Optimización Semidefinida
Clase 03 - El método simlex

Segundo Cuatrimestre 2021

Facultad de Ciencias Exactas y Naturales, UBA

Conceptos generales de optimización

- Muchos métodos de optimización se basan en un principio simple que consiste en comenzar en una solución factible cualquiera e intentar moverse a otra solución factible cercana de forma que se reduzca el costo de la función a minimizar. Si no existe ninguna solución cercana que permita mejorar la función, hemos alcanzado un mínimo local.
- En general, un mínimo local no tiene por qué ser un mínimo global, una función podría tener varios mínimos locales, y el mínimo global ser solo uno de ellos.
- En programación lineal un mínimo local es también global, dado que estamos minimizando una función convexa sobre un conjunto convexo.

Propiedad fundamental de programación lineal

El método Simplex es uno de los métodos más usados en la práctica para resolver problemas de programación lineal y se basa en aprovechar una propiedad adicional de la programación lineal:

Si existe un mínimo, éste se alcanza en un vértice del poliedro de puntos factibles.

Veremos que

- si estamos parados en un vértice, alcanza verificar si la función objetivo decrece al movernos a alguno de los vértices vecinos.
- si no decrece en ninguno de los vértices vecinos, hemos encontrado un mínimo local y por lo tanto global.

Esquema del método simplex

Obtenemos el siguiente algoritmo para minimizar una funcional $\boldsymbol{c} \cdot \boldsymbol{x}$ sobre un poliedro P:

1. Elegir un vértice $\boldsymbol{x}$ del poliedro P.
2. Para cada uno de los vértices vecinos $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s\}$ de $\boldsymbol{x}$ verificar si la función objetivo mejora en esos puntos.
3. Si no mejora en ninguno de los puntos, $\boldsymbol{x}$ es un óptimo global y finalizamos el procedimiento.
4. Si encontramos $\boldsymbol{y}_j$ tal que $\boldsymbol{c} \cdot \boldsymbol{y}_j < \boldsymbol{c} \cdot \boldsymbol{x}$, tomamos $\boldsymbol{x} = \boldsymbol{y}_j$ y volvemos a comenzar.

Problema en forma estándar

Para todo el desarrollo del método simplex, vamos a considerar el problema en forma estándar

$$\begin{aligned} \text{minimizar:} \quad & \boldsymbol{c} \cdot \boldsymbol{x} \\ \text{sujeto a:} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geq 0. \end{aligned}$$

con $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ matriz de rango $m \leq n$.

Soluciones adyacentes

Dos soluciones básicas de un conjunto de restricciones lineales en $\mathbb{R}^n$ se dicen *adyacentes* si existen $n-1$ restricciones linealmente independientes que están activas en ambas soluciones.

Si las dos soluciones son factibles, llamamos *arista* del conjunto factible al segmento que las une.

Ejercicio

Sea $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i \cdot \boldsymbol{x} \geq b_i, 1 \leq i \leq m\}$ *un poliedro y* $\boldsymbol{u}$, $\boldsymbol{v}$ *dos soluciones básicas factibles adyacentes, con* $\boldsymbol{a}_i \cdot \boldsymbol{u} = \boldsymbol{a}_i \cdot \boldsymbol{v} = b_i$ *para* $1 \leq i \leq n-1$ *y* $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}\}$ *vectores linealmente independientes. Probar que el segmento* $L = \{\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v} \mid 0 \leq \lambda \leq 1\}$ *que une* $\boldsymbol{u}$ *y* $\boldsymbol{v}$ *verifica*

$$L = \{\boldsymbol{z} \in P \mid \boldsymbol{a}_i \cdot \boldsymbol{z} = b_i, 1 \leq i \leq n-1\}.$$

Direcciones básicas

- Para movernos de una solución dada a una adyacente, podríamos reemplazar alguna de las variables básicas por una variable no-básica y calcular la nueva solución básica.
- Sin embargo, vamos a ver que al elegir cuál es la variable no-básica que vamos a convertir en básica, esto ya determina la dirección en la que debemos movernos para desplazarnos a una solución adyacente.

Direcciones básicas

- Para movernos de una solución dada a una adyacente, podríamos reemplazar alguna de las variables básicas por una variable no-básica y calcular la nueva solución básica.
- Sin embargo, vamos a ver que al elegir cuál es la variable no-básica que vamos a convertir en básica, esto ya determina la dirección en la que debemos movernos para desplazarnos a una solución adyacente.
- En efecto, al permitir una nueva variable no-nula, tendremos un sistema de $m+1$ ecuaciones y m incógnitas, cuyas soluciones forman una recta afín.

Coordenadas de una solución básica

Sea $\boldsymbol{x}$ una solución básica factible de un problema en forma estándar y sean $B = \{B(1), \ldots, B(m)\}$ los índices de las variables básicas.

