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

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

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

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

Si $\boldsymbol{A}$ 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:

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

Matrices simétricas

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

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

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

Proposición

Si $\boldsymbol{A}$ es simétrica, entonces

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

Teorema espectral para matrices simétricas

Teorema

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

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

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

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

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

con $\boldsymbol{U}$ matriz ortogonal y $\boldsymbol{T}$ triangular superior. Si $\boldsymbol{X}$ es simétrica, $\boldsymbol{T}$ 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 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ son *congruentes* si existe una matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ inversible tal que

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^T.$$

La relación de congruencia es una relación de equivalencia.

Signatura de una matriz

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

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

donde

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

Si $\boldsymbol{A} \in \mathcal{S}^n$, podemos escribir $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ con $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, la matriz con los autovalores de $\boldsymbol{A}$ en la diagonal, y $\boldsymbol{U}$ 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

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

obtenemos

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

Ley de inercia

Podemos escribir a la matriz $\boldsymbol{A}$ como

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

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

Ley de inercia

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

Teorema

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

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

Ley de inercia (continacuón de la demostración)

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

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

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

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_k \boldsymbol{v}_k \neq 0,$$

entonces

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

Ley de inercia (continación de la demostración)

Por lo tanto,

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

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

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 $\boldsymbol{S}$.

Diagonalización por congruencia

Dada una matriz simétrica $\boldsymbol{A} \in \mathcal{S}^n$ podemos obtener una matriz diagonal $\boldsymbol{D}$ congruente a $\boldsymbol{A}$ 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 $\boldsymbol{D}$ diagonal que sea congruente a

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

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

Matrices definidas positivas

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

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

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

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

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

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

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

Caracterización de matrices semidefinidas positivas

Dado un conjunto $S \subset \{1, \ldots, n\}$, definimos el menores principales 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 un matriz simétrica, las siguientes propiedades son equivalentes:

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

Caracterización de matrices definidas positivas

Para matrices definidas positivas, obtenemos equivalencias similares.

Teorema

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

1. $\boldsymbol{A}$ *es definida positiva* ($\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ *para todo* $\boldsymbol{x} \in \mathbb{R}^n$*),*
2. *todos los autovalores de* $\boldsymbol{A}$ *son positivos,*
3. $\boldsymbol{A} = \boldsymbol{M}^T\boldsymbol{M}$ *para alguna matriz* $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ *no singular,*
4. *todos los menores principales de* $\boldsymbol{A}$ *son positivos,*
5. *el polinomio característico de* $\boldsymbol{A}$ *tiene signos alternados: si* $\chi_{\boldsymbol{A}}(t) = t^n + a_{n-1}t^{n-1} + a_1 t + a_0$*, entonces* $a_i a_{i+1} < 0$ *para todo* $0 \leq i \leq n-1$ *(definiendo* $a_n = 1$*).*

Menores principales de cabeza

Si $S = \{1, 2, \ldots, k\}$, 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 $\boldsymbol{A}$, los autovalores de

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

con $\boldsymbol{y} \in \mathbb{R}^n$ y $a \in \mathbb{R}$ están entrecruzados con los autovalores de $\boldsymbol{A}$ (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 $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ admite una descomposición $\boldsymbol{QR}$ y puede escribirse como $\boldsymbol{C} = \boldsymbol{QR}$ con $\boldsymbol{Q}$ unitaria y $\boldsymbol{R}$ triangular superior del mismo rango que $\boldsymbol{C}$. Luego

$$\boldsymbol{A} = \boldsymbol{C}^T\boldsymbol{C} = (\boldsymbol{QR})^T\boldsymbol{QR} = \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{QR} = \boldsymbol{R}^T\boldsymbol{R}$$

y tomando $\boldsymbol{L} = \boldsymbol{R}^T$, obtenemos $\boldsymbol{A} = \boldsymbol{LL}^T$.

Factorización de Cholesky

Si $\boldsymbol{C}$ es no singular, podemos elegir $\boldsymbol{R}$ 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 $\boldsymbol{A}$ es definida positiva si y solo si existe una matriz trinagular inferior $\boldsymbol{L} \in \mathbb{R}^{n \times n}$ con valores positivos en la diagonal tal que $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$.

Conos

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

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

Conos

Decimos que un cono es *puntiagudo* si

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

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

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

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

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

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

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

Separación

Tenemos el siguiente resultado de separación.

