0. Задача на линейното оптимиране. Примери. Канонична форма.

Предмет на *математическото оптимиране* е изучаването на следната задача: да се намери оптималната стойност (минимум или максимум) на някаква функция $f(x)$ за стойности на аргумента x, които принадлежат на някакво подмножество X на пространството $\mathbb{R}^n$.

Настоящият курс е посветен изцяло на един основен дял на математическото оптимиране, наречен *линейно оптимиране*. Линейното оптимиране е един от най-добре развитите клонове на приложната математика. Много практически задачи могат да се формулират като задачи на линейното оптимиране, а линейните модели се използуват най-вече в икономическият анализ и планирането.

Задачата на линейното оптимиране се състои в оптимизиране на линейна функция (наречена целева функция) върху множество от линейни ограничения, които могат да са както равенства, така и неравенства по отношение на променливите.

Теоретичните основи на линейното оптимиране се полагат с изучаването на линейните неравенства, което може да се проследи назад във времето до една работа на Фурие от 1826 г. По-късно много математици доказват редица частни случаи на най-важния резултат на линейното оптимиране – т.нар. теорема за двойнственост. Приложната страна на теорията започва да се разработва през 1939 г. от руския математик Канторович, който пръв забелязва практическата важност на някои класове линейни оптимизационни задачи и пръв дава алгоритъм за тяхното решаване. За съжаление, години наред работите на Канторович остават непознати на Запад и незабелязани на Изток. Силен тласък в развитието на приложната страна на линейното оптимиране дават работите на Данциг, който през 1947 г. разработва симплекс метода за решаване на линейни задачи, които възникват в планирането на въздушни операции на военно-въздушните сили на САЩ. За първи път симплекс методът е публикуван през 1951 г. и си остава най-широко използваният алгоритъм за решаване на линейни задачи. По-късно, с развитието на компютърните технологии, се създават редица програмни реализации на този и други алгоритми за решаване на линейни задачи.

През същата 1951 г., през която Данциг публикува симплекс метода, в своя публикация Купман показва, че линейното оптимиране представлява подходящ модел за анализ на класически икономически теории. През 1975 г. Шведската Кралска Академия на Науките присъжда Нобеловата награда за икономика на Канторович и Купман “за техния принос в теорията за оптималното разпределение на ресурси”. Както се вижда, въпросната Академия разглежда работата на Данциг като твърде математическа за присъждане на наградата по икономика (а знаем, че не се присъжда Нобелова награда за математика).

Целта на курса е да развие основната теория на линейното оптимиране и да представи симплекс метода за решаване на линейни задачи. Изложението ще бъде едновременно алгоритмично и геометрично, като ще бъдат дадени икономически интерпретации на някои аспекти на линейното оптимиране.

За да дадем най-обща представа за задачите на линейното оптимиране за начало ще разгледаме три стандартни примера на такива задачи.

0.1. Примери за линейни задачи

Пример 1 (Транспортна задача). Фирма трябва да транспортира дадена стока (например електроенергия) от m производителя до n потребителя. Да предположим, че a_i единици от стоката се предлагат от i-я производител, $i = 1, 2, \dots, m$, а b_j единици се търсят от j-я потребител, $j = 1, 2, \dots, n$. Да предположим, че общото количество от стоката предлагано от производителите е равно на общото количество търсено от потребителите, т.е., $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$. Ако цената на превоза на единица от стоката от производителя i до потребителя j е c_{ij}, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$, колко единици от стоката би трябвало да се превозят между всяка двойка производител—потребител, така че общите транспортни разходи да бъдат минимални?

Ако означим с x_{ij} единиците от стоката превозвани от производителя i до потребителя j, можем да формулираме тази задача като следната задача на линейното оптимиране:

$$\text{да се минимизира} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}$$

$$\text{при ограничения} \quad \sum_{j=1}^{n} x_{ij} = a_i\,, \quad i = 1, 2, \dots, m\,, \tag{0.1}$$

$$\sum_{i=1}^{m} x_{ij} = b_j\,, \quad j = 1, 2, \dots, n\,, \tag{0.2}$$

$$x_{ij} \geqslant 0\,, \quad i = 1, 2, \dots, m\,, \quad j = 1, 2, \dots, n\,.$$

Целевата функция, която се минимизира представлява общите транспортни разходи. Всяко от ограниченията равенства (0.1) отговаря на изискването общото количество на стоката, превозено от производителя i до всички потребители да бъде равно на количеството предлагано от този производител. По подобен начин, ограниченията равенства (0.2) отговарят на изискването, че търсеното количество b_j от потребителя j се обезпечава от количествата превозени до него от всички производители. Да отбележим, че ограниченията за неотрицателност на превозените количества са съществени, т.к. в противен случай фирмата може да спести пари като превози отрицателни количества по някои от пътищата.

Транспортната задача е задача на линейното оптимиране, която има специфична структура: коефициентите на променливите в ограниченията равенства (0.1) и (0.2) са или нули, или единици. Както ще видим по-късно, транспортната задача е частен случай на т.нар. задачи за потоци в мрежи. Тя също така добре илюстрира защо линейните задачи, които се решават в практиката често са с голяма размерност, т.е. имат много променливи и много ограничения. Ако всяко от m и n е равно на 100, множеството от ограничения на горната задача се задава от 200 уравнения на 10 000 неотрицателни променливи. Благодарение на специфичната им структура много големи транспортни задачи (с няколко десетки милиона променливи) могат да бъдат решени за разумно компютърно време.

Пример 2 (Задача за диета). Така наричаме задачата за определяне на най-евтина диета, която да задоволява определени минимални дневни потребности от калории и хранителни съставки като белтъчини, калций, желязо и витамини. Да допуснем, че имаме n на брой различни храни в наличност и че диетата трябва да задоволява дневният минимум от m на брой хранителни съставки. Да означим с c_j единичната цена на j-та храна, с b_i минималната дневна потребност от i-та хранителна съставка и с a_{ij} количеството от хранителната съставка i осигурявано от единица от храната j.

Ако положим x_j да бъде броят единици от храната j включени в диетата, то най-евтината диета може да се намери при решаване на линейната задача:

$$
\begin{aligned}
&\text{да се минимизира} \quad && \sum_{j=1}^{n} c_j x_j \\
&\text{при ограничения} \quad && \sum_{j=1}^{n} a_{ij} x_j \geqslant b_i\,, \quad i = 1, 2, \ldots, m\,, \\
& && x_j \geqslant 0\,, \quad j = 1, 2, \ldots, n\,.
\end{aligned}
$$

Пример 3 (Задача за оптимално разпределение на ресурси или още задача за смесено производство). Даден производител е в състояние да произведе n на брой различни стоки използвайки за целта m на брои различни ресурси. Това може да са часове за изработване, време за работа на различни машини, или пък количества от различни материали. Нека означим с c_j печалбата (доходът минус производствените разходи) която се получава от произведена единица от стоката j. Нека b_i бъде количеството от наличен ресурс i и a_{ij} бъде количеството от ресурса i използвано за производството на единица от стоката j. Задачата, която стои пред производителя е да се направи план на производството (т.е. да се определи какво количество от всяка стока да бъде произведено), така че общата печалба да бъде максимална.

Ако означим с x_j броят единици от стоката j които се произвеждат, то тази линейна задача може да се формулира като:

$$
\begin{aligned}
&\text{да се максимизира} \quad && \sum_{j=1}^{n} c_j x_j \\
&\text{при ограничения} \quad && \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i\,, \quad i = 1, 2, \ldots, m\,, \\
& && x_j \geqslant 0\,, \quad j = 1, 2, \ldots, n\,.
\end{aligned}
$$

0.2. Канонична форма на задачата на линейното оптимиране

Както видяхме от разгледаните примери, в линейните задачи имаме променливи, чиито стойности трябва да се определят по някакъв оптимален начин. Целта винаги е да се минимизира или максимизира някаква линейна функция на тези променливи. Тази функция се нарича целева функция. В допълнение на целевата функция, примерите имаха също така и ограничения. Наистина, някои от тях бяха много прости, като например изискването за неотрицателност на променливите, докато други бяха по-сложни. Във всички случаи обаче разглежданите ограничения бяха равенства или неравенства, свързани с някакви линейни комбинации на променливите.

От казаното е ясно, че в общия си вид задачите на линейното оптимиране могат да бъдат за търсене на максимум, или на минимум. Множеството им от ограничения може да се задава както с равенства, така и с неравенства във всяка от посоките ($\leqslant$ и $\geqslant$). Върху някои от променливите може да е наложено условие за неотрицателност, върху други не.

За да се улесни и уеднакви подхода към задачите на линейното оптимиране, те се записват в една от следните две форми

$$\text{чрез неравенства:}\quad \min \quad \begin{array}{l} z = c^{\mathrm{T}} x \\ Ax \leqslant b \\ x \geqslant 0 \end{array} \qquad \text{или в канонична форма:} \quad \min \quad \begin{array}{l} z = c^{\mathrm{T}} x \\ Ax = b \\ x \geqslant 0\,, \end{array}$$

където A и в двата случая означава $(m \times n)$ матрица, състояща се от координатите на векторите, задаващи ограниченията, $c \in \mathbb{R}^n$ е векторът на целевата функция, $b \in \mathbb{R}^m$ е векторът на дясната част и $x \in \mathbb{R}^n$ е n-мерен вектор на променливите.

Тези две форми са напълно еквивалентни една на друга и всяка линейна задача може да се приведе в тях като се използват следните прости преобразувания:

• Всяка *свободна променлива*, т.е. променлива x_j, върху която в задачата няма наложено условие за неотрицателностр може да се замени с двойка неотрицателни променливи, $x'_j \geqslant 0$ и $x''_j \geqslant 0$ като се запише

$$x_j = x'_j - x''_j\,.$$

• Посоката на неравенство може да се промени чрез умножаване на двете му страни с (-1).

• Неравенство от вида

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i$$

може да се превърне в равенство чрез добавяне на нова неотрицателна променлива x_{n+i}, наречена *допълнителна променлива*

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j, \quad x_{n+i} \geqslant 0.$$

На допълнителните променливи често може да се даде икономическа или физическа интерпретация. Например, ако добавим допълнителна променлива x_{n+i} към i-то неравенство в задачата за смесено производство, за да го превърнем в равенство, то в този случай x_{n+i} представлява количеството от ресурса i, което остава неизразходвано при смесеното производство.

• Равенство от вида

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \text{ може да се замени с двойка неравенства } \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i, \quad \sum_{j=1}^{n} a_{ij} x_j \geqslant b_i.$$

• Накрая, максимизирането на линейната функция $c^{\mathrm{T}} x$ е еквивалентно на минимизирането на функцията $-c^{\mathrm{T}} x$.

Да забележим че транспортната задача е в канонична форма, докато задачите от другите два примера са по-скоро в първата форма – чрез неравенства.

По-нататък в курса често ще считаме, че задачата на линейното оптимиране, която разглеждаме, вече е приведена в канонична форма.

1. Канонично многостенно множество. Върхове и базисни допустими решения.

Както вече казахме, всяка задача на линейното оптимиране може да се приведе в т.нар. канонична форма:

$$
\begin{aligned}
\min \quad & z = c^{\mathrm{T}} x \\
& Ax = b\,, \\
& x \geqslant 0\,,
\end{aligned}
$$

където A означава $(m \times n)$ матрица, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$ е n-мерен вектор на променливите. По-нататък ще работим с тази форма на линейната задача и ще я наричаме задача канонична задача.

Съвкупност от числени стойности на променливите $(x_1, \ldots, x_n)$ се нарича *решение* на задачата. Решение $(x_1, \ldots, x_n)$ се нарича *допустимо*, ако удовлетворява всички ограничения и се нарича *оптимално*, ако освен че е допустимо, в него целевата функция достига желаната оптимална стойност. Множеството от всички допустими решения на задачата се нарича *допустимо множество.*

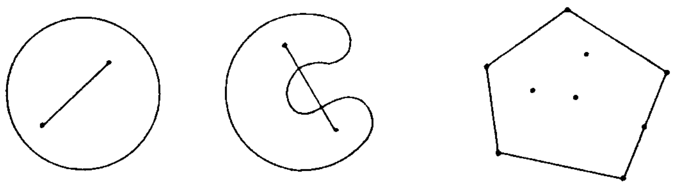
За да се направят разбираеми теорията на линейното оптимиране и методите, използвани за решаване на линейни задачи, от съществено значение е да се разгледа геометричния смисъл на тези задачи и да се даде алгебрична характеризация на включените в тях геометрични обекти.

За целта са ни необходими първо няколко дефиниции.

1.1. Дефиниции

Дефиниция 1.1. Множество $C \subseteq \mathbb{R}^n$ се нарича *изпъкнало* ако заедно с всеки две точки x', x'' принадлежащи на C, всички точки от вида $x(\lambda) = \lambda x' + (1 - \lambda)x''$, където $0 \leqslant \lambda \leqslant 1$, също принадлежат на C.

С други думи, едно множество е изпъкнало, ако заедно с всеки две точки от множеството, в него се съдържа и отсечката, която ги съединява (вж. Фиг. 1.1).



Фигура 1.1.

Дефиниция 1.2. Точката x се нарича *изпъкнала комбинация* на точките $x_1, \ldots, x_N$ ако

$$
x = \sum_{i=1}^{N} \lambda_i x_i, \quad \lambda_i \geqslant 0 \text{ за всяко } i \quad \text{и} \quad \sum_{i=1}^{N} \lambda_i = 1\,.
$$

Множеството от всички изпъкнали комбинации на две точки е отсечката, която ги съединява. Множеството от изпъкналите комбинации на три точки в общо положение е триъгълникът с върхове в тях, на четири – образуваната от тях пирамида, и т.н.

Дефиниция 1.3. *Екстремна точка* на изпъкнало множество C е точка $x \in C$, която не може да бъде представена като изпъкнала комбинация на (две) други точки на C.

Например, екстремните точки на триъгълника са върховете му, а екстремните точки на кълбото $\{x \in \mathbb{R}^n : \|x\| \leqslant 1\}$ са точките от сферата $\{x \in \mathbb{R}^n : \|x\| = 1\}$.

Дефиниция 1.4. Множеството $H = \{x \in \mathbb{R}^n : a^{\mathrm{T}}x = \beta\}$, където $a \in \mathbb{R}^n$, $a \neq 0$, и $\beta \in \mathbb{R}$ се нарича *хипер-равнина*. Множеството $\overline{H} = \{x \in \mathbb{R}^n : a^{\mathrm{T}}x \leqslant \beta\}$ се нарича *затворено полупространство.* Хипер-равнината H, определяща полупространството $\overline{H}$ се нарича *гранична хипер-равнина* на това полупространство.

Векторът a в горната дефиниция е ортогонален на хипер-равнината H и се нарича нейна *външина нормала*, тъй като сочи към външността на $\overline{H}$. За да видим това, нека y, $z \in H$ и $w \in \overline{H}$. Тогава

$$a^{\mathrm{T}}(y - z) = a^{\mathrm{T}}y - a^{\mathrm{T}}z = \beta - \beta = 0\,,$$

т.е., a е ортогонален на всички вектори, успоредни на H и

$$a^{\mathrm{T}}(w - z) = a^{\mathrm{T}}w - a^{\mathrm{T}}z \leqslant \beta - \beta = 0$$

т.е. a образува тъп ъгъл с всеки вектор, който сочи към вътрешността на $\overline{H}$.

Дефиниция 1.5. Множество $S_a \subseteq \mathbb{R}_n$ се нарича *афинно множество*, или *афинно подпространство*, ако за всеки две точки $x', x'' \in S_a$, всички точки от вида $x(\lambda) = \lambda x' + (1 - \lambda)x''$, където $-\infty < \lambda < \infty$, също принадлежат на S_a.

Да забележим, че за разлика от дефиницията на изпъкнало множество, при дадени две точки в афинно множество, имаме че цялата права, която минава през тях лежи в това множество, а не само отсечката между тях.

Да отбележим също така, че афинното множество S_a представлява линейно подпространство S транслирано с вектор y, т.е., $S_a = \{y + x : x \in S\}$. Казваме, че S_a е *успоредно* на S.

Всяка хипер-равнина в $\mathbb{R}^n$ представлява $(n-1)$-мерно афинно множество, като дефиницията за размерност е следната:

Дефиниция 1.6. *Размерността* на линейно подпространство S, и на всяко афинно подпространство S_a, успоредно на него, е равна на най-големия брой линейно независими вектори в S. *Размерността* на произволно множество $D \subseteq R^n$ е равна на най-малката размерност на афинно подпространство, което съдържа D.

Дефиниция 1.7. *Многостенно множество* е множество образувано чрез пресичането на краен брой затворени полупространства и хипер-равнини. Ако такова множество е непразно и ограничено, то се нарича *многостен.*

Лесно се вижда, че хипер-равнините и затворените полупространства са изпъкнали множества и че сечението на изпъкнали множества е изпъкнало множество. Следователно, всяко многостенно множество е изпъкнало множество. Сега е ясно, че множеството от допустими решения на канонична задача $P = \{x \in \mathbb{R}^n : Ax = b, x \geqslant 0\}$ е изпъкнало многостенно множество, тъй като е сечение на хипер-равнините определени от равенствата

$$a_1^{\mathrm{T}} x = b_1, \ldots, a_m^{\mathrm{T}} x = b_m \tag{1.1}$$

и полупространствата, определени от неравенствата

$$-e_1^{\mathrm{T}} x \leqslant 0, \ldots, -e_n^{\mathrm{T}} x \leqslant 0 \tag{1.2}$$

където a_i^{T} е i-я ред на матрицата A и e_i^{T} е i-я ред на $n \times n$ единичната матрица. Многостенното множество P често ще наричаме *канонично* множество.

Преди да разгледаме някои от важните свойства на многостенните множества, ни трябват още две дефиниции:

Дефиниция 1.8. *Опорна хипер-равнина* на изпъкнало множество C е хипер-равнина H, такава че $H \cap C \neq \emptyset$ и $C \subseteq \overline{H}$.

Дефиниция 1.9. Нека P е многостенно множество и нека H бъде негова опорна хипер-равнина. Сечението $F = P \cap H$ се нарича *стена* на P.

Интересуват ни три вида стени на P.

Дефиниция 1.10. *Връх, ръб* и *страна* на d-мерно многостенно множество се нарича негова стена, която има размерност съответно нула, едно и $d - 1$.

Ако разгледаме каноничното множество $P = \{x \in \mathbb{R}^n : Ax = b, x \geqslant 0\}$ очевидно е, че всяка от страните му съответства на сечение на P с хипер-равнина, определена с някое от равенствата (1.1) или с гранична хипер-равнина на полупространство, определено с някое от неравенствата (1.2). Въпреки това, не всички такива сечения определят страни P, тъй като някои от равенствата могат да са *излишни*, в смисъл че отстраняването им от дефиницията на P не го променя. На фигура 1.2 в описанието на многостена $P_1 \subset \mathbb{R}^3$ може да участва и полупространството $\overline{H}_2 = \{x \in \mathbb{R}^3 : (0,1,0)x \leqslant 2\}$. Въпреки това, сечението на новото полупространство с P не определя страна (с размерност 2), а ребро (с размерност 1) – отсечката $[(0,2,0),(2,2,0)]$. Интуитивно е ясно, че причината за това, е че новото полупространство е излишно в описанието на P_1.

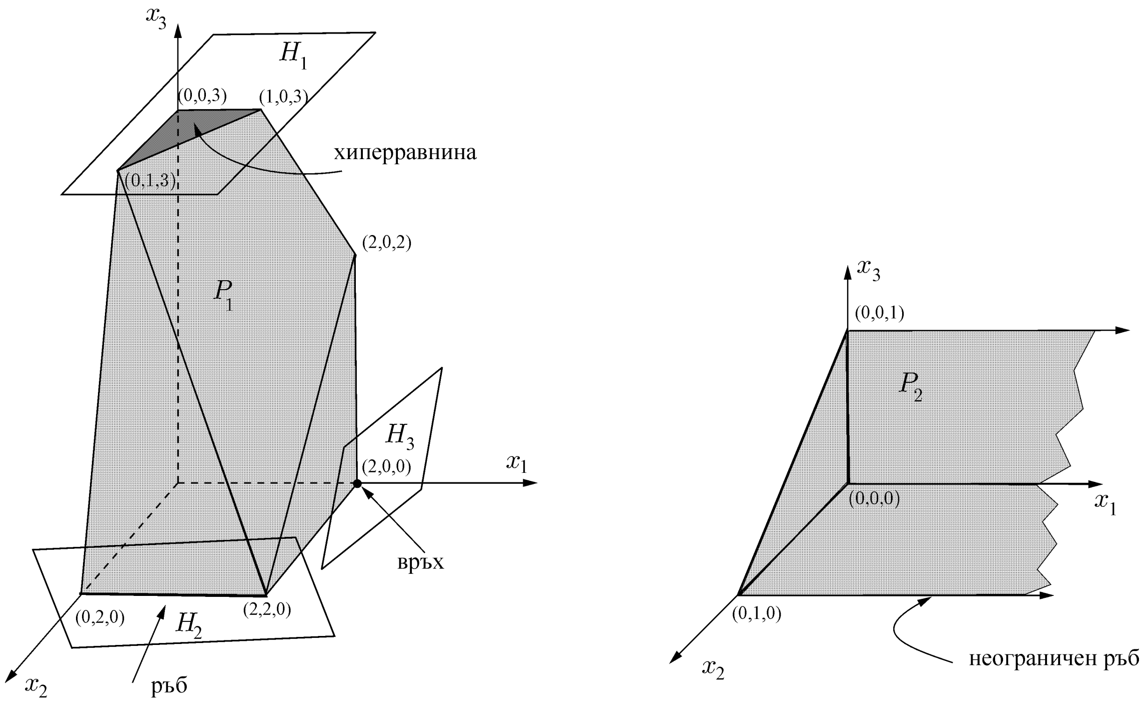
Върховете на многостенно множество P са екстремни точки на P, както ще докажем в следната

Лема 1.1. *Ако $\overline{x}$ е връх на P, то $\overline{x}$ е екстремна точка на P.*

Доказателство. Нека $\overline{x}$ е връх на P. Това означава, че съществува опорна за P хипер-равнина $H = \{x \in \mathbb{R}^n : a^{\mathrm{T}} x = \beta\}$, такава че $\{\overline{x}\} = P \cap H$, т.е. $a^{\mathrm{T}} \overline{x} = \beta$ и $a^{\mathrm{T}} x \leqslant \beta$ за всяко $x \in P$. Да допуснем, че $\overline{x}$ не е екстремна точка, т.е., че съществуват точки $x', x'' \in P$, $x' \neq x''$ и число $0 < \lambda < 1$, за които $\overline{x} = \lambda x' + (1-\lambda)x''$. Следователно, $\beta = a^{\mathrm{T}} \overline{x} = a^{\mathrm{T}}(\lambda x' + (1-\lambda)x'') =$

$\lambda a^{\mathrm{T}}x' + (1-\lambda)a^{\mathrm{T}}x''$. Тъй като $0 < \lambda < 1$ равенството е възможно само ако $a^{\mathrm{T}}x' = a^{\mathrm{T}}x'' = \beta$, т.е. само ако $x', x'' \in H$, което е в противоречие с това, че сечението на H с P се състои само от точката $\overline{x}$. ■

Ръбовете на P са или отсечки, които свързват *съседни* върхове, или са лъчи, които излизат от даден връх. На фигура 1.2 многостенното множество P_2 е неограничено и сечението на P_2 с опорната за множеството хипер-равнина $H = \{x \in \mathbb{R}^3 : (0,1,0)^{\mathrm{T}}x = 1\}$ е неограниченият ръб $(0,1,0) + \lambda(1,0,0)$, който има за начало върха $(0,1,0)$ и за направляващ вектор $(1,0,0)$.



Фигура 1.2.

