

PROGRAMACIÓN LINEAL Y NO LINEAL

Iker Caballero Bragagnini

Tabla de contenido

INTRODUCCIÓN A LA PROGRAMACIÓN LINEAL Y NO LINEAL	2
LOS CONCEPTOS BÁSICOS PARA LA PROGRAMACIÓN LINEAL Y NO LINEAL	6
LAS PROPIEDADES BÁSICAS DE LA PROGRAMACIÓN LINEAL.....	11
EL MÉTODO SIMPLEX: NOCIONES BÁSICAS	23
EL MÉTODO SIMPLEX: ALGORITMO MATRICIAL Y TABULAR	31
LA DUALIDAD DE LOS PROGRAMAS LINEALES.....	40
LOS PROBLEMAS DE TRANSPORTE Y DE FLUJOS DE REDES.....	48
EL PROBLEMA DE ÁRBOL DE EXPANSIÓN DE COSTE MÍNIMO	59
LAS PROPIEDADES BÁSICAS DE LA OPTIMIZACIÓN NO RESTRINGIDA.....	62
LOS MÉTODOS DE DESCENSO BÁSICOS	76
LAS CONDICIONES DE MINIMIZACIÓN RESTRINGIDA	78
LOS MÉTODOS DUALES Y DE CORTE DE PLANOS	92

Introducción a la programación lineal y no lineal

- El concepto de optimización está integrado en el análisis de decisiones complejas y en problemas de asignación. Utilizando la filosofía de optimización en estas decisiones, uno enfoca el problema, el cual involucra la selección de valores usando un número de variables, a través de un solo objetivo diseñado para cuantificar el rendimiento y la calidad de la decisión
 - El objetivo se intenta maximizar o minimizar sujeta a restricciones que pueden limitar la selección de valores de las variables, por lo que cualquier si un problema se puede aislar y caracterizar con un objetivo, la optimización es un buen marco de análisis
 - No suele ser posible representar todas las complejidades e interacciones, objetivos y restricciones de una situación, por lo que este marco se tiene que considerar como una aproximación
- Una de las aplicaciones más importantes de la programación lineal y no lineal y de la optimización se da en el campo de la investigación operativa (OR)
 - La investigación operativa es una disciplina que trata de sobre como decidir (toma de decisiones) científicamente el mejor diseño de un sistema real complejo
 - En general, se trabaja bajo recursos escasos y sistemas muy complejos, intentando buscar la mejor solución (la más económica, la que tenga menos riesgo, etc.)
 - Las etapas para poder solucionar un problema de investigación operativa son las siguientes:
 - La primera etapa es la definición de un problema, lo cual es la explicación detallada del problema o sistema en el mundo real. Se tienen que identificar los componentes y atributos del sistema, y normalmente se tiende a simplificar y realizar aproximaciones
 - La segunda etapa es la modelización, la cual consiste en formular un problema matemático de programación una vez se tienen todos los datos. Primero se tiene que definir las variables objetivo (representando los elementos del sistema a modelar sobre los cuales se quiere decidir), después la función objetivo (cuantifica y mide la calidad de las decisiones) y finalmente se escogen las restricciones (las limitaciones que los valores de las variables de decisión tienen que cumplir)

$$\max/\min (\text{objective function}) \quad s. t.$$

(constraints)

- La tercera etapa consiste en encontrar soluciones al modelo matemático a través de técnicas de optimización
 - La cuarta etapa es la validación del modelo, la cual consiste en comprobar si la solución es correcta. De no ser así, se debe modificar el modelo y solucionarlo otra vez
 - Finalmente, la última etapa consiste en la implementación del problema en el mundo real
- Existen varios tipos de modelos de investigación operativa, y se pueden clasificar según los siguientes criterios:
- Según el tipo de variable de decisión, los modelos pueden ser continuos (si $x \in \mathbb{R}^n$), discretos (si $x \in \mathbb{Z}^n$ o $x \in \{0,1\}$) o mixtos
 - Según el tipo de función objetivo y/o restricciones, los modelos pueden ser lineales o no lineales
 - Según si el modelo tiene restricciones o no, se pueden clasificar como modelos restringidos o no restringidos
 - Según si se conocen los datos del problema de manera exacta o no, los modelos pueden ser deterministas o estocásticos (con variables aleatorias)
 - Según si el modelo tiene solución, se pueden clasificar en modelos factibles (con una o infinitas soluciones), modelos infactibles (de modo que $\{Ax = b, x \geq 0\}$) y los modelos ilimitados (donde la función objetivo tiende a $-\infty$ o a ∞)
- Los modelos de programación lineal cumplen una serie de suposiciones implícitas:
- La contribución de una variable de decisión a la función objetivo y/o a las restricciones es directamente proporcional al valor de las variables de decisión (suposición de proporcionalidad)
 - El efecto de una variable a la función objetivo o en las restricciones es siempre la misma, independientemente del resto de las variables del problema (suposición de adición)
 - Las variables de decisión x del problema lineal pueden tomar cualquier valor real (suposición de divisibilidad)

- Se conocen todos los datos y parámetros del problema de forma exacta (suposición de determinismo)
- Existen tres grandes tipos de problemas: los problemas de programación lineal, los problemas no restringidos y los problemas restringidos (en donde las últimas dos partes forman los problemas de programación no lineal)
 - La programación lineal es el mecanismo más natural para formular un gran número de problemas, y este tipo de problemas se caracteriza por tener funciones lineales de incógnitas
 - El objetivo es una función lineal de las incógnitas y las restricciones son ecuaciones lineales o desigualdades lineales sobre las incógnitas
 - La popularidad de los modelos de programación lineal reside en la fase de formulación del análisis más que en la fase de solución. Un gran número de restricciones y objetivos en la práctica son lineales
 - Otra razón suele ser que es menos difícil de definir, por lo que, si una función objetivo no es puramente lineal por definición, es más fácil definirla linealmente que usar un modelo complejo
 - Los problemas no restringidos no tienen propiedades estructurales, pero tienen aplicabilidad debido a que hay veces en las que se considera realista no tener restricciones y a que los problemas restringidos se pueden convertir fácilmente en no restringidos
 - Si, por ejemplo, se considera la optimización para todas las variables, entonces es probable que no haya restricciones, dado que estas no son más que delimitaciones de foco artificiales
 - Cuando hay restricciones con igualdad, en verdad solo se limitan los grados de libertad, de modo que se puede reconstruir la función objetivo de modo que se tengan los grados de libertad reales (sustituyendo incógnitas y eliminando restricciones)
 - Mucha de la teoría y algoritmos para solucionar problemas restringidos tienen su motivación y se verifican para el caso no restringido (antes que el restringido)
 - La mayoría de problemas en la práctica, sin embargo, se formulan como problemas restringidos. Eso suele deberse a que problemas complejos no se puede tratar considerando todas las opciones posibles, sino separándolos en subproblemas con sus respectivas restricciones

- Por lo tanto, es común encontrarse con problemas matemáticos de programación no lineal restringida. El problema general de programación matemática se puede formular de la siguiente manera:

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && h_i(\mathbf{x}) = 0, && i = 1, 2, \dots, m \\ &&& g_j(\mathbf{x}) \leq 0, && j = 1, 2, \dots, r \\ &&& \mathbf{x} \in S. \end{aligned}$$

- En este problema, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ es el vector n -dimensional de incógnitas y f , h_i y g_j son funciones de variable real. Además, el subconjunto S es un subconjunto de un espacio n -dimensional y f es la función objetivo
- Una medida de complejidad de los problemas de programación es el tamaño, medido en términos del número de incógnitas o del número de restricciones, y se pueden distinguir tres clases de problemas:
 - Los problemas de pequeña escala son problemas con cinco o menos incógnitas y restricciones; los problemas de mediana escala son problemas de cinco a cien incógnitas y restricciones; y los problemas de gran escala son problemas de más de cien incógnitas y restricciones
 - El tamaño del problema también determina el método de resolución de este, de modo que problemas pequeños se pueden hacer a mano y problemas más grandes con ordenador
 - Mucha de la teoría básica de asociada con la optimización se centra en obtener condiciones suficientes y necesarias que un punto de solución satisface (más que en el cálculo)
- Las características más importantes del uso de ordenadores en programación matemática es la capacidad de realizar operaciones repetitivas de manera eficiente, por lo que muchos de los algoritmos diseñados para resolver estos problemas son iterativos por naturaleza
 - Al buscar un vector que resuelva el problema de programación, normalmente se selecciona un vector inicial \mathbf{x}_0 y el algoritmo genera un vector mejorado \mathbf{x}_1 , y el proceso se repite para encontrar más vectores mejorados
 - Por lo tanto, se tiene una secuencia de puntos mejorados $\{\mathbf{x}_i\}_{i=0}^{\infty}$ que se acerca a un punto solución \mathbf{x}^* . Dependiendo del algoritmo

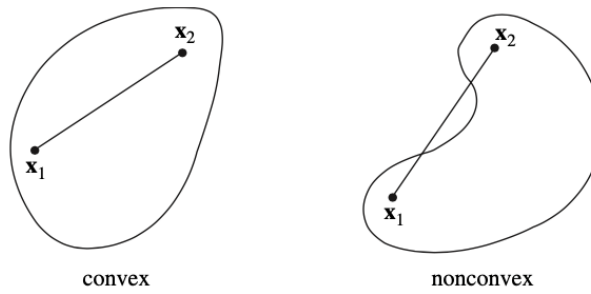
y del tipo de problema la secuencia es finita o infinita, y, en la práctica, el problema se soluciona cuando un punto está lo suficientemente cerca del óptimo

- La teoría de los algoritmos iterativos se puede dividir en tres aspectos: la creación de los mismos algoritmos, el análisis de convergencia real y el análisis de convergencia local
 - El análisis de convergencia real se basa en verificar si el algoritmo genera una secuencia de puntos que converge a un punto de solución, mientras que el análisis de convergencia local se centra en la velocidad a la que la secuencia del algoritmo converge al punto de solución

Los conceptos básicos para la programación lineal y no lineal

- Algunos conceptos básicos de cálculo y topología que se utilizan para el desarrollo de resultados en programación lineal y no lineal son los siguientes:
 - Una secuencia de vectores x_0, x_1, x_2, \dots , denotada por $\{x_k\}_{k=0}^{\infty}$, converge a un límite x si $|x_k - x| \rightarrow 0$ cuando $k \rightarrow 0$ (dada una $\varepsilon > 0$, existe una N tal que si $k \geq N$, entonces $|x_k - x| < \varepsilon$), y esta convergencia se denota como $x_k \rightarrow x$
 - Un punto x es un punto límite de la secuencia x_k si hay una subsecuencia de $\{x_k\}_{k=0}^{\infty}$ convergente a x . Por lo tanto, x es un punto límite de $\{x_k\}_{k=0}^{\infty}$ si hay un subconjunto \mathcal{K} de enteros positivos tal que $\{x_k\}_{k \in \mathcal{K}}$ es convergente a x
 - Una esfera alrededor de x es un conjunto de la forma $\{y : |y - x| < \varepsilon\}$ para alguna $\varepsilon > 0$. Esta esfera también se suele conocer como la vecindad de x de radio ε
 - Un subconjunto S de E^n es abierto si alrededor de cualquier punto en S existe una esfera que está contenida en S . Equivalentemente, S es abierto si, dada una $x \in S$, existe una $\varepsilon > 0$ tal que $|y - x| < \varepsilon$
 - El interior de cualquier S en E^n es el conjunto de puntos $x \in S$ que son el centro de alguna esfera contenida en S , y se denota por \dot{S} . El interior de un conjunto siempre es abierto, dado que es el conjunto abierto más grande contenido en S
 - Un conjunto P es cerrado si cualquier punto que esté arbitrariamente cerca al conjunto P es un miembro de P . En otras palabras, P es cerrado si $x_k \rightarrow x$ con $x_k \in P$ implica que $x \in P$

- La clausura de cualquier conjunto P en E^n es el conjunto cerrado más pequeño conteniendo P , y es denotado por \bar{S}
 - La frontera de un conjunto es la parte de la clausura que no está en el interior
- Un conjunto es compacto si el conjunto es cerrado y si está contenido dentro de una esfera de un radio finito (está acotado)
 - Un resultado importante demostrado por Weierstrass es que, si S es un conjunto compacto y $\{x_k\}_{k=0}^{\infty}$ es una secuencia en donde cada elemento pertenece a S , entonces $\{x_k\}$ tiene un punto límite en S (hay una subsecuencia que converge a un punto en S)
- Un conjunto de funciones de valores reales f_1, f_2, \dots, f_m en E^n puede entenderse como una función vectorial $\mathbf{f} \equiv (f_1, f_2, \dots, f_m)$, la cual asigna a un vector $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))$ en E^m para todo vector $\mathbf{x} \in E^n$
 - Una función de este tipo se denomina continua si todas las funciones que lo componen son continuas
 - Si cada componente de \mathbf{f} es continua en algún conjunto abierto de E^n , entonces se escribe $\mathbf{f} \in C$. Si, además, cada componente de la función tiene derivadas parciales continuas de orden p , se escribe $\mathbf{f} \in C^p$
- Si g es una función de valores reales de variable real, la notación $g(x) = O(x)$ significa que $g(x)$ va a cero al menos tan rápido como lo hace x . De manera más precisa, existe una $K \geq 0$ tal que $\left| \frac{g(x)}{x} \right| \leq K$ cuando $x \rightarrow 0$
 - La notación $g(x) = o(x)$ significa que $g(x)$ va a cero más rápido de lo que lo hace x , o que $K > 0$
- Los conceptos relacionados a los conjuntos convexos son importantes para la teoría de optimización, por lo que se comentan algunas de las definiciones y de los resultados más importantes
 - Un conjunto C en E^n se denomina convexo si para cualquier $x_1, x_2 \in C$ y cualquier número real $\alpha \in (0,1)$, el punto $\alpha x_1 + (1 - \alpha)x_2 \in C$
 - Esta definición se puede interpretar de manera geométrica, de modo que un conjunto es convexo si, dados dos puntos en conjunto, cualquier punto en el segmento que une a estos dos puntos también es un miembro del conjunto



- Si $C \in E^n$ es un conjunto convexo y β es un número real, el conjunto $\beta C \equiv \{x : x = \beta c, c \in C\}$ es convexo

$$x_1 = \beta c_1 \text{ \& \> } x_2 = \beta c_2$$

$$\Rightarrow \alpha x_1 + (1 - \alpha)x_2 = \alpha\beta c_1 + (1 - \alpha)\beta c_2 =$$

$$= \beta[\alpha c_1 + (1 - \alpha)c_2] = \beta c \in \beta C$$

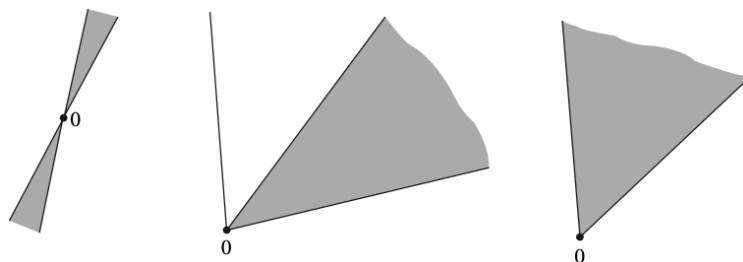
- Si C y D son conjuntos convexos, entonces el conjunto $C + D \equiv \{x : x = c + d, c \in C, d \in D\}$ es convexo

$$x_1 = c_1 + d_1 \text{ \& \> } x_2 = c_2 + d_2$$

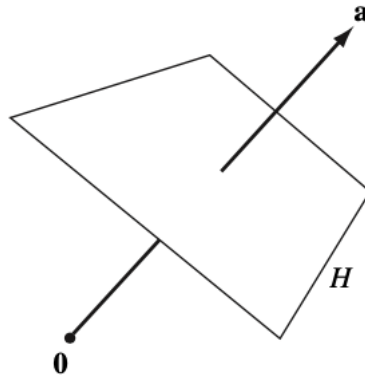
$$\Rightarrow \alpha x_1 + (1 - \alpha)x_2 = \alpha(c_1 + d_1) + (1 - \alpha)(c_2 + d_2) =$$

$$= \alpha c_1 + (1 - \alpha)c_2 + \alpha d_1 + (1 - \alpha)d_2 = c + d \in C + D$$

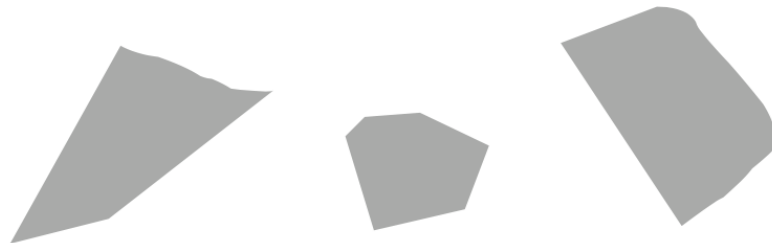
- La intersección de cualquier colección de conjuntos convexos es también convexa
- Siendo S un subconjunto de E^n , el casco convexo o *convex hull* de S , denotado como $conv(S)$, es el conjunto que es la intersección de todos los conjuntos convexos conteniendo S
 - El casco convexo cerrado de S se define como la clausura de $conv(S)$
- Un conjunto C es un cono si $x \in C$ implica que $\alpha x \in C$ para toda $\alpha > 0$, y un cono que es convexo se denomina cono convexo



- La propiedad básica es que si un punto x pertenece a un cono, entonces la línea de en medio desde el origen hasta el punto (sin contar el origen mismo) también tienen que pertenecer al cono
- Los tipos de conjuntos convexos más importantes son los hiperplanos, los cuales dominan toda la teoría de la optimización y se usan en diversos métodos. Finalmente, también se revisa el concepto de punto extremo de un conjunto convexo, el cual es vital para la programación lineal
 - Un conjunto V en E^n es una variedad lineal si, dado $x_1, x_2 \in V$, se cumple que $\lambda x_1 + (1 - \lambda)x_2 \in V$ para todos los números reales λ
 - La diferencia entre una variedad lineal y un conjunto convexo es que, en una variedad lineal, todos los puntos de la línea entre dos puntos deben pertenecer al conjunto, mientras que en un conjunto convexo solo los puntos entre los dos extremos del segmento deben pertenecer al conjunto
 - Por lo tanto, en tres dimensiones, las variedades lineales son los puntos, las líneas, los planos bidimensionales y el espacio entero
 - La dimensión de una variedad lineal en E^n se puede encontrar haciendo una translación (o desplazamiento) para que contenga el origen (a través de crear un conjunto nuevo que contenga el origen) y, después, determinando la dimensión del conjunto resultante, el cual es un subespacio de E^n
 - Un hiperplano en E^n es una variedad lineal con $n - 1$ dimensiones
 - Los hiperplanos generalizan el concepto de plano bidimensional en un espacio tridimensional y se pueden entender como las variedades lineales más grandes dentro de un espacio
 - Siendo a es un vector columna n -dimensional diferente de cero y c un número real, el conjunto $H \equiv \{x \in E^n : a'x = c\}$ es un hiperplano en E^n
 - Debido a que H es una variedad lineal, y suponiendo que $x_1 \in H$, se puede hacer una translación para $-x_1$ y así obtener un conjunto $M = H - x_1$ que será un subespacio lineal de E^n . Este subespacio consiste de todos los vectores x que cumplen la igualdad $a'x = 0$ (todos los vectores que son ortogonales a a), de modo que, por álgebra lineal básico, las dimensiones del subespacio son $n - 1$
 - Siendo H un hiperplano en E^n , existe un vector n -dimensional y una constante c tal que $H = \{x \in E^n : a'x = c\}$



- Siendo \mathbf{a} un vector diferente a cero en E^n y c un número real, se pueden definir el semiespacio positivo cerrado o *positive closed half plane* $H_+ \equiv \{\mathbf{x} : \mathbf{a}'\mathbf{x} \geq c\}$ y el semiespacio negativo cerrado o *negative closed half plane* $H_- \equiv \{\mathbf{x} : \mathbf{a}'\mathbf{x} \leq c\}$
 - Además, también se pueden definir de manera análoga a la anterior el semiespacio positivo abierto o *positive open half space* $\dot{H}_+ \equiv \{\mathbf{x} : \mathbf{a}'\mathbf{x} > c\}$ y el semiespacio negativo abierto o *negative open half space* $\dot{H}_- \equiv \{\mathbf{x} : \mathbf{a}'\mathbf{x} < c\}$
 - Estos semiespacios son conjuntos convexos y la unión de H_+ y H_- conforman el espacio completo
- Un conjunto que se puede expresar como la intersección de un número finito de semiespacios se denomina polígono convexo



- Los polítopos convexos son los conjuntos que se obtienen como la familia de soluciones a un sistema de inecuaciones lineales de la forma $A\mathbf{x} \leq \mathbf{b}$, dado que cada desigualdad individual define un semiplano y la familia de solución es la intersección de estos

$$A\mathbf{x} \leq \mathbf{b} = \begin{cases} \mathbf{a}'_1\mathbf{x} \leq b_1 \\ \mathbf{a}'_2\mathbf{x} \leq b_2 \\ \dots \\ \mathbf{a}'_m\mathbf{x} \leq b_m \end{cases}$$

- Un polígono acotado que no sea vacío se denomina poliedro

- Un punto x en un conjunto convexo C es un punto extremo de C si no existen dos puntos distintos x_1 y x_2 en C tal que $x = \alpha x_1 + (1 - \alpha)x_2$ para alguna $\alpha \in (0,1)$
 - Una variedad lineal que consista de más de un punto no tiene puntos extremos
- Un conjunto convexo cerrado y acotado en E^n es igual a un casco convexo cerrado de sus puntos extremos
- Un poliedro convexo puede ser descrito como una intersección acotada de un número finito de semiespacios cerrados o como el casco convexo de un número finito de puntos
 - Un poliedro convexo es un polítopo acotado, y como también es la intersección de semiespacios, entonces un poliedro convexo también es cerrado. Por lo tanto, cualquier poliedro convexo es también un casco convexo de sus puntos extremos
 - Como cualquier polítopo tiene un número finito de puntos extremos, un poliedro convexo es igual al casco convexo de un número finito de puntos

Las propiedades básicas de la programación lineal

- Un programa lineal (LP) es un problema de optimización en el que la función objetivo es lineal en las incógnitas y las restricciones consisten de igualdades y desigualdades lineales
 - Aunque la forma de las restricciones puede diferir de un problema a otro, cualquier problema de programación lineal se puede transformar en la siguiente forma estándar:

$$\begin{array}{ll}
 \text{minimize} & c_1x_1 + c_2x_2 + \dots + c_nx_n \\
 \text{subject to} & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\
 & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\
 & \cdot \qquad \qquad \qquad \cdot \\
 & \cdot \qquad \qquad \qquad \cdot \\
 & \cdot \qquad \qquad \qquad \cdot \\
 & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \\
 \text{and} & x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0,
 \end{array}$$

- En este caso, las a_{ij} , b_i y c_j son constantes reales fijas y las x son variables reales. Siempre se asume que las ecuaciones han sido multiplicadas por la unidad negativa, de modo que $b_i \geq 0$ para toda i

- En notación vectorial, el problema se puede plantear de la siguiente manera:

$$\begin{aligned} & \text{minimize} \quad \mathbf{c}^T \mathbf{x} \\ & \text{subject to} \quad \mathbf{Ax} = \mathbf{b} \quad \text{and} \quad \mathbf{x} \geq \mathbf{0} \end{aligned}$$

- En este último caso \mathbf{x} es un vector columna n -dimensional, \mathbf{c}^T es un vector fila n -dimensional, \mathbf{A} es una matriz $m \times n$ y \mathbf{b} es un vector columna m -dimensional
- Suponiendo que se tienen restricciones genéricas $f_i(\mathbf{x}) \leq b_i$ y $g_j(\mathbf{x}) \leq b_j$, algunas de las restricciones más típicas o lógicas son las siguientes:

- Si una de dos restricciones se tiene que cumplir obligatoriamente, se selecciona un valor M lo suficientemente grande para que $\forall \mathbf{x}, f_1(\mathbf{x}) \leq b_1 + M$ y $f_2(\mathbf{x}) \leq b_2 + M$. Tomando una variable binaria $y \in \{0,1\}$, se utilizan las siguientes dos restricciones:

$$\begin{aligned} f_1(\mathbf{x}) &\leq b_1 + My & f_2(\mathbf{x}) &\leq b_2 + M(1 - y) \\ y &\in \{0,1\} \end{aligned}$$

- Si se han de cumplir k de las N restricciones del modelo, se selecciona un valor M lo suficientemente grande para que $\forall \mathbf{x} \ \& \ \forall i = 1, 2, \dots, N: f_i(\mathbf{x}) \leq b_i + M$. Tomando N variables binarias $y_i \in \{0,1\}$, se utilizan las siguientes restricciones:

$$\begin{aligned} f_1(\mathbf{x}) &\leq b_1 + My_1 & \dots & & f_N(\mathbf{x}) &\leq b_N + My_N \\ \sum_{i=1}^N y_i &= N - k & y_i &\in \{0,1\} \text{ for } i = 1, 2, \dots, N \end{aligned}$$

- Si se quiere que una restricción se cumpla si se cumple otra (como implicación), entonces se selecciona un valor M lo suficientemente grande y se puede modelar como $f_i(\mathbf{x}) \leq b_i$ y $g_j(\mathbf{x}) \geq b_j$. Tomando una variable binaria $y \in \{0,1\}$, se utilizan las siguientes dos restricciones:

$$\begin{aligned} f_i(\mathbf{x}) &\leq b_i + M(1 - y) & -g_j(\mathbf{x}) &\leq b_j + My \\ y &\in \{0,1\} \end{aligned}$$

- Si se quiere que una restricción pueda tomar N valores posibles, entonces se puede modelar usando diferentes valores α_i para $i = 1, 2, \dots, N$ y tomando N variables binarias $y_i \in \{0,1\}$:

$$f_i(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i \quad \sum_{i=1}^N y_i = 1$$

$$y_i \in \{0,1\} \text{ for } i = 1, 2, \dots, N$$

- Existen muchas maneras de modelar un mismo problema, de modo que lo ideal suele ser transformar un problema a la manera estándar a través de varios métodos
 - Para poder transformar un problema de programación lineal con restricciones con desigualdades en uno de forma estándar, es posible utilizar variables de superávit y de holgura
 - Considerando un problema con restricciones en forma desigualdades lineales de menor o igual, se pueden expresar en términos de igualdades a través de sumar variables de holgura o *slack variables* y_1, y_2, \dots, y_m (las cuales son positivas). Por lo tanto, considerando el problema con $n + m$ incógnitas, se tiene el modelo estándar y la matriz $m \times (n + m)$ describe las igualdades lineales como $[A|I]$ (las columnas de la matriz de coeficientes es una partición de las n columnas originales y la matriz identidad $m \times m$)

$$\begin{array}{ll}
 \text{minimize} & c_1 x_1 + c_2 x_2 + \dots + c_n x_n \\
 \text{subject to} & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \\
 & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \\
 & \cdot \\
 & \cdot \\
 & \cdot \\
 & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m \\
 \text{and} & x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0,
 \end{array}$$

$$\begin{array}{ll}
 \text{minimize} & c_1 x_1 + c_2 x_2 + \dots + c_n x_n \\
 \text{subject to} & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + y_1 = b_1 \\
 & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + y_2 = b_2 \\
 & \cdot \\
 & \cdot \\
 & \cdot \\
 & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + y_m = b_m \\
 \text{and} & x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0, \\
 \text{and} & y_1 \geq 0, y_2 \geq 0, \dots, y_m \geq 0.
 \end{array}$$

- Si las restricciones del problema anterior cambian a desigualdades de mayor a igual, entonces se pueden restar variables de superávit o *surplus variables* y_1, y_2, \dots, y_m (las cuales

son positivas) y tener un enfoque igual al anterior para transformar el problema a la forma general

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \geq b_i$$

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n - y_i = b_i$$

- Si el programa lineal se da en forma estándar excepto porque una o más variables tienen una superior o una cota inferior diferente de cero, se puede transformar el problema de las siguientes formas:

- Si hay una cota superior del tipo $0 \leq x_i \leq u_i$, entonces se añaden variables de holgura para obtener una restricción del tipo $x_i + y_i = u_i$
- Si hay una cota inferior del tipo $l_i \leq x_i$, entonces se añaden variables de exceso y_i para obtener una restricción del tipo $x_i - s_i = l_i$
- Otro modo de poder lidiar con una cota inferior del tipo $l_i \leq x_i$ es haciendo una sustitución de variables $\tilde{x}_i = x_i - l_i$, por lo que el problema se podría expresar de la siguiente manera:

$$\begin{array}{ll} \min \mathbf{c}^T (\tilde{\mathbf{x}} + \mathbf{l}) & \min \mathbf{c}^T \tilde{\mathbf{x}} + \mathbf{c}^T \mathbf{l} \\ A(\tilde{\mathbf{x}} + \mathbf{l}) = \mathbf{b} & A\tilde{\mathbf{x}} = \mathbf{b} - A\mathbf{l} \\ \tilde{\mathbf{x}} \geq \mathbf{0} & \tilde{\mathbf{x}} \geq \mathbf{0} \end{array}$$

- Si el programa lineal se da en forma estándar excepto porque una o más variables no requieren la condición de no negatividad, llamadas variables libres, el problema se puede transformar a forma estándar con una de las dos siguientes técnicas:

