Записки
по Математическо оптимиране 1

за специалности
Приложна математика и Статистика

на Факултет по математика и
информатика
на СУ "‘Св. Климент Охридски"’

доц. д-р Надя Златева

§0. Кратко описание на предмета, целите и източниците на курса по Математическо оптимиране 1. Някои основни понятия

Предмет на *математическото оптимиране* е изучаването на следната задача: да се намери оптималната стойност (минимум или максимум) на някаква функция наречена *целева функция*

$$f(x_1, \dots, x_k)$$

при следните ограничения

$$g_i(x_1, \dots, x_k) \lesseqqgtr b_i, \ i = 1, \dots, m,$$

където f, g_i, $i = 1, \dots, m$, са функции на k променливи, а b_i, $i = 1, \dots, m$, са реални числа. Знакът $\lesseqqgtr$ означава, че функционалното ограничение е от един от следните три вида: неравенство $\leqslant$, равенство $=$ или неравенство $\geqslant$.

Тази математическа задача се нарича *задача на математическото оптимиране* или още *оптимизационна задача.*

Всяка точка $\mathbf{x} \in \mathbb{R}^k$, чиито координати удовлетворяват всичките m ограничения на оптимизационната задача се нарича *допустима точка* или *допустимо решение*. *Оптимално решение* е такава допустима точка, в която целевата функция приема най-добрата си стойност (оптимум) сред стойностите си във всички допустими точки. Оптималната стойност на целевата функция, която се търси, е *минимум* или *максимум* и зависи от задачата. Например това може да бъде максимума на печалбата или минимума на загубите при дадено производство.

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 използват най-вече в икономическия анализ и планирането.

§0.1. Обща задача на линейното оптимиране

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.

Следователно общият вид на задача на линейното оптимиране е следният

$$
(L)\qquad
\begin{aligned}
&\min / \max\ z(\mathbf{x}) = \sum_{j=1}^{k} c_j x_j,\\
&\sum_{j=1}^{k} a_{ij} x_j \lesseqgtr b_i,\ i = 1, \ldots, m,\\
&x_j \geqslant 0,\ j \in J \subseteq \{1, 2, \ldots, k\},
\end{aligned}
$$

където за удобство неравенствата отнасящи се само до неотрицателния знак на някои от променливите се записват отделно от останалите ограничения.

§0.2. Кратък исторически преглед на линейното оптимиране

Теоретичните основи на линейното оптимиране се полагат с изучаването на системи линейни неравенства, което може да се проследи назад във времето до една работа на Фурие от 1826 г. По-късно много математици доказват различни частни случаи на най-важния резултат на линейното оптимиране, който и ние ще разгледаме — т. нар. *силна теорема за двойнственост*.

Приложната страна на теорията започва да се разработва през 1939 г. от руския математик Леонид 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 Джордж Данциг, който през 1947 г. разработва симплекс метода за решаване на линейни задачи, които възникват в планирането на въздушните 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 Тялинг Купманс показва, че линейните задачи представляват подходящ модел за анализ на класически икономически теории. През 1975 г. Шведската Кралска Академия на Науките присъжда Нобеловата награда за икономика на Канторович и Купманс „за техния принос в теорията за оптималното разпределение на ресурси“. Става ясно, че Академията разглежда работата на Данциг като твърде математическа за присъждане на наградата по икономика (а знаем, че не се присъжда Нобелова награда за математика).

След като Канторович дава математическата формулировка на общата задача на ЛО, а Данциг разработва симплекс метода за решаването ѝ, редица задачи биват разпознати като задачи на ЛО. Важни примери за такива задачи са транспортната задача, поставена от Хичкок (1941) и задачата за диета, поставена от Стиглър (1945). Конкретни модели на тези и други линейни задачи ще бъдат разгледани на упражненията. Много практически задачи могат да се формулират като задачи на линейното оптимиране, а линейните модели се използват най-вече в икономическия анализ и планирането.

