UNIVERSIDAD DE PANAMÁ

FACULTAD DE CIENCIAS NATURALES, EXACTAS Y TECNOLOGÍA

ESCUELA DE MATEMÁTICA

OPTIMIZACIÓN CONVEXA

PRESENTADO POR:

YARUZBETH M. DE MCBARNETTE

TRABAJO DE GRADUACIÓN PARA OPTAR POR EL TÍTULO DE LICENCIADA EN MATEMÁTICA

2016

AGRADECIMIENTOS

A Dios Todo Poderoso

Por darme la vida, por bendecirme de muchas maneras, por las pruebas difíciles y obstáculos que me ha ayudado a superar, por no soltarme la mano y ayudarme a salir adelante.

A mi madre, Ana Celia Martínez

Por su apoyo incondicional y por su sacrificio, por sus consejos, por estar siempre a mi lado, por enseñarme la perseverancia, por guiar siempre mí camino. Eres el motor que impulsa mi vida. Te amo.

A mi Esposo Amado

Gracias por depositar tu confianza en mí para salir adelante en esta meta trazada desde el principio, por las experiencias vividas las cuales las guardo en el fondo de mi corazón, nuevamente gracias por ser mi compañero y mi gran amigo, Dios te bendiga, mi amor.

Estimado Profesor

Eterna gratitud al profesor Eloy Rico por todo el tiempo que me ha dedicado durante la dirección de esta tesis. Ha estado siempre dispuesto a colaborar aportando su gran experiencia y conocimientos. Gracias muchas gracias.

INDICE

INTRODUCCIÓN

La Optimización ha jugado una excelente y destacada participación en los variados descubrimientos y aplicaciones que se han alcanzado con el uso de las teorías desarrollas en Matemática, sobre todo en los últimos cincuenta años. Por ejemplo, el diseño de software ha facilitado enormemente el fortalecimiento de la industria manufacturera, la economía y las finanzas. Muchos procesos industriales de fabricación de productos de alta tecnología médica para el diagnóstico de padecimientos del cuerpo humano son ahora en esta época, mucho mejor estudiados y diagnosticados, gracias al uso de equipos de ondas que fueron modelados bajo esquemas de modelos matemáticos de optimización.

La industria comercial aeroespacial puede también controlar más eficientemente, el flujo de pasajeros que diariamente se están moviendo entre los principales aeropuertos del mundo. La determinación de horarios de vuelos de llegadas y de salidas diariamente, bajo una administración científica, le ahorra a esta industria millones de dólares cuando debe planificar todos estos vuelos con diferentes horarios líneas aéreas.

Este trabajo lo hemos concentrado en el desarrollo de las principales ideas de la Optimización Convexa, como una herramienta matemática útil en la Optimización, tanto en lo que se refiere a la linealidad como también en la no linealidad. Desde los años 90, se han reportado con el uso de la optimización convexa, muchas aplicaciones en áreas como los sistemas automáticos de control, cálculo y procesamiento de señales, comunicaciones y redes, diseño de circuitos electrónicos, análisis y modelado de datos, estadísticas y las finanzas. Optimización convexa también ha encontrado una amplia aplicación en optimización combinatoria y optimización global, donde se utiliza para encontrar límites a el valor óptimo, así como soluciones aproximadas.

En esta tesis le damos una pequeña visión a la optimización matemática, centrándome en especial, sobre la optimización y las funciones convexas. El concepto de convexidad es fundamental en el análisis y resolución de los problemas de optimización. Consideramos que el análisis de la convexidad de conjuntos así como los diferentes tipos de convexidad de funciones son instrumentos básicos para la teoría de la optimización matemática.

En el primer capítulo hacemos un estudio de los conjuntos y funciones convexas, factor principal que sustenta el análisis posterior y lo referente a la convexidad. Se proponen definiciones y propiedades básicas que le dan fortaleza a la teoría. Temas como combinación convexa de puntos, cápsula o cobertura convexa, epígrafo de una función convexa entre ellos. El capítulo dos lo concentramos en abordar las condiciones de optimalidad para un programa convexo, en donde un problema de Programación Lineal y algunos otros No Lineales son clasificados como problemas de optimización convexa.

CAPÍTULO 1

CONJUNTO Y FUNCIÓN CONVEXA

1.1 Aspectos generales

En este capítulo nos concentraremos en sentar las bases matemáticas de un conjunto y una función convexa. Muchas de ellas son consideradas fundamentales y clásicas, tales como cápsula convexa, hiperplano de separación, el epígrafo e hipografo de una función convexa entre otros.

Definición (Conjuntos Afines) Un conjunto $C \subseteq \mathbb{R}^n$ es afín si la recta que pasa para dos puntos distintos en C se encuentra en C, es decir, si por cualquier $x_1, x_2 \in C$ y $\theta \in \mathbb{R}$, tenemos $\theta x_1 + (1-\theta)\ x_2 \in C$. En otras palabras, C contiene la combinación lineal de dos puntos cualesquiera de él, siempre que los coeficientes (escalares) en la combinación lineal sumen uno.

Esta idea se puede generalizar para más de dos puntos del conjunto. Nos referimos a un punto de la forma $\theta_1 x_1 + \ldots + \theta_k x_k$, donde $\theta_1 + \ldots + \theta_k = 1$, como una combinación afín de los puntos $x_1, \ldots, x_k \in C$. Usando inducción, se puede demostrar que un conjunto afín contiene todas las combinaciones afín de sus puntos: Esto es si C es un conjunto afín $x_1, \ldots, x_k \in C$ y $\theta_1 + \ldots + \theta_k = 1$, entonces el punto $\theta_1 x_1 + \ldots + \theta_k x_k$ también pertenece a C.

Si C es un conjunto afín y $x_0 \in C$, entonces el conjunto

$$V = C - x_0 = \{x - x_0\ /\ x \in C\}$$

es un sub-espacio, es decir, cerrado bajo la suma y la multiplicación escalar. Para ver esto, supongamos $v_1, v_2 \in V$ y $\alpha, \beta \in \mathbb{R}$.

Entonces tenemos $v_1 + x_0 \in C$ y $v_2 + x_0 \in C$, y así:

$\alpha v_1 + \beta v_2 + x_0 = \alpha(v_1 + x_0) + \beta(v_2 + x_0) + (1 - \alpha - \beta)x_0 \in C$ ya que C es afín, y $\alpha + \beta + (1 - \alpha - \beta) = 1$. Llegamos a la conclusión de que $\alpha v_1 + \beta v_2 \in V$, desde que $\alpha v_1 + \beta v_2 + x_0 \in C$.

Por lo tanto, el conjunto afín C se puede expresar como

$$C = V + x_0 = \{v + x_0 \ / \ v \ \in V\}$$

es decir, como un sub-espacio más un desplazamiento. El sub-espacio V asociado con el conjunto afín C no depende de la elección de x_0 entonces así x_0 puede ser elegido como cualquier punto en C.

Definiremos la dimensión de un conjunto afín C como la dimensión del sub-espacio $V = C - x_0$, donde x_0 es cualquier elemento de C.

Ejemplo 1.1: El conjunto solución de un sistema lineal de ecuaciones $C = \{x \ / \ Ax = b\}$, donde $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$, es un conjunto afín. Para mostrar esto, supongamos que $x_1, x_2 \in C$, es decir, $A\,x_1 = b$, $A\,x_2 = b$. Entonces para cualquier θ, tenemos

$$\begin{aligned} A((\theta x_1) + (1-\theta)\,x_2) &= A\theta x_1 + (1-\theta)A\,x_2 \\ &= \theta b + (1-\theta)b \\ &= b \end{aligned}$$

Lo que demuestra que la combinación afín $\theta x_1 + (1-\theta)\, x_2$ está contenida en C. El sub-espacio asociado con el conjunto afín C es el espacio nulo de A. También tenemos lo contrario: cada conjunto afín se puede expresar como el conjunto de soluciones de un sistema de ecuaciones lineales.

El conjunto de todas las combinaciones afines de puntos de algún conjunto $C \subseteq \mathbb{R}^{\mathrm{n}}$ se denomina casco afin de C, y se denota por aff C; tal que

aff $C = \{ \theta_1 x_1 + . . . + \theta_k x_k \ / \ x_1 ,. . . , x_k \ \in C, \in, \ \theta_1 + . . . + \theta_k = 1\}$

El casco afín es el conjunto afín más pequeño que contiene C , en el siguiente sentido: si S es un conjunto afín con $C \subseteq S$, entonces aff $C \subseteq S$.

1.2 Conjuntos convexos

Un conjunto C es convexo si el segmento de línea entre dos puntos cualesquiera en C se encuentra en C, es decir, si por cualquier $x_1, \ x_2 \in C$ y cualquier escalar θ con $0 \leq \theta \leq 1$ se tiene que:

$$\theta x_1 + (1-\theta)\, x_2 \ \in C$$

En términos más sencillos, un conjunto es convexo, si cada punto del conjunto puede ser visto por todos los demás puntos a lo largo de una trayectoria recta y sin obstáculos entre estos, donde “sin obstáculos” significa ojo en el conjunto. Cada conjunto afín también convexo, ya que contiene la totalidad de la línea entre dos puntos distintos en él mismo, y por lo tanto también el segmento de línea entre los puntos. La Figura1 ilustra algunos ejemplos de convexos.

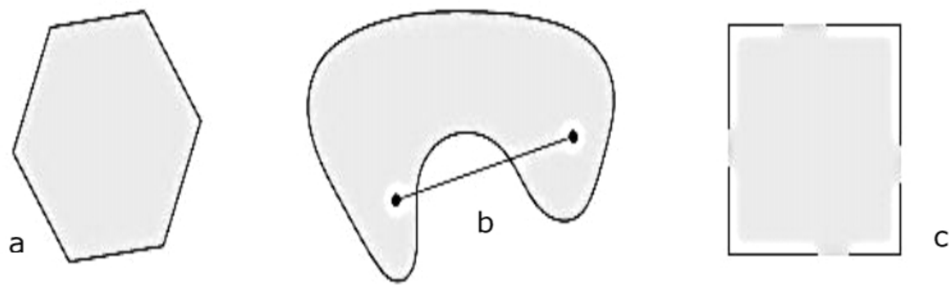


Figura 1

Ejemplos de conjuntos convexos y no convexos

(a) El hexágono, que incluye su límite, es convexo. (b) La figura de forma de un riñón no es un conjunto convexo, ya que el segmento de línea entre los dos puntos (en el conjunto se muestra como puntos) no está contenido en el conjunto. (c) La plaza contiene algunos puntos de los límites, pero no a otros, y no es convexa.

Llamamos combinación convexa de puntos $x_1,\ x_2,. . .,x_k$, en todo punto de la forma $\theta_1 x_1 + \theta_2 x_2 + . . . + \ \theta_k \theta_R$, tal que $\theta_1 + \theta_2 + . . . + \theta_k = 1, \theta_i \geq 0$ para todo $i = 1\ R\ \cdot\cdot\cdot\ k$

Al igual que el conjunto afín, se puede demostrar que un conjunto es convexo si y sólo si contiene todos las combinaciones convexas de sus puntos. Una combinación convexa de puntos puede ser considerada como una mezcla o promedio ponderado de los puntos, x_i con θ_i la fracción de x_i sobre la mezcla.

1.3 Cápsula o casco convexo

La cápsula (casco) convexa de un conjunto C que denota por $conv\,C$, es el conjunto de todas las combinaciones convexas de puntos en C definido como

$$conv\,C = \{\theta_1 x_1 + \ldots + \theta_k x_k \ / \ x_i \in C, \ \theta_i \geq 0, i = 1 \ldots k, \theta_1 + \ldots + \theta_k = 1\}$$

Otros nombres dados a la cápsula convexa son cobertura, envolvente etc.

Como su nombre lo sugiere, la cápsula convexa $conv\,C$ es siempre convexa. Además es el conjunto más pequeño convexo que contiene C. Si B es cualquier otro conjunto convexo que contiene C, entonces $conv\,C \subseteq B$. La figura 2 lo ilustra

Los cascos convexas de dos conjuntos en $\mathbb{R}^2$

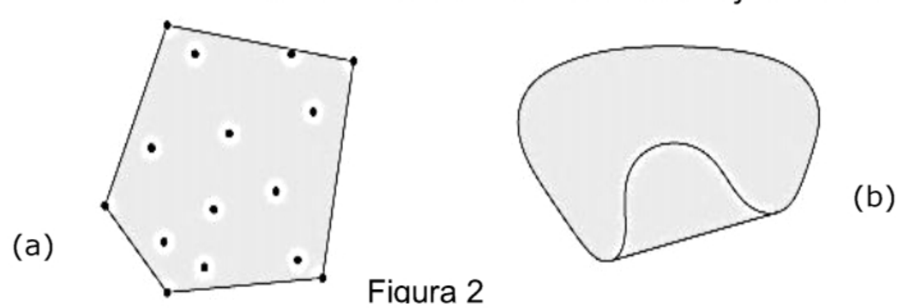


Figura 2

(a) La cápsula convexa de un conjunto de quince puntos (se muestra en forma de puntos) es el pentágono (se muestra sombreado). (b) La envolvente convexa del conjunto con forma de riñón.

Retomando lo expresado antes, la idea de una combinación convexa puede ser generalizada para incluir tantas sumas infinitas, integrales y más generales, las distribuciones de probabilidad. Para ello, sean $\theta_1, \theta_2, \ldots . \in \mathbb{R}$ que satisfacen:

$$\theta_i \geq 0 \quad , \quad i = 1,2, \ldots ; \quad \sum_{i=1}^{\infty} \theta_i = 1$$

Sean $x_1, x_2, \ldots \in C$ donde $C \subseteq \mathbb{R}^n$ convexo. Entonces,

$$\sum_{i=1}^{\infty} \theta_i x_i \in C$$

si la serie es convergente. Generalizando, supongamos ahora que $p: \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que $p(x) \geq 0$ para todo $x \in C$ y además $\int_C p(x)dx = 1$, en $C \subseteq \mathbb{R}^n$, donde C es un conjunto convexo. Entonces

$$\int_C p(x)xdx \in C$$

si la integral existe.

En términos más general, supongamos $C \subseteq \mathbb{R}^n$ es convexo y x un vector aleatorio con $x \in C$ con probabilidad uno. Entonces $\mathrm{E}(x) \in C$ (esperanza matemática) .

Por ejemplo, supongamos que la variable aleatoria x sólo toma dos valores x_1 y x_2, con probabilidad $p(x = x_1) = \theta$ y probabilidad $p\,(x = x_2) = 1 - \theta$, donde $0 \leq \theta \leq 1$. Entonces $E(x) = \theta x_1 + (1 - \theta)\, x_2$, y regresamos con esto a la simple combinación convexa de dos puntos.

1.4 Conos

Un conjunto C es llamado cono, o un homogéneo no negativo, si para cada $x \in C$ y $\theta \geq 0$, tenemos $\theta x \in C$. Un conjunto C es llamado cono convexo, si este es convexo y cono, lo cual significa que para cualquier $x_1, x_2, \in C$ y $\theta_1, \theta_2 \geq 0$, tenemos: $$\theta_1 x_1 + \theta_2 x_2 \in C$$

Los puntos de esta forma se pueden describir geométricamente como si fuese una pieza de un pastel de dos dimensiones, con 0 en el vértice y los bordes que pasan por x_1 y x_2. (Ver figura 3)

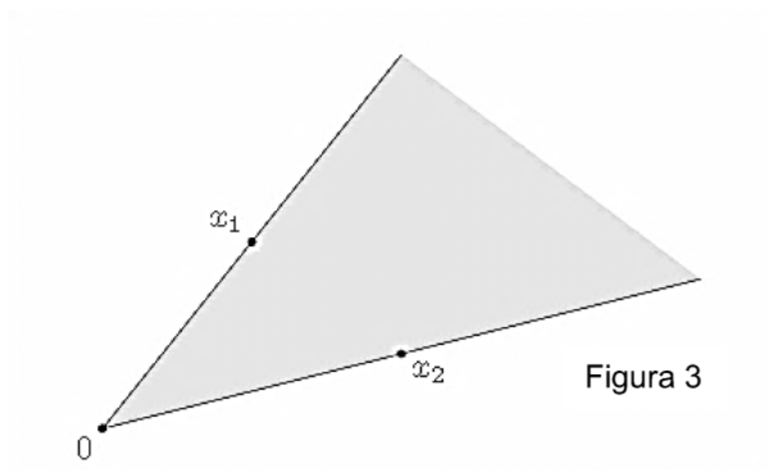


Figura 3

La figura en forma de una pieza de pastel muestra todos los puntos de la forma $\theta_1 x_1 + \theta_2 x_2$, donde $\theta_1 \theta_2 \geq 0$. El vértice de la figura (que corresponde a $\theta_1 = \theta_2 = 0$) es 0; sus bordes (que corresponden a $\theta_1 = 0$ o $\theta_2 = 0$) pasa a través de los puntos x_1 y x_2.

Un punto de la forma $\theta_1 x_1 + \ldots + \theta_k x_k$ con $\theta_1, \ldots, \theta_k \geq 0$ se llama una combinación cónica (o una combinación lineal no negativa) de $x_1, \ldots, x_k$.

Si para cada i, x_i se encuentran en un cono convexo C, entonces cada combinación cónica de x_i está en C . Por el contrario, un conjunto C es un cono convexo si y sólo si contiene todas las combinaciones cónicas de sus elementos. Al igual que las combinaciones convexas (o afín), la idea de la combinación cónica puede ser generalizado a sumas infinitas e integrales.

El casco cónico de un conjunto C es el conjunto de todas las combinaciones cónicas de puntos en C, es decir,

$$\{\theta_1 x_1 + \ldots + \theta_k x_k \ / \ x_i \in C,\ \theta_i \geq 0, i = 1, \ldots, k\}$$

que es también el cono convexo más pequeño que contiene C (Ver figura 4).