- Suponiendo que una variable x_i no tiene por qué cumplir con una restricción $x_i \geq 0$, esta puede tomar valores positivos o negativos y se puede hacer una sustitución $x_i = u_i - v_i$, en donde $u_i \geq 0$ y $v_i \geq 0$. De este modo, las condiciones de no negatividad de las variables se mantienen y el problema se expresa en términos de $n + 1$ variables. Esta técnica es un poco redundante, dado que añadir una constante a u_i y a v_i a la vez no afecta a x_i (por lo que la representación de x_i no es única)
- El segundo enfoque es eliminar x_i con una de sus restricciones. Cogiendo una de las m ecuaciones en donde x_i no tiene coeficiente, se puede expresar como una combinación lineal de las otras variables más una constante, y se puede sustituir en el problema para expresarlo en términos de otras variables. De este

modo, se tiene $n - 1$ variables y $m - 1$ restricciones y se puede determinar x_i una vez se encuentra la solución

$$\begin{aligned} \text{minimize} \quad & x_1 + 3x_2 + 4x_3 \\ \text{subject to} \quad & x_1 + 2x_2 + x_3 = 5 \\ & 2x_1 + 3x_2 + x_3 = 6 \\ & x_2 \geq 0, \quad x_3 \geq 0 \end{aligned}$$

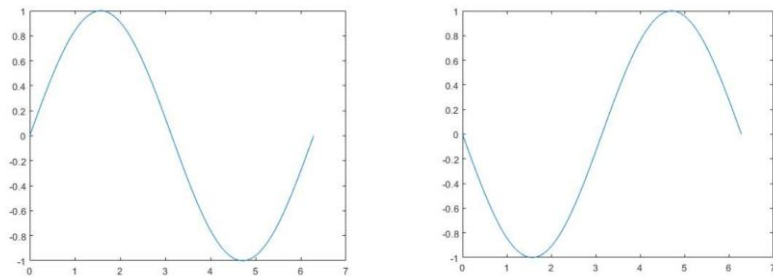
$$x_1 = 5 - 2x_2 - x_3$$

$$\begin{aligned} \text{minimize} \quad & x_2 + 3x_3 \\ \text{subject to} \quad & x_2 + x_3 = 4 \\ & x_2 \geq 0, \quad x_3 \geq 0 \end{aligned}$$

- Si el programa lineal se da en forma estándar excepto porque se maximiza en vez de minimizar, entonces se puede multiplicar la función objetivo por -1 :

$$\begin{array}{ll} \min \mathbf{c}^T \mathbf{x} & \max -\mathbf{c}^T \mathbf{x} \\ A\mathbf{x} = \mathbf{b} & A\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} & \mathbf{x} \geq \mathbf{0} \end{array}$$

- El óptimo \mathbf{x}^* de ambos problemas coincide porque el máximo de una función es equivalente al mínimo del opuesto de la función



- Considerando el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$, en donde \mathbf{x} es un vector n -dimensional, \mathbf{b} es un vector m -dimensional y A es una matriz $m \times n$, si de las n columnas de A se selecciona un conjunto de m columnas linealmente independientes, se puede formar una submatriz B de tamaño $m \times m$
 - Debido a que la matriz es cuadrada y sus vectores columna son linealmente independientes, esta es invertible y existe una única solución para cualquier matriz \mathbf{b} $m \times 1$ (teorema de equivalencia) y se puede establecer la siguiente ecuación

$$B\mathbf{x}_B = \mathbf{b}$$

- En este caso, x_B es una matriz $m \times 1$ o un vector m -dimensional, pero si se define $x \equiv [x'_B | \mathbf{0}']$, de modo que los primeros m componentes del vector son los mismos que en x_B y los otros componentes son cero, se obtiene una solución para el sistema $Ax = b$, ya que este será compatible y determinado
- La matriz A , por lo tanto, puede entenderse como la adjunción de las matrices B , que contiene vectores linealmente independientes como columnas (las variables), y una matriz N , formada de aquellos vectores que son combinación lineal de los de B

$$\min 2x_1 + 3x_2 \quad s.t. \quad \begin{aligned} x_1 + x_2 + s_1 &= 12 \\ -2x_1 + x_2 + s_2 &= 2 \\ x_1 + 3x_2 + s_3 &= 30 \\ x_1, x_2, s_1, s_2, s_3 &\geq 0 \end{aligned}$$

$$\Rightarrow B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 1 \\ -2 & 1 \\ 1 & 3 \end{bmatrix}$$

- La matriz B no siempre está compuesta de los vectores que representan las mismas variables, sino que múltiples vectores y variables pueden ser linealmente independientes entre ellos y, por tanto, haber diversas combinaciones
- Dado un conjunto de m ecuaciones lineales simultáneas en n incógnitas $Ax = b$ y siendo B una submatriz invertible $m \times m$ hecha de columnas de A , si todos los $n - m$ componentes de x no asociadas con columnas de B se igualan a cero, la solución del conjunto resultante de ecuaciones se denomina solución básica a $Ax = b$ con respecto a la base B
 - Los componentes de x asociados con las columnas de B se llaman variables básicas, ya sean las variables originales o aquellas creadas artificialmente para obtener la forma estándar del problema
 - En esta definición B se denomina base porque sus columnas son linealmente independientes que se pueden considerar como vectores columna que forman una base para el espacio E^m
 - La solución básica corresponde a una expresión para la b es una combinación lineal de estos vectores base

$$x_B = B^{-1}b$$

- En general, $Ax = b$ puede no tener soluciones básicas, pero se evitan trivialidades y dificultades no esenciales haciendo suposiciones elementales sobre la estructura de la matriz A
 - Primero se suele asumir que $n > m$, por lo que el número de variables x_i excede el número de restricciones de igualdad, y después se suele asumir que las filas de A son linealmente independientes, correspondientes a las m ecuaciones (dado que si no fueran independientes habría contradicción entre restricciones o redundancias que se podrían eliminar)
 - Formalmente, se hace la siguiente suposición formal: la matriz A $m \times n$ tiene $m < n$, y las m filas de A son linealmente independientes (correspondiendo a la independencia lineal de las m ecuaciones), por lo que el rango de la matriz es m . Esto hace que el sistema siempre tenga una solución (que el sistema sea consistente) y al menos una única solución básica
- Las variables básicas en una solución básica no necesariamente son diferentes de cero, y si una o más variables básicas en la solución básica tienen un valor de cero, la solución se denomina solución básica degenerada y estas variables se las llama variables básicas no factibles
 - En una solución básica no degenerada, las variables básicas y la base B pueden ser inmediatamente identificadas a partir de los componentes positivos de la solución (dado que no hay ninguno que sea igual a cero)
 - Hay ambigüedad asociada a la solución básica degenerada porque las variables básicas iguales a cero y las variables no básicas igualadas a cero se pueden intercambiar
- Hasta ahora no se han considerado las restricciones de positividad de las variables, pero definiciones equivalentes aplican cuando se consideran este tipo de restricciones (el procedimiento anterior no cambia)
 - Por lo tanto, el siguiente sistema de restricciones representa las restricciones del programa lineal en su forma estándar:

$$Ax = b$$

$$x \geq 0$$

- La región factible se define como el conjunto de puntos $C \equiv \{x : Ax = b, x \geq 0\}$, que es el conjunto de vectores que satisface el sistema de restricciones anterior

- Un vector \mathbf{x} que satisface el sistema de restricciones anterior se denomina factible para estas restricciones. Una solución factible para las restricciones que es también una solución básica se denomina solución básica factible, y si esta es una degenerada, se denomina solución básica degenerada factible
- A partir de todas estas nociones, se puede establecer la importancia principal de las soluciones factibles básicas al resolver programas lineales. El método de demostración y el teorema en si son igual de importantes, dado que la demostración representa el comienzo del desarrollo del método Simplex y el teorema muestra que solo es necesario considerar soluciones básicas factibles al optimizar porque el óptimo siempre reside en este
 - Considerando un problema de programación lineal en forma estándar, se puede definir el concepto de solución óptima

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T \mathbf{x} \\ \text{subject to} & \mathbf{Ax} = \mathbf{b} \quad \text{and} \quad \mathbf{x} \geq \mathbf{0} \end{array}$$

- Una solución factible a las restricciones que consiga el valor mínimo de la función objetivo sujeta a estas restricciones se denomina solución óptima factible. Si la solución es básica, entonces es una solución básica óptima factible
- El teorema fundamental de la programación lineal expresa que, dado un programa lineal en su forma estándar en donde A es una matriz $m \times n$ de rango m , se dan cumplen las siguientes proposiciones:
 - Si existe una solución factible, también existe una solución básica factible
 - Si existe una solución óptima factible, también existe una solución básica óptima factible
- La demostración de la primera proposición es la siguiente:
 - Denotando las columnas de A por $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ y suponiendo que $\mathbf{x} = (x_1, x_2, \dots, x_n)$ es una solución factible, la solución se puede expresar en términos de las columnas de A

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n = \mathbf{b}$$

- Si se asume que los p primeras variables x_i son mayores a cero, entonces se puede obtener la siguiente igualdad:

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_p = \mathbf{b}$$

- Si $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ son linealmente independientes, entonces $p \leq m$. Si $p = m$, la solución es básica (resuelve $B\mathbf{x}_B = \mathbf{b}$) y la prueba está completa, mientras que si $p < m$, entonces como el rango de A es m , se pueden encontrar $m - p$ vectores de los $n - p$ vectores restantes de modo que el conjunto de m vectores sea linealmente independiente. Asignando un valor nulo a las $m - p$ variables correspondientes, se puede encontrar una solución básica factible degenerada
- Si $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ son linealmente dependientes, entonces hay una combinación lineal no trivial de estos vectores que es cero. Por lo tanto, hay constantes y_1, y_2, \dots, y_p tal que $y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \dots + y_p\mathbf{a}_p = \mathbf{0}$, por lo tanto, multiplicando la ecuación por ε y sustrayéndola, se obtiene la siguiente igualdad, la cual se mantiene para toda ε :

$$(x_1 - \varepsilon y_1)\mathbf{a}_1 + (x_2 - \varepsilon y_2)\mathbf{a}_2 + \dots + (x_p - \varepsilon y_p)\mathbf{a}_p = \mathbf{b}$$

- Denotando $\mathbf{y} = (y_1, y_2, \dots, y_p, 0, 0, \dots, 0)$, entonces $\mathbf{x} - \varepsilon\mathbf{y}$ es la solución de las igualdades lineales para toda ε . Si $\varepsilon = 0$, entonces la solución es la solución factible original \mathbf{x} , pero si es diferente de cero, entonces los componentes pueden incrementar, reducirse o mantenerse constante dependiendo de si y_i es negativa, positiva o nula. Como se asume que al menos una y_i es positiva, entonces al menos un componente decrecerá cuando ε incremente, y ε se puede incrementar hasta que uno o más componentes son cero

$$\varepsilon = \min \left\{ \frac{x_i}{y_i} : y_i > 0 \right\}$$

- Con este valor de ε , la solución dada por $\mathbf{x} - \varepsilon\mathbf{y}$ es factible y tiene como mucho $p - 1$ variables positivas. Si se repite el proceso de manera necesaria, se pueden eliminar las variables positivas hasta quedarse con una solución factible correspondiente al número de columnas independientemente lineal y el primer caso aplica
- La demostración de la segunda proposición es el siguiente:
- Siendo $\mathbf{x} = (x_1, x_2, \dots, x_n)$ una solución factible óptima y habiendo p variables positivas x_1, x_2, \dots, x_n , existen dos casos como antes: uno para la independencia lineal (el que es igual al enfoque utilizado en el otro problema) y otro para la dependencia, en donde no hay independencia lineal

- Considerando el valor de la solución $\mathbf{c}^T \mathbf{x} - \varepsilon \mathbf{c}^T \mathbf{y}$ de $\mathbf{x} - \varepsilon \mathbf{y}$ y un valor muy pequeño de ε , $\mathbf{x} - \varepsilon \mathbf{y}$ es una solución factible para valores positivos o negativos de ε y por tanto $\mathbf{c}^T \mathbf{y} = 0$. Si $\mathbf{c}^T \mathbf{y} \neq 0$ y ε es de pequeña magnitud y de signo adecuado para hacer a $\mathbf{c}^T \mathbf{x} - \varepsilon \mathbf{c}^T \mathbf{y}$ más pequeña que $\mathbf{c}^T \mathbf{x}$ mientras se mantiene su factibilidad. Esto viola la suposición de optimalidad de \mathbf{x} , por lo que $\mathbf{c}^T \mathbf{y} = 0$
- Habiendo establecido que la nueva solución óptima factible con menos componentes positivos es también óptima, el resto se puede demostrar como en el apartado anterior de la primera proposición
- El teorema reduce la tarea de resolver programas lineales al de buscar en las soluciones básicas factibles, y como se tienen n variables y m restricciones, como mucho hay $\binom{n}{m}$ soluciones básicas (correspondiente al número de maneras para seleccionar m de n columnas) y hay un número finito de posibilidades

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

- Por lo tanto, el teorema fundamental proporciona una técnica de búsqueda finita de soluciones óptimas para el problema
- Los resultados anteriores tienen interpretaciones interesantes en términos de la teoría de conjuntos convexos que permite obtener una derivación alternativa del teorema fundamental y un entendimiento geométrico del resultado
 - Un punto \mathbf{x} en un conjunto convexo C se denomina un punto extremo de C si no hay dos puntos distintos \mathbf{x}_1 y \mathbf{x}_2 en C tal que $\mathbf{x} = \alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2$ para alguna $0 < \alpha < 1$
 - Por lo tanto, un punto extremo es un punto que no está estrictamente dentro de un segmento que conecta dos puntos del conjunto
 - Siendo A una matriz $m \times n$ de rango m , \mathbf{b} un vector m -dimensional y K un polítopo consistiendo de todos los vectores n -dimensionales \mathbf{x} satisfaciendo $A\mathbf{x} = \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$, entonces un vector \mathbf{x} es un punto extremo de K si, y solo si, \mathbf{x} es una solución básica factible para el sistema de restricciones
 - Suponiendo que $\mathbf{x} = (x_1, x_2, \dots, x_m, 0, 0, \dots, 0)$ es una solución básica factible y $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ son las primeras m columnas de la matriz A y son linealmente independientes, entonces se cumple la siguiente igualdad:

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_m \mathbf{a}_m = \mathbf{b}$$

- Suponiendo que \mathbf{x} se puede expresar como una combinación lineal de dos puntos $\mathbf{x} = \alpha \mathbf{y} + (1 - \alpha) \mathbf{z}$ (en donde $\alpha \in (0,1)$ e $\mathbf{y} \neq \mathbf{z}$), los componentes de \mathbf{x} , \mathbf{y} y \mathbf{z} no son negativos y $\alpha \in (0,1)$, entonces los últimos $n - m$ componentes de \mathbf{y} y \mathbf{z} son cero (dado que los $n - m$ componentes son cero para \mathbf{x}), por lo tanto, se cumplen las siguientes ecuaciones:

$$y_1 \mathbf{a}_1 + y_2 \mathbf{a}_2 + \cdots + y_m \mathbf{a}_m = \mathbf{b}$$

$$z_1 \mathbf{a}_1 + z_2 \mathbf{a}_2 + \cdots + z_m \mathbf{a}_m = \mathbf{b}$$

- Como los vectores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ son linealmente independientes, entonces $\mathbf{x} = \mathbf{y} = \mathbf{z}$ y por tanto \mathbf{x} debe ser un extremo del polígono K (debido a que \mathbf{x} no se puede expresar como combinación lineal de otros dos puntos, al ser iguales)
- De manera converso, si se asume que \mathbf{x} es en punto extremo de K , que los primeros k componentes de \mathbf{x} no son cero y que $x_i > 0$ para $i = 1, 2, \dots, k$, entonces se cumple la siguiente igualdad:

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_k \mathbf{a}_k = \mathbf{b}$$

- Como para que \mathbf{x} sea una solución básica es necesario que los vectores columna $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ sean linealmente independientes, entonces se puede demostrar por contradicción que no hay una combinación lineal de los vectores no trivial que sea igual a cero, suponiendo que los vectores si son linealmente dependientes:

$$\mathbf{y} \equiv (y_1, y_2, \dots, y_k, 0, 0, \dots, 0)$$

$$\text{If lin. dep. then } y_1 \mathbf{a}_1 + y_2 \mathbf{a}_2 + \cdots + y_k \mathbf{a}_k = \mathbf{0}$$

$$\Rightarrow \exists \varepsilon \text{ such that } \mathbf{x} + \varepsilon \mathbf{y} \geq 0 \text{ and } \mathbf{x} - \varepsilon \mathbf{y} \geq 0$$

$$\Rightarrow \mathbf{x} = \frac{1}{2}(\mathbf{x} + \varepsilon \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \varepsilon \mathbf{y}) \geq 0$$

$$\Rightarrow \mathbf{x} \text{ is a lin. comb. of two different vectors}$$

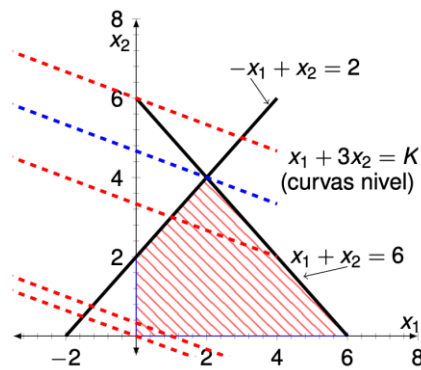
$$\Rightarrow \mathbf{x} \text{ is not extreme point in } K$$

$$\Rightarrow \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \text{ must be lin. indep.}$$

$$\Rightarrow \mathbf{x} \text{ is a basic feas. sol.}$$

- Por lo tanto, la correspondencia entre los extremos del polítopo K y las soluciones básicas factibles permite demostrar ciertas propiedades geométricas de este al definir el conjunto de restricciones en el problema de programación lineal, dando así una versión alternativa del teorema fundamental de la programación lineal
 - Si el conjunto convexo K correspondiente al conjunto de restricciones $Ax = b$ y $x \geq 0$ no es vacío, entonces tiene como mínimo un punto extremo. Esto se puede demostrar a partir de la primera proposición del teorema fundamental y del teorema de equivalencia entre soluciones básicas factibles y puntos extremos
 - Si hay una solución óptima finita para un problema de programación lineal, entonces hay al menos una solución óptima finita que es un punto extremo del conjunto de restricciones K . Esto se puede demostrar a partir de que, como la matriz A es invertible, entonces el sistema $Bx_B = b$ es consistente y tiene una única solución (por lo que el número de soluciones básicas es finito), y los puntos extremos de K son un subconjunto de estas soluciones básicas
 - El conjunto K correspondiente al conjunto de restricciones posee como mucho un número finito de puntos extremos. Esto se puede demostrar a partir de que, como la matriz A es invertible, entonces el sistema $Bx_B = b$ es consistente y tiene una única solución (por lo que el número de soluciones básicas es finito), y los puntos extremos de K son un subconjunto de estas soluciones básicas
 - Si el polítopo convexo K correspondiente al conjunto de restricciones está acotado, entonces K es un poliedro convexo, de modo que K consiste de puntos que son combinaciones convexas de un número finito de otros puntos
- Por lo tanto, se puede dar una versión alternativa del teorema fundamental de la programación lineal. Dado un programa lineal en su forma estándar en donde A es una matriz $m \times n$ de rango m , se dan cumplen las siguientes proposiciones:
 - Si existe una solución factible, también existe una solución básica factible
 - Si existe una solución óptima factible, también existe una solución básica óptima factible

- Las soluciones básicas factibles son los vértices del polítopo convexo (la región factible)
- Por lo tanto, si existe una solución óptima del problema, ésta se encuentra en uno de los puntos extremos del poliedro que forma el sistema de ecuaciones lineales de las restricciones
- En el caso bidimensional (y hasta tridimensional con el *software* apropiado), una manera de obtener las soluciones óptimas es a través del método gráfico
 - Este método se basa en trazar curvas de nivel para diferentes valores de la función objetivo en el gráfico de la región factible hasta encontrar el punto más bajo o alto (dependiendo de si es un problema de minimización o maximización, respectivamente)



- Los ejes de coordenadas del gráfico se corresponden a las variables de decisión del problema, mientras que las rectas (o los planos en tres dimensiones) representan las restricciones expresadas como igualdad
- Para cada restricción, el semiplano que representan la región factible de soluciones se determina a partir del signo de las desigualdades de las restricciones
- La solución gráfica de un problema lineal es útil para entender los fundamentos de la programación lineal
 - Además, permite obtener una intuición de como funciona el método Simplex, que es el método de optimización lineal más utilizado (y disponible para más dimensiones)

El método Simplex: nociones básicas

- La idea del método Simplex es proceder a partir de una solución básica factible (un punto extremos) del conjunto de restricciones de un problema estándar de

programación lineal de modo que se decrezca el valor de la función objetivo hasta que llegue a un mínimo

- Los resultados anteriores expresan que solo es necesario considerar soluciones básicas factibles para encontrar una solución óptima factible
 - Por lo tanto, solo es necesario tener en cuenta los puntos extremos del poliedro formado por las restricciones, ya que estos son soluciones básicas, y si hay un óptimo, estará en los puntos extremos
- El algoritmo del Simplex intenta moverse entre los puntos extremos del poliedro hasta que no haya un punto extremo que optimicé más la función objetivo
- Para poder entender el algoritmo Simplex es necesario entender el concepto de pivotar en un conjunto de ecuaciones lineales simultáneas, habiendo dos interpretaciones para el pivote
 - La primera interpretación considera un sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$ en donde $m \leq n$. En el espacio E^n , la colección de m ecuaciones lineales tiene que satisfacerse por \mathbf{x} , por lo que si \mathbf{a}^i denota la fila i de la matriz A , las ecuaciones se pueden expresar de la siguiente manera:

$$\begin{array}{rcl}
 a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n & = & b_1 \\
 a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n & = & b_2 \\
 \cdot & & \cdot \\
 \cdot & & \cdot \\
 \cdot & & \cdot \\
 a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n & = & b_m
 \end{array}
 \qquad
 \begin{array}{rcl}
 \mathbf{a}^1\mathbf{x} & = & b_1 \\
 \mathbf{a}^2\mathbf{x} & = & b_2 \\
 \cdot & & \cdot \\
 \cdot & & \cdot \\
 \cdot & & \cdot \\
 \mathbf{a}^m\mathbf{x} & = & b_m
 \end{array}$$

- Si $m < n$, entonces el sistema es consistente y las ecuaciones son linealmente independientes, pero no hay una única solución sino infinitas soluciones. Por lo tanto, si se adjuntan $n - m$ ecuaciones linealmente independientes al sistema, entonces $n = m$ y el sistema es consistente con una única solución
- Se puede especificar $n - m$ ecuaciones de la forma $\mathbf{e}^k\mathbf{x} = 0$ (en donde \mathbf{e}^k es un vector unitario k), por lo que $x_k = 0$ y se obtendría una solución básica del sistema de ecuaciones original. Se pueden obtener diferentes soluciones básicas imponiendo ecuaciones de esta forma especial diferentes
- Debido a que las ecuaciones son linealmente independientes, es posible reemplazar una ecuación por un múltiplo de esta sumada a cualquier combinación lineal de otras ecuaciones en el sistema (operaciones elementales de la matriz)

- Por lo tanto, se pueden utilizar los esquemas gaussianos de reducción de la matriz a su forma escalonada reducida. Es fácilmente demostrable, por lo tanto, que el sistema de ecuaciones se puede convertir, a través de operaciones elementales, en el siguiente sistema, llamado sistema en forma canónica:

$$\begin{array}{rcll}
 x_1 & + y_{1,m+1}x_{m+1} + y_{1,m+2}x_{m+2} + \cdots + y_{1,n}x_n & = & y_{10} \\
 x_2 & + y_{2,m+1}x_{m+1} + y_{2,m+2}x_{m+2} + \cdots + y_{2,n}x_n & = & y_{20} \\
 & \cdot & & \cdot \\
 & \cdot & & \cdot \\
 & \cdot & & \cdot \\
 x_m & + y_{m,m+1}x_{m+1} + \cdots + y_{m,n}x_n & = & y_{m0}
 \end{array}$$

- Las variables x_1, x_2, \dots, x_m son las variables básicas y $x_{m+1}, x_{m+2}, \dots, x_n$ son no básicas (de las n existentes en el sistema). Por lo tanto, una solución básica se puede dar como $x = (y_0 | 0)$ donde y_0 es un vector m -dimensional y 0 es un vector de ceros $(n - m)$ -dimensional

$$y_0 = (x_1 = y_{10}, x_2 = y_{20}, \dots, x_m = y_{m0})$$

$$0 = (x_{m+1} = 0, x_{m+2} = 0, \dots, x_n = 0)$$

- Dado un sistema como el visto, es posible saber la forma del sistema cuando se intercambian las variables básicas por las no básicas
- Si se desea reemplazar una variable básica x_p (para $1 \leq p \leq m$) por la variable no básica x_q , entonces esto se logra si, y solo si, y_{pq} no es cero, lo cual se cumple si se divide la fila p por y_{pq} para así obtener un coeficiente unitario para x_q en la ecuación p y sustrayendo un múltiplo adecuado de la fila p para cada una de las filas para obtener un coeficiente de cero para x_q en todas las otras ecuaciones

$$\begin{array}{rcl}
 x_1 & + & x_4 + x_5 - x_6 = 5 \\
 x_2 & + & 2x_4 - 3x_5 + x_6 = 3 \\
 x_3 & - & x_4 + 2x_5 - x_6 = -1
 \end{array}$$

x_1	x_2	x_3	x_4	x_5	x_6	
1	0	0	①	1	-1	5
0	1	0	2	-3	1	3
0	0	1	-1	2	-1	-1

x_1	x_2	x_3	x_4	x_5	x_6	
1	0	0	1	1	-1	5
-2	1	0	0	-5	3	-7
1	0	1	0	3	-2	4

x_1	x_2	x_3	x_4	x_5	x_6	
3/5	1/5	0	1	0	-2/5	18/5
2/5	-1/5	0	0	1	-3/5	7/5
-1/5	3/5	1	0	0	-1/5	-1/5

x_1	x_2	x_3	x_4	x_5	x_6	
1	-1	-2	1	0	0	4
1	-2	-3	0	1	0	2
1	-3	-5	0	0	1	1

$$x_4 = 4, \quad x_5 = 2, \quad x_6 = 1$$

- Esto transforma la columna q en una columna de ceros excepto en la entrada p de la columna, y no afecta las columnas de otras variables básicas. Denotando los coeficientes del nuevo sistema por y'_{ij} , se cumplen las siguientes fórmulas:

$$\begin{cases} y'_{ij} = y_{ij} - \frac{y_{pj}}{y_{pq}} y_{iq} \text{ for } i \neq p \\ y'_{pj} = \frac{y_{pj}}{y_{pq}} \end{cases}$$

- La segunda interpretación se basa en expresar \mathbf{b} como combinación lineal de los vectores columna \mathbf{a}_i de la matriz A , dado que el conjunto de ecuaciones simultáneas se puede interpretar como una ecuación vectorial en E^m

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{b}$$

- Si $m < n$ y los vectores \mathbf{a}_i lapsan E^m , entonces no hay una única solución, si no que hay soluciones infinitas, pero el vector \mathbf{b} tiene una única representación como combinación lineal de un subconjunto de estos vectores que sean linealmente independientes (dado que $B\mathbf{x}_B = \mathbf{b}$ tiene solución única). La solución correspondiente con $n - m$ variables x_i que se igualan a cero será la solución básica
- Si se supone que se tiene un sistema de la forma canónica, se puede formar una tabla de la siguiente forma:

$$\begin{array}{cccccccccc}
1 & 0 & \cdots & 0 & y_{1,m+1} & y_{1,m+2} & \cdots & y_{1n} & y_{10} \\
0 & 1 & \cdot & 0 & y_{2,m+1} & y_{2,m+2} & \cdots & y_{2n} & y_{20} \\
0 & 0 & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & \cdot & 1 & y_{m,m+1} & y_{m,m+2} & \cdots & y_{mn} & y_{m0}
\end{array}$$

- En el caso de los primeros m vectores, estos forman una base, y los otros vectores se pueden expresar como una combinación lineal de los vectores base. De este modo, se puede obtener la siguiente igualdad:

$$\mathbf{a}_j = y_{1,j}\mathbf{a}_1 + y_{2,j}\mathbf{a}_2 + \cdots + y_{m,j}\mathbf{a}_m$$

where j is a column for a non – basic variable

- Por lo tanto, la tabla se puede considerar como una representación de los vectores \mathbf{a}_j en términos de los vectores base (la columna j es la representación del vector \mathbf{a}_j). En particular, la expresión para \mathbf{b} en términos de la base se da en la última columna
- Para cambiar el vector \mathbf{a}_p para $1 \leq p \leq m$ por un vector \mathbf{a}_q , se tienen que aplicar las ecuaciones de pivote, que son las mismas encontradas en la interpretación anterior
 - Dado que los primeros m vectores en los que se reemplaza \mathbf{a}_p por \mathbf{a}_q son linealmente independientes, estos son una base y todos los vectores se pueden expresar en términos de esa base
 - Cualquier vector \mathbf{a}_j puede expresarse en términos de la tabla original a través de la igualdad anterior. Para unos vectores \mathbf{a}_q , \mathbf{a}_p y \mathbf{a}_j se obtienen las siguientes igualdades:

$$\begin{aligned}
\mathbf{a}_q &= \sum_{\substack{i=1 \\ i \neq p}}^m y_{i,q}\mathbf{a}_i + y_{p,q}\mathbf{a}_p \Rightarrow \mathbf{a}_p = \frac{1}{y_{p,q}}\mathbf{a}_q - \sum_{\substack{i=1 \\ i \neq p}}^m \frac{y_{i,q}}{y_{p,q}}\mathbf{a}_i \\
\Rightarrow \mathbf{a}_j &= \sum_{\substack{i=1 \\ i \neq p}}^m \left(y_{i,j} - \frac{y_{i,q}}{y_{p,q}}y_{p,j} \right) \mathbf{a}_i + \frac{y_{p,j}}{y_{p,q}}\mathbf{a}_q
\end{aligned}$$

