תכנון ליניארי: 9 במרץ, 2000

המרצה: יובל רבני

1 תכנון ליניארי

בעיית התכנון הליניארי, בהצגה כללית, היא למצוא

מקסימום (או מינימום) של $c_1^T x_1 + c_2^T x_2 + c_3^T x_3$ תחת האילוצים

$$A_{11}x_1 + A_{12}x_2 + A_{13}x_3 \geq b_1$$
$$A_{21}x_1 + A_{22}x_2 + A_{23}x_3 \leq b_2$$
$$A_{31}x_1 + A_{32}x_2 + A_{33}x_3 \geq b_3$$

$x_1 \geq 0$, $x_2 \leq 0$, x_3 לא מוגבל בסימן

באשר $c_1, x_1 \in \mathrm{R}^{n_1}$, $c_2, x_2 \in \mathrm{R}^{n_2}$, $c_3, x_3 \in \mathrm{R}^{n_3}$, $b_1 \in \mathrm{R}^{m_1}$, $b_2 \in \mathrm{R}^{m_2}$, $b_3 \in \mathrm{R}^{m_3}$, ו-$A_{ij} \in \mathrm{R}^{m_i \times n_j}$.

וקטור ב-$\mathrm{R}^{n_1+n_2+n_3}$ המקיים את האילוצים של התוכנית הליניארית נקרא פיתרון אפשרי. תוכנית שיש לה פיתרון אפשרי נקראת תוכנית אפשרית. עבור בעיית מקסימיזציה (ההגדרות לבעיית מינימיזציה דומות) פיתרון אפשרי, שערכו מירבי נקרא פיתרון אופטימלי. אם לכל ממשי יש פיתרון אפשרי שערכו גדול יותר, התוכנית לא חסומה.

מקובל לעבוד עם הצגות מוגבלות יותר, למשל, הצגה סטנדרטית:

$$\begin{aligned} \min \quad c^T x \quad & \text{subject to} \\ Ax &= b \\ x &\geq 0. \end{aligned}$$

קל להסב כל תוכנית ליניארית P לתוכנית שקולה P' בהצגה סטנדרטית. השקילות היא בכך שיש התאמה בין הפיתרונות האפשריים של P' לבין הפיתרונות האפשריים של P, כך שהסדר המושרה על ידי פונקצית המטרה של כל תוכנית נשמר על ידי ההתאמה. על מנת לבצע את ההתאמה, נפעיל את הכללים הבאים:

- $\max c^T x \Leftrightarrow \min -c^T x$;

- $Ax \geq b \Leftrightarrow Ax - s = b,\ s \geq 0$;
- $Ax \leq b \Leftrightarrow Ax + s = b,\ s \geq 0$;
- $x \leq 0 \Leftrightarrow y = -x \geq 0$;
- x לא מוגבל בסימן $\Leftrightarrow x = x^+ - x^-,\ x^+, x^- \geq 0$.

מקובלות גם הצגות אחרות, למשל הצגה קנונית:

$$\begin{array}{lll} \min & c^T x & \text{subject to} \\ & Ax \geq b & \end{array}$$

קל לעבור מהצגה סטנדרטית להצגה קנונית: נחליף כל שיוויון בשני אי-שיוויונות.

2 פאונים

$\mathcal{H}$-פאון ב-R^n הוא חיתוך של מספר סופי של חצאי מרחב. כלומר, זו קבוצת נקודות $P \subset \mathrm{R}^n$ המוגדרת על ידי $P = P(A,b) = \{x \in \mathrm{R}^n \mid Ax \leq b\}$. נבחין בין פאון כללי לפאון חסום.

$C \subset \mathrm{R}^n$ הוא חרוט אםם $x \in C \Leftarrow \lambda x \in C$ לכל $\lambda \geq 0$. $\mathcal{H}$-חרוט הוא $\mathcal{H}$-פאון מהצורה $P(A,0)$.

$x \in \mathrm{R}^n$ היא קומבינציה קמורה של $x_1, x_2, \ldots, x_m \in \mathrm{R}^n$ אםם קיימים $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathrm{R}$, $\sum_i \lambda_i = 1$, $\forall i \lambda_i \geq 0$, כך ש-$x = \sum_i \lambda_i x_i$. עבור קבוצה $S \subset \mathrm{R}^n$, הסגור הקמור של S, המסומן $conv(S)$, הוא הקבוצה

$$conv(S) = \{x \in \mathrm{R}^n \mid \text{קומבינציה קמורה של נקודות } S \text{ } x\}.$$

הסגור החרוטי של S, המסומן $cone(S)$, הוא הקבוצה

$$cone(S) = \{x \in \mathrm{R}^n \mid \exists x_1, x_2, \ldots, x_m \in S, \exists \lambda_1, \lambda_2, \ldots, \lambda_m \in \mathrm{R}, \forall i, \lambda_i \geq 0, x = \sum_i \lambda_i x_i\}.$$

$P \subset \mathrm{R}^n$ הוא $\mathcal{V}$-פאון אםם $P = conv(V) + cone(Y)$ עבור קבוצות סופיות $V \subset \mathrm{R}^n$ ו-$Y \subset \mathrm{R}^n$. ($+$ מציין סכום מינקובסקי.) P חסום אםם $P = conv(V)$ עבור קבוצה סופית $V \subset \mathrm{R}^n$. P הוא $\mathcal{V}$-חרוט אםם $P = cone(Y)$ עבור קבוצה סופית $Y \subset \mathrm{R}^n$.

