leq z_i \quad \text{Para cada } i = 1, \dots, n \\ & -x_i \leq z_i \quad \text{Para cada } i = 1, \dots, n \end{array}$$

Entonces en cualquier solución factible $z_i \geq |x_i|$ para cada i , si en la solución $z_i > |x_i|$ para algún i entonces la solución no es óptima, ya que $c_i \geq 0$, lo que implica que cualquier solución óptima $z_i = |x_i|$ para cada i no funciona para maximización ni si $c_i < 0$ para algún i .

Luego, para la segunda generalización se observa que $|x_i| = \max\{x_i, -x_i\}$ lo que implica que minimizando $|x_1|$ es igual a minimizar el $\max\{x_i, -x_i\}$, entonces el mínimo del $\max_{i=1, \dots, m} (c_i^T x + d_i)$ si se evalúa sujeto a $Ax \geq b$. Cabe notar que $\max_{i=1, \dots, m} (c_i^T x + d_i)$ es igual al valor más pequeño de z sujeto a $z \geq c_i^T x + d_i$ por cada i .

Entonces el problema es equivalente a

$$\begin{array}{ll} \text{mín} & z \\ & Ax \geq b \\ & z \geq c_i^T x + d_i \quad \text{Para cada } i = 1, \dots, m \end{array}$$

Las funciones de costo anteriores por partes son lineales y convexas, pero ¿Qué sucede cuando no son convexas?

Dado los puntos $a_1 < a_2 < \dots < a_k$ y la función lineal por parte $f(x)$ especificada por los puntos $(a_i, f(a_i))$, se quiere minimizar $f(x)$ dado x , ¿Cómo se puede calcular el valor de la función objetivo $f(x)$?

Para cualquier $x \in [a_i, a_{i+1}]$ se puede encontrar un único λ_i, λ_{i+1} tal que $x = \lambda_i a_i + \lambda_{i+1} a_{i+1}$ y $\lambda_i + \lambda_{i+1} = 1$ y $\lambda_\ell = 0$ si $i \neq \ell \neq i + 1$. Entonces $f(x) = \sum_{i=1}^k \lambda_i f(a_i)$, lo que implica que se calcula $\lambda_1, \dots, \lambda_k$ s.a. $x = \sum_{i=1}^k \lambda_i a_i, \lambda_i + \lambda_{i+1} = 1$ para algún i y $\lambda_\ell = 0$ si $i \neq \ell \neq i + 1$.

Entonces, ¿Cómo se modela que a lo más dos λ_i consecutivos sean no-cero?

Mediante una variable binaria y_i tal que $y_i = 1$ si $a_i \leq x \leq a_{i+1}$ y $y_i = 0$ si no.

$$\begin{aligned}
\text{mín } \sum_{i=1}^k \lambda_i f(a_i) \\
x &= \sum_{i=1}^k \lambda_i \cdot a_i \\
\sum_{i=1}^k \lambda_i &= 1 \\
\lambda_1 &\leq y_1 \\
\lambda_i &\leq y_{i-1} + y_i \quad \text{para } i = 2, \dots, k-1 \\
\lambda_k &\leq y_{k-1} \\
\sum_{i=1}^{k-1} y_i &= 1 \\
\lambda_i &\geq 0 \\
y_i &\in 0, 1
\end{aligned}$$

¿Por qué funciona? Se supone que $y_{i'} = 1$ (todas las otras son $y_\ell = 0$), entonces $\lambda_{i'}$ pueden ser positivos y $\lambda_{i'+1}$ pueden ser positivos, pero $\lambda_j \leq 0$ para todos los j con $i' \neq j \neq i'+1$ ($y_{i'}$ aparece sólo en las restricciones para $\lambda_{i'}$ y $\lambda_{i'+1}$), lo que implica que $\sum_{i=1}^k \lambda_i f(a_i) = \lambda_{i'} f(a_{i'}) + \lambda_{i'+1} f(a_{i'+1}) = \underbrace{f(\lambda_{i'} \cdot a_{i'} + \lambda_{i'+1} \cdot a_{i'+1})}_{=x}$

3.17. Sudoku

El juego consiste en una tabla de 9x9, en donde en algunas celdas un número del $\{1, \dots, 9\}$ es dado. Entonces el objetivo del juego es agregar números desde $\{1, \dots, 9\}$ en cada celda tal que en cada fila cada número aparezca una vez al igual que en cada columna. Cabe destacar que la tabla se subdivide en 9 bloques de 3x3, en cada uno de esos bloques, el número aparece exactamente una vez.

El modelo queda como sigue:

▪ **Variables:**

- x_{ijk} para cada $i, j, k \in \{1, \dots, 9\}$: En donde $x_{ijk} = 1$ si en la posición de la fila i , columna j , el número k existe, $x_{ijk} = 0$, si no.

▪ **Restricciones:**

- $\sum_{k=1}^9 x_{ijk} = 1$ para cada i, j (Cada celda tiene sólo una entrada)
- $\sum_{i=1}^9 x_{ijk} = 1$ para cada j, k (El número k aparece solamente una vez en la columna j)
- $\sum_{j=1}^9 x_{ijk} = 1$ para cada i, k (El número k aparece solamente una vez en la fila i)
- $\sum_{i=x+1}^{x+3} \sum_{j=y+1}^{y+3} x_{ijk} = 1$ para cada k y para cada $x \in \{0, 3, 6\}$ y cada $y \in \{0, 3, 6\}$ El número k aparece solamente una vez en cada bloque de 3x3)
- $x_{ijk} = 1$ si en la posición (i, j) el número k existe en el sudoku entregado al comienzo.

3.18. Restricciones disyuntivas

Sean dos restricciones dadas $a^T x \geq b$ y $\bar{a}^T x \geq \bar{b}$ con $a \geq 0$ y $\bar{a} \geq 0$ con $x \geq 0$. Entonces se pide que al menos una de las dos restricciones se debe satisfacer, entonces se crea la variable binaria y en donde si toma el valor 1, la primera restricción debe ser satisfecha y si tiene valor 0, la segunda restricción debe ser satisfecha, quedando entonces:

$$\begin{aligned}
a^T x &\geq yb \\
\bar{a}^T x &\geq (1-y)\bar{b} \\
y &\in 0, 1
\end{aligned}$$

Entonces, una aplicación sería la siguiente:

Las variables se definen como x, w_1, w_2 , en donde se crea la restricción $x = \min\{w_1, w_2\}$, luego se modela como $x \leq w_1$ y $x \leq w_2$, $x \geq w_1$ o $x \geq w_2$.

Entonces, en general el conjunto de restricciones $a_i^T x \geq b_i$ para $i = 1, \dots, K$, en donde al menos $k < K$ de ellas está satisfecha, luego se introduce la variable binaria y_i para cada restricción $a_i^T x \geq y_i b_i$ para cada $i = 1, \dots, K$, en donde $\sum_{i=1}^k y_i \geq K$.

Luego si se restringe el rango de valores, se tiene la variable x , en donde se supone que se quiere restringir x , en donde deberá tomar valores sólo del conjunto $\{a_1, \dots, a_m\}$, para ello se introducen variables binarias y_1, \dots, y_m , por ende, la intuición queda como: $y_i = 1 \Leftrightarrow x = a_i$, en donde $x = \sum_{j=1}^m a_j y_j$ con $\sum_{j=1}^m y_j = 1$ y $y_j \in \{0, 1\}$.

4. Geometría

Para comenzar a estudiar Geometría, debemos interiorizarnos en los Problemas de Programación Lineal (PPL) en donde se pueden encontrar las siguientes características:

- Función objetivo:
 - Maximizar $c^T x$
 - Minimizar $c^T x$
- Restricciones:
 - $a_i^T x \leq b_i$, en donde $i \in I$
 - $a_j^T x \geq b_j$, en donde $j \in J$
 - $a_k^T x = b_k$, en donde $k \in K$
- Variables:
 - $x \in \mathbb{R}^n$

En donde el objetivo principal es encontrar x , el cual cumpla todas las restricciones necesarias del problema de programación lineal, maximice o minimice según sea el caso y a la vez este sea factible.

Ejemplo: En una fábrica de pintura se tienen dos colores que son los más pedidos por los clientes, ellos son *Blanco Nieve* y *Negro Brillante*. La fábrica quiere maximizar la utilidad que les produce vender galones de *Blanco Nieve* y *Negro Brillante* según la cantidad vendida, siendo la utilidad 8 CLP y 7 CLP, respectivamente. Cabe destacar que esta fábrica posee dos máquinas, en la primera se pueden producir máximo 18 [Watts] y en la segunda 24 [Watts] diariamente, teniendo en cuenta que producir 1 galón de *Blanco Nieve* significan 2 [Watts] en la primera máquina y 4 [Watts] en la segunda. Al mismo tiempo, producir un galón de *Negro Brillante* significa 3 [Watts] en ambas máquinas.

$$\begin{aligned} & \text{maximize } 8x_1 + 7x_2 \\ & 2x_1 + 3x_2 \leq 18 \\ & 4x_1 + 3x_2 \leq 24 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

Dado el PPL anterior, se puede inferir que la región es convexa y que su óptimo está en un punto extremo. Pero ¿Esto siempre sucede? Eso es lo que se probará mediante las principales definiciones, lemas, corolarios y ejemplos.

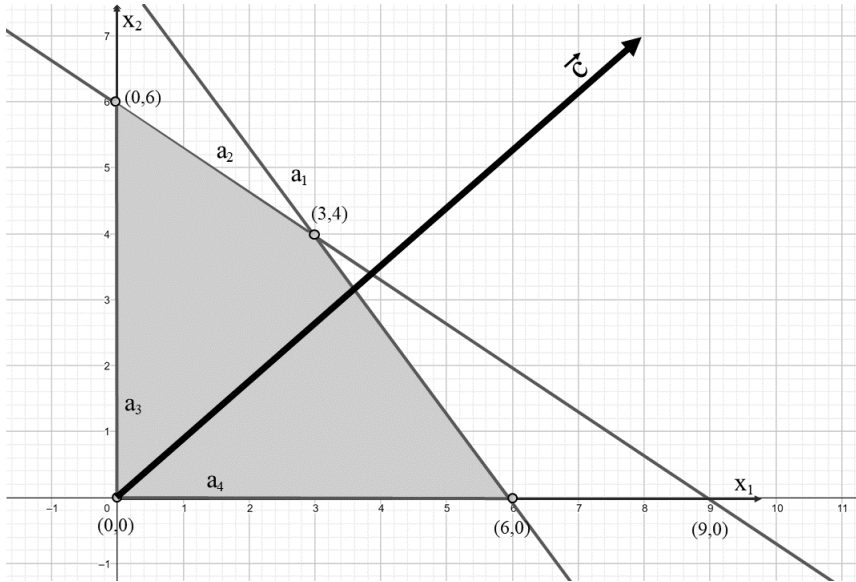


Figura 1: a_1 corresponde a la restricción $4x_1 + 3x_2 \leq 24$, a_2 corresponde a $2x_1 + 3x_2 \leq 18$, a_3 corresponde a $x_1 \geq 0$, a_4 corresponde a $x_2 \geq 0$. El vector $c = (8, 7)$ corresponde a la dirección de crecimiento de la función objetivo *i.e.* el gradiente de esta. El punto $(6, 0)$ es factible, pero el punto $(9, 0)$ no lo es, pues no cumple a_1 . Se puede intuir que el punto $(3, 4)$ es el punto óptimo dado el dibujo, se abordará este tema más adelante.

4.1. Formas especiales de los PPLs

Los PPLs pueden ser de diversas formas, ya sea en función objetivo, restricciones, naturaleza de variables, entre otros aspectos, es por ello que se presenta la forma canónica en la que se pueden trabajar de mejor forma.

Lema 3. *Dado un PPL genérico, es posible representarlo de una forma canónica, la cual es: $\text{mín } c^T x$ s.a. $Ax \geq b$*

Demostración. Lo anterior se logra simplificando tipos de variables y restricciones, es decir, teniendo en cuenta un PPL arbitrario, se encuentra uno equivalente de la siguiente forma:

1. Si la función objetivo es $\text{máx } c^T x$, se cambia mín mediante la multiplicación de -1 quedando esto de la siguiente forma $\text{mín } -c^T x$.
2. Se reemplaza cada restricción de la forma $a_i^T x \leq b_i$ por su equivalente $-a_i^T x \geq -b_i$.
3. Se reemplaza cada restricción de la forma $a_i^T x = b_i$ por sus dos restricciones equivalentes $a_i^T x \geq b_i$ y $-a_i^T x \geq -b_i \Leftrightarrow a_i^T x \leq b_i$.

□

Por lo anterior, de ahora en adelante todos los PPL serán de la forma $\text{mín } c^T x$ s.a. $Ax \geq b$.

4.2. Puntos factibles

Teniendo el PPL $\text{mín } c^T x$ s.a. $Ax \geq b$, se dice que el punto $x \in \mathbb{R}^n$ es **factible** si cumple con todas las restricciones dadas.

Al mismo tiempo, se dice que el punto $x \in \mathbb{R}^n$ es **óptimo** aparte de cumplir la condición de ser factible, cumple que por cada $x' \in \mathbb{R}^n$, se tiene $c^T x \leq c^T x'$ si el problema fuese de minimización. Es por lo anterior que al resolver un PPL se pueden encontrar las siguientes posibles soluciones:

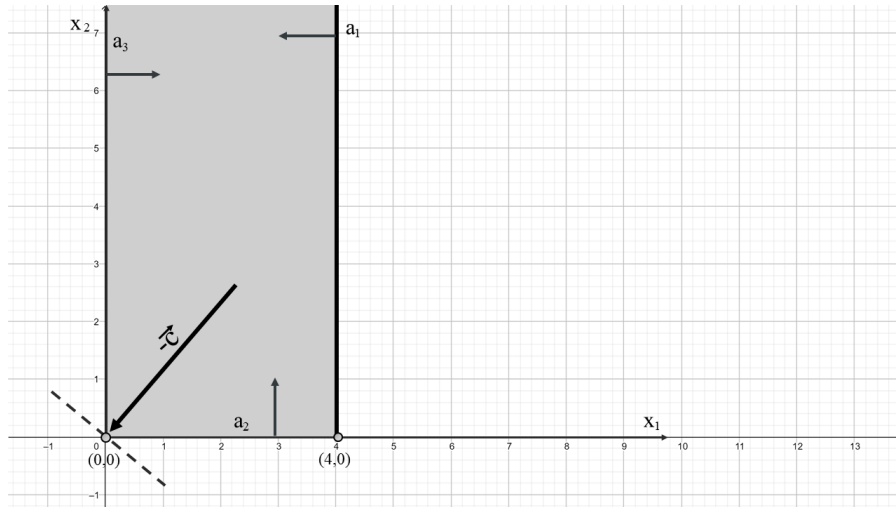


Figura 2: a_1 corresponde a la restricción $-x_1 \geq -4 \Leftrightarrow x_1 \leq 4$, a_2 corresponde a $x_2 \geq 0$, a_3 corresponde a $x_1 \geq 0$. El vector $-c = (-1, -2)$ corresponde a la dirección de crecimiento de la función objetivo *i.e.* el gradiente de esta. Se puede observar como la solución es única y corresponde al punto $(0, 0)$ con valor óptimo 0.

1. Existe **un solo punto** $x^* \in \mathbb{R}^n$ tal que $c^T x^* \leq c^T x$ por cada x factible, por ejemplo:

$$\begin{aligned} & \text{mín } x_1 + x_2 \\ & -x_1 \geq -4 \Leftrightarrow x_1 \leq 4 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

Existen **muchos puntos** $x^* \in \mathbb{R}^n$ tales que $c^T x^* \leq c^T x$ por cada x factible, por ejemplo:

$$\begin{aligned} & \text{min } x_1 \\ & -x_1 \geq -4 \rightarrow x_1 \leq 4 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

2. Por cada $C \in \mathbb{R}$ existe un punto factible x tal que $c^T x \leq C$, es decir, no acotado, por ejemplo:

$$\begin{aligned} & \text{min } -x_2 \\ & -x_1 \geq -4 \rightarrow x_1 \leq 4 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

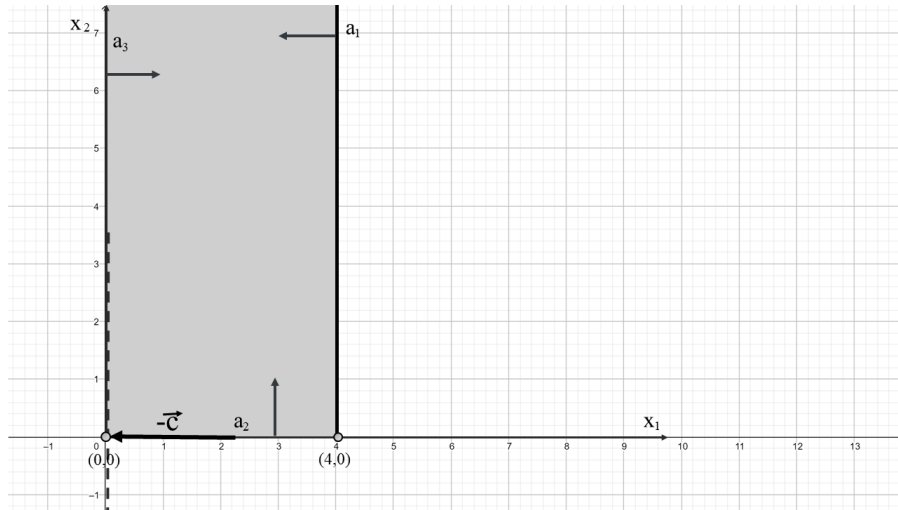


Figura 3: a_1 corresponde a la restricción $-x_1 \geq -4 \Leftrightarrow x_1 \leq 4$, a_2 corresponde a $x_2 \geq 0$, a_3 corresponde a $x_1 \geq 0$. El vector $-c = (-1, 0)$ corresponde a la dirección de crecimiento de la función objetivo *i.e.* el gradiente de esta. Se puede observar como la solución es la recta $x_1 = 0$ con valor óptimo 0. Esto implica que cualquier punto de la forma $(0, x_2)$ es óptimo.

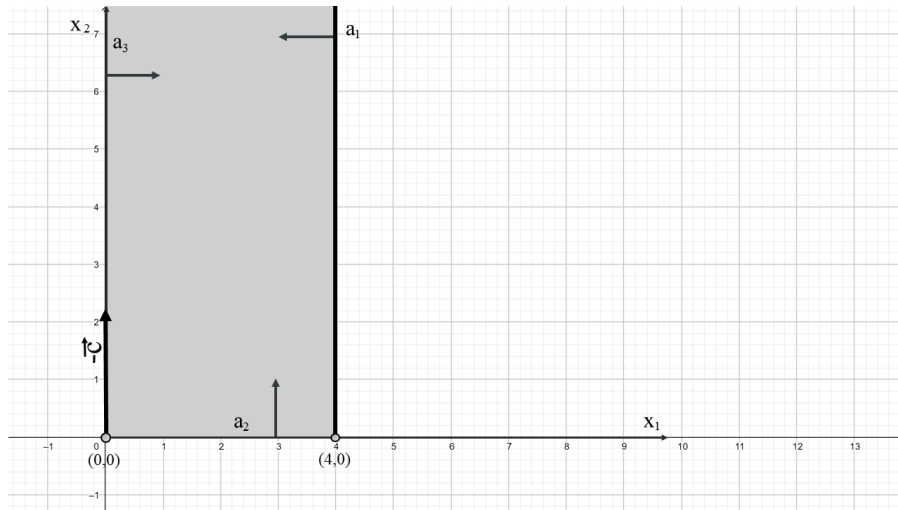


Figura 4: a_1 corresponde a la restricción $-x_1 \geq -4 \Leftrightarrow x_1 \leq 4$, a_2 corresponde a $x_2 \geq 0$, a_3 corresponde a $x_1 \geq 0$. El vector $-c = (0, 1)$ corresponde a la dirección de crecimiento de la función objetivo *i.e.* el gradiente de esta. Se puede observar como a cualquier solución en la recta $x_1 = 0$, se puede encontrar una solución mejor. Por lo tanto, el valor óptimo es $-\infty$.

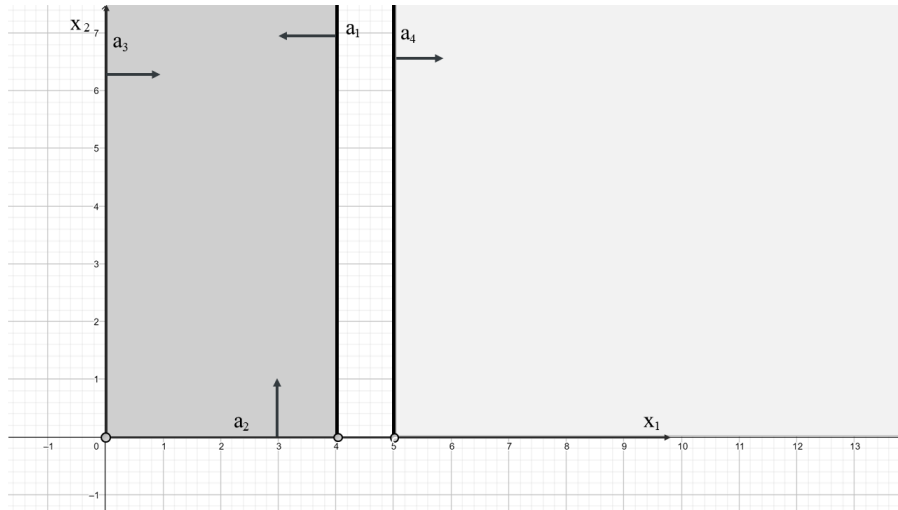


Figura 5: a_1 corresponde a la restricción $-x_1 \geq -4 \Leftrightarrow x_1 \leq 4$, a_2 corresponde a $x_2 \geq 0$, a_3 corresponde a $x_1 \geq 0$, a_4 corresponde a $x_1 \geq 5$. Se puede observar que la intersección de las restricciones es vacía, lo que genera que el poliedro sea infactible para cualquier función objetivo dada.

3. No hay punto $x \in \mathbb{R}^n$ factible, por ejemplo:

$$\begin{aligned} & \min -x_2 \\ & -x_1 \geq -4 \rightarrow x_1 \leq 4 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \\ & x_1 \geq 5 \end{aligned}$$

Definición 4. Sea $a \in \mathbb{R}^n$ y $b \in \mathbb{R}$:

- $\{x \in \mathbb{R}^n | a^T x \leq b\}$ es llamado **Semiespacio**, el cual gracias a sus intersecciones generan un poliedro.
- $\{x \in \mathbb{R}^n | a^T x = b\}$ es llamado **Hiperplano**, el cual representa a puntos con el mismo valor de la función objetivo.

Gracias a las definiciones anteriores, es posible verificar la construcción de un poliedro a partir de ellos como se describe a continuación:

Definición 5. $P \subseteq \mathbb{R}^n$ es un **poliedro** si es la intersección de un número finito de Semiespacios, es decir, $P = \{x \in \mathbb{R}^n | Ax \geq b\}$ para una matriz $A \in \mathbb{R}^{m \times n}$ y un vector $b \in \mathbb{R}^m$.

Definición 6. Si P puede describirse mediante una matriz $A \in \mathbb{Z}^{m \times n}$ y un vector $b \in \mathbb{Z}^m$, es llamado **Poliedro racional**. Cabe destacar que en este curso, sólo se verán Poliedros de este tipo.

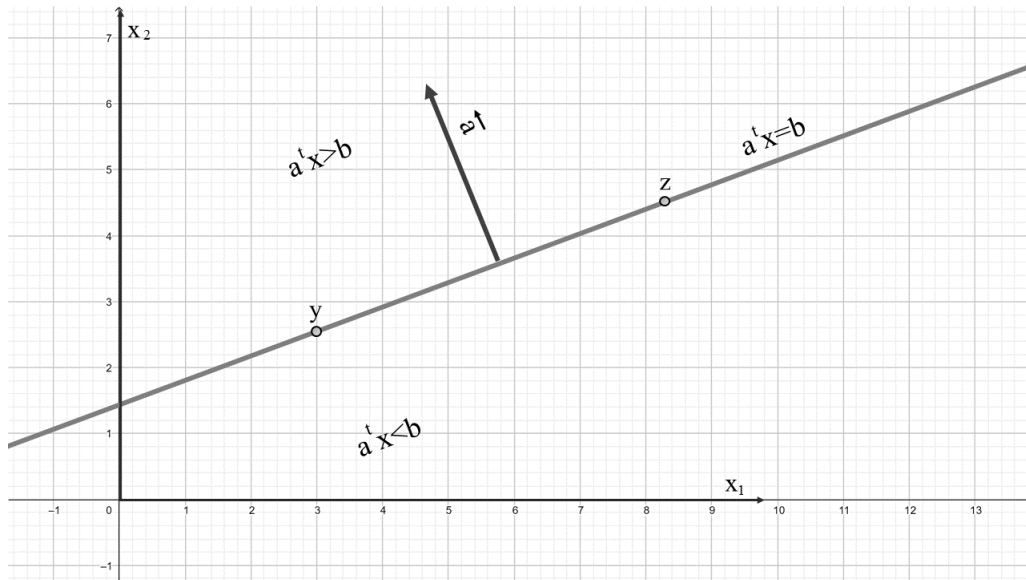


Figura 6: El Vector a es perpendicular a la restricción $a^T x = b$, pues como y y z están en el hiperplano, implica que $a^T y = b$ y $a^T z = b$, por lo tanto, $a^T (y - z) = 0$. Esto quiere decir que a es ortogonal al vector conformado por $(y - z)$.

Definición 7. Un conjunto $S \subseteq \mathbb{R}^n$ es **acotado** si existe una constante K tal que para todo $x \in S$ se tiene que $|x_i| \leq K$ para cada $i = 1, \dots, n$.

Ejemplo: El siguiente poliedro es acotado, por ejemplo, escogiendo $K=6$

$$\begin{aligned}
 P &= (x_1, x_2) \\
 2x_1 + 3x_2 &\leq 18 \\
 4x_1 + 3x_2 &\leq 24 \\
 x_1 &\geq 0 \\
 x_2 &\geq 0
 \end{aligned}$$

Ejemplo: El siguiente poliedro es no acotado

$$\begin{aligned}
 P &= (x_1, x_2) \\
 2x_1 + 3x_2 &\geq 18 \\
 4x_1 + 3x_2 &\geq 24 \\
 x_1 &\geq 0 \\
 x_2 &\geq 0
 \end{aligned}$$

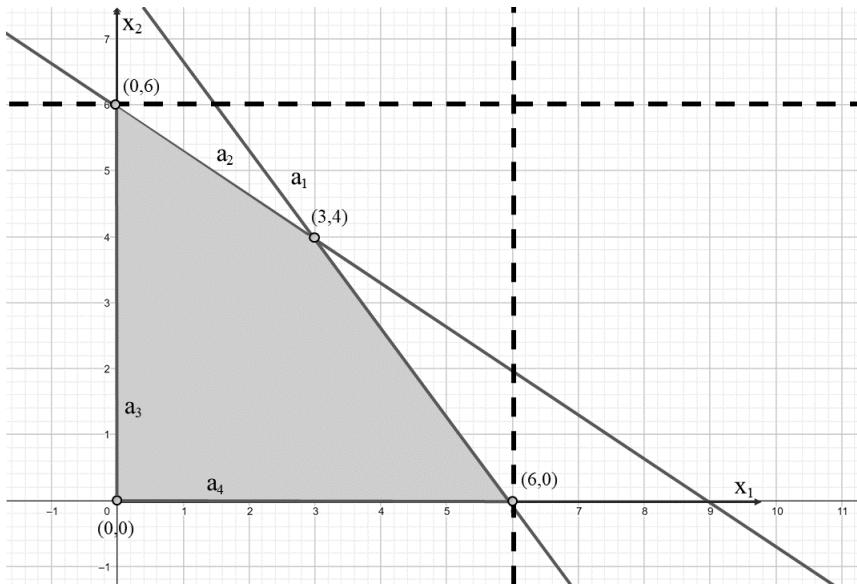


Figura 7: a_1 corresponde a la restricción $4x_1 + 3x_2 \leq 24$, a_2 corresponde a $2x_1 + 3x_2 \leq 18$, a_3 corresponde a $x_1 \geq 0$, a_4 corresponde a $x_2 \geq 0$. Las líneas punteadas representan gráficamente que $|x_1| \leq 6$ y $|x_2| \leq 6$ para cualquier x factible en el poliedro.

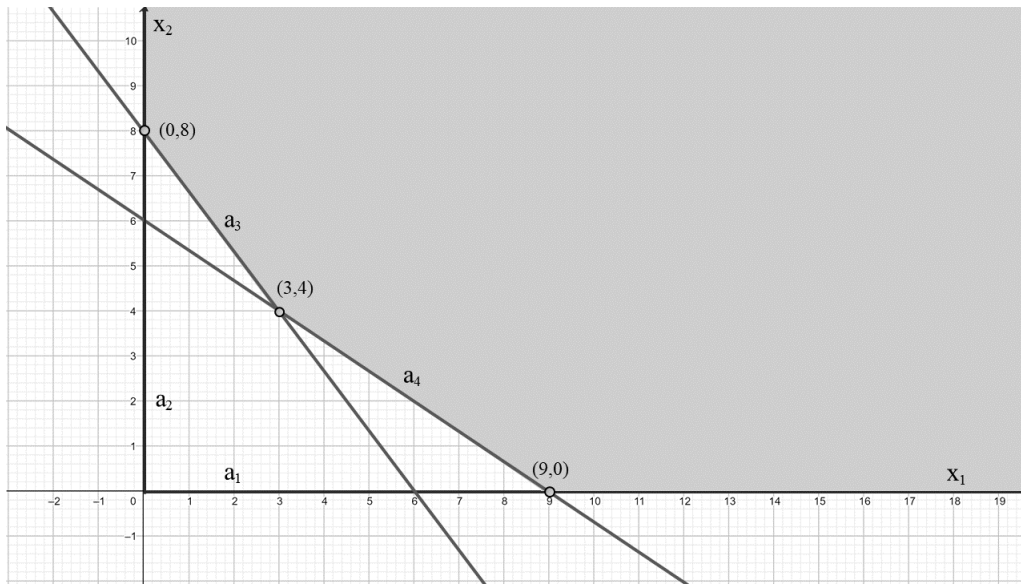


Figura 8: a_3 corresponde a la restricción $4x_1 + 3x_2 \geq 24$, a_4 corresponde a $2x_1 + 3x_2 \geq 18$, a_2 corresponde a $x_1 \geq 0$, a_1 corresponde a $x_2 \geq 0$. Se puede observar que no existe un K tal que $|x_i| \leq K$ para cada $i = 1, \dots, n$, pues siempre que se elija una cota superior, se puede encontrar un x factible que la supere.

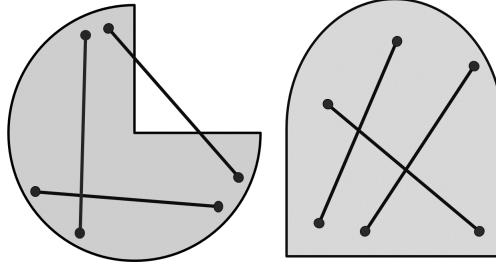


Figura 9: Se puede notar que el Pacman no es convexo, pues hay parte de la combinación convexa que queda fuera de este, y que el fantasma es convexo, ya que toda combinación convexa pertenece a este.

En particular, como se ve en los ejemplos anteriores los poliedros pueden ser acotados o no acotados dependiendo de sus restricciones, en particular en el primer ejemplo se verifica un polígono.

Definición 8. Un **polígono** es un poliedro acotado.

Algunas características esenciales de los poliedros son:

Definición 9. Un conjunto $S \subset \mathbb{R}^n$ es **convexo** si para cualquier $x, y \in S$ y cualquier $\lambda \in [0, 1]$ se tiene que $\lambda x + (1 - \lambda)y \in S$.

Definición 10. Sean $x_1, x_2, \dots, x_K \in \mathbb{R}^n$ y $\lambda_1, \dots, \lambda_K \in \mathbb{R}_+$ tal que $\sum_{i=1}^K \lambda_i = 1$.

- El vector $y = \sum_{i=1}^K \lambda_i x_i$ es una **combinación convexa** de vectores x_1, \dots, x_K .
- La **envoltura convexa** de los vectores x_1, \dots, x_K es un conjunto de todas las combinaciones convexas de estos vectores, denotado como $\text{Convconv}\{x_1, \dots, x_K\}$.

Teorema 11. Gracias a las definiciones anteriores, se verifica que:

1. Sea I un conjunto y $S_i \subseteq \mathbb{R}^n$ sea convexo para cada $i \in I$. Entonces $S := \bigcap_{i \in I} S_i$ es convexo.
2. Todo Semiespacio es convexo.
3. Todo poliedro es convexo.

Demostración. Para cada punto se verifica que:

1. Sea $x, y \in S$ y $\lambda \in [0, 1]$. Entonces $z := \lambda x + (1 - \lambda)y \in S_i$ para cada $i \in I$. Así, $z \in S$.
2. Sea $a \in \mathbb{R}^n$ y $b \in \mathbb{R}$ y considerando $H := \{z \in \mathbb{R}^n \mid a^T z \leq b\}$. Sea $x, y \in H$, es decir, $a^T x \leq b$ y $a^T y \leq b$. Sea $z := \lambda x + (1 - \lambda)y$. $a^T z = a^T(\lambda x + (1 - \lambda)y) \leq \lambda b + (1 - \lambda)b = b$ y así $z \in H$.
3. Gracias al punto 1 y 2, se ve la implicancia del punto 3.

□

4.3. Puntos extremos

En las secciones anteriores se ha hablado de los poliedros y sus características principales, pero ahora se podrá profundizar en otros contenidos tales como el reconocimiento y utilidad de sus puntos extremos.

Definición 12. Sea P un poliedro. Un punto $x \in P$ es un **punto extremo** de P si no se pueden encontrar dos vectores $y, z \in P$ con $y \neq x \neq z$ y un escalar $\lambda \in [0, 1]$ tal que $x = \lambda y + (1 - \lambda)z$.

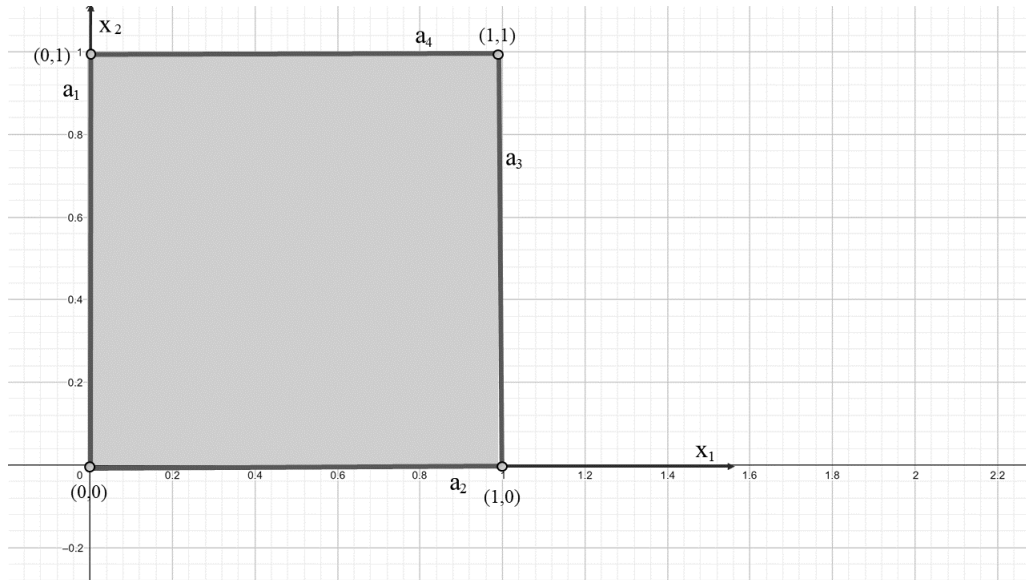


Figura 10: a_1 corresponde a la restricción $x_1 \geq 0$, a_2 corresponde a $x_2 \geq 0$, a_3 corresponde a $x_1 \leq 1$, a_4 corresponde a $x_2 \leq 1$.

Definición 13. Sea $P \subseteq \mathbb{R}^n$ un poliedro. Un vector $x \in P$ es un **vértice** de P si existe cualquier $c \in \mathbb{R}^n$ tal que $c^T x < c^T y$ para todos los y que satisfagan $y \in P$ y $y \neq x$.

Definición 14. Se tiene que si el vector x satisface $a_i^T x = b_i$ para cualquier i en I se dice que la restricción es i **activa** en x .

Ejemplo:

Se tiene el siguiente poliedro:

$$\begin{aligned}
 P &= (x_1, x_2) \\
 x_1 &\leq 1 \\
 x_2 &\leq 1 \\
 x_1, x_2 &\geq 0
 \end{aligned}$$

Agregando ahora la restricción $2x_1 \geq 0$, se verifica que el punto $(0, 1/2)$ tiene n restricciones activas, pero no es un vértice, ya que la restricción es linealmente dependiente con otra ya existente.

Definición 15. Un conjunto de restricciones son **linealmente independientes** si sus vectores correspondientes a_i son linealmente independientes.

- Cabe destacar que en los vértices hay n restricciones activas linealmente independientes.

Gracias a la cantidad de restricciones linealmente independientes, se pueden verificar otras características de los poliedros, por ejemplo, reconocer sus soluciones básicas y soluciones básicas factibles (s.b.f.).

Definición 16. Sea P un poliedro definido por restricciones con igualdad e inecuaciones, un punto $x^* \in \mathbb{R}^n$ es **solución básica** si

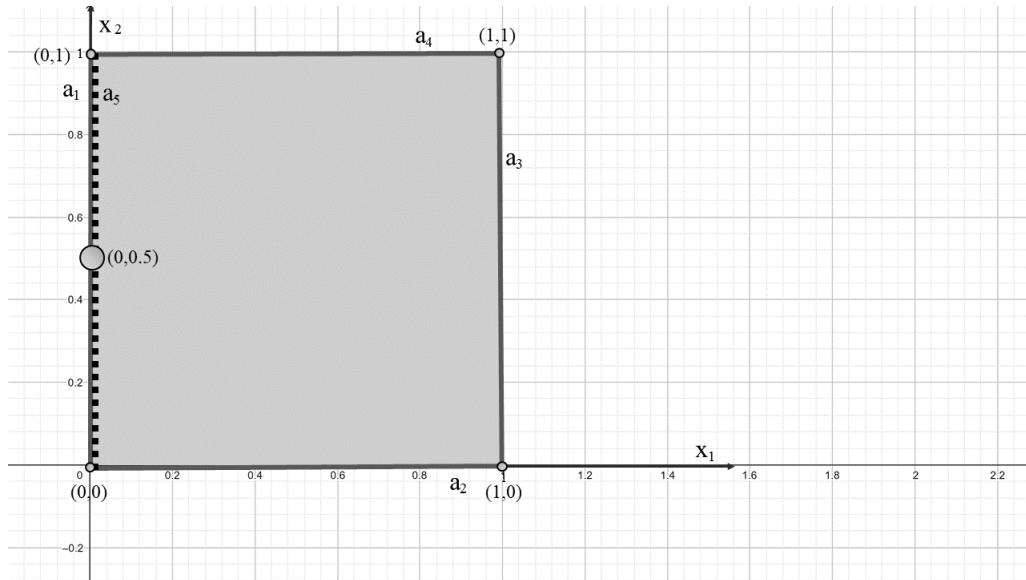


Figura 11: Ahora se le agrega la restricción a_5 la cual corresponde a $2x_1 \geq 0$.

- Hay n restricciones linealmente independientes que son activas en x^* .
- Todas las restricciones de igualdad son activas.

Es **Solución básica factible** si adicionalmente $x^* \in P$.