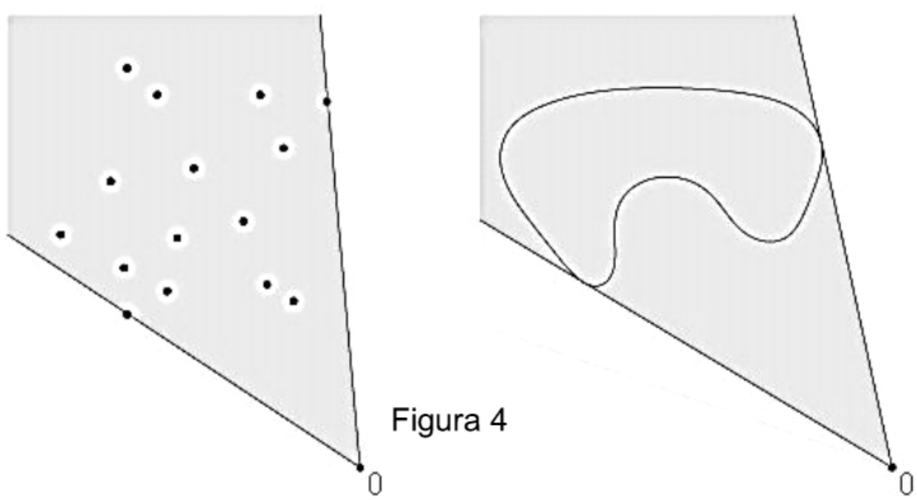


Figura 4

Las coberturas cónicas (que se muestra sombreado) de los dos conjuntos de la figura 2)

1.5 Hiperplanos y semi-espacios

Un hiperplano H es un conjunto de la forma

$$H = \{x \ / \ \mathrm{a}^T x = b\}$$

donde $a \in \mathbb{R}^n$, $a \neq 0$ y $b \in \mathbb{R}$. Analíticamente, es el conjunto solución de una ecuación lineal no trivial para los componentes de x y por lo tanto un conjunto afín. Desde el punto de vista geométrico, el hiperplano H se puede interpretar como el conjunto de puntos con un producto interno constante a un vector dado, "a" o como un hiperplano con un vector normal; "a" la constante $b \in \mathbb{R}$ determina el desfase del hiperplano desde el origen. Esta interpretación geométrica se puede comprender expresando al hiperplano en la forma

$$\{x \ / \ \mathrm{a}^T (x - x_0) = 0\}$$

donde x_0 es cualquier punto en el hiperplano es decir, cualquier punto que satisface en $a^T x_0 = b$.

Esta representación puede a su vez ser expresada como:

$$\{x \ / \ \mathrm{a}^T (x - x_0) = 0\} = x_0 + \mathrm{a}^{\perp}$$

donde $\mathrm{a}^{\perp}$ denota el complemento ortogonal de a, es decir, el conjunto de todos los vectores v ortogonales tal que:

$$\mathrm{a}^{\perp} = \{v \ /= \{ \mathrm{v} \ / \ \mathrm{a}^T v = 0\}$$

Esto demuestra que el hiperplano consiste en un desplazamiento x_0, más de todos los vectores ortogonales al vector a . Estas interpretaciones geométricas se ilustran en la figura 5.

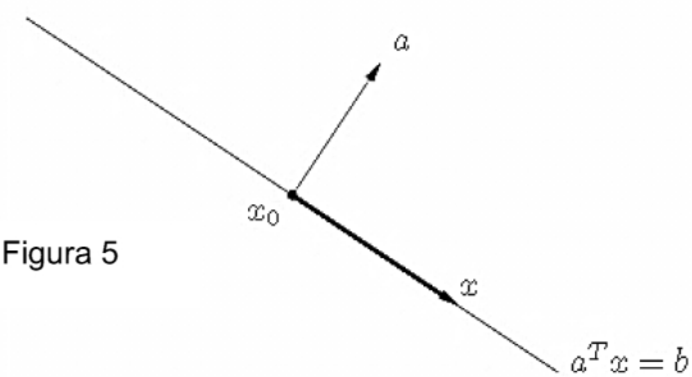


Figura 5

Hiperplano en $\mathbb{R}^2$, con un vector normal α y un punto x_0 en el hiperplano. Para cualquier punto x en el hiperplano, $x - x_0$ (que se muestra como en la flecha más oscura) es ortogonal en a.

Un hiperplano $\mathbb{R}^n$ se divide en dos semi-espacios. Un semi-espacio (cerrado) es un conjunto de la forma

$$\{x \,/\, \mathrm{a}^T x \leq b\} \qquad (1)$$

donde $a \neq 0$, es decir, el conjunto solución no trivial de una desigualdad lineal. Los Semi-espacios son convexos, pero no son afín. Esta situación se ilustra en la figura 6.

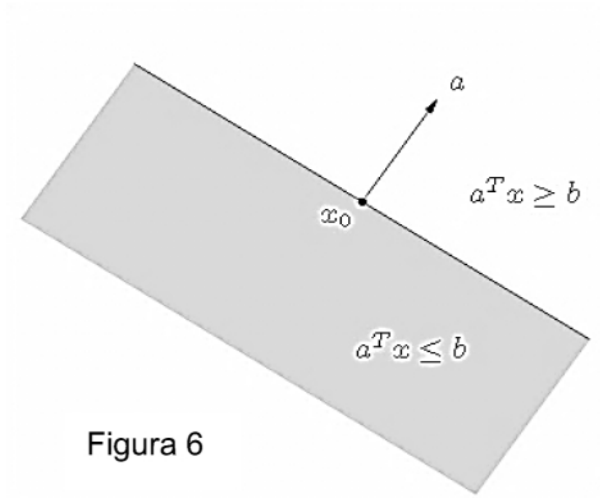


Figura 6

Un hiperplano definido por $\mathrm{a}^T x = b$ *en* $\mathbb{R}^2$ *determina dos semiespacios. El semi-espacio determinada por* $\mathrm{a}^T x \geq b$ *(no sombreado) es el semi-espacio que se extiende en la dirección* a*. El semi-espacio determinada por* $x \geq b$ *(que se muestra sombreada) se extiende en la dirección.*$-a$*. El vector de una es la normal exterior de este semi-espacio.*

El semi-espacio (1) también se puede expresar como

$$\{x \,/\; \mathrm{a}^T (x - x_0) \leq 0\} \qquad (2)$$

donde x_0 es un punto cualquiera del hiperplano asociado, es decir, cumple la

$$\mathrm{a}^T x_0 = b$$

representación (2) sugiere una interpretación geométrica sencilla. Consiste en el semi-espacio de x_0 más cualquier vector que hace un ángulo obtuso con un vector normal. Esto se ilustra en la figura 7.

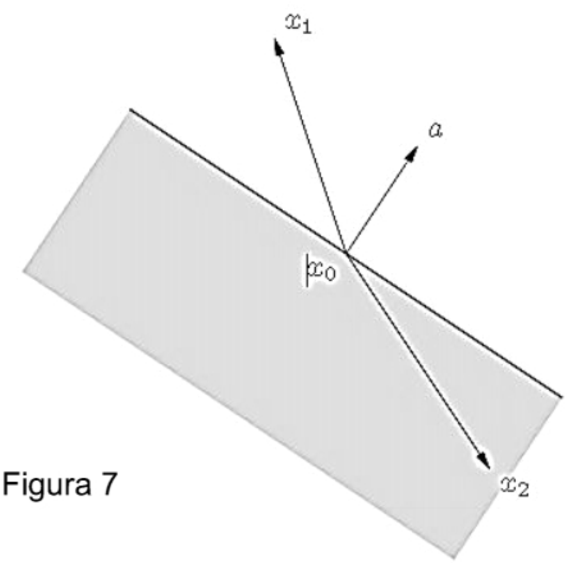


Figura 7

El conjunto sombreado es el semi-espacio determinada en $(x - x_0) \leq 0$*. El vector* $x_1 - x_0$ *forma un ángulo agudo con, por lo* x_1 *no está en el semiespacio. El vector* $x_2 - x_0$ *forma un ángulo obtuso con una, y también lo es en el semiespacio.*

La frontera del semi-espacio (1) es el hiperplano $\{x \;/\; \mathrm{a}^T x = b\}$. El conjunto $\{x \;/\; \mathrm{a}^T x < b\}$, que es el interior del semi-espacio $\{x \;/\; \mathrm{a}^T x \leq b\}$, el cual se llama el semi-espacio abierto.

1.6 Poliedros

Un poliedro se define como el conjunto solución de un número finito de igualdades y las desigualdades lineales; esto es, sea un P poliedro, entonces:

$$P = \{x \;/\; a_j^T x \leq b_{j,} \quad j = 1, \dots, m, \; c_j^T x = d_{j,}\, j = 1, \dots, p\} \qquad (3)$$

un poliedro es, pues, la intersección de un número finito de semi-espacios e hiperplanos. Los conjuntos afines por ejemplo, sub-espacios, hiperplanos, líneas;

rayos, segmentos de recta y semi-espacios son todos poliedros. Se demuestra fácilmente que los poliedros son conjuntos convexos.

Un poliedro acotado a veces se denomina un politopo, pero algunos autores utilizan la convención opuesta (es decir, politopo para cualquier conjunto de la forma (3), y poliedro cuando se está delimitada). Figura 8 muestra un ejemplo de un poliedro definido como la intersección de cinco semi-espacios.

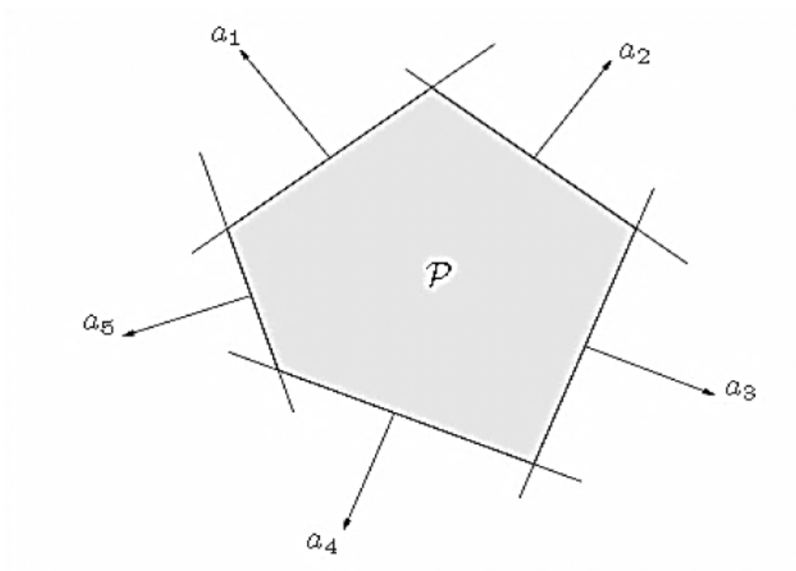


Figura 8

El poliedro $\mathcal{P}$ (que se muestra sombreada) es la intersección de cinco semi-espacios exteriores, con vectores normales, $a_1, \ldots, a_5$.

Puede encubrirse el poliedro, en forma matricial como:

$$P = \{x \;/\; Ax \leq b,\; Cx = d\} \tag{4}$$

por (3), donde

$$A = \begin{bmatrix} a_j^T \\ \cdot \\ \vdots \\ a_m^T \end{bmatrix}, \qquad\qquad C = \begin{bmatrix} c_1^T \\ \cdot \\ \vdots \\ c_p^T \end{bmatrix}$$

y el símbolo $\leq$ denota desigualdad vectorial o desigualdad a trozos en $\mathrm{R}^m : u \leq v$ significa $u_i \leq v_i$ para i = 1,…, m.

Ejemplo 1.2: El ortante no negativo es el conjunto de puntos con componentes no negativas y denotamos por $\mathbb{R}_+^n$, tal que:

$$\mathbb{R}_+^n = \{\, x \in \mathbb{R}^n \ / \ x_i \geq 0, i = 1, \dots, n\} = \{x \in \mathbb{R}^n \ / \ x \geq 0\}$$

note que $\mathbb{R}_+$ denota el conjunto de los números positivos tal que

$$\mathbb{R}_+ = \{x \in \mathbb{R} \,/\, x \geq 0\}.$$

el ortante no negativo es un poliedro y un cono, y por lo tanto es llamado cono poliédrico.

1.7 Algunos ejemplos importantes

En esta sección se describen algunos ejemplos importantes de conjuntos convexos. Comenzamos con algunos ejemplos sencillos.

• El conjunto vacío ∅, un solo punto $\{x_0\}$, y todo el espacio $\mathbb{R}^n$ son afines por lo tanto son conjuntos convexos.

• Cualquier línea es afín. Si se pasa a través de cero, que es un sub-espacio, por lo tanto, también es un cono convexo.

• Un segmento de línea es convexa, pero no es afín (a menos que se reduce a un punto).

• Un rayo, el cual tiene la forma $\{x_0 + \theta v \;/\; \theta \geq 0\}$, donde $v \neq 0$, es convexo, pero no afín. Este es un cono convexo si su base x_0 es 0.

• Cualquier sub-espacio es afín, y un cono convexo (por lo tanto es convexa)

1.8 Hiperplanos de separación y soporte

Se describe una idea que va a ser importante más adelante: el uso de hiperplanos o funciones afines para separar conjuntos convexos que no se intersecan. El resultado básico es el teorema de separación de hiperplano

1.8.1 Teorema de separación del hiperplano

Supongamos que C y D son dos conjuntos convexos cuya intersección es no vacía, es decir $C \cap D = \emptyset$. Entonces existe un $a \neq 0$ y b tales que $\mathrm{a}^T x \leq b$ para todo $x \in C$ y $\mathrm{a}^T x \geq b$ para todo $x \in D$. En otras palabras, la función afín en $\mathrm{a}^T x - b$ es no positiva en C y no negativa sobre D. El hiperplano $\{x \;/\; \mathrm{a}^T x = b\}$ se llama un hiperplano de separación para los conjuntos C y D, o bien que separar los conjuntos C y D. Esto se ilustra en la figura 9.

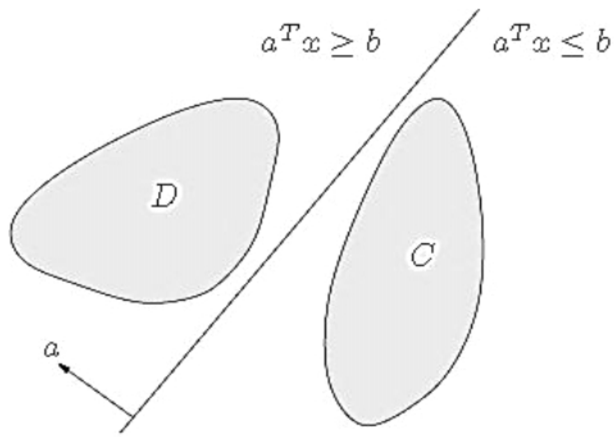


Figura 9

El hiperplano $\{x \ / \ a^T x = b\}$ *separa los conjuntos convexos disjuntos* C *y* D*. La función afín en* $a^T x - b$ *es no positivo en* C *y no negativo en* D

Prueba del teorema

Se desarrollará la prueba en un caso especial, supongamos que la distancia euclidiana entre C y D, que se define como

$$dist\ (C, D) = \inf\ \{\|u - v\|_2 \,/\, u \in C, v \in D\}$$

es positiva, además que existen puntos $c \in C$ y $d \in D$ que logran la mínima distancia, es decir, $\|c - d\|_2 = dist\ (C, D)$. Estas cumplen estas condiciones, por ejemplo, cuando C y D están cerradas y un conjunto es acotado.

Definimos las siguientes expresiones:

$$a = d - c, \quad \wedge \quad b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}$$

el objetivo es demostrar que la función afín

$$f(x) = \mathrm{a}^T x - b = (d-c)^T \left[x - \left(\tfrac{1}{2}\right)(d+c)\right]$$
$$= (d-c)^T \left[x - \left(\frac{1}{2}\right)(d+c)\right]$$

es no positiva en C y no negativa en D, esto es, que el hiperplano $\{x \,/\, \mathrm{a}^T x = b\}$ separa C y D. Este hiperplano es perpendicular al segmento de la línea entre c y d, y pasa a través de su punto medio, como se muestra en la figura 10.

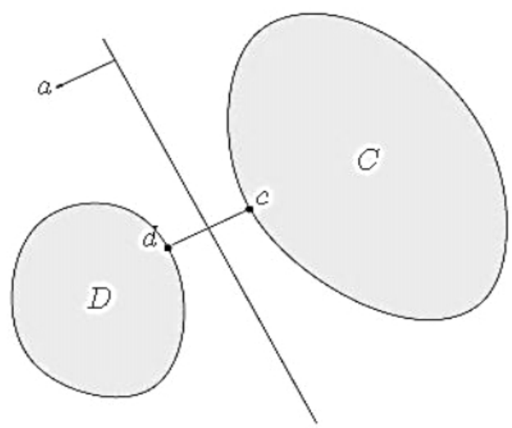


Figura 10

Construcción de un hiperplano de separación entre dos conjuntos convexos. Los puntos $c \in C$ y $d \in D$ son el par de puntos en los dos conjuntos que son los más cercanos el uno al otro. La separación de hiperplano es ortogonal, y biseca, el segmento de línea entre c y d.