משפט 1 (המשפט העיקרי עבור פאונים) P הוא $\mathcal{V}$-פאון אםם P הוא $\mathcal{H}$-פאון.

משפט 2 (המשפט העיקרי עבור חרוטים) $C \subset \mathrm{R}^n$ הוא $\mathcal{V}$-חרוט אםם C הוא $\mathcal{H}$-חרוט.

3 מתי תוכנית ליניארית היא אפשרית

למה 3 (**למת הטלה**) אם $C = P(A,0) \subset \mathrm{R}^n$ אזי $elim_k(C) = \{x - te_k \mid x \in C, t \in \mathrm{R}\}$ גם הוא $\mathcal{H}$-חרוט, ולכן גם $proj_k(C) = elim_k(C) \cap \{x \in \mathrm{R}^n \mid x_k = 0\}$ הוא $\mathcal{H}$-חרוט. בפרט, $elim_k(C) = P(A^{/k}, 0)$, באשר $A^{/k}$ היא המטריצה ששורותיה $A_{i\cdot}$ לכל i עבורו $A_{ik} = 0$, וכן $A_{ik}A_{j\cdot} + (-A_{jk})A_{i\cdot}$ לכל i,j עבורם $A_{ik} > 0$, $A_{jk} < 0$.

הוכחה. שורות $A^{/k}$ הן קומבינציות אי-שליליות של שורות A, ולכן אי-השיוויונות המתקבלים תקפים עבור נקודות C, כלומר, $C \subseteq P(A^{/k}, 0)$. כמו כן, ב-$A^{/k}$ בכל שורה האיבר ה-k-י הוא 0, לכן x_k לא מופיע באי-שיוויונות הללו, ולכן $elim_k(C) \subseteq P(A^{/k}, 0)$.

נראה הכלה גם בכיוון ההפוך: יהי $x \in P(A^{/k}, 0)$. נניח ש-$x_k = 0$ בה"כ. נראה שקיים $t \in \mathrm{R}$ עבורו $x - te_k \in C$. t כזה צריך לקיים $A(x - te_k) \leq 0$, כלומר

$$\max_i \left\{ \frac{1}{A_{ik}} A_{i\cdot}x \mid A_{ik} > 0 \right\} \leq t \leq \min_j \left\{ \frac{1}{-A_{jk}} (-A_{j\cdot})x \mid A_{jk} < 0 \right\}.$$

כדי למצוא t כזה, אגף שמאל צריך להיות קטן או שווה לאגף ימין. אם $A_{ik} > 0$ ו-$A_{jk} < 0$, תנאי זה שקול לתנאי: $(A_{ik}A_{j\cdot} + (-A_{jk})A_{i\cdot})x \leq 0$, שמתקיים עבור x. ∎

4 המשך (מתי תוכנית ליניארית היא אפשרית)

למה 4 (Farkas) תהי $A \in \mathrm{R}^{m \times n}$ ו-$b \in \mathrm{R}^m$. אזי בדיוק אחד משני המקרים הבאים מתקיים: או שקיים $x \in \mathrm{R}^n$ כך ש-$Ax \leq b$; או שקיים $c \in \mathrm{R}^m$ כך ש-$c \geq 0$, $c^TA = 0$, ו-$c^Tb < 0$.

הוכחה. שני המקרים לא יכולים להתקיים יחד, אחרת

$$0 = 0^Tx = (c^TA)x = c^T(Ax) \leq c^Tb < 0$$

וסתירה.

נגדיר $P = P(A,b)$ ו-$Q = P((-b,A),0)$ (באשר $(-b,A)$ היא מטריצה שהעמודה הראשונה שלה היא $-b$ והעמודות הבאות הן עמודות A). קיים $x \in \mathrm{R}^n$ עבורו $Ax \leq b$ אםם Q מכיל נקודה עבורה $x_0 > 0$ (לשם נוחות, נמספר את הקואורדינטות של R^{n+1} מ-0 עד n). נבצע אלימינציה לקואורדינטות 1 עד n על פי למת ההטלה. נקבל $\mathcal{H}$-חרוט $elim_1 elim_2 \cdots elim_n(Q)$. נזכור שב-$D^{/i}$ כל שורה היא קומבינציה חיובית של לכל היותר שתי שורות של D. כלומר, $D^{/i} = C_iD$ באשר C_i מטריצה שאיבריה אי-שליליים (ולכל היותר שניים בכל שורה חיוביים). מכאן,

$$elim_1 elim_2 \cdots elim_n(Q) = P((-b,A)^{/n/n-1/\cdots/1}, 0) = P(C_1C_2\cdots C_n(-b,A), 0).$$