Lema 17. Sea $x^* \in \mathbb{R}^n$ e $I = \{i | a_i^T x^* = b_i\}$ sea un conjunto de índices de restricciones activas en x^* , entonces los siguientes son equivalentes:

1. Hay n vectores en el conjunto $S = \{a_i | i \in I\}$ en donde son linealmente independientes.
2. $\text{span}(S) = \mathbb{R}^n$, es decir, el conjunto de combinaciones lineales de S genera \mathbb{R}^n .
3. El sistema lineal $a_i^T x = b_i$, $i \in I$, tiene una solución única.

Demostración. ■ Siendo n el número de columnas de A .

- 1. \Rightarrow 3. n vectores linealmente independientes en S . Entonces el rango de un sistema lineal $a_i^T x = b_i$, $i \in I$, es al menos n . \Rightarrow El sistema posee una única solución o no existe solución. x^* es solución \Rightarrow 3.
- 3. \Rightarrow 2. Supongamos que $\text{span}(S) \neq \mathbb{R}^n$. Luego, existe un vector $d \neq 0$ ortogonal a $\text{span}(S)$, es decir, $a_i^T d = 0$ por cada $i \in I$. Entonces $a_i^T (x + d) = a_i^T x + a_i^T d = b_i$ y así x no es una solución única.
- 2. \Rightarrow 1. $\text{span}(S) = \mathbb{R}^n$ implica que S contiene una base de tamaño n .

□

Ejemplo: Para aclarar las definiciones anteriores, se verifica el siguiente ejemplo, en donde

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 2 & 3 \end{pmatrix} x^* \geq \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

Luego, al descomponer la Matriz anterior por filas se obtiene:

$$a_1^T = (100), a_2^T = (010), a_3^T = (001), a_4^T = (123)$$

Luego de lo anterior, se genera S el cual contiene a las tres primeras filas de la Matriz anterior. Cabe destacar que la combinación lineal de los elementos de S genera el espacio en \mathbb{R}^3

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}, \text{span}(S) = \mathbb{R}^3$$

Generando el sistema lineal a continuación, se verifica la solución única:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$x^* = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Ya se ha hablado de las definiciones de punto extremo, vértice y solución básica factible como también su representación gráfica correspondiente, la cual converge en el siguiente Teorema:

Teorema 18. Sea $P = \{x | Ax \geq b\}$ un poliedro no vacío con $x^* \in P$. Entonces se cumplen las siguientes equivalencias:

1. x^* es un vértice.
2. x^* es un punto extremo.
3. x^* es una solución básica factible.

Por ende, desde ahora en adelante se tomará en cuenta que los puntos óptimos serán llamados de las tres formas anteriores de forma equivalente.

Demostración. ■ 1. \Rightarrow 2.

Sea $x^* \in P$ un vértice $\exists c \in \mathbb{R}^n$ tal que $c^T x^* < c^T y$ para todo $y \in P$ con $y \neq x^*$. Se asume por contradicción que $\exists y, z \in P$ con $y \neq x^* \neq z$ y $0 < \lambda < 1$ s.a. $x^* = \lambda y + (1 - \lambda)z$. $c^T x^* < c^T z$, lo cual implica que $c^T x^* < c^T(\lambda y + (1 - \lambda)z)$ y, por lo tanto, $x^* \neq \lambda y + (1 - \lambda)z$.

■ 2. \Rightarrow 3.

Se supone que x^* no es una solución básica factible. $I := \{i | a_i^T x^* = b_i\}$, lo cual implica que no existen n vectores linealmente independientes en $\{a_i | i \in I\}$, implicando entonces que existe $d \neq 0$ con $a_i^T d = 0$ para todo $i \in I$. Sea $y := x^* + \epsilon d$ y $z := x^* - \epsilon d$. Si $i \in I$: $a_i^T y = a_i^T z = b_i$. Si $i \notin I$: $a_i^T x^* > b_i$, lo cual implica que se puede elegir $\epsilon > 0$ tal que también $a_i^T y > b_i$ y $a_i^T z > b_i \forall i$ lo cual implica que $y, z \in P$ y $y + z = x^*$ y así x^* no es punto extremo.

■ 3. \Rightarrow 1.

Se quiere encontrar la función objetivo c tal que x^* sea el optimizador único para ese objetivo. Sea $I := \{i | a_i^T x^* = b_i\}$ el conjunto de sub índices que cumplen esa condición dada. Sea $c = \sum_{i \in I} a_i$ entonces

$$c^T x^* = \sum_{i \in I} a_i^T x^* = \sum_{i \in I} b_i$$

Para cualquier $x \in P$ y cualquier i se tiene $a_i^T x \geq b_i$, y así

$$c^T x = \sum_{i \in I} a_i^T x \geq \sum_{i \in I} b_i$$

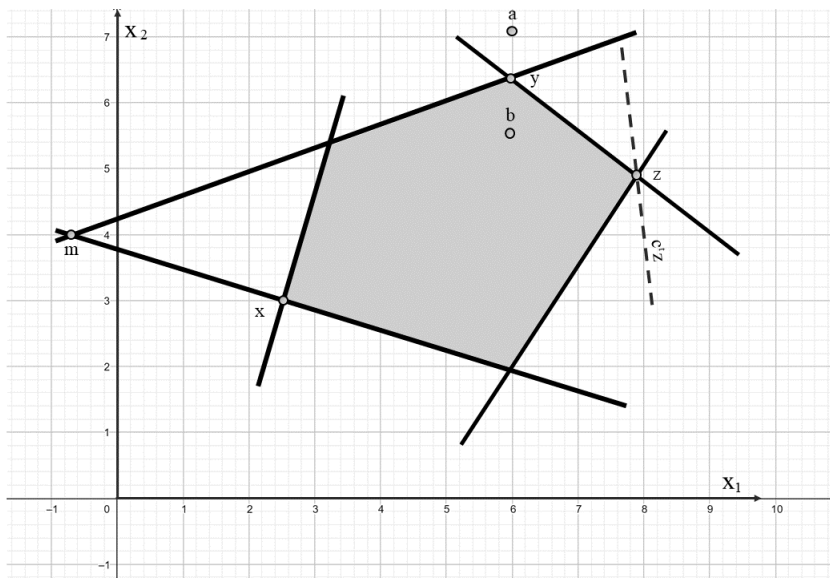


Figura 12: Se puede notar que x es solución básica factible, pues activa 2 restricciones linealmente independientes y cumple todas las restricciones dentro del poliedro, pero el punto m también activa 2 restricciones l.i. pero este no cumple todas las restricciones, por lo tanto, es sólo solución básica. El punto y es punto extremo, ya que la única manera de escribirlo como combinación convexa de puntos distintos de y , es con un punto dentro del poliedro como b y uno fuera de este como a . El punto z es vértice pues existe un c tal que $c^T z \leq c^T x$ para cualquier punto factible en el poliedro distinto de z .

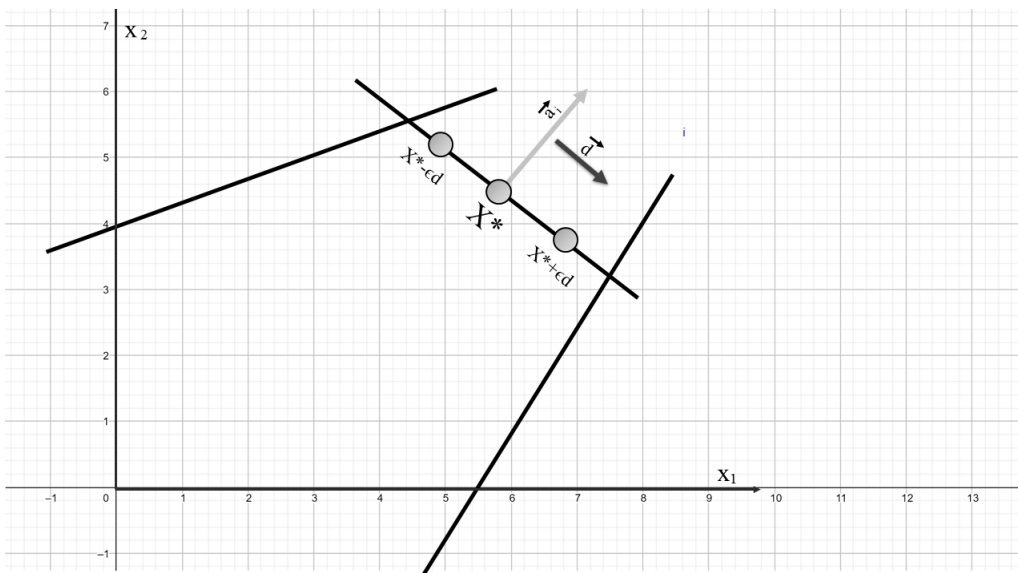


Figura 13: Representación gráfica de la demostración. x^* es el punto para el cual no hay n restricciones activas, pero es punto extremo. d es perpendicular a los a_i de las restricciones activas en x^* . El movimiento puede ser de x^* en la dirección de d y $-d$ donde los puntos resultantes son factibles en el poliedro y se podrá usar para definir una combinación convexa que entregue de resultado x^* obteniendo la contradicción.

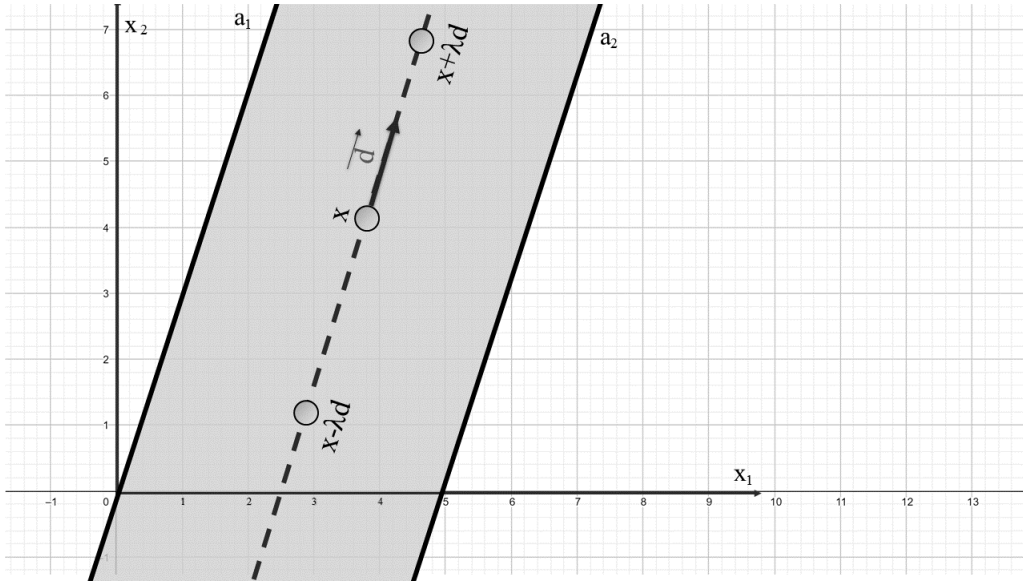


Figura 14: a_1 corresponde a la restricción $3x_1 - x_2 \geq 0$, a_2 corresponde a $3x_1 - x_2 \leq 15$, como son l.d., no hay suficientes restricciones para obtener una s.b.f.

Lo anterior implica que x^* es solución óptima para $\min_{x \in P} c^T x$, en donde, $c^T x = \sum_{i \in I} b_i$ si $a_i^T x = b_i$ para cada $i \in I$, por ende, x^* es solución única con esta propiedad $\Rightarrow x^*$ es vértice. □

Corolario 19. Dado un poliedro en \mathbb{R}^n el cual se define con m inecuaciones. Deben existir a lo más $\binom{m}{n} < \infty$ soluciones básicas.

Demostración. En cualquier solución básica hay n restricciones activas que son linealmente independientes. □

Definición 20. Un poliedro $P \subseteq \mathbb{R}^n$ **contiene una línea** si existe un vector $x \in P$ y un vector no cero $d \in \mathbb{R}^n$ tal que $x + \lambda d \in P$ para todos los escalares $\lambda \in \mathbb{R}$.

Gracias a las definiciones anteriores, se llega a la siguiente conclusión:

Teorema 21. Sea el poliedro $P = \{x \in \mathbb{R}^n | Ax \geq b\} \neq \emptyset$, las siguientes afirmaciones son equivalentes:

1. P posee al menos un punto extremo.
2. P no contiene una línea.
3. Existen n restricciones linealmente independientes.

Demostración.

2. \Rightarrow 1.

Sea $x \in P$, se define $I := \{i | a_i^T x = b_i\}$, $S := \{a_i | i \in I\}$ si $\text{rango}(S) = n$ entonces hay n vectores linealmente independientes $a_i, i \in I$.

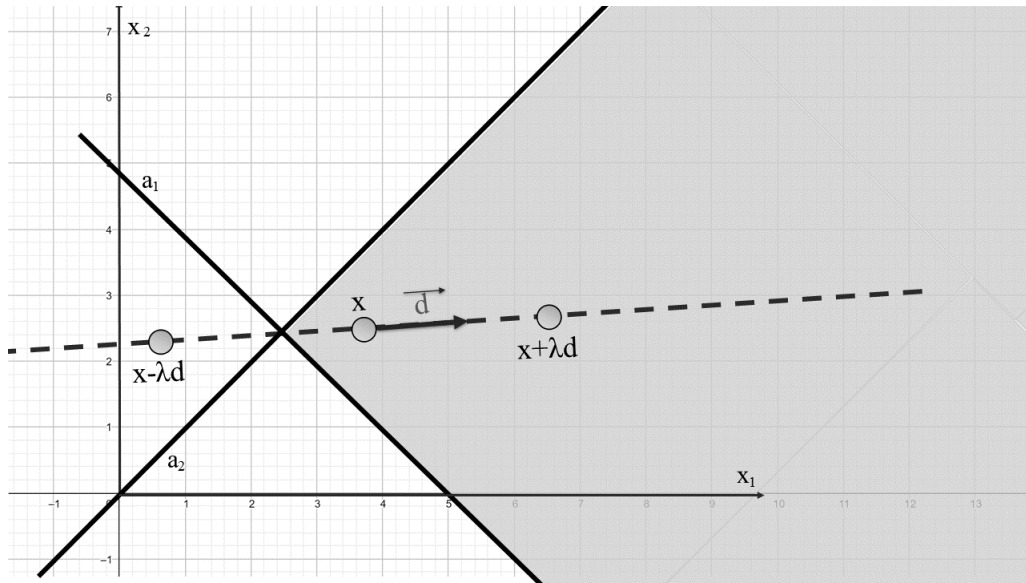


Figura 15: a_1 corresponde a la restricción $x_1 + x_2 \leq 5$, a_2 corresponde a $x_1 - x_2 \geq 0$. Como son l.i, se cruzan en algún momento y, por lo tanto, no permiten que exista una línea dentro del poliedro.

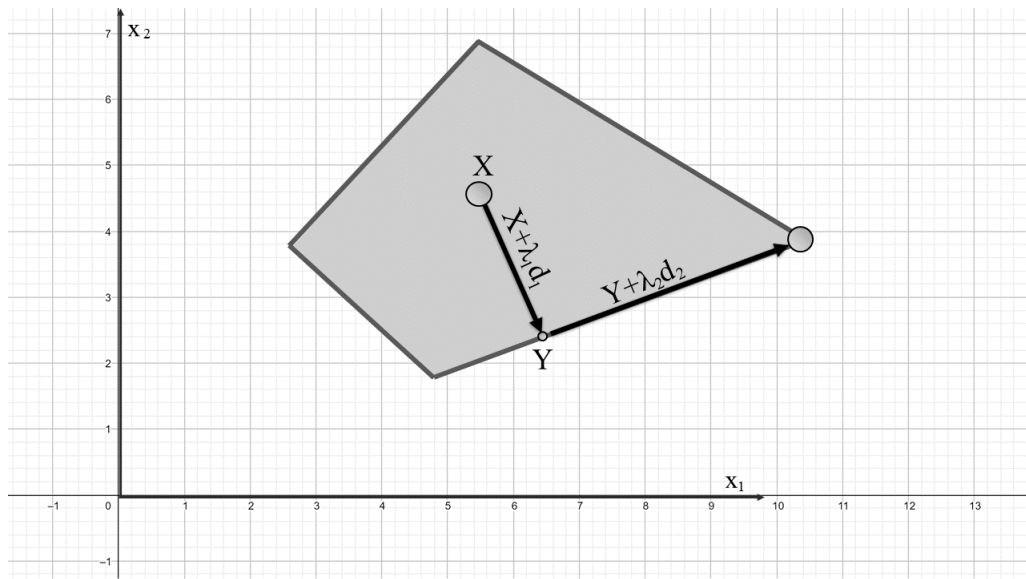


Figura 16: Se inicia en el punto arbitrario X dentro del poliedro. Se elige una dirección donde todas las restricciones actuales se mantienen activas, el movimiento sigue a lo largo de esta dirección hasta que alguna de las restricciones vaya a romperse. Aquí la cantidad de restricciones activas aumentó en al menos 1. Se repite el procedimiento hasta que se tengan n restricciones l.i. activas, en este punto se tendrá una s.b.f.

De otra forma, $\exists d \neq 0$ s.a. $a_i^T d = 0$ para cada i . Se consideran los puntos $y = x + \lambda d$ para $\lambda \in \mathbb{R}$ para todo $i \in I$, se tiene $a_i^T y = a_i^T x + \lambda a_i^T d = a_i^T x = b_i$ lo cual implica que si la restricción $i \in I$ es activa en x entonces es también activa en y . P no contiene líneas lo cual implica que $\exists \lambda^*$ y $j \notin I$ tal que $a_j^T(x + \lambda^* d) = b_j$. Se asume por contradicción que $a_j = \sum_{i \in I} \lambda_i a_i$, $a_j^T x \neq b_j$, $a_j^T(x + \lambda^* d) = b_j$ lo cual implica que $a_j^T d \neq 0$. Pero $a_i^T d = 0$ para cada $i \in I$ lo cual implica que $0 \neq a_j^T d = \sum_{i \in I} \lambda_i a_i^T d = 0$ lo cual implica que $\text{rango}(S \cup \{a_j\}) > \text{rango}(S)$. Se repite la construcción con $x + \lambda^* d$. Después de al menos n iteraciones se obtiene el set de vectores S' con $\text{rango}(S') = n$.

1. \Rightarrow 3.

Sea x un punto extremo lo cual implica que x es solución básica factible, es decir, hay n restricciones linealmente independientes que están activas en x .

3. \Rightarrow 2.

Sin pérdida de generalidad se asume que a_1, \dots, a_n son linealmente independientes. Se supone que P contiene líneas $\{x + \lambda d | \lambda \in \mathbb{R}\}$ donde $d \neq \emptyset$ lo cual implica que $a_i^T(x + \lambda d) \geq b_i$ para todos i y todos los $\lambda \in \mathbb{R}$. La restricción $a_i^T d < 0$ es violada si λ es muy grande. La restricción $a_i^T d > 0$ es violada si λ es muy pequeña (negativa) lo cual implica que $a_i^T d = 0$, así $\text{rango}(a_1, \dots, a_n) < n$, lo cual contradice que a_1, \dots, a_n son linealmente independientes. \square

Corolario 22. *Todo polígono posee al menos un punto extremo.*

Teorema 23. *Sea $P \subset \mathbb{R}^n$ un poliedro con al menos un punto extremo. Se considera el PPL $\min\{c^T x | x \in P\}$ y asume que la solución óptima existe. Entonces existe una solución óptima que es un punto extremo.*

Demostración.

Se asume $P = \{x \in \mathbb{R}^n | Ax \geq b\}$ y $v := \min\{c^T x | x \in P\}$, luego $Q := \{x \in \mathbb{R}^n | Ax \geq b, c^T x = v\}$ es también un poliedro. P no contiene líneas, lo cual implica que Q no contiene una línea, por ende, Q posee un punto extremo. Sea x^* es un punto extremo de Q . Se asume x^* no es punto extremo de P , por ende, $\exists y \in P, z \in P$ y $\lambda \in (0, 1)$ s.a. $x^* = \lambda y + (1 - \lambda)z$ lo cual implica que $v = c^T x^* = \lambda c^T y + (1 - \lambda)c^T z$

v es óptimo lo cual implica que $c^T y \geq v$ y $c^T z \geq v$. Si $c^T y > v$ entonces $v = c^T x^* = \lambda c^T y + (1 - \lambda)c^T z > v$ lo cual implica que $c^T y = v = c^T z$, por ende, $z \in Q$ y $y \in Q$, por ende, existe una contradicción, es decir, x^* es un punto extremo de P y $c^T x^* = v$. \square

Teorema 24. *Se considera el PPL $\min\{c^T x | x \in P\}$ y se supone que $P = \{x | Ax \geq b\}$ posee al menos un punto extremo. Entonces, el costo óptimo es igual a $-\infty$ o existe un punto extremo que sea óptimo.*

Demostración. ■ Sólo por el caso en que P es un polígono cada Semiespacio es cerrado, P es la intersección de un número finito de conjuntos cerrados lo cual implica que P es cerrado, luego como P es cerrado y acotado implica que P es compacto.

- $f(x) := c^T x$ es una función continua lo cual implica que $\{c^T x | x \in P\}$ es compacto, es decir, es igual a $[a, b]$ para algún $a, b \in \mathbb{R}$, donde a es el valor de la solución óptima y $f^{-1}(a) \neq \emptyset$ lo cual implica que $\exists x^*$ s.a. $c^T x^* = a$. \square

Pero, finalmente ¿De qué manera se pueden resolver estos PPLs?

- Se listan todas las soluciones básicas factibles.
- Se calcula $c^T w_i$ por cada solución básica factible w_i .
- Se selecciona la solución básica factible w^* s.a. $c^T w^* \leq c^T w_i$ por cada w_i .

Ejemplo:

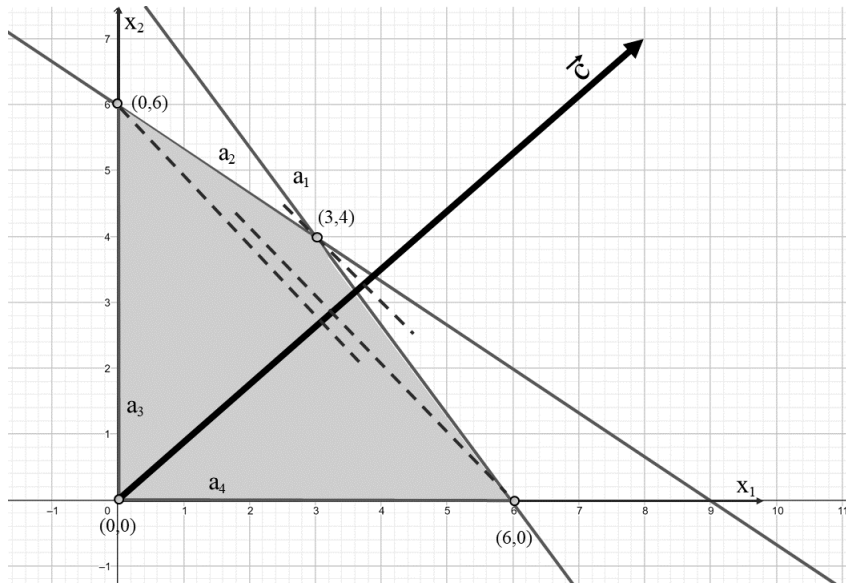


Figura 17: Siguiendo en el poliedro anterior, se tiene el vector $c = (8, 7)$ mueve el hiperplano $c^T x$ (línea punteada) hasta que no se pueda más dentro del poliedro. Se puede ver como este hiperplano muestra que la solución $(0,6)$ es peor que la solución $(6,0)$ y que esta es peor que $(3,4)$, la cual es la solución óptima del problema.

$$\begin{aligned}
 \max \quad & 8x_1 + 7x_2 \\
 & 2x_1 + 3x_2 \leq 18 \\
 & 4x_1 + 3x_2 \leq 24 \\
 & x_1 \geq 0 \\
 & x_2 \geq 0
 \end{aligned}$$

■ Siendo sus puntos extremos:

- $(0, 0) \rightarrow 0$
- $(0, 6) \rightarrow -42$
- $(6, 0) \rightarrow -48$
- $(3, 4) \rightarrow -52$, el cual es un punto extremo óptimo

■ Funciona solamente si un óptimo existe, es decir, el PPL no es acotado.

$$\begin{aligned}
 \min \quad & -8x_1 - 7x_2 \\
 & 2x_1 + 2x_2 \geq 8 \\
 & 2x_1 + x_2 \geq 6 \\
 & x_1, x_2 \geq 0
 \end{aligned}$$

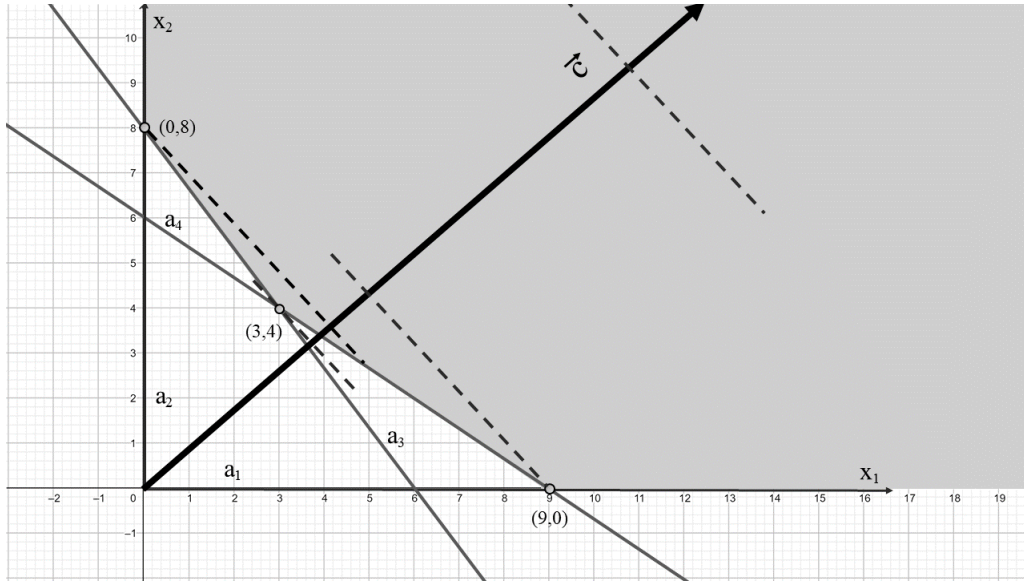


Figura 18: Vector $c = (8, 7)$ el cual mueve el hiperplano $c^T x$ (línea punteada) hasta que no se pueda más dentro del poliedro, en este caso no hay restricción que detenga el movimiento, es decir, es no acotado.

■ Siendo sus puntos extremos:

- $(3, 4) \rightarrow -52$
- $(0, 8) \rightarrow -56$
- $(9, 0) \rightarrow -72$

■ Pero

- $y = (3, 4)^T + \lambda(1, 1)^T$ es factible para cada $\lambda \geq 0$
- $c^T y = -8(3 + \lambda) + (-7)(4 + \lambda) = -52 - 15\lambda \Rightarrow$ Solución Óptima = $-\infty$.

4.4. Degenerancia

Como se explicó en los capítulos anteriores, una solución básica factible x posee n restricciones activas que sean linealmente independientes. Cabe destacar que también es posible que existan más de n restricciones activas en x , pero no todas ellas linealmente independientes.

Ejemplo :

$$\begin{aligned} 2x_1 + 3x_2 &\leq 18 \\ 4x_1 + 3x_2 &\leq 24 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \\ x_1 &\leq 6 \end{aligned}$$

En el ejemplo anterior el punto $(6, 0)$ es una solución básica que posee $3 > 2 = n$ restricciones activas.

Definición 25. Una solución básica $x \in \mathbb{R}^n$ es **degenerada** si más de n restricciones están activas en x . Cabe destacar que su estudio es relevante, ya que son problemáticas al momento de trabajar con ellas como se verificará más adelante.

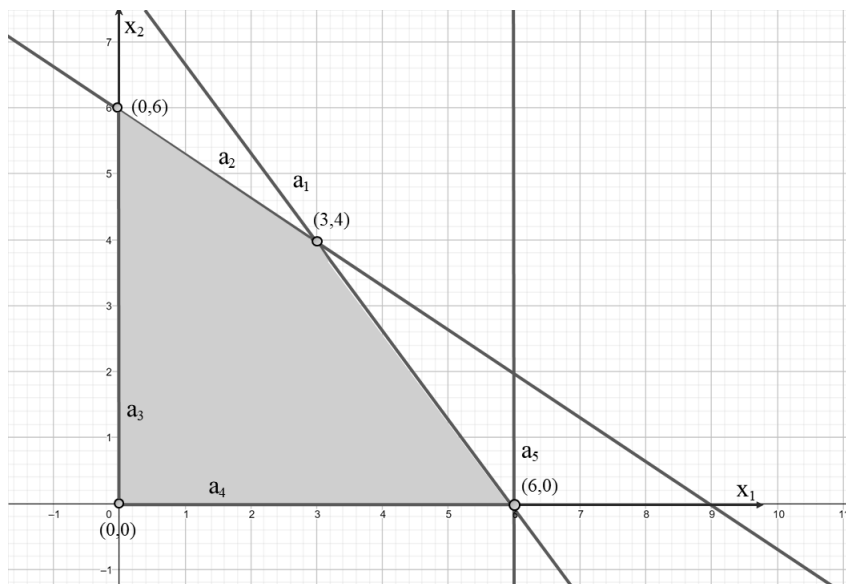


Figura 19: a_1 corresponde a la restricción $4x_1 + 3x_2 \leq 24$, a_2 corresponde a $2x_1 + 3x_2 \leq 18$, a_3 corresponde a $x_1 \geq 0$, a_4 corresponde a $x_2 \geq 0$. Si se agrega la restricción a_5 que corresponde a $x_1 \leq 6$ se obtiene que el punto $(6, 0)$ pasa a ser degenerado, pues tiene 3 restricciones activas, a_5 , a_4 y a_1 .

5. Método Simplex

El método Simplex es utilizado para resolver PPLs, el cual se puede realizar tanto a mano como con herramientas computacionales específicas para este objetivo. El algoritmo trabaja moviéndose en las soluciones básicas factibles del problema, es decir, sus vértices o puntos extremos como sigue:

1. Comienza en vértice arbitrario analizando esta solución según la función objetivo.
2. Si existe algún vértice adyacente que sea mejor al actual, el algoritmo se mueve hacia ese.
3. Se repite el proceso hasta encontrar el vértice óptimo.

5.1. PPLs en forma estándar

Hasta ahora se han estudiado PPLs de la forma $\min c^T x$ s.a. $Ax \geq b$, pero al momento de comenzar con Simplex, se requiere que el PPL esté en **forma estándar (FE)**: $\min\{c^T x : Ax = b, x \geq 0\}$.

Ejemplo:

$$\begin{array}{rcl} 2x_1 + 3x_2 & \leq & 18 \\ 4x_1 + 3x_2 & \leq & 24 \\ x_1, x_2 & \geq & 0 \end{array}$$

Se agregan *variables de Holgura*, las cuales sirven para cambiar las restricciones de desigualdades a igualdades sin que la restricción pierda sentido. Cabe destacar que la naturaleza de estas variables se define mayor o

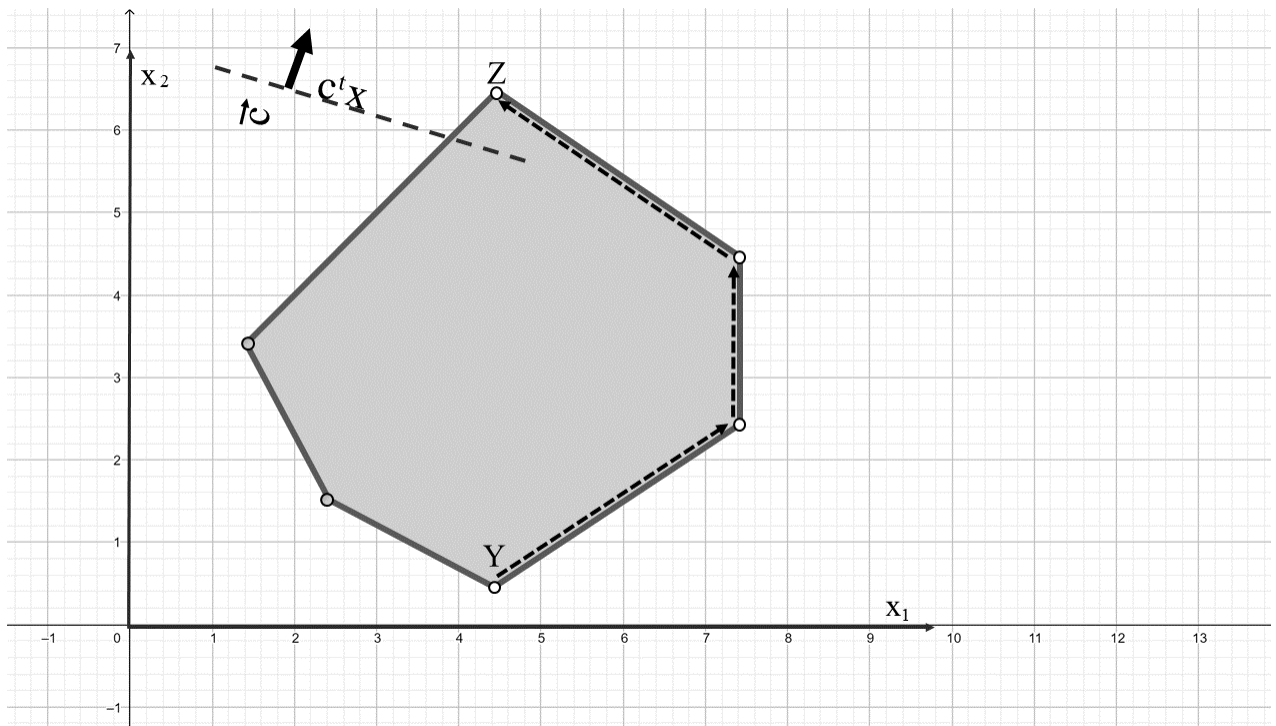


Figura 20: Se puede observar en la figura el movimiento desde un vértice arbitrario Y a vértice adyacente a través de un vector, el cual es la dirección del movimiento, hasta que se logra que el vértice es la solución óptima, en este caso, Z .

igual a cero como sigue:

$$\begin{array}{rcl} 2x_1 + 3x_2 + x_3 & = & 18 \\ 4x_1 + 3x_2 & + x_4 = & 24 \\ x_1, x_2, x_3, x_4 & \geq & 0 \end{array}$$

Ejemplo:

$$\begin{array}{rcl} -x_1 + 2x_2 & \leq & 8 \\ 3x_1 + 2x_2 & \geq & 0 \\ -x_1 + x_2 & \geq & -2 \\ x_1 + 4x_2 & \leq & 22 \\ x_2 & \geq & 0 \end{array}$$

Se agregan las variables de holgura x_3, \dots, x_6 como se explicó en el ejemplo anterior, se analiza la variable x_1 , la cual debido a su dominio podría ser negativa, es por ello que se incluyen dos nuevas variables x_1^+ y x_1^- , las cuales reemplazan a x_1 . Al mismo tiempo se deben agregar la naturaleza de variables como sigue $x_1^+ \geq 0$ y $x_1^- \geq 0$.

Teniendo el PPL resultante de la siguiente manera:

$$\begin{array}{rcl} -x_1^+ + x_1^- + 2x_2 + x_3 & = & 8 \\ 3x_1^+ - 3x_1^- + 2x_2 & - x_4 & = 0 \\ -x_1^+ + x_1^- + x_2 & - x_5 & = -2 \\ x_1^+ - x_1^- + 4x_2 & + x_6 & = 22 \\ x_1^+, x_1^-, x_2, x_3, x_4, x_5, x_6 & \geq & 0 \end{array}$$

Teniendo claros los conceptos anteriores, se puede llegar a la pregunta siguiente: ¿Cómo se ven los puntos extremos de esos tipos de PPLs?

Se asume que las filas de $A \in \mathbb{R}^{m \times n}$ son linealmente independientes, en donde, m son las filas y n las columnas, es decir, A posee rango m , lo cual implica que $m \leq n$, por ende, todo poliedro no vacío $P = \{x | Ax = b, x \geq 0\}$ posee un punto extremo.

5.1.1. Puntos extremos en PPLs en forma estándar

Si x es una solución básica, se tiene que existen n restricciones linealmente independientes de igualdad activas, lo cual implica que m restricciones de igualdad deben estar activas, pero entonces ¿Cómo se eligen las $n - m$ restricciones activas restantes?

1. Escoger m columnas linealmente independientes $A_{B(1)}, \dots, A_{B(m)}$, en donde, A_j es la j -ésima columna de A , a_i^T es i -ésima fila de A .
 - a) Las variables $x_B = (x_{B(1)}, \dots, x_{B(m)})$ son llamadas **variables básicas**, las cuales forman una **Base** $B = (B(1), \dots, B(m))$.
 - b) Todas las otras variables son llamadas **variables no básicas**, denotadas por x_N .
2. Se tiene $x_j := 0$ para todas las variables no básicas x_j en las $n - m$ restricciones activas restantes.
3. Finalmente se resuelve el sistema de ecuaciones $A_B x_B = b$ para los valores de $x_{B(1)}, \dots, x_{B(m)}$. Por cada solución básica x existe una base $B = (B(1), \dots, B(m))$ tal que $x = (x_B, x_N)$, x es factible si $x_B \geq 0$.

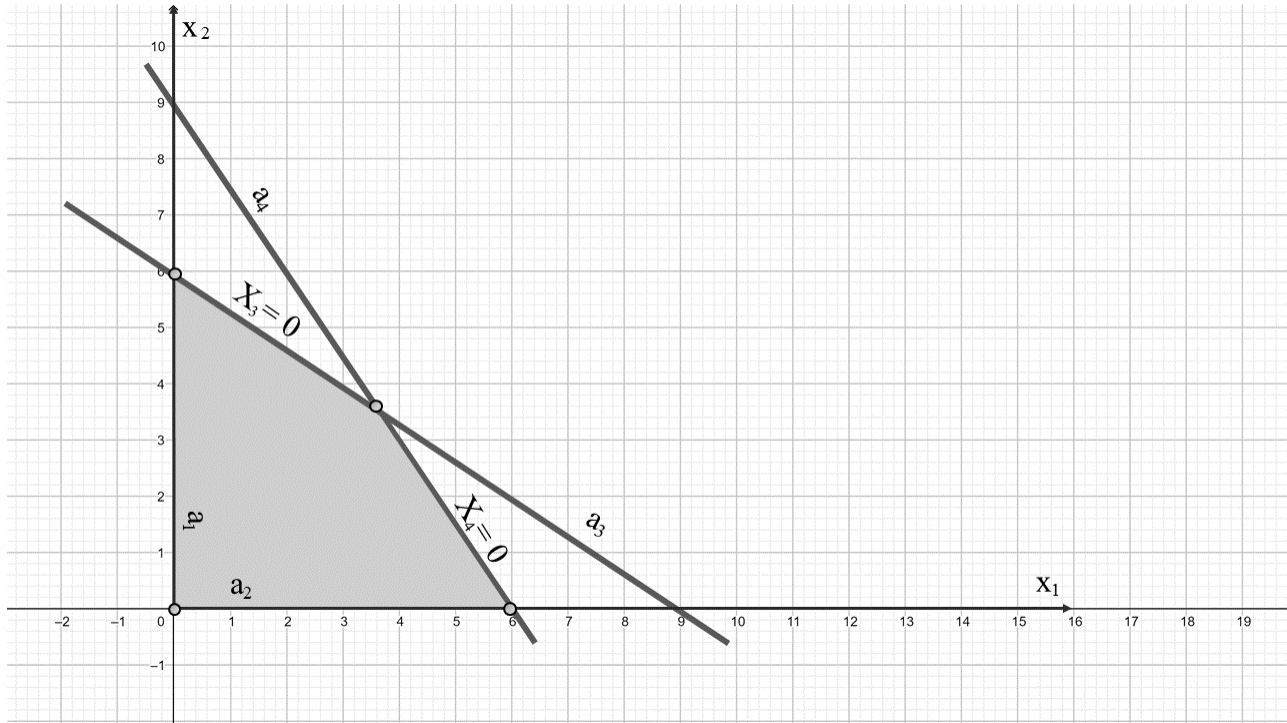


Figura 21: El poliedro original tiene 2 variables, por lo tanto como pertenece a \mathbb{R}^2 , es graficable. Al pasarlo a forma estándar aumenta su dimensionalidad por cada restricción tipo $a^t x$ que no sea igual a b , ya que estas restricciones tendrán variables de holgura, por lo tanto si continuamos con el mismo ejemplo, los puntos que ahora representan el poliedro en forma estándar pertenecerían a \mathbb{R}^4 , lo que claramente no es graficable. Para representar gráficamente el poliedro en forma estándar se inicia con el gráfico del poliedro en \mathbb{R}^2 , pero ahora en vez de tener relleno" dado por las desigualdades de las restricciones, se tiene relleno" dado por las variables de holgura al tener más grados de libertad, así que se debe indicar dónde estas valen 0. Estas variables valen 0 si se sitúan sobre su restricción asociada, es decir, que se tenía un punto que activaba la restricción en el poliedro original, entonces en el poliedro en forma estándar también la activa, y como este punto ahora incluye el valor de las variables de holgura, el valor de la variable de holgura asociada a la restricción que activa será 0. En este caso a_3 corresponde a la restricción $2x_1 + 3x_2 + x_3 = 18$ y sobre esta x_3 vale 0, a_4 corresponde a la restricción $4x_1 + 3x_2 + x_4 = 24$ y sobre esta x_4 vale 0, a_1 y a_2 corresponden a $x_1, x_2 \geq 0$ y $x_3, x_4 \geq 0$ no tiene representación gráfica.

