Formas cuadráticas semi-definidas

G.A. Raggio

21/05/19

Formas cuadráticas reales y sus matrices correspondientes: Consideramos formas cuadráticas a valores reales definidas en $\mathbb{R}^n$ dadas por

$$q(\boldsymbol{x}) := \sum_{j,k=1}^{n} q_{jk} x_j x_k \ , \ \ \boldsymbol{x} \in \mathbb{R}^n \ ,$$

donde la matriz cuadrada ($n \times n$) $\mathbb{Q}$ que tiene q_{jk} como entrada en la fila j y la columna k, es real y simétrica ($q_{jk} = q_{kj}$). Con el producto escalar canónico $\cdot$ en $\mathbb{R}^n$,

$$q(\boldsymbol{x}) = \boldsymbol{x} \cdot (\mathbb{Q}\boldsymbol{x}) \ .$$

La forma q y la correspondiente matriz $\mathbb{Q}$ se dicen positiva semi-definidas si $q(\boldsymbol{x}) \geq 0$ para todo $\boldsymbol{x} \in \mathbb{R}^n$; y, se dicen positiva definidas si $q(\boldsymbol{x}) > 0$ para todo $\boldsymbol{x} \in \mathbb{R}^n$ no nulo. La forma cuadrática q y su matriz asociada se dicen negativa semi-definida (resp. negativa definida) si $-q$ y su matriz asociada $-\mathbb{Q}$ son postiva semi-definidas (resp. positiva definidas). La matriz $\mathbb{Q}$ es diagonalizable en los reales ya que es real y simétrica. Claramente la forma será positiva semi-definida si y sólo si los autovalores de $\mathbb{Q}$ son todos no-negativos y positiva definida si estos autovalores son todos positivos[1]

Convexidad y el Hessiano: Recordamos que una función f a valores reales definida en un sub-conjunto convexo $\mathcal{K}$ de $\mathbb{R}^n$ se dice **convexa** si para todo par $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{K}$ con $\boldsymbol{x} \neq \boldsymbol{y}$, y todo λ real con $0 < \lambda < 1$ se tiene

$$f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \ .$$

En caso de que se tenga desigualdad estricta $<$ bajo las mismas condiciones, se dice que la función f es **estrictamente convexa**.

Recordamos que el Hessiano $\mathbb{H}$ de una función f definida en un abierto de $\mathbb{R}^n$ donde es dos veces diferenciable es la matriz cuadrada ($n \times n$) dada por

$$\mathbb{H}(\boldsymbol{x}) = (h_{jk}(\boldsymbol{x}))_{j,k=1}^{n} \ , \ \ h_{jk}(\boldsymbol{x}) := \left(\frac{\partial^2 f}{\partial x_j \partial x_k} \right)(\boldsymbol{x}) \ .$$

Teorema 1 *f definida en un abierto convexo $\mathcal{K} \subset \mathbb{R}^n$ es convexa si y sólo si el Hessiano $\mathbb{H}(\boldsymbol{x})$ es positivo semi-definido para todo $\boldsymbol{x} \in \mathcal{K}$.*

Teorema 2 *Si f está definida en un abierto convexo $\mathcal{K} \subset \mathbb{R}^n$ y el Hessiano $\mathbb{H}(\boldsymbol{x})$ es positivo definido para todo $\boldsymbol{x} \in \mathcal{K}$, entonces f es estrictamente convexa.*

[1] Hay una matriz real ortogonal $\mathbb{T}$ tal que $\mathbb{T}\mathbb{Q}\mathbb{T}^t = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, donde λ_j son los autovalores de $\mathbb{Q}$ repetidos de acuerdo a sus multiplicidades. Entonces, con $\boldsymbol{y} = \mathbb{T}\boldsymbol{x}$, $q(\boldsymbol{x}) = \boldsymbol{x} \cdot \mathbb{Q}\boldsymbol{x} = \boldsymbol{y} \cdot (\mathbb{T}\mathbb{Q}\mathbb{T}^{-1}\boldsymbol{y}) = \sum_{j=1}^{n} \lambda_j y_j$.

Ejemplo: Considere $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(\boldsymbol{x}) = |\boldsymbol{x}|^4$. Tenemos $f = h \circ g$ donde $h(t) := t^2$, $0 \le t \in \mathbb{R}$, y $g(\boldsymbol{x}) := |\boldsymbol{x}|^2$, $\boldsymbol{x} \in \mathbb{R}^n$. h es creciente y estrictamente convexa; g es estrictamente convexa[2]. Entonces si $\boldsymbol{x} \neq \boldsymbol{y}$ y $0 < \lambda < 1$, se tiene $f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) = h(g(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})) < h(\lambda g(\boldsymbol{x}) + (1-\lambda)g(\boldsymbol{y})) < \lambda h(g(\boldsymbol{x})) + (1-\lambda)h(g(\boldsymbol{y})) = \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y})$ donde la primera desigualdad es consecuencia de la concavidad estricta de g y de la propiedad de crecimiento de h, y la segunda desigualdad surge de la convexidad estricta de h. Para el Hessiano de f tenemos $H_{jk}(\boldsymbol{x}) = 4|\boldsymbol{x}|^2\delta_{jk} + 8x_jx_k$. $\mathbb{H}(\boldsymbol{x})$ es positivo definido para todo $\boldsymbol{x} \neq \boldsymbol{0}$ pero $\mathbb{H}(\boldsymbol{0}) = 0$.

