Optimización Numérica

Rodrigo Mendoza-Smith
rodrigo.mendoza@itam.mx

Clase 3 y 4: Conjuntos y funciones convexas II

Conos propios y desigualdades generalizadas

- Un cono $K \subset \mathbb{R}^n$ es *propio* si :
  - K es convexo.
  - K es cerrado.
  - K es sólido (el interior es no vacío).
  - K es puntiagudo (no contiene ninguna línea)
    - Si $x \in K$ y $-x \in K \Rightarrow x = 0$.
- Un cono propio define una *desigualdad generalizada* (orden parcial en $\mathbb{R}^n$).

$$x \preceq_K y \Leftrightarrow y - x \in K.$$

- También podemos definir un orden parcial estricto:

$$x \prec_K y \Leftrightarrow y - x \in \operatorname{int} K.$$

- Si $K = \mathbb{R}_+$, $\preceq_K \equiv \leq$ y $\prec_K \equiv <$.

Ejemplos

- Ortante no-negativo:
  - $K = \mathbb{R}^n_+ \subset \mathbb{R}^n$.
  - $\preceq_K \equiv \leq$ (Desigualdad por componentes).
- El cono positivo semidefinido:
  - $K = \mathbf{S}^n_+ \subset \mathbf{S}^n$.
  - $X \preceq_K Y$ ssi $Y - X \in \mathbf{S}^n_+$.
- Polinomios no-negativos en $[0,1]$:
  - $K = \{c \in \mathbb{R}^n : \sum_{i=0}^{n-1} c_i t^i \geq 0 \ \ \forall\, t \in [0,1]\}$.
  - $c, d \in \mathbb{R}^n$ satisfacen $c \preceq_K d$ ssi

$$\sum_{i=0}^{n-1} c_i t^i \leq \sum_{i=0}^{n-1} d_i t^i \ \ \forall\ t \in [0,1].$$

Propiedades de desigualdades generalizadas

Algunas propiedades de $\preceq_K$:

- Se preserva bajo sumas

$$a \preceq_K b, c \preceq_K d, \text{ entonces } a + c \preceq_K b + d.$$

- Transitividad
- Se preserva bajo multiplicación por escalares no-negativos.
- Es reflexiva.
- Es anti-simétrica $x \preceq_K y$, $y \preceq_K x \Rightarrow x = y$.
- Se preserva bajo límites:

$$a_i \preceq_K b_i, i \in \mathbb{N} \text{ y } (a_i, b_i) \xrightarrow[i \to \infty]{} (a, b) \text{ entonces } a \preceq_K b.$$

Propiedades de desigualdades generalizadas

Algunas propiedades de $\prec_K$:

- Si $a \prec_K b$, entonces $a \preceq_K b$.
- Si $a \prec_K b$ y $c \preceq_K d$, entonces $a + c \prec_K b + d$.
- Si $a \prec_K b$ y $\alpha > 0$, entonces $\alpha a \prec_K \alpha b$.
- Si $a \prec_K b$ y $c \preceq_K d$ son suficientemente pequeños, $a + c \prec_K b + d$.

Elementos mínimos y minimales

- $(\mathbb{R}, \leq)$ genera un *orden lineal*: Siempre $x \leq y$ o $y \leq x$ si $x, y \in \mathbb{R}$.
- No es el caso para $\preceq_K$ y $\prec_K$.
- $x \in S$ es el *mínimo* de S si

$$x \preceq_K y \text{ para todo } y \in S.$$

  (es único si existe)
- $x \in S$ es un elemento *minimal* de S si

$$y \in S, y \preceq_K x \Rightarrow y = x.$$

- $x \in S$ es un elemento mínimo de S ssi $S \subset x + K$.
- $x \in S$ es un elemento minimal de S ssi $(x - K) \cap S = \{x\}$.

Elementos mínimos y minimales

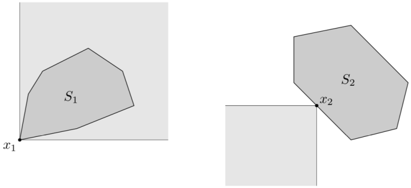


Figure 2.17 *Left.* The set S_1 has a minimum element x_1 with respect to componentwise inequality in $\mathbf{R}^2$. The set $x_1 + K$ is shaded lightly; x_1 is the minimum element of S_1 since $S_1 \subseteq x_1 + K$. *Right.* The point x_2 is a minimal point of S_2. The set $x_2 - K$ is shown lightly shaded. The point x_2 is minimal because $x_2 - K$ and S_2 intersect only at x_2.