Lema

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

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

Ejemplos

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

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

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

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

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:

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

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

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

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

Desplazamiento máximo

Recordemos que para las variables x_i no-básicas,

- o bien $d_i = 1$ si es la variable que pasa a ser variable básica,
- o bien $d_i = 0$.

En ambos casos $d_i \geq 0$, por lo tanto podemos restringirnos a mirar las variables básicas en la fórmula anterior. Obtenemos

$$\theta^\star = \min_{\{i=1,\ldots,m \mid d_{B(i)}<0\}} \left(-\frac{x_{B(i)}}{d_{B(i)}} \right). \tag{1}$$

Más aún, como supusimos que todas las soluciones son no-degeneradas, $x_i > 0$ para todas las variables básicas, y por lo tanto obtenemos siempre $\theta^\star > 0$.

Ejemplo (cont.)

Continuamos el ejemplo. Obtuvimos

$$\boldsymbol{x} = (1, 1, 0, 0), \quad \boldsymbol{d} = (-3/2, 1/2, 1, 0) \quad \text{y} \quad \bar{c}_3 = -3.$$

Como $\bar{c}_3$ es negativo, podemos disminuir el costo de la función objetivo desplazándonos en esta dirección. Es decir, consideramos los vectores

$$\boldsymbol{x} + \theta \boldsymbol{d},$$

con $\theta > 0$.

La única coordenada que decrece al desplazarnos es x_1, por lo que si tomamos cualquier valor de $\theta > 0$ tal que la primera coordenada se mantenga no-negativa, estaremos dentro del poliedro.

Ejemplo (cont.)

El máximo valor de θ que podemos tomar es

$$\theta^\star = -\frac{x_1}{d_1} = \frac{2}{3},$$

y nos desplazamos a

$$\boldsymbol{y} = \boldsymbol{x} + \frac{2}{3}\boldsymbol{d} = \left(0, \frac{4}{3}, \frac{2}{3}, 0\right).$$

Las columnas correspondientes a las coordenadas no-nulas son $\boldsymbol{A}_2 = (1, 0)$ y $\boldsymbol{A}_3 = (1, 3)$, y son linealmente independientes. Por lo tanto $\boldsymbol{y}$ es una nueva solución básica factible. Como queríamos minimizar el valor de x_1, y tenemos la restricción $x_1 \geq 0$, hemos encontrado el costo óptimo.

Ejemplo (cont.)

Para verificación, calculamos el costo reducido $\bar{c}_4$ por la fórmula

$$\bar{c}_4 = c_4 - \boldsymbol{c}_B \cdot (\boldsymbol{B}^{-1}\boldsymbol{A}_4) = 1 - \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \frac{4}{3}.$$

Vemos que movernos en la dirección básica correspondiente va a incrementar el valor de la función objetivo y por lo tanto no debemos desplazarnos en esa dirección.

Nueva solución básica factible

Una vez que determinamos $\theta^\star$, suponiendo que es finito, nos movemos a una nueva solución factible

$$\boldsymbol{y} = \boldsymbol{x} + \theta^\star \boldsymbol{d}.$$

Por la construcción que hicimos de $\boldsymbol{d}$ y $\theta^\star$, si ℓ es el índice que minimiza la fórmula (1) de $\theta^\star$ (es decir, $\theta^\star = -x_{B(\ell)}/d_{B(\ell)}$), se cumple

$$x_{B(\ell)} = 0 \quad \text{y} \quad x_j = \theta^\star.$$

Obtenemos que x_j reemplaza a $x_{B(\ell)}$ en el conjunto de variables básicas. El nuevo conjunto de índices de las variables básicas es $\bar{B} = \{\bar{B}(1), \ldots, \bar{B}(m)\}$, con

$$\bar{B}(i) = \begin{cases} B(i), & i \neq \ell, \\ j, & i = \ell. \end{cases}$$

Nueva solución básica factible

Llamamos $\bar{\boldsymbol{B}}$ a la nueva matriz base, formada por las columnas de $\boldsymbol{A}$ correspondientes a las nuevas variables básicas.

Teorema

1. *Las columnas $\boldsymbol{A}_{B(i)}$, $i \neq \ell$, y $\boldsymbol{A}_j$ son linealmente independientes y, por lo tanto, $\bar{\boldsymbol{B}}$ es una matriz base.*
2. *El vector $\boldsymbol{y} = \boldsymbol{x} + \theta^\star \boldsymbol{d}$ es una solución básica factible con matriz base asociada $\bar{\boldsymbol{B}}$.*

Demostración. Ver [Bertsimas and Tsitsiklis, 1997, Teorema 3.2].

