Optimización Semidefinida

Clase 11 - Distancia euclídea y rango mínimo

Segundo Cuatrimestre 2021

Facultad de Ciencias Exactas y Naturales, UBA

Distancias euclídeas

Consideramos un grafo $G = (N, E)$ con nodos $N = \{1, \ldots, n\}$, aristas $E \subset N \times N$, y pesos no negativos $D = \{d_{ij}\} \in \mathbb{R}_+^E$, que representan distancias entre los nodos.

Decimos que (G, D) es k-realizable si podemos ubicar los nodos de G en puntos $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^k$ de forma tal que las distancias euclídeas entre los nodos respeten las longitudes dadas:

$$\exists \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^k : \|\boldsymbol{v}_i - \boldsymbol{v}_j\| = d_{ij} \quad \forall \{i, j\} \in E.$$

Decimos que (G, D) es realizable (a secas) si existe k tal que (G, D) es k-realizable.

Grafo completo

Para el caso de un grafo completo tenemos la siguiente caracterización.

Teorema

Sea $G = K_n$ un grafo completo, con pesos D, y sea $\boldsymbol{A} \in \mathcal{S}^n$ la matriz

$$\boldsymbol{A} = \begin{pmatrix} 0 & d_{21}^2 & d_{31}^2 & \dots & d_{n1}^2 \\ d_{21}^2 & 0 & d_{32}^2 & \dots & d_{n2}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1}^2 & d_{n2}^2 & d_{n3}^2 & \dots & d_{nn}^2 \end{pmatrix}$$

El grafo G es realizable si y solo si $\boldsymbol{A}$ es semidefinida negativa en el espacio ortogonal al vector $\boldsymbol{e} = (1, 1, \dots, 1)$. Es decir, si

$$\boldsymbol{y}^T \boldsymbol{A} \boldsymbol{y} \leq 0 \quad \text{para todo } \boldsymbol{y} \in \mathbb{R}^n \text{ tal que } \sum_{i=1}^{n} y_i = 0.$$

Demostración

Demostramos solo $\Rightarrow$.

Suponemos que existen vectores $\boldsymbol{v}_i \in \mathbb{R}^k$ para algún $k \geq 1$ tales que $d_{ij} = \|\boldsymbol{v}_i - \boldsymbol{v}_j\|$. Consideramos ahora la matriz $\boldsymbol{X}$ de productos internos

$$\boldsymbol{X} = \begin{pmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle & \dots & \langle \boldsymbol{v}_1, \boldsymbol{v}_n \rangle \\ \langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle & \dots & \langle \boldsymbol{v}_2, \boldsymbol{v}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{v}_n, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_n, \boldsymbol{v}_2 \rangle & \dots & \langle \boldsymbol{v}_n, \boldsymbol{v}_n \rangle \end{pmatrix} = (\boldsymbol{v}_1 \dots \boldsymbol{v}_n)^T (\boldsymbol{v}_1 \dots \boldsymbol{v}_n),$$

que es semidefinida positiva. Como $a_{ij} = \|\boldsymbol{v}_i - \boldsymbol{v}_j\|^2 = \langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle + \langle \boldsymbol{v}_j, \boldsymbol{v}_j \rangle - 2\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$, tenemos

$$\boldsymbol{A} = \mathrm{diag}(\boldsymbol{X}) \cdot \boldsymbol{e}^T + \boldsymbol{e} \cdot \mathrm{diag}(\boldsymbol{X}) - 2\boldsymbol{X}.$$

Por lo tanto, para $\boldsymbol{y} \perp \boldsymbol{e}$, $\boldsymbol{y}^T \boldsymbol{A} \boldsymbol{y} = -2\boldsymbol{y}^T \boldsymbol{X} \boldsymbol{y} \leq 0$.

Grafos no completos

En general, para grafos no completos determinar si el grafo es realizable equivale a buscar una solución factible de un problema SDP.