En primer lugar, mostramos que f no es negativa en D. La prueba de que f no es positivo en C es similar "a" haciendo el intercambio de C y D y considerando $-f$. Suponga exista un punto $u \in D$:

$$f(u) = (d-c)^T \left[u - \left(\tfrac{1}{2}\right)(d+c)\right] < 0 \tag{5}$$

Podemos expresar $f(u)$ como

$$f(u) = (d-c)^T \left[u - d + \left(\frac{1}{2}\right)(d-c)\right]$$

$$= (d-c)^T (u-d) + \left(\frac{1}{2}\right)\|d-c\|_2^2$$

Vemos que de (5) se implica que $(d-c)^T (u-d) < 0$. Ahora observamos que,

$$\frac{d}{dt}\|d + t(u-d) - c\|_2^2 \Big|_{t=0} = 2(d-c)^T (u-d) < 0$$

Así, para algún $t > 0$, con $t \leq 1$, se tiene

$$\|d + t(u-d) - c\|_2^2 < \|d-c\|_2^2$$

es decir, el punto $d + t(u-d)$ está más cercano a c que a d. Puesto que D es convexo y contiene d y u, tenemos $d + t(u-d) \in D$. Pero esto es imposible, ya que d se supone que es el punto en D que está más cerca a C

1.8.2 La separación estricta

El hiperplano de separación que se construyó anteriormente satisface también la condición más fuerte tal que en $A^T x < b$ para todo $x \in C$ y en $A^T x > b$ para todo

$x \in D$. Esto se llama estricta separación de los conjuntos C y D. Ejemplos sencillos muestran que, en general, conjuntos disjuntos convexos no necesitan ser estrictamente separables por un hiperplano. Para muchos casos especiales, sin embargo, la separación estricta puede ser establecida.

Ejemplo 1.3: (La separación estricta de un punto y un conjunto convexo cerrado). Sea C un conjunto convexo cerrado y $x_0 \notin C$. Entonces existe un hiperplano que separa estrictamente x_0 de C.

Para ver esto, consideremos el conjunto C y $B\ (x_0,\ \mathcal{E})$ que los dos se cortan para algún $\mathcal{E} > 0$.

Por el teorema de separación de hiperplano, existe $a \neq 0$ y b tal que $\mathrm{a}^T x \leq b$ para $x \in C$ y en $\mathrm{a}^T x \geq b$ para $x \in B(x_0, \mathcal{E})$.

Usando a $B(x_0, \mathcal{E}) = \{x_0 + u \ / \ \|u\|_2 \leq \ \mathcal{E}\}$, la segunda condición se puede expresar como

$$\mathrm{a}^T\ (x_0 + u) \geq b \ \text{ para todos } \ \|u\|_2 \leq \ \mathcal{E}$$

la u que minimiza el lado izquierdo es $u = -\mathcal{E}a/\|a\|_2$, utilizando este valor que tenemos

$$\mathrm{a}^T\, x_0 - \mathcal{E}\|a\|_2 \geq b$$

por lo tanto la función afín

$$f(x) = \mathrm{a}^T\, x - b\ \mathcal{E}\|a\|_2\, /2$$

es negativa en C y positiva en x_0.

Como consecuencia inmediata que ya establecimos antes; un conjunto convexo cerrado es la intersección de todos los semi-espacios que lo contienen. Además, C cerrado y convexo, y S la intersección de todos los semi-espacios que contienen C. Obviamente $x \in C \Rightarrow x \in S$. Para demostrar el recíproco, supongamos que existe $x \in S$, $x \notin C$.

Por el teorema de separación estricta, resulta que existe un hiperplano que separa estrictamente x de C, es decir, existe un semi-espacio que contiene a C, pero no a x. En otras palabras, $x \notin S$.

1.8.3 Reciproco del teorema separación del hiperplano

El recíproco del teorema de hiperplano de separación (es decir, la existencia de un hiperplano de separación implica que los conjuntos C y D no se interceptan no se cumple al menos, que se impongan restricciones adicionales en, C o D, aun más allá de la convexidad. Como un simple contraejemplo c, considerare a $C = D = \{0\} \subseteq \mathbb{R}$. En este caso en el hiperplano $x = 0$ se separa C y D.

Añadiendo condiciones sobre C y D diversos recíprocos de teoremas de separación pueden ser derivados. Un ejemplo muy sencillo es el siguiente. Supongamos que C y D son conjuntos convexos, con C abierto, sea una función f afín que es no positiva en C y no negativa en D.

Los conjuntos, C y D son disjuntos. Para ver esto en primer lugar observamos que f debe ser negativa en C, ya que si f fuese nula en un punto de C entonces f debe tomar valores positivos cercanos el punto, lo cual es una contradicción. Pero entonces C y D deben ser disjuntos ya que f es negativa en C y no negativa en D. Poner esta reciproca al teorema de hiperplano de separación, se obtiene el siguiente resultado: Para cualquier par de conjuntos convexos C y D, en la que al menos uno de los dos está abierto, son disjuntos si y sólo si existe un hiperplano de separación.

Ejemplo 1.4: (Teorema de alternativas para desigualdades lineales estricta). Derivamos las condiciones necesarias y suficientes para la resolución de un sistema de desigualdades lineales estrictas de la forma

$$Ax < b \qquad (6)$$

estas desigualdades son infactibles si y sólo si los conjuntos (convexo)

$$C = \{b - \mathrm{A}x \ / \ x \in \ \mathbb{R}^n \}, \qquad D = \mathbb{R}^m_{++} = \{y \in \mathbb{R}^m \ / y > 0\}$$

no se intersecan note que el conjunto D es abierto y C es un conjunto afín. Por lo tanto, por el resultado anterior, C y D son disjuntos si y sólo si existe un hiperplano de separación, es decir, existe un vector no nulo $\lambda \in \mathbb{R}^m$ y $\mu \in \mathbb{R}$ tal que λ^T y $\leq \mu$ en C y λ^T y $\geq \mu$ en D.

Cada una de estas condiciones se puede simplificar. La primera significa que

$$\lambda^T (b - Ax) \leq \mu \text{ para todo } x \in C \ .$$

Ello implica que $A^T \lambda = 0$ y $b\lambda^T \leq \mu$. La segunda desigualdad significa que $\lambda^T y \geq \mu$ para todo $y > 0$. Esto implica $\mu \leq 0$ y $\lambda \geq 0, \lambda \neq 0$.

Poniendo todo junto, nos encontramos con que el conjunto de desigualdades estrictas (6) es infactible si y sólo si existe $\lambda \in \mathbb{R}^m$ tal que

$$\lambda \neq 0, \qquad \lambda \geq 0, \qquad A^T \lambda = 0, \qquad \lambda^T b \leq 0 \tag{7}$$

este es también un sistema de desigualdades lineales y ecuaciones lineales en la variable de $\lambda \in \mathbb{R}^m$.

Decimos que (6) y (7) forman un par de alternativas: para cualquier valores de A y b, exactamente uno de esto tiene solución.

1.8.4 Hiperplanos de apoyo o soporte

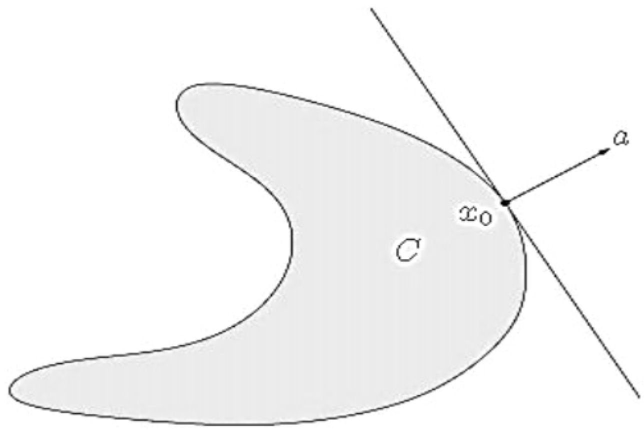
Supongamos que $C \subseteq \mathbb{R}^n$ y x_0 es un punto a su frontera $bd\ C$, es decir,

$$x_0 \in bd\ C = cl\ C \backslash int\ C$$

si $a \neq 0$ satisface en $a^T x \leq a^T x_0$ para todo $x \in C$, entonces el hiperplano $\{x \ / \ a^T x = a^T x_0\}$ se llama un hiperplano de soporte a C en el punto x_0 . Esto es equivalente a decir que el punto x_0 y el conjunto C están separados por el hiperplano $\{x \ / \ a^T x = a^T x_0\}$.

La interpretación geométrica es que el hiperplano $\{x \ / \ a^T x = a^T x_0\}$ es tangente a C en x_0, y el semiespacio $\{x \;/\; a^T x \leq \mathrm{a}^T x_0\}$ contiene C, tal como se aprecia la figura 11.

Figura 11



El hiperplano $\{x \;\;/\; a^T x = a^T x_0\}$ *soporta* C *en* x_0.

Un resultado básico, llamado el teorema del hiperplano de soporte, establece que para cualquier conjunto C, convexo no vacío y cualquier $x_0 \in bd\, C$, existe un hiperplano de soporte para C en x_0. Este teorema se puede demostrar fácilmente a partir del teorema del hiperplano de separación.

Se distinguen para apreciar este resultado dos casos. Si el interior de C no es vacío, el resultado se sigue inmediatamente, aplicando el teorema del hiperplano de separación para los conjuntos $\{x_0\}$ y $int\, C$. Si el interior de C está vacío, entonces C debe pertenecer en un conjunto afín de dimensión menor que n, y cualquier hiperplano que contiene ese conjunto afín contiene C y x_0, y es un hiperplano de apoyo (trivial).También hay un reciproco parcial del teorema de hiperplano de soporte: Si un conjunto está cerrado, tiene interior no vacío, y tiene un hiperplano de soporte en cada punto en su frontera, entonces es convexo.

1.9 Funciones convexas

Definición Una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa si $dom\, f$ es un conjunto convexo y si para todo $x, y \in dom\, f$ y θ con $0 \leq \theta \leq 1$, se tiene

$$f(\theta_x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \quad (8)$$

geométricamente, esta desigualdad significa que el segmento de línea entre $(x, f(x))$ y $(y, f(y))$, la cual es la cuerda de x a y, está por encima de la gráfica de f (ver figura 12).

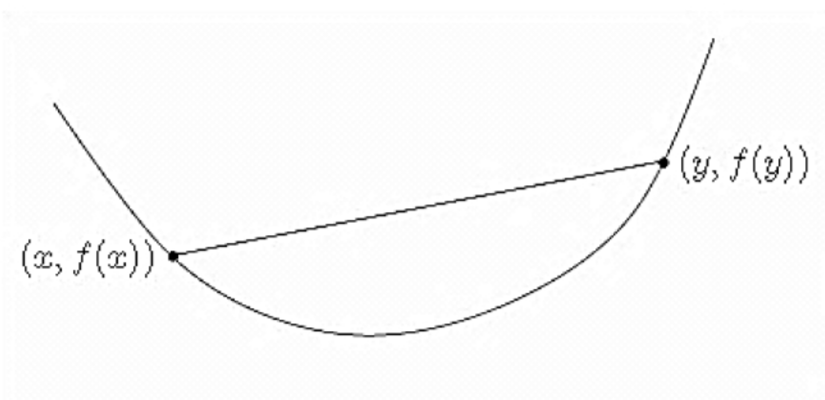


Figura 12

Gráfica de una función convexa. El acorde (es decir, el segmento de línea) entre cualesquiera dos puntos en el gráfico se encuentra por encima de la gráfica.

Una función f es estrictamente convexa si se mantiene la desigualdad tiene en (8) siempre que $x \neq y$ y $0 < \theta < 1$. Decimos f es cóncava si $-f$ es convexa, y estrictamente cóncava si $-f$ es estrictamente convexa.

En una función afín siempre tenemos la igualdad en (8), por lo que toda función afín es convexa y cóncava. Recíprocamente, cualquier función que es convexa y cóncava es afín.

Una función es convexa si y sólo si ella es convexa cuando está restringida a cualquier línea que interseca su dominio. En otras palabras, f es convexa si y sólo si para todo $x \in dom\, f$ y todo v, la función $g(t) = f(x + tv)$ es convexa tal que $g = \{t \,/\, x + tv \ \in dom\, f\}$.

Esta propiedad es muy útil, ya que permite comprobar si una función es convexa limitándola a una línea.

El análisis de las funciones convexas es un campo bien desarrollado, lo cual no será nuestra meta en este trabajo. Un simple resultado, por ejemplo, es que es una función convexa es continua en el interior relativo de su dominio; esta puede tener discontinuidades en su frontera relativa.

1.10 Condiciones de primer orden

Desarrollaremos condiciones de convexidad asumiendo diferenciabilidad en la función. Supongamos que f es diferenciable es decir, su gradiente ∇f existe en cada punto en $dom\, f$, el cual es abierto. Entonces f es convexa si y sólo si $dom\, f$ es convexa y además:

$$f(y) \geq f(x) + \nabla f(x)^T \,(y - x) \qquad (9)$$

Se verifica para todo $x, y \in dom\, f$. Esta desigualdad se ilustra en la figura 13

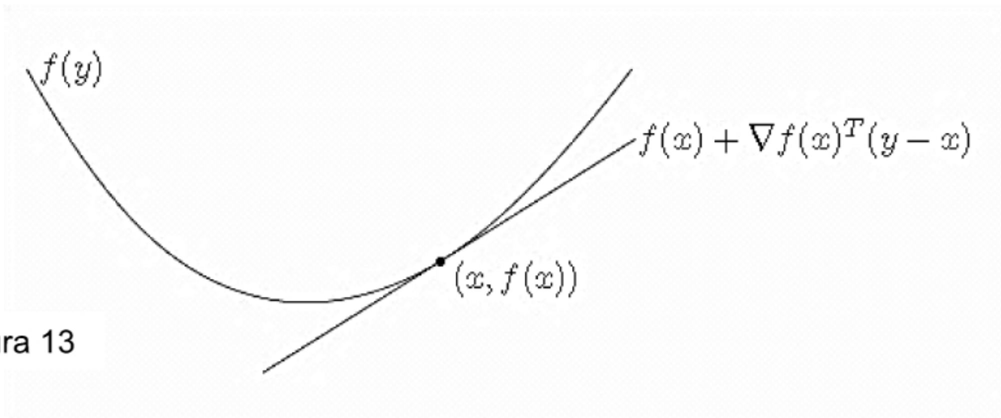


Figura 13

Si f es convexa y diferenciable, entonces $f(x) + \nabla f(x)^T(y - x) \leq f(y)$ para todo $x, y \in$ dom f.

La función afín de y dada por $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ es la primera aproximación de Taylor aproximación de f cercana a x. La desigualdad (9) establece que para una función convexa, la aproximación de Taylor de primer orden es, de hecho, un estimado global mínimo de la función. Inversamente, si la aproximación de Taylor de primer orden de una función es siempre un mínimo estimado global de la función, entonces la función es convexo.

La desigualdad (9) muestra que a partir de la información local acerca de una función convexa (es decir, su valor y su derivada en un punto), podemos obtener información global (es decir, un mínimo estimador global en ese punto). Esta es quizás la propiedad más importante de las funciones convexas, y explica algunas de las propiedades claves de funciones convexas y los problemas de optimización convexa.

Un caso particular en, la desigualdad (9) muestra que si $\nabla f(x) = 0$, entonces para todo $y \in dom\, f, f(y) \geq f(x)$, es decir, x es un minimizador global de la función f.

La convexidad estricta también puede ser caracterizada por una condición de primer orden, y f es estrictamente convexa si y sólo si $dom\, f$ es convexo y para $x, y \in dom\, f$, $x \neq y$, tenemos que:

$$f(y) > f(x) + \nabla f(x)^T (y - x) \tag{10}$$

para las funciones cóncavas tenemos la caracterización correspondiente: f es cóncava si y sólo si el $domf$ es convexa, y:

$$f(y) \leq f(x) + \nabla f(x)^T (y - x)$$

para todo $x, y \in dom\, f$.

A continuación desarrollaremos la demostración condición de convexidad de primer orden.

Prueba de la condición de primer orden de Convexidad

Para probar (9), consideramos en primer lugar el caso $n = 1$: esto es en lo reales $\mathbb{R}$ demuestra que una función diferenciable $f: \mathbb{R} \rightarrow \mathbb{R}$ es convexa si y sólo si

$$f(y) \geq f(x) + f'(x)(y - x) \tag{11}$$

para todo x y y en $dom\, f$.

Supongamos primero que f es convexa y $x, y \in dom\, f$. Puesto que $dom\, f$ es convexo (es un intervalo), se concluye que para todo $0 < t \leq 1, x + t(y - x) \in dom f$, y por convexidad de f,

$$f\big(x + t(y - x)\big) \leq (1 - t)f(x) + tf(y)$$

si dividimos ambos lados por t, la expresión anterior obtenemos lo siguiente:

$$f(y) \geq f(x) + \frac{f\big(x + t(y - x)\big) - f(x)}{\mathrm{t}}$$

y tomando el límite cuando $t \to 0$ se obtiene (11).

Para mostrar la condición suficiencia, asumamos que la función f satisface (11) para todo x, y en $dom\, f$. Escoja cualquier $x \neq y$, y $0 \leq \theta \leq 1$, y sea $z = \theta x + (1 - \theta)y$. Aplicando la expresión (11) se obtiene el doble

$$f(x) \geq f(z) + f'(z)(x - z) \quad ; \quad f(y) \geq f(z) + f'(z)(y - z)$$

multiplicando la primera desigualdad por θ y la segunda por $(1 - \theta)$, y sumando esto nos da:

$$\theta f(x) + (1 - \theta)f(y) \geq f(z)$$

lo que demuestra que f es convexa.

Ahora probemos el caso general, con $f: \mathbb{R}^n \to \mathbb{R}$. Sea $x, y \in \mathbb{R}^n$, y considere f restringida a la línea que pasa a través de ellos, es decir, la función definida por $g(t) = f(ty + (1 - t)x)$, por lo que $g'(t) = \nabla f(ty + (1 - t)x)^T\, (y - x)$.

Asumamos en primer lugar que f es convexa, lo cual implica g es convexa, así por el argumento anterior tenemos que $g(1) \geq g(0) + g'(0)$, lo que significa

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

supongamos ahora que esta desigualdad se verifica para cualquier x y y, tal que $ty + (1-t)x \in dom\, f$ y y $\bar{t}y + (1-\bar{t})x \in \mathrm{dom}\, f$ tenemos

$$f(ty + (1-t)x) \geq f(\bar{t}y + (1-\bar{t})x) + \nabla f(\bar{t}y + (1-\bar{t})x)^T (y - x)(\mathrm{t} - \bar{t})$$

es decir, $g(t) \geq g(\bar{t}) + g'(\bar{t})(\mathrm{t} - \bar{t})$, por lo que vemos que implica que g es convexa.