- A partir de los coeficientes de la última igualdad para \mathbf{a}_j , se pueden extraer los coeficientes de la nueva tabla, los cuales son fórmulas idénticas que en la interpretación anterior

$$\begin{cases} y'_{ij} = y_{ij} - \frac{y_{pj}}{y_{pq}} y_{iq} \text{ for } i \neq p \\ y'_{pj} = \frac{y_{pj}}{y_{pq}} \end{cases}$$

- Con esta interpretación, se pueden realizar las mismas operaciones que con la interpretación anterior para poder conseguir intercambiar los vectores base por los vectores columna
- Anteriormente se ha visto que solo se necesita considerar las soluciones básicas factibles del conjunto de soluciones formado por $A\mathbf{x} = \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$ y que la operación de pivote puede generar nuevas soluciones básicas a partir de soluciones viejas intercambiando una variable básica por una no básica. No obstante, se tienen que cumplir ciertas condiciones y hay resultados útiles que simplifican los cálculos
 - Cualquier solución básica factible del conjunto de soluciones formado por $A\mathbf{x} = \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$ es una solución básica factible no degenerada
 - Esta suposición simplifica sustancialmente muchos argumentos de la programación lineal, y se usa en el método Simplex para que el método siempre funcione (a no ser que se modifique adecuadamente para estos casos)
 - Suponiendo que se tiene una solución básica factible $\mathbf{x} = (x_1, x_2, \dots, x_m, 0, 0, \dots, 0)$, lo cual hace que se cumpla la ecuación $x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_m \mathbf{a}_m = \mathbf{b}$, la suposición de no degeneración hace que $x_i > 0$ para toda $i = 1, 2, \dots, m$
 - Suponiendo, además, que se puede hacer una representación del vector \mathbf{a}_q como $\mathbf{a}_q = y_{1q} \mathbf{a}_1 + y_{2q} \mathbf{a}_2 + \dots + y_{mq} \mathbf{a}_m$, entonces se puede multiplicar \mathbf{a}_q por $\varepsilon \geq 0$ y restándole el producto a \mathbf{b} se obtiene una combinación lineal de $m + 1$ vectores con $\varepsilon \geq 0$ para \mathbf{b}

$$(x_1 - \varepsilon y_{1q}) \mathbf{a}_1 + (x_2 - \varepsilon y_{2q}) \mathbf{a}_2 + \dots + (x_m - \varepsilon y_{mq}) \mathbf{a}_m + \varepsilon \mathbf{a}_q = \mathbf{b}$$
 - Cuando $\varepsilon = 0$, se obtiene la solución básica factible anterior. Cuando ε incrementa, el coeficiente de \mathbf{a}_q incrementa, y para una ε lo suficientemente pequeña, la ecuación da una ecuación factible pero no básica

- Los coeficientes de los otros vectores incrementarán o decrecerán linealmente cuando ε incremente. Si hay algún decrecimiento, se podría fijar ε igual al valor correspondiente al primer lugar donde uno o más coeficientes desaparecen

$$\varepsilon = \min_i \{x_i/y_{iq} : y_{iq} > 0\}$$

- En este caso se tendría una nueva solución factible básica con el vector \mathbf{a}_q reemplazando al vector \mathbf{a}_p donde p corresponde al índice minimizador en $\varepsilon = \min_i \{x_i/y_{iq} : y_{iq} > 0\}$. Si el mínimo en $\varepsilon = \min_i \{x_i/y_{iq} : y_{iq} > 0\}$ se consigue en más de un solo índice i , entonces la nueva solución es degenerada y cualquiera de los vectores con componente cero puede ser extraído de la base
 - Si ninguna de las y_{iq} son positivas, entonces todos los coeficientes en $(x_1 - \varepsilon y_{1q})\mathbf{a}_1 + \dots + \varepsilon \mathbf{a}_q = \mathbf{b}$ incrementarán (o se mantendrán constantes) cuando ε incremente, y no se puede obtener una solución básica factible. No obstante, en este caso hay soluciones factibles para las restricciones con coeficientes arbitrariamente grandes, por lo que la región no está acotada
 - En conclusión, se ha deducido que dada una solución básica factible y un vector arbitrario \mathbf{a}_q , hay una solución básica factible nueva teniendo \mathbf{a}_q en su base y uno de los vectores originales sustituido, o el conjunto de soluciones factibles no está acotado
- Se puede considerar el cálculo descrito anteriormente en términos de una tabla. Asumiendo que se tienen unas restricciones $A\mathbf{x} = \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$ se tiene una tabla de la siguiente forma:

\mathbf{a}_1	\mathbf{a}_2	\mathbf{a}_3	\dots	\mathbf{a}_m	\mathbf{a}_{m+1}	\mathbf{a}_{m+2}	\dots	\mathbf{a}_n	\mathbf{b}
1	0	0	\dots	0	$y_{1,m+1}$	$y_{1,m+2}$	\dots	y_{1n}	y_{10}
0	1	0		0	$y_{2,m+1}$	$y_{2,m+2}$		\cdot	y_{20}
0	0	1		\cdot	\cdot	\cdot		\cdot	\cdot
\cdot	\cdot	\cdot		\cdot	\cdot	\cdot		\cdot	\cdot
\cdot	\cdot	\cdot		\cdot	\cdot	\cdot		\cdot	\cdot
\cdot	\cdot	\cdot		\cdot	\cdot	\cdot		\cdot	\cdot
0	0	\cdot		1	$y_{m,m+1}$	$y_{m,m+2}$	\dots	y_{mn}	y_{m0}

- Esta tabla puede resultar de diversas operaciones de pivote aplicadas a la tabla original, pero de todos modos esta representa una solución con base $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$. Asumiendo que $y_{10}, y_{20}, \dots, y_{m0}$ no son negativos, entonces la solución básica correspondiente $y_{10} = x_1, \dots, y_{m0} = x_m$ es factible

- Si se quiere usar \mathbf{a}_q (para $q > m$) en la base y mantener la factibilidad, se tiene que determinar que elemento de la columna q se utilizará como pivote (de manera que se puede determinar que vector en la base se sustituirá) utilizando $\varepsilon = \min_i \{x_i/y_{iq} : y_{iq} > 0\}$ y calculando las *ratios* $x_i/y_{iq} = y_{i0}/y_{iq}$ para $i = 1, 2, \dots, m$. Se selecciona la *ratio* no negativa más pequeña y se pivota en la y_{iq} correspondiente

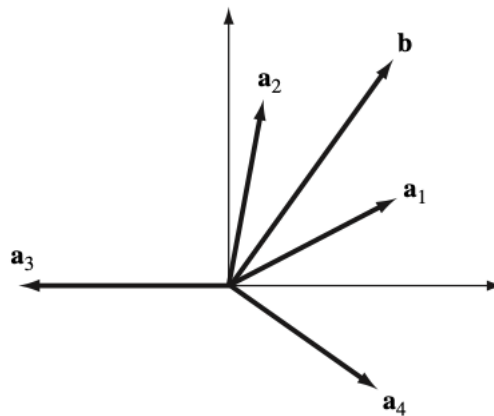
\mathbf{a}_1	\mathbf{a}_2	\mathbf{a}_3	\mathbf{a}_4	\mathbf{a}_5	\mathbf{a}_6	\mathbf{b}	
1	0	0	2	4	6	4	
0	1	0	1	2	3	3	
0	0	1	-1	2	1	1	$\mathbf{x} = (4, 3, 1, 0, 0, 0)$

$$4/2 = 2, \quad 3/1 = 3, \quad 1/-1 = -1$$

\mathbf{a}_1	\mathbf{a}_2	\mathbf{a}_3	\mathbf{a}_4	\mathbf{a}_5	\mathbf{a}_6	\mathbf{b}	
1/2	0	0	1	2	3	2	
-1/2	1	0	0	0	0	1	
1/2	0	1	0	4	4	3	$\mathbf{x} = (0, 1, 3, 2, 0, 0)$

- De manera alternativa, se puede justificar el método de selección del pivote a través de un enfoque dual que está basado en las filas de la tabla más que en sus columnas
 - Si se decide pivotar en y_{pq} , primero se divide la fila p por el elemento de pivote y_{pq} para hacer que este elemento sea la unidad. Con tal de que y_{p0} se mantenga positivo, se tiene que tener $y_{pq} > 0$
 - Después, se sustraen los múltiplos de la fila p a todas las otras filas con tal de obtener ceros en la columna q . En el proceso, los nuevos elementos en la última columna tienen que mantenerse no negativos si el pivote se selecciona correctamente
 - La operación entero es restar a la fila i la fila p una cantidad de y_{iq}/y_{pq} veces, lo cual permite obtener una nueva solución de la siguiente fórmula:
 - Como esta no puede ser negativa, $x_p/y_{pq} \leq x_i/y_{iq}$ y, por lo tanto, se escoge la p como el índice i que minimiza x_i/y_{iq}
- A partir de las dos interpretaciones del pivotaje y de los puntos extremos (desarrolladas algebraicamente) se pueden dar dos interpretaciones geométricas

- La primera es el espacio de actividad, definido como el espacio en donde se representa x (el cual es el más natural a considerar). En este espacio, la región factible se muestra directamente como un conjunto convexo y las soluciones básicas factibles son los puntos extremos
- Se define como puntos extremos adyacentes los puntos que están en un borde común, de modo que se comparten todas las x_B y todas las $x_N = 0$ menos una de sus variables. Para moverse de un punto extremo a otro, solo hace falta intercambiar una variable básica por una no básica
- La segunda es el espacio de requerimientos, definido como el espacio donde se representa A y b



El método Simplex: algoritmo matricial y tabular

- Aunque las transformaciones de pivote asociadas con el método Simplex son más fáciles de visualizar en formato de tabla, dado que se pone atención en los elementos individuales, es mucho mejor utilizar la interpretación matricial
 - Las relaciones vector-matriz que existen en las filas y columnas de la tabla permiten incrementar el entendimiento del método y a un método mucho más ventajoso en términos computacionales
 - El método se basa en utilizar nociones de vectores y matrices con tal de aplicar el método Simplex de manera mucho más sencilla, dado que se ignoran cálculos que no sean estrictamente necesario
 - A partir del problema de minimización estándar, se pueden expresar los siguientes elementos como matrices y vectores partidos (se adjuntan matrices y vectores)

$$\mathbf{A} = [\mathbf{B}, \mathbf{D}]$$

$$\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_D), \quad \mathbf{c}^T = [\mathbf{c}_B^T, \mathbf{c}_D^T]$$

- Por lo tanto, el programa lineal se puede expresar de la siguiente manera:

$$\begin{aligned} &\text{minimize} \quad \mathbf{c}_B^T \mathbf{x}_B + \mathbf{c}_D^T \mathbf{x}_D \\ &\text{subject to} \quad \mathbf{B}\mathbf{x}_B + \mathbf{D}\mathbf{x}_D = \mathbf{b} \\ &\quad \mathbf{x}_B \geq \mathbf{0}, \quad \mathbf{x}_D \geq \mathbf{0}. \end{aligned}$$

- La solución básica (que se asume que también es factible) correspondiente a la base B es $\mathbf{x} = (\mathbf{x}_B, \mathbf{0})$, en donde $\mathbf{x}_B = B^{-1}\mathbf{b}$ y $\mathbf{x}_D = \mathbf{0}$

- Para cualquier valor \mathbf{x}_D el valor necesario de \mathbf{x}_B se puede expresar en términos de $\mathbf{x}_D = \mathbf{0}$ de la siguiente manera:

$$\mathbf{x}_B = B^{-1}\mathbf{b} - B^{-1}\mathbf{D}\mathbf{x}_D$$

- Esta expresión se puede sustituir en la función de costes para obtener la siguiente expresión:

$$\begin{aligned} z &= \mathbf{c}_B'(B^{-1}\mathbf{b} - B^{-1}\mathbf{D}\mathbf{x}_D) + \mathbf{c}_D'\mathbf{x}_D = \\ &= \mathbf{c}_B'B^{-1}\mathbf{b} - \mathbf{c}_B'B^{-1}\mathbf{D}\mathbf{x}_D + \mathbf{c}_D'\mathbf{x}_D = \\ &= \mathbf{c}_B'B^{-1}\mathbf{b} + (\mathbf{c}_D' - \mathbf{c}_B'B^{-1}\mathbf{D})\mathbf{x}_D \end{aligned}$$

- Debido a que esta última expresión permite expresar la función de costes en términos de las variables no básicas \mathbf{x}_D , entonces el vector de coste relativo \mathbf{r}_D' es el siguiente:

$$\mathbf{r}_D' = \mathbf{c}_D' - \mathbf{c}_B'B^{-1}\mathbf{D}$$

- Este vector, en consecuencia, expresa el cambio en la función objetivo cuando se incrementa (o se disminuye) una unidad de una de las variables no básicas
- Esto hace ver que, en un problema de minimización, si algún componente de \mathbf{r}_D' es negativo, es más óptimo incluir este en la base (aumentar su valor para que no sea nulo), mientras que, si es positivo, lo mejor es que se quede fuera (dado que aumentaría el valor de la función objetivo y lo que se quiere es minimizar)

- Si se está tratando con un problema de maximización, la lógica se invierte, dado que se intenta aumentar lo máximo el valor de esta función objetivo
- El algoritmo del método Simplex en forma matricial contiene los siguientes pasos:
 - El primer paso consiste en expresar la tabla del Simplex de forma canónica a partir de la tabla que se puede formar con A , \mathbf{b} y con los coeficientes en la función objetivo \mathbf{c}'

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{c}^T & 0 \end{array} \right] = \left[\begin{array}{c|c|c} \mathbf{B} & \mathbf{D} & \mathbf{b} \\ \hline \mathbf{c}_B^T & \mathbf{c}_D^T & 0 \end{array} \right]$$

\mathbf{a}_1	\mathbf{a}_2	\mathbf{a}_3	\mathbf{a}_4	\mathbf{a}_5	\mathbf{a}_6	\mathbf{b}
2	1	1	1	0	0	2
1	2	3	0	1	0	5
2	2	1	0	0	1	6

$$\mathbf{c}^T = [-3, -1, -3, 0, 0, 0].$$

- Debido a que la primera tabla no está en la forma canónica deseada, se puede transformar en la siguiente tabla en forma canónica:

$$\mathbf{T} = \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{B}^{-1}\mathbf{D} & \mathbf{B}^{-1}\mathbf{b} \\ \hline \mathbf{0} & \mathbf{c}_D^T - \mathbf{c}_B^T\mathbf{B}^{-1}\mathbf{D} & -\mathbf{c}_B^T\mathbf{B}^{-1}\mathbf{b} \end{array} \right]$$

- En este caso, la parte inferior de esta tabla es una fila compuesta por los elementos en $\mathbf{0}$, en $\mathbf{c}_D^T - \mathbf{c}_B^T\mathbf{B}^{-1}\mathbf{D}$ y en $-\mathbf{c}_B^T\mathbf{B}^{-1}\mathbf{b}$
- La única complicación en el método es encontrar la inversa de la matriz B , pero esta se puede encontrar fácilmente adjuntando a la derecha una matriz I del mismo tamaño e intentando hacer que la matriz B a la izquierda se transforme en la matriz identidad a través de operaciones elementales de fila aplicadas en ambas matrices a la vez para transformarla primero en su forma de fila escalonada y después en la matriz identidad

$$[B|I] \Rightarrow [I|B^{-1}]$$

- También se puede encontrar por expansión de cofactores, calculando el determinante de los menores M_{ij} , sus respectivos cofactores C_{ij} , la matriz adjunta $adj(B)$ y dividiéndola por el determinante de B

$$C_{ij} = (-1)^{i+j} \det(M_{ij}) \Rightarrow [C_{ij}]' = adj(B)$$

$$\Rightarrow B^{-1} = \frac{1}{\det(B)} adj(B)$$

- Una vez con la tabla en forma canónica, el segundo paso consiste en fijar como solución corriente $\mathbf{y}_0 = \mathbf{x}_B = B^{-1}\mathbf{b}$ y calcular los coeficientes del coste relativo $\mathbf{r}'_D = \mathbf{c}'_D - \mathbf{c}'_B B^{-1}D$

Variable	\mathbf{B}^{-1}			\mathbf{x}_B
4	1	0	0	2
5	0	1	0	5
6	0	0	1	6

- Esto se puede hacer fácilmente calculando primero $\boldsymbol{\lambda}' = \mathbf{c}'_B B^{-1}$ y después $\mathbf{r}'_D = \mathbf{c}'_D - \boldsymbol{\lambda}' D$

$$\boldsymbol{\lambda}' = [0, 0, 0] \mathbf{B}^{-1} = [0, 0, 0]$$

$$\mathbf{r}'_D = \mathbf{c}'_D - \boldsymbol{\lambda}' D = [-3, -1, -3]$$

- Si $\mathbf{r}'_D \geq \mathbf{0}$, entonces el algoritmo se detiene y la solución corriente \mathbf{y}_0 es la óptima. De no ser así, el algoritmo sigue al siguiente paso
- El tercer paso consiste en determinar que vector \mathbf{a}_q de la matriz A (qué vector columna de coeficientes de una variable) entrará a la base
 - Para ello, se selecciona la variable q o el vector \mathbf{a}_q con el coeficiente c_q más negativo dentro del vector \mathbf{c}_B
 - Una vez escogido el vector, se calcula $\mathbf{y}_q = B^{-1}\mathbf{a}_q$, el cual será el vector \mathbf{a}_q expresado en términos de la base actual (antes de introducir el nuevo vector \mathbf{a}_q)

Variable	\mathbf{B}^{-1}			\mathbf{x}_B	\mathbf{y}_2
4	1	0	0	2	①
5	0	1	0	5	2
6	0	0	1	6	2

- Cada fila del vector columna \mathbf{y}_q representa una variable básica \mathbf{x}_B de la base
- En el cuarto paso, se calculan las *ratios* $x_{B,i}/y_{iq}$ (donde i indica la fila o la variable básica en el vector \mathbf{y}_q) para toda $y_{iq} > 0$ y se determina qué vector columna o variable básica deja la base
 - Si $y_{iq} \leq 0$ para toda i , entonces el problema no está acotado (se puede reducir el valor de la función objetivo infinitamente) y el algoritmo Simplex se detiene. Esto ocurre porque si $r'_q < 0$ pero $y_{iq} \leq 0$, entonces el coste relativo de aumentar la variable q siempre será negativo o nulo, y por tanto se puede reducir la función objetivo de manera indefinida aumentando el valor de esa variable
 - En el caso de un problema no acotado, entonces se puede expresar la solución como $\mathbf{x} + \alpha \mathbf{d}$, donde \mathbf{d} es un vector de la dirección en la que la función objetivo se reduce indefinidamente
 - En este caso, el criterio se basa en escoger la variable o vector columna \mathbf{a}_p de la base \mathbf{B} tal que su *ratio* $x_{B,i}/y_{iq}$ sea la mínima

$$\min_i \{x_{B,i}/y_{iq} : y_{iq} > 0\}$$

- Finalmente, el quinto paso consiste en actualizar la base B con el nuevo vector columna \mathbf{a}_q o la nueva variable y calcular la nueva inversa B^{-1} con tal de actualizar la solución a $\mathbf{y}_1 = \mathbf{x}_B = B^{-1}\mathbf{b}$

Variable	\mathbf{B}^{-1}				\mathbf{x}_B
2	1	0	0	0	2
5	-2	1	0	0	1
6	-2	0	1	0	2

- Para poder actualizar B^{-1} se realiza el mismo procedimiento de adjuntar matrices, pero con la nueva B que incluye \mathbf{a}_q (transformándola primero en su forma de fila escalonada y después en la matriz identidad) o por expansión de cofactores (calculando el determinante de los menores, sus respectivos cofactores, la matriz adjunta del mismo tamaño formada por los cofactores y dividiendo esta matriz adjunta por el determinante de B)

$$[B|I] \Rightarrow [I|B^{-1}]$$

$$C_{ij} = (-1)^{i+j} \det(M_{ij}) \Rightarrow [C_{ij}]' = \text{adj}(B)$$

$$\Rightarrow B^{-1} = \frac{1}{\det(B)} \text{adj}(B)$$

- También es necesario actualizar D , c'_B y c'_D con tal de poder realizar los cálculos correctamente
- Una vez hecho esto, se vuelve al segundo paso y se aplica el algoritmo hasta el quinto paso. El algoritmo, entonces, se detendrá en el segundo paso una vez se tenga la solución óptima

$$\lambda^T = [-1, 0, 0] \mathbf{B}^{-1} = [-1, 0, 0]$$

$$r_1 = -1, \quad r_3 = -2, \quad r_4 = 1.$$

Variable	\mathbf{B}^{-1}			\mathbf{x}_B	\mathbf{y}_3
2	1	0	0	2	1
5	-2	1	0	1	①
6	-2	0	1	2	-1

Variable	\mathbf{B}^{-1}			\mathbf{x}_B
2	3	-1	0	1
3	-2	1	0	1
6	-4	1	1	3

- Si no se puede obtener una solución básica factible inicial de forma inmediata o trivial (es decir, no es sencillo ver qué vectores columna son linealmente independientes), es posible utilizar variables artificiales con tal de encontrar una solución básica factible inmediata
 - Una solución básica factible normalmente es inmediata para programas lineales con desigualdades \leq en las restricciones y con $\mathbf{b} \geq \mathbf{0}$, dado que se pueden utilizar variables de holgura para poder convertirlas en restricciones de igualdad

$$\begin{array}{l} Ax \leq \mathbf{b} \\ x \geq \mathbf{0} \end{array} \Rightarrow \begin{array}{l} Ax = \mathbf{b} \\ x \geq \mathbf{0} \end{array}$$

- No obstante, en otras condiciones no siempre es fácil obtener una solución, por lo que se necesitan métodos para poder inicializar el procedimiento Simplex

- Debido a que a través de los procedimientos anteriormente vistos se puede transformar cualquier $Ax \leq b$ en $Ax = b$ cuando $b \geq 0$, se puede considerar un programa lineal auxiliar que se basa en introducir variables artificiales a_i para $i = 1, 2, \dots, m$ al problema original y resolver un problema de la siguiente forma:

$$\begin{array}{ll} \min c'x & \\ \text{s. t. } Ax = b & \\ x \geq 0 & \end{array} \Rightarrow \begin{array}{ll} \min \sum_{i=1}^m a_i & \\ \text{s. t. } Ax + a = b & \\ x \geq 0, a \geq 0 & \end{array}$$

- El vector a es un vector de m variables artificiales, una para cada restricción, las cuales tienen que ser no negativas

$$a' = (a_1, a_2, \dots, a_m)$$

- Si hay una solución factible en el problema original, el problema auxiliar tiene un valor mínimo de cero con $a = 0$ (porque $Ax = b$ y se hace que $a = 0$ necesariamente). Si no hay una solución factible, entonces el valor mínimo del problema auxiliar será mayor a cero
- Debido a que $b \geq 0$, la primera solución básica factible es $a = b$ del programa auxiliar, por lo que se puede aplicar el método anterior e ir cambiando de puntos extremos hasta encontrar el óptimo
- Si el valor mínimo de $\sum_{i=1}^m a_i$ es cero, entonces la solución básica final tendrá que $a = 0$ y se pueden eliminar estas variables de la solución. Pero si alguna $a_i > 0$, entonces el problema original no es factible
- No obstante, no es necesario añadir variables artificiales para cada restricción, sino que solo se añaden las variables artificiales a aquellas restricciones específicas con tal de obtener una solución trivial
 - Por lo tanto, en la función objetivo no se suman tantas variables artificiales como restricciones haya, sino que solo se suman tantas como sea necesario para encontrar una solución básica factible
 - No solo basta con añadir la variable, sino que también se tiene que hacer una transformación de la restricción (multiplicando o dividiendo por un número)
- A partir de esto, se puede resolver un problema lineal general usando un método de dos fases: la primera fase, en donde se introducen las variables artificiales, y la segunda fase, en donde se usa la solución básica

factible encontrada para poder resolver el problema original con el algoritmo

- En la primera fase, se añaden las variables artificiales y se resuelve el problema con el mismo procedimiento anteriormente descrito hasta encontrar una solución básica factible $\mathbf{x}_B \geq \mathbf{0}$ que no contenga las variables artificiales (que no estén en la base). Por lo tanto, se usa el procedimiento visto hasta que eso ocurra
- En la segunda fase, se utiliza esta solución básica factible $\mathbf{x} = [\mathbf{x}_B | \mathbf{x}_D]$ para comenzar el procedimiento anterior, pero omitiendo las variables artificiales y volviendo a considerar la función objetivo original (no aquella con las variables artificiales)
- Aunque el algoritmo para matrices es intuitivo, conlleva bastantes cálculos y álgebra lineal para su obtención manual, de modo que se puede seguir el siguiente algoritmo basado en una tabla de forma canónica
 - El primer paso consiste en expresar la tabla del Simplex de forma canónica a partir de la tabla que se puede formar con A , \mathbf{b} y con los coeficientes en la función objetivo \mathbf{c}'

\mathbf{a}_{m+1}	\mathbf{a}_{m+2}	\cdots	\mathbf{a}_j	\cdots	\mathbf{a}_n	\mathbf{a}_1	\mathbf{a}_2	\cdots	\mathbf{a}_m	\mathbf{b}
$y_{1,m+1}$	$y_{1,m+2}$	\cdots	y_{1j}	\cdots	y_{1n}	1	0	\cdots	0	y_{10}
\cdot	\cdot		\cdot		\cdot	0	1		\cdot	\cdot
\cdot	\cdot		\cdot		\cdot	\cdot	\cdot		\cdot	\cdot
\cdot	\cdot		\cdot		\cdot	\cdot	\cdot		\cdot	\cdot
$y_{i,m+1}$	$y_{i,m+2}$	\cdots	y_{ij}	\cdots	y_{in}	0	0		\cdot	y_{i0}
\cdot	\cdot		\cdot		\cdot	\cdot	\cdot		\cdot	\cdot
\cdot	\cdot		\cdot		\cdot	\cdot	\cdot		\cdot	\cdot
$y_{m,m+1}$	$y_{m,m+2}$	\cdots	y_{mj}	\cdots	y_{mn}	0	0		1	y_{m0}
r_{m+1}	r_{m+2}	\cdots	r_j	\cdots	r_n	0	0	\cdots	0	$-z_0$

- Esta tabla es equivalente a la tabla propuesta en notación matricial, aunque el orden se cambie entre la matriz básica y la no básica
- Una vez con la tabla en forma canónica, el segundo paso consiste en fijar como solución corriente $\mathbf{y}_0 = \mathbf{x}_B = (x_1 = y_{10}, x_2 = y_{20}, \dots, x_m = y_{m0})'$ y escoger la columna o variable no básica que entra en la base y aquella que sale

	\mathbf{a}_1	\mathbf{a}_2	\mathbf{a}_3	\mathbf{a}_4	\mathbf{a}_5	\mathbf{a}_6	\mathbf{b}
	②	1	1	1	0	0	2
	1	2	③	0	1	0	5
	2	2	1	0	0	1	5
\mathbf{r}^T	-3	-1	-3	0	0	0	0

- Para poder determinar que variable o columna se introducirá en la base, se escoge el coste relativo más negativo para las variables no básicas (se selecciona la variable q o el vector \mathbf{a}_q con el coeficiente c_q más negativo dentro del vector)
- Se calculan las *ratios* $x_{B,i}/y_{iq}$ (donde i indica la fila o la variable básica en el vector \mathbf{a}_q) para toda $y_{iq} > 0$ y se determina qué vector columna o variable básica deja la base
- Si $y_{iq} \leq 0$ para toda i , entonces el problema no está acotado (se puede reducir el valor de la función objetivo infinitamente) y el algoritmo Simplex se detiene. Esto ocurre porque si $r'_q < 0$ pero $y_{iq} \leq 0$, entonces el coste relativo de aumentar la variable q siempre será negativo o nulo, y por tanto se puede reducir la función objetivo de manera indefinida aumentando el valor de esa variable
- En el caso de un problema no acotado, entonces se puede expresar la solución como $\mathbf{x} + \alpha \mathbf{d}$, donde \mathbf{d} es un vector de la dirección en la que la función objetivo se reduce indefinidamente
- En este caso, el criterio se basa en escoger la variable o vector columna de la base \mathbf{B} tal que su *ratio* $x_{B,i}/y_{iq}$ sea la mínima

$$\min_i \{x_{B,i}/y_{iq} : y_{iq} > 0\}$$

- Finalmente, el tercer paso consiste en actualizar la base B con el nuevo vector columna \mathbf{a}_q o la nueva variable
 - Para poder actualizar la matriz de variables básicas (la base), primero se selecciona un valor de pivote y_{pq} , el cual será aquel valor en la columna q que entraba en la base (representando \mathbf{a}_q) y en la fila p que sale de la base representando (representando \mathbf{a}_p)

	\mathbf{a}_1	\mathbf{a}_2	\mathbf{a}_3	\mathbf{a}_4	\mathbf{a}_5	\mathbf{a}_6	\mathbf{b}
	②	1	1	1	0	0	2
	1	2	③	0	1	0	5
	2	2	1	0	0	1	5
\mathbf{r}^T	-3	-1	-3	0	0	0	0

- Una vez se ha escogido la cantidad de pivote, entonces se aplican las fórmulas de pivote para cada una de las filas. Por lo tanto, se aplica la primera fórmula para los coeficientes de todas las filas i

diferentes a p y la segunda fórmula solo se aplica a la fila de pivote p

$$\begin{cases} y'_{ij} = y_{ij} - \frac{y_{pj}}{y_{pq}} y_{iq} & \text{for } i \neq p \\ y'_{pq} = \frac{y_{pj}}{y_{pq}} & \text{for } i = p \end{cases}$$