Teorema

Un grafo con pesos (G, D) es realizable si y solo si el siguiente problema SDP tiene solución

$$
\begin{array}{ll}
\textit{existe:} & \boldsymbol{X} \in \mathcal{S}^n \\
\textit{sujeto a:} & x_{ii} + x_{jj} - 2x_{ij} = d_{ij}^2 \quad \forall \{i, j\} \in E, \\
 & \boldsymbol{X} \succeq 0
\end{array}
$$

Más aún, (G, D) es k-realizable si existe una solución $\boldsymbol{X}$ de rango a lo sumo k.

Demostración

Si $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^k$ es una realización de (G, D), entonces la matriz de Gram

$$\boldsymbol{X} = (\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle) = (\boldsymbol{v}_1 \ldots \boldsymbol{v}_n)^T (\boldsymbol{v}_1 \ldots \boldsymbol{v}_n)$$

es una solución del problema, de rango a lo sumo k.

Recíprocamente, si $\boldsymbol{X}$ es una solución del problema, podemos encontrar una descomposición

$$\boldsymbol{X} = \boldsymbol{V}^T \boldsymbol{V}, \quad \boldsymbol{V} \in \mathbb{R}^{k \times n}$$

y tomamos como $\boldsymbol{v}_i$ las columnas de $\boldsymbol{V}$.

Grafos completos revisitados

Utilizando el último teorema obtenemos otra caracterización para grafos completos.

Teorema

Sea $G = K_n$ grafo completo, con pesos D, y sea $\boldsymbol{X} \in \mathcal{S}^{n-1}$ la matriz definida por

$$\begin{aligned} x_{ii} &= d_{in}^2, \quad \forall 1 \leq i \leq n-1 \\ x_{ij} &= \frac{d_{in}^2 + d_{jn}^2 - d_{ij}^2}{2}, \quad \forall 1 \leq i \neq j \leq n-1. \end{aligned}$$

Entonces K_n es k-realizable si y solo si $\boldsymbol{X} \succeq 0$ y $\operatorname{rank} \boldsymbol{X} \leq k$.

Idea de la demostración. Si el problema SDP tiene solución factible, mediante una traslación, podemos suponer $\boldsymbol{v}_n = \boldsymbol{0}$.

El problema de la partición

Como vimos, determinar si un grafo (G, D) es realizable, puede resolverse eficientemente por optimización semi-definida. Sin embargo, determinar si es k-realizable para un k dado es un problema mucho más difícil.

Veremos un caso simple en el que el problema es NP-completo.

Consideramos el siguiente problema para el cual se sabe que es NP-completo.

El problema de la partición.

Dada una secuencia de números naturales $a_1, \ldots, a_n \in \mathbb{N}$, determinar si los números pueden separarse en dos conjuntos con la misma suma. Es decir, si existe $\boldsymbol{\epsilon} \in \{\pm 1\}^n$ tal que

$$\epsilon_1 a_1 + \cdots + \epsilon_n a_n = 0.$$

Grafo ciclo

Un grafo (G, E) se llama ciclo o cíclico si las aristas forma un ciclo de longitud n. Es decir, podemos suponer que las aristas son $(i, i+1)$ para $1 \leq i \leq n-1$ y $(n, 1)$.

Teorema

Dado un grafo ciclo (G, E) con pesos naturales $d \in \mathbb{N}^E$, decidir si (G, D) es 1-realizable es un problema NP-completo.

Demostración.

Para una instancia $a_1, \ldots, a_n \in \mathbb{N}$ del problema de la partición, consideramos el grafo ciclo (G, E) con pesos $d_{i(i+1)} = a_i$.

Si (G, D) es 1-realizable, con $v_i \in \mathbb{R}$, definimos $\epsilon_i = 1$ si $v_{i+1} > v_i$ y $\epsilon_i = -1$ si $v_{i+1} < v_i$ y obtenemos una partición de los a_i.

Rango 1 y rango 2

Puede demostrarse la siguiente propiedad:

Un grafo ciclo es realizable si y solo si es 2-realizable.

Ejercicio. Ver geométricamente que puedo llevar una realización 3D de un grafo ciclo a una realización 2D.

Concluimos que es posible contestar eficientemente si un grafo ciclo es realizable en un plano o una recta, pero determinar en cuál de los dos es un problema NP-completo.