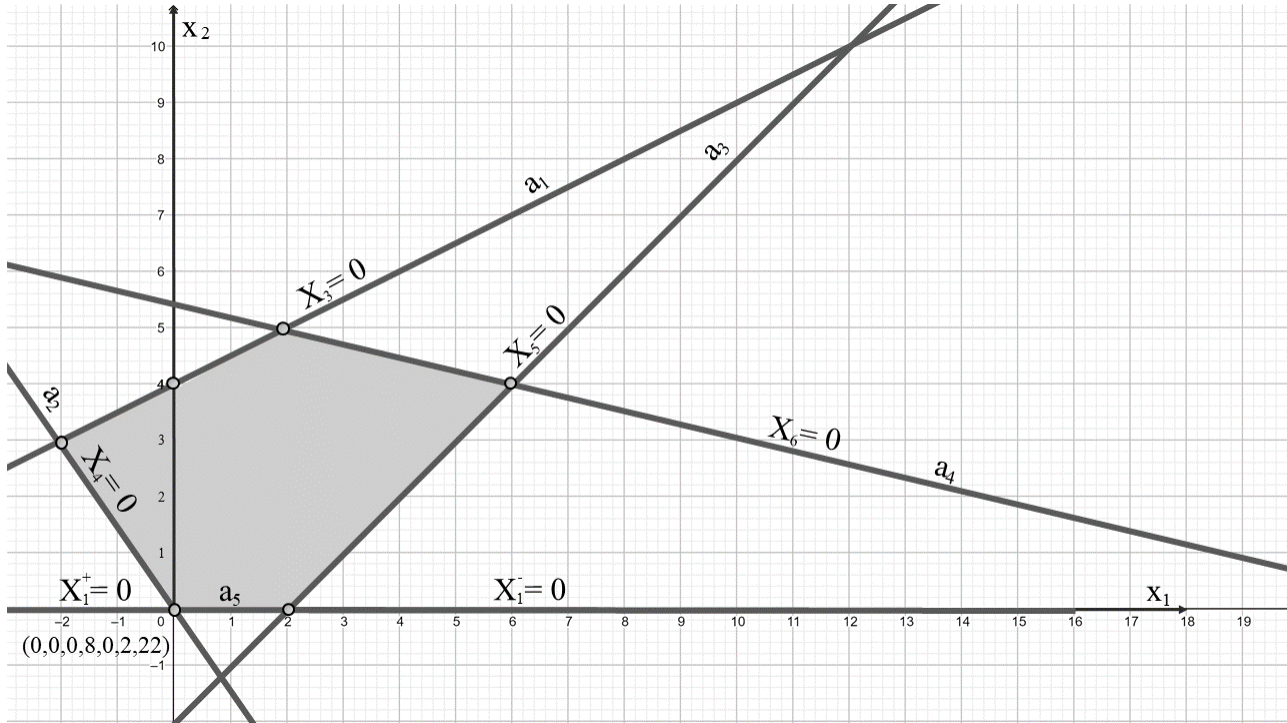


Figura 22: a_1 corresponde a la restricción $-x_1^+ + x_1^- + 2x_2 + x_3 = 8$, a_2 corresponde a la restricción $3x_1^+ - 3x_1^- + 2x_2 - x_4 = 0$, a_3 corresponde a la restricción $-x_1^+ + x_1^- + x_2 - x_5 = -2$ y a_4 corresponde a la restricción $x_1^+ - x_1^- + 4x_2 + x_6 = 22$, se deben considerar además las restricciones de positividad de las variables $x_1^+, x_1^-, x_2, x_3, x_4, x_5, x_6 \geq 0$. Se agregan las variables x_1^+ y x_1^- , donde su resta representa la libertad de la variable original. Esta representación es un cambio de variable, el cual corresponde a $x_1^+ - x_1^- = x_1$ y se tiene que x_1^+ vale 0 en el cuadrante negativo y x_1^- vale 0 en el cuadrante positivo de x_1 . El punto $(0, 0, 0, 8, 0, 2, 22)$ es factible en este poliedro. Una manera de obtenerlo es reemplazando el valor del punto en el poliedro original, es decir $(0,0)$, en las restricciones del poliedro en forma estándar, y así despejar los valores de las variables de holgura. Este método tiene el problema de requerir un punto en el poliedro original, cuando este no sea el caso se utilizará el método que se mostrará más adelante.

Ejemplo: Si se verifica el mismo PPL anterior se tiene, variables básicas x_1^+, x_2, x_3, x_4 y variables no básicas x_1^-, x_5, x_6 . Entonces $x_N = 0$, por ende, $x_1^- = x_5 = x_6 = 0$ para x_B se resuelve:

$$A_B x_B = b$$

$$\Leftrightarrow \begin{pmatrix} -1 & 2 & 1 & 0 \\ 3 & 2 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ 1 & 4 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^+ \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \\ -2 \end{pmatrix}$$

En donde, $x_1^+ = 6, x_2 = 4, x_3 = 10, x_4 = 18$, por ende, sus vértices son $(6, 0, 4, 10, 18, 0, 0)$.

Ejemplo: Bases $x_B = (x_1^+, x_3, x_4, x_5)$ y $x_N = (x_1^-, x_2, x_6)$

$$\begin{pmatrix} -1 & 1 & 0 & 0 \\ 3 & 0 & -1 & 0 \\ -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^+ \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \\ -2 \\ 22 \end{pmatrix}$$

En donde, $x_1^+ = 22, x_3 = 30, x_4 = 66, x_5 = -20$, es decir, entrega la solución básica $(22, 0, 0, 30, 66, -20)$, pero no es solución básica factible.

Ejemplo: Bases $x_B = (x_2, x_3, x_4, x_5)$ y $x_N = (x_1^+, x_1^-, x_6)$

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 2 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 4 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \\ -2 \\ 22 \end{pmatrix}$$

En donde, $x_2 = 4, x_3 = 8, x_4 = 5,5, x_5 = 3,5$. Finalmente la solución es: $x = (0, 0, 4, 8, 5,5, 3,5, 0)$.

5.2. Moverse a un nuevo vértice

Si se encontrara en un vértice x que no fuera la solución óptima, ¿Hacia dónde se debería mover para mejorar la solución? Eso es lo que se podrá verificar en el siguiente capítulo.

Definición 26. Sea $x \in P$. Un vector d es una dirección factible de x si existe un $\theta > 0$ tal que $x + \theta d \in P$.

¿Cómo se encuentra un d tal que se pueda mover de un vértice a otro?

Ahora se debe mover desde x hacia un nuevo vértice $x + \theta d$, aumentando así el valor de la variable no básica x_j con $d_j = 1$. Luego dejar todas las otras variables no básicas en 0 con $d_i = 0$ para cada variable no básica $x_i \neq x_j$, en donde, d_B denota los componentes de d que corresponden a variables en x_B , d_N denota los componentes de d que corresponden a variables en x_N .

¿Cómo se ve d ?

$$d_N = (0, 0, \dots, 0, \underset{\text{Lugar que corresponde a } x_j}{1}, 0, \dots, 0) \text{ y } d_B = (d_{B(1)}, \dots, d_{B(m)}). \text{ Se define } d_B = -A_B^{-1}A_j.$$

¿Por qué?

Sabemos que $Ax = b$ y requerimos $b = A(x + \theta d) = Ax + \theta Ad$. Por ende, se necesita que $Ad = 0$, $0 = Ad = \sum_{i=1}^n A_i d_i = A_j + \sum_{i=1}^m A_{B(i)} d_{B(i)} = A_B d_B + A_j$, lo cual implica que $d_B = -A_B^{-1}A_j$. En el algoritmo de Simplex, PPL en **Forma estándar**: $\min\{c^T x : Ax = b, x \geq 0\}$, en donde A tiene m filas y n columnas, la base posee m columnas linealmente independientes $(A_{B(1)}, \dots, A_{B(m)}) = A_B$, por ende, su solución básica es $x: x_B = A_B^{-1}b, x_N = 0$, en donde x es factible con $x_B \geq 0$, por ende, el punto x se moverá

a $x + \theta d$.

El aumento de la variable no básica x_j implica $d_j = 1$. Mantener las otras variables no básicas en cero implica $d_i = 0$ para cada variable no básica $x_i \neq x_j$.

lamamos d^j al vector j -ésima dirección básica de x . **Ejemplo:**

$$\begin{aligned} \text{mín } & -4x_1 - 3x_2 \\ & 2x_1 + 2x_2 + x_3 & = 8 \\ & 2x_1 + x_2 & + x_4 & = 6 \\ & 20x_1 + x_2 & & + x_5 = 60 \\ & & & & x_i \geq 0 \quad \forall i = 1, \dots, 5 \end{aligned}$$

Se comienza con una base $B = (3, 4, 5)$, en donde $x_B = (x_3, x_4, x_5)$ y $x_N = (x_1, x_2)$ con $A_B = I = A_B^{-1}$

$$x_B = A_B^{-1}b = I \cdot \begin{pmatrix} 8 \\ 6 \\ 60 \end{pmatrix} = \begin{pmatrix} 8 \\ 6 \\ 60 \end{pmatrix}$$

Se ve que $x = (0, 0, 8, 6, 60)^T$ en donde se mueve desde allí en dirección $d = d^2$, $d = (d_B | d_N)$ y $d_N = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ($d_1 = 0, d_2 = 1$)

$$\begin{aligned} Ad &= 0 \\ \Rightarrow A_B d_B + A_N d_N &= A_B d_B + A_2 \cdot 1 = 0 \\ \Rightarrow d_B &= -A_B^{-1}A_2 = -I \cdot \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ -1 \end{pmatrix} \\ \Rightarrow d &= \begin{pmatrix} 0 \\ 1 \\ -2 \\ -1 \\ -1 \end{pmatrix} \end{aligned}$$

5.2.1. Degenerancia y Forma estándar

Cada base B posee una solución básica x , pero ¿Puede la solución básica x tener bases diferentes?

Ejemplo (en Forma estándar):

$$\begin{aligned} & 2x_1 + 3x_2 + x_3 & = 18 \\ & 4x_1 + 3x_2 & + x_4 & = 24 \\ & & x_2 & + x_5 = 6 \\ & & & & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{aligned}$$

Base $B = (1, 2, 4)$ y luego $A_B^{-1}x = b$ lleva a $\begin{pmatrix} 2 & 3 & 0 \\ 4 & 3 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} 18 \\ 24 \\ 6 \end{pmatrix}$, lo cual implica que $x_B =$

$(x_1, x_2, x_4) = (0, 6, 6)$ y $x_N = (x_3, x_5) = (0, 0)$ con m restricciones de igualdad son activas y $n - m$ restricciones de no-negatividad y de variables no básicas son activas, en donde, 1 restricción de no negatividad de variables básicas es activa, $m + (n - m) + 1 = n + 1$ restricciones activas con el punto $(0, 6, 0, 6, 0)$ es degenerado, por ende, un punto extremo es degenerado si la variable básica es 0, usualmente posee diferentes bases, en el caso del punto $(0, 6, 0, 6, 0)$ el cual tiene la base $B = (2, 3, 4)$ y luego $x_B = (x_2, x_3, x_4) = (6, 0, 6)$ y $x_N = (x_1, x_5) = (0, 0)$, al mismo tiempo el punto $(0, 6, 0, 6, 0)$ tiene la base $B = (2, 4, 5)$ y luego $x_B = (x_2, x_4, x_5) = (6, 6, 0)$ y $x_N = (x_1, x_3) = (0, 0)$.

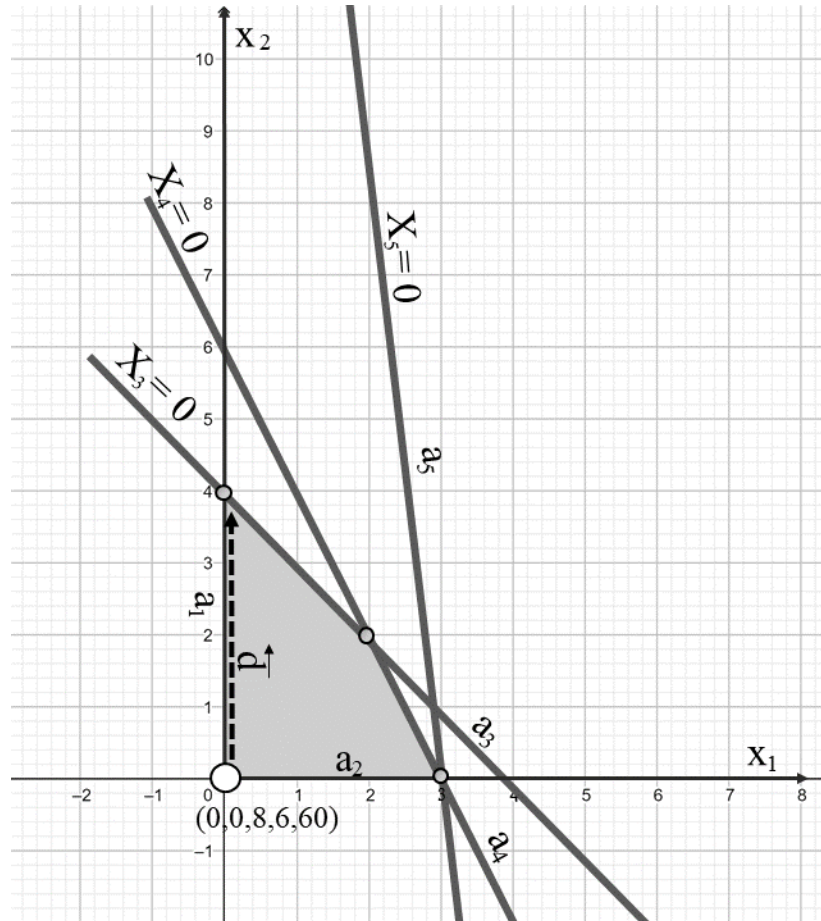


Figura 23: a_3 corresponde a la restricción $2x_1 + 2x_2 + x_3 = 8$, a_4 corresponde a la restricción $2x_1 + x_2 + x_4 = 6$, a_5 corresponde a la restricción $-x_1^+ + x_1^- + x_2 - x_5 = -2$ y a_4 corresponde a la restricción $20x_1 + x_2 + x_5 =$, se deben considerar además las restricciones de positividad de las variables $x_1, x_2, x_3, x_4, x_5, \geq 0$. Se puede observar la dirección factible anteriormente mencionada d permite trasladarse a un vértice adyacente, esta se conoce como dirección factible.

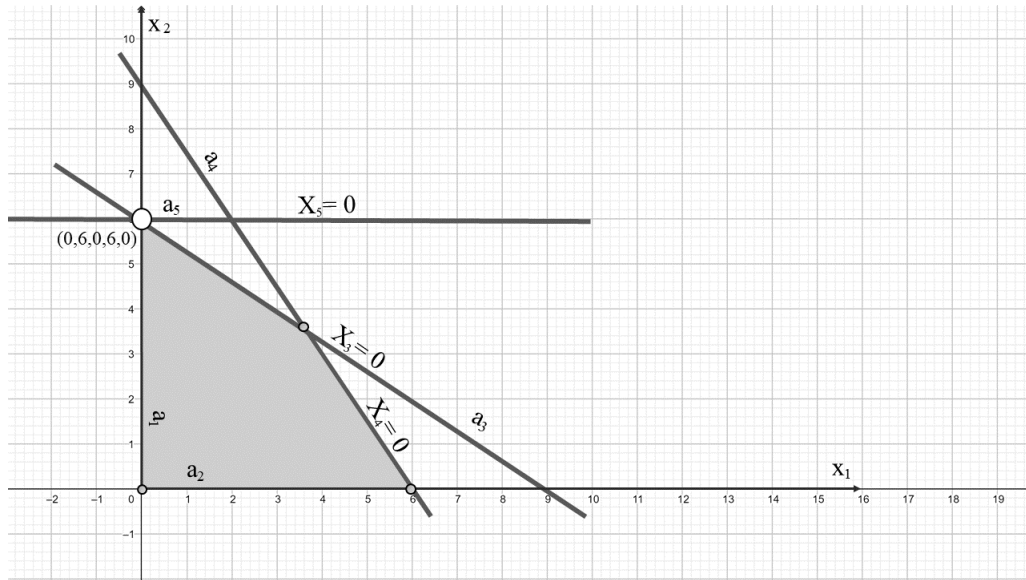


Figura 24: a_3 corresponde a la restricción $2x_1 + 3x_2 + x_3 = 18$, a_4 corresponde a la restricción $4x_1 + 3x_2 + x_4 = 24$, a_5 corresponde a la restricción $x_2 + x_5 = 6$, a_1 y a_2 corresponden a $x_1, x_2 \geq 0$, $x_3, x_4, x_5 \geq 0$. En el punto $(0, 6, 0, 6, 0)$ hay 6 restricciones activas de las 5 necesarias para caracterizar una s.b.f., de las cuales 3 restricciones activas vienen dadas por el poliedro y 3 restricciones activas vienen dadas por las restricciones de no negatividad a_1 , a_3 y a_5 . Ese punto es degenerado, ya que normalmente se tienen $n - m = 2$ variables iguales a 0 pero ahora $n - m = 3$, por lo tanto, una de nuestras variables básicas será igual a 0, y como puede ser cualquiera de esas 3 para definir la base, se tienen 3 bases distintas para el mismo punto. El punto $(0, 6)$ en el poliedro en forma normal también sería degenerado, ya que hay más de $n = 2$ restricciones activas en ese punto para el poliedro normal.

5.2.2. Cambio de costos al moverse en dirección d^j

¿Es buena idea moverse en la dirección d^j ? ¿Qué pasa con la función de costos si se mueve desde x a $x + \theta d^j$?

Si $\theta = 1$ el nuevo costo es $c^T(x + \theta d^j) = c^T x + c^T d^j$, lo cual implica que los costos crecen a $c^T d^j = (c_B^T d_B^j + c_j) = c_j - c_B^T A_B^{-1} A_j =: \bar{c}_j$ donde $c_B = (c_{B(1)}, \dots, c_{B(m)})$.

Definición 27. Sea B la base de la solución básica x . Se llama a $\bar{c}_j := c_j - c_B^T A_B^{-1} A_j$ los **costos reducidos** de la variable x_j .

- Si $\bar{c}_j < 0$ la función objetivo crece al moverse en d^j .

Lema 28. Sea B la base de una solución básica x e $i \in [m]$. Entonces $\bar{c}_{B(i)} = 0$.

Se vuelve al ejemplo:

$$\begin{aligned} \text{mín } & -4x_1 - 3x_2 \\ & 2x_1 + 2x_2 + x_3 = 8 \\ & 2x_1 + x_2 + x_4 = 6 \\ & 20x_1 + x_2 + x_5 = 60 \\ & x_i \geq 0 \quad \forall i = 1, \dots, 5 \end{aligned}$$

En donde, $x_B = (x_3, x_4, x_5)$, $x_N = (x_1, x_2)$ y así $c_B = (0, 0, 0)$, $c_N = (-4, -3)$

$$\bar{c} = (\bar{c}_1, \bar{c}_2, \bar{c}_3, \bar{c}_4, \bar{c}_5) = (-4, -3, 0, 0, 0) - (0, 0, 0) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 10 & 1 & 0 & 0 & 1 \end{pmatrix} = (-4, -3, 0, 0, 0)$$

$\bar{c}_N = (-4, -3)$, $\bar{c}_B = (0, 0, 0)$ (como el Lema prometió), lo cual implica que si se mueve en dirección d^1 o d^2 se mejora la función objetivo. Costos reducidos \bar{c} : Si $\bar{c} < 0$ los costos mejoran cuando se mueve en d^j . Si $\bar{c} \geq 0$ ninguna de las direcciones d^j mejora la función objetivo.

Teorema 29. Sea x solución básica factible con vector de costos reducidos \bar{c} :

- Si $\bar{c} \geq 0$ luego x es óptimo.
- Si x es óptimo y no degenerado entonces $\bar{c} \geq 0$.

Demostración. a) Se asume que $\bar{c} \geq 0$, sea y una solución factible y $d = y - x$, $Ax = Ay = b$ y así $Ad = 0$. Sea B la base de x , con $N = [n] \setminus B$ $0 = Ad = A_B d_B + \sum_{i \in N} A_i d_i$, es decir, $d_B = -\sum_{i \in N} A_B^{-1} A_i d_i$. ¿Cómo cambia el costo cuando se cambia de x a y ? $c^T y - c^T x = c^T (y - x) = c^T d$, en donde, $c^T d = c_B^T d_B + \sum_{i \in N} c_i d_i = c_B^T (-\sum_{i \in N} A_B^{-1} A_i d_i) + \sum_{i \in N} c_i d_i = \sum_{i \in N} (c_i - c_B^T A_B^{-1} A_i) d_i = \sum_{i \in N} \bar{c}_i d_i$, por cada $i \in N$ se tiene que $x_i = 0$ y $y_i \geq 0$, lo cual implica que $d_i \geq 0$, por ende, $\bar{c}_i d_i \geq 0$ para todo $i \in N$, lo que implica que $c^T d \geq 0$, por ende, $0 \leq c^T d = c^T (y - x)$, por ende, $c^T x \leq c^T y$, finalmente x es óptimo.

□

Demostración. b) Se supone que $\bar{c}_j < 0$ para algún j , entonces x_j debe ser una variable no básica. $x_B > 0$ con x es no degenerada y $x_N = 0$ y $d_N^j \geq 0 \Rightarrow \exists \theta > 0$ s.a. $x + \theta d^j \geq 0$, por ende, $A(x + \theta d^j) = Ax + \theta A d^j = Ax = b$, lo que implica que el vector $x + \theta d^j$ es factible con $c^T(x + \theta d^j) = c^T x + \bar{c}_j \theta < c^T x$ y así x no es óptimo.

□

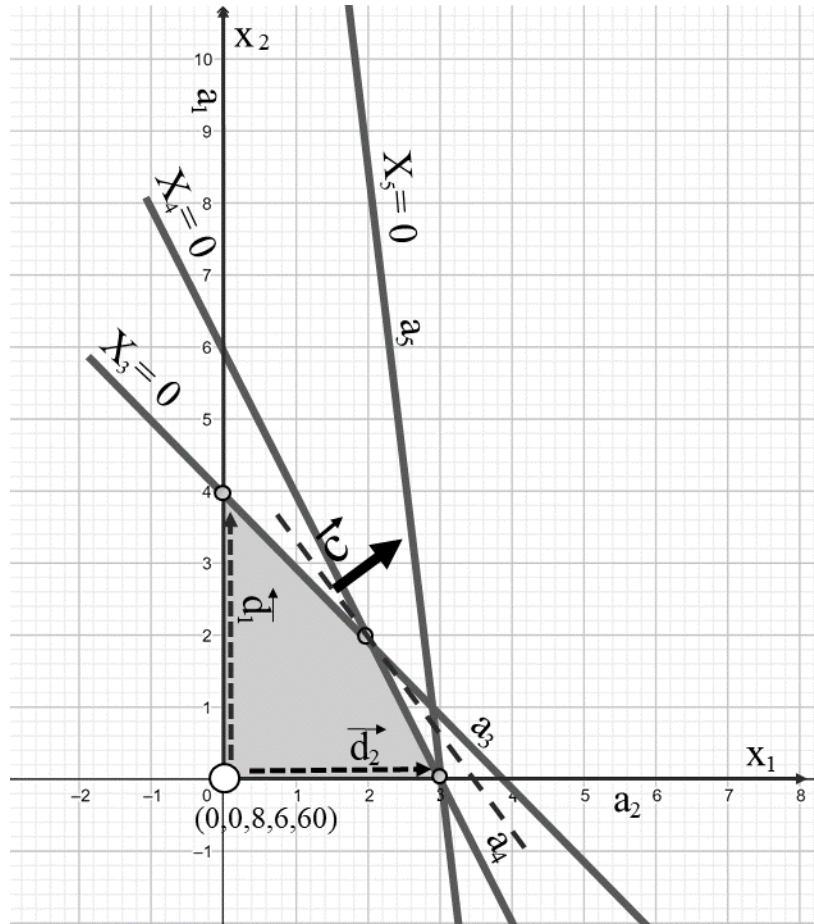


Figura 25: La función objetivo es $\text{mín } -4x_1 - 3x_2$, lo que implica que el vector de mejora de la función objetivo es en dirección $(4, 3)$. Las direcciones desde el origen son d_1 y d_2 , con las cuales se puede desplazar a alguno de los 2 puntos adyacentes, y a favor del crecimiento de la función objetivo, pues el óptimo mejora tanto aumentando x_1 como x_2 , cumpliendo lo que señala el teorema de los costos reducidos positivos.

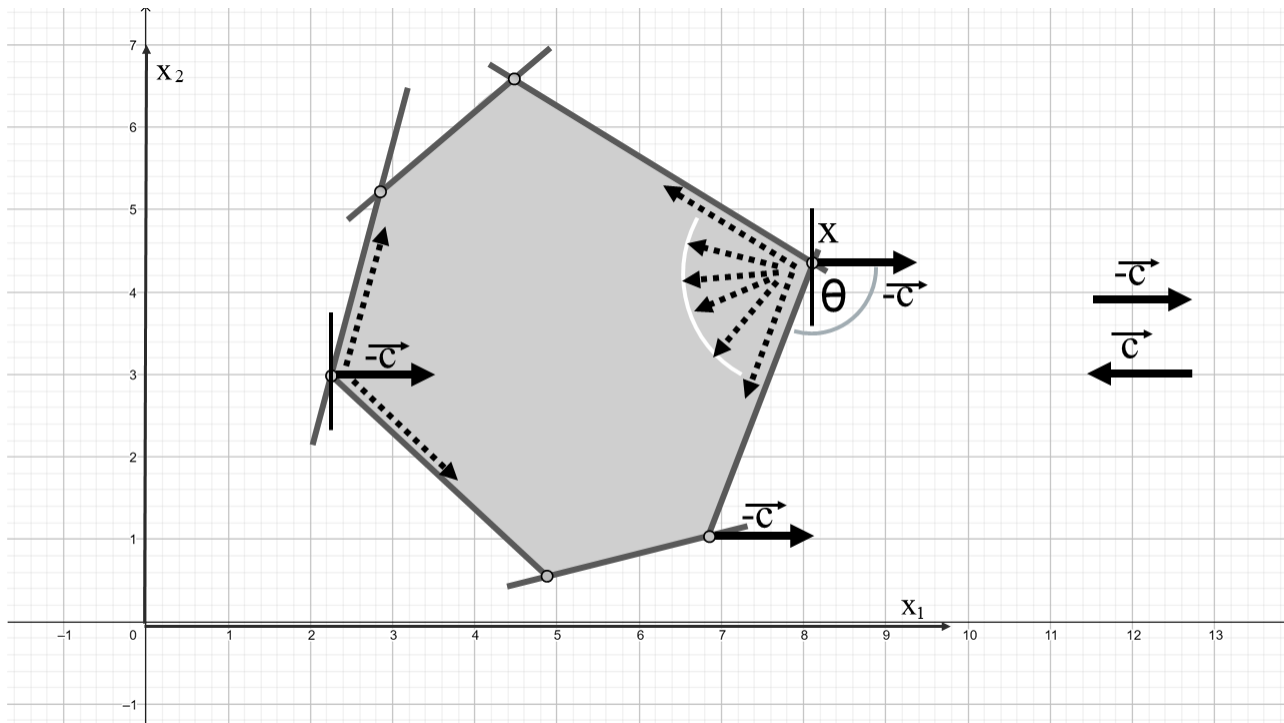


Figura 26: Una forma gráfica de observar el teorema anterior, si $\bar{c} \geq 0$ luego x es óptimo, es con el siguiente ejemplo. Se tiene la función objetivo $\min c^t x$, entonces el gradiente de esta es c , pero esa no es la dirección de crecimiento del problema, si no que es $-c$. Dado esto, para mejorar el valor óptimo, se debe mover siguiendo su vector de crecimiento. Por lo tanto, en el instante en que no hay ninguna dirección factible a favor de este, ese punto será el óptimo. Matemáticamente se puede observar al calcular el producto punto de $-c^t d$, lo cual sería $|c||d| \cos(\lambda)$ siendo λ el ángulo entre ambos vectores. Este producto punto será positivo, es decir, se moverá la dirección de crecimiento, si θ se encuentra en los cuadrantes a la derecha de la recta vertical que se colocó en el punto, por lo tanto, si λ se encuentra a la izquierda de esta recta, entonces el producto punto sería negativo y no hay direcciones para mejorar el valor óptimo. Desarrollando ese resultado se obtiene que $-c^t d \leq 0$, lo que es equivalente a $c^t d \geq 0$.

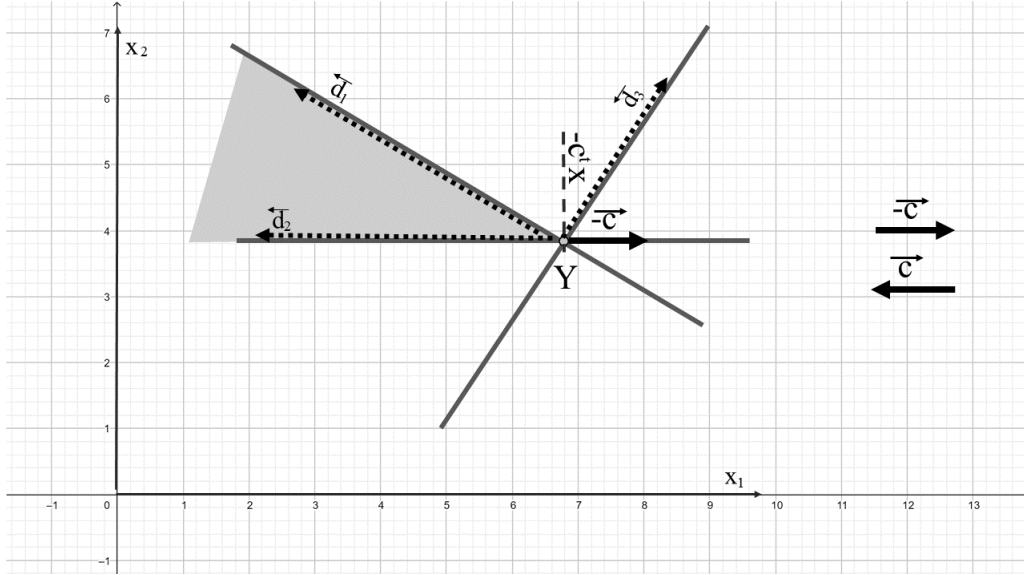


Figura 27: Sea el punto Y , el cual es el óptimo, por lo tanto, debería cumplir el criterio $c^T d \geq 0$ para toda dirección en el cono de direcciones factible del punto Y . El problema es que existe una dirección factible en específico, d_3 , donde se tiene que $c^T d \leq 0$, lo que implica que utilizando esa dirección aún se puede mejorar la función objetivo. Esta dirección tiene un tema en específico, el cual es que corresponde a moverse a través de la restricción degenerada, pero esto no es posible, ya que esos puntos no son factibles dentro del poliedro. Es por esto que no se utiliza este criterio en puntos con restricción degenerada.

Definición 30. Una base B es **óptima** si se cumple que $A_B^{-1}b \geq 0$ y $\bar{c}^T = c^T - c_B^T A_B^{-1}A \geq 0$.

- Cabe destacar que una base óptima implica una solución básica factible que es óptima.
- Sea x una solución básica factible, la cual se mueve en una dirección d^j con $\bar{c}_j < 0$ ¿Cuánto se deberá mover? Se deberá mover lo más lejos posible en donde el costo disminuya en d^j , es decir, se mueve al punto $x + \theta^* d^j$ donde $\theta^* = \max\{\theta \geq 0 \mid x + \theta d^j \in P\}$.
- Pero entonces la pregunta es ¿Cómo se calcula θ^* ? Por cada θ es $x + \theta d^j \in P$ $A(x + \theta d^j) = Ax + \theta \underbrace{A d^j}_{=0} = b$ para cualquier θ con restricciones no negativas. Teniendo así los siguientes casos:
 - Si $d^j \geq 0$ así $x + \theta d^j \geq 0$ para cualquier $\theta \geq 0$ y así el PPL es no acotado, se debe terminar el algoritmo que se está ocupando para encontrar el óptimo.
 - Si $d^j \not\geq 0$: $\theta^* = \max\{\theta \geq 0 \mid x_i + \theta d_i^j \geq 0 \forall i\}$, por cada i con $d_i^j < 0$ es necesario que $x_i + \theta d_i^j \geq 0 \Leftrightarrow \theta d_i^j \geq -x_i \Leftrightarrow \theta \leq -\frac{x_i}{d_i^j}$, lo cual implica que $\theta^* = \max\{\theta \geq 0 \mid x_i + \theta d_i^j \geq 0 \forall i\} = \min_{\{i \mid d_i^j < 0\}} -\frac{x_i}{d_i^j}$.

Por ende, se mueve a $y := x + \theta^* d^j$ generando la nueva base, en donde x_j entra a la base. Alguna variable en la base antigua se vuelve cero cuando se mueve desde x a $x + \theta^* d^j$, por ende, se saca esta variable de la base, resultando la base \bar{B} . Si todas las soluciones básicas factibles no degeneradas $\theta^* > 0$ y exactamente una variable de la base se vuelve cero.

Teorema 31. El punto y es solución básica factible para base \bar{B} .

Pero teniendo en cuenta todo lo anterior, ¿Cómo funciona Simplex? Mediante iteraciones, las cuales se conforman por pasos que se detallan a continuación:

1. Se comienza con la base $B = (B(1), \dots, B(m))$, matriz $A_B = (A_{B(1)}, \dots, A_{B(m)})$ se define la solución básica factible $x: x_B = (x_{B(1)}, x_{B(2)}, \dots, x_{B(m)}) = A_B^{-1}b$ y $x_N = 0$.
2. Se calculan los costos reducidos $\bar{c}_j = c_j - c_B^T A_B^{-1} A_j$ por cada variable no básica x_j :
 - a) Si todos $\bar{c}_j \geq 0$, entonces se encuentra allí el óptimo.
 - b) Escoger algún j con $\bar{c}_j < 0$.
3. Calcular $d := d_B^j = -A_B^{-1} A_j$. Si $d \geq 0$ entonces el óptimo es $-\infty$.
4. Escoger el índice ℓ tal que:

$$-\frac{x_{B(\ell)}}{d_\ell} = \min_{\{i|d_i < 0\}} -\frac{x_{B(i)}}{d_i} := \theta^*$$
5. Se genera una nueva base reemplazando $x_{B(\ell)}$ con x_j .
6. Nueva solución básica factible y :
 - a) $y_j = \theta^*$.
 - b) $y_{B(i)} = x_{B(i)} + \theta^* d_i$ para $i \neq \ell$.
 - c) $y_i = 0$ para $i \notin B \cup \{j\}$.
7. Regresar al paso 1.

■ Se vuelve al ejemplo:

$$\begin{aligned} \text{mín } & -4x_1 - 3x_2 \\ 2x_1 + 2x_2 + x_3 & = 8 \\ 2x_1 + x_2 + x_4 & = 6 \\ 20x_1 + x_2 + x_5 & = 60 \\ x_i & \geq 0 \quad \forall i = 1, \dots, 5 \end{aligned}$$

Se tiene $B = (3, 4, 5)$, $A_B = I$ y $x_B = A_B^{-1} \begin{pmatrix} 8 \\ 6 \\ 60 \end{pmatrix} = \begin{pmatrix} 8 \\ 6 \\ 60 \end{pmatrix}$, $x_N = (0, 0)$.

- Con $x = (0, 0, 8, 6, 60)^T$, $x_B = (x_3, x_4, x_5)$ y $x_N = (x_1, x_2)$, $c_N = (-4, -3)$. En donde, $\bar{c} = c^T - c_B^T A_B^{-1} A = (-4 \ -3 \ 0 \ 0 \ 0)$. Se selecciona $j = 1$, por ende, $d = -A_B^{-1} A_1 = (-2 \ -2 \ -20)$, por ello, corresponde a mover en dirección $d^1 = (1, 0, -2, -2, -20)$ para $x + \theta^* d^1$. En donde, $\theta^* = \min_{\{i|d_i^j < 0\}} -\frac{x_i}{d_i} = \min\{\frac{8}{2}, \frac{6}{2}, \frac{60}{20}\} = 3$. Luego, escoger $\ell = 3$, entonces x_1 entra a la base, mientras $x_{B(3)} = x_5$ sale de la base, por ello se mueve a $x + 3 \cdot d^1 = (3, 0, 2, 0, 0)$, lo que implica que $\bar{B} = (1, 3, 4)$, $x_{\bar{B}} = (x_1, x_3, x_4)$, $x_{\bar{N}} = (x_2, x_5)$.

$$A_{\bar{B}} = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 1 \\ 20 & 0 & 0 \end{pmatrix} \Rightarrow A_{\bar{B}}^{-1} = \begin{pmatrix} 0 & 0 & 0,05 \\ 1 & 0 & -0,1 \\ 0 & 1 & -0,1 \end{pmatrix}$$

Teniendo así los costos reducidos como siguen:

$$\bar{c} = (-4 \ -3 \ 0 \ 0 \ 0) - (-4, 0, 0) \begin{pmatrix} 0 & 0 & 0,05 \\ 1 & 0 & -0,1 \\ 0 & 1 & -0,1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = (0 \ -2,8 \ 0 \ 0 \ 0,2)$$

- Se selecciona $j = 2, x_2$ entra a la base.

