א. שמיר
עתק ראשון

ניתוח מערכות

א. שמיר

הקדמה

הספר מיועד לשמש בהוראת ניתוח מערכות למהנדסים. לכן מושם בו הדגש על הבנת הטכניקות ומשמעותן, תוך ויתור מסוים על הוכחות מתמטיות.

מהדורה ראשונה הופיעה בשנת 1971, ומהדורה שניה שכללה מספר שיפורים הודפסה שנים אחדות לאחר מכן. הספר שימש במשך כל השנים בהוראת המקצוע "ניתוח מערכות" בפקולטה להנדסה אזרחית בטכניון. במהדורה הנוכחית נוסף הפרק השביעי, "תכנות דינמי", וכן הרחבות בפרקים אחרים.

הנושא של ניתוח מערכות מקיף גישה מחשבתית לבעיות ושיטות לפתרונן. בספר יש טיפול בשיטות המתמטיות ולכן הוא באפיו סכני. הגישה לבעיות ובחירת הדרך לתקיפתן, שהם החלקים היותר קשים והיותר חשובים של הנושא, אלה נלמדים בהרצאות ובתרגול.

הספר מיועד לסטודנט הלומד את הנושא, וכן כספר עזר למהנדס העוסק בניתוח מערכות.

א. שמיר

1980

תוכן הענינים

תוכן עניינים (המשך)

עמוד

תוכן עניינים (המשך)

עמוד

תוכן ענינים (המשך)

עמוד

תוכן ענינים (המשך)

עמוד

פרק ראשון

מבוא

המושג "ניתוח מערכות" משמש בשטחי עבודה שונים לתיאור רבדים שונים. בנושא המחשבים משמש המושג לתיאור הפעילויות הקשורות בתיכנון ותכנות של מערכות תוכנה (software) לשימושים מרעיים או מינהליים. מהנדס האלקטרוניקה מתכוון בו לניתוח פעולתן של מערכות אלקטרוניות, ואילו העוסקים בבקרה מתארים בו את ניתוח התנהגותן של מערכות בקרה-מכניות, פניאומטיות או חשמליות.

כאן הכוונה לניתוח של מערכות הנדסיות במטרה לשפר את התכן ו/או את התפעול שלהן. במובן הצר ניתן להגדיר ניתוח זה כתהליך שבו עבור בעית הנדסית נתונה בוחרים מבין כל האלטרנטיבות האפשריות את זו שהיא הטובה ביותר לפי קריטריון נבחר. באשר מסתפקים במובן צר זה קשה לומר מה ההבדל בין ניתוח מערכות לבין חקר ביצועים. במידה רבה זו שאלה של שימוש במונחים כלליים לתיאור תכנים מסויימים. אנשים עשויים להתכוון לתכנים שונים בהשתמשם במושגים שווים, ולהיפך - להתכוון לתכנים שווים בהשתמשם בשמות כלליים שונים. מעבר לכל ספק הוא שבין אם משתמשים במושג "ניתוח מערכות" או "חקר ביצועים" שמור בהם מקום נכבד לטכניקות של אופטימיזציה.

למרות שעיקרם של הפרקים הבאים מוקדש לטכניקות אלה, לא מופיעה המלה "אופטימיזציה" בשם הספר. הסיבה העיקרית לכך היא שהכוונה היתה להרגיש את האספקטים הכלליים יותר של ניתוח מערכות הנדסיות, עליהם נעמוד להלן. כמו כן מדגיש הרבר את העובדה שהטיפול בטכניקות האופטימיזציה נעשה תוך בסיון לאזן בין פשטות ההצגה לבין ההעמקה המתימטית, כלומר, יש לעתים ויתור על הוכחות מתימטיות כלליות לטובת הדגמה אינטואיטיבית יותר של הטכניקות. הדבר נעשה מהכרה שספר זה מיועד בעיקר למהנדסים העוסקים בניתוח מערכות, ופחות לאנשים שעיסוקם בטכניקות האופטימיזציה עצמן.

ניתוח מערכות הנדסיות הוא צירוף של גישה כללית מסויימת לבעיות הנדסיות ושל אוסף טכניקות לטיפול בבעיות אלה ולפתרונן. מתר זו שיטה איכותית-מחשבתית, ומאידך - שיטה כמותית, המסתמכת על תורת הכלכלה, סטטיסטיקה, מתמטיקה ותורת התכנות המתימטי. כמו כן קשור הנושא קשר אמיץ לשימוש במחשבים, ויכול היה להתפתח לרמתו הנוכחית רק לאחר שהמחשבים הפכו כלי מקובל וזול יחסית.

השיטה שהיתה מקובלת בהנדסה, ועדיין הינה נפוצה מאוד, היא להציע ולבדוק מספר אלטרנטיבות לפתרון בעיה הנדסית. בתהליך זה מפעיל המהנדס את דמיונו היוצר, את נסיונו ואת הידע שלו, מציע מספר פתרונות אפשריים ובודק את טיבו של כל אחד לאור קריטריונים מסויימים שהוא מציב, אשר אחד הנפוצים בהם הוא המחיר. אין לוותר על הדמיון היוצר, הנסיון ההנדסי והידע של המהנדס גם כאשר מתכוונים להפעיל טכניקות אופטימיזציה מתוחכמות ולהשתמש במחשב. להיפך, הצורך לבנות מודל מתימטי של המערכת ההנדסית-כלכלית מחייב הבנה מעמיקה של כל צדדיה. מה שיכול המהנדס לכלול בצורה ספציפית בתוך אלטרנטיבה כלשהי המוצעת על ידו כפתרון אפשרי אחד של הבעיה ההנדסית, עליו לנסח כחוק כללי עבור כל הפתרונות האפשריים של בעית אופטי-מיזציה. הניסוח של מודל מתימטי למערכת הנדסית הוא שלב בו נחקרת המערכת בצורה המעמיקה ביותר, תוך יישום הנסיון והידע ההנדסיים.

ניסוח בעית האופטימיזציה הוא שלב קריטי, שכן לאחר שנוסח מודל האופטימיזציה של בעיה הנדסית על כל מרכיביה קבוע הפתרון באופן חד-משמעי. מכאן אין להסיק שבכל מקרה נדע גם למצוא פתרון זה בצורה מפורשת. להיפך, אלגוריתמים מתימטיים המבטיחים את מציאתו של הפתרון האופטימלי קיימים רק עבור בעיות אופטימיזציה בעלות מבנה מיוחד. במקרים אחרים אנו פונים לטכניקות פחות אלגנטיות אשר אולי אינן מבטיחות כי הפתרון האופטימלי יימצא, אבל לפחות מבטיחות כי באמצעותן יש סיכוי טוב למצוא פתרון שיהיה טוב מכל מה שאפשר למצוא בשיטה אחרת. השימוש במחשב גם מבטיח כי לפחות ניתן לבדוק תמורת סכום סביר מספר אלטרנטיבות גדול בהרבה ממה שניתן היה לבדוק בעבודה ידנית תמורת אותו סכום.

התהליך של ניסוח מודל לאופטימיזציה של מערכת הנדסית ופתרונו הוא ברוב המקרים איטרטיבי. פרט לבעיות פשוטות ביותר יש לעשות הנחות והזנחות מסויימות בעת בנית המודל. באופן תיאורטי יש לנסח את בעית האופטימיזציה תוך נאמנות מירבית לתיאור מדויק של המערכת וכלי שים-לב למורכבותו של המודל המתימטי המתקבל, וליכולתנו או אי יכולתנו לפתור אותו. זה מבטיח אמנם מודל נאמן אבל עם זאת אולי חסר פתרון. יש לכן לפשט את המודל עד שמגיעים למצב בו הוא ניתן לפתרון, ועם זאת התוצאות המתקבלות אינן מעוותות עד כדי כך שהן חסרות משמעות מעשית. מציאת המודל המתאים היא תהליך איטרטיבי בו מניחים הנחות מסויימות, מנסים לפתור, מתאימים את ההנחות ער שמשיגים פתרון, בודקים את הפתרון, מתקנים את המודל במידה שהתוצאות אינן מעשיות, וחוזר חלילה. על כל פנים, יש להישמר מהנטיה להתאים מראש את ניסוח הבעיה, תוך עיוות המציאות, לטכניקה מסויימת של פתרון, רק משום שזו מוכרת היטב או שקיימת תוכנית מחשב לביצועה.

שתי מטרות עיקריות ללימור הנושא של ניתוח מערכות הנדסיות, והן:

1. לפתח את הגישה לניסוח הבעיות של פתרון אופטימלי לבעיות הנדסיות, תוך הדגשת הרקע בו מתעוררות הבעיות ההנדסיות - האקראיות של תופעות הטבע והחברה, האספקטים הכלכליים, החברתיים, האסתטיים והאחרים של פרויקטים הנדסיים.

2. ללמוד את הטכניקות לפתרון בעיות אלה.

את המטרה הראשונה ניתן להשיג בעיקר על ידי עשייה, ופחות על ידי לימוד פורמלי. שני הפרקים הבאים מוקדשים לכמה אספקטים כלליים השייכים למטרה הראשונה - רקע כלכלי ותורת ההחלטות. שאר הספר מוקדש ללימוד כמה טכניקות אופטימיזציה. השגתה של המטרה הראשונה תוגשם בעת לימוד פרקים אלה, על ידי טיפול בדוגמאות הלקוחות מבעיות הנדסיות ממשיות.

הטכניקות הזוכות לטיפול בפרקים אלה הן: שיטות קלסיות, תכנות ליניארי ותכנות דינמי. זהו חלק קטן בלבד ממכלול הטכניקות של התכנות המתמטי. הפרק הרביעי מציג את הבעיה הכללית של התכנות המתמטי, את סיווג הבעיות לפי הרכבן ומבנן, ומונה כמה מהטכניקות הנוספות המקובלות לפתרון. בפרקים 3 ו-4 נדונות בעיות החלטה רב-קריטריוניות, ומוצגות בקצרה מספר גישות לפתרונן.

מ ק ו ר ו ת

Aguilar, R.J., "Systems Analysis and Design in Engineering Architecture, Construction and Planning", Prentice-Hall, 1973.

De Neufville, R. and Stafford, J.H., "Systems Analysis for Engineers and Managers", McGraw-Hill, 1971.

Stark, R.M. and Nicholls, R.L., "Mathematical Foundations of Design", McGraw-Hill, 1972.

פרק שני

רקע כלכלי

ייצור (Production) מוגדר כפעילות אשר מטרתה התמדת משאבים (Resources) המופיעים בצורה, במקום ובזמן מסויימים למשאבים בצורה ו/או במקום ו/או בזמן אחרים היותר מתאימים לצריכה, או לשלב אחר בייצור.

המשאבים ה"נכנסים" לייצוד נקראים תשומות (Inputs) ואילו אלה היוצאים ממנו - תפוקות (Outputs).

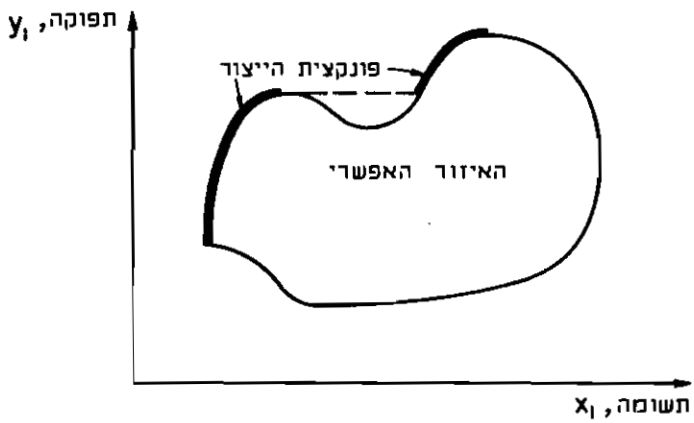
2.1 פונקצית הייצוד

מערכת משתמשת בתשומות $\underline{x} = (x_1,\ldots,x_n)$ כדי לייצר את התפוקות $\underline{y} = (y_1,\ldots,y_m)$. קוראים לזוג (x,y) אפשרי (Feasible) או נקודה אפשרית אם הוא אפשרי מבחינה טכנולוגית. קבוצת כל הנקודות האפשריות מהוות את האיזור האפשרי (Feasible Region). היותו אפשרי אינו עושה את הייצור של $\underline{y}$ בעזרת $\underline{x}$ לרצוי או יעיל. מכל הנקודות האפשריות מעניינות אותנו הנקודות היעילות (Efficient Points) המוגדרות כדלקמן:

נניח שקיימות שתי נקודות אפשריות $(\underline{x}_1;\underline{y}_1)$ $(\underline{x}_1;\underline{y}_2)$ המשתמשות באותן תשומות $\underline{x}_1$ כדי לייצר תפוקות שונות - $\underline{y}_1$ ו-$\underline{y}_2$ בהתאמה. אם כל רכיב של הוקטור $\underline{y}_2$ גדול יותר מהרכיב המתאים לו בוקטור y_1 אזי הנקודה $(\underline{x}_1;\underline{y}_2)$ יעילה יותר מהנקודה $(\underline{x}_1;\underline{y}_1)$. באופן דומה הנקודה $(\underline{x}_3;\underline{y}_3)$ יעילה יותר מהנקודה $(\underline{x}_4;\underline{y}_3)$. אם כל רכיב של $\underline{x}_4$ קטן יותר מהרכיב המתאים של $\underline{x}_3$. הביטויים "גדול יותר" ו-"קטן יותר" משמשים כאן במובן הטכנולוגי גרידא, כלי קשר לערך כלכלי. כדוגמה פשוטה תשמש הבעיה של אגירת מי שטפונות באמצעות סכר אגירה עשוי עפר. התשומה היא נפח העפר שבסכר והתפוקה - כמות המים הנאגרת. בהנחה שיש אופנים שונים לכנות את הסכר, הרי שבין כל הסכרים בעלי נפח עפר שווה יהיה הסכר היעיל זה שייצור אגם כעל הנפח הגדול ביותר. בין כל הסכרים המסוגלים לאגום נפח מים נתון יהיה הסכר

היעיל ביותר זה שיהיה בעל נפח העפר הקטן ביותר. הגדרת היעילות נעשתה בלי להתחשב בערך הכלכלי של המים הנאגרים על ידי הסכר, כלומר, בלי שים לב האם "כדאי" בכלל לבנות את הסכר היעיל הנדון.

הנקודה $(\underline{x}_o,\underline{y}_o)$ תקרא <u>יעילה</u> (Efficient) אם אין באיזור האפשרי אף נקודה אחת המייצרת $\underline{y}_o$ בעזרת $\underline{x}$ עבורו קיים $(\underline{x} < \underline{x}_o)$, או המשתמשת ב-$\underline{x}_o$ לייצור $\underline{y}$ עבורו קיים $(\underline{y} > \underline{y}_o)$, כאשר אי-השוויונות מבטאים "טוב-יותר" במשמעות הטכנולוגית. <u>פונקצית הייצור</u> (Production Function) היא המקום הגיאומטרי של כל הנקודות היעילות. להבהרה נסתכל במקרה כו יש רק תשומה אחת, x_1, ותפוקה אחת y_1 כמתואר בציור 2.1.



<u>ציור 2.1</u>: האיזור האפשרי ופונקצית הייצור.

פונקציית הייצור מורכבת מכל הנקודות עבורן אי אפשר למצוא נקודה אחרת בעלת x_1 קטן יותר באותו y_1 או y_1 גדול יותר באותו x_1.
פונקצית הייצור מסומנת כ-

$$g(\underline{x},\underline{y}) = g(x_1,\ldots,x_n\ ,\ y_1,\ldots,y_m) = 0 \qquad (2\text{-}1)$$

ברור שהפונקציה לא חייבת להיות רציפה, ובודאי לא בעלת נגזרות חלקות ורציפות.

עם זאת, לצורך המשך הדיון נניח, כפי שנהוג לעשות, כי הפונקציה רציפה ויש לה שתי נגזרות חלקיות כיחס לכל אחד מרכיביה.

כאשר תהליכי הייצור תלויים בגורמים אקראיים אי אפשר לחזות את $\underline{y}$ בבטחון מלא. לכל היותר אפשר לחשב את פונקצית הפירוס של כל אחד מרכיבי התפוקה. במקרה זה דנים בזוג הוקטורים $(\underline{x},\underline{w})$, כאשר $\underline{w}$ הוא וקטור התכונות הסטטיסטיות של $\underline{y}$. כדוגמא אפשר לתאר כל תפוקה, y_i, על ידי התוחלת שלה E_i והואריאנס שלה σ_i^2, אזי,

$$\underline{w} = (E_1,\sigma_1^2,\ldots,E_m,\sigma_m^2) \qquad (2\text{-}2)$$

כעת קשה יותר להגדיר מהי התפוקה ה"גדולה יותר" בין שתיים עבורן ידוע $\underline{w}$. אפשר למשל לדרוש שכדי ש-$\underline{w}_1$ תהיה תפוקה "גדולה יותר" מ-$\underline{w}_2$ דרוש שכל התוחלות בראשונה תהיינה גדולות יותר ואילו כל הואריאנסים קטניס יותר מאשר כשניה.

כאשר דנים כתשומות ותפוקות של תקופות זמן שונות יש לקחת כחשבון וקטור של וקטורים, למשל,

$$(\underline{x}^1,\underline{y}^1,\ldots,\underline{x}^t,\underline{y}^t,\ldots,\underline{x}^T,\underline{y}^T) \qquad (2\text{-}3)$$

כאשר הציון העילי מציין את התקופה. כעת קשה עוד יותר להגדיר איזו משתי אלטרנטיבות "גדולה יותר".

מכל שנאמר לעיל ברור כי הרעיון של פונקצית הייצור, יותר משהוא יכול לשמש באופן מעשי לאבחנה בין אלטרנטיבות שונות הוא כלי מחשבתי. ואמנם לצורך ההמשך נניח שיש בידינו פונקציית ייצור, שהיא רציפה ובעלת נגזרות חלקיות ביחס לכל אחד מרכיביה.

2.2 ניתוח שולי ותנאי האופטימליות

נניח שבכל נקודת תשומה-תפוקה $(\underline{x},\underline{y})$ הנמצאת על פונקצית הייצור $g(\underline{x},\underline{y})=0$ אפשר לחשב ערכה של פונקציה סקלרית חד-ערכית המבטאת ערכו הכלכלי של הייצור המתואר על ידי הנקודה. פונקציה זו, שהיא פונקצית המטרה ותסומן $f(\underline{x},\underline{y})$, תבטא

בדרך כלל את ההכנסה הנקייה מהייצור הנדון, כלומר, ההפרש בין ההכנסה הכוללת הנובעת מהתפוקות $\underline{y}$ לבין סה"כ עלות התשומות $\underline{x}$ הדרושות לייצורן. אנו דוצים כעת למצוא את הנקודה שעל פונקצית הייצור בה מקבלת פונקצית המטרה ערך מקסימלי.

נניח כי שתי הפונקציות גזירות ונשתמש בטכניקה של כופלי לגרנז'. טכניקה זו מוסברת בפירוט רק בפרק מאוחר יותר, אלא שלרוב הקוראים יש בודאי רקע מספיק כדי להבין את ההסבר שלהלן. הקורא הנזקק להסכר המפורט יפנה לסעיף 5.4.

בכדי למצוא את הנקודה $(\underline{x},\underline{y})$ הנותנת מכסימום של $f(\underline{x},\underline{y})$ בכפיפות לאילוץ $g(\underline{x},\underline{y})=0$ יוצרים את פונקצית לגרנז'

$$F(\underline{x},\underline{y},\lambda) = f(\underline{x},\underline{y}) + \lambda g(\underline{x},\underline{y}) \tag{2-4}$$

ומאפסים את הנגזרות החלקיות כיחס לכל אחד מן המשתנים

$$\begin{aligned} \frac{\partial F}{\partial x_i} &= 0 = \frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} \qquad i = 1,\dots n \\ \frac{\partial F}{\partial y_j} &= 0 = \frac{\partial f}{\partial y_j} + \lambda \frac{\partial g}{\partial y_j} \qquad j = 1,\dots,m \end{aligned} \tag{2-5}$$

על ידי יצירת מנות מתאימות מתקבל,

$$\left.\begin{aligned} \frac{\partial f/\partial x_i}{\partial f/\partial y_j} &= \frac{\partial g/\partial x_i}{\partial g/\partial y_j} \\ \frac{\partial f/\partial x_i}{\partial f/\partial x_h} &= \frac{\partial g/\partial x_i}{\partial g/\partial x_h} \qquad i,h = 1,\dots,n \\ \frac{\partial f/\partial y_j}{\partial f/\partial y_k} &= \frac{\partial g/\partial y_j}{\partial g/\partial y_k} \qquad j,k = 1,\dots,m \end{aligned}\right\} \tag{2-6}$$

בהנחה, כמובן, שהנגזרות במכנים אינן מתאפסות. בנקודות האופטימום מתקיים האילוץ $g(\underline{x},\underline{y})=0$, ולכן גם הדיפרנציאל השלם מקיים $\underline{dg}=0$. ניקח כעת את רכיביו המכוונים של הדיפרנציאל באופן שבכל פעם רק שניים מהם שונים מאפס.
למשל,

$$\underline{dg} = \frac{\partial g}{\partial x_i} dx_i + \frac{\partial g}{\partial y_j} dy_j = 0 \qquad (2\text{-}7)$$

ומזה,

$$\frac{\partial g/\partial x_i}{\partial g/\partial y_j} = -\frac{\partial y_j}{\partial x_i} \qquad (2\text{-}8)$$

באופן דומה מתקבל,

$$\frac{\partial g/\partial x_i}{\partial g/\partial x_h} = -\frac{\partial x_h}{\partial x_i} \qquad (2\text{-}9)$$

$$\frac{\partial g/\partial y_j}{\partial g/\partial y_k} = -\frac{\partial y_k}{\partial y_j} \qquad (2\text{-}10)$$

הכנסת קשרים אלה למשוואה (2-6) נותן עבור נקודת האופטימום את התנאים,

$$\left.\begin{aligned} \frac{\partial f/\partial x_i}{\partial f/\partial y_j} &= -\frac{\partial y_j}{\partial x_i} \\ \frac{\partial f/\partial x_i}{\partial f/\partial x_h} &= -\frac{\partial x_h}{\partial x_i} \\ \frac{\partial f/\partial y_j}{\partial f/\partial y_k} &= -\frac{\partial y_k}{\partial y_j} \end{aligned}\right\} \quad \begin{aligned} &i,h = 1,\ldots,n \\ &j,k = 1,\ldots,m \end{aligned} \qquad (2\text{-}11)$$

פירוש הנגזרות

נניח מחירי יחידה קבועים לתשומות ותפוקות ונכתוב את פונקציית ההכנסות הנקיות בצורה המפורשת

$$f(\underline{x};\underline{y}) = \sum_j p_j y_j - \sum_i c_i x_i \qquad (2\text{-}12)$$

כאשר p_j הוא מחיר היחידה של התפוקה y_j ו-c_i עלות היחידה של התשומה x_i. במקרה זה הנגזרות $\partial f/\partial x_i = -c_i$ ו-$\partial f/\partial y_j = p_j$ הן ערכים קבועים. במקרה הכללי $c_i=c_i(x_i)$ ו-$p_j=p_j(y_j)$ ואז ערכי הנגזרות משתנים בהתאם לערכי התשומות והתפוקות.

אנו מעוניינים בנגזרות בנקודת האופטימום, ועבורן נגדיר:

$$\frac{\partial f}{\partial y_j} = MB_j \qquad (2\text{-}13)$$

כאשר MB_j היא <u>ההכנסה השולית</u> (Marginal Benefit) של התפוקה y_j. זו ההכנסה ליחידת תפוקה כאשר התפוקה נמצאת ברמת האופטימום. באופן דומה נגדיר:

$$-\frac{\partial f}{\partial x_i} = MC_i \qquad (2\text{-}14)$$

כאשר MC_i היא <u>העלות השולית</u> (Marginal Cost) של התשומה x_i, כלומר, העלות ליחידת התשומה ה-i בנקודת האופטימום. בדומה

$$\frac{\partial y_j}{\partial x_i} = MP_{ij} \qquad (2\text{-}15)$$

MP_{ij} - <u>היעילות השולית</u> (Marginal Productivity) של התשומה x_i ביחס לתפוקה y_j, כלומר, בכמה תגדל התפוקה y_j אם נגדיל את התשומה x_i ביחידה אחת מעל ערכה באופטימום. בין שתי תשומות נגדיר

$$-\frac{\partial x_h}{\partial x_i} = MRS_{hi} \qquad (2\text{-}16)$$

MRS_{hi} - שיעור התחלופה השולי (Marginal Rate of Substitution) של התשומה x_h כיחס לתשומה x_i; כלומר, בכמה יהיה צריך להגדיל את התשומה x_h אם התשומה x_i תוקטן ביחידה אחת מערכה באופטימום. אם רוצים שכל שאר התשומות והתפוקות תשארנה ללא שינוי. כין שתי תפוקות נגדיר

$$-\frac{y_k}{y_j} = MRT_{kj} \qquad (2\text{-}17)$$

MRT_{kj} - שיעור ההתמרה השולי (Marginal Rate of Transformation) של התפוקה y_k כיחס לתפוקה y_j, כלומר, בכמה תקטן התפוקה y_k אם תוגדל התפוקה y_j ביחידה אחת מעל ערכה באופטימום כעוד כל שאר התפוקות וכל התשומות נשארות ללא שינוי.

כל ההגדרות מתיחסות לערכים השוליים, כלומר, העלות או ההכנסה מן היחידה האחרונה ליד נקודת האופטימום ולא לערכים הממוצעים. תוך שימוש בהגדרות אלה מתקבל ממשוואות (2-11) כי באופטימום חייבים להתקיים התנאים הבאים, הנקראים תנאי האופטימליות:

$$\frac{MC_i}{MB_j} = MP_{ij} \qquad i = 1,\dots,n \quad j = 1,\dots,m \qquad (2\text{-}18)$$

$$\frac{MC_i}{MC_h} = MRS_{hi} \qquad i,h = 1,\dots,n \qquad (2\text{-}19)$$

$$\frac{MB_j}{MB_k} = MRT_{kj} \qquad j,k = 1,\dots,m \qquad (2\text{-}20)$$

המשמעות של תנאים אלה הגיונית ופשוטה למדי. למשל: נניח כי בנקודה מסויימת שעל פונקצית הייצור שיעור ההתמרה השולי כין שתי תפוקות הוא $MRT_{kj} = q$. פירוש הדבר כי בנקודת ייצור זו, אם נגדיל את התפוקה ה-j ביחידה יש להקטין את התפוקה ה-k ב-q יחידות - אם רוצים כי כל שאר התפוקות וכל התשומות תשארנה ללא שינוי. הנקודה הנידונה תהיה אופטימלית רק אם ההכנסה השולית של התפוקה ה-j גדולה בדיוק פי q מזו של התפוקה ה-k, בהתאם לתנאי (2-20). אם זה אינו

קיים, למשל ההכנסה השולית של התפוקה ה-j גדולה פי t מזו של התפוקה ה-k, כאשר t>q, כדאי "לזוז" מן הנקורה הנדונה בכוון של הגדלת התפוקה ה-j של חשבון התפוקה ה-k, ובכך יגדלו ההכנסות הנקיות. באופן דומה ניתן לפרש את שאר התנאים.

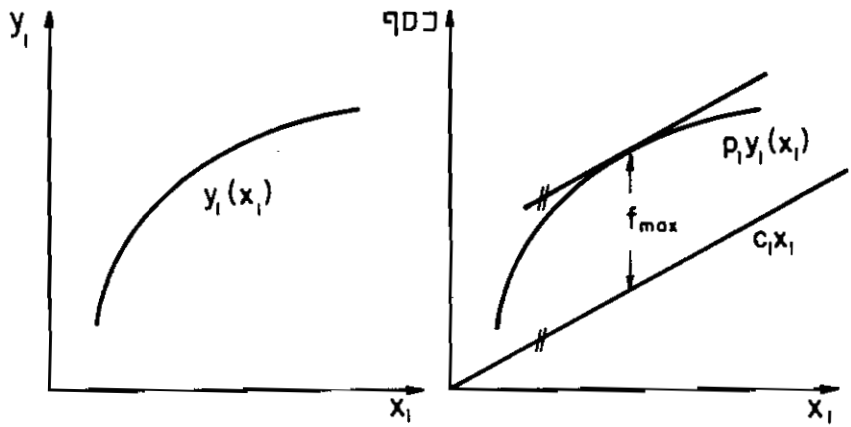
<u>דוגמא</u>: תשומה אחת ותפוקה אחת. פונקצית הייצור $g(x_1,y_1) = 0$, או $y_1 = y_1(x_1)$ מתוארת בציור 2.2 משמאל. מימין מתוארים מרכיבי הפונקציה

$$f(x_1;y_1) = p_1y_1 - c_1x_1$$

בה הונח, לשם הפשטות, כי p_1 ו-c_1 קבועים ואינם תלויים ב-y_1 וב-x_1. כאן קיים $MC_1 = c_1$ ו-$MB_1 = p_1$. לפי תנאי האופטימליות דרוש כי בנקודת האופטימום יתקיים:

$$MP_{1,1} = \frac{MC_1}{MB_1} \Longrightarrow \frac{dy_1}{dx_1} = \frac{c_1}{p_1}$$

כלומר, האופטימום הוא בנקודה בה אם התשומה x_1 גדלה ביחידת, על התפוקה y_1 לגידול ב-(c_1/p_1) יחידות. הגדלת x_1 ביחידה עולה c_1. התפוקה גדלה בשיעור (c_1/p_1) וגורמת להכנסה נוספת בשיעור $p_1(c_1/p_1) = c_1$. בנקודת האופטימום מתאזנת ההשקעה הנוספת על ידי הכנסה נוספת השווה לה, כלומר, אין הצדקה להשקיע יותר (או, מאותם שיקולים, פחות) מאשר עבור x_1 של נקודת האופטימום. באופן גרפי נראה הדבר בברור בציור 2.2 מימין: ההכנסות הנקיות, f, הן ההפרש האנכי שבין עקום ההכנסות הכוללות $p_1y_1(x_1)$ לבין עקום העלויות c_1x_1. האחרון הוא קו ישר, שכן c_1 קבוע. f יקבל ערך מקסימלי כאשר המשיק לעקום $p_1y_1(x_1)$ מקביל לקו c_1x_1. כלומר שיפועו c_1.



ציור 2.2: פונקצית הייצור (משמאל) ותנאי האופטימליות (מימין) לתשומה אחת ותפוקה אחת.

2.3 מתימטיקה של מימון

בסעיף זה ניתן כמה מן הנוסחאות המשמשות להשוואה בין "זרמים" שונים של כסף. הן בסעיפי ההוצאות והן בהכנסות מופיעים סכומים חד-פעמיים וסכומים שנתיים. לכסף בזמנים שונים יש ערך שונה. הדבר נובע משתי סיבות:

א. בזמנים שונים אפשר לנצל את הכסף למטרות שונות.

ב. על פני זמן ניתן להגדיל סכום כסף על ידי השקעתו בריבית השוררת בשוק.

בסיבה (א), הקובעת את "ערכו האמיתי" של הכסף לא נטפל כאן. נעסוק פשוט בהשוואת ערך הכסף בזמנים שונים בהתחשב באפשרות להשקיעו.

סימון:

סכום שנתי בסוף כל שנה	A
שער הריבית, בשבר עשרוני	i
מספר השנים	N
ערך נוכחי של הכסף	P
ערך עתידי בסוף N שנים	F

ריבית מצטברת על השקעה

אם משקיעים היום סכום P בריבית i למשך N שנים יתקבל כסוף התקופה סכום,

$$F = P(1+i)^N \qquad (2\text{-}21)$$

ערכו הנוכחי של הסכום F שהצטבר במשך N שנים לפי ריבית i הוא,

$$P = \frac{F}{(1+i)^N} \qquad (2\text{-}22)$$

סכומים שנתיים

בסוף כל שנה מכניסים A לריבית מצטברת לפי שער i. לאחר N שנים יהיה הסכום,

$$F = A + A(1+i)^2 + A(1+i) + \ldots + A(1+i)^{N-1} = A\frac{(1+i)^N-1}{i} \qquad (2\text{-}23)$$

הערך הנוכחי של סכומים שנתיים

מהשוואת הערך העתידי של השקעה חד-פעמית P לפי משוואה (2-21), עם הערך העתידי של הכנסת סכומים שנתיים A לפי משוואה (2-23) מתקבל

$$P(1+i)^N = A\frac{(1+i)^N-1}{i}$$

או, לאחר סידור

$$P = A\frac{(1+i)^N-1}{i(1+i)^N} = vA \qquad (2\text{-}24)$$

כאשר v הוא מקדם הערך הנוכחי של סדרת הסכומים השנתיים, לדוגמה כמה ערכים של v:

i	3	3	4	4	6	6	8	8
N	10	50	10	50	10	50	10	50
v	8.53	25.73	8.11	21.48	7.36	15.76	6.71	12.23

טבלאות שלימות ניתן למצוא בספרים רבים.

אפשר במובן, במקום להפוך סדרת סכומים שנתיים לערך נוכחי, להפוך את הסכום החד-פעמי המושקע היום לסדרה שוות-ערך של סכומים שנתיים על ידי היפוך משוואה (2-24)

$$A = \frac{P}{v} = \frac{i(1+i)^N}{(1+i)^N-1} \cdot P \qquad (2\text{-}25)$$

המקדם (1/v) נקרא מקדם החזר ההון.

מקורות

Baumol, W.J., "Economic Theory and Operations Analysis", Prentice-Hall, 1965.

Maass, A. et al, "Design of Water Resources Systems", Harvard University, 1962 (Chapters 2 and 3).

James, L.D. and Lee, R.R., "Economics of Water Resources Planning", McGraw-Hill, 1971.

ורשבסקי, א., "כלכלה הנדסית", מכלול, 1978.

פרק שלישי

מיסודות תורת ההחלטות

אנו פונים אל תורת ההחלטות כבואנו לטפל בבעיה של בחירה בין אלטרנטיבות. בשם "אלטרנטיבה" אנו קוראים לכל פתרון אפשרי של הבעיה ההנדסית. אלטרנטיבות יכולות להיות שונות זו מזו בעיקרי הגישה כולה או בפרטים בלבד. מתפקידו של המהנדס להפעיל את דמיונו היוצר ולבסות את כל הפתרונות האפשריים לבעיה, ותוך כך הוא יכול להציע פתרונות השונים מאוד זה מזה. לאחר מכן, כאשר הפתרון הבסיסי קיים, יש עדיין תחום רחב של אפשרויות הנבדלות זו מזו בגדלי המשתנים. שיטות האופטימיזציה נכנסות לפעולה בשלב בו כבר הוצע הפתרון כאופן עקרוני, אבל עדיין נותרה הבעיה של קביעת גדלי המשתנים. בשלב האופטימיזציה אנו זקוקים לפונקציית מטרה חד-משמעית, המצמידה ערך מספרי לכל אלטרנטיבה, כאשר ערכים אלה מהווים את הבסיס לבחירה בין האלטרנטיבות. אבל גם בשלב של בחינת פתרונות בעלי אופי שונה אנו זקוקים לקריטריונים לבחירה. בפרק זה נציג כמה מן הבעיות הקשורות לבחירה בין אלטרנטיבות. ההצגה בהכרח קצרה ואינה ממצה את הנושא. הקורא המעונין מופנה לספרות, ובעיקר לספריהם של Von Neuman and Morgenstern ושל Keeney and Raiffa.

3.1 מדידה

מדידה (Measurement) היא הפעולה של הצמדת תווית זיהוי או ערך מספרי לכל אלטרנטיבה. נסמן באותיות Q,P,B,A תוויות או ערכים מספריים כלשהם ונבדוק מהן הסקלות השונות הנוצרות כאשר קיימים תנאים מסויימים:

3.1.1 סקלת זיהוי (Nominal Scale)

סקלת זיהוי מקיימת שלושה תנאים

Either $A=B$ or $A\neq B$ (3-1)

If $A=B$ then $B=A$ (3-2)

If $A=B$ and $B=C$ then $A=C$ (3-3)

דוגמא לסקלת זיהוי יכולה להיות קבוצת שרטוטים המסומנים או ממוספרים בצורה כלשהי. המובן של הערך המספרי הוא כשל "שם", ואמנם, סקלת זיהוי ניתן ליצור בלי שימוש בספרות, אלא באמצעות אותיות או סימנים. סקלה זו היא ה"נמוכה" ביותר, ואין בה בל ביטוי ל"טיב" האלטרנטיבה המזוהה. שתי הצעות לבניית מבנה מחומרים שונים תיקראנה לפי סקלה זו פשוט בשמות שונים, בלי כל נסיון לתאר את תכונותיה של כל הצעה מבחינת המחיר, יציבות המבנה וכדומה.

3.1.2 סקלת סדר (Ordinal Scale)

סקלה זו מקיימת בנוסף לתנאים (3-1) עד (3-3), גם,

$$\text{If} \quad A > B \quad \text{then} \quad B \not> A \tag{3-4}$$

$$\text{If} \quad A > B \quad \text{and} \quad B > C \quad \text{then} \quad A > C \tag{3-5}$$

סימן אי השוויון מסמן כיוון של תכונה מסויימת, בלי ליחס ערך מספרי לתכונה. דוגמה לסקלת סדר היא סידור אבנים לפי משקלן, באשר משתמשים לשם כך במאזניים שאינן רושמות. אפשר בצורה זו לסדר את האבנים בסדר עולה של משקלן, אבל אי-אפשר לדעת מה משקלה, ביחידות כלשהן, של כל אחת מהן, וכן בכמה או פי כמה כבדה אבן אחת מאחרת. דוגמה לשימוש בסקלת סדר יהיה מניין אלטרנטיבות שונות לבניין סבר אגירה לפי נפח המים הנאגר בכל אלטרנטיבה, בלי לציין במפורש את הנפח הנאגר. היות וסקלת הסדר קובעת רק סדר יחסי אי אפשר לבצע פעולות אריתמטיות על הקבוצה.

3.1.3 סקלת מרווחים (Interval Scale)

סקלה זו מקיימת את התנאים (3-1) עד (3-5) ופריטיה מזוהים על ידי מספרים ממשיים. לדוגמא ישמש סידור אלטרנטיבות לפי הפרש המחיר בינן לבין אלטרנטיבה מסויימת שנבחרה בבסיס להשוואה. הפרשי המחירים כין זוגות אלטרנטיבות הם בעלי משמעות, אבל אי אפשר להגיד שאחת עולה פי כך ובך מאחרת, שבן אין בסקלה נקודת אפס ידועה. סידור אלטרנטיבות לפי סקלת מרווחים לעיתים קרובות קל הרבה יותר מאשר בסקלת היחס, אשר תתואר להלן.

3.1.4 סקלת יחס (Ratio Scale)

בנוסף לחמשת התנאים הראשונים מקיימת סקלה זו גם

$$\text{If} \quad A = P \quad \text{and} \quad B > 0 \quad \text{then} \quad A+B>P \tag{3-6}$$

$$A + B = B + A \tag{3-7}$$

$$\text{If} \quad A = P \quad \text{and} \quad B = Q \quad \text{then} \quad A+B = P+Q \tag{3-8}$$

$$(A+B) + C = A + (B+C) \tag{3-9}$$

תנאי (3-6) קובע את מקומו של האפס בסקלה. אי לכך רק היחידות בהן נמדדות התכונות (כלומר, קנה המידה) יכולות להקבע בצורה רצונית (אלפי ₪ במקום ₪, מטרים במקום ס"מ וכדומה).

סקלות היחס נפוצות מאוד בהנדסה, אך נדירות במדעי החברה. הכלכלה, המאחדת משתנים הנדסיים. חברתיים, פסיכולוגיים, אסתטיים ואחרים, עומדת בפני הבעיה של יצירת סקלות יחס, שכן, כפי שנראה להלן, פונקציות המטרה בהן נטפל יהוו למעשה סקלת יחס עבור האלטרנטיבות השונות.

3.1.5 סקלות רב מימדיות

עד עתה דנו בסידור אלטרנטיבות בהתאם לתכונה אחת בלבד. המצב הוא כדרך כלל שכוחניס כל אלטרנטיבה בהתאם למספר רב של תכונות - פיזיקליות, אסתטיות, כלכליות ואחרות. יכולתנו לדרג בצורה חד-משמעית אלטרנטיבות כאלה מותנה בקיום סקלת סדר, לכל הפחות. תכונה (3-5) נקראת <u>טרנזיטיביות</u>, וקיומה סבטיח שלא יווצר מעגל סגור של העדפות. כאשר מדרגים אלטרנטיבות לפי תכונה אחת שלהן בלבד מהווה (3-5) תנאי הגיוני. הדבר שונה באשר דנים בסקלות רב-מימריות. אפשר ליצור דוגמאות הסורות כי ארם עשוי להעדיף את אלטרנטיבה A על B כאשר שתיים אלו בלבד מוצגות בפניו, את B על C כאשר אלו בלבר מוצגות בפניו, ואילו כניגוד כביכול להגיון, יעריף את C על A כאשר תעמורנה רק שתיים אלה בפניו. לדוגמה ישמש הסיפור ההיפותטי הבא:

אדם רוצה לכנות גג מעל שטח מסוים. מוצעות לו שלוש אלטרנטיבות. כיסוי ברזנט, כיסוי מפח גלי וכיסוי מאסבסט גלי. מחיר הברזנט -.100 ₪, הפח -.150 ₪ והאסבסט -.200 ₪ (מחירים אלה כוללים הן את מחיר המבנה והן את הוצאות האחזקה על פני השנים). בין הברזנט לפח כוחר האיש בפח, בטענה ש"כדאי" להשקיע 50 ₪ נוספות כדי לקבל מבנה יציב יותר. בין הפח לאסבסט שוב בוחר האיש ביקר יותר, כלומר, באסבסט, בטענה ש"כדאי" להשקיע 50 ₪ נוספות כדי לשפר את מראיתו של הגג ולהקטין את הרעש הנגרם על ידי הגשם הנופל. כאשר מציגים לפניו את הברזנט לעומת האסבסט ייתכן בתחלט שיבחר בברזנט, בטענה ששניהם מכסים את השטח, ולא כדאי להכפיל את ההוצאה רק כדי לקבל יתר יציבות ויופי.

לא נוכל לטעון נגדו טענה של חוסר הגיון או חוסר עקביות, שכן הוא משווה את האלטרנטיבות בסקלה רב-מימדיות בה הרכיבים הם מחיר, יציבות, יופי ורעש. לו היינו יכולים להתמיר את ערכי היציבות, היופי והרעש ליחידות של כסף היתה הבעיה הופכת חד-מימדית. הקושי הוא במובן מציאת קני המידה הכספיים של ערכים כגון יופי, רעש וכדומה. להלן נראה שאפילו במקרים בהם הכסף הוא מימד יחידי להערכת אלטרנטיבה הרי נוכחותם של תנאים אקראיים גורמת לסקלה ליהפך לרב-מימדית, ובכך שוב הופכות אלטרנטיבות לבלתי-שקילות.

לפני שנמשיך נתעכב לרגע על אחת מהדרכים שהוצעו בכדי להתיר מצבים של חוסר טרנזיטיביות ביחסו של הפרט אל אלטרנטיבות שונות. כשעוסקים בפרויקטים ציבוריים אפשר לפנות אל ציבור המחליטים ולבחור אלטרנטיבה על סמך הצבעה. למרות שמבחינה מעשית מורה הצכעה זו לרוב על בחירה חד-משמעית בין אלטרנטיבות, הרי שמבחינה תיאורטית לפחות יש לה החסרון שאף היא יכולה להראות תכונה של חוסר-טרנזיטיביות. להדגמה נניח שיש שלושה מצביעים: א', ב', ו-ג', ושלוש אלטרנטיבות להצבעה: B,A ו-C. א' מעדיף אותן בסדר C<B<A ב' בסדר A<C<B, ואילו ג' בסדר B<A<C. אם נעמיד להצבעה את A לעומת B יצביעו א' ו-ג' בעד A ואילו ב' בעד B, כלומר - נבחרה A. כשנעמיד את B לעומת C יצביעו א' ו-ב' בעד B ואילו ג' בעד C, כלומר - נבחרה B. מכאן ניתן היה להסיק שהיות ו-A הועדפה על B ו-B הועדפה על C, הרי שבהצבעה בין A לבין C צריכה A להכחר. קל להיווכח שההיפך קורה: ב' ו-ג' מצביעים דוקא בעד C ורק א' בעד A - כאשר רק שתיים אלו מועמדות להצבעה. שוב נוצר מעגל סגור של העדפות. הפעם ב"הצבעה פומבית". (קוריוז: חלק מן הטענות נגד הדירוג במצער הפזמונים הישראלי מבוססות על האפשרות לקבלת מעגלי העדפה סגורים כגון זה שהוזכר לעיל).

בסעיף 4.5 נחזור לעסוק בבעיות החלטה רב-קריטריוניות.

3.2 החלטות בתנאי אי-ודאות ואי-ידיעה

החלטות יכולות להתקבל בתנאים של ודאות (Certainty). אי-ודאות (Risk) ואי-ידיעה (Uncertainty). ודאות היא המקרה בו ידוע מראש בבטחון מלא מה תהיה התוצאה של כל החלטה אפשרית. אי-ודאות הוא המקרה בו התוצאה מושפעת לא רק מן ההחלטה אלא גם ממצב הטבע (State of Nature) - מושג המבטא את השפעתם של גורמים שאינם בשליטתנו - כאשר ההסתברות של כל אחד מן המצבים האפשריים של הטבע ידועה מראש. אי-ידיעה היא המקרה בו ידוע אמנם כי התוצאה מושפעת על ידי גורמים שאין למחליט שליטה עליהם, אלא שההסתברות לתחולת האירועים הקראיים אינה ידועה, ולפעמים אפילו אין יודעים כבירור מהם האירועים עצמם.

על מנת לטפל בהחלטות נסדר את כל התוצאות האפשריות במטריצה, אשר תיקרא מטריצת התועלת (Payoff Matrix). כל שורה של המטריצה מתאימה לאחת מן ההחלטות האפשריות, וכל אחת מן העמודות - לאחד ממצבי הטבע האפשריים. הערכים במטריצה יהיו ההכנסות הנקיות של כל קומבינציה החלטה-מצב. לדוגמה נתאר לעצמנו כעיה בה אפשר להחליט אחת משלוש החלטות שתסומנה א', ב' ו-ג'. לאחר שהחלטנו יסתבר כי הטבע נמצא במצב 1 או 2. מטריצת התוצאות, המתארת את התועלת של כל קומבינציה היא:

התועלת כאשר הטבע

	במצב 1	כמצב 2
החלטה א'	130	400
החלטה ב'	140	260
התלטה ג'	80	80

לו ידענו מראש כי הטבע יהיה כמצב 1 היינו בוחרים ב-(ב'), לו ידענו שיהיה במצב 2 היינו בוחרים ב-(א'). במלים אחרות, כאשר יש לנו רק עמודה אחת, אין כל קושי כבחירת האלטרנטיבה הטובה ביותר. אנו רוצים לכן בשיטה שתיצור מכל העמודות של מטריצת התועלת עמודה אחת שבעזרתה נחליט. קיימים אופנים שונים לבניית עמודה כזו, ואנו נתעכב על כמה מהן.

3.2.1 עקרון *maxmin* של התועלת

לפי עקרון זה יש לבחור את אותה האלטרנטיבה המבטיחה את הערך הגבוה ביותר של התוצאה הנמוכה ביותר האפשרית. כדוגמא שלנו יש לבחור לפי קריטריון זה באלטרנסיבה (ב'), המבטיחה לפחות 140, בעוד ש-(א') מבטיחה רק 130 ו-(ג') מבטיחה רק 80.

שיטה זו שמרנית ביותר. היא אינה לוקחת בחשבון את הרווחים הפוטנציאליים הגבוהים. לו למשל היה רשום עבור החלטה א' ומצב טבע 2 שהתועלת היא 10^6 במקום 400, היתה הבחירה לפי הקריטריון הנוכחי עריין מעריפה את ב'. כלומר, במטרה להבטיח מינימום הכנסות הגבוה ב-10 יחידות אנו מוותרים על הכנסה פוטנציאליה עצומה.

לשיטה יש גם פגם לוגי, המתבטא באופן הבא: החסרת גורל קבוע מהתועלות של מצב טבע מסוים (טור) יכולה לגרום לשינוי בבחירה. למשל, החסרת 135 מהטור השמאלי נותנת שם סדר תועלת חדש: 265, 125, 55-, בהתאמה. לפי הנתונים החדשים יש כעת לבחור באלטרנטיבה א' במקום ב- ב', כי הראשונה מבטיחה לפחות 130 בעוד השניה מבטיחה רק 125. הפגם הלוגי הוא בכך שקנסנו במידה שווה את כל האלטרנטיבות על ידי חיסור הגורל הקבוע, וקנס כזה אינו צריך לגרום לשינוי בהחלטה.

3.2.2 עקרון *minmax* של ההפסד הפוטנציאלי

כדי להשתמש בעיקרון זה יוצרים מטריצה חדשה בה כל איבר הוא ההפרש בין כל תועלת לבין התועלת המכסימלית האפשרית באותו מצב טבע (באותו טור). המטריצה החדשה משקפת את ההפסד הפוטנציאלי עבור כל מצב טבע בין האלטרנטיבה הנדונה לאלטרנטיבה הנותנת את התועלת המכיסמלית במצב הטבע הזה. לדוגמא שלנו מקבלים

ההפסד הפוטנציאלי, כאשר הטבע

	במצב 1	במצב 2
החלטה א'	10	0
החלטה ב'	0	140
החלטה ג'	60	320

מספרים אלה הם ההפרש בין הרווח המתקבל מהחלטה בלשהי לבין הרווח שהיה מתקבל אילו היינו בוחרים את האלטרנטיבה הטובה ביותר עבור מצב הטבע שאמנם התקבל, כלומר, אילו ידענו מראש מה יהיה מצב הטבע. כעת בוחרים את אותה האלטרנטיבה בה הגדול שבין ההפסדים הפוטנציאליים הוא הקטן ביותר, במקרה שלנו את אלטרנטיבה א' - בה ההפסד הפוטנציאלי המכסימלי הוא 10, לעומת 140 ב- ב' ו-320 ב- ג'.

לשיטה זו יש חסרון לוגי, המוסבר כדלקמן: נניח שנוסיף אלטרנטיבה ד', אשר תוצאותיה 300 במצב 1 ו-50 במצב 2. טבלת ההפסדים הפוטנציאליים הכוללת את האלטרנטיבה החדשה היא:

ההפסד הפוטנציאלי, כאשר הטבע

	במצב 1	במצב 2
החלטה א'	170	0
החלטה ב'	160	140
החלטה ג'	220	320
החלטה ד'	0	350

כעת ההחלטה נופלת על ב', בה ההפסד הפוטנציאלי המכסימלי הוא 160, לעומת 170, 320 ו-350 באחרות. נשים לב כי אלטרנטיבה ד' נופלת הן מ-א' והן מ-ב' לפי קריטריון ה-*minmax* עצמו, ועם זאת היא גדמה לשינוי בהחלטה! זהו פגם שכן אלטרנטיבה אשר בעצמה אינה "מועמדת" כלל יכולה לגרום לשינוי ב"צמרת".

קיימות שיטות נוספות הדומות לשתיים שהוצגו. אלא שכולן סובלות מפגמים לוגיים דומים, ואין בהן גם אחת שתעמוד בכל המבחנים. איננו טוענים כי הקריטריונים לעיל חסרי ערך. הבעייה היא שאין קריטריון העומר בכל המבחנים הלוגיים ובמבחן השימוש המעשי. אפשר לכן להשתמש באחד או יותר מהם בעת בחינת אלטרנטיבות, בתנאי שלקריטריון יש משמעות בתנאים הספציפיים של הבעיה. למשל, המצבים בהם אי אפשר כלל להסתכן, ומותר מבחינה כלכלית להסתמך מראש רק על התועלת המינימלית העשויה להתקבל - יש להשתמש בעיקרון ה-*maxmin* למרות שמרנותו.

השיטות שתוארו לעיל מתאימות לתנאים של אי-ידיעה, בה כאמור אין ההסתברויות למצבי הטבע השונים ידועות. כאשר ההחלטות נעשות בתנאים של אי-ודאות, כלומר ידועות ההסתברויות, המצב נוח יותר, כפי שיוסבר להלן.

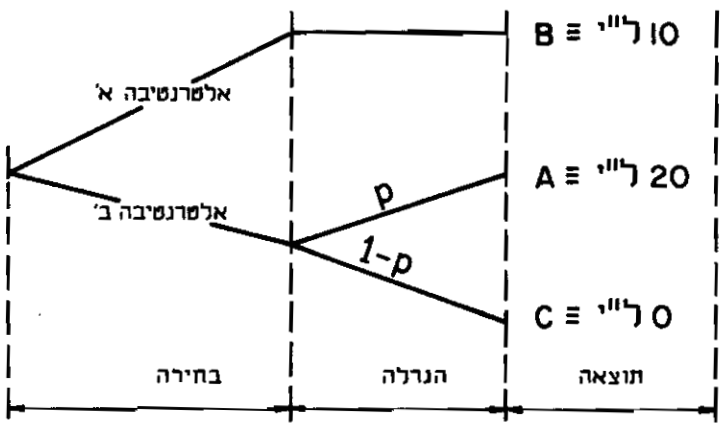
3.2.3 השיטה של פון נוימן ומורגנשטרן

פון נוימן ומורגנשטרן הניחו שמתקיימות שתי אקסיומות ביחס לדירוג האלטרנטיבות:

(א) קיימת סקלת סדר.

(ב) המחליט יכול לציין את העדפתו בין אלטרנטיבה בטוחה לבין קומבינציה אקראית מסויימת של שתי אלטרנטיבות אחרות.

כדי להבהיר את האקסיומה השניה נתאר לעצמנו שמציעים לאדם מסויים את הבחירה בין שתי אלטרנטיבות: אם יבחר בראשונה יקרה אירוע B, שהוא - האיש יקבל 10 ל׳. אם יבחר באלטרנטיבה השניה תקוים הגרלה בה בהסתברות p תהיה התוצאה C: האיש לא יקבל מאומה, ובהסתברות (1-p) תהיה התוצאה A: האיש יקבל 20 ל׳. בצורה גרפית ניתן המצב לתיאור בצורה הבאה.



ציור 3-1: בחירה בין אלטרנטיבה דטרמיניסטית לבין הגרלה.

כעת אנו מבקשים מהמחליט לתת לנו את ערכה של ההסתברות p (החייבת במובן לקיים $0 \le p \le 1$) בה יהיה אדיש בין אלטרנטיבה א' ל-ב'. לא נתפלא אם תשובתו תהיה $p = 0.5$. אלא תשובה זו אינה בעלת ערך מוחלט. למשל, אם אין פרוטה בכיסו, והכסף שיקבל ישמש לו לאכול את הארוחה הראשונה מזח שלושה ימים ייתכן בהחלט שתשובתו תהיה $p = 0.90$ או אפילו $p = 1.0$. המשמעות היא שלתוצאה הבטוחה של 10 ₪ יש "ערך" העולה בהרבה על הגרלה בעלת סיכוי שווה לקבלת 0 ₪ או 20 ₪, שכן 10 ₪ מבטיחות ארוחה, בעוד שהאיש ישאר רעב אם תוצאת ההגרלה תהיה C.
המסקנה: קביעת הערך של p היא סובייקטיבית ותלויה במצבו של המחליט, ביחסו להגרלות בכלל (האם הוא באופיו מהמר או שמרן) ובעוד גורמים, שעל אחדים מהם נתעכב בהמשך ביתר אריכות.

בעת קביעת ערכו של p עבורו המחליט אדיש בין האלטרנטיבה הדטרמיניסטית לאקראית הוא משתמש ראשית <u>בתוחלת</u> של האחרונה. כדוגמה לעיל, אם נקבע $p = 0.50$ הרי שתוחלת האלטרנטיבה האקראית היא $0.5\times20 + 0.5\times0 =$ 10 ₪. בעוד שאם נקבע $p = 0.10$ התוחלת היא $0.9\times20 + 0.1\times0 =$ 18 ₪. אלא שהתוחלת אינה המדד היחיד בעל החשיבות של האלטרנטיבה האקראית. נדגים: נניח כי מישהו קבע $p = 0.50$ בדוגמה לעיל.
האם היה קובע אותו ערך עבור ההגרלה בין האירועים

D = (הוא משלם 1000 ₪) ו- E = (הוא מקבל 1020 ₪)

נשים לב שעבור $p = 0.5$ מחקבלת אותה תוחלת כמו קודם כי $0.5\times(-1000) + 0.5\times1020 =$ 10 ₪. אלא שהפעם טווח התוצאות גדל באופן ניכר, כלומר, הואריאנס גדל. הפעם יתחשב האיש בגודלה של ה"קופה" העומדת לרשותו - האם יוכל בכלל לשלם 1000 ₪ אם ידרש לכך לפי תוצאות ההגרלה. גודלם של הסכומים המופיעים בהגדלה, יחסית לתחום אפשרויותיו של המחליט. הוא גורם קובע בהחלטתו. מה שיכולה להרשות לעצמה חברה גדולה - ל"שחק" בסכומים גדולים - אין הפרט יכול.

למרות כל ההסתייגויות נשאר העיקרון בר תוקף. על המחליט עצמו, זה שיצטרך בסופו של דבר לבחור בין האלטרנטיבות, לקבוע בהתאם לתנאים הספציפיים של הבעייה את ערכו של p. ערך זה חייב לשקף את כל השיקולים הסוביקטיביים ואין טענה שהערך הוא אבסולוטי ובלתי ניתן לשינוי. להיפך - דוקא כוחה של השיטה בכך שהיא משקפת תנאים ספציפיים.

3.2.4 תועלת וחישובה

פון נוימן ומורגנשטרן השתמשו במושג תועלת (Utility) במובן שונה מזה שניתן למילה קודם לכן. למושג זה יש עבר עשיר בכלכלה. בהשתמשם במושג לצורך חדש גרמו פון נוימן ומורגנשטרן למידה רבה של מבוכה ולמספר עצום של ויכוחים על משמעותה האמיתית של המילה. לא ניכנס לויכוח עצמו, ונאמר רק כי בסעיף הנוכחי נשתמש במושג "תועלת" במובן שנתנו לו אבות השיטה, והוא: תועלת היא סקלה לדירוג אלטרנטיבות.

כדי להמחיש נשתמש בדוגמה. נניח שקיימות שלוש אלטרנטיבות, שתסומנה A,B ו-C, שהן דטרמיניסטיות, כלומר, תוצאתה של כל אחת ידועה בוודאות מראש. נניח גם כי האדם שהוא מחליט ההחלטות מעדיף את A על B, את B על C וכן את A על C, כלומר נשמרת הטרנ זיטיביות. אנו מניחים כעת (וזו האקסיומה השניה של השיטה) כי אפשר למצוא p $(0 \le p \le 1)$ כזה שעבורו יהיה המחליט אדיש בין האלטרנטיבה B לבין אלטרנטיבה חדשה, בה תהיה התוצאה אקראית: A בהסתברות p או C בהסתברות $(1-p)$. קיומו של p כזה מאפשר לבחור בין אלטרנטיבה דטרמיניסטית להגרלה, ובקביעתו משתתפים כל השיקולים של הבוחר, גם אלה שאינם ניתנים לכיטוי כמותי מפורש.

לכל אחת מן האלטרנטיבות היסודיות, הדטרמיניסטיות, ניחס ערך של תועלת. ערך זה יבוטא ביחידות מספריות כלשהן, אשר נהוג בספרות לקרוא להן "יחידות תועלת", או Utils. ערכים אלה קובעים סקלת סדר בלבד, ואין ליחס לערכים משמעות שמעבר לזה. לפי ערכי התועלת של כל אלטרנטיבה אפשר לומר מהו סדר ההעדפה, אבל לא "בכמה" או "פי כמה" עדיפה אלטרנטיבה אחת על רעותה.

נחזור לדוגמה, נייחס למשל תועלת של 1 לאלטרנטיבה A ותועלת של 0 לאלטרנטיבה C, דבסמן $U(A) = 1$, $U(C) = 0$. אפשרית כל בחירה אחרת המקיימת $U(A) > U(C)$. נניח שהמחליט קבע את הערך $p = 2/3$. התועלת של B תחושב לכן בעזרת הנוסחה,

$$U(B) = p \times U(A) + (1-p) \times U(C) = \frac{2}{3} \times 1 + \frac{1}{3} \times 0 = \frac{2}{3}$$

נשים לב כי התועלת של B שווה בדיוק לתוחלת התועלת של ההגרלה.

נחזור לדוגמה של הסעיף הקודם, בה האפשרויות היו: A - לקבל 20 ₪, B - לקבל 10 ₪, ו-C - לקבל 0 ₪. ארם עם כסף רב בכיסו עשוי היה לבחור בערך p=0.5. עבורו התועלת של קבלת 10 ₪ שווה לתועלת של הגרלה שהתוחלת שלה 10 ₪. לעומתו עשוי היה אדם רעב וחסר פרוטה לבחור בערך p=0.90, ובכך ליחס ל-B תועלת של

$$U(B) = p \times U(A) + (1-p) \times U(C) = 0.90 \times 1 + 0.1 \times 0 = 0.9$$

התוחלת ביחידות כספיות, של הגרלה זו עם p=0.9 היא

$$E(B) = 0.90 \times (20) + 0.10 \times (0) = ₪18$$

כלומר, האיש חסר הפרוטה אדיש בין קבלת 10 ₪ בבטחון לבין הגרלה אשר בה תוחלת התוצאה הכספית היא 18 ₪. אנו רואים, אם כן, כי התועלת אינה בהכרח פונקציה לינארית של התוחלת הכספית.

אם אמנם משתמשים בערכים של תועלת, אזי הקריטריון של בחירה בין אלטרנטיבות הוא תוחלת התועלת. ברגע שעברנו לערכי תועלת, במקום להשתמש ביחידות פיזיקליות (כגון משקל, שטח וכו') או ביחידות כסף, הכנסנו כבר לחשבון את כל הפירוס של התוצאות האפשריות - את התוחלת, סטית התקן, הטווח וכל מומנט אחר של הפירוס שהוא בעל חשיבות. תוחלת התועלת כבר טומנת בתוכה את השקפתו הסובייקטיבית של מקבל ההחלטה על פירוס כל התוצאות האפשריות של כל אלטרנטיבה שיבחר בה.

במידת מה עוקפת שיטת התועלת את הבעיה במקום לתקוף אותה ישירות (לא שקיימים הטוענים כי אפשר בכלל לתקוף את הבעיה ישירות). במקום לנטות ליחס בצורה אובייקטיבית ערכי הערפה לאלטרנטיבות בהן יש הגרלות, כלומר, שלתוצאותיהן יש פירוס הסתברותי, השיטה "מחזירה את הכדור" לידיו של האדם המקבל בסופו של דבר את ההחלטות. עליו לספק מראש את ערכי התועלת, או את ערבי של כל אלטרנטיבה - בעזרתם מחשבים את התועלת.

פונקצית התועלת המתקבלת כשיטה זו היא טרנזיטיבית, ומבטיחה שלא יווצרו מעגלי הערפה סגורים.

יש להעיר כי בפרויקטים ציבוריים מניחים לעתים, ורבר זה נות מאוד בעבודה, כי התועלת היא פונקציה ליניארית של הכסף. בחינה מרוקדקת יותר מראה כי לעקום

המתאר תועלת כפונקציה של ההכנסות יש התנהגות המתוארת בצורה עקרונית בציור 3-2.

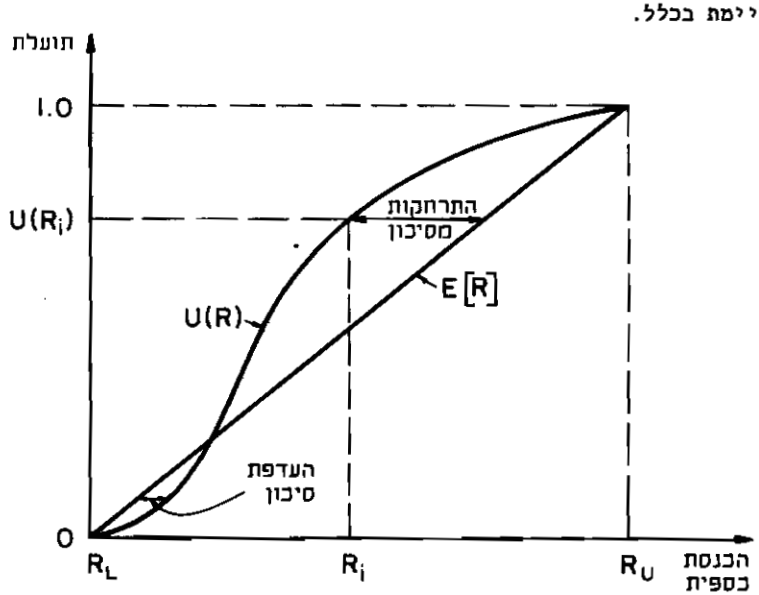
נהוג לבחור גבול תחתון ועליון של ההכנסה האפשרית R_L ו-R_U בהתאמה. נוח ליחס את התועלויות $U(R_L) = 0$, $U(R_U) = 1.0$. עבור ערכים שונים של R_i בונים את העקום.

$$U(R_i) = p_i U(R_U) + (1 - p_i)U(R_L)$$

כאשר המחליט קובע עבור כל ערך רצוי של R_i את ערכו של p_i. אם בוחרים $U(R_L) = 0$ ו-$U(R_U) = 1.0$ מתקבל $U(R) = p$. התוחלת של ההגרלה כין R_U בהסתברות ובין p בהסתברות $(1-p)$ היא

$$E[R_i] = pR_U + (1-p)R_L$$

פונקציה זו מתוארת על ידי הקו הישר $E[R]$ שבציור 3-2. ההפרש האופקי $E[R]-U[R]$ מבטא את מידת <u>ההתרחקות מסיכון</u> (Risk Aversion). כאשר הפרש זה שלילי הוא מבטא <u>העדפת סיכון</u>. העדפה זו באה לידי ביטוי בדרך כלל רק בערכים הנמוכים של R, אם היא קיימת בכלל.



<u>ציור 3.2</u>: תועלת כפונקציה של הכנסה כספית.

מקורות

Benjamin and Cornell, C.A., "Probability, Statistics and Decision for Civil Engineers", McGraw-Hill, 1970.

Keeney, R.L., and Raiffa, H., "Decisions with Multiple Objectives", John Wiley and Sons, 1976.

Luce, R.D. and H. Raiffa, "Games and Decisions", John Wiley and Sons, 1957.

Von Neumann, J. and O. Morgenstern, "Theory of Games and Economic Behavior", Princeton University Press, 1953.

פרק רביעי

מבוא לאופטימיזציה

אופטימיזציה היא נושאו של התכנות המתמטי (Mathematical Programming) ענף של המתמטיקה שהתפתח בעיקר בעידן המחשבים. שיטות אופטימיזציה מסויימות היו קיימות עוד קודם לכן, אבל כפי שנראה בפרקים הבאים מהווה השימוש במחשב תנאי הכרחי ברוב בעיות האופטימיזציה המעשיות בהן נתקלים.