1.11 Condiciones de segundo orden

Supongamos ahora que f es doblemente diferenciable, es decir, su derivada de

Hessiana o segunda derivada $\nabla^2 f$ existe en cada punto en $dom\, f$, el cual es abierto. Entonces f es convexo si y sólo si $dom\, f$ es convexa y su Hessiana es semi-definida positiva para todo $x \in dom\, f$, esto es;

$$\nabla^2 f(x) \geq 0$$

para una función en $\mathbb{R}$, esto se reduce a la simple condición de $f''(x) \geq 0$ (y $dom\, f$ convexo, y un intervalo), lo que significa que la derivada es no decreciente. La condición $\nabla^2 f(x) \geq 0$ puede interpretarse geométricamente como el requisito de que la gráfica de la función tiene curvatura positiva (hacia arriba) en x.

Del mismo modo, f es cóncava si y sólo si $dom\, f$ es convexa y $\nabla^2 f(x) \leq 0$ para

todo $x \in dom\, f$. La convexidad estricta puede ser parcialmente caracteriza por condiciones de segundo orden. Si $\nabla^2 f(x) > 0$ para todo $x \in dom\, f$, entonces f es estrictamente convexa. La inversa, sin embargo, no es cierto, por ejemplo, la función $f: \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = x^4$ es estrictamente convexa, pero tiene segunda derivada cero en $x = 0$.

Ejemplo 1.5: Funciones cuadráticas. Considere la función cuadrática $f: \mathbb{R} \rightarrow \mathbb{R}$, con $dom\, f = \mathbb{R}^n$, dada por

$$f(x) = (1/2)x^T Px + \mathrm{q}^T x + r$$

con $P \in S^n$ (matrices cuadradas), $q \in \mathbb{R}^n$, y $r \in \mathbb{R}$. Puesto que $\nabla^2 f(x) = P$ para todo x, f es convexa si y sólo si $P \geq 0$ (y cóncava si y sólo si $P \leq 0$).

Para las funciones cuadráticas, la convexidad estricta es fácilmente caracterizada:f es estrictamente convexa si y sólo si $P > 0$ (y estrictamente cóncava si y sólo si $P < 0$).

Observación: El requisito de separar que al $dom\, f$ sea convexo no puede ser descartado del primer o segundo orden para las caracterizaciones de convexidad y concavidad. Por ejemplo, la función $f(x) = (1/x^2)$, con

$dom\, f = \{x \in \mathbb{R} \ / \ x \neq 0\}$, satisface $f''(x) > 0$ para todo $x \in dom\, f$, pero vemos que f no es una función convexa.

1.12 Propiedades especiales de funciones convexas

1.12.1 Conjuntos subnivel

El conjunto $\propto$- subnivel de una función $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ se define como

$$C_\propto = \{x \in domf / f(x) \leq \propto\}$$

los conjuntos subnivel de una función convexa son convexos, para cualquier valor de $\propto$. La prueba es inmediata a partir de la definición de la convexidad:

si $x, y \in C_\propto$, entonces $f(x) \leq \propto$ y $f(y) \leq \propto$, y por lo tanto $f(\Theta x + (1 - \Theta)y) \leq \propto$ para $0 \leq \Theta \leq 1$, y por lo tanto $\Theta x + (1 - \Theta)y \leq$ y $\in C_\propto$.

Lo contrario no es cierto: una función puede tener todos sus conjuntos de subnivel convexo, pero no es una función convexa. Por ejemplo, $f(x) = e^x$ es no convexa en $\mathbb{R}$ (de hecho, es estrictamente cóncava), pero todos sus conjuntos de supranivel son convexos.

Si f es cóncava, entonces su conjunto $\propto$-supranivel, dada por $\{x \in domf \ / \ f(x) \geq \propto\}$, es un conjunto convexo. El conjunto subnivel es a menudo una buena propiedad manera de establecer la convexidad de un conjunto, expresándola como un conjunto subnivel de una función convexa, o como el conjunto supranivel de una función cóncava.

Ejemplo 1.6: Las medias geométricas y aritméticas de $x \in \mathbb{R}_+^n$, son, se definen como:

$$G(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n} \quad , \quad A(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(donde se toma $0^{1/n} = 0$ para definir a definición de G).

Los estados de desigualdad media aritmética-geométrica establece que $G(x) \leq A(x)$.

Supongamos que $0 \leq \propto \leq 1$, y consideremos el conjunto

$$\{x \in \mathbb{R}_+^n \ / \ G(x) \geq \propto A(x)\}$$

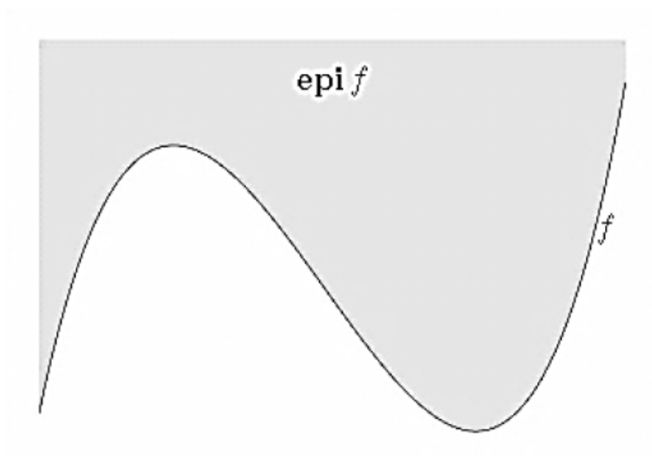
es decir, el conjunto de vectores con media geométrica, al menos, tan grande como un factor $\propto$ veces la media aritmética. Este conjunto es convexo, ya que es el conjunto 0- supranivel de la función $G(x) - \propto A(x)\}$, que es cóncava. De hecho, el conjunto es positivamente homogéneo, por lo que es un cono convexo.

1.12.2 Epígrafo / Hipografo de una Función Convexa

La gráfica de una función $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ se define como

$$\{(x, f(x)) \ / \ x \in dom\, f\}$$

que es un subconjunto de $\mathbb{R}^{n+1}$. El epígrafe de una función $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ denotado como $\operatorname{epi} f$ define como $\operatorname{epi} f = \{(x, t) \ / \ x \in \operatorname{dom} f, \ f(x) \leq t\}$ que es un subconjunto de 'Epi' (significa “por encima de” modo epígrafo signifique por encima de la gráfica). Esta definición se ilustra en la figura 14.



Epígrafe de una función f, se muestra sombreada. El límite inferior, se muestra más oscuro, es la gráfica de f .

El vínculo entre los conjuntos convexos y funciones convexas es a través de ser epígrafo, esto es una función convexa si y sólo si su epígrafo es un conjunto convexo. Una función es cóncava si y sólo si su hipografo, definido como $\text{hypo}\, f$, tal que:

$$\text{hypo}\, f = \{(x, t) \,/\, \ t \le f(x)\}$$

es un conjunto convexo.

1.12.3 La Desigualdad de Jensen

La desigualdad básica (8), es decir, $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-)f\theta(y)$ a veces se llama desigualdad de Jensen. Se extiende fácilmente a combinaciones convexas de más de dos puntos: Si f es convexa, $x_1, \dots, x_k \in \mathrm{dom}\, f$, y $\theta_1, \dots, \theta_k \geq 0$ con $\theta_1 + \dots + \theta_k = 1$, entonces

$$f(\theta_1 x_1 + \dots + \theta_k x_k) \leq \theta_1 f(x_1) + \dots + \theta_k f(x_k)$$

al igual que en el caso de conjuntos convexos, la desigualdad se extiende a infinita sumas, integrales, y valores esperados. Por ejemplo, si $p(x) \geq 0$ en $S \subseteq \mathrm{dom}\, f$, $\int_S p(x)\, dx = 1$,

entonces

$$f\left(\int_S p(x)\, x\, dx \right) \leq \int_S f(x) p(x)\, dx$$

proporcionado existen las integrales. En el caso más general, podemos tomar cualquier probabilidad de medir con el apoyo en $\mathrm{dom}\, f$. Si x es una variable aleatoria tal que $x \in \mathrm{dom}\, f$ con probabilidad uno, y f es convexa, entonces tenemos

$$f(\mathrm{E}x) \leq \mathrm{E}\, f(x) \qquad (12)$$

Siempre y cuando existan las expectativas. Podemos recuperar la desigualdad básica (8) de esta forma general, al tomar la variable aleatoria x tener soporte $\{x_1, x_2\}$, con prob $(x = x_1) = \theta$, prob $(x = x_2) = 1 - \theta$. Así, la desigualdad (12) caracteriza convexidad: Si f no es convexa, hay una variable aleatoria x, con $x \in \operatorname{dom} f$ con probabilidad uno, tal que $f(\mathrm{E}x) > \mathrm{E}\, f(x)$.

Todas estas desigualdades se llaman ahora la desigualdad de Jensen, aunque la desigualdad estudiado por Jensen fue muy simple

$$f\left(\frac{x + y}{2}\right) \leq \frac{f(x) + f(y)}{2}$$

1.13 Algunos tipos de funciones convexas

1.13.1 Funciones cuasi-convexa

Una función $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, se llama cuasi-convexa si su dominio y todos sus conjuntos de subnivel

$$S_{\propto} = \{x \in \operatorname{dom} f \ / \ f(x) \leq \propto\}$$

para $\propto \in \mathbb{R}$, son convexos. Una función que es a la vez cuasi-convexa y cuasi-cóncava se llama cuasi-lineal. Si una función f es casi lineal, entonces su dominio, y cada conjunto de nivel $\{x \ / \ f(x) = \propto\}$ es convexo.

Para una función de $\mathbb{R}$, cuasi-convexa requiere que cada conjunto subnivel debe haber un intervalo (incluyendo, posiblemente, un intervalo infinito). Un ejemplo de función cuasi-convexo en $\mathbb{R}$ se ilustra en la figura 16

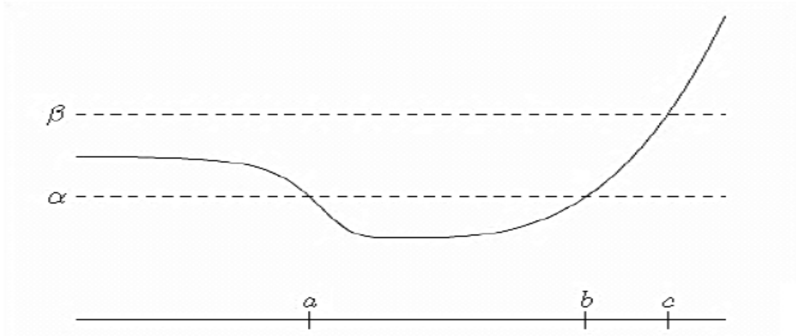


Figura 16

Una función cuasi-convexo en $\mathbb{R}$. Para cada α el α −subnivel establece S_α es convexa, es decir, un intervalo. El subnivel establece S_α es el intervalo $[a, b]$. El conjunto de subnivel S_β es el intervalo $(-\infty, c]$.

Ejemplo 1.7: Algunos ejemplos en $\mathbb{R}$:

• Logaritmo. $\log x$ en $\mathbb{R}_{++}$ es cuasi-convexa (y cuasi-cóncava, por lo tanto, cuasi-lineal).

• Función de techo. $\operatorname{ceil}(x) = \inf \{z \in Z \,/ \geq x\}$ es cuasi-convexo (y cuasi-cóncava).

Estos ejemplos muestran que las funciones cuasi-convexo pueden ser cóncava, o discontinua. Damos ahora algunos ejemplos en $\mathbb{R}^n$.

Ejemplo 1.8: Longitud de un vector. Se define la longitud de $x \in \mathbb{R}^n$ como el mayor índice de un componente distinto de cero, es decir,

$$f(x) = \max \{i \,/\; x_i \neq 0\}$$

(Se define la longitud del vector cero a ser cero.) Esta función es cuasi-convexo en $\mathbb{R}^n$, ya que sus conjuntos de subnivel son sub-espacios:

$$f(x) \leq \propto \Leftrightarrow \; x_i = 0 \text{ para } i = \lfloor \propto \rfloor + 1, \ldots, n.$$

Ejemplo 1.9: Considere $f: \mathbb{R}^2 \longrightarrow \mathbb{R}$, con $\operatorname{dom} f = \mathrm{R}_+^2$ y $f(x_1, x_2) = x_1 x_2$. Esta función no es ni convexa ni cóncava desde su Arpillera

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

es indefinida, sino que tiene una positiva y una negativa de valores propios . La función f es cuasi-cóncava, sin embargo, ya que los conjuntos de subnivel

$$\{ x \in \mathbb{R}_+^2 \; / \; x_1, x_2 \geq \propto \}$$

son conjuntos convexos para todo $\propto$. (Tenga en cuenta, sin embargo, que f no es cuasi-cóncava en $\mathbb{R}^2$).

Ejemplo 1.10: Función lineal fraccional. Sea f la siguiente función.

$$f(x) = \frac{a^T x + b}{c^T x + d}$$

es cuasi-convexa, y cuasi-cóncava, es decir, cuasi-lineal.

Su conjunto de $\propto$ -subnivel es

$$S_\propto = \{x \, / \, \mathrm{c}^T x + d > 0, (\mathrm{a}^T x + b)/(\mathrm{c}^T x + d) \leq \propto\}$$

$$= \{x \, / \, \mathrm{c}^T x + d > 0, \; \mathrm{a}^T x + b \leq \propto (\mathrm{c}^T x + d)\}$$

que es convexa, ya que es la intersección de un semi-espacio abierto y un semi-espacio cerrado.

Ejemplo 1.11: Función de relación de distancia. Supongamos que $a, b \in \mathbb{R}^n$, y definir

$$f(x) = \frac{\|x-a\|_2}{\|x-b\|_2}$$

es decir, la relación de la distancia euclidiana a una a la distancia a, b.

Entonces f es convexa en el semi-espacio $\{x \, / \, \|x-a\|_2 \leq \|x-b\|_2\}$. Para ver esto, consideremos el conjunto $\propto$ -subnivel de f, con $\propto \leq 1$ ya que $f(x) \leq 1$ en el semi-espacio $\{x \, / \, \|x-a\|_2 \leq \|x-b\|_2\}$. Este conjunto subnivel es el conjunto de puntos que satisface

$$\|x-a\|_2 \leq \propto \|x-b\|_2$$

la cuadratura de ambos lados, y reordenando los términos, vemos que esto es equivalente a

$$(1-\propto^2)x^T x - 2(a - \propto^2 b)^T x + a^T a - \propto^2 b^T b \leq 0$$

esto describe un conjunto convexo (de hecho una bola euclídea) si $\propto \leq 1$.

1.13.2 Máximo puntual

Si f_1 y f_2 son funciones convexas entonces su máximo f puntual, definido por

$$f(x) = \max \{ f_1(x), f_2(x)\}$$

con $\operatorname{dom} f = \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, también es convexa. Esta propiedad se comprueba fácilmente: si $0 \leq \theta \leq 1$ y $x, y \in \operatorname{dom} f$. entonces

$$f(\theta x + (1-\theta)y) = \max\ \{ f_1\ (\theta x + (1-\theta)y, f_2(\theta x + (1-\theta)y)\}$$

$$\leq \max\ \{\theta\, f_1(x) + (1-\theta) f_1(y), \theta f_2(x) + (1-\theta) f_2(y)\}$$

$$\leq \theta \max\ \{ f_1(x), f_2(x)\} + (1-\theta)\max\{ f_1(y), f_2(y)\}$$

$$= \theta f(x) + (1-\theta) f(y)\}$$

que establece convexidad de f. Se demuestra fácilmente que si $f_1,\ldots, f_m$ son convexos, entonces su máximo punto a punto

$f(x) = \max\ \{ f_1(x),\ldots, f_m(x)\}$ también es convexa.

Ejemplo 1.12: Funciones lineal a tramos. La función

$$f(x) = \max\ \{a_1^T x + b_1, \ldots, a_L^T x + b_L\}$$

Define una función (con L o menos regiones) lineal a trozos (de verdad, o afín) es convexa, ya que es la máxima puntual de funciones afines.

Lo contrario también puede ser mostrado: cualquier función convexa lineal a tramos con L o menos regiones pueden ser expresadas en este formulario.

1.13.3 Composición de funciones

En esta sección se examinan las condiciones en $\mathrm{h} : \mathbb{R}^k \to \mathbb{R}$ y $\mathrm{g} : \mathbb{R}^n \to \mathbb{R}^k$ que

garantizan convexidad o concavidad de su composición $f = \mathrm{h} \circ \mathrm{g} : \mathbb{R}^k \to \mathbb{R}$, definido por

$$f(x) = h\big(g(x)\big), \qquad \mathrm{dom}\, f \ = \{x \ \in \mathrm{dom}\, g \ / \ g(x) \in \mathrm{dom}\, h\}$$

1.13.3a Composición escalar

Consideremos en primer lugar el caso $k = 1$, por lo que $h : \mathbb{R} \to \mathbb{R}$ y $g : \mathbb{R}^n \to \mathbb{R}$. Podemos restringir nosotros mismos para el caso $n = 1$ (desde convexidad está determinada por el comportamiento de un funcionar en líneas arbitrarias que se cruzan su dominio).

Para descubrir las reglas de composición, se comienza suponiendo que h y g son dos veces diferenciable, con $\mathrm{dom}\, g = \mathrm{dom}\, h = \mathrm{R}$. En este caso, la convexidad de f se reduce a $f'' \geq 0$ (es decir, $f''(x) \geq 0$ para todo $x \in \mathbb{R}$).