- Se calculan d y θ , obteniendo así $d = -A_B^{-1}A_2 = (-0,05, -1,9, -0,9)^T$ y $\theta^* = \min_{\{i|d_i < 0\}} -\frac{x_i}{d_i} = \min\{\frac{3}{0,05}, \frac{2}{1,9}, \frac{0}{0,9}\} = 0$, luego $\theta^* = 0$. Por ende, implica que el punto es $(3, 0, 2, 0, 0)$, es por ello que corresponde moverse a:

$$\begin{pmatrix} 3 \\ 0 \\ 2 \\ 0 \\ 0 \end{pmatrix} + \theta^* \begin{pmatrix} -0,05 \\ 1 \\ -1,9 \\ -0,9 \\ 0 \end{pmatrix}$$

Si se está en el mismo punto, pero con x_2 entrando a la base y x_4 saliendo de ella, implica que $x_B = (x_1, x_2, x_3)$, $x_N = (x_4, x_5)$, lo cual pasa solamente cuando x es degenerado.

Teorema 32. *Se asume $P \neq \emptyset$, toda solución básica factible es no degenerada, y el algoritmo es iniciado con una solución básica factible. Luego este concluye después de un número finito de iteraciones. Cuando se termina, hay dos posibilidades:*

1. Se tiene una base óptima y una solución básica factible que es óptima.
2. Se encuentra un vector d el cual satisface $Ad = 0$, $d \geq 0$, y $c^T d < 0$, y así el PPL es no acotado, es decir, el óptimo es $-\infty$.

Demostración. 1. Si el algoritmo termina en el paso 2, es óptimo, ya que $\bar{c} \geq 0$.

2. Si el algoritmo termina en el paso 3, es una solución básica factible x y se encuentra la dirección $d^j \geq 0$. Con $Ad^j = 0$, lo cual implica que $x + \theta d^j \in P$ para todos $\theta > 0$, $c^T d^j = \bar{c}_j < 0$, por ende, el costo de $x + \theta d^j$ es $c^T(x + \theta d^j) = c^T x + \underbrace{\theta c^T d^j}_{< 0}$, lo cual implica que el PPL es no acotado.

3. Finalización: Para no degenerancia siempre se cumple que $\theta^* > 0$, el costo del nuevo punto $x + \theta d^j$ es $c^T(x + \theta d^j) = c^T x + \underbrace{\theta c^T d^j}_{< 0} < c^T x$, lo cual implica que en cada pivote el valor objetivo decrece estrictamente, lo cual implica que ningún vértice se visita dos o más veces, es decir, un número finito de iteraciones.

□

5.3. Problemas degenerados

Gracias a la definición de degenerancia aprendida en el capítulo anterior, se puede ahora proceder a comprender lo que es un problema degenerado, es decir, aquel que posee la nueva solución igual al antiguo problema, pero diferente base, lo cual implica que no hay progreso.

- *Lo anterior sucede si $d_i^j < 0$, pero variable básica $x_i = 0$*
- El algoritmo Simplex puede ciclar y estar siempre en la misma solución, para evitar aquello, se hace lo siguiente:
 - Se selecciona variable que entre a la base.
 - Se selecciona la variable que salga de la base.
- Regla de Bland
 1. Buscar el menor j para el cual $\bar{c}_j < 0$ y dejar que x_j entre a la base.
 2. De todas las variables x_i sale de la base, la variable x_i con menor i .

Generando así un número finito de iteraciones. Cabe destacar que existen otras reglas para pivotear que garantizan el número finito de iteraciones.

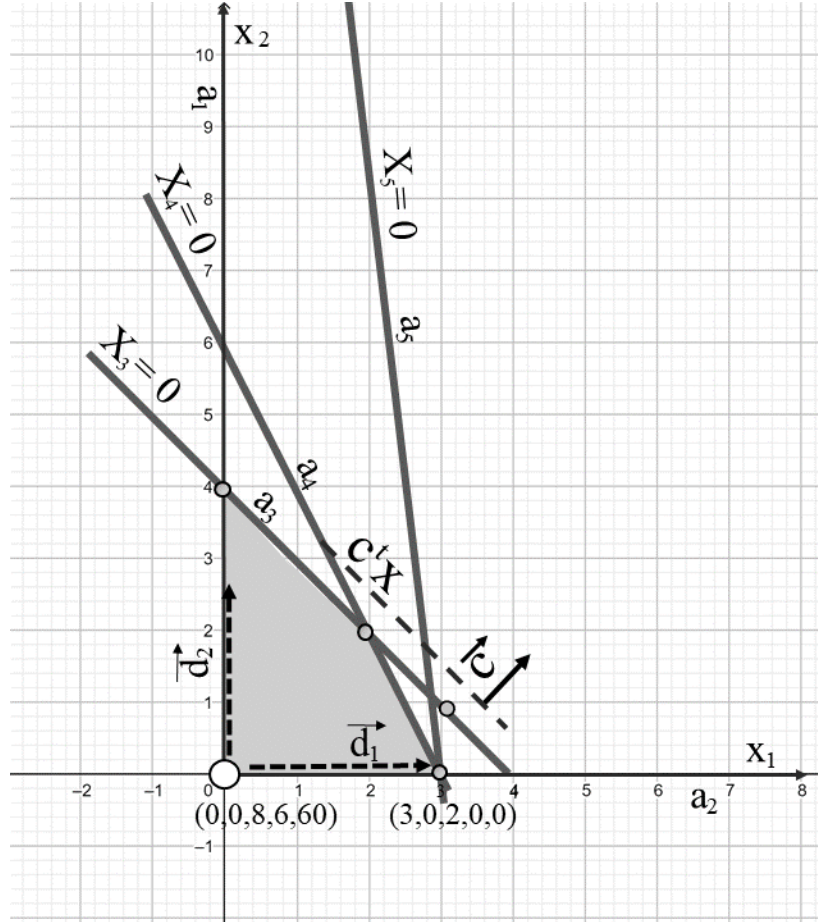


Figura 28: La función objetivo es $\min -4x_1 - 3x_2$. a_3 corresponde a la restricción $2x_1 + 2x_2 + x_3 = 8$, a_4 corresponde a la restricción $2x_1 + x_2 + x_4 = 6$, a_5 corresponde a la restricción $-x_1 + x_2 + x_5 = -2$ y a_2 corresponde a la restricción $20x_1 + x_2 + x_5 = 60$, se deben considerar además las restricciones de positividad de las variables $x_1, x_2, x_3, x_4, x_5, \geq 0$. En este ejemplo se puede apreciar mejor el caso de degenerancia en el algoritmo simplex. Los costos reducidos dan como opción moverse en dirección de d_1 o d_2 . En primera instancia, se moverá en d_1 , de esta manera x_1 entra a la base y se observa que x_3 sale de esta, llegando al punto $(3, 0, 2, 0, 0)$. Se inicia la segunda iteración, y según los costos reducidos se obtiene que x_2 entra a la base, así que se sigue esa dirección, pero se tiene que al calcular cuanto moverse obtenemos que θ es 0, es por esto que se debe tener cuidado con puntos degenerados.

5.4. Encontrar solución básica factible inicial

Simplex requiere solución básica factible para comenzar, pero la pregunta es ¿Cómo se encuentra? ¡Fácil, con otro PPL!

- Problema dado:

$$\begin{aligned} & \text{mín } c^T x \\ \text{s.a. } Ax &= & = b \\ & x \geq 0 \end{aligned}$$

Se asume que $b \geq 0$, lo cual puede ser logrado multiplicando las filas por -1 .

- Problema nuevo/artificial:

$$\begin{aligned} & \text{mín } \sum_{i=1}^m y_i \\ \text{s.a. } Ax + y &= & = b \\ & x \geq 0 \\ & y \geq 0 \end{aligned}$$

Si x es una solución factible para el problema dado, implica que $(x, 0)$ es solución factible para el nuevo problema con costo cero, es decir, la solución óptima al nuevo problema tiene costo cero (x^*, y^*) , es decir, el problema es factible.

- **Fase I:**

- Resolver el problema artificial comenzando por el vértice $x = 0, y = b$, el cual es solución básica factible. La base inicial consiste de todas las variables y .
- Sea (x, y) la solución final al nuevo problema:
 - Si $y \neq 0$ el problema dado es infactible.
 - Si $y = 0, x$ es solución factible al problema entregado.
- Sea $(x, 0)$ solución básica factible óptima del nuevo problema con base B , si B contiene solo variables x , implica que x es solución básica con base B del problema dado. Se asume que B contiene sólo $k < m$ variables x . En $(x, 0)$ sólo las variables $x_{B(1)}, \dots, x_{B(k)}$ pueden ser no-cero y $\sum_{i=1}^k A_{B(i)} x_{B(i)} = b$, con columnas $A_{B(1)}, \dots, A_{B(k)}$ linealmente independientes, se pueden encontrar los índices $B'(k+1), \dots, B'(m)$ tal que $A_{B(1)}, \dots, A_{B(k)}, A_{B'(k+1)}, \dots, A_{B'(m)}$ sean linealmente independientes. $B' = (B(1), \dots, B(k), B'(k+1), \dots, B'(m))$ es base de x en el problema dado.

- **Fase II:**

- Resolver problema dado.
- Comenzar con x y base B' .

5.5. Simplex: Método Full Tableau

En general, el algoritmo Simplex se puede ejecutar de diversas formas, en este capítulo en particular se procederá a detallar la forma en la que se utiliza el Tableau, el cual es una tabla en donde se ordenan las matrices necesarias para pivotar y llegar a la solución óptima de un PPL.

- Inicialización: Tableau para solución básica factible con base B .

$-c_B^T x_B$	\bar{c}_1	\dots	\bar{c}_n
$x_{B(1)}$	$ $	$ $	$ $
\dots	$A_B^{-1}A_1$	\dots	$A_B^{-1}A_n$
$x_{B(m)}$	$ $	$ $	$ $

$-c_B^T A_B^{-1}b$	$c^T - c_B^T A_B^{-1}A$
$A_B^{-1}b$	$A_B^{-1}A$

En la columna cero: $A_B^{-1}b$ se sitúan los valores de variables básicas, \bar{c} son los costos reducidos. $-c_B^T x_B$ es el valor objetivo negativo.

- Paso pivote:
 1. Si $\bar{c} \geq 0$ hay que detenerse, en caso contrario, se debe escoger j con $\bar{c}_j < 0$.
 2. Se considera $u := A_B^{-1}A_j$ (j -ésima columna del *Tableau*). Si $u \leq 0$ entonces el óptimo es $-\infty$.
 3. Para cada i con $u_i > 0$, se calcula $x_{B(i)}/u_i$. Sea ℓ el índice de una fila que minimice el ratio, la columna $A_{B(\ell)}$ posee base y la columna A_j entra a la base. Nueva base \bar{B} : $\bar{B}(\ell) = j$ y $\bar{B}(i) = B(i)$ para cada $i \neq \ell$.
 4. Se divide la ℓ -ésima fila por u_ℓ .
 5. Se suman los múltiplos de ℓ -ésima fila para todas las otras filas, aparte de u_ℓ en donde la columna j -ésima se hace cero incluyendo a la fila cero.

$-c_{\bar{B}}^T A_{\bar{B}}^{-1}b$	$c^T - c_{\bar{B}}^T A_{\bar{B}}^{-1}A$
$A_{\bar{B}}^{-1}b$	$A_{\bar{B}}^{-1}A$

Ejemplo:

- Solución inicial: $(0, 0, 0, 20, 20, 20)$. Luego $B(1) = 4$, $B(2) = 5$, $B(3) = 6$. $A_B = I = A_B^{-1}$, $c_B = 0$ y así $\bar{c} = c$.
- $\bar{c}_1 < 0$, se selecciona el pivote
- $\ell = 2 \min x_{B(i)}/u_i$ ($\ell = 3$ también)
- Por ende, A_1 entra a la base, A_5 deja la base.

Definición 33. Un vector $u \in \mathbb{R}^n$ es lexicográficamente positivo si la primera entrada no-cero es positiva: $u >_{lex} 0$. Para $u, v \in \mathbb{R}^n$ se tiene $u >_{lex} v$ ssi $u - v >_{lex} 0$.

Regla lexicográfica de pivoteo

1. Se escoge una variable entera x_j de forma arbitraria dentro de las que cumplan $\bar{c}_j < 0$. Sea $u := A_B^{-1}A_j$ (j -ésima columna de *Tableau*).
2. Por cada i con $u_i > 0$: Se divide la i -ésima fila del *Tableau* de Simplex por u_i , luego se escoge el más pequeño de forma lexicográfica de la fila ℓ , por ende, la variable básica $x_{B(\ell)}$ sale de la base.

Teorema 34. *Suponga que al iniciar este método todas las filas, aparte de la fila cero son lexicográficamente positivas. Cuando se sigue la regla lexicográfica de pivoteo:*

- *Toda fila aparte de la fila cero se queda lexicográficamente positiva.*
- *Fila cero crece de forma estricta en su forma lexicográfica en cada iteración, lo cual implica que el Simplex termina después de un número finito de pasos.*
- *Inicialmente todas las filas necesitan ser lexicográficamente positivas.*
- *Si no es el caso en la solución dada, entonces se deben reordenar filas y columnas de A s.a. x_1, \dots, x_m las bases iniciales, lo cual implica que el Tableau resultante posea filas lexicográficamente positivas.*

*	*
$A_B^{-1}b \geq 0$	$I A_B^{-1}A_{m+1}\dots A_B^{-1}A_n$

6. Dualidad

Dado un PPL, ¿Cómo se puede argumentar fácilmente que la solución no puede ser mejor que otra? ¿Existirá una manera de encontrar la solución a un costo menor? Para lo anterior, se utiliza Dualidad. Asociado a cada problema lineal o primal (P), existe otro problema de programación lineal denominado problema dual (D), que posee propiedades y formas de enfrentarse similar, pero importantes a destacar. Se destaca, por ejemplo el siguiente PPL:

$$(P) \quad \begin{aligned} \text{mín } z(x_1, x_2) &= -3x_1 + x_2 \\ x_1 + x_2 &\geq 0 & (3) \\ x_1 - 2x_2 &\geq -6 & (4) \\ x_1 + x_2 &\leq 5 & (5) \\ 5x_1 - x_2 &\leq 20 & (6) \\ -3x_1 + x_2 &\geq -40 & (7) \\ x_2 &\geq 0 \end{aligned}$$

Suponiendo que se encuentra la siguiente solución, no necesariamente óptima: $(4, 0)$, en donde la función objetivo vale -12 ¿Cuán lejos se encuentra del óptimo? Sea x^* la solución óptima y $z^* := z(x_1^*, x_2^*)$ con $z^* \leq -12$ la cota inferior : (7)

$$-3x_1 + x_2 \geq -40$$

Entonces $z^* \geq -40$. Luego, se calcula (3) + (4) -5× (5)

$$-3x_1 - 6x_2 \geq -31$$

$$\Rightarrow z^* \geq -31$$

$$\Rightarrow -31 \leq -3x_1 - 6x_2 \leq -3x_1 + x_2 = z(x_1, x_2), \forall x_1, x_2 \in P$$

Para buscar otra cota, se realiza el siguiente cálculo: $-3 \times (5)$

$$-15 \leq -3x_1 - 3x_2 \leq -3x_1 + x_2 = z(x_1, x_2)$$

Obteniendo así $z^* \geq -15$. Entonces se sabe que $-15 \leq z^* \leq -12$, con lo anterior se concluye que el valor óptimo que se tenía -12 no era tan malo. Se comienza a formalizar la aseveración anterior, multiplicando

las restricciones por las variables: y_1, y_2, y_3, y_4, y_5

$$\begin{array}{r|l} x_1 + x_2 \geq 0 & \cdot y_1 \geq 0 \\ x_1 - 2x_2 \geq -6 & \cdot y_2 \geq 0 \\ x_1 + x_2 \leq 5 & \cdot y_3 \leq 0 \\ 5x_1 - x_2 \leq 20 & \cdot y_4 \leq 0 \\ -3x_1 + x_2 \geq -40 & \cdot y_5 \geq 0 \end{array}$$

Esto lleva a lo siguiente:

$$y_1(x_1 + x_2) + y_2(x_1 - 2x_2) + y_3(x_1 + x_2) + y_4(5x_1 - x_2) + y_5(-3x_1 + x_2) \geq 0y_1 + (-6)y_2 + 5y_3 + 20y_4 - 40y_5$$

$$\Leftrightarrow x_1 \cdot (y_1 + y_2 + y_3 + 5y_4 - 3y_5) + x_2(y_1 - 2y_2 + y_3 - y_4 + y_5) \geq 0 \cdot y_1 - 6y_2 + 5y_3 + 20y_4 - 40y_5$$

En donde se busca que el término que acompaña a x_1 en el lado izquierdo sea igual a -3 y el término que acompaña a x_2 en el lado izquierdo sea menor o igual a 1 :

$$x_1 \cdot \underbrace{(y_1 + y_2 + y_3 + 5y_4 - 3y_5)}_{=-3} + x_2 \underbrace{(y_1 - 2y_2 + y_3 - y_4 + y_5)}_{\leq 1} \leq -3x_1 + x_2 \quad \forall x_1, x_2 \in P$$

Entonces se sabe que:

$$z^* \geq 0 \cdot y_1 - 6y_2 + 5y_3 + 20y_4 - 40y_5$$

Por ende, si se tiene el siguiente PPL

$$\begin{array}{l} y_1 + y_2 + y_3 + 5y_4 - 3y_5 = -3 \\ y_1 - 2y_2 + y_3 - y_4 + y_5 \leq 1 \\ y_1, y_2, y_5 \geq 0, \quad y_3, y_4 \leq 0 \end{array}$$

Teniendo en cuenta que la mejor cota inferior lleva al PPL siguiente

$$(D) \quad \begin{array}{l} \text{máx } w(y_1, y_2, y_3, y_4, y_5) = 0y_1 - 6y_2 + 5y_3 + 20y_4 - 40y_5 \\ y_1 + y_2 + y_3 + 5y_4 - 3y_5 = -3 \\ y_1 - 2y_2 + y_3 - y_4 + y_5 \leq 1 \\ y_1, y_2, y_5 \geq 0, \quad y_3, y_4 \leq 0 \end{array}$$

Entonces sea w^* la solución óptima del PPL (D), se observa lo siguiente:

$$w(y_1, y_2, y_3, y_4, y_5) \leq w^* \leq z^* \leq z(x_1, x_2) \quad \forall \text{ factible } (x_1, x_2), (y_1, y_2, y_3, y_4, y_5)$$

Se tiene que $x^* = (4, 0)$ y $z^* = -12$, así existe una solución $y^* = (0, 0, 0, -\frac{3}{5}, 0)$ y $w(y^*) = -12$, lo cual implica que x^* y y^* son óptimos para (P) y (D).

$$\blacksquare \text{ Primal: } \text{Mín } (-3 \ 1) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \text{ matriz: } \begin{pmatrix} 1 & 1 \\ 1 & -2 \\ 1 & 1 \\ 5 & -1 \\ -3 & 1 \end{pmatrix} \text{ y } \begin{pmatrix} 0 \\ -6 \\ 5 \\ 20 \\ -40 \end{pmatrix}$$

■ Dual: $\text{Max } (0 \ -6 \ 5 \ 20 \ -40) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}$, matriz: $\begin{pmatrix} 1 & 1 & 1 & 5 & -3 \\ 1 & -2 & 1 & -1 & 1 \end{pmatrix} y \begin{pmatrix} -3 \\ 1 \end{pmatrix}$

Es importante observar que la matriz del primal es la transpuesta del dual, por ejemplo, el vector costos del primal, si se transpone, es el vector b del dual.

6.1. Duales para PPL's arbitrarios

Antes de comenzar, se debe tener en claro que para formular problemas duales se debe asumir que el Primal posee una matriz A , vector de costos c^T y vector b , entonces el Dual posee una matriz A^T , vector de costos b^T y vector c .

Dicho lo anterior, siempre que se tenga un problema Primal en mínimo y se quiera formular el Dual, se puede hacer mediante la siguiente tabla. Cabe destacar que las restricciones del Primal, corresponden a las variables del Dual y viceversa. También es importante mencionar que si el Primal no se encuentra en mínimo, entonces se puede utilizar la tabla al revés sin problemas.

PRIMAL	Minimizar		Maximizar	DUAL
restricciones	$\geq b_i$		≥ 0	variables
	$\leq b_i$		≤ 0	
	$= b_i$		Libre	
variables	≥ 0		$\leq c_j$	restricciones
	≤ 0		$\geq c_j$	
	Libre		$= c_j$	

Ejemplo numérico:

Primal (P):

$$\begin{aligned}
 & \text{mín } x_1 + 2x_2 + 3x_3 \\
 & \text{s.a.} \\
 & -x_1 + 3x_2 = 5 \quad \text{---} > y_1 \text{ (Libre)} \\
 & 2x_1 - x_2 + 3x_3 \geq 6 \quad \text{---} > y_2 \geq 0 \\
 & \quad \quad \quad x_3 \leq 4 \quad \text{---} > y_3 \leq 0 \\
 & \quad \quad \quad x_1 \geq 0 \\
 & \quad \quad \quad x_2 \leq 0 \\
 & \quad \quad \quad x_3 \text{ Libre}
 \end{aligned}$$

Dual (D):

$$\begin{aligned}
 & \text{máx } 5y_1 + 6y_2 + 4y_3 \\
 & \text{s.a.} \\
 & -1y_1 + 2y_2 \leq 1 \\
 & 3y_1 - 1y_2 \geq 2 \quad y_1 \text{ (Libre)} \\
 & \quad \quad \quad 3y_2 + 1y_3 = 3 \\
 & \quad \quad \quad y_2 \geq 0 \\
 & \quad \quad \quad y_3 \leq 0
 \end{aligned}$$

Ejemplo matricial:

- **Primal (P):** $\text{mín } c^T x \text{ s.a. } Ax = b, x \geq 0$
- **Dual (D):** $\text{máx } y^T b \text{ s.a. } A^T y \leq c$

6.2. Dualidad débil

Continuando con el ejemplo anterior, se verifica que siempre se cumple lo siguiente:

Teorema 35. *Se considera el PPL con la matriz A , función de costos c y vector b en el Primal. Sea x una s.b.f. para el Primal e y una s.b.f. para el Dual, entonces:*

$$y^T b \leq c^T x$$

Demostración. Sean a_1^T, \dots, a_m^T las filas de A y A_1, \dots, A_m las columnas de A . Se define

$$u_i := y_i(a_i^T x - b_i)$$

$$v_j := (c_j - y^T A_j)x_j$$

Siendo i la i -ésima restricción del (P):

- $a_i^T x \geq b_i$, entonces $y_i \geq 0 \Rightarrow u_i \geq 0$
- $a_i^T x \leq b_i$, entonces $y_i \leq 0 \Rightarrow u_i \geq 0$
- $a_i^T x = b_i$, entonces y_i Libre $\Rightarrow u_i = 0$

Lo anterior implica que $u_i \geq 0$. De la misma forma, se procede a analizar $v_j \geq 0$:

$$\sum_i u_i = y^T Ax - y^T b$$

$$\sum_j v_j = c^T x - y^T Ax$$

Teniendo así: $0 \leq \sum_i u_i + \sum_j v_j = (y^T Ax - y^T b) + (c^T x - y^T Ax) = c^T x - y^T b$ □

Corolario 36. ▪ *Si el costo óptimo del (P) es $-\infty$, entonces el (D) debe ser infactible.*

- *Si el costo óptimo del (D) es $+\infty$, entonces el (P) debe ser infactible.*

Demostración. Se tiene como supuesto que el costo óptimo de (P) es $-\infty$ y el (D) posee una s.b.f y . Por Dualidad débil, y satisface $b^T y \leq c^T x$ para cada s.b.f. del (P) x . Pero existe una s.b.f del (P) \bar{x} con $c^T \bar{x} < b^T y$. Por ende, existe contradicción. □

6.3. Dualidad fuerte

Teorema 37. *Si un PPL posee solución finita óptima:*

- *El Dual posee solución finita óptima*
- *Sus respectivos costos óptimos son iguales*

Demostración. Se considera el problema en forma estándar:

$$\begin{aligned} &\text{mín } c^T x \\ &\text{s.a.} \\ &Ax = b \\ &x \geq 0 \end{aligned}$$

Se ejecuta Simplex con una regla anti-ciclos. Existe una s.b.f. x para la base B que es óptima, en donde $x_B := A_B^{-1}b$.

■ **Ejemplo:**

$$\begin{aligned} & \min 3x_1 + x_2 \\ \text{s.a. } & x_1 + x_2 - x_3 = 2 \\ & 2x_1 - x_2 - x_4 = 0 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

Solución óptima: $(2/3, 4/3, 0, 0)$

Base óptima: $\{x_1, x_2\}$

Costos reducidos son no-negativos, así $c^T - c_B^T A_B^{-1} A \geq 0$, es decir, $(0, 0, \frac{5}{3}, \frac{2}{3})$

Luego, se define la solución dual $y^T := c_B^T A_B^{-1} \Rightarrow y^T = (3 \ 1) \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & -1/3 \end{pmatrix} = (5/3 \ 2/3)$

Entonces $c^T - c_B^T A_B^{-1} A \geq 0 \Leftrightarrow c^T \geq c_B^T A_B^{-1} A \Leftrightarrow c^T \geq y^T A$ y así y es factible para el problema dual, siendo este:

$$\begin{aligned} & \text{máx } 2y_1 \\ \text{s.a. } & y_1 + 2y_2 \leq 3 \\ & y_1 - y_2 \leq 1 \\ & -y_1 \leq 0 \Leftrightarrow y_1 \geq 0 \\ & -y_2 \leq 0 \Leftrightarrow y_2 \geq 0 \end{aligned}$$

- $y^T b = c_B^T A_B^{-1} b = c_B^T x_B = c^T x$
- $2y_1 = 10/3 = 3x_1 + x_2$
- Dualidad débil implica y es óptimo para el problema dual.

□

■ Posibilidades para primal-dual

En la siguiente tabla se muestra la relación entre primal y dual según su naturaleza, por ejemplo, si el primal es Infactible, el dual puede ser no acotado o infactible como se ve en la siguiente tabla:

		Dual		
		Óptimo finito	No acotado	Infactible
Primal	Óptimo finito	sí	no	no
	No Acotado	no	no	sí
	Infactible	no	sí	sí

- **Ejemplo:** Primal y dual infactible, cabe destacar que lo son debido a que sus restricciones no son l.i.

$$(P) \begin{cases} \min & x_1 + 2x_2 \\ & x_1 + x_2 = 1 \\ & 2x_1 + 2x_2 = 3 \end{cases} \Rightarrow (D) \begin{cases} \max & y + 3y_2 \\ & y_1 + 2y_2 = 1 \\ & y_1 + 2y_2 = 2 \end{cases}$$

6.4. Holgura complementaria

Definición 38. Sea x e y s.b.f. para primal y dual, respectivamente:

- i -ésima restricción en primal es *activa en x* si $a_i^T x = b_i$
- j -ésima restricción en el dual es *activa en y* si $y^T A_j = c_j$

Ejemplo:

- Primal:

$$\begin{aligned} & \min 3x_1 + x_2 \\ \text{s.a. } & x_1 + x_2 - x_3 = 2 \\ & 2x_1 - x_2 - x_4 = 0 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

- Dual:

$$\begin{aligned} & \text{máx } 2y_1 \\ \text{s.a. } & y_1 + 2y_2 \leq 3 \\ & y_1 - y_2 \leq 1 \\ & y_1 \geq 0 \\ & y_2 \geq 0 \end{aligned}$$

Dada una solución óptima del primal $x = (2/3, 4/3, 0, 0)$, $x_1 > 0$ y $x_2 > 0$, lo cual implica que la primera y segunda restricción del dual son activas. Solución óptima del dual satisface $y_1 + 2y_2 = 3$ y $y_1 - y_2 = 1$. Solución óptima del dual es $y = (5/3, 2/3)$, $y_1 > 0$ y $y_2 > 0 \Rightarrow$ Primera y segunda restricción del primal son activas.

Teorema 39. Sea x e y s.b.f del primal y dual, respectivamente. Entonces x e y son ambas óptimos ssi:

$$\begin{aligned} y_i(a_i^T x - b_i) &= 0 \quad \forall i \\ (c_j - y^T A_j)x_j &= 0 \quad \forall j \end{aligned}$$

Demostración. Se asume x y y óptimos:

- Se define $u_i = y_i(a_i^T x - b_i)$ y $v_j = (c_j - y^T A_j)x_j$. $u_i \geq 0$ y $v_j \geq 0 \quad \forall i, j$. $\sum_i u_i + \sum_j v_j \stackrel{\text{prueba de igualdad dbil}}{=} c^T x - y^T b \stackrel{\text{dualidad fuerte}}{=} 0$ lo que implica que $u_i = 0$ y $v_j = 0$ para todo i, j .
- Si $u_i = 0$ y $v_j = 0 \quad \forall i, j$ implica que $0 = \sum_i u_i + \sum_j v_j \stackrel{\text{prueba de igualdad dbil}}{=} c^T x - y^T b$, entonces $c^T x = y^T b$ lo que implica que x, y son óptimos por dualidad débil.

□

6.5. Lema de Farkas

Usualmente verificar la factibilidad de un PPL es costoso y/o complicado, es por ello que se buscaron formas más fáciles y eficientes de hacerlo, una de ellas fue el Lema de Farkas, con el cual se puede verificar fácilmente la infactibilidad de un PPL.

Lema de Farkas: Sea $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$. Entonces una de las siguientes se cumple:

1. Existe un $x \geq 0$ tal que $Ax = b$

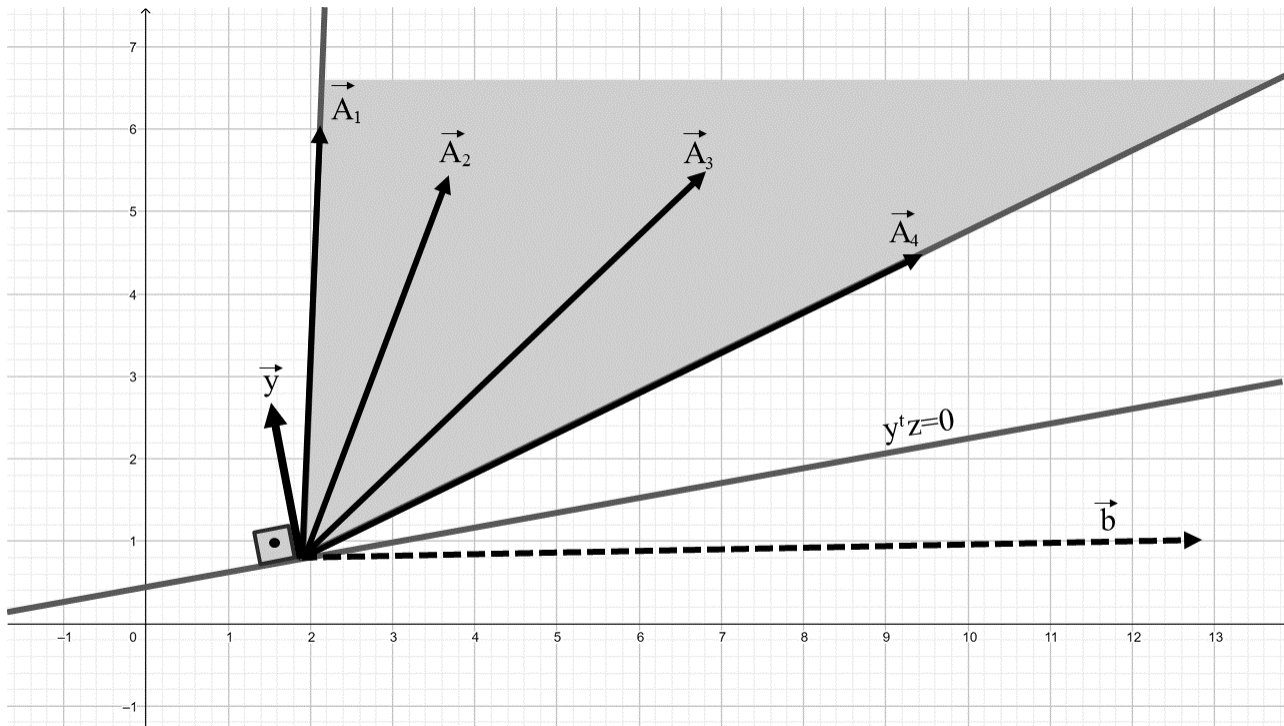


Figura 29: Gráficamente se puede notar que el producto punto entre los vectores A_i y el vector y es positivo, y que el producto punto entre el vector y y el vector b es negativo, por lo tanto, se tiene que el vector b no se puede obtener con combinación lineal de los vectores A_i , es decir, no pertenece al conjunto señalado con color plomo. Finalmente, dado lo anterior se tiene que se puede encontrar un hiperplano que los separa $y^t z = 0$ para una variable z .

2. Existe un vector y tal que $y^T A \geq 0$ y $y^T b < 0$.

- Prueba: Si (1) se cumple: Sea y un vector con $A^T y \geq 0$. Entonces $y^T b = y^T Ax \geq 0$ (con $x \geq 0$) y así (2) es falsa. En cambio, si (1) no se cumple, se consideran dos problemas (P) y (D) como sigue:

$$(P) \begin{cases} \text{máx} & 0 \cdot x \\ & Ax = b \\ & x \geq 0 \end{cases} \Rightarrow (D) \begin{cases} \text{mín} & b^T y \\ & y^T A \geq 0 \\ & y \text{ libre} \end{cases}$$

Primal es infactible, lo que implica que dual es no acotado o infactible. $y = 0$ es solución factible del dual, lo que implica que el dual es no acotado, es decir, $OPT = -\infty$, por ende, existe una solución dual y con $y^T A \geq 0$ y $y^T b = b^T y < 0$.

6.6. Simplex Dual

Se considera el PPL primal en forma estándar:

$$\begin{aligned} & \min c^T x \\ & \text{s.a. } Ax = b \\ & \quad x \geq 0 \end{aligned}$$

Y su respectivo dual:

$$\begin{aligned} & \max y^T b \\ & \text{s.a. } A^T y \leq c \end{aligned}$$

Dada la base B y solución del primal $x_B = A_B^{-1}b$, $x_N = 0$, es factible si $x_B = A_B^{-1}b \geq 0$ y es óptima si $\bar{c} = c^T - c_B^T A_B^{-1}A \geq 0 \Leftrightarrow c^T \geq c_B^T A_B^{-1}A$.

Luego se define la solución del dual $y^T := c_B^T A_B^{-1}$, $\bar{c} \geq 0 \Leftrightarrow c^T \geq c_B^T A_B^{-1}A = y^T A \Leftrightarrow y$ es factible para el problema dual. Y $c^T x = c_B^T x_B = c_B^T A_B^{-1}b = y^T b$ implica que el primal y dual poseen el mismo valor de la función objetivo, lo que implica que si $\bar{c} \geq 0$, existe una solución al dual con el mismo costo que x . Si adicionalmente $A_B^{-1}b \geq 0$, implica que x es factible y que tanto x como y son óptimos. Pero entonces ¿Cómo es una iteración del Simplex dual?

1. Se comienza con la matriz B s.a. $\bar{c} = c^T - c_B^T A_B^{-1}A \geq 0$.
2. Si $A_B^{-1}b \geq 0$, se tiene una base óptima. Por ende, las soluciones del primal y dual correspondientes son: $x_B = A_B^{-1}b$, $x_N = 0$, $y^T := c_B^T A_B^{-1}$. Se debe parar, en caso contrario, se elige ℓ con $x_{B(\ell)} < 0$.
3. Se considera ℓ -ésima fila del Tableau del PPL $x_{B(\ell)}, v_1, \dots, v_n$. Si $v_i \geq 0$ para todo i , entonces el óptimo dual es $+\infty$ y el primal es infactible. Se debe parar.
4. Para cada i tal que $v_i < 0$, se calcula $\bar{c}_i/|v_i|$, en donde sea j el índice de la columna que corresponde al ratio menor. La columna $A_{B(\ell)}$ sale de la base y la columna A_j entra a la base.
5. Se agrega a cada fila del Tableau un múltiplo de la ℓ -ésima fila para que v_j se convierta en 1 y todas las otras entradas de la columna pivote sean 0.

Ejemplo: Sólo la segunda fila posee entradas negativas en la columna cero-ésima ->Segunda fila es la fila pivote, $v_2 < 0$ and $v_3 < 0$, todos los otros v_i son no negativos. $\bar{c}_2/|v_2| = 3$, $\bar{c}_3/|v_3| = 10/3$ ->Segunda columna es la pivote. Por ende, se genera así un nuevo Tableau. El algoritmo puede ciclar, es por ello que se ocupa la regla anti-ciclo de Bland, en la cual se elige cualquier fila ℓ con $x_{B(\ell)} < 0$ para que sea fila pivote. Luego de ello, se determina por cada columna con $v_i < 0$, dividir todas las entradas por $|v_i|$ para finalmente escoger de forma lexicográfica la columna más pequeña. Si hay un empate, se escoge el menor índice.

6.7. Multiplicador de Lagrange

En capítulos anteriores se enseñó a derivar el dual mediante la utilización de la tabla de conversión, pero existe otra forma de encontrar el problema dual desde un primal y esta es mediante el multiplicador de Lagrange.

Supongamos que se quiere resolver el siguiente problema: $\min x_1^2 + x_2^2$

s.a. $x_1 + x_2 = 1$

Se comienza introduciendo el multiplicador de Lagrange y y forma la relajación Lagrangeana:

$$\min \underbrace{x_1^2 + x_2^2 + y(1 - x_1 - x_2)}_{L(x_1, x_2, y)}$$

Por cada y , se obtiene el problema relajado $\min_{x_1, x_2} L(x_1, x_2, y)$ y se tiene

$$\min_{x_1, x_2} L(x_1, x_2, y) \leq \min_{x_1, x_2: x_1 + x_2 = 1} x_1^2 + x_2^2$$

Lo anterior implica que $\min_{x_1, x_2} L(x_1, x_2, y)$ es la cota inferior del óptimo por cada y . Cabe destacar que por cada y la solución óptimo de $\min L(x, y, y)$ es la cota inferior del óptimo del problema original.