Algortimo del método simplex

Mediante el procedimiento descripto, pasamos de una solución básica factible a otra solución básica factible con menor costo. Obtenemos la siguiente iteración del método simplex.

1. Comenzamos con una solución básica factible $\boldsymbol{x}$, correspondiente a variables básicas con índices $I = \{B(1), \ldots, B(m)\}$.
2. Tomamos la submatriz $\boldsymbol{B}$ de $\boldsymbol{A}$ formada por las columnas correspondientes a las variables básicas, y calculamos los costos reducidos

$$\bar{c}_j = c_j - \boldsymbol{c}_B \cdot (\boldsymbol{B}^{-1}\boldsymbol{A}_j), \quad j \notin I.$$

   Si todos los costos son no-negativos, $\boldsymbol{x}$ es una solución óptima y terminamos. Si no, elegimos algún j tal que $\bar{c}_j < 0$.

Algortimo del método simplex

3. Calculamos $\boldsymbol{u} = -\boldsymbol{d}_B = \boldsymbol{B}^{-1}\boldsymbol{A}_j$. Si $\boldsymbol{u}$ no tiene ninguna componente positiva, tomamos $\theta^\star = \infty$, el costo óptimo es $-\infty$ y el algoritmo termina.
4. Si alguna componente de $\boldsymbol{u}$ es positiva, tomamos

$$\theta^\star = \min_{\{i=1,\dots,m|u_i>0\}} \frac{x_{B(i)}}{u_i}.$$

5. Sea ℓ un índice para el cuál se alcanza el mínimo. Construimos una nueva base reemplazando $B(\ell)$ por j. La nueva variable básica $\boldsymbol{y}$ queda definida por

$$\begin{cases} y_j = \theta^\star, & \\ y_{B(i)} = x_{B(i)} - \theta^\star u_i, & 1 \leq i \leq m, i \neq l \\ y_i = 0 & \text{para los índices de variables no-básicas.} \end{cases}$$

Puntos extremales y soluciones óptimas

En base a todo lo que vimos es fácil verificar que el método simplex descripto resuelve correctamente el problema de programación lineal.

Completando el método para considerar soluciones degeneradas, podemos demostrar la siguiente propiedad de los problemas de programación lineal.

Teorema

Consideremos el problema de programación lineal de minimizar una funcional $\boldsymbol{c} \cdot \boldsymbol{x}$ sobre un poliedro P. Si P tiene al menos un punto extremal entonces o bien el costo óptimo es $-\infty$ o bien existe un punto extremal que es óptimo.

Existencia de puntos extremales

En general los poliedros pueden no tener puntos extremales. Por ejemplo, un semiplano en $\mathbb{R}^2$ no tiene puntos extremales. Una condición muy simple para determinar si un poliedro tiene puntos extremales es la existencia o no de líneas rectas incluidas en el poliedro.

Decimos que un poliedro $P \subset \mathbb{R}^n$ contiene una recta si existe un vector $\boldsymbol{x} \in P$ y una dirección $\boldsymbol{d} \in \mathbb{R}^n$ tales que $\boldsymbol{x} + \lambda \boldsymbol{d} \in P$ para todo $\lambda \in \mathbb{R}$.

Existencia de puntos extremales

Teorema

Dado un poliedro $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i \cdot \boldsymbol{x} \geq b_i, i = 1, \ldots, m\}$, *las siguientes condiciones son equivalentes.*

1. *El poliedro* P *contiene al menos un punto extremal.*
2. *El poliedro* P *no contiene ninguna recta.*
3. *Existen* n *vectores linealmente independientes entre los vectores* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$.

Demostración. [Bertsimas and Tsitsiklis, 1997, Teorema 2.6].

En particular, cualquier poliedro acotado tiene puntos extremales y cualquier poliedro en forma estándar tiene puntos extremales, debido a que el ortante positivo $\{x_i \geq 0, 1 \leq i \leq n\}$ no contiene ninguna recta.

Referencias

Bertsimas, D. and Tsitsiklis, J. (1997).
Introduction to Linear Optimization.
Athena Scientific.

Horn, R. and Johnson, C. (1985).
Matrix Analysis.
Cambridge University Press.