4.1 בעיות אופטימיזציה וניסוחן

4.1.1 משתני החלטה ופרמטרים

משתנים נחלקים למשתני החלטה (Decision Variables) ולפרמטרים (Parameters). אנו נקרא משתנים לכל הגדלים היכולים לקבל ערכים שונים במצבים או בזמנים שונים, בבעיות הזהות מבחינה עקרונית. למשל: קוטר צינור, אורך מפתח, מספר המשאבות הפועלות ברגע מסוים, צריכת המים ברגע מסוים בנקודה מסויימת ברשת, משך הביצוע של שלב בפרויקט וכדומה.

משתני החלטה הם אלה שאת גודלם יש לקבוע בפתרון הבעיה. אגב, קביעת גודל אפס למשתנה החלטה, כלומר, הוצאתו מן הפתרון, היא אחת השיטות לבחירה בין משתנים.

במשוואות המתארות את בעית האופטימיזציה מופיעים מקדמים מספריים. חלקם קבועים ואינם נתונים לשינויים. אחרים משתנים, אבל איננו מתבוננים לקבוע את ערביהם באופן מפורש בבעית האופטימיזציה הנדונה. לאלו נקרא פרמטרים. הם אינם משתני החלטה, אבל ערביהם אינם קבועים מראש, אינם ידועים בדיוק מספיק או עשויים להשתנות בתנאים שונים, לדוגמא, אם אנו מבקשים לקבוע את המידות האופטימליות של קורה מבטון מזויין הנושאת עומס נתון, יופיעו במשוואות המתארות את התנהגות הקורה מקדמים המבטאים את תכונות הבטון והפלדה. תכונות אלה אינן משתני החלטה, שהרי אין אנו מבקשים לקבוע אותן בניסוח זה של בעיית האופטימיזציה. עם זאת, ירוע שאפשר להשתמש בחומרים שונים, בעלי תכונות שונות, וכי גם לאחר בחירת החומר משתנות תכונותיו עקב שינויים אקראיים בייצור החומרים ובהכנת הקורה.

לכן נקרא למקדמים המכטאים את התכונות בשם פרמטרים, ונבדוק את השפעת השתנותם (המכוונת או האקראית) על הפתרון האופטימלי. נשים לב שאפשר לנסח בעיית אופטי-מיזציה בה המידות תיקבענה מראש ויידרש למצוא את תכונות החומרים הנותנות פתרון אופטימלי. יתר-על-כן, אפשר לנסח בעיית אופטימיזציה בה הן המידות והן התכונות יהיו משתני החלטה. פתרונה של הבעיה האחרונה יהיה ככל הנראה קשה הרבה יותר מפתרון כל אחת משתי הבעיות שנוסחו קודם.

בעת ניסוח בעיות אופטימיזציה יש להגדיר את כל המשתנים, לחלקם למשתני החלטה ולפרמטרים, ולרשום את תחומי ההשתנות האפשריים של כל אחד. בשיטות אופטימיזציה אחדות קובע מספר משתני ההחלטה את דרגת הקושי של הפתרון, ואילו בשיטות אופטי-מיזציה אחדות מספר זה הוא בעל השפעה משנית בלבד.

4.1.2 אילוצים

באופן כללי אילוצים (Constraints) הם תנאים שעל הפתרון הנבחר לקיים. באופן ספציפי, אלה תנאים הנדרשים לגבי משתנה או קבוצת משתנים. בדרך כלל יש מספר גדול של פתרונות המקיימים את כל האילוצים. כל פתרון כזה נקרא פתרון אפשרי (Feasible Solution).

אילוצים יכולים להיות טכנולוגיים, חוקתיים, חברתיים, אסתטיים, כספיים וכדומה. אילוץ הוא תנאי שהפתרון חייב למלא. לכן צריך להזהר, ולקחת כאילוצים רק אותם תנאים שבאמת אינם ניתנים להפרה. לעיתים תכופות מדי מכניסים לניסוח הבעייה אילוצים שאינם ממש כאלה. למשל, בתכנון כלכלי של פרוייקט אסור להגביל שלב מסויים בו רק משום שהוא יקר מאוד כשלעצמו. יתכן שההוצאה הנוספת הכרוכה בביצוע שלב זה תהיה מוצדקת על ידי חסכון גדול ממנה בשלבים אחרים שהתאפשרו רק בגלל קיום השלב היקר; יש לתת את מחיר השלב, ולאפשר לשיטת האופטימיזציה לבטל אותו - אם אמנם הפרויקט הכולל שלב זה אינו אופטימלי.

יש מצבים בהם בחירת האילוצים וקביעת הערכים המספריים בהם אינה קלה כלל. לדוגמא, ישמשו הקריטריונים של איכות המים לשימושים שונים, של איכות האויר, או דרישות תקניות שונות ביחס לעמידתם של מבנים בעומסי רוח ורעידות אדמה. כל קביעה של קריטריונים אלה, המהווים אילוצים למערכות הנדסיות, הם בעלי משמעות מרחיקת לכת לגבי הפתרון האופטימלי. אפשר לבחון את השפעת האילוץ על ידי השוואת

שני פתרונות אופטימליים, האחד של הבעיה עם האילוץ והשני של בעיה הזהה בכל לראשונה, פרט לכך שהאילוץ הנדון בוטל. כאשר מספר האילוצים מסוג זה הוא רב, הופכת בדיקה זו ליקרה מדי, או לבלתי אפשרית כלל, אם כי במקרים מיוחדים אפשר לחשב את ההשפעה שתהיה לשחרור חלקי או מלא של האילוץ בלי שיהיה צורך לפתור את הבעיה מחדש. על כל פנים, בדיקת רגישות הפתרון האופטימלי לשינויים באילוצים היא בדרך כלל חלק חשוב מניתוח הבעיה.

4.1.3 פונקצית המטרה

פונקצית המטרה (Objective Function) היא ביטוי כמותי של טיב הפתרון. היא יכולה להיות ביטוי של "תועלת", ואז נרצה לבצע מקסימיזציה, או ביטוי של "נזק", ואז נרצה לבצע מינימיזציה.

היחידות בהן מבוטאת פונקצית המטרה יכולות להיות שונות, למשל, שטח, משקל, מספר מכוניות וכדומה. לעיתים קרובות אנו מבטאים ערכים כלכליים, כהם המכנה המשותף הוא כסף, ואז פונקצית המטרה היא ביחידות כספיות.

על פונקצית המטרה להיות טרנזיטיבית, כלומר: אם פתרון א' טוב מפתרון ב' לפי פונקצית המטרה, ופתרון ב' טוב מפתרון ג', אזי נדרש שלפי הפונקציה יהיה פתרון א' גם טוב מפתרון ג'. תכונה זו מבטיחה סולם עדיפויות חד-משמעי, ואין אפשרות ל"מעגלים סגורים" של העדפה. אך מה נעשה כאשר יש יותר מתכונה אחת רצויה בפתרון. למשל מחיר וערך אסתטי. לו היינו יכולים לתרגם את הערך האסתטי לערך כספי שקול, היתה לנו שוב פונקצית מטרה טרנזיטיבית פשוטה. לרוב איננו יכולים למצוא את שווה הערך של תכונה אחת מבוטא באמצעות תכונה אחרת, בעיקר כשמדובר בחיי אדם, באספקטים בריאותיים, בערכים אסתטיים וכדומה.

4.2 הבעיה הכללית של התכנות המתמטי

מצא $(x_1,\ldots,x_n)$ המקיימים את האילוצים,

$$g_i(\underline{x}) = g_i(x_1,\ldots,x_n) \quad \{\leq, =, \geq\} \quad b_i \qquad i=1,\ldots,m \qquad (4\text{-}1)$$

כך שפונקציית המטרה הנתונה

$$f(\underline{x}) = f(x_1,\ldots,x_n) \qquad (4\text{-}2)$$

תקבל ערך אקסטרמלי (מקסימום או מינימום).

לרוב קיימים גם <u>תנאים נוספים</u> (Side Conditions) כגון:

$$x_j \geq 0 \qquad j \in J_1 \qquad (4\text{-}3)$$

$$a_j \leq x_j \leq b_j \qquad j \in J_2 \qquad (4\text{-}4)$$

כאשר J_1 ו-J_2 הן קבוצות של האינדקס j המציינות את המשתנים עליהם חלים התנאים.

אפשר היה לכלול תנאים אלה באילוצים. הם נרשמים בנפרד בגלל הטיפול המיוחד לו הם זוכים בשיטת הפתרון. בכל פעם שסוג מסויים של אילוצים זוכה לטיפול נפרד בפתרון, נרשום אותם במפורש, ונקרא להם תנאים נוספים.

לעיתים נדרש גם שחלק מהמשתנים יהיו בדידים (Discrete). כל דרישה כזו ניתן לתרגם, על ידי שינוי ביחידות בהן נמדדים המשתנים, לתנאי מהצורה,

$$x_j \quad \text{integer} \qquad j \in J_3$$

בעיות כהן מופיע תנאי כזה דורשות טיפול מיוחד, המקשה לרוב את הפתרון במידה ניכרת.

האילוצים ביחד עם כל התנאים הנוספים, מגדירים במרחב ה-n מימדי, E^n, את קבוצת הנקודות המקיימות את כל התנאים. קבוצה זו נקראת <u>האיזור האפשרי</u> (Feasible Region). צורתו של האיזור האפשרי, כיחד עם צורת פונקציית המטרה, קובעים במידה רבה את שיטת האופטימיזציה.

הקושי החישובי בפתרון נקבע כשיטות האופטימיזציה השונות לפי דברים שונים. בשיטות אחדות קובע בעיקר מספר האילוצים, ומספר המשתנים רק באופן חלקי (תכנות ליניארי). בשיטות אחרות קובע בעיקר מספר המשתנים. כאשר מספר האילוצים מהווה גורם משני, אשר לפעמים אפילו תורם לחסכון במאמץ החישובי (תכנות דינמי).

לבעיה הכללית אין פתרון כללי. מבחינים במקרים מיוחדים של הבעיה הכללית, שעבורם קיימים אלגוריתמים לפתרון. מקרים אלה שימושיים מאד, אבל לא פעם נתקלים בבעיה שאינה מתאימה למקרים המיוחדים ואז יש לפנות לשיטות פחות אלגנטיות ולעתים מאולתרות.

4.3 אופטימום מקומי ואופטימום גלובלי

אנו קוראים אופטימום מקומי (Local Optimum) לנקודה בה ערך פונקצית המטרה טוב יותר מערכי פונקצית המטרה בכל הנקודות בסביבה הקרובה לנקודה (או לפחות שווה להם). בבעית מקסימיזציה רומה הדבר לפסגתה של גבעה, ממנה אפשר רק לרדת בכל כיוון שנבחר. האופטימום הגלובלי (Global Optimum) הוא הנקודה האפשרית בה ערך פונקצית המטרה טוב יותר מאשר בכל הנקודות האפשריות האחרות (או לפחות שווה להם). בבעית המקסימיזציה זו הפסגה הגבוהה ביותר מבין כל אלה שעליהן מותר להמצא.

רוב האלגוריתמים של התכנות המתמטי מבוססים על שיטה למציאת אופטימום מקומי. כאשר הפונקציות המופיעות באילוצים ובפונקצית המטרה מקיימות תנאים מסויימים, ואפשר להוכיח כי כל אופטימום מקומי (בעיקר אם יש רק אחד כזה) הוא גם האופטימום הגלובלי. אזי מבטיחה השיטה כי האופטימום הגלובלי יושג. כאשר תנאים אלה אינם מתקיימים מובטח לכל היותר שמגיעים לאופטימום מקומי.

4.4 תנאי קון-טאקר

התיאוריה של קון וטאקר (Kuhn-Tucker Theory) היא הרחבה של שיטת כופלי לגרנז' הקלסית (ראה סעיף 5.4) לבעיות אופטימיזציה עם אילוצי אי-שוויון.

כדי לפשט את ההצגה, נכתוב את בעית התכנות המתמטי הכללי שנוסחה במשוואות (4-1) ו-(4-2) בצורה שונה וקצת מוגבלת יותר כדלקמן:

$$\left.\begin{array}{lll} min\ f(x) & & \\ s.t.\quad g_i(\underline{x}) \le 0 & & i = 1,\dots m \\ x_j \ge 0 & & j \in J_1 \end{array}\right\} \qquad (4\text{-}6)$$

כאשר J_1 כוללת את האינדקסים של כל המשתנים המוגבלים להיות לא שליליים. עבור בעיה זו יוצרים פונקציה חדשה. הנקראת לעיתים פונקצית לגרנז' מוכללת, המוגדרת על ידי

$$F(\underline{x},\underline{\lambda}) = f(\underline{x}) + \sum_{i=1}^{m} \lambda_i g_i(\underline{x}) \qquad (4\text{-}7)$$

שבה $\underline{\lambda} = (\lambda_1,\dots,\lambda_m)$ הוא וקטור כופלי לגרנז'.

התיאוריה של קון וטאקר מוכיחה כי לבעיה (4-6) יש פתרון בנקודה $\underline{x}^*$ אם ורק אם קיים וקטור $\underline{\lambda}^*$ כך שלפונקציה $F(\underline{x},\underline{\lambda})$ יש אוכף בנקודה $(x^*,\underline{\lambda}^*)$, כלומר, קיים עבור כל $\underline{x} \ge \underline{0}$ ו-$\lambda \ge \underline{0}$ התנאי

$$F(\underline{x}^*,\underline{\lambda}) \le F(\underline{x}^*,\underline{\lambda}^*) \le F(\underline{x},\underline{\lambda}^*) \qquad (4\text{-}8)$$

כאשר,

$$\underline{x}^* \ge \underline{0} \qquad \underline{\lambda}^* \ge \underline{0} \qquad (4\text{-}9)$$

בנקודת אוכף זו, אם מחזיקים את $\underline{\lambda}^*$ קבוע ומשנים את $\underline{x}$ מערכו $\underline{x}^*$ בכיוון כלשהו תמיד עולים (הבעיה היתה מינימיזציה על $\underline{x}$), ואילו אם מחזיקים את $\underline{x}^*$ קבוע ומשנים את $\underline{\lambda}$ מערכו $\underline{\lambda}^*$ בכיוון כלשהו תמיר יורדים.

אם כל הפונקציות הנדונות גזירות, אפשר להראות כי התנאים לעיל אקוויולנטיים לתנאים,

$$\left.\begin{aligned} &\frac{\partial F(\underline{x}^*,\underline{\lambda}^*)}{\partial x_j} \geq 0 \\ &x_j^* \frac{\partial F(\underline{x}^*,\underline{\lambda}^*)}{\partial x_j} = 0 \\ &x_j^* \geq 0 \end{aligned}\right\} \quad j = 1,\ldots,n \qquad (4\text{-}10)$$

$$\left.\begin{aligned} &\frac{\partial F(\underline{x}^*,\underline{\lambda}^*)}{\partial \lambda_i} \leq 0 \\ &\lambda_i^* \frac{\partial F(\underline{x}^*,\underline{\lambda}^*)}{\partial \lambda_i} = 0 \\ &\lambda_i^* \geq 0 \end{aligned}\right\} \quad i = 1,\ldots,m \qquad (4\text{-}11)$$

במקרים בהם לא נדרשת אי-שליליות עבור המשתנה x_j יש להחליף את התנאים (4-10) בתנאי היחיד,

$$\frac{\partial F(\underline{x}^*,\underline{\lambda}^*)}{\partial x_j} = 0 \qquad j = 1,\cdots,n \qquad (4\text{-}12)$$

במשוואות אלה מחושבות הנגזרות בנקודה $(\underline{x}^*,\underline{\lambda}^*)$.

צורה אחרת של תנאי קון-טאקר, אשר אפשר להראות כי היא אקוילנטית לזו שבמשוואות (4-10) ו-(4-11), היא כדלקמן: עבור הבעיה (4-6) מתקיימים בנקודת האופטימום $\underline{x}^*$ התנאים הבאים,

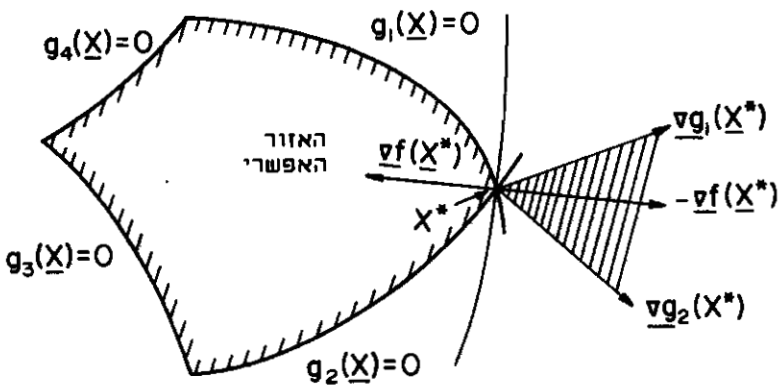
$$x_j^* \geq 0 \qquad j = 1,\ldots,n \qquad (4\text{-}13)$$

$$\lambda_i^* g_i(\underline{x}^*) = 0 \quad i = 1,\ldots,m \qquad (4\text{-}14)$$

$$\underline{\nabla f}(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \underline{\nabla g_i}(\underline{x}^*) = 0 \qquad (4\text{-}15)$$

$$\lambda_i^* \geq 0 \qquad i = 1,\ldots,m \qquad (4\text{-}16)$$

התנאי (4-13) מבטיח כי כל הפתרון אפשרי. (4-14) מבטיח שעבור כל אילוץ קיים אחד משני מצבים אפשריים - או שהאילוץ כובל ואז $g_i(\underline{x}^*)=0$ ו-$\lambda_i^* \neq 0$, או שהאילוץ אינו כובל ואז $\lambda_i^* = 0$. את התנאי (4-15) נסביר באופן גיאומטרי בעזרת ציור 4.1.



<u>ציור 4.1</u>: מובן גיאומטרי של תנאי קון-טאקר.

$\underline{\nabla g_i}(\underline{x}^*)$ הוא הגרדיאנט של האילוץ ה-i בנקודה $\underline{x}^*$. היות והאילוצים ב-(4-6) דורשים $g_i(\underline{x}) \leq 0$ הרי שהגרדיאנט פונה אל מחוץ לאיזור האפשרי. הגרדיאנט $\underline{\nabla f}(\underline{x}^*)$ פונה אל תוך האיזור האפשרי, בכיוון העליה של $f(\underline{x})$, שכן אנו מכצעים מינימיזציה. עבור הדוגמה שבציור 4.1 קיים לפי (4-14) $\lambda_3^* = \lambda_4^* = 0$, שכן האילוצים המתאימים אינם כובלים. מתוך (4-15) דרוש כי באופטימום יתקיים

$$-\underline{\nabla f}(\underline{x}^*) = \lambda_1^* \underline{\nabla g_1}(\underline{x}^*) + \lambda_2^* \underline{\nabla g_2}(\underline{x}^*) \qquad (4\text{-}17)$$

עבור $\lambda_1^*, \lambda_2^* > 0$. הפירוש הגיאומטרי הוא שעל וקטור הגרדיאנט השלילי של פונקציית המטרה להיות בקונוס הקמור המוגדר על ידי הגרדיאנטים של האילוצים הפעילים בנקודת האופטימום. הקונוס הקמור, המקווקו בציור 4.1, הוא האיזור המוגדר על ידי קומבינציה ליניארית עם מקדמים לא שליליים, של וקטורי הגרדינטים.

קל לראות מדוע על תנאי זה להתקיים. נתאר לעצמנו כי $-\nabla \underline{f}(\underline{x}^*)$ כציור 4.1 לא היה בתוך הקונוס הקמור המקווקו אלא היה למשל מעל $\nabla \underline{g}_1(x^*)$. במצב כזה היה לוקטור $-\nabla \underline{f}(x^*)$ רכיב לאורך האילוץ $g_1(\underline{x}) = 0$ ככיוון הרחק מחיתוך האילוצים. ברור כי אז לא היה חיתוך האילוצים הנקודה האופטימלית, שכן כדאי להתרחק מנקורה זו עד שמתקיים תנאי (4-15).

במקדה שבנקודת האופטימום יש רק אילוץ פעיל אחד פירוש התנאי (4-15) הוא כי על $-\nabla \underline{f}(\underline{x}^*)$ להתלכד עם הגרדיאנט של האילוץ - מה שברור מבחינה גיאומטרית.

יש לציין כי על מנת שבנקודת אופטימום מאולץ יתקיימו תנאי קון-טאקר יש לדרוש מן האילוץ לקיים תנאי מגביל (Constraint Qualification) המוסבר בצורה גיאומטרית בעזרת ציור 4.1 כדלקמן: על הקונוס הקמור של וקטורי הגרדיאנטים השליליים של האילוצים (תמונת הראי של האיזור המקווקו בציור) להיות כולו בתוך האיזור האפשרי, לפתות בסביבה הקרובה של נקודת האופטימום. היות ותנאי זה מתקיים בדרך כלל עבור בעיות מעשיות, ורק בדוגמאות מלאכותיות שהוכנו לצורך זה הוכח שייתכנו מצבים בהם אינו מתקיים, מניחים כי הוא קיים מבלי לברוק.

תנאי קון-טאקר מהווים חלק מהתיאוריה הבסיסית של תורת התכנות המתמטי. נוסף לכך הם מהווים את הבסיס של כמה שיטות חישוביות לפתרון כעיות אופטימיזציה לא ליניארית. שיטות אלה מתבססות על מעבר מנקודה אחת לשניה, באשר בכל אחת נבדק האם מתקיימים תנאי קון-טאקר. אם כן - זהו הפתרון המבוקש. אם לאו - מוצאים בעזרת וקטורי הגרדיאנטים של האילוצים ושל פונקציית המטרה ביוון רצוי (כיוון שבו ישתפר הערך של פונקציית המטרה) שהוא גם אפשרי (כלומר, נמצא בקונוס הקמור של וקטודי הגרדיא-נטים השליליים של האילוצים הפעילים). מתקדמים בכיוון זה לנקודה חדשה ושוב מבצעים את הבדיקה לקיום תנאי קון-טאקר.

4.5 סיווג בעיות אופטימיזציה לפי הטכניקות לפתרונן

ניתן לסווג בעיות אופטימיזציה בצורות שונות. אפשר למשל לערוך חלוקה לפי אחד או יותר מהקריטריונים הבאים:

א. בעיות בהן כל המשתנים דטרמיניסטיים לעומת בעיות בהן לפחות חלק מהמשתנים אקראי.

ב. בעיות בהן הזמן מופיע כמשתנה הקובע את השלכים לעומת בעיות בהן לזמן אין תפקיד.

ג. חלוקה בין בעיות תכן לבין בעיות של תכנון ופקוח על הביצוע ולבין בעיות של תפעול.

אנו נסווג בעיות אופטימיזציה לפי המבנה המתמטי שלהן, כלומר, לפי הטכניקה המתאימה לפתרונן ונמנה רק את הטכניקות החשובות ביותר. במקרים אחדים נזכיר גם באם יכולה הטכניקה לטפל במשתנים אקראיים - שכן זהו למעשה המצב תמיד בפרויקטים הנדסיים. אקראיותם של משתנים ופרמטרים מסויימים היא לעיתים בעלת חשיבות משנית בלבד, אך לעיתים יש לאקראיות של תהליכי הטבע השפעה מכרעת.

4.5.1 תכנות ליניארי (Linear Programming, or LP)

כאשר פונקציית המטרה וכל האילוצים ליניאריים ואין נדרשים תנאים נוספים ממשתני ההחלטה - פרט לכך שיהיו לא-שליליים - מתקבלת בעיית התכנות הליניארי

$$max(\text{or } min)\ f(\underline{x}) = \sum_{j=1}^{n} c_j x_j \qquad (4\text{-}18)$$

$$\text{s.t. } g_i(\underline{x}) = \sum_{j=1}^{n} a_{ij} x_j \{\leq\ ;\ =\ ;\ \geq\} b_i \qquad i = 1,\ldots,m \qquad (4\text{-}19)$$

$$x_j \geq 0 \qquad j = 1,\ldots,n \qquad (4\text{-}20)$$

לבעיה זו ולטכניקת הפתרון שלה מוקדש בהמשך פרק שלם. הטכניקה מפותחת היטב ותוכניות מחשב לפתרונה זמינות ביותר. קשיים מתעוררים כאשר חלק מן המקדמים המופיעים באילוצים ו/או בפונקצית המטרה (c_j-ו b_i, a_{ij}) הם משתנים אקראיים. קיימות שיטות לפתרון מקרים מיוחדים של תכנות ליניארי סטוכסטי, אך אלו מוגבלות ביכולתן ואינן כלליות. קושי אחר מתעורר כאשר נדרש כי כל משתני ההחלטה או חלקם יקבלו ערכים כדידים בלבד. זהו הנושא של תכנות ליניארי בשלמים (Integer LP) ושל תכנות ליניארי מעורב (Mixed Integer LP), בהתאמה.

4.5.2 תכנות ריבועי (Quadratic Programming, QP)

כאשר כל האילוצים ליניאריים ופונקצית המטרה היא צורה ריבועית, כלומר, מופיעים בה משתני ההחלטה רק בחזקה ראשונה, בחזקה שניה ובמכפלות של שני משתנים לכל היותר, מתקבלת בעית התכנות הריבועי

$$max(\text{or } min)\ f(\underline{x}) = \sum_{j=1}^{n} c_j x_j + \underline{x}^T \underline{\underline{D}} \underline{x} \qquad (4\text{-}21)$$

$$\text{s.t.}\quad g_i(\underline{x}) = \sum_{j=1}^{n} a_{ij} x_j \ \{\leq\ ;\ =\ ;\ \geq\} b_i \qquad i = 1,\ldots,m \qquad (4\text{-}22)$$

$$x_j \geq 0 \qquad j = 1,\ldots,n \qquad (4\text{-}23)$$

כאשר $\underline{\underline{D}}$ היא מטריצה ריבועית סימטרית מסדר n (ראה נספח א'). קיימות כמה טכניקות לפתרון, שהנפוצות ביותר הן זו של Beale ושני הואריאנטים של הטכניקה של Wolfe (ראה למשל פרק 2 בספרם של Kunzi et al.). הטכניקות מבטיחות כי יושג האופטימום הגלובלי רק במקרים הבאים: בעית מינימיזציה בה $\underline{\underline{D}}$ מוגדרת חיובית (או, בטכניקות מסויימות מספיק שתהיה מוגדרת חיובית למחצה), או בעית מקסימיזציה בה $\underline{\underline{D}}$ מוגדרת שלילית (או שלילית למחצה). ראה סעיף 15.א' בנספח א' להבהרת תכונות אלה של מטריצות.

4.5.3 תכנות פריק (Separable Programming)

בעיית התכנות הפריק מוגדרת על ידי

$$max(\text{or } min)\ f(\underline{x}) = \sum_{j=1}^{n} f_j(x_j) \qquad (4\text{-}24)$$

$$\text{s.t. } g_i(\underline{x}) = \sum_{j=1}^{n} g_{ij}(x_j)\ \{\leq\ ;\ =\ \geq\} b_i \qquad i = 1,\ldots,m \qquad (4\text{-}25)$$

$$x_j \geq 0 \qquad j = 1,\ldots,n \qquad (4\text{-}26)$$

הן פונקצית המטרה והן האילוצים הם סכום של פונקציות, כל אחת תלויה במשתנה החלטה אחד בלבד. הפונקציות $f_j(x_j)$ ו-$g_{ij}(x_j)$ אינן חייבות להיות ליניאריות (כאשר בולן ליניאריות הופכת הבעיה לתכנות ליניארי).

לבעיה זו אין טכניקת פתרון כללית, אבל אפשר לחקוף אותה על ידי ליניאריזציה בקטעים של כל הפונקציות ושימוש איטרטיבי בתכנות ליניארי, ראה למשל פרק 4 בספרו של Hadley, "Nonlinear and Dynamic Programming".

4.5.4 תכנות גיאומטרי (Geometric Programming, or GMP)

בעיית התכנות הגיאומטרי מוגדרת על ידי

$$min\ f(\underline{x}) = \sum_{p=1}^{k} c_p [\prod_{j=1}^{n} x_j^{a_{pj}}] \qquad (4\text{-}27)$$

$$\text{s.t. } g_i(\underline{x}) = \sum_{p=1}^{k} d_{ip} [\sum_{j=1}^{n} x_j^{b_{ipj}}] \leq 1 \qquad i = 1,\ldots,m \qquad (4\text{-}28)$$

$$x_j \geq 0 \qquad j = 1,\ldots,n \qquad (4\text{-}29)$$

$$c_p \geq 0 \qquad p = 1,\ldots,k \qquad (4\text{-}30)$$

$$d_{ip} \geq 0 \qquad i = 1,\ldots,m \;;\; p = 1,\ldots,k \qquad (4\text{-}31)$$

פתרונה של בעיה זו ושימושיה הוא נושאו של הספר "Geometric Programming" מאת Duffin, Peterson and Zener. כמו כן אפשר למצוא טיפול בבעיה בספרים של Zangwill ושל Wilde and Beightler.

4.5.5 תכנות דינמי (Dynamic Programming, or DP)

תכנות דינמי היא טכניקה המשמשת לפתרון בעיות אופטימיזציה בעלות מבנה מיוחד. אי אפשר לתת ניסוח מתמטי אחיד לבעיות אלה. יאמר רק כי אלה הן בעיות רב שלביות (Multi-Stage) או בעיות שניתן לפרק אותן למספר כעיות משנה בעלות מבנה עקרוני זהה, ולטפל בהן בזו אחר זו - כאילו היו שלבים עוקבים בבעיה המקורית. העקרון המונח ביסודה של השיטה, הנקרא עקרון האופטימליות (Principle of Optimality) הוא פשוט למדי, ועם זאת טובלת השיטה מקשיים חישוביים חמורים בטיפול בבעיות רב-מימדיות. ניתן לפתור בתכנות דינמי בעיות דטרמיניסטיות וסטוכסטיות כאחת.

ספרים רבים נכתבו על הטכניקה והשימוש כה. הקורא המעונין יפנה לספריהם של Bellman (הנחשב לאבי השיטה), של Nemhauser ושל Larson. פרק 7 מוקדש לתכנות דינמי.

4.5.6 שיטות קלסיות

השיטות הקלסיות כוללות איפוס נגזרות, שיטת כופלי לגרנז' לבעיות המאולצות על ידי שוויונות וחשבון וריאציות לאופטימיזציה של אינטגרלים. לשיטות אלה מוקדש פרק בהמשך.

4.5.7 תכנות לא-ליניארי (Nonlinear Programming, or NLP)

בקטגוריה זו נכללות כל הבעיות בהן חלק מהפונקציות אינו ליניארי, אך באופן מצומצם יותר אנו מתכוונים לאותן בעיות לא-ליניאריות שאינן ניתנות לפתרון בעזרת אחת מהטכניקות שכבר נמנו לעיל או בשיטה אחרת המיוחדת להן. לא נמנה כאן את כל השיטות שפותחו לטיפול בבעיה הלא-ליניארית הכללית, זו שנוסחה במשוואות (4-1) ו-(4-2), שכן יש עשרות רבות כאלו. נציין רק את חלוקת הטכניקות לשני סוגים עיקריים:

א. שיטות של חיפוש ישיר (Direct Search Technique)

כל אחת משיטות אלה היא איסטרטגיה לחתקדמות מפתרון אחד למשנהו, במטרה להגיע לפתרון שהוא טוב יותר מכל אלה שנקבעו ומכל אלה שבסביבתו הקרובה. שיטות אלו משתמשות רק בערך של פונקצית המטרה בנקודות עוקבות של החיפוש על מנת להחליט על כיוון ההתקדמות המבטיח ביותר. חלקן של שיטות אלה מרשות בעת החיפוש בדיקה רק של נקודות בהן הפתרון אפשרי, ואילו אחרות מרשות בריקת נקודות שאינן מקיימות את האילוצים - תוך מתן קנס אל אי-קיומם - בכוונה להגיע בסופו של דבר לפתרון אפשרי משופר.

ב. שיטות חיפוש (Search Techniques)

גם שיטות אלה הן איסטרטגיות להתקדמות מפתרון אחד לפתרון טוב ממנו, אלא שהן משתמשות בנוסף לערך של פונקצית המטרה גם בתכונות אחרות שלה כגון הגרדיאנט או המטריצה ההסיאנית שלה (ראה משוואות 5-6 ו-5-8). השיטה המוכרת ביותר אולי היא שיטת הגרדיאנט. בה מתקדמים מרחק מה בכיוון הגרדיאנט של פונקצית המטרה, ומחשבים אותו שוב בנקורה החדשה, מתקדמים בכיוון החדש וכך הלאה. התיפוש נפסק כאשר אי אפשר לשפר יותר, וזה קורה בנקודת אופטימום מקומי.

הספרות בנושא זה רבה מאוד. אך כנקודת התחלה ללימודו אפשר לפנות לאחד או יותר מן הספרים הבאים: של Zangwill, של Mangasarian או של Fiacco and McCormick.

4.5.8 סימולציה (Simulation)

סימולציה אינה טכניקה ישירה לאופטימיזציה, אבל היות ומשתמשים בה ככלי לשיפור פתרונן של בעיות מסויימות - בעיקר של כעיות תפעול מפעלים הנדסיים - מן הראוי להזכיר אותה כאן. בסימולציה הכוונה כאן לטכניקה של ביצוע ניסויים

בעזרת מחשב על מורל מתמטי של המערכת. במהלכם של ניסויים אלה משנים את ערכיהם של משתני ההחלטה, לומרים מן הפתרונות המושגים כמטרה לשפר את הפתרון. הסימולציה מהווה תחליף נוח וזול יחסית לביצוע ניסויים אלה על גבי המערכת ההנדסית עצמה. עצם הצורך לבנות, לתכנת ולהריץ במחשב מודל מתמטי מלמד הרבה על המערכת ההנדסית עצמה. לאחר שרוכשים נסיון ואימון במורל ניתן לבדוק בו גם מצבים שעריין לא קרו באבטיפוס אבל אולי הם חזויים באפשרות לעתיר.

בסימולציה ניתן לטפל גם במשתנים אקראיים, כאשר השיטה לכך נקראת לפעמים "שיטת מונטה קרלו". בשיטה זו מייצרים נתונים אקראיים - כביכול, הנבחרים כך שמבחינה סטטיסטית הם רומים למשתנים האקראיים עצמם. נתונים אלה משמשים לביצוע ניסויים מספריים כמודל פעמים רבות, לקבלת מרגם מייצג של התוצאות. ניתוח סטטיסטי של התוצאות נותן את הבסיס להערכה טובה של האלטרנטיבה הנבחנת.

4.6 בעיות החלטה רב-קריטריוניות

בעית ההחלטה הרב-קריטריונית היא

$$max\{\underline{f}(\underline{x}) = [f_1(\underline{x}),\ldots,f_p(\underline{x})]\} \qquad (4\text{-}32)$$

בכפיפות לאילוצים שיכתבו בצורה הכללית ביותר

$$\underline{x} \in X \qquad (4\text{-}33)$$

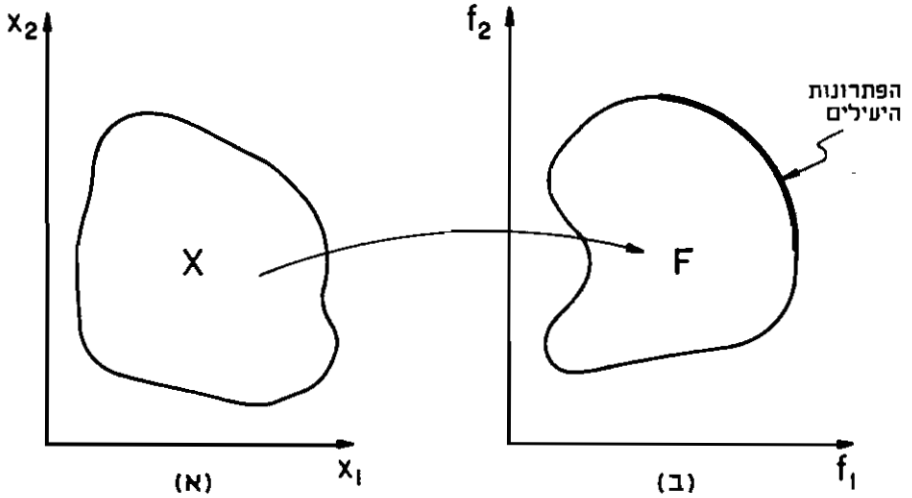
משמעות (4-32) היא שברצוננו להביא למקסימום כל אחת מ-p פונקציות המטרה בפני עצמה. יש אולי מקרים שבהם המטרות (קריטריונים) אינן מתחרות זו בזו, ואז אפשר להשיג בבת אחת מקסימום של כולן. אך זהו מקרה נדיר - ובלתי מעניין. אותנו מעסיקות במציאות בעיות שבהן יש תחרות בין המטרות וצריך לוותר מן ההישג באחת על מנת להשיג יותר באחרת.

כאשר מטרה מנוסחת למינימום ניתן להפכה למקסימיזציה, שכן $min f = max(-f)$.
לנוחות הדיון נניח לכן, מכלי לוותר על כלליות, כי כל הפונקציות הן למקסימיזציה.

כל מטרה מבוטאת ביחידותיה שלה: יחידות של כסף, כמות פיזית, זמן, מספר אנשים וכדומה. אחת השיטות לטיפול בבעיות רב-קריטריוניות היא לבטא את כל המטרות באותן יחידות, כלומר להביא אותן למכנה משותף. היחידה המקובלת לצורך זה היא הכסף. בשיטה זו מכטאים כל מטרה ביחידות של כסף, ואז מותר לחבר את ערכי כל המטרות ולקבל פונקצית מטרה יחידה. אלא שלשיטה זו מספר חסרונות, וביניהם שהיא דורשת מתן ביטוי כלכלי לערכים שאין אנו מוכנים להתייחס אליהם התייחסות כלכלית בלבד, כמו למשל איכות הסביבה או מספר המובטלים במשק. אנו מעדיפים לכן להשאיר את המטרות ביחידותיהן המקוריות ולטפל בהן כמאפיינים רבים של הפתרון. במצב זה לא נוכל יותר לדבר על פתרון אופטימלי אלא על <u>פתרון פשרה</u> (Compromise Solution) זהו פתרון אשר בו יש איזון המקובל על מקבל ההחלטה (Decision Maker) בין המטרות השונות.

בעשור האחרון התפתחו גישות ושיטות רבות לנושא. תקצר כאן היריעה מלכסות את הנושא, ואנו רק נסקור מספר מושגי יסוד וגישות המתייחסות לתכנות מתימטי רב-מטרתי. אי לכך, לא נטפל בבעיות שבהן יש לבחור בין מספר אלטרנטיבות נתונות לפי מספר קריטריונים, אלא בבעיות שבהן יש אילוצים (4-33) ו-p פונקציות מטרה (4-32).

בציור 4.2 מתוארים סכמטית (א) האזור האפשרי X, במרחב ההחלטות (Decision Space) ו-(ב) האזור האפשרי, F, במרחב המטרות (Objective Space), שהוא המיפוי של X למרחב של p המטרות. האזור האפשרי X הוא n - מימדי והאזור האפשרי F הוא p-מימדי, כאשר כבעיות מעשיות $p << n$. מספר המטרות אינו עולה לרוב על שלוש עד חמש. בציור 4.2 $n = p = 2$.



ציור 4.2: האזור האפשרי: (א) במרחב ההחלטות.
(ב) במרחב המטרות.

בציור 4.2(ב) מודגש אוסף הפתרונות היעילים (Efficient) הנקראים גם פתרונות לא-נשלטים (Non-Dominated), פתרונות לא-נחותים (Non-Inferior) או חזית היעילות (Efficiency Fronfier). זהו אוסף כל הפתרונות אשר מהם אי-אפשר לנוע באופן המשפר מטרה אחת מבלי שאחרת "תתקלקל". בכך דומים פתרונות אלה לפונקצית הייצור שהוזכרה בסעיף 2.1.

סביר הוא כי אם ידוע לנו אוסף הפתרונות היעילים (ואין זה חמיד קל או אפילו אפשרי "לייצר" את כל הפתרונות היעילים) הרי שפתרון הפשרה ייבחר מתוכם. אין זה הגיוני שנבחר נקודה לא יעילה, שכן מנקודה כזו ניתן לשפר לפחות פונקצית מטרה אחת מבלי לגרום לירידה בערכן של האחרות.

לאחר שניבחר את פתרון הפשרה, מכיון הנקודות היעילות, נוכל לאמר (בדיעבד) מהו השקלול היחסי שלנו בין המטרות, מבוטא ב-"כך וכך יחידות של מטרה i שוות יחידה אחת של מטרה j". השקלול שווה לשיפוע המשיק לחזית היעילות בנקודה של פתרון הפשרה הנבחר.

ניצור פונקצית מטרה יחידה, $f_e(\underline{x})$, ע"י שקלול כל המטרות:

$$f_e(\underline{x}) = \sum_{k=1}^{p} w_k f_k(\underline{x}) \tag{4-34}$$

נהוג לבחור את המשקלות, w_k, כך שיתקיים $\sum_{k=1}^{p} w_k = 1$. עבור ערכים כלשהם יתן $max\ f_e(\underline{x})$ נקודה על חזית היעילות. ע"י שינוי פרמטרי של המשקלות ניתן לייצר את כל חזית היעילות, אך תהליך זה יהיה בדרך כלל יקר מדי.

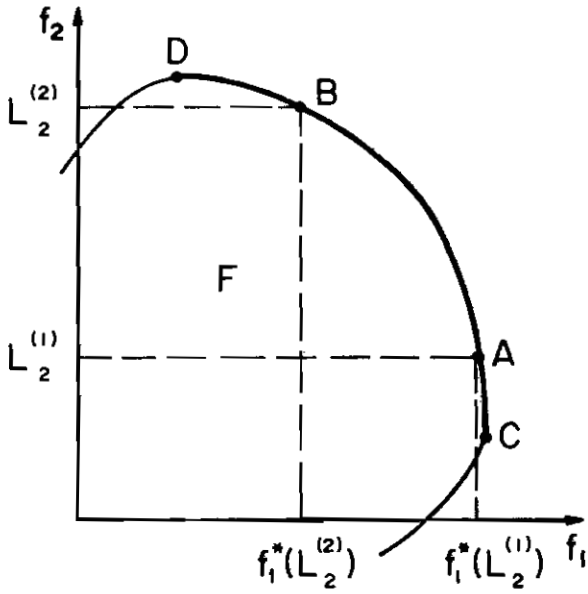
דרך שונה להתוות חזית היעילות היא שיטת האילוצים, אשר בה פותרים שורה של בעיות שצורתן:

$$max\ f_\ell(\underline{x}) \tag{4-35}$$

$$s.t. \quad \underline{x} \in X \tag{4-36}$$

$$f_k(\underline{x}) \geq L_k \qquad k = 1,\ldots,p\ ;\ k \neq \ell \tag{4-37}$$

L_k הם חסמים תחתונים על ערכי המטרות השונות מלבד זו המובאת למקסימום. ע"י שינוי פרמטרי של החסמים L_k ניתן לייצר את חזית היעילות, שכן עבור מערך אחד של ערכי החסמים תיתן $max\ f_\ell(\underline{x})$ נקודה יעילה, כמתואר בציור 4.3.



<u>ציור 4.3</u>: התווית חזית היעילות בשיטת האילוצים

נקודה A מתקבלת כ-$max\ f_1(\underline{x})$ כפוף ל-$f_2(\underline{x}) \geq L_2^{(1)}$ ונקורה B מתקבלת כ-$max\ f_1(\underline{x})$ כפוף ל-$f_2(\underline{x}) \geq L_2^{(2)}$, כאשר $L_2^{(1)}$ ו-$L_2^{(2)}$ הם ערכים פרמטריים של החסם.

הנקודות C ו-D הן קצות חזית היעילות. ב-C הגיעה $f_1(\underline{x})$ לערכה הגבוה ביותר וב-D הגיעה $f_2(\underline{x})$ לערכה הגבוה ביותר. כמקרה של שתי מטרות בלבד, כמתואר בציור, מתקבל הערך הגרוע (הנמוך) ביותר של כל מטרה באותה נקורה בה המטרה השניה מגיעה למקסימום. כאשר יש יותר משתי מטרות עדיין נכון הדבר שכל מטרה תגיע לערכה הגרוע ביותר באחת מן הנקודות שבהן מגיעות המטרות האחרות כל אחת לערכה המקסימלי, אלא שאי אפשר לרעת מראש באיזו מהן יהיה הדבר.

ע"י שינוי מודרג של L_k אפשר לייצר מספר מספיק של נקודות על פני חזית היעילות בכדי לתארה - בדרך גרפית במקרים דו - או אפילו תלת-מימדיים ובצורה טבולרית כאשר יש יותר מטרות.

מידע על חזית היעילות יסייע למקבל ההחלטות, שכן יוכל ללמוד על יחסי התחלופה בתחומיה השונים ויוכל לקבל לאורן החלטה או להנחות את המשך הניתוח.

שיטת השקלול ושיטת האילוצים דומות בכך ששתיהן יוצרות את חזית היעילות או חלקים ממנה. לכן הן נקראות שיטות יוצרות (Generating Methods). שיטות אלה שואפות לתת בידי מקבל ההחלטות מידע רב והולך בתהליך הניתוח, ולקבל ממנו הנחיות להמשך הניתוח (ערכי החסמים), עד שהוא מרוצה מן הפתרון.

אלו רק שתיים מן השיטות הרבות לטיפול בבעיות רב-קריטריוניות. המעונין יפנה לספרם של Keeney and Raiffa אשר בו הסבר של הגישות והטכניקות לטיפול בבעיות כאלה.

מקורות

Avriel, M., "Nonlinear Programming", Prentice-Hall, 1976.

Bellman, R., "Dynamic Programming", Princeton University Press, 1957.

Duffin, R.J., E.L. Peterson and C.M. Zener, "Geometric Programming", John Wiley and Sons, 1967.

Fiacco, A.V., and G.P. McGormick, "Nonlinear Programming: Sequential Unconstrained Minimization Techniques", John Wiley and Sons, 1968.

Hadley, G., "Linear Programming", Addison-Wesley, 1962.

Hadley, G., "Nonlinear and Dynamic Programming", Addison-Wesley, 1964.

Keeney, R.L., and Raiffa, H., "Decisions with Multiple Objectives", John Wiley and Sons, 1976.

Mangasarian, O.L., "Nonlinear Programming", McGraw-Hill, 1969.

Nemhauzer, G.L., "Introduction to Dynamic Programming", John Wiley and Sons, 1966.

פרק חמישי

שיטות אופטימיזציה קלאסיות

5.1 הקדמה

בפרק זה נדון בשיטות אופטימיזציה אחדות הנקראות קלאסיות. חשם נובע מן העובדה ששיטות אלה היו קיימות לפני עידן המחשבים. השיטות הן אנליטיות בעיקרן, אם כי יש בהן אספקטים חישוביים המצדיקים שימוש במחשב, או אפילו מחייבים אותו, באופן שבלי השימוש בו אין אפשרות מעשית להשיג את הפתרון המבוקש. שיטות אלה מוגבלות לבעיות בהן האילוצים מופיעים כשוויונות בלבד, ויש להן חשיבות מעשית רק לבעיות קטנות יחסית. עם זאת, יש לשיטות הקלאסיות משמעות תיאורטית חשובה, הבאה אף לידי ביטוי בשיטות אופטימיזציה מודרניות.

5.2.1 אקסטרמום בהעדר אילוצים

לפונקציה $f(\underline{x})$, המוגדרת בתחום הסגור X של המרחב ה-n מימדי E^n יש מקסימום מוחלט בנקודה $\underline{x}^*$ אם $f(\underline{x}) \le f(\underline{x}^*)$ עבור כל $\underline{x} \in X$. באשר מתקיים רק אי השוויון המקסימום נקרא מדויק (strict). המקסימום המוחלט נקרא גם מקסימום גלובלי (global). כאשר כיוון אי השוויון מתהפך מקבלים מינימום.

לפונקציה $f(\underline{x})$ יש מקסימום מקומי (local), הנקרא גם מקסימום יחסי, בנקודה $\underline{x}_o$ אם אפשר למצוא כך, שעבור כל $\underline{x}$ בתחום $0 < |\underline{x}-\underline{x}_o| < \varepsilon$ קיים $f(\underline{x}) \le f(\underline{x}_o)$. המקסימום הוא חזק אם רק אי השוויון מתקיים. באשר כיוון השוויון מתהפך מתקבל מינימום.

נשים לב כי המקסימום של $f(\underline{x})$ הוא המינימום של $-f(\underline{x})$.

הגישה הקלאסית מאפשרת מציאת נקודות של אקסטרמום מקומי. דרושות בדיקות נוספות כדי לבחור מתוכן את האקסטרמום הגלובלי. כדי למצוא את נקודות האקסטרמום בעזרת השיטה הקלאסית צריך להניח כי לפונקציה יש נגזרות ראשונות רציפות ביחס לכל המשתנים.

כדי למצוא את נקודות האקסטרמום המקומי מאפסים את כל הנגזרות החלקיות של $f(\underline{x})$:

$$\left.\frac{\partial f(\underline{x})}{\partial x_j}\right|_{x=x_o} = 0 \qquad j = 1,\ldots,n \tag{5-1}$$

או בסימון וקטוריאלי

$$\underline{\nabla f(\underline{x}_o)} = 0 \tag{5-2}$$

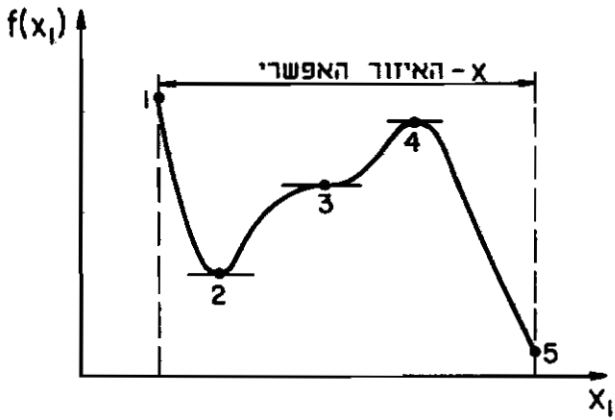
הערכים $\underline{x}_o = (x_{o_1},\ldots,x_{o_n})$ מתקבלים כפתרון של מערכת של n המשוואות (5-1) ב-n נעלמים. צורת המשוואות תלויה בצורתה של $f(\underline{x})$ והן כמובן יכולות להיות לא ליניאריות.

כאשר המשוואות ליניאריות מתקבל פתרון חד-משמעי, בתנאי כמובן, שאין בהן משוואות תלויות או סותרות. במלים אחרות, מתקבלת נקודה אחת $\underline{x}_o$ המקיימת את (5-1). כאשר המשוואות אינן ליניאריות פותרים את מערכת המשוואות (5-1) ברדך כלל בשיטה איטרטיבית, למשל בשיטת ניוטון-רפסון. כמקרה זה יכולות להתקבל נקורות אחדות המקיימות את (5-1). כולן נקראות <u>נקורות סטציוברית</u>. יכולות להיות ביניהן נקודות מקסימום ומינימום, אכל לא כולן בהכרח בקורות אקסטרמום מקומי. גם נקודות אוכף מקיימות את (5-1). על מנת לבחור את הנקודה המבוקשת בין אלה שהתקבלו בפתרון המערכת (5-1) יש להשתמש בתנאי הנגזרת השניה (ראה להלן), או פשוט לחשב את הערך של הפונקציה f בכל אחת מן הנקודות ולבחור את הנקודה הנותנת את הערך האקסטרמלי.

אם X הוא כל המרחב E^n, כלומר אין אילוצים, ול-$f(\underline{x})$ יש אקסטרמום גלובלי בנקודה $\underline{x}^*$ בך ש-$|\underline{x}^*|$ הוא סופי, אזי האקסטרמום הוא גם מקומי. במקרה זה אפשר לאתר את האקסטרמום הגלובלי על ידי בחירתו מבין נקודות האסטרמום המקומי. ברור כי במקרה שהאקסטרמום מתקבל כאשר $|\underline{x}^*| \to \infty$ הוא אינו אקסטרמום מקומי. כמו למשל המקסימום של $f(x) = x_1^2$ ב- $|x_1| \to \infty$.

5.2.2 אקסטרמום קונבחנת אולן בגם

במקרה המאולץ, כאשר X היא תת-קבוצה של E^n, כלומר, רק חלק מהמרחב ה-n מימדי, המוגדר על ידי אילוצים על המשתנים $(x_1,\ldots,x_n)$ אזי אפילו אם האקסטרמום הוא בנקודה $\underline{x}^*$ כך ש-$|\underline{x}^*|$ סופי, אין כטחון שזה גם אקסטרמום מקומי. ציור 5.1 מדגים את המקרה תחד-מימדי בו נקודות האקסטרמום הגלובלי מתקבלות כקצות התחום X בנקודות שאינן סטציונריות.



ציור 5.1: נקודות סטציונריות ונקודות אקסטרמום גלובליות כתחום מוגבל.

בציור מופיעות נקודה של מינימום מקומי (2), נקודת אוכף (3) ומקסימום מקומי (4). המקסימום הגלובלי על פני X הוא בנקודה (1), והמינימום הגלובלי בנקודה (5), שתיהן אינן נקודות סטציונריות.

כדי למצוא את האקסטרמום הגלובלי על פני תת-קבוצה X, יש לחפש בנוסף לנקודות סטציונריות שבתוך הקבוצה, גם על שפותיה. אם נסמן ב-$g_i(\underline{x}) = b_i$ $i = 1,\ldots,m$ את שפות הנקודה X, אזי יש לחפש על השפות את נקודות האקסטרמום לפי,

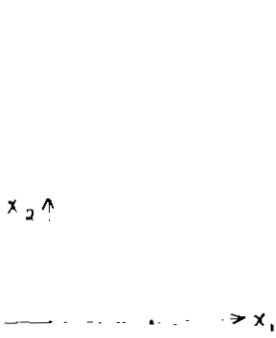
$$\left.\frac{\partial f(\underline{x})}{\partial x_j}\right|_{g_i(\underline{x}) = b_i} = 0 \qquad i = 1,\ldots,m \qquad (5\text{-}3)$$

<u>דוגמא:</u>

$$\min f = (x_1-1)^2 + (x_2+1)^2$$

כאשר

$$X = \{\underline{x} = (x_1,x_2) \mid x_1 \geq 0, \quad x_2 \geq 0\}$$



שפות הקבוצה, שהיא הרביע הראשון של מערכת הצירים, הן

$$g_1(x_1,x_2) = x_1 = 0$$

$$g_2(x_1,x_2) = x_2 = 0$$

הנקודה הסטציונרית היחידה מתקבלת כדלקמן:

$$\left.\begin{aligned} \frac{\partial f}{\partial x_1} &= 0 = (x_1 - 1) \\ \frac{\partial f}{\partial x_2} &= 0 = (x_2 + 1) \end{aligned}\right\} \Longrightarrow (x_1=1 \; ; \; x_2=-1)$$

היות והנקודה אינה נמצאת בקבוצה X (ברביע הראשון), היא אינה האקסטרמום המבוקש. נבדוק לפי (3-5) על השפות,

$$\left.\frac{\partial f}{\partial x_1}\right|_{x_2=0} = 0 = (x_1-1) \Longrightarrow (x_1=1 \; ; \; x_2=0)$$

$$\left.\frac{\partial f}{\partial x_2}\right|_{x_1=0} = 0 = (x_2+1) \Longrightarrow (x_1=0 \; ; \; x_2=-1)$$

הנקודה השניה אינה בקבוצה X, ולכן נותרת רק הראשונה. אין להסתפק בבדיקה זו, שכן על פני כל אחת משתי שפות הקבוצה האיזור מוגבל. למשל, על השפה $x_2 = 0$ התחום מוגבל ל-$x_1 \geq 0$. לכן יש לבדוק בנוסף לשתי הנקודות שנמצאו גם את שפות הקבוצות $(x_1 = 0, \; x_2 \geq 0)$ ו-$(x_1 \geq 0, \; x_2 = 0)$.

בדוגמה שלנו שתי השפות מתלכדות בנקודה $(x_1 = 0,\ x_2 = 0)$. יש אם כן, שתי נקודות בהן יש לחשב את f בדי למצוא את האקסטרמום: הראשית והנקודה $(x_1 = 1,\ x_2 = 0)$. החישוב נותן:

$$f(0,0) = (0-1)^2 + (0+1)^2 = 2$$

$$f(1,0) = (1-1)^2 + (0+1)^2 = 1$$

כלומר, המינימום הוא בנקודה (1,0) וערכו f=1.

בסיכום, כאשר האיזור ה-n מימדי מוגכל יש לחפש את הנקודות הסטציונריות שבתוכו ואת הנקודות שעל שפותיו בהן הנגזרות החלקיות בכיוון השפה מתאפסות. החלק השני מוביל לשורה של בעיות (n-1) מימדיות, אחת עבור כל שפה. היות וגם השפות מוגבלות כל בעיה (n-1) מימדית מובילה לשורה של בעיות (n-2) מימדיות, וכך הלאה - עד שמגיעים למימד אחד בו יש לתחום נקודה או נקודות קצה. מכל זה מסתבר כי החישובים במקרה של כמה משתנים בלתי תלויים וכמה אילוצים הם רבים ומורכבים למדי.

5.3 תנאי הנגזרות השניות

כאשר מפתחים את הפונקציה בטור טיילור סביב נקודת האקסטרמום המקומי x_{o_1} מקבלים עבור המקרה החד-מימדי.

$$f(x_{o_1}+\Delta x_1) = f(x_{o_1}) + \frac{df}{dx_1}\Big|_{x_{o_1}} \cdot \Delta x_1 + \frac{1}{2}\frac{d^2f}{dx^2}\Big|_{x_{o_1}} \cdot \Delta x_1^2 + \ldots \qquad (5\text{-}4)$$

היות ובנקודת האקסטרמום מתאפסת הנגזרת הראשונה קוכע סימנה של הנגזרת השניה האם האקסטרמום הוא מינימום, נקודת פיתול או מקסימום כדלקמן,

$$\left.\begin{array}{ll} \left.\frac{d^2f}{dx_1^2}\right|_{x_{o_1}} > 0 & \text{מינימום} \\ \left.\frac{d^2f}{dx_1^2}\right|_{x_{o_1}} = 0 & \text{נקודת פיתול} \\ \left.\frac{d^2f}{dx_1^2}\right|_{x_{o_1}} < 0 & \text{מקסימום} \end{array}\right\} \qquad (5\text{-}5)$$

במקרה הרב-מימרי האקסטרמום הוא בנקורה $\underline{x}_o$, ואת וקטור התוספות הרצוניות בסביבת האקסטרמום נסמן,

$$\underline{\Delta x}^T = [\Delta x_1, \ldots, x_n]$$

$\underline{\Delta x}^T$ הוא וקטור שורה, $\underline{\Delta x}$ הוא וקטור העמודה המתאים, (ראה נספח א'). פיתוח לפי טיילור בסביבת $\underline{x}_o$ נותן,

$$f(\underline{x}_o+\underline{\Delta x}) = f(\underline{x}_o) + \underline{\Delta x}^T \cdot f_x(\underline{x}_o) + \frac{1}{2!}\,\underline{\Delta x}^T \cdot \underline{\underline{H}}_f(\underline{x}_o)\underline{\Delta x} + \ldots \qquad (5\text{-}6)$$

כאשר,

$$f_x(\underline{x}_o) = \left\{\begin{array}{c} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{array}\right\}_{\underline{x}_o} \qquad (5\text{-}7)$$

הוא וקטור הנגזרות החלקיות, מחושב בנקורות האקסטרמום $\underline{x}_o$, ובן

$$\underline{\underline{H}}_f(\underline{x}_o) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{\underline{x}_o} \qquad (5\text{-}8)$$

היא מטריצת הנגזרות השניות <u>המטריצה ההסיאנית</u> (Hessian Matrix) מחושבת בנקודת האקסטרמום $\underline{x}_o$.

היות ו-$\underline{x}_o$ היא נקודה סטציונארית, קיים,

$$f_x(\underline{x}_o) = \underline{0} \qquad (5\text{-}9)$$

ולכן קובע האיבר המכיל את $\underline{\underline{H}}_f(\underline{x}_o)$ האם האקסטרמום הוא מינימום, מקסימום או נקורה סטציונרית אחרת. היות ו-$\Delta\underline{x}$ הוא וקטור של תוספות רצוניות, נותנת $\underline{\underline{H}}_f(\underline{x}_o)$ את התשובה כדלקמן:

$\underline{\underline{H}}_f(\underline{x}_o)$ מוגדרת חיובית (Positive Definite) $\Longleftarrow$ מינימום

$\underline{\underline{H}}_f(\underline{x}_o)$ מוגדרת שלילית (Negative Definite) $\Longleftarrow$ מקסימום

ראה נספח א' להבהרת תכונות אלו של המטריצות. בהרבה מקרים מעשיים קובעים איזוהי נקודת האקסטרמום המבוקשת, מבין כל הנקודות המקיימות את המשוואות (5-1), על ידי חישוב ערך הפונקציה בכל הנקודות ובחירה בהתאם לערך האקסטרמלי המבוקש. תנאי הנגזרות השניות משמש לכן בעיקר כתיאוריה של אופטימיזציה ופחות כמכשיר שימושי כפתרון בעיות.

5.4 שיטת כופלי לגרנז'

נתונה בעית האופטימיזציה

$$\text{ext}\, f = f(x_1,\ldots,x_n) \tag{5-11}$$

$$\text{s.t.} \quad g_i(x_1,\ldots,x_n) = b_i \qquad i = 1,\ldots,m \tag{5-12}$$

כאשר m=n, ויש פתרון למערכת המשוואות (5-12) הבעיה הופכת טריביאלית, שכן יש רק נקודה אפשרית אחת וערכה של f בנקודה זו הוא האופטימום המבוקש. כאשר m>n או שאין פתרון אפשרי כלל, או שיש אילוצים מיותרים, התלויים באחרים, באופן שלאחר טיפול בהם ניתן לצמצם את המספר ל-n אילוצים כלתי-תלויים. המקרה המעניין אותנו הוא כאשר m<n, כלומר, ב-(5-12) יש פחות משוואות מאשר משתנים. מספר הפתרונות של מערכת זו הוא אינסופי, ומהם רוצים לבחור את הטוב ביותר.

נסמן ב-$\underline{x}_o$ את הנקודה המבוקשת. היות ובנקודה זו יש ל-f אופטימום מקומי הדיפרנציאל השלם שלה צריך להתאפס, כלומר,

$$\sum_{j=1}^{n} \frac{\partial f}{\partial x_j} dx_j = 0 \tag{5-13}$$

אלא שהפעם וקטור התוספות $\underline{dx}$ אינו כלשהו, ועליו לקיים תנאים מסויימים, שכן הגרדיאנט צריך להיות מכוון בכיוון שלו האילוצים.

לכל אחד מן האילוצים אפשר לרשום גם כן את הדיפרנציאל השלם שלו

$$\sum_{j=1}^{n} \frac{\partial g_i}{\partial x_j} dx_j = 0 \qquad i = 1,\ldots,m \tag{5-14}$$

נכפול כל אחת מהמשוואות (5-14) במקדם כלשהו λ_i, נסכום על פני כל האילוצים, ונחבר את התוצאה ל-(5-13): מתקבל

$$\sum_{j=1}^{n} \frac{\partial f}{\partial x_j} dx_j + \sum_{i=1}^{m} \lambda_i \sum_{j=1}^{n} \frac{\partial g_i}{\partial x_j} dx_j = 0$$

לאחר החלפת סדר הסכימה וסידור מתקבל

$$\sum_{j=1}^{n} [\frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j}]dx_j = 0 \qquad (5\text{-}15)$$

להבהרת המשך הפיתות נעבור למקרה בו יש שלושה משתנים ושני אילוצים. עבור בעיה זו יש להציב ב-(5-15): (n=3) ו-(m=2).
פירוט המשוואה (5-15) הוא

$$\begin{aligned} &(\frac{\partial f}{\partial x_1} + \lambda_1 \frac{\partial g_1}{\partial x_1} + \lambda_2 \frac{\partial g_2}{\partial x_1})dx_1 \\ &+ (\frac{\partial f}{\partial x_2} + \lambda_1 \frac{\partial g_1}{\partial x_2} + \lambda_2 \frac{\partial g_2}{\partial x_2})dx_2 \\ &+ (\frac{\partial f}{\partial x_3} + \lambda_1 \frac{\partial g_1}{\partial x_3} + \lambda_2 \frac{\partial g_2}{\partial x_3})dx_3 = 0 \end{aligned} \qquad (5\text{-}16)$$

אם g_1 ו-g_2 הן פונקציות בלתי תלויות אפשר לקבוע את ערכי λ_1 ו-λ_2 כך שהביטויים בשני הסוגריים הראשונים יתאפסו. ערכי λ_1 ו-λ_2 מתקבלים פשוט מפתרון שתי משוואות ליניאריות הומוגניות. נותר כעת במשוואה (5-16) רק האיבר האחרון. היות ו-dx_3 היא תוספת רצונית [תמיד אפשר לקבוע אותה בצורה כלשהי ולהתאים את dx_1 ו-dx_2 כך שיתקיימו התנאים שהוזכרו לאחר משוואה (5-13)] הרי על מנת שהאיבר ישווה לאפס דרוש שהביטוי בסוגריים יתאפס. התנאים לאופטימום הם לכן

$$\frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \qquad j = 1,\ldots,n \qquad (5\text{-}17)$$

או בצורה אחרת

$$\frac{\partial}{\partial x_j} [f + \sum_{i=1}^{m} \lambda_i g_i] = 0 \qquad j = 1,\ldots,n \qquad (5\text{-}18)$$

כלומר, על מנת למצוא את האופטימום המאולץ של f יש לחשב את האופטימום הבלתי מאולץ של פונקציה חדשה

$$F(\underline{x},\underline{\lambda}) = f(\underline{x}) + \sum_{i=1}^{m} \lambda_i g_i(\underline{x})$$

שנהוג לרשום אותה בצורה שונה במקצת,

$$F(\underline{x},\underline{\lambda}) = f(\underline{x}) + \sum_{i=1}^{m} \lambda_i [g_i(\underline{x}) - b_i] \qquad (5\text{-}19)$$

שגם היא נותנת את הנגזרות המופיעות ב-(5-17). פונקציה זו נקראת <u>פונקציית לגרנז'</u>, ואילו המקדמים λ_i נקראים <u>כופלי לגרנז'</u> (Lagrange Multipliers).

האופטימום המבוקש הוא במרחב ה-(n+m) מימדי של $(\underline{x},\underline{\lambda})$ והוא בלתי מאולץ שכן נדרש

$$\left.\begin{aligned} \frac{\partial F}{\partial x_j} &= \frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \qquad & j = 1,\ldots,n \\ \frac{\partial F}{\partial \lambda_i} &= g_i - b_i = 0 \qquad & i = 1,\ldots,m \end{aligned}\right\} \qquad (5\text{-}20)$$

מתקבלות (n+m) משוואות במספר זהה של נעלמים. במקרה שהמשוואות אינן ליניאריות מושג הפתרון בשיטה איטרטיבית כלשהי, ואז ייתכן שתהיינה נקודות פתרון אחדות שמתוכן יש לבחור את המבוקשת לפי ערך הפונקציה f בכל אחת.