1.2. Екстремни точки и базисни допустими решения

Ще покажем, че ако канонична задача на линейното оптимиране има крайно оптимално решение, то тя има оптимално решение, което е връх на каноничното множество P от всички допустими решения. За целта ще дадем алгебрична характеризация на върховете на P, като с P, както казахме, ще означаваме множеството $\{x \in \mathbb{R}^n : Ax = b, x \geqslant 0\}$.

Теорема 1.1. *Точка $x \in P = \{x \in \mathbb{R}^n : Ax = b, x \geqslant 0\}$ е екстремна точка на P тогава и само тогава, когато стълбовете на матрицата A, съответстващи на положителните координати на x са линейно независими.*

Доказателство. Без ограничение на общността, да допуснем, че първите p координати на x са положителни и че последните $(n-p)$ координати на x са нули. Ако разделим x така, че $x = \binom{\overline{x}}{0}$, $\overline{x} > 0$, и означим първите p стълба на A с $\overline{A}$, то имаме, че $Ax = \overline{A}\overline{x} = b$.

Да допуснем че стълбовете на $\overline{A}$ *не са* линейно независими. Тогава, съществува вектор $\overline{w} \neq 0$, такъв че $\overline{A}\overline{w} = 0$. Оттук, $\overline{A}(\overline{x} \pm \varepsilon\overline{w}) = \overline{A}\overline{x} = b$ и за достатъчно малки положителни ε, $(\overline{x} \pm \varepsilon\overline{w}) \geqslant 0$. Следователно, точките

$$y' = \begin{pmatrix} \overline{x} + \varepsilon\overline{w} \\ 0 \end{pmatrix} \qquad \text{и} \qquad y'' = \begin{pmatrix} \overline{x} - \varepsilon\overline{w} \\ 0 \end{pmatrix}$$

принадлежат на P. Тъй като $x = \frac{1}{2}(y' + y'')$, то x не може да бъде екстремна точка на P. Значи, ако x е екстремна точка, то стълбовете на $\overline{A}$ са линейно независими.

Сега да предположим, че x не е екстремна точка. Това означава, че $x = \lambda y' + (1-\lambda)y''$ за някои y', $y'' \in P$, $y' \neq y''$ и $0 < \lambda < 1$. Тъй като и двете точки x и y' са в P, $A(x - y') = Ax - Ay' = b - b = 0$. Тъй като и двете числа λ и $1 - \lambda$ са строго положителни, последните $(n-p)$ координати на y' и y'' са нули, (тези координати при x са равни на нула). Следователно, последните $(n-p)$ координати на вектора $x - y'$ трябва да са нули, а измежду първите p има поне една ненулева (в противен случай $x = y' = y''$ и имаме противоречие). Това означава, че векторът $x - y'$ е ненулев, откъдето стълбовете на $\overline{A}$ са линейно зависими. Следователно, че ако стълбовете на $\overline{A}$ са линейно независими, то x е екстремна точка. ■

Да припомним, че ако A е $(m \times n)$ матрица, то *рангът* на A е равен на максималният брой линейно независими вектор-редове или вектор-стълбове на A и се означава с $r(A)$. Казваме, че A има пълен ранг по редове, ако $r(A) = m$.

Когато A няма пълен ранг по редове, то или системата линейни уравнения $Ax = b$ няма решение и P е празното множество, или някои от уравненията в системата са *излишни*. Във втория случай излишните уравнения могат да се отстранят едно по едно, докато се стигне до система от уравнения, чиято матрица има пълен ранг по редове. От тук нататък, без ограничение на общността предполагаме, че матрицата A има пълен ранг по редове. В този случай, еквивалентна характеризация на върховете на P може да се даде посредством т.нар. *базисни решения*.

Дефиниция 1.11. Нека B е неособена $m \times m$ матрица, съставена от m (линейно независими) стълба на A. Ако всички координати на x съответстващи на стълбове, неучастващи в B, наречем *небазисни променливи* и положим на нула и системата линейни уравнения $Ax = b$ разрешим относно останалите координати на x, които ще наречем *базисни променливи*, тогава полученото x се нарича *базисно решение* спрямо *базис* (базисна матрица) B. Също така ще използваме термина *базис* и означението $\mathbf{B}$ за обозначаване на множеството от базисните променливи и на множеството от индексите на тези променливи.

Да забележим, че ако положим небазисните променливи равни на нула, получаме система от m уравнения с m неизвестни

$$Bx_B = b\,,$$

която има единствено решение относно базисните променливи x_B,

$$x_B = B^{-1}b\,.$$

Терминът базис идва от това, че стълбовете на B образуват базис за линейното пространство, породено от стълбовете на A и на $Ax = b$ може да се гледа като на изразяване на b като линейна комбинация на стълбовете на A.

Ако базисното решение x относно базис B е неотрицателно, то се нарича *базисно допустимо решение*, т.к. в този случай то принадлежи на допустимото множество P.

Следното представлява характеризация на върховете на P.

Следствие 1.1. *Ако множеството P е непразно, следните твърдения са еквивалентни*
(а) *$\overline{x}$ е връх на P;*
(б) *$\overline{x}$ е екстремна точка на P;*
(в) *$\overline{x}$ е базисно допустимо решение.*

Доказателство. (а) $\Longrightarrow$ (б) доказахме в Лема 1.1. За (б) $\Longrightarrow$ (в) получаваме от Теорема 1.1 линейната независимост на стълбовете от $\overline{A}$, които после допълваме до базисна матрица B с някои от другите стълбове на A. Остава да покажем (в) $\Longrightarrow$ (а). За целта нека $\overline{x}$ е базисно допустимо решение с базис B, т.е. $\overline{x}_j = 0$ за $j \notin \mathbf{B}$, а $\overline{x}_B$ е единственото решение на системата $Bx_B = b$. Да положим

$$a_j = \begin{cases} 0, & j \in \mathbf{B} \\ -1, & j \notin \mathbf{B} \end{cases}$$

Ако $x \in P$, то $a^{\mathrm{T}}x \leqslant a^{\mathrm{T}}\overline{x} = 0$, т.е. хипер-равнината $H = \{x \in \mathbb{R}^n : a^{\mathrm{T}}x = 0\}$ е опорна за P и $\overline{x} \in H \cap P$. Ако допуснем, че за някое допустимо решение $x \in P$ е изпълнено, че $a^{\mathrm{T}}x = 0$, то $x_j = 0$ за всяко $j \notin \mathbf{B}$, а x_j за $j \in \mathbf{B}$ се получават като решения на системата $Bx_B = b$ и следователно са равни на съответните координати на $\overline{x}$. Следователно, $x = \overline{x}$ и сечението на H и P се състои само от точката $\overline{x}$, която по дефиниция е връх на P. ■

Това следствие ни позволява да използваме термините връх, екстремна точка и базисно допустимо решение като обозначаващи един и същи обект.

Следствие 1.2. *Канонично множество P има само краен брой върхове.*

Доказателство. Следва от предишното следствие и от факта, че има само краен брой начини да се изберат m линейно независими "базисни" стълба измежду n-те стълба на A. Ясно е, че горната граница на броят върхове на P е $\binom{n}{m} = n!/(m!(n-m)!)$. ■

2. Теорема за представяне на канонично многостенно множество.

Нека $P = \{x \in \mathbb{R}^n : Ax = b, x \geqslant 0\}$ е канонично многостенно множество. Върховете на каноничното множество P са важни обекти. Ако P е ограничено, т.е. P е многостен, то произволна точка от P може да се представи като изпъкнала комбинация на върховете на P (вж. Следствие 2.1 по-долу). Когато P е неограничено, представянето на произволна точка от P е малко по-сложно и за него ни е нужна следната дефиниция.

Дефиниция 2.1. *Посока* в P е ненулев вектор $d \in \mathbb{R}^n$, такъв че за всяка точка $x_0 \in P$ лъчът $\{x \in \mathbb{R}^n : x = x_0 + \lambda d,\ \lambda \geqslant 0\}$ лежи изцяло в P.

На фигура 1.2 посока в многостенното множество P_2 е $d = (1, 0, 0)$.

Очевидно P е неограничено, тогава и само тогава, когато в P има посока. Характеризация на посоките в P дава следната

Лема 2.1. Вектор $d \neq 0$ е посока в P тогава и само тогава, когато

$$Ad = 0 \qquad \text{и} \qquad d \geqslant 0\,.$$

Доказателство. Нека $d \neq 0$ е посока в P. За фиксирана точка $x_0 \in P$ имаме, че $x_0 + \lambda d \in P$ за всяко положително λ. Това означава че за всяко положително λ имаме $A(x_0 + \lambda d) = b$ и тъй като $Ax_0 = b$, то $\lambda Ad = 0$, откъдето $Ad = 0$. Ако допуснем, че измежду координатите на d има отрицателна, за големи λ съответната координата на $x_0 + \lambda d$ също ще бъде отрицателна, което противоречи на това че $x_0 + \lambda d \in P$ и следователно има само неотрицателни координати.

Обратно, ако $d \neq 0$ удовлетворява условията на твърдението, то за произволна точка $x_0 \in P$ имаме, че $A(x_0 + \lambda d) = Ax_0 + \lambda Ad = Ax_0 = b$ и $x_0 + \lambda d \geqslant 0$ за всяко неотрицателно λ. ■

Вече можем да формулираме и докажем следната теорема за представяне на точките на канонично многостенно множество.

Теорема 2.1 (Теорема за представяне). *Всяка точка $x \in P$ може да се представи като*

$$x = \sum_{i \in I} \lambda_i v_i + d\,, \tag{2.1}$$

където $V := \{v_i : i \in I\}$ е множеството от върховете на P,

$$\sum_{i \in I} \lambda_i = 1\,, \quad \lambda_i \geqslant 0 \text{ за всички } \ i \in I,$$

и, или d е посока в P, или $d = 0$.

Доказателство. Ще докажем теоремата с индукция по p, броят на положителните координати на x. Резултатът очевидно е верен за $p = 0$ (x е началото $(0, 0, \ldots, 0)$, което очевидно е връх). Сега да предположим, че твърдението е вярно за точките с по-малко от p положителни координати и да разгледаме случая, когато x има p положителни координати.

Ако x е връх, теоремата очевидно е вярна, тъй като $x = v_i$ за някое $i \in I$. Следователно, да предположим, че x не е връх. Тогава съществува вектор $w \neq 0$, такъв че $w_i = 0$ ако $x_i = 0$ и такъв че $Aw = 0$ (вж. Теорема 1.1). Имаме три случая за разглеждане.

Случай (а): w има и положителни, и отрицателни координати.

Да разгледаме точките $x(\theta) = x + \theta w$ върху правата през x определена от w. Да означим съответно с θ' и θ'' най-малката положителна и най-голямата отрицателна стойност на θ, за която $x(\theta)$ има най-малко още една нулева координата повече от x. Очевидно точките $x' = x(\theta')$ и $x'' = x(\theta'')$ лежат в P и според индукционното предположение имат желаното представяне. Точката x, която лежи на отсечката, определена от x' и x'', можем да представим като

$$x = \mu x' + (1-\mu)x''$$

за някое $\mu \in (0,1)$, което можем да намерим от

$$\begin{aligned} x &= \mu x' + (1-\mu)x'' \\ &= \mu(x + \theta' w) + (1-\mu)(x + \theta'' w) \\ &= x + (\mu\theta' + (1-\mu)\theta'')w, \end{aligned}$$

откъдето $(\mu\theta' + (1-\mu)\theta'')w = 0$, но w е ненулев вектор и следователно $\mu\theta' + (1-\mu)\theta'' = 0$. Оттук намираме $\mu = -\theta''/(\theta' - \theta'')$. Да отразим, че точките x' и x'' имат желаното представяне,

$$\begin{aligned} x &= \mu x' + (1-\mu)x'' \\ &= \mu\left(\sum_{i\in I}\lambda_i' v_i + d'\right) + (1-\mu)\left(\sum_{i\in I}\lambda_i'' v_i + d''\right) \\ &= \sum_{i\in I}\left(\mu\lambda_i' + (1-\mu)\lambda_i''\right)v_i + \mu d' + (1-\mu)d''\,, \end{aligned}$$

където

$$\begin{aligned} &\lambda_i' \geqslant 0 \text{ и } \lambda_i'' \geqslant 0 \quad \text{за всяко } i \in I, \qquad \sum_{i\in I}\lambda_i' = \sum_{i\in I}\lambda_i'' = 1\,, \\ &Ad' = Ad'' = 0, \quad d' \geqslant 0 \quad \text{и} \quad d'' \geqslant 0\,. \end{aligned}$$

Тъй като $0 < \mu < 1$, ако положим

$$\begin{aligned} &\lambda_i := \mu\lambda_i' + (1-\mu)\lambda_i'' \geqslant 0 \quad \text{за всяко } i \in I, \qquad \sum_{i\in I}\lambda_i = 1\,, \\ &d := \mu d' + (1-\mu)d'' \geqslant 0 \quad \text{и} \quad Ad = 0\,, \end{aligned}$$

получаваме, че x има желаното представяне.

Случай (б): $w \leqslant 0$.

Дефинираме x' както в случай (а). Точката x записваме като

$$x = x' + \theta'(-w) \qquad \text{където} \qquad \theta' > 0.$$

Тъй като според индукционното предположение x' може да се представи в желаната форма и тъй като $(-w)$ е посока в P, то очевидно x има желаното представяне.

Случай (в): $w \geqslant 0$.

Доказателството е аналогично на това, направено за случай (б), като заменим x', θ' и $-w$ съответно с x'', $-\theta''$ и w.

Като следствие получаваме:

Следствие 2.1. *Ако P е ограничено множество (т.е. P е многостен), то всяко $x \in P$ може да се представи като изпъкнала комбинация на върховете на P.*

3. Фундаментални теореми на линейното оптимиране

Ще докажем две теореми, които са в основата на разработването на алгоритми (и в частност на симплекс алгоритъма) за решаване на линейни задачи. По-точно казано, тези теореми разкриват значението, което имат върховете на канонично множество $P = \{x \in \mathbb{R}^n : Ax = b, x \geqslant 0\}$, т.е. базисните допустими решения, за тези методи.

Теорема 3.1. *Ако P е непразно множество, то P има поне един връх.*

Доказателство. Тъй като по допускане $P \neq \emptyset$, то в P има поне една точка x. Тя има представянето, доказано в Теорема 2.1. Ако допуснем, че множеството V от върховете на P е празното множество, то от теоремата за представяне, $x = d$, където $d \geqslant 0$, $Ad = 0$. Тъй като x е в P, $b = Ax = Ad = 0$, т.е. $P = \{x \in \mathbb{R}^n : Ax = 0, x \geqslant 0\}$ и следователно, началото $(0, 0, \ldots, 0) \in P$. Но началото очевидно е връх на P, откъдето получаваме противоречие с допускането, че множеството от върхове на P е празно. ■

Теорема 3.2. *Ако P е непразно множество, то или минималната стойност на функцията $z(x) = c^{\mathrm{T}}x$ за $x \in P$ се достига във връх на P, или z не е ограничена отдолу върху P.*

Доказателство. Трябва да разгледаме два случая:

Случай (а): В P има посока d, такава че $c^{\mathrm{T}}d < 0$. В този случай P е неограничено множество и стойността на $z \to \infty$ по посоката d. Наистина, ако вземем произволна точка x от непразното множество P, то по дефиницията за посока всички точки от лъча $x + \lambda d$ са в P. Стойностите на z върху този лъч са $z(x + \lambda d) = c^{\mathrm{T}}(x + \lambda d) = c^{\mathrm{T}}x + \lambda c^{\mathrm{T}}d$ и когато $\lambda \to +\infty$ клонят към $-\infty$ поради $c^{\mathrm{T}}d < 0$.

Случай (б): В P няма посока d, такава че $c^{\mathrm{T}}d < 0$. С други думи, за всички посоки d в P е изпълнено, че $c^{\mathrm{T}}d \geqslant 0$. В този случай трябва да разгледаме само точките, които се представят като изпъкнала комбинация на върховете v_i на P, тъй като дори P да е неограничено, всяка точка $x \in P$ според Теоремата за представяне е от вида $x = \hat{x} + d$, където

$$\hat{x} = \sum_{i \in I} \lambda_i v_i, \qquad \sum_{i \in I} \lambda_i = 1, \quad \text{и} \quad \lambda_i \geqslant 0 \text{ за всяко } i \in I\,,$$

а d е посока в P. Стойността на z в x е $z(x) = z(\hat{x} + d) = c^{\mathrm{T}}(\hat{x} + d) = c^{\mathrm{T}}\hat{x} + c^{\mathrm{T}}d \geqslant c^{\mathrm{T}}\hat{x}$ и следователно е ограничена отдолу от $c^{\mathrm{T}}\hat{x}$. Но

$$c^{\mathrm{T}}\hat{x} = c^{\mathrm{T}}\left[\sum_{i \in I} \lambda_i v_i\right] = \sum_{i \in I} \lambda_i c^{\mathrm{T}} v_i \geqslant \min_{i \in I}\{c^{\mathrm{T}} v_i\}\,.$$

Следователно, минимумът на z се достига във връх на P. ■

Тази теорема стои в основата на решаването на линейни задачи. Тя показва, че е необходимо да се разглеждат само върховете на P, т.е. само базисните допустими решения, като кандидати за оптимално решение. Също така тя показва, че трябва да се следи и дали в P има посоки, по които $z \to -\infty$.

4. Симплекс метод. Геометрична мотивация. Алгоритъм.

От Теорема 3.2 е ясно, че за да решим каноничната задача на линейното оптимиране,

$$\begin{array}{ll} \text{да се минимизира} & z(x) = c^{\mathrm{T}} x \\ \text{при ограничения} & Ax = b\,, \\ & x \geqslant 0\,, \end{array} \tag{4.1}$$

където A означава $(m \times n)$ матрица, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$, е достатъчно като кандидати за оптимално решение да разгледаме само върховете на каноничното множество $P = \{x : Ax = b, x \geqslant 0\}$ от допустими решения на задачата. За големи стойности на m и n обаче да се намерят всички върхове на P е непрактично, тъй като P може да има $\binom{n}{m} = n!/[m!(n-m)!]$ на брой базисни допустими решения. Очевидно се нуждаем от по-систематичен подход. Такъв подход прилага симплекс методът, разработен от Джордж Данциг през 1947 г. На практика, симплекс методът се оказва толкова успешен, че става един от най-известните и един от най-използваните методи за компютърно реализиране на числените пресмятания.

4.1. Геометрична мотивация

Идеята на симплекс метода наистина е много проста. Методът има две фази. *Първата фаза* намира връх на P (по Теорема 3.1 такъв винаги съществува, ако P е непразно множество). Същината на метода е във *втората фаза*, при която се преминава от връх на връх по ръбовете на P, като движението се извършва по ръбове, по които целевата функция z е строго намаляваща. По този начин методът генерира редица от върхове, в които целевата функция приема строго намаляващи стойности. Следователно, ако методът напусне даден връх, той никога не се връща в него и след краен брой стъпки (т.к. от Следствие 1.2, P има краен брой върхове) или ще достигне до връх, който е оптимално решение, или ще достигне до неограничен ръб, по който целевата функция z намалява до $-\infty$.

Задачата ни е да облечем горното просто геометрично описание на симплекс метода в алгебрична, и следователно, изчислителна форма.

Първо ще разгледаме втората фаза при която се предполага, че вече е намерен връх на P. Както ще видим по-късно (вж въпрос 6), алгоритъмът за тази втора фаза може да се използва и за решаването на задачата на първата фаза – намирането на началния връх.

Предполагаме, че редовете на матрицата A са линейно независими, т.е., че $r(A) = m$ и че $m < n$, така че задачата не е тривиална. Ако $r(A) < m$, то или системата от ограничения равенства е несъвместима, или някои от тях за излишни и могат да бъдат отстранени. Направените предположения за A гарантират, че има поне едно базисно решение и от стълбовете на A може да се избере базисна матрица B.

За да опишем итерацията на симплекс метода, нека за простота предположим, че координатите на текущия връх $\overline{x}$ са подредени така, че първите m от тях са базисните. Т.е., върхът $\overline{x}$ на P съответства на базисното допустимо решение

$$\overline{x} = \begin{bmatrix} \overline{x}_B \\ \overline{x}_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} \tag{4.2}$$

където матрицата A е разделена на $A = [B|N]$. По същия начин разделяме c^{T} на $c^{\mathrm{T}} = [c_B^{\mathrm{T}}, c_N^{\mathrm{T}}]$ в съответствие на направеното разделяне на $\overline{x}$ на базисна и небазисна част. Спрямо дадена базисна матрица B означаваме множеството от индекси на базисни променливи с $\mathbf{B}$ (в случая $\mathbf{B} = \{1, \ldots, m\}$), а множеството от индекси на небазисни променливи означаваме с $\mathbf{N}$ (в случая $\mathbf{N} = \{m+1, \ldots, n\}$).

Дефиниция 4.1. Базисно допустимо решение, на което една или повече базисни променливи са нули, се нарича *изродено*. В противен случай то се нарича *неизродено*.

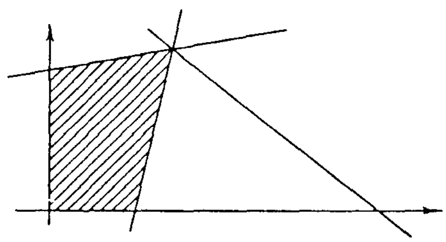
Ако базисното допустимо решение $\overline{x}$ (4.2) е неизродено, то лежи в сечението в $\mathbb{R}^n$ на m-те хипер-равнини, съответстващи на ограниченията равенства $Ax = b$ и на $(n-m)$-те хипер-равнини, съответстващи на изискването, че $(n-m)$-те небазисни променливи са равни на нула, т.е. $x_N = 0$. Да разгледаме матрицата

$$M = \begin{bmatrix} B & N \\ 0 & I \end{bmatrix}, \tag{4.3}$$

чиито редове са нормалите на тези n хипер-равнини. Тъй като B е неособена матрица, редовете на матрицата M са линейно независими и следователно M също е неособена. По този начин върхът $\overline{x}$ (4.2) е определен от пресичането на n линейно независими хипер-равнини.

Ако базисното допустимо решение $\overline{x}$ е изродено, то някои от базисните променливи $\overline{x}_B$ също са равни на нула. Следователно, повече от $(n-m)$ от ограниченията за неотрицателност $x_j \geqslant 0$ са удовлетворени като равенства и координатите на точката $\overline{x}$ удовлетворяват повече от n равенства.

Илюстрация на изродеността е дадена във Фигура 4.3. От тази илюстрация може да изг-



Фигура 4.3. Илюстрация на изродеността

лежда, че винаги има излишни ограничения във връх, който е изроден. Въпреки това, такъв е случаят само когато $n - m \leqslant 2$.

Когато базисното допустимо решение $\overline{x}$ е изродено, може да има огромен брой от базиси, асоциирани с върха $\overline{x}$. В действителност, ако $\overline{x}$ има $p < m$ положителни координати, то е възможно да има

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

"различни" базисни допустими решения, съответстващи на $\overline{x}$. Върхът $\overline{x}$ е един и същ за всяко от тях, но множествата от променливи, които се означават като базисни и небазисни са различни. Екстремен пример за изроденост се получава в така наречената задача за назначенията. Може да се покаже, че каноничното множество от допустими решения

$$P_k = \left\{ x_{ij},\ 1 \leqslant i,j \leqslant k : \sum_{j=1}^{k} x_{ij} = 1,\ 1 \leqslant i \leqslant k;\ \sum_{i=1}^{k} x_{ij} = 1,\ 1 \leqslant j \leqslant k;\ 0 \leqslant x_{ij},\ 1 \leqslant i,j \leqslant k \right\}$$

на тази специална “транспортна задача” има $k!$ върха, на всеки от които съответстват по $2^{k-1}k^{k-2}$ различни базиса. Следователно, при $k = 8$, всеки от $40\,320$ върха на P_8 има по $33\,554\,432$ различни базиса.

Нека да обърнем внимание, че докато на всяка базисна матрица B съответства точно едно базисно решение $\overline{x}$ (чиито базисни координати $\overline{x}_B$ са единственото решение на системата $Bx_B = b$, а небазисните му координати $\overline{x}_N$ са равни на нула), то на всяко изродено базисно допустимо решение могат да съответстват множество различни базиси.