Sea $\boldsymbol{B} = (\boldsymbol{A}_{B(1)} \cdots \boldsymbol{A}_{B(m)}) \in \mathbb{R}^{m \times m}$ la correspondiente matriz base.

En particular $x_i = 0$ para cada variable no-básica y podemos calcular las coordenadas correspondientes a variables básicas por la fórmula

$$\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}.$$

Direcciones básicas

Para movernos a una solución adyacente en la que una variable no-básica x_j pase a ser variable básica, debemos desplazarnos a un nuevo vector

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

con $\theta > 0$ y $\boldsymbol{d} = (d_1, \ldots, d_n)$ con $d_j = 1$ y $d_i = 0$ para todos los índices i distintos de j de variables no-básicas.

El vector $\boldsymbol{x}_B$ de variables básicas va a cambiar a $\boldsymbol{x}_B + \theta \boldsymbol{d}_B$ con $\boldsymbol{d}_B = (d_{B(1)}, \ldots, d_{B(m)})$.

Direcciones básicas

Como solo nos interesan las soluciones básicas factibles, debe cumplirse

$$\boldsymbol{A}(\boldsymbol{x} + \theta \boldsymbol{d}) = \boldsymbol{b},$$

y como $\boldsymbol{x}$ también es factible, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Por lo tanto debe cumplirse $\boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$ y esto nos permite calcular $\boldsymbol{d}_B$. En efecto,

$$\boldsymbol{0} = \boldsymbol{A}\boldsymbol{d} = \sum_{i=1}^{n} \boldsymbol{A}_i d_i = \sum_{i=1}^{m} \boldsymbol{A}_{B(i)} d_{B(i)} + \boldsymbol{A}_j = \boldsymbol{B}\boldsymbol{d}_B + \boldsymbol{A}_j,$$

donde $\boldsymbol{A}_i$ es la i-ésima columna de $\boldsymbol{A}$.

Como la matriz $\boldsymbol{B}$ es inversible, obtenemos

$$\boldsymbol{d}_B = -\boldsymbol{B}^{-1}\boldsymbol{A}_j.$$

Direcciones básicas

- Llamamos *j-ésima dirección básica* al vector $\boldsymbol{d}$ que acabamos de construir.
- Por construcción, las restricciones de igualdad se van a mantener si nos movemos en esta dirección.
- Para las condiciones de no-negatividad de las variables, recordemos que la variable no-básica x_j aumenta y las demás variables no-básicas se mantienen en 0, por lo tanto conservan la no-negatividad.
- Para las variables básicas, como estamos suponiendo que todas las soluciones son no-degeneradas, las variables básicas son todas positivas y podemos tomar θ suficientemente pequeño para que se sigan siendo positivas.

Costo reducido

Antes de calcular el valor de θ apropiado para movernos a una solución adyacente, calculamos cómo varía el costo de la función a optimizar al desplazarnos en la dirección de la j-ésima dirección básica.

Si $\boldsymbol{d}$ es la j-ésima dirección básica, la razón $\boldsymbol{c} \cdot \boldsymbol{d}$ de cambio del costo en la dirección $\boldsymbol{d}$ está dada por

$$\boldsymbol{c} \cdot \boldsymbol{d} = \boldsymbol{c}_B \cdot \boldsymbol{d}_B + c_j,$$

donde $\boldsymbol{c}_B = (c_{B(1)}, \ldots, c_{B(m)})$.

Costo reducido

Utilizando que $\boldsymbol{d}_B = -\boldsymbol{B}^{-1}\boldsymbol{A}_j$, hacemos la siguiente definición.

Definición

Sea $\boldsymbol{x}$ una solución básica asociada a una matriz base $\boldsymbol{B}$, y sea $\boldsymbol{c}_B$ el vector de costos de las variables básicas. Para cada j, definimos el *costo reducido* $\bar{c}_j$ de la variable x_j por la fórmula

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

El costo reducido nos dice cuánto varía la función de costo al movernos una unidad en la dirección de la j-ésima dirección básica. El término c_j indica el aumento del costo por unidad de la variable x_j y el último término es el costo de modificar las demás variables para compensar el cambio en la variable x_j.

Ejemplo

Consideremos el problema de programación lineal

$$\begin{aligned} \text{minimizar:} \quad & \boldsymbol{c} \cdot \boldsymbol{x} \\ \text{sujeto a:} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq 0, \end{aligned}$$

con

$$\boldsymbol{c} = (2, 0, 0, 0) \in \mathbb{R}^4, \quad \boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 3 & 4 \end{pmatrix} \in \mathbb{R}^{2\times 4} \quad \text{y} \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \in \mathbb{R}^2.$$

Ejemplo (cont.)