5.4.1 <u>המשמעות של כופלי לגרנז'</u>

נסמן ב-$\underline{x}^*$ את הנקודה בה יש ל-f(x) אופטימום גלובלי; $\underline{\lambda}^* = (\lambda_1^*,\ldots,\lambda_m^*)$ הוא וקטור כופלי לגרנז' המתאים ל-$\underline{x}^*$; $z^* = f(\underline{x}^*)$ הוא הערך האופטימלי. נחשב את הנגזרות החלקיות של z^* ביחס לכל אחד מה-b_i, בהנחה שהנגזרות קיימות בסביבת $\underline{x}^*$.

$$\frac{\partial z^*}{\partial b_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j^*} \frac{\partial x_j^*}{\partial b_i} \qquad i = 1,\ldots,m \qquad (5\text{-}21)$$

היות ובאופטימום מתקיימים האילוצים, הרי

$$g_k(\underline{x}^*) = b_k \qquad i = 1,\dots,m \tag{5-22}$$

גזירת משוואות אלה נותנת,

$$\frac{\partial g_k}{\partial b_i} = \delta_{ik} \begin{cases} =1 & \text{if} \quad i=k \\ =0 & \text{if} \quad i\neq k \end{cases} \tag{5-23}$$

מאידך אפשר לקבל ביטוי אחר עבור נגזרת זו

$$\frac{\partial g_k}{\partial b_i} = \sum_{j=1}^{n} \frac{\partial g_k}{\partial x_j^*} \frac{\partial x_j^*}{\partial b_i} \qquad k = 1,\dots,m \tag{5-24}$$

מהשוואת שני הביטויים עבור הנגזרות מתקבל

$$\sum_{j=1}^{n} \frac{\partial g_k}{\partial x_j^*} \frac{\partial x_j^*}{\partial b_i} - \delta_{ik} = 0 \qquad k = 1,\dots,m \tag{5-25}$$

נכפול את (5-25) ב-λ_k^* ונסכום על פני k

$$\sum_{k=1}^{m} \lambda_k^* \sum_{j=1}^{n} \frac{\partial g_k}{\partial x_j^*} \frac{\partial x_j^*}{\partial b_i} - \sum_{k=1}^{m} \lambda_k^* \delta_{ik} = 0 \tag{5-26}$$

נוסיף ביטוי זה, השווה לאפס, לצד ימין של (5-21). לאחר סידור מתקבל

$$\frac{\partial z^*}{\partial b_i} = \sum_{j=1}^{n} \left[\frac{\partial f}{\partial x_j^*} + \sum_{k=1}^{m} \lambda_k^* \frac{\partial g_k}{\partial x_j^*}\right] \frac{\partial x_j^*}{\partial b_i} - \sum_{k=1}^{m} \lambda_k^* \delta_{ik} \tag{5-27}$$

כל הסכום הראשון מתאפס, שכן כל אחד מהאיברים בסוגריים רבועים שווה לאפס מהיות $\underline{x}^*$ פתרון של משוואות (5-20). לכן נותר

$$\frac{\partial z^*}{\partial b_i} = -\sum_{k=1}^{m} \lambda_k^* \delta_{ik} = -\lambda_i^* \tag{5-28}$$

היות וכאשר יצרנו את משוואה (5-15) יכולנו לחסר את (5-14) מ-(5-13) במקום לחבר, הרי שב-(5-28) אין לסימן של λ_i^* משמעות מוחלטת. נוח לכן לכתוב

$$\frac{\partial z^*}{\partial b_i} = |\lambda_i^*| \qquad (5\text{-}29)$$

המשמעות היא ש-$|\lambda_i^*|$ נותן את השינוי בערך האופטימלי z^* עקב שינויים קטנים ב-b_i. את כיוון השינוי קל לקבוע מניסוח בעית האופטימיזציה עצמה. בבעית המקסימיזציה, למשל, צריכה הגדלה של b_i לתת שיפור, כלומר הגדלה, של הערך z^*. יש לשים לב ש-$|\lambda_i^*|$ הוא ערכה של הנגזרת החלקית של z^* ביחס ל-b_i בנקודת האופטימום. שינויים סופיים ב-b_i יתנו שינויים ב-z^* לפי שיפוע שונה במקצת בדרך כלל מהשיפוע הנתון על ידי λ_i^*.

כופלי לגרנז' נקראים גם <u>מחירי הצל</u> (Shadow Prices) של האילוצים.

אפשר להראות כי בתנאים מסויימים, המתקיימים בהרבה מקרים מעשיים, עבור הבעיה

$$\left.\begin{array}{ll} min & f(\underline{x}) \\ s.t. & g_i(\underline{x}) = b_i \end{array}\right\} \qquad (5\text{-}30)$$

מתקיים התנאי

$$F(\underline{x},\lambda^*) \geq F(\underline{x}^*,\underline{\lambda}^*) \geq F(\underline{x}^*,\underline{\lambda}) \qquad (5\text{-}31)$$

כלומר, יש ל-$F(\underline{x},\underline{\lambda})$ נקודת אוכף באופטימום $(\underline{x}^*,\underline{\lambda}^*)$ שהיא מינימום עבור $\underline{x}$ ומקסימום עבור $\underline{\lambda}$. תנאי זה קשור <u>בדואליות</u> (Duality) של בעיות אופטימיזציה, עליה נתעכב בהרחבה בפרק של התכנות הליניארי.

<u>דוגמא</u>

חשב את מידות התיבה המלבנית החסומה בתוך האליפסואיד

$$\left(\frac{x_1}{a_1}\right)^2 + \left(\frac{x_2}{a_2}\right)^2 + \left(\frac{x_3}{a_3}\right)^2 = 1$$

כך שפיאותיה מקבילות למישורי הצירים, באופן שנפחה יהיה מקסימלי.

<u>הפתרון</u>:

מידות התיבה $(2x_1;\ 2x_2;2x_3)$ ונפחה $8x_1x_2x_3$. נוותר על הקבוע ונרשום

$$max\ \ f(x_1,x_2,x_3) = x_1x_2x_3$$

$$g_1(x_1,x_2,x_3) = (\frac{x_1}{a_1})^2 + (\frac{x_2}{a_2})^2 + (\frac{x_3}{a_3})^2 = 1$$

ניצור את פונקצית לגרנז'

$$F(x_1,x_2,x_3,\lambda) = x_1x_2x_3 + \lambda[(\frac{x_1}{a_1})^2 + (\frac{x_2}{a_2})^2 + (\frac{x_3}{a_3})^2 - 1]$$

נגזור לפי כל המשתנים ונשווה לאפס

$$\frac{\partial F}{\partial x_1} = 0 = x_2x_3 + \frac{2\lambda x_1}{a_1^2}$$

$$\frac{\partial F}{\partial x_2} = 0 = x_1x_3 + \frac{2\lambda x_2}{a_2^2}$$

$$\frac{\partial F}{\partial x_3} = 0 = x_1x_2 + \frac{2\lambda x_3}{a_3^2}$$

$$\frac{\partial F}{\partial \lambda} = 0 = (\frac{x_1}{a_1})^2 + (\frac{x_2}{a_2})^2 + (\frac{x_3}{a_3})^2 - 1$$

פתרון המערכת של ארבע המשוואות נותן

$$x_1 = \frac{a_1}{\sqrt{3}};\quad x_2 = \frac{a_2}{\sqrt{3}};\quad x_3 = \frac{a_3}{\sqrt{3}};\quad \lambda = \frac{-a_1a_2a_3}{2\sqrt{3}} = -0.289a_1a_2a_3$$

ערך פונקצית המטרה הוא

$$z^* = max\ f = \frac{a_1a_2a_3}{3\sqrt{3}}$$

ונפח התיבה האופטימלית הוא $8a_1a_2a_3/3\sqrt{3}$.

זהו מקסימום ואין צורך להוכיח זאת שכן המינימום הוא אפס. נבדוק כעת את משמעות ערכו של λ. אם במקום b=1 היה רשום b = 1.01 היה מתקבל הפתרון

$$x_i = a_i\sqrt{1.01/3}$$

וערך פונקצית המטרה היה

$$a_1a_2a_3\,\frac{1.01\sqrt{1.01}}{3\sqrt{3}}$$

קירוב של הנגזרת החלקית של z* ביחס ל-b הוא

$$\frac{\Delta z^*}{\Delta b} = \frac{a_1a_2a_3(1.01\sqrt{1.01} - 1)}{0.01 \quad 3\sqrt{3}} = 0.29\ a_1a_2a_3$$

השווה כקרוב לערך המוחלט של λ. ברור לפי הבעיה שתוספת ל-b תגרום לעלית הערך של z*.

5.5 חשבון הוריאציות

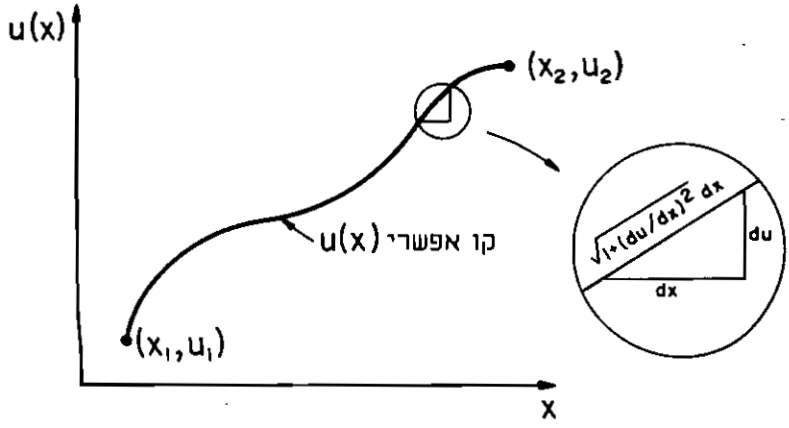
הבעיה היסודית של חשבון הוריאציות (Calculus of Variations) היא:

" מצא פונקציה, כך שאינטגרל מסויים המכיל באינטגרנד את הפונקציה ונגזרות מסויימות שלה מקבל ערך אקסטרמלי "

אנו נדון במציאת התנאים ההכרחיים שעל הפונקציה לקיים על מנת שהאינטגרל המסויים יהיה סטציונרי. תנאים מספיקים לקיום האקסטרמום המבוקש (מינימום או מקסימום) קשים בדרך כלל להוכחה מתימטית. אנו נסתפק בשיקולים פיזיקליים כדי לקבוע שאמנם קבלנו את האקסטרמום הדרוש.

דוגמא

נתונות הנקודות $(x_1;u_1)$ ו-$(x_2;u_2)$. מצא את משוואת הקו $u = u(x)$ המחבר שתי נקודות אלה, שהוא בעל האורך המינימלי (בין כל הקוים האפשריים), ראה ציור 5.2.



<u>ציור 5.2</u>: מציאת הקו הקצר כין שתי נקודות.

ברור מראש שהתשובה היא קו ישר המחבר את שתי הנקודות. נראה בהמשך איך מתקבלת תוצאה זאת בעזרת חשבון הוריאציות.

האורך הכולל של העקום יבוטא בעזרת האינטגרל

$$I = \int_{x_1}^{x_2} \sqrt{1+u_x^2}\, dx \qquad (5-32)$$

עם תנאי הגבול $u(x_1) = u_1$; $u(x_2) = u_2$.

הסימון התחתי x מבטא גזירה ביחס ל-x (כאן נגזרת שלמה, אבל כהמשך ישמש הסימון גם לנגזרות חלקיות).

אנו מחפשים פונקציה u(x) שעבורה I הוא מינימלי.

"הבעיה הפשוטה"

בחשבון הוריאציות מוגדרת "הבעיה הפשוטה" כדלקמן: מצא $u(x)$ כך שהאינטגרל

$$I = \int_{x_1}^{x_2} F(x,u,u_x)dx \qquad (5\text{-}33)$$

יהיה סטציונרי, עם תנאי הגבול $u(x_2) = u_2$; $u(x_1) = u_1$.

משפחת הקוים $u(x)$ מתוכם אנו מחפשים את הקו המבוקש צריכה לקיים תנאים אחדים, למשל, שאפשר לגזור את הפונקציה $u(x)$ (לפחות בקטעים). לדוגמא נרון בחיפוש $u(x)$ עבור I מינימלי.

פתרון

נניח כי $u(x)$ היא הפונקציה המבוקשת. נגדיר פונקציה $\eta(x)$ בעלת נגזרת רצופה כך ש-

$$\eta(x_1) = \eta(x_2) = 0$$

ונחבר אותה ל-$u(x)$, כשהיא כפולה בגודל כללי ε

$$u(x) + \varepsilon\eta(x)$$

פונקציה חדשה זו ממלאת את תנאי הגבול של הבעיה הנתונה.
האינטגרל

$$I(\varepsilon) = \int_{x_1}^{x_2} F(x\ ;\ u + \varepsilon\eta\ ;\ u_x + \varepsilon\eta_x)dx = \int_{x_1}^{x_2} F_\varepsilon dx \qquad (5\text{-}35)$$

הוא כעת פונקציה של ε כלבד, לאחר ש-$u(x)$ ו-$\eta(x)$ נקבעו. לאינטגרל זה יש מינימום כאשר $\varepsilon=0$, וזאת משום שהנחנו כי הפתרון המבוקש (זה הנותן את המינימום של I) הוא $u(x)$. כלומר,

$$\left.\frac{dI(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = 0 \qquad (5\text{-}36)$$

נגזור את F ביחס ל-ε ונקבל

$$\frac{dF_\varepsilon}{d\varepsilon} = \frac{\partial F_\varepsilon}{\partial(u+\varepsilon\eta)}\frac{\partial(u+\varepsilon\eta)}{\partial\varepsilon} + \frac{\partial F_\varepsilon}{\partial(u_x+\varepsilon\eta_x)}\frac{\partial(u_x+\varepsilon\eta_x)}{\partial\varepsilon} =$$

$$= \frac{\partial F_\varepsilon}{\partial u}\cdot\eta + \frac{\partial F_\varepsilon}{\partial u_x}\cdot\eta_x =$$

$$= \frac{\partial F}{\partial u}\cdot\eta + \frac{\partial F}{\partial u_x}\cdot\eta_x$$

נציב ביטוי זה בנגזרת של $I(\varepsilon)$ ונקבל

$$\left.\frac{dI(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = \int_{x_1}^{x_2}\left(\frac{\partial F}{\partial u}\eta + \frac{\partial F}{\partial u_x}\eta_x\right)dx = 0 \qquad (5\text{-}38)$$

את החלק השני של האינטגרציה אפשר לבצע בחלקים

$$\int_{x_1}^{x_2}\frac{\partial F}{\partial u_x}\frac{\partial\eta}{\partial x}dx = \left.\frac{\partial F}{\partial u_x}\eta\right|_{x_1}^{x_2} - \int_{x_1}^{x_2}\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right)\cdot\eta dx \qquad (5\text{-}39)$$

השתמשנו כסימני נגזרת חלקית גם במקומות בהם במקרה הנירון הנגזרת היא שלימה (משום ש-x הוא המשתנה הכלתי תלוי היחיד), וזאת כדי לרמוז על השיטה הכללית, שכדאה לפח למקרים של יותר משתנים בלתי תלויים.

בהתאם לתנאי הגבול של $\eta(x)$ מתאפס החלק הראשון מימין במשוואה האחרונה. אזי מקבלים

$$\left.\frac{dI(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = \int_{x_1}^{x_2}\left[\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right)\right]\eta dx = 0 \qquad (5\text{-}40)$$

היות ו-$\eta(x)$ היתה פונקציה רצונית, כלומר, בעלת צורה כלשהי, דרוש שהביטוי בסוגריים הרבועים יהיה שווה באופן זהותי לאפס על פני כל התחום x_1 עד x_2, וזאת על מנת שהאינטגרל יתאפס.

המשוואה הדיפרנציאלית שעל $u(x)$ לקיים, אם כן, היא

$$\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) = 0 \qquad (5\text{-}41)$$

<u>הנקראת משוואת אוילר.</u>

נשים לב כי במשוואה זו מופיעות נגזרות חלקיות של F ביחס ל-u וביחס ל-u_x. בגזירה זו מתיחסים אל u ואל u_x <u>כאילו היו משתנים בלתי-תלויים</u> בהם תלוי הביטוי הנתון F. הרבר יורגש שוב ברוגמאות שיובאו להלן. לאחר שגוזרים את F לפי u_x הגזירה הנוספת לפי x היא גזירה רגילה של הביטוי המתקבל.

את משוואת אוילר אפשר לפתח כרלקמן:

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right)\frac{dx}{dx} + \frac{\partial}{\partial u}\left(\frac{\partial F}{\partial u_x}\right)\frac{du}{dx}$$

$$+ \frac{\partial}{\partial u_x}\left(\frac{\partial F}{\partial u_x}\right)\frac{du_x}{dx} =$$

$$= \frac{\partial^2 F}{\partial x \partial u_x} + \frac{\partial^2 F}{\partial u \partial u_x}\frac{du}{dx} + \frac{\partial^2 F}{\partial u_x^2}\frac{d^2u}{dx^2} \qquad (5\text{-}42)$$

נצרף לזה את האבר האחרון במשוואת אוילר ונרשום

$$\frac{\partial^2 F}{\partial x \partial u_x} + \frac{\partial^2 F}{\partial u \partial u_x}\frac{du}{dx} + \frac{\partial^2 F}{\partial u_x^2}\frac{d^2u}{dx^2} - \frac{\partial F}{\partial u} = 0 \qquad (5\text{-}43)$$

או בצורה אחרת

$$\frac{\partial^2 F}{\partial u_x^2}\frac{d^2u}{dx^2} + \frac{\partial^2 F}{\partial u \partial u_x}\frac{du}{dx} + \left(\frac{\partial^2 F}{\partial x \partial u_x} - \frac{\partial F}{\partial u}\right) = 0 \qquad (5\text{-}44)$$

שהיא משוואה דיפרנציאלית מסדר שני (כאשר $\partial^2 F/\partial u_x^2 \neq 0$) ולכן אפשר לקיים את שני תנאי הגבול הנתונים. אפשר גם לכתוב את משוואת אוילר בצורה הבאה

$$\frac{1}{u_x}\left[\frac{\partial}{\partial x}\left(F - \frac{\partial F}{\partial u_x}\frac{du}{dx}\right) - \frac{\partial F}{\partial x}\right] = 0 \qquad (5\text{-}45)$$

ומכאן כמה מקרים מיוחדים:

1. כאשר $F = F(u,u_x)$, כלומר, x אינו מופיע בצורה מפורשת בביטוי של F אינטגרציה אחת נותנת

$$F - \frac{\partial F}{\partial u_x}\frac{du}{dx} = c \qquad (5\text{-}46)$$

2. כאשר $F = F(x,u_x)$, כלומר, u אינו מופיע בצורה מפורשת בביטוי של F, מקבלים מהמשוואה בצורתה המקורית אחרי אינטגרציה אחת

$$\frac{\partial F}{\partial u_x} = C \qquad (5\text{-}47)$$

<u>דוגמא:</u>

נחזור לבעיה של מציאת קו בעל אורך מינימלי המחבר את (x_1,u_1) עם (x_2,u_2)

$$min\ I = \int_{x_1}^{x_2} (1 + u_x)^{\frac{1}{2}}\, dx$$

$$\text{subject to } u(x_1) = u_1 \ ; \quad u(x_2) = u_2$$

היות ו-$F = F(u_x)$ אפשר להשתמש בצורה המיוחדת של משוואת אוילר

$$\frac{\partial F}{\partial u_x} = \frac{u_x}{(1+u_x^2)^{\frac{1}{2}}} = c_1$$

$$u_x^2 = c_1^2(1+u_x^2)$$

$$u_x = \frac{c_1}{\sqrt{1-c_1^2}} = c_2$$

אינטגרציה נותנת $u = c_3x + c_4$

כלומר, משוואת קו ישר. הקבועים נקבעים בעזרת תנאי הגבול, והתוצאה הסופית היא

$$u = \left(\frac{u_1-u_2}{x_1-x_2}\right)x + \left(u_1 - \frac{u_1-u_2}{x_1-x_2}x_1\right)$$

<u>תנאי גבול "טבעיים"</u>

נזכור כי בחרנו $\eta(x) = 0$ על הגבולות משום שהיה נתון $u(x_1) = u_1$ ו-$u(x_2) = u_2$. אם תנאים אלה אינם נתונים נוכל להציב דרישה אחרת ביחס לפונקציה F, כך שמשוואת אוילר תמשיך להיות התנאי ההכרחי לסטציונריות של I.

נשים לב כי אם $\partial F/\partial u_x = 0$ בגבולות $x = x_1$ ו-$x = x_2$, אזי הביטוי שקודם התאפס בגלל הבחירה המיוחדת של η ($\eta=0$ בגבולות) יתאפס גם עכשיו. התנאי $\partial F/\partial u_x = 0$ על הגבולות נקרא "תנאי הגבול הטבעי"(natural boundary condition)

<u>דוגמא:</u>

מצא את משוואת הקו בעל האורך הקצר ביותר המחבר את הנקודה $(x_1;u_1)$ עם הקו $x = x_2$. כמו מקודם נמצא מפתרון משוואת אוילר

$$u = c_1x + c_2$$

אלא שהפעם נדרוש בגבול $x = x_2$ שתנאי הגבול הטבעי יתקיים

$$\frac{\partial F}{\partial u_x} = 0 \Longrightarrow \frac{\partial}{\partial u_x}\left(\sqrt{1+u_x^2}\right) = \frac{u_x}{\sqrt{1+u_x^2}} = 0 \Longrightarrow u_x = 0$$

כלומר, בגבול $x=x_2$ מקבלים

$$u_x = c_1 = 0$$

ואז משוואת הקו בולו היא $u(x) = c_2 = u_1$.

<u>בעיות רב-מימדיות</u>

נדון לדוגמא במקדה בו יש שני משתנים בלתי-תלויים, x ו-y ושני משתנים התלויים בהם, $u(x,y)$ ו-$v(x,y)$. אנו רוצים למצוא $u(x,y)$ ו-$v(x,y)$ כך שהאינטגרל

$$I = \int\int_{(R)} F[x,y,u,v,u_x,v_x,u_y,v_y]dxdy \qquad (5\text{-}48)$$

יהיה סטציונרי, עם תנאי גבול מתאימים. R הוא תחום נתון במישור $(x;y)$, F היא פונקציה נתונה והסימונים התחתיים מסמנים נגזרות תלקיות ביחס למשתנים הבלתי-תלויים.

אפשר להראות כי $u(x,y)$ ו-$v(x,y)$ המקיימים תנאי זה הם הפתרונות של המערכת של שתי משוואות הדיפרנציאליות הבאות, שהן <u>משוואות אוילר</u>:

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right) - \frac{\partial F}{\partial u} = 0$$

$$(5\text{-}49)$$

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial v_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial v_y}\right) - \frac{\partial F}{\partial v} = 0$$

המקיימים את תנאי הגבול הנתונים. כמו קודם, גם כאן הגזירה החלקית ביחס ל- u_x וכו' נעשית כאילו היו אלה משתנים בלתי תלויים המופיעים בביטוי של F.

כפי שכבר נאמר לעיל, קיום משוואות אוילר הוא תנאי הכרחי לכך שהפונקציות המתקבלות גורמות לאינטגרל להיות אקסטרמלי. תנאים מספיקים קשים או בלתי ניתנים להוכחה בסקרים רבים. בבעיות שמקורן בעולם ההנדסי, קל במקרים רבים להיווכח אם התוצאה שנתקבלה היא אמנם המבוקשת.

אילוצים וכופלי לגרנז'

כדי למצוא את הפונקציה (או פונקציות) הגורמות לאינטגרל

$$\int_{x_1}^{x_2} F \cdot dx \qquad (5\text{-}50)$$

שיהיה סטציונרי, והנשמעות לאילוץ בעל הצורה

$$\int_{x_1}^{x_2} G \cdot dx = K \qquad (5\text{-}51)$$

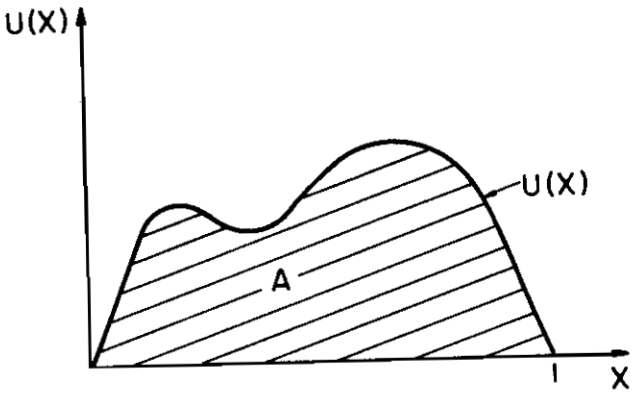
כאשר G היא פונקציה של u,x ו-u_x ברומה ל-F, ו-K קבוע נתון, יש לפתור את הבעיה הואריאציונית

$$I = \int_{x_1}^{x_2} [F + \lambda(G-K)]dx \qquad (5\text{-}52)$$

שבה λ הוא מקרם בללי (כופל לגרנז'), שאת ערכו יש לקבוע בנוסף לחישוב קבועי האינטגרציה.

דוגמא:

חשב את משוואתו של קו u(x) בעל האורך הקצר ביותר, העובר דרך הנקודות (0:0) ו-(1:0) והכולא שטח נתון A עם ציר ה-x כמתואר בשרטוט.



הגיסנח המתימטי:

$$min \int_0^1 \sqrt{1 + u_x^2}\, dx$$

$$u(0) = u(1) = 0$$

$$\text{subject to} \int_0^1 u dx = A$$

פונקצית לגרנז', שהיא האינטגרנד החדש:

$$F + (G-K) = \sqrt{1 + u_x^2} + \lambda(u-A)$$

משוואת אוילר המתקבלת היא

$$\frac{\partial}{\partial x}\left[\frac{\partial[F+\lambda(G-K)]}{\partial u_x}\right] - \frac{\partial[F+\lambda(G-K)]}{\partial u} = 0$$

$$\frac{\partial}{\partial x}\left[\frac{u_x}{\sqrt{1 + u_x^2}}\right] - \lambda = 0$$

$$u_x = \frac{\lambda x + c_1}{\sqrt{1-(\lambda x+c_1)^2}}$$

אינטגרציה נוספת (בלתי מסויימת) נותנת

$$u = \int \frac{\lambda x + c_1}{\sqrt{1-(\lambda x+c_1)^2}}\,dx = \frac{-\sqrt{1-(\lambda x+c_1)^2}}{\lambda} + c_2$$

או

$$(\lambda u - c_2)^2 + (\lambda x + c_1^2) = 1$$

שהיא משוואת מעגל. את שלושת הקבועים c_1, c_2 ו-λ מחשבים בעזרת התנאים

$$u(0) = u(1) = 1$$

וכן התנאי

$$\int_0^1 u\,dx = A$$

5.6 סיכום

לכל שיטות האופטימיזציה הקלאסית שהוזכרו יש בעיקר חשיבות עיונית, שכן הן מבהירות את התנאים השוררים בנקודות האופטימום המאולץ והבלתי מאולץ. עם זאת, אין לשכוח כי יש מקרים בהם השיטות הקלאסיות הן בעלות חשיבות מעשית. בבעיות לא גרולות מרי קורה לעיתים שאת מהשיטות שתוצגו היא היעילה ביותר לפתרון. כמו כן קורה לעיתים שבמסגרת בעית אופטימיזציה גדולה יש אפשרות לנסח בעית אופטימי-זציה בשיטה קלאסית עבור חלק ממשתני ההחלטה ולקבל קשר פשוט בין משתנים אלה החייב להתקיים על מנת שפתרון הבעיה הגדולה יהיה אופטימלי. פתרונות חלקיים כאלה מאפשרים לצמצם את מספר המשתנים בבעיה הגדולה ובכך מקלים על פתרונה. מומלץ לכן לא לשכוח את קיומן של השיטות הקלאסיות ולבדוק בכל מקרה את האפשרות להשתמש בהן. כאשר הרבר אפשרי, הן עשויות לחסוך עבודה רבה. יתר על כן, התוצאות מתקבלות בצורת קשרים פונקציונליים אנליטיים בין משתנים ולכן אפשר להסיק מהן מסקנות כלליות בקלות רבה יותר מאשר מתוצאות מספריות גרידא, כפי שמתקבל משיטות אופטימיזציה אחרות, אפילו הן חדישות ביותר.

מקורות

Avriel, M., "Nonlinear Programming", Prentice-Hall, 1976.

Hadley, G., "Nonlinear and Dynamic Programming", Addison-Wesley, 1964, (Chapter 3).

Hildebrand, F.B., "Advanced Calculus for Applications", Prentice-Hall, 1964, (Chapter 7).

Hildebrand, F.B., "Methods of Applied Mathematics", Prentice-Hall, 1952, (Chapter 2).

פרק ששי

תכנות ליניארי

6.1 הקדמה

תכנות ליניארי (Linear Programming) הוא מקרה פרטי של הבעיה הכללית של התכנות המתימטי, כאשר

$$g_i(x_1,\ldots,x_r) = \sum_{j=1}^{r} a_{ij}x_j\{\leq, =, \geq\}b_i \qquad i = 1,\ldots,m \qquad (6\text{-}1)$$

$$f(x_1,\ldots,x_r) = \sum_{j=1}^{r} c_j x_j \ \mathit{max}, \text{ or } \mathit{min} \qquad (6\text{-}2)$$

וכן

$$x_j \geq 0 \qquad j = 1,\ldots,r \qquad (6\text{-}3)$$

הדרישה לאי-שליליות משתני ההחלטה היא חלק אינטגרלי מניסוח בעית התכנות הלינארי, והיא תשמש כאחד מן היסודות של שיטת הפתרון. כאשר יש משתנה שאינו מוגבל בסימנו, כלומר הוא יכול להיות חיובי או שלילי, ניתן לרשום אותו כהפרש שבין שני משתנים שכל אחד מהם נדרש להיות חיובי: $x_j'\geq 0$ ו- $x_j''\geq 0$, $x_j = x_j'-x_j''$ ובדרך זו ניתן תמיד לחזור לניסוח שלעיל.

6.2 דוגמאות ופתרון גרפי

על מנת לראות כמה מתכונות הפתרון של בעית התכנות הליניארי נשתמש בדוגמה של שני משתני החלטה עם וריאציות שונות.

6.2.1 הבעיה המקורית

נתונה על ידי

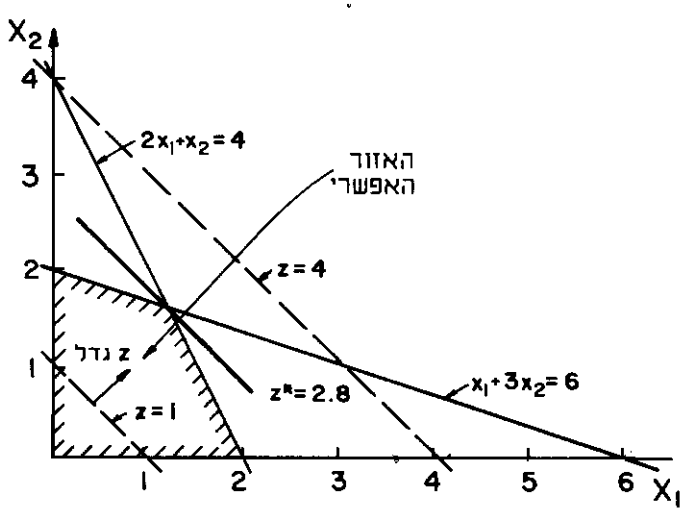
$$max \quad z = x_1 + x_2$$

$$\text{subject to} \quad 2x_1 + x_2 \le 4$$

$$x_1 + 3x_2 \le 6$$

$$x_1 , x_2 \ge 0$$

נשתמש בתיאור הגרפי הבא:



היות ועל שני משתני ההחלטה להיות לא-שליליים, יש לטפל רק ברביע הראשון.
נשרטט את הקו $2x_1+x_2= 4$. רק החלק שמשמאל לקו מקיים את אי-השוויון הראשון.
לגבי האילוץ השני, מקיים אותו רק החלק שמתחת לקו $x_1+3x_2 = 6$. האיזור האפשרי הוא החלק המקווקוו.
קוים שווי-פונקצית - המטרה הם קוים עבורם

$$z = x_1 + x_2 = constant$$

כלומר, קוים בשיפוע $-45°$. הקו $z=1$ נראה בציור לעיל. לקו זה יש נקודות בתוך האיזור האפשרי ועל גבולותיו, למשל, הנקודות $(0;1),(1;0)$ וכן $(0.5;0.5)$. הקו $z=4$ גם הוא מופיע בציור. לכל הנקורות שעל קו זה יש ערך של פונקצית המטרה, z, הגדול מערכה על הקו הקודם. היות ואנו רוצים מקסימיזציה, הנקודות על קו זה טובות יותר מאשר על הקו הראשון. אלא שאין אפילו נקורה אחת על הקו החדש הנמצאת בתוך או על גבולות האיזור האפשרי, לכן $z=4$ אינו אפשרי. ברור עתה כעת מה עלינו לעשות: להזיז את הקו $z=const.$ ימינה ככל האפשר תוך שמירה על כך שתהיה עליו לפחות נקודה אחת בתוך או על גבולות האיזור האפשרי. כציור ברור שהקו המבוקש הוא זה העובר דרך נקודת החיתוך של שני הקוים: הנקודה $(x_1^* = 1.2,\ x_2^* = 1.6)$. הערך של פונקצית המטרה על קו זה, שהוא הערך האופטימלי, הוא $z^* = 2.8$. תמיד נסמן בכוכב את הערך האופטימלי ואת ערך משתני ההחלטה בפתרון.

כפתרון האופטימלי שקיבלנו שני האילוצים <u>כובלים</u> (Binding), כלומר מתקיימים כשוויון. העובדה שבמקרה זה היתה נקודת האופטימום בחיתוך שני הקוים, כלומר, אפשר היה לפתור שתי משוואות בשני נעלמים, לקבל את נקורת החיתוך, להכניס אותה לפונקצית המטרה ולקבל את האופטימום - אינה צריכה להטעות. התוצאה אינה כללית מבחינה זו. במקרה הכללי מספר המשוואות אינו שווה למספר המשתנים, לכן פתרון מערכת המשוואות המתקבלת מהפיכת או-השוויונות לשוויונות אינו חד-משמעי, ולכן בודאי שאינו יכול לתת את הפתרון הנרדש. נשים רק לב כי הפתרון האופטימלי לא התקבל בתוך האיזור האפשרי אלא על גבולו, ודוקא בנקודת פינה של הגבול. זו כן תוצאה כללית, כפי שנראה גם להלן.

הקוים של $z=const$ שהופיעו בציור הם הטלים של "קוי-גובה" של מישור פונקצית המטרה הנמצא במימד השלישי הניצב למישור $(x_1;x_2)$ של הציור.

6.2.2 <u>מקרה שני</u>:

$$\max \quad z = 5x_1 + x_2$$

$$\text{subject to} \quad 2x_1 + x_2 \le 4$$

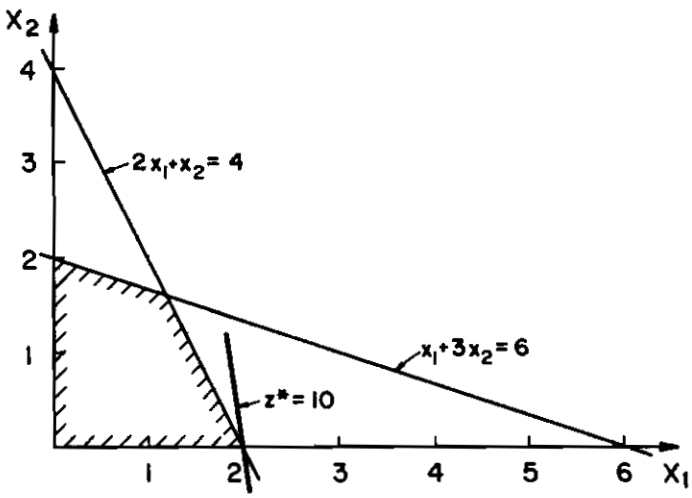
$$x_1 + 3x_2 \le 6$$

$$x_1 ,\ x_2 \ge 0$$

אגתם אילוצים, פונקצית המטרה שונה. הפתרון נראה בציור. שוב, הפתרון האופטימלי הוא בנקודת פינה של האיזור האפשרי, אם כי הפעם לא בחיתוך הקוים.

הפתרון הפעם הוא:

$$z^* = 10, \quad x_2^* = 0, \quad x_1^* = 2$$



6.2.3 <u>מקרה שלישי:</u>

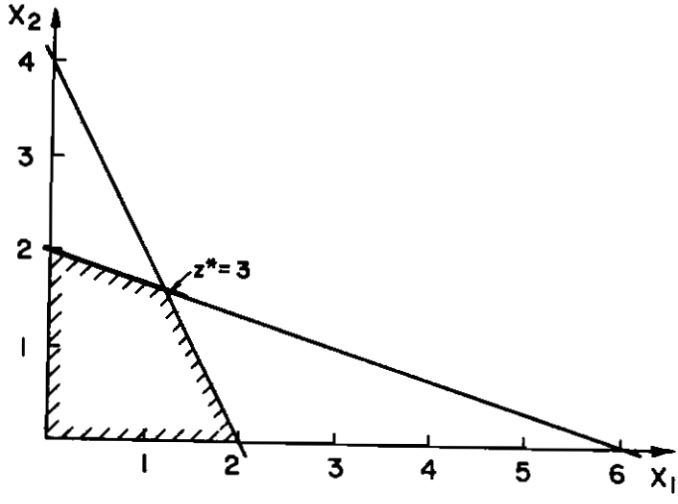
$$\max \quad z = 0.5\,x_1 + 1.5\,x_2$$

$$\text{subject to} \quad 2x_1 + x_2 \le 4$$

$$x_1 + 3x_2 \le 6$$

$$x_1\,,\ x_2 \ge 0$$

הפעם הפתרון הוא כל נקודה על הגבול העליון של האיזור האפשרי, כנראה בציור בעמוד הבא:



$z^* = 3$, והנקודה האופטימלית יכולה להיות כל נקודה על הקטע בין (0;2) לבין (1.2 ; 1.6), כולל את קצות הקטע. גם כאן יש פינה של האיזור האפשרי (במקרה זה שתי פינות) שהיא הפתרון האופטימלי.

6.2.4 <u>מקרה רביעי</u>:

$$\min \quad z = x_1 + x_2$$

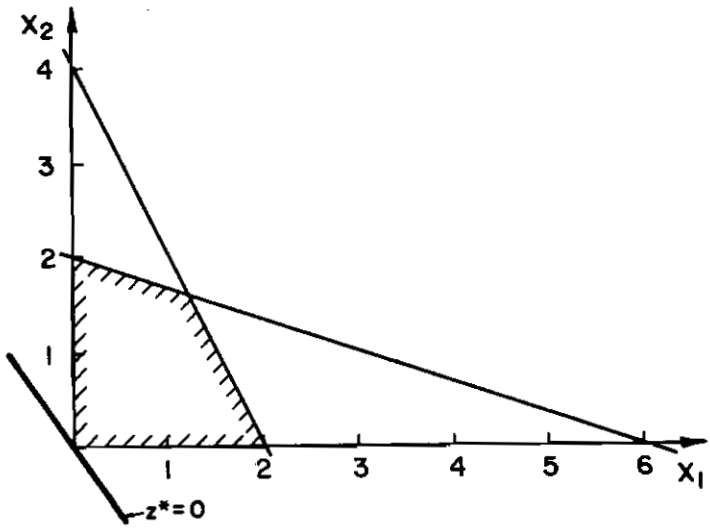
$$\text{subject to} \quad 2x_1 + x_2 \le 4$$

$$x_1 + 3x_2 \le 6$$

$$x_1 , x_2 \ge 0$$

הפתרון האופטימלי הוא הפעם:

$x_1^* = 0 \ , \ x_2^* = 0 \ , \ z^* = 0$



הבעיה רומה למקורית, אלא שהפעם מבצעים מינימיזציה. לכן משתדלים להזיז את הקו $z = constant$ שמאלה ולמטה ככל האפשר, תוך שמירה על התנאי שתהיה עליו לפחות נקודה אחת שהיא כתוך או על גבולות האיזור האפשרי. שוב, הפתרון האופטימלי מתקבל בנקורת פינה של האיזור האפשרי.

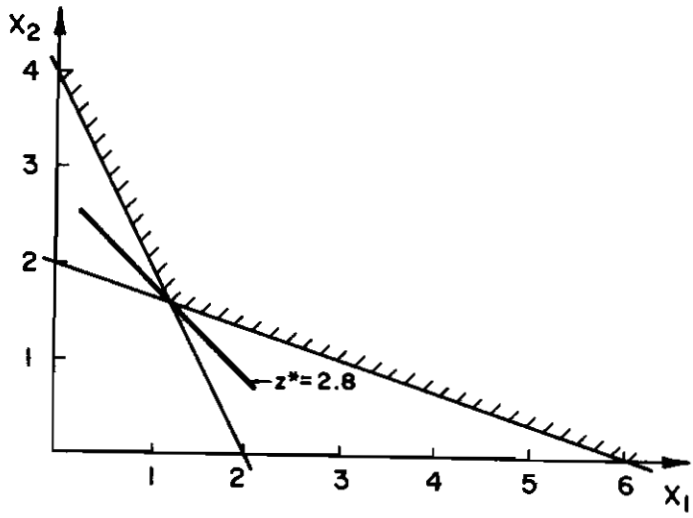
6.2.5 <u>מקרה חמישי</u>:

$$\min \quad z = x_1 + x_2$$

$$\text{subject to} \quad 2x_1 + x_2 \geq 4$$

$$x_1 + 3x_2 \geq 6$$

$$x_1 \ , \ x_2 \geq 0$$



הפעם האיזור האפשרי הוא אוסף חלק של הרביע הראשון שנמצא מעל שני הקוים, כלומר, הוא משתרע עד אינסוף בשני כווני הצירים. המינימום מוגדר וסופי, ושווה במקרה זה לפתרון האופטימלי של הבעיה המקורית, כלומר,

$$x_1^* = 1.2 \ , \quad x_2^* = 1.6 \ , \quad z^* = 2.8$$

6.2.6 <u>מקרה שישי:</u>

כמו מקרה חמישי, אלא $max\ z$. הפתרון במקרה זה לא מוגבל, כלומר, $z^* \to \infty$ כאשר ערכי המשתנים אינם מוגדרים.

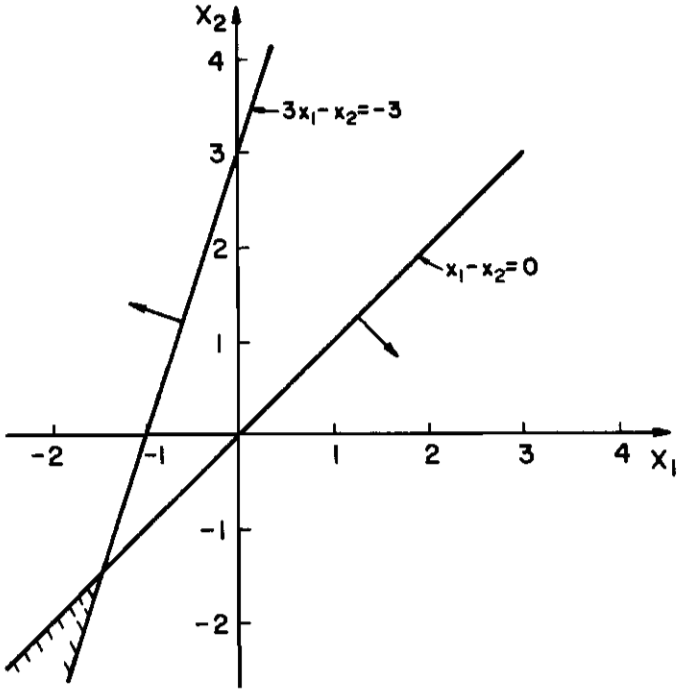
6.2.7 מקרה שביעי:

$$\max \quad z = x_1 + x_2$$

$$\text{subject to} \quad x_1 - x_2 \geq 0$$

$$3x_1 - x_2 \leq -3$$

$$x_1 , x_2 \geq 0$$



מן הציור נראה כי כדי לקיים את האילוץ הראשון יש להמצא מתחת לקו $x_1-x_2 = 0$, ואילו בדי לקיים את האילוץ השני יש להמצא מעל לקו $3x_1-x_2 = -3$. אין אף נקודה ברביע הראשון המקיימת תנאי זה. יש נקודות, באיזור המקווקו משמאל למטה, המקיימות את שני האילוצים, אבל לא את תנאי אי-השליליות של משתני ההחלטה, ולכן אינם פתרונות אפשריים. התשובה היא לכן שאין פתרון אפשרי לבעיה שהוצגה.

6.2.8 סיכום

א. כאשר היה איזור אפשרי, הוא היה קמור, בעל מקצועות ופינות. איזור הוא קמור כאשר הישר המחבר שתי נקורות בלשהן הנמצאות בתוך האזור נמצא כולו בתוך האזור. הדבר נכון גם במקרה ה- n-מימדי, אלא שאז האיזור תחום בפאות במקום במקצועות, כאשר הפאות הן מישורים n-מימדיים, וגם אז יש פינות.

ב. כאשר הערך האופטימלי, z^*, היה סופי, לפחות פינה אחת של האיזור האפשרי היתה הפתרון האופטימלי.

ג. פתרון בלתי מוגבל לא ייקרא פתרון אופטימלי.

6.3 שיקולים מתימטיים ראשונים

נחזור לניסוח בעית התכנות הליניארי כפי שניתנה בסעיף 6.1. כל וקטור המקיים את כל האילוצים (1-6) נקרא פתרון (Solution). אם הוא גם מקיים את תנאי אי-השליליות (3-6) הוא נקרא פתרון אפשרי (Feasible Solution). הפתרון האפשרי הגורם לפונקצית המטרה (2-6) לקבל ערך אקסטרמלי נקרא הפתרון האפשרי האופטימלי (Optimal Feasible Solution).

כדי לפשט את ההסברים בהמשך נתייחס לעת עתה, רק למקרה בו נדרשת מינימיזציה של פונקצית המטרה (2-6). ברור כי כאשר יש לבצע מקסימיזציה של (f) אפשר לקבל את הפתרון בעזרת מינימיזציה של (-f). מאוחר יותר נפתור ישירות בעיות מקסימיזציה, בלי להפכן קודם לבעית מינימיזציה. כמו כן נדרוש $b_i \geq 0$ $(i=1,\ldots,m)$. כאשר $b_i < 0$ יש לכפול את האילוץ כולו ב-(-1) כדי לקבל $b_i > 0$. נרשום כעת את בעית התכנות הליניארי בצורה המפורשת הבאה:

$$\left.\begin{array}{l} a_{11}x_1 + \dots \dots + a_{1r}x_r \le b \\ \vdots \qquad\qquad\qquad \vdots \\ a_{k1}x_1 + \dots \dots + a_{kr}x_r \ge b_k \\ \vdots \qquad\qquad\qquad \vdots \\ a_{m1}x_1 + \dots \dots + a_{mr}x_r = b_m \\ x_1 \qquad\qquad\qquad\qquad \ge 0 \\ \qquad\qquad\qquad\qquad x_r \ge 0 \\ \hline \min f = c_1x_1 + \dots \dots + c_r x_r \end{array}\right\} \qquad (6\text{-}4)$$

כאשר האילוץ הוא בעל הצורה $\sum_{j=1}^{r} a_{hj}x_j \le b_h$ נוסיף משמאל משתנה חדש, $x_{r+h} \ge 0$, כך שמתקיים

$$\sum_{j=1}^{r} a_{hj}x_j + x_{r+h} = b_h \qquad (6\text{-}5)$$

משתנה זה נקרא <u>משתנה חסר</u> (Slack Variable). כאשר האילוץ הוא בעל הצורה $\sum_{j=1}^{r} a_{kj}x_j \ge b_k$ נחסר משמאל משתנה חדש, $x_{r+k} \ge 0$ כך שמתקיים

$$\sum_{j=1}^{r} a_{kj}x_j - x_{r+k} = b_k \qquad (6\text{-}6)$$

משתנה זה נקרא <u>משתנה עודף</u> (Surplus Variable).

נתקבלו כעת m משוואות בצורה

$$\left.\begin{array}{lll} \sum_{j=1}^{r} a_{hj}x_j + x_{r+h} = b_h & & h = 1,\ldots,u \\ \sum_{j=1}^{r} a_{kj}x_j - x_{r+k} = b_k & & k = (u+1),\ldots,v \\ \sum_{j=1}^{r} a_{pj}x_j = b_p & & p = (v+1),\ldots,m \end{array}\right\} \qquad (6\text{-}7)$$

במשוואות אלה מופיעים n משתנים, r מקוריים ו-(n-r) חדשים כאשר $(r \le n \le r+m)$.
נרשום כעת את הבעיה בצורתה החדשה:

$$\left.\begin{array}{lll} min\ f = \sum_{j=1}^{n} c_j x_j & & \\ subject\ to\ \ g_i = \sum_{j=1}^{n} a_{ij}x_j = b_i & & i = 1,\ldots,m \\ x_j \ge 0 & & j = 1,\ldots,n \end{array}\right\} \qquad (6\text{-}8)$$

נשים לב כי:

א. <u>כל</u> המשתנים נדרשים להיות לא-שליליים.

ב. תרומת משתני החסר והעודף לפונקצית המטרה היא אפס, משום שמקדמיהם ב-f כולם אפס.

מערכת האילוצים היא בעלת הצורה $\underline{A}\underline{x} = \underline{b}$, כאשר

$$\underline{\underline{A}} = \begin{bmatrix} a_{11} & \dots & a_{1r} & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ a_{21} & \dots & a_{2r} & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{u1} & \dots & a_{ur} & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ a_{u+1,1} & \dots & a_{u+1,r} & 0 & 0 & \dots & 0 & -1 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{v,1} & \dots & a_{v,r} & 0 & 0 & \dots & 0 & 0 & \dots & -1 \\ a_{v+1,1} & \dots & a_{v+1,r} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,r} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \qquad (6-9)$$

$\underline{\underline{A}}$ היא $(m \times n)$ והפתרון כרגע כלל אינו חד-משמעי. מכיון כל הפתרונות אנו רוצים למצוא את זה הנותן אופטימום של f. נסמן ב-$(\underline{\underline{A}},\underline{b})$ את המטריצה המורחבת, המתקבלת מהוספת העמודה $\underline{b}$ כעמודה נוספת למטריצה $\underline{\underline{A}}$. נניח כעת שמתקיים $r(\underline{\underline{A}}) = r(\underline{\underline{A}},\underline{b})$, כלומר, שדרגת המטריצה $\underline{\underline{A}}$ שווה לדרגת המטריצה המורחבת (ראה נספח א'). אם תנאי זה אינו מתקיים המשוואות סותרות, ואין אפילו פתרון אחד. להלן נראה איך מטפלים במקרה זה. כמו כן נניח שמתקיים $r(\underline{\underline{A}}) = m$ פירוש הדבר שמספר העמודות, ולכן גם מספר המשתנים בבעיה, לפחות שווה למספר המשוואות. כמו כן, פירוש הדבר ש-$\underline{\underline{A}}$ אינה סינגולרית.

א. אם בהצבת הבעיה המקורית, לפני הוספת משתני חסר ועודף, נתקבלו m משוואות (לא אי-שוויונות) ב-n נעלמים, כך ש-$(m > n)$, הרי או שאין ככלל פתרון או שחלק מהמשוואות מיותרות, והן יכולות להיות מבוטאות כצירוף ליניארי של משוואות אחרות.

ב. אם בבעיה המקורית $(m=n)$ יש פתרון חד-משמעי, שהרי הנחנו $r(\underline{\underline{A}}) = m$. אם פתרון זה גם אפשרי, כלומר כל המשתנים אינם שליליים בפתרון, הרי שהפתרון הוא האופטי

ג. אם $(m < n)$ ו-$r(\underline{\underline{A}}) = m$ אזי אף משוואה אינה מיותרת. מקרה זה הוא המעניין אותנו, והוא גם הרגיל, שכן בו המשוואות אינן סותרות זו את זו וכל אחת מהן מוסיפה אינפורמציה חדשה, כלומר, הן בלתי תלויות זו בזו.

כאשר יש יותר משנים שלושה משתנים ו/או אילוצים, וזה המצב ברוב המקרים המעשיים, בלתי אפשרי להבטיח כי בעת כתיבת הבעיה לא תהיינה משוואות סותרות או תלויות. אנו מעונינים לכן באלגוריתם הלוקח בחשבון את האפשרות שאין כלל פתרון אפשרי (במלים אחרות - שיש אילוצים סותרים), ואת האפשרות שיש אילוצים מיותרים.

לשם כך מתחילים את תהליך הפתרון עם מטריצה $\underline{\underline{A}}$, שאמנם איננה מטריצת האילוצים המקוריים, המקיימת את תנאי $r(A) = m$. הדבר נעשה על ידי הוספת משתנים. מפעילים את האלגוריתם על המערכת החדשה ובתוך כך קובעים האם למערכת המקורית יש פתרון. אם כן - ממשיך האלגוריתם ומוביל לפתרון האופטימלי.
נשים לב לנקודות הבאות:

א. היות ונדרש כי משתני החסר והעודף יהיו לא-שליליים, ומקדמיהם בפונקציית המטרה הוצבו כאפסים הרי שיש התאמה חד-ערכית בין הפתרון של המערכת המורחבת לפתרון בעית האופטימיזציה המקורית.

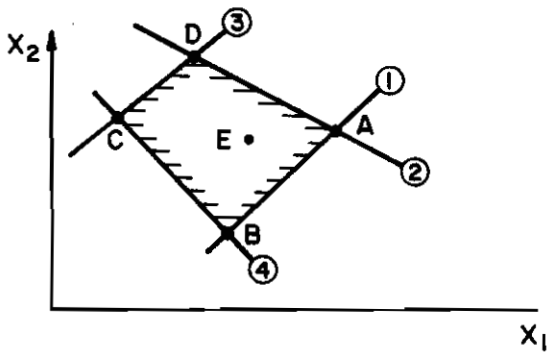
ב. הדרישה $r(\underline{\underline{A}}) = m$ נעשתה לצורך הדיון בלבד. מתחילים ממטריצה מורחבת המקיימת תנאי זה. אם בתוך תהליך הפתרון מסתבר שישנם אילוצים מיותרים במערכת המקורית, הדבר מתגלה בפתרון הסופי.

$\underline{\underline{A}}$ היא מטריצה $(m \times n)$. נבחר מתוכה m עמודות וניצור מהן מטריצה ריבועית $(m \times m)$. אם מטריצה זו אינה סינגולרית, כלומר הדטרמיננט שלה שונה מאפס, אזי המטריצה תיקרא <u>מטריצת בסיס</u>, ותסומן $\underline{\underline{B}}$ (ראה סעיף 11 בנספח א'). הוקטורים של מטריצת הבסיס מהווים <u>בסיס</u> (Basis). היות ו-$\underline{\underline{B}}$ אינה סינגולרית ניתן לבסא כל וקטור במרחב ה-m מימדי כצירוף ליניארי של וקטור הבסיס. המספר המקסימלי של בסיסים שאפשר ליצור מ-$\underline{\underline{A}}$ הוא

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!}$$

<u>פתרון בסיסי</u> (Basic Solution) מתקבל ע"י הצבת $(n-m)$ מן המשתנים $\underline{x}$ שווים לאפס ופתרון המשוואות $\underline{\underline{B}}\underline{x}_B = \underline{b}$ עבור שאר m המשתנים, המסומנים $\underline{x}_B$. חלק מן המשתנים האלה יכול אף הוא לקבל ערך אפס, אלא שזה אפס "מחושב" ולא אפס "מוגדר".

פתרונות בסיסיים מתאימים לנקודות פינה של האזור האפשרי. נדגים זאת בציור הבא:



כאן יש שני משתנים מקוריים, x_1 ו-x_2, וארבעה אילוצים. לאחר הוספת משתני חסר ועודף לכל האילוצים יהיו בסה"כ שישה משתנים. אי לכך: $m=4$ ו-$n=6$. בבסיס יהיו לכן 4 משתנים בלכד, ואילו השניים הנותרים ישוו לאפס. הנקודות היחידות במישור המתואר בציור אשר בהן שניים מתוך שישה המשתנים שווים (בהגדרה) לאפס נמצאות בחיתוכי האילוצים: הנקודות A,B,C ו-D שבציור. למשל, בנקודה A מתאפסים משתני החסר/עודף של אילוצים 1 ו-2, ואילו x_1,x_2 ומשתני החסר/עודף של אילוצים 3 ו-4 שונים מאפס. בנקודה הפנימית E כל ששת המשתנים שונים מאפס.

כבר ראינו בפתרונות הגרפיים כי האופטימום נמצא בנקודת פינה של האזור האפשרי. היות והפינות הן פתרונות בסיסיים הרי שהאופטימום יהיה בפתרון בסיסי. זאת נראה כהמשך בפתרון הנוסרי גם כן.

6.4 שיפור פתרון בסיסי אפשרי

ראינו בסעיף 6.1 כי בתכנות ליניארי האיזור האפשרי הוא מצולע קמור, וכי הפינות של המצולע הם פתרונות בסיסיים של מערכת האילוצים. נראה כעת כיצד ניתן לעבור מפתרון בסיסי אחד לשני טוב ממנו.

נתחיל בעזרת דוגמה. נתונה הבעיה הבאה, בה האילוצים כבר מופיעים כשוויונות:

$$\left.\begin{aligned} min\ z = x_1 + 6x_2 - 7x_3 + x_4 + 5x_5 \\ 5x_1 - 4x_2 + 13x_3 - 2x_4 + x_5 = 20 \\ x_1 - x_2 + 5x_3 - x_4 + x_5 = 8 \end{aligned}\right\} \qquad (6\text{-}19)$$

היות ויש חמישה משתנים ורק שני אילוצים, יש $\binom{5}{2} = 10$ פתרונות בסיסיים. נבחר למשל $x_2 = x_3 = x_4 = 0$ ונמצא את הפתרון הבסיסי (x_1,x_5). לשם כך נחסר את האילוץ השני מהראשון ונקבל

$$4x_1 - 3x_2 + 8x_3 - x_4 = 12$$

נחלק שורה זו ב-4

$$x_1 - 3/4\ x_2 + 2x_3 - 1/4\ x_4 = 3$$

ונחסר את התוצאה מהאילוץ השני. מתקבל

$$-1/4\ x_2 + 3x_3 - 3/4\ x_4 + x_5 = 5$$

את שתי המשוואות האחרונות נרשום בצורה

$$x_1 = 3 + 3/4\ x_2 - 2x_3 + 1/4\ x_4 \qquad (6\text{-}11)$$

$$x_5 = 5 + 1/4\ x_2 - 3x_3 + 3/4\ x_4 \qquad (6\text{-}12)$$

היות והצבנו $x_2 = x_3 = x_4 = 0$ מתקבל הפתרון הבסיסי האפשרי $(x_1 = 3,\ x_5 = 5)$. אלמלא היה הבסיס אפשרי נאלצים היינו לנסות בסיסים אחרים. ואילו בחנו את כולם ולא מצאנו אף בסיס אפשרי היינו מסיקים כי לבעיה אין פתרון אפשרי.
ביחד עם הפעולות שביצענו לקבלת המשוואות (6-11) ו-(6-12) צריך לבצע גם פעולות שמטרתן לאפס את מקדמי x_1 ו-x_5 בפונקציית המטרה. קל ביותר כעת פשוט להציב את הביטויים שקבלנו לעיל עבור x_1 ו-x_5 בפונקציית המטרה. לאחר קיבוץ איברים מתקבלת התוצאה:

$$z = 28 + 8x_2 - 24x_3 + 5x_4 \qquad (6\text{-}13)$$

כלומר, הערך חנוכחי של פונקצית המטרה הוא $z = 28$.
יש לציין כאן כי הדרך של בחירת בסיס כלשהו וביצוע הפעולות החישוביות בצורה הבלתי-מחייבת כביכול שתוארה לעיל, תוחלף מאוחר יותר בדרך שיטתית ומסודרת יותר.

נראה כעת כיצד לבצע את אותן הפעולות באמצעות טבלאות. נחזור לבעיה המקורית ונרשום את מטריצת הבעיה בצורה

$$\left(\begin{array}{ccccc|c} 5 & -4 & 13 & -2 & 1 & 20 \\ 1 & -1 & 5 & -1 & 1 & 8 \\ 1 & 6 & -7 & 1 & 5 & z \end{array}\right) \qquad (6\text{-}14)$$

לאחר הפעולות שצויינו לעיל מתקבלת המטריצה

$$\left(\begin{array}{ccccc|c} 0 & -1/4 & 3 & -3/4 & 1 & 5 \\ 1 & -3/4 & 2 & -1/4 & 0 & 3 \\ 0 & 8 & -24 & 5 & 0 & z-28 \end{array}\right) \qquad (6\text{-}15)$$

מהסתכלות בביטוי החדש עבור פונקצית המטרה, (6-13), ברור כי אם רוצים להקטין את ערכו של z יש להגדיל את ערכו של x_3 מערכו חנוכחי $x_3 = 0$, ולהשאיר את x_2 ו-x_4 בערך אפס. זאת נוכל לראות גם בשורת פונקצית המטרה בטבלה (6-15). נניח שהיינו קובעים $x_3 = 0.5$. אזי

$$x_1 = 3 - 2 \times 0.5 = 2 \quad ; \quad x_5 = 5 - 3 \times 0.5 = 3.5$$

הפתרון שהתקבל $(x_1 = 2,\ x_2 = 0,\ x_3 = 0.5,\ x_5 = 0,\ x_5 = 3.5)$ הוא אפשרי, שכן אף משתנה אינו שלילי, אבל אינו בסיסי, משום שיש יותר משני משתנים שערכם שונה מאפס. פונקצית המטרה של פתרון זה היא

$$z = x_1 - 7x_3 + 5x_5 = 2 - 7 \times 0.5 + 5 \times 3.5 = 16$$

ואמנם, ערך זה טוב יותר מהקודם, שהיה 28. כדאי, אם כן, להמשיך ולהגדיל את x_3. המגבלה היא שהפתרון חייב להיות אפשרי, כלומר, אסור שמשתנים יהיו שליליים.

הערך הגדול ביותר שיכול x_3 לקבל הוא זה שיגרום לאחד משני המשתנים האחרים הנמצאים בפתרון, x_1 או x_5 להגיע לערך אפס. נמצא את הערך המקסימלי של x_3 בעזרת משוואות (6-11) ו-(6-12).

$$x_1 = 3 - 2x_3 \geq 0 \Longrightarrow x_3 \leq 3/2$$

$$x_5 = 5 - 3x_3 \geq 0 \Longrightarrow x_3 \leq 5/3$$

קובעת הדרישה $x_3 \leq 3/2$. המסקנה: $x_3 = 3/2$, ואז $x_1 = 0$. x_5 יקבל אז את הערך $x_5 = 5-3\times3/2 = 1/2$. זהו פתרון בסיסי חדש, שכן בדיוק שני משתנים שונים מאפס. הערך החדש של פונקצית המטרה הוא:

$$z = -7x_3 + 5x_5 = -7 \times 3/2 + 5 \times 1/2 = -8$$

נחזור כעת לטפל במטריצות. ראינו במטריצה (6-15) כי כדאי להכניס את x_3 לפתרון, שכן המקדם שלו בפונקצית המטרה שלילי. אגב, לו היו עוד משתנים לא בסיסיים בעלי מקדמים שליליים בפונקצית המטרה, היה עלינו לבחור אחד מהם כדי להכניסו לבסיס. נהוג לבחוד כנכנס לבסיס את המשתנה הלא בסיסי בעל המקדם השלילי עם הערך המוחלט הגדול ביותר, כלומר את המקדם בעל הערך המינימלי. ראינו גם שבמקום x_3 הנכנס לבסיס יש להוציא מהבסיס משתנה אחר. לשם כך יצרנו את המנות של האבר החופשי חלקי האבר המתאים בעמודה של x_3 ב-(6-15), בחרנו את המנה הקטנה ביותר והוצאנו מהבסיס את המשתנה הבסיסי שבשורה שלו קרה הדבר, כלומר, המשתנה המופיע עם מקדם 1 בשורה זו ועם אפסים בשורות האחרות. המנות הן 5/3 ו-3/2. הקטנה יותר היא השניה, והמשתנה הבסיסי של השורה השניה הוא x_1. נרשום שוב את (6-15),

$$\left(\begin{array}{ccccc|c} 0 & -1/4 & 3 & -3/4 & 1 & 5 \\ 1 & -3/4 & 2 & -1/4 & 0 & 3 \\ 0 & 8 & -24 & 5 & 0 & z-28 \end{array}\right)$$

כאשר מקיפים במעגל את האיבר המהווה את הציר (Pivot) לפעולה הבאה. זה האבד בעמודה של המשתנה הנכנס לבסיס ובשורה של המשתנה העוזב אותו. כדי לקבל את הפתרון הבסיסי החדש, ואת ערך פונקצית המטרה בפתרון החדש, יש לבצע התמרה שתביא את המטריצה (6-15) למצב בו המקדם של הציר הוא 1 והמקדמים האחדים באותה עמודה

יתאפסו. התוצאה היא

$$\left[\begin{array}{ccccc|c} -3/2 & 7/8 & 0 & -3/8 & 1 & 1/2 \\ 1/2 & -3/8 & 1 & -1/8 & 0 & 3/2 \\ 12 & -1 & 0 & 2 & 0 & z+8 \end{array}\right] \qquad (6\text{-}16)$$

וכה רואים את הפתרון הבסיסי $(x_3 = 3/2,\ x_5 = 1/2)$ ואת הערך $z=-8$. נשים לב שכהתמרה מ-(6-15) ל-(6-16) "התקלקלה" העמודה הראשונה מעמודת יחירה, ואילו החמישית נשארה ללא שינוי.

6.5 מציאת הפתרון הבסיסי האופטימלי

נמשיך את התהליך על אותה רוגמה. אנו רואים שהמקדם של x_2 בשורה של z שלילי, כלומר, כראי להכניסו לבסיס. כדי למצוא את המשתנה החייב לעזוב את הבסיס נשתמש כתנאים:

$$x_5 = 1/2 - 7/8\ x_2 \geq 0 \implies x_2 \leq 4/7 \qquad (6\text{-}17)$$

$$x_3 = 3/2 + 3/8\ x_2 \geq 0 \implies x_2 \geq -4 \qquad (6\text{-}18)$$

הדרישה השניה אינה מגבלה כלל, שכן בלאו הכי נדרש $x_2 \geq 0$. לכן קובעת הדרישה הראשונה, ומכאן שעל x_5 לעזוב את הבסיס. הציר הוא אם כן, האיבר הראשון בעמודה השניה (7/8). לאחר הטרנספורמציה מתקבלת המטריצה החדשה,

$$\left[\begin{array}{ccccc|c} -12/7 & 1 & 0 & -3/7 & 8/7 & 4/7 \\ -1/7 & 0 & 1 & -2/7 & 3/7 & 12/7 \\ 72/7 & 0 & 0 & 11/7 & 8/7 & z+60/7 \end{array}\right] \qquad (6\text{-}19)$$

הפתרון הבסיסי הוא $(x_2=4/7,\ x_3=12/7)$. היות ואין מקדמים שליליים בפונקצית המטרה החדשה, הסתיים התהליך וזהו הפתרון האופטימלי, עכורו $z = z^* = -60/7$. לו היינו עוסקים כמקסימיזציה כמקום במינימיזציה, היה הקריטריון להכנסת משתנה

לבסיס מקדם חיובי בפונקצית המטרה. הקריטריון לבחירת המשתנה היוצא מהבסיס לא היה משתנה. התהליך היה מסתיים כאשר כל המקדמים בפונקצית המטרה היו שליליים.