נסמן $C = C_1C_2\cdots C_n$. C מטריצה שאיבריה אי-שליליים. כל אי-השיוויונות במערכת $C(-b,A)x \le 0$ הם מהצורה $\gamma_{i0}x_0 \le 0$, כי עשינו אלימינציה לכל שאר המשתנים. נניח כי $P = \emptyset$ ולכן $Q \subseteq \{x \in \mathrm{R}^{n+1} \mid x_0 \le 0\}$. לכן גם

$$elim_1 elim_2 \cdots elim_n(Q) \subseteq \{x \in \mathrm{R}^{n+1} \mid x_0 \le 0\}.$$

לכן המערכת $C(-b,A) \le 0$ חייבת להכיל אי-שיוויון מהצורה $\gamma_{i0}x_0 \le 0$ עם $\gamma_{i0} > 0$. תהי c^T השורה של C עבורה $c^T(-b,A) = (\gamma_{i0}, 0, 0, \ldots, 0)$. אזי $c^TA = 0$, ו-$c^Tb = -\gamma_{i0} < 0$. ■

5 פאות וקודקודים

ווקטורים $x_1, x_2, \ldots, x_k \in \mathrm{R}^n$ הם בלתי-תלויים אפינית אםם הפיתרון היחיד למערכת המשוואות $\sum_i \alpha_i x_i = 0$, $\sum_i \alpha_i = 0$, הוא $\alpha_i = 0$ לכל i. המימד של פאון P, שנסמנו ב-$dim(P)$, הוא k אםם המספר המירבי של נקודות ב-P, שהן בלתי-תלויות אפינית, הוא $k+1$.

אי-שיוויון ליניארי $c^Tx \le c_0$ תקף עבור פאון $P \in \mathrm{R}^n$ אםם הוא מתקיים לכל $x \in P$. קבוצה

$$F = P \cap \left\{x \in \mathrm{R}^n \mid c^Tx = c_0\right\},$$

באשר $c^Tx \le c_0$ תקף עבור P, היא שפה של P. $0^Tx = 0$, לכן P שפה של עצמו. אם $F \neq \emptyset$ נאמר ש-F תומכת ב-P, ואם גם $F \neq P$ נאמר ש-F שפה ממש.

כל שפה F היא פאון. אם $dim(F) = 0$ אזי F נקראת קודקוד. אם $dim(F) = dim(P) - 1$ אזי F נקראת פאה. נסמן ב-$vert(P)$ את קבוצת הקודקודים של P.

משפט 5 יהי $P \in \mathrm{R}^n$ פאון חסום. אזי:

1. $P = conv(vert(P))$.
2. אם $P = conv(V)$ אזי $vert(P) \subseteq V$.

הוכחה. יהי $P = conv(V)$. אם קיים $v \in V$ שהוא קומבינציה קמורה של ווקטורים מ-$V' = V \setminus \{v\}$, אזי $P = conv(V')$. נראה שאם $v \in V$ איננו קומבינציה קמורה של V' אזי v הוא קודקוד של P.

$$\begin{aligned}
v \notin conv(V') \quad &\Leftrightarrow \quad \nexists t \ge 0 \mid v = V't, 1^Tt = 1 \\
&\Leftrightarrow \quad \nexists t \ge 0 \mid \begin{pmatrix} 1^T \\ V' \end{pmatrix} t = \begin{pmatrix} 1 \\ v \end{pmatrix} \\
&\Leftrightarrow \quad \exists c \mid c^T \begin{pmatrix} 1^T \\ V' \end{pmatrix} \ge 0, c^T \begin{pmatrix} 1 \\ v \end{pmatrix} < 0 \\
&\Leftrightarrow \quad \exists(\beta, -b^T) = c^T \mid b^TV' \le \beta 1, b^Tv > \beta \\
&\Leftrightarrow \quad \exists \beta, b \mid \forall u \in V, u \neq v, b^Tu \le \beta \wedge b^Tv > \beta.
\end{aligned}$$

לכן v הוא קודקוד המוגדר על ידי אי-השיוויון התקף $b^T x \leq b^T v$.

מנגד, כל קודקוד v של P לא ניתן להצגה בתור קומבינציה קמורה של נקודות מ-$P \setminus \{v\}$, כי יש משוואה ליניארית המתקיימת ב-v אבל לא בשאר נקודות P. ■

$x \in P$ היא נקודה קיצונית של P אםם לא קיים וקטור $y \neq 0$ כך שגם $x - y \in P$ וגם $x + y \in P$. הראינו, ש-x נקודה קיצונית אםם היא קודקוד.

6 פיתרונות בסיסיים

משפט 6 נסמן $P = \{x \mid Ax = b, x \geq 0\}$. נניח שהתוכנית $\min c^T x \text{ s.t. } x \in P$ חסומה. אזי לכל $x \in P$ קיים קודקוד x' של P כך ש-$c^T x' \leq c^T x$.