Първото успешно решаване на линейна задача с компютър датира от 1952 г., след което симплекс методът, негови вариации, а и принципно различни алгоритми за решаване на линейната задача са кодирани за среден и висок клас машини.

§0.3. Кратко описание на курса

Целта на курса по Математическо оптимиране 1 е да запознае студентите от специалности Приложна математика и Статистика на ФМИ с основния клас задачи на математическото оптимиране – линейните задачи, със съответните им модели, както и с математически методи за тяхното решаване.

В курса ще бъде развита основната теория на линейното оптимиране и ще бъде представен симплекс метода за решаване на линейни задачи. Изложението ще бъде едновременно геометрично, аналитично и алгоритмично. Идеята е да се даде на студентите представа за математическия апарат на линейното оптимиране и с помощта на прости примери да се покаже сферата на приложение на методите за решаване на линейни задачи. След като получат основни знания за математическите основи на линейното оптимиране студентите биха могли да се занимават самостоятелно с практическо изследване на реални задачи и сами да повишават нивото на своята подготовка като изучават съответните специализирани публикации. Придобитите знания ще бъдат от полза и поради факта, че линейното оптимиране служи за основа и на много други методи на математическото оптимиране като например целочисленото оптимиране.

Въпреки че за решаване на линейни задачи съществува добре разработен софтуер, заблуждаващо е мнението, че не е необходимо да се изучава математическия апарат на линейното оптимиране, понеже компютърът може сам да „реши“ задачата. Не бива да се забравя, че компютърът решава това, което му е формулирано като задача. Следователно, ако студентът не е наясно с вида и възможностите на използвания модел, както и с преимуществата и недостатъците на съответния математически метод, то това неизбежно ще се отрази на качеството на полученото решение.

Полезни за доброто и бързо овладяване на преподавания материал ще бъдат вече придобитите знания от студентите в областта на линейната алгебра и теорията на графите.

§0.4. Източници на курса

Лекциите върху преподавания материал са базирани на следните основни източници:

[1] Х. Таха, Введение в исследование операций, Мир, Москва, 1985; Вильямс, Москва, 2005.

[2] D. Goldfarb, M. Todd. Chapter II. Linear programming, In: Handbooks in OR & MS, vol. 1 (G. L. Nemhauser et al. Eds.), Elsevier, 1989.

[3] W. L. Winston, Operations Research. Applications and Algorithms, 4th ed., Duxbury Press, 2004.

[4] М. Базара, К. Шетти, Нелинейное программирование. Теория и алгоритмы, Москва, Мир, 1982.

[5] P. Jensen, J. Bard, Operations Research. Models and Methods, John Wiley & Sons, 2003.

§0.5. Обща задача на линейното оптимиране

Както видяхме в § 0.1 общата задача на линейното оптимиране се състои в намирането на оптималната стойност на линейна функция (наречена целева функция) върху множество, зададено с линейни ограничения, които могат да са както равенства, така и неравенства по отношение на променливите.

Общият вид на задача на линейното оптимиране е следният

$$(L)\qquad \begin{aligned} &\min / \max \ z(\mathbf{x}) = \sum_{j=1}^{k} c_j x_j, \\ &\sum_{j=1}^{k} a_{ij} x_j \lesseqqgtr b_i, \ i = 1, \ldots, m, \\ &x_j \geqslant 0, \ j \in J \subseteq \{1, 2, \ldots, k\}, \end{aligned}$$

където за удобство неравенствата отнасящи се само до неотрицателния знак на някои от променливите се записват отделно от останалите ограничения.