- Una vez hecho esto, se vuelve al segundo paso y se aplica el algoritmo hasta el segundo paso. El algoritmo, entonces, se detendrá en el segundo paso una vez se tenga la solución óptima

$$\begin{array}{ccccccc} 2 & 1 & 1 & 1 & 0 & 0 & 2 \\ -3 & 0 & \textcircled{1} & -2 & 1 & 0 & 1 \\ -2 & 0 & -1 & -2 & 0 & 1 & 2 \\ -1 & 0 & -2 & 1 & 0 & 0 & 2 \end{array}$$

$$\begin{array}{ccccccc} \textcircled{5} & 1 & 0 & 3 & -1 & 0 & 1 \\ -3 & 0 & 1 & -2 & 1 & 0 & 1 \\ -5 & 0 & 0 & -4 & 1 & 1 & 3 \\ -7 & 0 & 0 & -3 & 2 & 0 & 4 \end{array}$$

$$\begin{array}{ccccccc} 1 & 1/5 & 0 & 3/5 & -1/5 & 0 & 1/5 \\ 0 & 3/5 & 1 & -1/5 & 2/5 & 0 & 8/5 \\ 0 & 1 & 0 & -1 & 0 & 1 & 4 \\ 0 & 7/5 & 0 & 6/5 & 3/5 & 0 & 27/5 \end{array}$$

- El método de dos fases del algoritmo de Simplex visto anteriormente también se puede aplicar utilizando el algoritmo tabular mostrado

La dualidad de los programas lineales

- Para cada programa lineal existe un programa lineal íntimamente relacionado llamado programa lineal dual. Ambos se construyen a partir del mismo coste subyacente y de los mismos coeficientes de las restricciones, pero de una manera diferente
 - Un programa lineal se denomina programa lineal primal si este es un programa expresado de la manera estándar, mientras que el programa lineal dual es un programa lineal de la siguiente forma:

	<i>Primal</i>		<i>Dual</i>
minimize	$\mathbf{c}^T \mathbf{x}$	maximize	$\boldsymbol{\lambda}^T \mathbf{b}$
subject to	$\mathbf{Ax} \geq \mathbf{b}$	subject to	$\boldsymbol{\lambda}^T \mathbf{A} \leq \mathbf{c}^T$
	$\mathbf{x} \geq \mathbf{0}$		$\boldsymbol{\lambda} \geq \mathbf{0}$

- En este caso A es una matriz $m \times n$, entonces \mathbf{x} y \mathbf{b} son vectores columna n -dimensional, \mathbf{c}' es un vector fila n -dimensional y $\boldsymbol{\lambda}'$ es un vector fila m -dimensional. Por lo tanto, \mathbf{x} es el vector de variables en el programa primal y $\boldsymbol{\lambda}$ el vector de variables en el programa dual
 - Este par de programas se denomina forma simétrica de dualidad, y se puede utilizar para definir el dual en un programa lineal. No obstante, también hay otras formas para definir el dual correspondiente a un primal diferente
- El rol del programa primal y del programa dual se pueden invertir, de modo que hace falta estudiar el proceso por el que se obtiene el dual a partir del primal
- Dado un programa lineal primal en su forma estándar, el programa dual se obtiene intercambiando el vector de restricciones \mathbf{b} por el de costes \mathbf{c} , transponiendo la matriz de coeficientes A , revertir la dirección de las restricciones y cambiando la minimización por la maximización

$$(1) \min \mathbf{c}'\mathbf{x} \text{ is changed by } \max \boldsymbol{\lambda}'\mathbf{b}$$

$$(2) \mathbf{Ax} \text{ is changed by } \boldsymbol{\lambda}'\mathbf{A}$$

$$(3) \mathbf{Ax} \geq \mathbf{b} \text{ is changed by } \boldsymbol{\lambda}'\mathbf{A} \leq \mathbf{c}'$$

- Dependiendo de cómo son las restricciones, se añade una u otra restricción de positividad, negatividad o ninguna. La regla es que la restricción sobre el valor de las variables depende de la igualdad o desigualdad en las restricciones del poliedro

$$\begin{array}{ll} \min \mathbf{c}'\mathbf{x} & \\ \text{if } \text{s.t. } \mathbf{Ax} = \mathbf{b} & \text{then } \max \boldsymbol{\lambda}'\mathbf{b} \\ \mathbf{x} \geq \mathbf{0} & \text{s.t. } \boldsymbol{\lambda}'\mathbf{A} \leq \mathbf{c}' \end{array}$$

(sym. form)

$$\begin{array}{ll} \min \mathbf{c}'\mathbf{x} & \\ \text{if } \text{s.t. } \mathbf{Ax} \geq \mathbf{b} & \text{then } \max \boldsymbol{\lambda}'\mathbf{b} \\ \mathbf{x} \geq \mathbf{0} & \text{s.t. } \boldsymbol{\lambda}'\mathbf{A} \leq \mathbf{c}' \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

(antisym. form)

$$\begin{array}{ll} \min \mathbf{c}'\mathbf{x} & \max \mathbf{\lambda}'\mathbf{b} \\ \text{if } \text{s.t. } \mathbf{Ax} \leq \mathbf{b} & \text{then } \text{s.t. } \mathbf{\lambda}'\mathbf{A} \leq \mathbf{c}' \\ \mathbf{x} \geq \mathbf{0} & \mathbf{\lambda} \leq \mathbf{0} \end{array}$$

- Si algunas de las variables \mathbf{x} en el programa primal son libres, las restricciones en el programa dual se vuelven $\mathbf{\lambda}'\mathbf{A} = \mathbf{c}'$
- Todas estas reglas de transformación en un programa dual no son aleatorias, sino que son consecuencia original de la definición original
- Aunque la relación entre el problema lineal primal y el problema lineal dual parece una relación formal basada en una definición arbitraria, el teorema de dualidad permite hacer una conexión más profunda
 - Para poder obtener una cota inferior cualquiera para un problema primal de la forma estándar $\min\{\mathbf{c}'\mathbf{x} : \mathbf{Ax} = \mathbf{b} \text{ \& } \mathbf{x} \geq \mathbf{0}\}$, se puede hacer una combinación lineal de restricciones que sea menor a la función objetivo
 - Haciendo una combinación lineal de las restricciones, es posible obtener una expresión que sea menor que la función objetivo del problema primal

$$\sum_i \lambda_i (\mathbf{a}'_i \mathbf{x}) = \mathbf{\lambda}' \mathbf{Ax} \leq \mathbf{c}' \mathbf{x}$$

- Debido a que $\mathbf{a}'_i \mathbf{x} = b_i$, se puede hacer una sustitución y así encontrar una constante que sea menor o igual a la función objetivo $\mathbf{c}'\mathbf{x}$, siendo así una cota inferior para la misma

$$\sum_i \lambda_i (\mathbf{a}'_i \mathbf{x}) = \sum_i \lambda_i b_i = \mathbf{\lambda}' \mathbf{b} \leq \mathbf{c}' \mathbf{x}$$

- Este método permite obtener una cota inferior cualquiera a través de escoger valores de $\mathbf{\lambda}$ arbitrariamente para que se cumpla la desigualdad. No obstante, no tiene por qué ser la cota inferior más alta (la más cercana al óptimo), de modo que se puede plantear un problema de optimización para $\mathbf{\lambda}'\mathbf{b}$ restringido a que se cumpla la desigualdad $\mathbf{\lambda}'\mathbf{A} \leq \mathbf{c}'$ (motivando así la definición del problema dual lineal)

- El lema de dualidad débil expresa que si x y λ son factibles para $\min\{c'x : Ax = b \text{ \& } x \geq 0\}$ y también para $\max\{\lambda'b : \lambda'A \leq c'\}$, respectivamente, entonces $c'x \geq \lambda'b$

- Si x y λ son factibles, entonces $\lambda'b = \lambda'A x \leq c'x$, dado que $x \geq 0, Ax = b$ y $\lambda'A \leq c'$
- El lema muestra como un vector factible para cualquiera de los problemas permite obtener una cota para el valor del otro problema. Los valores asociados con el problema primal son siempre más grandes o iguales a aquellos asociados con el problema dual, por lo que los valores primales permiten obtener una cota superior y los duales una cota inferior



- Una medida para poder medir que tan cerca están las cotas es el *optimality gap*, el cual es la diferencia entre la cota superior y la inferior y se puede dividir entre la cota inferior si se quiere este en porcentaje

$$opt. gap = \frac{Primal Value - Dual Value}{Dual Value}$$

- Como el problema primal busca un mínimo y el dual un máximo, ambos buscan encontrar un valor común para el otro, de modo que si x_0 y λ_0 son factibles para $\min\{c'x : Ax = b \text{ \& } x \geq 0\}$ y para $\max\{\lambda'b : \lambda'A \leq c'\}$, respectivamente, y $c'x_0 = \lambda_0'b$, entonces x_0 y λ_0 son óptimos para sus respectivos problemas
 - El corolario enseña como si un par de valores factibles puede encontrarse para los programas primal y dual con un mismo valor de la función objetivo, entonces ambos son óptimos
 - El teorema de dualidad de la programación lineal expresa que lo converso también es verdad, de modo que las dos regiones tienen un punto común
- Si alguno de los programas $\min\{c'x : Ax = b \text{ \& } x \geq 0\}$ o $\max\{\lambda'b : \lambda'A \leq c'\}$ tiene una solución finita óptima, también tiene el otro programa, y los valores correspondientes de la función objetivo son iguales. Si alguno de los problemas tiene una función objetivo no acotada, el otro problema no tiene solución factible

- Es posible demostrar el teorema de dualidad a través del método Simplex, de modo que se establece una relación entre los problemas duales y los problemas primales a través del método de optimización

- Suponiendo un problema lineal $\min\{\mathbf{c}'\mathbf{x} : A\mathbf{x} = \mathbf{b} \text{ \& } \mathbf{x} \geq \mathbf{0}\}$, se obtiene una solución factible básica óptima $\mathbf{x} = (\mathbf{x}_B, \mathbf{0})$ con base correspondiente B

- Es necesario solucionar el problema dual $\max\{\boldsymbol{\lambda}'A : \boldsymbol{\lambda}'A \leq \mathbf{c}'\}$ en términos de la base B
- Partiendo A en $A = [B, D]$, como $\mathbf{x}_B = B^{-1}\mathbf{b}$ es óptimo, el vector del coste relativo \mathbf{r} será no negativo en cada componente. Por lo tanto, se obtiene la siguiente desigualdad:

$$\mathbf{r}'_D = \mathbf{c}'_D - \mathbf{c}'_B B^{-1}D \geq 0 \Rightarrow \mathbf{c}'_D \geq \mathbf{c}'_B B^{-1}D$$

- Definiendo $\boldsymbol{\lambda}' \equiv \mathbf{c}'_B B^{-1}$, se puede ver como esta $\boldsymbol{\lambda}$ es la solución al problema dual cuando la base B es la la óptima

$$\boldsymbol{\lambda}'A = [\boldsymbol{\lambda}'B, \boldsymbol{\lambda}'D] = [\mathbf{c}'_B, \mathbf{c}'_B B^{-1}D] \leq [\mathbf{c}'_B, \mathbf{c}'_D] = \mathbf{c}'$$

$$\Rightarrow \boldsymbol{\lambda}'A \leq \mathbf{c}' \Rightarrow \text{feasible for dual}$$

$$\boldsymbol{\lambda}'\mathbf{b} = \mathbf{c}'_B B^{-1}\mathbf{b} = \mathbf{c}'_B \mathbf{x}_B \Rightarrow \text{optimal value is the same}$$

$$\Rightarrow \boldsymbol{\lambda} \text{ is solution to dual problem}$$

- Suponiendo que el programa lineal $\min\{\mathbf{c}'\mathbf{x} : A\mathbf{x} = \mathbf{b} \text{ \& } \mathbf{x} \geq \mathbf{0}\}$ tiene una solución factible básica óptima correspondiente a la base B , entonces el vector $\boldsymbol{\lambda}$ satisface $\boldsymbol{\lambda}' = \mathbf{c}'_B B^{-1}$ es una solución óptima el programa dual $\max\{\boldsymbol{\lambda}'\mathbf{b} : \boldsymbol{\lambda}'A \leq \mathbf{c}'\}$ y los valores de ambas funciones objetivo son iguales
- En cualquier punto del algoritmo Simplex se puede obtener el vector $\boldsymbol{\lambda}' = \mathbf{c}'_B B^{-1}$, que, aunque sea solución solo si B es la base óptima, tiene una interpretación económica como el vector de multiplicadores Simplex
- Como para cualquier base B se puede construir un vector columna \mathbf{a}_j a partir de los de la base, y cada vector columna tiene un coste asociado c_j , también se puede crear este coste como combinación lineal de los costes asociados a los vectores columna de la base

- Por lo tanto, $\lambda_j \in \lambda' = c'_B B^{-1}$ se puede interpretar como el precio sombra de aumentar una unidad del vector a_i de la base B
 - La optimalidad del primal corresponde a la situación en donde cada vector a_1, a_2, \dots, a_n es más barato cuando se construye desde la base óptima B que cuando se compra a su propio precio (se está dispuesto a pagar hasta λ_j para cada variable)
- Los valores óptimos de las variables duales en un programa lineal pueden ser interpretados como precios sombra, lo cual facilita la realización de un análisis de sensibilidad. No obstante, es posible analizar más profundamente cambios en otros elementos del problema
 - El análisis de sensibilidad permite analizar como varía la solución óptima actual de un problema lineal ante cambios en determinados parámetros o bien al considerar nuevas variables y/o restricciones
 - Aunque se desarrollan métodos para estudiar rangos de valores que no afectan la optimalidad o factibilidad de la solución y como varían los valores de la solución óptima, las decisiones se tienen que tomar considerando no solo la función objetivo, sino las variables y otros elementos del problema
 - Suponiendo que en un programa lineal $\min\{c'x : Ax = b \text{ \& } x \geq 0\}$ la base óptima es B con la solución correspondiente $(x_B, 0)$, donde $x_B = B^{-1}b$, se tiene una solución óptima dual $\lambda' = c'_B B^{-1}$
 - Asumiendo que la solución no es degenerada, pequeños cambios en el vector b no causaran que la base óptima B cambie, dado que no afecta a $r'_D = c'_D - c'_B B^{-1}D$ (que es el criterio para cambiar variables dentro de la base). Lo único que cambia es la solución óptima x (por el cambio en x_B) y el valor de la función objetivo,
 - Por lo tanto, para $b + \Delta b$, la solución óptima, su cambio y el cambio de la función objetivo vendrán dados por la siguiente expresión:

$$x_B = B^{-1}(b + \Delta b) \Rightarrow x_B = B^{-1}b + B^{-1}\Delta b$$

$$\Rightarrow \Delta x_B = B^{-1}\Delta b \Rightarrow x = (x_B + \Delta x_B, 0)$$

$$z = c'x = c'_B x_B + c'_B B^{-1}\Delta b \Rightarrow \Delta z = c'_B B^{-1}\Delta b = \lambda' \Delta b$$

- Los resultados anteriores muestran como $\lambda = c'_B B^{-1}$ es la sensibilidad de la función objetivo con respecto a pequeños cambios en el vector b
 - La interpretación del vector dual λ está íntimamente relacionada a su interpretación como vector de multiplicadores del Simplex: como λ_j es el precio del vector unitario e_j cuando se construye a partir de la base B , este mide directamente mide el cambio en la función objetivo debido a un cambio en el componente j del vector b
 - En consecuencia, λ_j se puede interpretar como el precio marginal (precio sombra) del componente b_j , dado que b_j se cambia a $b_j + \Delta b_j$ el valor de la solución óptima por $\lambda_j \Delta b_j$
- Suponiendo que en un programa lineal $\min\{c'x : Ax = b \text{ \& } x \geq 0\}$ tiene una solución óptima $x = (x_B, 0)$, los cambios en el vector c (en c_B o en c_D) pueden afectar a su optimalidad
 - La factibilidad se ve afectada porque c_B y c_D determinan los costes relativos $r'_D = c'_D - c'_B B^{-1} D$, los cuales determinan si una la base sigue siendo óptima ($r'_D \geq 0$) o no
 - Por lo tanto, se pueden calcular los rangos de vectores c'_D y c'_B para los cuales la solución óptima x se mantiene invariante (asumiendo que uno de los dos se mantiene constante) realizando un sistema con los componentes de la siguiente suma:

$$r'_D = c'_D - c'_B B^{-1} D \geq 0 \Rightarrow c'_D + c'_B B^{-1} D \geq 0$$

- También es posible estudiar como afecta añadir una nueva variable o una nueva restricción al programa lineal
 - Cuando se añade una nueva variable \tilde{x} , todo el problema se ve afectado (función objetivo, restricciones y variables de decisión), pero puede no afectar al óptimo. Por lo tanto, es necesario calcular r'_D para la solución óptima actual y valorar si es posible mejorar incluyéndola en la base o no
 - En este caso, la nueva variable se tomaría como una variable no básica (nula) en la solución óptima, por lo que se tiene que calcular $r'_{D_{\tilde{x}}}$ (el coste relativo de la variable no básica \tilde{x})

- Cuando se añade una nueva restricción, la base óptima actual tendrá un elemento adicional. La opción más natural en este caso es incluir la variable de holgura como variable básica
- Si la solución óptima anterior satisface la nueva restricción de la forma $x_i \leq b_j$, no hay cambio de óptimo y la base actual continúa siendo factible y óptima. Si no la cumple, la solución es factible en el problema dual pero no es factible en el primal
- Las soluciones óptimas de los programas primales y duales satisfacen una relación adicional que tiene interpretación económica, la cual se puede postular para cualquier par de problemas duales: el teorema de holgura complementaria
 - Siendo x y λ las soluciones factibles para el problema primal y dual, respectivamente, $\min\{c'x : Ax = b \text{ \& } x \geq 0\}$ y $\max\{\lambda'b : \lambda'A \leq c'\}$, una condición necesaria y una suficiente de que ambas soluciones son óptimas es que, para cada i , se cumplan las siguientes proposiciones:

$$x_i > 0 \Rightarrow \lambda'a_i = c_i \quad \lambda'a_i < c_i \Rightarrow x_i = 0$$

- En este teorema, la forma de dualidad es la forma simétrica
- Si las dos condiciones se mantienen, entonces se cumple que $(\lambda'A - c')x = 0$. En consecuencia, $\lambda'b = c'x$ y ambas soluciones son óptimas (el valor de la función objetivo del dual y el primal, que son cota inferior y superior respectivamente son iguales)
- De manera converso, si dos soluciones son óptimas, el teorema de dualidad muestra que $\lambda'b = c'x$ y, por tanto, que se cumple que $(\lambda'A - c')x = 0$. Como cada componente de x no es negativo y cada componente de $\lambda'A - c'$ no es positivo (es cero o negativo), las condiciones se deben mantener
- Siendo x y λ las soluciones factibles para el problema primal y dual, respectivamente, $\min\{c'x : Ax \geq b \text{ \& } x \geq 0\}$ y $\max\{\lambda'b : \lambda'A \leq c' \text{ \& } \lambda \geq 0\}$, una condición necesaria y una suficiente de que ambas soluciones son óptimas es que, para toda i y j , se cumplan las siguientes proposiciones:

$$x_i > 0 \Rightarrow \lambda'a_i = c_i \quad \lambda'a_i < c_i \Rightarrow x_i = 0$$

$$\lambda_j > 0 \Rightarrow a_jx = b_j \quad a_jx > b_j \Rightarrow \lambda_j = 0$$

- En este teorema, la forma de dualidad es la forma antisimétrica

- Estas condiciones no solo implican lo anterior, sino que implican que se mantiene $\lambda(Ax - b) = 0$ si las dos condiciones se cumplen, y que, si ambas soluciones son óptimas, el teorema de dualidad muestra que $\lambda'b = c'x$ y que también se cumple que $\lambda(Ax - b) = 0$. Por lo tanto, se puede expresar el teorema de la siguiente manera:

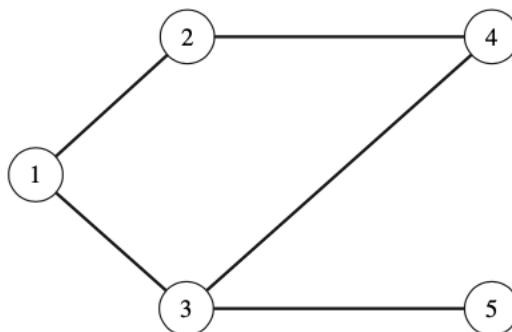
$$(c - \lambda'^* A)x^* = 0 \quad \& \quad \lambda^*(Ax^* - b) = 0$$

- La versión del teorema para el problema antisimétrico permite hacer una interpretación económica
 - Si el precio sombra λ_j es nulo, entonces no se estaría dispuesto a obtener más cantidad de b_j , ya que se tiene la cantidad óptima y eso hace que $a_jx > b_j$. No obstante, si es diferente de cero, entonces quiere decir que se estaría dispuesto a obtener más cantidad de b_j y eso implica que $a_jx = b_j$ (no puede ser menor, por las restricciones impuestas)
 - Si la cantidad x_i es nula, entonces quiere decir que su coste c es mayor al coste de obtener la cantidad “de manera externa”, $\lambda'a_i$, y eso hace que $\lambda'a_i \leq c_j$. No obstante, si es diferente de cero, entonces quiere decir que el coste de obtener la cantidad “de manera externa” es igual al coste de “producir” esas cantidades y eso hace que $\lambda'a_i = c_j$ (no puede ser mayor, por las restricciones impuestas)
 - Si las restricciones son de desigualdades \geq en el el problema primal, entonces la dirección de las interpretaciones es la contraria para las desigualdades, ya que se tendría que interpretar λ_j como cuanto se está dispuesto a pagar para reducir la cantidad de b_j

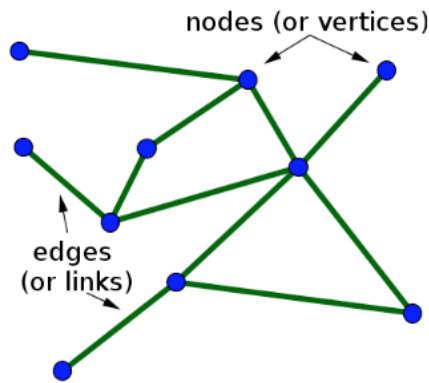
Los problemas de transporte y de flujos de redes

- Existen varios problemas que, por su estructura especial, se pueden representar por el problema de transporte y otros problemas canónicos vistos anteriormente. De este modo, la estructura especial permite desarrollar algoritmos de solución eficientes
 - Uno de esos tipos de problemas con estructura especial son los problemas de transporte, de los cuales ya se han hablado anteriormente
 - Este tipo de problemas tiene problemas relacionados importantes como el problema de asignación

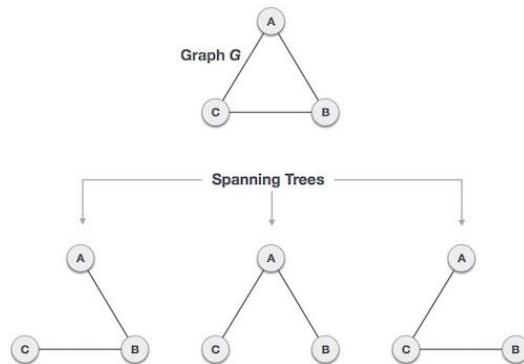
- Este tipo de problemas son importantes porque representan áreas de aplicaciones que surgen frecuentemente y por la riqueza de la teoría asociada a este, la cual proporciona ideas para nuevos desarrollos
- Otro de esos tipos de problemas con estructura especial son los problemas de flujos de redes, los cuales se pueden considerar uno de los tipos de problemas más importantes en programación lineal
 - El desarrollo de reglas especializadas para el método Simplex (que aprovechan la estructura especial de los problemas de flujo de redes) permiten resolver de manera muy eficiente este tipo de problemas. Estas reglas permiten resolver de forma eficiente problemas muy difíciles donde el Simplex estándar sería incapaz
 - Este tipo de problemas se aplica en campos como logística, transporte, redes eléctricas y redes de telecomunicaciones
- Para poder desarrollar un algoritmo eficiente para un problema de flujo de redes, es necesario utilizar conceptos básicos de la teoría de grafos
 - Un grafo consiste en una colección finita de elementos llamados nodos con un subconjunto de pares desordenados de nodos llamados arcos



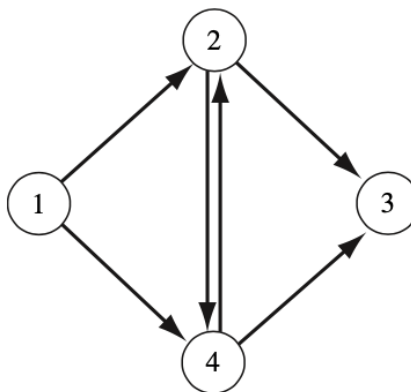
- Los nodos de un grafo normalmente se designan por círculos y se enumeran, y un arco entre los nodos i y j se representa por un par desordenado (i, j) y una línea entre esos nodos
 - Una cadena o *chain* entre los nodos i y j es una secuencia de arcos conectándolos, de modo que la secuencia tiene la forma $(i, k_1), (k_1, k_2), \dots, (k_m, j)$. Si se especifica una dirección de movimiento de la secuencia (es decir, de un nodo i a un nodo j), la secuencia se denomina camino o *path*. Si una cadena va de un nodo i a el mismo nodo i , esta se denomina
- Un grafo está conectado si hay una cadena entre cualquier par de nodos, y un grafo es un árbol si está conectado y no tiene ciclos



- Muchas veces se considera un árbol dentro de un grafo G , el cual es un árbol hecho de un subconjunto de arcos de G , el cual se denomina árbol de expansión si toca todos los nodos de G . Se puede demostrar que un grafo está conectado si, y solo si, contiene un árbol de expansión



- El interés en el contexto de problemas de flujos de redes está en los grafos direccionales, en el cual un sentido de orientación se da a cada arco



- En este caso un arco se considera un par ordenado de nodos (i, j) y se dice que el arco es del nodo i al j , lo cual se indica con una flecha en cada arco (apuntando desde i a j)

- Cuando se trabaja con grafos direccionales, algunos pares de nodos pueden tener un aro en ambas direcciones entre ellos
- Las nociones de caminos y ciclos pueden aplicarse directamente a grafos direccionales, y además se puede añadir que un nodo j es alcanzable desde i si hay un camino del nodo i al j
- Además de la representación visual de un grado direccional, otro método común de representación es en términos de la matriz de incidencia nodo-arco de un grafo, la cual tiene los nodos como filas y los arcos como columnas

	(1, 2)	(1, 4)	(2, 3)	(2, 4)	(4, 2)
1	1	1			
2	-1		1	1	-1
3			-1		
4		-1		-1	1

- En la columna debajo del arco (i, j) , un 1 se posiciona en el nodo i y un -1 se posiciona en el nodo j . Por lo tanto, se muestra las direcciones de los arcos y toda la información de la estructura del grado se contiene en esta matriz
- Un grafo es una manera efectiva de representar la estructura comunicativa entre nodos, pero cuando hay la posibilidad de que haya flujos en los arcos, este grafo direccional se denomina red
- Un flujo en un arco direccional (i, j) es un número $x_{ij} \geq 0$, y estos flujos deben satisfacer un criterio de conservación en cada nodo. Específicamente, a no ser que el nodo sea un nodo productor o uno demandante, un flujo no se puede crear o perder en un nodo, por lo que el flujo total que entra en un nodo debe ser igual al flujo total que sale de este

$$\sum_{j=1}^n x_{ij} - \sum_{k=1}^n x_{ki} = 0$$

- La primera suma es el flujo total que sale de i , y la segunda suma es el flujo total que entra en i
- En muchas aplicaciones, se considera que hay unos nodos productores y unos nodos demandantes, de modo que estos nodos pueden crear o perder flujos. Los nodos se pueden clasificar en nodos de origen (donde se produce el material o desde donde se envía), nodos de destino (donde ha de llegar el material) y nodos intermedios (nodos por donde pasa el material)

- Una vez introducidos los conceptos necesarios de la teoría de grafos, es posible presentar el problema de flujos de redes de coste mínimo, el cual generaliza un poco el problema de transporte
 - El problema de flujos de redes de coste mínimo consiste en determinar los flujos $x_{ij} \geq 0$ en cada arco de la red de modo que el flujo neto en cada nodo sea b_i mientras que se minimiza el coste total. En términos matemáticos, este problema se puede representar de la siguiente manera:

$$\begin{aligned} \min_{x_{ij}} \quad & \sum_{(i,j) \in A} c_{ij} x_{ij} \\ \text{s. t.} \quad & \sum_{j=1}^n x_{ij} - \sum_{k=1}^n x_{ki} = 0 \text{ for } \forall i \in N \\ & \sum_{i=1}^n b_i = 0 \\ & l_{ij} \leq x_{ij} \leq u_{ij} \text{ for } (i,j) \in A \end{aligned}$$