6.6 פתרון לא חסום

ראינו כי בעת איתור הציר היה עלינו לבחור את המשתנה הבסיסי אשר בשורת שלו התקבלה המנה המינימלית של הצד הימני חלקי האבר הנמצא בעמודה של המשתנה הנכנס לבסיס, בתנאי שאבר זה אינו שלילי. הצד הימני תמיד חיובי, שכן הוא ערכו של משתנה בסיסי. כאשר האבר בעמודה של המשתנה הנכנס שלילי, אין הגבלה על ערכו של המשתנה הנכנס. מכאן משתמעת מסקנה חשובה. לו היה מתקבל במהלך האיטרציות כי האבר הראשון של העמודה השניה ב-(6-16) גם הוא שלילי, היו האילוצים (6-17) ו-(6-18) מקבלים את הצורה,

$$x_5 = 1/2 + 7/8\ x_2 \geq 0 \Longrightarrow x_2 \geq -4/7 \qquad (6\text{-}20)$$

$$x_3 = 3/2 + 3/8\ x_2 \geq 0 \Longrightarrow x_2 \geq -4 \qquad (6\text{-}21)$$

יכולנו כעת להגדיל את x_2 ללא גבול, מבלי לגרום לאחד מהמשתנים הבסיסיים הנוכחים, x_3 ו-x_5, להפוך שליליים. תוך הגדלה זו של x_2 היתה פונקצית המטרה קסנה כל הזמן, ללא גבול. כמלים אחרות, הפתרון אינו חסום כי אפשר לקבל $z \to (-\infty)$. מצאנו אם כן את הסימן לכך שפתרון אינו חסום והוא:

אם מתגלה באחת האיטרציות כי כל האברים הם שליליים בעמודה כלשהי בה המקדם כפונקצית המטרה הוא שלילי (עבור מינימיזציה, וחיובי עבור מקסימיזציה), אזי הפתרון אינו חסום.

נשים לב שהפתרון אינו חסום כאשר הרבר קורה בעמודה כלשהי בה המקדם בפונקצית המטרה שלילי, ולאו דוקא בזו כעלת הערך המינימלי.

6.7 פתרון מנוון

לו היה האיבר הימני באילוץ הראשון של הבעיה הנתונה 8 כמקום 20, היה מתקבל לאחר חיסור האילוץ השני מהראשון.

$$4x_1 - 3x_2 + 8x_3 - x_4 = 0$$

או :

$$x_1 - 3/4x_2 + 2x_3 - 1/4x_4 = 0$$

כלומר, במטריצה (6-15) היה מופיע בשורה השנייה מימין 0 במקום 3. הפתרון הבסיסי ($x_1 = 0$; $x_5 = 8$) הוא מנוון, שכן אחד המשתנים בו התאפס. ברצותנו להכניס את x_3 לבסיס היה נדרש לקיים את האילוץ,

$$x_1 = 0 - 2x_3 \geq 0 \Longrightarrow x_3 \leq 0 \Longrightarrow x_3 = 0$$

כל שינוי של x_3 היה גורם ל- x_1 להיות שלילי, ולכן שינוי כזה אסור. המסקנה: למרות שהמקדם של x_3 בפונקציית המטרה שלילי הוא יכול להכנס לבסיס רק עם ערך אפס. כלומר, עוברים מהפתרון מנוון אחד (בו $x_1 = 0$) לשני (בו $x_3 = 0$), ו- z אינו משתנה. זה עלול להוליך לחוג סגור (Cycling), אשר בו הפתרון חוזר באופן מחזורי על אותם בסיסים. באופן מעשי נדיר מקרה זה ביותר, והדוגמאות המוכרות הן מלאכותיות ברובן. עם זאת מכילות תכניות המחשב הקיימים אמצעים להתגבר על מקרים כאלה.

6.8 דוגמאות של מהלך הפתרון

להלן הפתרון המספרי של הבעיה שנפתרה בצורה גרפית בסעיף 6.2.1.

$$min \quad z = -x_1 - x_2$$

$$s.t. \quad 2x_1 + x_2 \leq 4$$

$$x_1 + 3x_2 \leq 6$$

את בעית המקסימיזציה הפכנו למינימיזציה על ידי היפוך הסימן. את התנאים x_1 , $x_2 \geq 0$ לא רשמנו בניסוח הבעיה, וזאת בדי להדגיש שתנאים אלה הם חלק אינטגרלי מכל בעית תכנות ליניארי, ולכן גם אם אינם רשומים במפורש הם חייבים להתמלא.

אחרי הוספת משתני חסר יהיו האילוצים

$$2x_1 + x_2 + x_3 = 4$$

$$x_1 + 3x_2 + x_4 = 6$$

המטריצה היא

$$\left[\begin{array}{cccc|c} 2 & 1 & 1 & 0 & 4 \\ 1 & 3 & 0 & 1 & 6 \\ -1 & -1 & 0 & 0 & z \end{array}\right]$$

הפתרון הבסיסי הוא $(x_3 = 4 \; ; \; x_4 = 6)$ ואילו $x_1 = x_2 = 0$. זו פינת האיזור האפשרי המהווה את ראשית הצירים. (ראה ציור בסעיף 6.2.1). ערך פונקצית המטרה הוא $z = 0$. היות ויש שני מקדמים שליליים שווים בפונקצית המטרה, נבחר אחד מהמשתנים. נהוג לקחת את הראשון מהם, כאן את x_1 , כנכנס לבסיס. המשתנה היוצא יבחר לפי

$$min \{ \frac{4}{2} , \frac{6}{1} \} = 2$$

כלומר , x_3 עוזב את הבסיס. הציר הוא האיבר 2 בשורה הראשונה. לאחר ההתמרה מקבלים

$$\left[\begin{array}{cccc|c} 1 & 1/2 & 1/2 & 0 & 2 \\ 0 & 5/2 & -1/2 & 1 & 4 \\ 0 & -1/2 & 1/2 & 0 & z+2 \end{array}\right]$$

הפתרון הבסיסי החדש הוא $(x_1 = 2 \; ; \; x_5 = 4)$, שהוא בפינה הימנית התחתונה של האיזור האפשרי, כפי שהוא מתואר כסעיף 6.2.1. כאן $z = -2$. הציר החדש מסומן, ולאחר ההתמרה מקבלים

$$\left[\begin{array}{cccc|c} 1 & 0 & 3/5 & -1/5 & 6/5 \\ 0 & 1 & -1/5 & 2/5 & 8/5 \\ 0 & 0 & 2/5 & 1/5 & z+14/5 \end{array}\right]$$

זה הפתרון האופטימלי ($x_1^* = 8/5 = 1.6$; $x_2^* = 6/5 = 1.2$), עבורו $z^* = 14/5 = 2.8$, כפי שקבלנו בפתרון הגרפי.

נסקור את מהלך הפתרון : עברנו מפינה אחת של האיזור האפשרי לפינה שכנה, בה הפתרון טוב יותר, עד שהתברר שאין פינה שכנה טובה יותר. לא עברנו אף פעם דרך האיזור האפשרי, אלא נשארנו על גבולו ועברנו מפינה אחת לשכנתה.

לו היינו בוחרים להכניס בשלב הראשון את x_2 במקום את x_1 היה התהליך מסתכם במטריצות הבאות :

$$\left[\begin{array}{cccc|c} 2 & 1 & 1 & 0 & 4 \\ 1 & 3 & 0 & 1 & 6 \\ -1 & -1 & 0 & 0 & z \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 5/3 & 0 & 1 & -1/3 & 2 \\ 1/3 & 1 & 0 & 1/3 & 2 \\ -2/3 & 0 & 0 & 1/3 & z+2 \end{array}\right]$$

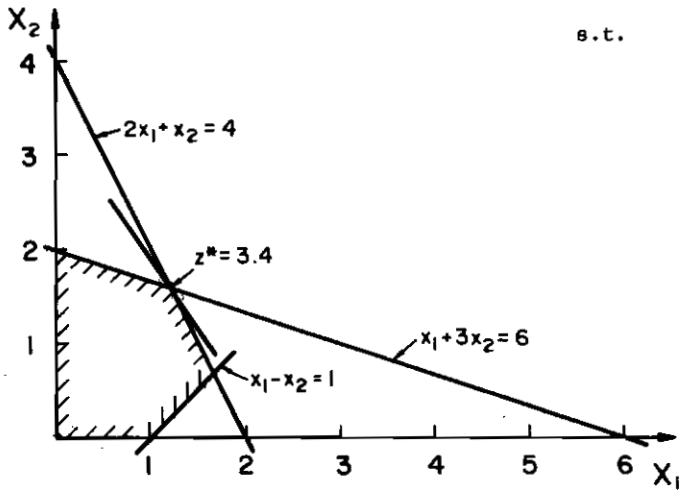
$$\left[\begin{array}{cccc|c} 1 & 0 & 3/5 & -1/5 & 6/5 \\ 0 & 1 & -1/5 & 2/5 & 8/5 \\ 0 & 0 & 2/5 & 1/5 & z+14/5 \end{array}\right]$$

בריוק אותו פתרון כמו קודם, אלא שהפעם עברנו מן הראשית לפינה העליונה של האיזור האפשרי ואחר-כך למינה לפתרון האופטימלי. נשים לב <u>שכל</u> אברי המטריצה הסופית שווים לאלה שהתקבלו קודם.

נראה כעת שבחירת המשתנה הנכנס לבסיס לפי המקדם בעל הערך המוחלט הגדול ביותר אינה בהכרח הטובה ביותר. הרוגמה הבאה דומה לקודמת, אלא עם אילוץ נוסף ובשינוי המקדם של x_1 בפונקצית המטרה :

$$max\ z = 1.5x_1 + x_2$$

$$\begin{aligned} s.t. \quad & 2x_1 + x_2 \le 4 \\ & x_1 + 3x_2 \le 6 \\ & x_1 - x_2 \le 1 \\ & x_1 , x_2 \ge 0 \end{aligned}$$



הפתרון הגרפי הוא ($x_1^* = 1.2$; $x_2^* = 1.6$; $z^* = 3.4$). את הפתרון המספרי נבצע בטבלאות. הפעם נבצע מקסימיזציה, לצורך תרגול, ונבחר את המשתנה הנכנס לפי המקדם החיובי הגדול ביותר.

הוספת משתני חסר :

$$max \quad z = 1.5x_1 + x_2$$

$$\begin{aligned} 2x_1 + x_2 + x_3 \qquad\qquad &= 4 \\ x_1 + 3x_2 \qquad + x_4 \qquad &= 6 \\ x_1 - x_2 \qquad\qquad + x_5 &= 1 \end{aligned}$$

$$\left[\begin{array}{ccccc|c} 2 & 1 & 1 & 0 & 0 & 4 \\ 1 & 3 & 0 & 1 & 0 & 6 \\ 1 & -1 & 0 & 0 & 1 & 1 \\ 3/2 & 1 & 0 & 0 & 0 & z \end{array}\right] \qquad \begin{array}{l} x_1 = x_2 = 0 \\ x_3 = 4, x_4 = 6, x_5 = 1 \\ z = 0 \end{array}$$

$$\left[\begin{array}{ccccc|c} 0 & 3 & 1 & 0 & -2 & 2 \\ 0 & 4 & 0 & 1 & -1 & 5 \\ 1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 5/2 & 0 & 0 & -3/2 & z-3/2 \end{array}\right] \qquad \begin{array}{l} x_2 = x_5 = 0 \\ x_1 = 1, x_3 = 2, x_4 = 5 \\ z = 1.5 \end{array}$$

$$\left[\begin{array}{ccccc|c} 0 & 1 & 1/3 & 0 & -2/3 & 2/3 \\ 0 & 0 & -4/3 & 1 & 5/3 & 7/3 \\ 1 & 0 & 1/3 & 0 & 1/3 & 5/3 \\ 0 & 0 & -5/6 & 0 & 1/6 & z-19/6 \end{array}\right] \qquad \begin{array}{l} x_3 = x_5 = 0 \\ x_1 = 5/3, x_2 = 2/3, x_4 = 7/3 \\ z = 19/6 \end{array}$$

$-1/5$

$$\left[\begin{array}{ccccc|c} 0 & 1 & 13/15 & 2/5 & 0 & 8/5 \\ 0 & 0 & -4/5 & 3/5 & 1 & 7/5 \\ 1 & 0 & 3/5 & -1/5 & 0 & 6/5 \\ 0 & 0 & -7/10 & -1/10 & 0 & z-17/5 \end{array}\right] \qquad \begin{array}{l} x_3^* = x_4^* = 0 \\ x_1^* = 6/5, x_2^* = 8/5, x_5^* = 7/5 \\ z^* = 17/5 = 3.4 \end{array}$$

ברור מהתאור הגרפי שעברנו מהראשית אל הנקודה הסופית נגד כיוון השעון, פינה אחר פינה. במסלול זה יש, נוסף לראשית ולפתרון הסופי שתי נקודות ביניים. לו בחרנו להכניס את x_2 <u>לפני</u> x_1 , למרות שהמקדם שלו בטבלה המקורית קטן יותר, היינו חוסכים איטרציה אחת, כפי שניווכח בפתרון המספרי הבא :

$$\left[\begin{array}{ccccc|c} 2 & 1 & 1 & 0 & 0 & 4 \\ 1 & 3 & 0 & 1 & 0 & 6 \\ 1 & -1 & 0 & 0 & 1 & 1 \\ 3/2 & 1 & 0 & 0 & 0 & z \end{array}\right] \qquad \begin{array}{l} x_1 = x_2 = 0 \\ x_3 = 4, x_4 = 6,\ x_5 = 1 \\ z = 0 \end{array}$$

$$\left[\begin{array}{ccccc|c} 5/3 & 0 & 1 & -1/3 & 0 & 2 \\ 1/3 & 1 & 0 & 1/3 & 0 & 2 \\ 4/3 & 0 & 0 & 1/3 & 1 & 3 \\ 7/6 & 0 & 0 & -1/3 & 0 & z-2 \end{array}\right] \qquad \begin{array}{l} x_1 = x_4 = 0 \\ x_2 = 2, x_3 = 2, x_4 = 3 \\ z = 2 \end{array}$$

$$-\frac{1}{5} \longrightarrow \left[\begin{array}{ccccc|c} 1 & 0 & 3/5 & -1/5 & 0 & 6/5 \\ 0 & 1 & 13/15 & 2/5 & 0 & 8/5 \\ 0 & 0 & -4/5 & 3/5 & 1 & 7/5 \\ 0 & 0 & -7/10 & -1/10 & 0 & z-17/5 \end{array}\right] \qquad \begin{array}{l} x_3^* = x_4^* = 0 \\ x_1^* = 6/5, x_2^* = 8/5,\ x_5^* = 7/5 \\ z^* = 17/5 = 3.4 \end{array}$$

<u>מסקנה</u>: <u>קצב</u> השיפור של פונקציית המטרה אינו בהכרח הקריטריון הטוב ביותר לבחירת המשתנה הנכנס לבסיס. נשים לב כי בפתרון הראשון יצא x_5 מהבסיס באיטרציה הראשונה, ואילו באיטרציה האחרונה הוא חזר לבסיס. לו יכולנו לצפות קדימה באיטרציה הראשונה ולדעת ש-x_5 יתזור לבסיס, היינו יכולים אולי להשתמש באינפורמציה זו כדי להחליט להכניס את x_2 לפני x_1. ואמנם, תוכניות מחשב לפתרון תכנות ליניארי משתמשות בדרך כלל בקריטריונים מורכבים יותר לבחירת המשתנה הנכנס לבסיס, בנסיון לבטל איטרציות מיותרות ובכך לייעל את הפתרון.

ראוי לציין כי הפתרון הבסיסי הראשון התקבל בשתי הדוגמאות בצורה פשוטה מאד, משום שבמטריצה הראשונה כבר היתה מטריצת יחידה. המשתנים בבסיס הראשון היו כולם משתני חסר. זהו תמיד המצב כאשר כל האילוצים הם (≤). כאשר יהיו אילוצים מן הסוגים (≥) או (=) לא תתקבל מטריצת יחידה לאחר הוספת משתני חסר ועודף.

6.9 שיטת הסימפלקס המעודכן

פתרון התכנות הליניארי בשיטת הסימפלקס (Simplex method) היא למעשה זו שהשתמשנו בה לעיל משיקולים אינטואיטיביים. שיטת החישוב המקובלת היא הסימפלקס המעודכן (Revised Simplex).

לאחר הוספת משתני חסר ועודף, מקבלת בעית התכנות הליניארי את הצורה הבאה:

$$min\ z = \underline{c}\ \underline{x} \tag{6-23}$$

$$s.t.\ \underline{\underline{A}}\ \underline{x} = \underline{b} \tag{6-24}$$

$$\underline{x} \geq \underline{0} \tag{6-25}$$

כאשר $\underline{\underline{A}}$ היא $(m\times n)$, כה $\underline{a}_j$ $(j=1,\ldots,n)$ היא העמודה ה-j. עמודות של $\underline{\underline{A}}$ מרכיבות את מטריצת הבסיס $\underline{\underline{B}} = (\underline{b}_1,\ldots,\underline{b}_m)$, שהיא $(m\times m)$, לא סינגולרית. סדר האינדקסים $(1,\ldots,m)$ הוא לפי הופעת הוקטורים ב-$\underline{\underline{B}}$, ואין קשר בין אינדקס זה לאינדקס j של העמודה במטריצה $\underline{\underline{A}}$.

כל עמודה $\underline{a}_j$ שאינה בבסיס יכולה להכתב בצורה:

$$\underline{a}_j = y_{1j}\underline{b}_1 + \ldots + y_{mj}\underline{b}_m = \sum_{i=1}^{m} y_{ij}\underline{b}_i = \underline{\underline{B}}\ \underline{y}_j \tag{6-26}$$

כאשר

$$\underline{y}_j = \underline{\underline{B}}^{-1}\ \underline{a}_j \tag{6-27}$$

ראה נספח א', משוואות (52-א') ו-(59-א') ו-(70.1,2-א').

הפתרון הבסיסי הנוכחי הוא

$$\underline{x}_B = (x_{B_1},\ldots,x_{B_m}) = \underline{\underline{B}}^{-1}\ \underline{b} \tag{6-28}$$

כל בחירה של בסיס $\underline{\underline{B}}$ נותנת $\underline{x}_B$ אחר, כאשר כל המשתנים האחרים, שאינם בבסיס, שווים לאפס. x_{B_i} מתייחס לעמודה ה-i של $\underline{\underline{B}}$, אבל היות וב-$\underline{\underline{B}}$ סדר העמודות שונה מן הסדר ב-$\underline{\underline{A}}$ אין האינדקס i נותן את מספרו הסידורי של המשתנה במשוואות $\underline{\underline{A}}\ \underline{x} = \underline{b}$.

וקטור מחירי המשתנים הבסיסיים הוא $\underline{c}_B = (c_{B_1},\ldots,c_{B_m})$, כאשר יש התאמה בסדר לזה של הבסיס $\underline{\underline{B}}$. הערך הנוכחי שלפונקציית המטרה הוא

$$z = \underline{c}_B\ \underline{x}_B$$

בתהליך האיטרציות אנו בוחרים, לפי קריטריון שיוצג להלן, וקטור שאינו נמצא בבסיס ומכניסים אותו לבסיס במקום וקטור הנבחר לעזוב את הבסיס. הוקטור $\underline{a}_j$ שאינו בבסיס יכול להכנס לבסיס ולהחליף בו את $\underline{b}_r$ בתנאי שבמשוואה (6-26) קיים עבור זוג זה $y_{rj} \neq 0$. אם התנאי מתקיים, ומכניסים את $\underline{a}_j$, אפשר לבטא את הוקטור $\underline{b}_r$ בעזרת הבסיס החדש בצורה

$$\underline{b}_r = \frac{1}{y_{rj}}\ \underline{a}_j - \sum_{\substack{i=1\\ i\neq r}}^{m} \frac{y_{ij}}{y_{rj}}\ \underline{b}_i \qquad (6\text{-}29)$$

הפתרון הבסיסי הקודם היה $\underline{\underline{B}}\ \underline{x} = \underline{b}$, או

$$\sum_{i=1}^{m} x_{B_i}\ \underline{b}_i = \underline{b} \qquad (6\text{-}30)$$

נחליף כעת את $\underline{b}_r$ בעזרת (6-29). מתקבל

$$\sum_{\substack{i=1\\ i\neq r}}^{m} x_{B_i}\ \underline{b}_i + [\frac{1}{y_{rj}}\ \underline{a}_j - \sum_{\substack{i=1\\ i\neq r}}^{m} \frac{y_{ij}}{y_{rj}}\ \underline{b}_i]x_{B_r} = \underline{b} \qquad (6\text{-}31)$$

או, לאחר סידור המשוואה

$$\sum_{\substack{i=1\\ i\neq r}} (x_{B_i} - x_{B_r}\frac{y_{ij}}{y_{rj}})b_i + \frac{x_{B_r}}{y_{rj}}\ \underline{a}_j = \underline{b} \qquad (6\text{-}32)$$

הערכים של המשתנים הבסיסיים החדשים מתקבלים מפתרון משוואה זו. על מנת שהפתרון יהיה אפשרי דרוש שכל המשתנים כו יהיו לא שליליים, כלומר, דרוש שיתקיימו התנאים

$$x_{B_i} - x_{B_r}\frac{y_{ij}}{y_{rj}} \geq 0 \qquad i \neq r \tag{6-33}$$

$$\frac{x_{B_r}}{y_{rj}} \geq 0 \tag{6-34}$$

x_{B_i} (כולל x_{B_r}) הם הערכים של המשתנים הבסיסיים בפתרון הכסיסי הקודם, שגם בו דרשנו שיהיה אפשרי, ולכן הם מקיימים: $x_{B_i} \geq 0 \quad (i=1,\ldots,m)$ כדי שיתקיים (6-34) דרוש $y_{rj} \geq 0$.

אם רוצים לעבור מפתרון בסיסי אחד לשני יש לדרוש שכאשר וקטור נכנס לבסיס, וערך המשתנה המתאים לו מקבל ערך חיובי כעוד שקודם היה אפס, באותו זמן יעזוב משתנה אחר את הכסיס, כלומר יתאפס. הרכר נעשה אם דורשים שבדיוק אחד מהאילוצים (6-33) יתקיים כשוויון, וכל האחרים עדיין יתקיימו. אותם אילוצים מהצורה (6-33) שבהם $y_{ij} < 0$ מתקיימים עבור כל עדך חיובי של x_{B_r} כלומר, הם אינם מציבים מגבלה על ערכו. באותם האילוצים שבהם $y_{ij} > 0$ נחלק בו ונקבל

$$\frac{x_{B_i}}{y_{ij}} - \frac{x_{B_r}}{y_{rj}} \geq 0 \qquad (y_{ij} > 0 \ , \ i \neq r) \tag{6-35}$$

יש לקבוע את r, המשתנה העוזב את הבסיס, כך שאילוץ (6-35) מתקיים בו כשוויון, שכן אז יתאפס המשתנה הבסיסי המתאים. זה קורה אם בוחרים את r לפי

$$\frac{x_{B_r}}{y_{rj}} = \min_i \left\{ \frac{x_{B_i}}{y_{ij}} \,\middle|\, y_{ij} > 0 \right\} = \theta \tag{6-36}$$

ותמיד קיים $\theta \geq 0$.

לאחר שהוחלף $\underline{b}_r$ ב-$\underline{a}_j$ מתקבל כסיס חדש, שיסומן $\hat{\underline{\underline{B}}}$ (מה שסומן בנספח א' ב-$\underline{\underline{B}}_a$),

$$\hat{\underline{\underline{B}}} = (\underline{b}_1,\ldots,\underline{b}_{r-1},\underline{a}_j,\underline{b}_{r+1},\ldots,\underline{b}_m) \tag{6-37}$$

הפתרון הבסיסי החדש הוא

$$\hat{\underline{x}}_B = \hat{\underline{\underline{B}}}^{-1}\,\underline{b} \tag{6-38}$$

הנתון גם על ידי המשוואות

$$\hat{x}_{B_i} = x_{B_i} - x_{B_r}\frac{y_{ij}}{y_{rj}} \qquad i \neq r \tag{6-39}$$

$$\hat{x}_{B_r} = \frac{x_{B_r}}{y_{rj}} \tag{6-40}$$

הערך החדש של פונקצית המטרה הוא $\hat{z} = \hat{\underline{c}}_B\,\hat{\underline{x}}_B$, באשר המחירים הם

$$\hat{c}_{B_i} = c_{B_i} \qquad i \neq r \tag{6-41}$$

$$\hat{c}_{B_r} = c_j \tag{6-42}$$

נשתמש במשוואות (6-39) עד (6-42) כדי לבטא את ערכה החדש של פונקצית המטרה בצורה

$$\hat{z} = \sum_{\substack{i=1\\ i\neq r}}^{m} c_{B_i}\left(x_{B_i} - x_{B_r}\frac{y_{ij}}{y_{rj}}\right) + c_j\frac{x_{B_r}}{y_{rj}} \tag{6-43}$$

על התנאי $i\neq r$ אפשר לוותר, משום שהוא תורם אפס לסכום, שכן הביטוי בסוגריים מתאפס עבור $i=r$. לכן

$$\hat{z} = \sum_{i=1}^{m} c_{B_i}\,x_{B_i} - \frac{x_{B_r}}{y_{rj}}\sum_{i=1}^{m} c_{B_i}\,y_{ij} + \frac{x_{B_r}}{y_{rj}}\,c_j \tag{6-44}$$

נגדיר כעת שורה של גרלים חדשים, אחר לכל עמודה j של משתנה לא בסיסי באיטרציה הקודמת, זו כה היה הבסיס $\underline{x}_B$,

$$z_j = \sum_{i=1}^{m} c_{B_i} y_{ij} \tag{6-45}$$

אזי משוואה (6-44) מקבלת את הצורה

$$\hat{z} = z + \frac{x_{B_r}}{y_{rj}} (c_j - z_j) = z + \theta(c_j - z_j) \tag{6-46}$$

אם ברצוננו לבצע מינימיזציה יש לבחור את הוקטור הנכנס a_j מבין אלה המקיימים את התנאי

$$\theta(c_j - z_j) < 0 \tag{6-47}$$

לשם כך דרוש

$$c_j - z_j < 0 \tag{6-48}$$

שכן $\theta \geq 0$. מהסתכלות במשוואה (6-46) נראה שכדאי היה לבחור את הוקטור הנכנס a_j לפי הקריטריון

$$\frac{x_{B_r}}{y_{rj}} (c_j - z_j) = \min_k \{\frac{x_{B_r}}{y_{rk}} (c_k - z_k) \mid (c_k - z_k) < 0\} \tag{6-49}$$

כדי להשתמש במשוואה (6-49) יש לבדוק את כל הוקטורים שאינם בבסיס; עבור כל וקטור המקיים את (6-48) יש לבצע את החישובים הדרושים להכנסתו לבסיס, לבחור את הוקטור העוזב לפי (6-36) ואז לכחור את הוקטור הנכנס לפי (6-49). מסתבר שחישובים אלה רבים מדי, ואינם מצדיקים את המאמץ החישובי. במקום זה נהוג לבחור כנכנס לבסיס את הוקטור $\underline{a}_j$ לפי הקריטריון הפשוט יותר

$$c_j - z_j = \min_k \{(c_k - z_k) \mid (c_k - z_k) < 0\} \tag{6-50}$$

כאשר לא נותר מחוץ לבסיס אף וקטור המקיים את התנאי (6-48) הפתרון הבסיסי הנוכחי הוא האופטימלי.

<u>הערה</u>: כאשר מבצעים מקסימיזציה יש להפוך את כיוון אי-השוויון ב-(6-48), (6-49) ו-(6-50) ולדרוש ב-(6-49) ו-(6-50) מקסימום במקום מינימום.

פתרון לא חסום

הפתרון הבסיסי הנוכחי הוא

$$\sum_{i=1}^{m} \underline{b}_i x_{B_i} = \underline{b} \qquad (6\text{-}51)$$

הוקטור שאינו בבסיס $\underline{a}_j$ נתון על ידי

$$\underline{a}_j = \sum_{i=1}^{m} y_{ij} \underline{b}_i \qquad (6\text{-}52)$$

נוסיף ונחסר באגף השמאלי של (6-51) θ כפול הוקטור $\underline{a}_j$, כאשר $\theta > 0$ הוא מקדם,

$$\sum_{i=1}^{m} \underline{b}_i x_{B_i} - \theta \underline{a}_j + \theta \underline{a}_j = \underline{b} \qquad (6\text{-}53)$$

בהצבת ערכו של $\underline{a}_j$ ממשוואה (6-52) מתקבל

$$\sum_{i=1}^{m} (x_{B_i} - \theta y_{ij}) \underline{b}_i + \theta \underline{a}_j = \underline{b} \qquad (6\text{-}54)$$

היות ודרשנו $\theta > 0$, הרי שאם $y_{ij} < 0$ עבור $(i=1,\ldots,m)$ יהיה המקדם הראשון תמיד חיובי, ועבור $\theta > 0$ כלשהו מתקבל פתרון אפשרי. פתרון זה לא יהיה בסיסי, שכן יש בו $(m+1)$ וקטורים. לכן, אם עכור וקטור לא-בסיסי כלשהו קיים התנאי (6-47) וכל ה-y_{ij} המתאימים שליליים, אפשר להגריל את θ במשוואה (6-46) בלי גבול, תוך שיפור מתמיד של פונקצית המטרה החרשה z. זהו, אם כן, המקרה של פתרון בלתי חסום.

פתרון מנוון

אם כאיטרציה כלשהי הפתרון הכסיסי מנוון, כלומר, $x_{B_i} = 0$, עבור אחד או יותר משתנים, אזי ייתכן שבאיטרציה הבאה, בה ייכנס הוקטור r לבסיס, יתקבל $x_{B_r} = 0$, וזאת, לפי משוואה (6-36), אם קיים $y_{rj} > 0$. במקרה כזה $\theta = 0$, ואין שיפור בערך פונקצית המטרה באיטרציה זו.

במקרה כזה ייתכן, לפחות באופן תיאורטי, שהאלגוריתם לא יתכנס לפתרון אופטימלי, אלא ייכנס לחוג סגור (Cycling), בו לאחר מספר צעדים נחזור לטבלה המקורית. יש דרכים להתגבר על הבעיה, על ידי הוספת קריטריונים לבחירה חד-משמעית של הוקטורים הנכנסים ועוזבים את הבסיס במקרים שבהם יש "תיקו" בין שני וקטורים לפי הקריטריונים שנזכרו לעיל. רוב האלגוריתמים הסטנדרטיים לוקחים את האפשרות בחשבון, למרות שבכל הנסיון הרב שנצטבר בשימוש בתכנות ליניארי, לא דווח אפילו על מקרה מעשי אחד שהדבר קרה. כדי להדגיש שהאפשרות קיימת לא רק באופן תיאורטי נבנו שתיים או שלוש דוגמאות מלאכותיות ועליהן מדווח בספרות (למשל בפרק 7 בספרו של Gass).

התכנסות לפתרון האופטימלי

היות ויש מספר סופי של פתרונות בסיסיים לכל בעיה, והיות והאלגוריתם של שיטת הסימפלקס מבטיח שבכל איטרציה הפתרון הבסיסי טוב מהקודם (פרט, כמובן, למקרה של ניוון), הרי שמובטח כי האלגוריתם מתכנס אל הפתרון האופטימלי במספר סופי של צעדים.

שינויים מסויימים בקריטריונים להחלפת וקטורים בבסיס יכולים להבטיח כי גם כאשר יש פתרון מנוון לא ייכנס האלגוריתם לחוג סגור. הוספתם מבטיחה כי האלגוריתם מתכנס תמיד לפתרון האופטימלי. שוב נזכיר כי בשל מורכבותם של שינויים אלה, ומשום שחוג סגור אינו מתקבל בכל הנראה בבעיות מעשיות, הובנסו השינויים בקריטריונים רק לחלק מתוכניות המחשב הקיימות.

סיכום השיטה

אם יש בידינו פתרון בסיסי $\underline{x}_B$, המתאים למטריצה הבסיסית $\underline{B}$, וידועים הערכים (המחושבים) y_j ו-(c_j-z_j) עבור כל המשתנים הלא-בסיסיים, מהלך האיטרציה הבאה הוא:

1. בדוק את ערכי c_j-z_j: עבור בעיית מינימיזציה:

   1.1 אם $c_j - z_j \geq 0$ עבור כל j, הפתרון הבסיסי הוא האופטימלי.

   1.2 $c_j - z_j < 0$ עבור j כלשהו, שכו $y_{ij} \leq 0$ עבור כל i, הפתרון בלתי חסום.

1.3 $c_j - z_j < 0$ עבור אחד או יותר ערכים של j, ובכל אחד מהם לפחות y_{ij} אחד חיובי, בוחרים אחד מה-a_j להכנסה לבסיס.

2. אם 1.3 מתקיים בוחרים את $\underline{b}_r$ היוצא מהבסיס לפי (6-36).

3. מבצעים את המעבר לפתרון הבסיסי החדש $\underline{x}_B$ לפי משוואות (6-39) ו-(6-40), מחשבים את הוקטורים $\hat{\underline{y}}_j$ ואת ערכי $(\hat{c}_j-\hat{z}_j)$ החדשים וחוזרים ל-(1).

באופן מעשי, עושים לרוב את (1.3) לפי הקריטריון (6-50), ורק אחר כך בורקים עבור הוקטור הנכנס בלבד את (1.2) ולא עבור כל אלה המקיימים $(c_j - z_j < 0)$.

<u>דוגמה</u>

נתונה הבעיה

$$max \quad z = 3x_1 + 2x_2 + x_3$$

$$s.t. \quad 3x_1 + 2x_2 + 2x_3 = 8$$

$$3x_1 + 4x_2 + x_3 = 7$$

נניח שבאיטרציה בלשהי הבסיס הוא

$$\underline{x}_B = (x_{B1}, x_{B2}) = (x_3, x_2)$$

יש לשים לב לסדר הוקטורים המקוריים, x_j, בבסיס: למשל $x_{B_1} = x_3$. המטריצה הבסיסית המתאימה, וההפכית שלה, הן

$$\underline{\underline{B}} = \begin{pmatrix} 2 & 2 \\ 1 & 4 \end{pmatrix} \qquad \underline{\underline{B}}^{-1} = \frac{1}{6}\begin{pmatrix} 4 & -2 \\ -1 & 2 \end{pmatrix}$$

שוב יש לשים לב כי סדר הוקטורים בבסיס מתאים לזה שב-$\underline{x}_B$. הפתרון הבסיסי באיטרציה זו הוא

$$\underline{x}_B = \underline{\underline{B}}^{-1}\,\underline{b} = \frac{1}{6}\begin{pmatrix} 4 & -2 \\ -1 & 2 \end{pmatrix} \cdot \begin{Bmatrix} 8 \\ 7 \end{Bmatrix} = \begin{Bmatrix} 3 \\ 1 \end{Bmatrix}$$

בדיקה:

$$\underline{\underline{B}}^{-1}\underline{\underline{B}} = \frac{1}{6}\begin{bmatrix} 4 & -2 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ 1 & 4 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad ✓$$

כלומר, $x_2 = 1$, $x_3 = 3$.

נחשב כעת את שאר הסבלה,

$$\underline{y}_1 = \underline{\underline{B}}^{-1}\,\underline{a}_1 = \frac{1}{6}\begin{pmatrix} 4 & -2 \\ -1 & 2 \end{pmatrix} \cdot \begin{Bmatrix} 3 \\ 3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1/2 \end{Bmatrix}$$

$$\underline{c}_B = (c_3\ ,\ c_2) = (1\ ,\ 2)$$

$$z = \underline{c}_B\,\underline{x}_B = (1\ 2)\begin{Bmatrix} 3 \\ 1 \end{Bmatrix} = 5$$

$$z_1 = \underline{c}_B\,y_1 = (1\ \ 2)\begin{Bmatrix} 1 \\ 1/2 \end{Bmatrix} = 2$$

נבדוק האם כדאי להכניס את x_1 לבסיס:

$$c_1 - z_1 = 3 - 2 = 1$$

היות ונתקבל $(c_1 - z_1 > 0)$ והבעיה היא מקסימיזציה, אפשר לשפר את הפתרון על ידי הכנסת x_1 לבסיס (ראה משוואות 6-46 עד 6-48).

כדיקה עבור הוקטור העוזב את הבסיס, x_r, לפי משוואה (6-36):

$$\frac{x_{B_r}}{y_{r1}} = \min_i \left\{ \frac{x_{B_r}}{y_{ij}} \;\middle|\; y_{ij} > 0 \right\} = \min \left\{ \begin{array}{l} \dfrac{x_{B_1}}{y_{11}} = \dfrac{3}{1} = 3 \\[2ex] \dfrac{x_{B_2}}{y_{21}} = \dfrac{1}{1/2} = 2 \end{array} \right\} = 2 \Longrightarrow r = 2$$

הערך החדש של פונקצית המטרה היהי לפי משוואה (6-46)

$$\hat{z} = z + \frac{x_{B_1}}{y_{r1}}(c_1 - z_1) = 5 + 2(3-2) = 7$$

נבדוק:

המטריצה הבסיסית החדשה וההפכית שלה הן

$$\hat{\underline{\underline{B}}} = \begin{pmatrix} 2 & 3 \\ 1 & 3 \end{pmatrix} \qquad \hat{\underline{\underline{B}}}^{-1} = \frac{1}{3}\begin{pmatrix} 3 & -3 \\ -1 & 2 \end{pmatrix}$$

יש לשים לב שהסדר בבסיס הוא עכשיו $\underline{x}_B = (x_3 \ , \ x_1)$. הפתרון הבסיסי הוא

$$\hat{\underline{x}}_B = \underline{\underline{B}}^{-1}\,\underline{b} = \frac{1}{3}\begin{pmatrix} 3 & -3 \\ -1 & 2 \end{pmatrix} \cdot \begin{Bmatrix} 8 \\ 7 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}$$

כלומר: $(x_1 = 2, x_3 = 1)$.

ערך פונקצית המטרה הוא

$$\hat{z} = \hat{\underline{c}}_B\,\hat{\underline{x}}_B = (1 \quad 3)\begin{Bmatrix} 1 \\ 2 \end{Bmatrix} = 7$$

כפי שקבלנו קודם בעזרת משוואה (6-46).

כדי לעבור לאיטרציה הבאה נסיר את הסימון (^) מעל הערכים החדשים שעכשיו הופכים לערכים הנוכחיים לאיטרציה הבאה, כלומר $\underline{x}_B = (x_3 \ , \ x_1)$

התהליך נותן:

$$\underline{y}_2 = \underline{\underline{B}}^{-1}\,\underline{a}_2 = \frac{1}{3}\begin{pmatrix} 3 & -3 \\ -1 & 2 \end{pmatrix} \cdot \begin{Bmatrix} 2 \\ 4 \end{Bmatrix} = \begin{Bmatrix} -2 \\ \cancel{4}\ 2 \end{Bmatrix}$$

$$z_2 = \underline{c}_B\,\underline{y}_2 = (1 \quad 3) \cdot \begin{Bmatrix} -2 \\ 2 \end{Bmatrix} = 4$$

$$c_2 - z_2 = 2 - 4 = -2 < 0$$

הפתרון הנוכחי הוא לכן אופטימלי.

דוגמה לפתרון לא חסום

נשנה את הסימנים בוקטור $\underline{a}_1$ של הדוגמה הקודמת,

$$\max \quad z = 3x_1 + 2x_2 + x_3$$

$$-3x_1 + 2x_2 + 2x_3 = 8$$

$$-3x_1 + 4x_2 + x_3 = 7$$

אם הפתרון הבסיסי הנוכחי הוא $\underline{x}_B = (x_3 , x_2)$, כמו שהיה קודם, יתקבל

$$\underline{y}_1 = \underline{B}^{-1} \underline{a}_1 = \frac{1}{6} \begin{pmatrix} 4 & -2 \\ -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} -3 \\ -3 \end{pmatrix} = \begin{Bmatrix} -1 \\ -1/2 \end{Bmatrix}$$

$$z_1 = \underline{c}_B \underline{y}_1 = (1 \quad 2) \cdot \begin{Bmatrix} -1 \\ -1/2 \end{Bmatrix} = -2$$

$$c_1 - z_1 = 3 - (-2) = 5 > 0$$

המסקנה היא שניתן לשפר. הפתרון בלתי חסום כי כל ה-y_{i1} שליליים.

6.10 משתנים מלאכותיים. חלוקת הפתרון לשני שלבים

כדי להתחיל את תהליך הפתרון, רצוי שמטריצת הבסיס הראשון תהיה מטריצת יחידה. כאשר זה קיים נתונה מיד ההפכית של המטריצה הבסיסית, ובן $\underline{x}_B = \underline{b}$. כאשר כל האילוצים בבעיה הם אי-שוויונות בצורה (≤), קיים הדבר לאחר הוספת משתני החסר. כאשר חלק מהאילוצים הם שוויונות או אי-שוויונות בעלי הצורה (≥) אין הדבר מתקיים. כדי שבכל זאת נוכל להתחיל עם מטריצת יחידה, מוסיפים משתנים נוספים לבעיה, הנקראים משתנים מלאכותיים (Artificial Variables). למשל, אם אילוץ הוא בעל הצורה,

$$\sum_{j=1}^{r} a_{kj} x_j \geq b_k$$

נהפוך אותו לשוויון לפי משוואה (6-6) על ידי חיסור משתנה עודף

$$\sum_{j=1}^{r} a_{kj}\, x_j - x_{r+k} = b_k \qquad (6\text{-}6)$$

אלא שכעת אין בשורה ה-k איבר (+1) שבעמודה שלו כל האיברים האחרים אפסים, כפי שנדרש עבור מטריצת היחידה. משוואה (6-9) מתארת את המצב, כאשר בשורות (u+1) עד m, אלה המתאימות לאילוצי שוויון ואי-שוויונות בעלי הצורה (≥), אין איברים עבור מטריצת יחידה. לכן נוסיף כעת משתנים מלאכותיים ונקבל בצורה חדשה למשוואות (6-7)

$$\left.\begin{array}{lll}
\sum_{j=1}^{r} a_{hj}\, x_j + x_{r+h} = b_h & & h = 1,\dots,u \\
\sum_{j=1}^{r} a_{kj}\, x_j - x_{r+k} + x_{a,\ell} = b_k & & \begin{array}{l} k = (u+1),\dots,v \\ \ell = 1,\dots,(v-u) \end{array} \\
\sum_{j=1}^{r} a_{pj}\, x_j + x_{a,q} = b_p & & \begin{array}{l} p = (v+1),\dots,m \\ q = (v-u+1),\dots,(m-u) \end{array}
\end{array}\right\} \qquad (6\text{-}55)$$

המשתנים $x_{a,1}$ עד $x_{a(m-u)}$ הם המשתנים המלאכותיים, ואנו דורשים

$$x_j \geq 0 \;\; ; \;\; j = 1,\dots,m \;\; ; \;\; x_{a,j} \geq 0 \;\; ; \;\; j = 1,\dots,(m-u)$$

יש לנו כעת מטריצת יחידה (m×m) ואפשר להתחיל את תהליך הפתרון בצורה נוחה. בעית האיטרציה הראשונה נפתרה, אלא שנוצרה בעיה חדשה. אם לא נדרוש אלא $x_{a,j} \geq 0$ בנוסף ל-$x_j \geq 0$ יכול להתקבל פתרון לא-אפשרי. למשל, אם יתקבל בפתרון הסופי של התהליך, זה שהוא כביכול אופטימלי, כי באילוץ ה-k קיים

$$x_{r+k} \leq x_{a,\ell} \Longrightarrow (x_{r+k} - x_{a,\ell}) < 0$$

הרי שקיים

$$\sum_{j=1}^{r} a_{kj}\, x_j = b_k + (x_{r+k} - x_{a,\ell}) \leq b_k$$

כלומר, האילוץ ה-k אינו מתקיים, שכן נדרש בו (≥), והפתרון לכן אינו אפשרי, היות ולא כל האילוצים מתקיימים. הפתרון לבעיה זו הוא לדרוש שעל מנת שהפתרון הבסיסי יהיה אפשרי, לא יופיעו בו משתנים מלאכותיים, כלומר, כל המשתנים המלאכותיים חייבים להתאפס. יש שתי וריאציות לשיטה בה נוקטים:

א. לתת קנס גבוה על משתנים מלאכותיים בפונקצית המטרה. בבעית מקסימיזציה נותנים למשתנים אלה מקדם שלילי, הגבוה בכמה סדרי גודל מכל המקדמים האחרים.

ב. חלוקת התהליך לשני שלבים: בראשון שואפים לאפס את סכום המשתנים המלאכותיים ובשני פועלים על פונקצית המטרה המקורית.

6.10.1 שיטת הקנסות

את משוואות (6-7), לפני הוספת המשתנים המלאכותיים נרשום בצורה

$$min \quad z = \underline{c}\,\underline{x} \tag{6-56}$$

$$s.t. \quad \underline{\underline{A}}\,\underline{x} = \underline{b} \tag{6-57}$$

נוסיף כעת את המשתנים המלאכותיים הדרושים ונקבל את האילוצים

$$\underline{\underline{A}}\,\underline{x} + \underline{\underline{M}}\,\underline{x}_a = \underline{b} \tag{6-58}$$

כאשר $\underline{\underline{M}}$ היא מטריצה (m×m), שבה על האלכסון מופיע 1 בשורות (u+1) ועד m, וכל שאר האיברים (כולל על האלכסון בשורות 1 עד u) אפסים. פונקצית המטרה החדשה תהיה

$$min \quad z = \underline{c}\,\underline{x} + \sum_j K\, x_{a,j} \tag{6-59}$$

כאשר K מספר גדול מאד, למשל, פי 10^3 מהמספר הגדול ביותר ב-$\underline{c}$. הפתרון יתחיל עם מטריצה המכילה משתנים מלאכותיים והם יעזבו אחד אחד את הבסיס שכן יש קנס כבד על המצאותם בו. היות ובחרנו K גדול בהשוואה לכל ה-c_j אנו בטוחים שוקטור של משתנה מלאכותי לא יחזור לבסיס.

ומה קורה אם במהלך האיטרציות מתקבל שהפתרון הבסיסי האופטימלי (כביכול) עדיין מכיל משתנים מלאכותיים, אחד או יותר? התשובה ברורה: אין פתרון אפשרי לבעיה. מצאנו אם כן, דרך בה ניתן תמיד להתחיל את האיטרציות, בלי לכרר לפני כן האם בכלל קיים פתרון אפשרי. (אגב, אם כל האילוצים בבעית המינימיזציה הם רק אי-שוויונות מהצורה $\leq$ ברור שיש פתרון אפשרי אחד, והוא הטריביאלי $\underline{x} = \underline{0}$). אם הפתרון האופטימלי של הניסוח המורחב מכיל משתנים מלאכותיים, המסקנה היא שאין פתרון אפשרי. אם באיטרציה כלשהי יוצא אחרון המשתנים המלאכותיים מהבסיס, הרי הפתרונות הבסיסיים הבאים הם אפשריים והתהליך נמשך. השיטה מאפשרת לנו לכן לגלות מתי אין פתרון אפשרי, וכן מובילה אותנו מפתרון התחלתי נוח, כזה שיש בו מטריצת יחידה מוכנה מראש, עד שכאשר מתקבל הפתרון האפשרי הראשון, יש לנו עבורו מטריצת היחידה.

<u>דוגמא</u>

$$
\begin{aligned}
max \quad & z = 3x_1 + 2x_2 + x_3 + 5x_4 \\
s.t. \quad & 3x_1 + 4x_2 + 5x_3 + 6x_4 \leq 5 \\
& 2x_1 + 6x_2 + x_3 + 5x_4 \geq 6 \\
& x_1 + x_2 + 5x_3 + x_4 = 7 \\
& x_j \geq 0 \quad , \quad j = 1,2,3,4
\end{aligned}
$$

אחרי הוספת משתני חסר ועודף מתקבלות המשוואות

$$
\begin{bmatrix} 3 & 4 & 5 & 6 & 1 & 0 \\ 2 & 6 & 1 & 5 & 0 & -1 \\ 1 & 1 & 5 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{Bmatrix} = \begin{Bmatrix} 5 \\ 6 \\ 7 \end{Bmatrix}
$$

חסרים שני משתנים מלאכותיים. לאחר הוספתם מתקבל,

$$\begin{bmatrix} 3 & 4 & 5 & 6 & 1 & 0 & 0 & 0 \\ 2 & 6 & 1 & 5 & 0 & -1 & 1 & 0 \\ 1 & 1 & 5 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \{\underline{x}\} = \begin{Bmatrix} 5 \\ 6 \\ 7 \end{Bmatrix}$$

כאשר,

$$\underline{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_{a,1}, x_{a,2}) \geq \underline{0}$$

פונקצית המטרה היא,

$$max \quad z = 3x_1 + 2x_2 + x_3 + 5x_4 - Kx_{a,1} - Kx_{a,2}$$

ברור שכאן חסרנו מפונקצית המטרה את תרומת המשתנים המלאכותיים, שכן אנו עוסקים במקסימיזציה, בניגוד למשוואה (6-59). עכור פתרון ידני אין צורך להציב ערך מספרי עבור K.

הפתרון הבסיסי הראשון הוא:

$$\underline{x}_B = (x_5, x_{a,1}, x_{a,2}) = (5,6,7)$$

ובהתאמה

$$\underline{c}_B = (0, -K, -K)$$

בשלב זה, שהוא ההתחלתי, קיים

$$\underline{y}_j = \underline{\underline{B}}_1^{-1} \underline{a}_j = \underline{\underline{I}} \underline{a}_j = \underline{a}_j$$

כעת לפי

$$c_j - z_j = c_j - \underline{c}_B \underline{y}_j$$

מקבלים

$$c_1 - z_1 = 3 + 3K$$

$$c_2 - z_2 = 2 + 7K$$

$$c_3 - z_3 = 1 + 6K$$

$$c_4 - z_4 = 5 + 6K$$

$$c_6 - z_6 = -K$$

נכחר להכניס לבסיס את a_2 שכן בו המקדם של K הוא הגדול ביותר. בחירת הוקטור העוזב

$$\frac{x_{B_r}}{y_{rj}} = \min_i \left\{ \frac{x_{B_i}}{y_{i2}} \,\middle|\, y_{12} > 0 \right\} = \min \left\{ \begin{array}{l} \frac{x_{B_1}}{y_{12}} = \frac{5}{4} \\ \frac{x_{B_2}}{y_{22}} = \frac{6}{6} \\ \frac{x_{B_3}}{y_{32}} = \frac{7}{1} \end{array} \right\} = 1 \Longrightarrow r = 2$$

עוזב הוקטור השני בבסיס, שהוא $x_{a,1}$.

<u>איטרציה שניה</u>

$$\underline{x}_B = (x_5, x_2, x_{a,2})$$

$$\underline{c}_B = (0,\ 2,\ -K)$$

נחשב את ההפכית החדשה בשיטת המכפלה

$$\underline{y}_2 = \underline{\underline{B}}_1^{-1}\ \underline{a}_2 = \underline{a}_2 = \begin{Bmatrix} 4 \\ 6 \\ 1 \end{Bmatrix} \qquad \underline{\eta}_2 = \begin{Bmatrix} -4/6 \\ 1/6 \\ -1/6 \end{Bmatrix}$$

$$\underline{\underline{B}}^{-1} = \underline{\underline{E}}_1\ \underline{\underline{B}}^{-1} = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix} \quad ; \quad \underline{x}_B = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix} \cdot \begin{Bmatrix} 5 \\ 6 \\ 7 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \\ 6 \end{Bmatrix}$$

$$\underline{y}_1 = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix} \begin{Bmatrix} 3 \\ 2 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 5/3 \\ 1/3 \\ 2/3 \end{Bmatrix}$$

$$c_1 - z_1 = c_1 - \underline{c}_B \, \underline{y}_1 = 3 - (0,2,-K) \cdot \begin{Bmatrix} 5/3 \\ 1/3 \\ 2/3 \end{Bmatrix} = 7/3 + 2K/3$$

$$\underline{y}_3 = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix} \begin{Bmatrix} 5 \\ 1 \\ 5 \end{Bmatrix} = \begin{Bmatrix} 13/3 \\ 1/6 \\ 29/6 \end{Bmatrix}$$

$$c_3 - z_3 = c_3 - \underline{c}_B \, \underline{y}_3 = 1 - (0,2,-K) \begin{Bmatrix} 13/3 \\ 1/6 \\ 29/6 \end{Bmatrix} = 2/3 + 29K/6$$

$$\underline{y}_4 = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix} \cdot \begin{Bmatrix} 6 \\ 5 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 8/3 \\ 5/6 \\ 1/6 \end{Bmatrix}$$

$$c_4 - z_4 = c_4 - \underline{c}_B \, \underline{y}_4 = 5 - (0,2,-K) \cdot \begin{Bmatrix} 8/3 \\ 5/6 \\ 1/6 \end{Bmatrix} = 10/3 + K/6$$

$$\underline{y}_6 = \underline{\underline{B}}_1^{-1} \, \underline{a}_6 = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix} \cdot \begin{Bmatrix} 0 \\ -1 \\ 0 \end{Bmatrix} = \begin{Bmatrix} 2/3 \\ -1/6 \\ 1/6 \end{Bmatrix}$$

$$c_6 - z_6 = c_6 - \underline{c}_B\, y_6 = 0 - (0,2,-K)\begin{Bmatrix} 2/3 \\ -1/6 \\ 1/6 \end{Bmatrix} = 1/3 + K/6$$

$$\underline{y}_{a,1} = \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix}\begin{Bmatrix} 0 \\ 1 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -2/3 \\ 1/6 \\ -1/6 \end{Bmatrix}$$

$$c_{a,1} - z_{a,1} = c_{a,1} - \underline{c}_B\, \underline{y}_{a,1} = -K - (0,2,-K)\cdot\begin{Bmatrix} -2/3 \\ 1/6 \\ -1/6 \end{Bmatrix} = -1/3 - 7K/6$$

נכנס לבסיס x_3, שעבודו המקדם של K בביטוי $(c_j - z_j)$ הוא הגדול ביותר מבין כל המשתנים הלא בסיסיים.

בחירת הוקטור העוזב תיעשה לפי:

$$\frac{x_{B_r}}{y_{r,3}} = \min_i\left\{\frac{x_{B_i}}{y_{i,3}} \,\middle|\, y_{i,3} > 0\right\} = \min\left\{\begin{array}{l} \dfrac{x_{B_1}}{y_{1,3}} = \dfrac{1}{13/3} = \dfrac{1}{13} \\[2ex] \dfrac{x_{B_2}}{y_{2,3}} = \dfrac{1}{1/6} = 6 \\[2ex] \dfrac{x_{B_3}}{y_{3,3}} = \dfrac{6}{29/6} = \dfrac{36}{29} \end{array}\right\} = \frac{3}{13} \Longrightarrow r = 1$$

כלומר, עוזב המשתנה הראשון בבסיס, שהוא x_5.

<u>איטרציה שלישית</u>

$\underline{x}_B = (x_3, x_2, x_{a,2})$ $\qquad$ $\underline{c}_B = (1,2,-K)$

נחשב את ההפכית החדשה על ידי מכפלה, כדלקמן:

$$\underline{y}_3 = \begin{Bmatrix} 13/3 \\ 1/6 \\ 29/6 \end{Bmatrix} \qquad \underline{\eta}_3 = \begin{Bmatrix} 3/13 \\ -1/26 \\ -29/26 \end{Bmatrix}$$

$$\underline{\underline{B}}_3^{-1} = \underline{\underline{E}}\ \underline{\underline{B}}_2^{-1} = \begin{bmatrix} 3/13 & 0 & 0 \\ -1/26 & 1 & 0 \\ -29/26 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2/3 & 0 \\ 0 & 1/6 & 0 \\ 0 & -1/6 & 1 \end{bmatrix} = \begin{bmatrix} 3/13 & -2/13 & 0 \\ -1/26 & 5/26 & 0 \\ -29/26 & 15/26 & 1 \end{bmatrix}$$

הפעם נחשב את $\underline{x}_B$ ואת $\underline{y}_j$ בבת אחת, כדלקמן:

$$[\underline{x}_B, \underline{y}_1, \underline{y}_4, \underline{y}_5, \underline{y}_6, \underline{y}_{a,1}] = \underline{\underline{B}}_3^{-1} [\underline{b}, \underline{a}_1, \underline{a}_4, \underline{a}_5, \underline{a}_6, \underline{a}_{a,1}] =$$

$$= \begin{bmatrix} 6/26 & -4/26 & 0 \\ -1/26 & 5/26 & 0 \\ -29/26 & 15/26 & 1 \end{bmatrix} \begin{bmatrix} 5 & 3 & 6 & 1 & 0 & 0 \\ 6 & 2 & 5 & 0 & -1 & 1 \\ 7 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} =$$

$$= \frac{1}{26} \begin{bmatrix} 6 & 10 & 16 & 6 & 4 & -4 \\ 25 & 7 & 19 & -1 & -5 & 5 \\ 127 & -31 & -73 & -29 & -15 & 15 \end{bmatrix}$$

וכעת את $z_j = \underline{c}_B \underline{y}_j$ בבת אחת

$$(z_1, z_4, z_5, z_6, z_{a,1}) = \underline{c}_B [\underline{y}_1, \underline{y}_4, \underline{y}_5, \underline{y}_6, \underline{y}_{a,1}] =$$

$$= (1, 2, -K) \begin{bmatrix} 10/26 & 16/26 & 6/26 & 4/26 & -4/26 \\ 7/26 & 19/26 & -1/26 & -5/26 & 5/26 \\ -31/26 & -73/26 & -29/26 & -15/26 & 15/26 \end{bmatrix} =$$

$$= [(24/26+31K/26) \;;\; (54/26+73K/26) \;;\; (4/26+29K/26) \;;$$

$$(-6/26+15K/26) \;;\; (6/26-15K/26)]$$

ולבסוף את $c_j - z_j$

$$[(c_1 - z_1) \;;\; (c_4 - z_4) \;;\; (c_5 - z_5) \;;\; (c_6 - z_6) \;;\; (c_{a,1} - z_{a,1})] =$$

$$= [(3 - 24/26 - 31K/26) \;;\; (5 - 54/26 - 73K/26) \;;\; (0 - 4/26 - 29K/26) \;;$$

$$(0 + 6/26 - 15K/26) \;;\; (-K - 6/26 + 15K/26)] =$$

$$= [(54/26 - 31K/26) \;;\; (76/26 - 73K/26) \;;\; (-4/26 - 29K/26) \;;$$

$$(6/26 - 15K/26) \;;\; (-6/26 - 11K/26)]$$

הסתכלות בתוצאה זו מראה שהגענו כביכול לאופטימום, שכן כל המקדמים של K שליליים, ובבעיית מקסימיזציה זה מורה על המקסימום. המסקנה היא שאין לבעיה פתרון אפשרי, שכן בפתרון האחרון עדיין יש משתנה מלאכותי, והוא $x_{a,2}$. אגב, אפשר לראות שאין פתרון אפשרי מחיסור האילוץ השלישי מהראשון; מתקבל,

$$2x_1 + 3x_2 + 5x_4 \le -2$$

היות ודרוש $x_j \ge 0$ אי אפשר לקיים אילוץ חדש זה.

6.10.2 חלוקה לשני שלבים

באופן עקרוני אין הבדל בין שיטת הקנסות לבין השיטה שתתואר להלן, בה מחלקים את הפתרון לשני שלבים. בעבורה ידנית אפשר להשתמש בשתיהן. תוכניות מחשב משתמשות בשיטה השניה. לפי שיטה זו רושמים פונקציית מטרה נוספת למקורית, שהיא מינימיזציה של סכום המשתנים המלאכותיים. בשלב הראשון פועלים על פי פונקציית מטרה זו, ורק כאשר מגיעים למינימום שלה עוברים לפעול לפי פונקציית המטרה המקורית.

<u>דוגמה</u>: נחזור לדוגמה של סעיף 6.4 ונראה כיצד משמשים המשתנים המלאכותיים להתחלת הפתרון בצורה מסודרת ללא צורך לבחור בסיס התחלתי:

$$min \quad z = x_1 + 6x_2 - 7x_3 + x_4 + 5x_5$$

$$s.t. \quad 5x_1 - 4x_2 + 13x_3 - 2x_4 + x_5 = 20$$

$$x_1 - x_2 + 5x_3 - x_4 + x_5 = 8$$

כדי להתחיל את הפתרון נוסיף משתנה מלאכותי בכל אחד מהאילוצים. כמו כן נוסיף פונקצית מטרה חדשה: מינימום סכום המשתנים המלאכותיים

$$min \quad z = x_1 + 6x_2 - 7x_3 + x_4 + 5x_5$$

$$min \quad w = x_6 + x_7$$

$$s.t. \quad 5x_1 - 4x_2 + 13x_3 - 2x_4 + x_5 + x_6 = 20$$

$$x_1 - x_2 + 5x_3 - x_4 + x_5 + x_7 = 8$$

כדי להתחיל את הפתרון יש לאפס את המקדמים של x_6 ו-x_7 בפונקציה w, שכן מתחילים עם עמודות של מטריצת יחידה במקומות אלה. התוצאה היא,

$$min \quad w-28 = -6x_1 + 5x_2 - 18x_3 + 3x_4 - 2x_5$$

אותה מקבלים על ידי הצבת ערכי x_6 ו-x_7 בפונקציה w, מבוטאים בעזרת המשתנים המקוריים, שאינם בבסיס הראשון. הבסיס הראשון הוא לכן (x_6,x_7), הערך של פונקצית המטרה החדשה הוא w=28, ואילו של המקורית הוא z=0. לפונקצית המטרה החדשה אפשר לקרוא "מידת אי-האפשריות" של הפתרון הנוכחי (The Infeasibility Form).

המשך הפתרון ייעשה בעזרת הטבלאות כדלקמן:

Tableau 0

בסיס	ממשיים					מלאכותיים		
$\underline{x}_B$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$\underline{b}$
x_6	5	-4	13	-2	1	1	0	20
x_7	1	-1	5	-1	1	0	1	8
z	1	6	-7	1	5	0	0	0
w	-6	5	-18	3	-2	0	0	-28

Tableau 1

$\underline{x}_B$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$\underline{b}$
x_3	5/13	-4/13	1	-2/13	1/13	1/13	0	20/13
x_7	-12/13	7/13	0	-3/13	8/13	-5/13	1	4/13
z	48/13	50/13	0	-1/13	72/13	7/13	0	140/13
w	12/13	-7/13	0	3/13	-8/13	18/13	0	-4/13

Tableau 2

$\underline{x}_B$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$\underline{b}$
x_3	1/2	-3/8	1	-1/8	0	1/8	-1/8	3/2
x_5	-12/8	7/8	0	-3/8	1	-5/8	13/8	4/8
z	12	-1	0	2	0	4	-9	8
w	•	•	•	•	•	1	1	0

נשלם השלב הראשון (Phase I), כי $w = w_{min} = 0$.

כל המשתנים המלאכותיים יצאו מהבסיס, ונתקבל פתרון בסיסי ראשון של הבעיה המקורית. כעת עוברים לשלב השני (Phase II), בו פועלים על פונקצית המטרה המקורית, z. עד עתה עברה z את אותן טרנספורמציות כמו שאר שורות המטריצה אבל בצורה בלתי פעילה: לא נבחנו בה ערכי המקדמים כדי להחליט על הוקטור הנכנס לבסיס ולא נערכה בה הבריקה למציאת הוקטור היוצא. בשלב השני אפשר לוותר על אותו חלק של הטבלה המשמש לעמורות של המשתנים המלאכותיים ולשורה של w (ועם זאת נכלול כאן את המקדמים בשורת z של המשתנים המלאכותיים, אשר יידרשו לנו בסעיף מאוחר יותר). הטבלו הבא הוא:

Tableau 3

$\underline{x}_B$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$\underline{b}$
x_3	-1/7	0	1	-2/7	3/7	·	·	12/7
x_2	-12/7	1	0	-3/7	8/7	·	·	4/7
z	72/7	0	0	11/7	8/7	23/7	-50/7	60/7

הפתרון האופטימלי הוא אם כן $z^* + 60/7 = 0$ או $z^* = -60/7$. נשים לב כי איננו מתייחסים כלל למקדמים של המשתנים המלאכותיים בשורת z בבדיקת האופטימליות.

<u>דוגמה:</u>

במפעל מייצרים ארבעה מוצרים. בייצור של כל אחד מבצעים שלוש פעולות. מספר השעות של כל פעולה הדרושות לייצור יחידה של כל מוצר והרווח נטו ממכירת יחידה נתונות בטבלה הבאה:

	רווח נטו ליחידה	שעות דרושות לייצור יחידה אחת של המוצר		
		פעולה 1	פעולה 2	פעולה 3
מוצר א'	12	5	2	3
מוצר ב'	5	1	3	2
מוצר ג'	15	9	4	5
מוצר ד'	10	12	1	10

לרשות המפעל עומד מספר מוגבל של שעות עבודה בכל פעולה: 1500 שעות בחודש של פעולה מס' 1, 1000 שעות בחודש של פעולה 2 ו-800 שעות בחודש של פעולה 3. המפעל גם מתחייב מראש לייצר בחודש לפחות 40 יחידות של מוצר א', 130 יחידות מוצר ב', 30 יחידות של מוצר ג' ו-10 יחידות של מוצר ד'.

מהי תוכנית הייצור החודשית האופטימלית?