Criterios en términos de menores principales: Dada una matriz cuadrada $(n \times n)$ $\mathbb{A}$, sus **menores principales** son las determinantes de las matrices que se obtienen a partir de $\mathbb{A}$ tachando ciertas filas y sus correspondientes columnas. Para formalizar, dada $\mathbb{A}$, y un natural ℓ con $1 \le \ell < n$, denotamos con

$$\mathbb{A}^{[j_1,j_2,\cdots,j_\ell]}\ , \quad \text{donde } j_1, j_2, \cdots, j_\ell \text{ son } \ell \text{ naturales en } \{1,2,\cdots,n\} \text{ que satisfacen } j_1 < j_2 < \cdots j_\ell\ ,$$

a la matriz cuadrada $(n-\ell) \times (n-\ell)$ que se obtiene de $\mathbb{A}$ tachando las filas y columnas $j_1, j_2, \cdots, j_\ell$. El número

$$\det(\mathbb{A}^{[j_1,j_2,\cdots,j_\ell]})$$

es un menor principal de orden $n - \ell$. El menor principal de orden n es $\det(\mathbb{A})$. La cantidad de menores principales de orden k $(1 \le k \le n)$ es $\begin{pmatrix} n \\ n-k \end{pmatrix}$.

Ejemplo: $n = 3$, $\mathbb{A} = (a_{jk})_{j,k=1}^3$.
$\ell = 1$, menores principales de orden 2:

$$\mathbb{A}^{[1]} = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}\ ,\ \ \mathbb{A}^{[2]} = \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix}\ ,\ \ \mathbb{A}^{[3]} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\ ;$$

$\ell = 2$, menores principales de orden 1:

$$\mathbb{A}^{[1,2]} = a_{33}\ ,\ \ \mathbb{A}^{[1,3]} = a_{22}\ ,\ \ \mathbb{A}^{[2,3]} = a_{11}\ .$$

Teorema 3 *La matriz cuadrada real y simétrica $\mathbb{A}$ es positiva semi-definida si y sólo si todos sus menores principales son no-negativos.*

Teorema 4 *(Criterio de Sylvester) La matriz cuadrada $(n \times n)$ real y simétrica $\mathbb{A}$ es positiva definida si y sólo si los menores principales*

$$\det(\mathbb{A}^{[2,3,4,\cdots,n]}) = a_{11}\ ,\quad ,\ \det(\mathbb{A}^{[3,4,\cdots n]})\ ,\ \cdots\ \det(\mathbb{A}^{[n-1,n]})\ ,\ \det(\mathbb{A}^{[n]})\ ,\ \det(\mathbb{A})$$

son todos positivos.

Comentarios:

1. En el magnifico criterio de Sylvester (no hay que calcular todos los menores principales sino sólo n) los menores principales involucrados son aquellos "esquinados arriba a la izquierda[3]" construidos

[2] $\nabla g(\boldsymbol{x}) = 2\boldsymbol{x}$, $g_{x_j,x_k} = 2\delta_{j,k}$ y el Hessiano es 2 veces la identidad.
[3] en inglés se usa el término leading (o corner) principal minors

a partir del elemento a_{11} engrosando sucesivamente agregando la segunda fila y columna, luego la tercera fila y columna, etc.

$$a_{11} \ \to \ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \to \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \to \cdots .$$

No hay nada sacrosanto en la elección de a_{11}, se puede empezar con cualquier elemento diagonal. Por ejemplo en el caso $n = 3$

$$a_{22} \ \to \ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \to \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} .$$

O bien

$$a_{22} \ \to \ \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} \to \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} .$$

Cuando la matriz tiene elementos nulos puede ser que la elección del elemento diagonal inicial simplifique las determinantes a calcular.
Ejemplo:

$$\mathbb{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1/2 \end{pmatrix} .$$

$$\det(a_{11}) = 1 \ , \ \ \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = 2 \ , \ \ \det(\mathbb{A}) = 0 \ .$$

Después de 3 menores deducimos que $\mathbb{A}$ no es positiva definida. Pero

$$\det(a_{22}) = 2 \ , \ \ \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} = 0 \ , \ \ .$$

Despues de dos menores principales sabemos que $\mathbb{A}$ no es positiva definida.
Queda todavía la posibilidad de que la matriz sea positiva semi-definida. Los menores principales que faltan son:

$$a_{33} = 1/2 \ , \ \det \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix} = 1/2 \ ,$$

y, ahora si, sabemos que $\mathbb{A}$ es positiva semi-definida.

2. Hay $\sum_{\ell=0}^{n-1} \begin{pmatrix} n \\ n-k \end{pmatrix} = (1+1)^n - 1 = 2^n - 1$ menores principales y esto crece exponencialmente con n. Pero hay sólo n menores principales "esquinados". Un protocolo eficiente para decidir si un dada matriz cuadrada es positiva semi-definida empezaría entonces calculando los menores principales esquinados. Si todos estos son positivos o alguno (pero no todos) resulta ser negativo: ¡felicitaciones!, ya se sabe que la matriz es postiva definida o indefinida. Sino, si los esquinados son no-negativos, se calculan los $2^n - n - 1$ menores principales que faltan. Cuando hay menores principales nulos del Hessiano la cuestión de la convexidad estricta debe analizarse por separado.

3. La matriz $\mathbb{A}$ es negativa semi-definida (resp. definida) si $-\mathbb{A}$ es positiva semi-definida (resp. definida). Los menores de $-\mathbb{A}$ son

$$\det((-\mathbb{A})^{[j_1, j_2, \cdots, j_\ell]}) = (-1)^{n-\ell} \det(\mathbb{A}^{[j_1, j_2, \cdots, j_\ell]}) \ ,$$

y hay que corregir los signos: $\mathbb{A}$ es negativa semi-definida si los menores de orden impar son no-positivos y los de orden par son no-negativos.