Да се върнем към симплекс метода и за да улесним изложението, да допуснем, че базисното допустимо решение $\overline{x}$ (4.2) е неизродено. Това означава, че $\overline{x}$ се получава като сечение на n линейно независими хипер-равнини. Ако отстраним една от последните $(n-m)$ хипер-равнини, остават $(n-1)$ хипер-равнини и една степен на свобода. Точките в $\mathbb{R}^n$, чиито координати удовлетворяват тези $(n-1)$ уравнения образуват ръб, с начало $\overline{x}$, а геометрично, това е сечението на тези $(n-1)$ хипер-равнини. Тъй като по този начин можем да отстраним всяка една от последните $(n-m)$ хипер-равнини, то имаме точно $(n-m)$ ръба (т.е. едномерни стени) на P, излизащи от $\overline{x}$. Направляващите вектори на тези ръбове са последните $(n-m)$ вектор-стълба на обратната на матрицата от нормалите на активните ограничения M:

$$M^{-1} = \begin{bmatrix} B^{-1} & -B^{-1}N \\ 0 & I \end{bmatrix}. \tag{4.4}$$

Да проверим горното твърдение. За $q > m$ (небазисно) означаваме q-ия стълб на M^{-1} с η_q. Т.е. $\eta_q := M^{-1}e_q$ (където e_q е q-ия стълб на $n \times n$ единичната матрица). Понеже $M\eta_q = M(M^{-1}e_q) = Ie_q = e_q$, то η_q е ортогонален на всички вектор-редове на M, различни от q-ия, и следователно, той е ортогонален на нормалите на всички хипер-равнини, които се пресичат в $\overline{x}$, освен на тази, която съответства на $x_q = 0$. Това означава, че векторът η_q е успореден на сечението на $(n-1)$ линейно независими хипер-равнини, съответстващи на $Ax = b$ и $x_k = 0$, $k > m$, $k \neq q$. Твърдим, че направляващият за ръба вектор η_q е допустима посока, т.е. за достатъчно малки $\theta > 0$, точките от вида

$$x(\theta) = \overline{x} + \theta\eta_q \tag{4.5}$$

остават допустими (т.е. в P). За да докажем твърдението е необходимо само да проверим дали координатите са неотрицателни. Да означим q-ия стълб на A с a_q и да разпишем (4.5) покоординатно. Получаваме, че за достатъчно малки $\theta > 0$

$$\begin{aligned} &x_j(\theta) = 0,\ j > m, j \neq q, \\ &x_q(\theta) = \theta > 0 \\ &x_i(\theta) = \overline{x}_i - \theta B^{-1}a_q \geqslant 0,\ i = 1, 2, \ldots, m \end{aligned} \tag{4.6}$$

като последните m неравенства се дължат на факта, че $\overline{x}$ е неизродено и всичките му базисни координати са положитетелни.

От (4.6) е ясно, че движението по ръба η_q съответства на нарастването на една от небазисните променливи, а именно x_q, докато останалите небазисни променливи остават фиксирани на нула.

На всяка итерация от втората фаза симплекс методът е изправен пред следния проблем: намирайки се във върха $\overline{x}$ по кой от $(n-m)$-те ръба, излизащи от него да тръгне. Както вече казахме, методът търси измежду ръбовете такъв, по който целевата функция z намалява. Целевата функция $z = c^{\mathrm{T}}x$ намалява при движение по ръб с направляващ вектор η_j, ако нейният градиент c сключва тъп ъгъл с направлението на ръба η_j. За да се определи има ли такива ръбове и кои са те, за всяко $j > m$ (небазисните индекси) се пресмятат т.нар. *относителни оценки* или *редуцирани цени*:

$$\overline{c}_j = c^{\mathrm{T}}\eta_j = c_j - c_B^{\mathrm{T}}B^{-1}a_j, \quad j > m\,.$$

Ако $\overline{c}_j < 0$, то градиентът на целевата функция $z = c^{\mathrm{T}}x$ сключва тъп ъгъл с направлението на ръба η_j и z намалява при движение по този ръб (т.е. когато θ нараства). Терминът редуцирана цена е свързан със съответната небазисна променлива, която нараства при движение по ръба. Това е така, защото $\overline{c}_j$ представлява изменението на целевата функция z при изменение с единица на небазисната променлива x_j, при положение, че всички останали небазисни променливи остават фиксирани на нула: да забележим, че от (4.5) за $q = j$, следва че

$$z(x(\theta)) = c^{\mathrm{T}}x(\theta) = c^{\mathrm{T}}\overline{x} + \theta c^{\mathrm{T}}\eta_j = z(\overline{x}) + \theta\overline{c}_j\,.$$

Въпреки, че симплекс методът може да тръгне по произволен ръб, по който целевата функция намалява, обичайното правило е да се избере ръбът, съответстващ на най-малката отрицателна редуцирана цена.

Да отбележим, че редуцираните цени, съответстващи на базисни променливи са равни на нула, и че тези, съответстващи на небазисни променливи могат да се получат като първо се пресметне векторът на *симплексните множители* $\pi^{\mathrm{T}} = c_B^{\mathrm{T}}B^{-1}$, а след това се "обезценят" всички небазисни стълбове, т.е.,

$$\overline{c}_j = c_j - \pi^{\mathrm{T}}a_j, \quad j > m\,.$$

В следващата Лема 4.1 ще покажем, че всяка точка $x \in P$ лежи вътре в в конуса, породен от дадено базисно допустимо решение $\overline{x}$ и векторите по ръбовете η_j, излизащи от $\overline{x}$. В неизродения случай тези вектори са истински направляващи вектори по ръбовете. В изродения случай някои от тях могат да бъдат недопустими. Направляващ вектор η_q е недопустим в точка $\overline{x}$, където $\overline{x}_i = 0$, ако неговата i-та координата η_{iq} е отрицателна. Да напомним, че

Дефиниция 4.2. *Множество $C \subset \mathbb{R}^n$ се нарича конус, ако за всяка точка $x \in C$ и всяко неотрицателно число λ, точката $\lambda x \in C$.*

За множеството $\{y \in \mathbb{R}^k : y = D\alpha, \alpha \geqslant 0\}$, където D е $(k \times p)$ матрица, а $\alpha \in \mathbb{R}^p$ (което е от вида, разглеждан в Лема 4.1), чрез проверка на дефиницията покажете, че е конус. Покажете, също така, че множеството е изпъкнало.

Лема 4.1. *Ако $\overline{x}$, зададено с (4.2), е базисно допустимо решение, всяко $x \in P$ може да се представи като*

$$x = \overline{x} + \sum_{j=m+1}^{n} x_j\eta_j, \qquad x_j \geqslant 0, \quad j = m+1, \ldots, n,$$

където η_j е j-ия стълб на M^{-1} (4.4).

Доказателство. Нека $x \in P$ е произволна точка от допустимото множество P. Да разделим нейните координати на базисна и небазисна част, спрямо базиса B на $\overline{x}$, т.е. $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$. Тъй като $x \in P$, то координатите му удовлетворяват системата

$$\begin{aligned} & Ax = b, \iff \\ & [\,B\,|\,N\,] \begin{bmatrix} x_B \\ x_N \end{bmatrix} = b, \iff Bx_B + Nx_N = b, \iff \\ & x_B + B^{-1}Nx_N = B^{-1}b, \iff x_B = B^{-1}b - B^{-1}Nx_N, \iff \\ & x_B = \overline{x}_B - B^{-1}Nx_N. \end{aligned} \tag{4.7}$$

Като използваме, че $\overline{x}_N = 0$, получаваме

$$x = \begin{bmatrix} x_B \\ x_N \end{bmatrix} = \begin{bmatrix} \overline{x}_B \\ \overline{x}_N \end{bmatrix} + \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix} x_N = \overline{x} + \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix} x_N.$$

От $x \in P$ също така имаме че $x_N \geqslant 0$, т.е. $x_j \geqslant 0,\ j = m+1, \dots, n$,. Остава да забележим, че матрицата $\begin{bmatrix} -B^{-1}N \\ I \end{bmatrix}$ е точно матрицата от последните $(n-m)$ стълба на M^{-1} и нейните вектор-стълбове са направляващите вектори η_j, $j > m$. Следователно, като разпишем вектора, получен от матричното умножение като сума на вектори, получаваме

$$x = \overline{x} + \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix} x_N = \overline{x} + \sum_{j=m+1}^{n} x_j \eta_j, \qquad x_j \geqslant 0, \quad j = m+1, \dots, n,$$

което трябваше да се докаже. ■

Теорема 4.1. (Критерий за оптималност) *Базисното допустимо решение $\overline{x}$ (4.2) е оптимално решение за линейната задача* (4.1) *ако всички редуцирани цени (относно неговия базис) са неотрицателни.*

Доказателство. За произволно $x \in P$, от Лема 4.1 имаме, че $x = \overline{x} + \sum_{j=m+1}^{n} x_j \eta_j$, $x_j \geqslant 0, \quad j = m+1, \dots, n$. Стойността на целевата функция z в x е

$$z(x) = z(\overline{x} + \sum_{j=m+1}^{n} x_j \eta_j) = c^{\mathrm{T}}(\overline{x} + \sum_{j=m+1}^{n} x_j \eta_j) = c^{\mathrm{T}}\overline{x} + \sum_{j=m+1}^{n} (c^{\mathrm{T}}\eta_j) x_j = z(\overline{x}) + \sum_{j=m+1}^{n} \overline{c}_j x_j. \tag{4.8}$$

Тъй като $x_j \geqslant 0, \quad j = m+1, \dots, n$, ако всички редуцирани цени $\overline{c}_j$ са неотрицателни, следва че $z(x) \geqslant z(\overline{x})$ за всяко $x \in P$, т.е. $\overline{x}$ е оптимално решение. ■

От (4.7) и (4.8) следва, че при базис $\mathbf{B}$ каноничната задача (4.1) е еквивалентна на задачата

$$\begin{aligned} &\text{да се минимизира} && z(x) = z(\overline{x}) + \sum_{j \in \mathbf{N}} \overline{c}_j x_j \\ &\text{при ограничения} && x = \overline{x} + \sum_{j \in \mathbf{N}} x_j \eta_j, \\ &&& x \geqslant 0, \end{aligned} \tag{4.9}$$

а (4.9) се нарича *базисен вид* на задачата (4.1) спрямо базис B.

В неизродения случай е вярна и обратната на Теорема 4.1. Докато изродено базисно допустимо решение може да бъде оптимално дори когато някои от редуцираните цени са отрицателни, т.к. съответните им направления по ръбовете могат да са недопустими.

Следните са непосредствени следствия от (4.8).

Следствие 4.1. *Базисно допустимо решение* $\overline{x}$ *е единственото оптимално решение на* (4.1) *ако всички небазисни редуцирани цени са строго положителни.*

Следствие 4.2. *Ако* $\overline{x}$ *дадено с* (4.2) *е оптимално базисно допустимо решение с небазисни редуцирани цени* $\overline{c}_{j_1} = \overline{c}_{j_2} \ldots = \overline{c}_{j_k} = 0$*, то всяко допустимо решение* $x \in P$ *от вида*

$$x = \overline{x} + \sum_{i=1}^{k} x_{j_i} \eta_{j_i} \tag{4.10}$$

също е оптимално.

Ако оптимално базисно допустимо решение е изродено и редуцираните цени, съответстващи на някои от небазисните променливи, са нули, от Следствие 4.2 не следва, че оптималното решение не е единствено. Това е така, защото в изродения случай $\overline{x}$ може да е единствената точка от вида (4.10) която е в множеството P, поради недопустимостта на посоките по ръбовете η_{j_i} в (4.10).

След като в $\overline{x}$ е намерен и избран ръб на намаляване η_q с редуцирана цена $\overline{c}_q < 0$, следващата задача на итерацията по симплекс метода включва движение по този ръб до връх, съседен на $\overline{x}$. Това се постига чрез увеличаване на небазисната променлива x_q, т.е. чрез увеличаване на θ в (4.5), докато една от базисните променливи стане равна на нула.

Ако положим

$$w_q := B^{-1} a_q \tag{4.11}$$

от (4.5) и (4.6) следва, че $x(\theta) \geqslant 0$ тогава и само тогава, когато $\overline{x}_B - \theta w_q \geqslant 0$ и $\theta \geqslant 0$. Получаваме следната

Теорема 4.2. (Критерий за неограниченост на целевата функция) *Ако* $\overline{c}_q < 0$ *и* w_q *от* (4.11) *е вектор с неположителни координати* ($w_q \leqslant 0$)*, то линейната задача* (4.1) *е неограничена:* $x(\theta)$ *е допустимо за всяко* $\theta \geqslant 0$ *и* $z(x(\theta)) \to -\infty$ *когато* $\theta \to \infty$*. В този случай,* $d = \eta_q$ *е посока в* P*, за която* $c^{\mathrm{T}} d = \overline{c}_q < 0$*.*

Ако векторът w_q има положителна координата, това означава, че можем да се движим известно време по ръба η_q като оставаме в допустимото множество P. Най-голямата стъпка θ, която можем да направим по ръба с направляващ вектор η_q, докато поддържаме $x(\theta) \geqslant 0$ се определят от т.нар. "тест за минимално отношение":

$$\theta = \overline{x}'_q = \min\left\{ \frac{\overline{x}_i}{w_{iq}} : w_{iq} > 0, 1 \leqslant i \leqslant m \right\} = \frac{\overline{x}_p}{w_{pq}}\,, \tag{4.12}$$

като с w_{iq} означаваме i-та координата на вектора w_q. Числото w_{pq}, при което се достига минимума се нарича *ключово число* и то винаги е положително. Използвахме знака прим, за да покажем, че $\overline{x}'_q$ е стойността на q-та променлива в новата точка $\overline{x}'$, която получаваме в края на стъпката с дължина θ по ръба η_q. Координатите на $\overline{x}'$ съгласно (4.5) и (4.12) са

$$\begin{aligned} &\overline{x}'_j = 0, \quad j > m, j \neq q\,,\\ &\overline{x}'_q = \theta = \overline{x}_p / w_{pq}\,, \\ &\overline{x}'_i = \overline{x}_i - w_{iq}\theta = \overline{x}_i - w_{iq}\overline{x}_p / w_{pq}\,, \quad i = 1, \ldots, m\,. \end{aligned} \tag{4.13}$$

Да забележим, че в $\overline{x}'$ променливата x_p, е тази, която първа става нула като увеличаваме θ, т.е. координатите на $\overline{x}'$ са такива, че q-та вече е ненулева (със стойност $\theta > 0$), а p-та е равна на нула.

Ще докажем, че така полученото $\overline{x}'$ е връх на P.

Лема 4.2. *$\overline{x}'$ с координати (4.13) е връх (базисно допустимо решение) на P с базисна матрица B', която се различава от базисната матрица B на $\overline{x}$, по това, че един от нейните стълбове, стълба a_p, се заменя със стълба a_q, т.е.*

$$B' = B + (a_q - a_p)e_p^{\mathrm{T}}\,.$$

Доказателство. Точката $\overline{x}'$ очевидно принадлежи на P (движейки се по ръба η_q поддържаме $A\overline{x}' = b$, а $\overline{x}' \geqslant 0$ следва от (4.13). Да забележим, че стълбовете на матрицата A, съответстващи на положителните координати $\overline{x}'$ са включени в стълбовете на матрицата B'. Следователно, достатъчно е да покажем, че стълбовете на матрицата B' са линейно независими (вж. Следствие 1.1 и неговото доказателство).

Да допуснем противното, т.е., че стълбовете a_i, $i \in \mathbf{B}' = \{1, \ldots, m\} \setminus \{p\} \cup \{q\}$ са линейно зависими. В този случай съществуват числа β_i, $i \in \mathbf{B}'$, не всичките равни на нула и такива, че $\sum\limits_{i \in \mathbf{B}'} \beta_i a_i = 0$, или

$$\sum_{i \in \mathbf{B}, i \neq p} \beta_i a_i + \beta_q a_q = 0. \tag{4.14}$$

От $a_q = BB^{-1}a_q = Bw_q$ имаме, че a_q се изразява като линейна комбинация на стълбовете на B като коефициентите на линейната комбинация са координатите на вектора w_q, т.е. $a_q = \sum\limits_{i \in \mathbf{B}} w_{iq} a_i$. Да заместим това представяне на a_q в (4.14)

$$\sum_{i \in \mathbf{B}, i \neq p} \beta_i a_i + \beta_q \sum_{i \in \mathbf{B}} w_{iq} a_i = 0 \iff \sum_{i \in \mathbf{B}, i \neq p} (\beta_i + \beta_q w_{iq}) a_i + \beta_q w_{pq} a_p = 0.$$

Векторите a_i, $i \in \mathbf{B} = \{1, \ldots, m\}$, участващи в горната линейна комбинация са линейно независими (това са вектор стълбовете на базисната матрица B), следователно коефициентите в нея са равни на нули. От $\beta_q w_{pq} = 0$ и $w_{pq} > 0$ следва, че $\beta_q = 0$, което пък влече $\beta_i = 0$, $i \in \mathbf{B}, i \neq p$. Това означава че всички коефициенти β_i, $i \in \mathbf{B}'$ са нули, с което получаваме търсеното противоречие. ■

За да завършим итерацията на симплекс метода ни остава само да сменим върха $\overline{x}$ с върха $\overline{x}'$, да сменим базисната матрица B с B' и да направим q-та променлива базисна, а p-та променлива небазисна.

Да забележим, че ако минимумът в теста за минимално отношение (4.12) се достига за повече от една базисна координата, то следващият връх $\overline{x}'$ е изродено базисно допустимо решение, което има повече от един базис и следователно, изборът на напускаща базиса променлива е нееднозначен.

Като обобщим горните разсъждения, получаваме

Теорема 4.3. *Ако $\overline{c}_q < 0$ и w_q от (4.11) има положителна координата, то $\overline{x}'$ с координати (4.13) е друго базисно допустимо решение, за което стойността на целевата функция $z(\overline{x}') = c^{\mathrm{T}}\overline{x}' = c^{\mathrm{T}}\overline{x} + \theta\overline{c}_q = z(\overline{x}) + \theta\overline{c}_q$ е строго по-малка от $z(\overline{x})$ когато θ в (4.12) е положително, т.е. когато базисното допустимо решение $\overline{x}$ е неизродено.*

Вече можем да дадем симплекс метода в алгоритмична форма.

4.2. Алгоритъм на симплекс метода

(0) Нека е дадено базисно допустимо решение $\overline{x}$ на линейната задача (4.1), съответстващо на базисна матрица $B = [a_{j_i}, \dots, a_{j_m}]$. Нека $\boldsymbol{B} = \{j_1, \dots, j_m\}$ означава индексното множество от базисни променливи, т.е. x_{j_i} означава i-та базисна променлива.

(1) Пресмятаме *симплексните множители* като решим относно π системата

$$B^{\mathrm{T}}\pi = c_B$$

и пресмятаме *редуцираните цени*

$$\overline{c}_j = c_j - \pi^{\mathrm{T}} a_j, \quad \text{за всяко } j \notin \boldsymbol{B}.$$

(2) Проверка за оптималност: ако $\overline{c}_j \geqslant 0$, за всяко $j \notin \boldsymbol{B}$, КРАЙ; текущото решение е оптимално.

(3) Определяне на небазисна променлива x_q за влизане в базиса; т.е. намиране на ръб на намаляване чрез избиране на

$$q \in \{j \notin \boldsymbol{B} : \overline{c}_j < 0\}\,.$$

(4) Проверка за неограничен ръб: намиране на w_q чрез решаване на системата

$$Bw_q = a_q\,.$$

Ако $w_q \leqslant 0$, КРАЙ; има неограничен ръб в допустимото множество, по който $z \to -\infty$.

(5) Определяне на базисна променлива x_{j_p} за излизане от базиса: пресмятане на

$$\theta = \frac{\overline{x}_{j_p}}{w_{pq}} = \min_{1\leqslant i\leqslant m}\left\{\frac{\overline{x}_{j_i}}{w_{iq}} : w_{iq} > 0\right\}.$$

(6) Обновяване на текущото решение и на базисната матрица B: полагаме

$$\overline{x}_q \leftarrow \theta = \overline{x}_{j_p}/w_{pq}\,,$$

$$\overline{x}_{j_i} \leftarrow \overline{x}_{j_i} - \theta w_{iq}, \quad 1 \leqslant i \leqslant m\,,$$

$$B \leftarrow B + (a_q - a_{j_p})e_p^{\mathrm{T}}\,,$$

$$\boldsymbol{B} \leftarrow \boldsymbol{B} \setminus \{j_p\} \cup \{q\}\,,$$

$$j_p \leftarrow q\,,$$

и преминаваме на стъпка (1).

5. Изроденост и зацикляне на симплекс алгоритъма. Алгоритъм на Бленд, избягващ зациклянето.

Във въпрос 4 изложихме симплекс метода за решаване на каноничната линейна задача

$$\begin{array}{ll} \text{да се минимизира} & z(x) = c^{\mathrm{T}}x \\ \text{при ограничения} & Ax = b\,, \\ & x \geqslant 0\,, \end{array} \tag{5.1}$$

където A означава $(m \times n)$ матрица, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$.

Въпреки, че в описанието на итерацията на симплекс метода предположихме, че базисното допустимо решение $\overline{x}$ е неизродено, изродеността на практика не представлява пречка за симплекс алгоритъма. Единственото, което може да се получи в случай на изроденост на $\overline{x}$, е в стъпка (5) на алгоритъма (вж. въпрос 4), базисната променлива x_{j_p} да има нулева стойност $\overline{x}_{j_p} = 0$, което като резултат дава, че предприетата стъпка ще бъде с нулева дължина ($\theta = 0$), т.е. в действителност върхът $\overline{x}$ няма да се промени. Това се случва, тъй като "тръгвайки" да се движим по ръба η_q незабавно отиваме в ограничението $x_{j_p} \geqslant 0$. Въпреки, че на стъпка (6) на алгоритъма координатите на върха $\overline{x}$ не се променят, променя се неговият базис. Тъй като текущият връх $\overline{x}$ и следователно, стойността на целевата функция z в $\overline{x}$ не се променят, теоретично е възможно симплекс методът да зацикли, т.е. да преминава безкрайно през редица от базиси, на един и същи изроден връх, и да не го напусне. На практика това не представлява проблем, тъй като съществуват правила за избор на влизаща и излизаща от базиса променлива, които предотвратяват зациклянето. Пример за такова правило е правилото на Бленд, което се състои в това винаги да се избират за влизащи и излизащи от базиса, в случай че има повече от един кандидат, променливите с най-малък индекс. Като кандидат за влизане в базиса се разглежда всяка променлива x_q с отрицателна редуцирана цена, а като кандидат за влизаща в базиса променлива се разглежда всяка променлива x_{j_p} при която се достига минимува в теста за минимално отношение в стъпка (5). При спазване на правилото на Бленд, симплекс методът приключва след краен брой итерации.