נסמן ב-x_j את מספר היחידות של המוצר ה-j שיש לייצר בחודש. הבעיה היא:

$$
\begin{array}{llrcl}
max & z = 12x_1 + 5x_2 + 15x_3 + 10x_4 & & & \\
s.t. & 5x_1 + x_2 + 9x_3 + 12x_4 & \leq & 1500 \\
 & 2x_1 + 3x_2 + 4x_3 + x_4 & \leq & 1000 \\
 & 3x_1 + 2x_2 + 5x_3 + 10x_4 & \leq & 800 \\
 & x_1 & \geq & 40 \\
 & x_2 & \geq & 130 \\
 & x_3 & \geq & 30 \\
 & x_4 & \geq & 10 \\
 & x_j & \geq & 0
\end{array}
$$

בשלושת האילוצים הראשונים נוסיף משתני חסר, בארבעת הנותרים משתני עודף ומשתנים מלאכותיים. נרשום גם את w (סכום המשתנים המלאכותיים) בעזרת המשתנים האחרים. בצורה טבולרית מתקבל

Tableau 0

$\underline{x}_B$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\underline{b}$
x_5	5	1	9	12	1											1500
x_6	2	3	4	1		1										1000
x_7	3	2	5	10			1									800
x_{12}	1							-1				1				40
x_{13}		1							-1				1			130
x_{14}			1							-1				1		30
x_{15}				1							-1				1	10
z	12	5	15	10												0
w	-1	-1	-1	-1				1	1	1	1					-210

Tableau 1

$\underline{x}_B$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\underline{b}$
x_5		1	9	12	1			5				-5				1300
x_6		3	4	1		1		2				-2				920
x_7		2	5	10			1	3				-3				680
x_1	1							-1				1				40
x_{13}		1							-1				1			130
x_{14}			1							-1				1		30
x_{15}				1							-1				1	10
z		5	15	10				12				-12				480
w		-1	-1	-1					1	1	1	1	~~1~~			-170

X

Tableau 2

$\underline{x}_B$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\underline{b}$
x_5			9	12	1			5	1			-5	-1			1170
x_6			4	1		1		2	3			-2	-3			530
x_7			5	10			1	3	2			-3	-2			420
x_1	1							-1				1				40
x_2		1							-1				1			130
x_{14}			1							-1				1		30
x_{15}				1							-1				1	10
z			15	10				12	5			-12	-5			1130
w			-1	-1						1	1	1	1			-40

Tableau 3

$\underline{x}_B$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\underline{b}$
x_5				12	1			5	1	9		-5	-1	-9		900
x_6				1		1		2	3	4		-2	-3	-4		410
x_7				10			1	3	2	5		-3	-2	-5		270
x_1	1							-1				1				40
x_2		1							-1				1			130
x_3			1							-1				1		30
x_{15}				1							-1				1	10
z				10				12	5	15		-12	-5	-15		1580
w				-1							1	1	1	1		-10

Tableau 4

$\underline{x}_B$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\underline{b}$
x_5					1			5	1	9	12	-5	-1	-9	-12	780
x_6						1		2	3	4	1	-2	-3	-4	-1	400
x_7							1	3	2	5	10	-3	-2	-5	-10	170
x_1	1							-1				1				40
x_2		1							-1				1			130
x_3			1							-1				1		30
x_4				1							-1				1	10
z								12	5	15	10	-12	-5	-15	-10	1680
w												1	1	1	1	0

-2/5

Tableau 5

$\underline{x}_B$	1	2	3	4	5	6	7	8	9	10	11	$\underline{b}$
x_5					1		-9/5	-2/5	-13/5		-6	474
x_6						1	-4/5	2/5	7/5		-7	264
x_{10}							1/5	3/5	2/5	1	2	34
x_1	1							-1				40
x_2		1							-1			130
x_3			1				1/5	3/5	2/5		2	64
x_4				1							-1	10
z							-3	3	1		-20	2190

-2

-1

Tableau 6

$\underline{x}_B$	1	2	3	4	5	6	7	8	9	10	11	$\underline{b}$
x_5					1		-5/3		-7/9	2/3	-14/3	1490/3
x_6						1	-14/15		17/15	-5/3	-25/3	860/3
x_8							1/3	1	2/3	5/3	10/3	170/3
x_1	1						1/3		2/3	5/3	10/3	290/3
x_2		1							-1			130
x_3			1							-1		30
x_4				1							-1	10
z							-4		-3	-5	-30	2360

הפתרון האופטימלי הוא:

$$z^* = 2360$$

$$x_1^* = 290/3 = 96\frac{2}{3}$$

$$x_2^* = 130$$

$$x_3^* = 30$$

$$x_4^* = 10$$

בתוכנית אופטימלית זו מייצרים $x_8^* = 56\frac{2}{3}$ יחידות של מוצר א' מעל למינימום הנדרש לפי ההתחייבות. לפי תוכנית ייצור זו מנצלים את כל שעות העבודה של פעולה 3 (800 שעות) ואילו $x_5^* = 496\frac{2}{3}$ שעות של פעולה 1 ו- $x_6^* = 286\frac{2}{3}$ שעות של פעולה 2 אינן סנוצלות. לאחר חישוב Tableau 4 מסתבר שהגענו לפתרון אפשרי. שכן w=0. כבר בטבלה זו לא היה צורך לחשב את העמודות הסתאיסות למשתנים המלאכותיים, והרבר נעשה רק לשם שלימות ההצגה. כשלב זה עוברים לבחור את הוקטור הנכנס לבסיס לפי המקדמים בשורה של z, בעוד שבשלבים הקורמים הבחירה היתה לפי הסקדמים של w. הבעיה היא מקסימיזציה ב-z, ולכן בוחרים את המקרם החיובי הגדול ביותר בכל איטרציה, והתהליך מסתיים כאשר כל המקדמים שליליים.

מן הראוי לציין כי הרשינו למשתני ההחלטה לקבל ערכים שאינם שלימים, למרות שאין משמעות לחלקי יחידה. זה באילו הנחנו כי הייצור רציף ונמשך חדשים רבים, ולמשל במשך שלושה חורשים יש לייצר 290 יחידות של מוצר 1, שזהו מספר שלם. יכולנו גם לטעון כי בכעיות מעשיות ניתן ל"עגל" קצת את הפתרון, כאילו שקבלנו $x_1^* = 97$.

6.11 הבעיה הדואלית ומשמעותה

6.11.1 ניסוח סימטרי של הבעיה הדואלית

נתונה בעית התכנות הליניארי

$$max \quad z = \sum_{j=1}^{n} c_j x_j \qquad (6\text{-}60)$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij}\, x_j \le b_i \qquad i = 1,\dots,m \tag{6-61}$$

$$x_j \ge 0 \qquad j = 1,\dots,n \tag{6-62}$$

זו תיקרא <u>הבעיה הראשונית</u> (Primal). נשים לב כי כל האילוצים מופיעים כאי-שוויונות בסימן ($\le$). תמיד אפשר לכתוב אילוצים בצורה זו, אם אין דורשים $b_i > 0$. גם אילוצי שוויון אפשר להביא בצורה זו, על ידי כתיבת שני אילוצים. למשל אם נתון האילוץ

$$\sum_{j=1}^{n} a_{kj}\, x_j = b_k$$

אפשר במקומו לרשום שני אילוצים

$$\sum_{j=1}^{n} a_{kj}\, x_j \le b_k$$

$$\sum_{j=1}^{n} -a_{kj}\, x_j \le -b_k$$

לבעיה הראשונית צמודה בעית תכנות ליניארי, הנקראת <u>הדואלית</u> (Dual) בה המשתנים הם π_i, וצורתה:

$$min \quad v = \sum_{i=1}^{m} b_i\, \pi_i \tag{6-63}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} a_{ij}\, \pi_i \ge c_j \qquad j = 1,\dots,n \tag{6-64}$$

$$\pi_i \ge 0 \tag{6-65}$$

המקדמים a_{ij}, b_i ו-c_j הם אותם כמו בבעיה הראשונית.

<u>דוגמא</u>

לבעיה הראשונית

$$\begin{aligned} max \quad z &= x_1 + 2x_2 \\ s.t. \quad x_1 + x_2 &\le 6 \\ 2x_1 + x_2 &\le 10 \\ -2x_1 + x_2 &\le 1 \\ x_1 \,,\, x_2 &\ge 0 \end{aligned}$$

צמודה הבעיה הדואלית

$$\begin{aligned} min \quad v &= 6\pi_1 + 10\pi_2 + \pi_3 \\ s.t. \quad \pi_1 + 2\pi_2 - 2\pi_3 &\ge 1 \\ \pi_1 + \pi_2 + \pi_3 &\ge 2 \\ \pi_1 \,,\, \pi_2 \,,\, \pi_3 &\ge 0 \end{aligned}$$

6.11.2 <u>הקשרים בין פתרונות הראשונית והדואלית</u>

אפשר להראות כי:

1. $z^* = v^*$, כלומר המקסימום של z שווה למינימום של v.

2. מתוך פתרון אחת הבעיות אפשר לקבל את פתרונה השלם של השניה. זה כולל את ערך פונקצית המטרה, לפי (1) לעיל, וכמו כן מתוך הטבלה הסופית של הבעיה הפתורה אפשר לבנות את הטבלה הסופית של הבעיה הצמודה.

נשים לב כי הדואלית של הדואלית היא הראשונית. השמות ראשונית ודואלית ניתנו לבעיות לפי מובנן הכלכלי, כפי שיוסבר להלן, אבל לפעמים הבעיה המוצגת לפתרון מקבלת דווקא את הצורה של משוואות (6-63) עד (6-65). אזי זו הראשונית והדואלית נתונה במשוואות (6-60) עד (6-62).

היות וכמות החישובים בפתרון בעית תכנות ליניארי תלויה בעיקר בגודל המטריצה עליה עובדים, ומימדיה נקבעים לפי מספר האילוצים בבעיה, הרי שרצוי לצמצם את

מספר האילוצים במידת האפשר. בבעיה הדואלית מספר האילוצים שווה למספר המשתנים בבעיה הראשונית, ולהיפך, מספר האילוצים בראשונית שווה למספר המשתנים בדואלית. אפשר לכן לבחור לפתור את זו משתיהן בה מספר האילוצים קטן יותר. נוסף לערך המעשי הזה של השימוש בדואלית יש לה משמעות כלכלית חשובה. לפני שנסביר משמעות זאת, ניתן את משוואות הקשר בין הטכלאות האופטימליות של שתי הבעיות.

את הפתרון האופטימלי של הראשונית נכתוב בצורה

$$x^*_{B_i} = A_i + \sum_j B_{ij}\, x_j \qquad (6\text{-}66)$$

$$z^* = z^* + \sum_j D_j\, x_j \qquad (6\text{-}70)$$

כאשר הסכום הוא על פני המשתנים x_j שאינם בבסיס האופטימלי. הפתרון האופטימלי של הדואליות נתון על ידי הקשרים:

$$\pi^*_{B_k} = E_k + \sum_\ell F_{k\ell}\, \pi_\ell \qquad (6\text{-}68)$$

$$v^* = v^* + \sum_\ell G_\ell\, \pi_\ell \qquad (6\text{-}69)$$

<u>כאשר</u>:

1. יש קשר זוגות כין משתני הראשונית והדואלית, כדלקמן:

   לכל משתנה צמוד משתנה החסר או העודף המופיע בבעיה הצמודה באילוץ המתאים (ראה כדוגמה להלן).

2. x_j אינו בבסיס האופטימלי של הראשונית, ואילו π_k הצמוד לו כן מופיע בבסיס האופטימלי של הדואלית. מאידך, x_i נמצא בבסיס האופטימלי של הראשונית, ואילו π_ℓ הצמוד לו אינו בבסיס האופטימלי של הדואלית.

3. קיימים הקשרים הבאים בין המקדמים:

$$\left.\begin{aligned} E_k &= -D_j \\ F_{k\ell} &= -B_{ij} \\ G_\ell &= A_i \end{aligned}\right\} \qquad (6\text{-}70)$$

כל משתנה דואלי מתאים לאילוץ של הראשונית - לאותו אילוץ אשר בו משתנה החסר/עודף הוא בן זוגו של הדואלי. הדואלי של אילוץ שוויון הוא בן זוגו של המשתנה המלאכותי של אילוץ השוויון.

נחזור לדוגמה. נרשום שתי הבעיות עם משתני חסר ועודף. הראשונית:

$$\begin{aligned} max \quad z &= x_1 + 2x_2 \\ x_1 + x_2 + x_3 &= 6 \\ 2x_1 + x_2 + x_4 &= 10 \\ -2x_1 + x_2 + x_5 &= 1 \\ x_j &\geq 0 \end{aligned}$$

הדואלית:

$$\begin{aligned} min \quad v &= 6\pi_1 + 10\pi_2 + \pi_3 \\ \pi_1 + 2\pi_2 - 2\pi_3 - \pi_4 &= 1 \\ \pi_1 + \pi_2 + \pi_3 - \pi_5 &= 2 \\ \pi_i &\geq 0 \end{aligned}$$

הזוגות הם:

$$(x_1,\pi_4)\ ,\ (x_2,\pi_5)\ ,\ (\pi_1,x_3)\ ,\ (\pi_2,x_4)\ ,\ (\pi_3,x_5)$$

למשל, π_4 הוא משתנה העודף באילוץ הדואלי שנוצר מן העמודה של x_1 בראשונית: x_3 הוא משתנה החסר באילוץ הראשוני שנוצר מן העמודה של π_1 בדואלית, וכדומה. פתרון הראשונית:

Tableau 0

$\underline{x}_B$	1	2	3	4	5	$\underline{b}$
x_3	1	1	1			6
x_4	2	1		1		10
x_5	-2	1			1	1
z	1	2				0

Tableau 1

$\underline{x}_B$	1	2	3	4	5	$\underline{b}$
x_3	3		1		-1	5
x_4	4			1	-1	9
x_2	-2	1			1	1
z	5				-2	-2

Tableau 2

x_B	1	2	3	4	5	$\underline{b}$
x_1	1		1/3		-1/3	5/3
x_4			-4/3	1	1/3	7/3
x_2		1	2/3		1/3	13/3
z			-5/3		-1/3	-31/3

נרשום את הפתרון האופטימלי לפי משוואות (6-66) ו-(6-67).

$$x_1^* = \frac{5}{3} - \frac{1}{3}x_3 + \frac{1}{3}x_5$$

$$x_2^* = \frac{13}{3} - \frac{2}{3}x_3 - \frac{1}{3}x_5$$

$$x_4^* = \frac{7}{3} + \frac{4}{3}x_3 - \frac{1}{3}x_5$$

$$z^* = \frac{31}{3} - \frac{5}{3}x_3 - \frac{1}{3}x_5$$

פתרון גרפי של הבעיה מופיע בציור 6.1.

זוגות המשתנים של הבעיה הראשונית והדואלית הם:

x_{B_i} : x_1 x_2 x_4

π_ℓ : π_4 π_5 π_2

π_{B_k} : π_1 π_3

x_j : x_3 x_5

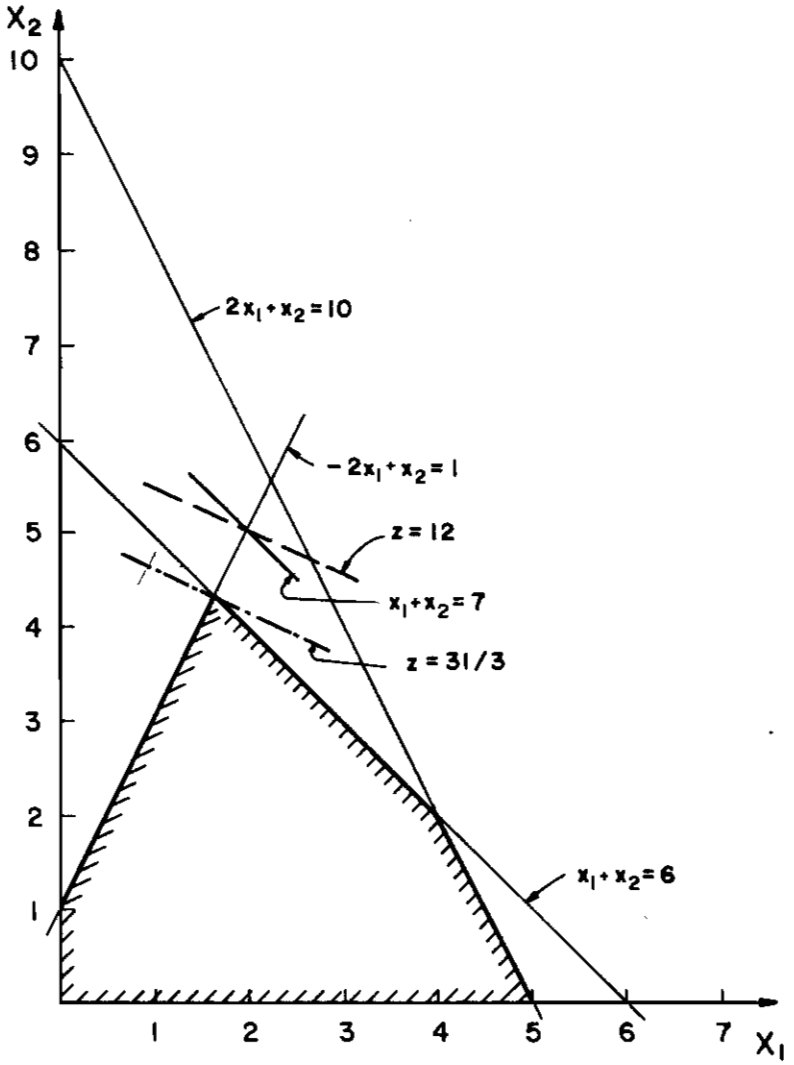
נרשום את פתרון הדואלית לפי משוואות (6-68),(6-69) ו-(6-70)

$$\pi_1^* = \frac{5}{3} + \frac{1}{3}\pi_4 + \frac{2}{3}\pi_5 - \frac{4}{3}\pi_2$$

$$\pi_3^* = \frac{1}{3} - \frac{1}{3}\pi_4 + \frac{1}{3}\pi_5 + \frac{1}{3}\pi_2$$

$$v^* = \frac{31}{3} + \frac{5}{3}\pi_4 + \frac{13}{3}\pi_5 + \frac{7}{3}\pi_2$$

לא הוכחנו את הקשרים כין הפתרונות האופטימליים (ההוכחה נמצאת למשל בפרק 5 בספרו של Gass). ולכן נחשב לדוגמה שלעיל גם את הפתרון האופטימלי של הדואלית ונראה כי הקשרים אמנם מתקיימים.



<u>ציור 6.1</u>: פתרון גרפי של הבעיה הראשונית.

Tableau 0

$\underline{\pi}_B$	1	2	3	4	5	6	7	$\underline{b}$
π_6	1	2	-2	-1		1		1
π_7	1	1	1		-1		1	2
v	6	10	1					0
w	-2	-3	1	1	1			3

Tableau 1

$\underline{\pi}_B$	1	2	3	4	5	6	7	$\underline{b}$
π_2	1/2	1	-1	-1/2		1/2		1/2
π_7	1/2		2	1/2	-1	-1/2	1	3/2
v	1		11	5		-5		-5
w	-1/2		-2	-1/2	1	3/2		3/2

Tableau 2

$\underline{\pi}_B$	1	2	3	4	5	6	7	$\underline{b}$
π_2	3/4	1		-1/4	-1/2	1/4	1/2	5/4
π_3	1/4		1	1/4	-1/2	-1/4	1/2	3/4
v	-7/4			9/4	11/2	-9/4	-11/2	-53/4
w						1	1	0

Tableau 3

$\underline{\pi}_B$	1	2	3	4	5	$\underline{b}$
π_1	1	4/3		-1/3	-2/3	5/3
π_3		-1/3	1	1/3	-1/3	1/3
v		7/3		5/3	13/3	-31/3

הבריקה מראה כי הקשרים בין הפתרונות האופטימליים מתקיימים.

6.11.3 ניסוח כללי של הדואליות

את קשרי הראשונית והדואלית ניתן להגדיר בצורה כללית יותר מזו שניתנה בתחילת הסעיף הנוכחי, על מנת לאפשר אילוצי שוויון ומשתנים שאינם מוגבלים בסימנם (הנקראים גם "משתנים חפשיים" - Free Variables):

$$max \quad z = \sum_{j=1}^{n} c_j x_j \tag{6-60a}$$

$$s.t. \quad \sum_{j=1}^{n} a_{ij} x_j = b_i \qquad i \in E \tag{6-61a}$$

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \qquad i \in \bar{E} \tag{6-61b}$$

$$x_j \ge 0 \qquad j \in P \tag{6-62a}$$

$$x_j \gtrless 0 \qquad j \in \bar{P} \tag{6-62b}$$

צמורה הדואלית

$$min \quad v = \sum_{i=1}^{m} b_i \pi_i \qquad (6\text{-}63a)$$

$$s.t. \quad \sum_{i=1}^{m} a_{ij} \pi_i \geq c_j \qquad j \in P \qquad (6\text{-}64a)$$

$$\sum_{j=1}^{m} a_{ij} \pi_i = c_j \qquad j \in \bar{P} \qquad (6\text{-}64b)$$

$$\pi_i \geq 0 \qquad i \in \bar{E} \qquad (6\text{-}65a)$$

$$\pi_i \gtreqless 0 \qquad i \in E \qquad (6\text{-}65b)$$

E היא קבוצת אילוצי השוויון בראשונית ו-$\bar{E}$ קבוצת אילוצי אי השוויון. P היא קבוצת המשתנים הלא שליליים בראשונית ו-$\bar{P}$ קבוצת המשתנים שאינם מוגבלים בסימנם. לקבוצת אילוצי השוויון בראשונית (E) מתאימים משתנים רואליים שאינם מוגבלים בסימנם; לקבוצת אילוצי השוויון בדואלית ($\bar{P}$) מתאימים משתני הראשונית שאינם מוגבלים בסימנם.

6.11.4 משמעותה של הדואלית

נראה את הראשונית כבעיית ייצור בה כל משוואה מבטאת את השימוש בתשומה של המשאב הנריר שכמותו מוגבלת ל-b_i לייצור התוצרים השונים j ככמויות x_j. המקדמים c_j הם הרווח נטו ליחידת תוצר j, והמקדמים a_{ij} הם כמות התשומה ה-i לייצור יחידת תוצר j. הבעיה הראשונית היא לכן: עבור n תפוקות מיוצרות בעזרת m תשומות, מצא מקסימום של

$$\sum_{j=1}^{n} \left(\frac{\text{רווח נטו מתוצר } j}{\text{תפוקה } j}\right) (\text{תפוקה } j) = (\text{סה"כ רווח נטו}) \qquad (6\text{-}71)$$

בכפיפות לאילוצים

$$\sum_{j=1}^{n} \left(\frac{\text{תשומה } i}{\text{תפוקה } j}\right) (\text{תפוקה } j) \le (\text{סה"כ תשומה } i) \qquad i=1,\dots,m \tag{6-72}$$

$$(\text{תפוקה } j) \ge 0 \tag{6-73}$$

הבעיה הדואלית היא: מצא מינימום של

$$v = \sum_{i=1}^{m} (\text{סה"כ תשומה } i)\ \pi_i \tag{6-74}$$

בכפיפות לאילוצים

$$\sum_{i=1}^{m} \left(\frac{\text{תשומה } i}{\text{תפוקה } j}\right) \pi_i \ge \left(\frac{\text{רווח נטו מתוצר } j}{\text{תפוקה } j}\right) \tag{6-75}$$

$$\pi_i \ge 0$$

לפי המימרים של המשוואות דרוש בי ל-π_i יהיו המימדים

$$\left(\frac{\text{רווח נטו מתוצר } j}{\text{תשומה } i}\right) \tag{6-76}$$

ואז ל-v יהיה המימד של סה"כ הרווח נטו, כלומר, כמו של z. ניסוח מילולי של שתי הבעיות יהיה:

<u>הבעיה הראשונית</u>: עבור תשומות מוגבלות (b_i), ולפי הרווח נטו מכל תוצר (c_j) כמה יש לייצר מכל תוצר (x_j) כך שהרווח הכולל נטו יהיה מקסימלי.

<u>הבעיה הדואלית</u>: עבור תשומות מוגבלות (b_i) וחסם תחתון על הרווח של כל תוצר (c_j) איזה מחירי יתידה יש ליחס לכל תשומה (π_i) כך שסה"כ ערך התשומות הדרושות יהיה מינימלי.

למשתנים הדואליים π_i קוראים מחירי צל (Shadow Prices או Imputed Prices "מחירים מיוחסים", כאשר הכוונה היא שהם מיוחסים לתשומות). π_i גם נותן בכמה ישתפר הפתרון האופטימלי אם האילוץ ה-i ישוחרר ביחידה, כלומר,

$$\pi_i^* = \frac{\Delta z^*}{\Delta b_i} \qquad (6\text{-}77)$$

נחזור לפתרון הגרפי של הבעיה הראשונית, ונכדוק את השפעתם של שינויים בערכי b_i. אם במקום האילוץ הראשון כצורתו המקורית היה רשום

$$x_1 + x_2 \le 7$$

כלומר $\Delta b_1 = 1$, בעוד שאר האילוצים נשארו כפי שהיו, אפשר לראות בציור 6.1 מה יקרה. הקו המתאר אילוץ זה יזוז מקביל לעצמו כלפי מעלה עד שיחתוך את הקו

$$-2x_1 + x_2 = 1$$

(בנקודה $x_1=2$; $x_2=5$) שבה יהיה הפתרון האופטימלי. ערכו של פתרון זה הוא 12, כלומר שיפור מהערך הקורם בשיעור $\Delta z^* = 12 - 31/3 = 5/3$, השווה בריוק לערך של π_1 שחישבנו קודם. באופן דומה קל לראות שעבור $\Delta b_1 = -1$, כלומר כאשר $b_1 = 5$, מתקבל הפתרון האופטימלי $z^* = 26/3$, ושוב קיים

$$\pi_1^* = \frac{\Delta z^*}{\Delta b_1} = \frac{26/3 - 31/3}{-1} = \frac{-5/3}{-1} = \frac{5}{3}$$

הזזת האילוץ השני $2x_1 + x_2 \le 10$ ביחידה לשני הכיוונים לא תשנה את הפתרון האופטימלי, שכן אילוץ זה אינו כוכל בפתרון האופטימלי המקורי. לכן

$$\pi_2^* = \frac{\Delta z^*}{\Delta b_2} = 0$$

כפי שאמנם קיבלנו.

בדיקה רומה עבור האילוץ השלישי תראה כי עכור $\Delta b_3 = 1$ מתקבל,

$$\frac{\Delta z^*}{\Delta b_3} = \frac{32/2 - 31/3}{1} = \frac{1}{3} = \pi_3^*$$

Δb = תוספת (תמיד גד)
Δz^* = שינוי (לאו דוקא "שיפור")

כמובן שמדובר רק בשינויים שאינם גורמים להחלפת הבסיס האופטימלי. שינוי בסיס אופטימלי קורה כאשר עקב "תנועת" אחד האילוצים, תוך השארת שאר האילוצים במקומם, מתקבל הפתרון האופטימלי בנקודת החיתוך של אילוצים אחרים. ערכיהם המספרי של z^* ושל המשתנים הבסיסיים משתנים כל זמן הזזת האילוץ, גם כאשר הבסיס האופטימלי אינו משתנה. להלן נדאה כיצד לחשב את תחום השינוי של כל b_i שכמסגרתו אין הבסיס האופסימלי משתנה.

מבחינה מעשית יש חשיבות רבה למחירי הצל. היות וידוע שיחול שיפור בשיעור π_i^* אם האילוץ ה-i "ישוחדר" ביחידה, אפשר להסיק מכך כמה כדאי לשלם תמורת תוספת יחידה לכמות המשאב ה-i. כאשר דנים בכעית ייצור, הרי שאם אפשר לרכוש תוספת של יחידה אחת של התשומה ה-i כמחיר הקטן מ-π_i^* כדאי לעשות זאת, שכן הרווח נטו יגדל כתוצאה מתוספת זו בתשומה ביותר מאשר מחירה (מחיר זה של התשומה הוא המחיר השולי, שכן מדובר בתוספת של יחידה על הכמות הקיימת כבר). מאידך, בבעית מינימיזציה של עלויות, כאשר האילוץ ה-i מכטא למשל את ההכרח לייצר לפחות b_i יחידות תפוקה מסויימת, יהיה מחיר הצל המתאים שווה להקטנת העלות הכוללת שתתקבל מ"שחרור" אילוץ זה ביחידה, כלומר, <u>מהקטנת</u> b_i ביחידה. כאשר הראשונית מנוסחת לפי משוואות (6-60) עד (6-62) יהיו המשתנים הדואליים חמיד לא שליליים. בניסוח כלשהו של הראשונית תהיה משמעות סימני המשתנים הדואליים כדלקמן:

פונקצית המטרה	אילוץ	π_i^*	Δz^* עבור $\Delta b_i=+1$
$max\ z$	$\le b_i$	>0	>0
	$\ge b_i$	<0	<0
	$= b_i$	>0	>0
	$= b_i$	<0	<0
$min\ z$	$\le b_i$	<0	<0
	$\ge b_i$	>0	>0
	$= b_i$	>0	<0
	$= b_i$	<0	<0

בכדי לראות מיקומם בטבלו האופטימלי ומשמעותם של המשתנים הדואליים המתאימים לאילוצי שוויון נחזור לדוגמה של סעיף 6.4, אשר הטבלו האופטימלי שלה ניתן בסעיף 6.10.2: (ראה עמ' 120)

$$\min \quad z = x_1 + 6x_2 - 7x_3 + x_4 + 5x_5$$

$$\text{s.t.} \quad 5x_1 + 4x_2 + 13x_3 - 2x_4 + x_5 = 20$$

$$x_1 - x_2 + 5x_3 - x_4 + x_5 = 8$$

לאילוץ הראשון הוכנס משתנה מלאכותי x_6 ולשני x_7. בטבלו האופטימלי הופיעו כשורת z מתחת למשתנים אלה הערכים 23/7 ו-50/7- בהתאמה. הפתרון האופטימלי הוא:

$$x_2^* = 4/7 \qquad x_3^* = 12/7 \qquad z^* = -60/7$$

תוספת 1+ לצד ימין של האילוץ הראשון מביאה לפתרון אופטימלי

$$x_2^* = -1/7 \qquad x_3^* = 11/7 \qquad z^* = -83/7$$

השיפור בפונקצית המטרה הוא 23/7, השווה למשתנה הדואלי של האילוץ הראשון. נשים לב כי כחישוב זה התרנו ל-x_2 להפוך שלילי, דבר הנוגר את דרישת אי-השליליות, רק לפשטות החישוב של $\Delta z^*/\Delta b_1$. תוצאה זהה היתה מתקבלת לו הגדלנו את צד ימין של האילוץ, למשל, ב-0.1 במקום ב-1.0, שכן אז היה מתקבל $\Delta z^* = 23/70$, ו-x_2 לא היה הופך שלילי.

תוספת 1+ לצד ימין של האילוץ השני מביאה לפתרון האופטימלי

$$x_2^* = 17/7 \qquad x_3^* = 16/7 \qquad z^* = -10/7$$

ואמנם, פונקצית המטרה "התקלקלה" בשעור 50/7, כפי שהיה חזוי על ידי $\pi_2^* = -50/7$.

להסבר משמעות המשתנים הדואליים נחזור לדוגמה של המפעל המייצר ארבעה מוצרים, סעיף 6.10.2. לאחר שנכתוב את האילוצים במתכונת של משוואות (6-61), תוך הפיכת סימני ארבעת האילוצים האחרונים ניתן לכתוב את הדואלית לפי משוואות (6-63) עד (6-65) כדלקמן:

$$\min \quad v = 1500\pi_1 + 1000\pi_2 + 800\pi_3 - 40\pi_4 - 130\pi_5 - 30\pi_6 - 10\pi_7$$

$$\text{s.t.} \quad \begin{aligned} 5\pi_1 + 2\pi_2 + 3\pi_3 - \pi_4 &\geq 12 \\ \pi_1 + 3\pi_2 + 2\pi_3 - \pi_5 &\geq 5 \\ 9\pi_1 + 4\pi_2 + 5\pi_3 - \pi_6 &\geq 15 \\ 12\pi_1 + \pi_2 + 10\pi_3 - \pi_7 &\geq 10 \end{aligned}$$

$$\pi_i \geq 0 \qquad i = 1,\ldots,7$$

בפתרון האופטימלי של הבעיה הראשונית היו האילוצים הבאים כובלים: השביעי - לכל היותר 800 שעות של פעולה 3, התשיעי - לכל הפחות 130 יחידות של מוצר 2, העשירי - לכל הפחות 30 יחידות של מוצר 3, והאחר-עשר - לכל הפחות 10 יחידות של מוצר 4. לאילוצים אלה מתאימים בבעיה הדואלית לפי הסדר π_3, π_5, π_6 ו-π_7, אשר, בהיותם מחירי הצל של האילוצים הכובלים, יופיעו בבסיס האופטימלי של הבעיה הדואלית. את ערכיהם יש לקרוא מתוך השורה האחרונה בטבלו הסופי של הבעיה הראשונית לפי הקשרים (6-70). מתקבל,

$$\pi_3^* = 4 \qquad \pi_5^* = 3 \qquad \pi_6^* = 5 \qquad \pi_7^* = 30$$

חישוב v^* בעזרת ערכים אלה נותן,

$$v^* = 800\times4 - 130\times3 - 30\times5 - 10\times30 = 2360$$

ואמנם, התקבל $v^* = z^*$. כמו כן קל להיווכח על ידי הצבה כי כל אילוצי הדואלית מתקיימים, במקרה זה כשוויונות. הדבר קורה משום שבפתרון הראשונית הופיעו בבסיס האופטימלי כל המשתנים המקוריים של הבעיה הראשונית (x_1 עד x_4), ולכן בני-זוגם, שהם משתני העודף בבעיה הדואלית כולם מתאפסים בפתרון האופטימלי.

על סמך הערכים האופטימליים של המשתנים הדואליים, שהם מחירי הצל של הבעיה הראשונית ניתן להסיק:

א. תוספת של שעת עבודה אחת בחודש בפעולה 3, כלומר שינוי מ-800 שעות ל-801 שעות, תביא לשיפור של 4 יחידות בפונקציית המטרה, כלומר, שעבור $b_3=801$ היה מתקבל $z^*=2364$.

בשים לב כי דוקא משום שהאילוץ השלישי הוא הכובל קבלנו כפתרון האופטימלי של הראשונית ייצור של מוצר א' מעל למינימום הדרוש. הדבר מוסבר על ידי בדיקת היחסים (c_j/a_{3j}) כלומר היחס בין המחירים לבין המקדמים הטכנולוגיים באילוץ השלישי. עם כל יחידה נוספת של b_3 אפשר לייצר (1/3) יחידה של x_1 אשר תמורתה תהיה $12 \times (1/3)=4$. לעומת זה נותן יחס המקדמים עבור שאר המוצרים,

$$c_2/a_{32} = 5/2 \quad ; \quad c_3/a_{33} = 3 \quad ; \quad c_4/a_{34} = 1$$

$c_1/a_{31} = 4$ הוא אם כן היחס הגבוה ביותר אשר פירושו הניצול הטוב ביותר של כל שעת עבודה נוספת של פעולה 3. יחס זה גם שווה למחיר הצל שקבלנו.

ב. המסקנה ממחירי הצל של האילוצים על מינימום מספר היחידות המיוצרות של מוצרים ב,ג ו-ד היא למשל: לו היינו משחררים את האילוץ על ייצור מוצר מספר כ'ביחידה ("שחרור" כאילוץ מהצורה ≥ מתבטא על ידי הקטנת ה-b המתאים) היה הפתדון משתפר ב-3 יחידות, כלומר, עבור $b_5 = 129$ היה מתקבל $z^* = 2363$. הדבר מוסבר בצורה הבאה: הקטנת b_5 ביחידה היתה מאפשרת הקטנת הייצור האופטימלי של מוצר ב' ביחידה. זה אמנם מקטין את z ב-5 יחידות. אבל היות ועקב הקטנה זו משתחררות 2 שעות של פעולה 3, הרי שאפשר להגדיל את הייצור של מוצר א' ב-2/3 יחידות המכניסות $12 \times (2/3) = 8$ יחידות, התוספת נטו לפונקצית המטרה היא $8-5=3$. באופן רומה ניתן להבין את מחירי הצל של שני האילוצים האחרים על ייצור מינימום מספר יחידות של מוצרים ג' ו-ד'.

ג. היות ומחיר הצל של האילוץ השלישי הוא 4, משתמע מכך שאם אפשר לקנות שעת עבודה אחת נוספת בחורש של פעולה 3 כמחיר הקסן מ-4 יחידות כדאי לעשות זאת, שכן התוספת לפונקצית המטרה על כל שעה נוספת כזו תכסה את ההוצאה ותותיר רווח. ברור שמסקנה זו נכונה עבור קנית מספר מסוים של שעות פעולה 3, שכן מעבר לערך מסוים של b_3, בהנחה שכל שאר המקדמים בבעיה נשארים כשהיו, חדל אילוץ זה להיות כובל ומחיר הצל החדש שלו הוא אפס. הטיפול בשאלה זו נדחה לסעיף 6.12.2.

ד. נניח שאפשר לקנות את אותם מוצרים שמייצר המפעל גם ממקורות אחרים. לפי מחירי הצל כדאי למפעל הנדון לקנות את המוצרים ב' ג' ו-ד' במקום לייצרם, אם מחיריהם בשוק קטנים מ-5,3 ו-30 יחידות בהתאמה. קנייתם במחירים אלה מאפשרת שחרור שעות עבודה לייצור מוצר א' באופן שההכנסה הנוספת ממנו יותר מאשר מכסה את הוצאות קנית המוצרים האחרים. כמה מוצרים מכל סוג כדאי לקנות היא שאלה רומה לזו שנדונה לעיל, ואשר תזכה לטיפול מפורט בסעיף 6.12.2.

<u>הערה:</u> במספור העמודים נפלה שגיאה. <u>עבור לעמוד 143.</u>

ה. מחיר הצל נותן, בהתאם למשוואה (6-76), את חלקו של הרווח נטו מהתוצר ה-j שיכול להיות מיוחס ליחידת התשומה ה-i. בהתאם לזה ניתן למשל לייחס $\pi_3^* = 4$ יחידות רווח לכל יחידה של התשומה השלישית (פעולה 3) בייצור המוצר א', בעוד $\pi_1^* = \pi_2^* = \pi_4^* = 0$ פירושם שאף תשומה אחרת אינה "תורמת" לרווח זה. ואמנם, היות ודרושות 3 שעות של פעולה 3 לייצור יחידת מוצר א', מתקבל $a_{33}\pi_3^*=3\times4=12$ השווה בדיוק ל- c_1 - המקדם של x_1 בפונקציית המטרה.

6.12 מבחני רגישות: תחומי התקפות של הפתרון לשינויים בנתונים

שלושה סוגי מקדמים מופיעים בבעית התכנות הליניארי:

א. המקדמים הטכנולוגיים a_{ij} - המבטאים את התמרת התשומות לתפוקות.

ב. האילוצים b_i - המבטאים בדרך כלל את כמות המשאבים המצויה בידי המתכנן והדרושים כתשומות.

ג. ה"מחירים" c_i - המבטאים את העלות או את ההכנסות הנקיות של התפוקות השונות.

לעיתים קרובות קיימת אי-ודאות ביחס לערכם המספרי של לפחות חלק מהמקדמים האלה. חשוב לכן לבדוק את רגישותו של הפתרון האופטימלי לשינויים בכל אחד מסוגי המקדמים.

אם רוצים לבדוק מהו הפתרון האופטימלי באשר חלק מהמקדמים או כולם מקבלים ערך שונה מזה שהיה להם בפתרון הראשון של הבעיה, אפשר כמובן לפתור את הבעיה מחדש, כאילו אין בידינו כבר פתרון אחד. שיטה זו אינה יעילה, משום שיש תחום מסוים בו לא חל שינוי בבסיס האופטימלי (כלומר, אותם אילוצים נשארים כובלים) וקל לחשב את הפתרון החדש מן הישן. נבדוק לבן מהו תחום השינוי המותר עבור כל סוג מקדמים בו לא יהיה שינוי <u>במדיניות</u> האופטימלית, כלומר, אותם משתנים ישארו בבסיס ורק ערכיהם וערך פונקצית המטרה ישתנה.

6.12.1 שינויים בוקטור המחירים

נבדוק בכמה אפשר לשנות את וקטור המחירים בכיוון כלשהו בלי לשנות את הבסיס האופטימלי. במהלך השינוי הזה לא "יזוז" הפתרון האופטימלי ממקומו ואילו ערך פונקצית המטרה האופטימלי כן ישתנה.

בניח שעבור הבעיה

$$\begin{aligned} max \quad z &= \underline{c}\,\underline{x} \\ s.t. \quad \underline{\underline{A}}\,\underline{x} &= \underline{b} \\ \underline{x} &\geq \underline{0} \end{aligned} \tag{6-78}$$

הבסיס האופטימלי הוא $\underline{x}_B = \underline{\underline{B}}^{-1}\,\underline{b}$ ומתאים לו $z^* = \underline{c}_B\,\underline{x}_B$, ואנו משנים את וקטור המחירים על ידי הוספת וקטור $\underline{\Delta c}$ כעל כיוון כלשהו, כלומר, המקדמים ב-$\underline{\Delta c}$ אינם בהכרח פרופורציוניים לאלו של $\underline{c}$. וקטור המחירים החדש הוא

$$\underline{c}^+ = \underline{c} + \alpha\underline{\Delta c} \tag{6-79}$$

כאשר α הוא גורל לא-שלילי כלשהו. נבדוק באופן גרפי מה קורה עבור שני משתנים. כעזרת ציור 6.2.

הפתרון האופטימלי המקורי הוא בפינה 1 של האיזור האפשרי. וקטור המחירים המקורי הוא $\underline{c}$, קוים שווי z ניצבים לו, והפתרון האופטימלי הוא z^*. הכיוון של וקטור השינויים, $\underline{\Delta c}$ נראה בציור. עבור $\alpha = \alpha_1$ מתקבל בחיבור וקטורי

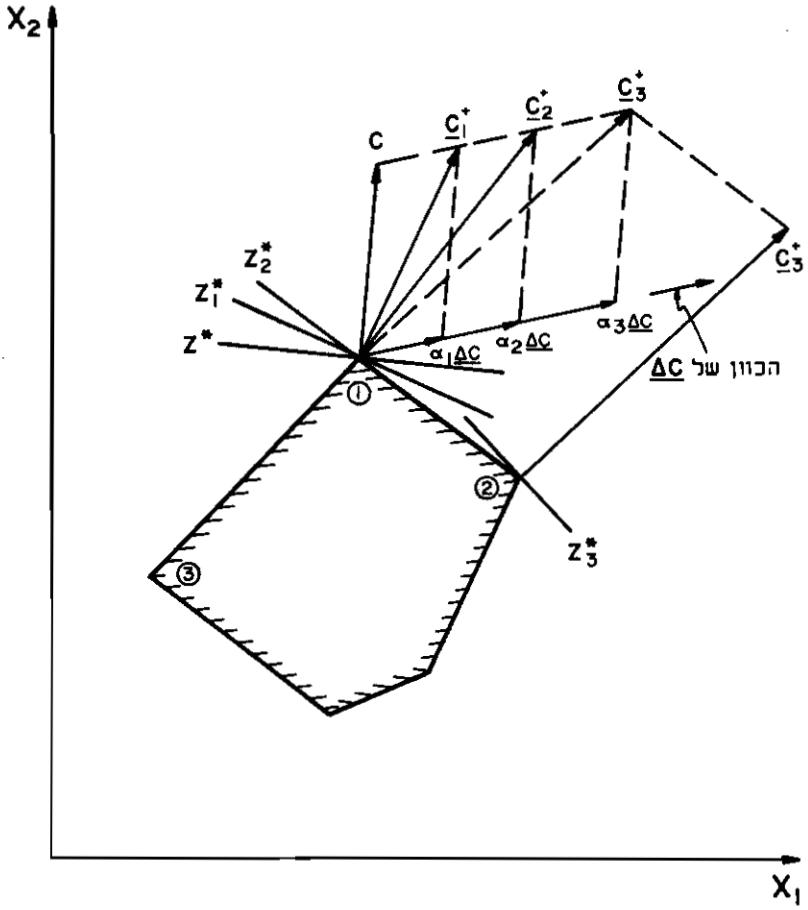
$$\underline{c}_1^+ = \underline{c} + \alpha_1\,\underline{\Delta c}$$

הפתרון האופטימלי עודנו בפינה 1 עם ערך אופטימלי z_1^*. כאשר נגדיל את α עד לערך α_2 יהיה הקו

$$\underline{c}_2^+ = \underline{c} + \alpha_2\,\underline{\Delta c}$$

ניצב לפאת האיזור האפשרי המחברת את הפינות 1 ו-2. הערך האופטימלי הוא z_2^* (מתקבל כניצב ל- $\underline{c}_2^+$), והכסיס האופטימלי הוא כל נקורה על הפאה. כאשר נגדיל את α מעבר α_2 יעבור הפתרון האופטימלי לפינה 2, בפי שנראה בציור עכור $\alpha = \alpha_3$.

ברור מהציור שלו בחרנו $\underline{\Delta c}$ קרוב בכיוונו ל-$\underline{c}$, היינו יכולים להגדיל את α ללא גבול בלי לגרום לשינוי הבסיס האופטימלי. באופן ספציפי, כל עוד $\underline{\Delta c}$ נמצא בתוך האיזור התחום על ידי הניצבים לפיאות האיזור האפשרי בנקודה 1, אפשר להגדיל את α כרצוננו בלי לגרום לשינוי בבסיס האופטימלי.



<u>ציור 6.2</u>: שינויים בוקטור המחירים.

לבעינת דו-מימדיות ניתן לאתר את תחום התקפות לשינויים בוקטור המחירים ע"י "סיבוב" פונקצית המטרה עד שהיא חופפת את אחד מן האילוצים הנחתכים בנקודת האופטימום.

חישוב אנליטי של תחום התקפות לשינויים בוקטור המחירים נעשה בשיטת הסימפלקס המעודכן. החישוב ייעשת כדלקמן: בבעית מינימיזציה יהיו בפתרון האופטימלי כל ערכי $(c_j - z_j)$ חיוביים - ראה משוואות (6-48) עד (6-50). כאשר מחליפים את $\underline{c}$ ב-$\underline{c}^+$ יהיו הערכים החדשים $(c_j^+ - z_j^+)$. שינוי בבסיס האופטימלי יחול כאשר לפחות אחד מערכים אלה יהפוך שלילי. נסמן ב-$\Delta\underline{c}_B$ את וקטור השינויים במחירים של המשתנים הבסיסיים בפתרון האופטימלי (יש לשים לב שסדר האיברים בוקטור זה מתאים לסדר המשתנים בבסיס האופטימלי), וב-Δc_j את רכיב השינוי במחיר של המשתנה הלא-בסיסי ה-j. אזי, לפי (6-45)

$$c_j^+ - z_j^+ = (c_j+\alpha\Delta c_j) - (\underline{c}_B+\alpha\underline{\Delta c}_B)\underline{y}_j = (c_j-z_j)+\alpha(\Delta c_j-\underline{\Delta c}_B\underline{y}_j) \qquad (6\text{-}80)$$

כאשר $\underline{y}_j = \underline{\underline{B}}^{-1}\,\underline{a}_j$, ו-$\underline{\underline{B}}$ היא מטריצת הבסיס של הבעיה המקורית, לפני השינויים במחירים. כעת, אם מתקיים

$$\Delta c_j - \underline{\Delta c}_B\,\underline{y}_j > 0 \qquad (6\text{-}81)$$

עבור כל ה-x_j שאינם בבסיס אפשר להגדיל את α ללא גבול בלי לגרום לשינוי בבסיס האופטימלי. אם, לעומת זאת, יש בין ערכים אלה גם שליליים אפשר להגדיל את α רק עד שעבור אחד מהמשתנים יתקבל $c_j^+ - z_j^+ = 0$, שכן מעבר לערך זה של יהיה גודל זה שלילי וידרוש שינוי בבסיס. הערך הקריטי של α מחושב לפי

$$\alpha_{cr} = \min_j \left\{ \frac{-(c_j-z_j)}{c_j-\underline{\Delta c}_B y_j} \;\middle|\; \Delta c_j - \underline{\Delta c}_B y_j < 0 \right\} \qquad (6\text{-}82)$$

בבעית מקסימיזציה יהיו בפתרון האופטימלי כל ערכי $(c_j - z_j)$ שליליים, והערך הקריטי של α יתקבל כאשר אחד מערכי $(c_j^+ - z_j^+)$ המתושבים לפי (6-80) יהפוך חיובי, כלומר

$$\alpha_{cr} = \min_j \left\{ \frac{-(c_j - z_j)}{\Delta c_j - \underline{\Delta c}_B y_j} \middle| \Delta c_j - \underline{\Delta c}_B y_j > 0 \right\} \qquad (6\text{-}83)$$

<u>רוגמא:</u>

(המשך דבר)

נחזור לכעיה הראשונית של סעיף 6.11 עכורה יש שם פתרון מספרי וכן פתרון גרפי, המופיע בציור 6.1, ונכדוק השפעת שינויים ב- c_1.

וקטור המחירים המקורי הוא: $\underline{c} = (c_1, c_2) = (1,2)$. נברוק בבת-אחת השפעת הגדלתו והקטנתו של c_1. לשם כך נרשום,

$$\underline{c}^+ = \underline{c} + \alpha \underline{\Delta c} = (1,2) + \alpha(\pm 1, 0)$$

כאשר α גודל לא-שלילי. נחשב את הערך הקריטי של α לפי משוואה (6-83). לנוחיות הקורא נעתיק כאן את הטכלו הסופי של פתרון הבעיה המקורית, ממנו יש לשאוב את האינפורמציה לחישובים.

$\underline{x}_B$	1	2	3	4	5	$\underline{b}$
x_1	1		1/3		-1/3	5/3
x_4			-4/3	1	1/3	7/3
x_2		1	2/3		1/3	13/3
z			-5/3		-1/3	-31/3

הבסיס האופטימלי הוא (x_1, x_4, x_2) ומתאים לו וקטור מחירים מקורי $\underline{c}_B = (1,0,2)$ ווקטור שינויי מחירים $\underline{\Delta c}_B = (\pm 1, 0, 0)$. בעזרת וקטור זה נחשב את הערכים $(\Delta c_j - \underline{\Delta c}_B y_j)$ עבור כל המשתנים שאינם בבסיס, כאשר הוקטורים $\underline{y}_j$ הם אלה המופיעים בטבלו האחרון.

$$\Delta c_3 - \underline{\Delta c}_B\, \underline{y}_3 = 0 - (\pm 1,0,0) \begin{Bmatrix} 1/3 \\ -4/3 \\ 2/3 \end{Bmatrix} = \mp 1/3$$

$$\Delta c_5 - \underline{\Delta c}_B\, y = 0 - (\pm 1,0,0) \begin{Bmatrix} -1/3 \\ 1/3 \\ 1/3 \end{Bmatrix} = \pm 1/3$$

מסתבר כי כאשר $\Delta c_1 = +1$ מתקיים רק עכור x_5 חנאי אי-השלילות הקובע גבול קריטי ל-α לפי משוואה (6-83). אין צורך לחפש מינימום, שכן יש רק ערך אחד ל-α והוא:

$$\alpha_{cr} = \frac{-(c_5-z_5)}{\Delta c_5-\underline{\Delta c}_B\, \underline{y}_5} = \frac{-(-1/3)}{1/3} = 1$$

משמעות התוצאה היא וקטור המחירים הגבולי עבורו יישאר הבסיס האופטימלי ללא שינוי הוא

$$\underline{c}^+ = \underline{c} + \alpha_{cr}\underline{\Delta c} = (1,2) + 1 \times (1,0) = (2,2)$$

כל עוד $c_1 \le 2$ לא יחול שינוי בבסיס. התוצאה מוסברת מעיון בציור 6.1. עם הגדלת c_1 גדל שיפוע הקוים שווי z. הערך הגבולי מתקבל כאשר הקוים מקבילים לאילוץ $x_1 + x_2 = 6$. היות והמקדם של x_2 כפונקציית המטרה הוא $c_2 = 2$, דרוש שיהיה גם $c_1 = 2$ כדי שהקוים יהיו מקבילים. אם מגדילים את c_1 מעבר לערך זה ייכנס לבסיס x_5 שהוא משתנה החסר של האילוץ $-2x_1 + x_2 \le 1$ שכן הפתרון האופטימלי החדש כבר לא יהיה על הקו $-2x_1 + x_2 = 1$.

עכור $\Delta c_1 = -1$ מוגבל α רק בגלל x_3, וערכו

$$\alpha_{cr} = \frac{-(c_3-z_3)}{\Delta c - \underline{\Delta c}_B\, \underline{y}_3} = \frac{-(-5/3)}{1/3} = 5$$

וקטור המחירים הגבולי הוא

$$\underline{c}^{+} = \underline{c} + \alpha_{cr}\underline{\Delta c} = (1,2) + 5(-1,0) = (-4,2)$$

כל עוד $c_1 \geq -4$ לא יחול שינוי בבסיס. מעיון בציור 6.1 ברור ש- $c_1 = -4$ נותן קוים שווי z המקבילים לאילוץ $-2x_1 + x_2 = 1$ והקטנת c_1 מתחת לערך קריטי זה תגרום לשינוי בבסיס האופטימלי. הפתרון האופטימלי החדש יהיה בפינת האיזור האפשרי (0,1) ובו נמצא בבסיס x_3 - משתנה החסר של האילוץ $x_1 + x_2 \leq 6$ אשר חדל להיות כובל.

כל עוד $\alpha < \alpha_{cr}$ הבסיס נשאר אותו, ערכם של המשתנים הבסיסיים אף הם אינם משתנים, ואת הערך החדש של z^* אפשר לחשב על ידי $z^* = \underline{c}^{+}\underline{x}_B$, באשר $\underline{x}_B$ הוא הפתרון הבסיסי המקורי. כאשר $\alpha > \alpha_{cr}$ שיטת הפתרון היא כדלקמן: מתחילים מן הטבלה הסופית של הפתרון המקורי, כאשר ערכי $(c_j - z_j)$ מוחלפים ב-$(c_j^{+} - z_j^{+})$ וממשיכים את התהליך האיטרטיבי עד לקבלת הפתרון האופטימלי.

כאשר $\underline{c}^{+}$ שונה מאד מ- $\underline{c}$, ייתכן שכראי להתחיל את הפתרון מההתחלה, כאילו היתה זו בעיה חדשה, ולא להסתמך על הפתרון האופטימלי המקורי.

6.12.2 שינויים בוקטור האילוצים

אנו מעונינים לבחון מהו תחום השינויים בוקטור האילוצים $\underline{b}$ אשר בו אין שינוי בהרכב המשתנים שבבסיס האופטימלי. במלים אחרות, אנו מחפשים את תחום השינויים בוקטור $\underline{b}$ אשר בתוכו לא יחול שינוי בנקודת חיתוך האילוצים המהווה את הפתרון האופטימלי. ברור שבמהלך השינוי "יזוז" הפתרון האופטימלי ממקומו, ואף ערכה של פונקצית המטרה ישתנה.

בדוגמאות רו-מימריות ניתן לבחון את תחום התקפות של הפתרון לשינויים בוקטור $\underline{b}$ בדרך גרפית, ע"י הזזת כל אילוץ במקביל לעצמו עד שנקודת הפתרון האופטימלי מתקבלת בחיתוך חדש של אילוצים ו/או צירי המערכת. לבעיות רב-מימריות ניתן לקבל תשובה אנליטית רק בשיטת הסימפלקס המעודכן.

נניח שעבור הבעיה המוגדרת ב-(6-87) יש בידינו פתרון בסיסי אופטימלי $\underline{x}_B = \underline{\underline{B}}^{-1}\,\underline{b}$.
נוסיף כעת לוקטור האילוצים $\underline{b}$ תוספת רצונית $\theta\underline{\Delta b}$, כאשר רכיבי $\underline{\Delta b}$ הם כלשהם, כלומר, לא כהכרח פרופורציוניים לאלו של $\underline{b}$. וקטור האילוצים החדש הוא

$$\underline{b}^{+} = \underline{b} + \theta\,\underline{\Delta b} \qquad (6\text{-}84)$$

כאשר θ הוא גודל לא שלילי כלשהו.

נמצא את הערך המקסימלי של θ עבורו הכסיס האופטימלי הנוכחי עדיין אפשרי. נבתון את הביטוי עבור הבסיס האופטימלי, שערכם המספרי של המשתנים בו משתנה תוך כדי הגדלת θ, כל עוד אין החלפה של הבסיס האופטימלי.

$$\underline{x}_B^{+} = \underline{\underline{B}}^{-1}\,\underline{b}^{+} = \underline{\underline{B}}^{-1}\,\underline{b} + \theta\,\underline{\underline{B}}^{-1}\,\underline{\Delta b} = \underline{x}_B + \theta\,\underline{p} \qquad (6\text{-}85)$$

כאשר $\underline{p} = \underline{\underline{B}}^{-1}\underline{\Delta b}$ ו- $\underline{x}_B$ הוא הפתרון הבסיסי האופטימלי הנוכחי. אם $\underline{p} > \underline{0}$ כלומר אין אף $p_i < 0$, אפשר להגדיל את θ בלי גבול ו-$\underline{x}_B^{+}$ ישאר תמיד אפשרי ואופטימלי - משום שהמחירים לא השתנו ולכן אין צורך להחליף משתנים ככסיס.

אם $p_i < 0$ עכור אחד או יותר מערכי i, אזי הערך הקריטי של θ, זה שיגרום לשינוי בבסיס האופטימלי, הוא

$$\theta_{cr} = \min_{i}\left\{ \frac{-x_{B_i}}{p_i} \,\middle|\, p_i < 0 \right\} \qquad (6\text{-}86)$$

או

$$\theta_{cr} = \infty \quad \text{if} \quad \underline{p} = \underline{\underline{B}}^{-1}\,\underline{\Delta b} > \underline{0} \qquad (6\text{-}87)$$

חישוב הפתרון האופטימלי לאחר הכנסת השינוי $\theta\,\underline{\Delta b}$ יעשה כדלקמן: אם $\theta \le \theta_{cr}$ לפי (6-86), מחשבים את ערביהם החדשים של המשתנים הבסיסיים $\underline{x}_B = \underline{\underline{B}}^{-1}\,\underline{b}^{+}$ ואת ערכה החדש של פונקצית המטרה $(z^*)^{+} = \underline{c}_B\,\underline{x}_B^{+}$. אם $\theta > \theta_{cr}$ הופך הבסיס האחרון לבלתי-אפשרי. ניתן להתחיל את כל הפתרון מתדש, עם וקטור האילוצים החדש $\underline{b}^{+}$.
קיימת גם אפשרות להשתמש בפתרון האופטימלי האחרון ולהמשיך ממנו. היות והפתרון הבסיסי החדש $\underline{x}_B^{+}$ אינו אפשרי, יש להשתמש באלגוריתם המרשה במהלבו פתרונות לא

אפשריים. לא התעכבנו על אלגוריתם זה (שיטת הסימפלקס הדואלי, Dual Simplex Algorithm), ולכן נאמר רק כי האפשרות להמשיך מהפתרון האחרון קיימת.

<u>דוגמא:</u>

נשתמש שוב בבעיה הראשונית של סעיף 6.11, שפתרונה הגרפי מופיע בציור 6.1, ונבדוק השפעת שינויים ב-b_1.

וקטור האילוצים המקורי הוא $\underline{b} = (6,10,1)$. נבדוק בבת אחת הגדלתו והקטנתו של b_1. לשם כך נרשום

$$\underline{b}^+ = \underline{b} + \theta\underline{\Delta b} = (6,\ 10,\ 1) + \theta(\pm 1,\ 0,\ 0)$$

כאשר θ גודל לא-שלילי, שאת ערכו הקריטי יש לקבוע לפי משוואה (6-86). מטריצת הבסיס האופטימלי וההפכית שלה הן:

$$\underline{\underline{B}} = [a_1, a_4, a_2] = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ -2 & 0 & 1 \end{pmatrix} \quad ; \quad \underline{\underline{B}}^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 0 & -1 \\ -4 & 3 & 1 \\ 2 & 0 & 1 \end{pmatrix}$$

נחשב את הוקטור $\underline{p}$ המוגדר ב-(6-85).

$$\underline{p} = \underline{\underline{B}}^{-1}\,\underline{\Delta b} = \frac{1}{3}\begin{pmatrix} 1 & 0 & -1 \\ -4 & 3 & 1 \\ 2 & 0 & 1 \end{pmatrix}\begin{Bmatrix} \pm 1 \\ 0 \\ 0 \end{Bmatrix} = \left(\pm\frac{1}{3},\ +\frac{4}{3},\ \pm\frac{2}{3}\right)$$

עבור $\Delta b_1 = +1$ מתקבל לפי (6-86)

$$\theta_{cr} = \min_i \left\{ \frac{-x_{B_i}}{p_i} \;\middle|\; p_i < 0 \right\} = \frac{-x_{B_2}}{p_2} = \frac{-7/3}{-4/3} = \frac{7}{4}$$

כאשר $x_{B_2} = x_4$, שכן לפי הסדר בבסיס האופטימלי x_4 הוא המשתנה השני. המסקנה היא שניתן להגדיל את b_1 עד לערך $b_1^+ = b_1 + \theta_{cr}\Delta b_1 = 6 + 7/4 = 31/4$ בלי שיחול שינוי בבסיס האופטימלי. עיון בציור 6.1 מראה כי עבור ערך זה של עובר האילוץ הראשון בצורתו החדשה, $x_1 + x_2 \le 31/4$, ررך נקודת החיתוך של האילוץ השני והשלישי $(x_1 = 9/4 \; ; \; x_2 = 22/4)$. הגדלת b_1 מעבר לערך הקריטי לא תגרום לשינוי נוסף בפתרון, שכן האילוץ הראשון לא יהיה כובל. כאשר $b_1 = 31/4$ עוזב x_4 - משתנה החסר של האילוץ הראשון - את הבסיס. עבור $\Delta b_1 = -1$ מתקבל לפי (6-86)

$$\theta_{cr} = \min_i \left\{ \frac{-x_{B_i}}{p_i} \right\} = \min \left\{ \begin{matrix} -x_{B_1}/p_1 \\ -x_{B_3}/p_3 \end{matrix} \right\} = \min \left\{ \begin{matrix} (-5/3)/(-1/3) \\ (-13/3)(-2/3) \end{matrix} \right\} = 5$$

כאשר הערך הקריטי, אשר התקבל עבור $x_{B_1} = x_1$ הוא

$$b_1^+ = b_1 + \theta_{cr}\,\Delta b_1 = 6 + 5(-1) = 1$$

עיון בציור 6.1 מורה כי כאשר $b_1 = 1$ עובר האילוץ הראשון ررך הנקורה $(x_1 = 0; \; x_2 = 1)$, בה מתאפס x_1. עבור $0 \le b_1 < 1$ הבסיס האופטימלי מכיל את x_1, x_2 ו-x_6 (משתנה החסר של האילוץ השלישי אשר כעת אינו כובל יותר). בסיס זה הוא מנוון שכן $x_1 = 0$.

6.12.3 שינויים במטריצה הטכנולוגית

אם עושים שינוי ב-$\underline{a}_j$, העמודה ה-j של המטריצה הטכנולוגית $\underline{\underline{A}}$ השייכת למשתנה x_j שאינו בבסיס האופטימלי, יש לערוך בריקה על המחיר החדש של המשתנה הלא-בסיסי הנרון לפי

$$c_j - z_j = c_j - \underline{c}_B \underline{y}_j^+ = c_j - \underline{c}_B \underline{\underline{B}}^{-1} \underline{a}_j^+ \qquad (6\text{-}88)$$

כאשר

$$\underline{a}_j^+ = \underline{a}_j + \phi\, \underline{\Delta a}_j \qquad (6\text{-}89)$$

הוא הוקטור עם השינויים. לפי הסימן של (c_j-z_j) אפשר לקבוע האם יש להכניס את המשתנה הנדון לבסיס. אם לא - הפתרון האופטימלי אינו משתנה, לא בערכו ולא בערכי המשתנים שבו. אם כן - מכניסים את המשתנה לבסיס וממשיכים באיטרציות עד לפתרון האופטימלי החדש.

אם עושים שינויים בוקטור a_j של משתנה בסיסי המצב מסובך הרבה יותר. שינוי כזה יכול: (א) לגרום למטריצת הבסיס $\underline{\underline{B}}$ ליהפך לסינגולרית, או (ב) לגרום לכך שהבסיס האחרון אינו אפשרי, או (ג) לגרום לצורך כשינוי בבסיס, עקב שינויים ב-(c_j-z_j) של המשתנים שאינם בבסיס. בדרך כלל מומלץ להתחיל במקרה כזה את הפתרון מההתחלה.

6.12.4 הוספת משתנים לבעיה

לעתים מתברר לאחר פתרון הבעיה כפי שהוצבה במקור, כי רוצים להוסיף משתנים נוספים, וזאת בלי להוסיף אילוצים. נניח שנוסף למשתנים המקוריים $(x_1,\ldots,x_n)$ נוסיף, לאחר שכבר התקבל הפתרון האופטימלי, משתנה חדש x_{n+1}, ואתו עמודה חדשה למטריצה $\underline{\underline{A}}$, $\underline{a}_{n+1}$ ומקדם c_{n+1} בפונקציית המטרה. בעזרת הבסיס האופטימלי נחשב

$$c_{n+1} - z_{n+1} = c_{n+1} - \underline{c}_B \underline{y}_{n+1} = c_{n+1} - \underline{c}_B \underline{\underline{B}}^{-1} \underline{a}_{n+1} \qquad (6\text{-}90)$$

לפי הסימן של גודל זה נקבע האם צריך להכניס את x_{n+1} לבסיס. אם לא - אין שינוי בפתרון האופטימלי. אם כן - מכניסים אותו לבסיס וממשיכים באיטרציות עד לקבלת הפתרון האופטימלי החדש.

6.12.5 הוספת אילוצים לבעיה

כמו במקרה של משתנים, כך גם מתברר לפעמים כי יש להוסיף אילוצים לבעיה לאחר שכבר נפתרה. הכוונה כאן שאין מכניסים משתנים חדשים, פרט כמובן למשתני חסר או עודף המוכנסים יחד עם האילוצים הנוספים.

הפתרון של הבעיה המורחבת תמיד יספק את אילוצי הבעיה המקורית, לכן הפתרון החדש אינו יכול לשפר את הפתרון המקורי, אלא לכל היותר להיות שווה לו. אפשריים שני מקרים:

(א) הפתרון האופטימלי המקורי מקיים גם את האילוצים החדשים - אין שינוי כפתרון.

(ב) הפתרון האופטימלי המקורי איבו מקיים את האילוצים החדשים - אפשר ליצור כסיס חדש מן הכסיס האופטימלי הנוכחי בתוספת האילוצים החדשים ולהמשיך את האיטרציות. הדבר פשוט למדי כאשר מוסיפים רק אילוץ אחד. כאשר מוסיפים יותר אילוצים יש לשקול את האפשרות להתחיל את כל הפתרון מחדש.

6.12.6 ריצות פרמטריות

כין תוכניות המחשב הקיימות, המיועדות לפתרון כעית התכנות הליניארי (ראה סעיף 6.14), יש כאלה המאפשרות כיצוע ריצות פרמטריות ובדיקות רגישות כצורה נוחה למשתמש. כקלט מספק המשתמש את הוקטורים $\underline{\Delta b}$, $\underline{\Delta c}$ ו-$\underline{\Delta a_j}$ וכן את שורת הערכים של α, θ ו-Φ (ראה משוואות (6-79), (6-84) ו-(6-89)), ואת סדר השינויים המכוקשים, והתוכנית מכצעת שורה של ריצות עוקבות. התוכנית כורקת כמקרה זה האם יש שינוי בבסיס האופטימלי, ומחשבת את הפתרון החדש כדרך היעילה כיותר.