מסקנה 7 אם התוכנית חסומה קיים פיתרון אופטימלי x^* שהוא קודקוד.

הערה. לתוכנית בהצגה קנונית אין בהכרח קודקוד, אפילו אם היא חסומה. למשל,

$$\min\{x_2 \mid x \in \mathrm{R}^2 \wedge x_2 \geq 1 \wedge -x_2 \geq -2\}.$$

הוכחת המשפט 6. יהי $x \in P$. נניח ש-x איננו קודקוד. כלומר, קיים $y \neq 0$ עבורו $x - y, x + y \in P$. לכן חייב להתקיים $A(x - y) = A(x + y) = b$, $Ay = 0$. בנוסף, אם $x_j = 0$ אזי גם $y_j = 0$ כי $x_j - y_j \geq 0$ וגם $x_j + y_j \geq 0$. נניח בה"כ ש-$c^T y \leq 0$, אחרת נחליף את y ב-$-y$. אם $c^T y = 0$, אזי גם y וגם $-y$ עומדים בדרישות שלנו. נבחר אחד מהם בו יש איבר שלילי.

נתבונן ב-$x + \lambda y$ עבור $\lambda > 0$. $c^T(x + \lambda y) = c^T x + \lambda c^T y \leq c^T x$.

אם $y \geq 0$, אזי $x + \lambda y$ פיתרון אפשרי לכל λ, ועל-פי הבחירה של y מתקיים $c^T y < 0$. לכן התוכנית לא חסומה בניגוד להנחה. נניח, אם כן, שיש אינדקס j עבורו $y_j < 0$. ניקח

$$\lambda = \min_{j \mid y_j < 0} \left\{ \frac{x_j}{-y_j} \right\}.$$

נקבל $x + \lambda y \in P$. בנוסף, ל-$x + \lambda y$ יש יותר איברים שהם 0 מאשר ל-x. נחליף את x ב-$x + \lambda y$ ונחזור על התהליך. אפשר לעשות זאת לכל היותר n פעמים. ■

משפט 8 נסמן $P = \{x \mid Ax = b, x \geq 0\}$. עבור $x \in P$ תהי A_x המטריצה שעמודותיה עמודות A עבורן האיבר המתאים ב-x שונה מ-0. אזי x קודקוד אםם עמודות A_x בלתי-תלויות ליניארית.

הוכחה. נניח ש-x איננו קודקוד. אזי קיים $y \neq 0$ כך ש-$x - y, x + y \in P$. $Ay = 0$, לכן עמודות A_y תלויות ליניארית. אבל, כפי שראינו, אם $x_j = 0$ אזי גם $y_j = 0$ ולכן עמודות A_y הן תת-קבוצה של עמודות A_x, ולכן גם עמודות A_x תלויות ליניארית.

נניח שעמודות A_x תלויות ליניארית, אזי קיים $y \neq 0$ כך ש-$A_x y = 0$. נרחיב את y ל-R^n על ידי הוספת איברים עם ערך 0 במקומות המתאימים לעמודות A שהושמטו ב-A_x. נקבל שקיים $y \in \mathrm{R}^n$, $y \neq 0$, שמקיים $y_j = 0$ אם $x_j = 0$. באופן דומה להוכחת המשפט הקודם, יש $\lambda > 0$ עבורו $x - \lambda y, x + \lambda y \in P$. ■

יהי x קודקוד של $P = \{x \mid Ax = b, x \geq 0\}$, באשר $A \in \mathrm{R}^{m \times n}$. נניח $rank(A) = m$ (אחרת או שיש אילוצים מיותרים, או שאין פיתרון). נסמן $B = \{j \mid x_j > 0\}$. נניח בתחילה ש-$|B| = m$. נסמן ב-A_B את המטריצה שעמודותיה הן עמודות A המתאימות לאינדקסים ב-B. באופן דומה נסמן ב-x_B את הווקטור המתקבל מ-x על ידי לקיחת האיברים עם אינדקס ב-B. ראינו שעמודות A_B בלתי-תלויות ליניארית. נסמן $N = \{1, 2, \ldots, n\} \setminus B$. אזי אפשר להציג את x על ידי $x_B = A_B^{-1} b > 0$, $x_N = 0$. אם $|B| < m$ אפשר להוסיף עמודות בלתי-תלויות של A ולקבל הצגה זהה עם $|B| = m$, אבל אז חלק מאיברי x_B יהיו 0. נקרא ל-B בסיס, ול-x הנתון בצורה זו פיתרון אפשרי בסיסי (שימו לב שאפשר לפעמים לייצג קודקוד בתור מספר פיתרונות אפשריים בסיסיים שונים, בהתאם לאופן בו משלימים לבסיס). כמו כן, שימו לב שלא כל בסיס מגדיר קודקוד (כי $A_B^{-1} b$ לא בהכרח אי-שלילי).