Теорема 5.1. *Симплекс методът приключва след краен брой итерации, ако влизащата и излизащата променлива от базиса винаги се избират с най-малкия индекс.*

Доказателство. Достатъчно е да покажем, че при спазване на горното правило, алгоритъмът не зацикля, което ще направим като допуснем, че се образува цикъл и ще докажем, че това води до противоречие. И така, да допуснем, че въпреки спазването на правилото на Бленд се получава цикъл и че зациклянето е в самото начало на алгоритъма, в изродения връх $\overline{x}$. Нека $(\mathbf{B_0}, \mathbf{B_1}, \ldots, \mathbf{B_k})$ означава редица от базиси на $\overline{x}$, през които зацикля методът. Т.е. симплекс методът генерира следната редица от базиси на $\overline{x}$:

$$\mathbf{B_0}, \mathbf{B_1}, \ldots, \mathbf{B_k}, \mathbf{B_0}, \mathbf{B_1}, \ldots. \tag{5.2}$$

Ще казваме, че дадена променлива е *непостоянна*, ако тя участва в някой от тези базиси и не участва в някой друг базис. Нека x_p бъде непостоянната променлива с най-голям индекс и нека $\mathbf{B}$ означава базиса, напуснат от x_p. Нека x_q е променливата, влизаща в базиса $\mathbf{B}$ на мястото на x_p. Нека означим с B базисната матрица, съответстваща на базиса $\mathbf{B}$ и с $\mathbf{N}$ множеството от небазисни индекси, т.е. $\mathbf{N} := \{1, 2, \ldots, n\} \setminus \mathbf{B}$. Нека x е произволно решение на системата $Ax = b$. Разделяме $A = [B|N]$ и също както при (4.7) получаваме

$$x_B = \overline{x}_B - B^{-1}Nx_N \quad \sim \quad x_i = \overline{x}_i - \sum_{j\in\mathbf{N}} w_{ij}x_j, i \in \mathbf{B}. \tag{5.3}$$

Да напомним, че w_{ij} е i-та координата на вектора $w_j := B^{-1}a_j$, където a_j е j-ия стълб на A. а стойността на целевата функция в x имаме

$$\begin{aligned} z(x) &= c^{\mathrm{T}}x = c_B^{\mathrm{T}}x_B + c_N^{\mathrm{T}}x_N = c_B^{\mathrm{T}}(\overline{x}_B - B^{-1}Nx_N) + c_N^{\mathrm{T}}x_N = \\ &= c_B^{\mathrm{T}}\overline{x}_B + (c_N^{\mathrm{T}} - c_B^{\mathrm{T}}B^{-1}N)x_N = z(\overline{x}) + \sum_{j\in\mathbf{N}} \overline{c}_j x_j, \end{aligned} \tag{5.4}$$

където, както и преди, с $\overline{c}_j$ означаваме редуцираната цена на x_j.

Тъй като в базиса $\mathbf{B}$ влизащата променлива е x_q, а напускащата го е x_p, то $q \in \mathbf{N}$ и $p \in \mathbf{B}$ и следователно променливата x_q също е непостоянна променлива (не участва в базиса $\mathbf{B}$, но участва в следващия го базис). Оттук $q < p$.

Нека сега означим с $\mathbf{B}^*$ базиса, измежду базисите в (5.2) (очевидно различен от $\mathbf{B}$), в който променливата x_p влиза в базиса. Да означим с $\mathbf{N}^*$ множеството от небазисните спрямо базиса $\mathbf{B}^*$ индекси, т.е. $\mathbf{N}^* + \{1, \ldots, n\} \setminus \mathbf{B}^*$. Като направим горните разсъждения относно базиса $\mathbf{B}^*$ при разделянето на $A = [B^*|N^*]$, за решение x на системата $Ax = b$, имаме че

$$x_{B^*} = \overline{x}_{B^*} - (B^*)^{-1}N^*x_{N^*}, \tag{5.5}$$

а стойността на целевата функция в x е

$$z(x) = c^{\mathrm{T}}x = z(\overline{x}) + \sum_{j\in\mathbf{N}^*} \overline{c}_j^* x_j = z(\overline{x}) + \sum_{j=1}^{n} \overline{c}_j^* x_j, \tag{5.6}$$

където с $\overline{c}_j^*$ е означена редуцираната цена на x_j спрямо базиса $\mathbf{B}^*$, а последното равенство следва от това, че тъй като редуцираните цени на базисните променливи са нули можем да ги допишем в сумата.

От (5.3) забелязваме, че ако зададем някакви стойности на променливите x_N и пресметнем стойностите на променливите x_B от (5.3), то полученото $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$ е решение на системата $Ax = b$.

Без да се интересуваме от неотрицателност на координатите, търсим решение на системата $Ax = b$, което се получава като зададем $x_q = y$ за произволно $y \in \mathbb{R}$ и $x_j = 0$ за $j \in \mathbf{N}$, $j \neq q$. От (5.3) получаваме $x_i = \overline{x}_i - w_{iq}y$ за $i \in \mathbf{B}$.

Стойността на целевата функция в така полученото x според (5.4) е $z(x) = z(\overline{x}) + \overline{c}_q y$, а според (5.6) е $z(x) = z(\overline{x}) + \overline{c}_q^* y + \sum_{i\in\mathbf{B}} \overline{c}_i^*(\overline{x}_i - w_{iq}y)$. Като приравним тези два израза получаваме

$$(\overline{c}_q - \overline{c}_q^* + \sum_{i\in\mathbf{B}} \overline{c}_i^* w_{iq})y = \sum_{i\in\mathbf{B}} \overline{c}_i^* \overline{x}_i.$$

Тъй като това равенство трябва да е в сила за всяко $y \in \mathbb{R}$, то коефициентът пред y (а също така и дясната част на равенството), трябва да е нула. Оттук,

$$\overline{c}_q - \overline{c}_q^* + \sum_{i \in \mathbf{B}} \overline{c}_i^* w_{iq} = 0.$$

Тъй като x_q влиза в базиса $\mathbf{B}$, това означава, че $\overline{c}_q < 0$. Тъй като x_q не е влизащата променлива в $\mathbf{B}^*$ (влизаща там беше x_p) и $q < p$, то по правилото на Бленд за избор на влизаща в базиса променлива, $\overline{c}_q^* \geqslant 0$. Следователно,

$$\sum_{i \in \mathbf{B}} \overline{c}_i^* w_{iq} > 0,$$

което означава, че съществува индекс $r \in \mathbf{B}$, такъв че

$$\overline{c}_r^* w_{rq} > 0. \tag{5.7}$$

Следователно, $\overline{c}_r^* \neq 0$ (тъй като за базисните променливи редуцираните цени са нула) и $r \in \mathbf{N}^*$. Получаваме, че променливата x_r е непостоянна (участва в базиса $\mathbf{B}$ и не участва в базиса $\mathbf{B}^*$) откъдето $r \leqslant p$. В действителност, $r < p$, тъй като $\overline{c}_p^* w_{pq} < 0$ ($\overline{c}_p^* < 0$, понеже x_p влиза в базиса $\mathbf{B}^*$, а $w_{pq} > 0$, понеже x_p напуска базиса $\mathbf{B}$, за да влезе x_q, и следователно w_{pq} е ключовото число).

Това, че $r < p$ влече, че $\overline{c}_r^* \geqslant 0$, защото в противен случай, съгласно критерия за избор на променлива с най-малък индекс, x_r щеше да е влизащата променлива в $\mathbf{B}^*$. Следователно, (5.7) влече, че

$$w_{rq} > 0.$$

Тъй като всеки от базисите се отнася до един и същи връх, то всяка непостоянна променлива е с нулева стойност във всеки от тези базиси. В частност,

$$\overline{x}_r = 0.$$

От двете последни следва, че x_r е била кандидат за напускаща базиса $\mathbf{B}$ променлива и понеже $r < p$, то по правилото на Бленд, тя е трябвало да бъде избрана за напускане на базиса преди променливата x_p. Получихме търсеното противоречие и с това доказахме теоремата. ■

Задачи, при които симплекс методът минава през изродени върхове са често срещани. Ако се разпечатат и разгледат стойностите на целевата функция след всяка итерация, това че методът е преминал многократно през изроден връх се разпознава по факта, че една и съща стойност се повторя неколкократно (преди симплекс методът да намери ръб, по който да тръгне), а след това стойността на целевата функция отново намалява.

Спазването на правилото на Бленд гарантира, че ако текущият изроден връх не е оптимално решение, то след краен брой итерации методът ще го напусне и ще отиде в съседен връх, където стойността на целевата функция ще е строго по-малка.

6. Реализации на симплекс алгоритъма.

Във въпрос 4 се запознахме със симплекс метода за решаване на каноничната линейна задача

$$\begin{array}{ll} \text{да се минимизира} & z(x) = c^{\mathrm{T}}x \\ \text{при ограничения} & Ax = b\,, \\ & x \geqslant 0\,, \end{array} \tag{6.1}$$

където A означава $(m \times n)$ матрица, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$. Сега ще се спрем на някои негови приложни реализации.

6.1. Таблична форма.

При всяка итерация по симплекс метода се прави избор на влизаща и излизаща от базиса променлива и последваща смяна на базиса и базисната матрица. Един от начините да се организира итерацията е да се вложат данните за текущото решение в една голяма матрица, наречена *симплексна таблица*, която може да се генерира директно от данните на задачата.

При зададени матрица A, дясна част b, и вектор на целевата функция c, началната таблица е просто следната по-голяма матрица:

$$\tilde{\tilde{T}} = \left[\begin{array}{c|c} b & A \\ \hline 0 & c^{\mathrm{T}} \end{array}\right].$$

Таблицата $\tilde{\tilde{T}}$ е с $m+1$ реда, които ще номерираме от 1 до $m+1$ и $n+1$ стълба, които ще номерираме от 0 до n.

Нека $\overline{x}$ е текущото базисно допустимо решиние. Ако е необходимо преномерираме променливите така, че $x_1, \ldots, x_m$ да бъдат базисните променливи. Тогава базисната матрица B се състои от първите m стълба на A, а последните $(n-m)$ стълба образуват $m \times n - m$ подматрицата N. По подобен начин разделяме координатите на вектора на целевата функция c на $[c_B, c_N]$, и координатите на променливите x на $[x_B, x_N]$. След преномерирането на променливите в таблицата имаме, че първите стълбове са базисните, т.е.

$$\tilde{\tilde{T}} = \left[\begin{array}{c|c|c} b & B & N \\ \hline 0 & c_B^{\mathrm{T}} & c_N^{\mathrm{T}} \end{array}\right].$$

Целта ни е пред базисните променливи да получим единичната матрица на мястото на матрицата B. Чрез елементарни (Гаусови) преобразувания на горната част на таблицата получаваме

$$\tilde{T}(\overline{x}) = \left[\begin{array}{c|c|c} B^{-1}b & I & B^{-1}N \\ \hline 0 & c_B^{\mathrm{T}} & c_N^{\mathrm{T}} \end{array}\right].$$

Матрицата, която преобразува B в I очевидно е матрицата B^{-1}, с други думи направените елементарни преобразувания са еквивалентни на умножение с матрицата B^{-1} на горната част на таблицата. За да завършим, умножаваме с c_B^{T} горната част на таблицата и я вадим от долната:

$$T(\overline{x}) = \left[\begin{array}{c|c|c} B^{-1}b & I & B^{-1}N \\ \hline -c_B^{\mathrm{T}}B^{-1}b & 0 & c_N^{\mathrm{T}} - c_B^{\mathrm{T}}B^{-1}N \end{array}\right].$$

Това е окончателния вид на симплексната таблица за базисното допустимо решение $\overline{x}$. Тя съдържа цялата информация, необходима на симплексната итерация. Най-вляво горе са базисните координати $\overline{x}_B = B^{-1}b$ (небазисните координати са $\overline{x}_N = 0$). Текущата стойност на целевата функция е $c^{\mathrm{T}}\overline{x} = c_B^{\mathrm{T}}B^{-1}b$ и може да се намери в долния ляв ъгъл на таблицата с обратен знак.

По важното е, че от таблицата можем да преценим дали базисното допустимо решение $\overline{x}$ е оптимално или не като разгледаме координатите на вектора на редуцираните цени $\overline{c}_N = c_N^{\mathrm{T}} - c_B^{\mathrm{T}}B^{-1}N$ който се намира в долния десен ъгъл на таблицата.

Ако $\overline{c}_N \geqslant 0$, то $\overline{x}$ удовлетворява критерия за оптималност и следователно е оптимално решение, а оптималната стойност на целевата функция е $z(\overline{x}) = c_B^{\mathrm{T}}B^{-1}b$.

Ако измежду координатите на $\overline{c}_N$ има отрицателни, то $\overline{x}$ не удовлетворява критерия за оптималност стойността на целевата функция още може да се намали. Всяка отрицателна координата на вектора $\overline{c}_N$ съответства на ръб, по който целевата функция намалява. Обичайната стратегия е да се избере най-малката отрицателна координата на $\overline{c}_N$.

Да допуснем, че най-отрицателната редуцирана цена е $\overline{c}_q$. Тогава влизащата в базиса променлива е q-та. За да определим на мястото на коя от базисните променливи ще влезе, ни е необходимо да определим θ от теста за минимално отношение:

$$\theta = \min\left\{\frac{(B^{-1}b)_i}{(B^{-1}a_q)_i} = \frac{\overline{x}_i}{w_{iq}} : (B^{-1}a_q)_i = w_{iq} > 0, 1 \leqslant i \leqslant m\right\} = \frac{\overline{x}_p}{w_{pq}} = \frac{(B^{-1}b)_p}{(B^{-1}a_q)_p}\,, \tag{6.2}$$

където векторът $w_q = B^{-1}a_q$ се намира в q-ия стълб на таблицата, а a_q е q-ия стълб на N.

Стълбовете $B^{-1}b$ и $B^{-1}a_q$, чиито координати участват в теста за минимално отношение също се намират в таблицата като стълбът $B^{-1}b$ вече казахме са намира горе в ляво, а стълбът $w_q = B^{-1}a_q$ стои над избраната най-малка координата на вектора $\overline{c}_N$, а именно над $\overline{c}_q$.

Минимумът се взима само по положителните координати на w_q. Ако в стълба на w_q няма положителни координати, то от текущия връх $\overline{x}$ излиза неограничен ръб, по който целевата функция намалява до $-\infty$.

В противен случай се преминава към съседно базисно допустимо решение $\overline{x}'$, чиято базисна матрица B' се получава като старият p-ти стълб на B, а именно a_p се замени със стълба a_q. Симплексната таблица на новото базисно допустимо решение $T(\overline{x}')$ се получава директно от таблицата $T(\overline{x})$ на предходното, като стъпка (6) на симплекс алгоритъма се заменя с т.нар. "завъртане". Завъртането прави следното:

(1) дели p-ия ред на $T(\overline{x})$ на $t_{p,q}$ ((p, q)-ия елемент на $T(\overline{x})$) и

(2) за $1 \leqslant i \leqslant m + 1$, $i \neq p$ вади от i-ия ред произведението на t_{iq} с новия p-ти ред, при което се получава нула в q-ия стълб на новия i-ти ред.

Завъртането създава таблицата $T(\overline{x}')$ на новото базисно допустимо решение във същия вид като използва само елементите на таблицата $T(\overline{x})$ на предходното базисно допустимо решение. Така при всяка итерация се създава симплексна таблица за текущото решение, от която се черпи всичката необходима информация - оптималност, неограниченост, критерии за край на алгоритъма.

Често именно табличната форма на симплекс метода се нарича симплекс метод, тъй като това е формата в която симплекс методът е бил описан при своето създаване. Въпреки, че тази форма е приемлива за малки "учебни" примери, тя не е подходяща за решаване на големи задачи и на задачи, в които матрицата A е с някаква структура (например с много нули), които често възникват в практиката. Това е така, тъй като завъртането в таблицата обикновено разрушава структурата на матрицата A. Друг недостатък е, че завъртането генерира всички стълбове на $m \times (n-m)$ матрицата $B^{-1}N$ на всяка итерация докато само един от нейните стълбове, стълба $B^{-1}a_q$, е необходим за теста за минимално отношение.

6.2. Модифициран симплекс метод.

Да оставим настрана табличната форма на симплекс метода и да погледнем дадения във въпрос 4 алгоритъм на симплекс метода, за да установим кои пресмятания са ни наистина необходими. Намирайки се във върха $\overline{x}$ с базисна матрица B итерацията решава (в зависимост от $\overline{c}$ и θ) коя променлива да влезе, и коя да излезе от базиса. Необходимите за това пресмятания, отразени в алгоритъм, изложен във въпрос 4, обикновено се наричат *модифициран симплекс метод*, за да се прави разлика с табличната форма на метода. За да извършим необходимите пресмятания, на стъпки (1), (4) и (5) от алгоритъма трабва да се решат относно π^{T}, w_q и $\overline{x}_B$ следните системи линейни уравнения:

$$\pi^{\mathrm{T}} = c_B B^{-1}, \quad w_q = B^{-1}a_q, \quad \text{и} \quad \overline{x}_B = B^{-1}b. \tag{6.3}$$

Да забележим, че матрицата от коефициенти във всяка от трите системи е матрицата B^{-1}.

При първоначалните реализации на модифицирания симплекс метод матрицата B^{-1} се съхранява и поддържа в явен вид като се обновява при всяка смяна на базиса.

При последвалите реализации матрицата B^{-1} се намира в явен вид за началния връх, а в следващите върхове се процедира по следния начин: Следващата базисна матрица (която се получава като p-ия стълб на B се замени със стълба a_q) може да се получи след умножение отдясно на B с матрица, която ще означим с E, т.е.

$$B' = BE\,,$$

където

$$E^{-1} = \begin{bmatrix} 1 & & w_{1q} & & \\ & \ddots & \vdots & & \\ & & w_{pq} & & \\ & & \vdots & \ddots & \\ & & w_{mq} & & 1 \end{bmatrix} \tag{6.4}$$

Лесно се проверява, че умножаването на B отдясно с E оставя всички стълбове на B непроменени с изключение на p-ия, който се трансформира в $Bw_q = a_q$, както е необходимо.

За да обновим матрицата B^{-1}, да забележим, че

$$(B')^{-1} = (BE)^{-1} = E^{-1}B^{-1},$$

където

$$E^{-1} = I - \frac{(w_q - e_p)e_p^{\mathrm{T}}}{w_{pq}} = \begin{array}{c} \begin{bmatrix} 1 & & -w_{1q}/w_{pq} & & \\ & \ddots & \vdots & & \\ & & 1/w_{pq} & & \\ & & \vdots & \ddots & \\ & & -w_{mq}/w_{pq} & & 1 \end{bmatrix} \\ \uparrow \\ \text{стълб } p \end{array} \tag{6.5}$$

е обратната матрица на матрицата E от (6.4).

Тъй като матрицата B от своя страна може да се получи чрез замяна на стълбовете на I, със стълбовете на B, следва, че всяка обратна на базисна матрица можем да изразим *във вид на произведение* като

$$B^{-1} = E_k^{-1} E_{k-1}^{-1} \cdots E_1^{-1}, \tag{6.6}$$

където всяка от матриците E_i^{-1} е от вида (6.5) и е прието да се нарича елементарна матрица. Обновяването на обратната на базисната матрица става просто чрез добавяне на нова елементарна матрица в произведението.

Ясно е, че вместо да се съхранява цялата матрица E_i^{-1}, достатъчно е да се съхранява само стълба на E_i^{-1}, който се различава от стълб на единичната матрица и неговото място в E_i^{-1}. Също така е препоръчително след натрупването на m на брой елементарни матрици, матрицата B^{-1} да се преизчислява, а използваните за нейното преизчисляване елементарни матрици да се изтриват. По този начин редицата от елементарни матрици се поддържа до редица от не повече от m матрици. Това спестява изчислетелно време на следващите итерации, спестява място за съхранение и намалява грешките от закръгляване.

Още по-нов и съвременен подход като се използуват обичайните методи на числената линейна алгебра системите (6.3) да се разглеждат като три системи с една и съща матрица на коефициентите B:

$$\pi B = c_B, \quad B w_q = a_q, \quad B x_B = b. \tag{6.7}$$

Тогава стандартното разлагане на матрицата $B = LU$, където L е долна триъгълна, а U горна триъгълна матрица води до лесно намиране на решенията на трите системи. Елементите на разлагането L' и U' на следващата базисна матрица $B' = L'U'$ не се преизчисляват, а се получават от L и U чрез умножение с елементарни матрици от вида (6.4).

6.3. Изкуствени променливи и фаза I на симплекс метода

Една от добрите характеристики на симплекс метода, че той може да се използва за намиране на базисно допустимо решение на каноничната задача (6.1). Прилагането на симплекс метода за решаването на тази задача за допустимост се отнася до т.нар. фаза I (вж. въпрос 4), докато прилагането му за намиране на опитимално решение на (6.1) беше наречено фаза II и разгледано детайлно във въпрос 4. Задачата за намиране на базисно допустимо решение за задачата (4.1) може да се дефинира като следната (изкуствена) минимизационна задача

$$\begin{array}{ll} \text{да се минимизара} & \displaystyle\sum_{i=1}^{m} y_i \\ \text{при ограничения} & Ax + y = b \quad (b \geqslant 0)\,, \\ & x \geqslant 0\,, \quad y \geqslant 0\,. \end{array}$$

Тази задача има очевидно базисно допустимо решение $x = 0$, $y = b$, с базис, състоящ се изцяло от променливите y и базисна матрица I. Следователно незабавно можем да приложим фаза II на симплекс метода, използвайки това базисно допустимо решение за начален връх. Променливите y_i, $i = 1, \ldots, m$, се наричат *изкуствени променливи* и целта на горната минимизационна задача е да ги сведе до нулева стойност. Тази задача винаги има крайно оптимално решение, тъй като целевата й функция е ограничена отдолу върху допустимото й множество. Ако изходната задача има допустимо решение (т.е. допустимото множество P не е празното), то тогава това е възможно. В този случай симплекс методът ще приключи с намирането на базисно допустимо решение, за което всичките $y_i = 0$. Ако това решение е изродено, то всяка изкуствена променлива, която е останала в базиса може или да се смени с някоя небазисна x-променлива, или да бъде отстранена, заедно с излишното уравнение, към което е асоциирана. По този начин, след решаването на изкуствената задача се получава базисно допустимо решение, включващо само x-променливи, което се явява базисно допустимо решение за изходната задача (6.1) и с него може да започне изпълнението на фаза II на симплекс метода за нея.

По точно казано, ако $y_i = 0$ е k-та базисна променлива и $e_k^{\mathrm{T}} B^{-1} a_j \neq 0$, то y_i може да се замени с небазисната променлива x_j. Ако $e_k^{\mathrm{T}} B^{-1} a_j = 0$ за всички $j \notin \boldsymbol{B}$, в оригиналната система линейни уравнения $Ax = b$, k-то е излишно и k-ия ред може да се отстрани от матрицата A. Ако изходната задача (6.1) няма допустимо решение (т.е. множеството й от допустими решения P е празното множество), то фаза I ще приключи със строго положителна стойност на целевата функция на задачата с изкуствените променливи.

Друг, често използван подход, е да се комбинират фаза I и фаза II в една фаза, като се минимизира върху множеството от ограничения на изходната задача (6.1) следната целева функция

$$z = \sum_{j=1}^{n} c_j x_j + M \sum_{i=1}^{m} y_i$$

където числото M се избира толкова голямо, че всички изкуствени променливи от даден момент нататък да бъдат сведени до нула, в случай че изходната задача има допустимо решение. Този подход още се нарича M-задача.

7. Двойнственост в линейното оптимиране

В този въпрос ще въведем и разгледаме важното понятие за двойнственост в линейното оптимиране. В частност ще покажем, че с всяка линейна задача може да се асоциира друга линейна задача, наречена нейна двойнствена, която е тясно свързана с условията за оптималност на решението на оригиналната задача. Ще дадем икономическа интерпретация на променливите на двойнствената задача и ще дадем пример как може да се интерпретира двойнствената задача.

7.1. Двойнственост и оптималност

Както обикновено, разглеждаме каноничната задача на линейното оптимиране

$$
\begin{array}{lll}
(P) & \text{да се минимизира} & z(x) = c^{\mathrm{T}}x \\
 & \text{при ограничения} & Ax = b\,, \\
 & & x \geqslant 0\,,
\end{array}
\tag{7.1}
$$

където A означава $(m \times n)$ матрица, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$.

Ще наричаме горната задача още *права задача* и ще асоциираме с нея друга задача на линейното оптимиране (D), която ще наричаме *двойнствена* на (P):

$$
\begin{array}{lll}
(D) & \text{да се максимизира} & v(y) = b^{\mathrm{T}}y \\
 & \text{при ограничения} & A^{\mathrm{T}}y \leqslant c\,.
\end{array}
\tag{7.2}
$$

Да забележим, че двойнствената задача има същите данни като правата, а именно A, b, и c, че векторът на променливите $y \in \mathbb{R}^m$ и че двойнствената в някакъв смисъл се явява транспонирана версия на правата, при която минимизацията е заменена с максимизация.

