Resumen de Matrices, Funciones y Optimización

Bernardita Vial

Segundo semestre de 2015

1 Matrices y concavidad de funciones

1.1 Menores principales y matriz negativa definida

- $\triangle_r$, **menores principales** de orden r de matriz $A_{n\times n}$:
  - $\triangle_r$ son los determinantes de las sub-matrices de A obtenidas al borrar $n-r$ filas y columnas de A de igual índice (submatrices de orden $r\times r$ porque se borran todas menos r).
  - ejemplo, si $n=3$ y $r=1$, un menor principal resulta de borrar las primeras dos filas y columnas; otro de borrar las dos últimas filas y columnas, y otro de borrar las primeras y terceras.
- D_r, **menores principales líderes** de orden r de $A_{n\times n}$:
  - D_r son los determinantes de las sub-matrices obtenidas al borrar **las últimas** $n-r$ filas y columnas de A de igual índice (submatrices de orden $r\times r$ porque se borran todas menos las r primeras filas y columnas).
  - en ejemplo, el menor principal líder es el que resulta de borrar las dos últimas filas y columnas.
- La matriz A es **negativa semidefinida** si $(-1)^r\triangle_r\geq 0$ y **negativa definida** si $(-1)^r D_r>0$
- La matriz A es **positiva semidefinida** si $\triangle_r(\mathbf{x})\geq 0$, y **positiva definida** si $D_r(\mathbf{x})>0$

1.2 Funciones

- Una función es una relación de X a Y en que cada elemento de X se relaciona con un único elemento de Y.
  - Si $f:X\rightarrow Y$, escribimos $y=f(x)$.
  - Veremos funciones reales $f:S\rightarrow\mathbb{R}$.
- f se dice **cóncava** si:
$$f(\lambda x+(1-\lambda)x')\geq\lambda f(x)+(1-\lambda)f(x')$$
para todo $x,x'\in S$ y $\lambda\in[0,1]$ (estricta si $>$ estricto para todo $\lambda\in(0,1)$ con $x\neq x'$).
- f se dice **convexa** si $-f$ es cóncava
- Si f es (dos veces) diferenciable, concavidad se puede expresar en términos de la matriz de segundas derivadas de f (Hesiano, que denotaremos por $H_f(\mathbf{x})$):
  - Una función f es cóncava ssi $H_f(\mathbf{x})$ es negativa semidefinida en todo su dominio (cóncava estricta si $H_f(\mathbf{x})$ negativa definida).

- Una función f es convexa ssi $H_f(\mathbf{x})$ es positiva semidefinida en todo su dominio (cóncava estricta si $H_f(\mathbf{x})$ positiva definida).

- f se dice **cuasi-cóncava** si:

$$f(\lambda x + (1-\lambda)x') \geq \min\{f(x), f(x')\}$$

para todo $x, x' \in S$ y $\lambda \in [0,1]$ (estricta si $>$ estricto para todo $\lambda \in (0,1)$ con $x \neq x'$).

- Similarmente, f se dice **cuasi-convexa** si:

$$f(\lambda x + (1-\lambda)x') \leq \max\{f(x), f(x')\}$$

para todo $x, x' \in S$ y $\lambda \in [0,1]$

- Si f es (dos veces) diferenciable, cuasi-concavidad se puede expresar en términos de la matriz horlada de segundas derivadas (que denotaremos por $\overline{H_f}(\mathbf{x})$):

$$\overline{H}_f(\mathbf{x}) = \begin{bmatrix} 0 & f_1(\mathbf{x}) & \dots & f_n(\mathbf{x}) \\ f_1(\mathbf{x}) & f_{11}(\mathbf{x}) & \dots & f_{1n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_n(\mathbf{x}) & f_{n1}(\mathbf{x}) & \dots & f_{nn}(\mathbf{x}) \end{bmatrix}$$

  - Condición necesaria para f cuasicóncava: $(-1)^r\,\overline{D}_r(\mathbf{x}) \geq 0$, donde $\overline{D}_r$ es el menor principal líder de orden r de la matriz $\overline{H}_f(\mathbf{x})$
  - Condición suficiente para f cuasicóncava: $\overline{H}_f(\mathbf{x})$ negativa definida, i.e., $(-1)^r\,\overline{D}_r(\mathbf{x}) > 0$

2 Optimización

Revisaremos condiciones necesarias y suficientes para un óptimo al analizar el problema de maximización de una función $f : \mathbb{R}^n \to \mathbb{R}$, posiblemente sujeta a restricciones expresadas a través de funciones g_j. Supondremos que f es doblemente diferenciable y que cada g_j es diferenciable.

Si $\mathbf{x}^*$ resuelve el problema de maximización de f dentro de un conjunto A, decimos que $\mathbf{x}^* = \arg\max\limits_{\mathbf{x}\in A} f(\mathbf{x})$

El valor de la función evaluada en el óptimo es $f(\mathbf{x}^*)$.