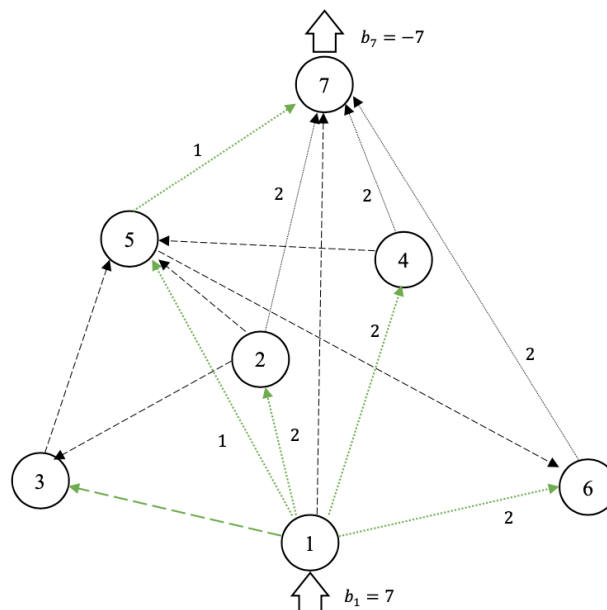
- Considerando una red de n nodos, y correspondiendo a cada nodo i una oferta disponible b_i (si $b_i < 0$ entonces hay demanda), se asume que la red es balanceada, de modo que la suma de ofertas disponibles es nula

$$\sum_{i=1}^n b_i = 0$$

- Asociado a cada arco $(i,j) \in A$ hay un número c_{ij} representando el coste unitario por cada unidad de flujo en cada arco. Además, hay un límite inferior l_{ij} y un límite superior u_{ij} de unidades de flujo para cada arco
- El problema de transporte es un caso especial de este problema, correspondiendo a una red con arcos que van solo de nodos productores a nodos demandantes
- Este problema es claramente un problema lineal, en donde la matriz de coeficientes A de las restricciones de flujo es la matriz de incidencias
 - Para una columna (i,j) hay un 1 en la fila del nodo i y un -1 en la fila del nodo j , de modo que la suma de cada fila (de cada nodo) será nula (el vector cero) y eso hace que la matriz tenga un rango

igual a $n - 1$ (todos los nodos menos uno, que se puede expresar como combinación lineal)

- Debido a que se tiene un grado de libertad, se puede eliminar cualquier fila de A para obtener una matriz de coeficientes con rango igual a aquel de la original, de modo que cualquier submatriz A_T $(n - 1) \times (n - 1)$ es una matriz triangular invertible y un árbol de expansión G corresponde a una base
- Dada una base, la solución básica correspondiente puede ser encontrada sustituyendo hacia atrás usando la estructura triangular
 - En este proceso se intenta encontrar una ecuación que tenga una sola variable básica indeterminada correspondiente a un flujo de arco indeterminado x_{ij}
 - Esta ecuación se resuelve para x_{ij} , y entonces se encuentra otra ecuación de este tipo. En términos de redes, uno primero busca el final de un árbol de expansión que corresponde a la base (un nodo que solo toque un arco en el árbol, un nodo hoja) y entonces se determina el flujo de este arco por la producción o demanda de ese nodo (y se sigue el proceso hacia atrás)
- El método Simplex derivado anteriormente se puede generalizar para el problema de flujos de redes de coste mínimo, siguiendo así los siguientes pasos:
 - El primer paso consiste en obtener una solución básica factible al problema



- Normalmente se da una solución factible o se puede encontrar por inspección, habiendo siempre $n - 1$ arcos básicos (tantos arcos como nodos haya menos uno) conformando la base B de tamaño $m \times (n - 1)$. Esto quiere decir que todos los otros arcos se distribuirán en no básicos D_l que alcanzan su límite inferior y no básicos D_u que alcanzan su límite superior
 - Lo recomendable cuando se tiene una solución básica es realizar la red correspondiente a esa solución factible, con tal de visualizar mejor los pasos siguientes
 - Si se puede escoger una base cualquiera a partir de una solución factible y de unos arcos que cumplan $l_{ij} < x_{ij} < u_{ij}$ (los cuales se tienen que incluir en la base obligatoriamente), siempre se escogerán aquellos arcos que después permitan obtener una variable dual para cada nodos
- A partir de establecer a qué conjunto pertenece cada arco, se pueden calcular los multiplicadores del Simplex (o las variables duales) para cada nodo i . Esto es equivalente a resolver las siguientes ecuaciones:

$$\lambda_i - \lambda_j = c_{ij} \text{ for } (i, j) \in A \Rightarrow c_{ij} - \lambda_i + \lambda_j = 0 \text{ for } (i, j) \in A$$

- La segunda expresión es mucho más fácil de utilizar, dado que solo hace falta colocar los subíndices del arco en el mismo orden para cada variable dual
- Estas ecuaciones se resuelven escogiendo arbitrariamente el valor de uno de los multiplicadores, dado que se tiene un grado de libertad. Por lo tanto, se determinan los valores de todas las variables duales (para cada nodo)

$$\begin{cases} 4 - y_1 + y_2 = 0 \\ 6 - y_1 + y_3 = 0 \\ 4 - y_1 + y_4 = 0 \\ 6 - y_1 + y_5 = 0 \\ 4 - y_1 + y_6 = 0 \\ 6 - y_2 + y_7 = 0 \end{cases} \Rightarrow \text{if } y_2 = 0 \text{ then } \begin{cases} y_1 = 4 \\ y_3 = -2 \\ y_4 = 0 \\ y_5 = -2 \\ y_6 = 0 \\ y_7 = -6 \end{cases}$$

- Esto se puede hacer porque cada arco (i, j) corresponde a una columna de A con un 1 en la fila i y un -1 en la fila j , de modo que se tiene relación con las restricciones del problema dual

$$\begin{array}{ll} \min_x c'x & \max_{\lambda} b'\lambda \\ \text{s.t. } Ax = b & \text{s.t. } A'\lambda \leq b \\ x_i \geq 0 & \lambda_i \text{ free} \end{array} \Rightarrow$$

- Una vez se obtienen los valores para todas las variables duales de todos los nodos, es posible calcular el coste reducido para cada arco no básico

$$r_{ij} = c_{ij} - \lambda_i + \lambda_j \text{ for } (i,j) \in A$$

- En el óptimo, los arcos en D_l tienen un coste positivo o nulo y los arcos en D_u uno negativo o nulo

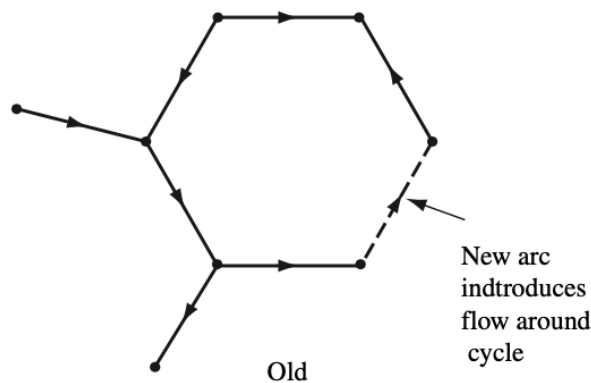
$$Optimal = \begin{cases} r_{ij} = c_{ij} - \lambda_i + \lambda_j \geq 0 & \text{for } (i,j) \in D_l \\ r_{ij} = c_{ij} - \lambda_i + \lambda_j \leq 0 & \text{for } (i,j) \in D_u \end{cases}$$

- Por lo tanto, si algún arco en D_l tiene coste negativo, es candidato a entrar en la base, y si algún arco de D_u tiene un coste positivo, también

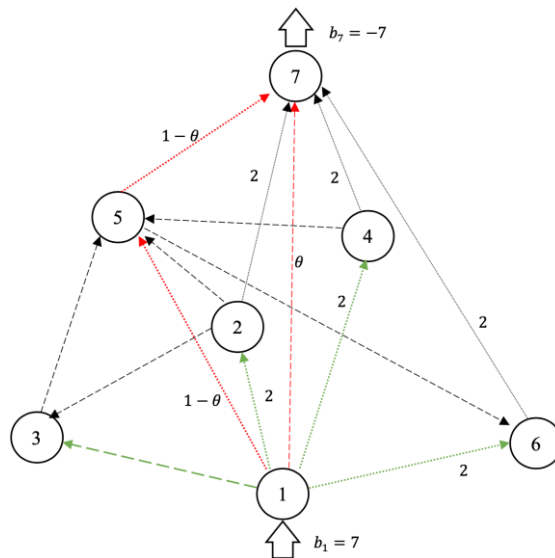
$$r_{ij} = c_{ij} - \lambda_i + \lambda_j < 0 \text{ for } (i,j) \in D_l \Rightarrow \text{enter } B$$

$$r_{ij} = c_{ij} - \lambda_i + \lambda_j > 0 \text{ for } (i,j) \in D_u \Rightarrow \text{enter } B$$

- Si la base no es óptima, se añade uno de los arcos candidatos a entrar a la base en la red de la solución factible, lo cual producirá un ciclo



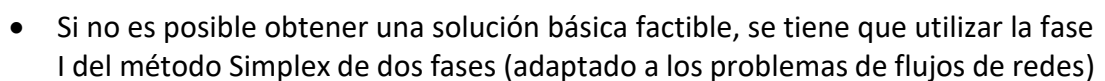
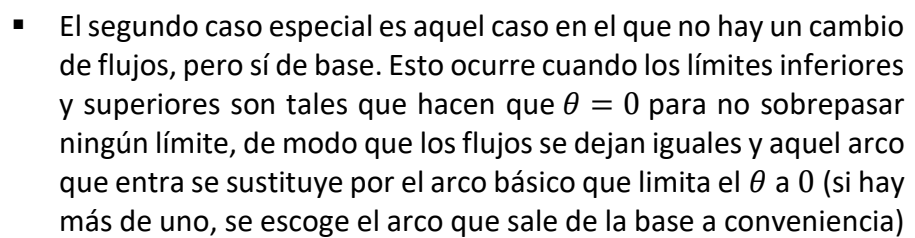
- Para poder cambiar la base con este nuevo ciclo, se introduce un flujo $\theta > 0$ en el nuevo arco que se introduce, pero como la producción total no varía, esto también hace que se modifiquen los flujos de los arcos que forman el ciclo, por lo que se tendrán que añadir o restar unidades de flujo al flujo de la solución factible
- Para saber si añadir o restar unidades en un arco del ciclo, es necesario tener en cuenta la dirección del arco que se añade y la de los otros. Los flujos se suman cuando siguen el mismo sentido de ciclo que el arco añadido, pero se restan si van en sentido de ciclo contrario

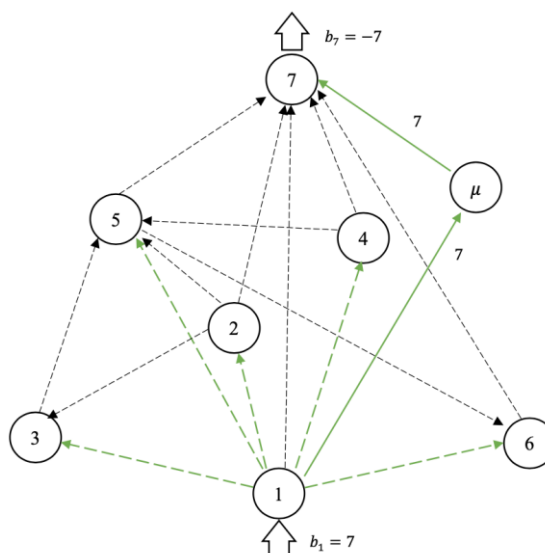


- Al plantear las sumas y restas correspondientes para cada flujo, se escoge un valor de θ que sea el más pequeño que haga que uno de los flujos de algún arco del ciclo sea nulo. Para ello, se utilizan los límites inferiores y superiores de unidades de flujo

$$l_{ij} \leq x_{ij} \pm \theta \leq u_{ij} \text{ for } (i,j) \text{ in the cycle}$$

- Además, también se tiene que tener en cuenta el límite superior e inferior de los arcos al plantear decidir qué arcos entran a la base y el valor del cambio en los flujos, ya que pueden ocurrir dos casos especiales
 - El primer caso especial es aquel caso en el que hay un cambio de flujo, pero no un cambio de base. Este caso ocurre cuando al arco que se añade (y crea un ciclo) se le asigna un valor de θ igual a su límite superior o inferior, haciendo que si los otros arcos tienen flujos que no llegan a estos límites, la base no cambie, pero los flujos sí lo hagan (también cambia la composición de los conjuntos de arcos no básicos, dependiendo del límite)





- En este caso, la fase I consistirá en añadir un nodo artificial μ que permita transportar la producción total a través de sus arcos hasta los nodos demandantes. Los arcos tendrán un límite inferior $l_\mu = 0$ y un límite superior $u_\mu = \infty$ (se puede transportar cualquier cantidad), y estos tienen un coste c_μ igual a 1
 - Los arcos originales se fijan a su cota inferior l_{ij} , de modo que esto permite saber cuantos arcos artificiales conectar al nodo artificial (los flujos mínimos originales determinarán si algún nodo original necesita un arco artificial o no), aunque también se pueden conectar a todos (pero es mucho más complicado trabajar así)
 - Como se ha añadido un nodo artificial, la base constará de $n - 1$ arcos (tanto originales como artificiales). Estos se pueden escoger a conveniencia si es que no hay $n - 1$ arcos artificiales, dado que estos conformarían la base porque $x_{ij} < u_{ij}$ para todo arco $(i, j) \in A_\mu$
- A la vez, se fijan los costes de todos los otros arcos a cero, de modo que se hace un procedimiento similar a como se hacía con el Simplex estándar (se cambiaba la función objetivo a la suma de las variables artificiales)

$$\min_{x_{ij}} \sum_{(i,j) \in A} c_{ij} x_{ij} \Rightarrow \min_{x_{ij}} \sum_{(i,j) \in A_\mu} x_{ij}$$

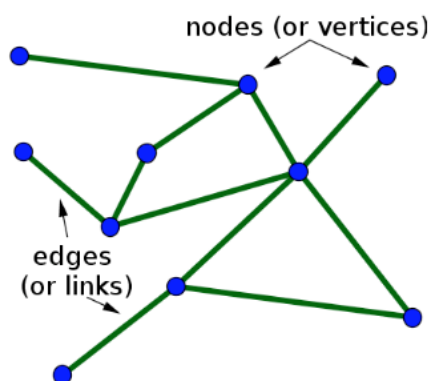
- Esto hace que el objetivo ahora sea el de hacer que los flujos de los arcos artificiales sean nulos (se eliminan para que la función objetivo sea nula)

$$\min_{x_{ij}} \sum_{(i,j) \in A_{\mu}} x_{ij} = 0 \Rightarrow x_{ij} = 0 \text{ for } (i,j) \in A_{\mu}$$

- Una vez se fijan los flujos de los arcos originales y de los artificiales, se pueden calcular los multiplicadores del Simplex (o variables duales) y los costes relativos para todos los arcos no básicos
 - Como se ha fijado el valor mínimo para los flujos en los arcos originales, solo se tendrá que ver qué arcos tienen un coste negativo para poder entrar a la base. Si no hay arcos con un coste negativo, entonces el problema es infactible
- A partir de este punto, se puede realizar el proceso explicado anteriormente hasta que los flujos de los arcos artificiales sean nulos y los costes relativos indiquen que la base es óptima
 - Cuando la base es óptima, los flujos de los arcos artificiales son nulos y por tanto se pueden eliminar los arcos y los nodos artificiales del problema, pudiendo así comenzar el algoritmo con una solución factible

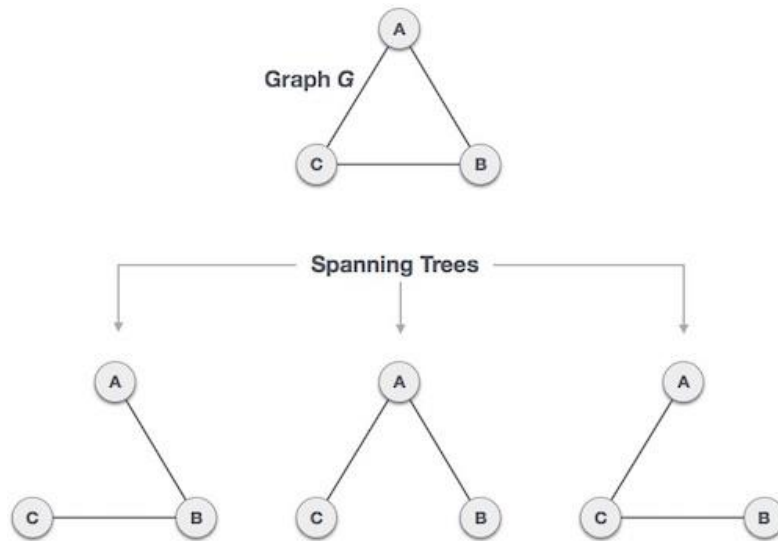
El problema de árbol de expansión de coste mínimo

- Para resolver algunos tipos de problemas especiales se utilizan los conceptos de redes no direccionales y de árbol de expansión, por lo que resultados para estos también son de vital importancia
 - Una red no direccional es un grafo $G = (\mathcal{N}, \mathcal{A})$ de n nodos y m arcos, en los cuales un arco es un par desordenado de elementos de \mathcal{N}

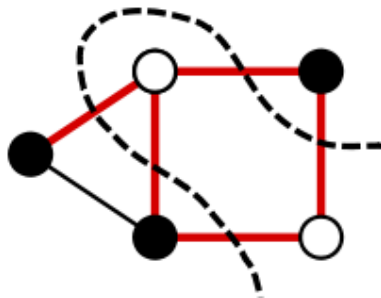


- En otras palabras, para dos elementos $i, j \in \mathcal{N}$, el arco del nodo i al j (denotado como (i, j)) es lo mismo que un arco del nodo j al i (denotado como (j, i))

- Un árbol de expansión T de un grafo no direccional G es un subgrafo T de G tal que \mathcal{N}' se expande a todo G (si $\mathcal{N} = \mathcal{N}'$). Por lo tanto, es un subgrafo conectado y acíclico de G que se expande a todos los nodos



- En otras palabras, un grafo conectado con $n - 1$ arcos y n nodos es un árbol de expansión de G . Consecuentemente, un grafo puede tener diferentes árboles de expansión
- Siendo T cualquier árbol, cualquier arco (i, j) en T es un arco del árbol, mientras que cualquier arco que no pertenezca a T es un arco fuera del árbol
- Siendo T cualquier árbol de expansión y a un arco fuera del árbol, añadir el arco a al árbol de expansión crea un único ciclo C , y eliminar cualquier arco del ciclo C crea un árbol de expansión T'
- Un conjunto F de nodos es un corte o *cut* en G si existe una partición $\{\mathcal{S}, \mathcal{N} - \mathcal{S}\}$ de G tal que $F = \mathcal{A}(\mathcal{S}, \mathcal{N} - \mathcal{S})$, y se dice que los arcos de F cruzan esta partición



- En otras palabras, si (i, j) es un arco del árbol, entonces $T - (i, j)$ se divide en dos conjuntos de nodos \mathcal{S} y $\mathcal{N} - \mathcal{S}$, y los arcos en la partición $(\mathcal{S}, \mathcal{N} - \mathcal{S})$ constituyen un corte
- Un camino o *path* es un grafo no vacío $P = (\mathcal{N}, \mathcal{A})$ de la siguiente forma:

$$\mathcal{N} = \{x_0, x_1, \dots, x_k\} \quad \mathcal{A} = \{x_0x_1, x_1x_2, \dots, x_{k-1}x_k\}$$

where all x_i are distinct

 - Los nodos x_0 y x_1 están vinculados por P y se denominan nodos finales o finales, mientras que los nodos x_1, \dots, x_{k-1} se denominan nodos interiores de P
 - El número de arcos en un camino determina su longitud, y un camino de una longitud k se denota como P^k
- Una importante clase de problemas cuyas aplicaciones están en la frontera entre la programación lineal y entera son los problemas de flujos de redes de coste mínimo
 - El problema del árbol de expansión de mínimo coste se basa en encontrar el árbol de expansión con el mínimo coste total (asociado a sus nodos)
 - En este caso, cada arco tiene un coste asociado c_{ij} y el coste total $\sum_i \sum_j c_{ij}$ es la suma de los costes de cada arco
 - A partir del concepto de corte de un grafo y de camino de un grafo, es posible derivar condiciones de optimalidad y teoremas para árboles de expansión de coste mínimo
 - Las condiciones de optimalidad de un corte expresan que, para cualquier arco $(i, j) \in T$ se da que $c_{ij} \leq c_{kl}$ para cualquier arco (k, l) contenido en el corte formado al eliminar (i, j) de T
 - El teorema derivado de esta condición es que un árbol de expansión T es un árbol de expansión de coste mínimo si, y solo si, satisfacen las condiciones de optimalidad de un corte
 - Al ser una condición suficiente, solo es necesario demostrar la primera parte del teorema. Si se diera el caso que $c_{ij} \geq c_{kl}$ para un arco (k, l) en el corte formado al eliminar (i, j) de T , entonces $T' = T - (i, j) + (k, l)$ tendría un coste menor al árbol de expansión T y no sería mínimo
 - Las condiciones de optimalidad del camino expresan que, para cualquier arco $(k, l) \in G$ que no sea del árbol, se da que $c_{ij} \leq c_{kl}$

para todo arco (i, j) contenido en el camino de T conectando los nodos k y l

- El teorema derivado de esta condición es que un árbol de expansión T es un árbol de expansión de coste mínimo si, y solo si, satisfacen las condiciones de optimalidad del camino
- Al ser una condición suficiente, solo es necesario demostrar la primera parte del teorema. Si se diera el caso que $c_{ij} \geq c_{kl}$ para un arco (k, l) fuera del árbol, entonces $T' = T - (i, j) + (k, l)$ tendría un coste menor al árbol de expansión T y no sería mínimo
- Una observación importante es que, siendo (k, l) un arco en $\mathcal{A} - T$, entonces (i, j) está contenido en el ciclo C que se forma en $T + (k, l)$ si, y solo si, (k, l) está en el corte formado por $T - (i, j)$
 - Si (k, l) está en el corte formado por $T - (i, j)$, si se añade (i, j) entonces se puede crear un ciclo. De manera conversas, si existe un ciclo C en $T + (k, l)$, se puede ver que hay algún arco $(i, j) \in T$ que está dentro de C y que, si se elimina, entonces se forma un corte del cual el arco (k, l) es perteneciente
 - Esto hace que las condiciones de optimalidad de un corte y del camino sean equivalentes
- Los dos algoritmos más importantes para poder resolver el problema de árbol de expansión de coste mínimo son el algoritmo de Kruskal y el de Prim
 - El algoritmo de Kruskal es un algoritmo de búsqueda codiciosa y sigue los siguientes pasos:
 - Primero, se ordenan los arcos a_1, a_2, \dots, a_m en orden no decreciente (de menor a mayor) dependiendo de su coste asociado y se fija $t = 0$
 - Después, se fija $i = 1$ y se suma $T + a_i$. Si $T + a_i$ no tiene ningún ciclo, entonces se fija $T = T + a_i$. De haber un ciclo, entonces se fija $i = i + 1$ y se repite el paso
 - El algoritmo finaliza cuando se han probado todos los arcos de 1 a m , de manera que se obtiene un árbol de expansión mínimo que satisface las condiciones de optimalidad del camino

Las propiedades básicas de la optimización no restringida

- En los problemas de optimización no restringidos, se suelen considerar problemas de la forma $\min\{f(x): x \in \Omega\}$, en donde f es una función de valores reales y el conjunto factible Ω es un subconjunto de E^n

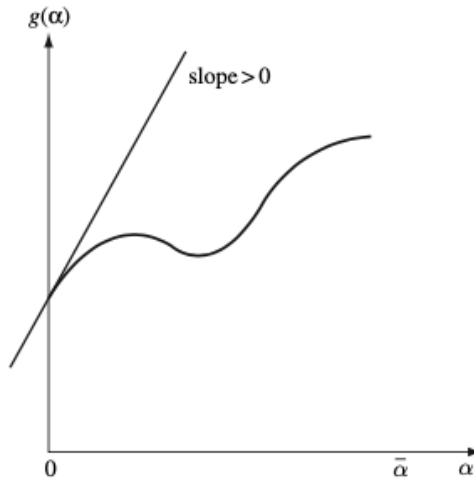
$$\min f(x)$$

$$\text{subject to } x \in \Omega$$

- Especialmente, se considera el caso en el que $\Omega = E^n$, correspondiendo al caso sin restricción alguna. Sin embargo, se pueden considerar casos donde Ω es un subconjunto simple particular de E^n
 - En cualquier caso, como f es una función no lineal, esta puede tener mínimos y máximos
- El teorema de Weierstras permite ver que si f es una función continua y Ω es un conjunto compacto, entonces existe una solución para el problema de minimización $\min\{f(x): x \in \Omega\}$
 - El teorema de Weierstras expresa que una función continua definida en un conjunto compacto S tiene un punto mínimo en S , de modo que existe una $x \in S$ tal que para toda $x \in S$ se da que $f(x) \geq f(x^*)$
 - Este resultado es muy importante y se tiene que mantener en mente durante el desarrollo de las siguientes secciones
- Existen dos tipos de puntos de solución: los puntos mínimos locales y los puntos mínimos globales
 - Un punto $x^* \in \Omega$ se denomina punto mínimo relativo o punto mínimo local de f sobre Ω si existe una $\varepsilon > 0$ tal que $f(x) \geq f(x^*)$ para toda $x \in \Omega$ y $|x - x^*| < \varepsilon$. Si $f(x) > f(x^*)$ para toda $x^* \neq x \in \Omega$ y $|x - x^*| < \varepsilon$, entonces x^* es un punto mínimo relativo estricto (único) de f sobre Ω
 - Un punto $x^* \in \Omega$ se denomina punto mínimo global de f sobre Ω si $f(x) \geq f(x^*)$ para toda $x \in \Omega$. Si $f(x) > f(x^*)$ para toda $x^* \neq x \in \Omega$, entonces x^* es un punto mínimo global estricto (único) de f sobre Ω
- Las condiciones de primer orden son las condiciones necesarias para la existencia de una solución al problema de minimización planteado anteriormente. Estas no son más que una extensión de las condiciones de las derivadas de funciones univariantes

- Al formular un problema como $\min\{f(\mathbf{x}):\mathbf{x} \in \Omega\}$, se busca un punto mínimo global de f sobre Ω por definición. No obstante, desde el punto de vista teórico y computacional, la práctica muestra como hay muchas situaciones en las que uno se tiene que conformar con un punto mínimo relativo
 - Al derivar las condiciones necesarias o al buscar puntos mínimos a través de procedimientos numéricos, la comparación de los valores en puntos cercanos es todo lo posible y, por tanto, uno suele poner su atención en mínimos relativos
 - Las condiciones globales y las soluciones globales, en general, pueden ser encontradas solo si el problema posee propiedades de convexidad concretas que garantizan que cualquier mínimo relativo es un mínimo global
- Para derivar las condiciones necesarias satisfechas por un punto mínimo relativo \mathbf{x}^* , la idea básica es considerar el movimiento lejos del punto en una dirección dada
 - Con respecto a una dirección dada, la función objetivo se pueden considerar una función univariante, la cual será el parámetro definiendo el movimiento en esta dirección
 - Por lo tanto, dada una $\mathbf{x} \in \Omega$, un vector $\mathbf{d} \neq \mathbf{0}$ es una dirección factible en \mathbf{x} si existe una $\bar{\alpha} > 0$ tal que $\mathbf{x} + \bar{\alpha}\mathbf{d} \in \Omega$ para toda α que cumpla $0 \leq \alpha \leq \bar{\alpha}$
- Siendo Ω un subconjunto de E^n y $f \in C^1$ una función en Ω , si \mathbf{x}^* es un punto mínimo relativo de f sobre Ω , entonces para cualquier $\mathbf{d} \neq \mathbf{0} \in E^n$ que sea factible en la dirección en \mathbf{x}^* , se tiene que $\nabla_{\mathbf{x}}f(\mathbf{x}^*)\mathbf{d} \geq 0$
 - Para cualquier α que cumpla $0 \leq \alpha \leq \bar{\alpha}$, el punto $\mathbf{x}(\alpha) \equiv \mathbf{x}^* + \alpha\mathbf{d} \in \Omega$, así que definiendo una función $g(\alpha) \equiv f(\mathbf{x}(\alpha))$, se puede ver que g tiene un mínimo relativo en $\alpha = 0$. Haciendo una expansión de primer orden lineal, se puede ver que si $g'(0) < 0$, entonces para valores suficientemente pequeños de $\alpha > 0$, el lado derecho es negativo y se contradice la naturaleza mínima de $g(0)$. Por lo tanto, $g'(0) = \nabla_{\mathbf{x}}f(\mathbf{x}^*)\mathbf{d} \geq 0$

$$g(\alpha) - g(0) = g'(0)\alpha + o(\alpha)$$



$g'(0) < 0$ & small values of α :

$$g(\alpha) - g(0) < 0 \Rightarrow g'(0) = f'(x(0))x'(0) = \nabla_x f(x^*)d \geq 0$$

- Un caso particular importante es cuando x^* está en el interior de Ω (como sería el caso de $\Omega = E^n$), ya que en estos casos hay direcciones factibles emanando en cualquier dirección desde x^* y, por lo tanto, $\nabla_x f(x^*)d \geq 0$ para toda $d \neq 0 \in E^n$. Esto causa que $\nabla_x f(x^*)d = \nabla_x f(x^*) = 0$
- Siendo Ω un subconjunto de E^n y $f \in C^1$ una función en Ω . Si x^* es un punto mínimo relativo de f sobre Ω y si x^* es un punto interior de Ω , entonces $\nabla_x f(x^*) = 0$. Esto hace que se plantee un sistema de n ecuaciones y n incógnitas que puede ser resuelto para encontrar una solución
- Igual que las condiciones de primer orden se derivan de una aproximación de primer orden, condiciones adicionales se pueden obtener considerando aproximaciones de mayor orden, tales como las condiciones de segundo orden
 - Siendo Ω un subconjunto de E^n y $f \in C^2$ una función sobre Ω , si x^* es un punto mínimo relativo de f sobre Ω , entonces para toda $d \in E^n$ que sea una dirección factible en x^* se cumple que $\nabla_x f(x^*)d \geq 0$ y que, si $\nabla_x f(x^*)d = 0$, entonces $d'F(x^*)d \geq 0$ (para $d \neq 0$)
 - Definiendo $x(\alpha) \equiv x^* + \alpha d \in \Omega$ y $g(\alpha) \equiv f(x(\alpha))$ para cualquier α que cumpla $0 \leq \alpha \leq \bar{\alpha}$, se puede hacer una expansión de segundo orden asumiendo que $g'(0) = \nabla_x f(x^*)d = 0$. Si $g''(0) < 0$, entonces para valores suficientemente pequeños de $\alpha > 0$, el lado derecho es negativo y se contradice la naturaleza mínima de $g(0)$. Por lo tanto, $g''(0) = d'F(x^*)d \geq 0$

$$g(\alpha) - g(0) = g'(0)\alpha + \frac{1}{2}g''(0)\alpha^2 + o(\alpha^2) =$$

$$g'(0) = 0, g''(0) < 0 \text{ \& small values of } \alpha:$$

$$g(\alpha) - g(0) < 0 \Rightarrow g''(0) = \mathbf{d}'\mathbf{F}(\mathbf{x}^*)\mathbf{d} \geq 0$$

