

Optimización Semidefinida

Clase 04 - Matrices definidas positivas

Segundo Cuatrimestre 2021

Facultad de Ciencias Exactas y Naturales, UBA

Motivación

Consideremos una sucesión de vectores definida recursivamente por

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k], \quad \mathbf{x}[0] = \mathbf{x}_0$$

El sistema tiende a 0 para todo vector inicial si y solo todos los autovalores de \mathbf{A} tienen módulo menor que 1.

Dada una matriz \mathbf{A} con autovalores $\lambda_1, \dots, \lambda_n$, los autovalores de $\mathbf{A} - \alpha I$ son $\lambda_1 - \alpha, \dots, \lambda_n - \alpha$.

Si \mathbf{A} es simétrica, todos sus autovalores son reales, y podemos calcular su radio espectral calculando el mayor y menor autovalor.

En este caso, podemos plantear el problema de hallar el menor autovalor como un problema de optimización:

$$\min\{\lambda : \lambda \text{ autovalor de } \mathbf{A}\} = \sup_{\alpha \in \mathbb{R}}\{\alpha : \mathbf{A} - \alpha I \succeq 0\}.$$

Matrices simétricas

Notamos \mathcal{S}^n al espacio de matrices simétricas

$$\mathcal{S}^n = \{A \in \mathbb{R}^{n \times n} \mid a_{ij} = a_{ji}, \forall 1 \leq i, j \leq n\}.$$

Es un subespacio vectorial de $\mathbb{R}^{n \times n}$ de dimensión $\frac{n(n+1)}{2}$.

Proposición

Si A es simétrica, entonces

- 1 $x^* A x \in \mathbb{R}$ para todo $x \in \mathbb{C}^n$,
- 2 todos los autovalores de A son reales,
- 3 autovectores correspondientes a autovalores distintos son perpendiculares,
- 4 $S^T A S$ es simétrica para toda $S \in \mathbb{R}^{n \times n}$.

Teorema espectral para matrices simétricas

Teorema

Cualquier matriz simétrica $\mathbf{A} \in \mathcal{S}^n$ admite una descomposición

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^T,$$

con $\mathbf{U} \in \mathbb{R}^{n \times n}$ matriz ortogonal y $\mathbf{D} \in \mathbb{R}^{n \times n}$ matriz diagonal, con los autovalores $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{R}$ de \mathbf{A} en la diagonal.

Idea de la demostración. Para cualquier matriz $\mathbf{X} \in \mathbb{R}^{n \times n}$ podemos calcular la descomposición de Schur

$$\mathbf{X} = \mathbf{U} \mathbf{T} \mathbf{U}^T,$$

con \mathbf{U} matriz ortogonal y \mathbf{T} triangular superior. Si \mathbf{X} es simétrica, \mathbf{T} resulta diagonal. Ver [Horn and Johnson, 1985, Teorema 4.1.5].

Congruencia de matrices

En el caso de matrices simétricas existe una diagonalización más simple que la diagonalización por matrices semejantes, que es especialmente útil en problemas de programación semidefinida.

Definición

Dos matrices $A, B \in \mathbb{R}^{n \times n}$ son *congruentes* si existe una matrix $S \in \mathbb{R}^{n \times n}$ inversible tal que

$$A = SBS^T.$$

La relación de congruencia es una relación de equivalencia.

Signatura de una matriz

Dada una matriz simétrica $\mathbf{A} \in \mathcal{S}^n$ definimos la *signatura* de \mathbf{A} como la terna

$$(n_+, n_-, n_0),$$

donde

- n_+ la cantidad de autovalores positivos de \mathbf{A} ,
- n_- la cantidad de autovalores negativos,
- n_0 es la cantidad de autovalores nulos.

Si $\mathbf{A} \in \mathcal{S}^n$, podemos escribir $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ con $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, la matriz con los autovalores de \mathbf{A} en la diagonal, y \mathbf{U} unitaria. Suponemos que los autovalores positivos son los primeros, luego los autovalores negativos y finalmente los autovalores cero.