La segunda derivada de la función composición $f = \mathrm{h} \circ \mathrm{g}$ está dada por

$$f''(x) = h''\big(g(x)\big)\, g'(x)^2 + h'\big(g(x)\big)\, g''(x) \tag{13}$$

Ahora supongamos que, por ejemplo, que g es convexa (para $g'' \geq 0$) y h es convexa y no decreciente (de modo $h'' \geq 0$ y $h \geq 0$). Se deduce de (13) que $f'' \geq 0$, es decir, f es convexa. De manera similar, la expresión (13) da los resultados:

f es convexa si h es convexa y no decreciente, y g es convexa,

f es convexa si h es convexa y no creciente, y g es cóncava, (14)

f es cóncava si h es cóncava y no decreciente, y g es cóncava,

f es cóncava si h es cóncava y no creciente, y g es convexa.

Estas declaraciones son válidas cuando las funciones g y h son dos veces diferenciables y tienen dominios que son todas $\mathbb{R}$. Resulta que las reglas de composición muy similares mantener en el caso general $n > 1$, sin asumir la diferenciabilidad de h y g o que $\operatorname{dom} g = \mathbb{R}^n$ y $\operatorname{dom} h = \mathbb{R}$:

f es convexa si h es convexa, $\tilde{h}$ es no decreciente, y g es convexa,

f es convexa si h es convexa, $\tilde{h}$ se no creciente, y g es cóncava,

f es cóncava si h es cóncava, $\tilde{h}$ es no decreciente, y g es cóncava, (15)

f es cóncava si h es cóncava, $\tilde{h}$ su no creciente, y g es convexa.

Aquí $\tilde{h}$ denota la extensión valor ampliada de la función h , que asigna el ∞ valor $(-\infty)$ para no puntos en $\operatorname{dom} h$, para h convexa (cóncava). La única diferencia entre estos resultados y los resultados en (14), es que se requiere que la extensión de valor ampliada de la función de extensión $\tilde{h}$ puede no ser creciente o no decreciente, sobre todo de $\mathbb{R}$.

Para entender lo que esto significa, supongamos h es convexa, por lo que $\tilde{h}$ toma el valor ∞ $\operatorname{dom} h$ exterior. Para decir que $\tilde{h}$ es no decreciente significa que para cualquier $x, y \in \mathrm{R}$, con $x < y$, tenemos $\tilde{h}(x) \leq \tilde{h}(y)$. En particular, esto significa que si $y \in \operatorname{dom} h$, entonces $x \in \operatorname{dom} h$. En otras palabras, el dominio de h se extiende infinitamente en la dirección negativa; esto es, R o un intervalo de la forma $[(-\infty, a)$ o $(-\infty, A)]$. De una manera similar, decir que h es convexa y $\tilde{h}$ es no creciente significa que h es no creciente y $\operatorname{dom} h$ se extiende infinitamente en la dirección positiva. (Esto se ilustra en la figura 17)

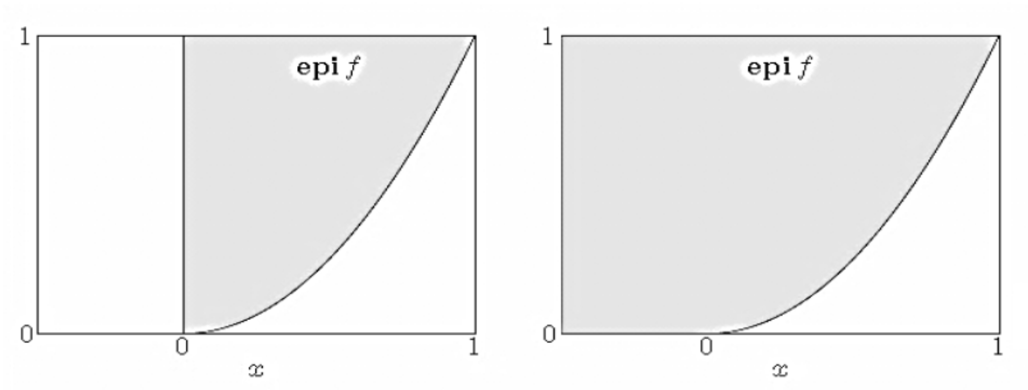


Figura 17

Izquierda. La función x^2, con dominio$\mathbb{R}_+$, es convexa y no decreciente en su dominio, pero su extensión valor ampliada no es no decreciente. Derecha. La función max,$\{x, 0\}^2$ con el dominio $\mathbb{R}$, es convexo, y su extensión valor extendida no es decreciente

Ejemplo 1.13: Algunos ejemplos simples ilustrarán las condiciones en las que aparecen h en los teoremas de composición.

• La función $h(x) = \log x$, con $\operatorname{dom} h = \mathbb{R}_{++}$, es cóncava y satisface $\tilde{h}$ no decreciente.

• La función $h(x) = x^{1/2}$, con $\operatorname{dom} h = \mathbb{R}_{++}$, es cóncava y satisface la condición $\tilde{h}$ no decreciente.

• La función $h(x) = x^{3/2}$, con $\operatorname{dom} h = \mathrm{R}_+$, es convexa pero no satisface la condición $\tilde{h}$ no decreciente. Por ejemplo, tenemos $\tilde{h}\,(-1) = \infty$, pero $\tilde{h}\,(-1) = 1$

• La función $h(x) = x^{3/2}$ para $x \geq 0$, $h(x) = 0$ para $x < 0$, con $\mathrm{dom}\, h = \mathbb{R}$, es convexa y no satisfacer la condición $\tilde{h}$ no decreciente.

Los resultados de la composición (15) se pueden probar directamente, sin asumir la diferenciabilidad, o el uso de la fórmula (13). Como ejemplo, vamos a probar el siguiente teorema de composición: si g es convexa, h es convexa, y $\tilde{h}$ es no decreciente, entonces $f = \mathrm{h} \circ \mathrm{g}$ es convexa. Supongamos que $x, y \in \mathrm{dom}\, f$, y $0 \leq \theta \leq 1$. Como $x, y \in \mathrm{dom}\, f$, tenemos que $x, y \in \mathrm{dom}\, g$ y $g(x), g(y) \in \mathrm{dom}\, h$. Desde $\mathrm{dom}\, g$ es convexa, llegamos a la conclusión de que $\Theta x + (1 - \Theta) y \in \mathrm{dom}\, g$, y de convexidad de g, que tienen.

$$g(\theta x + (1 - \theta) y) \leq \theta g(x) + (1 - \theta) g(y) \qquad (16)$$

como $g\,(x),\ g\,(y) \in \mathrm{dom}\, h$, llegamos a la conclusión de que $\theta g(x) + (1 - \theta) g(y) \in \mathrm{dom}\, h$ es decir, el lado derecho de (16) es en $\mathrm{dom}\, h$. Ahora usamos el supuesto de que $\tilde{h}$ es no decreciente, lo que significa que su dominio se extiende infinitamente en la dirección negativa. Desde el lado derecho de (16) es en $\mathrm{dom}\, h$, llegamos a la conclusión de que el lado izquierdo, es decir, $g(\theta x + (1 - \theta) y) \in \mathrm{dom}\, h$.

Esto significa que $\theta x + (1 - \theta) y \in \mathrm{dom}\, f$.

En este punto, hemos demostrado que $\mathrm{dom}\, f$ es convexa.

Ahora, utilizando el hecho de que $\tilde{h}$ es no decreciente y la desigualdad (16), obtenemos

$$h(g(\theta x + (1 - \theta)y) \leq h(\theta g(x) + (1 - \theta)g(y)) \qquad (17)$$

de convexidad de h, tenemos

$$h(\theta g(x) + (1 - \theta)g(y)) \leq \theta h g(x) + (1 - \theta)\, h\big(g(y)\big) \qquad (18)$$

poner (17) y (18) en conjunto, tenemos

$$h(g(\theta x + (1 - \theta)y)) \leq \theta h(g(x)) + (1 - \theta)\, h\big(g(y)\big)$$

lo que demuestra el teorema de composición.

Ejemplo 1.14: resultados de la composición simples.

• Si g es convexa entonces $\exp\ g(x)$ es convexa.

• Si g es cóncava y positivo, entonces $\log g(x)$ es cóncava.

• Si g es cóncava y positivo, a continuación, $1/g(x)$ es convexa.

• Si g es convexa y no negativa y $p \geq 1$, entonces $g(x)^p$ es convexa.

• Si g es convexa entonces $\log(-g(x))$ es convexa en $\{x \ / \ g\ (x) \ < 0\,\}$.

Observación: El requisito de que la monotonicidad se mantiene para la extensión de amplio valor $\tilde{h}$, y no sólo la función h, no se pueden quitar. Por ejemplo, considere la función $g(\mathrm{x}) = x^2$, con $\mathrm{dom}\ g = \mathrm{R}$, y $h(x) = 0$, con

$\mathrm{dom}\, h = [1, 2]$. Aquí g es convexa, y h es convexa y no decreciente.

Pero la función $f = \mathrm{h} \circ g$, dada por

$$f(x) = 0, \qquad \mathrm{dom}\, f = \left[-\sqrt{2, -1}\,\right] \cup \left[1, \sqrt{2}\,\right]$$

no es convexa, desde su dominio no es convexa. Aquí, por supuesto, la función $\tilde{h}$ no es decreciente.

1.13.3b Composición vectorial

Pasamos ahora al caso más complicado cuando $k \geq 1$. Suponer

$$f(x) = h(g(x) = h(g_1(x), \dots, g_k(x))$$

con $h : \mathbb{R}^k \to \mathbb{R},\ g_i : \mathbb{R}^n \to \mathbb{R}$. Una vez más, sin pérdida de generalidad se puede suponer $n = 1$. Al igual que en el caso $k = 1$, empezamos asumiendo todas las funciones son dos veces diferenciable, con $\mathrm{dom}\, g = \mathbb{R}$ y $\mathrm{dom}\, h = \mathbb{R}^k$, con el fin de descubrir las reglas de composición. Tenemos

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x) \tag{19}$$

que es el análogo del vector (13). Una vez más la cuestión consiste en determinar las condiciones bajo que $f''(x) \geq 0$ para todo $x($ o $f''(x) \leq 0$ para todo x de concavidad). A partir de (19) que pueden derivar muchas reglas, por ejemplo:

f es convexa si h es convexa, h es no decreciente en cada argumento, y g_i son convexos,

f es convexa si h es convexa, h es no creciente en cada argumento, y g_i son cóncavas,

f es cóncava si h es cóncava, h es no decreciente en cada argumento, y g_i son cóncavas.

Como en el caso escalar, los resultados de la composición similares tienen, en general, con $n > 1$, ninguna suposición de diferenciabilidad $\mathrm{h} \circ g$, y dominios generales. Para que los resultados generales, la condición de monotonicidad de h debe mantener durante la extensión de amplio valor $\tilde{h}$.

Para entender el significado de la condición de que la extensión de amplio valor $\tilde{h}$ ser monótona, consideramos el caso donde $h : \mathbb{R}^k \to \mathbb{R}$ es convexa, y $\tilde{h}$ no decreciente, es decir, cada vez que $\mathcal{U} \leq \mathcal{V}$, tenemos $\tilde{h}\,(\mathcal{U}) \leq \tilde{h}\,(\mathcal{V})$. Esto implica que si $\mathcal{V} \in \operatorname{dom} h$, entonces también lo es $\mathcal{U}$: el dominio de h debe extenderse infinitamente en direcciones al $-\mathbb{R}_+^k = \operatorname{dom} h$.

1.13.4 Perspectiva de una función

Si $f : \mathbb{R}^n \to \mathbb{R}$, entonces la perspectiva de f es la función $g : \mathbb{R}^{n+1} \to \mathbb{R}$ definida por

$$g(x, t) = tf(x/t)$$

con dominio

$$\operatorname{dom} g = \{(x, y) \;/\; x/t \;\in \operatorname{dom} f, \qquad t > 0\}$$

la operación perspectiva conserva convexidad: Si f es una función convexa, entonces que así es su función perspectiva g. Del mismo modo, si f es cóncava, entonces también lo es g .

Esto se puede probar varias maneras, por ejemplo, la verificación directa de la definición la desigualdad.

Damos una breve prueba aquí usando epígrafes y el mapeo perspectiva sobre $\mathbb{R}^{n+1}$ (que también le explicará el nombre "Perspectiva"). Para $t > 0$ tenemos

$$(x, t, s) \in \operatorname{epi} g \iff tf(x/t) \leq s,$$

$$\iff f(x/t) \leq s/t$$

$$\iff (x/t, s/t) \;\in\; \operatorname{epi} f$$

por lo tanto $\operatorname{epi} g$ es la imagen inversa de $\operatorname{epi} f$ bajo el mapeo perspectiva que

toma $(u\,, v, w)$ a $(u\,, w)/v)$. De ello se desprende que $\operatorname{epi} g$ es convexa, por lo que la función g es convexa.

Ejemplo 1.15: norma euclídeana al cuadrado. La perspectiva de la función convexa. $f(x) = x^T x$ en $\mathbb{R}^n$ en $g(x, t) = t(x/t)^T\,(x/t) = \frac{x^T x}{t}$ que es convexa en (x, t) para $t > 0$.

Podemos deducir convexidad de g utilizando otros métodos. En primer lugar, podemos expresar g como la suma de las funciones cuadrática sobre funciones lineales x_i^2/t . También podemos expresar g como un caso especial de la matriz fraccional función $x^T(tI)^{-1}$.

1.13.5 La función conjugada

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$. La función $f^* : \mathbb{R}^n \rightarrow \mathbb{R}$, definida como

$$\sup_{x \in \operatorname{dom} f} (y^T x - f(x))$$

se denomina el conjugado de la función f. El dominio de la función de conjugado consiste de $y \in \mathbb{R}^n$ por que el supremo es finito, es decir, para los que la diferencia $y^T x - f(x)$ está acotado superiormente en $\operatorname{dom} f$. (Esta definición se ilustra en la figura 18).

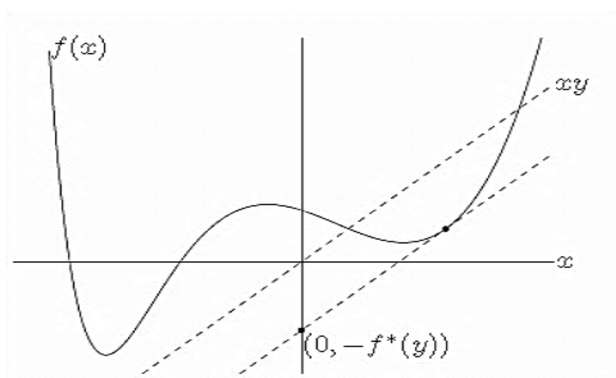


Figura 18

Una función $f: \mathbb{R} \longrightarrow \mathbb{R}$, y un valor de $y \in \mathbb{R}$. La conjugada función $f^+(y)$ es la diferencia máxima entre el yx función lineal y $f(x)$, como se muestra por la línea discontinua en la figura. Si f es diferenciable, este se produce en un punto x donde $f'(x) = y$.

Vemos inmediatamente que f^* es una función convexa, ya que es el punto a punto supremo de una familia de convexos (de hecho, afín) funciones de y.

Esto es cierto si f es convexa. (Tenga en cuenta que cuando f es convexa, el subíndice $x \in \operatorname{dom} f$ no es necesaria ya que, por convención, $y^T x - f(x) = -\infty$ para $x \notin \operatorname{dom} f$).

Comenzamos con algunos ejemplos sencillos, y luego describimos algunas reglas para conjugar funciones. Esto nos permite derivar una expresión analítica para el conjugado de muchas de las funciones convexas comunes.

Ejemplo 1.16: Derivamos los conjugados de algunas funciones convexas en $\mathbb{R}$

• Función afín. $f(x) = ax + b$. Como una función de $x, yx - ax - b$, está acotado si y sólo si $y = a$, en cuyo caso es constante. Por lo tanto el dominio de la función conjugada f^* es el singleton $\{a\}$ y $f^*(a) = -b$.

• Logaritmo negativo. $f(x) = -\log x$, con $\operatorname{dom} f = \mathbb{R}_{++}$. La función de $xy + \log x$ es sin límites por encima de si $y \geq 0$ y alcanza su máximo en $x = -1/y$ de lo contrario. Por lo tanto $\operatorname{dom} f^* = \{y \,/\, y < 0\} = -\mathrm{R}_{++}$ y $f^*(y) = -\log(-y) - 1$ para $y \geq 0$.

• Exponencial. $f(x) = e^x$ $.xy - e^x$ no está acotado si $y < 0$. Para $y > 0, xy - e^x$ alcanza su máximo en $x = \log y$, por lo que tenemos que f * (y) $f^*(y) = y \log y - y$. Para $y = 0$, $f^*(y) = \sup_x -e^x = 0$.

CAPÍTULO 2

OPTIMIZACIÓN CONVEXA

2.1. Problema de optimización

Este capítulo desarrollará los principales fundamentos matemáticos de la optimización convexa. Antes, introducimos algunos conceptos básicos necesarios que ayudaran a formalizar las ideas básicas.

Consideremos el siguiente problema de optimización

$$\begin{aligned} &\text{Minimizar} \quad f_0(x) \\ &\text{Sujeto a} \quad f_i(x) \leq 0, \quad i = 1, . \ . \ . , m \\ &\qquad\qquad\quad h_j(x) = 0, \quad j = 1, . \ . \ . , p \end{aligned} \tag{20}$$

que describiré el problema de encontrar una x que minimiza $f_0(x)$ entre todos las x que satisfacen las condiciones:

$$f_i(x) \leq 0, \quad i = 1, . \ . \ . , m, \text{ y } \ h_j(x) = 0, \quad j = 1, . \ . \ . , p.$$

Llamamos $x \in \mathbb{R}^n$ la variable de optimización y la función $f_0: \mathbb{R}^n \longrightarrow \mathbb{R}$ la función objetivo, función de los costos. Las desigualdades $f_i(x) \leq 0,$ se denominan restricciones de desigualdad, y las funciones correspondientes $f_i: \mathbb{R}^n \longrightarrow \mathbb{R}$ se llaman las funciones de restricciones de desigualdad.