7 שיטת הסימפלקס (Dantzig 1947)

נתחיל מבסיס B ופיתרון אפשרי בסיסי x. אפשר להניח שהתוכנית נתונה כך ש-$b \geq 0$ ו-A_B היא מטריצת היחידה (אחרת נכפיל את האילוצים עם אגף ימין שלילי ב--1 ונבצע אלימינצית גאוס על A_B).

$$\begin{aligned} \min \quad & c_B^T x_B + c_N^T x_N \quad \text{s.t.} \\ & A_B x_B + A_N x_N = b \\ & x_B, x_N \geq 0. \end{aligned}$$

$$x_B = A_B^{-1} b - A_B^{-1} A_N x_N$$

ולכן

$$c_B^T x_B + c_N^T x_N = c_B^T \left(A_B^{-1} b - A_B^{-1} A_N x_N \right) + c_N^T x^N$$

$$= \quad c_B^T A_B^{-1} b + \left(c_N^T - c_B^T A_B^{-1} A_N \right) x_N.$$

(כלומר, אפשר לכתוב את פונקצית המטרה כך שלא תהיה תלוייה ב-x_B.)

נסמן

$$\tilde{c}^T = c_N^T - c_B^T A_B^{-1} A_N.$$

אם קיים j עבורו $\tilde{c}_j < 0$, אזי הגדלה של x_j תקטין את פונקצית המטרה. בתמורה נצטרך לשנות איברים של x_B כדי לשמור על פיזיביליות. זיכרו כי A_B היא מטריצת יחידה, לכן כל איבר ב-x_B משפיע על אילוץ אחד, ולכן קל לתקן את הפיתרון כך שיהיה פיזיבילי. כמו כן, אם כל האיברים ב-x_B דורשים הגדלה, אזי אפשר להגדיל אותם ואת x_j בלי הגבלה, ולכן התוכנית לא חסומה במקרה זה. לכן, נניח שחלק מהאיברים ב-x_B דורשים הקטנה. נגדיל את x_j ונשנה בהתאם את האיברים ב-x_B עד שאחד מהם לפחות יתאפס. נחליף בבסיס את האיבר שהתאפס ב-j.

אם אין j עבורו $\tilde{c}_j < 0$, אזי הפיתרון האפשרי הבסיסי שלנו הוא אופטימלי.

נדגים את אלגוריתם הסימפלקס על תוכנית הקלט הבאה:

$$\begin{aligned} \min \quad & 19x_1 - 13x_3 + x_4 + 3x_6 \quad \text{s.t.} \\ & 3x_1 - 4x_3 + x_4 + 4x_5 = 6 \\ & -4x_1 + x_2 + x_3 + 2x_5 = 1 \\ & 5x_1 - 2x_3 - x_5 + x_6 = 1 \\ & x \geq 0 \end{aligned}$$

מקובל לרשום את התוכנית בטבלה כזו:

$$\begin{array}{|cccccc|c|} \hline 19 & 0 & -13 & 1 & 0 & 3 & 0 \\ \hline 3 & 0 & -4 & 1 & 4 & 0 & 6 \\ -4 & 1 & 1 & 0 & 2 & 0 & 1 \\ 5 & 0 & -2 & 0 & -1 & 1 & 1 \\ \hline \end{array}$$

השורה העליונה מכילה את מקדמי פונקצית המטרה, באשר האיבר בפינה הימנית הוא מינוס האיבר החופשי. השורות הבאות מכילות את מקדמי האילוצים, באשר העמודה הימנית היא צד ימין של המשוואות.

נבחר לבסיס B את $\{2,4,6\}$. שימו לב שתת-המטריצה A_B נתונה כבר כמטריצת יחידה. פונקצית המטרה לא נתונה בפורמט הדרוש, כי היא תלוייה במשתנים הבסיסיים. נתקן זאת על ידי החסרת שורת האילוצים הראשונה + $3 \times$ שורת האילוצים השלישית מפונקצית המטרה (שימו לב לנוחות שבייצוג האיבר החופשי על ידי ההופכי). נקבל טבלה כזו:

1	0	-3	0	-1	0	-9
3	0	-4	1	4	0	6
-4	1	1	0	2	0	1
5	0	-2	0	-1	1	1

כלומר, פונקצית המטרה המקורית שקולה לפונקציה: $x_1-3x_3-x_5+9$ בתחום האופטימיזציה.

כעת נבחר איבר להכניס לבסיס. צריך עמודה שבה המקדם בפונקצית המטרה שלילי, למשל עמודה 5 (נכניס לבסיס את האינדקס 5). צריך להוציא מהבסיס אחד האיברים שישנם כרגע. הגדלה של x_5 מחייבת שינוי ערכי המשתנים הבסיסיים x_2, x_4, x_6 על מנת לשמור על פיזיביליות. כל אילוץ (שורה בטבלה) מכתיב מגבלות עבור משתנה בסיסי אחר. על פי האילוץ הראשון, ניתן להגדיל את x_5 עד $\frac{6}{4}$ מבלי להפוך את x_4 לשלילי. האילוץ השני מכתיב הגדלה של לכל היותר $\frac{1}{2}$ מבלי להפוך את x_2 לשלילי. האילוץ השלישי אינו מכתיב דבר, כי המקדם של x_5 הוא שלילי, ולכן הגדלה של x_5 דורשת הגדלה גם של x_6. שימו לב, שאם כל המקדמים של x_5 הם שליליים, התוכנית לא חסומה.