Figure: Fuente: Boyd & Vandenberghe (Convex Optimization)

Hiperplanos separadores

Teorema (Separación de hiperplanos)

Supongamos C y D son no-vacíos y convexos. Si $C \cap D = \emptyset$, entonces existe $a \neq 0$ y b tal que $a^T x \leq b$ $\forall x \in C$ y $a^T x \geq b$ $\forall x \in D$.

- $\{x : a^T x = 0\}$ es el hiperplano separador.
- (Demostración para caso especial)
- ¿ Es cierto el converso?

Hiperplanos de soporte

Si $C \subset \mathbb{R}^n$ y $x_0 \in \text{bd}\, C$. Si $a \neq 0$ satisface $a^T x \leq a^T x_0$ para todo $x \in C$, entonces $\{x : a^T x = a^T x_0\}$ es el hiperplano de soporte a C en x_0.

Teorema (Hiperplanos de soporte)

Para todo conjunto convexo C no vacío y todo $x_0 \in \text{bd}\, C$ existe un hiperplano de soporte a C en x_0.

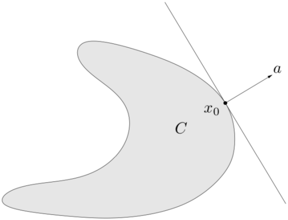


Figure 2.21 The hyperplane $\{x \mid a^T x = a^T x_0\}$ supports C at x_0.

Figure: Fuente: Boyd & Vandenberghe (Convex Optimization)

Conos duales

- Sea K un cono. El conjunto,

  $$K^* = \{y : x^T y \geq 0 \ \ \forall x \in K\}$$

  es el cono dual de K.
  - K^* es un ________.
  - K^* es siempre convexo (aún cuando K no lo es).
  - ¿Cómo se ve el cono dual?

Ejemplos

- (Subespacio) El cono dual de un subespacio $V \leq \mathbb{R}^n$ es $V^\perp$.
- (El ortante no-negativo) El cono $\mathbb{R}^n_+$ es su propio dual (es *auto-dual*).
- (El cono positivo-semidefinido) $\mathbf{S}^n_+$ es auto-dual (con $\langle X, Y \rangle = \mathrm{tr}(XY)$.

$$\mathrm{tr}(XY) \geq 0 \ \forall \ X \succeq 0 \Leftrightarrow Y \succeq 0.$$

Algunas propiedades

- K^* es convexo y cerrado.
- $K_1 \subset K_2$ implica $K_2^* \subset K_1^*$.
- Si el interior de K no es vacío, K^* es puntiagudo.
- Si la cerradura de K es puntiaguda, el interior de K^* es no vacío.
- K^{**} es la cerradura de la envoltura convexa de K.

Desigualdades generalizadas (duales)

Si K es propio, K^* es propio e induce la desigualdad generalizada dual $\preceq_{K^*}$.

- $x \preceq_K y$ ssi $\lambda^T x \leq \lambda^T y$ para todo $\lambda \succeq_{K^*} 0$.
- $x \prec_K y$ ssi $\lambda^T x < \lambda^T y$ para todo $\lambda \succeq_{K^*} 0$, $\lambda \neq 0$.

Elementos mínimos (caracterización dual)

Sea $S \subset \mathbb{R}^n$ (no necesariamente convexo), K un cono propio.

- x es el *elemento mínimo* de S c.r.a. $\preceq_K$ ssi para todo $\lambda \succ_{K^*} 0$, x es el único minimizador de $\lambda^T z$ sobre $z \in S$.
- Para todo $\lambda \succ_{K^*} 0$, $\{z : \lambda^T(z - x) = 0\}$ es un hiperplano de soporte estricto a S en x.