За да се напише двойнствената задача на задача на линейното оптимиране в общ вид (не само в канонична форма), се постъпва по следния начин: дадената задача се привежда в канонична форма посредством правилата, изложени във въпрос 0 и по горния начин се образува нейната двойнствена.

Като приложим този подход към двойнствената задача (D) и опростим, получаваме

Лема 7.1. *Двойнствената задача на двойнствената задачата (D) е правата задача (P).*

Доказателство. Канонизираме (D) като положим $y = y^+ - y^-$, $y^+ \geqslant 0$, $y^- \geqslant 0$ и въведем допълнителни променливи $s \geqslant 0$, пишем двойнствената на каноничната на (D) по горното правило и опростяваме, при което получаваме (P).

$$
\begin{array}{llcll}
(D) & \max b^{\mathrm{T}}y & \Longrightarrow & \min -b^{\mathrm{T}}(y^+ - y^-) & \Longrightarrow \\
 & A^{\mathrm{T}}y \leqslant c & & A^{\mathrm{T}}(y^+ - y^-) + s = c & \\
 & & & y^+ \geqslant 0, y^- \geqslant 0, s \geqslant 0 &
\end{array}
$$

като положим $u := [y^+, y^-, s] \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n$ получаваме следната канонична задача, на която пишем двойнствената, чийто вектор на променливите означаваме с $x' \in \mathbb{R}^n$

$$
\begin{array}{lclcrclcll}
\min[-b, b, 0]^{\mathrm{T}}u & \Longrightarrow & \max c^{\mathrm{T}}x' & \Longrightarrow & \multicolumn{3}{l}{\max c^{\mathrm{T}}x'} & \Longrightarrow & (P) & \min c^{\mathrm{T}}x \\
[A^{\mathrm{T}}| - A^{\mathrm{T}}|I^{\mathrm{T}}]u = c & & & & Ax' & \leqslant & -b & & & Ax = b \\
u \geqslant 0 & & [A| - A|I]x' \leqslant \begin{bmatrix} -b \\ b \\ 0 \end{bmatrix} & & -Ax' & \leqslant & b & & & x \geqslant 0, \\
 & & & & x' & \leqslant & 0 & & &
\end{array}
$$

като последното получаваме след комбинирането на първите две матрични неравенства в $Ax' = -b$ и полагането $x := -x'$. ■

Задачите, (P) и (D) се наричат *двойка спрегнати задачи.*

Двойнствената задача е тясно свързана с критерия за оптималност на базисно допустимо решение за правата задача. Нека $x^* = \begin{bmatrix} x_B^* \\ x_N^* \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$, $x^* \geqslant 0$ е базисно допустимо решение за (P), за което е в сила критерият за оптималност:

$$\bar{c}_N^{\mathrm{T}} = c_N^{\mathrm{T}} - c_B^{\mathrm{T}} B^{-1} N \geqslant 0.$$

Да положим $y^{*\mathrm{T}} := c_B^{\mathrm{T}} B^{-1}$. Критерият за оптималност е еквивалентен на

$$c_N^{\mathrm{T}} \geqslant y^{*\mathrm{T}} N,$$

а очевидно $c_B^{\mathrm{T}} = c_B^{\mathrm{T}} B^{-1} B = y^{*\mathrm{T}} B$. Оттук

$$c = [c_B^{\mathrm{T}}, c_N^{\mathrm{T}}] \geqslant [y^{*\mathrm{T}} B, y^{*\mathrm{T}} N] = y^{*\mathrm{T}}[B|N] = y^{*\mathrm{T}} A = A^{\mathrm{T}} y^*,$$

т.е. y^* е допустимо решение на задачата (D). Да забележим, че така положеното $y^{*\mathrm{T}}$ всъщност е векторът на симплексните множители $\pi^{\mathrm{T}} = c_B^{\mathrm{T}} B^{-1}$ пресмятан в стъпка (1) на симплекс метода (вж. въпрос 4).

Следователно, от критерия за оптималност на базисно допустимо решение за правата задача след полагане получаваме допустимо решение за двойнствената задача. Имаме следната

Теорема 7.1 (Слаба теорема за двойнственост). *Ако x е допустимо решение за правата задача (P) и y е допустимо решение за двойнствената задача (D), то $b^{\mathrm{T}} y \leqslant c^{\mathrm{T}} x$.*

Доказателство. От това, че x е допустимо за правата задача $Ax = b$ и $x \geqslant 0$. За произволно $y \in \mathbb{R}^m$ имаме $y^{\mathrm{T}} b = y^{\mathrm{T}} Ax = A^T yx$. Ако y е допустимо за двойнствената задача $A^{\mathrm{T}} y \leqslant c$ и от $x \geqslant 0$, получаваме, че $A^{\mathrm{T}} yx \leqslant c^{\mathrm{T}} x$. Следователно $b^{\mathrm{T}} y \leqslant c^{\mathrm{T}} x$. ■

Слабата теорема за двойнственост всъщност твърди, че стойността на целевата функция в допустимо решение за правата (двойнствената) задача дава горна (долна) граница на стойността на целевата функция на другата задача в произволно нейно допустимо решение (включително и оптималното). Като следствие от теоремата получаваме

Следствие 7.1. *Ако x^* е допустимо решение за правата задача (P), а y^* е допустимо решение за двойнствената задача (D), за които $c^{\mathrm{T}} x^* = b^{\mathrm{T}} y^*$, то x^* и y^* са оптимални решения на съответните задачи.*

Доказателство. Тъй като по слабата теорема за двойнственост, за всяко допустимо решение x на (P) имаме, че $c^{\mathrm{T}} x \geqslant b^{\mathrm{T}} y^* = c^{\mathrm{T}} x^*$, то x^* е оптимално за (P).

Тъй като по слабата теорема за двойнственост, за всяко допустимо решение y на (D) имаме, че $b^{\mathrm{T}} y \leqslant c^{\mathrm{T}} x^* = b^{\mathrm{T}} y^*$, то y^* е оптимално за (D). ■

Съществуват ли обаче допустими решения x^* и y^* за съответните задачи, които да удовлетворяват хипотезите на това следствие? Отговор на този въпрос дава:

Теорема 7.2 (Силна теорема за двойнственост). (а) *Ако едната от двойката спрегнати задачи (P) и (D) има крайно оптимално решение, то крайно оптимално решение има и другата задача, като* $\min c^{\mathrm{T}} x = \max b^{\mathrm{T}} y$.

(б) *Ако едната от двойката спрегнати задачи има неограничена целева функция, то допустимото множество на другата задача е празно.*

Доказателство. От Лема 7.1 е ясно, че всяка една от задачите (P) и (D) може да се разглежда като права задача. Поради това достатъчно е да намерим оптимално решение x^* на правата задача (P) и допустимо решение y^* на (D), които да удовлетворяват $c^{\mathrm{T}}x^* = b^{\mathrm{T}}y^*$ и след това да използваме Следствие 7.1 за да получим оптималността на y^*. Нека $x^* = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$ е оптимално базисно допустимо решение на (P), получено по симплекс метода и да положим $y^{*\mathrm{T}} = c_B^{\mathrm{T}}B^{-1}$. Вече видяхме, че y^* е допустимо за двойнствената задача. Проверяваме, че

$$c^{\mathrm{T}}x^* = [c_B^{\mathrm{T}}, c_N^{\mathrm{T}}] \begin{bmatrix} x_B^* \\ x_N^* \end{bmatrix} = [c_B^{\mathrm{T}}, c_N^{\mathrm{T}}] \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} = c_B^{\mathrm{T}}B^{-1}b = y^{*\mathrm{T}}b = b^{\mathrm{T}}y^*\,,$$

което приключва доказателството на (а).

(б) следва непосредствено от Теорема 7.1. ■

Доказателството на теоремата показва, че векторът от симплексните множители, съответстващ на оптимално решение x^* на правата задача, всъщност представлява оптимално решение y^* на двойнствената задача. Наистина, на всяка итерация на симплекс метода, симплексните множители образуват вектор y, за който $c^{\mathrm{T}}x = b^{\mathrm{T}}y$, но освен в случая, когато всички редуцирани цени са неотрицателни, y не е допустим за двойнствената задача. Тоест, симплекс алгоритъмът на всяка итерация поддържа допустимостта на решението за правата задача и това, че $c^{\mathrm{T}}x = b^{\mathrm{T}}y$, като се стреми да постигне допустимост на y за двойнствената задача.

Да отбележим също, че обратното на (б) не винаги е вярно, т.е. ако допустимото множество на едната от двойката спрегнати задачи е празно, то от това НЕ следва че другата задача е с неограничена целева функция; възможно е и двете задачи да имат празно допустимо множество.

Добре известна и много полезна теорема за системи от равенства и неравенства е

Теорема 7.3 (Лема на Фаркаш). *Системата*

(I) $\quad Ax = b\,, \quad x \geqslant 0\,,$

е съвместима, тогава и само тогава, когато е несъвместима системата

(II) $\quad y^{\mathrm{T}}A \leqslant 0, \quad b^{\mathrm{T}}y > 0.$

Доказателство. Да разгледаме следната двойката спрегнати задачи

$$\textbf{(P)}\quad \min\ 0^{\mathrm{T}}x,\quad Ax = b\,,\quad x \geqslant 0\,, \qquad\qquad \textbf{(D)}\quad \max\ b^{\mathrm{T}}y,\quad y^{\mathrm{T}}A \leqslant 0\,.$$

Нека (I) е съвместима и нека x е нейно решение. Да допуснем, че (II) също е съвместима и нека y е нейно решение. Тогава

$$0 < b^{\mathrm{T}}y = y^{\mathrm{T}}b = y^{\mathrm{T}}Ax \leqslant 0,$$

което води до противоречие.

Нека сега (II) е несъвместима, т.е. за всяко y, такова че $y^{\mathrm{T}}A \leqslant 0$ имаме, че $b^{\mathrm{T}}y \leqslant 0$. Това означава, че $(\mathbf{D})$ има непразно допустимо множество ($y = 0$ е допустимо решение) и че целевата ѝ функция е ограничена отгоре върху допустимото ѝ множество. Следователно, (D) има крайно оптимално решение. От силната теорема за двойнственост, (P) също има крайно оптимално решение, в частност системата (I) е съвместима. ■

Лемата на Фаркаш, датираща от 1902 г. стои в основата на развитието на линейното оптимиране и често се използва за доказване на Теоремата за двойнственост, вместо както ние постъпихме. Геометрично, тя твърди, че точно едно от следните твърдения е вярно:

(I) b е в изпъкналия конус C, породен от стълбовете на A; или

(II) съществува вектор y, който сключва остър ъгъл с b, но с нито един от векторите в C.

Дотук видяхме, че на всяка линейна задача съответства двойнствена линейна задача, като слабата теорема за двойнственост, нейното следствие и силната теорема за двойнственост са в сила за двойката спрегнати задачи.

Често при изследване на двойнствеността се разглежда следната права задача

$$(P') \qquad \begin{array}{l} \min c^{\mathrm{T}} x \\ Ax \geqslant b\,, \\ x \geqslant 0\,. \end{array}$$

За да намерим нейната двойнствена (D'), както вече казахме, привеждаме (P') в канонична форма, на получената канонична задача пишем двойнствената и опростяваме:

$$(P') \quad \begin{array}{l} \min c^{\mathrm{T}} x \\ Ax \geqslant b\,, \\ x \geqslant 0\,. \end{array} \implies \begin{array}{l} \min c^{\mathrm{T}} x \\ Ax - x' = b\,, \\ x \geqslant 0\,,\ x' \geqslant 0. \end{array} \implies \begin{array}{l} \max b^{\mathrm{T}} y \\ A^{\mathrm{T}} y \leqslant c\,, \\ -I^{\mathrm{T}} y \leqslant 0. \end{array} \implies (D') \quad \begin{array}{l} \max b^{\mathrm{T}} y \\ A^{\mathrm{T}} y \leqslant c, \\ y \geqslant 0. \end{array}$$

Двойката спрегнати задачи (P') и (D') е "симетрична", в смисъл, че и двете задачи са с неотрицателни променливи и с неравенства от вида "$\geqslant$" в минимизационната задача и от вида "$\leqslant$" в максимизационната задача.

Ще докажем още две теореми, които характеризират оптималните решения на тази двойка спрегнати задачи.

Теорема 7.4 (Условия за допълнителност). *Нека x и y са допустими решения съответно на правата (P') и двойнствената задача (D'). Необходими и достатъчни условия те да са оптимални решения на съответните задачи са*

$$(c^{\mathrm{T}} - y^{\mathrm{T}} A)x = 0 \tag{7.3}$$

и

$$y^{\mathrm{T}}(Ax - b) = 0\,. \tag{7.4}$$

Доказателство. За допустимото x за (P') и за допустимото y за (D') имаме

$$s \equiv Ax - b \geqslant 0, \quad x \geqslant 0, \quad \text{и} \quad w^{\mathrm{T}} \equiv c^{\mathrm{T}} - y^{\mathrm{T}} A \geqslant 0,\ y \geqslant 0\,. \tag{7.5}$$

Следователно,

$$0 \leqslant (c^{\mathrm{T}} - y^{\mathrm{T}} A)x = c^{\mathrm{T}} x - y^{\mathrm{T}} Ax \leqslant c^{\mathrm{T}} x - y^{\mathrm{T}} b\,,$$

или, което е еквивалентно

$$c^{\mathrm{T}}x \geqslant y^{\mathrm{T}}Ax \geqslant y^{\mathrm{T}}b. \tag{7.6}$$

Ако условията (7.3) са в сила, то неравенствата в горната редица се превръщат в равенства и оптималността на x и y идва от Следствие 7.1.

Обратно, ако x и y са оптимални, от силната теорема за двойнственост следва, че $c^{\mathrm{T}}x = y^{\mathrm{T}}b$, откъдето в (7.6) имаме равенства вместо неравенства и, следователно, условията (7.3) и (7.4) са удовлетворени. ■

За двойката спрегнати задачи (P') и (D'), поради неотрицателността на правите и двойнствените променливи x и y, както и на допълнителните променливи s и w (вж. (7.5)) условията за допълнителност (7.3) и (7.4) могат да се изкажат в следната по-полезна форма.

$$\begin{array}{lll} w_j \equiv (c - A^{\mathrm{T}}y)_j = 0 & \text{или } x_j = 0\,, & \text{за всяко } j = 1, \dots, n\,, \\ s_i \equiv (Ax - b)_i = 0 & \text{или } y_i = 0\,, & \text{за всяко } i = 1, \dots, m\,. \end{array} \tag{7.7}$$

От условията за допълнителност в този им вид, е ясно, че допустими решения x на правата (P') и y двойнствената задача (D') са оптимални, тогава и само тогава, когато

(1) променлива x_j (съотв. y_i)е нула в една от задачите, когато съответната й допълнителна променлива w_j (съотв. s_i) е строго положителна (т.е., съответното ограничение се удовлетворява като строго неравенство) и

(2) допълнителна променлива w_j (съотв. s_i) е нула (т.е., съответното ограничение неравенство се удовлетворява като равенство) в една от задачите, когато съответната й променлива x_j (съотв. y_i) е положителна в другата задача.

Да отбележим, че за двойката спрегнати задачи (P) и (D), която разглеждахме първоначално, от значение е само условието (7.3), тъй като (7.4) е вярно за всяко допустимо решение x на правата задача (P).

7.2. Икономическа интерпретация на двойнствеността

Казахме вече, че двойнствената задача възниква естествено от критерия за оптималност на правата задача. Да допуснем, че координатите b_i на вектора дясна част b на правата задача (P) представляват количества от някакви рерурси i, $i = 1, \dots, m$. Да предположим, че

$$x^* = \begin{pmatrix} x_B^* \\ 0 \end{pmatrix} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$$

е неизродено оптимално базисно допустимо решение на (P). От предположението за неизроденост $(x_B^* > 0)$ следва, че малка промяна Δb на вектора b не би променила оптималния базис B (тъй като $\overline{c}_N = c_N - c_B B^{-1} N$ не се влияе от изменение на b, то малко изменение в дясната част b няма да повлияе на оптималността, а само на допустимостта на решението). Следователно, ако b се замени с $b + \Delta b$, новото оптимално решение за малко Δb става

$$\hat{x}^* = \begin{pmatrix} \hat{x}_B^* \\ 0 \end{pmatrix} = \begin{bmatrix} B^{-1}(b + \Delta b) \\ 0 \end{bmatrix},$$

а оптималната стойност на целевата функция се изменя с

$$\Delta z = c_B^{\mathrm{T}} B^{-1} \Delta b = y^{*\mathrm{T}} \Delta b,$$

където $y^{*\mathrm{T}} = c_B^{\mathrm{T}} B^{-1}$ и от доказателството на Теорема 7.2, y^* е оптимално решение на двойнствената задача (D). Оттук е ясно, че оптималната стойност на двойнствената променлива y_i^* може да се разглежда като неявна цена, или цена в сянка, на i-ия ресурс, тъй като y_i^* представлява изменението в оптималната стойност на целевата функция при изменение с единица на този ресурс. Тази икономическа интерпретация е много полезна понеже тя показва кои са ценните ресурси - тези, които водят до подобряване на оптималната стойност на целевата функция z.

Да отбележим, че от условията за допълнителност (7.7) имаме, че неявната цена на ресурс е нула, ако този ресурс не е използван напълно в оптималното решение.

Тези неявни цени също така са полезни, когато при вече намерено оптимално решение трябва да се реши дали да се включи ново производство (или да се започне нова дейност). Като пример да разгледаме задачата за диета (вж. въпрос 0). Задачата беше да се определи най-евтина диета като се удовлетворят минималните дневни потребности от различни хранителни съставки. Да предположим, че имаме в наличност n на брой различни храни и че диетата трябва да задоволява дневният минимум от m на брой хранителни съставки. Нека c_j е единичната цена на j-та храна, с b_i минималната дневна потребност от i-та хранителна съставка и с a_{ij} количеството от хранителната съставка i осигурявано от единица от храната j.

Ако положим x_j да бъде броят единици от храната j включени в диетата, то най-евтината диета се намира при решаване на линейната задача:

$$\begin{array}{ll} \text{да се минимизира} & \displaystyle\sum_{j=1}^{n} c_j x_j \\ \text{при ограничения} & \displaystyle\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i\,, \quad i = 1, 2, \ldots, m\,, \\ & x_j \geqslant 0\,, \quad j = 1, 2, \ldots, n\,. \end{array}$$

Да предположим, че след като сме решили задачата и сме получили оптималното решение, на пазара се е появила нова храна $k = n+1$, която първоначално не е била включена в диетата. Възниква въпросът да се включи ли тази нова храна в диетата или не. За да отговорим на този въпрос, нека количеството от хранителната съставка i, осигурявано от единица от новата храна бъде a_{ik} и нека единичната цена на храната бъде c_k. Тъй като оптималната стойност y_i^* на i-та двойнствена променлива може да се интерпретира като неявна цена на единица от i-та хранителна съставка, то хранителните вещества, осигурявани от единица от новата храна имат следната цена $\sum_{i=1}^{m} y_i^* a_{ik}$. Следователно, ако $c_k < \sum_{i=1}^{m} y_i^* a_{ik}$, то новата храна си заслужава да бъде закупена и включена в диетата (y^* не е допустимо за новото ограничение); в противен случай, текущата оптимална диета си остава оптимална (y^* остава допустимо). В първия случай, прилагането на симплекс метода за ре-оптимизация води до незабавно избиране на новата храна за влизане в базиса (включване в диетата). Тази операция в модифицирания симплекс метод съответства на пресмятане на редуцираната цена на новата променлива (храна).

За да видим, че не само двойнствените променливи, а и цялата двойнствена задача има икономическа интерпретация, да забележим, че задачата за диета е във формата на правата задача (P). Следователно, нейната двойнствена е във формата на (D). Можем да разгледаме двойнствената на задачата за диета като задача на конкурент на бакалина, от който този, който прави диетата, купува продуктите. Нека този конкурент е дрогерист, който продава хранителните съставки в чист вид – на хапчета, които съдържат само желязо, или само някакъв витамин. За да продаде такива хапчета на правещия диетата, дрогеристът трябва да даде на тези хапчета цени $y_1, \ldots, y_m$, които да са неорицателни и да са конкурентни на цените на бакалина, т.е. да удовлетворяват

$$\sum_{i=1}^{m} y_i a_{ij} \leqslant c_j, \quad j = 1, \ldots, n\,.$$

Тъй като минималната дневна потребност от веществото i е b_i, дрогеристът ще се опита при горните ограничения да максимизира $\sum_{i=1}^{m} b_i y_i$; т.е. да реши двойнствената задача (D).

8. Двойнствен симплекс метод

Разглеждаме каноничната задача на линейното оптимиране

$$
\begin{array}{lll} (P) & \min z(x) = c^{\mathrm{T}}x & \text{и нейната двойнствена} \quad (D) \ \max v(y) = b^{\mathrm{T}}y \\ & Ax = b & \qquad\qquad\qquad\qquad\qquad A^{\mathrm{T}}y \leqslant c, \\ & x \geqslant 0 & \end{array} \tag{8.1}
$$

където A означава $(m \times n)$ матрица, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$. Нека $\overline{x} = \begin{bmatrix} \overline{x}_B \\ \overline{x}_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$ е базисно решение за задачата (P). Ще казваме, че $\overline{x}$ е *псевдо оптимално*, ако редуцираните цени на небазисните му променливи са неотрицателни (т.е. $\overline{c}_N = c_N - c_B^{\mathrm{T}}B^{-1}N \geqslant 0$, т.е. чиито вектор от симплексни множители $y^{\mathrm{T}} = c_B^{\mathrm{T}}B^{-1}$ е допустим за задачата (D)). Да отбележим, че псевдо оптимално базисно решение $\overline{x}$ може да не е оптимално решение на (P), защото може да не е допустимо (т.е. може да не удовлетворява условието $\overline{x} \geqslant 0$) и, че ако псевдо оптимално решение $\overline{x}$ е и допустимо, то е оптимално решение на (P) (вж. теоремите за двойнственост).

Нека сега псевдо оптималното базисно решение $\overline{x}$ не е допустимо за (P), т.е. $\overline{x} \not\geqslant 0$. Такова решение може да се получи например, когато се добави $(m+1)$-во ограничение неравенство към допустимото множество на (P). Да положим $k = m+1$ и да добавим ограничението $\displaystyle\sum_{j=1}^{n} a_{kj}x_j \leqslant b_k$ след като задачата (P) е вече решена. Ако новото ограничение се удовлетворява от текущото оптимално решение $\overline{x}$, няма какво повече да правим, тъй като то си остава оптимално. Ако обаче новото ограничение не се удовлетворява от $\overline{x}$, т.е. ако $\displaystyle\sum_{j=1}^{n} a_{kj}\overline{x}_j > b_k$, то новото ограничение може да се превърне в равенство $\displaystyle\sum_{j=1}^{n} a_{kj}x_j + s_k = b_k$ с добавяне на неотрицателна допълнителна променлива $s_k \geqslant 0$ и да бъде добавено към ограниченията на (P). Ясно е, че оптималният базис за оригиналната задача (P) и новата допълнителна променлива образуват базис за новата задача. За този базис е изпълнен критерият за оптималност (небазисните променливи и редуцираните им цени не се променят), но базисното решение, което му съответства $(\overline{x}, \overline{s}_k)$ е недопустимо тъй като в него стойността на новата допълнителна променлива е отрицателна $(\overline{s}_k = b_k - \displaystyle\sum_{j=1}^{n} a_{kj}\overline{x}_j < 0)$.

Двойнственият симплекс метод разработен от Лемке през 1954 г. е създаден, за да решава точно такава ситуация. Също както симплекс метода (който вече ще наричаме *прав симплекс метод*) той решава задачата (P) като се движи от базисно решение към съседно на него базисно решение, но вместо да поддържа на всяка итерация допустимостта на текущото решение, той поддържа неговата псевдо оптималност и спира при намиране на допустимо решение.