האילוץ הקריטי הוא, אם כן, האילוץ השני. האיבר בעמודה 5 ובשורה 2 (של האילוצים) נקרא פיבוט. על פי אילוץ זה, נגדיל את x_5 ל-$\frac{1}{2}$ ובתמורה נקטין את x_2 ל-0. בבסיס, נחליף את 2 ב-5. עדכון הטבלה בהתאם נקרא צעד פיבוט. נשתמש בשורת האילוץ השני כדי לבצע את העדכון. נוסיף לשורת מקדמי פונקצית המטרה $\frac{1}{2}$ משורה זו (על מנת לאפס את המקדם של x_5). נעדכן את שורות האילוצים כדלקמן: משורה 1 נחסיר פעמיים שורה 2. לשורה 3 נוסיף חצי שורה 2. את שורה 2 נחלק ב-2. נקבל את הטבלה הבאה:

-1	$\frac{1}{2}$	$-\frac{5}{2}$	0	0	0	$-\frac{17}{2}$
11	-2	-6	1	0	0	4
-2	$\frac{1}{2}$	$\frac{1}{2}$	0	1	0	$\frac{1}{2}$
3	$\frac{1}{2}$	$-\frac{3}{2}$	0	0	1	$\frac{3}{2}$

על מנת להשלים את תאור האלגוריתם, צריך להגדיר במדוייק את הכלל לפיו בוחרים את העמודה להכניס לבסיס (מבין כל העמודות האפשריות), וכן צריך להגדיר את הכלל לפיו בוחרים את השורה הקובעת איזה איבר מוציאים מהבסיס, במקרה שיש יותר מאילוץ קריטי אחד. הכללים האילו נקראים ביחד כלל הפיבוט. הוצעו בספרות כללים שונים. הכלל המקובל לבחירת עמודה הוא לבחור את העמודה שבה מקדם פונקצית המטרה הוא השלילי ביותר.

הערות:

1. אם יש כבר בבסיס איברים שהם 0, יכול להיות שלא נוכל להגדיל את x_j ואז ייתכן שניכנס ללולאה אינסופית שבה נחליף איבר k ב-j ו-j ב-k לסירוגין. זה יכול לקרות בכללי פיבוט מקובלים, ומימוש האלגוריתם מחייב הוספת בדיקה ללולאה אינסופית ויציאה ממנה.
2. יש כלל פיבוט (Bland 1977) שמונע לולאות: נבחר עמודה j מינימלית (האינדקס מינימלי), ושורה ℓ עבורה האינדקס של האיבר k שייצא מהבסיס הוא מינימלי. באופן מעשי זה כלל גרוע, שאינו מוביל לביצועים טובים של האלגוריתם.
3. לרוב הכללים אפשר להוכיח שבמקרה הגרוע ביותר נבצע מספר אקספוננציאלי של צעדי פיבוט.
4. באופן מעשי, מספר צעדי הפיבוט בדר"כ ליניארי במספר האילוצים m וכנראה גדל לוגריתמית במספר המשתנים n.
5. לא ידוע אם יש כלל פיבוט עבורו מספר צעדי הפיבוט חסום פולינומית במימדי הבעייה. שאלה זו קשורה בהשערת Hirsch: האם זה נכון שעבור סידור כלשהו של m על-מישורים ב-$\mathbb{R}^d$, אורך המסלול הקצר ביותר בין כל שני קודקודים בסידור הוא לכל היותר $m-d$ (צעד במסלול מעביר אותנו מקודקוד לקודקוד שכן, שמחובר אליו בצלע של הסידור).
6. גיל קלעי הראה כלל פיבוט רנדומי שמבטיח מספר תת-אקספוננציאלי של צעדי פיבוט. זה גם החסם הטוב ביותר הידוע לגבי השערת Hirsch.

נותרנו עם השאלה איך למצוא בסיס התחלתי. נניח שתוכנית הקלט היא:

$$\begin{aligned} \min \quad & c^T x \quad \text{s.t.} \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

באשר A מטריצה $m \times n$, שדרגתה m. נוסיף משתנים $x'_1, x'_2, \ldots, x'_m$, ונתבונן בתוכנית:

$$\begin{aligned} \min \quad & \textstyle\sum_{i=1}^m x'_i \quad \text{s.t.} \\ & Ax + Ix' = b \\ & x, x' \geq 0 \end{aligned}$$

באשר I מטריצת יחידה $m\times m$. הפיתרון האופטימלי הוא 0 אסם התוכנית המקורית אפשרית. כאן אפשר לקחת את העמודות של x' כבסיס, ואת $x'=b$ כפיתרון אפשרי בסיסי. נמצא אופטימום, ואם הוא 0 אפשר להשתמש בפיתרון הבסיסי האופטימלי כפיתרון בסיסי אפשרי בתוכנית המקורית.