Las primeras dos columnas de $\boldsymbol{A}$ son $\boldsymbol{A}_1 = (1, 2)$ y $\boldsymbol{A}_2 = (1, 0)$. Como son linealmente independientes podemos elegir x_1 y x_2 como variables básicas. La matriz base correspondiente es

$$\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}.$$

Tomamos $x_3 = x_4 = 0$, y calculamos los valores de x_1 y x_2: $x_1 = 1$ y $x_2 = 1$. Luego $\boldsymbol{x} = (1, 1, 0, 0)$ es una solución básica factible no-degenerada.

Ejemplo (cont.)

Construimos una dirección básica $\boldsymbol{d}$ correspondiente a aumentar el valor de la variable x_3. Tenemos $d_3 = 1$ y $d_4 = 0$. Las coordenedas de $\boldsymbol{d}$ correspondientes a las variables básicas las obtenemos por la fórmula

$$\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} d_{B(1)} \\ d_{B(2)} \end{pmatrix} = \boldsymbol{d}_B = -\boldsymbol{B}^{-1}\boldsymbol{A}_3 = -\begin{pmatrix} 0 & \frac{1}{2} \\ 1 & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -\frac{3}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Por lo tanto $\boldsymbol{d} = \left(-\frac{3}{2}, \frac{1}{2}, 1, 0\right)$ y el costo de movernos en esta dirección es

$$\bar{c}_3 = \boldsymbol{c} \cdot \boldsymbol{d} = -\frac{3}{2}c_1 + \frac{1}{2}c_2 + c_3 = -3.$$

Costos reducidos de las variables básicas

Calculamos el costo reducido de las variables básicas aplicando la definición.
Consideramos $x_{B(i)}$.

Como $\boldsymbol{B}^{-1}\boldsymbol{B} = \boldsymbol{I}$,

$$\boldsymbol{B}^{-1}\boldsymbol{A}_{B(i)} = \boldsymbol{e}_i,$$

el i-ésimo vector canónico. Por lo tanto,

$$\bar{c}_{B(i)} = c_{B(i)} - \boldsymbol{c}_B \cdot \boldsymbol{B}^{-1}\boldsymbol{A}_{B(i)} = c_{B(i)} - \boldsymbol{c}_B \cdot \boldsymbol{e}_i = c_{B(i)} - c_{B(i)} = 0,$$

y comprobamos que el costo reducido de las variables básicas es 0.

Costos reducidos y soluciones óptimas

El siguiente resultado confirma la idea intuitiva que tenemos de los costos reducidos y es la clave del método simplex.

Teorema

Sea $\boldsymbol{x}$ una solución básica factible asociada a una matriz base $\boldsymbol{B}$, y sea $\bar{\boldsymbol{c}}$ el vector de costos reducidos.

1. *Si $\bar{\boldsymbol{c}} \geq 0$, entonces $\boldsymbol{x}$ es una solución óptima.*
2. *Si $\boldsymbol{x}$ es una solución óptima no-degenerada, entonces $\bar{\boldsymbol{c}} \geq 0$.*

Desplazamiento máximo

Por último, si encontramos una dirección básica en la cual movernos para reducir el costo, debemos calcular el mayor valor de θ que podemos tomar, lo que equivale a determinar

$$\theta^\star = \max\{\theta \geq 0 \mid \boldsymbol{x} + \theta \boldsymbol{d} \in P\}.$$

Vamos a deducir una fórmula para $\theta^\star$. Como $\boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$, tenemos que

$$\boldsymbol{A}(\boldsymbol{x} + \theta \boldsymbol{d}) = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

para todo θ, por lo tanto las restricciones de igual se cumplen siempre.

El punto $\boldsymbol{x} + \theta \boldsymbol{d}$ solo puede salirse del poliedro si alguna de las coordenadas se vuelve negativa.

Desplazamiento máximo

Distinguimos dos casos:

1. Si $\boldsymbol{d} \geq 0$ (es decir, $d_i \geq 0$ para todo $1 \leq i \leq n$), entonces $\boldsymbol{x} + \theta \boldsymbol{d} \geq 0$ para todo $\theta \geq 0$. El vector $\boldsymbol{x} + \theta \boldsymbol{d}$ nunca se vuelve no-factible, y tomamos $\theta^\star = +\infty$. El costo óptimo es $-\infty$.
2. Si $d_i < 0$ para algunos i, la restricción $x_i + \theta d_i > 0$ nos da la restricción
$$\theta \leq -\frac{x_i}{d_i}.$$
Estas restricciones se deben cumplir para todo i tal que $d_i < 0$. Por lo tanto el mayor valor de θ que podemos tomar es
$$\theta^\star = \min_{\{i | d_i < 0\}} \left(-\frac{x_i}{d_i} \right).$$

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

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

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.