За да развием двойнствения симплекс метод, да допуснем, че решаваме правата задача (P) и че текущият базис $\mathbf{B}$ на псевдо оптималното решение $\overline{x}$ се състои от първите m променливи. Следователно, $\overline{x}_B = B^{-1}b$, $\pi^{\mathrm{T}} = c_B^{\mathrm{T}}B^{-1}$, и $\overline{c}_N^{\mathrm{T}} = c_N^{\mathrm{T}} - \pi^{\mathrm{T}}N \geqslant 0$. Ако $\overline{x}_B \not\geqslant 0$, векторът $\overline{x} = \begin{bmatrix} \overline{x}_B \\ 0 \end{bmatrix}$ съответства на псевдо оптимален, но недопустим връх на многостенното множество от допустими решения на задачата (P).

Да предположим, че базисната координата $\overline{x}_p < 0$. Ясно е, че има смисъл да се премине към съседно базисно решение (допустимо или не), за което $\overline{x}_p = 0$, чрез замяната на базисната променлива x_p с небазисна променлива x_q. Изборът на x_q се управлява от изискването да поддържаме превдо оптималност. За да определим кои от $(n-m)$-те съседни върхове на текущия връх са псевдо оптимални ще използваме следната:

Лема 8.1 (модификационна формула на Шърман-Морисън-Удбъри).
(а) *Ако M е $n \times n$ неособена матрица и u и v са два вектора в $\mathbb{R}^n$, то $M + uv^{\mathrm{T}}$ е неособена матрица, тогава и само тогава, когато $w \equiv 1 + v^{\mathrm{T}} M^{-1} u \neq 0$.*
(б) *Нещо повече, в този случай, $(M + uv^{\mathrm{T}})^{-1} = M^{-1} - (1/w) M^{-1} uv^{\mathrm{T}} M^{-1}$.*

Доказателство. Тъй като $M + uv^{\mathrm{T}} = (I + uv^{\mathrm{T}} M^{-1}) M$, (а) следва от факта, че $I + uv^{\mathrm{T}} M^{-1}$ има $n-1$ собствени стойности, които са равни на единица и една, която е равна на $1 + v^{\mathrm{T}} M^{-1} u$. Обновяващата формула (б) лесно се проверява посредством умножение с $M + uv^{\mathrm{T}}$. ■

За неособената матрица M^{-1} дадена с (4.4), обратна на матрицата от нормалите на активните в $\overline{x}$ ограничения, дадена с (4.3) (вж. въпрос 4) имаме, че

$$c^{\mathrm{T}} M^{-1} = (c_B^{\mathrm{T}}, c_N^{\mathrm{T}}) \begin{bmatrix} B^{-1} & -B^{-1}N \\ 0 & I \end{bmatrix} = (c_B^{\mathrm{T}} B^{-1}, c_N^{\mathrm{T}} - c_B^{\mathrm{T}} B^{-1} N) = (\pi^{\mathrm{T}}, \overline{c}_N^{\mathrm{T}}).$$

Ако p-та базисна променлива (в случая x_p) бъде заменена с небазисната променлива x_q това е еквивалентно на замяната на q-ия ред на M (реда e_q^{T}) с реда e_p^{T}; т.е. M става $M' = M + e_q(e_p - e_q)^{\mathrm{T}}$. Като използваме Лема 8.1 и това, че $e_q^{\mathrm{T}} M^{-1} = e_q^{\mathrm{T}}$ получаваме

$$M'^{-1} = M^{-1} - \frac{M^{-1} e_q \left(e_p^{\mathrm{T}} M^{-1} - e_q^{\mathrm{T}}\right)}{e_p^{\mathrm{T}} M^{-1} e_q}.$$

Като умножим двете страни на израза с c^{T} получаваме следните формули за пресмятане на новите симплексни множители и новите небазисни редуцирани цени:

$$\pi' = \pi + \tau u_p\,,$$

$$\overline{c'}_j = \overline{c}_j - \tau w_{pj}\,, \quad j > m\,,\ j \neq q\,,$$

и

$$\overline{c}'_p = -\tau\,,$$

където сме положили

$$u_p^{\mathrm{T}} = e_p^{\mathrm{T}} B^{-1} \ \text{ и } \tau = \overline{c}_q / w_{pq}\,.$$

Да отбележим, че u_p^{T} е p-ия ред на B^{-1} и че, както и преди, $w_j = B^{-1} a_j$, a_j е j-ия стълб на матрицата A, а w_{ij} е i-та координата на вектора w_j.

От рекурентната връзка между редуцираните цени следва, че за да бъде новият вектор от небазисните редуцирани цени $\overline{c}'$ с неотрицателни координати, q трябва да се избере така, че

$$0 \leqslant -\tau = -\overline{c}_q / w_{pq} \leqslant -\overline{c}_j / w_{pj}\,, \qquad \text{за всяко } w_{pj} < 0,\ j > m\,.$$

Ако $w_{pj} \geqslant 0$ за всяко небазисно j, то $u_p^{\mathrm{T}} A$ е вектор с неотрицателни координати и уравнението $u_p^{\mathrm{T}} Ax = u_p^{\mathrm{T}} b$ не може да има неотрицателно решение x, поради $u_p^{\mathrm{T}} b = e_p^{\mathrm{T}} B^{-1} b = \overline{x}_p < 0$. Това означава, че линейната задача (P) е с празно допустимо множество.

Вече можем да дадем "модифицираната" версия на двойнствения симплекс метод.

8.1. Алгоритъм на двойнствения симплекс метод

(0) Нека е дадено превдо оптимално базисно решение $\overline{x}$ на каноничната задача (P), с базисна матрица $B = [a_{j_1}, \ldots, a_{j_m}]$. Нека $\boldsymbol{B} = \{j_1, \ldots, j_m\}$ означава индексното множество от базисните променливи. Пресмятаме симплексните множители за $\overline{x}$ като първо решим системата $B^{\mathrm{T}} \pi = c_B$, и след това ги използваме, за да пресметнем небазисните редуцирани цени $\overline{c}_j = c_j - \pi^{\mathrm{T}} a_j$, за всяко $j \notin \boldsymbol{B}$.

(1) Проверка за допустимост: Ако $\overline{x}_B \geqslant 0$, КРАЙ; текущото решение $\overline{x}$ е допустимо и, следователно, оптимално. В противен случай

(2) Определяне на базисна променлива x_{j_p}, която да напусне базиса: избор на

$$j_p \in \{j_i \in \boldsymbol{B} : \overline{x}_{j_i} < 0\} \,.$$

(3) Проверка за празно допустимо множество: Пресмятаме u_p като решим системата $B^{\mathrm{T}} u_p = e_p$ относно u_p и пресмятаме $w_{pj} = u_p^{\mathrm{T}} a_j$, за всяко $j \notin \boldsymbol{B}$. Ако $w_{pj} \geqslant 0$ за всяко $j \notin \boldsymbol{B}$, КРАЙ; задачата (P) е с празно допустимо множество.

(4) Определяне на небазисна променлива x_q за влизане в базиса: намиране на

$$-\tau = \min\{-\overline{c}_j / w_{pj} : w_{pj} < 0,\ j \notin \boldsymbol{B}\} = -\overline{c}_q / w_{pq} \,.$$

(5) Обновяване на редуцираните цени: полагане на

$$\overline{c}_j \leftarrow \overline{c}_j - \tau w_{pj} \,, \quad j \notin \boldsymbol{B} \,,\ j \neq q \,,$$

$$\overline{c}_{j_p} \leftarrow -\tau \,.$$

(6) Обновяване на решението и на базисната матрица B: Пресмятаме w_q като решим

$$B w_q = a_q \,.$$

Полагаме

$$x_q \leftarrow \theta = \overline{x}_{j_p} / w_{pq} \,,$$

$$x_{j_i} \leftarrow x_{j_i} - \theta w_{iq} \,, \quad \text{за } 1 \leqslant i \leqslant m,\ i \neq p$$

$$B \leftarrow B + (a_q - a_{j_p}) e_p^{\mathrm{T}} \,,$$

$$\boldsymbol{B} \leftarrow \boldsymbol{B} \cup \{q\} \setminus \{j_p\} \,,$$

$$j_p \leftarrow q \,,$$

и отиваме на стъпка (1).

На всяка итерация двойнствения симплекс метод изисква абсолютно същото количество работа, както подобно реализираната версия на правия симплекс метод.

Ако задачата (P) е неизродена се движим от недопустимо базисно псевдо оптимално решение към съседно на него псевдо оптимално базисно решение, което, ако е допустимо е оптимално, а ако не е допустимо повтаряме двойнствения симплекс метод като за краен брой итерации или ще получим оптимално решение на (P), или ще покажем, че допустимото й множество е празно. В случай на изроденост на (P) можем да прибегнем до алгоритъма на Бленд, избягващ зациклянето.

Решаването на правата задача (P) посредством двойнствения симплекс метод математически е еквивалентно на решаването на двойнствената й задача (D) посредством правия симплекс метод. Наистина, за задачата (P) да предположим, че $c \geqslant 0$, да въведем изцяло изкуствен базис от изкуствените променливи $u \in \mathbb{R}^m$

$$(P) \quad \begin{array}{l} \min c^{\mathrm{T}}x \\ Ax + u = b \\ x \geqslant 0, u \geqslant 0 \end{array} \quad \text{и да образуваме нейната двойнствена } (D) \quad \begin{array}{l} \max b^{\mathrm{T}}y \\ A^{\mathrm{T}}y \leqslant c \\ y \leqslant 0. \end{array} \tag{8.2}$$

Превръщаме последната в задача за минимум, полагаме $z := -y$ и въвеждаме допълнителни променливи $s \geqslant 0$, за да получим еквивалентната на нея

$$(D) \quad \begin{array}{l} \min b^{\mathrm{T}}z \\ -A^{\mathrm{T}}z + s = c \\ z \geqslant 0, s \geqslant 0. \end{array}$$

Имаме двойката спрегнати канонични задачи

$$(P) \quad \begin{array}{l} \min c^{\mathrm{T}}x \\ Ax + u = b \\ x \geqslant 0, u \geqslant 0 \end{array} \quad \text{и} \quad (D) \quad \begin{array}{l} \min b^{\mathrm{T}}z \\ -A^{\mathrm{T}}z + s = c \\ z \geqslant 0, s \geqslant 0. \end{array}$$

с базисно псевдо оптимално решение $\overline{x} = (0, b)$ с базисни променливи u за (P) и базисно допустимо решение $\overline{z} = (0, c)$ с базисни променливи s за (D) съответно. Симплексните таблици за съответните базисни решения са

$$T(\overline{x}) = \left[\begin{array}{c:c:c} b & A & I \\ \hdashline 0 & c & 0 \end{array}\right] \qquad T(\overline{z}) = \left[\begin{array}{c:c:c} c & -A^{\mathrm{T}} & I \\ \hdashline 0 & b & 0 \end{array}\right]$$

Да разгледаме преобразување чрез правия симплекс метод на симплексната таблица $T(\overline{z})$: от реда с редуцираните цени трябва да изберем отрицателна редуцирана цена, т.е. $b_p < 0$, а след това променливата, напускаща базиса трябва да изберем по теста за минимално отношение

$$\min\{\frac{c_j}{-a_{pj}} : -a_{pj} > 0\} = \min\{\frac{-c_j}{a_{pj}} : a_{pj} < 0\},$$

което съответства на избора на ключов елемент от двойнствения симплекс метод в симплексната таблица $T(\overline{x})$.

9. Анализ за чувствителност. Параметрично оптимиране.

Разглеждаме, както обикновено, каноничната задача

$$
\begin{array}{lll}
(P) & \text{да се минимизира} & z(x) = c^{\mathrm{T}}x \\
 & \text{при ограничения} & Ax = b\,, \\
 & & x \geqslant 0\,,
\end{array}
\tag{9.1}
$$

където A означава $(m \times n)$ матрица, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$.

Във въпрос 8 видяхме как можем да получим оптимално решение на (P) след добавяне на ново ограничение, използвайки оптималното решение на задачата преди добавянето на това ограничение и двойнствения симплекс метод. Във въпрос 7 обяснихме, че оптималните симплексни множители (т.е. оптималното решение на двойнствената задача) отразяват измененията в оптималната стойност на целевата функция при малки изменения в дясната страна на ограниченията в случая на неизроденост на оптималното базисно допустимо решение.

Сега ще изследваме как по-общи изменения във вектора дясна част b или във вектора на целевата функция c влияят на вече полученото оптимално решение. Такива изследвания се наричат още *анализ за чувствителност* или *следоптимален анализ.*

Нека $\overline{x} = \begin{bmatrix} \overline{x}_B \\ \overline{x}_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$ е базисно решение за задачата (P) с базисна матрица B. $\overline{x}$ е допустимо, ако $\overline{x}_B = B^{-1}b \geqslant 0$ и $\overline{x}$ е псевдо оптимално, ако $\overline{c}_N^{\mathrm{T}} = c_N^{\mathrm{T}} - c_B^{\mathrm{T}}B^{-1}N \geqslant 0$. Ясно е, че при непроменяща се базисна матрица B, изменения във вектора b биха повлияли само на допустимостта на базисното решение $\overline{x}$, а изменения във вектора c биха повлияли само на неговата псевдо оптималност.

9.1. Изменения в b.

Да разгледаме фамилията канонични задачи с параметър в дясната част:

$$
\begin{array}{ll}
(P_t) & \min z(t) = c^{\mathrm{T}}x \\
 & Ax = b + td \\
 & x \geqslant 0\,,
\end{array}
$$

където $d \in \mathbb{R}^m$ е фиксиран ненулев вектор, а $t \in \mathbb{R}$ е реален параметър.

Твърдение 9.1. *Множеството от стойности на параметъра t, за които задачата (P_t) има допустими решения е интервал $[\alpha, \beta]$ (с крайни или безкрайни граници).*

Доказателство. Множеството $M := \{(x,t) | Ax - td = b;\ x \geqslant 0\}$ е многостенно множество в $\mathbb{R}^{n+1}$ Следователно проекцията му върху оста на t е интервал. ■

Твърдение 9.2. *Ако за една стойност $t_0 \in [\alpha, \beta]$, задачата (P_{t_0}) има крайно оптимално решение, то за всяко $t \in [\alpha, \beta]$ задачата (P_t) има крайно оптимално решение.*

Доказателство. Според силната теорема за двойнственост, (D_{t_0}) - двойнствената на задачата (P_{t_0}) също има крайно оптимално решение. Множество от ограничения на двойнствената задача (D_t) е $A^{\mathrm{T}}y \leqslant c$ и очевидно не зависи от стойностите на t. Следователно, за всяко

$t \in [\alpha, \beta]$ двойнствената задача (D_t) има допустимо решение. Тъй като за всяко $t \in [\alpha, \beta]$ задачата (P_t) има допустимо решение, то целевата функция на (D_t) е ограничена отгоре върху допустимото й множество, следователно (D_t) има крайно оптимално решение. От силната теорема за двойнственост, за всяко $t \in [\alpha, \beta]$ задачата (P_t) има крайно оптимално решение. ■

Ако B е базисната матрица на оптимално базисно допустимо решение $\overline{x}$ за някаква стойност на $t = t_0$, то интервалът $I_B = [\underline{t}, \overline{t}]$ за който базисното решение $\overline{x} = \begin{bmatrix} \overline{x}_B \\ \overline{x}_N \end{bmatrix} = \begin{bmatrix} \overline{b}^{\mathrm{T}} + t\overline{d}^{\mathrm{T}} \\ 0 \end{bmatrix}$, където $\overline{b} = B^{-1}b$ и $\overline{d} = B^{-1}d$ остава допустимо (и следователно оптимално) се определя от

$$\underline{t} = \max\{\max_{i\in\mathbf{B}}\{-\overline{b}_i/\overline{d}_i : \overline{d}_i > 0\},\ -\infty\} \leqslant t \leqslant \min\{\min_{i\in\mathbf{B}}\{-\overline{b}_i/\overline{d}_i : \overline{d}_i < 0\},\ \infty\} = \overline{t}\,. \tag{9.2}$$

В интервала I_B координатите на оптималното решение зависят линейно от t, но базисната матрица B и оптималното решение на двойнствената задача $y^{\mathrm{T}} = c_B^{\mathrm{T}}B^{-1}$ остават същите.

Ако $\overline{x}$ е оптимално базисно допустимо решение на (P_t) за всяко t в интервала $[\underline{t}, \overline{t}]$, то или съществува съседно базисно допустимо решение, което е оптимално за стойности на t в някакъв интервал $[\underline{\underline{t}}, \underline{t}]$, такъв че $-\infty < \underline{\underline{t}}$, или (P_t) е с празно допустимо множество за всяко $t \in (-\infty, \underline{t})$. Новото базисно допустимо решение се получава чрез итерация по двойнствения симплекс метод, при която се извежда от базиса променливата x_i, за която $\underline{t} = -\overline{b}_i/\overline{d}_i$ в (9.2), а $\underline{\underline{t}}$ се определя като се използва новия базис. Ако в симплексната итерация се достигне до празно допустимо множество, то (P_t) е с празно допустимо множество за всяко $t < \underline{t}$. По аналогичен начин се работи за $t \geqslant \overline{t}$.

Твърдение 9.3. *Оптималната стойност $z^*(t)$ е изпъкнала, на части линейна функция на t в интервала $[\alpha, \beta]$.*

Доказателство. Нека I_B е интервала на изменение на t, за който B е базисната матрица на оптималното решение $\overline{x}$. Следователно,

$$z^*(t) = c^{\mathrm{T}}\overline{x} = c_B^{\mathrm{T}}(B^{-1}b + tB^{-1}d)$$

е линейна функция на t върху интервала I_B. Броят на интервалите I_B е краен, тъй като имаме краен брой базисни матрици.

За да покажем, че $z^*(t)$ е изпъкнала, да разгледаме две стойности t', $t'' \in [\alpha, \beta]$. Нека съответните им оптимални решения са x', x''. Да образуваме изпъкналата комбинация на двете решения $x := \lambda x' + (1-\lambda)x''$ $(0 \leqslant \lambda \leqslant 1)$.

Умножаваме $Ax' = b + t'd$ с λ, а $Ax'' = b + t''d$ умножаваме с $1-\lambda$ и събираме, при което получаваме че $Ax = b + td$ за $t := \lambda t' + (1-\lambda)t''$. Следователно x е допустимо решение за (P_t). Оттук $z^*(t) \leqslant c^{\mathrm{T}}x$, т.е. $z^*(\lambda t' + (1-\lambda)t'') \leqslant \lambda c^{\mathrm{T}}x' + (1-\lambda)c^{\mathrm{T}}x'' = \lambda z^*(t') + (1-\lambda)z^*(t'')$ и следователно z^* е изпъкнала функция на $t \in [\alpha, \beta]$. ■

9.2. Изменения в c.

Да разгледаме фамилия от канонични задачи с параметър в целевата функция:

$$\begin{array}{ll} (P_t) & \min z(t) = (c+td)^{\mathrm{T}}x \\ & Ax = b \\ & x \geqslant 0\,, \end{array} \tag{9.3}$$

където $d \in \mathbb{R}^n$ е фиксиран ненулев вектор, а $t \in \mathbb{R}$ е реален параметър.

Твърдение 9.4. *Множеството от стойности на параметъра t, за които задачата (P_t) има крайно оптимално решение е интервал $[\alpha, \beta]$ (с крайни или безкрайни граници).*

Доказателство. От силната теорема за двойнственост е ясно, че (P_t) има крайно оптимално решение, тогава и само тогава, когато допустимото множество на (D_t) е непразно и целевата й функция $b^{\mathrm{T}}y$ е ограничена отгоре върху него. Последното е винаги в сила, ако допустимото множество на (P_t) (което не зависи от t!) не е празното. Допустимото множество на двойнствената задача (D_t) е множеството от тези $y \in \mathbb{R}^m$, за които $A^{\mathrm{T}}y \leqslant c + td$. Множеството $\{(x, t) | A^{\mathrm{T}}y \leqslant c + td\}$ е многостенно множество в $\mathbb{R}^{m+1}$. Следователно, проекцията му върху оста на t е интервал. ■

Да допуснем, че при $t = t_0$, $\overline{x}$ е оптимално базисно допустимо решение на задачата (P_{t_0}). Искаме да определим интервал $\underline{t} \leqslant t \leqslant \overline{t}$ за параметъра t, такъв че текущата базисна матрица остава оптимална. Нека базисната матрица на $\overline{x}$ е B и нека c и d бъдат разбити на базисни и небазисни части c_B, d_B и c_N и d_N, съответно. Матрицата B ще бъде оптимална докато небазисните редуцирани цени остават неотрицателни, т.е. докато $(c_N + td_N)^{\mathrm{T}} - (c_B + td_B)^{\mathrm{T}} B^{-1} N \geqslant 0$. Като положим $\overline{c}_N^{\mathrm{T}} = c_N^{\mathrm{T}} - c_B^{\mathrm{T}} B^{-1} N$ и $\overline{d}_N^{\mathrm{T}} = d_N^{\mathrm{T}} - d_B^{\mathrm{T}} B^{-1} N$, това означава $t\overline{d}_N^{\mathrm{T}} \geqslant -\overline{c}_N^{\mathrm{T}}$. От тази система неравенства за t получаваме

$$\underline{t} = \max\{\max\{-\overline{c}_j/\overline{d}_j : \overline{d}_j > 0,\ j \notin \boldsymbol{B}\},\ -\infty\} \leqslant t \leqslant \min\{\min\{-\overline{c}_j/\overline{d}_j : \overline{d}_j < 0,\ j \notin \boldsymbol{B}\},\ \infty\} = \overline{t}\,. \tag{9.4}$$

Ако $t_0 = 0$ и изберем $d = e_j$, то $[c_j + \underline{t}, c_j + \overline{t}]$ дава *интервала на изменение* на j-ия коефициент на целевата функция, за който оптималното решение, получено при $t = 0$ остава оптимално, при положение, че всички останали данни на задачата не се изменят.

Ако $\overline{x}$ е оптимално базисно допустимо решение на (P_t) за всяко t в интервала $[\underline{t}, \overline{t}]$, то или съществува съседно базисно допустимо решение, което е оптимално за стойности на t в някакъв интервал $[\underline{\underline{t}}, \underline{t}]$, такъв че $-\infty < \underline{\underline{t}}$, или $z^*(t)$ е неограничена отдолу за всяко $t \in (-\infty, \underline{t})$. Новото базисно допустимо решение се получава чрез итерация по правия симплекс метод, при която се въвежда в базиса променливата x_j, за която $\underline{t} = -\overline{c}_j/\overline{d}_j$ в (9.4), а $\underline{\underline{t}}$ се определя като се използва новия базис. Ако в симплексната итерация се достигне до неограничен ръб, то $z^*(t)$ е неограничена отдолу за всяко $t < \underline{t}$. По аналогичен начин се работи за $t \geqslant \overline{t}$.

Твърдение 9.5. *Оптималната стойност $z^*(t)$ е вдлъбната, на части линейна функция на t в интервала $[\alpha, \beta]$.*

Доказателство. Нека I_B е интервала на изменение на t, за който B е базисната матрица на оптималното решение $\overline{x}$. За $t \in I_B := [\underline{t}, \overline{t}]$

$$z^*(t) = (c^{\mathrm{T}} + td^{\mathrm{T}})\overline{x} = c_B^{\mathrm{T}} B^{-1} b + td_B^{\mathrm{T}} B^{-1} b),$$

т.е. оптималната стойност на целевата функция е линейна функция на t върху интервала I_B.