Ley de inercia

Definiendo la matriz diagonal no-singular real

$$\mathbf{D} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{n_+}}, \sqrt{-\lambda_{n_++1}}, \dots, \sqrt{-\lambda_{n_++n_-}}, 1, \dots, 1),$$

obtenemos

$$\mathbf{\Lambda} = \mathbf{D} \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & -1 & & & & \\ & & & & \ddots & & & \\ & & & & & -1 & & \\ & & & & & & 0 & \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{pmatrix} \mathbf{D}.$$

Ley de inercia

Podemos escribir a la matriz \mathbf{A} como

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \mathbf{S} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & -1 & & & \\ & & & & \ddots & & \\ & & & & & -1 & \\ & & & & & & 0 & \ddots \\ & & & & & & & & 0 \end{pmatrix} \mathbf{S}^T = \mathbf{S}\mathbf{I}(\mathbf{A})\mathbf{S}^T,$$

donde $\mathbf{S} = \mathbf{U}\mathbf{D}$ es una matriz no-singular e $\mathbf{I}(\mathbf{A})$ es la matriz de inercia de \mathbf{A} .

Ley de inercia

Utilizando esta escritura, la ley de inercia nos permite determinar fácilmente si dos matrices simétricas son congruentes.

Teorema

Dos matrices simétricas $A, B \in \mathcal{S}^n$ son congruentes si y solo si tienen la misma signatura.

Demostración. Si dos matrices A, B tienen la misma signatura, utilizando la construcción anterior, obtenemos que A y B son congruentes a la misma matriz de inercia. Como la relación de congruencia es una relación de equivalencia, las matrices A y B son congruentes.

Ley de inercia (continuación de la demostración)

Para la otra implicación, suponemos \mathbf{A} y \mathbf{B} congruentes, con $\mathbf{A} = \mathbf{SBS}^T$, para alguna matriz no-singular $\mathbf{S} \in \mathbb{R}^{n \times n}$. Como las matrices congruentes tienen el mismo rango, $n_0(\mathbf{A}) = n_0(\mathbf{B})$ y alcanza ver que $n_+(\mathbf{A}) = n_+(\mathbf{B})$. Notamos $k = n_+(\mathbf{A})$.

Sean $\mathbf{v}_1, \dots, \mathbf{v}_k$ autovectores ortonormales de \mathbf{A} correspondientes a los autovalores positivos $\lambda_1(\mathbf{A}), \dots, \lambda_k(\mathbf{A})$ y sea $S_+(\mathbf{A}) = \langle \mathbf{v}_1, \dots, \mathbf{v}_k \rangle$.

La dimensión de $S_+(\mathbf{A})$ es k , y si

$$\mathbf{x} = \alpha_1 \mathbf{v}_1 + \dots + \alpha_k \mathbf{v}_k \neq 0,$$

entonces

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \lambda_1(\mathbf{A}) |\alpha_1|^2 + \dots + \lambda_k(\mathbf{A}) |\alpha_k|^2 > 0.$$

Ley de inercia (continuación de la demostración)

Por lo tanto,

$$\mathbf{x}^T (\mathbf{S} \mathbf{B} \mathbf{S}^T) \mathbf{x} = (\mathbf{S}^T \mathbf{x})^T \mathbf{B} (\mathbf{S}^T \mathbf{x}) > 0,$$

luego $\mathbf{y}^T \mathbf{B} \mathbf{y} > 0$ para todo vector no nulo $\mathbf{y} \in \langle \mathbf{S}^T \mathbf{v}_1, \dots, \mathbf{S}^T \mathbf{v}_k \rangle$, que es un espacio de dimensión k . Por lo tanto, $n_+(\mathbf{B}) \geq k = n_+(\mathbf{A})$ y realizando el mismo razonamiento intercambiando \mathbf{A} por \mathbf{B} , obtenemos $n_+(\mathbf{A}) = n_+(\mathbf{B})$.

En particular, cualquier matriz simétrica es congruente a una matriz diagonal con valores 0, +1 o -1 en la diagonal, y la cantidad de cada uno de estos valores en la diagonal es invariante, no depende de la matriz \mathbf{S} .

Diagonalización por congruencia

Dada una matriz simétrica $\mathbf{A} \in \mathcal{S}^n$ podemos obtener una matriz diagonal \mathbf{D} congruente a \mathbf{A} por eliminación gaussiana (realizando simultáneamente eliminación en filas y columnas).

Por lo tanto, podemos calcular eficientemente tanto la diagonalización por congruencia como la signatura de la matriz.

Ejemplo

Calcular una matriz \mathbf{D} diagonal que sea congruente a

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 6 \\ 0 & 0 & 1 \\ 6 & 1 & 1 \end{pmatrix}.$$

Calcular los autovalores de \mathbf{A} y verificar la ley de inercia.

Matrices definidas positivas

Una matriz $\mathbf{A} \in \mathbb{R}^{n \times n}$ se llama *semidefinida positiva* si es simétrica y

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$$

para todo $\mathbf{x} \in \mathbb{R}^n$. Si además $\mathbf{x}^T \mathbf{A} \mathbf{x} = 0$ solo para $\mathbf{x} = 0$, decimos que la matriz es *definida positiva*.