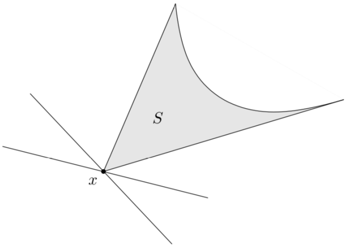
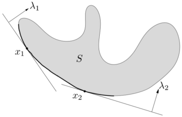


Figure: x es el mínimo de S c.r.a. $\mathbb{R}^2_{\geq 0}$ (Fuente: Boyd & Vandenberghe, Convex Optimization)

Elementos minimales (caracterización dual)

Sea $S \subset \mathbb{R}^n$ (no necesariamente convexo), K un cono propio.

- Si $\lambda \succ_{K^*} 0$ y x minimiza $\lambda^T z$ sobre $z \in S$, entonces x es minimal.
- El converso no es cierto: Un punto x puede ser minimal en S, pero no un minimizador de $\lambda^T z$ sobre $z \in S$ para todo λ.



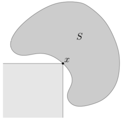


Figure 2.25 The point x is a minimal element of $S \subseteq \mathbf{R}^2$ with respect to $\mathbf{R}^2_+$. However there exists no λ for which x minimizes $\lambda^T z$ over $z \in S$.

Figure: Conjunto de elementos minimales con respecto a $\mathbb{R}^2_{\geq 0}$, y ejemplo que muestra que el converso no es cierto. (Fuente: Boyd & Vandenberghe, Convex Optimization)

Funciones convexas

Funciones convexas

Una función $f : \mathbb{R}^n \to \mathbb{R}$ es convexa si $\operatorname{dom} f$ es convexo y $\forall x, y \in \operatorname{dom} f$ y θ en $\theta \in [0, 1]$ tenemos

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y). \qquad (1)$$

- La función es estrictamente convexa si la desigualdad es estricta cuando $x \neq y$ y $\theta \in (0, 1)$.
- f es cóncava si $-f$ es convexa.
- Función es afín ssi es convexa y cóncava.
- Definimos la *extensión* de f como:

$$\hat{f}(x) = \begin{cases} f(x) & x \in \operatorname{dom}(f) \\ \infty & x \notin \operatorname{dom}(f) \end{cases}$$

Condiciones de primer orden

Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable

- f es convexa ssi $\mathrm{dom}(f)$ es convexo y
$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \tag{2}$$
$\forall x, y \in \mathrm{dom}(f)$
- (Demostración)
- La aproximación de Taylor de primer orden es un *sub-estimador* global de la función.
- Información local $(f(x), \nabla f(x)) \Rightarrow$ Información global (sub-estimador global).
- Si $\nabla f(x) = 0$, entonces __________.

Condiciones de segundo orden

Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es dos veces diferenciable ($\nabla^2 f$ existe en todo punto de $\mathrm{dom}(f)$).

- f es convexa ssi $\mathrm{dom}(f)$ es convexo y
$$\nabla^2 f \in \mathbf{S}^n_+. \tag{3}$$
- Demostración: Tarea

Ejemplos de funciones $f : \mathbb{R}^n \to \mathbb{R}$ convexas (o cóncavas) usadas en ML

- Normas: $f(x) = \|x\|_p \; p \geq 1$.
- Max: $f(x) = max_i x_i$.
- Log-sum-exp: $f(x) = \log(\sum_{i=1}^n e^{x_i})$.
  - Modelos gráficos probabilísticos.
- Media geométrica: $f(x) = (\prod_{i=1}^n x_i)^{1/n}$.
  - Aprendizaje métrico
- Log-determinante: $f(X) = \log \det(X) \; (X \in \mathbf{S}_{++}^n)$.
  - Procesos Gaussianos

El epigrafo

El epigrafo de una función $f : \mathbb{R}^n \to \mathbb{R}$ se define como:

$$\text{epi}(f) = \{(x, t) : x \in \text{dom}(f), f(x) \leq t\}. \tag{4}$$

- f es convexa ssi epi(f) es un conjunto convexo.
- El epigrafo nos permite usar resultados de convexidad de conjuntos, pero en funciones.
- Por ejemplo,

$$(y, t) \in \text{epi}(f) \Rightarrow \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^{-1} \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \tag{5}$$

e.g. El hiperplano definido por $[\nabla f(x), -1]$ soporta a epi(f) en $(x, f(x))$.