Las ecuaciones $h_i(x) = 0,$ se llaman las restricciones de igualdad, así como las funciones $h_i: \mathbb{R}^n \longrightarrow \mathbb{R}$ son las funciones de restricciones de igualdad. Si no hay restricciones (es decir, $m = p = 0$) se dice que el problema (20) es un problema sin restricciones.

El conjunto de puntos por la que se define el objetivo y todas las funciones de restricción.

Se denota por D tal que

$$D = \bigcap_{i=o}^{m} dom\, f_i \ \cap \ \bigcap_{i=1}^{p} dom\, h_i$$

y se llama el dominio del problema de optimización (20). Un punto $x_0 \in D$ se dice, siempre que se cumplan sus restricciones

$$f_i(x_0) \leq 0, \quad i = 1,.\ .\ ., m, \quad y \quad h_j(x_0) = 0, \quad j = 1,.\ .\ ., p$$

El problema (20) se denomina factible, si existe al menos un punto factible en D, y de lo contrario es factible. El conjunto de todos los puntos factibles se llama el conjunto factible.

El valor óptimo del problema (20) se define como

$$p^* = \inf\ \{f_0(x)\ | f_i(x) \leq 0 \quad i = 1,.\ .\ ., m,\ h_i(x) = 0 \quad i = 1,.\ .\ ., p\ \}$$

que p^* asuma los valores extendidos $\pm\infty$. Si el problema es no factible, tenemos $p^* = \infty$ (siguiendo la convención estándar que el ínfimo del conjunto vacío es ∞). Si hay puntos factibles x_k con $f_0(\, x_k) \longrightarrow -\infty$ cuando $k \longrightarrow \infty,$, entonces $p^* = -\infty$, y decimos que el problema (20) no está acotado.

2.1.2 Puntos óptimos y localmente óptimos

Recordando lo anterior x^* es un punto óptimo, o bien resuelve el problema (20), si x^* es factible y además que se $f_0(x^*) = p^*$. El conjunto de todos los puntos óptimos es el conjunto x_{0pt} y es tal que

$$x_{0pt} = \{x \,/\, f_i(x) \leq 0, \quad i = 1,. . .,m, \; h_i(x) = 0, \quad i = 1,. . .,p, \qquad f_0(x) = p^* \}$$

Si existe un punto óptimo para el problema (20), decimos que el valor óptimo se logra o se consigue, y el problema (20), tiene solución. Si x_{0pt} está vacía, se dice que el valor óptimo no se alcanza o no se consigue. (Esto siempre ocurre cuando el problema no está acotado). Un punto x viables con $f_0(x) \leq p^* + \in$ (donde $\in > 0$) se llama ∊-subóptima, y el conjunto de todos los puntos

∊-subóptimas se llama ∊-subóptima conjunto para el problema (20).

Decimos un punto x es factible a nivel local óptima si hay un r >0 tal que

$$f_0(x) = \inf \; \{f_0(z) \,/\, f_i(z) \leq 0, \; i = 1,. \; .,m, \; h_i(z) = 0, \; \; i = 1,..,p, \|z - x\|_2 \leq r\}$$

o, en otras palabras, x resuelve el problema de optimización

$$\begin{aligned} \text{Minimizar} \quad & f_0(z) \\ \text{Sujeto a} \quad & f_i(z) \leq 0, \quad i = 1,. . .,m \\ & h_j(z) = 0, \quad j = 1,. . .,p \\ & \|z - x\|_2 \leq r \end{aligned}$$

con la variable z . En términos generales, esto significa que x minimiza a f_0 más lugares en el conjunto factible. El término óptimo global "se utiliza a veces para indicar" óptima y distinguir entre "localmente óptima" y "óptima". Sin embargo, óptima significará óptima a nivel global.

Si x es factible y $f_i(x) = 0,$ es decir la restricción de desigualdad $f_i(x) \leq 0,$ es activa en x. Por otro lado, si $f_i(x) < 0,$ la restricción $f_i(x) \leq 0,$ es inactiva. (Las restricciones están activas en todos los puntos posibles.) Decimos que una restricción es redundante si la eliminación no cambia el conjunto factible.

Ejemplo 1.17: Se ilustran estas definiciones con unos simples problemas de optimización sin restricciones con variables $x \in R$ y $\operatorname{dom} f_0 = \mathbb{R}_{++}$.

- $f_0(x) = {}^1\!/_x : p^* = 0,$ pero no se consigue el valor óptimo.

- $f_0(x) = -\log x : p^* = -\infty,$ por lo que este problema no está acotado a continuación.
- $f_0(x) = x \log x : p^* = -{}^1\!/_e$, lograda en el punto óptimo (única)$x^* = {}^1\!/_e$.

2.2 Óptimos locales y globales

Una propiedad fundamental de los problemas de optimización convexa es que cualquier óptimo localmente también es óptimo global. Para ver esto, supongamos que x es localmente óptima para un problema de optimización convexa, es decir, x es factible y $f_0(x) = \inf\ \{f_0(z)\ /\ z\ \ \text{factible}\ \ , \|z - x\|_2 \leq r\}$ (21)

para algunos r >0. Ahora supongamos que x no es óptima global, es decir, hay un posible y tal que $f_0(y)$< $f_0(x)$. Evidentemente $\|z - x\|_2 \leq r$, ya que de lo contrario $f_0(x) \leq\ f_0(y)$.

Considere el punto z dada por

$$z = (1-\theta)x + \theta y, \qquad \theta = \frac{r}{\|z - x\|_2}$$

Entonces tenemos $\|z - x\|_2 = {}^{\mathrm{r}}\!/_{2} < \mathrm{r}$, y por convexidad del conjunto factible, z es factible. Por convexidad de f_0 tenemos

$$f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y) < f_0(x)$$

lo que contradice (21). Por lo tanto no existe y factible con $f_0(y) < f_0(x)$, es decir, x es óptima global.

2.3 Expresando el problema en su forma estándar

Nos referimos a (20) como un problema de optimización en forma estándar. El problema en su forma estándar, asumimos por convención que el lado derecho de las restricciones de desigualdades y de igualdades son cero. Esto siempre se puede arreglar, restando cualquier cero lado derecho: representamos la igualdad restricción $g_i(x) = \tilde{g}_i(x)$ por ejemplo, como $h_i(x) = 0$ h, donde

$h_i(x) = g_i(x) - \widetilde{g_i}(x)$. De la misma manera que expresamos las desigualdades de la forma $f_i(x) \geq 0$ como $-f_i(x) \leq 0$.

Ejemplo 2.1 Restricciones encajonadas. Considere el problema de optimización

$$\begin{array}{ll} \text{Minimizar} & f_0(x) \\ \text{Sujeto a} & l_i \leq x_i \leq v_i, \quad i = 1, \ . \ . \ ., n \end{array}$$

donde $x \in \mathrm{R}^n$. Las restricciones se denominan variables acotadas (ya que cada variable x_i tiene límites inferior y superior para cada x_i) o las restricciones de caja (puesto que el conjunto factible es una caja).

Podemos expresar este problema en forma estándar como

$$\begin{array}{ll} \text{Minimizar} & f_0(x) \\ \text{Sujeto a} & l_i - x_i \leq 0, \quad i = 1, \ . \ . \ ., n \\ & x_i - v_i \leq 0, \quad i = 1, \ . \ . \ ., n \end{array}$$

Hay $2n$ funciones de restricciones de desigualdad:

$f_i(x) = l_i - x_i, \quad i = 1, \ . \ . \ ., n$ y $f_i(x) = x_i - v_i, \quad i = n + 1, \ . \ . \ ., 2n$

2.3.1 Problemas de maximización

Nos concentramos en el problema de minimización para (20). Podemos resolver el problema de maximización:

$$\begin{array}{ll} \text{Maximizar} & f_0(x) \\ \text{Sujeto a} & f_i(x) \leq 0, \quad i = 1, . \ . \ ., m \\ & h_j(x) = 0, \quad j = 1, . \ . \ ., p \end{array} \tag{22}$$

mínimizando la función $-f_0(x)$ sujeto a las restricciones dadas. Por esta correspondencia podemos definir todos los términos anteriormente para el problema de maximización (22). Por ejemplo el valor óptimo de (22) se define como

$$p^* = \sup \ \{f_0(x) \ / \ f_i(x) \leq 0, \quad i = 1,. \ . \ ., m, \ h_i(x) = 0, \qquad i = 1,. \ . \ ., p \}$$

y un punto x factible es ϵ-subóptima si $f_0(x) \geq p^* - \epsilon$. En un problema de maximización, la función objetivo considerada es a veces llamada el nivel de utilidad o satisfacción en lugar del costo.

2.4 Los problemas equivalentes

Utilizaremos la noción de equivalencia de los problemas de optimización en una manera informal. Llamamos a dos problemas equivalentes, si a partir de la solución de uno, la solución del otro se encuentra fácilmente, y viceversa. (Es posible, pero complicada, dar una definición formal de la equivalencia).

Como un simple ejemplo, consideremos el problema

$$\begin{aligned} \text{Minimizar} \quad & \tilde{f}(x) = \alpha_0 \ f_0(x) \\ \text{Sujeto a} \quad & \tilde{f_i}(x) = \alpha_i \ f_i(x) \leq 0, \quad i = 1, . \ . \ ., m \\ & \tilde{h_j}(x) = \beta_j \ h_j(x) = 0, \quad j = 1, . \ . \ ., p \end{aligned} \qquad (23)$$

donde $\alpha_i \ > 0, \quad i = 0, . \ . \ ., m$ y $\beta_j \ \neq 0, \quad j = 1, . \ . \ ., p$. Este problema se obtiene a partir del problema de forma estándar (20), la ampliación de la restricción objetiva y la función desigualdad son constantes positivas, y la ampliación de las funciones de restricciones de igualdad por constantes distintos de cero. Como resultado de ello, los conjuntos factibles del problema (23) y el problema (20) son idénticos. Un punto x es óptima para el problema (20) si y sólo si es óptima para el problema a escala (23), por lo que se dicen que los dos problemas son equivalentes. Los dos problemas (20) y (23) no son los mismos, ya que el objetivo de restricción y funciones difieren. Ahora describiremos algunas transformaciones generales que producen problemas equivalentes.

2.4.1 Cambio de variables

Supongamos que $\emptyset: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es uno-a-uno, con la imagen que cubre el dominio del problema $\mathcal{D}$, es decir, $\emptyset(\mathrm{dom}\,\emptyset) \supseteq \mathcal{D}$. Se definen las funciones $\widetilde{f_\iota}$ y $\widetilde{h_\iota}$ como

$$\widetilde{f_\iota}(x) = f_i(\emptyset(z)), \quad i = 0, \quad . \, . \quad ., m, \qquad \widetilde{h_\iota}(x) = h_i(\emptyset(z)), \quad i = 1, \quad . \, . \quad ., p$$

Consideremos ahora el problema

$$\begin{array}{ll} \text{Minimizar} & \widetilde{f_\iota}(z) \\ \text{Sujeto a} & \widetilde{f_\iota}(z) \leq 0, \quad i = 1, . \, . \, ., m \\ & \widetilde{h_\iota}(z) = 0, \quad i = 1, . \, . \, ., p \end{array} \tag{24}$$

con la variable z . Nosotros decimos que el problema en forma estándar (20) y el problema (24) están relacionados por el cambio de variable o sustitución de la variable $x = \emptyset(z)$.

Los dos problemas son claramente equivalente: si x resuelve el problema (20), a continuación, $z = \emptyset^{-1}(x)$ resuelve el problema (24); si z resuelve el problema (25), entonces $x = \emptyset(z)$ resuelve el problema (20).

2.4.2 Transformación de la función objetiva y las funciones de restricción

Supongamos que $\psi_0: \mathbb{R} \longrightarrow \mathbb{R}$ es monótona creciente,

$\psi_1, . . ., \psi_m: \mathbb{R} \longrightarrow \mathbb{R}$ satisface

$\psi_i(u) \leq 0$ si y sólo si $u \leq 0$, y $\psi_{m+1}, . . ., \psi_{m+p}: \mathbb{R} \longrightarrow \mathbb{R}$ satisfacer $\psi_i(u) = 0$ si y sólo si $u = 0$. Definimos las funciones $\widetilde{f_\iota}$ y $\widetilde{h_\iota}$ como las composiciones

$$\widetilde{f_\iota}(x) = \psi_i(\, f_i(x)), \quad i = 0, \quad . \quad . m, \qquad \widetilde{h_\iota}(x) = \psi_{m+i}(h_i(z)), \quad i = 1, \quad . . \quad ., p$$

Evidentemente, el problema asociada

$$\begin{array}{ll} \text{Minimizar} & \widetilde{f_0}(x) \\ \text{Sujeto a} & \widetilde{f_\iota}(x) \leq 0, \quad i = 1, . . ., m \\ & \widetilde{h_\iota}(x) = 0, \quad i = 1, . . ., p \end{array}$$

y el problema de forma estándar (20) son equivalentes y, de hecho, los grupos son factibles idénticas, y los puntos óptimos son idénticos.

Ejemplo 2.2: Problemas de mínimos cuadrados y norma minima. Como un ejemplo sencillo. Considerar el problema de minimización de la norma euclidiana sin restricciones

$$\text{Minimizar} \qquad \|Ax - b\|_2 \tag{25}$$

con la variable $x \in \mathbb{R}^n$. Dado que la norma es siempre no negativa, podemos tan bien resolver el problema

$$\text{Minimizar} \qquad \|Ax - b\|_2^2 = (Ax - b)^T(Ax - b) \tag{26}$$

en el que se minimiza el cuadrado de la norma euclidiana. Los problemas (25) y (26) son claramente equivalentes; los puntos óptimos son los mismos. Los dos problemas no son lo mismos. Por ejemplo, el objetivo de (25) no es diferenciable en ningún x con $Ax - b = 0$, mientras que el objetivo en (26) es diferenciable para todo x (de hecho, cuadrática).

2.4.3 Variables de holgura

Una simple transformación se basa en la observación de que $f_i(x) \leq 0$ si y sólo si hay un $s_i \geq 0$ que satisface $f_i(x) + s_i = 0$. Usando esta transformación se obtiene el problema

$$\begin{array}{ll} \text{Minimizar} & f_0(x) \\ \text{Sujeto a} & f_i(x) + s_i = 0, \quad i = 1, . . . , m \\ & h_i(x) = 0, \quad i = 1, . . . , p \\ & s_i \geq 0, \quad i = 1, . . . , m \end{array} \tag{27}$$

donde las variables son $x \in \mathbb{R}^n$ y $s \in \mathbb{R}^m$. Este problema tiene $n + m$ variables, m restricciones de desigualdad (las restricciones de no negatividad en sí), y $m + p$ restricción de igualdad. El nuevo s_i variable se llama la variable de holgura asociada con la desigualdad original restricción $f_i(x) \leq 0$. Introduciendo variables de holgura reemplaza cada restricción de desigualdad con una restricción de igualdad, y una restricción de no negatividad. El problema (28) es equivalente al problema forma estándar original (20). En efecto, si (x, s) es factible para el problema (27), entonces x es factible para el original problema, ya que si $-f_i(x) \geq 0$. Por el contrario, si x es factible para el problema original, entonces (x, s) es factible para el problema (27), donde se toma si $-f_i(x)$. Del mismo modo, x es óptima para el problema original (20) si y sólo si (x, s) es óptima para la problema (2.8), donde $s_i = -f_i(x)$.

2.4.4 La Eliminación de restricciones de igualdad

Si podemos parametrizar explícitamente todas las soluciones de las restricciones de igualdad

$$h_i(x) = 0, \qquad i = 1,. \ . \ ., p \tag{28}$$

utilizando algún parámetro $z \in \mathbb{R}^k$, entonces podemos eliminar las restricciones de igualdad desde el problema, de la siguiente manera. Supongamos que la función $\emptyset: \mathbb{R}^k \longrightarrow \mathbb{R}^n$ es tal que x satisface (28) si y sólo si hay algún $z \in \mathbb{R}^k$ tal que $x = \emptyset\ (z)$. El problema de optimización a continuación,

$$\text{Minimizar} \quad \widetilde{f_0}(z) = f_0(\emptyset(z)$$

$$\text{Sujeto a} \quad \widetilde{f_\iota}(x) = f_i(\emptyset(z)) \leq 0, \quad i = 1, . \ . \ ., m$$

es equivalente al problema original (20). Este problema transformado tiene variable $z \in \mathbb{R}^k, m$ restricciones de desigualdad, y sin restricciones de igualdad. Si

z es óptima para el problema transformado, entonces $x = \emptyset\,(z)$ es óptima para el problema original. Por el contrario, si x es óptima para el problema original, a continuación, (puesto que x es factible) hay al menos un Z tal que $x = \emptyset\,(z)$. Cualquier z es óptima para la transformado problema.

2.4.5 La eliminación de restricciones de igualdad lineales

El proceso de eliminación de las variables puede ser descrito de manera más explícita, y fácilmente se lleva a cabo numéricamente, cuando las restricciones de igualdad son lineales, es decir, tiene la forma $Ax = b$. Si $Ax = b$ es inconsistente, es decir,$b \notin \mathbb{R}\,(A)$, entonces el problema original es inviable. Asumiendo que esto no es el caso, sea x_0 denotar cualquier solución de la igualdad limitaciones. Sea $F \in \mathbb{R}^{n\times k}$ cualquier matriz con $\mathbb{R}\,(F) = \mathcal{N}(A)$, por lo general, la solución de las ecuaciones lineales $Ax = b$ está dada por $Fz + x_0$, donde $z \in \mathbb{R}^k$. (Nosotros puede elegir F ser de rango completo, en cuyo caso tenemos $k = n - \operatorname{rank} A$) Sustituyendo $x = Fz + x_0$ en el problema original produce el problema

$$\begin{array}{ll} \text{Minimizar} & f_0(Fz + x_0) \\ \text{Sujeto a} & f_i(Fz + x_0) \leq 0, \quad i = 1, . . . , m \end{array}$$

con la variable z , que es equivalente al problema original, no tiene restricciones de igualdad, $\operatorname{rank} A$ y menos variables.