- Estas condiciones se denominan condiciones necesarias de primer y segundo orden. De este modo, todos los puntos que las cumplan pueden ser un punto mínimo relativo, pero no es suficiente como para decir si de verdad lo es
 - Para poder encontrar el vector o el conjunto de vectores direccionales \mathbf{d} , se tiene que utilizar $\nabla_{\mathbf{x}}f(\mathbf{x}^*)\mathbf{d} \geq 0$, de modo que estos vectores serán tales que cumplan esta desigualdad
 - Todos los puntos que cumplen la condición $\nabla_{\mathbf{x}}f(\mathbf{x}^*) = 0$ se denominan puntos estacionarios. Estos pueden ser un punto extremo o no, ya que estas condiciones son solo necesarias, pero no suficientes como para asegurarlo
- Igual que antes, el caso en donde el punto minimizador es un punto interior de Ω (como cuando $E^n = \Omega$) es importante y se obtiene un corolario similar al anterior
- Siendo Ω un subconjunto de E^n y $f \in C^2$ una función sobre Ω , si \mathbf{x}^* es un punto mínimo relativo de f sobre Ω , entonces para toda $\mathbf{d} \in E^n$ que sea una dirección factible en \mathbf{x}^* se cumple que $\nabla_{\mathbf{x}}f(\mathbf{x}^*)\mathbf{d} \geq 0$ y $\mathbf{d}'\mathbf{F}(\mathbf{x}^*)\mathbf{d} \geq 0$ para toda $\mathbf{d} \neq \mathbf{0}$
 - Estas serán las condiciones necesarias de primer y segundo orden que más se usan en la práctica. Estas condiciones permiten obtener candidatos para mínimos locales o relativos, pero no permiten asegurar o garantizar que en verdad son mínimos relativos
- Usando una versión más fuerte para la segunda condición, se puede obtener un conjunto de condiciones que implican que el punto \mathbf{x}^* es un mínimo relativo estricto (único). Estas condiciones aplican solo a problemas sin restricciones o a problemas donde el punto mínimo está dentro de la región factible
- Siendo $f \in C^2$ una función definida en una región en la que el punto \mathbf{x}^* es un punto interior y suponiendo que $\nabla_{\mathbf{x}}f(\mathbf{x}^*) = \mathbf{0}$ y $\mathbf{F}(\mathbf{x}^*)$ es una matriz definida positiva, entonces \mathbf{x}^* es un punto mínimo relativo estricto (único) de f (para $\mathbf{d} \neq \mathbf{0}$)

- Estas dos condiciones se denominan condiciones suficientes de primer y segundo orden. De este modo, todos los puntos que las cumplan serán un punto mínimo relativo
- Como $\mathbf{F}(\mathbf{x}^*)$ es una definida positiva, entonces existe una $a > 0$ tal que para toda $\mathbf{d} \neq \mathbf{0}$ se cumple que $\mathbf{d}'\mathbf{F}(\mathbf{x}^*)\mathbf{d} \geq a|\mathbf{d}|^2$. Por lo tanto, a través de la expansión de Taylor y para valores pequeños de $|\mathbf{d}|$, se puede demostrar que $\mathbf{d}'\mathbf{F}(\mathbf{x}^*)\mathbf{d} > 0$

$$f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) = \frac{1}{2}\mathbf{d}'\mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(|\mathbf{d}|^2) \geq (a/2)|\mathbf{d}|^2 + o(|\mathbf{d}|^2)$$

For small values of $|\mathbf{d}|$:

$$\Rightarrow f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) = \frac{1}{2}\mathbf{d}'\mathbf{F}(\mathbf{x}^*)\mathbf{d} \geq \left(\frac{a}{2}\right)|\mathbf{d}|^2 > 0$$

- Hay puntos extremos mínimos que no verifican estas condiciones suficientes (y, por tanto, tampoco las necesarias). No obstante, si las cumplen, es suficiente como para decir que el punto es un mínimo relativo
- La condición de que la matriz Hessiana sea semidefinida o positiva se puede comprobar de tres maneras diferentes:
- Se puede utilizar directamente la condición expresada sobre $\mathbf{b}'\mathbf{F}(\mathbf{x})\mathbf{b}$, en donde, si se saben las dimensiones de $\mathbf{F}(\mathbf{x})$, se puede derivar una expresión cuadrática que dependa de \mathbf{b} . Las dimensiones de \mathbf{b} dependen de las de $\mathbf{F}(\mathbf{x})$

$$\nabla_x f(\mathbf{x}^*)\mathbf{d} \geq 0$$

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \Rightarrow \mathbf{d}'\mathbf{F}(\mathbf{x})\mathbf{d} = 2b_1^2 + 2b_1b_2 + b_2^2 > 0$$

- Se puede utilizar el método de menores, de modo que si todos son mayores a cero (no solo los principales), la matriz es semidefinida positiva. Esta es definida positiva si todos los menores principales son mayores a cero
- Se pueden calcular los valores propios de la matriz Hessiana y comprobar si son todos positivos. Esto se hace mediante el determinante de la expresión $\mathbf{F}(\mathbf{x}) - \lambda\mathbf{I}$

$$|\mathbf{F}(\mathbf{x}) - \lambda\mathbf{I}| = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = (2-\lambda)(1-\lambda) - 1 \Rightarrow \lambda = \frac{3 \pm \sqrt{5}}{2} \geq 0$$

- Con tal de desarrollar una teoría dirigida a la caracterización de puntos óptimos globales, y no solo locales, es necesario introducir suposiciones de convexidad y resultados relacionados con estas

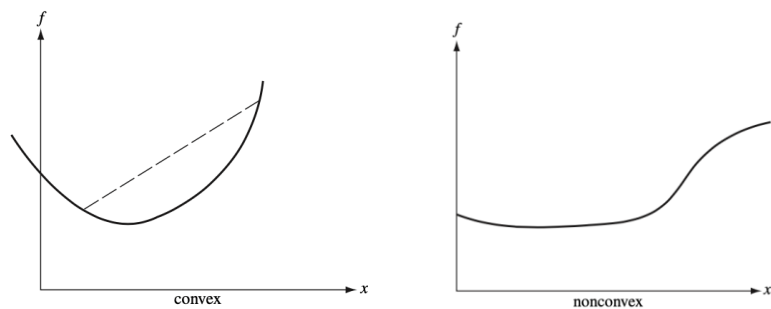
- Una función f definida en un conjunto convexo Ω se denomina convexo si para toda $x_1, x_2 \in \Omega$ y cualquier $\alpha \in [0,1]$, se da la siguiente desigualdad:

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)$$

- Si para cualquier $\alpha \in [0,1]$ y $x_1 \neq x_2$ se mantiene la siguiente desigualdad, entonces f se denomina estrictamente convexa

$$f(\alpha x_1 + (1 - \alpha)x_2) < \alpha f(x_1) + (1 - \alpha)f(x_2)$$

- Geométricamente, una función es convexa si la línea que junta dos puntos de su gráfico siempre está por encima del gráfico o de la función



- Una función g definida en un conjunto convexo Ω se cóncava si la función $f = -g$ es convexa
 - La función g es estrictamente cóncava si $-g$ es estrictamente convexa
- Siendo f_1 y f_2 funciones convexas en el conjunto convexo Ω , la función $f_1 + f_2$ es convexa en Ω
 - Siendo $x_1, x_2 \in \Omega$ y para cualquier $\alpha \in [0,1]$, se pueden obtener los siguientes resultados:

$$f_1(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f_1(x_1) + (1 - \alpha)f_1(x_2)$$

$$f_2(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f_2(x_1) + (1 - \alpha)f_2(x_2)$$

$$\Rightarrow f_1(\alpha x_1 + (1 - \alpha)x_2) + f_2(\alpha x_1 + (1 - \alpha)x_2)$$

$$\leq \alpha[f_1(x_1) + f_2(x_1)] + (1 - \alpha)[f_1(x_2) + f_2(x_2)]$$

- Siendo f una función convexa en el conjunto convexo Ω y $a > 0$, la función af es convexa en Ω

- Siendo $x_1, x_2 \in \Omega$ y para cualquier $\alpha \in [0,1]$, se pueden obtener los siguientes resultados:

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)$$

$$\Rightarrow af(\alpha x_1 + (1 - \alpha)x_2) \leq a[\alpha f(x_1) + (1 - \alpha)f(x_2)]$$

$$\Rightarrow af(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha[af(x_1)] + (1 - \alpha)[af(x_2)]$$

- A partir de esta proposición y la anterior, se puede ver como cualquier combinación positiva $a_1f_1 + a_2f_2 + \dots + a_mf_m$ de funciones convexas es convexa

- Siendo f una función convexa en un conjunto convexo Ω , el conjunto $\Gamma_c \equiv \{x : x \in \Omega, f(x) \leq c\}$ es un conjunto convexo para cualquier número real c

- Siendo $x_1, x_2 \in \Gamma_c$ y para cualquier $\alpha \in [0,1]$, se pueden obtener los siguientes resultados:

$$f(x_1) \leq c \text{ \& } f(x_2) \leq c$$

$$\Rightarrow \alpha f(x_1) \leq \alpha c \text{ \& } (1 - \alpha)f(x_2) \leq (1 - \alpha)c$$

$$\Rightarrow \alpha f(x_1) + (1 - \alpha)f(x_2) \leq \alpha c + (1 - \alpha)c = c$$

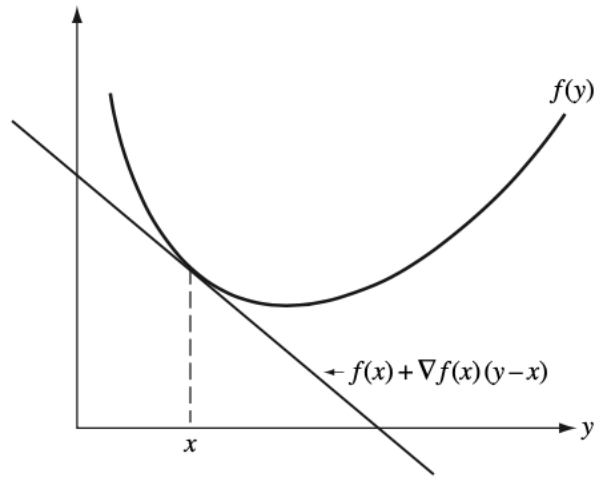
$$\Rightarrow f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2) \leq c$$

$$\Rightarrow \alpha x_1 + (1 - \alpha)x_2 \in \Gamma_c$$

- Debido a que la intersección de conjuntos convexas también es convexa, el conjunto de puntos que simultáneamente satisfacen $f_1(x) \leq c_1, f_2(x) \leq c_2, \dots, f_m(x) \leq c_m$, donde f es una función convexa, define un conjunto convexo (esto es muy utilizado en programación matemática, dado que el conjunto de restricciones se suele expresar de esta forma)

- Si la función f no solo es convexa, sino que además es diferenciable, entonces la convexidad se puede caracterizar de una forma diferente utilizando la diferenciación

- Siendo $f \in C^1$, esta es una función convexa sobre un conjunto convexo Ω si, y solo si, $f(y) \geq f(x) + \nabla_x f(x)(y - x)$ para toda $x, y \in \Omega$



- Suponiendo que f es convexa, para toda $\alpha \in (0,1]$ se pueden obtener las siguientes desigualdades:

$$\begin{aligned}
 f(\alpha y + (1 - \alpha)x) &\leq \alpha f(y) + (1 - \alpha)f(x) \\
 \Rightarrow f(x + \alpha(y - x)) &\leq \alpha(f(y) - f(x)) + f(x) \\
 \Rightarrow \frac{f(x + \alpha(y - x)) - f(x)}{\alpha} &\leq f(y) - f(x) \\
 \Rightarrow \lim_{\alpha \rightarrow 0} \frac{f(x + \alpha(y - x)) - f(x)}{\alpha} &\leq \lim_{\alpha \rightarrow 0} f(y) - f(x) \\
 \Rightarrow \nabla f(x)(y - x) &\leq f(y) - f(x) \\
 \Rightarrow f(y) &\geq f(x) + \nabla_x f(x)(y - x)
 \end{aligned}$$

- Suponiendo que $f(y) \geq f(x) + \nabla_x f(x)(y - x)$ para toda $x, y \in \Omega$ y que $\alpha \in [0,1]$, se puede fijar $x = \alpha x_1 + (1 - \alpha)x_2$ y (alternativamente) $y = x_1$ o $y = x_2$ para obtener las siguientes desigualdades:

$$\begin{aligned}
 &\begin{cases} f(x_1) \geq f(x) + \nabla_x f(x)(x_1 - x) \\ f(x_2) \geq f(x) + \nabla_x f(x)(x_2 - x) \end{cases} \\
 \Rightarrow &\begin{cases} \alpha f(x_1) \geq \alpha[f(x) + \nabla_x f(x)(x_1 - x)] \\ (1 - \alpha)f(x_2) \geq (1 - \alpha)[f(x) + \nabla_x f(x)(x_2 - x)] \end{cases} \\
 \Rightarrow &\alpha f(x_1) + (1 - \alpha)f(x_2) \geq f(x) + \nabla_x f(x)[\alpha x_1 + (1 - \alpha)x_2 - x] \\
 \Rightarrow &\alpha f(x_1) + (1 - \alpha)f(x_2) \geq f(\alpha x_1 + (1 - \alpha)x_2)
 \end{aligned}$$

- Siendo $f \in C^2$, esta es una función convexa sobre un conjunto convexo Ω que contiene al menos un punto interior si, y solo si, $\mathbf{x}'\mathbf{F}(\mathbf{x})\mathbf{x} \geq 0$ para cualquier vector diferente de cero \mathbf{x} toda $\mathbf{x} \in \Omega$

- Por el teorema de Taylor, se puede hacer una expansión de $f(\mathbf{x})$ para cualquier $\alpha \in [0,1]$ y ver que, si $\mathbf{x}'\mathbf{H}(\mathbf{x})\mathbf{x} \geq 0$, entonces la función f es convexa

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla_{\mathbf{x}}f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})'\mathbf{F}(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x})$$

$$\Rightarrow \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})'\mathbf{F}(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \geq \nabla_{\mathbf{x}}f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

$$\Rightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla_{\mathbf{x}}f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

- Suponiendo que la hessiana no es positiva semidefinida en algún punto $\mathbf{x} \in \Omega$, por continuidad de la Hessiana se puede asumir (sin pérdida de generalidad) que \mathbf{x} es un punto interior de Ω . Como existe una $\mathbf{y} \in \Omega$ tal que $(\mathbf{y} - \mathbf{x})'\mathbf{F}(\mathbf{x})(\mathbf{y} - \mathbf{x}) < 0$, y como se puede asumir, por la continuidad de la hessiana, que se puede seleccionar una \mathbf{y} tal que $(\mathbf{y} - \mathbf{x})'\mathbf{F}(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) < 0$, se puede ver como la función no es convexa si la hessiana no es semidefinida positiva
- Es importante ver, a partir de la demostración, que $(\mathbf{y} - \mathbf{x}) \in \Omega$ porque $\mathbf{y} \in \Omega$, de modo que se puede renombrar $(\mathbf{y} - \mathbf{x})$ como un vector $\mathbf{x} \in \Omega$ y obtener la misma condición. La cuestión es que se utilice un vector $\mathbf{x} \in \Omega$ cuyos componentes sean las variables de la función
- La matriz Hessiana es la generalización a E^n del concepto de curvatura de mucha función, por lo que una hessiana definida positiva es la generalización de una curvatura positiva
 - Una función es localmente convexa si su matriz Hessiana positiva semidefinida positiva en una región pequeña, y una función es localmente estrictamente convexa si su matriz Hessiana es definida positiva
 - Los resultados de suficiencia de segundo orden vistos anteriormente requieren que la función sea localmente convexa en el punto \mathbf{x}^* . Por lo tanto, aunque la teoría de optimización local, derivada solo en términos de cálculo elemental, está muy íntimamente relacionado a la convexidad (al menos localmente)
 - En consecuencia, la teoría de optimización local y global no son desarrollos paralelos si no complementarios e interactivos. Los

resultados basados en la convexidad aplican hasta a problemas no convexos en una región cerca de la solución, y conversamente, los resultados locales aplican a el punto mínimo global

- Existen tres resultados clásicos en relación a la minimización y a la maximización de funciones convexas que son muy útiles para encontrar la solución en problemas de optimización

- Siendo f una función convexa definida en un conjunto convexo Ω , el conjunto Γ donde f alcanza un mínimo es convexo, y cualquier mínimo relativo de f es global

- Si f no tiene mínimos relativos, el teorema se demuestra automáticamente. Si c_0 es un mínimo de f , entonces $\Gamma \equiv \{x : f(x) \leq c_0, x \in \Omega\}$ y este conjunto es convexo como el que se vio en una proposición anteriormente, que demuestra que se cumple el teorema

- Suponiendo que $x^* \in \Omega$ es un mínimo relativo de f pero que hay otro punto $y \in \Omega$ tal que $f(y) < f(x^*)$. En la línea o en la recta $\alpha y + (1 - \alpha)x^*$ para cualquier $\alpha \in (0,1)$, se puede obtener la siguiente desigualdad que contradice el hecho de que x^* es un punto mínimo relativo:

$$f(\alpha y + (1 - \alpha)x^*) \leq \alpha f(y) + (1 - \alpha)f(x^*) < f(x^*)$$

- Se puede parafrasear este teorema diciendo que, para funciones convexas, todos los puntos mínimos están localizados juntos (en un conjunto convexo) y todos los mínimos relativos son mínimos globales

- Siendo $f \in C^1$ una función convexa y diferenciable en un conjunto convexo Ω , si hay un punto $x^* \in \Omega$ tal que, para toda $y \in \Omega$, se da $\nabla_x f(x^*)(y - x^*) \geq 0$, entonces x^* es un punto mínimo global de f sobre Ω

- Debido a que $y - x^*$ es una dirección factible en x^* , la condición dada es equivalente a la condición necesaria de primer orden vista anteriormente. La demostración de esta proposición es inmediata, dado que se puede obtener la siguiente desigualdad a partir de uno de los resultados anteriormente vistos:

$$f(y) \geq f(x^*) + \nabla_x f(x^*)(y - x^*) \geq f(x^*)$$

- Siendo f en una función convexa definida en el conjunto cerrado y convexo Ω , si f tiene un máximo sobre Ω , entonces este se alcanza en un punto extremo de Ω

- Mucha parte del desarrollo práctico para la resolución de problemas de programación sin restricciones es el análisis y el diseño de algoritmos para resolver este tipo de problemas, de modo que es necesario revisar conceptos básicos para ello
 - Aunque los algoritmos varían mucho en motivación, aplicación y análisis, todos son algoritmos iterativos de descenso
 - Por iterativo, uno se refiere a un algoritmo que genera series de puntos, en donde cada punto se calcula en base a los puntos anteriores a este
 - Que el algoritmo sea de descenso se refiere a que mientras se genera cada punto por el algoritmo, el valor correspondiente de alguna función (evaluado en su punto más reciente) es menor
 - Un algoritmo iterativo es iniciado por especificando un punto inicial. Si un algoritmo garantiza una secuencia de puntos convergentes a una solución si para unos puntos iniciales arbitrarios, entonces el algoritmo se denomina algoritmo convergente global
 - No todos los algoritmos tienen esta propiedad deseable. Muchos de los algoritmos más importantes para resolver programas no lineales no son globalmente convergentes en su forma más pura, y ocasionalmente genera secuencias que o no convergen o no todos los puntos convergentes
 - Sin embargo, es posible modificar esos algoritmos añadiendo dispositivos especiales para garantizar convergencia global
 - Normalmente se piensa en un algoritmo como un mapeado, por lo que dado un punto x en un espacio X , el resultado de un algoritmo aplicado a x es un nuevo punto. Operado iterativamente, un algoritmo se aplica repetidamente a nuevos puntos que genera para producir toda una secuencia de puntos
 - Por lo tanto, una definición preliminar es que un algoritmo A es un mapeado tomando puntos en un espacio X en otros puntos en X . Operado iterativamente, el algoritmo A iniciado en $x_0 \in X$ generaría la secuencia $\{x_k\}$ definido por $x_{k+1} = A(x_k)$
 - En la práctica, un mapeado A puede estar explícitamente definido por una expresión matemática simple o puede estar definido implícitamente por un programa de ordenador complejo, por ejemplo. Dado un vector de insumo, ambos definen un resultado correspondiente, por lo que se puede generalizar aún más la definición de algoritmo

- Un algoritmo A es un mapeado definido en un espacio X que asigna a cada punto $x \in X$ un subconjunto de X
 - El término espacio se puede interpretar vagamente. Normalmente X será un espacio vectorial E^n , pero puede ser solo un subconjunto de E^n o un espacio métrico más concreto
 - El aspecto más importante de esta definición es que el mapeado A , más que ser un mapeado de un punto a un punto, es un mapeado de un punto a un conjunto de X
 - Un algoritmo A genera una secuencia de puntos de la siguiente manera: dado un punto $x_k \in X$, el algoritmo permite obtener $A(x_k)$, el cual es un subconjunto de X , y se selecciona arbitrariamente un punto x_{k+1} . De este modo, dado un punto inicial x_0 , el algoritmo genera secuencias a través de la iteración $x_{k+1} \in A(x_k)$
 - Como es obvio, a diferencia del caso en el que el algoritmo es un mapeado de un punto a un punto, no es posible predecir cual será el siguiente punto solo x_0 . Este grado de incertidumbre se ha diseñado para reflejar la incertidumbre que se puede tener en la práctica (específicamente para los detalles de un algoritmo)
- Con tal de describir la idea de que es un algoritmo de descenso, primero se tiene que definir el conjunto solución Γ del espacio X . La idea básica de una función de descenso es que para puntos fuera del conjunto solución, un solo paso del algoritmo resulta en un decrecimiento del valor de la función de descenso
 - Siendo $\Gamma \subset X$ un conjunto de solución dado y siendo A un algoritmo sobre X , una función de valores reales Z sobre X es una función de descenso para Γ y A si satisface las siguientes condiciones:

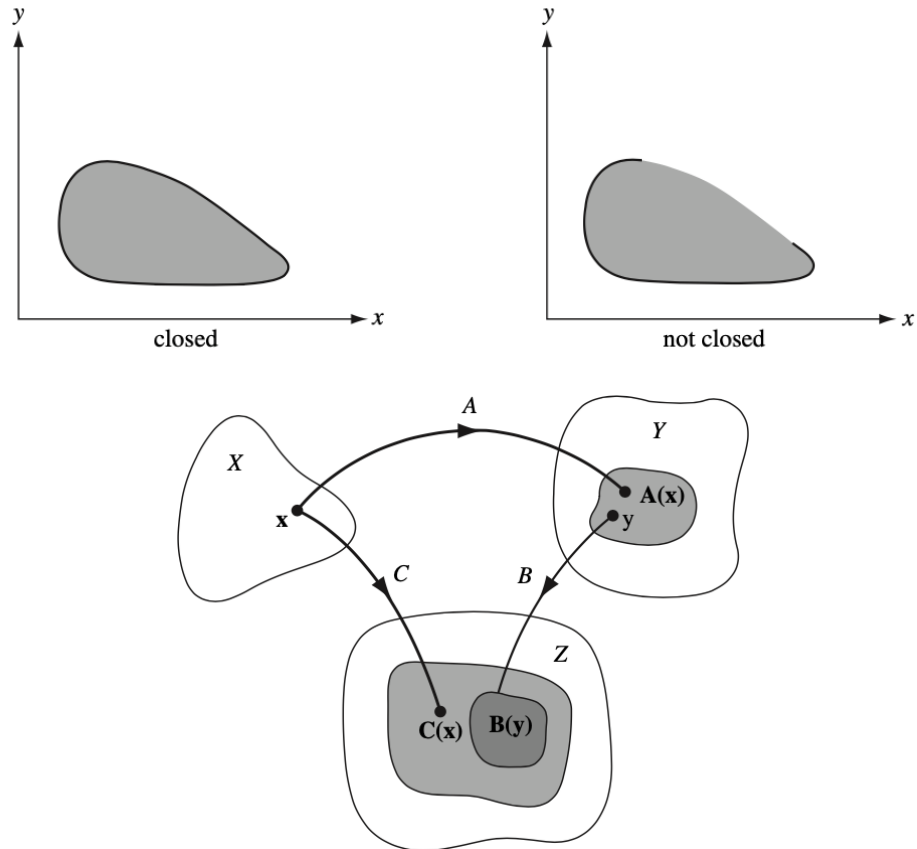
$$x \notin \Gamma \ \& \ y \in A(x) \Rightarrow Z(y) < Z(x)$$

$$x \in \Gamma \ \& \ y \in A(x) \Rightarrow Z(y) \leq Z(x)$$

- Hay muchas maneras de definir un conjunto de solución, un algoritmo y una función de descenso. Un marco natural para el problema $\min\{f(x): x \in \Omega\}$ es hacer que Γ sea el conjunto de puntos minimizadores y definir el algoritmo A en Ω de tal manera que f decrezca en cada paso y, por lo tanto, sirve como una función de descenso

- Otra posibilidad para problemas sin restricciones es dejar que Γ sea el conjunto de puntos x satisfaciendo $\nabla_x f(x) = \mathbf{0}$, por lo que se podría diseñar un algoritmo para el cual $|\nabla_x f(x)|$ sirve como una función de descenso o para el cual $f(x)$ sirve como función del descenso

○ MAPEADOS CERRADOS



- El teorema de la convergencia global se usa para establecer la convergencia en las siguiente situación general: existe un conjunto de solución Γ y los puntos se generan acorde al algoritmo $x_{k+1} \in A(x_k)$, haciendo que cada punto nuevo decrezca estrictamente una función de descenso Z hasta que se llega a Γ
 - Un algoritmo adecuado se encuentra cuando se generan puntos tal que cada punto nuevo estrictamente reduce el valor de la función objetivo
 - Entonces, bajo condiciones apropiadas, la secuencia converge a un conjunto de solución, y el teorema de convergencia global establece las condiciones técnicas por las cuales la convergencia está garantizada
 - Siendo A un algoritmo sobre X , Γ el conjunto solución y suponiendo que dado un punto x_0 la secuencia $\{x_k\}_{k=0}^{\infty}$ se genera satisfaciendo $x_{k+1} \in A(x_k)$, si se cumplen las siguientes condiciones, el límite de cualquier subsecuencia convergente de $\{x_k\}$ es una solución:

- Todos los puntos x_k están contenidos en un conjunto compacto $S \subset X$
- Existe una función continua Z sobre X tal que si $x \notin \Gamma$, entonces $Z(y) < Z(x)$ para toda $y \in A(x)$
- Existe una función continua Z sobre X tal que si $x \in \Gamma$, entonces $Z(y) \leq Z(x)$ para toda $y \in A(x)$
- El mapeado A es cerrado en puntos fuera del conjunto solución Γ

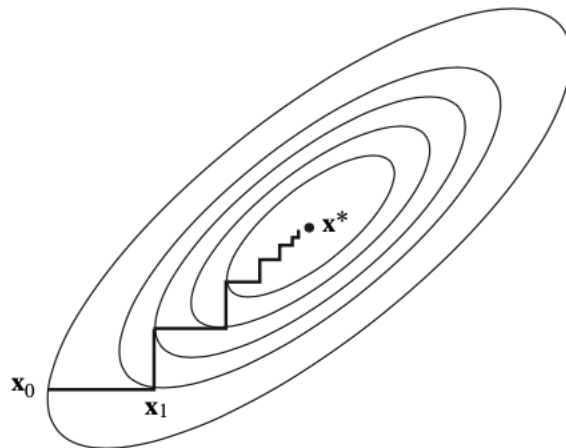
Los métodos de descenso básicos

- Existen técnicas básicas que se usan para resolver problemas de minimización no restringida de manera iterativa. Estas son importantes para la aplicación práctica porque ofrecen alternativas para obtener soluciones y establecen mesetas de referencia en cuanto a la dificultad de implementación
 - Hay una estructura subyacente fundamental para la mayoría de algoritmos de descenso: se comienza en un punto inicial, se determina (acorde a una regla fija) la dirección del movimiento, y se mueve en esa dirección a un mínimo (relativo) de la función objetivo en esa línea
 - En el nuevo punto, una nueva dirección se determina, y el proceso se repite
 - Las diferencias principales entre algoritmos es la regla fija que se aplica para determinar las direcciones de movimiento sucesivas. Una vez la selección se hace, todos los algoritmos hacen que uno se mueva al punto mínimo de la línea correspondiente
 - El proceso de determinar el punto mínimo en una línea dada se denomina búsqueda de línea. Para funciones no lineales que no pueden ser minimizadas analíticamente, este proceso se realiza buscando el punto mínimo, de una manera inteligente, a lo largo de la línea
 - Las técnicas de búsqueda de línea, que en su mayoría son procedimientos para resolver problemas de minimización unidimensionales, forman las bases de los algoritmos de programación no lineal
 - Los problemas de dimensiones mayores se solucionan ejecutando una secuencia de búsquedas de línea