За да покажем, че $z^*(t)$ е вдлъбната, да разгледаме две стойности t', $t'' \in [\alpha, \beta]$. Да образуваме изпъкналата им комбинация $t = \lambda t' + (1 - \lambda)t''$ $(0 \leqslant \lambda \leqslant 1)$. Нека x е оптимално решение на (P_t). Очевидно x е допустимо решение за (P_t) за всяко $t \in [\alpha, \beta]$. Оттук, $z^*(\lambda t' + (1-\lambda)t'') = (c^{\mathrm{T}} + \lambda t' d^{\mathrm{T}} + (1 - \lambda)t'' d^{\mathrm{T}})x = \lambda(c^{\mathrm{T}} + t' d^{\mathrm{T}})x + (1 - \lambda)(c^{\mathrm{T}} + t'' d^{\mathrm{T}})x \geqslant \lambda z^*(t') + (1 - \lambda)z^*(t'')$ и z^* е вдлъбната функция на $t \in [\alpha, \beta]$. ■

10. Оптимиране в мрежи. Обхващащи дървета

Много линейни задачи могат да се разглеждат като задачи за минимизиране на разходите за придвижване на ресурси, разположени в някои от възлите на мрежа, така че да се задоволят потребностите от тези ресурси в други възли на мрежата. Такива задачи се наричат *задачи за потоци в мрежи.* Те образуват най-важния частен случай на линейните задачи. Транспортните, електрическите и комуникационните мрежи са очевидни примери за сферата на тяхното приложение. По-малко очевидни, но също толкова важни са приложенията в управлението на ресурсите, финансовото планиране и др.

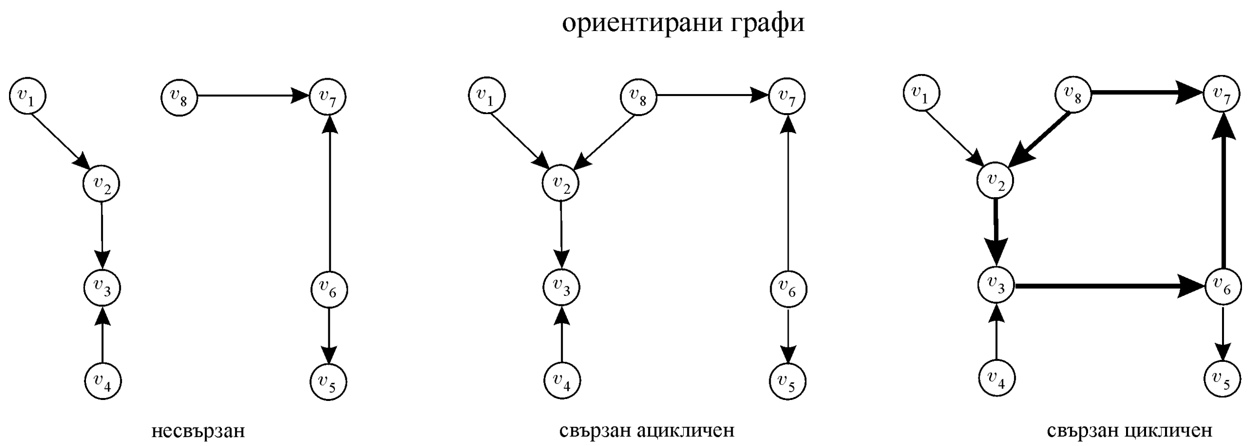
10.1. Графи.

Ще припомним някои основни дефиниции и резултати от теория на графите, тъй като мрежите могат да се разглеждат като графи.

Граф G е двойка $G = (V, E)$, където V е множество с краен брой елементи, наречени *върхове*, а E е множество с краен брой елементи, които са двойки различни върхове, наречени *ребра.* Ако ребрата на E са **наредени** двойки, то те се наричат *дъги*, а G се нарича *ориентиран граф.* Ако $e = (i, j)$ е дъга от E, казваме, че i е *опашка* на e, а j е *глава* на e и пишем $i = t(e)$, $j = h(e)$.

Път в ориентирания граф G е редица от върхове на G от вида $P = (i_0, i_1, \ldots, i_l)$, такава че между всеки два съседни върха има дъга, която може да е от вида $e_k = (i_{k-1}, i_k)$ (*права* дъга), или от вида $e_k = (i_k, i_{k-1})$ (*обратна* дъга) за $1 \leqslant k \leqslant l$. Пътищата могат да се разглеждат като редица от еднопосочни улици – движението на автомобили е разрешено само в една от посоките, докато пешеходците могат да се движат свободно и в двете посоки. Пътят P дефиниран по-горе е *от* i_0 *до* i_l и е с *дължина* l. Път P, който удовлетворява и условието $i_0 = i_l$, се нарича *цикъл.*

Граф G се нарича *свързан* ако всеки два негови върха могат да се свържат с път, и се нарича *ацикличен*, ако не съдържа цикъл. Граф $H = (W, F)$ се нарича *подграф* на G, ако $W \subseteq V$ и $F \subseteq E$ и се нарича *обхващащ* подграф, ако $W \equiv V$.



Фигура 10.4.

10.2. Мрежа и мрежова задача.

Оттук нататък ще работим само със ориентиран свързан граф, който ще означаваме с $G = (V, E)$. Да предположим, че върховете му са възли, в които имаме недостиг или излишък на някакъв ресурс, а дъгите му са транспортни маршрути, по които ресурса може да се транспортира. Нека $\hat{b} = (b_i)_{i \in V}$ е вектор, чиито координати удовлетворяват *условието за баланс*

$$\sum_{i \in V} b_i = 0, \tag{10.1}$$

като във възлите с $b_i < 0$ считаме че има недостиг на количество $|b_i|$, а във възлите с $b_i > 0$ имаме излишък на количество b_i от ресурса.

Целта е да се преразпредели ресурса от възлите с излишък към възлите с недостиг като придвижването му се извършва по дъгите $e \in E$ съгласно техните посоки. Променливите на задачата са количествата ресурс, които се движат по всяка дъга. Тоест за всяка дъга $e = (i, j) \in E$, с $x_e \equiv x_{ij}$ ще означаваме количеството ресурс което преминава от възел i до възел j директно по дъгата $e = (i, j)$. За да бъдем улеснени във вземането на решение по кои дъги да придвижим ресурса, нека ни е даден вектор с цените за превоз на единица от ресурса по всеки от маршрутите $c = (c_e)_{e \in E}$.

Целта ни е да минимизираме съвкупните транспортни разходи, които правим при преразпределянето на ресурса за задоволяване на потребностите във върховете.

Както вече казахме, ограниченията върху променливите идват от това че трябва да се осигурим потребностите във всеки връх(т.е. да покрием недостига или да се освободим от излишъка там). Нека разгледаме фиксиран връх, да кажем $k \in V$. Сумарно, количеството ресурс, влизащо във върха k е

$$\sum_{(i,k) \in E} x_{ik} = \sum_{e \in E : h(e) = k} x_e,$$

а сумарно количеството ресурс излизащо от върха k е

$$\sum_{(k,j) \in E} x_{kj} = \sum_{e \in E : t(e) = k} x_e.$$

Разликата между тези две количества ще бъде това количество, което остава във върха k и което би трябвало да е равно на недостигащото или на излишното количество в k, т.е.

$$\sum_{(k,j) \in E} x_{kj} - \sum_{(i,k) \in E} x_{ik} = b_k.$$

Накрая, количеството ресурс, което придвижваме по всяка дъга трябва да е неотрицателно, т.е. $x_e \geqslant 0,\ \forall e \in E$ (в противен случай би било възможно ресурсът да се движи в погрешната посока).

Така достигаме до следната линейна минимизационна задача

$$(\hat{T}) \qquad \begin{array}{ll} \text{да се минимизира} & \displaystyle\sum_{e \in E} c_e x_e \\ \text{при ограничения} & \displaystyle\sum_{e : t(e) = i} x_e - \sum_{e : h(e) = i} x_e = b_i\,, \quad i \in V, \\ & x_e \geqslant 0\,, \quad e \in E\,. \end{array}$$

Ориентиран свързан граф $G = (V, E)$, с чиито върхове са асоциирани числа b_i, $i \in V$ удовлетворяващи условието за баланс (10.1) се нарича *мрежа*, а количествата $x_e, e \in E$, асоциирани с дъгите му се наричат *поток*. С други думи, горната задача може да се разглежда като задача в мрежа за намиране на поток с минимална цена, който да задоволява определени потребности във възлите на мрежата и поради това се нарича *мрежова задача*.

Да приведем задачата $(\hat{T})$ в матрична форма.

За целта, да означим с $\hat{A}$ *матрицата на инцидентност* между върховете и дъгите на G, т.е. $\hat{A}$ има ред за всеки връх и стълб за всяка дъга, като

$$\hat{a}_{ie} = \begin{cases} +1 & \text{ако } t(e) = i\,, \\ -1 & \text{ако } h(e) = i\,, \\ 0 & \text{в противен случай.} \end{cases}$$

Можем да запишем $(\hat{T})$ като следната канонична задача

$$\begin{array}{lll} (\hat{T}) & \text{да се минимизира} & c^{\mathrm{T}} x \\ & \text{при ограничения} & \hat{A}x = \hat{b}\,, \\ & & x \geqslant 0\,. \end{array}$$

Да забележим, че в задачата $(\hat{T})$ матрицата на ограниченията $\hat{A}$ не е с линейно независими редове (с пълен ранг по редове). Наистина, всеки стълб на $\hat{A}$ има една $+1$ и една -1. Следователно, сумата на редовете на $\hat{A}$ дава нулевия вектор и те са линейно зависими, т.е. ако означим с $\overline{a}_i$ i-ия ред на $\hat{A}$ за $i \in V$, то $\sum\limits_{i \in V} \overline{a}_i = 0$.

Сега можем да обясним и защо наложихме условието за баланс (10.1) върху координатите на $\hat{b}$. Причината е, че това условие е необходимо условие за това задачата $(\hat{T})$ да има допустимо решение, тъй като ако $\hat{x}$ е допустимо решение за $(\hat{T})$, то за всяко $i \in V$ имаме $\overline{a}_i \hat{x} = b_i$. Сумираме и получаваме $0 = \left(\sum\limits_{i \in V} \overline{a}_i \right) \hat{x} = \sum\limits_{i \in V} b_i$.

Това, че редовете на $\hat{A}$ са линейно зависими, означава, че измежду тях има излишни и съответните им ограничения те могат да бъдат премахнати от задачата. Както ще видим, достатъчно е да премахнем само един, произволно избран ред.

Нека фиксираме произволен връх r във V, който ще наричаме *корен*. Нека A и b бъдат $\hat{A}$ и $\hat{b}$, от които сме изтрили реда, съответстващ на върха r. Тогава $(\hat{T})$ е еквивалентна на задачата

$$\begin{array}{lll} (T) & \text{да се минимизира} & c^{\mathrm{T}} x \\ & \text{при ограничения} & Ax = b\,, \\ & & x \geqslant 0 \end{array}$$

в смисъл, че допустимите им множества съвпадат. Наистина, ако $\hat{x}$ е допустимо за $(\hat{T})$, то очевидно е допустимо и за (T), а ако x е допустимо за (T) от $\sum\limits_{i \in V} \overline{a}_i = 0$ имаме, че $\overline{a}_r = -\sum\limits_{i \in V, i \neq r} \overline{a}_i$. Оттук, $\overline{a}_r x = -\sum\limits_{i \in V, i \neq r} \overline{a}_i x = -\sum\limits_{i \in V, i \neq r} b_i = b_r$ като второто равенство идва от това, че x е допустимо за (T), а последното от условието за баланс (10.1).

10.3. Обхващащи дървета и базисни решения.

Трябва ни следния резултат от теория на графите, който ще оставим без доказателство.

Лема 10.1. *Нека $H = (V, F)$ е обхващащ подграф на свързания ориентиран граф $G = (V, E)$ и $|V| = m$. Тогава следните са еквивалентни:*
(i) $|F| = m - 1$ *и* H *е свързан;*
(ii) $|F| = m - 1$ *и* H *е ацикличен;*
(iii) H *е свързан и ацикличен;*
(iv) H *е минимално свързан – отстраняването на произволна дъга го прави несвързан;*
(v) H *е максимално ацикличен – добавянето на произволна дъга създава цикъл.*

Ако за H е в сила някое от тези еквивалентни условия, казваме че H е *обхващащо дърво* за G.

Ще използваме горния резултат, за да покажем, че рангът на A е $m - 1$, където $m = |V|$. В същото време ще дадем характеризация на неособените подматрици на A от ред $m - 1$ (базисните матрици).

Теорема 10.1. *Нека $G = (V, E)$ е ориентиран свързан граф и $|V| = m$. Нека $\hat{A}$ е неговата матрица на инцидентност, $r \in V$ е произволен връх и нека A е получена от $\hat{A}$ след отстраняването на реда с индекс r.*

Тогава A има пълен ранг по редове $m-1$, и ако B е квадратна $(m-1)\times(m-1)$ подматрица на A, то B е неособена, тогава и само тогава, когато нейните стълбове съответстват на дъги на обхващащо дърво за G.

Доказателство. $\Longleftarrow$) Според Лема 10.1(iv) всеки свързан граф притежава обхващащо дърво, което се получава чрез премахване на дъги от G, докато полученият подграф стане минимално свързан, а според Лема 10.1(ii) той е с $m - 1$ върха. Следователно съществува поне една $(m-1) \times (m-1)$ подматрица на A, стълбовете на която съответстват на обхващащо дърво за G. Да означим коя да е такава подматрица с B. Това, че B е неособена директно следва от

Лема 10.2. *Нека $H = (V, F)$ е обхващащо дърво за G и нека B е съответната му подматрица на A. Тогава редовете и стълбовете на B могат да се пренаредят така, че да образуват горна триъгълна матрица с ненулеви елементи по диагонала. В частност, B е неособена.*

Доказателство. С индукция по $|V| = m$. За $m = 2$, B е 1×1 матрица, т.е. число, което е ± 1. Нека твърдението е вярно за $m < k$ и да разгледаме случая $m = k$.

За всеки връх $i \in V$ дефинираме *степен* на i относно H, която е равна на броя на влизащите и излизащите от i дъги от F, т.е. на броя на $e \in F$ такива, че $t(e) = i$ или $h(e) = i$. Сумата от степените на върховете в H е $2|F| = 2(m-1)$, тъй като началото и края на всяка дъга се броят към някой от върховете. Всеки връх има степен поне 1, тъй като H е свързан (т.е. до всеки връх достига поне една дъга). Оттук следва, че има поне два върха със степен 1, а всички такива върхове ще наричаме *листа.*

Да вземем листо $i \in V$, което е различно от r (такова винаги съществува, защото листата са поне две) и нека $e \in F$ е дъгата (единствената), която го свързва в H. Да разгледаме графа H', който се получава, като се махнат върха i и дъгата e:

$$H' := (V \setminus \{i\}, F \setminus \{e\}).$$

H' има $m-1$ върха и $m-2$ дъги, ацикличен е, тъй като се получава чрез отстраняване на елементи от ацикличния граф H. От Лема 10.1(ii) той е обхващащо дърво за графа

$$G' = (V \setminus \{i\}, E \setminus \{e\}).$$

Според индукционната хипотеза, можем да наредим върховете и дъгите на H' така, че съответната им матрица B' да бъде горно триъгълна с ненулеви елементи по диагонала. Да добавим върха i като последен ред и дъгата e като последен стълб, за да получим

$$B = \begin{pmatrix} B' & u \\ 0 & \pm 1 \end{pmatrix} \tag{10.2}$$

за някакъв вектор u, чиито координати са в множеството $\{0, 1, -1\}$ и B е пренаредена в желаната форма. С това приключва индукцията и доказателството на лемата. ■

$\Longrightarrow$) Нека сега B е неособена подматрица на A от ред $m-1$ и да допуснем, че дъгите, съответстващи на стълбовете ѝ образуват цикъл в G. Лесно се вижда, че стълбовете на A, които съответстват на дъги, образуващи цикъл в G, са линейно зависими. Наистина, ако означим с $P(Q)$ правите (обратните) дъги на цикъла, имаме че

$$\sum_{e \in P} a_e - \sum_{e \in Q} a_e = 0\,.$$

Следователно, направеното допускане води до противоречие. ■

Показахме, че в задачата (T) матрицата A е с пълен ранг по редове. Задачата е канонична и има базисно решение (съответстващо на минимално обхващащо дърво). Намирането на базисните координати $\overline{x}_B$ на базисно решение $\overline{x}$ с базисна матрица B става чрез решаване на системата $B\overline{x}_B = b$, а небазисните му координати, разбира се, са нули. Тъй като базисната матрица B може да се пренареди във във вида (10.2), то решаването на $B\overline{x}_B = b$ относно $\overline{x}_B$ става без да е необходимо да извършваме деления (т.к. диагоналните елементи на B са ± 1) и без да е необходимо да извършваме умножения (т.к. извън диагоналните ненулеви елементи на B са ± 1). С други думи, системата $B\overline{x}_B = b$ се решава чрез проста редица от събирания и изваждания. Следователно, ако векторът b е целочислен, базисното решение $\overline{x}$ също ще бъде целочислено.

Следствие 10.1. *Ако b_i, $i \in V$ са цели числа, то всяко базисно решение x има целочислени координати, като $x_e \neq 0$ само за $e \in F$ за множеството от дъги F на някое обхващащо дърво на G.*

11. Мрежов симплекс метод

Ще разгледаме реализацията на правия симплекс метод за решаване на мрежовата задача

$$(\hat{T}) \qquad \begin{array}{ll} \text{да се минимизира} & c^{\mathrm{T}}x \\ \text{при ограничения} & \hat{A}x = \hat{b} \\ & x \geqslant 0\,, \end{array}$$

където $\hat{A}$ е матрицата на инцидентност на свързан ориентиран граф $G = (V, E)$, векторът $\hat{b} = (b_i)_{i \in V}$ е такъв че удовлетворява условието за баланс $\sum_{i \in V} b_i = 0$, $c = (c_e)_{e \in E}$, $x = (x_e)_{e \in E}$.

Адаптирането на правия симплекс метод за мрежови задачи се нарича мрежов симплекс метод. Той има по-различно описание и някои важни специфични свойства.

На намирането на начално базисно допустимо решение и на това какво се случва при изроденост на базисното допустимо решение, тук няма да се спираме. За решаване и на двата проблема съществуват специфични мрежови техники. И така, предполагаме, че имаме начално неизродено базисно допустимо решение $\overline{x}$ с базисна матрица B.

Основната идея е да се представи базисната матрица B и съответстващото ѝ обхващащо дърво $H = (V, F)$ така, че ефикасно да се намери решение на линейните системи $B^{\mathrm{T}}\pi = c_B$ и $Bw = a_e$. Тъй като B може да бъде пренаредена така че да стане триъгълна, тези системи могат бързо и лесно да бъдат решени.

Нека дъгата $e = (u, v) \in E \setminus F$, т.е. x_e е небазисна променлива. Тъй като H е свързан, от u до v съществува път $P_{u,v} = P_e$ в H от дъги в F. Добавяме към този път дъгата e и получаваме цикъл $P_e \cup \{e\}$. От доказателството на посоката $\Longrightarrow$ на Теорема 10.1 е ясно, че можем да използваме този цикъл, за да намерим решението w на системата $Bw = a_e$. Ако означим с $P(Q)$ множеството от правите (обратните) дъги в цикъла, намираме, че $w_f = +1$, ако $f \in P$; $w_f = -1$ ако $f \in Q$; и $w_f = 0$ в противен случай.

За бързото и лесно намиране на пътя P_e за всяко $i \in V$, различно от корена r пазим неговия *предшественик* $p(i)$ – върха следващ i в единствения път $P_{i,r}$ от i до r в H (такъв път винаги съществува, т.к. H е свързан, а ако допуснем, че съществува и друг ще получим цикъл в H, което е в противоречие с това, че H е ацикличен), пазим и $d(i)$ – дължината на пътя $P_{i,r}$. Ако дъгата свързваща i и $p(i)$ в този път е права дъга, пишем $+p(i)$, а ако е обратна дъга пишем $-p(i)$.

Системата $B^{\mathrm{T}}\pi = c_B$ може да се запише още като

$$\pi_i - \pi_j = c_f \quad \text{за } f = (i, j) \in F,$$

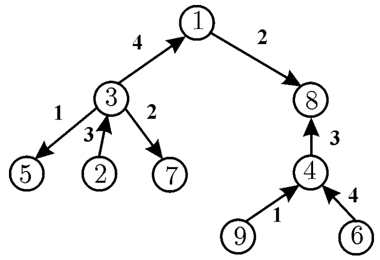
където $\pi_r \equiv 0$. Тази система решаваме относно π-та последователно като наредим върховете на обхващащото дърво H, започвайки от корена r така, че $p(i)$ да се среща винаги преди i. Такава наредба се нарича *преднаредба* на дървото. Съществуват множество преднаредби. Да фиксираме една от тях и да означим с $s(i)$ върха следващ i в нея.

За описание на дървото H използваме векторите d, p, s.

За илюстрация на казаното, да разгледаме следния

Пример 11.1. Нека $V = \{1, 2, \ldots, 9\}$ с корен $r = 1$, като H е обхващащото дърво, показано на Фигура 11.5 като $\overline{x}_{35} = \overline{x}_{94} = 1$, $\overline{x}_{37} = \overline{x}_{18} = 2$, $\overline{x}_{23} = \overline{x}_{48} = 3$, и $\overline{x}_{31} = \overline{x}_{64} = 4$. Дървото H е представено посредством трите вектора p, d и s с дължина $|V|$ дадени в Таблица 11.1.

Стъпките на итерация по правия симплекс метод за решаване на мрежовата задача са следните:



Фигура 11.5.

Таблица 11.1

i	1	2	3	4	5	6	7	8	9
$p(i)$	–	$+3$	$+1$	$+8$	-3	$+4$	-3	-1	$+4$
$d(i)$	0	2	1	2	2	3	2	1	3
$s(i)$	3	7	5	9	2	–	8	4	6

(1) Решаваме системата $B^{\mathrm{T}}\pi = c_B$, или $\pi_i - \pi_j = c_f$ за $f = (i,j) \in F$, където $\pi_r \equiv 0$. За целта, пресмятаме π-та в такава преднаредба, че $\pi_{|p(i)|}$ е вече известно, когато пресмятаме π_i. (По-нататък ще видим, че е по-изгодно да обновяваме вектора π, а не да го преизчисляваме.)

В нашия пример, пресмятаме π_3, π_5, π_2, π_7, π_8, π_4, π_9, и тогава π_6.

(2) Проверка на критерия за оптималност: $\overline{c}_e = c_e - \pi_i + \pi_j \geqslant 0$ за всяка небазисна дъга $e = (i,j) \in E \setminus F$. Ако всички неравенства са удовлетворени, то текущото базисно допустимо решение е оптимално; КРАЙ.

(3) В противен случай, избираме небазисна дъга $e = (u,v) \in E \setminus F$, за която имаме $\overline{c}_e = c_e - \pi_u + \pi_v < 0$.

(4) Проверка за неограниченост на целевата функция: решаваме системата $Bw = a_e$ относно w. За целта трябва да намерим път $P_{u,v} = P_e$ от u до v в H, като използваме векторите p и d. Ако в пътя няма прави дъги (координатите на w са неположителни), то тогава имаме неограничн ръб в допустимото множество на задачата, по който целевата функция намалява неограничено; КРАЙ. В противен случай отиваме на стъпка (5).

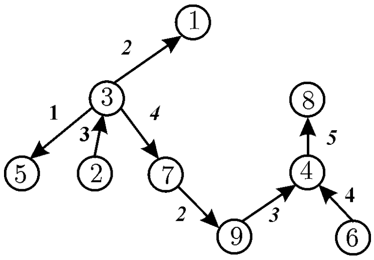
В нашия пример, да предположим, че $e = (u,v) = (7,9)$. Тъй като $d(9) > d(7)$, намираме $p(9) = +4$; следователно $(9,4)$ е обратна дъга в пътя. Сега $d(7) = d(4)$, но $7 \neq 4$. Така че намираме $p(7) = -3$, $p(4) = +8$, откъдето $(3,7)$ и $(4,8)$ са обратни дъги в пътя. Тъй като $3 \neq 8$, намираме $p(3) = +1$, $p(8) = -1$, така че $(3,1)$ и $(1,8)$ са прави дъги в пътя и тъй като двата под-пътя от u и от v вече се срещнаха, значи сме намерили целия път P_e от u до v.