קיימות תוכניות בהן אפשר לקבל כפלט (כעזרת פקודה מתאימה) את תחום אי-רגישותו של הפתרון האופטימלי לשינויים בוקטורים $\underline{b}$ ו-$\underline{c}$. הכוונה כאן לשינויים הנערכים אחר אחד בכל אחד מרכיבי $\underline{b}$ או $\underline{c}$, ולא לשינויים סימולטניים (שכן מספר הקומכינציות האפשריות לשינויים סימולטניים הוא גדול מאד).

6.13 משתנים תחומים

לעתים קרוכות נדרשים משתנים להיות בתחום מסוים. למשל

$$0 < d_j \le x_j \le u_j \qquad (6\text{-}91)$$

שייקרא <u>משתנה תחום</u> (Bounded Variable). למעשה, כל המשתנים בתכנות ליניארי תחומים, אם נרשה ל-d_j לקבל את הערך אפס, ו/או ל-u_j את הערך אינסוף, כאשר אין אילוץ מהצורה (6-91) מופיע במפורש.

ניתן היה לכלול אילוצים כאלה עם האחרים - ראה למשל את פתרון בעיית הייצור בסעיף 6.10 שנעשה בצורה זו. קיימת אפשרות לייעל את הפתרון, על ידי טיפול מיוחד באילוצי תחום.

באילוצי גבול תחתון

$$0 < d_j \le x_j$$

קל יחסית לטפל, שכן אפשר להגדיר משתנים חדשים

$$x'_j = x_j - d_j \qquad (6\text{-}92)$$

עבורם נדרש פשוט $x'_j \ge 0$. את וקטור האילוצים של הבעיה המקורית מחליפים בוקטור

$$\underline{b}' = \underline{b} - \sum_{j=1}^{r} d_j \underline{a}_j = \underline{b} - \underline{\underline{A}}\underline{d} \qquad (6\text{-}93)$$

כאשר $\underline{\underline{A}}$ היא מטריצת האילוצים, שאינה כוללת את אילוצי הגבול התחתון, ו-$\underline{d}$ הוא וקטור הגבולות התחתונים. פונקצית המטרה החדשה היא

$$z' = \sum_{j=1}^{r} c_j x'_j = \sum_{j=1}^{r} c_j (x_j - d_j) = \sum_{j=1}^{r} c_j x_j - \sum_{j=1}^{r} c_j d_j = \qquad (6\text{-}94)$$

$$= z - \sum_{j=1}^{r} c_j d_j \ldots$$

פותרים, אם כן, את הבעיה

$$\left.\begin{array}{lll} max. \ (or\ min) & z' = \sum_{j=1}^{r} c_j x'_j & \\ s.t. & \underline{\underline{A}}\ \underline{x}'\{\le\ ;\ =\ ;\ \ge\}\underline{b}' & \\ & \underline{x}' \ge \underline{0} & \end{array}\right\} \qquad (6\text{-}95)$$

ולאחר קבלת פתרונה מחשבים את המשתנים האופטימליים של הבעיה המקורית מתוך

$$\underline{x}^* = (\underline{x}')^* + \underline{d} \qquad (6\text{-}96)$$

ואת ערך פונקציית המטרה המקורית מתוך

$$z^* = (z')^* + \sum_{j=1}^{n} c_j d_j \qquad (6\text{-}97)$$

<u>דוגמא</u>

נראה ביצד ליישם שיטה זו בעזרת הדוגמא של סעיף 6.10, הבעיה של המפעל המייצר ארבעה מוצרים. ארכעת האילוצים האחרונים היו אילוצי גבול תחתון, כדלקמן

$$x_1 \geq 40 \quad ; \quad x_2 \geq 130 \quad ; \quad x_3 \geq 30 \quad ; \quad x_4 \geq 10$$

נגדיר לפי (6-93) משתנים חדשים

$$x_1' = x_1 - 40 \; ; \quad x_2' = x_2 - 130 \; ; \quad x_3' = x_3 - 30 \; ; \quad x_4' = x_4 - 10$$

עכורם נדרוש $x_j' \geq 0$. בבעיה החדשה יהיו רק שלושה אילוצים, ווקטור הצד הימני החדש, לפי (6-94) הוא

$$\underline{b}' = \underline{b} - \underline{\underline{A}}\underline{d} = \begin{Bmatrix} 1500 \\ 1000 \\ 800 \end{Bmatrix} - \begin{pmatrix} 5 & 1 & 9 & 12 \\ 2 & 3 & 4 & 1 \\ 3 & 2 & 5 & 10 \end{pmatrix} \begin{Bmatrix} 40 \\ 130 \\ 30 \\ 10 \end{Bmatrix} = \begin{Bmatrix} 780 \\ 400 \\ 170 \end{Bmatrix}$$

הבעיה החדשה היא

$$\begin{aligned} max \quad z' &= 12x_1' + 5x_2' + 15x_3' + 10x_4' \\ s.t. \quad & 5x_1' + x_2' + 9x_3' + 12x_4' \leq 780 \\ & 2x_1' + 3x_2' + 4x_3' + x_4' \leq 400 \\ & 3x_1' + 2x_2' + 5x_3' + 10x_4' \leq 170 \\ & x_1' , \; x_2' , \; x_3' , \; x_4' \geq 0 \end{aligned}$$

Tableau 0

$\underline{x}'_B$	1	2	3	4	5	6	7	$\underline{b}'$
x'_5	5	1	9	12	1	0	0	780
x'_6	2	3	4	1	0	1	0	400
x'_7	3	2	5	10	0	0	1	170
z'	12	5	15	10	0	0	0	0

Tableau 1

$\underline{x}'_B$	1	2	3	4	5	6	7	$\underline{b}'$
x'_5	-2/5	-13/5	0	-6	1	0	-9/5	474
x'_6	2/5	7/5	0	-7	0	1	-4/5	264
x'_3	3/5	2/5	1	2	0	0	1/5	34
z'	3	1	0	-20	0	0	-3	510

Tableau 2

$\underline{x}'_B$	1	2	3	4	5	6	7	$\underline{b}'$
x'_5	0	-7/3	2/3	-14/3	1	0	-5/3	1490/3
x'_6	0	17/15	-5/3	-25/3	0	1	-14/15	860/3
x'_1	1	2/3	5/3	10/3	0	0	1/3	170/3
z'	0	-3	-5	-30	0	0	-4	680

הפתרון האופטימלי של הבעיה החדשה הוא

$$(z')^* = 680 \; ; \quad (x_1')^* = 170/3 \; ; \quad (x_5')^* = 1490/3 \; ; \quad (x_6') = 860/3$$

וכן $(x_2')^* = (x_3')^* = (x_4')^* = (x_7')^* = 0$

פתרון הבעיה המקורית הוא לכן

$$z^* = 680 + \sum_{j=1}^{4} c_j \, d_j = 680+40\times12+130\times5+30\times15+10\times10 = 2360$$

$$x_1^* = 170/3 + 40 = 290/3 = 96\,\frac{2}{3}$$

$$x_2^* = 130 \; ; \quad x_3^* = 30 \; ; \quad x_4^* = 10$$

ואותו פתרון כמו בסעיף 6.10 הושג במאמץ תישובי קטן הרבה יותר.

נשים לב כי את המשתנים הדואליים של הררישות לייצור המינימלי קיבלנו הפעם מתחת לעמורות של המשתנים עצמם. משתנים דואליים אלה מבטאים את מתיר הצל של האילוץ $x_j' \geq 0$, הזהה לאילוץ $x_j \geq d_j$ שבבעיה המקורית, עבור המשתנים שאינם בבסיס האופטימלי.

באילוצי גבול עליון מהצורה

$$x_j < u_j$$

הטיפול מורכב יותר, אבל עדיין כדאי להפריד אילוצים אלה מן האחרים, לטפל בהם בנפרד ובכך לייעל את הפתרון. רוב תוכניות המחשב הקיימות מטפלות בצורה מיוחדת באילוצי תחום, ולכן כדאי לציין אותם עבור התוכנית בתור שכאלה (למשל: נתוני BOUNDS בתוכנית MPSX/370 ראה סעיף 6.14).

6.14 תוכניות מחשב לתיכנות ליניארי

קיימות תוכניות מחשב מוכנות לפתרון בעית התיכנות הליניארי עבור רוב המחשבים. התוכניות נבדלות זו מזו בגודל הבעיות אותן ניתן לפתור, בשיטת הפתרון ובעיקר בנוחיות ובגמישות שהן מקנות למשתמש בהן. היות ובהרבה מקרים מהווה התכנות הליניארי רק חלק מהניתוח הכולל של מערכות מורכבות, הרי שאפשרות לשלב טיפול בנתונים וחישובים מבינים או מסכמים ביחד עם הפתרון עצמו היא בעלת חשיבות ראשונה במעלה. את רוב התוכניות הקיימות ניתן לשלב כשגרה (Subroutine) בתוכנית אחרת. טבלה המסכמת את הנתונים העיקריים (שפת התכנות, שיטת הפתרון וגודל הבעיה הניתנת לפתרון) מופיעה בספרם של Kunzi et al. (עמ' 154-161 שם).

אחת מהנפוצות שבתכ ניות היא המערכת MPSX/370, המיועדת למחשבים מסדרת 370 של IBM. התוכנית מסוגלת לפתור בעיות עם אלפי אילוצים ומספר בלתי מוגבל של משתנים. התכנית APEX III פועלת על מחשבי CDC. השימוש בשתי התכניות דומה מאוד, שכן תוכננו להתאמה מלאה. המשתמש בתוכנית זו כותב תוכנית בקרה בעזרת פקודות מקרו. כל פקודה כזו מפעילה שגרה או שגרות בעלות תפקיד מסוים. על המשתמש לספק את הנתונים על גבי כרטיסים מנוקבים או על סרט מגנטי במתכונת הנדרשת על ידי התוכנית. יש לתת שם לכל אילוץ, לכל משתנה, לפונקצית המטרה ולוקטור האילוצים. בכל אילוץ יש לציין את צורתו: שוויון או אי-שוויון עם כיוונו. במערך הנתונים יש לתת במפורש רק את המקדמים השונים מאפס - התוכנית מאפסת את השאר. כמו כן אין צורך להוסיף משתני תסר, עודף ומלאכותיים - התוכנית מוסיפה אותם לפי סוג האילוץ. בעזרת פקודות המקרו כונה המשתמש את המטריצה הטכנולוגית, את וקטור האילוצים ואת פונקצית המטרה עבור הבעיה שהוא רוצה לפתור מתוך מערך הנתונים שהכתיב.

פקורות מקרו אחרות עורכות בדיקות בנתונים ומדפיסות סיכומים וטבלאות. בין השאר ניתן להדפיס "תמונה" של המטריצה, עם כל השמות של השורות והעמודות, כאשר בכל מקום שיש מקדם שונה מאפס מופיע סמל, ולכל סדר גודל של ערכי המקדמים מתאים סמל מיוחד. הסיכומים נותנים את ספירת המקדמים השונים מאפס לפי שורות ועמודות. הסיכומים וה"תמונה" מאפשרים אימות הנתונים. בעית האימות קשה ביותר, שכן כמות הנתונים בבעיות מעשיות לרוב גדולה מאור. לפני שמסתמכים על הפתרון אליו הגיעה התוכנית יש לוודא כי כל המקדמים מופיעים במקומם הנכון וכי ערכיהם נכונים. הפלט של פקודות הבדיקה מקל על עבודה זו, אבל אינו מבטל את הצורך בה. פקודות נוספות קוראות לשגרת הפתרון ומדפיסות את הפתרון - או את כולו או חלקים ממנו

לפי הדרישה. קיימות במערכת גם פקודות לביצוע ריצות פרמטריות, לחישוב תחום אי-הרגישות של הפתרון לשינויים בוקטור האילוצים או בוקטור המחירים, לאגירת הנתונים והפתרון של בעיה כדי שאפשר יהיה לחזור אליה בעתיד, לבצע שינויים בנתונים ולהריצה מחדש.

המערכת גם מאפשרת קישור בין שגרות הכתובות ב-FORTRAN לבין שאר התוכניות. קישור זה נעשה דרך מערכי נתונים מיוחדים בהם ניתן לאגור תוצאות של חלק אחד מהמערכת במטרה לשלוף אותם לתוך תוכניות אחרות.

השימוש בתוכניות מחשב אלה, ובתכניות דומות של מחשבים אחרים, פשוט למדי שכן נכתבו עבור המשתמש שאינו בקי במחשב עצמו אך מכיר את הבעיה הנפתרת.

המעונין יפנה אל ה"מדריך למשתמש" של התוכנית, אשר בו מוסבר כיצד להכין את תוכנית הכקרה, מהן הפונקציות העומדות לרשות המשתמש וכיצד להכין את הנתונים.

6.15 נושאים מיוחדים

פרק זה מוקדש להצגתם של כמה נושאים מיוחדים הקשורים בתיכנות ליניארי. הכוונה רק לתאר בקצרה את הבעיות ולהצביע באופן כללי על הגישה לפתרון. דיון מעמיק הוא מעבר להיקף הנכלל כאן, וניתן למצוא אותו בספרות.

6.15.1 בעית התובלה

בעית התובלה (Transportation Problem) היא סוג מיוחד ונפוץ למדי של בעית התיכנות הליניארי. בגלל מבנה המיוחד, ניתנת בעית התובלה לפתרון בשיטה יעילה יותר מאשר השיטה הכללית לפתרון בעיות תיכנות ליניארי.

ניסוח כללי של בעית התובלה היא כדלקמן: בכל אחד מ-m מקורות ישנה כמות מסויימת של מוצר. נדרש להעביר ~~כלל~~ כדלקמן אחר מ-n יעדים כמות מסויימת של המוצר, כאשר מחירי הובלת יחידת המוצר מכל מקור לכל יעד ידועה, באופן שסך כל הוצאות ותובלה תהיינה מינימליות.

על מנת שלבעיה יהיה פתרון אפשרי נדרש שסך כל הכמויות כמקורות יעלה על סך כל הדרישות ביעדים. ברור משיקולים אינטואיטיביים כי לכל יעד תובא בדיוק הכמות הנדרשת שם, שכן הכאת עודף רק גוררת הוצאה נוספת. נסמן ב-a_i את הכמות המצויה במקור ה-i, ב-d_j את הכמות הנדרשת כיעד ה-j, ב-x_{ij} את הכמות המועברת מהמקור ה-i ליעד ה-j, וב-c_{ij} את מחיר הובלת יחידת מוצר מהמקור ה-i ליעד ה-j. בעית התובלה היא לכן:

$$min \qquad z = \sum_{i,j} c_{ij}\, x_{ij} \tag{6-98}$$

$$s.t. \qquad \sum_{j=1}^{n} x_{ij} \le a_i \qquad i = 1,\dots,m \tag{6-99}$$

$$\sum_{i=1}^{m} x_{ij} = b_j \qquad j = 1,\dots,n \tag{6-100}$$

קל להפוך את האילוצים (6-99) לשוויונות על ידי הוספת משתני חסר. במקום לעשות זאת נניח כי עבור כל אחד מהמקורות מוגדר יעד פיקטיבי נוסף והוא המחסן לאיחסון העודף הנותר במקור. לכמות המובלת מכל מקור למחסן שלו אפשר לייחס מחיר הובלה של אפס, או, אם הדבר תואם את המציאות, לחייב כמות זאת בדמי האחסנה. לכן לא נפגעת כלליות הבעיה כשהאילוצים (6-99) נרשמים כשוויונות. את האילוצים (6-100) ניתן היה לכתוב ($\ge b_j$) אך ברור מראש שלא נוביל ליערים יותר מן הנדרש, שכן המטרה היא מינימיזציה של העלות.

לבעית התובלה אין פתרון אפשרי אלא אם מתקיים

$$\sum_{i=1}^{m} a_i \ge \sum_{j=1}^{n} b_j$$

המטריצה של בעית תובלה ל-m מקורות ו-n יעדים צורתה לכן

$$\underline{\underline{A}} = \begin{array}{c} \left[\begin{array}{ccccc} \underline{1}_n & 0 & 0 & \cdots & 0 \\ 0 & \underline{1}_n & 0 & \cdots & 0 \\ 0 & 0 & \underline{1}_n & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \underline{1}_n \\ \underline{\underline{I}}_n & \underline{\underline{I}}_n & \underline{\underline{I}}_n & & \underline{\underline{I}}_n \end{array}\right] \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \\ \\ \end{array}\right\} \text{m - שורות} \\ \left.\begin{array}{l} \\ \end{array}\right\} \text{n - שורות} \end{array} \end{array}$$

$\underline{1}_n$ הוא וקטור שורה כאורך n; $\underline{\underline{I}}_n$ היא מטריצת יחידה (n×n). $\underline{\underline{A}}$ היא לכן בעלת (m+n) שורות ו-(m n) עמודות.

את בעית התובלה נהוג להציג בצורת טבלה: לכל מקור מתאימה שורה, ולכל יעד - עמודה. בתוך כל משבצת נרשם המחיר c_{ij} של הובלת יחידה ממקור i ליעד j; לכל שורה מצוין כשר המקור, a_i, ולכל עמודה דרישת היעד הכוללת, b_j.

יעד j / מקור i	1	2	. . .	n	a_i
1	$c_{1,1}$	$c_{1,2}$	. . .	$c_{1,n}$	a_1
2	$c_{2,1}$	$c_{2,2}$	. . .	$c_{2,n}$	a_2
. . .	. . .	. . .		. . .	. . .
m	$c_{m,1}$	$c_{m,2}$	. . .	$c_{m,n}$	a_m
b_j	b_1	b_2	. . .	b_n	

ה'פ'ן m ו- n
ר'ש בלה.

<u>דוגמה</u>: ארבעה מקורות וארבעה יעדים. הנתונים:

יעד j / מקור i	1	2	3	4	a_i
1	2	5	6	1	10
2	5	2	3	5	5
3	1	3	1	3	6
4	3	1	4	2	7
b_j	8	6	5	9	

כאן $\sum_{i=1}^{4} a_i = \sum_{j=1}^{4} b_j = 28$. האילוצים:

$$
\begin{array}{l}
\text{מקורות} \\
\\
\\
\\
\text{יעדים} \\
\\
\\
\\
\end{array}
\left[\begin{array}{cccccccccccccccc}
1 & 1 & 1 & 1 & & & & & & & & & & & & \\
 & & & & 1 & 1 & 1 & 1 & & & & & & & & \\
 & & & & & & & & 1 & 1 & 1 & 1 & & & & \\
 & & & & & & & & & & & & 1 & 1 & 1 & 1 \\
1 & & & & 1 & & & & 1 & & & & 1 & & & \\
 & 1 & & & & 1 & & & & 1 & & & & 1 & & \\
 & & 1 & & & & 1 & & & & 1 & & & & 1 & \\
 & & & 1 & & & & 1 & & & & 1 & & & & 1
\end{array}\right]
\begin{array}{c}
\le \\ \le \\ \le \\ \le \\ = \\ = \\ = \\ =
\end{array}
\begin{bmatrix} 10 \\ 5 \\ 6 \\ 7 \\ 8 \\ 6 \\ 5 \\ 9 \end{bmatrix}
$$

סדר המשתנים במטריצה הוא $x_{1,1}$, $x_{1,2}$, $x_{1,3}$, $x_{1,4}$, $x_{2,1}$... וכו'. עלויות ההובלה נתונות בטבלאות שלעיל.

ניתן כמובן לפתור את בעית התובלה בתכנות לינארי רגיל. משום שלמטריצה יש מבנה מיוחד, ומשום שכל מקדמיה הם 1 ניתן לפתור את בעית התובלה בשיטה יעילה יותר. בשיטת פתרון זו מבחינים בשני שלבים:

(א) מציאת פתרון אפשרי ראשון

(ב) שיפור הפתרון בצורה איטרטיבית עד לקבלת הפתרון האופטימלי.

כאשר קיים $\sum_{i=1}^{n} a_i \geq \sum_{j=1}^{m} b_j$ יש תמיד פתרון אפשרי, אותו קל למצוא. השיטה הפשוטה ביותר למציאת פתרון אפשרי ראשון היא זו הנקראת "שיטת הפינה הצפון-מערבית" בה מתחילים מן הפינה השמאלית העליונה של הטבלה ומשבצים כמויות בתאים כך שיתמלאו אילוצי המקורות והיעדים. ניתן להתחיל בכל פינה אחרת של הטבלה, וצורת ההצבה אינה משתנה בעקרון. נדגים לכן כשמתחילים מהפינה הצפון-מערבית.

אם $a_1 < b_1$ מציבים $x_{1,1} = a$, מבטלים את השורה הראשונה מן הטבלה ורושמים (b_1-a_1) במקום בתחתית העמודה הראשונה. אם $b_1 < a_1$ מציבים $x_{1,1} = b_1$, מבטלים את העמודה הראשונה מן הטבלה ורושמים (a_1-b_1) במקום a_1 בקצה השורה הראשונה.

ממשיכים כעת עם הטבלה המצומצמת בדיוק באותו אופן. כלומר, בפינה השמאלית העליונה מכניסים את הערך הקטן יותר - של העמודה הראשונה או של השורה הראשונה של הטבלה המצומצמת. שוב מבטלים שורה או עמודה וכך ממשיכים עד הפינה הנגדית (הדרום-מזרחית) של הטבלה.

בשיטה זו מתקבל עבור הדוגמה הפתרון האפשרי הבא:

יעד j / מקור i	1	2	3	4	a_i
1	8	2			10
2		4	1		5
3			4	2	6
4				7	7
b_j	8	6	5	9	

בשיטה זו אין כל התחשבות בעלויות, ולכן אין להניח כי הפתרון האפשרי שנמצא קרוב לאופטימלי. שיטה אחרת למציאת פתרון אפשרי ראשון נקראת שיטת הקנסות או שיטת ווגל (Penalty Method; Vogel's Method), ששלביה כדלקמן:

(א) לכל שורה במטריצה מציינים את הערך המוחלט של ההפרש כין שני המחירים הנמוכים כיותר באותה שורה, וכנ"ל לגכי כל עמודה. ערכיה אלה מבטאים כעין קנס על אי-שימוש במחיר הנמוך ביותר.

(ב) בוחרים את השורה או העמודה בעלת הקנס הגבוה ביותר. בשורה או עמודה זו בוחרים את התא בעל המחיר הנמוך ביותר ומשכצים כו $x_{i,j} = min(a_i;b_j)$. ע"י שיבוץ זה מתמלא האילוץ של השורה או העמודה המתאימה.

(ג) מבטלים מן המטריצה את השורה או העמודה אשר בה התקיים האילוץ של a_i או b_j, תוך החסרת ה-$x_{i,j}$ ששובץ מן ה-a_i או ה-b_j שנותר בטבלה. מתקבלת מטריצה מוקטנת. אם היא מכילה רק שורה אחת או עמודה אחת משכצים כה את האברים הנותרים. בכך מסתיים התהליך ויש בידינו פתרון אפשרי. אחרת חוזרים לשלב (א) וממשיכים כתהליך.

נחזור לדוגמה. בעמורה הנוספת משמאל ובשורה הנוספת בראש הטבלה מופיעים הקנסות:

	1	1 ~~2~~	2 ~~1~~	1	
1	2	5	6	1	10
1	5	2	3	5	5
2	1	3	1	3	6
1	3	1	4	2	7
	8	6	5	9	

עמודה ~~2~~ 3 ושורה 3 הן בעלות הקנס הגבוה כיותר. נבחר, למשל, את שורה 3 ובה את התא שבעמודה 1, שכן הוא בעל המחיר הנמוך כיותר. לו כחרנו את התא בעמודה 3 בשורה זו או לו בחרנו עמורה 3 ובה את התא בשורה ~~הכבעבת~~ היתה התוצאה הסופית
הקיימת

של השיבוץ לפי שיטת הקנסות זהה - כפי שיכול הקורא להיווכח בעצמו. נציב כעת

$$x_{3,1} = min(a_3;b_1) = min(6,8) = 6$$

בכך נתמלא האילוץ של השורה השלישית ואותה נמחק מן הטבלה. בתחתית העמודה הראשונה נרשום כעת 2 = 6-8.

	1	1	1	1	
1	2	5	6	1	10
1	5	2	3	5	5
1	3	1	4	2	7
	2	6	5	9	

$x_{1,4} = min(9;10) = 9$

מכאן ואילך בקטרון הסיבוי
אם בלתי אם
כאלמנט הבא, במקום

(6,1)

בראש הטבלה ובעמודה השמאלית מופיעים כעת קנסות חדשים. נבחר לפיהם את העמודה הראשונה ובה את התא הראשון.(*) נשבץ שם

$$x_{1,1} = min(10;2) = 2$$

נבטל את העמודה הראשונה, נרשום בשורה הראשונה 8 = 2-10, ונקבל

	1	1	1	
4	5	6	1	8
1	2	3	5	5
1	1	4	2	7
	6	5	9	

$x_{1,4} = 8$

	1	1	1	
3	2	5	6	1
1	5	2	3	5
2	3	1	4	7
	2	6	5	

$x_{1,1} = 1$

(*) יכולנו לשים שם 1 ולא את התא הבא.

השלבים הנותרים:

	1	1	3	
1	2	3	5	5
1	1	4	2	7
	6	5	1	

$x_{4,4} = 1$

	2	1	1	
1	5	2	3	5
2	3	1	4	7
	1	6	5	

$x_{4,2} = 6$

	1	1		
1	2	3	5	
3	1	4	6	
	6	5		

$x_{4,2} = 6$

	2	1	
2	5	3	5
1	3	4	1
	1	5	

$x_{2,3} = 5$

$\boxed{3}$	1
1	

$x_{4,1} = 1$

השיבוץ הנותר: $x_{2,3} = 5$. הפתרון האפשרי המתקבל הוא:

מקור i \ יעד j	1	2	3	4	a_i
1	2 (1)			8 (9)	10
2			5		5
3	6				6
4	(1)	6		1 (0)	7
b_j	8	6	5	9	

בפתרון האפשרי שהתקבל יש תאים "מלאים" ו"ריקים". כהמשך עוברים לתהליך האיטרטיבי של שיפור הפתרון. בתהליך זה מכצעים העברות של כמויות כין תאים תוך שמירה על קיום האילוצים. צעדי התהליך הם:

(א) לכל שורה מציבים ערך α_i ולכל עמודה ערך β_j בשיטה הכאה:

- הצב $\alpha_i = 0$ עבור שורה כלשהי (למשל, הראשונה)
- לכל התאים <u>המלאים</u> $\alpha_i + \beta_j = c_{i,j}$

(ב) לכל תא ריק מחשבים $d_{i,j} = c_{i,j} - (\alpha_i + \beta_j)$

(ג) אם כל $d_{i,j} \geq 0$ התהליך הסתיים והפתרון אופטימלי. אם לא המשך לצעד (ד). אם יש $d_{i,j} = 0$ קיים פתרון אופטימלי אלטרנטיכי.

(ד) בחר את ה-$d_{i,j}$ הקטן כיותר. לתא זה מצא לולאה סגורה הבנויה קטעים אפקיים ואנכיים אשר בה כל "שבירה" (מאפקי לאנכי) היא בתא מלא.

(ה) בין כל התאים המרוחקים מספר אי-זוגי של "שבירות" בלולאה מן התא (i,j) שנבחר מצא זה שבו $x_{k,\ell}$ ששובץ הוא הקטן ביותר.

(ו) "העבר" את $x_{k,\ell}$ זה על פני הלולאה: חסר אותו בתא (k,ℓ) שישאר עתה ריק, חכר אותו ל-x בכל תא כ"שבירות" הראשונה, השלישית, החמישית וכו', וחסר אותו מ-x בכל תא ב"שבירות" השניה, הרביעית וכו'. לאחר העברה זו תזור לצעד (א) לעיל.

להדגמת התהליך נשתמש בפתרון האפשרי הראשון של הדוגמה אשר התקבל בשיטת הפינה הצפון-מערבית. את הטבלאות נבנה במתכונת הבאה. בכל תא יירשם $c_{i,j}$ בפינה הימנית העליונה. את $x_{i,j}$ של התאים המלאים נרשום בפינה השמאלית התחתונה של כל תא, כשהוא מוקף בעיגול. באותה פינה נרשום עבור התאים הריקים את $d_{i,j}$; את המינימלי (השלילי) שביניהם נקיף בריבוע. הלולאה הנבחרת מסומנת בקוים מתכרים. ערכי α_i כתובים משמאל לשורות וערכי β_j בראש העמורה.

β_j

α_i \ β_j	2	5	6	8
0	2 / (8)	5 / (2)	6 / 0	1 / $\boxed{-7}$
-3	5 / 6	2 / (4)	3 / (1)	5 / 0
-5	1 / 4	3 / 3	1 / (4)	3 / (2)
-6	3 / 7	1 / 2	4 / 4	2 / (7)

$\min_{i,j}(d_{ij})$

z = 61

$\boxed{c_{ij}}$

(x_{ij}) מקיים

$d_{ij} = c_{ij} - (\alpha_i + \beta_j)$: אז

לאחר הצבת $\alpha_1 = 0$ מתקבל מיד $\beta_1 = 2$ ו-$\beta_2 = -5$. מכאן $\alpha_2 = -3$, ואז $\beta_3 = 6$. הקורא יווכח כי תמיד ניתן לחשב ערכי α_i ו-β_j לפי ההנחיות שלעיל, אם כי לעיתים נותר חופש כחירה מסוים כחישוב חלק מן המקדמים. לאחר חישוב ערכי $d_{i,j}$ בתאים הריקים נמצא כי רק בתא (1,4) הערך הוא שלילי, ולכן מינימלי. הלולאה העוברת דרך תא זה, שבה ה"שבירות" האחרות כתאים מלאים מופיעה בטכלה. בין ערכי x_{ij} שבמרחק אי-זוגי של שבירות נמצא $min(2;1;2) = 1$. את הכמות הזאת "נעביר" על פני הלולאה, תוך החסרה והוספה לסרוגין.

שאר השלבים:

α_i \ β_j	2	5	-1	1
0	2 / (8)	5 / (1)	6 / 7	1 / (1)
-3	5 / 6	2 / (5)	3 / 7	5 / 7
2	1 / -3	3 / -4	1 / (5)	3 / (1)
1	3 / 0	1 / $\boxed{-5}$	4 / 4	2 / (7)

איטרציה 2

z = 54

α_i \ β_j	2	0	-1	1
0	2 / (8)	5 / 5	6 / 7	1 / (2)
2	5 / 1	2 / (5)	3 / 2	5 / 2
2	1 / $\boxed{-3}$	3 / 1	1 / (5)	3 / (1)
1	3 / 0	1 / (1)	4 / 4	2 / (6)

איטרציה 3

z = 49

	2	0	0	1
0	2 / 2	5 / 5	6 / 6	1 / 8
3	5 / 3	2 / 2	3 / 5	5 / 4
-1	1 / 6	3 / 4	1 / 2	3 / 3
1	3 / 0	1 / 6	4 / 3	2 / 1

פתרון אופטימלי

z* = 41

	2	0	2	1
0	2 / 7	5 / 5	6 / 4	1 / 3
2	5 / 1	2 / 5	3 / -1	5 / 2
-1	1 / 1	3 / 4	1 / 5	3 / 3
1	3 / 0	1 / 1	4 / 1	2 / 6

איטרציה 4

z = 46

הפתרון האופטימלי שהתקבל זהה לפתרון האפשרי הראשון של שיטת הקנסות, מה שמורה שכראי לבחור פתרון אפשרי בשיטת הקנסות. היות ובטבלה האחרונה $d_{4,4} = 0$ הרי שקיים פתרון אופטימלי אלטרנטיבי, אותו אפשר לקבל ע"י יצירת לולאה עבור $d_{4,4}$ ו"העברות" כפי שנעשה לעיל.

6.15.2 פירוק בעיות גדולות

לבעיות תיכנות ליניארי גדולות יש לעיתים קרובות צורה כדלקמן:

$$max\ z = \sum_{j=1}^{r} \underline{c}_j\ \underline{x}_j \qquad (6\text{-}101)$$

s.t.

$$\begin{bmatrix} \underline{\underline{A}}_1 & \underline{\underline{A}}_2 & \underline{\underline{A}}_3 & \cdots & \underline{\underline{A}}_r \\ \underline{\underline{D}}_1 & \underline{\underline{0}} & \underline{\underline{0}} & & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{D}}_2 & \underline{\underline{0}} & & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{D}}_3 & & \underline{\underline{0}} \\ \vdots & \vdots & \vdots & & \vdots \\ \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{0}} & & \underline{\underline{D}}_r \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \underline{x}_3 \\ \vdots \\ \underline{x}_r \end{bmatrix} = \begin{bmatrix} \underline{b}_o \\ \underline{b}_1 \\ \underline{b}_2 \\ \vdots \\ \underline{b}_r \end{bmatrix} \qquad (6\text{-}102)$$

$$\underline{x}_j \geq 0 \qquad j = 1,\ldots,r$$

כאשר $\underline{\underline{A}}_j$ ו-$\underline{\underline{D}}_j$ הן מטריצות מלבניות. בקבוצת האילוצים המתאימה למטריצה $\underline{\underline{D}}_j$ מופיעים רק המשתנים הנמצאים בקבוצה $\underline{x}_j$, ואילו באילוצים מופיעים כל המשתנים:

$$\sum_{j=1}^{r} \underline{A}_j \, \underline{x}_j = \underline{b}_o$$

$\underline{c}_j$ הוא תת-הוקטור של $\underline{c}$ המתייחס למשתנים בקבוצה $\underline{x}_j$.

קיימות שיטות אחרות ליעול הפתרון של בעיות בעלות מבנה כזה, המבוססות כולן על עיקרון הפירוק (Decomposition). לפי אחת מן השיטות יוצרים בעיה ראשית, המכילה את האילוצים הראשונים - אלו המורכבים מהמטריצות $\underline{\underline{A}}_j$ - בתוספת r אילוצים נוספים, ו-r בעיות משניות, שכל אחת מכילה את האילוצים של אחד ה-$\underline{\underline{D}}_j$ כלבד. תהליך הפתרון הוא איטרטיבי, כדלקמן: פותרים את הראשית, מפתרונה יוצרים פונקצית מטרה עבור כל אחת מהבעיות המשניות, פותרים כל אחת מהן לחוד ומפתרונן יוצרים פונקצית מטרה ואילוצים חדשים לבעיה הראשית, פותרים אותה וחוזר חלילה, עד שהתהליך מתכנס לפתרון האופטימלי. תהליך הפירוק, למרות אופיו האיטרטיבי, יעיל יותר מאשר פתרון כל הבעיה בבת-אחת. היעילות היחסית עולה במובן ככל שבבעיה הראשונה מופיעים פחות אילוצים מן הסוג הראשון.

הספר "Optimization Theory for Large Systems", Lasdon מוקדש ברובו לנושא.

6.15.3 תיכנות פריק

בעיית התיכנות הפריק (Separable Programming). כפי שהוצגה בסעיף 4.5.3 תיא:

$$max \text{ (or } min) \quad f(\underline{x}) = \sum_{j=1}^{n} f_j(x_j) \tag{6-103}$$

$$\text{s.t.} \quad g_i(\underline{x}) = \sum_{j=1}^{n} g_{ij}(x_j)\{\leq ; = ; \geq\}b_i \tag{6-104}$$

$$i = 1,\ldots,m$$

$$x_j \geq 0 \qquad j = 1,\ldots,n \tag{6-105}$$

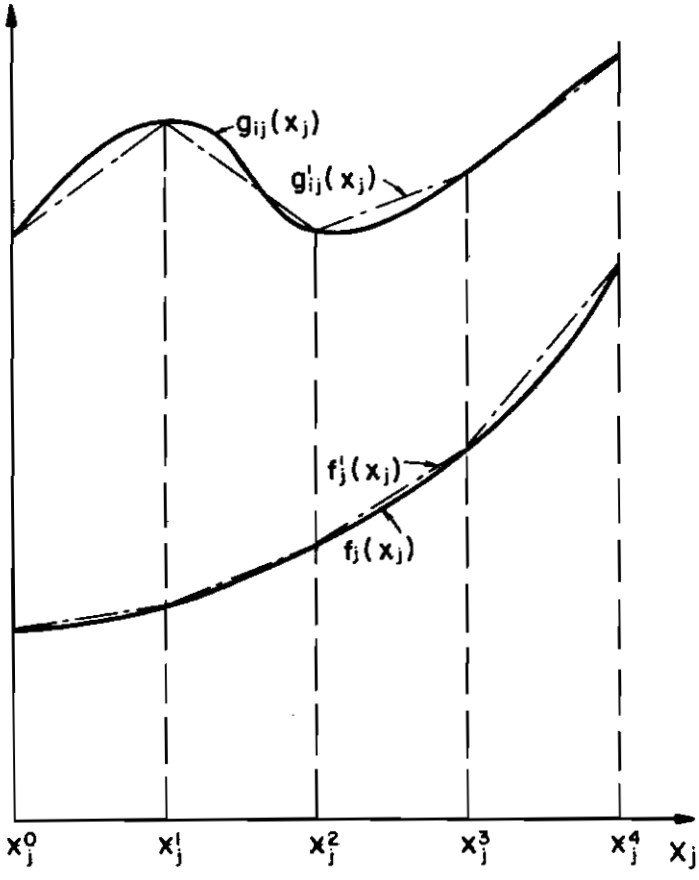
שבה הפונקציות $f_j(x_j)$ ו-$g_{ij}(x_j)$ אינן חייבות להיות ליניאריות, אך כל אחת מהם תלויה במשתנה החלטה אחד בלבד. מן הראוי לציין כי במקרים רבים לפחות חלקן של הפונקציות ליניאריות, ודבר זה מקל על הפתרון, כפי שיובן כשתוצג השיטה. שיטת הפתרון מתבססת על ליניאריזציה בקטעים של כל אחת מהפונקציות הלא ליניאריות. לנוחיות הדיון נניח כי כל הפונקציות אינן ליניאריות. שיטת הליניאריזציה תוסבר בעזרת ציור 6.3.

בציור מתוארות בקו מלא פונקציות $f_j(x_j)$ ו-$g_{ij}(x_j)$ כלשהן. בוחרים לאורך x_j נקודות חלוקה שתסומנה x_j^k. מספר הנקודות נתון לשיקול, וכן הרווחים ביניהם אינם חייבים להיות שווים. בקו מרוסק מתוארות הפונקציות המקורבות $f_j'(x_j)$ ו-$g_{ij}'(x_j)$, שהן ליניאריות בקטעים. בנקודות החלוקה שנבחרו x_j^k קיים

$$f_j'(x_j^k) = f_j(x_j^k) \quad ; \quad g_{ij}'(x_j^k) = g_{ij}(x_j^k) \tag{6-106}$$

ערך הפונקציות המקורבות בנקודות ביניים תחושבנה כדלקמן: עבור

$$x_j^k \leq x_j \leq x_j^{k+1}$$



<u>ציור 6.3</u>: ליניאריזציה בקטעים בשיטת ה-λ-יות.

אפשר לרשום

$$x_j = \lambda x_j^k + (1-\lambda)x_j^{k+1} \qquad (6\text{-}107)$$

ואז הערכים של הפונקציות המקורבות בנקודה זו הן

$$f_j'(x_j) = \lambda f_j(x_j^k) + (1-\lambda)f_j(x_j^{k+1}) \qquad (6\text{-}108)$$

$$g_{ij}'(x_j) = \lambda g_{ij}(x_j^k) + (1-\lambda)g_{ij}(x_j^{k+1}) \qquad (6\text{-}109)$$

וזאת בהסתמך על כך שהשוויונות שב-(6-106), (6-108) ו-(6-109) נכונים עבור

$$x_j^k \le x_j \le x_j^{k+1}$$

ניתן לרשום ביטוי כללי יותר עבור הפונקציות המקורבות בכל תחום הקיום של x_j כדלקמן:

$$f_j'(x_j^k) = \sum_{k=o}^{K_j} \lambda_j^k f_j(x_j^k \qquad (6\text{-}110)$$

$$g_{ij}'(x_j^k) = \sum_{k=o}^{K_j} \lambda_j^k g_{ij}(x_j^k \qquad (6\text{-}111)$$

כאשר K_j הוא מספרה של נקודת החלוקה האחרונה שנבחרה עבור המשתנה x_j ($K_j = 5$ כדוגמה של ציור 6.3). יש לדרוש קיום התנאי נוסף

$$\sum_{k=o}^{K_j} \lambda_j^k = 1 \qquad (6\text{-}112)$$

וכי לכל היותר שניים מה-λ_k יהיו שונים מאפס ושניים אלה חייבים להיות שכנים - כלומר, ערכי האינדקס k שונה ביחידה. לנוחיות ההמשך נכתוב תנאי זה בצורה הסימבולית

$$\lambda_j^k \ , \ \lambda_j^{k+1} \geq 0 \qquad (6\text{-}113)$$

שימוש כליניאריזציה זו מאפשר כתיבה מחדש של הבעיה המקורית - משוואות (6-103) עד (6-105) כדלקמן:

$$max(\text{or } min) \quad f'(x) = \sum_{j=1}^{n} \sum_{k=o}^{K_j} \lambda_j^k f_j(x_j^k) \qquad (6\text{-}114)$$

$$\text{s.t.} \quad g_i'(\underline{x}) = \sum_{j=1}^{n} \sum_{k=o}^{K_j} \lambda_j^k g_{ij}(x_j^k)\{\leq \ ; = \ ; \geq\} b_i \qquad (6\text{-}115)$$

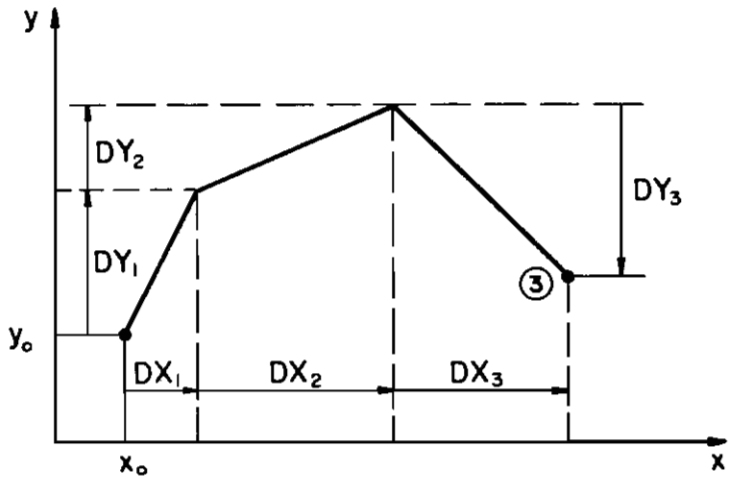
$$i = 1,\ldots,m$$

$$\sum_{k=o}^{K_j} \lambda_j^k = 1 \qquad j = 1,\ldots,n \qquad (6\text{-}116)$$

$$\lambda_j^k \ , \ \lambda_j^{k+1} \geq 0 \qquad j = 1,\ldots,n \qquad (6\text{-}117)$$

(6-117) הוא ביטוי לדרישה שלכל היותר שניים מערכי λ_j עבור כל x_j שונים מאפס ושניים אלה שכנים. הבעיה כעת ליניארית במשתנים λ_j^k בתוספת הדרישה המיוחדת (6-117). מבחינה חישובית הקושי הוא בכך שגדל מספר המשתנים. ככל שרוצים לדייק יותר בתיאור הפונקציות הלא-ליניאריות יש לקחת מספר גדול יותר של נקודות חלוקה. מספר האילוצים בבעיה החדשה הוא m - $(m+n)$ המקוריים בתוספת n האילוצים (6-116), אחד לכל x_j. מספר האילוצים לא גדל בצורה ניכרת, אבל הגדלת מספר המשתנים ותוספת הדרישה (6-177) מקשה על הפתרון.

דרך שונה לניסוח הפונקציה הלינארית בקטעים, נקראת "שיטת ה-Δ-ות" (The Delta Method); היא מוסברת באמצעות ציור 6.4.



ציור 6.4: ליניאריזציה בקטעים של פונקציות בשיטת ה-Δ-ות.

הפונקציה מתחילה בנקודה $(x_o;y_o)$. מכל נקודה מוגדרים אינקרמנטים (רלתות) לנקודה הבאה, היכולים להיות חיוביים או שליליים. נקודה כלשהי על הפונקציה הלינארית בקטעים נתונה ע"י

$$x = x_o + DX_1 * d_1 + DX_2 * d_2 + ... + DX_r * d_r \qquad (6\text{-}118)$$

כאשר x נמצא במרווח ה-i של מתקיים

$$\left.\begin{array}{l} d_1 = d_2 = ... d_{i-1} = 1 \\ \qquad 0 \le d_i \le 1 \\ d_{i+1} = d_i = ... = d_r = 0 \end{array}\right\} \qquad (6\text{-}119)$$

הערך של $y(x)$, המבטא אילוץ או פונקצית מטרה, הוא:

$$y = y_o + DY_1 * d_1 + DY_2 * d_2 + ... + DY_r * d_r \qquad (6\text{-}120)$$

עם ערכי d_i המתאימים ל-x הנרון.

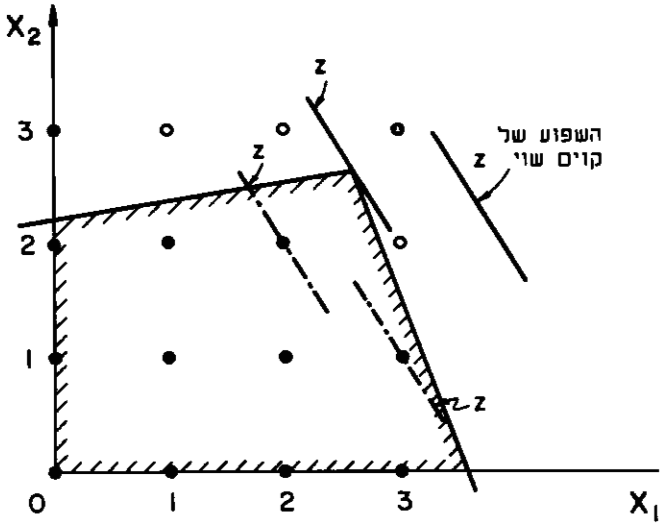
רוב חבילות התוכנה הגדולות משתמשות בשיטה זו לתכנות פריק. לשם כך מוגדרים המשתנים d_i כסדרה מיוחדת (S.O.S. = Special Ordered Set) המקיימת את התנאי (6-119) והמשתמש בונה באמצעותם את הפונקציה הלינארית בקטעים. התכנית דואגת באופן אוטומטי לקיום התנאי (6-119) עבור קבוצת המשתנים d_i שהוגדרו כסדרה מיוחדת.

אם הבעיה אינה קמורה (מינימום של פונקציה קמורה על פני תחום קמור), אזי הפתרון המתקבל עלול להיות אופטימום לוקלי בלבד, שכן הפתרון מסתיים בנקודה בה אי אפשר לשפר באופן מקומי על ידי החלפת משתנים בבסיס. אפשר עם זאת לנסות נקודות התחלה שונות עבור הפתרון, ולבחור בין הפתרונות המתקבלים את הטוב ביותר. אם הפונקציות המקורבות יוצרות תיכנות קמור, כלומר, מינימיזציה של פונקציה קמורה על פני תחום קמור, מובטחת השגת הפתרון האופטימלי הגלובלי (בתחום הדיוק של הליניאריזציה כמובן).

6.15.4 תיכנות ליניארי בשלמים

לעיתים יש בבעיות תיכנות לינארי משתני החלטה היכולים לקבל רק ערכים שלמים. מתקבל אז תיכנות ליניארי בשלמים (Integer Linear Programming). אם רק חלק מן המשתנים צריך להיות שלם מתקבל תכנות מעורב בשלמים (MIP = Mixed Integer Programming).

אם הערך שיקבל המשתנה השלם גדול מספיק, ניתן אולי להרשות לו להיות רציף במהלך הפתרון ורק בסוף "לעגל" את התוצאות לערכים השלמים הקרובים ביותר אך נמוכים מן הערכים שנתקבלו בפתרון האופטימלי. כאשר המספרים קטנים יותר ייתכן בהחלט שהפתרון המעוגל כלפי מטה אינו הפתרון האופטימלי. להדגמה ישמש ציור 6.5. בציור מתואר האיזור המוגבל על ידי שני אילוצים הליניאריים. רק הנקודות שבתוכו מהוות את האיזור האפשרי אם נדרש כי שני המשתנים יהיו שלמים. הפתרון של הבעיה הרציפה מתקבל כפינה הימנית העליונה, בחיתוך שני האילוצים וערכו z_1. היות ובנקודה זו המשתנים אינם שלמים אפשר לנסות לעגל כלפי מטה. הנקודה האפשרית הקרובה ביותר היא $(x_1 = x_2 = 2)$, וערך פונקצית המטרה בה z_2. אלא שהוחמצה בדרך זו הנקודה $(x_1 = 3 ; x_2 = 1)$ בה ערך פונקציה z^* הגדול מ-z_2. אפשר היה לנסות יותר נקודות בסביבת האופטימום הרציף אלא שבבעיה רב מימדית הדבר מורכב מאד.



ציור 6.5: בעית תיכנות ליניארי בשלמים.

השיטות המקובלות כיום לפתרון בעיות עם משתנים שלמים הן שיטות של סעוף וחיסום (Branch and Bound). מתחילים מפתרון הבעיה הרציפה, בלי אילוצי השלמות. הערך האופטימלי של פונקצית המטרה בבעיה הרציפה מהווה חסם "טוב ביותר" (חסם עליון לבעית מקסימיזציה, חסם תחתון לבעית מינימיזציה) על הפתרון, שכן כל הוספת אילוץ שלמות על משתנה רק "תקלקל" את הפתרון.

הפתרון הרציף מהווה צומת התחלתי להמשך הפתרון, כאשר כל פתרון חדש מתקבל ע"י הצבת משתנה הנדרש להיות שלם, אך אינו שלם עדיין, בערך שלם. אם משתנה כזה נמצא במרווח כלשהו בין שני ערכים שלמים שמותר לו לקבל הרי שאפשר להסתעף מן הצומת הנוכחי לשני ענפים - כל אחד ע"י הצבת המשתנה הנדון באחד מן הערכים השלמים שבקצות התחום בו הוא נמצא כרגע. חישובי עזר מאפשרים לחשב את החסם העליון ו/או התחתון על הפתרון האופטימלי העומד בדרישות השלמות אשר יכול להתקבל בהליכה מכל צומת כזה. חסמים אלה משמשים כקריטריונים לבחירת הצומת ממנו ממשיכים בסיעוף.

הכללים המשמשים בסעוף וחיסום הם בחלקם היוריסטיים, ולכן יעילות תכנית המחשב תלויה בטיבם של כללים אלה.

ראוי לדעת כי בעוד שניתן לפתור בעיות תכנות ליניארי רציף עם מאות ואלפי אילוצים ומשתנים הרי שמספר המשתנים השלמים בהם ניתן לטפל בזמן מחשב סביר אינו עולה על כמה עשרות.

6.15.5 שימושים מיוחדים של תכנות ליניארי בשלמים

משתנים שלמים משמשים גם לניסוח תנאים לוגיים בתכנות הליניארי. לשם כך מגדירים משתנים בוליאניים היכולים לקבל רק ערכים 0 או 1. יהיו I_1 ו-I_2 משתנים כאלה. הטבלה הבאה נותנת משמעותם של אילוצים שונים על הסכום/הפרש של משתנים אלה:

	אילוץ	משמעות
(6-121)	$I_1 + I_2 \le 1$	אחד מהם יכול להופיע, אך לא שניהם ביחד
(6-122)	$I_1 + I_2 = 1$	אחד מהם חייב להופיע, אך לא שניהם ביחד
(6-123)	$I_1 + I_2 \ge 1$	לפחות אחד מהם חייב להופיע
(6-124)	$I_1 + I_2 = 2$	שניהם חייבים להופיע
(6-125)	$I_1 - I_2 \le 0$	אם I_1 מופיע גם I_2 חייב להופיע
(6-126)	$I_1 - I_2 = 0$	או אף אחד או שניהם
(6-127)	$I_1 - I_2 \ge 0$	אם I_1 אינו מופיע אסור ל-I_2 להופיע

נסמן כעת ב-x_j משתנה רציף, ב-I_j משתנה לוגי (בוליאני), וב-U_j גבול עליון של המשתנה המתאים x_j. נדגים את השימוש במשתנים הלוגיים ליצירת תנאים מסוימים.

(א) x_1 ו-x_2 אינם יכולים להופיע ביחד בפתרון

$$\left.\begin{aligned} x_1 &\le I_1 U_1 \\ x_2 &\le I_2 U_2 \end{aligned}\right\} \qquad (6\text{-}128)$$

$$I_1 + I_2 \le 1 \qquad (6\text{-}129)$$

האילוץ הראשון מחייב $I_1=1$ כאשר $x_1>0$. האילוץ השני מתייב $I_2=1$ כאשר $x_2>0$. האילוץ השלישי אינו מאפשר לשני המשתנים הלוגיים להופיע ביחד בפתרון ולכן מונע מ-x_1 ו-x_2 להיות שונים מאפס בעת ובעונה אחת. הפתרון $x_1 = x_2 = 0$ אפשרי.

(ב) <u>x_1 חייב לקבל אחד מסדרת ערכים מסוימים:</u> למשל הערכים 0, A_1, A_2 או A_3.

$$x_1 - I_1A_1 - I_2A_2 - I_3A_3 = 0 \qquad (6\text{-}130)$$

$$I_1 + I_2 + I_3 \le 1 \qquad (6\text{-}131)$$

(ג) <u>מן המשתנים x_1, x_2 ו-x_3 יכולים להופיע לכל היותר שניים בפתרון</u>

$$x_1 \le I_1U_1 \qquad (6\text{-}132)$$

$$x_2 \le I_2U_2 \qquad (6\text{-}133)$$

$$x_3 \le I_3U_3 \qquad (6\text{-}134)$$

$$I_1 + I_2 + I_3 \le 2 \qquad (6\text{-}135)$$

(ד) <u>אין להכניס את x_2 לפתרון ער ש-x_1 הגיע לגבולו העליון U_1, ואין להכניס את x_3 ער ש-x_2 הגיע לגבולו העליון, U_2.</u>

$$x_1 \ge I_1U_1 \qquad (6\text{-}136)$$

$$x_2 \ge I_2U_2 \qquad (6\text{-}137)$$

$$x_2 \le I_1U_2 \qquad (6\text{-}138)$$

$$x_3 \le I_2U_3 \qquad (6\text{-}139)$$

האילוץ הראשון מכתיב $I_1=0$ כל עוד x_1 קטן מגבולו העליון, ומאפשר $I_1=1$ כאשר x_1 מגיע לגבולו העליון. כך האילוץ השני ביחס למשתנה הלוגי I_2. האילוץ השלישי מבטיח $I_1=1$ כאשר $x_2 \geq 0$. כך האילוץ הרביעי מבטיח $I_2 = 1$ כאשר $x_3 \geq 0$.

(ה) <u>עלות המשתנה x_1 נתונה ע"י $(c_0+c_1 x_1)$ (Fixed Charge Problem)</u>

$$x_1 \leq I_1 U_1 \tag{6-140}$$

ובפונקצית המטרה יופיע

$$C_0 I_1 + C_1 x_1 \tag{6-141}$$

כאשר $x_1=0$ יתקבל $I_1=0$ ואז תרומת x_1 לפונקצית המטרה מתאפסת. $x_1>0$ מכתיב $I_1=1$ ואז החלק הקבוע של תרומת x_1 לפונקצית המטרה, c_0, נוסף אף הוא.

(ו) <u>הפעלה סלקטיבית של אילוצים</u>

אנו רוצים שרק אחד מן האילוצים:

$$a_{1,1}x_1 + a_{1,2}x_2 \leq b_1 \tag{6-142}$$

$$a_{2,1}x_1 + a_{2,2}x_2 \leq b_2 \qquad \text{או} \tag{6-143}$$

יהיה פעיל. נכתוב:

$$a_{1,1}x_1 + a_{1,2}x_2 \leq b_1 + I_1 L_1 \tag{6-144}$$

$$a_{2,1}x_1 + a_{2,2}x_2 \leq b_2 + I_2 L_2 \tag{6-145}$$

$$I_1 + I_2 = 1 \tag{6-146}$$

האילוץ השלישי מכתיב שאחד מן המשתנים הלוגיים יהיה שווה ל-1 והשני יהיה שווה לאפס. L_1 ו-L_2 הם מספרים גדולים מספיק, כך שכאשר $I_1=1$ האילוץ הראשון אינו כובל עבור הערכים הרלבנטיים של המשתנים x_1 ו-x_2, וכך לגבי L באילוץ השני. אי לכך $I_1=1$ הופך את האילוץ הראשון לבלתי פעיל, והיות ובאותו זמן $I_2=0$ הרי שהאילוץ השני פעיל.

ע"י הכתבת תנאי שונה ביחס למשתנים הלוגיים יכולנו ליצור מצבים אחרים, כגון: או ששני האילוצים פעילים או אף אחד מהם.

6.15.6 תיכנות ליניארי סטוכסטי

נדירים הם המקרים בהם כל המקדמים המופיעים בבעית תיכנות ליניארי ידועים מראש בוודאות. סטיות במקדמים הטכנולוגיים יכולות לנבוע מתנודות אקראיות בתהליכי הייצור, או מהתפתחות בתהליכים אלה שאינה חזויה מראש. וקטור האילוצים המבטא את כמויות החשומות. נתון לשינויים עקב תנאי טבע אקראיים. ואילו המחירים מתבססים על תחזיות השוק, אשר ברוב המקרים אינן יכולות להיות מדוייקות. מבחני הרגישות שתוארו בסעיף 6.12 משמשים לבדיקת השפעתם של שינויים אפשריים במקדמים. מבחנים אלו יכולים לספק הנחיות על מה שיש לעשות אם מתברר כי מקדם מסוים שונה מהערך שניתו לו בניסוח המקורי. בכך אין אבל תשובה לשאלה העיקרית שהיא: כיצד יש לתכנן מראש כאשר ידוע שחלק מן המקדמים (או כולם) בבעית תיכנות ליניארי הם משתנים אקראיים בעלי פונקציות פירוס הסתברות ידועות.

נניח כי ידוע מראש מה יהיו ערכיהם של כל המקדמים בבעית התיכנות הליניארי, או כי אנו יכולים להמתין עד שיתברר מהם ערכים אלה. אזי אפשר לפתור את כעית התיכנות הליניארי הדטרמיניסטי ולקבל את הערך האופטימלי $z^*(\underline{A},\underline{b},\underline{c})$ ואת וקטור משתני ההחלטה האופטימליים $\underline{x}^*(\underline{A},\underline{b},\underline{c})$. z^* ו-$\underline{x}^*$ הם פונקציה של המשתנים האקראיים $\underline{A},\underline{b}$ ו-$\underline{c}$ ולכן הם בעצמם משתנים אקראיים. לו יכולנו לחשב בצורה מפורשת את פונקצית הפירוס של z^* ושל $\underline{x}^*$ היה הדבר מאפשר בחירה של פתרון לפי קריטריון רצוני (למשל שקלול רצוי של הממוצע והואריאנס של z^*). חישוב מפורש של פונקציות הפירוס של z^* ו-$\underline{x}^*$ המתאים לו אפשרי רק במקרים פשוטים, שאינם מספיקים לטיפול בהרבה בעיות מעשיות. אנו נותרים איפוא, עם הבעיה של קביעת הפתרון א-פריורי בלי שנוכל לחשב את פונקציות הפירוס של התוצאות.

נתאר בקצרה את העקרונות של כמה מן השיטות שפותחו לפתרון בעית התיכנות הליניארי הסטוכסטי.

א. שימוש בממוצעים: מקובל להציב עבור כל מקרם את הערך הממוצע שלו, ולפתור כאילו היתה הבעיה דטרמיניסטית. היות ושיטה זו אינה מורה על ההסתברות והגורל של הסטיות האפשריות, נהוג לבצע בריקות רגישות (סעיף 6.12) על פתרון הממוצעים ולהסיק מהן מסקנות.

ב. שימוש בערכים פסימיים: כאשר אי אפשר להסתכן במימוש תוכנית פעולה שיש לה הסתברות גבוהה לחריגה מן האילוצים או לסטיות גרולות של ערך פונקצית המטרה מזה שתוכנן נוקטים בשיטה שמרנית ומציבים ערכים "פסימיים" עבור המשתנים. ערכים אלה נבחרים כך שאמנם הפתרון האופטימלי אינו כל כך טוב כמו לפי הממוצעים למשל, אכל יש הסתברות נמוכה יותר לחריגה מהאילוצים ולסטיות גרולות מערך פונקצית המטרה המתוכנן.

ג. אילוצים הסתברותיים: שיטה זו (CCLP=Chance Constrained Linear Programming)

מטפלת בבעיות כהן וקטור האילוצים $\underline{b}$ אקראי. הבעיה הנפתרת היא:

$$\left.\begin{array}{ll} max \text{ (or } min) & z = \underline{c}\,\underline{x} \\ \text{s.t.} & P_r[\underline{\underline{A}}x \le \underline{b}] \ge \alpha \\ & \underline{x} \ge \underline{0} \end{array}\right\} \qquad (6\text{-}147)$$

כלומר, הפתרון המתקבל מבטיח כי האילוצים יתקיימו לפחות כהסתברות נתונה α.

את האילוצים ההסתברותיים מחליפים באילוצים דטרמיניסטיים, ע"י מציאת ערכי b_i' כך שמתקיים

$$Pr[\sum_{j=1}^{n} a_{ij}x_j \le b_i'] = \alpha \qquad \forall i \qquad (6\text{-}148)$$

ופתרון הבעיה עם האילוצים הדטרמיניסטיים

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i' \qquad \forall i \qquad (6\text{-}149)$$

ד. <u>פתרון בשני שלבים:</u> גם שיטה זו מטפלת בוקטור $\underline{b}$ אקראי ופוחרת את הבעיה

$$\left.\begin{array}{ll} max \ \ (\text{or } min) & z = \underline{c}\ \underline{x} + \underline{d}\ \underline{y} \\ \text{s.t.} & \underline{\underline{A}}\ \underline{x} + \underline{\underline{D}}\ \underline{y} = \underline{b} \\ & \underline{x} \geq \underline{0}\ ,\ \underline{y} > \underline{0} \end{array}\right\} \qquad (6\text{-}150)$$

כאשר בשלב ראשון בוחרים את הוקטור $\underline{x}$ ומחכים עד שיתברר ערכו של $\underline{b}$. אז מחליטים על $\underline{y}$ כך שיתקיימו האילוצים ויתקבל אופטימום של z. מניחים שעבור כל בחירה של $\underline{x}$ בשלב הראשון אפשר למצוא $\underline{y}$ אפשרי.

ה. <u>מינימיזציה של הואריאנס:</u> כאשר לוקטור המחירים $\underline{c}$ פירוס נורמלי רב-משתני עם ממוצעים $\underline{m}$ ומטריצת קו-וואריאנס $\underline{\underline{V}}$ אזי ל-z* יש פירוס נורמלי עם ממוצע $\underline{m}\ \underline{x}$ וואריאנס $\underline{x}^T\underline{\underline{V}}\underline{x}$. במקרה זה אפשר ליצור בעית תיכנות ריבועי כדלקמן

$$\left.\begin{array}{ll} min & V[z] = \underline{x}^T\ \underline{\underline{V}}\ \underline{x} \\ \text{s.t.} & \underline{\underline{A}}\ \underline{x} \leq \underline{b} \\ & \underline{x} \ \geq \ 0 \\ & \underline{m}\ \underline{x} \geq z_o \end{array}\right\} \qquad (6\text{-}151)$$

המבטיחה לפחות הכנסה של z_o תוך מינימיזציה של הואריאנס של z.

ו. <u>שיטת נקודת החלוקה ה-α:</u> עבור המקרה שהוגדר ב-(ה') לעיל אפשר ליצור בעיח תיכנות מתימטי כדלקמן:

$$\left.\begin{array}{ll} max \ \ (\text{or } min) & F_\alpha(z = \underline{c}\ \underline{x}) \\ \text{s.t.} & \underline{\underline{A}}\ \underline{x} \leq \underline{b} \\ & \underline{x} \ \geq \ \underline{0} \end{array}\right\} \qquad (6\text{-}121)$$

כאשר F_α היא הנקודה z_o על פונקציית הפירוס של z המקיימת

$$P[z \le z_o] = \alpha$$

קיימות שיטות נוספות, אך כולן - גם אלה שהוצגו, מוגבלות רק למקרים מיוחדים של הבעיה הכללית של תיכנות ליניארי סטוכסטי, ורובן כרוך בקשיים חישוביים ניכרים.

מקורות

Dantzig, G.B., "Linear Programming and Extensions", Princeton University Press, 1963.

Ford, L.R. and Fulkerson, D.R., "Flows in Networks" Princeton University Press, 1962.

Gass, S.I., "Linear Programming", McGraw-Hill, 1958.

Hadley, G., "Linear Programming", Addison-Wesley, 1962.

Hadley, G., "Nonlinear and Dynamic Programming", Addison-Wesley, 1964.

Lasdon, L., "Optimization Theory for Large Systems" MacMillan Co., 1970.

צדר, א., "תורת הרשתות ותהליכים דינמיים", הוצאת דקל, 1978.

פרק שביעי

תכנות דינמי

7.1 מבוא

תכנות דינמי (Dynamic Programming) הוא שיטת אופטימיזציה המתאימה לפתרון בעיות בעלות מבנה מיוחד, אשר ניתן לראותו כמבנה סדרתי או רב-שלבי. השיטה מתבססת על עקרון יסודי אחד הנקרא עקרון האופטימליות (Principle of Optimality). העקרון מיוחס לאבי השיטה, ריצ'רד בלמן (Richard Bellman). שני ניסוחים אלטרנטיביים של העקרון הם :

(א) "סדרת החלטות אופטימלית בתהליך החלטה רב-שלבי היא בעלת התכונה שעבור החלטה כלשהי בשלב הראשון מהוות שאר ההחלטות סדרת החלטות אופטימליות ביחס למצב שנוצר כתוצאה מן ההחלטה הראשונה".

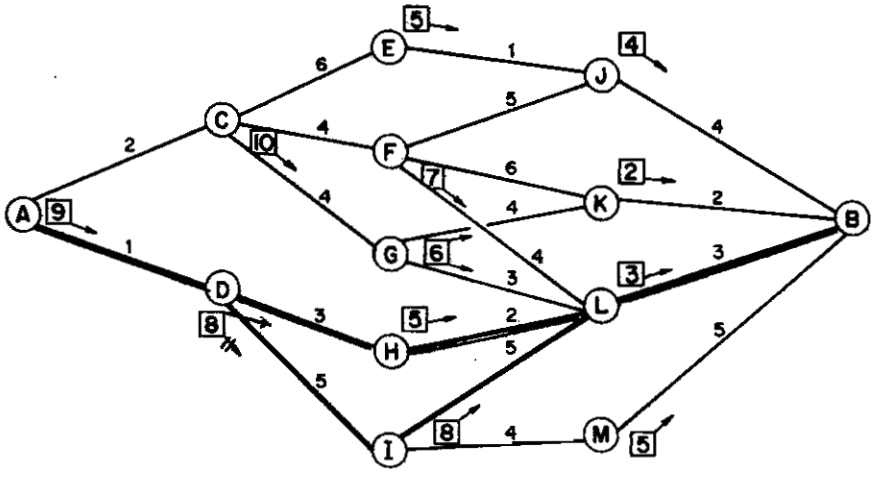
(ב) "סדרת החלטות אופטימלית כתהליך החלטה רב-שלבי היא בעלת התכונה שעבור מצב תחילי כלשהו ועבור החלטה כלשהי בשלב זה שאר ההחלטות מהוות סדרת החלטות אופטימלית עבור הבעיה הנותרת, אשר בה השלב והמצב הנובעים מן ההחלטה הראשונה מהווים תנאים תחיליים".