- Uno de los algoritmos más viejos y conocidos para minimizar una función de varias variables es el método de descenso más pronunciado o *steepest descent* (también llamado método del gradiente o *gradient method*)
 - Siendo f una función con primeras derivadas parciales continuas en E^n , el método del gradiente se define por el siguiente algoritmo iterativo:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla_x f(\mathbf{x})$$

- En este caso, α_k , el cual se llama longitud de pasa, es un escalar no negativo que minimiza $f[\mathbf{x}_k - \alpha_k \nabla_x f(\mathbf{x})]$ (de modo que se cumpla que $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$). El opuesto del gradiente $-\nabla_x f(\mathbf{x})$ hace de dirección de descenso o de búsqueda (del nuevo punto \mathbf{x}_{k+1})
- Desde el punto \mathbf{x}_k se busca en la dirección del gradiente negativo $-\nabla_x f(\mathbf{x})$ (en la dirección de búsqueda) a un punto mínimo en la línea, el cual es \mathbf{x}_{k+1}



- En términos formales, el algoritmo $\mathbf{A} : E^n \rightarrow E^n$ lo cual permite obtener $\mathbf{x}_{k+1} \in \mathbf{A}(\mathbf{x}_k)$ puede descomponerse en la forma $\mathbf{A} = \mathbf{S}\mathbf{G}$, donde $\mathbf{G} : E^n \rightarrow E^{2n}$ se define por $\mathbf{G}(\mathbf{x}) = (\mathbf{x}, -\mathbf{g}(\mathbf{x}))$, dando el punto inicial y la dirección de la línea de búsqueda y $\mathbf{S} : E^{2n} \rightarrow E^n$ es la línea de búsqueda definida anteriormente
- Se ha demostrado que \mathbf{S} es un algoritmo cerrado si $\nabla_x f(\mathbf{x}) \neq \mathbf{0}$, y es claro que \mathbf{G} es continua, de modo que se cumple el corolario visto anteriormente para la convergencia global
 - Se define el conjunto de solución como el conjunto de puntos \mathbf{x} donde $\nabla_x f(\mathbf{x}) = \mathbf{0}$. Entonces $Z(\mathbf{x}) = f(\mathbf{x})$ es una función de descenso para \mathbf{A} , dado que para $\nabla_x f(\mathbf{x}) \neq \mathbf{0}$ se cumple la siguiente desigualdad:

$$\lim_{0 \leq \alpha < \infty} f(\mathbf{x} - \alpha \mathbf{g}(\mathbf{x})) < f(\mathbf{x})$$

- Entonces, por el teorema de convergencia global, si la secuencia $\{\mathbf{x}_k\}$ está acotada, tendrá puntos límite y cada uno de estos será solución

Las condiciones de minimización restringida

- Existen condiciones necesarias y suficientes para los puntos de solución de problemas de minimización restringida. Estas condiciones caracterizan las soluciones y también permiten definir los multiplicadores de Lagrange y una matriz Hessiana, que son los fundamentos para el análisis de problemas restringidos
 - Normalmente se lidia con problemas no lineales generales de la siguiente forma:

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & h_1(\mathbf{x}) = 0 \quad g_1(\mathbf{x}) \leq 0 \\ & h_2(\mathbf{x}) = 0 \quad g_2(\mathbf{x}) \leq 0 \\ & \vdots \quad \quad \quad \vdots \\ & h_m(\mathbf{x}) = 0 \quad g_p(\mathbf{x}) \leq 0 \\ & \mathbf{x} \in \Omega \subset E^n, \end{array}$$

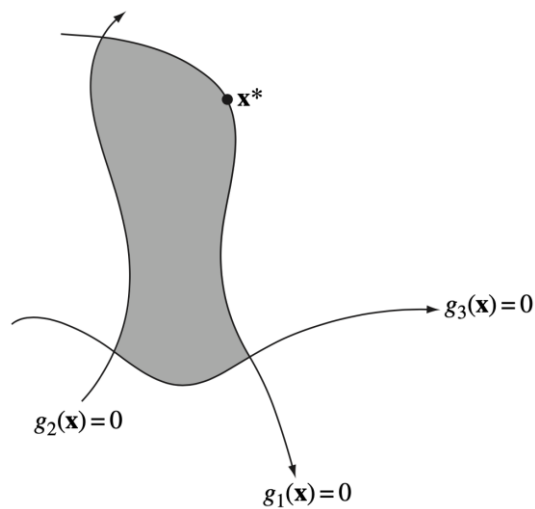
- En este programa, $m \leq n$ y las funciones f, h_i y g_j para $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, p$ son continuas y se asume que poseen derivadas parciales de segundo orden
- Para simplificar la notación, se utilizan vectores y funciones vectoriales $\mathbf{h} = (h_1, h_2, \dots, h_m)$ y $\mathbf{g} = (g_1, g_2, \dots, g_p)$

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \Omega. \end{array}$$

- Las restricciones $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ y $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ se conocen como restricciones funcionales, mientras que las restricciones como $\mathbf{x} \in \Omega$ se conocen como restricciones de conjunto. Un punto que satisface las condiciones funcionales se denomina punto factible
- Igual que antes, no se pone mucho énfasis en las restricciones de conjunto debido a que en la mayoría de casos $\Omega = E^n$ o la solución está en el interior de Ω
 - Un concepto que permite obtener una visión más completa y simplificar mucho el desarrollo teórico es el concepto de restricción activa. Una

restricción de desigualdad $g_i(\mathbf{x}) \leq 0$ es activa en un punto factible \mathbf{x} si $g_i(\mathbf{x}) = 0$ e inactiva si $g_i(\mathbf{x}) < 0$

- Por convención, se denomina activa a la restricción $h_i(\mathbf{x}) = 0$ para cualquier punto factible
- La restricción activas en un punto factible restringen el dominio de factibilidad de las vecindades de \mathbf{x} . Por lo tanto, al estudiar las propiedades de un mínimo local, la atención se puede enfocar solo en las restricciones activas



- Si las condiciones activas en el problema se supieran *a priori*, la solución sería un mínimo local en el problema definido ignorando las condiciones inactivas y tratando todas las restricciones activas como restricciones de igualdad. Por lo tanto, respecto a soluciones locales, el problema se podría entender como un problema con restricciones de igualdad únicamente
 - Esta observación sugiere que la mayoría de desarrollo teórico se puede derivar considerando restricciones de igualdad y después añadiendo elementos para tener en cuenta la selección de restricciones activas
- Antes de poder entender el concepto de plano tangente, es necesario saber la noción de curva. Una curva en una superficie S es una familia de puntos $\mathbf{x}(t) \in S$ continuamente parametrizados por t en $a \leq t \leq b$
- La curva es diferenciable si $(d/dt)\mathbf{x}(t)$ existe, y es doblemente diferenciable si $(d^2/dt^2)\mathbf{x}(t)$ existe
 - Una curva $\mathbf{x}(t)$ pasa por un punto \mathbf{x}^* si $\mathbf{x}^* = \mathbf{x}(t^*)$ para una t^* en $a \leq t^* \leq b$

- La derivada de la curva en \mathbf{x}^* se define como $(d/dt)\mathbf{x}(t^*)$, que es un vector en E^n
- Un conjunto de ecuaciones de igualdad sobre E^n define un subconjunto de E^n que se puede ver como una hipersuperficie S

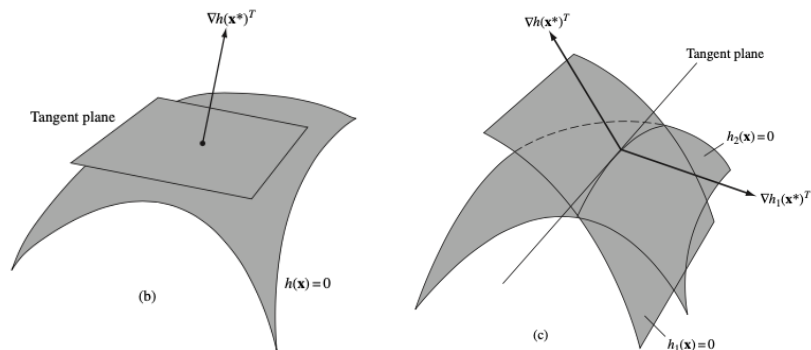
$$h_1(\mathbf{x}) = 0$$

$$h_2(\mathbf{x}) = 0$$

$$\vdots$$

$$h_m(\mathbf{x}) = 0$$

- Si las restricciones son regulares en cualquier punto, esta superficie S tiene dimensiones $n - m$. Si además las funciones $\mathbf{h}(\mathbf{x})$ pertenecen a C^1 (como se asume), la superficie S obtenida es suave o *smooth*
- Existe un plano tangente a un punto asociado a un punto en la superficie suave. Si se consideran todas las curvas diferenciables en S pasando por el punto \mathbf{x}^* , el plano tangente en \mathbf{x}^* se define como el conjunto de derivadas en \mathbf{x}^* de todas estas curvas diferenciables, el cual es un subespacio de E^n
- Para superficies definidas a través de un conjunto de restricciones como las anteriores, el problema de obtener una representación explícita para el plano tangente es fundamental
- Idealmente, se querría expresar el plano tangente en términos de las derivadas de las funciones h_i que define la superficie, por lo que se introduce el subespacio $M = \{\mathbf{y} : \nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$ e investigar para que condiciones M es el plano tangente en \mathbf{x}^*



- En este caso, $\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}^*)$ es la matriz jacobiana de las restricciones de igualdad del problema

$$J = \nabla_x \mathbf{h}(\mathbf{x}^*) = \begin{bmatrix} \nabla_x' h_1(\mathbf{x}^*) \\ \vdots \\ \nabla_x' h_m(\mathbf{x}^*) \end{bmatrix} = \begin{bmatrix} \frac{\partial h_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial h_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial h_m(\mathbf{x}^*)}{\partial x_n} \end{bmatrix}$$

- Un punto \mathbf{x}^* que satisface las restricciones $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ se denomina punto regular de las restricciones si los vectores gradientes $\nabla_x h_1(\mathbf{x}^*), \nabla_x h_2(\mathbf{x}^*), \dots, \nabla_x h_m(\mathbf{x}^*)$ son linealmente independientes

- Si \mathbf{h} es afín, de modo que $\mathbf{h}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, la regularidad es equivalente a tener un rango igual a m , y esta condición es independiente de \mathbf{x}
- Para poder demostrar si un punto es regular o no, se tiene que comprobar que la matriz jacobiana tiene rango completo, lo cual se puede hacer obteniendo la forma escalonada de la matriz

$$J(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial h_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial h_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial h_m(\mathbf{x}^*)}{\partial x_n} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

- En general, es posible caracterizar el plano tangente en términos de los gradientes de las restricciones en los puntos regulares. En un punto regular \mathbf{x}^* en una superficie S definida por $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, el plano tangente es igual a $M = \{\mathbf{y} : \nabla_x \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$
- Es importante reconocer que la condición de ser un punto regular no es una condición sobre la superficie en sí, sino en su representación en términos de \mathbf{h} . El plano tangente es definido independientemente de su representación, mientras que M si depende de su representación
- La derivación de condiciones necesarias y suficientes para que un punto sea un mínimo local con restricciones de igualdad es simple habiendo visto la representación del plano tangente, asumiendo solo que $f, \mathbf{h} \in C^2$
 - Siendo \mathbf{x}^* un punto regular de las restricciones $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ y un punto extremo local (un mínimo o un máximo) de f sujeta a estas restricciones, entonces toda $\mathbf{y} \in E^n$ satisfaciendo $\nabla_x \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}$ también tienen que satisfacer $\nabla_x f(\mathbf{x}^*)\mathbf{y} = 0$
 - Siendo \mathbf{y} cualquier vector en el plano tangente en \mathbf{x}^* y siendo $\mathbf{x}(t)$ cualquier curva suave en la superficie de las restricciones

pasando a través de \mathbf{x}^* con la derivada de \mathbf{y} en \mathbf{x}^* . Esto se traduce en lo siguiente:

$$\mathbf{x}(0) = \mathbf{x}^* \quad \dot{\mathbf{x}}(0) = \mathbf{y} \quad \mathbf{h}(\mathbf{x}(t)) = \mathbf{0}$$

$$\text{for } -a \leq t \leq a \text{ \& } a > 0$$

- Como \mathbf{x}^* es un punto regular, el plano tangente es idéntico al conjunto de vectores \mathbf{y} satisfaciendo $\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) \mathbf{y} = \mathbf{0}$. Por lo tanto, como \mathbf{x}^* es un punto extremo local de f , se respeta la siguiente igualdad:

$$\left. \frac{d}{dt} f(\mathbf{x}(t)) \right|_{t=0} = 0 \Leftrightarrow \nabla_{\mathbf{x}} f(\mathbf{x}^*) \mathbf{y} = 0$$

- Este lema dice que $\nabla_{\mathbf{x}} f(\mathbf{x}^*)$ es ortogonal al plano tangente
- Siendo \mathbf{x}^* un punto regular de las restricciones $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ y un punto extremo local (un mínimo o un máximo) de f sujeta a estas restricciones, entonces existe un vector $\boldsymbol{\lambda} \in E^m$ tal que $\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \boldsymbol{\lambda}' \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$
 - Esta es la condición necesaria de primer orden para problemas con restricciones de igualdad. Si un punto no es regular y/o no es factible (no cumple las ecuaciones e inecuación), entonces la condición necesaria no aplica
 - A partir del lema anterior, se puede concluir que el valor de un programa lineal como el siguiente es nulo:

$$\begin{array}{ll} \text{maximize} & \nabla f(\mathbf{x}^*) \mathbf{y} \\ \text{subject to} & \nabla \mathbf{h}(\mathbf{x}^*) \mathbf{y} = \mathbf{0} \end{array}$$

- Por lo tanto, por el teorema de dualidad de los programas lineales, el problema dual es factible. Específicamente, existe un vector $\boldsymbol{\lambda} \in E^m$ tal que $\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \boldsymbol{\lambda}' \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$
- Las condiciones necesarias de primer orden $\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \boldsymbol{\lambda}' \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ junto con las restricciones $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ dan un total de $n + m$ ecuaciones (generalmente no lineales) en las $n + m$ variables comprometiendo \mathbf{x}^* , $\boldsymbol{\lambda}$
 - Por lo tanto, las condiciones necesarias son un conjunto completo dado que, localmente, determinan una solución única (número igual de variables y de ecuaciones)

- A estas alturas, es necesario introducir la función Lagrangiana asociada al problema restringido, definida como $L(\mathbf{x}, \boldsymbol{\lambda}) \equiv f(\mathbf{x}) + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x})$. Las condiciones necesarias de primer orden, por tanto, se pueden expresar de la siguiente forma:

$$\nabla_{\mathbf{x}}L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0} \quad \nabla_{\boldsymbol{\lambda}}L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$$

- Esta última forma es la más sencilla en la práctica, dado que implica solo obtener el gradiente para cada una de las variables e igualarlas a cero
- Las condiciones necesarias de segundo orden se pueden resumir de la siguiente manera:

- Suponiendo que \mathbf{x}^* es un mínimo local de f sujeta a $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ y que \mathbf{x}^* es un punto regular de esas restricciones, entonces hay alguna $\boldsymbol{\lambda} \in E^m$ tal que se cumple la siguiente igualdad:

$$\nabla_{\mathbf{x}}L(\mathbf{x}, \boldsymbol{\lambda}) = \nabla_{\mathbf{x}}f(\mathbf{x}^*) + \boldsymbol{\lambda}'\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

- Si se denota por M el plano tangente $M = \{\mathbf{y} : \nabla\mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$, entonces la siguiente matriz es positiva semidefinida en M , de modo que $\mathbf{y}'L(\mathbf{x}^*)\mathbf{y} \geq 0$ para toda $\mathbf{y} \in M$ tal que $\mathbf{y} \neq \mathbf{0}$:

$$L(\mathbf{x}^*) = F(\mathbf{x}^*) + \boldsymbol{\lambda}'\nabla_{\mathbf{x}}H(\mathbf{x}^*)$$

- Si un punto no es regular y/o no es factible (no cumple las ecuaciones e inecuación), entonces las condiciones necesarias no aplican
- Por lo tanto, se puede ver que es necesario que \mathbf{x}^* sea un punto regular y la Hessiana puede ser solo semidefinida positiva. Para comprobar esto, es necesario definir una \mathbf{y} que satisficiera $\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}$, resolviendo un sistema de m variables y m ecuaciones para satisfacer la ortogonalidad, y trabajar con este vector o utilizar el método de comprobar los menores:

$$\begin{aligned} \nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0} &\Rightarrow \begin{cases} \sum_{j=1}^m \frac{\partial h_1(\mathbf{x})}{\partial x_j} y_j = 0 \\ \dots \\ \sum_{j=1}^m \frac{\partial h_m(\mathbf{x})}{\partial x_j} y_j = 0 \end{cases} \Rightarrow \mathbf{y} = (y_1, y_2, \dots, y_n)' \\ &\Rightarrow \mathbf{y}'L(\mathbf{x}^*)\mathbf{y} \geq 0 \end{aligned}$$

- La demostración de las condiciones necesarias de segundo orden son las siguientes:

- Para cualquier curva doblemente diferenciable en la superficie de restricciones S a través de \mathbf{x}^* (con $\mathbf{x}(0) = \mathbf{x}^*$) se obtiene la siguiente condición:

$$\left. \frac{d^2}{dt^2} f(\mathbf{x}(t)) \right|_{t=0} = \dot{\mathbf{x}}(0)' \mathbf{F}(\mathbf{x}^*) \dot{\mathbf{x}}(0) + \nabla_{\mathbf{x}} f(\mathbf{x}^*) \ddot{\mathbf{x}}(0) \geq 0$$

- Diferenciando la relación $\lambda' \mathbf{h}(\mathbf{x}(t)) = 0$ dos veces y sumando el resultado a la condición anterior, se obtiene la siguiente condición:

$$\dot{\mathbf{x}}(0)' \lambda' \mathbf{H}(\mathbf{x}^*) \dot{\mathbf{x}}(0) + \lambda' \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) \ddot{\mathbf{x}}(0) = 0$$

$$\Rightarrow \left. \frac{d^2}{dt^2} f(\mathbf{x}(t)) \right|_{t=0} = \dot{\mathbf{x}}(0)' \mathbf{L}(\mathbf{x}^*) \dot{\mathbf{x}}(0) \geq 0$$

- Debido a que $\dot{\mathbf{x}}(0)$ es arbitraria en M , entonces se demuestra el teorema
- Suponiendo que hay un punto \mathbf{x}^* satisfaciendo $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$, que existe alguna $\lambda \in E^m$ tal que $\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \lambda' \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ y que la matriz $\mathbf{L}(\mathbf{x}^*)$ es una definida positiva en $M = \{\mathbf{y} : \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) \mathbf{y} = \mathbf{0}\}$, entonces \mathbf{x}^* es un punto mínimo local estricto (único) de f sujeto a $\mathbf{h}(\mathbf{x}) = \mathbf{0}$
 - Estas condiciones son las condiciones suficientes de segundo orden para problemas con restricciones de igualdad. Por lo tanto, se puede ver que no se requiere que \mathbf{x}^* sea un punto regular y la Hessiana tiene que ser positiva definida, no solo semidefinida
 - Si \mathbf{x}^* no es punto mínimo relativo estricto (único), entonces existe una secuencia de puntos factibles $\{\mathbf{y}_k\}$ convergiendo a \mathbf{x}^* tal que para cada k , $f(\mathbf{y}_k) \leq f(\mathbf{x}^*)$
 - Expresando cada \mathbf{y}_k en la forma $\mathbf{y}_k = \mathbf{x}^* + \delta_k \mathbf{s}_k$ donde $\mathbf{s}_k \in E^n$, $|\mathbf{s}_k| = 1$ y $\delta_k > 0$ para cada k , entonces $\delta_k \rightarrow 0$ y la secuencia $\{\mathbf{s}_k\}$, siendo acotada, tiene que tener una subsecuencia convergente a alguna \mathbf{s}^* . Por conveniencia notacional, se asume que la secuencia $\{\mathbf{s}_k\}$ tiende a \mathbf{s}^*
 - Además, se tiene que $\mathbf{h}(\mathbf{y}_k) - \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$, y dividiendo por δ_k y permitiendo que $k \rightarrow \infty$, se puede ver como $\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) \mathbf{s}^* = \mathbf{0}$
- Una explicación más intuitiva del origen de las condiciones necesarias y suficientes para problemas con restricciones de igualdad se da a través de argumentos de factibilidad y mejora

- Usando series de Taylor de primer orden y suponiendo que se está en un punto factible de la región factible y que este se mueve en la dirección $\mathbf{y} \neq \mathbf{0}$, es posible explicar las condiciones que harían mover el punto actual a uno óptimo

- Para mantener la factibilidad de las restricciones de igualdad, es necesario que $\nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}$ porque se tiene que mantener que $\mathbf{h}(\mathbf{x} + \mathbf{y}) = \mathbf{0}$ (las restricciones se siguen cumpliendo, aunque se haga un incremento de \mathbf{y})

$$\mathbf{h}(\mathbf{x} + \mathbf{y}) \approx \mathbf{h}(\mathbf{x}) + \nabla \mathbf{h}(\mathbf{x})\mathbf{y} \quad \& \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

$$\Rightarrow \nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0} \Rightarrow \mathbf{h}(\mathbf{x} + \mathbf{y}) \approx \mathbf{0}$$

- Además, moverse en la dirección \mathbf{y} (de descenso) mejora el valor de la función objetivo, de modo que $\nabla f(\mathbf{x})\mathbf{y} < 0$ para que $f(\mathbf{x}) > f(\mathbf{x} + \mathbf{d})$

$$f(\mathbf{x}) > f(\mathbf{x} + \mathbf{d}) \quad \& \quad f(\mathbf{x} + \mathbf{d}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})\mathbf{y}$$

$$\Rightarrow f(\mathbf{x}) > f(\mathbf{x}) + \nabla f(\mathbf{x})\mathbf{y} \Rightarrow \nabla f(\mathbf{x})\mathbf{y} < 0$$

- Por lo tanto, una condición necesaria de optimalidad de un punto es que no exista una dirección \mathbf{y} que cumpla $\nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}$ y $\nabla f(\mathbf{x})\mathbf{y} < 0$

- Si $\nabla f(\mathbf{x})\mathbf{y} < 0$, entonces existe una dirección en la que se puede minimizar $f(\mathbf{x})$ manteniendo que $\nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}$, pero si se da el caso contrario, en donde $\nabla f(\mathbf{x})\mathbf{y} > 0$, se podría coger la dirección $-\mathbf{y}$ para que $f(\mathbf{x}) > f(\mathbf{x}) - \nabla f(\mathbf{x})\mathbf{y}$ y manteniendo la factibilidad debido a que $\nabla \mathbf{h}(\mathbf{x})\mathbf{y} = -\nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}$
- Por lo tanto, para que un punto \mathbf{x} se óptimo, es necesario que se cumpla que $\nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}$ y que $\nabla f(\mathbf{x})\mathbf{y} = 0$, lo cual es equivalente a decir que $\nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}$ y que $\nabla f(\mathbf{x}) = \mathbf{0}$ porque $\mathbf{y} \neq \mathbf{0}$
- Esto es lo mismo que decir que $\nabla f(\mathbf{x})$ y $\nabla \mathbf{h}(\mathbf{x})$ son vectores paralelos (ya que ambos resultan en el vector $\mathbf{0}$), lo cual se puede expresar con la condición necesaria de primer orden:

$$\nabla f(\mathbf{x}) = -\lambda \nabla \mathbf{h}(\mathbf{x}) \Rightarrow \nabla f(\mathbf{x}) + \lambda \nabla \mathbf{h}(\mathbf{x}) = \mathbf{0}$$

$$\text{for some } \lambda \in E^m$$

- Sabiendo que una dirección factible \mathbf{y} será aquella en donde se cumpla $\mathbf{h}(\mathbf{x}^* + \mathbf{y}) = \mathbf{0}$ y $\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}$, se puede ver que la lagrangiana también debe respetar esto:

$$L(\mathbf{x}^* + \mathbf{y}, \lambda^*) = f(\mathbf{x}^* + \mathbf{y}) + \lambda^* h(\mathbf{x}^* + \mathbf{y}) = f(\mathbf{x}^* + \mathbf{y})$$

- En consecuencia, por expansión de la serie de Taylor de la lagrangiana $L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$, es posible obtener la siguiente aproximación:

$$\begin{aligned} L(\mathbf{x}^* + \mathbf{y}, \lambda^*) &\approx L(\mathbf{x}^*, \lambda^*) + \nabla L(\mathbf{x}^*, \lambda^*) \mathbf{y} + \frac{1}{2} \mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y} = \\ &= f(\mathbf{x}^*) + \frac{1}{2} \mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y} \end{aligned}$$

- Por lo tanto, para que el punto \mathbf{x}^* sea un mínimo relativo, $L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$ debe cumplir que $L(\mathbf{x}^*, \lambda^*) \leq L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$ (ir en la dirección de descenso no mejora el valor de la función objetivo), por lo que $\mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y} \geq 0$

$$L(\mathbf{x}^*, \lambda^*) \leq L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \Rightarrow f(\mathbf{x}^*) \leq f(\mathbf{x}^*) + \frac{1}{2} \mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y}$$

$$\Rightarrow \mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y} \geq 0 \text{ for } \mathbf{y}: \nabla h(\mathbf{x}^*) \mathbf{y} = 0$$

- Si, en cambio, el punto \mathbf{x}^* es un mínimo estricto, entonces $L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$ debe cumplir que $L(\mathbf{x}^*, \lambda^*) < L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$ (ir en la dirección de descenso siempre incrementa el valor de la función objetivo), por lo que $\mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y} > 0$

$$L(\mathbf{x}^*, \lambda^*) \leq L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \Rightarrow f(\mathbf{x}^*) \leq f(\mathbf{x}^*) + \frac{1}{2} \mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y}$$

$$\Rightarrow \mathbf{y}' \nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*) \mathbf{y} \geq 0 \text{ for } \mathbf{y}: \nabla h(\mathbf{x}^*) \mathbf{y} = 0$$

- Al considerar problemas de optimización con restricciones de igualdad y de desigualdad, el enfoque que se tiene que utilizar es diferente, y es necesario utilizar las condiciones de Karush-Kuhn-Tucker, asumiendo solo que $f, \mathbf{h}, \mathbf{g} \in C^2$

- Normalmente se lidia con problemas no lineales generales de la siguiente forma:

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ &&& \mathbf{g}(\mathbf{x}) \leq \mathbf{0}. \end{aligned}$$

- Se asume que f y \mathbf{h} son como antes y que \mathbf{g} es una función p -dimensional
- Hay varias teorías sobre como enfocar este problema, basadas en varias condiciones de regularidad o calificaciones de restricciones,

las cuales se dirigen a obtener proposiciones definitivas generales de condiciones necesarias y suficientes. No obstante, estos resultados no se pueden obtener como extensiones de la teoría derivada para restricciones de igualdad

- Siendo \mathbf{x}^* un punto satisfaciendo $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ y $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$ y siendo J el conjunto de índices j para los cuales $g_j(\mathbf{x}^*) = 0$, entonces \mathbf{x}^* es un punto regular de las restricciones si los vectores gradientes $\nabla_{\mathbf{x}} h_j(\mathbf{x}^*)$ y $\nabla_{\mathbf{x}} g_j(\mathbf{x}^*)$ para $1 \leq i \leq m$ y $j \in J$ son linealmente independientes
 - Para poder demostrar si un punto es regular o no, se tiene que comprobar que la matriz jacobiana tiene rango completo, lo cual se puede hacer obteniendo la forma escalonada de la matriz

$$J(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial h_1(\mathbf{x}^*)}{\partial x_1} & \dots & \frac{\partial h_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m(\mathbf{x}^*)}{\partial x_1} & \dots & \frac{\partial h_m(\mathbf{x}^*)}{\partial x_n} \\ \frac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \dots & \frac{\partial g_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_J(\mathbf{x}^*)}{\partial x_1} & \dots & \frac{\partial g_J(\mathbf{x}^*)}{\partial x_n} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}$$

- Siguiendo la definición de restricciones activas dadas anteriormente, un punto \mathbf{x}^* es regular si los gradientes de las restricciones activas son linealmente independientes
- Si se considera un punto \mathbf{x}^* *ex ante*, solo se tiene que utilizar el criterio del jacobiano. Sin embargo, si no se da, entonces se tiene que plantear el problema de minimización restringida y ver cuales serían las restricciones con desigualdad activas para poder aplicar esto
- Siendo \mathbf{x}^* un mínimo relativo para el problema $\min\{f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \mathbf{0} \ \& \ \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$ y suponiendo que \mathbf{x}^* es un punto regular para las restricciones, entonces existe un vector $\boldsymbol{\lambda} \in E^m$ y un vector $\boldsymbol{\mu} \in E^p$ con $\boldsymbol{\mu} \geq \mathbf{0}$ tal que se cumplen las siguientes igualdades:

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \boldsymbol{\lambda}' \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}' \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

$$\boldsymbol{\mu}' \mathbf{g}(\mathbf{x}^*) = 0 \quad \boldsymbol{\mu} \geq \mathbf{0}$$

- Estas condiciones son las condiciones necesarias de primer orden para problemas con restricciones de desigualdad, llamadas condiciones de Karush-Kuhn-Tucker
- Si un punto no es regular y/o no es factible (no cumple las ecuaciones e inecuación), entonces las condiciones necesarias no aplican
- Las condiciones de segundo orden para problemas con restricciones de desigualdad, tanto las necesarias como las suficientes, se derivan esencialmente de la consideración del problema con restricciones de igualdad