Se tiene entonces el problema primal:

$$\begin{aligned} & \text{mín } c^T x \\ & \text{s.a. } Ax = b \\ & \quad x \geq 0 \end{aligned}$$

Sea x^* la solución óptima, se considera el problema relajado, en donde y es un vector y $g(y)$ el costo óptimo para el problema relajado:

$$\begin{aligned} & \text{mín } c^T + y^T(b - Ax) \\ & \text{s.a. } x \geq 0 \text{ y libre} \end{aligned}$$

$$g(y) = \text{mín}_{x \geq 0} [c^T x + y^T(b - Ax)] \leq c^T x^* + y^T(b - Ax^*) = c^T x^*$$

Por cada y se tiene una cota inferior, es por ello que se quiere encontrar el más grande de ellos:

$$\begin{aligned} & \text{máx } g(y) \\ & \text{s.a. } y \in \mathbb{R}^n \end{aligned}$$

$$g(y) = \text{mín}_{x \geq 0} (c^T x + y^T(b - Ax)) = y^T b + \text{mín}_{x \geq 0} (c^T - y^T A)x$$

- Si $c^T - y^T A \geq 0$ entonces $\text{mín}_{x \geq 0} (c^T - y^T A)x = 0$
- Si $c^T - y^T A$ posee una entrada negativa, entonces entra $\text{mín}_{x \geq 0} (c^T - y^T A)x = -\infty$, lo que implica que $\text{mín}_{x \geq 0} (c^T - y^T A)x \in \{0, -\infty\}$, por ende, $g(y) \in \{y^T b, -\infty\}$, lo que implica que se satisface la búsqueda de y en donde $c^T - y^T A \geq 0$. Es por lo anterior que el problema dual queda de la siguiente forma:

$$\begin{aligned} & \text{máx } y^T b \\ & \text{s.a. } y^T A \leq c^T \Leftrightarrow A^T y \leq c \end{aligned}$$

6.8. Precios sombra

Imagine que se puedan cambiar los valores del vector b , ¿Qué pasaría con los costos? Para responder lo anterior, se analiza el siguiente ejemplo:

En una firma se quiere maximizar la utilidad obtenida por la venta de dos de sus productos estrella: P y W , en donde las utilidades de cada uno son 100 y 130, respectivamente. Se sabe además que el stock disponible de P es 16 y 15 de W y que la venta de ambos productos en conjunto no pueden superar las 21 unidades. Finalmente, se tiene que si la utilidad de W está en 1,5 y el de P en 1, al vender ambos productos en forma conjunta su utilidad máxima será de 27, en cambio, si la utilidad de W es de 0,3 y 0,5 para P , su máxima utilidad será de 9. Se plantea el PPL como sigue:

$$\begin{aligned} & \text{máx } 130W + 100P \\ & \text{s.a. } 1,5W + P \leq 27 \\ & \quad W + P \leq 21 \\ & \quad 0,3W + 0,5P \leq 9 \\ & \quad W \leq 15 \\ & \quad P \leq 16 \\ & \quad W, P \geq 0 \end{aligned}$$

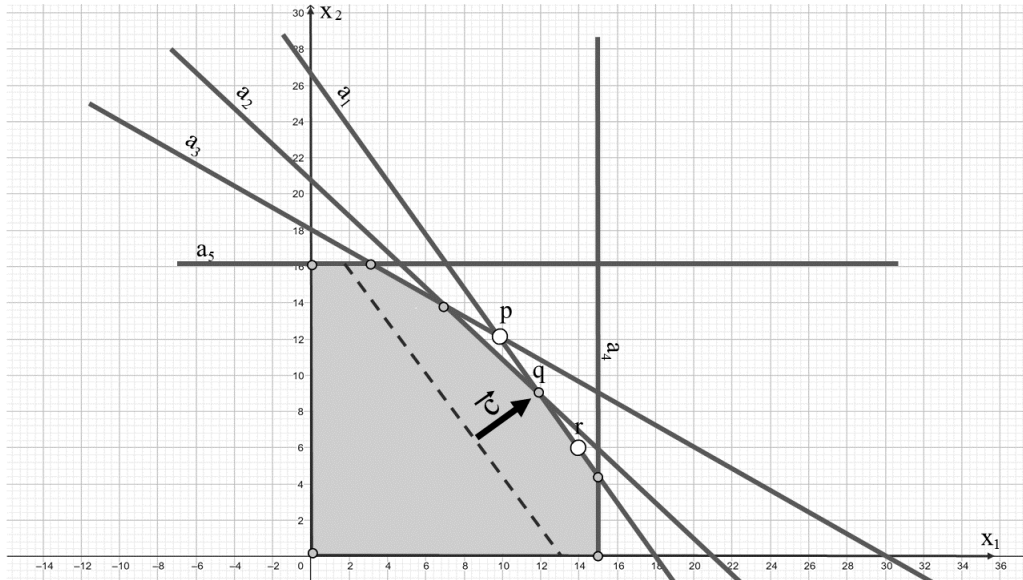


Figura 30: a_1 es la restricción $1,5W + P \leq 27$, a_2 es la restricción de interés en este problema $W + P \leq 21$, a_3 es la restricción $0,3W + 0,5P \leq 9$, a_5 es la restricción $P \leq 16$, a_6 es la restricción $W \leq 15$, el poliedro también incluye las restricciones $W, P \geq 0$ de no negatividad. Se puede apreciar gráficamente el óptimo del problema, que corresponde al punto $q=(12,9)$ con valor óptimo 2460. Si la restricción de interés aumenta en 1 unidad su b , el óptimo es el punto $p=(10,12)$ con valor óptimo 2450 y si disminuye en una unidad, el punto óptimo pasa a ser $r=(14,6)$ con valor óptimo 2420. Este cambio en la función objetivo sólo sucede con cambios unitarios del vector b .

La solución óptima es: $W = 12, P = 9$, en conjunto con su valor = 2460, por ende, una unidad extra $21 \rightarrow 22$, produce: $W = 10, P = 12$ y su valor = $2500 = 2460+40$.

En cambio, una unidad menos, es decir, $21 \rightarrow 20$, se tiene que $W = 14, P = 6$, y valor = $2420 = 2460-40$.

Entonces, cambiando la capacidad en pocas unidades: Cambia la función objetivo en 40\$ por cambio en unidad de capacidad. El precio sombra de la restricción en el caso anterior, es decir, el cambio del costo óptimo por unidad de cambio del vector b es 40, pero sólo en cambios de pocas unidades, ya que el comportamiento cuando las unidades son mayores es diferente como se a continuación:

Dos unidades extra, es decir, $21 \rightarrow 23$, producen: $W = 10, P = 12$ y el valor es $2500 = 2460+40$, en lugar de $2460+80$.

En cambio, dos unidades menos, es decir, $21 \rightarrow 19$, producen: $W = 15, P = 4$ y el valor será $2350 = 2460-110$, en lugar de $2460-240$.

Se considera el siguiente PPL:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.a.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Sea B la base óptima y $x = (x_B, x_N)$ con $x_B = A_B^{-1}b$ y $x_N = 0$, suponga que se cambia b_i a $b_i + \delta$ por un pequeño δ .

Se define entonces la nueva solución correspondiente a la misma base $x'_B := A_B^{-1}(b + \delta e_i)$, $c^T x' = c_B^T A_B^{-1}(b + \delta e_i) = \underbrace{c_B^T x_B}_{\text{costo anterior}} + c_B^T A_B^{-1} \delta e_i$

Así, el valor $c_B^T A_B^{-1} e_i$ es el *precio sombra* de la i -ésima restricción.

Se tiene entonces el problema dual:

$$\begin{aligned} & \max y^T b \\ \text{s.a. } & A^T y \leq c \end{aligned}$$

Se construye así, la solución del dual para la base B : $y^T := c_B^T A_B^{-1}$, lo que implica que es factible si la base B es óptima, ya que $c^T - c_B^T A_B^{-1} A \geq 0 \Leftrightarrow c^T \geq c_B^T A_B^{-1} A \Leftrightarrow c^T \geq y^T A$. Por lo anterior, el precio sombra correspondería a x y la i -ésima restricción es $y^T e_i = y_i$. Cabe destacar que la solución dual a la base B indica los precios sombra.

6.9. Interpretación dual de costos reducidos

Se considera el primal en forma estándar:

$$\begin{aligned} & \min c^T x \\ \text{s.a. } & Ax = b \\ & x \geq 0 \end{aligned}$$

Y el dual correspondiente:

$$\begin{aligned} & \max y^T b \\ \text{s.a. } & A^T y \leq c \end{aligned}$$

Dada la base B para, la solución primal es $x_B = A_B^{-1} b$, $x_N = 0$ es factible si $x_B = A_B^{-1} b \geq 0$, y óptimo si $\bar{c} = c^T - c_B^T A_B^{-1} A \geq 0 \Leftrightarrow c^T \geq c_B^T A_B^{-1} A$

Una prueba alternativa de lo anterior, si $\bar{c} \geq 0$ y x es factible, entonces $x = (x_B, x_N)$ es óptima. Se asume entonces que $\bar{c} \geq 0$ según la base B , se define el vector $y^T := c_B^T A_B^{-1}$, $\bar{c} \geq 0 \Leftrightarrow c^T \geq c_B^T A_B^{-1} A = y^T A \Leftrightarrow y$ es factible para el problema dual. Entonces, $c^T x = c_B^T x_B = c_B^T A_B^{-1} b = y^T b$, implica que las soluciones del primal y el dual poseen el mismo valor objetivo, por ende, x y y son óptimos.

7. Análisis de sensibilidad

Imagina que un arquitecto planea construir una casa con sus diversas características, ya sea tamaño de las habitaciones, cantidad de pisos, entre otras. Pero en un punto del proceso, el arquitecto piensa que la ventana ubicada en el tercer piso debió haber sido un poco más grande...se encuentra en un gran dilema: Desechar el plan completo o modificar el antiguo plan, detallando el pequeño cambio de la ventana. Claramente es más eficiente la segunda opción. Algo similar se hace con los PPL, pudiendo distinguir así los cambios en el valor de la función objetivo debido al aumento o disminución de algún elemento. Como ejemplo de lo anterior, se tiene el siguiente PPL:

$$\begin{aligned} & \min c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

Sea x^* la solución óptima para alguna base B . Se supone en este punto entonces que se cambian datos a este modelo, ¿Siguen x^* y B siendo óptimos? Si la respuesta es no, ¿Cómo se puede obtener una solución de forma rápida? ¿Por cuánto cambia el óptimo? ¿Por qué es esto relevante?

- Verificar existencia de errores.
- Incerteza de data.
- Robustez.
- Usualmente solucionar un PPL es costoso, por ende, si la data cambia, se pueden obtener valores de forma rápida sin incrementar en gran medida los costos.

Se asume que las filas de A son linealmente independientes dada una solución óptima x^* con la base B , teniendo entonces:

- $A_B^{-1}b \geq 0$ (Factibilidad)
- $c^T - c_B^T A_B^{-1}A \geq 0$ (Optimalidad: Costos reducidos no negativos)

Si luego del cambio, las condiciones se siguen satisfaciendo, entonces B siguen siendo óptima.

7.1. Cambio en el vector b

Se supone que la componente b_i de b se cambia en $b_i + \delta$, pero ¿La base actual sigue siendo óptima?

Ejemplo:

$$\begin{array}{rcl} \text{mín} & -10x_1 - 10x_2 & \\ & 2x_1 + 2x_2 + x_3 & = 8 \\ & 2x_1 + x_2 & + x_4 = 6 \\ & x_1, x_2, x_3, x_4 & \geq 0 \end{array}$$

La base $B = (1, 2)$ es óptima, correspondiente al punto: $(2, 2, 0, 0)$. Ahora es cuando el vector b se cambia, es por ello que se verifican las condiciones antes nombradas. En particular, la condición de optimalidad no se ve afectada, ya que b no forma parte de la condición. Para la condición de factibilidad se verifica que $A_B^{-1}b \geq 0$:

$$A_B = \begin{pmatrix} 2 & 2 \\ 2 & 1 \end{pmatrix} \Rightarrow A_B^{-1} = \begin{pmatrix} -\frac{1}{2} & 1 \\ 1 & -1 \end{pmatrix}$$

Entonces, se fija $b_2 = 6$, obteniendo:

$$A_B^{-1}b = \begin{pmatrix} -\frac{1}{2} & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} b_1 \\ 6 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Lo cual implica que $-\frac{1}{2}b_1 + 6 \geq 0$, así $b_1 \leq 12$. Al mismo tiempo se verifica que $b_1 - 6 \geq 0$, lo que implica que $b_1 \geq 6$, por ende, si $b_1 \in [6, 12]$, entonces la base actual B sigue factible y óptima.

Se fija ahora $b_1 = 8$:

$$A_B^{-1}b = \begin{pmatrix} -\frac{1}{2} & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 8 \\ b_2 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Lo cual implica que si $b_2 \in [4, 8]$, entonces la base actual B sigue factible y óptima. En general, se debe chequear $A_B^{-1}(b + \delta e_i) \geq 0$, en donde, $g = (g_1, g_2, \dots, g_m)$ es la i ésima columna de A_B^{-1} :

- Si $g_j < 0$: $x_{B(j)} + \delta g_j \geq 0$, entonces $\delta \leq -\frac{x_{B(j)}}{g_j}$
- Si $g_j > 0$: $x_{B(j)} + \delta g_j \geq 0$, entonces $\delta \geq -\frac{x_{B(j)}}{g_j}$.

Por ende, la base actual sigue siendo óptima si:

$$\max_{\{j|g_j>0\}} \left(-\frac{x_{B(j)}}{g_j} \right) \leq \delta \leq \min_{\{j|g_j<0\}} \left(-\frac{x_{B(j)}}{g_j} \right)$$

Si δ se encuentra fuera del rango, entonces B es infactible, pero los costos reducidos son ≥ 0 . Cabe destacar que el Simplex Dual, el cual es una variación de Simplex, puede partir de esta solución.

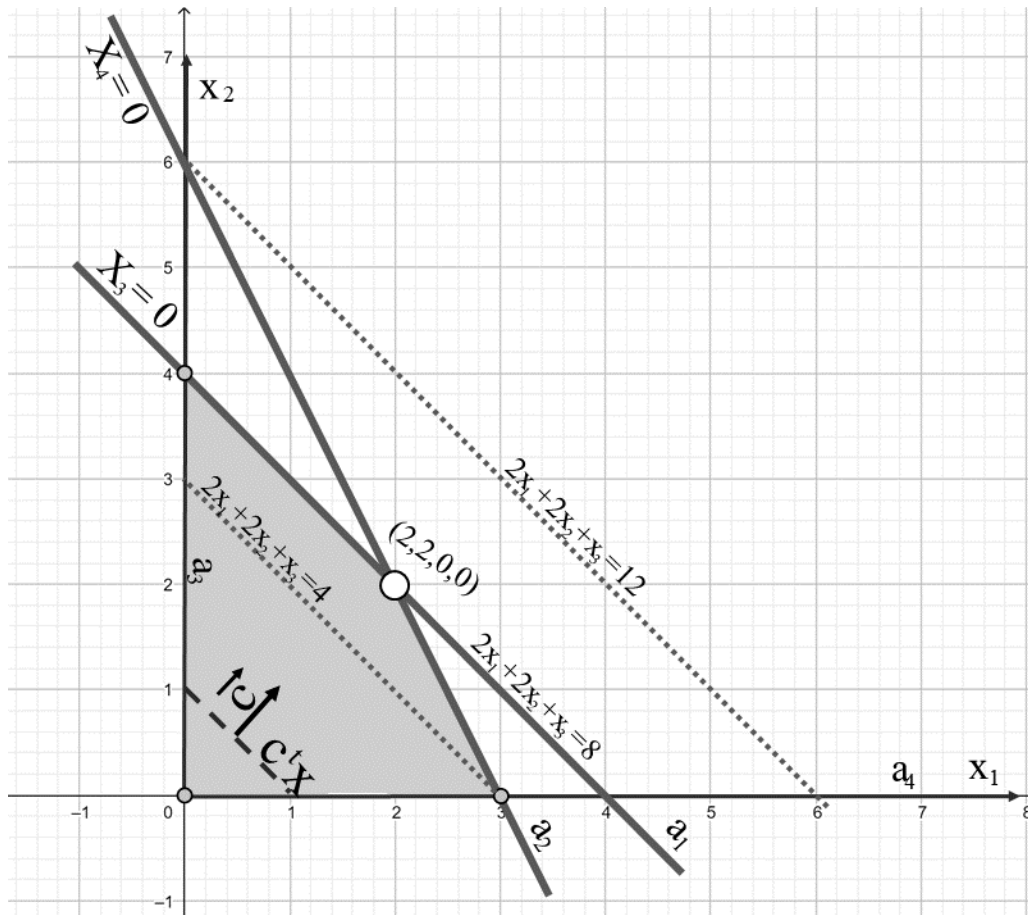


Figura 31: Se vuelve a un poliedro conocido. a_1 corresponde a $2x_1 + 2x_2 + x_3 = 8$, a_2 corresponde a $2x_1 + x_2 + x_4 = 6$, a_3 y a_4 corresponde a x_1 y $x_2 \geq 0$. Se puede apreciar que el punto óptimo es $(2, 2, 0, 0)$ con base óptima $(1, 2)$. Entonces, si fijamos b_2 ¿Entre qué valores de b_1 se mantiene esa base como óptima?, es decir, ¿Qué tanto se puede mover la restricción a_1 ? Se puede notar en el poliedro que el rango de b_1 es $(6, 12)$, esto se aprecia gráficamente, pues mientras las restricciones se sigan tocando dentro de la nueva región factible, la base continuará siendo factible y óptima.

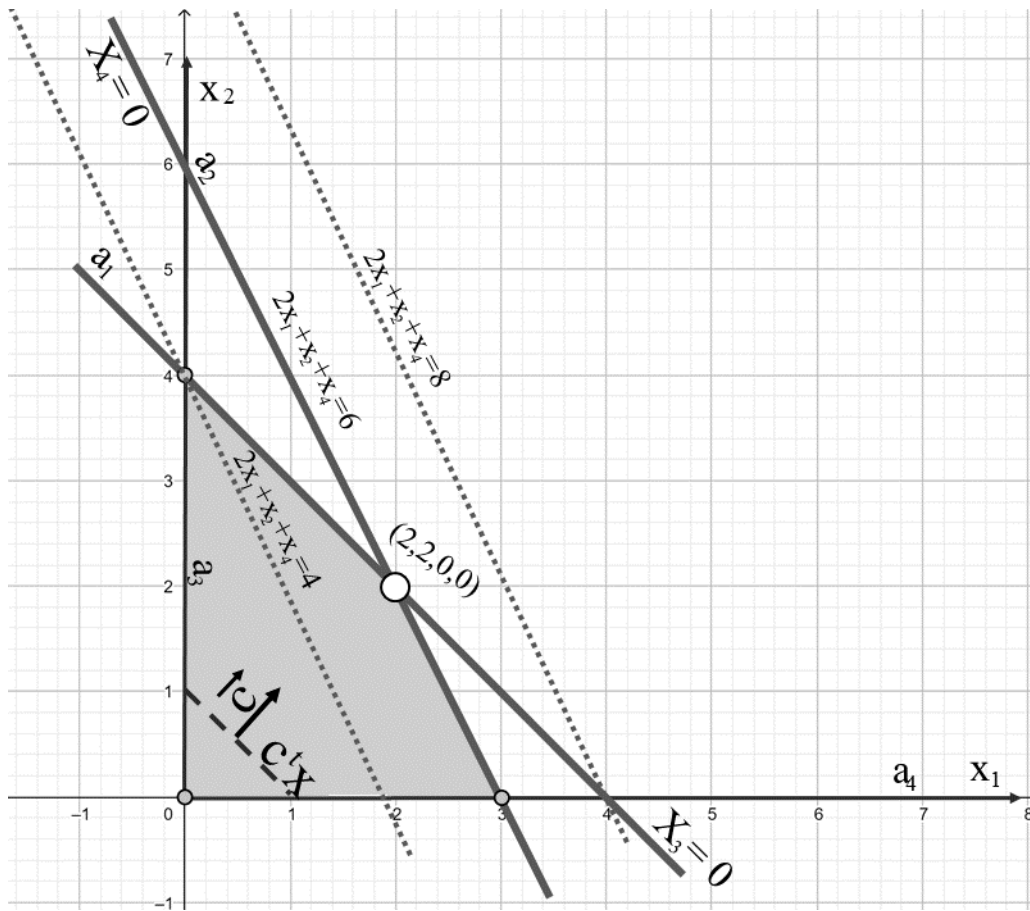


Figura 32: En este caso se fija b_1 , y se tiene que el rango para mover a_2 , es decir, los valores para b_2 de tal manera que la base siga siendo factible y óptima sería $(4, 8)$.

7.2. Cambios en el vector de costos

El coeficiente de costo c_j se convierte en $c_j + \delta$, pero ¿sigue la base actual siendo óptima?

- La factibilidad $A_B^{-1}b \geq 0$ no se ve afectada, ya que b no forma parte de la condición.
- La condición de optimalidad $c^T - c_B^T A_B^{-1}A \geq 0$, se ve afectada.

Ejemplo:

$$\begin{array}{rcl} \text{mín} & -10x_1 - 10x_2 & \\ & 2x_1 + 2x_2 + x_3 & = 8 \\ & 2x_1 + x_2 & + x_4 = 6 \\ & x_1, x_2, x_3, x_4 & \geq 0 \end{array}$$

En donde el punto $(2, 2, 0, 0)$ es óptimo y con $\bar{c} = (0, 0, 5, 0)$. Si se cambia c_1 :

$$\begin{aligned} \bar{c} &= (c_1, -10, 0, 0) - (c_1, -10) \begin{pmatrix} -\frac{1}{2} & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} \geq 0 \\ \Rightarrow & (0, 0, \frac{1}{2}c_1 + 10, -c_1 - 10) \geq 0 \\ \Rightarrow & c_1 \in [-20, -10] \end{aligned}$$

Si se cambia c_2 :

$$\begin{aligned} \bar{c} &= (-10, c_2, 0, 0) - (-10, c_2) \begin{pmatrix} -\frac{1}{2} & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} \geq 0 \\ \Rightarrow & (0, 0, -5 - c_2, 10 + c_2) \geq 0 \\ \Rightarrow & c_2 \in [-10, -5] \end{aligned}$$

Si se cambia c_3 :

$$\begin{aligned} \bar{c} &= (-10, -10, c_3, 0) - (-10, -10) \begin{pmatrix} -\frac{1}{2} & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} \geq 0 \\ \Rightarrow & (0, 0, c_3 + 5, 0) \geq 0 \\ \Rightarrow & c_3 \in [-5, \infty) \end{aligned}$$

En general, si se cambia c_j a $c_j + \delta$, si x_j es variable básica con $j = B(\ell)$:

- c_B se convierte en $c_B + \delta e_\ell$
- B sigue siendo óptima si $c_i - (c_B + \delta e_\ell)^T A_B^{-1}A_i \geq 0$ para cada variable no básica i , por ende, $c_i - c_B^T A_B^{-1}A_i \geq \delta e_\ell^T A_B^{-1}A_i$, entonces $c_i \geq \delta q_{\ell i}$ para cada variable no básica i donde $q_{\ell i}$ es la ℓ -ésima entrada de $A_B^{-1}A_i$, la cual se puede obtener mediante Tableau. Lo anterior implica que existe un rango para δ en donde B sigue siendo óptima.

Si x_j es una variable no básica

- c_B no cambia.
- Sólo se necesita chequear costos reducidos para x_j .
- Se desea que $(c_j + \delta) - c_B^T A_B^{-1}A_j \geq 0$, es decir, $\delta \geq -c_j + c_B^T A_B^{-1}A_j = -\bar{c}_j$. Si lo anterior es cierto, entonces la base actual sigue siendo óptima, en caso contrario se utiliza Simplex partiendo con B .

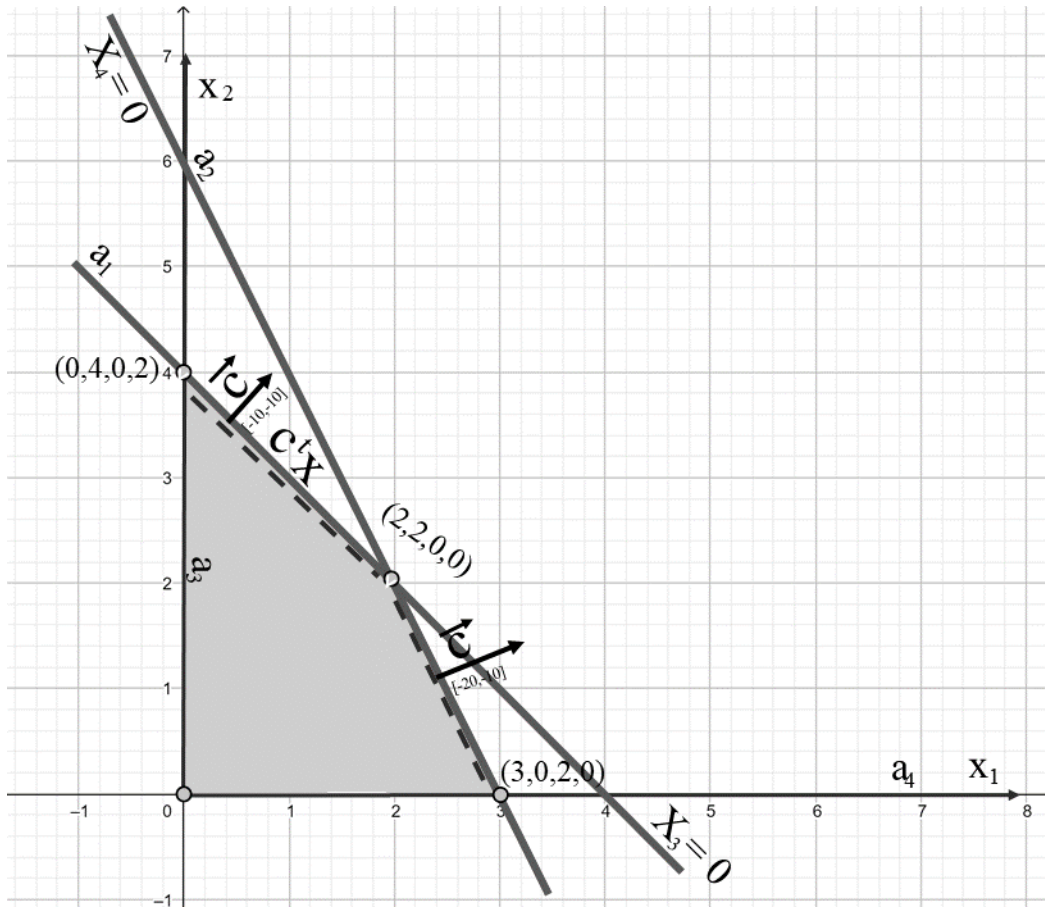


Figura 33: Continuando con el mismo poliedro. a_1 corresponde a $2x_1 + x_2 + x_3 = 8$, a_2 corresponde a $x_1 + x_2 + x_4 = 6$, a_3 y a_4 corresponden a las restricciones de no negatividad. Se puede apreciar que el punto óptimo es $(2, 2, 0, 0)$. Pero si se cambia la función objetivo, ¿Cuánto es el rango para que ese punto siga siendo solución óptima? Se nota que al cambiar la función objetivo no afecta la factibilidad del punto, así que sólo quedaría revisar la optimalidad, es decir, la condición de que los costos reducidos sean positivos. De ahí se desprende que el rango de c_1 sea $(-20, -10)$. Esto se puede apreciar geoméricamente al ver que la función objetivo sea perpendicular a una de las restricciones activas de $(2, 2, 0, 0)$ para cada uno de los casos límites, el cual es que el costo reducido sea igual a 0. De esto se logra obtener los valores de c_1 , los cuales corresponden a que $c_1 = -20$ para que sea perpendicular a a_2 y $c_1 = -10$ para que sea perpendicular a a_1 . Cabe destacar que se tienen infinitas soluciones cuando la función objetivo es perpendicular a la restricción, como se puede apreciar con el primer caso, donde tanto $(2, 2, 0, 0)$ como $(0, 4, 0, 2)$ entregan el mismo valor óptimo.

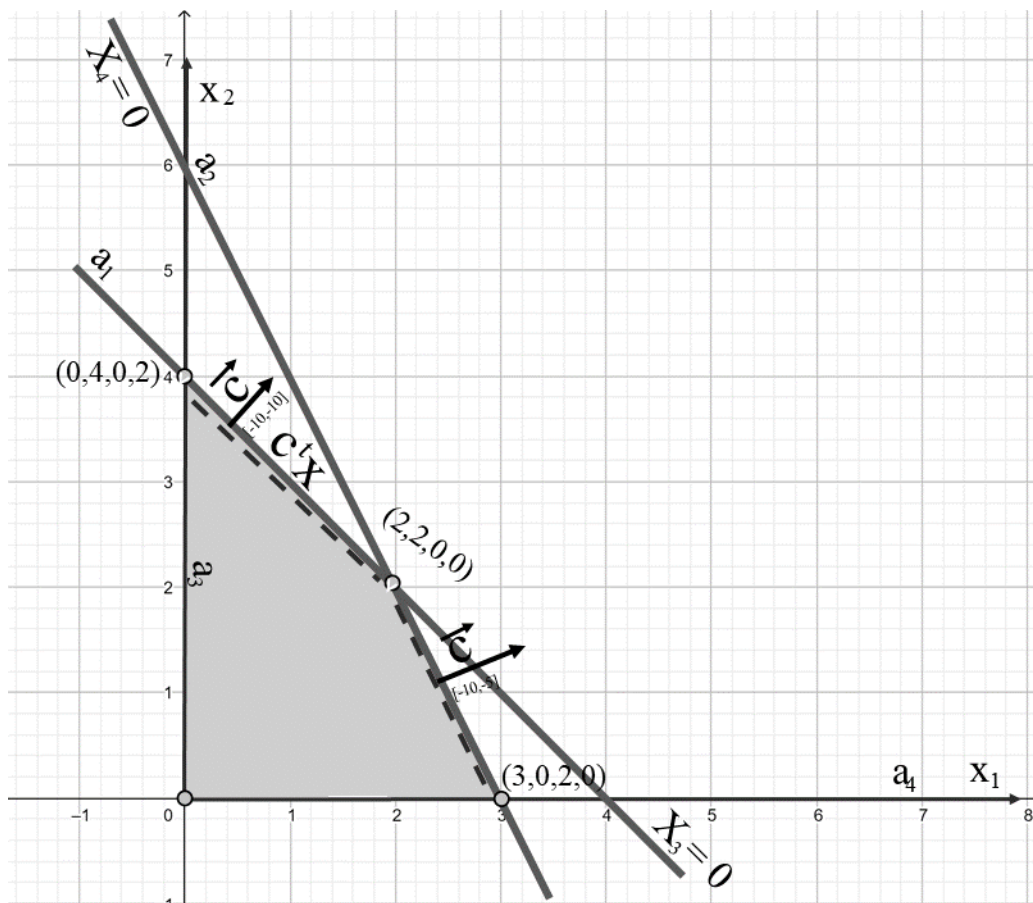


Figura 34: Continuando con el mismo poliedro. a_1 corresponde a $2x_1 + 2x_2 + x_3 = 8$, a_2 corresponde a $2x_1 + x_2 + x_4 = 6$, a_3 y a_4 corresponde a x_1 y $x_2 \geq 0$. Se puede también fijar c_1 y obtener el rango de c_2 , el cual es $[-10, -5]$, obtenido con el mismo procedimiento anterior. Teniendo que $c_2 = -10$ para que sea perpendicular a a_1 y $c_2 = -5$ para que sea perpendicular a a_2 . Cabe destacar que el producto punto igual a 0 (perpendicularidad entre vectores) no es único, pues es posible que haya más vectores con los cuales se obtiene ese resultado como se ve con el caso de fijar c_2 o c_1 y ver la perpendicularidad del vector c a la restricción a_2 , donde se obtiene $c = (-20, -10)$ en el primer caso y $c = (-10, -5)$ para el segundo.

7.3. Agregar nueva variable

Se agrega una nueva variable x_{n+1} , la cual corresponde a la columna A_{n+1} , se verifica que el nuevo PPL queda de la siguiente manera:

$$\begin{aligned} \text{mín} \quad & c^T x + c_{n+1} x_{n+1} \\ & Ax + A_{n+1} x_{n+1} = b \\ & x, x_{n+1} \geq 0 \end{aligned}$$

Pero ¿sigue siendo B óptima? Se sabe que $(x^*, 0)$ sigue siendo solución básica factible con base B , por ende, se verifica su optimalidad

$$\bar{c}_{n+1} = c_{n+1} - c_B^T A_B^{-1} A_{n+1} \geq 0$$

Si $\bar{c}_{n+1} < 0$ entonces se debe utilizar Simplex, empezando con la base B .

7.4. Agregar nueva restricción de igualdad

Se agrega la restricción $a_{m+1}^T x = b_{m+1}$:

Si x^* satisface la nueva restricción, entonces sigue siendo óptima. Pero se asume que la restricción es violada por x^* . Para lo anterior, se construye un problema primal auxiliar.

$$A_{\bar{B}} = \begin{pmatrix} A_B & 0 \\ a^T & -1 \end{pmatrix} \Rightarrow A_{\bar{B}}^{-1} = \begin{pmatrix} A_B^{-1} & 0 \\ a^T A_B^{-1} & -1 \end{pmatrix} \text{ Por ende, la solución nueva es } \bar{x} = A_{\bar{B}}^{-1} b = \begin{pmatrix} A_B^{-1} b \\ a^T A_B^{-1} b - b_{m+1} \end{pmatrix} = \begin{pmatrix} x^* \\ a^T x^* - b_{m+1} \end{pmatrix}$$

En donde, $\bar{x} \geq 0$ y así \bar{x} es factible. Al utilizar Simplex, se comienza con la base \bar{B} . Si la solución óptima es \bar{x}^* , entonces satisface $\bar{x}_{n+1}^* = 0$, el nuevo problema es factible, \bar{x}^* lleva la solución. En cambio, si la solución óptima \bar{x}^* satisface $\bar{x}_{n+1}^* > 0$, entonces el nuevo problema es infactible, asumiendo que M es lo suficientemente grande.

7.5. Cambios en entrada de A

La entrada a_{ij} en la matriz A se cambia a $a_{ij} + \delta$. Si la columna A_j no está en la base, la matriz A_B no cambia, lo que implica que la solución antigua sigue siendo factible, pero los costos reducidos de la columna A_j cambian, por ende, la condición de optimalidad queda como sigue $c_j - c_B^T A_B^{-1} (A_j + \delta e_i) \geq 0$, en donde si A_j es violada, puede ser traída a la base y, por ende, se debe continuar con Simplex. En cambio, si A_j se encuentra en la base, entonces los pasos se vuelven más complicados.

7.6. Dependencia global en b

En el punto de vista global, se tiene que $P(b) := \{x | Ax = b \text{ and } x \geq 0\}$, en donde interesan los valores para b tal que $P(b) \neq \emptyset$. Al mismo tiempo se define $S = \{b | P(b) \neq \emptyset\} = \{Ax | x \geq 0\}$ y así S es convexo. Es importante notar que el costo óptimo, se define como $F(b) = \text{mín}_{x \in P(b)} c^T x$, en donde si se asume que el Dual es factible, entonces $\{y | A^T y \leq c\} \neq \emptyset$, lo cual implica que el costo óptimo $F(b)$ es finito para cada $b \in S$.

Teorema 40. *El costo óptimo $F(b)$ es función convexa en S .*

Demostración. Sean $b^1, b^2 \in S$ y x^1, x^2 soluciones óptimas a los problemas siguientes:

$$\begin{aligned} \text{mín} \quad & c^T x \\ & Ax = b^1 \\ & x \geq 0 \end{aligned}$$

$$\begin{aligned} \text{mín} \quad & c^T x \\ & Ax = b^2 \\ & x \geq 0 \end{aligned}$$

Lo anterior implica que $F(b^1) = c^T x^1$ y $F(b^2) = c^T x^2$. Ahora considerando $\lambda \in [0, 1]$, se verifica que:

$$\begin{aligned} \text{mín} \quad & c^T x \\ & Ax = \lambda b^1 + (1 - \lambda)b^2 \\ & x \geq 0 \end{aligned}$$

En donde $\bar{x} = \lambda x^1 + (1 - \lambda)x^2$ es una solución factible y \bar{x} es la solución factible para el lado derecho de la ecuación $\lambda b^1 + (1 - \lambda)b^2$, lo cual implica que $F(\lambda b^1 + (1 - \lambda)b^2) \leq c^T \bar{x} = \lambda c^T x^1 + (1 - \lambda)c^T x^2 = \lambda F(b^1) + (1 - \lambda)F(b^2)$. \square

7.7. Dependencia global en c

Se tiene el poliedro no vacío $P = \{x | Ax = b\}$. Si c cambia, la factibilidad no se ve afectada. Para algún c en donde el óptimo es finito y para algún c en donde el PPL sea no acotado, ¿Cómo se pueden describir los valores respectivos de c ?

- Si el óptimo es finito, el dual es factible.
- Si el óptimo es infinito, el dual es infactible.

Se define $Q(c) := \{y | A^T y \leq c\}$ como el poliedro dual y el conjunto siguiente $T := \{c | Q(c) \neq \emptyset\}$. Entonces:

- Si $c \notin T$, implica que el dual es infactible, por ende, el primal es infactible o no acotado, pero dado que P no es vacío, entonces el primal es no acotado.
- Si $c \in T$, implica que el primal posee un óptimo finito.

Teorema 41. *El conjunto de todos los c para cada costo óptimo finito, es convexo.*

Demostración. Sea $c_1, c_2 \in T$, se quiere probar que $\lambda c_1 + (1 - \lambda)c_2 \in T$ para todo $\lambda \in [0, 1]$, lo cual implica que $\exists y_1, y_2$ s.a. $A^T y_1 \leq c_1$ y $A^T y_2 \leq c_2$. Entonces se construye una solución \bar{y} para $A^T \bar{y} \leq \lambda c_1 + (1 - \lambda)c_2$, en donde se define $\bar{y} := \lambda y_1 + (1 - \lambda)y_2$.

Quedando entonces $A^T \bar{y} = A^T(\lambda y_1 + (1 - \lambda)y_2) = \lambda A^T y_1 + (1 - \lambda)A^T y_2 \leq \lambda c_1 + (1 - \lambda)c_2$, lo cual implica que $Q(\lambda c_1 + (1 - \lambda)c_2) \neq \emptyset \Rightarrow \lambda c_1 + (1 - \lambda)c_2 \in T$ \square

Para cada $c \in T$, sea $G(c)$ la notación para el costo óptimo, se quiere describir: $G : T \rightarrow \mathbb{R}$, entonces sean $\bar{x}^1, \bar{x}^2, \dots, \bar{x}^N$ soluciones básicas factibles para P , implica que $G(c) = \min_{i=1, \dots, N} c^T \bar{x}_i$ para todos los $c \in T$, es decir, G es el mínimo conjunto finito de funciones lineales, lo que implica que G es una función lineal cóncava por partes.

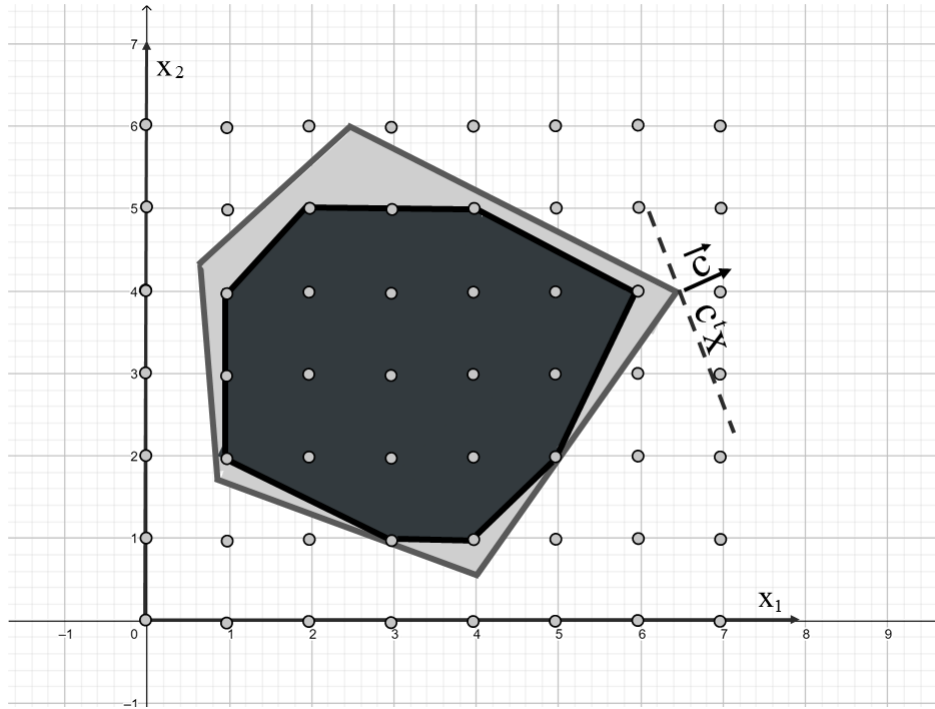


Figura 35: La región factible de un problema de optimización lineal entero, corresponde a los puntos enteros que son factibles en el problema de optimización lineal (PL) que se obtiene al quitarle la restricción que sus puntos sean enteros y, por lo tanto, el conjunto definido por el PE no es un poliedro.

8. Programación entera, Branch & Bound

8.1. Métodos de programación entera

Se ha visto hasta ahora que la programación lineal se puede representar con el siguiente PPL:

$$\begin{aligned} \text{mín } & c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

Considerando que las regiones factibles del PPL anterior, forman un poliedro en donde el óptimo estaría en un punto extremo. En cambio, en la programación lineal entera (PE), se ve lo siguiente:

$$\begin{aligned} \text{mín } & c^T x \\ & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{Z}^n \end{aligned}$$

La región factible en el caso de la PE no es un poliedro. Usualmente, los problemas de PE son difíciles de resolver en comparación a los PPL's.