Notamos $\mathbf{A} \succ 0$ a las matrices definidas positivas y $\mathbf{A} \succeq 0$ a las matrices semidefinidas positivas.

Si $\mathbf{A} = \mathbf{M}^T \mathbf{M}$ para $\mathbf{M} \in \mathbb{R}^{m \times n}$ cualquiera, entonces

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{M}^T \mathbf{M} \mathbf{x} = \|\mathbf{M} \mathbf{x}\|^2 \geq 0,$$

lo que nos da un amplio stock de matrices positivas semidefinidas.

Caracterización de matrices semidefinidas positivas

Dado un conjunto $S \subset \{1, \dots, n\}$, definimos el menor principal asociado a S como el determinante de la submatriz cuadrada $A_{S,S}$ formada por las filas y columnas de A con índices en S .

Teorema

Si $A \in \mathbb{R}^{n \times n}$ es una matriz simétrica, las siguientes propiedades son equivalentes:

- 1 A es semidefinida positiva ($x^T A x \geq 0$ para todo $x \in \mathbb{R}^n$),
- 2 todos los autovalores de A son no-negativos,
- 3 $A = M^T M$ para alguna matriz $M \in \mathbb{R}^{n \times n}$,
- 4 todos los menores principales de A son no-negativos.

Caracterización de matrices definidas positivas

Para matrices definidas positivas, obtenemos equivalencias similares.

Teorema

Si $\mathbf{A} \in \mathbb{R}^{n \times n}$ es una matriz simétrica, las siguientes propiedades son equivalentes.

- ① *\mathbf{A} es definida positiva ($\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ para todo $\mathbf{x} \in \mathbb{R}^n$),*
- ② *todos los autovalores de \mathbf{A} son positivos,*
- ③ *$\mathbf{A} = \mathbf{M}^T \mathbf{M}$ para alguna matriz $\mathbf{M} \in \mathbb{R}^{n \times n}$ no singular,*
- ④ *todos los menores principales de \mathbf{A} son positivos,*
- ⑤ *el polinomio característico de \mathbf{A} tiene signos alternados: si $\chi_{\mathbf{A}}(t) = t^n + a_{n-1}t^{n-1} + a_1t + a_0$, entonces $a_i a_{i+1} < 0$ para todo $0 \leq i \leq n-1$ (definiendo $a_n = 1$).*

Menores principales de cabeza

Si $S = \{1, 2, \dots, k\}$, decimos que el menor asociado es un *menor principal de cabeza*.

En el caso de matrices definidas positivas, podemos restringir la condición (4) a considerar solo los menores principales de cabeza.

Idea de la demostración. Usamos que para una matriz simétrica \mathbf{A} , los autovalores de

$$\hat{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & \mathbf{y} \\ \mathbf{y}^T & a \end{pmatrix},$$

con $\mathbf{y} \in \mathbb{R}^n$ y $a \in \mathbb{R}$ están entrecruzados con los autovalores de \mathbf{A} (ver [Horn and Johnson, 1985, Teorema 4.3.8]).

Factorización de Cholesky

En ambos casos, la condición (3) podemos restringirla a matrices triangulares inferiores. Recordemos que cualquier matriz $C \in \mathbb{R}^{n \times n}$ admite una descomposición QR y puede escribirse como $C = QR$ con Q unitaria y R triangular superior del mismo rango que C . Luego

$$A = C^T C = (QR)^T QR = R^T Q^T QR = R^T R$$

y tomando $L = R^T$, obtenemos $A = LL^T$.

Factorización de Cholesky

Si \mathbf{C} es no singular, podemos elegir \mathbf{R} con todos los valores en la diagonal positivos (y la descomposición de esta forma es única). Esto prueba el siguiente corolario, que nos da la *descomposición de Cholesky* de una matriz definida positiva.

Corolario

Una matriz \mathbf{A} es definida positiva si y solo si existe una matriz triangular inferior $\mathbf{L} \in \mathbb{R}^{n \times n}$ con valores positivos en la diagonal tal que $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.

Un subconjunto \mathcal{C} de un espacio vectorial real V es un *cono convexo* si es cerrado por combinaciones lineales positivas. Es decir,

- ① Si $\mathbf{x} \in \mathcal{C}$ y $a \geq 0$, entonces $a\mathbf{x} \in \mathcal{C}$.
- ② Si $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, entonces $\mathbf{x} + \mathbf{y} \in \mathcal{C}$.

En este apunte siempre que hablemos de *conos* vamos a referirnos a conos convexos.