Minimización de rango

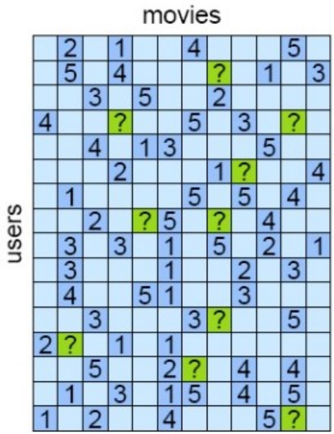
Como vimos, un problema interesante en optimización es el problema de la minimización de rango, que podemos plantear en la forma

$$\begin{aligned} \text{minimizar:} \quad & \operatorname{rank} \boldsymbol{X} \\ \text{sujeto a:} \quad & \boldsymbol{X} \in \mathcal{C}, \end{aligned}$$

donde la matriz $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ es la variable de decisión, y $\mathcal{C}$ es un conjunto convexo. Como la función a optimizar tiene valores enteros, este problema en general no es un problema convexo. En el ejemplo anterior vimos que en general es un problema NP-completo.

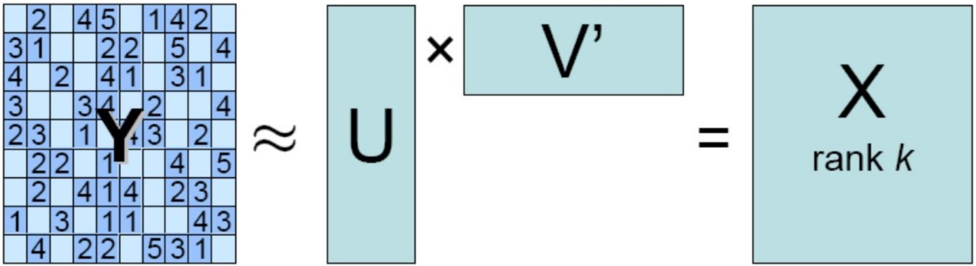
Ejemplo - El problema de Netflix

- Algunos usuarios califican algunas de las películas que vieron.
- Los puntajes son números enteros entre 1 y 5.
- Queremos predecir los puntajes para cada par usuario / película.
- Es decir, queremos completar una matriz incompleta como la de la figura.



Ejemplo - El problema de Netflix

Si pensamos que hay unos pocos perfiles de usuarios arquetípicos (amante de las películas de terror, románticas, comedias, etc.) y cada usuario es una combinación lineal de esos perfiles, podemos factorizar la matriz:



Rango y valores singulares

- Para una matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ cuadrada, el rango de $\boldsymbol{A}$ es igual a la cantidad de autovalores no-nulos.
- Para una matriz $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, el rango de $\boldsymbol{A}$ es igual a la cantidad de valores singulares no-nulos.

Heurística

- Como no podemos minimizar el rango eficientemente, minimizamos la suma de los valores singulares.
- Para $\boldsymbol{A} \succeq 0$, la suma de los valores singulares es igual a $\mathrm{Tr}(\boldsymbol{A})$, que es una función lineal en los coeficientes de $\boldsymbol{A}$.

Norma nuclear - Heurística

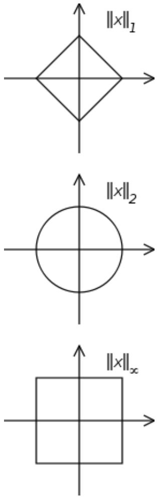
En $\mathbb{R}^{m\times n}$ definimos la norma nuclear

$$\|\boldsymbol{X}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i .$$

Como los valores singulares son todos no-negativos,

$$\|\boldsymbol{X}\|_* = \|\boldsymbol{\sigma}\|_1 .$$

Comparando las curvas de nivel distintas normas, vemos que minimizando la norma-1 obtenemos en general vectores esparsos con gran cantidad de 0's. Por este motivo es una buena elección para obtener matrices de rango bajo.