(5) Определяне на напускаща дървото дъга f: това е правата дъга в пътя $(w_f = +1)$ с най-малък поток, която намираме като при намирането на пътя P_e на стъпка (4) сравняваме потока по всяка новопоявила се в него права дъга с текущия минимален поток по правите му дъги.

За нашия пример $f = (1,8)$.

(6) Обновяване на обхващащото дърво (на базиса)

$$F \leftarrow (F \cup \{e\}) \setminus \{f\}\,.$$



Фигура 11.6.

Таблица 11.2

i	1	2	3	4	5	6	7	8	9
$p(i)$	–	$+3$	$+1$	-9	-3	$+4$	-3	-4	-7
$d(i)$	0	2	1	4	2	5	2	5	3
$s(i)$	3	7	5	6	2	8	9	–	4

Обновяване на текущото решение $\overline{x}$:

$$\overline{x}_e \leftarrow \overline{x}_f\,,$$

$$\overline{x}_g \leftarrow \begin{cases} \overline{x}_g - \overline{x}_e & g \text{ е права дъга в пътя } P_e, \\ \overline{x}_g + \overline{x}_e & g \text{ е обратна дъга в пътя } P_e, \\ \overline{x}_g & \text{в противен случай}\,, \end{cases}$$

За нашия пример, $\overline{x}_{79} \leftarrow 2$, $\overline{x}_{37} \leftarrow 4$, $\overline{x}_{31} \leftarrow 2$, $\overline{x}_{18} \leftarrow 0$, $\overline{x}_{48} \leftarrow 5$ и $\overline{x}_{94} \leftarrow 3$, а останалите не се променят.

Обновяване на π: нека $f = (i, j)$; имаме две възможности

а) $d(i) < d(j)$, в който случай:

$$\pi_j \leftarrow \pi_j - \overline{c}_e, \quad k \leftarrow s(j)$$

и докато $d(k) > d(j)$,

$$\pi_k \leftarrow \pi_k - \overline{c}_e\,, \quad k \leftarrow s(k);$$

б) $d(j) < d(i)$, в който случай:

$$\pi_i \leftarrow \pi_i + \overline{c}_e, \quad k \leftarrow s(i)$$

и докато $d(k) > d(i)$,

$$\pi_k \leftarrow \pi_k + \overline{c}_e\,, \quad k \leftarrow s(k);$$

В нашия пример, вадим $\overline{c}_e$ от π_8, а след това от π_4, π_9 и π_6.

Накрая, трябва да обновим представянето на дървото, т.е. трябва да обновим векторите p, d и s. За това също съществуват правила на прехода, но те са сравнително сложни и няма да ги излагаме в детайли. Основното е, че ако $d(i) < d(j)$, под-дървото под j вече виси под дъгата $e = (u, v)$.

За нашия пример, резултатът е дървото, показано на Фигура 11.6, чието представяне е дадено в Таблица 11.2.

12. Генериране на стълб. Задача за едномерен разкрой

Ще разгледаме подход за прилагане на симплекс метода при решаването на линейни задачи с голяма размерност. Нека е дадена канонична линейна задача

$$(P) \qquad \min \quad \begin{aligned} & z = c^{\mathrm{T}} x \\ & Ax = b\,, \\ & x \geqslant 0\,, \end{aligned} \tag{12.1}$$

където A е $m \times n$ матрица, такава че стълбовете на A са известни, но само неявно (те могат да са с астрономически голям брой). Въпреки това е известен *вида* на стълбовете и те могат да се генерират, когато това стане необходимо в течение на симплекс алгоритъма – оттук идва и "генериране на стълб" като наименование на този подход.

За да го илюстрираме, ще работим с класически пример на такава задача – задачата за едномерен разкрой.

Да предположим, че фирма за производство на хартия разполага с широки ролки хартия с ширина W. Клиентите ѝ обаче търсят хартия с по-малка ширина. Да предположим, че за да задоволи това търсене, фирмата трябва да произведе b_i ролки с ширина w_i, $i = 1, 2, \dots, m$. По-тесните ролки се получават чрез разрязване на широка ролка по различни *начини*: например, широка ролка с ширина $W = 70$ см може да бъде разрязана на три ролки с ширина $w_1 = 17$ см и на една ролка с ширина $w_2 = 15$ см, като се бракува ширина от 4 см. Би трябвало да намерим всички такива начини на разкрояване и да образуваме матрица $A = \{a_{ij}\}$, в която a_{ij} означава колко ролки с ширина w_i се получават при j-ия начин на разрязване, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. Начинът на разрязване от нашия пример се описва със стълба $(3, 1, 0, \dots, 0)^{\mathrm{T}}$ на A. Ако в (P) положим c_j да бъдат равни на 1 за всяко j, то (P) е задачата да се определи минималният брой широки ролки ($z = \sum_j x_j$), които трябва да се нарежат, за да се задоволи търсенето на b_i тесни ролки с ширина w_i за всяко i. В действителност, бихме искали да решим съответната *целочислена задача*, където x_j – броят големи ролки, нарязани по j-ия начин, е естествено число. Въпреки това, решаването на линейната нецелочислена задача (P) след закръгляване често дава достатъчно точно решение на целочислената, или поне такъв е случаят, когато количествата b_i са разумно големи.

Решаването на (P) само по себе си обаче е значима изчислителна задача: дори когато m е сравнително малко число, броят на възможните начини на разкрояване n може да бъде огромен, така че да се образува изцяло матрицата от коефициенти A е непрактично. Въпреки това, както показват Гилмор и Гомори през 1961–1963 г., задачата (P) може ефективно да бъде решена със симплекс алгоритъма (вж. въпрос 4), като стълбовете на A се генерират в течение на алгоритъма тогава, когато това е необходимо, а не предварително.

Намирането на начално базисно допустимо решение е лесно. Наистина, нека начина i се състои в разрязване на широка ролка само на тесни ролки с ширина w_i. Тогава първите m стълба на A ще са от вида

$$a_i = \begin{bmatrix} 0 \\ \vdots \\ [W/w_i] \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{-та координата,}$$

(като $[\lambda]$ означава най-голямото естествено число, което не надминава λ) и очевидно образуват диагонална базисна подматрица на A.

Да предположим, че на дадена итерация разполагаме с базисно допустимо решение и искаме да продължим да прилагаме симплекс метода.

Пресмятаме симплексните множители π като решаваме системата $B^{\mathrm{T}}\pi = e$ (с e означаваме вектор със съответна размерност, всичките координати на който са равни на единица. Да напомним, че всичките c_j са единици, следователно $c_B = e$).

Следващата стъпка е да се пресметнат редуцираните цени

$$\bar{c}_j = 1 - \pi^{\mathrm{T}} a_j \tag{12.2}$$

за всяко j, за да се определи дали текущото базисно допустимо решение е оптимално, или да се избере за влизане в базиса небазисна променлива x_q, за която $\bar{c}_q < 0$. Това изглежда трудна задача, тъй като не знаем всички стълбове a_j на A. Въпреки това, благодарение на структурата на задачата, можем да извършим тази стъпка неявно.

Вектор $a = (\alpha_1, \alpha_2, \ldots, \alpha_m)^{\mathrm{T}} \in Z_+^m$ (всяко от α_i е естествено число) ще бъде стълб на A ако съответства на допустим начин на разкрояване, т.е. ако $w^{\mathrm{T}}a \leqslant W$, където $w = (w_1, w_2, \ldots, w_m)^{\mathrm{T}}$. Бихме искали да знаем дали редуцираната цена $1 - \pi^{\mathrm{T}}a$ е неотрицателна за всяко такова a, а ако това не е така, да намерим допустим вектор a, за който $1 - \pi^{\mathrm{T}}a$ е отрицателно. Следователно, решаваме спомагателната задача

$$\begin{aligned} (SP) \qquad \max \quad & \pi^{\mathrm{T}} a \\ & w^{\mathrm{T}} a \leqslant W\,, \\ & a \in Z_+^m\,. \end{aligned} \tag{12.3}$$

Тази задача е интересна сама по себе си и се среща в литературата като *задача за раницата.* (За π_i можем да си мислим като за стойност, а за w_i да си мислим като за тегло на i-ия предмет. Търсим кои и колко предмета да сложим в раницата, така че тя да има най-голяма стойност при условие, че тегло ѝ не надминава W.) Да отбележим, че (SP) е целочислена линейна задача.

Ако оптималната стойност на (SP) е най-много 1, тогава всички редуцирани цени са неотрицателни и текущото базисно допустимо решение е оптимално. В противен случай, оптималното решение a на (SP) дава стълб $a_q = a$, който определя за влизане в базиса съответната небазисна променлива x_q и итерацията по симплекс алгоритъма за решаване на (P) продължава по обичайния начин.

Показахме как можем да използваме симплекс метода за решаване на задачата за едномерен разкрой (P) без да знаем предварително всички стълбове на матрицата A като ги генерираме тогава, когато това е необходимо, като решения на спомагателната задача (SP).

Остава да се спрем накратко на това как може да бъде решена задачата (SP) чрез използването на рекурсивно динамично програмиране в случай че, както е и разумно, W и всички w_i са естествени числа. Нека с $f(v)$ означим оптималната стойност на (SP), когато дясната страна W е заменена с v. Тогава $f(v) = 0$ за $0 \leqslant v < w_{\min}$, където $w_{\min} = \min w_i$. Получаваме $f(W)$ като пресмятаме рекурсивно

$$f(v) = \max_{\substack{1 \leqslant i \leqslant m \\ w_i \leqslant v}} \{f(v - w_i) + \pi_i\}$$

за $v = w_{\min}, \dots, W$. Оптималното решение, за което се получава оптималната стойност $f(W)$ може лесно да се получи чрез връщане назад, ако пазим максимизиращия индекс на всяка стъпка.

Подобни методи за генериране на стълб, при които на всяка итерация се решават по-сложни спомагателни задачи, могат да се използват за решаване на по-общи задачи като например двумерна задача за разкрой (новите ролки са със зададени ширина и дължина), или когато има граница за броя на ножовете в едномерната задача за разкрой.

13. Принцип за декомпозиция

Интересуваме се от подход за решаването на линейна задача от вида

$$\begin{aligned} \min \quad & z = c^{\mathrm{T}} x \\ & A_0 x = b_0\,, \\ & A_1 x = b_1\,, \\ & x \geqslant 0\,, \end{aligned} \tag{13.1}$$

където ограниченията са разделени на "общи" ограничения $A_0x = b_0$ и "специални" ограничения $A_1x = b_1$, $x \geqslant 0$.

Например, специалните ограничения могат да представляват мрежова задача, или пък могат да бъдат разделени на групи от ограничения върху непресичащи се множества от променливи. Бихме искали ефективно да решим (13.1) като използваме факта, че линейни задачи, включващи само специални ограничения могат да бъдат решени много по-лесно. Ще видим, че това води отново до идеята за генериране на стълб, която разгледахме във въпрос 12. Методът, който се получава е известен като принцип за декомпозиция на Данциг и Волф от 1960 г.

Нека запишем задачата (13.1) във вида

$$\begin{aligned} (P) \qquad \min \quad & z = c^{\mathrm{T}} x \\ & A_0 x = b_0\,, \\ & x \in X\,, \end{aligned} \tag{13.2}$$

където

$$X = \{x \in \mathbb{R}^n : A_1 x = b_1,\ x \geqslant 0\}. \tag{13.3}$$

Предполагаме, че A_0 е $m_0 \times n$, а A_1 е $m_1 \times n$ матрица, а векторите c, x, b_0 и b_1 са със съгласувани размерности. Например, (P) може да представлява производствено–дистрибуторен модел с голяма размерност, при който $A_0x = b_0$ включва ограниченията, идващи от ресурсите за производство, докато $A_1x = b_1$ включва мрежови ограничения, идващи от системата за дистрибуция на продукцията.

Идеята е да се представи многостенното множество X от (13.3) посредством неговите върхове и някои от неговите посоки, както в Теорема 2.1. Ще ни трябва малко по-силен резултат от тази теорема, за който ни е необходима

Дефиниция 13.1. Посока d на X се нарича *екстремна* ако не може да се представи като неотрицателна комбинация на две различни (т.е. непропорционални) посоки, т.е.

$$d = \mu_1 d_1 + \mu_2 d_2, \qquad \mu_1 \geqslant 0,\ \mu_2 \geqslant 0,$$

за посоки d_1 и d_2 на X, влече че $d_1 = \alpha_1 d$ и $d_2 = \alpha_2 d$ за някои $\alpha_1, \alpha_2 \geqslant 0$.

Теорема 13.1. *Всяко $x \in X$ може да се представи като*

$$x = \sum_{i \in I} \lambda_i v_i + \sum_{j \in J} \mu_j d_j$$

където $\{v_i : i \in I\}$ е множеството от върховете, а $\{d_j : j \in J\}$ е множеството от екстремните посоки на X и $\sum_{i \in I} \lambda_i = 1$, $\lambda_i \geqslant 0$ за всяко $i \in I$, $\mu_j \geqslant 0$ за всяко $j \in J$. Обратно, всяко такова x лежи в X. Нещо повече, X има само краен брой екстремни посоки.

На доказателството на този резултат няма да се спираме, тъй като то е подобно на доказателството на Теорема 2.1. Важно следствие от тази теорема е

Следствие 13.1. *Нека стълбовете на V и D бъдат съответно всички върхове и екстремни посоки на X. Тогава*

$$X = \{V\lambda + D\mu : e^{\mathrm{T}}\lambda = 1,\ \lambda \geqslant 0,\ \mu \geqslant 0\}$$

(*с e означаваме вектор със съответна размерност, чиито координати са единици*).

Като използваме горното представяне на X, получаваме че задачата (P) е еквивалентна на задачата

$$\begin{array}{lll} (SP) & \min & (c^{\mathrm{T}}V)\lambda + (c^{\mathrm{T}}D)\mu \\ & & (A_0V)\lambda + (A_0D)\mu \;=\; b_0, \\ & & e^{\mathrm{T}}\lambda \;=\; 1, \\ & & \lambda \geqslant 0, \quad \mu \geqslant 0. \end{array} \tag{13.4}$$

Принципът за декомпозиция се състои в прилагане на симплекс метода за решаване на задачата (SP). За разлика от (P), задачата (SP) има само $m_0 + 1$ ограничения, но пък има астрономически брой стълбове – по един за всеки връх и за всяка екстремна посока на X, като при това те са известни само неявно. Така че за решаването на (SP) ще използваме техниката за генериране на стълб, разгледана във въпрос 12.

Да предположим, че имаме базисно допустимо решение $(\overline{\lambda}, \overline{\mu})$ на (SP), на което съответстват симплексни множители $\overline{\pi}_0$ (за първите m_0 ограничения) и $\overline{\sigma}$. Тук $\overline{\lambda}_i > 0$ влече това, че знаем съответния връх v_i на X, а $\overline{\mu}_j > 0$ влече това, че знаем екстремната посока d_j. Въпреки това, множеството от всички върхове и екстремни посоки не ни е известно, така че ще трябва да ги генерираме (както и съответните им стълбове в матрицата от ограничения в (SP)), когато това е необходимо.

Итерацията по симплекс метода за (SP) изисква да намерим връх v_i с редуцирана цена

$$c^{\mathrm{T}}v_i - \overline{\pi}_0^{\mathrm{T}}A_0v_i - \overline{\sigma} < 0 \tag{13.5}$$

или екстремна посока d_j с редуцирана цена

$$c^{\mathrm{T}}d_j - \overline{\pi}_0^{\mathrm{T}}A_0d_j < 0\,. \tag{13.6}$$

Ако такива няма, можем да заключим, че текущото решение $(\overline{\lambda}, \overline{\mu})$ е оптимално за (SP) и следователно $\overline{x} = V\overline{\lambda} + D\overline{\mu}$ е оптимално за (P).

Да разгледаме първо (13.5). Искаме да намерим връх v_i на X, такъв че стойността на линейната функция $(c^{\mathrm{T}} - \overline{\pi}_0^{\mathrm{T}}A_0)v_i$ е по-малка от $\overline{\sigma}$. Тъй като минимумът на линейна функция върху многостенно множество се достига във връх (ако функцията не е неограничена отдолу), то е естествено да разгледаме под-задачата

$$(SP)(\overline{\pi}_0) \quad \min\{(c^{\mathrm{T}} - \overline{\pi}_0^{\mathrm{T}}A_0)x : x \in X\}. \tag{13.7}$$

Да разгледаме възможните изходи от решаването на $(SP)(\overline{\pi}_0)$. Първо, ако тя има празно допустимо множество, то X е празното множество и изходната ни задача (P) също е с празно допустимо множество, като в този случай не бихме могли да имаме и текущо базисно допустимо решение на (SP).

Второ, $(SP)(\overline{\pi}_0)$ може да бъде неограничена. В този случай, прилагането на симплекс метода ще генерира направление η_q от някой връх v на X, за което $v+\theta\eta_q$ е в X за всяко $\theta \geqslant 0$ и $(c^{\mathrm{T}} - \overline{\pi}_0^{\mathrm{T}} A_0)\eta_q < 0$. Лесно се вижда, че η_q е екстремна посока на X, така че като положим $d_j = \eta_q$ получаваме (13.6). Разбира се, не е необходимо η_q да е посока за многостенното множество от допустими решения на (P), тъй като в X не участват ограниченията $A_0 x = b_0$.

И последно, $(SP)(\overline{\pi}_0)$ може да има крайно оптимално решение $\overline{x}$. Тогава според основните теореми на линейното оптимиране (по-специално Теорема 3.2 и нейното доказателство) имаме, че $(c^{\mathrm{T}} - \overline{\pi}_0^{\mathrm{T}} A_0)d \geqslant 0$ за всички посоки d и $\overline{x}$ може да се вземе като връх на X (симплекс метода намира такъв връх). Следователно, (13.6) не е вярно за никое j и нито един стълб, възникващ от екстремна посока не е кандидат за влизане в базиса. Ако $(c^{\mathrm{T}} - \overline{\pi}_0^{\mathrm{T}} A_0)\overline{x} \geqslant \overline{\sigma}$, тъй като минимизираме върху цялото X (и в частност върху всичките му върхове) (13.5) не е вярно за никое i, така че и от връх не възниква стълб като кандидат за влизане в базиса. Заключаваме, че $(\overline{\lambda}, \overline{\mu})$ е оптимално решение за (SP) и, следователно $x^* = V\overline{\lambda} + D\overline{\mu}$ е оптимално за (P). От друга страна, ако $(c^{\mathrm{T}} - \overline{\pi_0} A_0)\overline{x} < \overline{\sigma}$, полагайки $v_i = \overline{x}$ имаме (13.5).

Така във всеки от случаите, или доказваме оптималност или генерираме стълб за (SP) за влизане в базиса, в който случай итерацията по симплекс метода за (SP) може да продължи както обикновено. Като сумираме горните разсъждения, получаваме

Теорема 13.2. (а) *Ако $(SP)(\overline{\pi}_0)$ е неограничена, симплекс метода приложен за нея води до намирането на екстремна посока d_j, удовлетворяваща* (13.6)*, така че стълбът*

$$\begin{pmatrix} A_0 d_j \\ 0 \end{pmatrix}$$

отговаря на изискванията за влизане в базиса за задачата (SP).

(б) *Ако $(SP)(\overline{\pi}_0)$ има оптимално решение v_i и оптимална стойност по-малка от $\overline{\sigma}$, то стълбът*

$$\begin{pmatrix} A_0 v_i \\ 1 \end{pmatrix}$$

отговаря на изискванията за влизане в базиса за задачата (SP).

(в) *Накрая, ако $(SP)(\overline{\pi}_0)$ има оптимална стойност поне $\overline{\sigma}$ и ако $\overline{\pi}_1$ е оптимално решение на двойнствената задача на $(SP)(\overline{\pi}_0)$, то текущото базисно допустимо решение $(\overline{\lambda}, \overline{\mu})$ е оптимално за (SP), а $(\overline{\pi}_0, \overline{\sigma})$ е оптимално решение на двойнствената задача на (SP) и накрая $x^* = V\overline{\lambda} + D\overline{\mu}$ е оптимално за (P), а $(\overline{\pi}_0, \overline{\pi}_1)$ е оптимално решение на двойнствената на (P) задача.*

Доказателство. Предвид направените по-горе разсъждения, трябва да покажем само последната част (в). Тъй като (13.5) и (13.6) не са в сила за никое $i \in I$, $j \in J$, то $(\overline{\pi}_0, \overline{\sigma})$ е допустимо за двойнствената задача на (SP) и следователно $(\overline{\lambda}, \overline{\mu})$ и $(\overline{\pi}_0, \overline{\sigma})$ са оптимални решения съответно на (SP) и на нейната двойнствена задача. Оттук, оптималните стойности на двете задачи са равни: $c^{\mathrm{T}} V\overline{\lambda} + c^{\mathrm{T}} D\overline{\mu} = \overline{\pi}_0^{\mathrm{T}} b_0 + \overline{\sigma}$. Получаваме, че x^* е допустим за (P), тъй като удовлетворява $A_0 x^* = b_0$ и лежи в X според Теорема 13.1. Стойността на целевата функция на (P) в x^* е $c^{\mathrm{T}}x^* = c^{\mathrm{T}}V\overline{\lambda} + c^{\mathrm{T}}D\overline{\mu}$, т.е. е равна на оптималната стойност на (SP). Тъй като $\overline{\pi}_1$ е оптимално за двойнствената задача на $(SP)(\overline{\pi}_0)$, то е допустимо

$$\overline{\pi}_1^{\mathrm{T}}A_1 \leqslant c^{\mathrm{T}} - \overline{\pi}_0^{\mathrm{T}}A_0 \tag{13.8}$$

и стойността на целевата ѝ функция в него е

$$\begin{aligned} \overline{\pi}_1^{\mathrm{T}}b_1 &\geqslant \overline{\sigma} = (\overline{\pi}_0^{\mathrm{T}}b_0 + \overline{\sigma}) - \overline{\pi}_0^{\mathrm{T}}b_0 \\ &= (c^{\mathrm{T}}V\overline{\lambda} + c^{\mathrm{T}}D\overline{\mu}) - \overline{\pi}_0^{\mathrm{T}}b_0 \\ &= c^{\mathrm{T}}x^* - \overline{\pi}_0^{\mathrm{T}}b_0. \end{aligned} \tag{13.9}$$

От (13.8) имаме, че $(\overline{\pi}_0, \overline{\pi}_1)$ е допустимо за двойнствената на (P), а от (13.9) имаме, че стойността на целевата функция на двойнствената на (P) задача в него е по-голяма от стойността на целевата функция на (P) в x^*. От слабата теорема за двойнственост, имаме че x^* е оптимално за (P) и че $(\overline{\pi}_0, \overline{\pi}_1)$ е оптимално за двойнствената на (P). ■

Теоремата показва, че можем да решим (P) като решим със симплекс метода (SP). Прилагаме симплекс алгоритъма за задача, в която не всички коефициенти са предварително известни. Алгоритъмът приключва или с оптимално решение на (SP), и следователно с оптимално решение на (P), или с индикация за неограниченост, като в този случай лесно се вижда, че (P) също е неограничена. Доказателството показва също така, че когато алгоритъмът приключва с намирането на оптимално решение на $(SP)(\overline{\pi}_0)$ за някое $\overline{\pi}_0$, то оптималната стойност на $(SP)(\overline{\pi}_0)$ е точно $\overline{\sigma}$ и от условията за допълнителност имаме, че всички v_i с положителни $\overline{\lambda}_i$ също са оптимални решения на $(SP)(\overline{\pi}_0)$. Прилагането на симплекс метода за (SP) задачата явно показва как да комбинираме оптималните решения на $(SP)(\overline{\pi}_0)$ за да получим оптимално решение на (P), което в общия случай не е връх на X.