8.1.1. Planos cortantes

1. Se debe resolver la relajación lineal del problema de programación entera (PPE):

$$\begin{aligned} \text{mín} \quad & c^T x \\ & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{Z}^n \end{aligned}$$

Luego de ello, se encuentra la solución óptima x^* para el PPL.

$$\begin{aligned} \text{mín} \quad & c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

2. Si $x^* \in \mathbb{Z}^n$, entonces es óptimo para PPE, por ende, se debe parar.
3. Si lo anterior no se cumple, entonces se debe agregar una inecuación válida para el PPL que sea también válida para el PPE, pero no para x^* . Luego de lo anterior, se debe volver al paso 1.

Definición 42. Una **inecuación válida** o un **corte** para un PPE es una restricción lineal de la forma $\sum_{j=1}^n a_j x_j \leq b_j$, la cual es satisfecha por todas las soluciones enteras factibles.

8.1.2. Problema de la mochila

Dados los ítems I , cada ítem del tipo i posee un tamaño s_i y valor w_i . Para cada tipo $i \in I$ existen infinitos ítems disponibles, pero el tamaño de la mochila es U , por ende, el objetivo del problema es seleccionar un conjunto de ítems tal que se puedan llevar en la mochila. En otras palabras, se busca maximizar el valor total de los ítems seleccionados. Por ende, la formulación del PPE es:

$$\begin{aligned} \text{máx} \quad & \sum_{i \in I} w_i x_i \\ & \sum_i s_i x_i \leq U \\ & x_i \geq 0 \quad \forall i \in I \\ & x_i \in \mathbb{Z} \quad \forall i \in I \end{aligned}$$

Ejemplo: $U = 50$

Tipos de ítems	1	2	3	4	5
s	24	11	51	26	30
w	37	12	500	50	12

$$\begin{aligned} \text{máx} \quad & 37x_1 + 12x_2 + 500x_3 + 50x_4 + 41x_5 \\ \text{s.t.} \quad & 21x_1 + 11x_2 + 51x_3 + 26x_4 + 30x_5 \leq 50 \\ & x_i \geq 0 \quad \forall i \in I \\ & x_i \in \mathbb{Z} \quad \forall i \in I \end{aligned}$$

La solución óptima es $(1, 0, 0, 1, 0)$ con un valor de función objetivo de 87. Cabe destacar que la solución óptima de la relajación lineal es $(0, 0, 0, \dots, 98, 0, 0)$ con un valor de 490,196.

Se agrega el corte $x_3 \leq 0$, el cual es válido para todas las soluciones enteras tal que $s_3 > U$.

La nueva solución óptima de la relajación lineal es $(0, 0, 0, 1, \dots, 9, 0)$ con un valor de 96,154.

Se agrega el corte $x_4 \leq 1$, el cual es válido, ya que $2 \cdot s_4 > U$.

La solución óptima de la relajación lineal es $(1, 0, 0, 1, 0)$, por ende, su valor es 87, en donde se concluye que es la solución óptima para el PPE.

En general:

- Todas las inecuaciones $x_i \leq \lfloor \frac{s_i}{U} \rfloor$ son válidas.
- Se tiene el siguiente caso especial: $x_i \leq 0$ si $s_i < U$.

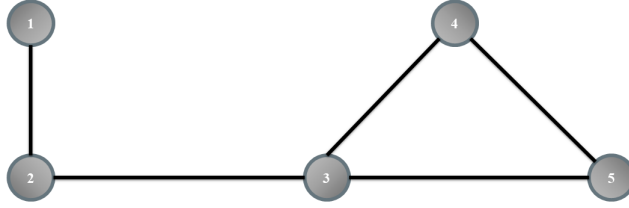


Figura 36: El problema consiste en construir un subconjunto V' del conjunto V de nodos del grafo, tal que ningún arco del grafo toque a 2 nodos del conjunto V' . Se puede notar que, observando el problema, la cardinalidad máxima del conjunto V' es 2 y que este conjunto podría estar compuesto por los nodos $(1,3)$ ó $(1,4)$ ó $(1,5)$ ó $(2,4)$ ó $(2,5)$.

8.1.3. Conjunto independiente

Se tiene el grafo no dirigido $G = (V, E)$: El objetivo que se persigue es seleccionar un subconjunto de los vértices $V' \subseteq V$ tal que ninguna arista E conecte dos vértices en V' , maximizando así $|V'|$. Por ende, la formulación del PPE es:

$$\begin{aligned} \text{máx} \quad & \sum_{i \in V} x_i \\ & x_i + x_j \leq 1 \quad \forall (i, j) \in E \\ & x_i \in \{0, 1\} \quad \forall i \in V \end{aligned}$$

Se tiene entonces que el valor de la solución óptima es 2.

Para la relajación lineal se tiene $x = (0,5, 0,5, 0,5, 0,5, 0,5)$ con un valor de 2,5. Fuera de los nodos $\{3, 4, 5\}$ sólo uno puede ser seleccionado.

Entonces se agrega la inecuación válida $x_3 + x_4 + x_5 \leq 1$, por ende, la nueva solución para la relajación lineal es $x = (1, 0, 1, 0, 0)$, la cual es el óptimo para el PPE.

Definición 43. Sea $G = (V, E)$ un grafo no dirigido. El set de nodos $V' \subseteq V$ es llamado **Clique** si para cualquier par de nodos $i, j \in V'$ existe una arista $(i, j) \in E$.

Para cualquier clique $V' \subseteq V$ se tiene una inecuación válida $\sum_{i \in V'} x_i \leq 1$.

Los ciclos poseen un número particular de nodos, en donde sólo dos de ellos se pueden seleccionar, entonces $\sum_{i=1}^5 x_i \leq 2$ es un corte válido. En general, sea i_1, \dots, i_{2k+1} el ciclo particular con $k \in \mathbb{N}$, entonces $\sum_{\ell=1}^{2k+1} x_{i_\ell} \leq k$ es un corte válido.

8.1.4. PPE general en forma estándar

Al igual que los PPL's, los PPE's también poseen forma estándar, la cual se detallará a continuación:

Ejemplo:

Sea x^* una solución óptima del PPL con $x^* \notin \mathbb{Z}^n$, en donde N son los índices de las variables no básicas de x^* . Se tiene que $x_i^* = 0$ para todo $i \in N$ en cualquier solución del PPE, una variable x_j con $j \in N$ debe ser diferente de cero.

8.1.5. Cortes de Gomory

Hasta ahora se han visto problemas específicos en donde se cortan planos, pero en la presente sección se revisará un método general para obtener cortes de planos, en donde, se especificará el primer algoritmo para resolver PPE's.

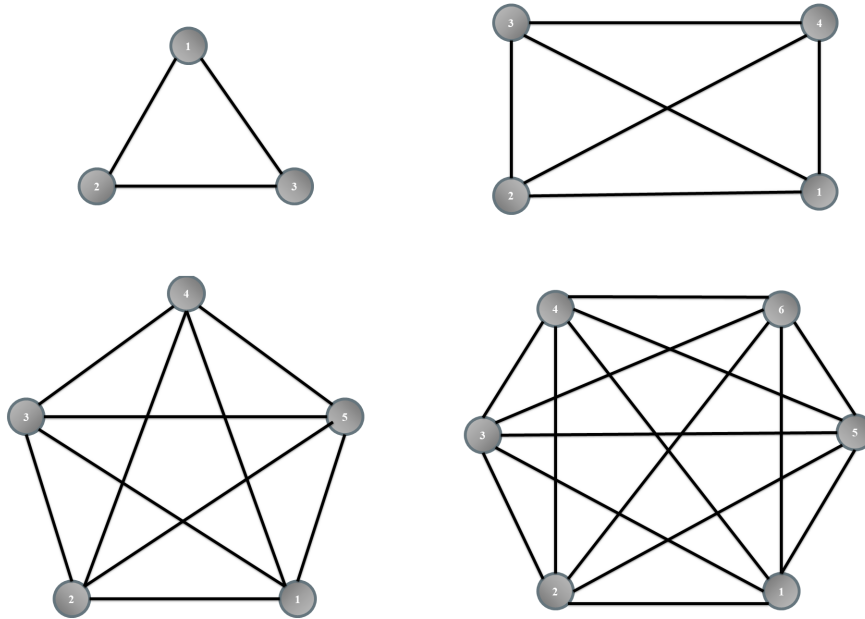


Figura 37: Un clique es un grafo no dirigido con la característica que todos los nodos de este tienen aristas que los conectan con el resto de los nodos del grafo.

Ejemplo:

$$\begin{aligned}
 \text{mín} \quad & x_1 - x_2 \\
 & -4x_1 + 6x_2 \leq 9 \\
 & x_1 + x_2 \leq 4 \\
 & x_1, x_2 \geq 0
 \end{aligned}$$

Pasamos el problema a forma estándar.

$$\begin{aligned}
 \text{mín} \quad & x_1 - 2x_2 \\
 \text{s.a.} \quad & -4x_1 + 6x_2 + x_3 = 9 \\
 & x_1 + x_2 + x_4 = 4 \\
 & x_i \geq 0 \quad \forall i
 \end{aligned}$$

Cabe destacar que todos los coeficientes en la matriz y en el vector b son enteros, por ende, se puede asumir que las variables de holgura son enteras.

Se resuelva la relajación lineal, en donde la solución óptima es $x^1 = (1,5, 2,5, 0, 0)$ con $B = \{1, 2\}$, en donde la solución entera debe satisfacer $Ax = b$, $A_B = \begin{pmatrix} -4 & 6 \\ 1 & 1 \end{pmatrix}$, $A_B^{-1} = \begin{pmatrix} -0,1 & 0,6 \\ 0,1 & 0,4 \end{pmatrix}$, es decir,

$$A_B^{-1}Ax = A_B^{-1}b \quad \begin{array}{rcl} x_1 & -0,1x_3 & +0,6x_4 = 1,5 \\ x_2 & +0,1x_3 & +0,4x_4 = 2,5 \end{array}$$

Se elige la igualdad tal que b sea no entera, luego en la segunda igualdad se tiene $x_3, x_4 \geq 0$, es decir, $x_2 + [0,1] x_3 + [0,4] x_4 \leq 2,5$, en donde todos los coeficientes y variables son enteras, por ende, $x_2 + [0,1] x_3 + [0,4] x_4 \leq [2,5]$, en donde, $x_2 \leq 2$ es el nuevo corte. Cabe destacar que al agregar un nuevo corte al PPL se debe comenzar otra vez desde el inicio.

En general, Sea x^* la solución óptima no entera, B la base, N los índices de todas las variables no básicas y x la solución entera arbitraria, entonces se tiene $Ax = b$, por ende, $A_B^{-1}Ax = A_B^{-1}b$, es decir, $A_B^{-1}A_Bx_B + A_B^{-1}A_Nx_N = A_B^{-1}b$, en donde, $x_B + A_B^{-1}A_Nx_N = A_B^{-1}b$ y $x_B^* = A_B^{-1}b$, $x_N^* = 0$. Cabe destacar que:

x^* posee componente fraccionariaes decir $A_B^{-1}b$ posee componente fraccionaria

Entonces $\bar{a}_{ij} := (A_B^{-1}A_j)_i$ y $\bar{b}_i := (A_B^{-1}b)_i$, sea i tal que \bar{b}_i es fraccionaria, se tiene que $x_{B(i)} + \sum_{j \in N} \bar{a}_{ij}x_j = \bar{b}_i$, en donde todos $x_j \geq 0$, entonces $x_{B(i)} + \sum_{j \in N} [\bar{a}_{ij}] x_j \leq x_{B(i)} + \sum_{j \in N} \bar{a}_{ij}x_j = \bar{b}_i$ y así cualquier solución del PPE $x \in \mathbb{Z}^n$, por ende, $x_{B(i)} + \sum_{j \in N} [\bar{a}_{ij}] x_j \leq [\bar{b}_i]$, la cual no se cumple por x^* , ya que $x_{B(i)}^* = \bar{b}_i > [\bar{b}_i]$ y $x_j^* = 0$ por cada $j \in N$. Se agrega el corte y se resuelve el nuevo PPL.

Ejemplo:

Se tiene el nuevo corte $x_2 \leq 2$, en donde se agrega la gualdad $x_2 + x_5 = 2$ con una nueva variable de holgura $x_5 \geq 0$

$$\begin{array}{rcl}
 & \text{mín} & x_1 - 2x_2 \\
 \text{s.a.} & -4x_1 + 6x_2 + x_3 & = 9 \\
 & x_1 + x_2 & + x_4 = 4 \\
 & x_2 & + x_5 = 2 \\
 & & x_i \geq 0 \forall i
 \end{array}$$

En donde la solución óptima es $(0,75, 2, 0, 1,25, 0, 0)$ con la base $B = \{1, 2, 4\}$, $A_B = \begin{pmatrix} -4 & 6 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, así se tiene

$$A_B^{-1} = \begin{pmatrix} -1/4 & 0 & 3/2 \\ 0 & 0 & 1 \\ 1/4 & 1 & -5/2 \end{pmatrix} \text{ lo que implica que } A_B^{-1}b = \begin{pmatrix} 0,75 \\ 2 \\ 1,25 \end{pmatrix}$$

$$A_B^{-1}A = \begin{pmatrix} 1 & 0 & -1/4 & 0 & 3/2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1/4 & 1 & -5/2 \end{pmatrix}$$

Entonces $x_1 - \frac{1}{4}x_3 + \frac{3}{2}x_5 = 0,75$, lo cual implica que $x_1 - [\frac{1}{4}]x_3 + [\frac{3}{2}]x_5 = [0,75]$, es decir, $x_1 - x_3 + x_5 \leq 0$. Se agrega la variable de holgura : $x_1 - x_3 + x_5 + x_6 = 0$, en donde se tiene el nuevo punto óptimo fraccionario $(1, 2, 1, 1, 0, 0)$, lo que implica que el óptimo es entero.

La inecuación $x_1 - x_3 + x_5 \leq 0$ en términos de variables x_1, x_2 , se tiene que $x_3 = 9 + 4x_1 - 6x_2$; $x_5 = 2 - x_2$; $x_1 - x_3 + x_5 \leq 0$, es decir, $x_1 - (9 + 4x_1 - 6x_2) + (2 - x_2) \leq 0$ con $-3x_1 + 5x_2 \leq 7$

8.1.6. Branch and bound

También llamado ByB es un algoritmo muy utilizado al momento de analizar programación entera, como se verifica en el siguiente ejemplo: **Ejemplo:**

$$\begin{array}{rcl}
 \text{mín} & x_1 - x_2 & \\
 & -4x_1 + 6x_2 & \leq 9 \\
 & x_1 + x_2 & \leq 4 \\
 & x_1, x_2 & \geq 0
 \end{array}$$

Sea $U := \infty$, U almacena valores de la mejor solución encontrada hasta el momento. Por ende, se resuelve el PPL, encontrando $x^{(1)} = (1,5, 2,5)$

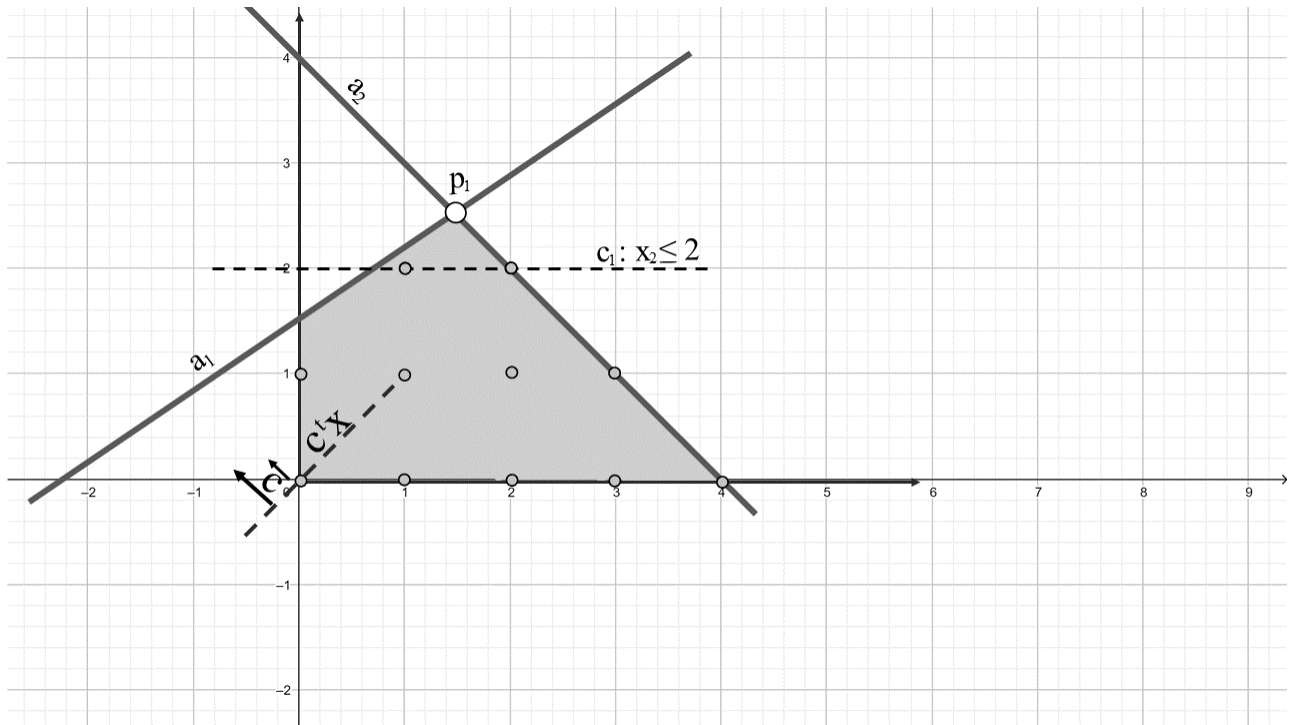


Figura 38: La función objetivo es $\text{mín } x_1 - 2x_2$. a_1 corresponde a $-4x_1 + 6x_2 \leq 9$, a_2 corresponde a $x_1 + x_2 \leq 4$, $x_1, x_2 \geq 0$ y son enteros. En la primera iteración de la resolución del PL obtenido al relajar el PE, se obtiene que p_1 es la solución óptima, pero esta no es entera en ninguna de sus coordenadas, así que se agrega la restricción $x_1 \leq 2$ obtenida por Corte de Gomory.

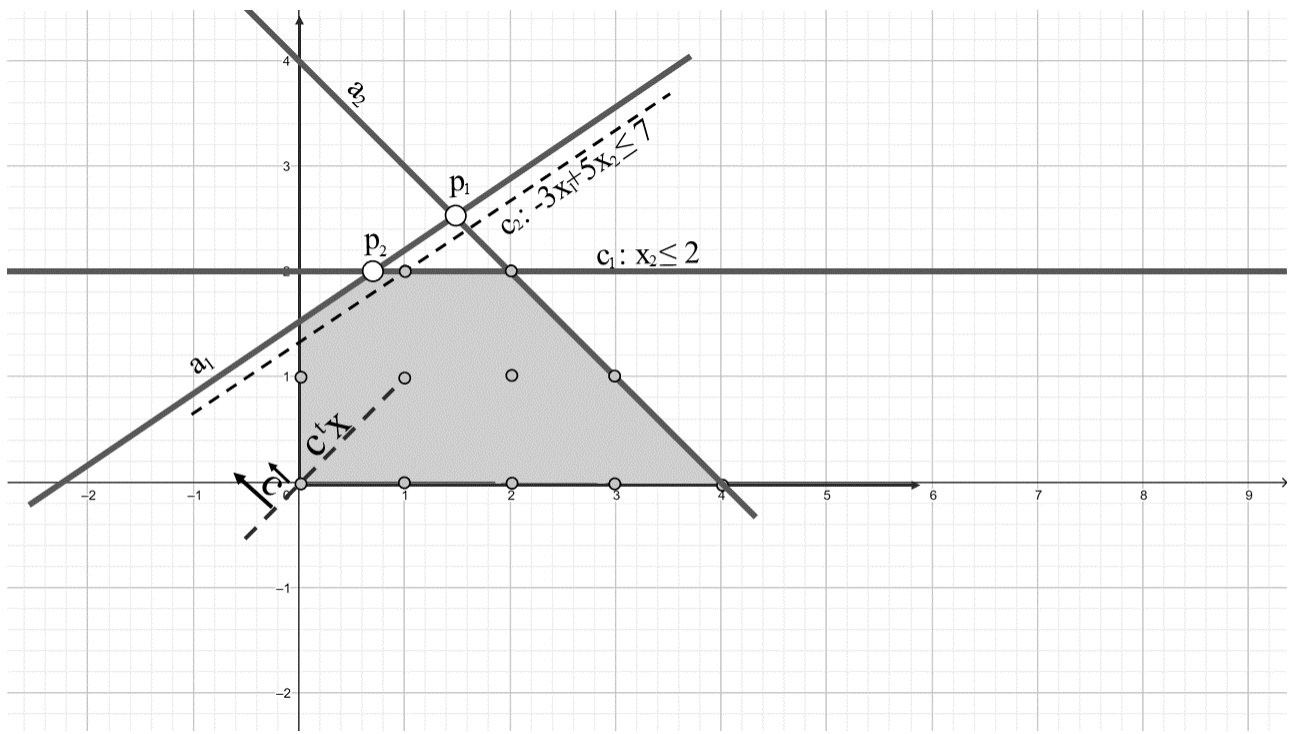


Figura 39: Se resuelve el nuevo PL y se obtiene que p_2 es la nueva solución óptima, pero esta tampoco es entera, así que se agrega la restricción $-3x_1 + 5x_2 \leq 7$.

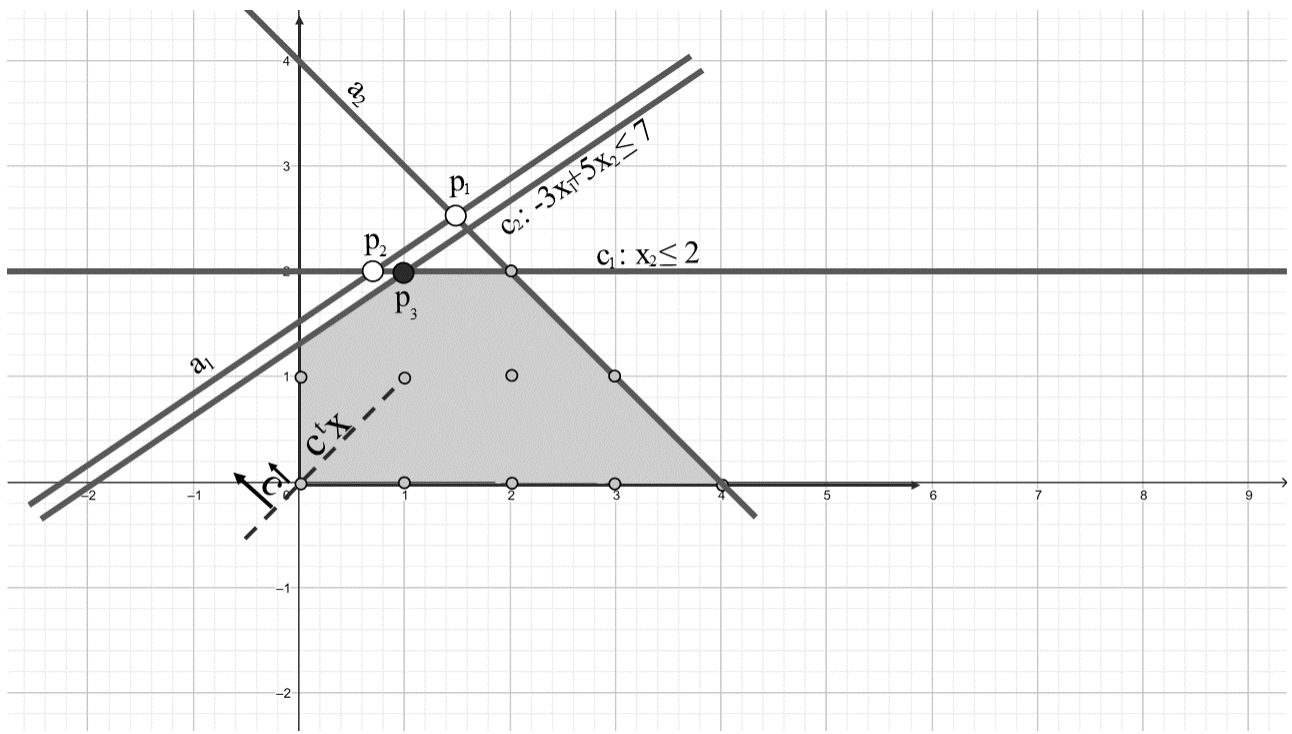


Figura 40: Finalmente, se resuelve el nuevo PL, y se obtiene el punto $P_3 = (1, 2)$ como solución óptima, y como este es entero, se ha encontrado la solución al PPE.

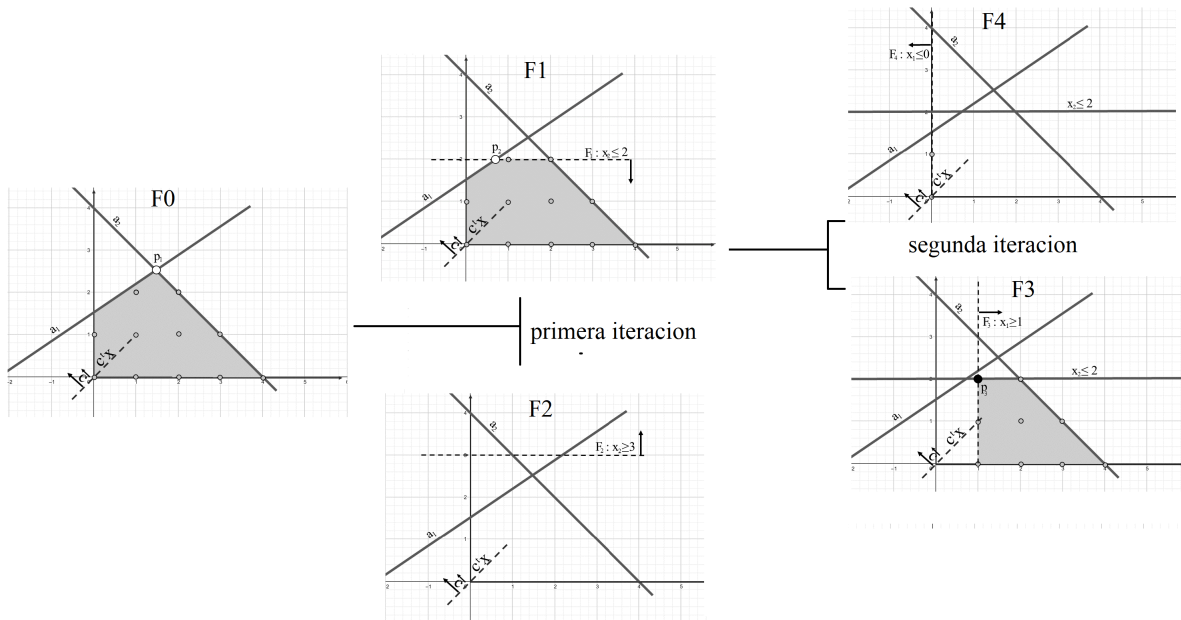


Figura 41: Se pueden apreciar los criterios de entrada de este algoritmo, como también los 3 criterios de salida, los cuales corresponden a infactibilidad, óptimo entero y finalmente un óptimo peor que el óptimo entero encontrado.

Luego $x_2^{(1)}$ es fraccionaria, pero debe ser entera, es por ello que se crean dos subproblemas: F_1, F_2 , entonces F_1 : Agrega la restricción $x_2 \geq 3$; F_2 : Agrega la restricción $x_2 \leq 2$.

Se resuelve el PPL de F_1 : Infactible, por ende, se elimina F_1 . Luego se resuelve el PPL de F_2 , llegando a $x^{(2)} = (\frac{3}{4}, 2)$.

Por lo anterior, se crean nuevamente dos subproblemas: F_3, F_4 . Específicamente, F_3 : Agrega la restricción $x_1 \geq 1$; F_4 : Agrega la restricción $x_1 \leq 0$.

Luego se resuelve el PPL para F_3 , por ende, $x^{(3)} = (1, 2)$ es entero con el valor -3 , entonces si se actualiza U , quería como $U := -3$. No se necesita explorar la región factible de F_3 , ya que la solución óptima fraccionaria es entera. Es decir, encontramos la primera solución entera.

Luego, se resuelve el PPL para F_4 , por ende, $x^{(4)} = (0, 3/2)$ con valor -3 . Entonces como $-3 \geq U$ no se considera el subproblema, ya que no puede obtener una mejor solución entera que la recientemente encontrada.

Se procede a describir el algoritmo de manera general para el siguiente PPE:

$$\begin{aligned} \text{mín} \quad & c^T x \\ & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{Z}^n \end{aligned}$$

Se comienza con $U := \infty$, por ende, se define F_0 para que sea el PPE, por ello se define F_0 para ser activo en cada iteración. Entonces en general:

1. Selecciona un subproblema activo F_i .
2. Resuelve la relajación lineal para F_i . Si F_i es infactible entonces se elimina F_i , en el caso contrario, sea x^* la solución óptima fraccionaria.

3. Si $c^T x^* \geq U$ entonces se elimina F_i , la cual es la mejor solución entera para F_i , la cual no puede ser mejor que U , definido como el valor de la mejor solución encontrada hasta ahora.
4. Si $x^* \in \mathbb{Z}^n$ entonces:
 - Si $c^T x^* < U$, se actualiza $U := c^T x^*$. En donde x^* es la mejor solución encontrada.
 - Se debe eliminar F_i .
5. Si $x^* \notin \mathbb{Z}^n$, sea x_i^* fraccionaria, se crean dos problemas $F_{i'}$ y $F_{i''}$ para luego añadirlos a la lista de subproblemas activos.
 - En $F_{i'}$ se agrega la restricción $x_i \leq \lfloor x_i^* \rfloor$
 - En $F_{i''}$ se agrega la restricción $x_i \geq \lceil x_i^* \rceil$
6. Se elimina F_i .

8.1.7. Branch and cut

Se mezcla Branch and Bound con cortes de planos, entonces al tener el problema se agregan los planos cortantes y luego se realiza ByB con normalidad. Se resuelve:

$$\begin{array}{ll} \text{mín} & x_1 - x_2 \\ & -4x_1 + 6x_2 \leq 9 \\ & x_1 + x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{array}$$

Obteniendo la solución $x^1 = (1,5, 2,5)$, el cual genera subproblemas:

- F_1 : Se agrega la restricción $x_2 \geq 3$; F_2 : Se agrega a restricción $x_2 \leq 2$.
- Se resuelve el PPL F_1 , lo cual da que es infactible.
- F_2 : Se agrega la restricción $-x_1 + x_2 \leq 1$, la cual es válida para cualquier solución entera, ya que $x_2 \leq 2$ y $(0, 2)$ fuera de la región factible.
- Se resuelve el PPL resultante, en donde, $x^3 = (1, 2)$ es entero, por ende, se debe detener la iteración.

8.1.8. Árbol de mínima expansión o Minimum Spanning Tree (MST)

Dado un grafo no direccionado que esté conectado $G = (V, E)$, en donde cada arista posee un costo de $c_e > 0$

Se tiene el objetivo de encontrar el grupo de aristas $E' \subseteq E$ tal que $G = (V, E')$ sea un árbol, en donde se minimice el costo total de E' , es decir, minimizar $c(E') := \sum_{e \in E'} c_e$.

Algunos ejemplos en la vida real para los cuales utilizar MST pueden ser la creación de calles para que el costo sea el menor posible al momento de conectar ciudades, conectar computadores para formar una red de trabajo en donde se quiera minimizar la cantidad de cable a instalar.

Formulación de programación entera 1: Se eliminan los subtours IP_{sub} , es decir, se eliminan todos los ciclos.

$$\begin{array}{l} \text{mín} \sum_{e \in E} c_e \cdot x_e \\ \sum_{e \in E} x_e = n - 1 \\ \sum_{e \in E(S)} x_e \leq |S| - 1, \forall S \subset V \text{ con } \emptyset \neq S \\ x_e \in \{0, 1\}, \forall e \in E \end{array}$$

Se considera el conjunto $S \subseteq V$, sea $E(S) = \{\{u, v\} \in E | u, v \in S\}$. Entonces si la solución contiene ciclos S , entonces la restricción para el conjunto S no se cumple, lo cual implica que cualquier solución produce un grafo sin ciclos y con $n - 1$ aristas, es decir, un árbol.

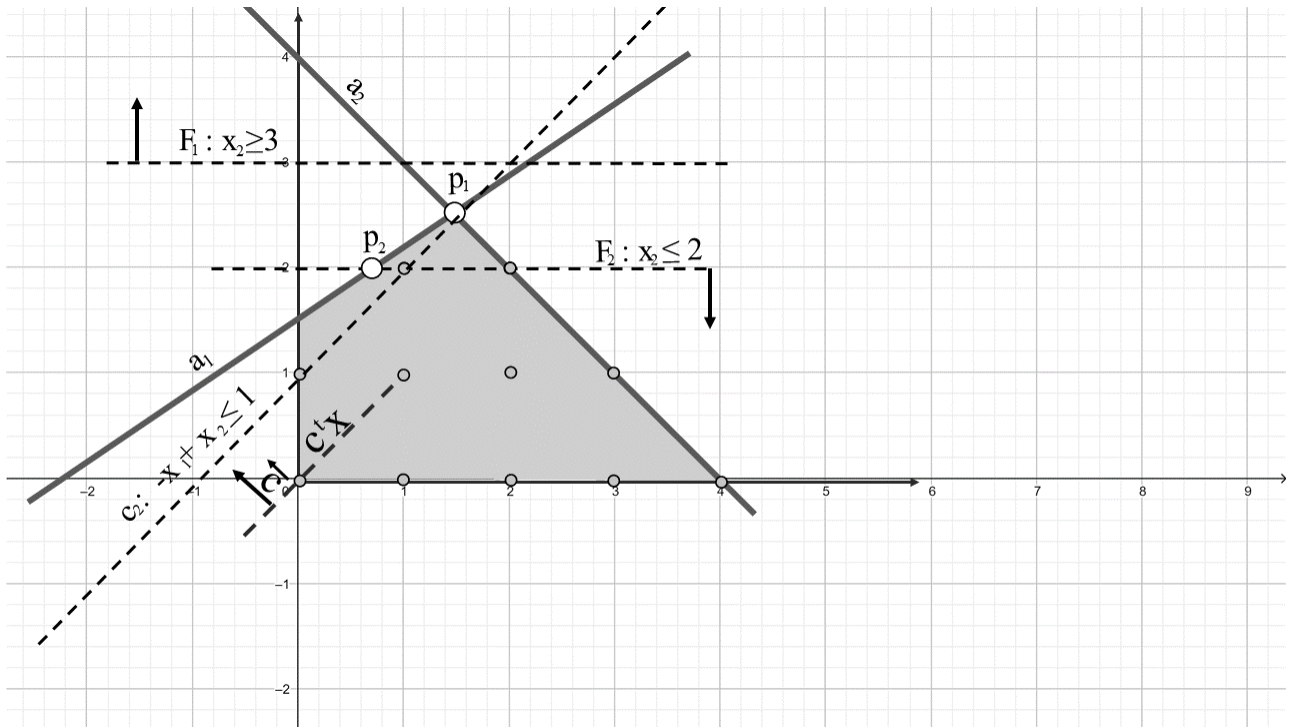


Figura 42: La función objetivo es $\min x_1 - 2x_2$. a_1 corresponde a $-4x_1 + 6x_2 \leq 9$, a_2 corresponde a $x_1 + x_2 \leq 4$, $x_1, x_2 \geq 0$ y son enteros. El algoritmo Branch and Cut es una mezcla de los 2 anteriores mencionados. En esta imagen se muestran los 2 subproblemas a crearse, F_1 y F_2 , como también el corte a realizarse después.

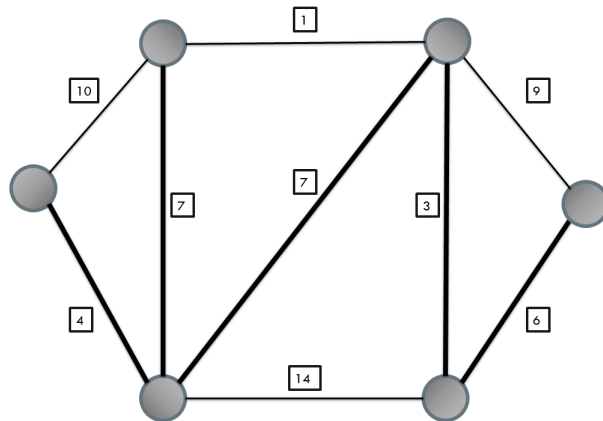


Figura 43: En esta figura se pueden apreciar 6 nodos y 9 aristas, cada una con su costo c_e sobre ellas. El árbol de peso mínimo corresponde a las aristas oscurecidas.

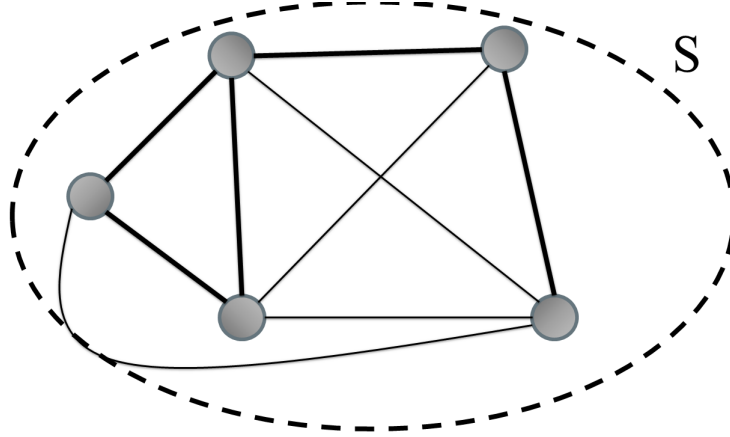


Figura 44: En este ejemplo se puede apreciar que, aunque haya una cantidad claramente mayor de aristas, se requieren al menos 4 para formar un árbol, y si el subconjunto tiene más de 4 aristas, estaría formando un ciclo.

Lema 44. Dado el grafo $G = (V, E)$. Para cualquier árbol $T = (V, E')$ con $E' \subseteq E$ y cualquier otro conjunto de vértices $S \subset V$ con $S \neq \emptyset$, se tiene que T contiene a lo más $|S| - 1$ aristas de $E(S)$.

Demostración. Supone que $|E' \cap E(S)| \geq |S|$ en donde $E' \cap E(S)$ está conectado, así T es un árbol y $E' \cap E(S)$ no contiene un ciclo. Cualquier subconjunto de $|S| - 1$ aristas de $|E' \cap E(S)|$ sin ciclos forma un árbol en S . Es importante destacar que $|E' \cap E(S)| \geq |S|$ implica que $E' \cap E(S)$ contiene un ciclo, por ende, la formulación modela un MST. \square

Formulación de programación entera 2: Formulación de corte IP_{cut}

$$\begin{aligned} & \text{mín} \sum_{e \in E} c_e \cdot x_e \\ & \sum_{e \in E} x_e = n - 1 \\ & \sum_{e \in \delta(S)} x_e \geq 1, \forall S \subset V \text{ con } \emptyset \neq S \neq V \\ & x_e \in \{0, 1\}, \forall e \in E \end{aligned}$$

Produce una solución conectada con $n - 1$ aristas, lo cual implica que es un árbol. Cualquier árbol contiene $n - 1$ aristas en total y al menos una arista desde $\delta(S)$ a cada S con $\emptyset \neq S \neq V$, es decir, la formulación modela un problema de MST. Pero ¿Cuál formulación es mejor? ¿Cuándo corresponde ocupar una formulación de programación entera?