Norma nuclear

Para analizar las propiedades de la norma nuclear, recordemos las siguientes propiedades:

- Para $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$,

$$\boldsymbol{A} \bullet \boldsymbol{B} = \operatorname{Tr}(\boldsymbol{A}^T \boldsymbol{B}) = \operatorname{Tr}(\boldsymbol{A}\boldsymbol{B}^T)$$

  es un producto interno en $\mathbb{R}^{n \times n}$, que define la norma

$$\|\boldsymbol{A}\|_F = \sqrt{\boldsymbol{A} \bullet \boldsymbol{A}} = \sqrt{\operatorname{Tr}(\boldsymbol{A}\boldsymbol{A}^T)}.$$

  Si $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{A}^T$, $\operatorname{Tr}(\boldsymbol{X}) = \|\boldsymbol{A}\|_F^2$.
- Desigualdad de Cauchy-Schwarz: $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|$
- Desigualdad MA-MG: $\|\boldsymbol{u}\| \|\boldsymbol{v}\| \leq \frac{1}{2}(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2)$

Norma nuclear

Lema

Para $\boldsymbol{X} \in \mathbb{R}^{n \times n}$,

$$\|\boldsymbol{X}\|_* = \min_{\boldsymbol{X}=\boldsymbol{U}\boldsymbol{V}^T} \|\boldsymbol{U}\|_F \|\boldsymbol{V}\|_F = \min_{\boldsymbol{X}=\boldsymbol{U}\boldsymbol{V}^T} \frac{1}{2}(\|\boldsymbol{U}\|^2 + \|\boldsymbol{V}\|^2).$$

Este y los siguientes resultados valen también para $\boldsymbol{X}$ rectangular.

Demostración. Tomamos una descomposición SVD de $\boldsymbol{X}$,

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{S}\boldsymbol{Q}^T,$$

con $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ unitaria, $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ diagonal, con los valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ en la diagonal y $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ unitaria.

Norma nuclear

Si $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^T$, con $\boldsymbol{U} \in \mathbb{R}^{n\times k}$, $\boldsymbol{V} \in \mathbb{R}^{n\times k}$, entonces

$$\begin{aligned}
\|\boldsymbol{X}\|_* &= \mathrm{Tr}(\boldsymbol{S}) = \mathrm{Tr}(\boldsymbol{P}^T\boldsymbol{U}\boldsymbol{V}^T\boldsymbol{Q}) \\
&= \boldsymbol{P}^T\boldsymbol{U} \bullet \boldsymbol{Q}^T\boldsymbol{V} \leq \|\boldsymbol{P}^T\boldsymbol{U}\|_F \|\boldsymbol{Q}^T\boldsymbol{V}\|_F \\
&= \|\boldsymbol{U}\|_F \|\boldsymbol{V}\|_F \leq \frac{1}{2}(\|\boldsymbol{U}\|_F + \|\boldsymbol{V}\|_F),
\end{aligned}$$

donde $\boldsymbol{P}^T\boldsymbol{U}, \boldsymbol{Q}^T\boldsymbol{V} \in \mathbb{R}^{n\times k}$.

Tomando $\boldsymbol{U} = \boldsymbol{P}\boldsymbol{S}^{\frac{1}{2}}$ y $\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{S}^{\frac{1}{2}}$, alcanzamos el mínimo.

Norma nuclear

Utilizando el lema, obtenemos una caracterización de la norma nuclear por un problema SDP.

Lema

Para cualquier matriz $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ *y* $t \in \mathbb{R}$, $\|\boldsymbol{X}\|_* \leq t$ *si y solo si existen* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *y* $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ *tales que*

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{pmatrix} \succeq 0 \quad \text{y} \quad \mathrm{Tr}(\boldsymbol{A}) + \mathrm{Tr}(\boldsymbol{B}) \leq 2t.$$

Demostración

Si $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{pmatrix} \succeq 0$, podemos factorizarla

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{pmatrix} = \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}^T & \boldsymbol{V}^T \end{pmatrix},$$

con $\boldsymbol{U} \in \mathbb{R}^{n\times(n+n)}$ y $\boldsymbol{V} \in \mathbb{R}^{n\times(n+n)}$.

Tenemos $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{U}^T$, $\boldsymbol{B} = \boldsymbol{V}\boldsymbol{V}^T$ y $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^T$.