8 דואליות

נתונה התוכנית הליניארית הבאה בהצגה סטנדרטית, שנסמנה (P):

$$z = \min \quad c^T x \quad \text{s.t.}$$
$$Ax = b$$
$$x \geq 0$$

נחפש חסם תחתון על z. נחפש חסם באופן הבא: ניקח קומבינציה ליניארית של שורות A, שחסומה על ידי c, ונסתכל על המשוואה התקפה המתקבלת מקומבינציה זו. בכתיב מטריצות, ניקח ווקטור y כלשהו עבורו $y^T A \leq c^T$, ונסתכל על המשוואה $y^T b = (y^T A)x$. $x \geq 0$ ולכן אי-השיוויון $y^T b = (y^T A)x \leq c^T x$ מתקיים לכל x אפשרי, ובפרט $y^T b \leq z$. החסם התחתון הטוב ביותר, שניתן לקבל בצורה זו, הוא $w = \max\{y^T b \mid y^T A \leq c^T\}$. זו תוכנית ליניארית. נסמנה (D). היא נקראת התוכנית הדואלית ל-(P).

בדיון לעיל הוכחנו את המשפט הבא:

משפט 9 (**דואליות חלשה**) $z \geq w$.

טענה 10 התוכנית הדואלית ל-(D) היא (P).

הוכחה.

$$\begin{aligned}
\max\{b^T y \mid A^T y \leq c\} &= \min\{-b^T y \mid A^T y + Is = c,\ s \geq 0\} \\
&= \min\{-b^T y^+ + b^T y^- \mid A^T y^+ - A^T y^- + Is = c,\ y^+, y^-, s \geq 0\}.
\end{aligned}$$

הדואלי הוא:

$$\max\{c^T(-x) \mid A(-x) \leq -b,\ -A(-x) \leq b,\ I(-x) \leq 0\} = \min\{c^T x \mid Ax = b,\ x \geq 0\}.$$

■

משפט 11 (**דואליות חזקה**) אם (P) או (D) אפשריות, אזי $w = z$.

הוכחה. די להראות $z \leq w$. נניח בה"כ ש-(P) אפשרית. אם (P) לא חסומה, אזי על פי דואליות חלשה $z = w = -\infty$. אחרת, יהי x^* פיתרון אופטימלי של (P), כלומר

$$Ax* = b,\ x^* \geq 0,\ c^T x^* = z.$$

נראה שקיים y עבורו $A^T y \leq c$ ו-$b^T y \geq z$. נניח בשלילה שאין y כזה. אזי למערכת אי-השיוויונות

$$\begin{pmatrix} A^T \\ -b^T \end{pmatrix} y \leq \begin{pmatrix} c \\ -z \end{pmatrix}$$

אין פיתרון. לכן, על פי הלמה של Farkas, קיים

$$\begin{pmatrix} x \\ \lambda \end{pmatrix} \geq 0$$

כך ש-

$$\begin{pmatrix} A^T \\ -b^T \end{pmatrix}^T \begin{pmatrix} x \\ \lambda \end{pmatrix} = 0$$

ו-

$$\begin{pmatrix} c \\ -z \end{pmatrix}^T \begin{pmatrix} x \\ \lambda \end{pmatrix} < 0.$$

כלומר, יש $x \geq 0,\ \lambda \geq 0$ עבורם $Ax = \lambda b$ ו-$c^T x < \lambda z$. נבחין בין שני מקרים:

מקרה 1: $\lambda = 0$. אזי יש x כך ש-$Ax = 0$ ו-$c^T x < 0$. אבל אז לכל $\alpha > 0$, $x^* + \alpha x$ הוא פיתרון אפשרי, וערכו $c^T(x^* + \alpha x) < z$, וזו סתירה.

מקרה 2: $\lambda \neq 0$. אז $\frac{1}{\lambda}x$ פיתרון אפשרי שערכו קטן ממש מ-z, ושוב סתירה. ■

9 כלל העודפיות המשלימה

נתונה תוכנית ליניארית בהצגה סטנדרטית (נסמנה (P)):

$$\begin{aligned} \min \quad & c^T x \quad \text{subject to} \\ & Ax = b \\ & x \geq 0. \end{aligned}$$

התוכנית הדואלית (נסמנה (D)) היא:

$$\begin{aligned} \max \quad & y^T b \quad \text{subject to} \\ & y^T A \leq c^T \end{aligned}$$

יהי x פיתרון אפשרי ל-(P), ויהי y פיתרון אפשרי ל-(D). על פי דואליות חלשה, $c^Tx \geq y^Tb$. ההפרש $c^Tx - y^Tb$ נקרא פער הדואליות. פער הדואליות הוא 0 אםם x פיתרון אופטימלי ל-(P) ו-y פיתרון אופטימלי ל-(D).