2.4.6 Presentación de restricciones de igualdad

También podemos introducir restricciones de igualdad y nuevas variables en un problema. En lugar de describir el caso general, que es complicado y no es muy esclarecedor, damos un ejemplo típico que será útil más adelante. Considere el problema

$$\begin{aligned} \text{Minimizar} \quad & f_0(A_0x+b_0) \\ \text{Sujeto a} \quad & f_i(A_ix+b_i) \leq 0, \quad i = 1,. . .,m \\ & h_i(x) = 0, \quad i = 1,. . .,p \end{aligned}$$

donde $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{k_i \times n}$ y $f_i: \mathbb{R}^{k_i} \longrightarrow \mathbb{R}$. En este problema el objetivo y funciones de restricción se dan como composiciones de las funciones f_i con transformaciones afín definidas por $A_ix + b_i$.Introducimos las variables nuevas $y_i \in \mathbb{R}^{k_i}$, así como nuevas restricciones de igualdad $y_i = A_ix + b_i$ para $i = 0,. . .,m$, y la forma del problema equivalente

$$\begin{aligned} \text{Minimizar} \quad & f_0(y_0) \\ \text{Sujeto a} \quad & f_i(y_i) \leq 0, \quad i = 1,. . .,m \\ & y_i = A_ix+b_i, \quad i = 0,. . .,m \\ & h_i(x) = 0, \quad i = 1,. . .,p \end{aligned}$$

este problema tiene $k_0 + . . . +k_m$ nuevas variables, $y_0 \in \mathbb{R}^{k_0},. . .,y_m \in \mathbb{R}^{k_m}$ y $k_0 + . . . +k_m$ nuevas restricciones de igualdad,

$$y_0 = A_0x+b_0, \quad . . . \quad .,y_m = A_mx+b_m$$

las limitaciones objetivas y la desigualdad en este problema son independientes, es decir, implican diferentes variables de optimización.

2.4.7 Optimización sobre algunas variables

Siempre tenemos:

$$\inf_{x,y} f(x,y) = \inf_{x,y} \tilde{f}\,(x)$$

donde $\widetilde{f_0}(x_1) = \inf_y f(x,y)$. En otras palabras, siempre podemos minimizar una función minimizada primero sobre algunas de las variables, y a continuación, reducir al mínimo el restante queridos. Este principio simple y general se puede utilizar para transformar los problemas en formas equivalentes. El caso general es engorroso para describir y no de iluminación, así que en su lugar describimos un ejemplo. Supongamos que la variable $x \in \mathbb{R}^n$ se reparte como $x = (x_1, x_2)$, con $x_1 \in \mathrm{R}^{n_1}$, $x_2 \in \mathrm{R}^{n_2}$ y $n_1 + n_2 = n$. Consideramos el problema

$$\begin{array}{ll} \text{Minimizar} & f_0 = (x_1, x_2) \\ \text{Sujeto a} & f_i(x_i) \leq 0, \quad i = 1,. \; . \; ., m_1 \\ & \widetilde{f_i}\,(x_2) \leq 0, \quad i = 1,. \; . \; ., m_2 \end{array} \tag{29}$$

en el que las restricciones son independientes, en el sentido de que cada función de restricción depende de x_1 o x_2,. En primer lugar, minimizar el exceso x_2 . Definir la función $\widetilde{f_0}$ de x_1 por

$$\widetilde{f_0}(x_1) = \inf \left\{ f_0(x_1, z) \, / \, \widetilde{f_i}(z) \leq 0, \qquad i = 1,. \; . \; ., m_2 \right\}$$

El problema (29) es entonces equivalente a

Minimizar $\widetilde{f_0}(x_1)$

Sujeto a $f_i(x_i) \leq 0, \quad i = 1,. \ . \ ., m_1$ (30)

Ejemplo 2.3: Minimización de una función cuadrática con restricciones sobre algunas variables. Considere la posibilidad de un problema con el objetivo cuadrático estrictamente convexo, con algunas de las variables sin restricciones:

Minimizar $x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2$

Sujetos a $f_i(x_1) \leq 0, \quad i = 1,. \ . \ ., m$

Donde P_{11} y P_{22} son simétricas. Aquí podemos minimizar analíticamente sobre x_2:

$$\inf_{x,y} \left(x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2\right) = x_1^T \left(P_{11} - P_{12} P_{22}^{-1} P_{12}^T\right) x_1$$

2.4.8 Forma del problema epígrafo

La forma epígrafo del problema del (20) es la siguiente

Minimizar t

Sujeto a $f_0(x) - t \leq 0$

$f_i(x) \leq 0, \quad i = 1,. \ . \ ., m$ (31)

$h_i(x) = 0,. \ . \ ., p$

con las variables $x \in \mathbb{R}^n$ y $t \in \mathbb{R}$. Podemos ver fácilmente que es equivalente a la problema original: (x, t) es óptima para (31) si y sólo si x es óptima para (20) y $t = f_0(x)$. Tenga en cuenta que la función objetivo del problema es una forma epígrafe función lineal de las variables x, t. El problema forma epígrafo (31) puede interpretarse geométricamente como un problema de optimización en el "espacio gráfico (x, t) : minimizamos t en el epígrafe de f_0, sujeto a las limitaciones de la x . (Esto se ilustra en la figura 19).

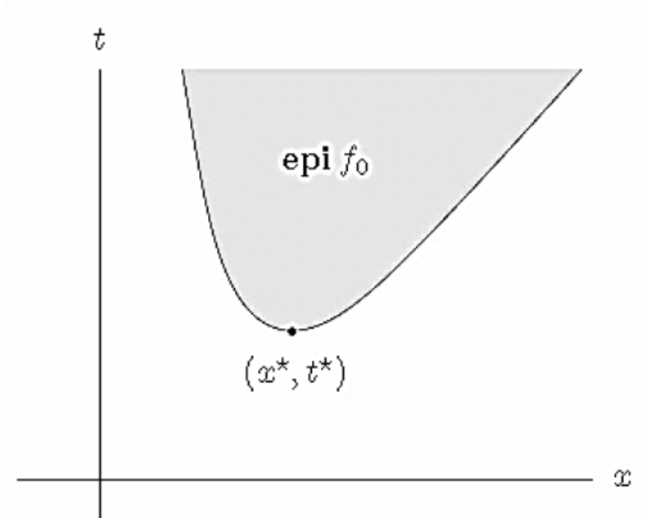


Figura 19

Interpretación geométrica del problema forma epígrafo, por un problema sin restricciones. El problema es encontrar el punto en el epígrafe (se muestra sombreado) que minimiza t, es decir, el punto 'más bajo' en el epígrafo. El punto óptimo es $(x^* t^*)$.

2.5 Problemas de optimización convexa

Un problema de optimización convexo es una de la forma:

$$
\begin{aligned}
&\text{Minimizar} && f_0(x) \\
&\text{Sujeto a} && f_i(x) \leq 0, \quad i = 1, . . . , m \qquad (32) \\
& && a_i^T x = b_i, \quad i = 1, . . . , p
\end{aligned}
$$

donde $f_0, . . . , f_m$, son funciones convexas. Comparando (32) con la norma general del problema de la forma (20), el problema convexo tiene tres requisitos adicionales:

- la función objetivo debe ser convexa,
- las funciones de restricción de desigualdad deben ser convexa,
- la igualdad restricción funciones $h_i(x) = a_i^T x - b_i$ debe ser afín.

Inmediatamente observamos una propiedad importante: El conjunto factible de problema optimización convexa es convexa, ya que es la intersección del dominio del problema:

$$\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i$$

que es un conjunto convexo, con m (convexos) conjuntos de subnivel $\{x \;/\; f_i(x) \leq 0\}$ y p hiperplanos $\{x \;/\; a_i^T x = b_i\}$. (Podemos suponer sin pérdida de generalidad que $a_i \neq 0$: si $a_i = 0$ y $b_i = 0$ para algún i, entonces la restricción de igualdad se pueden eliminar, si $a_i = 0$ y $b_i \neq 0$, la restricción de igualdad es inconsistente, y el problema es inviable.)

Por lo tanto, en un problema de optimización convexa, se minimiza un objetivo convexo la función sobre un conjunto convexo.

Si f_0 es cuasi-convexo vez de convexos, decimos que el problema (32) es un (estándar formulario) problema de optimización cuasi-convexo. Dado que los conjuntos de subnivel de un convexo o función cuasi-convexo son convexos, llegamos a la conclusión de que para un convexo o cuasi-convexo de problema optimización de los conjuntos de ϵ-subóptimas son convexos. En particular, el óptimo conjunto es convexo. Si el objetivo es estrictamente convexo, a continuación, el conjunto óptimo contiene por lo más un punto.

2.6 Un criterio de optimalidad para la función objetivo

Supongamos que la función objetivo f_0 de un problema de optimización convexa es diferenciable, de modo que para todo $x, y \in f_0$

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T (x - y) \tag{33}$$

sea X el conjunto factible, es decir,

$$X = \{x \;/\; f_i(x) \leq 0, i = 1,. \;. \;., m, h_i(x) = 0, \qquad i = 1,. \;. \;., p\}$$

entonces x es óptima si y sólo si $x \in X$ y

$$\nabla f_0(x)^T (y - x) \geq 0 \qquad \text{para todo} \qquad y \in X \tag{34}$$

este criterio de optimalizar puede ser entendido geométricamente: Si $\nabla f_0(x) \neq 0$, significa que $-\nabla f_0(x)$ define un hiperplano de soporte al conjunto factible en x

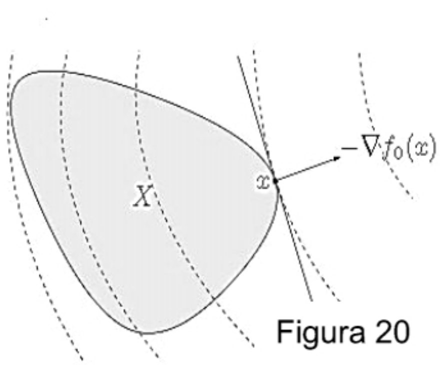


Figura 20

Interpretación geométrica de la condición de optimalizar. El conjunto factible X se muestra sombreado. Algunas curvas de nivel de f_0 se muestran como líneas discontinua. El punto x es óptima: $-\nabla f_0(x)$ define un hiperplano de soporte (que se muestra como una línea continua) para X en x.

Prueba de la condición de optimalidad

Primero supongamos $x \in X$ y satisface (34). Entonces, si $y \in X$ que tenemos, por (33), $f_0(y) \geq f_0(x)$. Esto demuestra x es un punto óptimo para (20).

A la inversa, supongamos que x es óptima, pero la condición (34) no se cumple, es decir, para algunos $y \in X$ tenemos

$$\nabla f_0(x)^T(y - x) < 0$$

considere el punto $z(t) = ty + (1 - t)x$, donde $t \in [0,1]x$ es un parámetro. Puesto que $z(t)$ es en el segmento de línea entre x y y, y el conjunto factible es convexa, $z(t)$ es factible.

Afirmamos que los pequeños t positivo tenemos $f_0(z(t)) < f_0(x)$, que demostrará que t no es óptima. Para mostrar esto, tenga en cuenta que

$$\frac{d}{dt} f_0(z(t)) \mid_{t=0} = \nabla f_0(x)^T(y - x) < 0$$

por lo que para pequeñas t positivos, que tenemos $f_0(z(t)) < f_0(x)$

2.6.1 Problemas con restricciones de igualdad sólo

Considere el caso donde hay restricciones de igualdad, pero sin restricciones de desigualdad, es decir,

$$\begin{array}{ll} \text{Minimizar} & f_0(x) \\ \text{Sujeto a} & Ax = b \end{array}$$

aquí el conjunto factible es afín. Suponemos que es no vacío, de lo contrario el problema no es factible. La condición de optimalizar para un x que es factible

$$\nabla f_0(x)^T(y-x) \geq 0$$

debe mantener para todo y satisfactoria $Ay = b$. Puesto que x es factible, cada y viable tiene la forma $y = x + v$ para algunos $v \in \mathcal{N}(A)$. La condición de optimalidad puede ser por lo tanto expresado como:

$$\nabla f_0(x)^T v \geq 0 \text{ para todo } v \in \mathcal{N}(A)$$

si una función lineal es no negativa en un sub-espacio, entonces debe ser cero en la sub-espacio, lo que se deduce que $\nabla f_0(x)^T v = 0$ para todo $v \in \mathcal{N}(A)$.

En otras palabras,

$$\nabla f_0(x)^T v = 0 \perp \mathcal{N}(A)$$

usando el hecho de que la $\mathcal{N}(A)^{\perp} = \mathcal{R}(A^T)$, esta condición de optimalidad puede ser expresada como $\nabla f_0(x) \in \mathcal{R}(A^T)$, es decir, existe una $v \in \mathbb{R}^p$, tal que

$$\nabla f_0(x) + A^T v = 0$$

junto con el requisito de $Ax = b$ (es decir, que x es factible).

2.7 Optimización cuasi-convexa

Recordemos que un problema de optimización cuasi-convexa tiene la forma estándar

$$\begin{array}{lll} \text{Minimizar} & f_0(x) & \\ \text{Sujeto a} & f_i(x) \leq 0, \quad i = 1,. \; . \; ., m & (35) \\ & Ax = b & \end{array}$$

donde la desigualdad de restricción de funciones $f_0,. \; . \; ., f_m$ son convexas, y el objetivo f_0 es cuasi-convexa (en lugar de convexa, como en un problema de optimización convexa). (Cuasi-convexa funciones de restricción pueden ser reemplazados con equivalente restricción convexa funciones, es decir, funciones de restricción que son convexas y tienen el mismo 0-subnivel establecida

En esta sección destacamos algunas diferencias básicas entre problemas de optimización convexa y cuasi-convexa, y también muestran cómo resolver una optimización cuasi-convexa problema se puede reducir a la solución de una secuencia de problemas de optimización convexa.

2.7.1 Soluciones localmente optimas y las condiciones de optimalizar

La diferencia más importante entre optimización convexa y cuasi-convexa es que un problema de optimización cuasi-convexa puede tener soluciones localmente óptimas que los no son (a nivel global) óptima. Este fenómeno puede ser visto incluso en el caso sencillo sin restricciones de la minimización de una función cuasi-convexa en$\mathbb{R}$, como el que se muestra en la figura 21.

Sin embargo, una variación de la condición de optimalizar (34) En el punto 6 hace sostenga por problemas de optimización cuasi-convexa con función objetivo diferenciable.

Sea X el conjunto factible para el problema de optimización cuasi-convexa (35). Lo se deduce de la condición de $f(y) \leq f(x) \Longrightarrow \nabla f(x)^T(y-x) \leq 0$ que x es óptima si

$$x \in X, \qquad \nabla f(x)^T(y-x) > 0 \ \text{ para todo } y \in X \,/\, \{x\} \tag{36}$$

hay dos diferencias importantes entre este criterio y el análogo uno (34) para la optimización convexa:

- La condición (36) sólo es suficiente para la optimalizar ejemplos sencillos muestran que no tiene que mantener por un punto óptimo. En contraste, la condición (34) es necesaria y suficiente para x para resolver el problema convexa.
- La condición (35) requiere que el gradiente de f_0 sea distinto de cero, mientras que la condición (34) no lo hace. En efecto, cuando $\nabla f_0(x) = 0$ en el caso convexo, la condición (34 se satisface, y x es óptima.

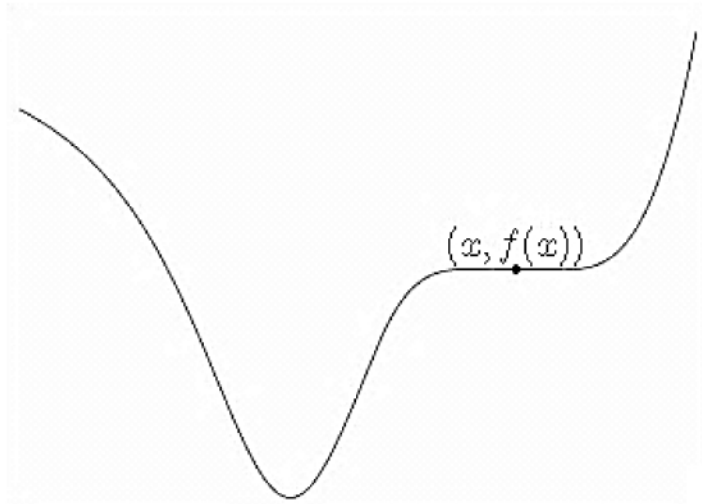


Figura 21

Una función cuasi-convexa f en $\mathbb{R}$, con un punto localmente óptima x que no es globalmente óptima. Este ejemplo muestra que la simple optimalidad condición $f'(x) = 0$, válido para funciones convexas, no se sostiene para las funciones *cuasi-convexa*.

2.7.2 Optimización cuasi-convexa a través de problemas de factibilidad convexas

Un enfoque general para optimización cuasi-convexa se basa en la representación de los conjuntos de subnivel de una función cuasi-convexa a través de una familia de las desigualdades convexas, como se describe en $f(x) \leq t \Leftrightarrow \emptyset_t(x) \leq 0$. Vamos $\emptyset_{\boldsymbol{t}}: \mathbb{R}^{\boldsymbol{n}} \longrightarrow \mathbb{R}, \boldsymbol{t} \in \mathbb{R}$, ser una familia de funciones convexas que satisface

$$f_0(x) \leq t \Leftrightarrow \emptyset_t(x) \leq 0$$

y también, para cada $x, \emptyset_t(x)$ es una función no creciente de t, es decir, $\emptyset_t(x) \leq \emptyset_t(x)$ cuando $s \geq t$.

Sea P^* denota el valor óptimo del problema de optimización cuasi-convexa (35). Si el problema de viabilidad.