Por lo tanto, $\|\boldsymbol{U}\|_F^2 + \|\boldsymbol{V}\|_F^2 = \operatorname{Tr}\boldsymbol{A} + \operatorname{Tr}\boldsymbol{B} \leq 2t$, y obtenemos

$$\|\boldsymbol{X}\|_* \leq \frac{1}{2}(\|\boldsymbol{U}\|^2 + \|\boldsymbol{V}\|^2) \leq t.$$

Demostración

Recíprocamente, si $\|\boldsymbol{X}\|_* \leq t$, y $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{S}\boldsymbol{Q}^T$ es una descomposición SVD, tomando $\boldsymbol{U} = \boldsymbol{P}\boldsymbol{S}^{\frac{1}{2}}$ y $\boldsymbol{V}^T = \boldsymbol{S}^{\frac{1}{2}}\boldsymbol{Q}^T$, obtenemos

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}^T & \boldsymbol{V}^T \end{pmatrix} = \begin{pmatrix} \boldsymbol{U}\boldsymbol{U}^T & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{V}\boldsymbol{V}^T \end{pmatrix} \succeq 0$$

y $\mathrm{Tr}(\boldsymbol{Z}) = \mathrm{Tr}(\boldsymbol{U}\boldsymbol{U}^T) + \mathrm{Tr}(\boldsymbol{V}\boldsymbol{V}^T) \leq 2t$.

Norma nuclear como problema SDP

Extendiendo los resultados anteriores a matrices rectangulares (ejercicio), obtenemos que la norma nuclear $\|\boldsymbol{X}\|_*$, $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, se corresponde con el valor óptimo del problema SDP

$$\begin{array}{ll} \text{minimizar:} & \dfrac{1}{2}\operatorname{Tr}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{pmatrix} \\ \text{sujeto a:} & \begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{pmatrix} \succeq 0 \end{array}$$

para matrices $\boldsymbol{A} \in \mathbb{R}^{m\times m}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n}$.

Problema SDP

Para el problema de Netflix, obtenemos que el problema de completar una matriz $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ para la cual solo se conocen algunas casillas m_{ij}, $(i,j) \in I$ de forma tal que la norma nuclear sea mínima es equivalente a resolver el problema

$$
\begin{aligned}
\text{minimizar:} \quad & \frac{1}{2}\operatorname{Tr}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{pmatrix} \\
\text{sujeto a:} \quad & x_{ij} = m_{ij}, (i,j) \in I, \\
& \begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{B} \end{pmatrix} \succeq 0
\end{aligned}
$$

para matrices $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, $\boldsymbol{A} \in \mathbb{R}^{m\times m}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n}$.

Más generalmente, podemos reemplazar la restricción $x_{ij} = m_{ij}, (i,j) \in I$, por $\boldsymbol{X} \in \mathcal{C}$, para $\mathcal{C}$ un conjunto convexo (definido por restricciones lineales).

Ejemplo: sumas de cuadrados

Determinar si un polinomio $f \in \mathbb{R}[x_1, x_2, x_3]$ homogéneo de grado 4 es suma de cuadrados, es equivalente a determinar si existe una matrix $\boldsymbol{A}$ semidefinida positiva tal que

$$f = \boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v},$$

para $\boldsymbol{v} = \left(x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3\right)$.

La escritura como suma de cuadrados se puede obtener factorizando $\boldsymbol{A} = \boldsymbol{X}^T \boldsymbol{X}$ y el rango de $\boldsymbol{A}$ nos dice la cantidad de polinomios linealmente independientes en la descomposición.

Si queremos estudiar el problema de hallar la menor cantidad de polinomios que aparecen en una descomposición, debemos minimizar el rango de la matriz $\boldsymbol{A}$.

Norma nuclear de matrices simétricas

Si trabajamos con matrices simétricas, minimizar la norma nuclear equivale a resolver el problema SDP:

$$\begin{array}{ll} \text{minimizar:} & \operatorname{Tr} \boldsymbol{X} \\ \text{sujeto a:} & \boldsymbol{X} \in \mathcal{C}, \\ & \boldsymbol{X} \succeq 0. \end{array}$$