Decimos que un cono es *puntiagudo* si

$$\boldsymbol{x} \in \mathcal{C} \text{ y } -\boldsymbol{x} \in \mathcal{C} \Rightarrow \boldsymbol{x} = \mathbf{0}.$$

El *cono dual* de un cono \mathcal{C} es

$$\mathcal{C}^* = \{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{x}^T \boldsymbol{y} \geq 0 \ \forall \boldsymbol{x} \in \mathcal{C}\}.$$

Orden parcial

Dado un cono puntiagudo \mathcal{C} en \mathbb{R}^n , podemos definir un orden parcial en \mathbb{R}^n por

$$\mathbf{x} \succeq \mathbf{y} \iff \mathbf{x} - \mathbf{y} \in \mathcal{C}$$

para $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Este orden satisface las siguientes propiedades:

- **reflexividad:** $\forall \mathbf{x} \in \mathbb{R}^n : \mathbf{x} \succeq \mathbf{x}$
- **antisimetría:** $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : \mathbf{x} \succeq \mathbf{y}, \mathbf{y} \succeq \mathbf{x} \Rightarrow \mathbf{x} = \mathbf{y}$
- **transitividad:** $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n : \mathbf{x} \succeq \mathbf{y}, \mathbf{y} \succeq \mathbf{z} \Rightarrow \mathbf{x} \succeq \mathbf{z}$
- **homogeneidad:** $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}_{\geq 0} : \mathbf{x} \succeq \mathbf{y} \Rightarrow \alpha \mathbf{x} \succeq \alpha \mathbf{y}$
- **aditividad:** $\forall \mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathbb{R}^n : \mathbf{x} \succeq \mathbf{y}, \mathbf{x}' \succeq \mathbf{y}' \Rightarrow \mathbf{x} + \mathbf{x}' \succeq \mathbf{y} + \mathbf{y}'$

Desigualdad estricta

En general, nos van a interesar conos convexos de dimensión completa, es decir conos con interior no vacío. En ese caso, podemos definir la desigualdad estricta

$$x \succ y \iff x - y \in \text{int } \mathcal{C}.$$

Tenemos el siguiente resultado de separación.

Lema

Sean $\mathcal{C} \subset \mathbb{R}^n$ un cono convexo cerrado y $x \in \mathbb{R}^n \setminus \mathcal{C}$ un punto fuera de \mathcal{C} . Existe un hiperplano que separa a $\{x\}$ de \mathcal{C} . Más aún, existe un vector no nulo $c \in \mathbb{R}^n$ tal que

$$\forall y \in \mathbb{R}^n : c \cdot y \geq 0 > c \cdot x.$$

Cono generado por un conjunto de puntos. El cono generado por un conjunto de puntos $A = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ es el menor cono que contiene a A . Es el cono

$$\text{cono } A = \left\{ \sum_{i=1}^N \alpha_i x_i \mid \alpha_1, \dots, \alpha_N \in \mathbb{R}_{\geq 0} \right\}.$$

El ortante no-negativo. El ortante no-negativo que utilizamos en programación lineal, definido por

$$\mathbb{R}_{\geq 0}^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_1, \dots, x_n \geq 0\}$$

es un cono puntiagudo convexo cerrado de dimensión completa.

Ejemplo: el cono de matrices semidefinidas positivas

Proposición

El conjunto \mathcal{S}_+^n de matrices semidefinidas positivas es un cono convexo.

Demostración. Podemos escribir a \mathcal{S}_+^n como intersección de infinitos semiespacios:



$$\mathcal{S}_+^n = \{\mathbf{A} \in \mathcal{S}^n \mid \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0 \ \forall \mathbf{x} \in \mathbb{R}^n\} = \bigcap_{\mathbf{x} \in \mathbb{R}^n} \{\mathbf{A} \in \mathcal{S}^n \mid \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0\},$$

y por lo tanto es un conjunto convexo cerrado. Para ver que forman un cono, si $\mathbf{A}, \mathbf{B} \succeq 0$ y $a \geq 0$, es claro que $a\mathbf{A} \succeq 0$ y también que

$$\mathbf{x}^T (\mathbf{A} + \mathbf{B}) \mathbf{x} = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{B} \mathbf{x} \geq 0$$

para todo $\mathbf{x} \in \mathbb{R}^n$.

Referencias

-  Bertsimas, D. and Tsitsiklis, J. (1997).
Introduction to Linear Optimization.
Athena Scientific.
-  Horn, R. and Johnson, C. (1985).
Matrix Analysis.
Cambridge University Press.