2.1 Optimización sin restricciones

Problema general:

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

(donde $\mathbf{x} = (x_1, ..., x_n)$; n es el número de variables)

- Si $\mathbf{x}^*$ óptimo local, $f(\mathbf{x})$ no puede aumentar dentro del conjunto factible. Entonces, debe ser cierto que satisface las siguientes condiciones de Primer Orden (CPO):

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_k} = 0 \text{ para todo } k = 1, ..., n$$

- $\frac{\partial f(\mathbf{x}^*)}{\partial x_k} = 0\ \forall k$ son **condiciones necesarias** para el óptimo. Una **condición suficiente** para que un $\mathbf{x}^*$ que satisface dichas condiciones sea realmente un óptimo (condición de Segundo Orden) es que f sea cóncava, esto es, $H_f(\mathbf{x})$ negativa semidefinida.

2.2 Optimización con restricciones de igualdad

Problema general:

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \text{ s.a. } m \text{ restricciones de la forma } g_j(\mathbf{x}) = b_j$$

(donde $\mathbf{x} = (x_1, ..., x_n)$; n es el número de variables y m de restricciones)

- Si $\mathbf{x}^*$ óptimo local, $f(\mathbf{x})$ no puede aumentar dentro del conjunto factible. Esto implica que $\exists\ \boldsymbol{\lambda} \in \mathbb{R}^m$ tal que

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_k} = \lambda_1 \frac{\partial g_1(\mathbf{x}^*)}{\partial x_k} + ... + \lambda_m \frac{\partial g_m(\mathbf{x}^*)}{\partial x_k} \text{ para todo } k = 1, ..., n$$

- Al escribir

$$\mathcal{L} = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j [b_j - g_j(\mathbf{x})]$$

obtenemos las siguientes **condiciones necesarias** (CPO, condiciones de Primer Orden) para el óptimo:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_k} &= 0 \text{ para todo } k = 1, ..., n \\ \frac{\partial \mathcal{L}}{\partial \lambda_j} &= 0 \text{ para todo } j = 1, ..., m \end{aligned}$$

ya que $\frac{\partial \mathcal{L}}{\partial x_k} = \frac{\partial f(\mathbf{x}^*)}{\partial x_k} - \left(\lambda_1 \frac{\partial g_1(\mathbf{x}^*)}{\partial x_k} + ... + \lambda_m \frac{\partial g_m(\mathbf{x}^*)}{\partial x_k}\right)$ y $\frac{\partial \mathcal{L}}{\partial \lambda_j} = b_j - g_j(\mathbf{x})$

- Las siguientes condiciones (de Segundo Orden) son **condiciones suficientes** para que un $\mathbf{x}^*$ que satisface dichas CPO sea realmente un óptimo: para todo $r \in \{m+1, ..., n\}$, se requiere que $(-1)^r\, \overline{D}_r(\mathbf{x}^*) > 0$, donde cada $\overline{D}_r(\mathbf{x}^*)$ es el menor principal líder de orden r del hesiano orlado definido como:

$$\overline{H}_{\mathcal{L}}(\mathbf{x}) = \begin{bmatrix} 0 & \nabla g(\mathbf{x})^T \\ \nabla g(\mathbf{x}) & H_{\mathcal{L}}(\mathbf{x}) \end{bmatrix}$$

(con $H_{\mathcal{L}}(\mathbf{x}^*)$ matriz de segundas derivadas de $\mathcal{L}$ respecto de $\mathbf{x}$, y $\nabla g(\mathbf{x})$ gradiente de g, o vector de primeras derivadas).

Esto es, los últimos $n - m$ menores principales líderes de $\overline{H}_{\mathcal{L}}(\mathbf{x}^*)$ alternan de signo, empezando con el signo de $(-1)^{m+1}$.

- Teorema de la envolvente: $\frac{\partial f(\mathbf{x}^*)}{\partial b_j} = \lambda_j$; es decir, λ_j indica cuánto aumentaría el valor máximo de la función objetivo si aumentara la constante b_j. Note que en el caso de restricciones de igualdad, λ_j podría ser positivo, negativo o cero.

(más general: si a es cualquier parámetro del problema, $\frac{\partial f(\mathbf{x}^*)}{\partial a} = \left.\frac{\partial \mathcal{L}}{\partial a}\right|_{\mathbf{x}^*, \boldsymbol{\lambda}^*}$; esto es, basta con derivar $\mathcal{L}$ y evaluar en el óptimo para encontrar $\frac{\partial f(\mathbf{x}^*)}{\partial a}$).

2.3 Optimización con restricciones de desigualdad

Problema general:

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \text{ s.a. } m \text{ restricciones de la forma } g_j(\mathbf{x}) \leq b_j$$

(donde $\mathbf{x} = (x_1, ..., x_n)$; n es el número de variables y m de restricciones)

- Si $\mathbf{x}^*$ óptimo local, $f(\mathbf{x})$ no puede aumentar dentro del conjunto factible. Esto implica que $\exists\ \boldsymbol{\lambda} \in \mathbb{R}^m_+$ tal que

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_k} = \lambda_1 \frac{\partial g_1(\mathbf{x}^*)}{\partial x_k} + ... + \lambda_m \frac{\partial g_m(\mathbf{x}^*)}{\partial x_k} \text{ para todo } k = 1, ..., n$$