מאוחר יותר נגדיר ונסביר את המושגים שלב, מצב והחלטה ; כאן אנו משתמשים בהם במובנם האינטואיטיבי הפשוט. ההגדרות שלעיל של סדרת החלטות אופטימלית נראות פשטניות; לאחר הדגמה והסבר יתברר כי נובעת מהן שיטת אופטימיזציה.

תכנות דינמי הוא בסופו של דבר שיטה פשוטה למדי. ועם זאת, מראה הנסיון כי ניתן להבין את יישומה הלכה למעשה רק לאחר תרגול מספיק. הצגת הנושא תיעשה כאן בהסתמך על דוגמאות פשוטות, ולאחר מכן יינתן הניסוח המתמטי.

7.2 דוגמה : מעבר דרך רשת

בציור 7.1 נראית רשת של דרכים. ליד כל קטע רשום הזמן (בשעות) הדרוש לעבור את הקטע מצומת לצומת. אדם הנמצא ב- A רוצה להגיע ל- B בזמן הקצר ביותר. בכדי לפשט את הניתוח נניח כי הוא חייב לנוע תמיד משמאל לימין.



ציור 7.1: מסלול אופטימלי דרך רשת מ- A ל- B.

הפתרון כנוי על השיקול האינטואיטיבי הבא. לו ידענו מהו ערך המסלול האופטימלי (זמן המעבר) מ- C ל- B ומ- D ל- B לא היה כל קושי לפתוד את הבעיה, שכן לכל ערך כזה היינו מוסיפים את הזמן הנדרש להגיע לצומת הנדון מ-A , היינו משוים את שתי התוצאות ובוחרים את הקטנה ביניהן. באופן דומה, מציאת המסלול האופטימלי מ- C ל- B קלה ביותר, אם יש בידינו ערך המסלול האופטימלי ל- B מ-E, F ו-G . ואילו מכל אחת מאלה היינו מוצאים את המסלול האופטימלי בקלות לו היה בידינו המסלול האופטימלי מן הנקודות J , K ו- L ל- B. כאן הגענו לבעיה טריביאלית, שכן מכל צומת יש רק דרך אחת להגיע ל- B , והיא לכן האופטימלית. אנחנו מתחילים לכן מן הצמתים הקרובים ביותר ל-B , הם J עד M , ורושמים ליד כל אחד (בתוך הריבוע) את ערך המסלול האופטימלי ממנו ל-B . לו היתה יותר מדרך

אחת מכל אחד מן הצמתים האלה ל-B היה עלינו לבחור את הקצרה ביותר ולרשום את הקצרה ביותר ולרשום את הזמן שלה ליד הצומת. לאחר שסיימנו עם הצמתים J עד M אנו "נסוגים" לקבוצה E עד I. עבור כל צומת אנחנו בוחנים את כל הדרכים היוצאות ממנו, ולכל אחת אנו מחברים את הזמן של הקטע ואת הערך הרשום ליד הנקודה שאליה מגיעים. למשל, עבור הנקודה F אנחנו משוים בין האפשרויות ללכת ל- J , K ו- L. התוצאה הקטנה כיותר היא בהליכה לכיוון L , וערכה 7. ערך זה נרשם (בריבוע) ליד הצומת F.

עלינו לבצע אותו חישוב עבור כל אחד מהצמתים E עד I, שכן אנו יודעים שנהיה חייבים לעבור דרך אחד מהם אך איננו יודעים דרך איזה מהם. אי לכך עבור הצומת ענינו על השאלה : אם יעבור המסלול האופטימלי דרך F אז החלק מ-F ל- B יקח 7 שעות והוא יחל מ- F ל- L.

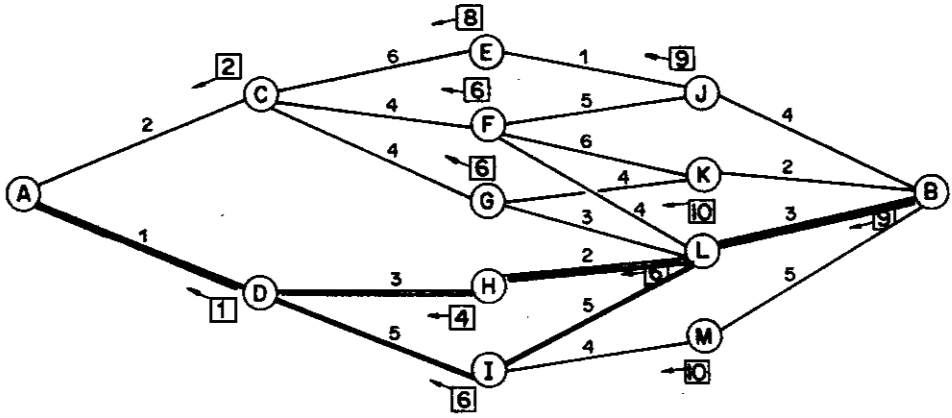
לאחר שסיימנו את הטיפול בקבוצת הצמתים הזו אנו שוב "נסוגים" שלב אחר, ומבצעים אותו חישוב עבור הצמתים C ו- D. למשל ב-C אנחנו משוים (6+5),(4+7) ו-(4+6),ומוצאים שהסכום האחרון הוא הקטן כיותר ולכן אם המסלול האופטימלי יעבור דרך C הוא ימשיך ל- G, ומ- C ל- B יקח 10 שעות.

ב-A אנחנו משוים (2+10) עם (1+8) ומוצאים שזמן המסלול האופטימלי הוא 9, ותחילתו מ- A ל-D . כעת אנחנו עוקבים אחר המסלול האופטימלי, תוך ליווי החיצים מצומת לצומת, ומקבלים A-D-~~Z~~-L-B.
H

מספר המסלולים האפשריים ברשת זו הוא 9. למראית עין בדקנו את כל המסלולים האפשריים על מנת למצוא את האופטימלי, אך למעשה אין הרבר כך. למשל, לאחר שסיימנו את הטיפול בצומת F הרי שכל מסלול שיעבור דרך F ימשיך משם דרך L. אי לכך, כאשר אנו מטפלים בצומת C אין אנו כוחנים כלל את המסלול C-F-J-B אשר אינו אופטימלי מ-F ל-B. בדוגמה שלפנינו כמות החישוכים בסריקה מקיפה של כל 9 המסלולים האפשריים לא היתה אולי גדולה בהרכה מזו של שיטת האופטימיזציה שתארנו, אכל זאת רק משום שהדוגמה כה פשוטה. כבעיות מורכבות יותר משיגה שיטת האופטימיזציה תסכון של סדרי גודל בכמות החישוב יחסית לסריקה מקיפה של כל האפשרויות.

נשים לב כי על מנת למצוא את המסלול האופטימלי מ- B ל- A מצאנו את המסלולים האופטימליים מכל צומת של הרשת ל- B, זאת למרות שלא נתבקשנו לכך במפורש.

את המסלול האופטימלי בין A ו- B יכולנו למצוא גם על ידי הליכה בכיוון הפוך, שכן מ- B ל- A המסלול האופטימלי יהיה זהה. חישוב זה הודגם בציור 7.2.



ציור 7.2: מסלול אופטימלי דרך רשת מ- B ל- A.

כמו קודם, המסלול האופטימלי ערכו 9 והוא עובר דרך אותם צמתים. אלא שהפעם מצאנו את המסלולים האופטימליים מכל צומת ל- A (או מ- A לכל צומת, כרצוננו) ולא מכל צומת ל-B .

נשים לב כי בעת הטיפול בצומת מסוים אנחנו משתמשים רק באינפורמציה "מקומית" : הזמן של הקטע הבא ועוד הזמן האופטימלי לשארית הדרך מן הצומת שאליו נגיע.

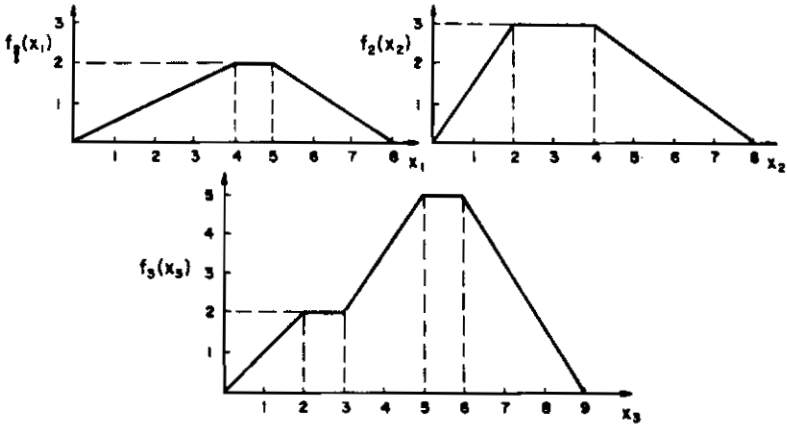
קבוצת הנקודות שעל קו אנכי ברשת מהוות שלב (Stage). למשל, הנקודות E, F, G , H ו- I מהוות שלב. את בעית המעבר דרך הרשת ניתן לנסח כך: יש למצוא בכל שלב את הנקודה שדרכה עובר המסלול האופטימלי. במסלול יש ארבעה קטעים, ולכן

עלינו לקבל ארבע החלטות - בכל מעבר משלב לשלב יש לבחור את הקטע שעליו עוברים. מקום הנקורות על כל שלב נקרא מצב (State).

7.3 דוגמה: הקצאת שעות עבודה למספר פעולות

לרשותנו מספר מוגבל של שעות עבודה, אותן יש לחלק בין שלוש פעולות. ההכנסה הנקייה מכל פעולה, f_i , היא פונקציה של מספר שעות העבודה המושקעות בה, x_i, והיא כלתי תלויה במספר שעות העבודה המושקעות בפעולות האחרות.

פונקציות ההכנסה הנקיה, $f_i(x_i)$, מתוארות בציור 7.3.



ציור 7.3: ההכנסות הנקיות מכל פעולה כפונקציה של שעות העבודה המושקעות.

הפונקציות נתונות ביחידות כלשהן; כאן x_i בשעות ו- f_i במאות ₪ .

הסתכלות בפונקציה $f_1(x_1)$ מורה שלו היתה זו הפעולה היחידה לביצוע, ולו עמדו לרשותנו, למשל, 10 שעות עבודה לא היינו מקצים לפעולה מס' 1 יותר מ-4 שעות, שכן לאחר מכן קטנה ההכנסה הנקייה. (הסבר : העלות של שעות העבודה הנוספות עולה על תוספת ההכנסה הנובעת מהן). באופן דומה, לא נקצה לפעולה מס' 2 יותר מ-2 שעות, ולפעולה מס' 3 יותר מ-5 שעות. אי לכך, אם עומדות לרשותנו בסה"כ 11 שעות עבודה או יותר הפתרון ברור : נקצה 11 שעות - 4 לפעולה מס' 1, 2 לפעולה מס' 2 ו-5 לפעולה מס' 3. ההכנסה הנקייה הכוללת תהיה אז 10.

אך ביצר נקצה את השעות אם עומדות לרשותנו פחות מ-11 שעות. בהסתכלות נראה שאת שעת העבודה הראשונה העומדת לרשותנו נקצה לפעולה מס' 2, שכן שם השיפוע של הפונקציה הוא הגדול מבין כל הפונקציות - הכנסה נקיה של 1.5 יחידות לשעת העבודה הראשונה, לעומת 0.5 בפעילות מס' 1, ו-1 בפעילות מס' 3. גם את השעה השניה העומדת לרשותנו נקצה לפעולה מס' 2. את השעה השלישית נקצה, לפי אותו שיקול, לפעולה מס' 3, וכך גם את השעה הרביעית. יכולנו להמשיך ולפתור את הבעיה בדרך פשוטה זו, אם כי בהמשך מסתבכים השיקולים במקצת בגלל החלק האפקי של $f_3(x_3)$ בקטע $x_3 = 2$ עד $x_3 = 3$. האפשרות לפתור בצורה כה פשוטה נובעת, כמובן, מצורת הפונקציות שבדוגמה. במקרה הכללי, של מספר רב של פעולות ופונקציות כלשהן היה פתרון אינטואיטיבי כזה קשה יותר או אף בלתי אפשרי.

ניתן לראות את הבעיה שלפנינו כבעית החלטה סדרתית : מתוך סה"כ השעות שלרשותנו נקצה ראשית לפעולה אחת, מן השעות הנותרות נקצה לפעולה שניה, ומן הכמות שנותרה לנו לאחר שתי הקצאות אלה נקצה לפעולה השלישית. היות וסדר ההקצאה איננו חשוב, שכן כל הפעולות שוות במעמדן, נבחר את הסדר הנתון : 1, 2, 3.

ברור שברצוננו להקצות כך שסה"כ ההכנסה הנקיה מכל הפעולות תהיה מקסימלית, ולכן חייבות ההקצאות לפעולות השונות להיות תלויות זו בזו. אי לכך נפעל בדרך הבאה. נכין עצמנו לענות על השאלה "אם תעמודנה לרשות פעולה מס' 1 s_1 שעות, מהי ההקצאה האופטימלית, $x_1^*(s_1)$, של שעות עבודה לפעולה זו, ומהי ההכנסה הנקיה הצפויה, $f_1^*(s_1)$". נשים לב כי f_1^* היא פונקציה של כמות השעות הנותרת העומדת לרשותנו לאחר הקצאה לשאר שתי הפעולות. הערך s_1 הוא פרמטר, המתאר את מצב המערכת בהגיענו לטפל בפעולה מס' 1, אותו צריך לשנות בתחום 0 עד 11, ולקבל את

הפונקציות $f_1^*(s_1)$ ו- $x_1^*(s_1)$.

כדוגמה הפשוטה שלפנינו ברור כי

$$x_1^*(s_1) = s_1 \text{ and } f_1^*(s_1) = f_1(s_1) \text{ for } s_1 \leq 4$$

$$x_1(s_1) = 4 \text{ and } f_1^*(s_1) = 2 \qquad \text{for } s_1 > 4 \tag{7-1}$$

כלומר, כל עוד עומדות לרשותנו עד 4 שעות עבודה נקצה את כולן לפעולה מס' 1 ונקבל הכנסה נקיה הנתונה ע"י הפונקציה $f_1(s_1)$, ואילו כאשר עומדות לרשותנו יותר מ-4 שעות עבודה נקצה רק 4 (והשאר נשארות ללא שימוש ואינן מהוות לגכינו לא רווח ולא הפסד) וההכנסה הנקיה תישאר 2.

נעבור כעת לטפל בצירוף של פעולות מס' 1 ו-2. שוב נשאל : "אם תעמודנה לרשות שתי הפעולות s_2 שעות עבורה, כיצד יש לחלקן כין שתי הפעולות כך שההכנסה הנקיה תהיה מקסימלית". s_2 הוא שוב פרמטר, שאותו יש לשנות באותו תחום של 0 עד 11. את ההקצאה האופטימלית כין שתי הפעולות, עכור כל ערך אפשרי של s_2, נקבל כאמצעות המשוואה הכאה :

$$f_2^*(s_2) = \max_{0 \leq x_2 \leq s_2} \{f_2(x_2) + f_1^*(s_2 - x_2)\} \text{ for } 0 \leq s_2 \leq 11 \tag{7-2}$$

עכור ערך מסוים של s_2 כתחום הנתון יש לאתר על פני כל ההקצאות האפשריות לפעולה מס' 2, כלומר $0 \leq x_2 \leq s_2$, את אותה הקצאה המביאה למקסימום את הסכום שכסוגריים. האבר הראשון בסכום הוא ההכנסה הנקיה מהקצאת x_2 לפעולה 2, והאכר השני הוא ההכנסה הנקיה האופטימלית מהכמות שתיוותר לאחר הקצאה זו, כלומר ההכנסה האופטימלית מפעולה מס' 1 כפונקציה של הכמות הנותרת $(s_2 - x_2)$. את הפונקציה הזו הכינונו כשלב הקודם, עכור כל ערך אפשרי של הארגומנט שלה, ולכן נוכל להשתמש כה כאן.

ההחלטה האופטימלית עבור כל ערך של s_2 תסומן $x_2^*(s_2)$, והערך האופטימלי הוא $f_2^*(s_2)$. נשים לכ ששני אלה הם פונקציה של s_2. הם מהווים תשובה לשאלה :

"אם תעמוד לרשות פעולות 1 ו-2 במות s_2 , אז תהיה ההכנסה הנקיה $f_2^*(s_2)$ וההקצאה האופטימלית לפעולה 2 תהיה $x_2^*(s_2)$." בכדי לקבל את ההקצאה האופטימלית לפעולה מס' 1 עלינו לחזור ל-(7-1) ולאתר שם את $x_1^*(s_2 - x_2^*(s_2))$.

לרשות שלוש הפעולות ביחד עומדת כמות שתסומן s_3. עבור ערכים שונים של כמות זו בתחום הרלבנטי נרשום את המשוואה שתיתן את ההקצאה האופטימלית לפעולה 3 כדלקמן:

$$f_3^*(s_3) = \max_{0 \le x_3 \le s_3} \{f_3(x_3) + f_2^*(s_3 - x_3)\} \text{ for } 0 \le s_3 \le 11 \qquad (7\text{-}3)$$

המשוואה הכללית עבור הפעולה ה-k תרשם :

$$f_k^*(s_k) = \max_{0 \le x_k \le s_k} \{f_k(\overset{\chi_k}{\cancel{s_k}}) + f^*_{k-1}(s_k - x_k)\} \text{ for } 0 \le s_k \le smax \qquad (7\text{-}4)$$

את החישובים נבצע בטבלאות. עבור $f_1^*(s_1)$ יש לנו את משוואה (7-1) ואין צורך בטבלה. החישובים של משוואה (7-2) מופיעים בטבלה 7.1, עבור $s_2 \le 6$, שבן עבור ערכים גבוהים יותר התוצאה תישאר בעינה.

בטבלה 7.2 מופיעים לדוגמה החישובים של משוואה (7-3) עבור הערכים $s_3 = 0,1,2,3,4,5,10,11$.

בשתי הטבלאות מסומנים הערכים של $f_k^*(s_k)$ ב-(*). למשל : $f_3^*(s_3 = 10) = 9.5$ וההקצאה האופטימלית היא $f_3^*(s_3 = 10) = 5$. אי לכך $s_2 = 10 - 5 = 5$, ומטבלה 7.1 ניתן לראות שעבור s_2 ההקצאה האופטימלית היא $x_2^*(s_2 = 5) = 2$. השארית היא $s_1 = 5 - 2 = 3$ וממשוואה (7-1) מתקבל $x_1^*(s_1 = 3) = 3$.

<u>טבלה 7.1</u>: הקצאה אופטימלית לפעולות מס' 1 ו-2.

s_2	x_2	s_2-x_2	$f_2(x_2)$	$f_1^*(s_2-x_2)$	$f_2+f_1^*$
0	0	0	0	0	0
1	0	1	0	0.5	0.5
	1	0	1.5	0	1.5*
2	0	2	0	1	1
	1	1	1.5	0.5	2
	2	0	3	0	3 *
3	0	3	0	1.5	1.5
	1	2	1.5	1	2.5
	2	1	3	0.5	3.5*
	3	0	3	0	3
4	0	4	0	2	2
	1	3	1.5	1.5	3
	2	2	3	1	4 *
	3	1	3	0.5	3.5
	4	0	3	0	3

s_2	x_2	s_2-x_2	$f_2(x_2)$	$f_1^*(s_2-x_2)$	$f_2+f_1^*$
5	0	5	0	2	2
	1	4	1.5	2	3.5
	2	3	3	1.5	4.5*
	3	2	3	1	4
	4	1	3	0.5	3.5
	5	0	2.25	0	2.25
6	0	6	0	2	2
	1	5	1.5	2	3.5
	2	4	3	2	5 *
	3	3	3	1.5	4.5
	4	2	3	1	4
	5	1	2.25	0.5	2.75
	6	0	1.5	0	1.5
>6	-	-	3	2	5 *

טבלה 7.2: הקצאה אופטימלית לשלוש הפעולות לדוגמה עבור מספר ערכים של s_3.

s_3	x_3	s_3-x_3	$f_3(x_3)$	$f_2^*(s_3-x_3)$	$f_3+f_2^*$
0	0	0	0	0	0
1	0	1	0	1.5	1.5*
	1	0	1	0	1
2	0	2	0	3	3 *
	1	1	1	1.5	2.5
	2	0	2	0	2
3	0	3	0	3.5	3.5
	1	2	1	3	4 *
	2	1	2	1.5	3.5
	3	0	2	0	2
4	0	4	0	4	4
	1	3	1	3.5	4.5
	2	2	2	3	5 *
	3	1	2	1.5	3.5
	4	0	3.5	0	3.5
5	0	5	0	4.5	4.5
	1	4	1	4	5
	2	3	2	3.5	5.5*
	3	2	2	3	5
	4	1	3.5	1.5	5
	5	0	5	0	5

s_3	x_3	s_3-x_3	$f_3(x_3)$	$f_2^*(s_3-x_3)$	$f_3+f_2^*$
10	0	10	0	5	5
	1	9	1	5	6
	2	8	2	5	7
	3	7	2	5	7
	4	6	3.5	5	8.5
	5	5	5	4.5	9.5*
	6	4	5	4	9
	7	3	3.33	3.5	6.85
	8	2	1.67	3	4.67
	9	1	0	1.5	1.5
	10	0	0	0	0
11	0	11	0	5	5
	1	10	1	5	6
	2	9	2	5	7
	3	8	2	5	7
	4	7	3.5	5	8.5
	5	6	5	5	10 *
	6	5	5	4.5	9.5
	7	4	3.33	4	7.33
	8	3	1.67	3.5	5.17
	9	2	0	3	3
	10	1	0	1.5	1.5
	11	0	0	0	0

7.4 ניסוח מתמטי

נעבור כעת לניסוח כללי של שיטת התכנות הדינמי. אנו דנים במערכת הניתנת לתיאור על ידי וקטור מצב (State Vector)

$$\underline{s} = (s_1 \ ,\ldots,s_m) \qquad (7\text{-}5)$$

בעל m רכיבים. הרכיבים השונים אינם חייבים להיות בעלי אותו מימד, והם יכולים, כבעיה מסוימת, לכלול למשל כמות מים כמאגר, מליחות המים, סכום הכסף המוקצב לחלק מסוים של פרויקט, ריכוז הבקטריות בשפכים בכניסה לתהליך טיפול מסוים או ביציאה ממנו וכדומה. כדוגמת הרשת שהוצגה בסעיף 7.2 היה המצב מקום הנקודה במימד האנכי. בדוגמה של הקצאת שעות העבודה כסעיף 7.3 היה המצב כמות שעות העבודה שנותרו לשאר הפעולות.

בבעיות בהן מטפל התכנות הדינמי אנו מבחינים כשלבים (Stages). השלב יכול לציין זמן מסוים, כאשר דנים בתהליך המשתנה בזמן השלב יכול גם לציין נקודה במערכת הפיזית, כמו למשל נקודות הצומת כרשת של סעיף 7.2, או נקודות כניסה ויציאה ממתקני טיהור שפכים, קואורדינטות של נקודות על תוואי של צינור או כביש וכדומה. כמקרים אחרים מבטא השלב אמצעי להפרדה כין ההחלטות על משתני החלטה שונים, כמו בדוגמה של הקצאת שעות עבודה לפעולות שונות. אמצעי זה, מלאכותי ככל שהוא נראה, מאפשר בתנאים מסוימים להחליף את הבעיה של כחירת כל משתני ההחלטה כיחד בסדרת בעיות קטנות יותר.

מצב המערכת כשלב ה- k יסומן ע"י הוקטור

$$\underline{s}(k) = [s_1(k),\ldots,s_m(k)] \quad k = 0 \ ,\ldots,K \qquad (7\text{-}6)$$

כאשר יש סה"כ (K+1) ~~מצבים של המערכת~~ בתים. $\underline{s}(0)$ נקרא המצב התחילי של המערכת ו-$\underline{s}(K)$ המצב הסופי. וקטור המצב של המערכת חייב לקיים אילוצים נתונים, הרשומים בצורה

$$s(k) \in \underline{S}(k) \qquad k = 0,\ldots,K \qquad (7\text{-}7)$$

כאשר $\underline{S}(k)$ היא קבוצת וקטורי המצבים האפשריים בשלב ה-k.

בכל אחד מהשלכים מוגדר <u>וקטור משתני החלטה</u> (הנקרא לעיתים <u>וקטור הבקרה</u>)

$$\underline{d}(k) = [d\ (k),\ldots,d_n(k)]\quad k = 0,\ldots,(K-1) \tag{7-8}$$

החייב לקיים את האילוצים

$$d(k) \in \underline{D}[\underline{s}(k),k] \qquad k = 0,\ldots,(K-1) \tag{7-9}$$

כאשר $\underline{D}[\underline{s}(k),k]$ הוא קבוצת וקטורי משתני ההחלטה אפשריים כשלב ה-k, התלוייה בדרך כלל לא רק בשלב אלא גם במצב $\underline{s}(k)$.

מעבר המערכת מהמצב $\underline{s}(k)$ בשלב ה-k למצב $\underline{s}(k+1)$ בשלב הבא מתואר על ידי <u>משוואת המעבר</u>, או <u>ההתמרה</u> (Transformation ;Law of Motion) :

$$\hat{\underline{s}}(k+1) = T_k[\hat{\underline{s}}(k)\ ;\ \underline{d}(k)]\quad k = 0,\ldots,K-1 \tag{7-10}$$

לכל אחר מן השלכים מוגדרת פונקצית תמורה

$$f_k[\hat{\underline{s}}(k)\ ;\ \underline{d}(k)] \qquad k = 0,\ldots,(K-1) \tag{7-11}$$

התלויה במצב וכהחלטה כאחת. פונקצית המטרה הכוללת, של כל השלבים, היא פריקה ובעלת הצורה

$$F = \sum_{k=0}^{K-1} f_k[\underline{s}(k)\ ;\ \underline{d}(k)] + f_K[\underline{s}(K)] \tag{7-12}$$

כאשר האיבר הנוסף הוא תרומת המצב הסופי, אשר ממנו אין מעברים נוספים, לפונקצית המטרה.

בעית האופטימיזציה אותה מיועד התכנות הדינמי לפתור היא

$$\underset{\underline{d}(k)}{ext} \quad \{ F = \sum_{k=0}^{K-1} f_k[\underline{s}(k) \; ; \; \underline{d}(k)] + f_K[\underline{s}(K)] \} \tag{7-13}$$

$$\text{subject to} \qquad \underline{s}(k) \in \underline{S}(k) \qquad k = 1,\ldots,K \tag{7-14}$$

$$\underline{d}(k) \in \underline{D}[\underline{s}(k),k] \quad k = 0,\ldots,K-1 \tag{7-15}$$

כאשר, לשם הקיצור, נכתוב לעיתים בהמשך פשוט $\underline{S}$ ו-$\underline{D}$ באילוצים, בהבנה שהכוונה לאילוצים כפי שנכתבו ב-(7-14) ו-(7-15).

אנו נקרא <u>מסלול</u> לשורה של מצבים אפשריים עוקבים. המטרה היא, אם כן, למצוא את המסלול שעבורו פונקצית המטרה F מקבלת ערך אקסטרמלי.

בבעיות בהן משתני המצב ו/או משתני ההחלטה הם רציפים יש צורך לקבוע עבור כל משתנה כזה עדכים בדידים שעבורם יבוצעו החישובים. בקביעת צפיפות רשת החלוקה לערבים בדידים יש לאזן בין שיקולים של ריוק לבין עלות החישוב העולה עם מספר נקודות החלוקה.

7.5 <u>פתרון על ידי חיפוש מקיף</u>

לפני שנעבור לניסוח שיטת הפתרון בתכנות דינמי נבחן פתרון באמצעות <u>חיפוש מקיף</u> (Total Enumeration), שהוא השיטה של בדיקת <u>כל</u> הפתרונות האפשריים. ראינו לדוגמה בסעיף 7.2 את האפשרות של חישוב כל המסלולים האפשריים מ-A ל-B ובחירת המסלול שנותן את האורך המינימלי. רק בבעיות קטנות יחסית ניתן הדבר לביצוע מבחינה מעשית.

אם בוחרים בשיטה של בדיקת כל המסלולים האפשריים יש לנקוט בצעדים הבאים :

(1) חשב $\underline{S}^i(1) = T_o[\underline{s}(0) \; , \; \underline{d}^i(0)]$

עבור $\underline{d}^i(0) \in \underline{D}(\underline{s}(0),0)$.

(2) בדוק כל $\underline{s}^i(1)$ המתקבל. אם הוא אפשרי, כלומר $\underline{s}^i(1) \in \underline{S}$ עבור ל-(3). אם לאו, עבור לבדיקת $\underline{s}^i(1)$ הבא, או, אם נבדקו כולם עבור ל-(4).

(3) חשב $g[\underline{s}^i(1)] = f_o[\underline{s}(0) \; ; \; \underline{d}^i(0)]$ ועבור לבדיקת $\underline{s}^i(1)$ הבא, או, אם נבדקו כולם עבור ל-(4).

(4) עבור כל $\underline{s}^i(1) \in \underline{S}(1)$ חשב

$$\underline{s}^j(2) = T\,[\underline{s}^i(1) \; , \; \underline{d}^j(1)]$$

עבור כל $\underline{d}^j(1) \in \underline{D}\,[,\underline{s}^i(1),1]$.

(5) חזור על צעדים (2) - (3) עבור כל $\underline{s}^j(2)$ ועבור כל $s^i(1)$. ככל מקרה ש-$\underline{s}^j(2)$ אפשרי, כלומר $\underline{S}^j(2) \in \underline{S}(2)$ חשב

$$g[\underline{s}^j(2)] = f_1[\underline{s}^i(1) \; ; \; \underline{d}^j(1)] + g\,[\underline{s}^i(1)]$$

(6) עבור לשלב הבא, ובאופן כללי, חשב

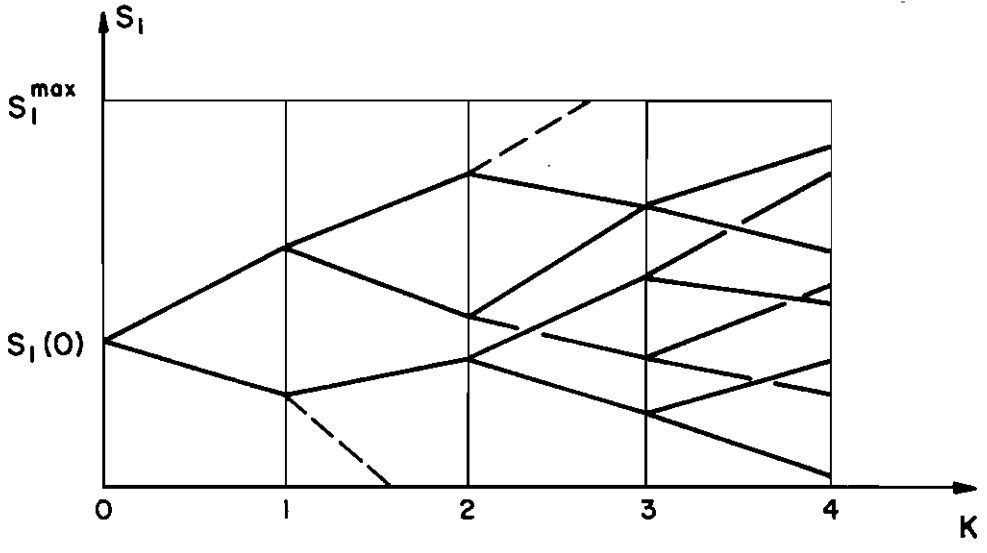
$$\underline{s}^p(k+1) = T_k[\underline{s}^q(k) \; , \; \underline{d}^p(\underline{s}^g(k),k)]$$

ועבור כל $\underline{s}^p(k+1) \in \underline{S}(k+1)$ חשב

$$g[\underline{s}^p(k+1)] = f_k[\underline{s}^g(k) \; ; \; \underline{d}^p(k)] + g[\underline{s}^q(k)]$$

וכך עד $k = K - 1$. לבסוף, מוסיפים את $f_K[\underline{s}(K)]$ ומקבלים עבור כל מסלול אפשרי את F. בוחרים את המסלול בעל F האקסטרמלי וזהו האופטימום המבוקש.

כדי להרגים את כמות החישובים ונפח הזכרון הדרוש לאחסנת התוצאות נבחן דוגמה בת $K = 4$ שלבים ובה משתנה מצב אחד, היכול לקבל שני ערכים אפשריים בכל מצב ובכל שלב.



<u>ציור 7.4</u>: הדגמה של חיפוש מקיף.

בדוגמא זו נדרש

$$0 < s_1(k) < s_1^{max}$$

ומשום כך התבטלו כמה מסלולים במהלך הפתרון. לולא ביטול המסלולים היה עלינו לחשב 2^4 מסלולים, כי מכל נקודה יוצאים שני מסלולים חדשים. באופן כללי, אם מספר השלבים הוא K ומספר הבקרות האפשריות בכל מצב וכל שלב הוא p, יהיו סה"כ p^K מסלולים, בלי להתחשב בביטול אפשרי של חלקים עקב אי קיום אילוצים. עבור K = 20 ו- p = 5 מתקבל מספר המסלולים $5^{20} \simeq 10^{14}$. גם אם חישוב מסלול אחד לוקח רק 10^{-6} שניות של זמן מחשב ידרשו 10^8 שניות, שהן למעלה מאלף שעות מחשב !

7.6 פתרון בתכנות דינמי

הבעיה לפתרון נתונה במשוואות (7-13), (7-14), (7-15). נגדיר כעת פונקציה חדשה של המצב בשלב ה- j :

$$F_k[\underline{s}(k)] = \underset{\underline{d}(k)}{ext} \left\{ \sum_{j=k}^{K-1} f_j[\underline{s}(j);\underline{d}(j)] + f_K[\underline{s}(K)] \right\} \qquad (7\text{-}16)$$

$$\text{s.t.} \qquad \underline{s}(j) \in \underline{S}(j) \qquad j = k+1,\ldots,K \qquad (7\text{-}17)$$

$$\underline{d}(j) \in \underline{D}[\underline{s}(j),j] \qquad j = k,\ldots,(K-1) \qquad (7\text{-}18)$$

אפשר לכתוב ישירות מעקרון האופטימליות את המשוואה הרקורסיבית

$$F_k[\underline{s}(k)] = \underset{\underline{d}(k)}{ext} \left\{ f_k[\underline{s}(k);\underline{d}(k)]+F_{k+1}[T_j[\underline{s}(k);\underline{d}(k)]] \right\} \qquad (7\text{-}19)$$

$$k = 1,\ldots,(K-1)$$

$$F_K[\underline{s}(K)] = f_K[\underline{s}(K)] \qquad (7\text{-}20)$$

וזהו הבסיס של החישוב הרקורסיבי. בבעיות בהן השלבים מבטאים רק סדר נוח בהחלטה ובבעיות תלויות בזמן שהן דטרמיניסטיות אפשר לבחור כרצוננו את כיוון החישוב, כלומר לסמן ב- $j = 1$ את השלב האחרון במקום את הראשון. למונחים ראשון ואחרון יש משמעות מוחלטת של מוקדם ומאוחר רק כבעיות התלויות בזמן. בבעיות אחרות ראשון ואחרון הם מונחים יחסיים ומורים רק על הסדר בו אנו פותרים. בבעיות סטוכסטיות, שהן בעיות התלויות בזמן אשר בהן התוצאה של החלטה תלויה לא רק במצב ממנו יוצאים ובהחלטה עצמה אלא גם במצב הטבע באותו זמן, חייב הפתרון להתקדם "אחורה", כלומר החישוב מתקדם מן השלב המאוחר למוקדם.

לפני תחילת החישוב לפי המשוואה הרקורסיבית (7-19) יש לבצע דיסקרטיזציה של המשתנים הרציפים. לאחר שהדבר נעשה מתבצע החישוב כדלקמן :

(1) מכניסים לטבלה את הערכים של $f_K[\underline{s}(K)]$ עבור כל נקודה של $\underline{s}(K)$.
ערכים אלה נקראים הערכים השיוריים (salvage values) שכן הם מבטאים את ערך

המערכת עם "סגירתה" כסוף אופק הזמן הנדון כפונקציה של מצבה ברגע הסגירה. הערכים מסומנים ב- $F_K[\underline{s}(K)]$.

(2) עבור כל ערך אפשרי של $\underline{s}(K-1)$ מבצעים את הפעולות הבאות:

עבור כל החלטה אפשרית, $\underline{d}(K-1) \in D[\underline{s}(K-1),K-1]$, מחשבים את המצב הנובע ממנה לפי $\underline{s}(K) = T_{K-1}[\underline{s}(K-1) ; \underline{d}(K-1)]$. אם מצב זה אפשרי, כלומר $\underline{s}(K) \in \underline{S}(K)$ מחשבים את תרומת המעבר לפי טרנספורמציה זו שהיא

$$f_{K-1}[\underline{s}(K-1) ; \underline{d}(K-1)]$$

ומוסיפים לה את הערך $F_K[\underline{s}(K)]$ של המצב $\underline{s}(K)$ המתאים, וזאת מתוך הטבלה שהוכנה ב-(1) לעיל. שומרים את ערך הסכום ועוברים להחלטה האפשרית הבאה. לאחר שנבדקו עבור המצב $\underline{s}(K-1)$ הנדון כל ההחלטות האפשריות מחושבת הפונקציה

$$F_{K-1}[\underline{s}(K-1)] = \underset{\underline{d}(K-1)}{ext} \{ f_{K-1}[\underline{s}(K-1); \underline{d}(K-1)] + F_K[\underline{s}(K)] \}$$

הערכים של $F_{K-1}[\underline{s}(K-1]$ ואותן החלטות הנותנות את האקסטרמום (מינימום או מקסימום) המבוקש $\underline{d}^*[\underline{s}(K-1)]$ נרשמים בטבלאות כפונקציה של המצב, ואילו ישמשו בשלבים הבאים.

(3) עבור לשלב הבא של החישוב, שלב (K-2), ועשה עבורו אותן פעולות כמו עבור השלב ה-(K-1). תוך כך השתמש בטבלה של $F_{K-1}[\underline{s}(K-1)]$ שהוכנה ב-(2).

(4) חזור על (3) עבור השלבים (K-3) עד 1. התוצאה הסופית תהיה טבלאות של $F[\underline{s}(1)]$ ושל $\underline{d}^*[\underline{s}(1)]$.

(5) בשלב ה-0, שבו $\underline{s}(0)$ ידוע, חשב עבור כל $\underline{d}(0) \in D[\underline{s}(\underline{0}),0]$ את הערך של

$$f_o[\underline{s}(0) ; \underline{d}(0)] + F[T_o[\underline{s}(0) ; \underline{d}(0)]]$$

ומבין כל הערכים מצא

$$F_o[\underline{s}(0)] = \underset{\underline{d}(0)}{ext} \{f_o[\underline{s}(0) ; \underline{d}(0)] + F_1[T_o[\underline{s}(0) ; \underline{d}(0)]]\}$$

הנגרם על-ידי $\underline{d}^*[\underline{s}(0)]$

$F_o[\underline{s}(0)]$ הוא הערך האופטימלי של פונקצית המטרה. על מנת לחשב את המסלול האופטימלי יש להמשיך בצעדים הבאים:

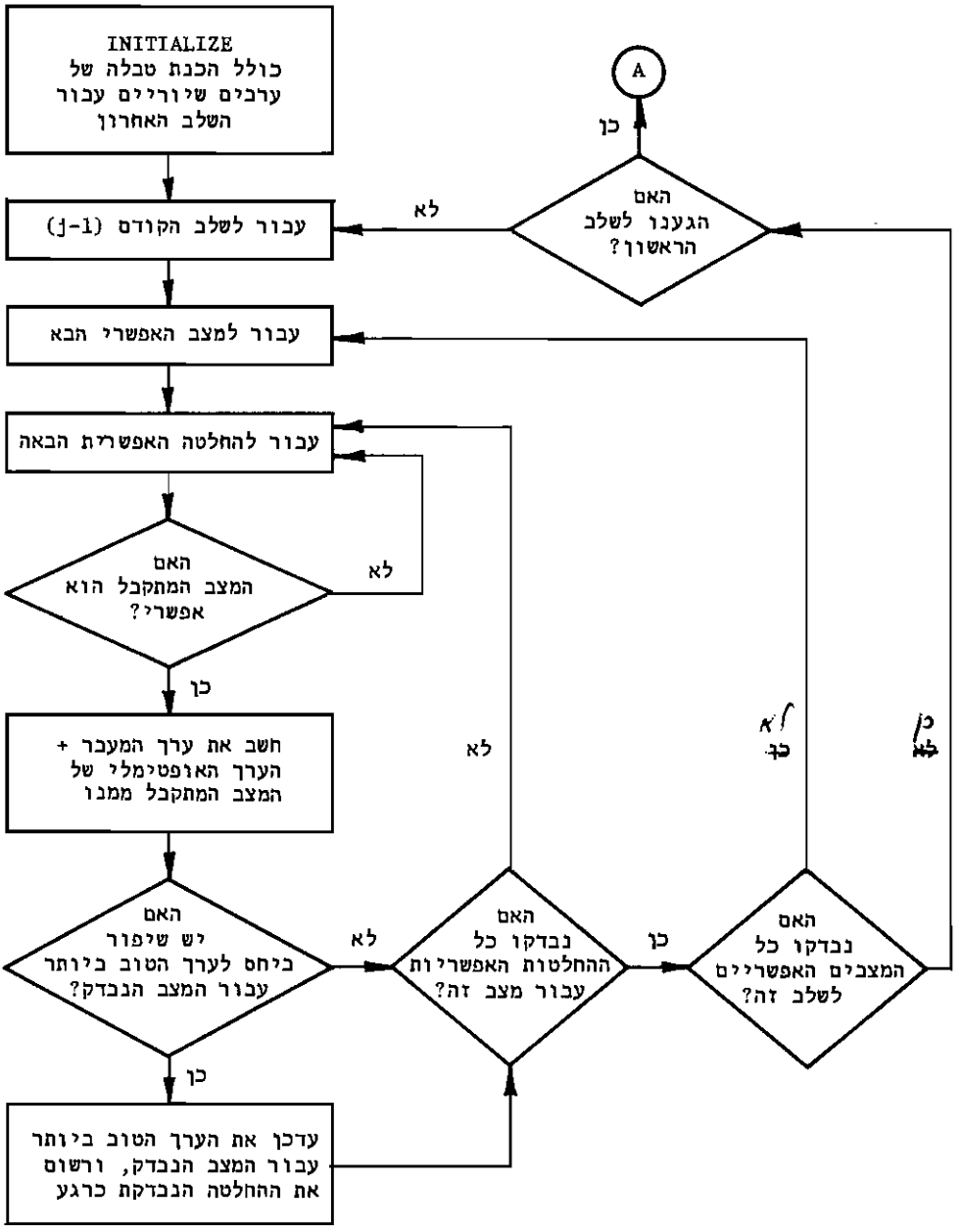
(6) הנקודה על המסלול האופטימלי בשלב ה-k תסומן $\underline{s}^*(k)$. את הנקודות האלה מוצאים באמצעות החישוב הרקורסיבי

$$\underline{s}^*(k) = T_{k-1}[\underline{s}^*(k-1) ; \underline{d}^*[s^*(k-1)]] \quad k = 1,\ldots,K$$

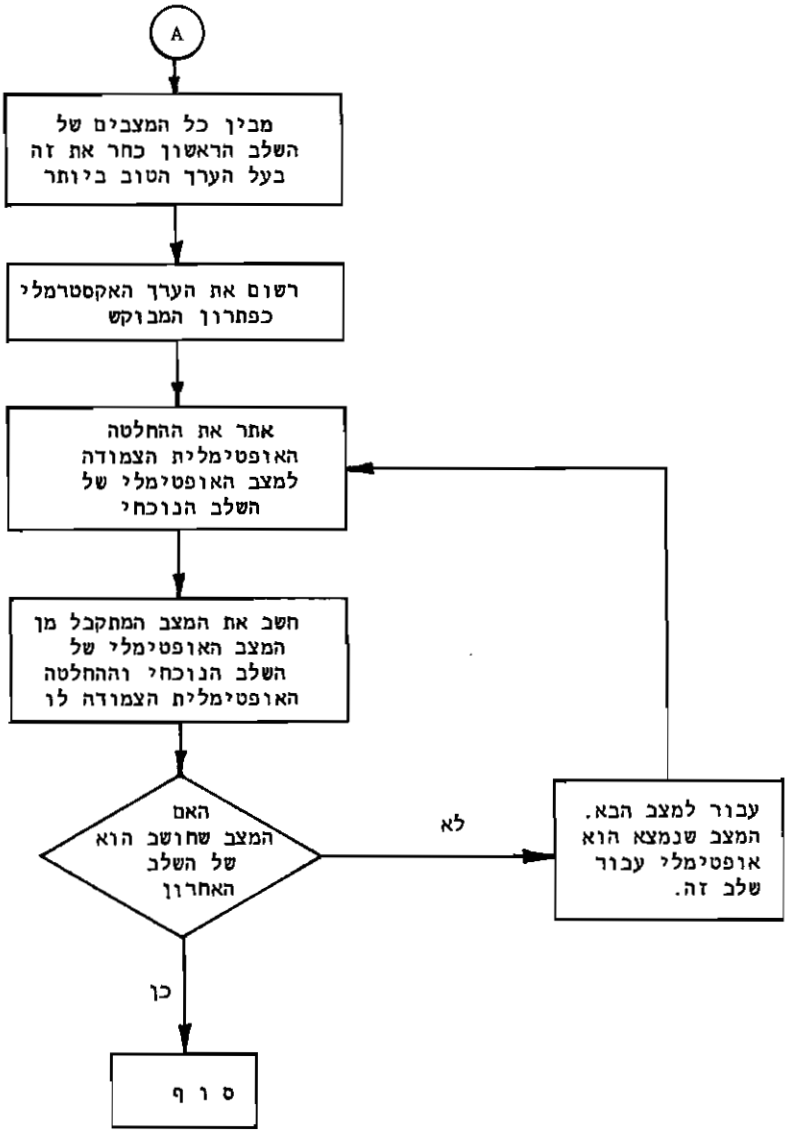
כאשר את ערכי $\underline{d}^*$ חישבנו בחלק הראשון של הפתרון. כלומר בהליכה "אחורה".

נשים לב שבנוסף לפתרון המבוקש קיבלנו גם את הערך האופטימלי עבור כל מצב בכל שלב, וכן "חץ" המאפשר לנו לחשב מכל מצב בכל שלב את המסלול האופטימלי עד לשלב הסופי, ה-K. <u>אין</u> בידינו מסלול אופטימלי מהשלב התחילי לכל נקודה.

התהליך מתואר בתזרים שבציור 7.5.



<u>ציור 7.5</u>: תזרים התהליך של תכנות דינמי.



ציור 7.5: תזרים התהליך של תכנות דינמי (המשך).

7.7 דרישות חישוב ואחסון, "קללת המימדיות"

כשעובדים על השלב ה-k דרוש לשמור בזכרון המהיר את טבלת $F_{k+1}[\underline{s}(k+1)]$.
אם מספר משתני המצב הוא m (כלומר לוקטור $\underline{s}$ יש m רכיבים), ולמשתנה המצב ה-i יש N_i ערכים בדידים, הרי שבטכלה של F_{k+1} יש

$$N_h = \prod_{i=1}^{m} N_i \qquad (7\text{-}21)$$

ערכים. תוך כדי החישוב מייצרים עוד טבלה בגודל כזה של F_k. סה"כ דרוש בזכרון המהיר מקום ל- $2N_h$ ערכים. למשל, עבור $m = 3$ ו- $N_i = 100$ דרושים 2×10^6 ערכים.
לאחר שגומרים לחשב את טבלת F_k אפשר להוציא לזכרון איטי (דיסק) את F_{k+1}, "להדיז" את F_k למקום של F_{k+1} ולחשב במקום בו היה F_k את F_{k-1}.

בזכרון המשני יש לאגור $KN_h = N_c$ ערכים. יש סה"כ N_c חישובים של F ושל d* וזמן החישוב סה"כ

$$T_c = N_c \Delta t_c \qquad (7\text{-}22)$$

למשל $K = 50$ (שלבים), $N_h = 10^6$ מקבלים

$$T_c = 5 \times 10^7 \Delta t_c$$

נניח שכל חישוב נמשך $\Delta tc = 5\times10^{-5} sec$ אזי

$$T_c \simeq 40\ min$$

אבל לחיפוש מקיף עבור 5 ערכי d בכל נקודה מספר המסלולים הוא $N = 5^{50} \cong 10^{35}$
וזמן החישוב, בהנחה שחישוב כל מסלול לוקח רק 5×10^{-5} שניות כמו קודם, יהיה

$$T = 5 \times 10^{-5} \times 10^{35} = 5 \times 10^{30} secs \cong 10^{23} years\ (!)$$

שיטת התכנות הדינמי סובלת מ"קללת המימדיות" (The Curse of Demensionality).
כפי שניתן לראות ממשוואות (7-21) ו-(7-22) נפח האחסון בזכרון המהיר וכמות

החישובים עולים באופן אקספוננציאלי עם מספר משתני המצב. אי לכך, למרות פשטותה העקרונית של השיטה היא אינה מעשית עבור בעיה עם יותר משלושה, ארבעה, חמישה משתני מצב. זאת למרות שקיימות מספר שיטות מיוחדות לטיפול בבעיות רב-מימדיות למשל זו של State Increment Dynamic Programming : Larson.

7.8 אינטרפולציה

אם מגיעים לטבלה של השלב הקודם בנקורה שאינה נקודת חלוקה יש לעשות אינטרפולציה של ערכי F_{k+1}. נהוג להשתמש באחת משתי השיטות הפשוטות: שימוש בערך הלקוח מנקורת החלוקה הקרובה ביותר או אינטרפולציה ליניארית בין הערכים של חלק מנקודות החלוקה היוצרות את המרווח שבו נמצאים או כולן. למשל - אינטרפולציה ליניארית בין שני הערכים השכנים במקרה של משתנה מצב אחד, חישוב מישור משולש דרך שלוש מארבע נקודות פינה של תחום דו-מימדי מלבני או התאמת מישור במינימום ריבועים דרך כל ארבעת הערכים בפינות ובאופן דומה למקרים הרב מימדיים. הבעיה אינה פשוטה ודורשת מחשבה בכל מקרה ספציפי. הקושי הוא ששיטות אינטרפולציה יותר מורכבות דורשות זמן חישוב רב יותר.

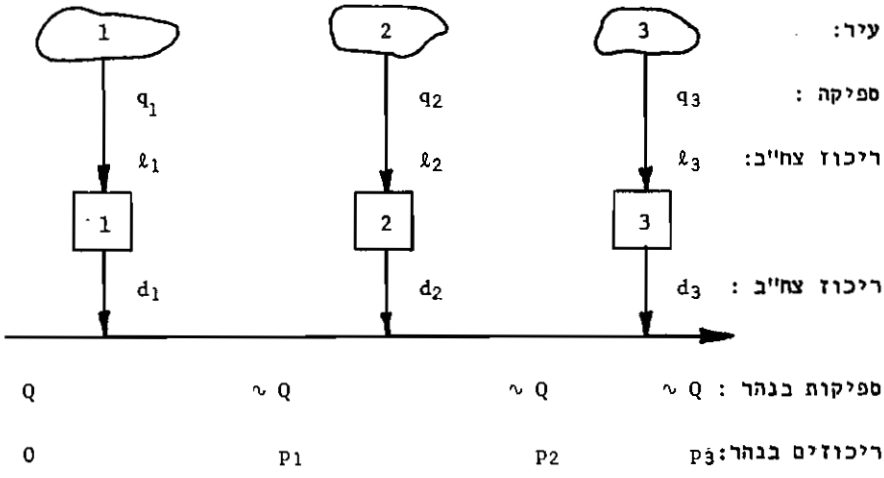
בעיה חדשה מתעוררת בעת חישוב המסלול האופטימלי. אם עכשיו "נופלים" בין נקורות חלוקה אפשר אולי להגדיר שם F באינטרפולציה, אבל לא חושב הערך של הבקרה האופטימלית שם. אם הבקרה רצופה ניתן אולי לעשות אינטרפולציה ברומה למה שנאמר לעיל ביחס לערכי הפונקציה גם עבור ערכי הבקרה האופטימלית. באשר הבקרה אינה רצופה ניתן לבחור את הבקרה של נקודת החלוקה הקרובה ביותר.

7.9 דוגמה : תכנון אזורי של סילוק שפכים לנהר

שלוש ערים ממוקמות לאורך נהר במתואר בציור 7.6. הן מסומנות $i = 1,2,3$. לכל אחת יש ספיקת שפכים נתונה q_i (מ"ק ליום) אותה היא רוצה להזרים לנהר. ריכוז הצח"ב בשפכים של כל עיר, ℓ_i (מ"ג לליטר) ידוע, והוא משתנה מעיר לעיר.

ריכוזי הצח"ב בשפכים הגולמיים גבוהים, והנהר לא יוכל לשאת את העומסים האלה בלי שריכוזי הצח"ב בו יעלו על המותר. לכן יש לתכנן מתקני טיפול, כמתואר בציור. ריכוז הצח"ב בקולחים מן המתקנים יסומנו d_i. עלות מתקן טיפול תלויה בספיקה העוברת בו ובריכוזי הצח"ב בכניסה וביציאה, והיא יכולה להיות שונה מעיר לעיר.

פונקציות העלות הן $f_i(q_i;\ell_i;d_i)$.



<u>ציור 7.6</u>: תכנון אזורי של סילוק שפכים לנהר.

הספיקה במעלה הנהר מהמוצא הראשון היא Q וריכוז הצח"ב בה אפס. ספיקה זו גדולה מאד יחסית ל- q_i , ולכן ניתן לחשב את ריכוז הצח"ב בכל קטע של הנהר כדלקמן:

$$p_1 = \frac{q_1 d_1}{Q} \tag{7-23}$$

$$p_2 = \frac{q_1 d_1 + q_2 d_2}{Q} \tag{7-24}$$

$$p_3 = \frac{q_1 d_1 + q_2 d_2 + q_3 d_3}{Q} \tag{7-25}$$

לריכוזים אלה אסור לעלות על ריכוז מקסימלי מותר, p_j^{max}, שהוא שונה לכל קטע של הנהר. מניחים (על מנת לפשט את הכעיה) כי לנהר אין תכונות של סיהור עצמי, כלומר ריכוז הצח"ב אינו משתנה תוך הזרימה בנהר.

לצורך פתרון כתכנות דינמי נראה את קטעי הנהר או הערים כשלכים.

משתני ההחלטה הם הריכוזים, d_i , הקוכעים את גודל המתקנים. הריכוזים כנהר, p_j, גם הם היו יכולים להילקח כמשתני החלטה (שכן הם תלויים בצורה חד משמעית בריכוזים d_i, ולהיפך) - אלא שזה פחות נוח, ועדיף לקחת את הריכוזים בנהר כמשתני המצב.

משתנה המצב הוא הריכוז כנהר בקטעים השונים. בעור שיש שלושה משתני החלטה, יש רק משתנה מצב אחד. משתנה המצב הוא אותו גורל (פיזי, כלכלי או אחר) שבתהליך הפתרון אומרים לגביו : "הערך האופטימלי של ... כפונקציה של משתנה המצב הוא המינימום (או המקסימום) על פני ..." וכו'.

פרמטרים יכולים להחשב כל הגדלים הפיזיים הנתונים, כלומר

$$q_i \; ; \; \ell_i \; ; \; p_j^{max}$$

וכן גרלים כלכליים, כלומר

$$f_i(q_i \; , \; \ell_i \; , \; d_i)$$

פונקצית המטרה: מינימום עלות כל המתקנים ביחר. כאן ראוי להעיר שההנחה היא שבעית המינימיזציה מנוסחת עכור רשות מרכזית הצריכה לנהל את כל האיזור ולממן את כל המפעלים. כהתעלם מכעיות האיכות בנהר רוצה כל עיר להזרים לנהר את השפכים ללא כל טיפול, שכן זה הזול ביותר מכחינתה. הרשות המרכזית רוצה לאזן את ההוצאות בין הערים כך שסה"כ ההוצאות יהיה מינימלי תוך עמידה באילוצים.
פונקצית המטרה היא לכן

$$min \quad F = \sum_{i=1}^{3} f_i(q_i \; ; \; \ell_i \; ; \; d_i) \tag{7-26}$$

כשהמינימיזציה על פני $d_i (i = 1,2,3)$ בכפיפות <u>לאילוצים</u>

$$p_j \le p_j^{max} \qquad j = 1,2,3 \tag{7-27}$$

כיטוי האילוצים לגבי משתני ההחלטה הוא

$$\sum_{i=1}^{j} \frac{q_i d_i}{Q} \le p_j^{max} \qquad j = 1,2,3 \tag{7-28}$$

<u>מהלך הפתרון בתכנות דינמי</u>

$\frac{q_1 \ell_1}{Q}$

(1) קבע ערכים כדידים עבור p_1 בתחום $0 \le p_1 \le p_1^{max}$ או $0 \le p_1 \le \ell_1$ מה שיותר כוכל.

(2) עבור כל ערך של p_1 חשב d_1 מתוך $d_1 = Qp_1/q_1$ ועבורו חשב $f_1(d_1)$. אגור את התוצאות בטבלה של $f_1^*(p_1)$. הסיכה לפשטות החישובים בשלב זה, שהם למעשה

$$f_1^*(p_1) = \min_{0 \le d_1 \le Qp_1/q_1} \{f_1(q_1;\ell_1;d_1)\} = f_1(q_1;\ell_1;Qp_1/q_1)$$

היא שברור אינטואיטיבית שעבור p_1 מסויים בנהר כדאי להרשות ל- d <u>לגדול</u> עד כדי הערך שגורם ל- p_1 לקבל את הערך הנדון וזאת משום ש- $f_1(q_1;\ell_1;d_1)$ הוא בודאי פונקציה עולה של d_1. (d_1 ין ⇐ f_1 עולה).
כאן בדאי להעיר שאם יש גם עלות הקשורה בריכוז בנהר (למשל נזקים לדייג או לתיירות כתוצאה מירידת האיכות בנהר) האופטימיזציה צריכה להעשות באופן מלא, כלומר :

$$f_1^*(p_1) = \min_{0 \le d_1 \le Qp_1/q_1} \{f_1(q_1;\ell_1;d_1) + p_1 \text{ נזקי } \} \tag{7-29}$$

$\frac{g_1 \ell_1 + g_2 \ell_2}{Q}$

שכן לא ברור מראש שכדאי לקחת d_1 בערכו המקסימלי.

(3) קבע ערכים בדידים עבור p_2 בתחום $0 \le p_2 \le p_2^{max}$ או $0 \le p_2 \le (\ell_2 + \ell_1)$
מה שיותר כובל.

(4) עבור כל ערך כצע

$$f_2^*(p_2) = \min_{0 \le d_2 \le Qp_2/q_2} \{ f_2(q_2;\ell_2;d_2) + f_1^* \left[p_2 - \frac{q_2 d_2}{Q} \right] \} \qquad (7\text{-}30)$$

החלק השני הוא $f_1^*(p_1)$ כאשר את p_1 מחשבים ממשוואת ההתמרה (טרנספורמציה) הפשוטה

$$p_2 = \frac{q_1 d_1 + q_2 d_2}{Q} = p_1 + \frac{q_2 d_2}{Q} \qquad (7\text{-}31)$$

במידה ואין בטבלה שהוכנה בשלב (2) ערך מתאים של p_1 יש לעשות אינטרפולציה.
את התוצאות רושמים בטבלה של $f_2^*(p_2)$, וכן רושמים את $d_2^*(p_2)$ המתאים שנתן את האופטימום עבור הערך המתאים של p_2.

(5) קבע ערכים בדידים עבור p_3 בתחום $0 \le p_3 \le p_3^{max}$
אן $0 \le p_3 \le \cancel{(\ell_3 + \ell_2 + \ell_1)}$ $(g_1\ell_1 + g_2\ell_2 + g_3\ell_3)/Q$ - מה שיותר כובל.

(6) עבור כל ערך של p_3 בצע

$$f_3^*(p_3) = \min_{0 \le d_3 \le Qp_3/q_3} \{ f(g_3;\ell_3;d_3) + f_2^* \left[p_3 - \frac{g_3 d_3}{Q} \right] \} \qquad (7\text{-}32)$$

לפי אותה שיטה כמו לעיל.
משיקולים אינטואיטיביים בראי ככל הנראה להרשות $p_3 = p_3^{max}$ שכן כמורר אין ערים נוספות, ומחיר המתקנים יורד עם עלית ה-p_j. אגב, זה אינו תופס

ביחס ל- p_2 ו- p_1 כי שם יש שיקול של איזון בין הערים. אם אמנם כך, מספיק ב-(7-32) לעשות $f_3^*(p_3^{max})$ פעם אחת ולא עבור על ערכי p_3. לו היה בפונקציית המטרה גם איבר נוסף - למשל נזקים על p_j גבוה - לא היה השיקול שלעיל תופס.

(7) כין כל ערכי $f_3^*(p_3)$ מצא את המינימלי וקרא לערך המתאים p_3^* (בהערה לעיל נאמר שזה כנראה ב-p_3^{max} עבור השאלה שלנו), מתאים לו $d_3^*(p_3^*)$ שהוא ערך אחר ממשתני ההחלטה או באופטימום.

חשב באמצעות ערך זה את

$$p_2^* = p_3^* - \frac{q_3 d_3^*(p_3^*)}{Q}$$

הכנס לטבלת $f_2^*(p_2)$ בערך זה p_2^* וראה מהו $d_2^*(p_2^*)$ המתאים - זה נותן את הריכוז כיציאה ממתקן 2 - שהוא משתנה ההחלטה השני באופטימום. חשב כעת

$$p_1^* = p_2^* - \frac{q_2 d_2^*(p_2^*)}{Q} \qquad (7\text{-}33)$$

ומצא $d_1^*(p_1^*)$ - שהוא משתנה ההחלטה האחרון באופטימום.

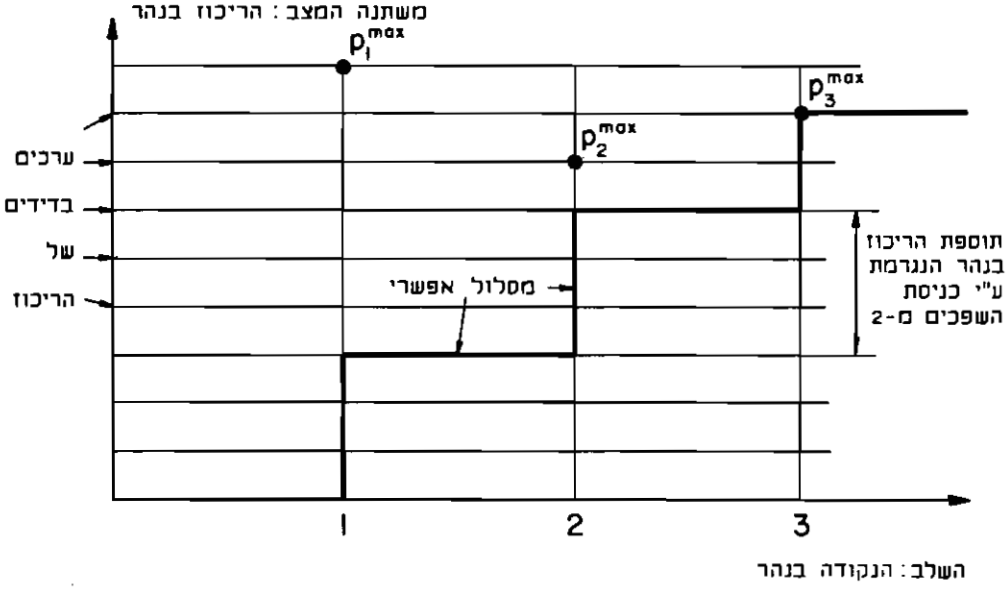
(8) חשב

$$F^* = \sum_{i=1}^{3} f_i^*(q_i;\ell_i;d_i^*)$$

לקבלת העלות הכוללת.

בכך נסתיים הפתרון - פרט אולי לבדיקות רגישות ביחס לפרמטרים השונים, שהן חשובות ביותר.

את הפתרון ניתן לתאר בצורה גרפית, כמתואר בציור 7.7.



ציור 7.7: תאור גרפי של פתרון הריכוז בנהר.

את הפתרון ניתן לראות במציאת מסלול דרך רשת, במרחב משתנה המצב, הנוצרת ע"י בחירת ערכים בדידים. הקפיצה האנכית בכל שלב היא תרומת השפכים של אותה כניסה לריכוז בנהר. את הקפיצה הזו ניתן לתרגם באופן ישיר לערך משתנה ההחלטה, d_i, של אותו שלב באמצעות משוואות המיהול בנהר.

מ ק ו ר ו ת

Bellman, R., "Dynamic Programming", Princeton University Press, 1957.

Dreyfus, S., "Dynamic Programming and the Calculus of Variations", Academic Press, 1965.

Hadley, G., "Nonlinear and Dynamic Programming", Addison-Wesley, 1964.

Larson, R.E., "State Increment Dynamic Programming", American Elsevier, 1968.

Nemhauser, G.L., "Introduction to Dynamic Programming", John Wiley and Sons, 1966.

נספח א': יסודות חשבון המטריצות.

1.א כ ל ל י

מטריצה (Matrix) היא קבוצת גדלים, מסודרים במערך מלבני בעל שורות ועמודות, שעבורה מוגדרים חוקים מסויימים של שוויון, חיבור וכפל. נסמן ב-a_{ij} את אכרי המטריצה $\underline{\underline{A}}$, אשר לה m שורות ו- n עמודות

$$\underline{\underline{A}} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots\cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots\cdots & a_{2n} \\ \vdots & & \vdots & \\ a_{m1} & a_{m2} & \cdots\cdots\cdots & a_{mn} \end{bmatrix} \qquad (\text{א-1})$$

וקטור (Vector) עמודה הוא מטריצה בעלת עמודה אחת בלבד, למשל

$$\underline{x} = \begin{Bmatrix} x_1 \\ \vdots \\ x_n \end{Bmatrix} \qquad (\text{א-2.1})$$

וקטור שורה הוא מטריצה בעלת שורה אחת בלבד, למשל

$$\underline{C} = [C_1, \ldots, C_m] \qquad (\text{א-2.2})$$

2.א דטרמינטים

מטריצה נקראת ריבועית כאשר $m = n$ ואז אומרים ש- $\underline{\underline{A}}$ היא מסדר (Order) n.
עבור מטריצה כזו מוגדר הדטרמינטנ (Determinant) כמספר המתקבל על ידי סכום כל המכפלות האפשריות, כשבכל אחת מופיע רק אלמנט אחד מכל שורה ומכל עמודה. לכל מכפלה כזו נקבע סימן חיובי או שלילי לפי כלל שיוסבר להלן. את הדטרמינט נסמן כך: $|\underline{\underline{A}}|$ הוא הדטרמינט של המטריצה (הריבועית) $\underline{\underline{A}}$. הדטרמינט מחושב כדלקמן:

$$|\underline{\underline{A}}| = \Sigma(\pm 1)\, a_{1i}\, a_{2j}\, a_{3k} \cdots a_{nr} \qquad (\text{א}-3)$$

כאשר לוקחים את הסכום על כל התמורות של האינדקס השני. למכפלה סימן חיובי כאשר מספר התמורות של האינדקסים השניים $(i,j,k, \ldots ,r)$ הוא זוגי ביחס לסדר $(1,2,3, \ldots ,n)$, ושלילי כאשר מספר התמורות בלתי זוגי. למשל, למכפלה בה שורת האינדקסים השניים היא $(2,1,3,4, \ldots ,n)$ יהיה סימן שלילי.