Relajación de PPL: Formulación de programación entera en donde se reemplaza $x_e \in \{0, 1\}$ por $x_e \geq 0$.

La relajación de PPL ideal es cuando esta es la envoltura convexa para todos los puntos enteros, es decir, todos los puntos extremos son enteros, por ende, el algoritmo Simplex encuentra una solución entera óptima de forma directa. Sea $IP :=$ La envoltura convexa de todas las soluciones enteras, $LP_A :=$ El poliedro de la relajación de programación lineal de una formulación de programación entera y $LP_B :=$ Poliedro de la relajación del programa lineal de otra formulación de un programa entero.

Definición 45. Sean A, B dos formulaciones de problemas de programación entera. Sea P_A y P_B los conjuntos factibles de la correspondientes relajaciones de problemas lineales, respectivamente. Luego, A es al menos tan fuerte como B si $P_A \subseteq P_B$.

Teorema 46. Sean P_{sub} y P_{cut} los conjuntos de puntos factibles de las relajaciones de programación lineal de IP_{sub} y IP_{cut} , respectivamente.

1. Se tiene $P_{sub} \subseteq P_{cut}$ y hay ejemplos en donde esta inclusión es estricta.

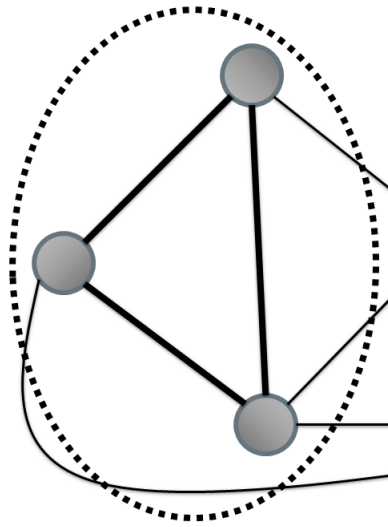


Figura 45: En el ejemplo anterior, este subconjunto del grafo estaría formando un ciclo, pues es un subcamino del grafo inicial donde el nodo final y el nodo inicial son el mismo.

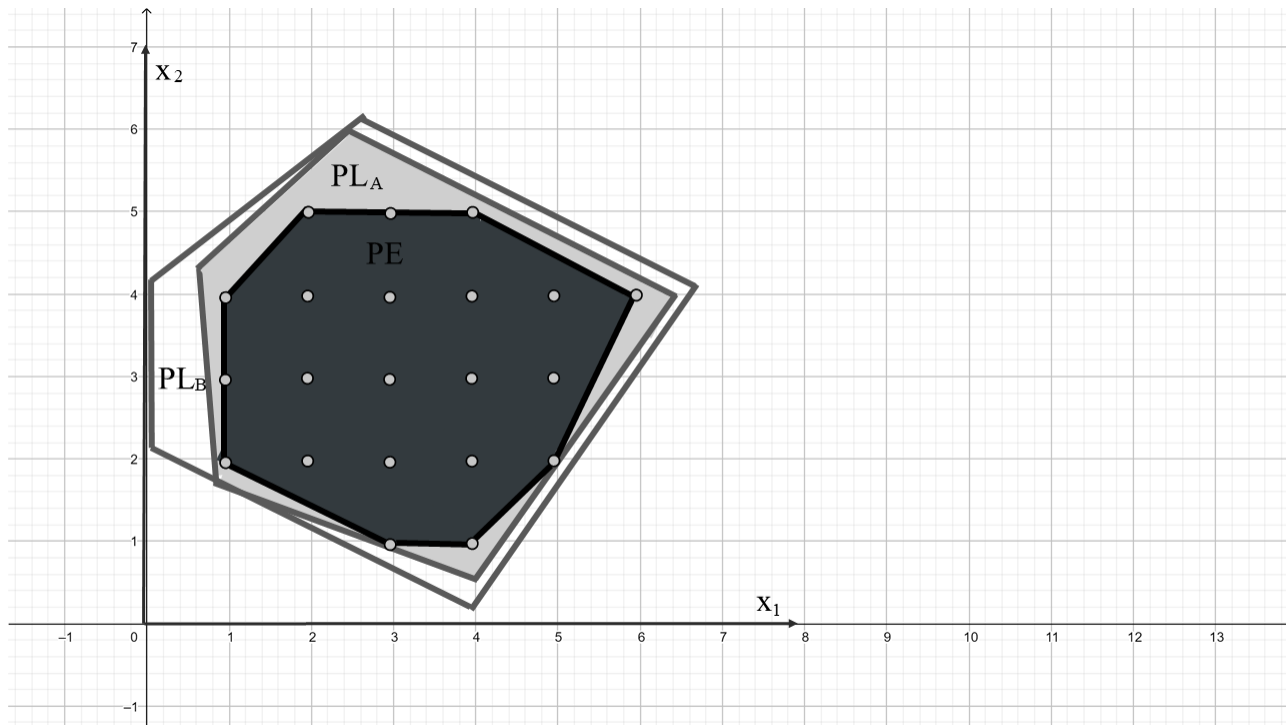


Figura 46: Se pueden apreciar la envoltura convexa de los puntos enteros PE o IP en inglés, como también las distintas formulaciones PL_A y PL_B , donde claramente la primera es mejor que la otra.

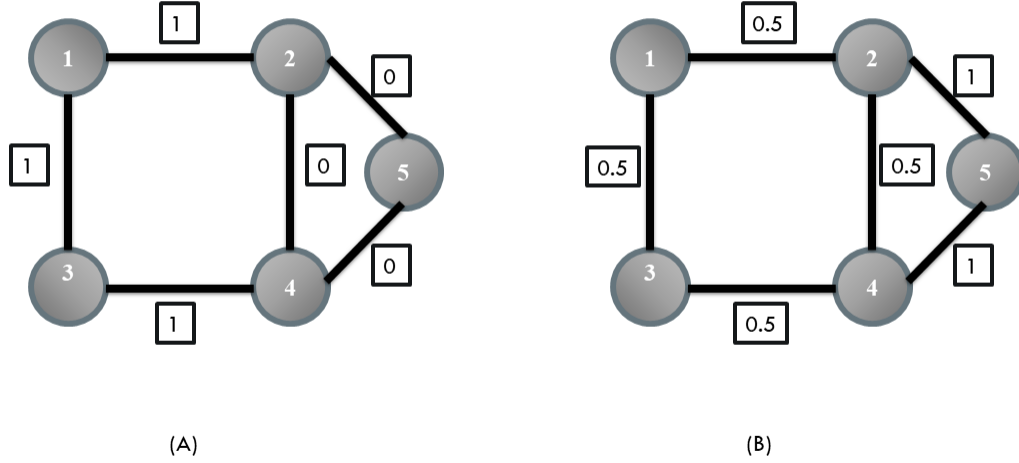


Figura 47: (A) muestra los costos, (B) muestra la solución fraccional a P_{cut} , pero se tiene que esta solución toma 2,5 aristas de $E(\{2, 4, 5\})$, pero la restricción en P_{sub} para $\{2, 4, 5\}$ dice que como máximo 2 de estos pueden ser elegidos (fraccionalmente), lo que implica que esta solución está en P_{cut} pero no en P_{sub} .

2. El poliedro P_{cut} puede tener puntos extremos fraccionarios.
3. P_{sub} es igual a la envoltura convexa de todas las soluciones enteras.

Resolviendo MST mediante IP/ LP_{sub} puede tomar mucho tiempo, por ejemplo, si $n \geq 265$, entonces más de 10^{80} desigualdades son aproximadamente el número de átomos en el universo. Es por lo anterior que hay algoritmos muchos más rápidos, por ejemplo, el algoritmo de Prim.

8.2. Algoritmo de Prim

Se toma un nodo arbitrario v con $V_1 := \{v\}$ y $E_1 := \emptyset$. Luego de ello se inicia con el árbol $T_1 = (V_1, E_1)$ y $k := 1$ cuando $k < n$.

Sean E' todas las aristas $\{u, v\}$ tal que $u \in V_k$ y $v \notin V_k$ y $e^* = \{u^*, v^*\}$ sea la arista en E' con el menor costo, se define $V_{k+1} := V_k \cup \{v^*\}$ y $E_{k+1} := E_k \cup \{e^*\}$ con $k := k + 1$, generando así un loop.

En cada iteración k hay dos conjuntos de vértices V_k y $V \setminus V_k$. Se toma la arista del costo mínimo que conecta a ambos.

Lema 47. *Propiedad de Corte:* Sea $G = (V, E)$ un grafo no dirigido y \bar{E} un conjunto de aristas tal que el grafo $(V, E \setminus \bar{E})$ tiene al menos dos componentes conectados. Sea e una arista con costo mínimo en \bar{E} . Entonces existe un MST T de G que contiene e . Si $\tilde{E} \subseteq E \setminus \bar{E}$ es un conjunto de aristas que esté contenido dentro de algún MST, entonces $\tilde{E} \cup \{e\}$ está contenido también en algún MST.

Teorema 48. *Se el grafo $G_n := (V_n, E_n)$ analizado mediante el Algoritmo de Prim es un MST*

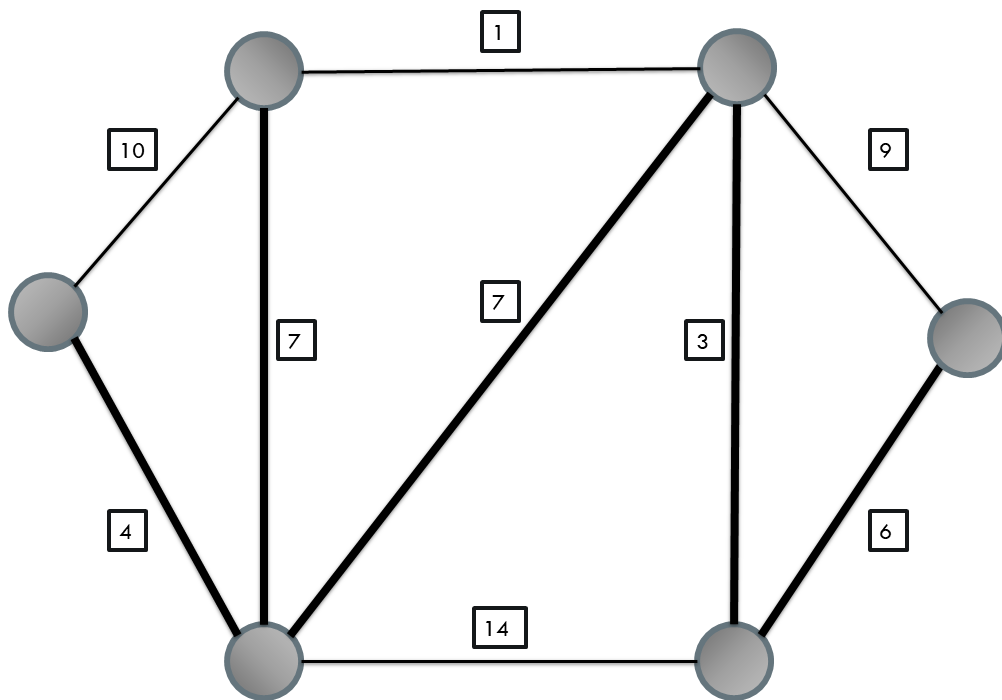


Figura 48: Se vuelve al ejemplo anterior para así ejemplificar otro algoritmo.

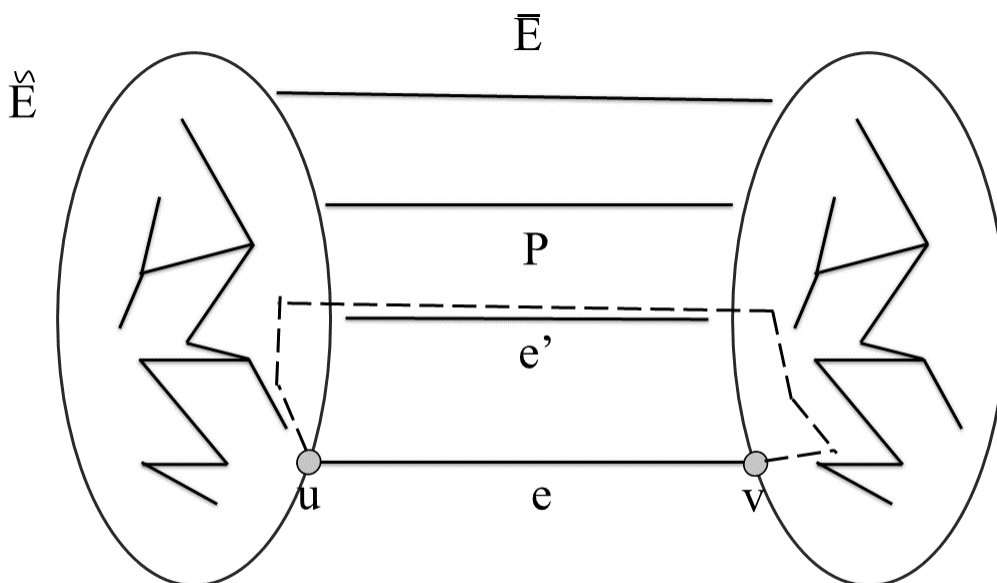


Figura 49: Representación gráfica de la propiedad de corte.

Demostración. Por inducción se verifica que para cada k el árbol (V_k, E_k) es parte de algún MST, es decir, existe un conjunto de aristas $\bar{E}_k \supseteq E_k$ tal que (V, \bar{E}_k) es un MST. Entonces en $k = 1$ con $E_1 = \emptyset$, suponiendo que esa afirmación es cierta para valores en donde $k < n$. Sea $e^* = \{u^*, v^*\}$ la arista que se agrega a E_k en la iteración k , es decir, $u^* \in V_k$ y $v^* \notin V_k$.

Se define $\tilde{E} := E_k$ y \bar{E} como todas aristas $e = \{u, v\}$ tal que $u \in V_k$ y $v \notin V_k$. Además, e^* es la arista con costo mínimo en \tilde{E} .

La propiedad de corte implica que $E_k \cup \{e^*\} = E_{k+1}$ está contenida en un MTS de G . \square

Lema 49. *Propiedad de Corte:* Sea $G = (V, E)$ un grafo no dirigido y \bar{E} un conjunto de aristas tal que el grafo $(V, E \setminus \bar{E})$ tiene al menos dos componentes conectados. Sea e una arista con costo mínimo en \bar{E} . Entonces existe un MST T de G que contiene e . Si \tilde{E} es un conjunto de aristas que esté contenido dentro de algún MST y $\tilde{E} \cap \bar{E} = \emptyset$, entonces $\tilde{E} \cup \{e\}$ está contenido también en algún MST.

Demostración. La segunda afirmación implica la primera cuando $\tilde{E} = \emptyset$. Ahora, la prueba de la primera se demuestra cuando E' sea el conjunto de aristas en $G' = (V, E')$, el cual es un MST para G y $\tilde{E} \subseteq E'$. Sea $e = \{u, v\}$, si $e \in E'$ se cumple, en caso contrario G' es un árbol de expansión para G , lo cual implica que hay un camino P desde u hacia v en G' . Al mismo tiempo, \bar{E} separa u y v , lo que implica que P debe contener la arista e' desde \bar{E} . Luego se define $E'' := E' \setminus \{e'\} \cup \{e\}$, con $G'' = (V, E'')$ es un árbol de expansión: Cuando se agrega e se cierra un ciclo que contiene e' , entonces $c(e) \leq c(e')$ y así $c(E'') \leq c(E')$ lo que implica que G'' es un MST y $\tilde{E} \cup \{e\} \subseteq E''$. \square

8.3. Algoritmo de Kruskal

En el mismo contexto de Grafos, se encuentra el algoritmo de Kruskal, el cual tiene por objetivo encontrar un árbol recubridor mínimo en un grafo, es decir, busca un subconjunto de aristas que, formando un árbol, incluyen todos los vértices y donde la suma de todas las aristas sea el mínimo.

El algoritmo se inicia con $E_0 := \emptyset$, en donde, se ordenan las aristas de forma decreciente por costos $c(e_1) \leq c(e_2) \leq \dots \leq c(e_m)$, para $k = 1$ hasta m . Sea $e_k = (u_k, v_k)$, si u_k y v_k están en diferentes subárboles de E_{k-1} , entonces $E_k := E_{k-1} \cup \{e_k\}$, en cambio, si no lo están se utiliza $E_k := E_{k-1}$.

Teorema 50. *Dado el Grafo $G = (V, E_n)$ analizado mediante el Algoritmo de Kruskal es un MST.*

Demostración. Se prueba mediante inducción que por cada k , el grafo $G_k = (V, E_k)$ es un árbol y está contenido en algún MST.

Se comienza con $k = 0$, asumiendo que la afirmación es verdadera para algún valor $k - 1$, si en la iteración k no se agrega una arista e_k entonces la afirmación es verdadera. Si en la iteración k se agrega una arista e_k , entonces G_k es un árbol, ya que u_k y v_k están en diferentes subárboles de G_{k-1} . Sean \bar{E} todas las aristas que conectan dos diferentes componentes de G_{k-1} , para todo $\bar{e} \in \bar{E}$ se tiene $c(\bar{e}) \geq c(e_k)$, ya que para cada arista $e_i = \{u_i, v_i\} \in \bar{E}$ con $c(e_i) < c(e_k)$, se sabe que u_i y v_i están en un mismo subárbol de G_i y así en el mismo subárbol de G_{k-1} , lo que implica que $e_i \notin \bar{E}$.

Sea $\tilde{E} := E_{k-1}$, entonces $\tilde{E} \cap \bar{E} = \emptyset$ por la propiedad de corte $E_k = E_{k-1} \cup \{e_k\}$ está contenida en algún Mst de G .

Dado el grafo conectado $G = (V, E_n)$, se asume que G no está conectado. Sea e la arista que conecta dos componentes de G , en donde e es considerado en alguna iteración de k , entonces $E_{k-1} \cup \{e\}$ no contiene un ciclo, pero $e \notin E_k$, por ende, existe una contradicción. \square

Cabe destacar que aunque se pueda modelar un MST como un PPE/PPL, esto no significa que el algoritmo Simplex sea el mejor algoritmo para resolverlo.

8.4. Matrices totalmente unimodulares

Dado el PPE siguiente:

$$\begin{aligned} \text{mín} \quad & c^T x \\ & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{Z}^n \end{aligned}$$

Si se resuelve mediante el algoritmo de planos cortantes, se tiene como primer paso resolver la relajación del problema lineal en donde se tiene un x^* , si $x^* \in \mathbb{Z}^n$ se debe parar.

Si se resuelve mediante Branch and bound, entonces primero se resuelve la relajación del problema lineal en donde se tiene un x^* , si $x^* \in \mathbb{Z}^n$ se debe parar.

Gracias a lo anterior, se puede verificar que $x^* \in \mathbb{Z}^n$ es el mejor caso. En los siguientes capítulos se verá cuándo sucede esto, es decir, cuándo sucede $x^* \in \mathbb{Z}^n$.

Dado un problema bipartito de matching en donde $G = (V, E)$ es un grafo no dirigido tal que $V = U \dot{\cup} W$ y así cualquier arista posee exactamente una conexión en U y otra en W , el peso w_e corresponde a cada arista $e \in E$. El objetivo es encontrar el conjunto de aristas $E' \subseteq E$ tal que cada vértice incide a lo más en una arista en E' , tal conjunto es llamado un *matching*. Cabe destacar que el objetivo es maximizar $\sum_{e \in E'} w_e$.

Formulación de programación entera lineal:

$$\begin{aligned} \text{máx} \quad & \sum_{e \in E} x_e \cdot w_e \\ & \sum_{e \in \delta(v)} x_e \leq 1 \quad \forall v \in V \\ & x_e \geq 0 \end{aligned}$$

Cabe notar que $x_e \geq 0$ reemplazó a $x_e \in \{0, 1\} \forall e \in E$.

Los PPE son difíciles de resolver, entonces qué sucede si se resuelve mediante Simplex la relajación de un problema lineal.

Teorema 51. *Cualquier solución básica factible x del PPL es entera, es decir, $x_e \in \mathbb{Z}$ para todo e .*

Algunas consecuencias del Teorema anterior son las siguientes:

- Que exista un óptimo de la relajación del problema lineal alcanzado en la solución básica factible implica que existe una solución óptima entera.
- El algoritmo Simplex encuentra una solución básica factible que es óptima, lo que implica que Simplex encuentra una solución óptima entera.
- La región factible de la relajación del problema lineal es la envoltura convexa de todas las soluciones básicas factibles, lo que implica que es igual a la envoltura convexa de todos los puntos enteros.

Demostración. Sea x solución básica factible y $F := \{e \in E \mid 0 < x_e < 1\}$ el conjunto de aristas fraccionales. Se asume que $F \neq \emptyset$, ya que de forma contraria la prueba no sería necesaria. Al mismo tiempo se asume que F posee un ciclo C . G es bipartito: C tiene longitud uniforme. Se divide C en dos C_1 y C_2 , lo que implica que cada vértice en C es incidente a exactamente una arista en C_1 y exactamente una arista en C_2 .

Se definen nuevas soluciones y, z . $y_e := \begin{cases} x_e & \text{if } e \notin C \\ x_e - \epsilon & \text{if } e \in C_1 \\ x_e + \epsilon & \text{if } e \in C_2 \end{cases}$ y $z_e := \begin{cases} z_e & \text{if } e \notin C \\ z_e + \epsilon & \text{if } e \in C_1 \\ z_e - \epsilon & \text{if } e \in C_2 \end{cases}$ Donde $\epsilon > 0$ tal que

$\epsilon \leq \min_{e \in C} x_e$, lo que implica que y, z son factibles y $x = \frac{1}{2}y + \frac{1}{2}z$ así, x no es s.b.f.

Se asume que F no tiene ciclos, lo que implica que cada componente conectado es un árbol, por ende, existen dos vértices u, v conectados por un camino P y u, v tienen sólo una arista en F incidente a ellos, es decir, misma prueba de arriba, dividiendo P en C_1, C_2 . \square

Definición 52. Sea un punto $x \in \mathbb{R}^n$ entero si $x \in \mathbb{Z}^n$. El poliedro es entero si todos los puntos extremos lo son.

Lo anterior implica que la región factible del problema lineal es entera dado el poliedro $\{x|Ax \geq b\}$. ¿Es entero? Se necesita la condición de matriz A para responder eso.

Definición 53. La matriz A es **totalmente unimodular** si cada matriz cuadrada de A posee un determinante 0, 1, o -1 .

Ejemplo:

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}, \text{ Determinantes de matrices cuadradas: } \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1, \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1, \\ \det \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} = 1, \det(1) = 1, \det(0) = 0, \det(-1) = -1, \text{ por ende, es totalmente unimodular.} \\ A' = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \end{pmatrix}, \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = -2, \text{ por ende, } A' \text{ no es totalmente unimodular.}$$

Lema 54. Sea A totalmente unimodular, entonces cada entrada de la matriz A es 0, 1, o -1 , por ende, A^T , $(A \ I)$ y $(A \ -A)$ también son totalmente unimodulares.

Teorema 55. Sea A una matriz totalmente unimodular. Entonces por cualquier entero del lado derecho b , el poliedro $P = \{x|Ax \leq b, x \geq 0\}$, $P' = \{x|Ax \geq b, x \geq 0\}$ y $P'' = \{x|Ax = b, x \geq 0\}$ son enteros.

Demostración. Sólo para P al agregar variables de holgura, siendo $Q = \{(x, s)|Ax + Is = b, x \geq 0, s \geq 0\}$, Q es entero si y sólo si P es entero.

Sea $A' = [A \ I]$ implica que A' es totalmente unimodular considerando la solución básica factible $x = (x_B, x_N)$ de Q para alguna base B . Sea $x_N = 0$, $x_B = (A'_B)^{-1}b$. Entonces por la regla de Cramer: $(x_B)_i = \frac{\det((A'_B)_i|b)}{\det(A'_B)}$ donde $(A'_B)_i|b$ es la matriz A'_B en donde la i -ésima columna es reemplazada por b , en donde cada entrada en $(A'_B)_i|b$ es entera, lo que implica que $\det((A'_B)_i|b)$ es entero. Entonces $\det(A'_B) \in \{1, -1\}$, implica que $(x_B)_i$ es entero, por ende, Q es integral al igual que P . \square

Se considera la matriz A del problema lineal, la cual se conforma por una columna A_e por cada arista $e \in E$, una fila a_v^T por cada vértice $v \in V$. $A_{v,e} = \begin{cases} 1 & \text{if } e \in \delta(v) \\ 0 & \text{otherwise} \end{cases}$

Teorema 56. La matriz A del programa lineal es totalmente unimodular.

8.4.1. El camino más corto

Se comienza con un grafo dirigido $G = (V, A)$ en donde el costo $c_{ij} > 0$ para cada arco $(i, j) \in A$, dos nodos s, t . Se debe encontrar el camino más corto desde s hasta t .

La forma de modelarlo es la siguiente:

$$\begin{aligned} & \min \sum_{(i,j) \in A} c_{ij} x_{ij} \\ & \sum_{\{j:(i,j) \in A\}} x_{ij} - \sum_{\{j:(j,i) \in A\}} x_{ji} = \begin{cases} 1 & \text{if } i = s \\ 0 & \text{if } i \in V \setminus \{s, t\} \\ -1 & \text{if } i = t \end{cases} \\ & x_{ij} \in \{0, 1\} \end{aligned}$$

Pero, ¿Cómo se escribiría esa matriz?

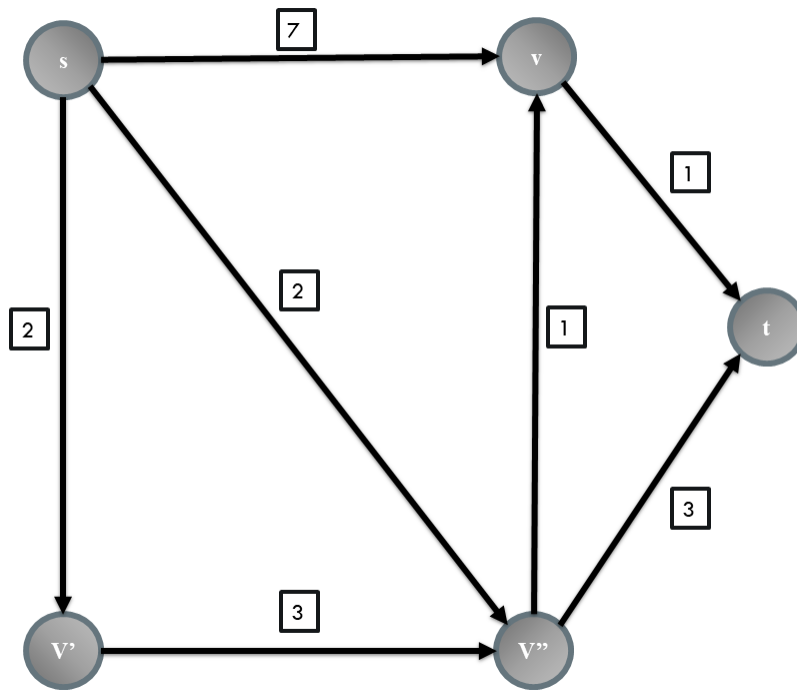


Figura 50: Grafo con nodo inicial s y nodo final t , donde cada arista tiene un costo asociado $c_{ij} > 0$. El objetivo es encontrar el camino más corto en este.

$$\begin{array}{l}
s \\
v \\
v' \\
v'' \\
t
\end{array}
\begin{array}{c}
\left(\begin{array}{cccccc}
1 & 1 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & -1 & 0 & 1 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & -1
\end{array} \right) \\
\underbrace{\hspace{10em}}_{\text{matriz de incidencia de } G}
\end{array}
\begin{array}{c}
\left(\begin{array}{c}
x_{sv} \\
x_{sv'} \\
x_{v'v''} \\
x_{sv''} \\
x_{v''v} \\
x_{v''t} \\
x_{vt}
\end{array} \right)
\end{array}
=
\begin{array}{c}
\left(\begin{array}{c}
1 \\
0 \\
0 \\
0 \\
-1
\end{array} \right)
\end{array}
\text{ La matriz de incidencia } A_G \text{ del}$$

grafo G tiene una fila por cada nodo i y una columna por cada arco (i', j') .

La entrada en la fila para el nodo i y la columna para el arco (i', j') es:

- 1 si $i = i'$, es decir, (i', j') puntos fuera de i .
- -1 si $i = j'$, es decir, (i', j') puntos en i .
- 0 de lo contrario.

Teorema 57. *La matriz de incidencia A_G de cualquier grafo dirigido G es totalmente unimodular.*

Demostración. Se considera una submatriz de M $t \times t$, en donde se quiere mostrar que $\det(M) \in \{0, 1, -1\}$. Si $t = 1$ entonces $\det(M) \in \{0, 1, -1\}$.

Por inducción, se supone que la afirmación es verdadera para todas las submatrices $(t-1) \times (t-1)$. Si M posee una columna de ceros, entonces $\det(M) = 0$. En cambio, si M tiene una columna e con exactamente una entrada diferente de cero:

$$M = \begin{pmatrix} 0 & & & \\ & 0 & & \\ * & 0 & & * \\ & 1/-1 & & \\ & 0 & & \end{pmatrix}$$

Se considera la matriz M' eliminando la fila v y la columna e desde

M . Entonces $\det(M) = \det(M')$ o $\det(M) = -\det(M')$. Por hipótesis de inducción $\det(M') \in \{0, 1, -1\}$. Si cada columna de M tiene exactamente dos entradas diferentes a cero, implica que cada columna tiene exactamente una entrada 1 y una -1 . Al sumar todas las filas da un vector de ceros, lo que implica que las filas de M son linealmente independientes, es decir, $\det(M) = 0$. □

Teorema 58. *El poliedro que corresponde a los problemas de camino mínimo es entero.*

El problema de costo mínimo de flujo comienza con un grafo dirigido G , en particular la oferta/demanda: $b: V \rightarrow \mathbb{R}$. Cabe destacar que:

- Si $b(i) > 0$ es un nodo de oferta.
- Si $b(i) < 0$ es un nodo de demanda.
- Si $b(i) = 0$ es un nodo de paso.

Los costos correspondientes son: $c: A \rightarrow \mathbb{R}$ y la capacidad máxima es $u: A \rightarrow \mathbb{R}_0^+ \cup \{\infty\}$.

$$\begin{aligned}
& \text{mín} \sum_{(i,j) \in E} c_{ij} x_{ij} \\
& \sum_{\{j:(i,j) \in E\}} x_{ij} - \sum_{\{j:(j,i) \in E\}} x_{ij} = b(i) \quad \forall i \in V \\
& x_{ij} \leq u_{ij} \quad \forall (i,j) \in E \\
& x_{ij} \geq 0 \quad \forall (i,j) \in E
\end{aligned}$$

Es importante notar que la matriz (A_G^I) es totalmente unimodular.

Teorema 59. *El poliedro correspondiente a la relajación del problema lineal del flujo mínimo es entero si todas las capacidades de las aristas son enteras y si todos los nodos de oferta/demanda son enteros.*

Como también, para el otro caso de la función objetivo, es decir:

Teorema 60. *El poliedro correspondiente a la relajación del problema lineal del flujo máximo es entero si todas las capacidades de las aristas son enteras.*

9. Optimización no lineal

Hasta ahora se han conocido las funciones y restricciones lineales, pero ahora se podrán ver los problemas no lineales de la forma:

$$\begin{aligned} & \text{mín } f(x) \\ & g_i(x) \leq 0 \forall i \in I \\ & x \in \mathbb{R}^n \end{aligned}$$

Con $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ (en donde están definidas para cada punto en \mathbb{R}^n .) Cabe destacar que la programación lineal es un caso específico, ya que $f(x) = c^T x$, $g_i(x) = a_i^T x - b$. Sea $\Omega := \{x \in \mathbb{R}^n | g_i(x) \leq 0 \text{ for each } i \in I\}$. Un punto x es **factible** si $x \in \Omega$.

Un ejemplo sería encontrar el punto $x = (x_1, x_2)$ más cercano al punto $(3, 4)$ tal ue la distancia entre x y $(0, 2)$ es a lo más 2 y $x \in \mathbb{R} \times (-\infty, 2]$.

$$\text{mín } f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 4)^2 \quad (8)$$

$$\text{s.t. } x_2 \leq 2 \quad (9)$$

$$x_1^2 + (x_2 - 2)^2 \leq 4 \quad (10)$$

Se reescribe como:

$$\text{mín } f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 4)^2 \quad (11)$$

$$\text{s.t. } g_1(x_1, x_2) = x_2 - 2 \leq 0 \quad (12)$$

$$g_2(x_1, x_2) = x_1^2 + (x_2 - 2)^2 - 4 \leq 0 \quad (13)$$

Definición 61. Sea un punto $x \in \mathbb{R}^n$ el óptimo global si $x \in \Omega$ y $f(x) \leq f(x')$ por cada $x' \in \Omega$.

Definición 62. Sea el punto $x \in \mathbb{R}^n$ es un óptimo local si $x \in \Omega$ y existe un conjunto abierto $\mathcal{N} \subseteq \mathbb{R}^n$ con $x \in \mathcal{N}$ (un **vecino** de x) tal que $f(x) \leq f(x')$ por cada $x' \in \mathcal{N} \cap \Omega$.

Pero ¿La solución global siempre existe? Si S es cerrado y acotado, sí.

Teorema 63 (Weierstrass). *El teorema de Weierstrass señala que sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continua y S un conjunto cerrado y acotado, entonces el problema $\text{mín}_{x \in S} f(x)$ admite un óptimo global, es decir, existe un valor $x^* \in S$ con la restricción de que $f(x^*) \leq f(x)$ para todos los $x \in S$.*

Por ejemplo, si f es continua, su región factible es cerrada y acotada, entonces por el teorema anterior implica que existe un mínimo.

Este tipo de problemas en general pueden ser muy complejos dependiendo de f y de g_i .

Definición 64. Una función $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa si $\forall x_1, x_2 \in C, \lambda \in [0, 1]$, se tiene $f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$. Cabe destacar que la función es cóncava si $-f$ es convexa.

Algunos ejemplos de lo anterior son $f(x) = x^2$, $f(x) = e^x$ y $f(x) = 2x$. Se asume que f y todos los $g_i, i \in I$ son convexos.

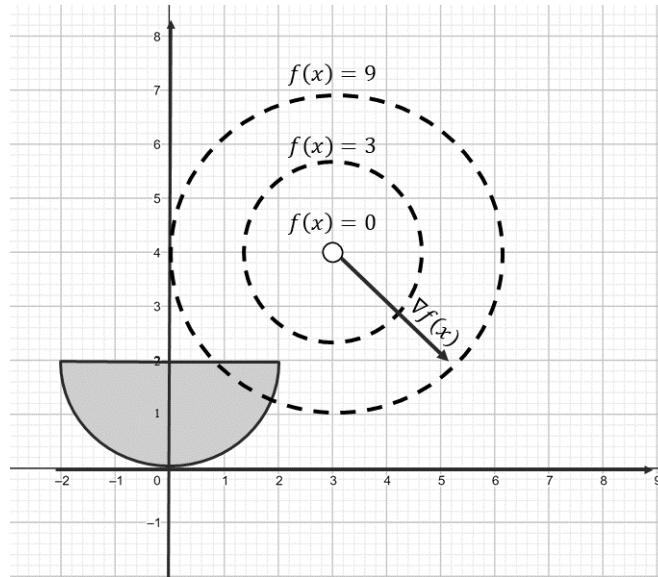


Figura 51: La región factible representada en amarillo, y las líneas rojas punteadas representan las curvas de nivel de la función objetivo, la intuición apunta a hacer lo más pequeño posible el círculo dentro del semi-círculo amarillo.

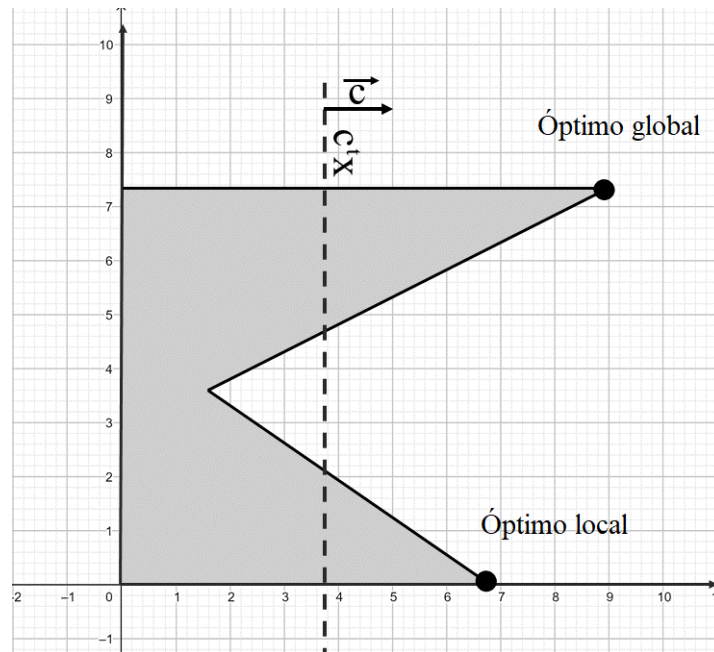


Figura 52: Se puede elegir un subconjunto del conjunto original y buscar una solución óptima dentro de este, lo cual se conoce como óptimo local. Si este punto también es óptimo para el conjunto original, se conoce como óptimo global.

Lema 65. Sea $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa y $K \in \mathbb{R}$. Entonces el conjunto $S_i = \{x \in \mathbb{R}^n : g_i(x) \leq K\}$ es convexo.

Demostración. Sea $x_1, x_2 \in S_i$ y $\lambda \in [0, 1]$, entonces:

$$\begin{aligned} g_i(\lambda x_1 + (1 - \lambda)x_2) &\leq \lambda g_i(x_1) + (1 - \lambda)g_i(x_2) \rightarrow \text{Convexidad de } g_i \\ &\leq \lambda K + (1 - \lambda)K \rightarrow x_1, x_2 \in S_i \\ &= K \end{aligned}$$

□

Entonces cada retricción por una función convexa g_i define un conjunto convexo. Así una familia de conjuntos convexos $C_i \subseteq \mathbb{R}^n$, $i \in I$, entonces $\bigcap_{i \in I} C_i$ es convexo. Todos los g_i son convexos, lo que implica que Ω es convexo.

¿Cuál es el beneficio de que Ω y f sean convexos?

Teorema 66. Se asume que Ω y f son convexos, entonces cada óptimo local es también un óptimo global.