Operaciones que preservan convexidad

- Sumas escaladas por elementos no-negativos: $f = \sum_{i=1}^{n} w_i f_i$.
  - El conjunto de funciones convexas es un cono convexo!
- Composición con funciones afines: $g(x) = f(Ax + b)$.
  - Si f es convexo, g es convexo.
- Máximos puntuales: $f(x) = \max\{f_1(x), f_2(x)\}$ con $\mathrm{dom}(f) = \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$.
- Supremos puntuales: $g(x) = \sup_{y \in S} f(x, y)$ ($f(x, y)$ convexo en x).
  - $f(X) = \lambda_{max}(X)$ es convexo (X simétrica).
- Composición: $h : \mathbb{R}^k \to \mathbb{R}$ y $g : \mathbb{R}^n \to \mathbb{R}^k$, $f = g \circ h$.

  $$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x) \text{ ,si } k = n = 1.$$

  Se pueden derivar casos. El caso general se razona de manera similar.

Operaciones que preservan convexidad II

- Minimización: $g = \inf_{y \in S} f(x,y)$ si $g(x) > -\infty \forall x$.
- Si $f : \mathbb{R}^n \to \mathbb{R}$, entonces la *función de perspectiva* $g : \mathbb{R}^{n+1} \to \mathbb{R}$ de f definida por

$$g(x,t) = tf(x/t) \text{ es convexa.} \tag{6}$$

  - $f(x) = \log(x)$ es convexa en $\mathbb{R}_{>0}$. Su perspectiva es

$$g(x,t) = t\log(t/x) \text{ La entropía relativa de } t \text{ y } x. \tag{7}$$

  - La entropía relativa para vectores $x, y \in \mathbb{R}^n_{>0}$:

$$H(u \mid v) = \sum_{i=1}^{n} u_i \log(u_i/v_i). \tag{8}$$

  - Entropía relativa + Función lineal = *Divergencia de Kullback-Leibler*

$$D_{KL}(u,v) = \sum_{i=1}^{n} (u_i \log(u_i/v_i) - u_i + v_i). \tag{9}$$

Funciones conjugadas

Sea $f : \mathbb{R}^n \to \mathbb{R}$, la función conjugada $f^* : \mathbb{R}^n \to \mathbb{R}$ está dada por:

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} (y^T x - f(x)), \tag{10}$$

El conjugado tiene dominio $\mathrm{dom}(f^*) = \{y : f^*(y) < \infty\}$.

- Supremo puntual de una familia de funciones convexas (en y), entonces convexo (incluso cuando f no es convexo).
- Minimización: $g = \inf_{y \in S} f(x, y)$ si $g(x) > -\infty$ para todo x.
- Ejemplos:
  - Log-determinante (en clase)
  - La función indicadora (en conjuntos no necesariamente convexos)
  - Log-sum-exp (Tarea)
  - Normas

Propiedades

- La desigualdad de Fenchel:

$$f(x) + f^*(y) \geq x^T y. \tag{11}$$

- Si f es convexo y epi(f) es cerrado, $f^* * = f$.
- *Transformada de Legendre:* (f convexa y diff) x^* maximiza $y^T x - f(x)$ ssi $y = \nabla f(x^*)$, por lo tanto si $y = \nabla f(x^*)$,

$$f^*(y) = (x^*)^T \nabla f(x^*) - f^*(x). \tag{12}$$

- Si $a > 0$ y $b \in \mathbb{R}$, el conjugado de $g(x) = af(x) + b$ es $g^*(y) = af^*(y/a) - b$.
- Si $f(u, v) = f_1(u) + f_2(v)$, $f * (w, z) = f_1^*(w) + f_2^*(z)$.