- Al escribir
$$\mathcal{L}=f\left(\mathbf{x}\right)+\sum_{j=1}^{m}\lambda_j\left[b_j-g_j\left(\mathbf{x}\right)\right]$$
obtenemos las siguientes **condiciones necesarias** (condiciones de KKT, condiciones de Karush-Kuhn-Tucker) para el óptimo:
$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial x_k} &= 0 \text{ para todo } k=1,...,n\\ \frac{\partial\mathcal{L}}{\partial\lambda_j} &\geq 0 \text{ y } \frac{\partial\mathcal{L}}{\partial\lambda_j}\lambda_j=0 \text{ para todo } j=1,...,m\end{aligned}$$
ya que $\frac{\partial\mathcal{L}}{\partial x_k}=\frac{\partial f(\mathbf{x}^*)}{\partial x_k}-\left(\lambda_1\frac{\partial g_1(\mathbf{x}^*)}{\partial x_k}+...+\lambda_m\frac{\partial g_m(\mathbf{x}^*)}{\partial x_k}\right)$ y $\frac{\partial\mathcal{L}}{\partial\lambda_j}=b_j-g_j\left(\mathbf{x}\right).$
  - Si una restricción j' no es activa en el óptimo, entonces $\frac{\partial\mathcal{L}}{\partial\lambda_{j'}}>0$ (esto es, $b_{j'}>g_{j'}\left(\mathbf{x}^*\right)$) pero el teorema de la envolvente indica que $\lambda_{j'}$ debe ser cero: la restricción no afecta, por lo que aumentar $b_{j'}$ no tiene ningún efecto sobre $f\left(\mathbf{x}^*\right)$.
  - Si una restricción j'' es activa, entonces $\frac{\partial\mathcal{L}}{\partial\lambda_{j''}}=0$ (esto es, $b_{j''}=g_{j''}\left(\mathbf{x}^*\right)$) pero $\lambda_{j''}$ debe ser no-negativa: si $\lambda_{j''}$ fuera negativa, entonces al reducir $b_{j''}$ aumentaría el valor máximo de f; eso no es posible, porque todos los $\mathbf{x}$ alcanzables con el nuevo $b_{j''}$ ya eran factibles antes de reducir $b_{j''}$ (reducir el conjunto factible no puede ser beneficioso).
- Es necesario analizar 2^m casos: distintas combinaciones de λ_j positivos o nulos.
  - Por ejemplo, con $m=2$ los casos posibles son:
    * $\lambda_1=\lambda_2=0$ (caso en que ninguna restricción es activa, equivalente a maximización sin restricciones)
    * $\lambda_1=0$ y $\lambda_2>0$ (caso en que solo la segunda restricción es activa, equivalente a maximización con la segunda restricción de igualdad)
    * $\lambda_1>0$ y $\lambda_2=0$ (caso en que solo la primera restricción es activa, equivalente a maximización con la primera restricción de igualdad)
    * $\lambda_1>0$ y $\lambda_2>0$ (caso con dos restricciones activas, equivalente a maximización con dos restricciones de igualdad).
  - Algunos de estos casos podrían descartarse al obtener una contradicción; en los puntos críticos que se obtienen en los casos restantes se podrían verificar las condiciones de segundo orden correspondientes, o sencillamente comparar el valor de f que se obtiene al evaluarla en cada uno de los puntos críticos para elegir aquel que entrega $f\left(\mathbf{x}^*\right)$ más alto.
- El caso de las **restricciones de no-negatividad** $x_k\geq 0$ es un caso particular, en que la restricción se puede reescribir como $-x_j\leq 0$ para proceder como antes. Se puede mostrar que en un caso como este las condiciones de KKT son equivalentes a:
$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial x_k} &\leq 0 \text{ y } \frac{\partial\mathcal{L}}{\partial x_k}x_k \text{ para todo } k \text{ con restricción } x_k\geq 0\\ \frac{\partial\mathcal{L}}{\partial x_k} &= 0 \text{ para todo } k \text{ sin restricción de no-negatividad}\\ \frac{\partial\mathcal{L}}{\partial\lambda_j} &\geq 0 \text{ y } \frac{\partial\mathcal{L}}{\partial\lambda_j}\lambda_j=0 \text{ para todo } j=1,...,m\end{aligned}$$
Esto es, si x_k tiene restricción de no-negatividad, entonces en el óptimo puede ocurrir que $\frac{\partial\mathcal{L}}{\partial x_k}<0$ si $x_k^*=0$, pero cuando $x_k^*>0$ sigue siendo necesario que $\frac{\partial\mathcal{L}}{\partial x_k}=0$ como antes.
- Para obtener el valor de $\mathbf{x}$ que **minimiza** $f\left(\mathbf{x}\right)$ dentro de un conjunto factible A basta con resolver el problema de maximización de $-f\left(\mathbf{x}\right)$ dentro del mismo conjunto A: $\arg\min_{\mathbf{x}\in A}\left(f\left(\mathbf{x}\right)\right)=\arg\max_{\mathbf{x}\in A}\left(-f\left(\mathbf{x}\right)\right)$