Demostración. Se supone que x es un óptimo local y $y \neq x$ es óptimo global con $f(y) < f(x)$. Se define $z(\theta) := \theta y + (1 - \theta)x$ por cada $\theta \in [0, 1]$, Ω es convexo, lo que implica que $z(\theta) \in \Omega$ para todos los $\theta \in [0, 1]$, entonces $f(z(\theta)) = f(\theta y + (1 - \theta)x) \leq \theta f(y) + (1 - \theta)f(x) < \theta f(x) + (1 - \theta)f(x) = f(x)$ por cada $\theta > 0$ el punto $x' = z(\theta)$ satisface que $f(x') < f(x)$, por ende, existe un punto x' arbitrariamente cercano a x . Tal que $f(x') < f(x)$, lo cual contradice que x es óptimo local. □

9.1. Dualidad para problemas no lineales

Se considera un problema de optimización no lineal como sigue:

$$\begin{aligned} &\text{mín } f(x) \\ &g_i(x) \leq 0 \quad \forall i \in I \\ &h_i(x) = 0 \quad \forall i \in E \\ &x \in \mathbb{R}^n \end{aligned}$$

Sea x^* la solución óptima y $p^* := f(x^*)$ sea el valor de la solución óptima del primal. Entonces se fija algún $\lambda_i \geq 0 \quad \forall i \in I$ y algún $\nu_i \in \mathbb{R} \quad \forall i \in E$ que sean multiplicadores lagrangianos. En donde λ_i es un multiplicador lagrangiano de restricción $f_i(x) \leq 0$ al igual que ν_i de restricción $h_i(x) = 0$.

$$\underbrace{\inf_{x \in \mathbb{R}^n} f(x) + \sum_{i \in I} \lambda_i g_i(x) + \sum_{i \in E} \nu_i h_i(x)}_{=: L(x, \lambda, \nu)}$$

Luego se define,

$$d(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$$

Se puede tomar el valor $-\infty$ si $L(x, \lambda, \nu)$ no es acotado por abajo, se tiene que $d(\lambda, \nu) \leq p^*$.

$$d(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu) \leq L(x^*, \lambda, \nu) = f(x^*) + \sum_{i=1}^m \lambda_i g_i(x^*) + \sum_{i=1}^p \nu_i h_i(x^*) \leq f(x^*) = p^*$$

Para cada (λ, ν) con $\lambda \geq 0$ se da la cota inferior $d(\lambda, \nu)$ para p^* , se dice que (λ, ν) es *dual factible* si $\lambda \geq 0$ y $d(\lambda, \nu) > -\infty$.

$$\begin{aligned} &\text{máx } d(\lambda, \nu) \\ &\lambda \geq 0 \\ &\nu \text{ libre} \end{aligned}$$

Sea (λ^*, ν^*) como la solución óptima, $d^* := d(\lambda^*, \nu^*)$ se tiene $d^* \leq p^*$ (Dualidad débil). La diferencia es $p^* - d^*$, es decir, *el gap de dualidad*.

En los problemas lineales, si el primal y dual son factibles entonces implica que siempre $p^* = d^*$ por dualidad fuerte, pero eso no es cierto para los problemas no lineales en general.

$$\begin{aligned} & \text{mín } f(x) \\ & g_i(x) \leq 0 \forall i \in I \\ & h_j(x) = 0 \forall j \in E \\ & x \in \mathbb{R}^n \end{aligned}$$

Sea x^* la solución óptima primal, $p^* := f(x^*)$, se sabe que la idea del dual es tratar de encontrar una cota inferior para p^* , por ende, se define

$$d(\lambda, \nu) = \underbrace{\inf_{x \in \mathbb{R}^n} f(x) + \sum_{i \in I} \lambda_i g_i(x) + \sum_{j \in E} \nu_j h_j(x)}_{=: L(x, \lambda, \nu)}$$

para $\lambda \geq 0$, luego $d(\lambda, \nu) \leq p^*$ por cualquier $\lambda \geq 0$ y $\nu \in \mathbb{R}$. Luego el lagrangiano dual es:

$$\begin{aligned} & \text{máx } d(\lambda, \nu) \\ & \lambda \geq 0 \\ & \nu \text{ libre} \end{aligned}$$

9.2. Dualidad fuerte implica...

Se asume que $d^* = p^*$ (*dualidad fuerte*) con soluciones óptimas x^* y (λ^*, ν^*) .

$$\begin{aligned} & f(x^*) = d(\lambda^*, \nu^*) \\ & = \inf_{x \in \mathbb{R}^n} L(x, \lambda^*, \nu^*) \\ & \leq L(x^*, \lambda^*, \nu^*) \\ & = f(x^*) + \sum_{i \in I} \lambda_i^* g_i(x^*) + \sum_{j \in E} \nu_j^* h_j(x^*) \\ & \leq f(x^*) \end{aligned}$$

Entonces $h_j(x^*) = 0$ para todos j , lo que implica que $\sum_{i=1}^m \lambda_i^* g_i(x^*) = 0$, cada término ≤ 0 , entonces $\lambda_i^* g_i(x^*) = 0$ para todos los i , así $\lambda_i^* > 0$, es decir, $g_i(x^*) = 0$ y $g_i(x^*) < 0$ entonces $\lambda_i^* = 0$ y x^* minimiza $L(x, \lambda^*, \nu^*)$ sobre x , es decir, $\nabla L(x^*, \lambda^*, \nu^*) = 0$ equivalente a $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p \nu_j^* \nabla h_j(x^*) = 0$.

10. Condiciones de Karush-Kuhn-Tucker (KKT)

Se considera el problema siguiente:

$$\begin{aligned} \text{(P)} \quad & \text{mín } f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0 && i \in I \\ & h_j(x) = 0 && j \in E \\ & x \in \mathbb{R}^n \end{aligned}$$

En donde el punto x satisface las condiciones de KKT si hay escalares $\{\lambda_i\}_{i \in I}$, $\{\nu_j\}_{j \in E}$ sujetos a :

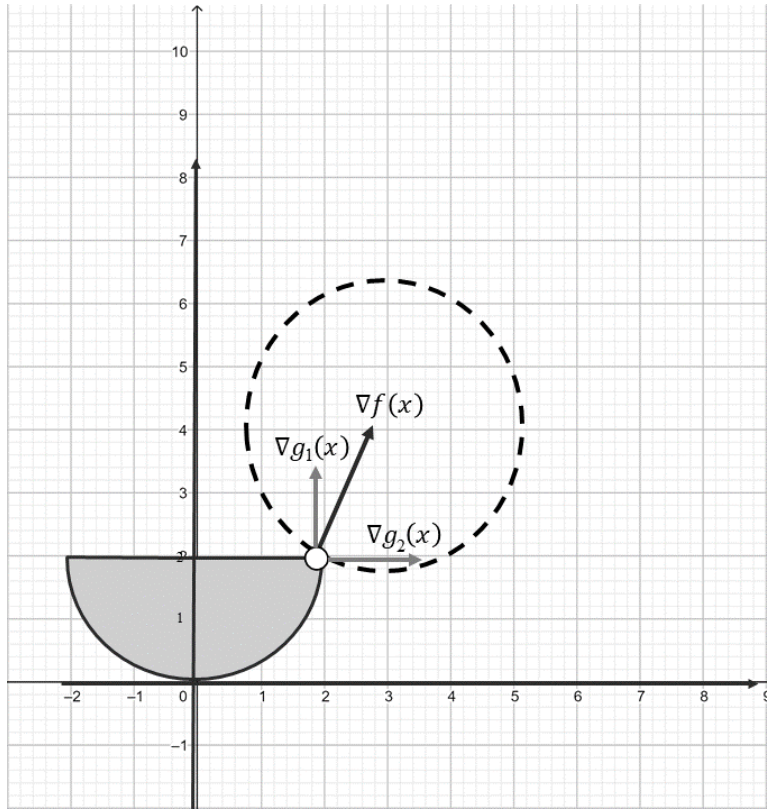


Figura 53: Se puede apreciar vectorialmente la condición de primer orden. Esta también se conoce como gradientes alineados.

$$-\nabla f(x) = \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) \quad \rightarrow \text{Cond. de Primer Orden} \quad (14)$$

$$g_i(x) \leq 0 \quad \forall i \in I \quad \rightarrow \text{Factibilidad del Primal} \quad (15)$$

$$h_j(x) = 0 \quad \forall j \in E$$

$$\lambda_i \geq 0 \quad \forall i \in I \quad \rightarrow \text{Factibilidad Dual}$$

$$\nu_j \in \mathbb{R} \quad \forall j \in E \quad (16)$$

$$\lambda_i g_i(x) = 0 \quad \forall i \in I \quad \rightarrow \text{Holgura Complementaria}$$

Si el punto x es óptimo y la Dualidad fuerte se cumple, entonces las condiciones de KKT son satisfechas.

Ejemplo:

$$\text{mín} \quad f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 4)^2 \quad (17)$$

$$\text{s.t.} \quad g_1(x_1, x_2) = x_2 - 2 \leq 0 \quad (18)$$

$$g_2(x_1, x_2) = x_1^2 + (x_2 - 2)^2 - 4 \leq 0 \quad (19)$$

El punto óptimo es: $x^* = (2, 2)$, en donde $\nabla f(x_1, x_2) = (2(x_1 - 3), 2(x_2 - 4))$, $-\nabla f(x^*) = (2, 4)$, luego $\nabla g_1(x_1, x_2) = (0, 1) = \nabla g_1(x^*)$, entonces $\nabla g_2(x_1, x_2) = (2x_1, 2(x_2 - 2))$, $\nabla g_2(x^*) = (4, 0)$.

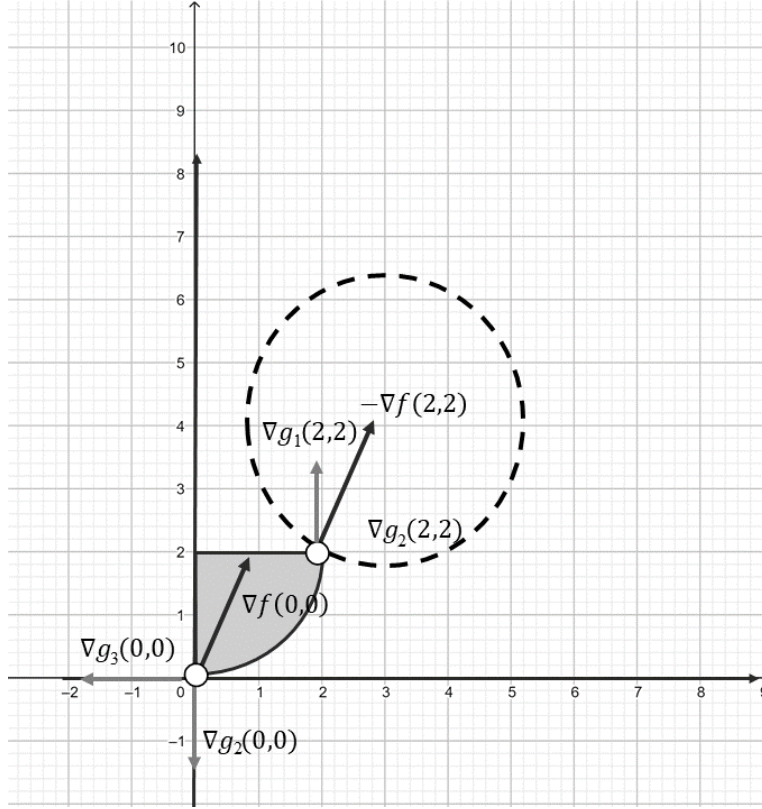


Figura 54: Ejemplo al agregar la restricción g_3 , se puede apreciar que la restricción no es activa en el punto $(2, 2)$.

Se puede expresar $(2, 4) = -\nabla f(x^*) = \lambda_1 \cdot \nabla g_1(x^*) + \lambda_2 \cdot \nabla g_2(x^*) = \lambda_1(0, 1) + \lambda_2(4, 0)$ con $\lambda_1 = 4$ y $\lambda_2 = 0,5$, así $g_1(x^*) = 0 = g_2(x^*)$, es decir, las restricciones son activas, $\lambda_1 > 0$ y $\lambda_2 > 0$, lo que implica que x^* satisface las condiciones de KKT. Ahora, si se agrega una restricción:

$$g_3(x_1, x_2) : -x_1 \leq 0 \quad (20)$$

Luego, x^* sigue siendo óptimo y si se tiene $\lambda_3 := 0$, entonces x^* se siguen cumpliendo las condiciones de KKT:

$$\begin{aligned}
 -\nabla f(x^*) &= \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) + \lambda_3 \nabla g_3(x^*) \\
 g_1(x^*) &= 0 \quad \lambda_1 > 0 < - - \text{ activa en } x^* \\
 g_2(x^*) &= 0 \quad \lambda_2 > 0 < - - \text{ activa en } x^* \\
 g_3(x^*) &< 0 \quad \lambda_3 = 0 < - - \text{ inactiva en } x^*
 \end{aligned}$$

La intuición es que se necesita que $-\nabla f(x^*)$ sea una combinación cónica de $\nabla g_i(x^*)$ con $g_i(x^*) = 0$. Entonces, sea x una combinación cónica de puntos y_1, \dots, y_k si $\exists \lambda_1, \dots, \lambda_k \geq 0$ sujeto a $x = \sum_i \lambda_i y_i$ para puntos que no sean óptimos no se pueden encontrar para este programa tal que $\lambda_1, \lambda_2, \lambda_3 \geq 0$. Por ejemplo: $x' = (0, 0)$: g_2 y g_3 son activos, pero $-\nabla f(x') = (6, 8)$, $\nabla g_2(x') = (0, -4)$, $\nabla g_3(x') = (-1, 0)$.

Ejemplo :

Teniendo un programa lineal como el siguiente:

$$\begin{aligned} & \text{mín } c^T x \\ \text{s.a. } & a_i^T x = b_i & i \in \{1, \dots, m\} \\ & x_j \geq 0 & j \in \{1, \dots, n\} \end{aligned}$$

Se reescribe como:

$$\begin{aligned} & \text{mín } f(x) = c^T x & (21) \\ \text{s.a. } & h_j(x) = a_j^T x - b_j = 0 & j \in \{1, \dots, m\} \quad (\nu_j) & (22) \\ \text{s.a. } & g_i(x) = -x_i \leq 0 & i \in \{1, \dots, n\} \quad (\lambda_i) & (23) \end{aligned}$$

Entonces $\nabla f(x) = c$, $\nabla h_j(x) = a_j$ y $\nabla g_i(x) = -e_i$. Entonces la primera condición sería:

$$-c = -\sum_{i=1}^n \lambda_i \cdot e_i + \sum_{j=1}^m \nu_j \cdot a_j$$

por cada $i = 1, \dots, n$

$$\lambda_i = \sum_{j=1}^m \nu_j \cdot a_{ji} + c_i$$

Se mantienen las condiciones.

$$\lambda_i = \sum_{j=1}^m \nu_j \cdot a_{ji} + c_i \quad j \in \{1, \dots, n\} \quad (24)$$

$$a_i^T x - b_i = 0 \quad i \in \{1, \dots, m\} \quad \text{--- } x \text{ primal factible} \quad (25)$$

$$x_j \geq 0 \quad j \in \{1, \dots, n\} \quad \text{--- } x \text{ primal factible} \quad (26)$$

$$\lambda_i \geq 0 \quad i \in \{1, \dots, n\} \quad \text{--- } \nu \text{ dual factibe} \quad (27)$$

$$\nu_j \in \mathbb{R} \quad j \in \{1, \dots, m\} \quad \text{--- } \nu \text{ dual factible} \quad (28)$$

$$x_i \lambda_i = 0 \quad i \in \{1, \dots, n\} \quad \text{--- holgura complementaria} \quad (29)$$

Pero, ¿Qué implica lo anterior?

Lema 67. *x* satisface las condiciones de KKT, lo que implica que es factible para el primal y que existe una solución factible para el dual ν que satisface la holgura complementaria con x , es decir, x y ν son óptimos, lo que implica que x satisface las condiciones de KKT con (λ, ν) , entonces el problema dual es como sigue:

$$\text{máx } \nu^T b \quad \sum_{j=1}^m a_{ji} \nu_j \leq c_i \quad \forall i = 1, \dots, n \quad \nu \text{ libre}$$

$0 \leq \lambda_i = \sum_{j=1}^m \nu_j \cdot a_{ji} + c_i$ lo que implica que $\sum_{j=1}^m a_{ji}(-\nu_j) \leq c_i$, entonces $-\nu$ es solución factible para el dual con las condiciones de holgura complementaria.

$$\begin{aligned} -\nu_j(a_j^T x - b_j) &= 0 \quad \forall j \\ (c_i - (-\nu)^T A_i)x_i &= 0 \quad \forall i \end{aligned}$$

Entonces, la primera condición se satisface, luego la segunda condición se satisface si $x_i = 0$, de otra forma $\lambda_i = 0$ y así $\sum_{j=1}^m a_{ji}(-\nu_j) = c_i$, por ende, x y $-\nu$ son factibles para primal y dual, satisfaciendo la holgura complementaria.

Se considera el problema siguiente:

$$\begin{array}{ll}
 \text{(P)} & \text{mín } f(x) \\
 \text{s.t.} & g_i(x) \leq 0 \quad i \in I \\
 & h_j(x) = 0 \quad j \in E \\
 & x \in \mathbb{R}^n
 \end{array}$$

Entonces el punto x satisface las condiciones de KKT si son escalares $\{\lambda_i\}_{i \in I}$, $\{\nu_j\}_{j \in E}$ sujeto a

$$-\nabla f(x) = \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) \quad \rightarrow \text{Cond. de Primer Orden} \quad (30)$$

$$\begin{array}{ll}
 g_i(x) \leq 0 & \forall i \in I \quad \rightarrow \text{Factibilidad primal} \\
 h_j(x) = 0 & \forall j \in E \quad (31)
 \end{array}$$

$$\lambda_i \geq 0 \quad \forall i \in I \quad \rightarrow \text{Factibilidad dual}$$

$$\nu_j \in \mathbb{R} \quad \forall j \in E \quad (32)$$

$$\lambda_i g_i(x) = 0 \quad \forall i \in I \quad \rightarrow \text{Holgura Complementaria}$$

Entonces para problemas lineales, sería bueno encontrar algo que nos garantice la relación entre KKT y optimalidad como:

- **Si x satisface KKT** + alguna condición que se cumpla, implicaría que x es óptimo.
- **Si x es óptimo** + alguna condición que se cumple, implicaría **que x satisface las condiciones de KKT**.

Teorema 68. *Sea x un punto que satisface las condiciones de KKT, f sea convexa, cada $g_i, i \in I$ sea convexa, cada h_j sea afín, es decir cóncava y convexa, entonces x es óptimo global.*

Pero, ¿La convexidad es necesaria para el teorema anterior? Sí. **Ejemplo:**

$$\begin{array}{ll}
 \text{mín} & f(x_1, x_2) = 3x_1 - x_2 \\
 \text{s.a.} & g_1(x_1, x_2) = 4 - x_1^2 - (x_2 - 2)^2 \leq 0 \\
 & g_2(x_1, x_2) = x_2 - 2 \leq 0 \\
 & g_3(x_1, x_2) = x_1 + x_2 - 5 \leq 0 \\
 & g_4(x_1, x_2) = -x_1 \leq 0 \\
 & g_5(x_1, x_2) = -x_2 \leq 0
 \end{array}$$

El punto $(2, 2)$ satisface las condiciones de KKT, pero no es óptimo. El óptimo es $(0, 0)$ y satisface las condiciones de KKT, lo que implica que si un punto satisface las condiciones de KKT, no es necesariamente óptimo. Cabe destacar que la región factible no es convexa. Entonces, ¿Se puede decir que si x es óptima entonces x satisface KKT? No.

Ejemplo:

$$\begin{array}{ll}
 \text{mín} & f(x_1, x_2) = -x_1 \\
 \text{s.a.} & g_1(x_1, x_2) = x_2 - (1 - x_1)^3 \leq 0 \\
 & g_2(x_1, x_2) = -x_2 \leq 0
 \end{array}$$

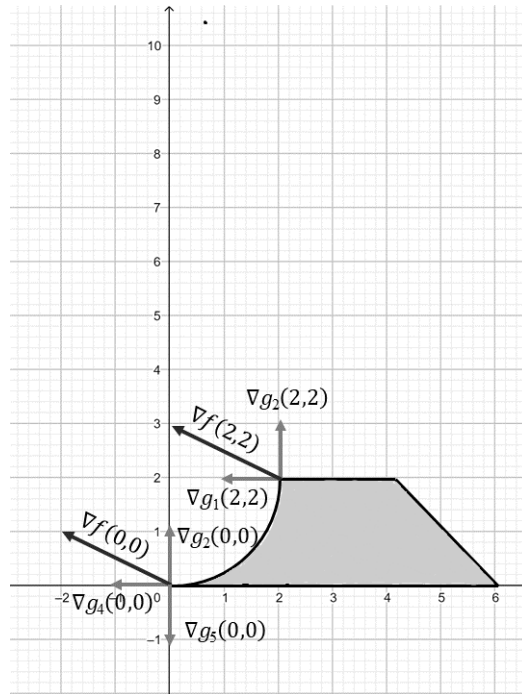
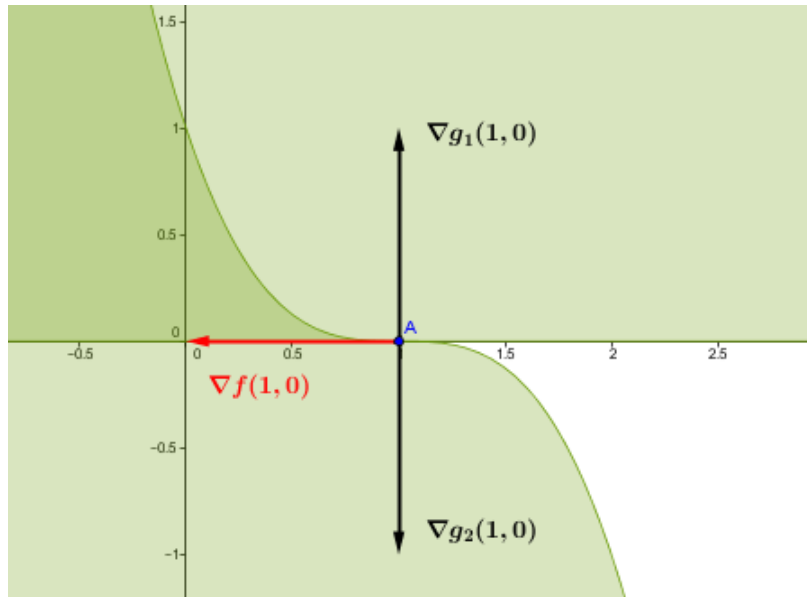


Figura 55: ¿Por qué no se cumplen las condiciones de KKT en este ejemplo? Por la convexidad.



- Si x es óptimo + alguna condición que se cumpla, implica que x **satisface las condiciones KKT**.
- ¿Si x es óptimo + región factible f es convexa, implica que satisface KKT? No.

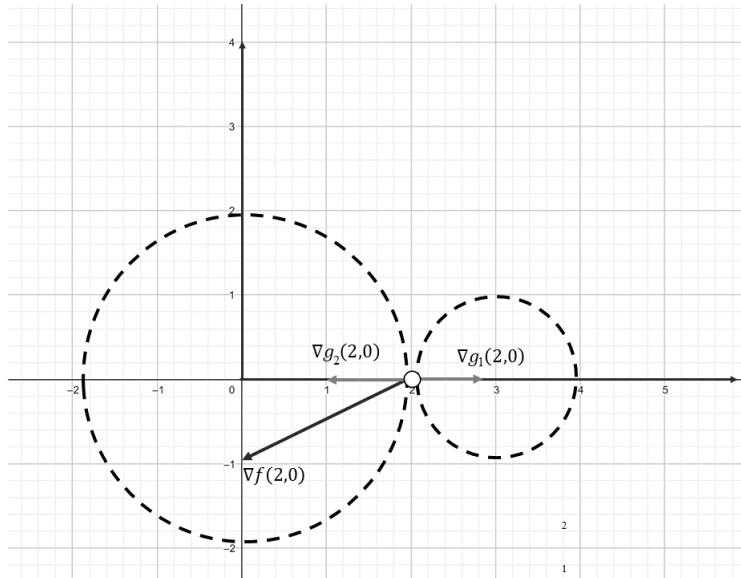


Figura 56: Aunque no se cumplan las condiciones de KKT, sigue siendo óptimo, ¿Qué se necesita para que un punto óptimo cumpla KKT?

Ejemplo:

$$\begin{aligned} \text{mín} \quad & f(x_1, x_2) = 2x_1 + x_2 \\ \text{s.a.} \quad & g_1(x_1, x_2) = x_1^2 + x_2^2 - 4 \leq 0 \\ & g_2(x_1, x_2) = (x_1 - 3)^2 + x_2^2 - 1 \leq 0 \end{aligned}$$

El punto $(2, 0)$ es óptimo, pero no satisface condiciones de KKT. Cabe destacar que la región factible no se comporta de buena forma, ya que, por ejemplo, posee sólo un punto convexo. cuando una región factible se comporta de buena forma, es aquí donde se verifica la condición de Slater, la cual asegura que existe un punto x tal que: $g_i(x) < 0$ por cada $i \in I$ $h_j(x) = 0$ por cada $j \in E$

Teorema 69. *Se asume que f es convexa, cada g_i e $i \in I$ son convexas, cada h_j y $j \in E$ son afín y se asume que la condición de Slater se cumple. Sea x^* una solución óptima, entonces x^* satisface las condiciones de KKT. Es importante notar que h_j es afín si y sólo si $h_j(x) = a_j^T x + b$ para algún $a_j \in \mathbb{R}^n$ y algún $b \in \mathbb{R}$.*

Definición 70. Una calificación de restricción es una condición que garantice que f, g_i sean convexas, h_j afín, x^* óptimo, lo que implica que x^* satisface las condiciones de KKT.

Hay muchas calificaciones de restricciones conocidas, pero algunas son más fuertes que otras.

Definición 71. Dado un punto x . El conjunto de activo $\mathcal{A}(x)$ es definido como $\mathcal{A}(x) = E \cup \{i \in I | g_i(x) = 0\}$.

Definición 72. Dado un punto x^* . La calificación de restricción de independencia lineal (CRIL) afirma que x^* y el set de gradientes de restricciones activas $\{\nabla g_i(x^*) | i \in \mathcal{A}(x^*) \cap I\} \cup \{\nabla h_j(x^*) | j \in \mathcal{A}(x^*) \cap E\}$ es linealmente independiente.

Teorema 73. *Se asume que f es convexa, cada g_i e $i \in I$ son convexas y cada h_j y $j \in E$ son afín. Entonces x^* sea una solución óptima tal que CRIL se cumple, entonces x^* satisface las condiciones de KKT.*

A modo de resumen del capítulo, se tiene que:

- Si f, g_i son convexas, h_i es afín y x satisface las condiciones KKT x es óptimo.
- Si f, g_i son convexas, h_i es afín, x es óptima y algunas restricciones se cumplen, implica que x satisface las condiciones de KKT.

11. Descenso más empuinado

En este capítulo se verificará la minimización de una función no lineal, pero sin restricciones para comenzar como el siguiente:

$$\begin{aligned} \text{mín} \quad & f(x) \\ & x \in \mathbb{R}^n \end{aligned}$$

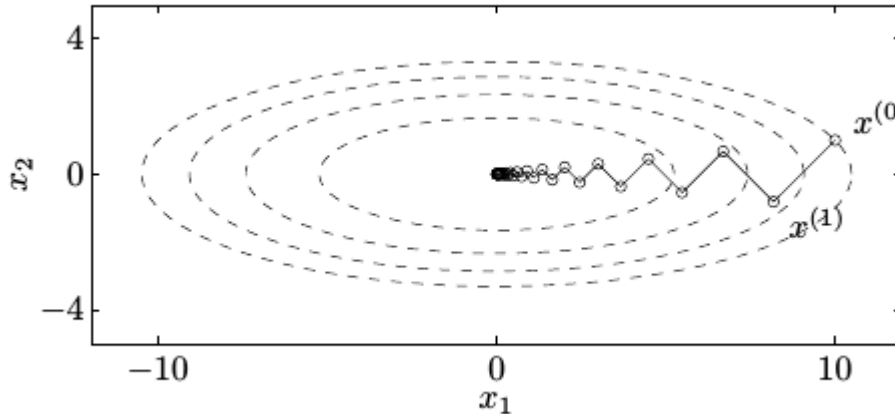
Se asume que f es diferenciable. Si x^* es óptima, entonces $\nabla f(x^*) = 0$. Se puede encontrar un x tal que $\nabla f(x) = 0$ resolviendo el sistema de ecuaciones.

Si f es muy complejo, se necesita otro algoritmo tal como comenzar en un punto $x^{(0)}$ para luego ir al punto $x^{(1)}$ con $f(x^{(1)}) < f(x^{(0)})$, luego a $x^{(2)}$ con $f(x^{(2)}) < f(x^{(1)})$ y así sucesivamente, obteniendo la secuencia $x^{(0)}, x^{(1)}, \dots$ con $\lim_{k \rightarrow \infty} f(x^{(k)}) = f(x^*)$, donde x^* es la solución óptima. Cabe destacar que el algoritmo debe para cuando $f(x^{(k)}) - f(x^*) < \epsilon$ para algún $\epsilon > 0$ específico.

Ejemplo:

$$\text{mín} \quad f(x) = \frac{1}{2}(x_1^2 + 10x_2^2)$$

En donde el óptimo: $x^* = 0$ y $f(x^*) = 0$.



Entonces $x^{(0)} = (10, 1)$, $x^{(1)} = (\frac{90}{11}, -\frac{9}{11})$, $x^{(k)} = (10(\frac{9}{11})^k, (-\frac{9}{11})^k)$, $f(x^{(k)}) = \frac{1}{2} \left(\left(10\left(\frac{9}{11}\right)^k\right)^2 + 10\left(\left(-\frac{9}{11}\right)^k\right)^2 \right) = \frac{1}{2} \left(100\left(\frac{9}{11}\right)^{2k} + 10 \cdot \left(-\frac{9}{11}\right)^{2k} \right) = 55\left(\frac{9}{11}\right)^{2k} = \left(\frac{9}{11}\right)^{2k} \cdot f(x^{(0)})$, $f(x^{(k+1)}) = \left(\frac{9}{11}\right)^2 f(x^{(k)}) < f(x^{(k)})$ con $\lim_{k \rightarrow \infty} x^{(k)} = 0$ y así $\lim_{k \rightarrow \infty} f(x^{(k)}) = 0$.

Entonces el algoritmo comienza en el punto $x \in \mathbb{R}^n$ para luego repetir los siguientes pasos:

- Determinar la dirección Δx .
- Escoger el tamaño del paso t .
- Actualizar $x := x + t\Delta x$.

Hasta que el criterio de detención sea satisfecho. Pero ¿Qué dirección Δx es buena?

Dirección: El objetivo debe ser decrecer en la dirección Δx .

El gradiente $\nabla f(x)$ apunta en la dirección del mejor ratio decreciente/creciente de f . Se elige $\Delta x := -\nabla f(x)$, lo que implica que el gradiente es descendiente.

Ejemplo:

$x_1^{(k)} = 10 \left(\frac{9}{11}\right)^k$, $x_2^{(k)} = \left(-\frac{9}{11}\right)^k$, entonces $f(x) = \frac{1}{2}(x_1^2 + 10x_2^2)$ con $-\nabla f(x) = -(x_1, 10x_2)$, luego $-\nabla f(x^{(k)}) = \left(-10 \left(\frac{9}{11}\right)^k, -10 \left(-\frac{9}{11}\right)^k\right)$, con $t := 2/11$ y $(x_1^{(k)}, x_2^{(k)}) + t \left(-10 \left(\frac{9}{11}\right)^k, -10 \left(-\frac{9}{11}\right)^k\right) = (x_1^{(k+1)}, x_2^{(k+1)})$.

Largo de paso t : Se elige t para maximizar el decrecimiento del objetivo: $t := \arg \min_{s \geq 0} \phi(s) := f(x + s\Delta x)$.

Ejemplo:

$x^{(k)} + t \cdot \Delta x = \left(\left(\frac{9}{11}\right)^k 10(1-t), \left(-\frac{9}{11}\right)^k (1-10t)\right)$ con $f(x^{(k)} + t\Delta x) = \frac{1}{2} \left(\left(\frac{9}{11}\right)^{2k} 100(1-t)^2 + 10 \left(-\frac{9}{11}\right)^{2k} (1-10t)^2\right) = \frac{1}{2} \left(\frac{9}{11}\right)^{2k} (100(1-t)^2 + 10(1-10t)^2)$, minimizado por $t = 2/11$.

Criterio de detención:

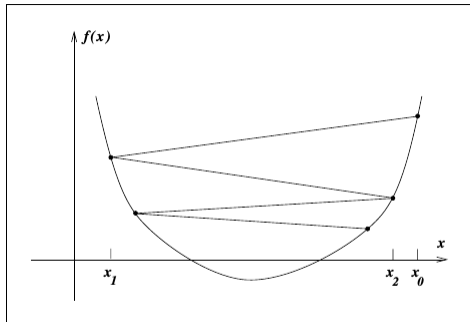
La intuición se centra en que $\nabla f(x^*) = 0$ para el punto óptimo x^* y $\|\nabla f(x^*)\|_2 = 0$. Entonces se debe detener si $\|\nabla f(x)\|_2 \leq \eta$ para algún valor pequeño $\eta > 0$.

Teorema 74. Se supone que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es continuamente diferenciable en un conjunto abierto \mathcal{N} conteniendo el conjunto $S = \{x \in \mathbb{R}^n | f(x) \leq f(x^0)\}$ donde x^0 es el punto de comienzo. Sea el gradiente ∇f Lipschitz continuo en \mathcal{N} , es decir, hay una constante $L > 0$ tal que

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L \|x - \tilde{x}\| \quad \forall x, \tilde{x} \in \mathcal{N}$$

Entonces $\lim_{k \rightarrow \infty} \|\nabla f(x^{(k)})\| = 0$.

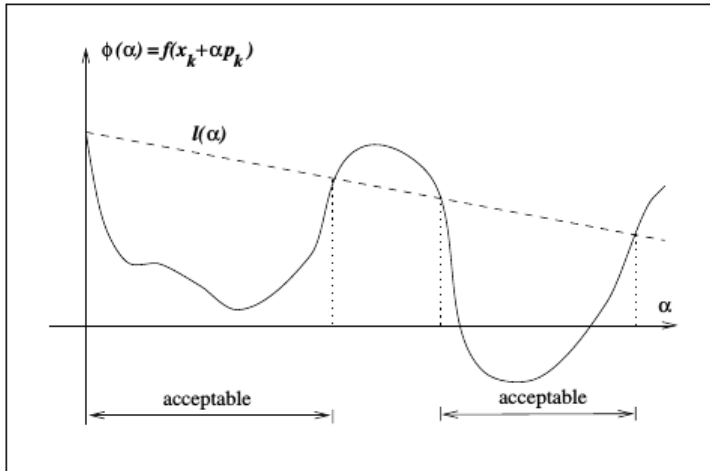
Entonces en el límite los algoritmos encuentran el punto x con $\nabla f(x) = 0$, el cual es el óptimo local si f y Ω son convexos, pero no óptimo general. El tamaño del paso $t := \arg \min_{s \geq 0} \phi(s) := f(x - s\nabla f(x))$ contiene al minimizador exacto t de $\phi(s)$ es difícil de encontrar, en particular si f no es convexo. Como primer intento, se requiere que $f(x - s\nabla f(x)) < f(x)$.



En donde el mínimo global es -1 , pero $\lim_{k \rightarrow \infty} f(x^{(k)}) = 0$. Cabe destacar que se necesita un decrecimiento suficiente, lo cual requiere que

$$\phi(s) \leq \phi(0) + c_1 s \phi'(s) \tag{33}$$

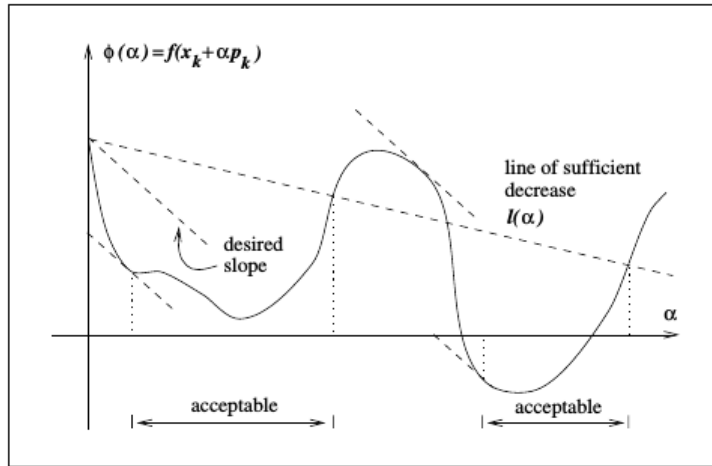
Para algún $c_1 \in (0, 1)$, típicamente $c_1 = 10^{-4}$.



Pero aún no es la mejor forma de hacerlo. Entonces se toma un largo de paso de s , el cual puede ser muy pequeño. Por ende, la intuición es $\phi'(s) = 0$ si s es minimizador local de ϕ , entonces se requiere que:

$$\phi'(s) \geq c_2 \phi'(0) \tag{34}$$

Para algún $c_2 \in (c_1, 1)$, típicamente $c_2 = 0,1$. Por ende, la intuición es que si $\phi'(0)$ es muy pequeño (negativo), entonces se debería seguir buscando en la misma dirección.



Las condiciones 33 y 34 son llamadas *condiciones de Wolfe*.

Lema 75. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuamente diferenciable. Se asume que f es acotada por abajo en el conjunto $\{x^{(k)} - s\nabla f(x) | s > 0\}$. Entonces si $0 < c_1 < c_2$ existen intervalos de largo de paso satisfaciendo las condiciones de Wolfe.

El Teorema de Convergencia también aplica si en cada iteración el nuevo punto $x' = x - s\nabla f(x) = \phi(s)$ satisface las condiciones de Wolfe, pero s no necesariamente minimiza $\phi(s)$.

La práctica del Backtracking consiste en escoger $\bar{s} > 0$, $\rho \in (0, 1)$, $c_1 \in (0, 1)$ y el conjunto $s := \bar{s}$, el cual se repite hasta que $\phi(s) \leq \phi(0) + c_1 s \phi'(s)$, con $s := \rho \cdot s$ y luego repetir.

El valor s satisface (33), el largo del paso s no es tan corto con $s = \rho \cdot s'$ para algún s' que no cumpla (33) y reemplace (34).

11.1. Método de Newton

Se asume que $f(x)$ es dos veces continua y diferenciable. Entonces se recuerda que en el teorema de Taylor se da el punto x , $f(x+p) \approx f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x) p =: m(p)$, luego se va al punto $x+p$ donde p minimiza $m(p)$.

Luego, el algoritmo comienza en el punto $x_0 \in \mathbb{R}^n$, $k := 0$ y luego se repite lo siguiente: Dado el punto x_k , sea $m_k(p) := f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p$, se toma p que minimiza $m_k(p)$, luego se define $x_{k+1} := x_k + p$, luego $k := k + 1$.

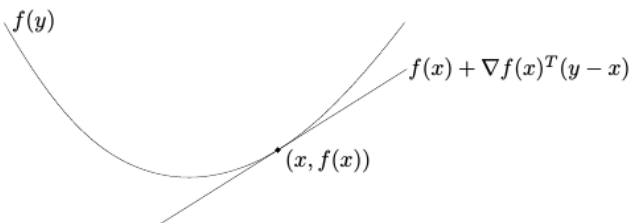
11.2. Ratio de convergencia

Definición 76. Sea f una función continuamente diferenciable. Luego f es β -suavizada si

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|$$

Teorema 77. Sea f convexa y β -suavizada en \mathbb{R}^n . Luego, el gradiente descendente con tamaño de paso $t = 1/\beta$ satisface

$$f(x^{(k)}) - f(x^*) \leq \frac{2\beta \|x^{(0)} - x^*\|^2}{k-1}$$



Luego, f es convexa, lo que implica que $f(y) \geq f(x) + \nabla f(x)^T (y - x)$, entonces $f(x) - f(y) \leq \nabla f(x)^T (x - y)$.

Definición 78. Una función f es α -fuerte convexa si satisface

$$f(x) - f(y) \leq \nabla f(x)^T (x - y) - \frac{\alpha}{2} \|x - y\|^2$$

Teorema 79. Sea f α -fuerte convexa y β -suave. Entonces el gradiente descendente con tamaño de paso $t = 1/\beta$ satisface

$$f(x^{(k+1)}) - f(x^*) \leq \frac{\beta}{2} \exp\left(-\frac{4k}{\beta/\alpha + 1}\right) \|x^{(0)} - x^*\|^2$$