צורה אחרת לחישוב הדטרמינט היא על ידי פיתוח לפי שורה או עמודה. למשל

$$|\underline{\underline{A}}| = \sum_{i=1}^{n} a_{ij}\, A_{ij} \qquad (\text{א}4-1)$$

הוא פיתוח לפי השורה ה-i. A_{ij} נקרא הקופקטור (Cofactor) של a_{ij} והוא נתון על ידי

$$A_{ij} = (-1)^{k+j}\, |\underline{\underline{M}}_{ij}| \qquad (\text{א}-5)$$

כאן $\underline{\underline{M}}_{ij}$ היא תת-מטריצה של $\underline{\underline{A}}$, המתקבלת ממנה על ידי מחיקת השורה ה-i והעמודה ה-j. פיתוח לפי העמודה ה-j נותן

$$|\underline{\underline{A}}| = \sum_{i=1}^{n} a_{ij}\, A_{ij} \qquad (\text{א}4-2)$$

פעולות של מטריצות והשפעתן על הדטרמיננט :

1. החלפת השורות בעמודות, כלומר השורה ה-i הופכת להיות העמורה ה-i. המטריצה המתקבלת נקראת <u>המטריצה המוחלפת</u> (Transfer) , $\underline{\underline{A}}^T$.
קיים $|\underline{\underline{A}}^T| = |\underline{\underline{A}}|$

2. החלפת שתי שורות זו בזו, או שתי עמורות זו בזה, משנה את סימן הדטרמיננט.

3. כפל של כל האברים בשורה אחת, או של כל האברים בעמודה אחת, בגורל קבוע k, שווה לכפל הדטרמיננט ב-k.

4. אם שורה אחת היא כפולה של שורה אחרת, או עמודה אחת היא כפולה של עמורה אחרת, הדטרמיננט של המטריצה מתאפס.

5. תוספת של k כפול שודה לשורה אחרת, או של k כפול עמודה לעמודה אחרת, אינה משנה את ערך הדטרמיננט.

<u>דרגת המטריצה</u> (Rank of the Matrix) היא הסדר של המטריצה הריבועית הגדולה ביותר שאפשר לסדר משורות ועמורות של המטדיצה המקורית, כך שהרטרמיננט שלה אינו מתאפס. אם $\underline{\underline{A}}$ היא $(m \times n)$, ואפשר, על ידי מחיקת שורות ועמודות ממנה, לקבל מטריצה $(p \times p)$, שהיא הגדולה ביותר עבורה הדטרמיננט אינו מתאפס, דרגתה של $\underline{\underline{A}}$ היא p, וזה נכתב $r(\underline{\underline{A}}) = p$.

מטריצה ריבועית היא <u>סינגולרית</u> (Singular) אם הדטרמיננט שלה שווה לאפס.

3.א <u>חיבור מטריצות</u>

חיבור של מטריצות מוגדר על ידי

$$\underline{\underline{A}} + \underline{\underline{B}} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{nm} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m} & \cdots & b_{mn} \end{bmatrix} =$$

$$= \begin{bmatrix} (a_{11}+b_{11}) & \cdots & (a_{1n}+b_{1n}) \\ \vdots & & \vdots \\ (a_{m1}+b_{m1}) & \cdots & (a_{mn}+b_{mn}) \end{bmatrix} \qquad (\text{א-6})$$

הפעולה מוגדרת רק כאשר מספר השורות והעמודות במטריצות המחוברות שווה, והמטריצה המתקבלת גם היא בעלת אותו מספר שורות ועמודות.

א.4 <u>כפל המטריצות</u>

כפל של מטריצות מוגדר על ידי

$$\underline{\underline{A}} \cdot \underline{\underline{B}} = \begin{bmatrix} a_{11} & \cdots\cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots\cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1\ell} \\ \vdots & & \vdots \\ b_{n1} & & b_{n\ell} \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{1i}b_{i1} & \cdots & \sum_{i=1}^{n} a_{1i}b_{i\ell} \\ \vdots & & \vdots \\ \sum_{i=1}^{n} a_{mi}b_{i\ell} & \cdots & \sum_{i=1}^{n} a_{mi}b_{i\ell} \end{bmatrix} \qquad (\text{א-7})$$

אם $\underline{\underline{A}}$ היא $(m \times n)$ ו-$\underline{\underline{B}}$ היא $(n \times p)$ אזי התוצאה היא מטריצה $(m \times p)$. המכפלה מוגדרת רק כאשר מספר העמודות במטריצה הראשונה, $\underline{\underline{A}}$, שווה למספר השורות במטריצה השניה, $\underline{\underline{B}}$. אם $m \neq p$ אזי המכפלה $\underline{\underline{B}}\ \underline{\underline{A}}$ אינה מוגדרת. כמובן, כאשר $m = n = p$ אפשר להגדיר גם את המכפלה השניה, אבל התוצאה המתקבלת היא שונה, כלומר, גם אם שתי המכפלות מוגדרות, הרי באופן כללי

$$\underline{\underline{A}}\ \underline{\underline{B}} \neq \underline{\underline{B}}\ \underline{\underline{A}} \qquad (\text{א-8})$$

הכפל הוא אסוציאטיבי, כלומר

$$\underline{\underline{A}}\ (\underline{\underline{B}}\ \underline{\underline{C}}) = (\underline{\underline{A}}\ \underline{\underline{B}})\ \underline{\underline{C}} \qquad (\text{א-9})$$

וריסטריבוטיבי

$$\underline{\underline{A}}\ (\underline{\underline{B}} + \underline{\underline{C}}) = (\underline{\underline{A}}\ \underline{\underline{B}}) + (\underline{\underline{A}}\ \underline{\underline{C}}) \qquad (\text{א-10})$$

כפל של מטריצה בוקטור נותן וקטור, למשל

$$\underline{\underline{A}}\ \underline{x} = \underline{b} \qquad (\text{א-11})$$

על מנת שהמכפלה תהיה מוגדרת דרוש שמספר העמודות ב- $\underline{\underline{A}}$ יהיה שווה למספר האברים של $\underline{x}$ ושל $\underline{b}$.

א.5 מטריצת היחידה

מטריצת היחידה מסדר m , $\underline{\underline{I}}_m$, היא מטריצה ריבועית $(m \times m)$, בה כל האברים על האלכסון הראשי הם 1 וכל האברים האחרים אפס.

$$\underline{\underline{I}}_m = \begin{pmatrix} 1 & 0 & \cdots\cdots & 0 & 0 \\ 0 & 1 & \cdots\cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots\cdots & 0 & 1 \end{pmatrix} = [\delta_{ij}] \qquad (\text{א-12})$$

כאשר δ_{ij} היא הדלתה לקרונקר (Kroneker's Delta).

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad i,j = 1, \ldots ,m \qquad (\text{א}-13)$$

א.6 מערכות של משוואות ליניאריות וכיטוין בעזרת מטריצות

את מערכת המשוואות הליניאריות הסימולטניות

$$\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots\ldots\ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots\ldots\ldots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots\ldots\ldots + a_{mn}x_n = b_m \end{array} \qquad (\text{א}-14)$$

ניתן לכתוב בעזרת מטריצות בצורה

$$\underline{\underline{A}} \cdot \underline{x} = \underline{b} \qquad (\text{א}-15)$$

כאשר $\underline{\underline{A}}$ היא מטריצה של המקדמים, a_{ij} , $\underline{x}$ הוא וקטור הנעלמים ו-$\underline{b}$ הוא וקטור האברים החופשיים, או בצורה מפורשת

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{Bmatrix} \qquad (\text{א}-16)$$

7.א פעולות אלמנטריות

לפעולות הבאות, המבוצעות על מטריצות, קוראים פעולות אלמנטריות :

א. החלפת שתי שורות, או שתי עמודות, זו בזו.
ב. כפל כל האברים של שורה, או של עמודה, בגודל קבוע K.
ג. הוספת K כפול שורה אחת לשורה אחרת, או K כפול עמודה אחת לעמודה אחרת.

אפשר להראות כי פעולות אלמנטריות המבוצעות בעת ובעונה אחת על מטריצת המקדמים, $\underline{\underline{A}}$, ועל וקטור האיברים החופשיים, $\underline{b}$, של מערכת המשוואות (14-א') אינה משנה את פתרונן. מטרתן של הפעולות האלמנטריות לשנות את צורת המשוואות, ולהביאן למצב בו ניתן לחשב את פתרונן.

8.א פעולות אלמנטריות באמצעות מטריצות עזר

את הפעולות האלמנטריות ניתן לבצע על ידי כפל מלפנים של המטריצה עליה פועלים במטריצה אחרת,הנקראת מטריצת עזר. לדוגמא, אם ברצוננו להוסיף λ כפול השורה השניה לשורה הראשונה של המטריצה

$$\underline{\underline{D}} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ d_{31} & d_{32} & \cdots & d_{3n} \end{pmatrix}$$

נכפול אותה מלפנים במטריצת העזר

$$\underline{\underline{E}}_1 = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\underline{\underline{E}}_1 \cdot \underline{\underline{D}} = \begin{pmatrix} (d_{11} + \lambda d_{12}) & (d_{12}+\lambda d_{22}) & \cdots\cdots & (d_{1n}+\lambda d_{2n}) \\ d_{21} & d_{22} & \cdots\cdots & d_{2n} \\ d_{31} & d_{32} & \cdots\cdots & d_{3n} \end{pmatrix}$$

באופן דומה, החלפת השורה הראשונה והשלישית של המטריצה $\underline{\underline{D}}$ יכולה להעשות על ידי כפל $\underline{\underline{D}}$ מלפנים במטריצת העזר,

$$\underline{\underline{E}}_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

אם ברצוננו להוסיף לשורה הראשונה λ כפול השניה, ואחר כך להחליף את השורה הראשונה (החדשה) בשורה השלישית, ניתן הדבר להעשות על ידי המכפלה החוזרת $\underline{\underline{E}}_2 \cdot \underline{\underline{E}}_1 \cdot \underline{\underline{D}}$. התוצאה תהיה,
ואמנם קל לראות כי

$$\underline{\underline{E}}_2 \cdot (\underline{\underline{E}}_1 \cdot \underline{\underline{D}}) = \underline{\underline{E}}_2 \begin{pmatrix} (d_{11}+\lambda d_{21}) & (d_{12}+\lambda d_{22}) \dots & (d_{1n}+\lambda d_{2n}) \\ d_{21} & d_{22} \quad \dots & d_{2n} \\ d_{31} & d_{32} \quad \dots & d_{3n} \end{pmatrix}$$

$$= \begin{pmatrix} d_{31} & d_{32} & \dots & d_{3n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ (d_{11}+\lambda d_{21} & (d_{12}+\lambda d_{22} & \dots & (d_{1n}+\lambda d_{2n} \end{pmatrix}$$

במקום לבצע את המכפלה בסדר $\underline{\underline{E}}_2 \cdot (\underline{\underline{E}}_1 \cdot \underline{\underline{D}})$, ניתן לבצעה כמוכן בסדר

$$(\underline{\underline{E}}_2 \cdot \underline{\underline{E}}_1) \cdot \underline{\underline{D}}$$

מתקבלת מטריצת עזר חדשה

$$\underline{\underline{E}}^{(2)} = \underline{\underline{E}}_2 \cdot \underline{\underline{E}}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & \lambda & 0 \end{pmatrix}$$

הכוללת את שתי הפעולות האלמנטריות, מבוצעות בסדר הרצוי, כלומר קודם תוספת λ כפול השורה השניה לראשונה ואחר כך החלפת השורה הראשונה והשלישית. הדבר ניתן להכללה למספר רב יותר של פעולות אלמנטריות. למשל, אם לאחר שתי הפעולות שנעשו לעיל, רוצים להוסיף γ כפול השורה השניה לשורה הראשונה (החדשה, זו שהיתה קודם שלישית), ניתן הדבר להעשות על ידי מכפלה מלפנים של $\underline{\underline{D}}$ במטריצת העזר החדשה $\underline{\underline{E}}^{(3)}$

$$\underline{\underline{E}}^{(3)} = \underline{\underline{E}}_3 \cdot \underline{\underline{E}}_2 \cdot \underline{\underline{E}}_1 = \underline{\underline{E}}_3 \cdot \underline{\underline{E}}^{(2)} = \begin{bmatrix} 1 & \gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & \lambda & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & \gamma & 1 \\ 0 & 1 & 0 \\ 1 & \lambda & 0 \end{bmatrix}$$

באופן כללי ניתן לסמן

$$\underline{\underline{E}}^{(p)} = \underline{\underline{E}}_p \cdot \underline{\underline{E}}_{p-1} \cdot \ldots \underline{\underline{E}}_2 \cdot \underline{\underline{E}}_1 = \prod_{i=1}^{p} \underline{\underline{E}}_{(p-i+1)} \qquad (\text{א-17})$$

כאשר כל מטריצה $\underline{\underline{E}}_j$ כוללת פעולה אלמנטרית, וסדר המכפלה נקבע על ידי סדר הפעולות הרצויות במטריצות עליה פועלים : קודם הפעולה של $\underline{\underline{E}}_1$, אחר כך של $\underline{\underline{E}}_2$ וכו'.

9.א <u>פתרון מערכת של משוואות ליניאריות סימולטניות</u>

עבור מערכת המשוואות הליניאריות הסימולטניות (14-א'), המבוטאות באמצעות מטריצות במשוואות (15-א') ובצורה מפורשת במשוואה (16-א') נתיחס למקרה בו $m = n$ ונגדיר את המטריצה המורחבת (Augmented Matrix), המתקבלת מהוספת העמודה $\underline{b}$ למטריצה $\underline{\underline{A}}$.

$$(\underline{\underline{A}},\underline{b}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{bmatrix}$$

למערכת של n משוואות ליניאריות ב-n נעלמים :

1. יש פתרון חד-משמעי אם $r(\underline{\underline{A}}) = r(\underline{\underline{A}},\underline{b}) = n$

2. יש אינסוף פתרונות, כאשר $r(\underline{\underline{A}}) = r(\underline{\underline{A}},\underline{b}) < n$

3. אין פתרון כאשר $r(\underline{\underline{A}}) < r(\underline{\underline{A}},\underline{b})$

9.1.א פתרון לפי Gauss

בעזרת פעולות אלמנטריות מביאים את המערכת (16-א') לצורה $\underline{\underline{h}} \cdot \underline{x} = \underline{b}'$, הנתונה על ידי

$$\begin{bmatrix} 1 & h_{12} & h_{13} & \dots & h_{1n} \\ 0 & 1 & h_{23} & \dots & h_{2n} \\ 0 & 0 & 1 & \dots & h_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots 1 & h_{n-1,n} \\ 0 & 0 & 0 & \dots 0 & 1 \end{bmatrix} \cdot \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-1} \\ x_n \end{Bmatrix} = \begin{Bmatrix} b'_1 \\ b'_2 \\ b'_3 \\ \vdots \\ b'_{n-1} \\ b'_n \end{Bmatrix} \qquad (\text{א-19})$$

$\underline{\underline{h}}$ היא מטריצה שלבית (Echelon Matrix), בה כל האברים מתחת לאלכסון הראשי אפס, על האלכסון הראשי 1, ומעל לאלכסון הראשי כעלי ערך כלשהו. b', הוא וקטור חדש, המתקבל מ-$\underline{b}$ על ידי הפעולות האלמנטריות הנעשות על המטריצה $\underline{\underline{A}}$ כדי להפכה ל-$\underline{\underline{h}}$. הפתרון מושג בדרך הבאה : $x_n = b'_n$. מציבים ערך זה כמשוואה ה-(n-1) ומקבלים:

$$x_{n-1} = b'_{n-1} - h_{n-1,n} b'_n$$

מציבים את ערכי x_n ו- x_{n-1} במשוואה ה-(n-2) , פותרים עבור x_{n-2} וכך ממשיכים בהצבה אחורנית (Back Substitution) עד לפתרון כל הנעלמים.

<u>דוגמא</u>

נתונה מערכת משוואות

$$\begin{aligned} 2x_1 + x_2 + 4x_3 &= 16 \\ 3x_1 + 2x_2 + x_3 &= 10 \\ x_1 + 3x_2 + 3x_3 &= 16 \end{aligned} \qquad (20\text{-א})$$

על מנת לקבל עבור x_1 מקדם 1 כשורה הראשונה ואפסים בשתי השורות האחרות נבצע את הפעולות הבאות: נכפול את השורה הראשונה ב-½, נחסר 3 פעמים השורה הראשונה (החדשה) מן השניה, נחסר את השורה הראשונה מהשלישית. הפעולות מכוצעות כמוכן גם על הצד הימני. התוצאה היא :

$$\left.\begin{aligned} x_1 + \frac{1}{2}x_2 + 2x_3 &= 8 \\ \frac{1}{2}x_2 - 5x_3 &= -14 \\ \frac{5}{2}x_2 + x_3 &= 8 \end{aligned}\right\} \qquad (21\text{-א})$$

על מנת לקבל עבור x_2 מקדם 1 כשורה השניה ואפס כשורה השלישית יש לכפול את השורה השורה השניה כ-2 ולחסר 5/2 כפול השורה השניה (החדשה) מן השלישית. התוצאה היא

$$\left.\begin{aligned} x_1 + \frac{1}{2}x_2 + 2x_3 &= 8 \\ x_2 - 10x_3 &= -28 \\ 26x_3 &= 8 \end{aligned}\right\} \qquad (22\text{-א})$$

הפתרון הוא כעת

$$x_3 = 78/26 = 3$$

$$x_2 = -28 + 10 \times 3 = 2$$

$$x_1 = 8 - \frac{1}{2} \times 2 - 2 \times 3 = 1$$

את שלוש הפעולות שהעבירו את משוואות (20-א') לצורה (21-א') אפשר לסכם כמטריצת עדר

$$\underline{\underline{E}}^{(3)} = \underline{\underline{E}}_3 \cdot \underline{\underline{E}}_2 \cdot \underline{\underline{E}}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -3/2 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} \qquad \text{(23-א)}$$

קל להיווכח כי המכפלה

$$\underline{\underline{E}}^{(3)} \cdot \underline{\underline{A}} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -3/2 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & 2 \\ 0 & \frac{1}{2} & -5 \\ 0 & 5/2 & 1 \end{bmatrix} \qquad \text{(24-א)}$$

אמנם נותנת את מטריצת המקדמים של (21-א'). את שתי הפעולות הנותרות ניתן גם כן לכטא בעזרת מטריצות עדר. וסך כל הפעולות כלול במטריצה $\underline{\underline{E}}^{(5)}$ הנתונה על ידי:

$$\underline{\underline{E}}^{(5)} = \underline{\underline{E}}_5 \cdot \underline{\underline{E}}_4 \cdot \underline{\underline{E}}^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5/2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -3/2 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -3 & 2 & 0 \\ 7 & -5 & 1 \end{bmatrix} \qquad \text{(25-א)}$$

כדי לפתור את מערכת המשוואות $\underline{\underline{A}}\ \underline{x} = \underline{b}$

יש לכפול כל צד מלפנים ב-$\underline{\underline{E}}^{(5)}$ כלומר

$$\underline{\underline{E}}^{(5)} \underline{\underline{A}}\ \underline{x} = \underline{\underline{E}}^{(5)} \underline{b} \qquad \text{(26-א)}$$

את המכפלה ניתן לבצע בבת אחת על $\underline{\underline{A}}$ ועל $\underline{b}$ בצורה,

$$\underline{\underline{E}}^{(5)}(\underline{\underline{A}},\underline{b}) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ -3 & 2 & 0 \\ 7 & -5 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 4 & 16 \\ 3 & 2 & 1 & 10 \\ 1 & 3 & 3 & 16 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & 2 & 8 \\ 0 & 1 & -10 & -28 \\ 0 & 0 & 26 & 78 \end{pmatrix} \qquad (\text{א}-27)$$

והתוצאה מכטאת בדיוק את משוואה (22-א').

9.2.א פתרון לפי Gauss-Jordan

השיטה היא למצוא מטריצה $\underline{\underline{E}}$ כזו, שעבורה

$$\underline{\underline{E}} \cdot \underline{\underline{A}} = \underline{\underline{I}} \qquad (\text{א}-28)$$

המטריצה $\underline{\underline{E}}$ המקיימת תנאי זה נקראת המטריצה ההפכית (Inverse Matrix) של $\underline{\underline{A}}$ וסימונה $\underline{\underline{A}}^{-1}$. פתרון מערכת המשוואות הוא אז :

$$\underline{\underline{A}}^{-1} \cdot \underline{\underline{A}} \cdot \underline{x} = \underline{\underline{A}}^{-1} \cdot \underline{b} \qquad (\text{א}-29)$$

כלומר

$$\underline{x} = \underline{\underline{A}}^{-1} \cdot \underline{b} \qquad (\text{א}-30)$$

דוגמא

נשתמש שוב במערכת הנתונה ב-(20-א'). שלוש הפעולות הראשונות, המסתיימות כמצב (21-א') יחזרו גם כאן, הפעולות עבור x_2 , המסתיימות במקדם 1 כשורה שניה ובאפסים כשורות האחרות, נתונות כמטריצות

$$\underline{\underline{E}}_6 \cdot \underline{\underline{E}}_5 \cdot \underline{\underline{E}}_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5/2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -5 & 1 \end{pmatrix} \qquad (\text{א}-31)$$

הפעלתה על המטריצה המורחבת של (21-א') נותנת

$$\begin{bmatrix} 0 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -5 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & 2 & 8 \\ 0 & \frac{1}{2} & -5 & -14 \\ 0 & 5/2 & 1 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 7 & 22 \\ 0 & 1 & -10 & -28 \\ 0 & 0 & 26 & 78 \end{bmatrix} \qquad (32\text{-א})$$

נותר כעת רק לחלק את המשוואה האחרונה ב-26, לחבר 10 כפול השורה השלישית לשניה, ולחסר 7 כפול השורה השלישית מהראשונה. פעולות אלה תסתכמנה במטריצות

$$\underline{\underline{E}}_9 \cdot \underline{\underline{E}}_8 \cdot \underline{\underline{E}}_7 = \begin{bmatrix} 1 & 0 & -7 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/26 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -7/26 \\ 0 & 1 & 10/26 \\ 0 & 0 & 1/26 \end{bmatrix} \qquad (33\text{-א})$$

כל הפעולות יחד מבוטאות במכפלה

$$\underline{\underline{E}}^{(9)} = \prod_{i=1}^{9} \underline{\underline{E}}_{(10-i)} = \begin{bmatrix} 1 & 0 & -7/26 \\ 0 & 1 & 10/26 \\ 0 & 0 & 1/26 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -5 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -3/2 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 3/26 & 9/26 & -7/26 \\ -8/26 & 2/26 & 10/26 \\ 7/26 & -5/26 & 1/26 \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 3 & 9 & -7 \\ -8 & 2 & 10 \\ 7 & -5 & 1 \end{bmatrix} \qquad (34\text{-א})$$

אפשר להיווכח על ידי בדיקה כי $\underline{\underline{E}}^{(9)} \cdot \underline{\underline{A}} = \underline{\underline{I}}$ או $\underline{\underline{E}}^{(9)} = \underline{\underline{A}}^{-1}$ בלומר קבלנו את המטריצה ההפכית.

פתרון המשוואות מתקבל לפי משוואה (30-א')

$$\underline{x} = \underline{\underline{A}}^{-1}\,\underline{b} = \frac{1}{26} \begin{bmatrix} 3 & 9 & -7 \\ -8 & 2 & 10 \\ 7 & -5 & 1 \end{bmatrix} \begin{Bmatrix} 16 \\ 10 \\ 16 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 2 \\ 3 \end{Bmatrix} \qquad (35\text{-א})$$

10.א חישוב המטריצה ההפכית

$\underline{\underline{A}}^{-1}$ קיימת רק עבור מטריצה $\underline{\underline{A}}$ ריבועית, אשר דרגתה שווה לסדר שלה. כלומר, $\underline{\underline{A}}^{-1}$ קיימת עבור $\underline{\underline{A}}$ שהיא $(n \times n)$, רק אם $r(\underline{\underline{A}}) = n$. אם מתקיים תנאי זה, אז יש למטריצה $\underline{\underline{A}}$ רק הפכית אחת, כלומר, יש רק מטריצה אחת $\underline{\underline{A}}^{-1}$ המקיימת $\underline{\underline{A}}^{-1}\cdot\underline{\underline{A}} = \underline{\underline{I}}$.
קל להראות כי המטריצה ההפכית של $\underline{\underline{A}}^{-1}$ היא $\underline{\underline{A}}$. נרשום לשם הנוחיות $\underline{\underline{B}} = \underline{\underline{A}}^{-1}$.
אזי $\underline{\underline{B}}\ \underline{\underline{A}} = \underline{\underline{I}}$, כלומר

$$\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \qquad (36\text{-א})$$

כאשר מבצעים את המכפלות מתקבל

$$\sum_{i=1}^{n} b_{ki}a_{i\ell} = \begin{cases} 1 \text{ if } k = \ell \\ 0 \text{ if } k \neq \ell \end{cases} \qquad (37\text{-א})$$

אם נבצע כעת את המכפלה ההפוכה $\underline{\underline{B}} \cdot \underline{\underline{A}}$ ונשתמש בתוצאה (37-א'), מתקבל

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{1n} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} \Sigma a_{1i}b_{i1} & \cdots & \Sigma a_{1i}b_{in} \\ \vdots & & \vdots \\ \Sigma a_{ni}b_{i1} & \cdots & \Sigma a_{ni}b_{in} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \qquad (38\text{-א})$$

ובכך הראינו כי

$$\underline{\underline{A}}^{-1} \cdot \underline{\underline{A}} = \underline{\underline{A}} \cdot \underline{\underline{A}}^{-1} = \underline{\underline{I}} \qquad (39\text{-א})$$

את חישוב המטריצה ההפכית אפשר לעשות במספר גדול מאד של שיטות, המתחלקות לשני סוגים עיקריים : שיטות ישירות ושיטות איטרטיביות. האחרונות מיועדות לפתרון מערכת של משוואות ליניאריות סימולטניות, כאשר בשלבי הביניים של הפתרון (בכל איטרציה) מתקבל פתרון מקורב, טוב יותר מקודמו. עבור הנושא של תיכנות ליניארי

מעניינות רק השיטות הישירות, אליהן משתייכת גם שיטת הדיאגונליזציה של גאוס (סעיף 9.1.א). שיטה נוספת לחישוב המטריצה ההפכית מבוססת על השיטה הידועה של Cramer לפתרון משוואות. נגדיר עבור מערכת המשוואות (14-א')

$$\underline{\underline{A}}_j = \begin{pmatrix} a_{11} & \cdots & a_{1,j-1} & b_1 & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & b_n & a_{n,j+1} & \cdots & a_{nn} \end{pmatrix} \qquad (40\text{-א})$$

מתקבלת מ- $\underline{\underline{A}}$ בהחלפת העמודה ה-j בוקטור $\underline{b}$. הפתרון הוא

$$x_j = |\underline{\underline{A}}_j| / |\underline{\underline{A}}| \qquad (41\text{-א})$$

הפתרון קיים כמובן רק כאשר $|\underline{\underline{A}}| \neq 0$.

כאשר משווים פתרון זה עם הפתרון $\underline{x} = \underline{\underline{A}}^{-1} \underline{b}$ מתקבל עבור המטריצה ההפכית

$$\underline{\underline{A}}^{-1} = \frac{1}{|\underline{\underline{A}}|} \operatorname{Adj} \underline{\underline{A}} \qquad (42\text{-א})$$

כאשר $\operatorname{Adj} \underline{\underline{A}}$ היא מטריצה המתקבלת מ- $\underline{\underline{A}}$ על ידי החלפת כל אלמנט בקופקטור שלו והחלפת המטריצה (ביצוע Transpose), כלומר

$$\operatorname{Adj} \underline{\underline{A}} = \begin{pmatrix} A_{11} & \cdots & A_{n1} \\ \vdots & & \vdots \\ A_{1n} & & A_{nn} \end{pmatrix} \qquad (43\text{-א})$$

הקופקטור מוגדר במשוואה (5-א'). נחשב בשיטה זו את המטריצה ההפכית של הדוגמה הנתונה במשוואה (20-א').

המטריצה המקורית

$$\underline{\underline{A}} = \begin{pmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix}$$

הקופקטור

$$A_{11} = (-1)^2 \begin{vmatrix} 2 & 1 \\ 3 & 3 \end{vmatrix} = 3;\ A_{21} = (-1)^3 \begin{vmatrix} 1 & 4 \\ 3 & 3 \end{vmatrix} = 9;\ A_{31}=(-1)^4 \begin{vmatrix} 1 & 4 \\ 2 & 1 \end{vmatrix} = -7;$$

$$A_{12} = (-1)^3 \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} = -8;\ A_{22} = (-1)^4 \begin{vmatrix} 2 & 4 \\ 1 & 3 \end{vmatrix} = 2;\ A_{32}=(-1)^5 \begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} = 10$$

$$A_{13} = (-1)^4 \begin{vmatrix} 3 & 2 \\ 1 & 3 \end{vmatrix} = 7;\ A_{23} = (-1)^5 \begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} = -5;\ A_{33}=(1)^6 \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} = 1$$

הדטרמיננט של $\underline{\underline{A}}$ הוא

$$|\underline{\underline{A}}| = \begin{vmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{vmatrix} = 12 + 1 + 36 - 8 - 9 - 6 = 26$$

והמטריצה ההפכית היא

$$\underline{\underline{A}}^{-1} = \frac{1}{26} \begin{pmatrix} 3 & 9 & -7 \\ -8 & 2 & 10 \\ 7 & -5 & 1 \end{pmatrix}$$

תוצאה זהה לזו שנתקבלה במשוואה (34-א').

11.א <u>פתרון בסיסי</u>

כתוב $\underline{\underline{A}}\ \underline{x} = \underline{b}$ כאשר $\underline{\underline{A}}$ היא $(m \times n)$ עם $n > m$, כלומר יש יותר משתנים ממשוואות, וכן ידוע כי $r(\underline{\underline{A}}) = m$, כלומר, דרגת המטריצה במספר המשוואות שבה.

נחלק את $\underline{\underline{A}}$ לשתי מטריצות : $\underline{\underline{B}}$ שהיא $(m \times m)$, ו- $\underline{\underline{R}}$ שהיא $[m\times (n - m)]$ כך שמתקיים $r(\underline{\underline{B}}) = m$. העמודות של $\underline{\underline{B}}$ לקוחות מ- $\underline{\underline{A}}$, אבל אינן מופיעות בהכרח באותו סדר כמו

ב-$\underline{\underline{A}}$. נסמן $\underline{\underline{A}} = (\underline{\underline{B}}\ ,\ \underline{\underline{R}})$.

נחלק גם את הוקטור $\underline{x}$, שבו n רכיבים, לשניים : $\underline{x}_B$, הכולל את הרכיבים המתאימים לעמודות המופיעות ב- $\underline{\underline{B}}$, ו- $\underline{x}_R$ הכולל את הרכיבים המתאימים לעמודות המופיעות ב- $\underline{\underline{R}}$. סדר הרכיבים ב- $\underline{x}_B$ מתאים לסדר הופעת העמודות המתאימות ב- $\underline{\underline{B}}$, ובך ב- $\underline{x}_R$ ביחס ל- $\underline{\underline{R}}$. אז נרשום:

$$\underline{\underline{A}}\ \underline{x} = \underline{\underline{B}}\ \underline{x}_B + \underline{\underline{R}}\ \underline{x}_R = \underline{b} \qquad (\text{א-}44)$$

היות ויש m משוואות בלתי תלויות (שהרי $r(\underline{\underline{A}}) = m$), אפשר לקבוע את ערכיהם של $(n - m)$ משתנים בצורה רצונית, ולפתור עבור שאר m המשתנים. נציב $\underline{x}_R = \underline{0}$. היות ובחרנו את $\underline{\underline{B}}$ כך ש- $r(\underline{\underline{B}}) = m$, קיימת $\underline{\underline{B}}^{-1}$ ואז מהמשוואה $\underline{\underline{B}}\underline{x}_B = \underline{b}$ מתקבל הפתרון

$$\underline{x}_B = \underline{\underline{B}}^{-1}\ \underline{b} \qquad (\text{א-}45)$$

$\underline{x}_B$ נקרא <u>פתרון בסיסי</u> (Basic Solution). הגדרה : כאשר בוחרים מטריצה לא סינגולרית של $(m \times m)$ מתוך $\underline{\underline{A}}$, שהיא $(m \times n)$ ודרגתה m, וכל המשתנים שאינם קשורים בעמודות של מטריצה זאת מושווים לאפס, הפתרון למערכת המשוואות החרשה נקרא פתרון בסיסי.

מספר הפתרונות הבסיסיים האפשריים, בתנאי שאפשר לבחור כל קומבינציה של עמודות עבור $\underline{\underline{B}}$ ולקיים $r(\underline{\underline{B}}) = m$, הוא,

$$\binom{n}{m} = \frac{n!}{m!\ (n-m)!} \qquad (\text{א-}46)$$

למשל, עבור מטריצה של 5 שורות ו-10 עמורות מתקבל

$$\binom{10}{5} = \frac{10!}{5!\ 5!}\ 252$$

פתרון בסיסי נקרא <u>מנוון</u> (Degenerate), כאשר אחד או יותר מהמשתנים הבסיסיים מקבל את הערך אפס.

12.א <u>תלות ליניארית של וקטורים</u>

נסמן ב-E^m את המרחב האויקלידי ה-m מימדי. הוקטור $\underline{a} \in E^m$ הוא קומבינציה ליניארית של $(\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k) \in E^m$ אם אפשר לכתוב

$$\underline{a} = \lambda_1\underline{a}_1 + \lambda_2\underline{a}_2 + \ldots + \lambda_k\,\underline{a}_k \qquad (\text{א-}47)$$

כאשר ה- λ_i סקלרים.

$(\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n) \in E^m$ <u>תלויים ליניארית</u> עם אפשר למצוא מקדמים שלא כולם אפס, כך שמתקיים

$$\lambda_1\underline{a}_1 + \lambda_2\underline{a}_2 + \ldots + \lambda_n\underline{a}_n = \underline{0} \qquad (\text{א-}48)$$

אם שוויון זה קיים <u>רק</u> עבור $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$ אזי הוקטורים <u>כלתי תלויים</u>, פירוש: הוקטורים תלויים ליניארית כאשר אחד מהם נתון על ידי צירוף ליניארי של וקטורים אחרים. נפרט את משוואה (48-א')

$$\left.\begin{array}{l} a_{11}\lambda_1 + a_{21}\lambda_2 + \ldots + a_{n1}\lambda_n = 0 \\ a_{12}\lambda_1 + a_{22}\lambda_2 + \ldots + a_{n2}\lambda_n = 0 \\ \vdots \qquad\qquad\qquad\qquad \vdots \qquad\quad \vdots \\ a_{n1}\lambda_1 + a_{n2}\lambda_2 + \ldots + a_{nn}\lambda_n = 0 \end{array}\right\} \qquad (\text{א-}49)$$

למשוואות אלה יש פתרון טריביאלי, $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$. פתרון נוסף, לא טריביאלי, קיים רק כאשר $|\underline{\underline{A}}| = 0$. כלומר, כאשר $r(\underline{\underline{A}}) = n$ יש פתרון יחיד, והוא הטריביאלי. כאשר $r(\underline{\underline{A}}) < n$ יש אינסוף פתרונות, וכל אחד מהם אפשר למצוא חלק מה- λ_i שאינם אפס.

אם $(\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n) \in E^m$, כאשר $n \geq 2$, היא קבוצת וקטורים התלויים ליניארית, כלומר הם מקיימים את משוואה (48-א') בה לא כל ה- λ_i שווים לאפס, אפשר לכתוב למשל

$$\underline{a}_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{\lambda_i}{\lambda_j} \underline{a}_i \quad ; \ (\lambda_j \neq 0) \qquad \text{(50-א)}$$

<u>ה ע ר ו ת</u>

1. כל תת-קבוצה של קבוצה בלתי תלויה, גם היא בלתי תלויה.

2. כל קבוצה המכילה תת-קבוצה תלויה, גם היא תלויה.

13.א <u>כסיס</u>

<u>בסיס</u> (Basis) ב-E^m הוא קבוצה של m וקטורים בלתי תלויים ליניארית, כך <u>שכל</u> וקטור כ-E^m יכול להכתב כצירוף ליניארי של הקבוצה. אומרים שהבסיס "מכסה" (spans) את כל המרחב ה-m מימדי.

הדוגמה הפשוטה ביותר לכסיס היא m וקטורי יחידה, כל אחד ככיוון ציר אחר :

$$\underline{e}_1 = \begin{Bmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{Bmatrix} \quad ; \quad \underline{e}_2 = \begin{Bmatrix} 0 \\ 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{Bmatrix} \quad \ldots \quad \underline{e}_m = \begin{Bmatrix} 0 \\ . \\ . \\ . \\ . \\ 0 \\ 1 \end{Bmatrix} \qquad \text{(51-א)}$$

אבל זה אינו הבסיס היחידי האפשרי במרחב ה-m מימדי, ויש אינסוף בסיסים אפשריים.

אם $\underline{a}_i$ הוא בסיס ב- E^m , אזי הביטוי של וקטור כלשהו $\underline{d} \in E^m$ בעזרת בסיס זה הוא חד-משמעי, כלומר, יש רק קבוצה אחת של מקדמים λ_i המקיימת

$$\underline{d} = \sum_{i=1}^{m} \lambda_i \cdot \underline{a}_i \qquad (\text{א-}52)$$

<u>הוכחה</u>: נניח שהדבר אינו נכון, אזי קיימת קבוצת מקדמים אחרת, μ_i , המקיימת

$$\underline{d} = \sum_{i=1}^{m} \mu_i \underline{a}_i$$

אז

$$\sum_{i=1}^{m} \lambda_i \underline{a}_i = \sum_{i=1}^{m} \mu_i \underline{a}_i$$

או

$$\sum_{i=1}^{m} (\lambda_i - \mu_i) \underline{a}_i = \sum_{i=1}^{m} \gamma_i \underline{a}_i = 0 \qquad (\text{א-}53)$$

כאשר ה-γ_i הם מקדמים חדשים. היות וה-$\underline{a}_i$ מהווים בסיס, הם בלתי תלויים. כדי לקיים את (53-א') דרוש לכן

$$\gamma_i = 0,\ i = 1, \ldots, m \Longrightarrow \mu_i = \lambda_i,\ i = 1, \ldots, m \qquad (\text{א-}54)$$

ובכך הוכחנו שביטוי $\underline{d}$ בעזרת הבסיס הוא חד-משמעי.

נרשום את (52-א') בצורת :

$$\sum_{i=1}^{m} \lambda_i \underline{a}_i - \underline{d} = \underline{0} \qquad (\text{א-}55)$$

אפשר כעת להחליף את הבסיס, כלומר, לקבל בסיס חדש בו מופיע $\underline{d}$, אם נוציא מהקבוצה במקומו וקטור כלשהו, $\underline{a}_j$, עבורו במשוואה (55-א') קיים $\lambda_i \neq 0$.
אם למשל, קיים $\lambda_m \neq 0$, נוכיח כי הקבוצה $\underline{d}$, $\underline{a}_{m-1}$,..., $\underline{a}_2$, $\underline{a}_1$ היא בסיס, כלומר קבוצה בלתי תלויה. אם הדבר אינו נכון אפשר לכתוב

$$\sum_{i=1}^{m-1} \mu_i \underline{a}_i + \mu_m \underline{d} = \underline{0}$$

נכניס כעת את הביטוי עבור $\underline{d}$ בעזרת הבסיס הקודם, ונקבל :

$$\sum_{i=1}^{m-1} \mu_i \underline{a}_i + \mu_m \sum_{i=1}^{m} \lambda_i \underline{a}_i = \sum_{i=1}^{m-1} (\mu_i + \mu_m\lambda_i) \underline{a}_i + \mu_m\lambda_m \underline{a}_m = \sum_{i=1}^{m} \rho_i \underline{a}_i = \underline{0}$$

כאשר ρ_i הם מקדמים חדשים. אם השוויון האחרון קיים נסתרת ההנחה ש- $(\underline{a}_1 , \underline{a}_2 ,..., \underline{a}_m)$ הוא בסיס. המסקנה היא: גם $(\underline{a}_1 , \underline{a}_2 ,..., \underline{a}_{m-1}, \underline{d})$ הוא בסיס.
כל וקטור ב- E^m יכול להרשם כצירוף ליניארי של הבסיס החדש. למשל, אם היה נתון עבור E^m $\underline{x}$ כי

$$\underline{x} = \sum_{i=1}^{m} \gamma_i \underline{a}_i \qquad \text{(56-א)}$$

והחלפנו בבסיס את $\underline{a}_m$ ב- $\underline{d}$, כך ש-

$$\underline{a}_m = \frac{1}{\lambda_m} \left[\underline{d} - \sum_{i=1}^{m-1} \lambda_i \underline{a}_i \right] \quad ; \quad (\lambda_m \neq 0) \qquad \text{(57-א)}$$

אזי

$$\underline{x} = \sum_{i=1}^{m-1} \gamma_i \underline{a}_i + \frac{\gamma_m}{\lambda_m}\left[\underline{d} - \sum_{i=1}^{m-1} \lambda_i \underline{a}_i\right] = \sum_{i=1}^{m-1}\left[\gamma_i - \frac{\gamma_m}{\lambda_m}\lambda_i\right]\underline{a}_i + \frac{\gamma_m}{\lambda_m}\underline{d} \qquad (58\text{-א})$$

<u>דוגמא:</u>

ב-E^3 מוגדר בסיס באמצעות וקטורי היחידה

$$\underline{e}_1 = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix} \qquad \underline{e}_2 = \begin{Bmatrix} 0 \\ 1 \\ 0 \end{Bmatrix} \qquad \underline{e}_3 = \begin{Bmatrix} 0 \\ 0 \\ 1 \end{Bmatrix}$$

ונתונים וקטורים

$$\underline{d} = \sum_{i=1}^{3} \lambda_i \underline{e}_i = 3\underline{e}_1 + 4\underline{e}_3$$

$$\underline{x} = \sum_{i=1}^{3} \gamma_i \underline{e}_i = 2\underline{e}_1 + \underline{e}_2 + 3\underline{e}_3$$

נבחר להוציא את $\underline{e}_1$ מהבסיס ולהכניס את $\underline{d}$ במקומו. הכיטוי של $\underline{x}$ בעזרת הבסיס החדש $(\underline{d}\ ,\ \underline{e}_2\ ,\ \underline{e}_3)$, לפי משוואה (59-א') הוא

$$\underline{x} = \frac{\gamma_1}{\lambda_1}\underline{d} + \sum_{i=2,3}\left[\gamma_i - \frac{\gamma_1}{\lambda_1}\lambda_i\right]\underline{e}_i =$$

$$= \frac{2}{3}\underline{d} + \left[1 - \frac{2}{3}\cdot 0\right]\underline{e}_2 + \left[3 - \frac{2}{3}\cdot 4\right]\underline{e}_3 = \frac{2}{3}\underline{d} + \underline{e}_2 + \frac{1}{3}\underline{e}_3$$

<u>הערה:</u> אי אפשר היה להחליף בבסיס את $\underline{e}_2$ ב- $\underline{d}$, משום ש- $\lambda_2 = 0$.

14.א חישוב שינויי המטריצה ההפכית בעזרת מכפלות

נתונה מטריצת הבסיס $\underline{\underline{B}}$, וידועה המטריצה ההפכית שלה $\underline{\underline{B}}^{-1}$. מחליפים כעת עמודה אחת , $\underline{b}_r$, של המטריצה $\underline{\underline{B}}$, בוקטור $\underline{a}$ ומתקבלת המטריצה $\underline{\underline{B}}_a$. מבוקשת המטריצה ההפכית $\underline{\underline{B}}_a^{-1}$. אפשר כמובן להתחיל את החישוב מחדש עם $\underline{\underline{B}}_a$, ולא להשתמש ב-$\underline{\underline{B}}^{-1}$. אפשרות זו אינה יעילה, ובמקומה משתמשים במכפלה של מטריצת עזר. שיטה זו של מעבר מהפכית של בסיס אחד להפכית של בסיס חדש, השונה מהקודם בעמודה אחת, נקראת PFI, or Product Form of the Inverse .

$$\underline{\underline{B}} = (\underline{b}_1 , \ldots , \underline{b}_{r-1} , \underline{b}_r , \underline{b}_{r+1} , \ldots , \underline{b}_m)$$

$$\underline{\underline{B}}_a = (\underline{b}_1 , \ldots , \underline{b}_{r-1} , \underline{a} , \underline{b}_{r+1} , \ldots , \underline{b}_m)$$

נבטא את $\underline{a}$ בעזרת הבסיס המקורי

$$\underline{a} = \sum_{i=1}^{m} y_i \underline{b}_i = \sum_{i=1}^{r-1} y_i \underline{b}_i + y_r \underline{b}_r + \sum_{i=r+1}^{m} y_i \underline{b}_i \qquad (\text{א-}59)$$

וכעת נבטא את $\underline{b}_r$ בעזרת הבסיס החדש

$$\underline{b}_r = \sum_{i=1}^{r-1} \frac{-y_i}{y_r} \underline{b}_i + \frac{1}{y_r} \underline{a} + \sum_{i=r+1}^{m} \frac{-y_i}{y_r} \underline{b}_i \qquad (\text{א-}60)$$

משוואה זו ניתן לכתוב בצורה

$$\underline{b}_r = \begin{bmatrix} b_{11} & \cdots & b_{1,r-1} & a_1 & b_{1,r+1} & \cdots & b_{1m} \\ \vdots & & & & & & \vdots \\ \underline{b}_{m1} & \cdots & b_{m,r-1} & a_m & b_{m,r+1} & \cdots & b_{m,m} \end{bmatrix} \begin{Bmatrix} -y_1/y_r \\ \vdots \\ -y_{r+1}/y_r \\ 1/y_r \\ -y_{r+1}/y_r \\ \vdots \\ -y_m/y_r \end{Bmatrix} \qquad (\text{א-}61)$$

או בכתיב מטריצות

$$\underline{b}_r = \underline{\underline{B}}_a \, \underline{\eta} \qquad (\text{א-}62)$$

כאשר $\underline{\eta}$ הוא הוקטור הרשום מימין ב-(61-א').

אם נגדיר כעת מטריצה לפי

$$\underline{\underline{E}} = (\underline{e}_1 , \underline{e}_2, \ldots, \underline{e}_{r-1} , \underline{\eta} , \underline{e}_{r+1} , \ldots , \underline{e}_m) \qquad (\text{א-}63)$$

כלומר, מטריצה יחידה מדרגה m, בה הוחלפה העמודה ה-r בוקטור $\underline{\eta}$ שהוגדר קודם, אזי

$$\underline{\underline{B}} = \underline{\underline{B}}_a \cdot \underline{\underline{E}}$$

נכפול משוואה זו מלפנים ב- $\underline{\underline{B}}_a^{-1}$ ומאחור ב- $\underline{\underline{B}}^{-1}$ ונקבל

$$\underline{\underline{B}}_a^{-1} \, \underline{\underline{B}} \, \underline{\underline{B}}^{-1} = \underline{\underline{B}}_a^{-1} \, \underline{\underline{B}}_a \, \underline{\underline{E}} \, \underline{\underline{B}}^{-1}$$

או

$$\underline{\underline{B}}_a^{-1} = \underline{\underline{E}}\ \underline{\underline{B}}^{-1} \qquad (\text{א-}65)$$

וזו הדרך לחישוב $\underline{\underline{B}}_a^{-1}$ מתוך $\underline{\underline{B}}^{-1}$ בעזרת המטריצה $\underline{\underline{E}}$.

אם כעת מבצעים שינויים נוספים, יוצרים וקטור $\underline{\eta}$ חדש, ממנו בונים מטריצה חדשה, ומתקבל

$$\underline{\underline{B}}_{a_2}^{-1} = \underline{\underline{E}}_2\ \underline{\underline{B}}_{a_1}^{-1} = \underline{\underline{E}}_2\ \underline{\underline{E}}_1\ \underline{\underline{B}}^{-1} \qquad (\text{א-}66)$$

וכך הלאה. לכל שינוי יש לזכור וקטור באורך m, וכן את המקום בו מבוצעת ההחלפה במטריצה הקודמת, כלומר את האינדקס r של מקום השינוי.

אפשר גם לכתוב עבור מטריצה כלשהי $\underline{\underline{C}}$ מסדר m כי

$$\underline{\underline{C}}^{-1} = \underline{\underline{E}}_m \cdot \underline{\underline{E}}_{m-1} \quad \cdots\cdots \quad \underline{\underline{E}}_2 \cdot \underline{\underline{E}}_1 \cdot \underline{\underline{I}} = \prod_{j=1}^{m} \underline{\underline{E}}_{m-j+1} \qquad (\text{א-}67)$$

כלומר, להחחיל ממטריצת היחידה $\underline{\underline{I}}$, השווה גם להפכית שלה, להחליף בכל שלב עמודה אחת שלה בעמודה המתאימה של $\underline{\underline{C}}$, לחשב את הוקטור $\underline{\eta}$, לבנות את המטריצה $\underline{\underline{E}}$ המתאימה, ואחרי m החלפות מבצעים m מכפלות מטריצות ומקבלים את $\underline{\underline{C}}^{-1}$.
כל מטריצת עזר נתונה על ידי

$$\underline{\underline{E}}_j = (\underline{e}_1\ ,\ \cdots\cdots\ ,\ \underline{e}_{j-1}\ ,\ \underline{\eta}_j\ ,\ \underline{e}_{j+1}\ ,\ \cdots\ ,\ \underline{e}_m) \qquad (\text{א-}68)$$

כאשר

$$\underline{\eta}_j = \left[\frac{-y_{1,j}}{y_{j,j}}\ ,\ \cdots\ ,\ \frac{-y_{j-1,j}}{y_{j,j}}\ ,\ \frac{1}{y_{j,j}}\ ,\ \frac{-y_{j+1,j}}{y_{j,j}}, \cdots, \frac{-y_{m,j}}{y_{j,j}} \right] \qquad (\text{א-}69)$$

כאשר

$$\underline{y}_j = \underline{\underline{B}}_{j-1}^{-1}\underline{a}_j \qquad (70.1\text{-א})$$

הוא וקטור המקדמים של ההחלפה במקום ה-j.

המקדם $y_{i,j}$ נתון על ידי כפל השורה ה-i של הבסיס האחרון, $\underline{\underline{B}}_{j-1}^{-1}$ בעמודה הנכנסת לבסיס במקום ה-j, $\underline{a}_j$, כלומר,

$$y_{i,j} = \sum_{k=1}^{m} (b_{j-1}^{-1})_{i,k}(a_j)_k \qquad (70.2\text{-א})$$

<u>דוגמא:</u>

נתונה $\underline{\underline{B}}_1 = (\underline{b}_1 , \underline{b}_2 , \underline{b}_3)$ וידוע כי

$$\underline{\underline{B}}_1^{-1} = \begin{pmatrix} 2 & 0 & 4 \\ 5 & 1 & 6 \\ 7 & 9 & 3 \end{pmatrix}$$

נחליף כעת את $\underline{b}_2$ בוקטור $\underline{a}_2 = (1 ,8 ,6)$ ונקבל $\underline{\underline{B}}_a = (\underline{b}_1 , \underline{a}_2 , \underline{b}_3)$ המקדמים הם

$$\underline{y}_2 = \underline{\underline{B}}_1^{-1}\ \underline{a}_2 = \begin{pmatrix} 2 & 0 & 4 \\ 5 & 1 & 6 \\ 7 & 9 & 3 \end{pmatrix} \begin{Bmatrix} 1 \\ 8 \\ 6 \end{Bmatrix} = \begin{Bmatrix} 26 \\ 49 \\ 97 \end{Bmatrix}$$

ולכן

$$\underline{\eta}_2 = \left(\frac{-26}{49} ; \frac{1}{49} ; \frac{-97}{49} \right)$$

ומטריצת העזר היא

$$\underline{\underline{E}} = \begin{pmatrix} 1 & -26/49 & 0 \\ 0 & 1/49 & 0 \\ 0 & -97/49 & 1 \end{pmatrix}$$

לכן ההפכית החדשה היא

$$\underline{\underline{B}}_a^{-1} = \underline{\underline{E}}\ \underline{\underline{B}}^{-1} = \begin{pmatrix} 1 & -26/49 & 0 \\ 0 & 1/49 & 0 \\ 0 & -97/49 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 4 \\ 5 & 1 & 6 \\ 7 & 9 & 3 \end{pmatrix} = \frac{1}{49} \begin{pmatrix} -32 & -26 & 40 \\ 5 & 1 & 6 \\ -142 & 344 & -435 \end{pmatrix}$$

נראה כעת כיצד מחשבים מטריצה הפכית של מטריצה נתונה בעזרת מכפלות של מטריצות עזר, בהתאם למשוואה (67-א').

<u>דוגמא:</u>

נתונה המטריצה של משוואות (20-א')

$$\underline{\underline{A}} = \begin{pmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix} = [\underline{a}_1\ ;\ \underline{a}_2\ ;\ \underline{a}_3]$$

ויש למצוא את המטריצה ההפכית $\underline{\underline{A}}^{-1}$. נתחיל עם מטריצת יחידה של (3×3), השווה גם להפכית שלה, $\underline{\underline{I}} = \underline{\underline{I}}^{-1}$. נכניס כעת לתוך $\underline{\underline{I}}$ את העמודה הראשונה של $\underline{\underline{A}}$, $\underline{a}_1$ ונקבל את המטריצה

$$\underline{\underline{A}}_1 = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

נחשב את הוקטור

$$\underline{y}_1 = \underline{\underline{I}}^{-1}\,\underline{a}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} 2 \\ 3 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 2 \\ 3 \\ 1 \end{Bmatrix}$$

וממנו את הוקטור

$$\underline{\eta}_1 = \left(\frac{1}{y_{1,1}} \; ; \; \frac{-y_{2,1}}{y_{1,1}} \; ; \; \frac{-y_{3,1}}{y_{1,1}} \right) = \left(\frac{1}{2} \; ; \; \frac{-3}{2} \; ; \; \frac{-1}{2} \right)$$

מטריצת העזר הראשונה מתקבלת מהכנסת וקטור זה בעמודה הראשונה (משום שהחלפנו את $\underline{a}_1$ בעמודה הראשונה של $\underline{\underline{I}}$) של מטריצת יחידה

$$E_1 = \begin{bmatrix} 1/2 & 0 & 0 \\ -3/2 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix}$$

וכעת ההפכית של $\underline{\underline{A}}_1$ היא $\underline{\underline{A}}_1^{-1} = \underline{\underline{E}}_1 \cdot \underline{\underline{I}}^{-1} = \underline{\underline{E}}_1$.

כעת נחליף את העמודה השניה של $\underline{\underline{A}}_1$ בעמודה השניה של $\underline{\underline{A}}$, כלומר נכניס במקום וקטור היחידה את הוקטור $\underline{a}_2$. מתקבלת המטריצה

$$\underline{\underline{A}}_2 = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 2 & 0 \\ 1 & 3 & 1 \end{bmatrix}$$

המשך החישובים, כמו קודם:

$$\underline{y}_2 = \underline{\underline{A}}_1^{-1}\,\underline{a}_2 = \begin{bmatrix} 1/2 & 0 & 0 \\ -2/3 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix} \begin{Bmatrix} 1 \\ 2 \\ 3 \end{Bmatrix} = \begin{Bmatrix} 1/2 \\ 1/2 \\ 5/2 \end{Bmatrix}$$

$$\underline{\eta}_2 = \left(\frac{-1/2}{1/2} \; ; \; \frac{1}{1/2} \; ; \; \frac{-5/2}{1/2} \right) = (-1 \; ; \; 2 \; ; \; -5)$$

$$\underline{\underline{E}}_2 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -5 & 1 \end{pmatrix}$$

$$\underline{\underline{A}}_2^{-1} = \underline{\underline{E}}_2 \, \underline{\underline{A}}_1^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -5 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1/2 & 0 & 0 \\ -3/2 & 1 & 0 \\ -1/2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 0 \\ -3 & 2 & 0 \\ 7 & -5 & 1 \end{pmatrix}$$

נותר כעת להכניס את $\underline{a}_3$ במקום וקטור היחידה בעמודה השלישית של $\underline{\underline{A}}_2$, ומתקבלת המטריצה מקורית $\underline{\underline{A}}$.

$$\underline{y}_3 = \underline{\underline{A}}_2^{-1} \, \underline{a}_3 = \begin{pmatrix} 2 & -1 & 0 \\ -3 & 2 & 0 \\ 7 & -5 & 1 \end{pmatrix} \cdot \begin{Bmatrix} 4 \\ 1 \\ 3 \end{Bmatrix} = \begin{Bmatrix} 7 \\ -10 \\ 26 \end{Bmatrix}$$

$$\underline{\eta}_3 = \left(\frac{-7}{26} \; ; \; \frac{-(-10)}{26} \; ; \; \frac{1}{26} \right) = \left(\frac{-7}{26} \; ; \; \frac{10}{26} \; ; \; \frac{1}{26} \right)$$

$$\underline{\underline{E}}_3 = \begin{pmatrix} 1 & 0 & -7/26 \\ 0 & 1 & 10/26 \\ 0 & 0 & 1/26 \end{pmatrix}$$

$$\underline{\underline{A}}^{-1} = \underline{\underline{E}}_3 \, \underline{\underline{A}}_2^{-1} = \begin{pmatrix} 1 & 0 & -7/26 \\ 0 & 1 & 10/26 \\ 0 & 0 & 1/26 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -3 & 2 & 0 \\ 7 & -5 & 1 \end{pmatrix} = \begin{pmatrix} 3/26 & 9/26 & -7/26 \\ -8/26 & 2/26 & 10/26 \\ 7/26 & -5/26 & 1/26 \end{pmatrix}$$

תוצאה זהה לתוצאות שהתקבלו בשיטות האחרות.

15.א מספרים אופייניים ומטריצות מוגדרות

נגביל את הדיון למטריצות ריבועיות סימטריות, כלומר, מטריצות $\underline{\underline{A}}$ בהן $a_{ij} = a_{ji}$. המספרים האופייניים (Characteristic Numbers) של מטריצה סימטרית $\underline{\underline{A}}$ הם הערכים λ המתקבלים מפתרון המערכת של n המשוואות

$$\left.\begin{array}{l} a_{11}\, x_1 + \ldots + a_{1\,n}\, x_n = \lambda x \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ a_{n1}\, x_1 + \ldots + a_{nn}\, x_n = \lambda x_n \end{array}\right\} \qquad (71\text{-א})$$

בכתיב מטריצות

$$\underline{\underline{A}}\,\underline{x} = \lambda \underline{x} \qquad (72\text{-א})$$

או

$$(\underline{\underline{A}} - \lambda \underline{\underline{I}})\underline{x} = \underline{0} \qquad (73\text{-א})$$

למערכת זו פתרון לא טריביאלי רק אם דטרמיננט המקדמים מתאפס, כלומר, אם קיים

$$\begin{vmatrix} (a_{11}-\lambda) & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & (a_{22}-\lambda) & \cdots\cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots\cdots & (a_{nn}-\lambda) \end{vmatrix} = 0 \qquad (74\text{-א})$$

λ הוא לכן פתרון של פולינום מדרגה n, הנקרא המשוואה האופיינית (Characteristic Equation) של המטריצה $\underline{\underline{A}}$. בדרך כלל יש n פתרונות של λ, $(\lambda_1,\ldots,\lambda_n)$, שהם המספרים האופיניים של המטריצה.

מטריצה נקראת מוגדרת חיובית (Positive Definite) אם ורק אם כל המספרים האופי-ניים שלה חיוביים, והיא מוגדרת שלילית (Negative Definite) כאשר כל המספרים האופיניים שליליים. מטריצה היא מוגדרת חיובית למחצה (Positive semi Definite) אם המספרים האופיניים כולם לא-שליליים, אם כי יכולים להיות מהם השווים לאפס.

נסמן ב-$\underline{\underline{D}}_m$ את המטריצה המתקבלת על ידי מחיקת כל האברים שאינם ב-m השורות הראשונות ו-m העמודות הראשונות של $\underline{\underline{A}}$. $\underline{\underline{D}}_m$ היא ריבועית וסימטרית.

אפשר לראות כי $\underline{\underline{A}}$ היא מוגדרת חיובית אם ורק אם מתקיים

$$|D_m| > 0 \qquad m = 1,\ldots,n \qquad (\text{א-}75)$$

כלומר, הדטרמינטים של כל המטריצות $\underline{\underline{D}}_m$ חיוביים. עבור m=1 מתקבל $|a_{11}| > 0$, עבור m=2 מתקבל $a_{11}\cdot a_{22} - a_{12}^2 > 0$ וכך הלאה, עד שעבור m=n מתקבל $|\underline{\underline{A}}| > 0$.

אם $\underline{\underline{A}}$ מוגדרת חיובית אזי המכפלה $\underline{x}^T\,\underline{\underline{A}}\,\underline{x}$ חיובית עבור בחירה כלשהי של רכיבי הוקטור $\underline{x}$, כלומר, הצורה הריבועית (Quadratic Form) $\underline{x}^T\,\underline{\underline{A}}\,\underline{x}$ קמורה (Convex)

מקורות לנספח א'

Hildebrand, F.B., "Methods of Applied Mathematics", Prentice-Hall, 1963.

Westlake, J.R., "A Handbook of Numerical Matrix Inversion and Solution of Linear Equations", Wiley, 1968.

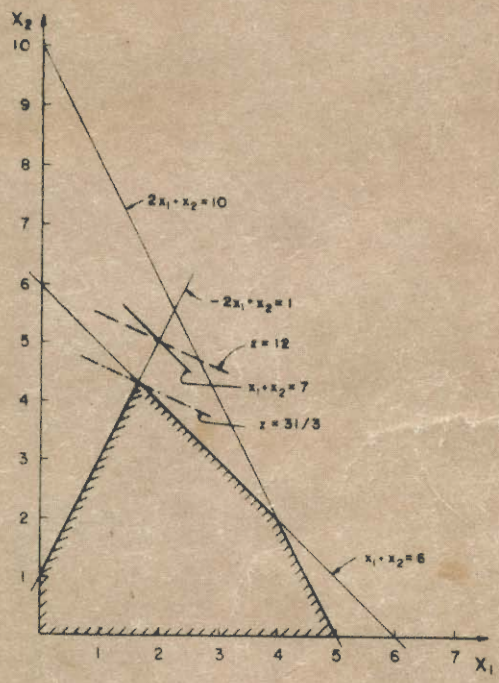
ניתוח מערכות - א. שמיר

רשימת תיקונים - 1.11.85

עמ'	שורה (ל=מלמטה)	כתוב	צריך להיות
34	מש. (4-12)		$j=1\ldots,n$ $\quad \frac{\partial F(\underline{x}^*,\underline{\lambda}^*)}{\partial x_j}=0$
35	בסוף	∇f ו- ∇g_i	∇f ו- $\underline{\nabla g_i}$
39	מש. (4-28)	$\sum_{j=1}^{n}$	$\prod_{j=1}^{n}$
48		מספר הסעיף 5.2	צריך להיות 5.2.1
49	5	$x_o=$	$\underline{x}_o=$
50	ראש העמוד		להוסיף: 5.2.2 אקסטרמום בנוסחות אילוצים
55	מש. (5-11)	ext z =	ext f=
60	6		חסר λ לפני הסוגריים ה-
60	10	a_2 במכנה	a_3
96	טבלו ראשון	הציר (3) אינו בכון	צ"ל (1) בשורה השלישית, עמודה ראשונה
98	מש. (6-26)	a_j	$\underline{a}_j$
101	מש. (6-43)	Z	$\hat{Z}$
101	מש. (6-44)	Z	$\hat{Z}$
107	ל-4	בצד ימין: 2- ו- 4	בצד ימין: 2- ו- 2
122	טבלו ראשון	בשורת W : 1 בעמודה 13	למחוק בעמודה 13 ולהכניס 1 בעמודה 9
124	טבלו ראשון	עמודה 8 שורה 2: 2/5 עמודה 11 שורה 6: 2 עמודה 9 שורת z: 1	-2/5 -2 -1

עמ'	שורה (ל=מלמטה)	כתוב	צריך להיות
162	טבלה אחרונה		יש להחליף בכל מקום m ב- n ולהיפך.
165	טבלה	שורה עליונה: 11 2 1	1 1 2 1
165	ל- 3	עמודה 2 ושורה 3	עמודה 3 ושורה 3
165	ל- 1	בחרנו עמודה 2 ובה את התא בשורה הרביעית	בחרנו עמודה 3, ובה את התא בשורה השלישית.
187	ציור 7.1	המסלול האופטימלי עובר דרך I	הוא צריך להיות A-D-H-L-B והחץ ליד צומת D , הצמוד למשבצת עם 8 בתוכה, צריך להורות אל H.
188	ל- 10	ומקבלים A-D-I-L-B	ומקבלים A-D-H-L-B
189	ציור 7.2	המסלול האופטימלי הוא B-L-I-D-A	המסלול הנכון הוא B-L-H-D-A
196	ל- 3	(K+1) מצבים של המערכת S(0)	(K+1) שלבים. (0) s
197	מש. (7-01)	S גדולה, פעמיים	s קטנה
197	מש. (7-11)	S גדולה	s קטנה
204	ציור 7.5		להחליף מקום "כן" ו"לא" הימניים
210	7	עבור p	עבור p_1
210	ל- 4	פונקציה עולה	פונקציה יורדת
211	מש. (7-32)	g_3d_3/Q	q_3d_3/Q
220	מטריצה תחתונה	שורה עליונה (d11+d12) (d12+d22)...(d1n+drn)	(d11+λd21) (d12+λd22) (d1n+λd2n)
229	מש' 41-א	$X_j=A_j/\lvert A\rvert$	$X_j=\lvert A_j\rvert/\lvert A\rvert$
166	הסבר מתחת לטבלה העליונה	נבחר לפיהם את	יכולנו לבחור לפיהם את העמודה הראשונה ובה את התא הראשון (שכן יכולנו לבחור גם עמודה רביעית תא ראשון). נשבץ שם
96.	טבלה אחרון	13/15	-1/5
97	" "	13/15	-1/5

"מכלול" בהוצאת אס"ט



מכלול