Encontrar x

$$\text{Sujeto a} \quad \emptyset_t(x) \leq 0$$

$$f_i(x) \leq 0, \quad i = 1, . . . , m$$

$$Ax = b \tag{37}$$

es factible, entonces tenemos $P^* \leq t$. A la inversa, si el problema (37) no es factible, a continuación, podemos concluir $P^* \geq t$. El problema (37) es un problema de viabilidad convexa, desde las funciones de restricciones de desigualdad son convexas, y las restricciones de igualdad son lineales. Así, podemos comprobar si el valor óptimo P^* de un cuasi-convexa problema de optimización es menor o mayor que un valor dado x mediante la resolución del problema de viabilidad convexa (37). Si el problema de viabilidad convexa es factible luego tenemos $P^* \leq t$, y cualquier punto x factible es factible para el problema cuasi-convexa y satisface $f_0(x) \leq t$. Si el problema de viabilidad convexa no es factible, entonces sabemos que $P^* \geq t$.

Esta observación se puede utilizar como la base de un algoritmo sencillo para resolver la optimización cuasi-convexa del problema (35) mediante bisección, resolver un problema factibilidad convexa en cada paso. Suponemos que el problema es factible, y empezamos a con un intervalo $[l, u]$ se sabe que contiene el valor óptimo P^*. A continuación, resolvemos el problema de viabilidad convexa en su punto medio $t = (l + u) \,/\, 2$, para determinar si el valor óptimo se encuentra en la parte inferior o superior del intervalo, y actualizar el intervalo en consecuencia. Esto produce un nuevo intervalo, que también contiene el valor óptimo, pero tiene la mitad de la anchura del intervalo inicial. Esto se repite hasta que la anchura del intervalo es lo suficientemente pequeño:

Algoritmo 1 Método de bisección para la optimización cuasi-convexa.

Dado $l \leq P^*, u \leq P^*$, la tolerancia $\varepsilon > 0$.

Repetir

1. $t = (l + u) \,/\, 2$
2. Resolver el problema de viabilidad convexa (37).
3. if (38) es posible, $u := t$, más $l := t$ hasta $u - l \leq \varepsilon$.

El intervalo $[l, u]$ está garantizado para contener P^*, es decir, tenemos $l \leq P^* \leq u$ en cada paso. En cada iteración el intervalo se divide en dos, es decir, dividida en dos partes, por lo que la longitud del intervalo después de k iteraciones es $2^{-k}(u - l)$, donde $u - l$ es la longitud del intervalo inicial. De ello resulta que exactamente $\lceil \log_2((u - l)/\varepsilon) \rceil$ se requieren iteraciones antes de que termine el algoritmo. Cada paso implica la resolución de la viabilidad convexa problema (37).

2.8 Problemas de optimización lineal

Cuando la función objetiva y las restricciones son todas afines, el problema se llama una programación lineal (PL). Un programa lineal general tiene la forma.

$$\begin{aligned} \text{Minimizar} \quad & c^T x + d \\ \text{Sujeto a} \quad & Gx \leq h \\ & Ax = b \end{aligned} \tag{38}$$

donde $G \in \mathbb{R}^{m \times n}$ y $A \in \mathbb{R}^{p \times n}$. Programas lineales son, por supuesto, problemas de optimización convexa.

Es común omitir la constante d en la función objetivo, ya que no se afectar al conjunto óptimo (o posible). Dado que podemos maximizar un objetivo afín $c^T x + d$, reduciendo al mínimo $-c^T x - d$ (que todavía es convexa), que también se refiere a un problema de maximización con el objetivo afín y las funciones de restricción como un PL. (La interpretación geométrica de un PL se ilustra en la figura 22). El posible conjunto de la PL (38) es un poliedro $\mathcal{P}$; el problema es reducir al mínimo el afín función $+d$ (o, de manera equivalente, la función lineal$c^T x$) sobre $\mathcal{P}$.

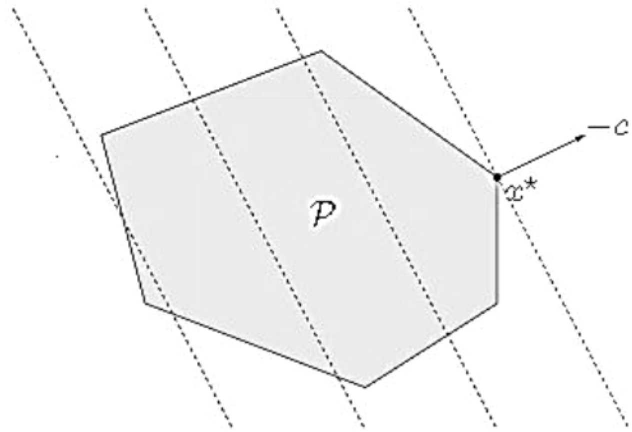


Figura 22

Interpretación geométrica de un LP. El conjunto $\mathcal{P}$ factible, que es un poliedro, es la sombra. El objetivo $c^T x$ es lineal, por lo que sus curvas de nivel hiperplanos son ortogonales a c (que se muestran como líneas discontinuas). El punto es x^* óptimo; es el punto $\mathcal{P}$ en la medida de lo posible en la dirección $-c$.

2.8.1 Estándar y la desigualdad constituyen los programas lineales

Dos casos especiales del PL (38) son tan ampliamente encontrados que han sido dado nombres distintos. En una forma estándar de PL se componen las únicas desigualdades twise no negatividad limitaciones $x \geq 0$:

$$\begin{array}{ll} \text{Minimizar} & c^T x \\ \text{Sujeto a} & Ax = b \\ & x \geq 0 \end{array} \tag{39}$$

si el PL no tiene restricciones de igualdad, se llama una forma de desigualdad PL, por lo general escrito como

$$\begin{array}{ll} \text{Minimizar} & c^T x \\ \text{Sujeto a} & Ax \leq b \end{array} \tag{40}$$

2.8.2 Convertir un PL a su forma estándar

Muchas veces es conveniente transformar un PL general (38) a uno en su forma estándar (39),en el caso de utilizar un algoritmo en su forma estándar. El primer paso es introducir variables de holgura si para las desigualdades, lo cual resulta en

$$\begin{aligned} \text{Minimizar} \quad & c^T x + d \\ \text{Sujeto a} \quad & Gx + s = h \\ & Ax = b \\ & s \geq 0 \end{aligned}$$

el segundo paso es expresar la variable x como la diferencia de dos variables no negativos las variables x^+ y x^- , es decir, $x = x^+ - x^-, x^+, x^- \geq 0$. Esto produce el problema:

$$\begin{aligned} \text{Minimizar} \quad & c^T x^+ - c^T x^- + d \\ \text{Sujeto a} \quad & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & x^+ \geq 0, \;\; x^- \geq 0, \;\; s \geq 0 \end{aligned}$$

lo cual es un PL en su forma estándar, con las variables x^+, x^-, y s.

2.9 Programación lineal fraccionada

El problema de minimización de una relación de las funciones afines más de un poliedro se llama programa lineal fraccional:

$$\begin{array}{ll} \text{Minimizar} & f_0(x) \\ \text{Sujeto a} & Gx \leq h \\ & Ax = b \end{array} \tag{41}$$

donde la función objetivo está dada por

$$f_0(x) = \frac{c^T x + d}{e^T x + f}, \qquad \mathrm{dom}\, f_0 = \{x \;/\; e^T x + f > 0\}$$

la función objetivo es cuasi-convexa (de hecho, casi lineal) para programas lineales fraccionales son problemas de optimización cuasi-convexa.

2.9.1 La transformación a un programa lineal

Si el conjunto factible

$$\{x \;/\; Gx \leq h, Ax = b, e^T x + f > 0\}$$

es no vacía, el programa lineal-fraccionada (4.32) puede ser transformado a un programa lineal equivalente:

$$\begin{array}{ll} \text{Minimizar} & c^T y + dz \\ \text{Sujeto a} & Gy - hz \leq 0 \\ & Ay - bz = 0 \\ & e^T x + fz = 1 \\ & z \geq 0 \end{array} \tag{42}$$

con variables $y, \ z$.

Para demostrar la equivalencia, observamos en primer lugar que si x es factible en (41), el par

$$y = \frac{x}{e^T x + f}, \qquad z = \frac{1}{e^T x + f},$$

es factible en (42), con el mismo valor objetivo $c^T y + dz = f_0(x)$. Resulta que el valor óptimo de (41) es mayor o igual que el valor óptimo de (42). A la inversa, si $(y, \ z)$ es factible en (42), con $z \neq 0$, entonces $x = y \,/\, z$ es factible en (41), con el mismo valor objetivo $f_0(x) = c^T y + dz$. Si $(y, \ z)$es factible en (42) con $z = 0$ y x_0 es factible (41), entonces $x = x_0 + ty$ es factible en (41) para todo $t \geq 0$.

Por otra parte, $\lim_{t \to \infty} f_0 = c^T y + dz$, para que podamos encontrar arbitrariamente puntos factibles en (41) con los valores objetivos cerca del valor objetivo de $(y, \ z)$. Llegamos a la conclusión de que el valor óptimo de (41) es menor o igual al valor óptimo de (42).

2.9.2 Generalización de una programación lineal fraccional

Una generalización del programa lineal-fraccionada (41) es la fraccional lineal generalizada programa en el que

$$f_0(x) = \max_{i=1,\dots,r} \frac{c_i^T x + d_i}{e_i^T x + f_i}, \qquad \mathrm{dom}\,\{x \ / \ e_i^T x + f_i > 0, \qquad i = 1,\dots,r\}$$

la función objetivo es la máxima puntual de las funciones r cuasi-convexa y tanto cuasi-convexa, por lo que este problema es cuasi-convexo. Cuando $r = 1$, se reduce a el programa lineal fraccional estándar.

Ejemplo 2.4: (Von Neumann problema de crecimiento). Consideraro una economía con n sectores y niveles de actividad $x_i > 0$ en el período actual y los niveles de actividad $x_i^+ > 0$ en el período siguiente. (En este problema sólo tenemos en cuenta un período.) Hay productos m que se consumen, y también producen, por la actividad: Un nivel de x actividad consume $\mathrm{B}x \in \mathbb{R}^m$, bienes produce bienes de Ax. Los bienes que se consumen en el próximo período no puede exceder de los bienes producidos en el período actual, es decir, $Bx^+ \leq Ax$. La tasa crecimiento en el sector i , en el período, está dada por x_i^+ / x_i.

El problema de crecimiento de Von Neumann es encontrar un nivel de actividad del vector x que maximiza la tasa mínima de crecimiento en todos los sectores de la economía. Este problema puede ser expresado como un problema lineal-fraccional generalizada:

$$\begin{array}{ll} \text{Maximizar} & \min_{i=1,\ldots,n} x_i^+ / x_i \\ \text{Sujeto a} & x^+ \geq 0 \\ & Bx^+ \leq Ax \end{array}$$

con dominio $\{(x, x^+) \ \mid x \succ 0\}$ hay que tener en cuenta que este problema es homogénea en x y, x^+ por lo que podemos reemplazar la restricción implícita $x > 0$ por la restricción explícitas $x \geq 1$.

2.10 Problemas de optimización cuadrática

El problema de optimización convexa (32) se llama un programa cuadrática (PC) si la función objetivo es (convexa) cuadrática, y las funciones de restricción son afín. Un programa cuadrático se puede expresar en la forma:

$$\begin{aligned} &\text{Minimizar} \quad \left(\frac{1}{2}\right) x^T P x + q^T x + r \\ &\text{Sujeto a} \quad Gx \leq h \\ &\qquad\qquad\quad Ax = b \end{aligned} \tag{43}$$

donde $P \in \mathbf{S}_+^n, G \in \mathbb{R}^{m \times \mathrm{n}}$ y $A \in \mathbb{R}$. En un programa de segundo grado, minimizamos una función cuadrática convexa sobre un poliedro, (como se ilustra en la figura 23).

Si el objetivo en (32), así como las funciones de restricción de desigualdad son (convexa) cuadrática, como en:

$$\begin{aligned} &\text{Minimizar} \quad \left(\frac{1}{2}\right) x^T P_0 x + q_0^T x + r_0 \\ &\text{Sujeto a} \quad \left(\frac{1}{2}\right) x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, . \ . \ . , m \\ &\qquad\qquad\quad Ax = b \end{aligned} \tag{44}$$

donde $P_i \in \mathbf{S}_+^n, i = 0,1 \ldots, m$ el problema se llama un programa cuadráticamente limitado (PCCL). En un PCCL, minimizamos una función cuadrática convexa sobre una región factible que es la intersección de elipsoides (cuando $P_i > 0$).

Programas cuadráticas incluyen programas lineales como un caso especial, mediante la adopción de $P = 0$ en (43). Programas cuadrática cuadráticamente restringidos incluyen programas de segundo grado (y por lo tanto también los programas lineales) como un caso especial, mediante la adopción de $P_i = 0$ en (4.5), para $i = 1, \ldots, m$

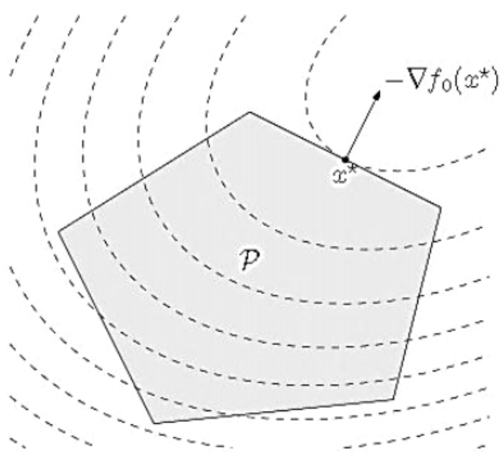


Figura 23

Ilustración geométrica de QP. El conjunto $\mathcal{P}$ factible, que es un poliedro,es mostrado a la sombra. Las curvas de nivel de la función objetivo, que es convexa cuadrática, se muestran como curvas de trazos. El x^* punto es óptimo.

2.11 Programación de cono de segundo orden

Un problema que está estrechamente relacionado con la programación cuadrática es el de segundo orden programa de cono (PCSO):

$$\begin{aligned} &\text{Minimizar} \quad f^T x \\ &\text{Sujeto a} \quad \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1,\ldots,m \qquad (45) \\ &\qquad Fx = g \end{aligned}$$

donde $x \in \mathbb{R}^{k\times n}$ es la variable de optimización, $A_i \in \mathbb{R}^{n_i\times n}$, y $F \in \mathbb{R}^{p\times n}$. Hacemos un llamado a la restricción de la forma:

$$\|Ax + b\|_2 \le c^T x + d$$

donde $A \in \mathbb{R}^{k\times n}$, una limitación del cono de segundo orden, ya que es el mismo que exigir la función afín $(Ax + b, c^T x + d)$ estar en el cono de segundo orden en $\mathbb{R}^{k+1}$.

Cuando $c_i = 0,\ i = 1,\ldots,m$, el PCSO (4.6) es equivalente a una PCCL (que se obtiene elevando al cuadrado cada una de las restricciones). Del mismo modo, si $A_i = 0,\ i = 1,\ldots,m$, entonces el PCSO (4.6) se reduce a un PL (general). Programas cono de segundo orden son, sin embargo, más general que PCSO (y por supuesto, PL).

CONCLUSIÓN

En este trabajo, hemos dado una pequeña visión de la optimización matemática, centrándonos en el caso particular, de la optimización convexa y sus conceptos fundamentales. Por lo tanto, presentamos las siguientes conclusiones a la que hemos llegado:

(1) se descubrieron numerosas aplicaciones importantes de los conjuntos convexos, principalmente en el campo de la optimización geométrica, lo que acrecentó el interés de esta teoría. Los dos capítulos están dedicados a estudiar los conceptos y resultados fundamentales de la Teoría de los Conjuntos y Optimización Convexos, así como las herramientas básicas que se utilizarán en el posterior desarrollo de la materia.

(2) Este tema de la tesis me ayudo a poder resolver problemas de optimización (programación matemática) de manera mucho más simple que con otros tipos de conjuntos.

(3) Se dificulto el desarrollo del tema ya que la mayoría de los libros son en inglés, y en la carrera desafortunadamente no se capacita con frecuencia a leer documentos técnicos en inglés. Algunos profesores lo hacen.

(4) Hablar de convexidad es hablar de interdisciplinaridad, es una noción básica en la geometría, pero también se usa ampliamente en otras áreas de las matemáticas. El uso de técnicas de convexidad aparece en muchas ramas de las matemáticas y las ciencias.

(5) Invito a los que leen la tesis a incursionar el tema, aplicándolo en otras áreas como: Análisis Funcional, Análisis Complejo, Cálculos de Variaciones, Ecuaciones Diferenciales, Matemática Discreta, Geometría Algebraica, Teoría de Probabilidad, Teoría de Códigos, Teoría de Grafos y Cristalografía, pero además encuentra aplicaciones importante en otras áreas fuera de las matemáticas como la Medicina, Economía, Física, Química, Biología, Ingeniería, Arquitectura y otras áreas del conocimiento y el pensamiento.

BIBLIOGRAFÍA

· Avriel, M. and Golany, B., Mathematical Programming for Industrial Engineers, Marcel Dekker, 1996.

· Bertsekas, D.P., Nonlinear Programming, Athena Scientific, 1995.

· Boyd, S. and Vandenberghe, L., Convex Optimization, Cambridge University Press 2004.

· Faigle, U., Kern, W. and Still, G., Algorithmic Principles of Mathematical Programming, Kluwer Academic, Dordrecht, 2002.

· Goberna, M.A: and López, M.A., Linear Semi-Infinite Optimization, Wiley, 1998.

· Hiriart-Urruty J-B. and Lemaréchal C., Convex Analysis and Minimization Algorithms I. Springer-Verlag Berlin, 1993.

· Mangasarian, O.L., Nonlinear Programming. Classics in Applied Mathematics, SIAM, 1994.

· Nash, S.G. and Sofer, A., Linear and Nonlinear Programming, McGraw-Hill, 1996.

· Nazareth, J.L., The Newton-Cauchy Framework, Springer, 1991.