נסמן $s^T = c^T - y^TA$. אזי

$$\begin{aligned} c^Tx - y^Tb &= c^Tx - y^TAx \\ &= (c^T - y^TA)x \\ &= s^Tx. \end{aligned}$$

משפט 12 (**כלל נעודפיות המשלימה**) יהי x פיתרון אפשרי ל-(P), ויהי y פיתרון אפשרי ל-(D). אזי התנאים הבאים שקולים:

1. אילו פיתרונות אופטימליים.
2. $s^Tx = 0$.
3. $s_jx_j = 0$ לכל j.
4. אם $s_j > 0$, אזי $x_j = 0$.

הוכחה. נניח שמתקיים תנאי 1. אזי לפי דואליות חזקה, $c^Tx = y^Tb$. אבל $y^TAx + s^Tx = c^Tx$. נציב את השיוויון הקודם באגף ימין ונקבל $y^TAx + s^Tx = y^Tb = y^TAx$. מכאן, $s^Tx = 0$, כלומר, מתקיים תנאי 2. אבל $x, s \geq 0$, וכן כל איבר בסכום s^Tx הוא 0, כלומר מתקיים 3. תנאי 3 גורר האופן מיידי את תנאי 4. כל הגרירות שהוכחנו מתקיימות גם בכיוון ההפוך, ולכן המשפט נכון. ■

לדוגמה, נתבונן בבעייה של מציאת זרימת מקסימום ממקור s לבור t ברשת זרימה. ניתן לבטא את הבעייה בתוכנית ליניארית (המשתנים הם ערכי הזרימה על הקשתות). התוכנית הדואלית דורשת מתן משקלות אי-שליליים על הקשתות, כך שהמרחק מ-s ל-t יהיה לפ-חות 1, והסכום של המשקלות, משוקללים על פי הקיבולים, יהי מינימלי. (שימו לב, זו לא הצגה סטנדרטית, אבל אפשר לתרגם את כלל העודפיות המשלימה להצגות אחרות.) לתוכנית הדואלית יש פיתרון בשלמים, המתאים לחתך מינימום (נשים משקל 1 על קשתות החתך, ו-0 על הקשתות האחרות). אם הקיבולים שלמים, אזי גם לבעיית הזרימה יש פיתרון בשלמים. מה אומר כלל העודפיות המשלימה? עבור זוג פיתרונות אופטימליים, אם משתנה חיובי ממש באחת התוכניות, אזי האילוץ המתאים בתוכנית האחרת מתקיים באופן הדוק. כלומר, כל מסלול שיש עליו זרימה חיובית הוא מסלול קל ביותר מ-s ל-t בדואלי, וכל קשת שיש עליה משקל חיובי בדואלי, הזרימה עליה שווה לקיבול.

10 הערות נוספות

שיטת הסימפלקס לא נותנת (לפחות על פי נידוע לנו עד עתה) אלגוריתם פולינומי לתכנון ליניארי. ישנם אלגוריתמים כאלה. בשנת 1979 הוכיח Khachian שאלגוריתם האליפסואיד (שפותח קודם לכן עבור תכנון קמור) פותר תוכניות ליניאריות בזמן פולינומי. התגלית זכתה לתהודה ציבורית גדולה (הניו-יורק טיימס הכתיר אותה בכותרת "הספוטניק המתמטי"). מהר מאוד התברר שלאלגוריתם האליפסואיד אין משמעות מעשית, כי זמן הריצה שלו גרוע מאוד בהשוואה לסימפלקס על בעיות "מהחיים". אף על פי כן, יש לאלגוריתם האליפסואיד חשיבות תיאורטית רבה. מתברר, שהאלגוריתם מסוגל לפתור בזמן פולינומי מגוון רחב של בעיות אופטימיזציה, בפרט, מינימיזציה של פונקציה קמורה על תחום קמור הנתון על ידי אוב הפרדה (כלומר, שגרה שבהינתן נקודה במרחב מסוגלת לזהות אם הנקודה בתוך התחום או לא, ואם לא היא מחזירה על-מישור המפריד בין הנקודה לתחום). (בתיאור זה נשמטו מספר מגבלות טכניות.) בשנת 1986 פרסם Karmarkar אלגוריתם פולינומי אחר לחלוטין לתכנון ליניארי. בעקבות זאת, פותחו אלגוריתמים נוספים דומים. האלגוריתמים נוסח Karmarkar נקראים interior point methids, ובאופן מעשי הם ברי-תחרות לסימפלקס. ניתוח זמן הריצה שלהם במקרה הגרוע הוא פולינומי במספר המשתנים, וליניארי במספר הביטים הדרוש לרשום את המקדמים של התוכנית הליניארית. יש לציין שהתלות של זמן הריצה בגודל הייצוג של התוכנית הליניארית קיים גם עבור אלגוריתם האליפסואיד. השאלה אם קיים אלגוריתם שמספר פעולות החשבון שהוא מבצע הוא תמיד פולינומי במספר המשתנים ובמספר האילוצים, ללא תלות בגודל המקדמים, היא שאלה פתוחה מרכזית בתחום.