- El plano tangente apropiado para este tipo de problemas es el plano tangente a las restricciones activas, que en este caso será el siguiente conjunto M

$$M = \{y : \nabla_x h(x^*)y = 0 \text{ \& } \nabla_x g_j(x^*)y = 0 \text{ for } \forall j \in J\}$$

$$\text{where } J = \{j : g_j(x^*) = 0\}$$

- Por analogía con la situación sin restricciones, la hipótesis requerida es que $L(x^*)$ sea definida positiva en el plano tangente M , pero esto no es suficiente con restricciones con desigualdades degeneradas (restricciones de desigualdad activas teniendo un cero asociado a sus multiplicadores de Lagrange) y se necesita que sea definida positiva en un subespacio más grande que M , el cual será M'

$$M' = \{y : \nabla_x h(x^*)y = 0 \text{ \& } \nabla_x g_j(x^*)y = 0 \text{ for } \forall j \in J \text{ \& } \mu_j > 0\}$$

$$\text{where } J = \{j : g_j(x^*) = 0\}$$

- Si todas las restricciones de desigualdad activas tienen multiplicadores de Lagrange estrictamente positivos (no hay desigualdades degeneradas), entonces el conjunto J incluye todas las desigualdades activas. En este caso, la condición suficiente es que la Lagrangiana sea positiva definida en $M' = M$, el plano de restricciones activas
- Suponiendo que las funciones $f, g, h \in C^2$ y que x^* es un punto regular de las restricciones de igualdad y de desigualdad, si x^* es un punto mínimo relativo para el problema anteriormente planteado, entonces existe un vector $\lambda \in E^m$ y un vector $\mu \in E^p \geq 0$ tal que se cumplen las condiciones necesarias de primer orden anteriores y $L(x^*)$ es una matriz semidefinida positiva en el subespacio M tal que se cumplan las siguientes condiciones:

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \lambda'^* \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) + \mu'^* \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

$$\mu' \mathbf{g}(\mathbf{x}^*) = 0 \quad \mu \geq 0$$

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \lambda' \mathbf{H}(\mathbf{x}^*) + \mu' \mathbf{G}(\mathbf{x}^*) \Rightarrow \mathbf{y}' \mathbf{L}(\mathbf{x}^*) \mathbf{y} \geq 0 \text{ for } \mathbf{y} \in M$$

- Estas condiciones son las condiciones necesarias de segundo orden para problemas con restricciones de desigualdad
- Si un punto no es regular y/o no es factible (no cumple las ecuaciones e inecuación), entonces las condiciones necesarias no aplican
- Para comprobar que la Hessiana sea semidefinida positiva, es necesario definir una \mathbf{y} que satisfaga $\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) \mathbf{y} = \mathbf{0}$ y $\nabla_{\mathbf{x}} g_j(\mathbf{x}^*) \mathbf{y} = 0$ para las restricciones que estén activas trabajar con este vector. De este modo, solo es necesario utilizar la matriz Jacobiana para las restricciones y multiplicarla por \mathbf{y} , el cual tendrá $m + J$ componentes (un componente por cada restricción activa)

$$\mathbf{J}(\mathbf{x}^*) \mathbf{y} = \mathbf{0} \text{ for } M \Rightarrow \begin{cases} \sum_{k=1}^m \frac{\partial h_1(\mathbf{x})}{\partial x_k} y_k = 0 \\ \dots \\ \sum_{j=1}^J \frac{\partial g_j(\mathbf{x})}{\partial x_j} y_j = 0 \end{cases}$$

$$\Rightarrow \mathbf{y} = (y_1, y_2, \dots, y_{m+J})' \Rightarrow \mathbf{y}' \mathbf{L}(\mathbf{x}^*) \mathbf{y} \geq 0$$

- Si \mathbf{x}^* es un mínimo relativo sobre las restricciones $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ y $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$, entonces también es un punto mínimo relativo con el problema para las restricciones activas tomadas como restricciones de igualdad (de manera trivial)
- Siendo $f, \mathbf{g}, \mathbf{h} \in C^2$ y suponiendo que \mathbf{x}^* satisface $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ y $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$, entonces el punto es un mínimo relativo estricto (único) del problema planteado anteriormente si existe un vector $\lambda \in E^m$, un vector $\mu \in E^p$ y una matriz Hessiana $\mathbf{L}(\mathbf{x}^*)$ que sea definida positiva en el subespacio M' tal que se cumplan las siguientes condiciones:

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \lambda' \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}^*) + \mu' \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

$$\mu \geq 0 \quad \mu' g(x^*) = 0$$

$$L(x^*) = F(x^*) + \lambda' H(x^*) + \mu' G(x^*) \Rightarrow y' L(x^*) y \geq 0 \text{ for } y \in M'$$

- Estas condiciones son las condiciones suficientes de segundo orden para problemas con restricciones de desigualdad. Por lo tanto, se puede ver que no es necesario que x^* sea un punto regular y la Hessiana tiene que ser positiva definida, no solo semidefinida
- Para comprobar que la Hessiana sea semidefinida positiva, es necesario definir una y que satisfaga $\nabla h(x^*)y = 0$ y $\nabla g_j(x^*)y = 0$ para las restricciones que estén activas trabajar con este vector. De este modo, solo es necesario utilizar la matriz Jacobiana para las restricciones y multiplicarla por y , el cual tendrá $m + J'$ componentes (un componente por cada restricción activa que además tenga $\mu > 0$):

$$J(x^*)y = 0 \text{ for } M' \Rightarrow \begin{cases} \sum_{k=1}^m \frac{\partial h_1(x)}{\partial x_k} y_k = 0 \\ \dots \\ \sum_{j=1}^{J'} \frac{\partial g_j(x)}{\partial x_j} y_j = 0 \end{cases}$$

$$\Rightarrow y = (y_1, y_2, \dots, y_{m+J'})' \Rightarrow y' L(x^*) y \geq 0$$

- Una explicación más intuitiva del origen de las condiciones necesarias y suficientes para problemas con restricciones de igualdad y desigualdad se da a través de argumentos de factibilidad y mejora de la misma manera que antes
 - Usando series de Taylor de primer orden y suponiendo que se está en un punto factible de la región factible y que este se mueve en la dirección $y \neq 0$, es posible explicar las condiciones que harían mover el punto actual a uno óptimo
 - Para mantener la factibilidad de las restricciones de igualdad, es necesario que $\nabla h(x)y = 0$ porque se tiene que mantener que $h(x + y) = 0$ (las restricciones se siguen cumpliendo, aunque se haga un incremento de y)

$$h(x + y) \approx h(x) + \nabla h(x)y \quad \& \quad h(x) = 0$$

$$\Rightarrow \nabla h(x)y = 0 \Rightarrow h(x + y) \approx 0$$

- Además, moverse en la dirección \mathbf{y} (de descenso) mejora el valor de la función objetivo, de modo que $\nabla f(\mathbf{x})\mathbf{y} < 0$ para que $f(\mathbf{x}) > f(\mathbf{x} + \mathbf{d})$

$$f(\mathbf{x}) > f(\mathbf{x} + \mathbf{d}) \text{ \& } f(\mathbf{x} + \mathbf{d}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})\mathbf{y}$$

$$\Rightarrow f(\mathbf{x}) > f(\mathbf{x}) + \nabla f(\mathbf{x})\mathbf{y} \Rightarrow \nabla f(\mathbf{x})\mathbf{y} < 0$$

- Por lo tanto, una condición necesaria de optimalidad de un punto es que no exista una dirección \mathbf{y} que cumpla $\nabla h(\mathbf{x})\mathbf{y} = 0$ y $\nabla f(\mathbf{x})\mathbf{y} < 0$

- Si $\nabla f(\mathbf{x})\mathbf{y} < 0$, entonces existe una dirección en la que se puede minimizar $f(\mathbf{x})$ manteniendo que $\nabla h(\mathbf{x})\mathbf{y} = 0$, pero si se da el caso contrario, en donde $\nabla f(\mathbf{x})\mathbf{y} > 0$, se podría coger la dirección $-\mathbf{y}$ para que $f(\mathbf{x}) > f(\mathbf{x}) - \nabla f(\mathbf{x})\mathbf{y}$ y manteniendo la factibilidad debido a que $\nabla h(\mathbf{x})\mathbf{y} = -\nabla h(\mathbf{x})\mathbf{y} = 0$
- Por lo tanto, para que un punto \mathbf{x} se óptimo, es necesario que se cumpla que $\nabla h(\mathbf{x})\mathbf{y} = 0$ y que $\nabla f(\mathbf{x})\mathbf{y} = 0$, lo cual es equivalente a decir que $\nabla h(\mathbf{x})\mathbf{y} = 0$ y que $\nabla f(\mathbf{x}) = 0$ porque $\mathbf{y} \neq 0$
- Esto es lo mismo que decir que $\nabla f(\mathbf{x})$ y $\nabla h(\mathbf{x})$ son vectores paralelos (ya que ambos resultan en el vector 0), lo cual se puede expresar con la condición necesaria de primer orden:

$$\nabla f(\mathbf{x}) = -\lambda \nabla h(\mathbf{x}) \Rightarrow \nabla f(\mathbf{x}) + \lambda \nabla h(\mathbf{x}) = 0$$

$$\text{for some } \lambda \in E^m$$

- Sabiendo que una dirección factible \mathbf{y} será aquella en donde se cumpla $h(\mathbf{x}^* + \mathbf{y}) = 0$ y $\nabla h(\mathbf{x}^*)\mathbf{y} = 0$, se puede ver que la lagrangiana también debe respetar esto:

$$L(\mathbf{x}^* + \mathbf{y}, \lambda^*) = f(\mathbf{x}^* + \mathbf{y}) + \lambda^* h(\mathbf{x}^* + \mathbf{y}) = f(\mathbf{x}^* + \mathbf{y})$$

- En consecuencia, por expansión de la serie de Taylor de la lagrangiana $L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$, es posible obtener la siguiente aproximación:

$$\begin{aligned} L(\mathbf{x}^* + \mathbf{y}, \lambda^*) &\approx L(\mathbf{x}^*, \lambda^*) + \nabla L(\mathbf{x}^*, \lambda^*)\mathbf{y} + \frac{1}{2}\mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*)\mathbf{y} = \\ &= f(\mathbf{x}^*) + \frac{1}{2}\mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \lambda^*)\mathbf{y} \end{aligned}$$

- Por lo tanto, para que el punto \mathbf{x}^* sea un mínimo relativo, $L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$ debe cumplir que $L(\mathbf{x}^*, \lambda^*) \leq L(\mathbf{x}^* + \mathbf{y}, \lambda^*)$ (ir en

la dirección de descenso no mejora el valor de la función objetivo), por lo que $\mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)\mathbf{y} \geq 0$

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*) \Rightarrow f(\mathbf{x}^*) \leq f(\mathbf{x}^*) + \frac{1}{2}\mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)\mathbf{y}$$

$$\Rightarrow \mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)\mathbf{y} \geq 0 \text{ for } \mathbf{y}: \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = 0$$

- Si, en cambio, el punto \mathbf{x}^* es un mínimo estricto, entonces $L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)$ debe cumplir que $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) < L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)$ (ir en la dirección de descenso siempre incrementa el valor de la función objetivo), por lo que $\mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)\mathbf{y} > 0$

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*) \Rightarrow f(\mathbf{x}^*) \leq f(\mathbf{x}^*) + \frac{1}{2}\mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)\mathbf{y}$$

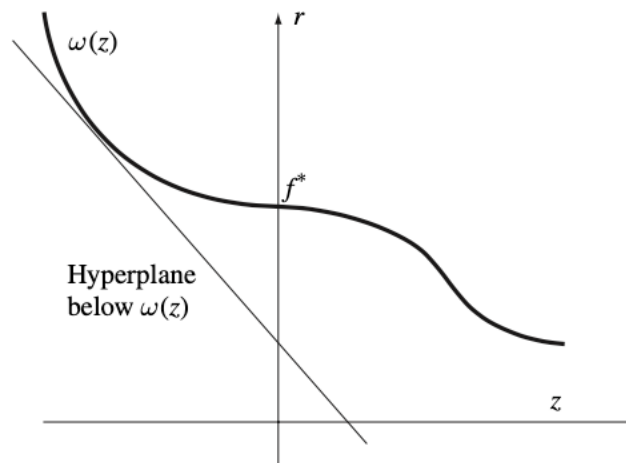
$$\Rightarrow \mathbf{y}'\nabla^2 L(\mathbf{x}^* + \mathbf{y}, \boldsymbol{\lambda}^*)\mathbf{y} \geq 0 \text{ for } \mathbf{y}: \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = 0$$

Los métodos duales y de corte de planos

- La dualidad en un problema de programación no lineal tiene su forma más útil cuando se define de manera global en términos de conjuntos y hiperplanos que tocan estos conjuntos. Esta teoría de dualidad se conoce como teoría de dualidad de Lagrange
 - La dualidad global permite dejar claro el rol de los multiplicadores de Lagrange como definidores de los hiperplanos que se pueden considerar como puntos duales en un espacio vectorial
 - La teoría proporciona una simetría entre los problemas primales y duales y esta simetría se considera perfecta en problemas convexos
 - Para problemas no convexos, esta simetría imperfecta se hace clara por el *gap* de dualidad, el cual tiene una interpretación geométrica simple
 - La teoría global permite especializarse en la teoría de dualidad local, la cual se puede usar incluso cuando no hay convexidad y que es central para entender la convergencia de los algoritmos duales
 - A diferencia de antes, en donde primero se consideran restricciones con igualdades, ahora se consideran primero las restricciones con desigualdades, considerando un problema como el siguiente:

$$\begin{aligned}
& \text{minimize} && f(\mathbf{x}) \\
& \text{subject to} && \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\
& && \mathbf{x} \in \Omega.
\end{aligned}$$

- En este caso, $\Omega \subset E^n$ es un conjunto convexo, el conjunto de funciones f y g está definido en Ω y la función g es p -dimensional
- El problema no es necesariamente convexo, pero se asume que hay un punto factible
- La función primal asociada a un problema como el anterior se define para $\mathbf{z} \in E^p$ como $\omega(\mathbf{z}) = \inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{z} \text{ \& } \mathbf{x} \in \Omega\}$, definida para que \mathbf{z} tome valores arbitrarios. Se entiende que $\omega(\mathbf{z})$ se define en el conjunto $D = \{\mathbf{z} : \mathbf{g}(\mathbf{x}) \leq \mathbf{z} \text{ for some } \mathbf{x} \in \Omega\}$
- Si el problema tiene una solución \mathbf{x}^* con valor $f^* = f(\mathbf{x}^*)$, entonces f^* es el punto del eje vertical en el espacio E^{p+1} donde la función primal pasa por ese eje. Si no tiene solución, entonces $f^* = \omega(\mathbf{z})$ es el punto de intersección
- El principio de dualidad se deriva de la consideración de todos los hiperplanos que están debajo de la función primal. Como se puede ver en el gráfico, el intercepto del hiperplano con el eje vertical siempre está por debajo o justo en el valor f^*



- Para expresar esta propiedad, se puede definir la función dual definida en el cono positivo en E^p , la cual, en general, no tiene por qué ser finita en el espacio E_+^p , pero la región en donde sí es finita es convexa. Esta se puede definir de la siguiente manera:

$$\phi(\mu) = \inf \{f(\mathbf{x}) + \mu' \mathbf{g}(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

- La función dual es cóncava en la región en la que es finita, lo cual se puede demostrar a través de la siguiente desigualdad, asumiendo que se está en la región finita:

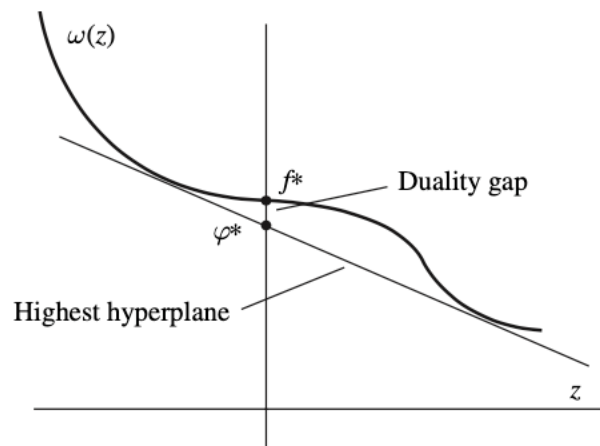
$$\begin{aligned}
\phi(\alpha\mu_1 + (1-\alpha)\mu_2) &= \inf\{f(x) + (\alpha\mu_1 + (1-\alpha)\mu_2)'g(x) : x \in \Omega\} \\
&\geq \inf\{\alpha f(x_1) + \alpha\mu_1'g(x_1) : x_1 \in \Omega\} \\
&\quad + \inf\{(1-\alpha)f(x_2) + (1-\alpha)\mu_2'g(x_2) : x_2 \in \Omega\} \\
&= \alpha\phi(\mu_1) + (1-\alpha)\phi(\mu_2)
\end{aligned}$$

- Se define el punto óptimo de la función dual como $\phi^* = \sup\{\phi(\mu) : \mu \geq \mathbf{0}\}$, en donde se entiende que la suprema se toma en la región en donde ϕ es finita

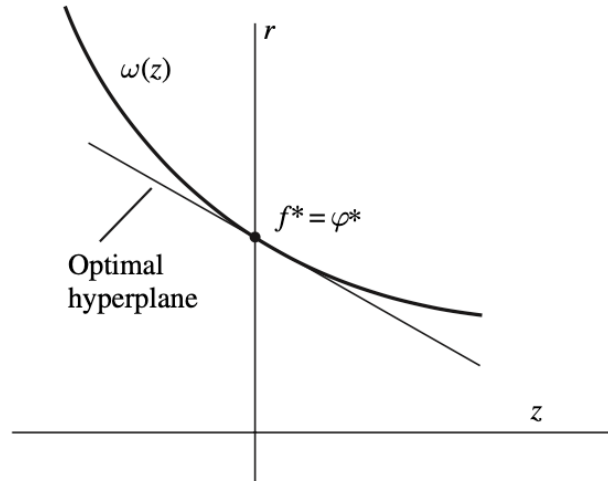
- El teorema de dualidad débil expresa que, siendo x un punto factible del problema primal y μ uno factible del dual, entonces $\phi^* \leq f^*$

$$\begin{aligned}
\mu \geq \mathbf{0} &\Rightarrow \phi(\mu) = \inf\{f(x) + \mu'g(x) : x \in \Omega\} \\
&\leq \inf\{f(x) + \mu'g(x) : g(x) \leq \mathbf{0} \text{ \& } x \in \Omega\} \\
&\leq \inf\{f(x) : g(x) \leq \mathbf{0} \text{ \& } x \in \Omega\} = f^* \\
&\Rightarrow \phi^* = \sup\{\phi(\mu) : \mu \geq \mathbf{0}\} \leq f^*
\end{aligned}$$

- Esto permite ver como la función dual permite obtener cotas inferiores para el valor óptimo del problema primal f^*
- La función dual tiene una interpretación geométrica fuerte que permite entender mejor el rol de los multiplicadores de Lagrange y los hiperplanos



- Introduciendo suposiciones de convexidad, el análisis hecho se puede fortalecer para dar una versión fuerte del teorema de dualidad, donde no hay un *gap* de dualidad cuando el intercepto está en f^*



- Para poder generalizar los resultados, se considera un problema que también incluyen restricciones del tipo $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, en donde f es una función convexa, \mathbf{h} es afín de dimensión m , \mathbf{g} es convexa de dimensión p y Ω es un conjunto convexo (suposiciones de convexidad global)

$$\begin{aligned} & \text{maximize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & && \mathbf{x} \in \Omega \end{aligned}$$

- En este caso, la función dual y su óptimo se pueden definir de la siguiente manera:

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf \{f(\mathbf{x}) + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}'\mathbf{g}(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

$$\phi^* = \sup \{\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\lambda} \in E^m, \boldsymbol{\mu} \in E^p, \boldsymbol{\mu} \geq \mathbf{0}\}$$

- El teorema de dualidad extrema se puede resumir en las siguientes dos proposiciones:
 - Suponiendo que en el problema planteado \mathbf{h} es regular con respecto a Ω , que hay un punto $\mathbf{x}_1 \in \Omega$ tal que $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ y $\mathbf{g}(\mathbf{x}) < \mathbf{0}$ y que el problema tiene una solución \mathbf{x}^* con valor $f(\mathbf{x}^*) = f^*$, entonces para toda $\boldsymbol{\lambda}$ y $\boldsymbol{\mu} \geq \mathbf{0}$ se cumple la siguiente desigualdad:

$$\phi^* \leq f^*$$

- Con estos supuestos, también existe alguna λ y alguna $\mu \geq 0$ tal que $\phi(\lambda, \mu) = f^*$ y, por tanto, $\phi^* = f^*$. Además, λ y μ son multiplicadores lagrangianos para el problema (cuando se está en el óptimo)
- La demostración de este teorema proviene del teorema de Lagrange de orden cero, en donde los multiplicadores del teorema permiten ver que debe haber una igualdad entre $\phi(\lambda, \mu)$, ϕ^* y f^*

$$\begin{aligned} f^* &= \max \{f(x) + \lambda' h(x) + \mu' g(x) : x \in \Omega\} = \\ &= \phi(\lambda, \mu) \leq \phi^* \leq f^* \Rightarrow \phi(\lambda, \mu) = \phi^* = f^* \end{aligned}$$

- En la práctica, las mecánicas de dualidad se llevan a cabo localmente, haciendo que las derivadas sean igual a cero, o moviéndose en la dirección de un gradiente. Esto hace que la teoría global se pueda remplazar por una teoría local más débil pero más útil
 - Como se ha hecho antes por conveniencia, primero se considera un problema de programación no lineal de la siguiente forma:

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{aligned}$$

- En este caso, $\mathbf{x} \in E^n$, $\mathbf{h}(\mathbf{x}) \in E^n$ y $f, \mathbf{h} \in C^2$, pero no se asume convexidad global
- Asumiendo que \mathbf{x}^* es un punto regular de las restricciones, entonces habrá un multiplicador de Lagrange correspondiente λ^* tal que se cumple la siguiente igualdad:

$$\nabla f(\mathbf{x}^*) + \lambda'^* \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

- Además, si \mathbf{x}^* es un punto regular de las restricciones, la matriz Hessiana $\mathbf{L}(\mathbf{x}^*)$ será semidefinida positiva en el subespacio tangente $M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*) \mathbf{y} = \mathbf{0}\}$

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \lambda'^* \mathbf{H}(\mathbf{x}^*)$$

- En este punto se introduce la suposición especial de convexidad local, necesaria para el desarrollo de la teoría de dualidad local. Específicamente, se asume que la Hessiana $\mathbf{L}(\mathbf{x}^*)$ es positiva definida en todo el espacio E^n , no solo en el subespacio M

- La suposición garantiza que la función lagrangiana es localmente convexa en \mathbf{x}^*
 - Con esta suposición, el punto \mathbf{x}^* no solo es una solución local al problema restringido planteado, también es una solución local al problema no restringido $\min \{f(\mathbf{x}) + (\boldsymbol{\lambda}^*)' \mathbf{h}(\mathbf{x})\}$, ya que satisface las condiciones suficientes de primer y segundo orden para un punto mínimo local
 - Además, para cualquier $\boldsymbol{\lambda}$ lo suficientemente cerca de $\boldsymbol{\lambda}^*$, la función $f(\mathbf{x}) + \boldsymbol{\lambda}' \mathbf{h}(\mathbf{x})$ tiene un punto mínimo local en algún punto \mathbf{x} cerca de \mathbf{x}^* . Esto ocurre porque la ecuación $\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}' \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ tiene solución en un \mathbf{x} cerca de \mathbf{x}^* cuando $\boldsymbol{\lambda}$ está cerca de $\boldsymbol{\lambda}^*$ porque $\mathbf{L}^*(\mathbf{x})$ es invertible y es positiva definida
 - Por lo tanto, localmente hay una sola correspondencia entre $\boldsymbol{\lambda}$ y \mathbf{x} a través de la solución a $\min \{f(\mathbf{x}) + \boldsymbol{\lambda}' \mathbf{h}(\mathbf{x})\}$ y es continuamente diferenciable
- Cerca de $\boldsymbol{\lambda}^*$ se define la función dual ϕ como $\phi(\boldsymbol{\lambda}) = \min_{\mathbf{x}} \{f(\mathbf{x}) + \boldsymbol{\lambda}' \mathbf{h}(\mathbf{x})\}$ (cerca de \mathbf{x}^*). A partir de esta definición, es posible demostrar que, localmente, el problema primal de minimización es equivalente al problema dual de maximización con respecto a $\boldsymbol{\lambda}$
 - Para ello, solo se tiene que optimizar con respecto a \mathbf{x} y, una vez encontrados los puntos óptimos, se mira la función en esos puntos (que será una función de $\boldsymbol{\lambda}$)
 - Suponiendo que el problema $\min \{f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$ tiene una solución local en \mathbf{x}^* con el valor correspondiente r^* y multiplicador de Lagrange $\boldsymbol{\lambda}^*$, que \mathbf{x}^* es un punto regular de las restricciones y que la Hessiana de la lagrangiana $\mathbf{L}^* = \mathbf{L}(\mathbf{x}^*)$ es definida positiva, entonces el siguiente problema dual tiene una solución local en $\boldsymbol{\lambda}^*$ con valor correspondiente r^* y \mathbf{x}^* como el punto correspondiente a $\boldsymbol{\lambda}^*$ en la definición de ϕ

$$\max_{\boldsymbol{\lambda}} \phi(\boldsymbol{\lambda})$$

- Para problemas que también incluyan restricciones de desigualdad (a parte de las de igualdad) solo se requiere una pequeña modificación a la teoría desarrollada, de modo que se considera el siguiente problema:

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ &&& \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \end{aligned}$$

- En este caso, $\mathbf{x} \in E^n$, $\mathbf{h}(\mathbf{x}) \in E^n$, $\mathbf{g}(\mathbf{x}) \in E^p$ y $f, \mathbf{h}, \mathbf{g} \in C^2$, pero no se asume convexidad global
- Suponiendo que \mathbf{x}^* es una solución local para el problema y que \mathbf{x} es un punto regular para las restricciones, entonces existen multiplicadores de Lagrange $\boldsymbol{\lambda}$ y $\boldsymbol{\mu} \geq \mathbf{0}$ tal que se cumplen las siguientes condiciones:

$$\nabla f(\mathbf{x}^*) + (\boldsymbol{\lambda}^*)' \nabla \mathbf{h}(\mathbf{x}^*) + (\boldsymbol{\mu}^*)' \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

$$(\boldsymbol{\mu}^*)' \mathbf{g}(\mathbf{x}^*) = 0$$

- Se impone la suposición de convexidad local sobre la Hessiana de la función Lagrangiana $L(\mathbf{x}^*)$ en todo el espacio

$$L(\mathbf{x}^*) = F(\mathbf{x}^*) + (\boldsymbol{\lambda}^*)' H(\mathbf{x}^*) + (\boldsymbol{\mu}^*)' G(\mathbf{x}^*) = \mathbf{0}$$

- Para $\boldsymbol{\lambda}$ y $\boldsymbol{\mu} \geq \mathbf{0}$ cerca de $\boldsymbol{\lambda}^*$ y $\boldsymbol{\mu}^*$ se define la función dual de la siguiente manera:

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x}} \{f(\mathbf{x}) + \boldsymbol{\lambda}' \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}' \mathbf{g}(\mathbf{x}) : \boldsymbol{\mu} \geq \mathbf{0}\}$$

- Para ello, solo se tiene que optimizar con respecto a \mathbf{x} y, una vez encontrados los puntos óptimos, se mira la función en esos puntos (que será una función de $\boldsymbol{\lambda}$ y $\boldsymbol{\mu}$)
- Por lo tanto, se puede ver como se puede hacer un desarrollo paralelo al desarrollo con restricciones de igualdad, y ver que ϕ consigue un máximo local con respecto a $\boldsymbol{\lambda}$ y $\boldsymbol{\mu} \geq \mathbf{0}$ en $\boldsymbol{\lambda}^*$ y $\boldsymbol{\mu}^*$

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \quad s. t. \quad \boldsymbol{\mu} \geq \mathbf{0}$$

- Debido a que la condición necesaria y suficiente para la optimalidad de la función dual es $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$ es $\nabla f(\mathbf{x}^*) + (\boldsymbol{\lambda}^*)' \nabla \mathbf{h}(\mathbf{x}^*) + (\boldsymbol{\mu}^*)' \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$, se puede expresar el problema dual de la siguiente manera:

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ s. t. \quad \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \quad \Rightarrow \quad \begin{aligned} \max_{\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ s. t. \quad \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0} \\ \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

- Este problema asume que las funciones son todas diferenciables y se conoce como el problema dual de Wolfe
- Con tal de obtener la formulación final, normalmente se intenta sustituir $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}$ en $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$, de modo que la función objetivo se simplifica, y se añade $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}$ como

restricción. Al final suele quedar un problema de optimización más simple (sobre solo algunos tipos de variables o con menos restricciones, por ejemplo)

- No es necesario incluir los multiplicadores de Lagrange de todas las restricciones de un problema en la definición de función dual. Esta característica se denomina dualidad parcial
 - En general, la suposición de convexidad local se puede definir con respecto a cualquier subconjunto de restricciones funcionales
 - Por lo tanto, con un problema con restricciones de ambos tipos (de los vistos anteriormente) se puede definir la siguiente función dual, la cual conseguirá un máximo local en λ^* :

$$\phi(\lambda) = \min_{\lambda} \{f(x) + \lambda'h(x) : g(x) \leq 0\}$$

- En este caso, el procedimiento para obtener la función es igual, pero quedará en función de las variables duales que se utilicen para definir Lagrangiana