Dualidad de Fenchel

Sean $f, g : \mathbb{R}^n \to \mathbb{R}$, f convexa, g cóncava tal que $f, -g > -\infty$. Sean

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{y^T x - f(x)\}$$
$$g_*(y) = \inf_{x \in \mathbb{R}^n} \{y^T x - g(x)\}$$

Entonces,

$$\inf_x (f(x) - g(x)) = \sup_p (g_*(p) - f^*(p)). \tag{13}$$

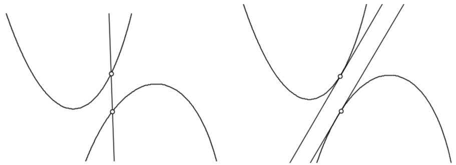


Figure: Fuente: Wikipedia

Funciones cuasi-convexas

Sean $f : \mathbb{R}^n \to \mathbb{R}$, es una función cuasi-convexa si su dominio y todos sus conjuntos de nivel $S_\alpha = \{x \in \mathrm{dom}(f) : f(x) \leq \alpha\}$ para $\alpha \in \mathbb{R}$ son convexos.

- Funciones convexas $\subset$ Funciones cuasi-convexas
- Funciones convexas $\neq$ Funciones cuasi-convexas
- f es cuasi-convexa ssi $\mathrm{dom}(f)$ es convexo y $\forall$ $x, y \in \mathrm{dom}(f)$ y $\theta \in [0, 1]$,

$$f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\}. \tag{14}$$

- *e.g*. La función $\mathrm{rank}(X)$ es cuasi-cóncava en $\mathbf{S}^n_+$. Entonces $\mathrm{rank}(X + Y) \geq \min\{\mathrm{rank}(X), \mathrm{rank}(Y)\}$.
- Si f es diferenciable, f es cuasi-convexa ssi $\mathrm{dom}(f)$ es convexo y $\forall x, y \in \mathrm{dom}(f)$, $f(y) \leq f(x) \Rightarrow \nabla f(x)^T(y - x) \leq 0$.
- Condiciones de segundo orden también existen!

Funciones log-cóncavas y log-convexas

Sean $f : \mathbb{R}^n \to \mathbb{R}$, es log-(cónvava/convexa) si $f(x) > 0$ para todo $x \in \mathrm{dom}(f)$ y $\log f$ es (cóncava/convexa).

- Log-cóncava ssi $\forall x, y \in \mathrm{dom}(f)$ y $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \geq f(x)^\theta f(y)^{1-\theta}. \qquad (15)$$

- Ejemplos:
  - La función de distribución acumulada Gaussiana $\Phi(x)$.
  - La función de distribución Gamma.
  - $\det(X)$ y $\det(X)/\operatorname{tr}(X)$en $\mathbf{S}^n_{++}$.
  - La distribución normal multivariada.
  - La distribución exponencial multivariada en $\mathbb{R}^n_+$.
  - La distribución uniforme en un conjunto convexo.
  - La distribución de Wishart (soporte en $\mathbf{S}^n_{++}$).
- Cerradura bajo suma, multiplicación, multiplicación por escalares, integración, convolución.

Monotonicidad con respecto a desigualdades generalizadas

Sea K un cono propio con desigualdad generalizada $\preceq_K$ y $f : \mathbb{R}^n \to \mathbb{R}$.

- f es K-no-decreciente si
$$x \preceq_K y \Rightarrow f(x) \leq f(y). \tag{16}$$
- f es K-creciente si
$$x \preceq_K y, x \neq y \Rightarrow f(x) < f(y). \tag{17}$$
- Si f es diferenciable y $\mathrm{dom}(f)$ es convexo, f es K-no-decreciente si
$$\nabla f(x) \succeq_{K^*} 0. \tag{18}$$
- (Demostración)

Convexidad con respecto a desigualdades generalizadas

Sea K un cono propio con desigualdad generalizada $\preceq_K$. Decimos que $f : \mathbb{R}^n \to \mathbb{R}$ es K-convexa si para todo x, y y $\theta \in [0, 1]$.

$$f(\theta x + (1-\theta)x) \preceq_K \theta f(x) + (1-\theta) f(y). \tag{19}$$

- f es *estrictamente* K-convexa si desigualdad es $\prec_K$ para todo $x \neq y$.

$$x \preceq_K y \Rightarrow f(x) \leq f(y). \tag{20}$$

- Muchos resultados de convexidad tienen extensiones a K-convexidad :)
- *Caracterización dual*: f es K-convexa ssi $\forall w \succeq_{K^*} 0$, $w^T f$ es estrictamente convexa.