Като пример за конкретна линейна задача в общ вид с две променливи да разгледаме следния

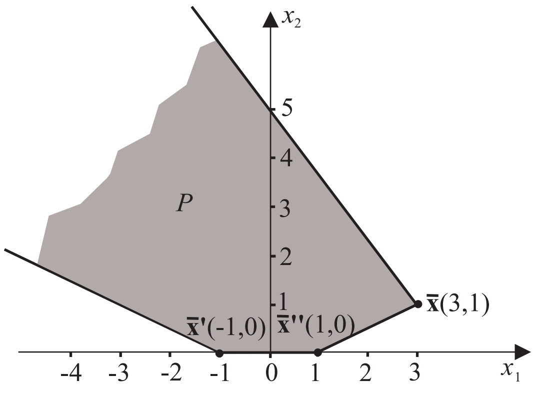
Пример 0.1. Да се намери

$$\max z(\mathbf{x}) = 3x_1 - 2x_2$$

при ограничения

$$\begin{aligned} x_1 - 2x_2 &\leqslant 1, \\ 4x_1 + 3x_2 &\leqslant 15, \\ x_1 + 2x_2 &\geqslant -1, \\ x_2 &\geqslant 0. \end{aligned}$$

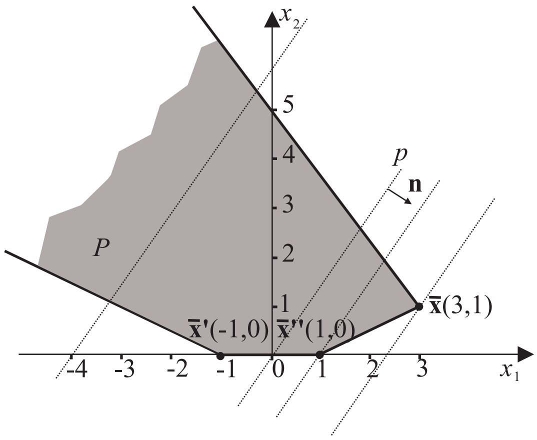
Нека P е допустимото множество на задачата. То се състои от всички вектори $\mathbf{x}(x_1, x_2)$, чиито координати удовлетворяват и четирите линейни неравенства. От геометрична гледна точка, множеството от точки, чиито координати са решения на линейно неравенство, е *полуравнина.* Следователно допустимото множество P е сечение на четири полуравнини и представлява защрихования неограничен многоъгълник с върхове в точките $\overline{\mathbf{x}}(3,1)$, $\overline{\mathbf{x}}'(-1,0)$ и $\overline{\mathbf{x}}''(1,0)$, който е изобразен на фиг. 0.1.



Фигура 0.1.

Интересуваме се от най-голямата стойност на линейната функция $z(\mathbf{x}) = 3x_1 - 2x_2$, която тя може да приеме за векторите $\mathbf{x}(x_1, x_2) \in P$. Да разгледаме уравнението $z(\mathbf{x}) = 0$. Негови решения са всички точки от правата $p : 3x_1 - 2x_2 = 0$, а точките от допустимото множество, в които z приема стойност 0, са точките от сечението на p с многоъгълника, което е отсечка, както се вижда на фиг. 0.2.

Нормалният вектор към правата p е векторът $\mathbf{n}(3, -2)$. Той указва посоката, при движение в която стойността на линейната функция z нараства (в нашия случай надясно). Тъй като решаваната задача е за максимум, ясно е че увеличаване на стойността на целевата функция ще получим при успоредно преместване на p надясно. Това преместване е смислено дотогава, докато сечението на правата с допустимото множество е непразно. В нашия случай можем да преместваме p надясно дотогава, докато сечението ѝ с допустимото множество стане върха $\overline{\mathbf{x}}(3, 1)$. За стойността на z в $\overline{\mathbf{x}}$ имаме $z(\overline{\mathbf{x}}) = 3x_1 - 2x_2 = 3.3 - 2.1 = 7$. По-нататъшно увеличение на стойностите на z не е възможно, тъй като при по-нататъшно местене на p надясно сечението ѝ с допустимото множество P ще бъде празно. Следователно максималната стойност на z върху допустимото множество е 7 и тя се постига във върха $\overline{\mathbf{x}}(3, 1)$, или с други думи казано, за стойности на променливите $x_1 = 3$ и $x_2 = 1$.



Фигура 0.2.

Ако вместо най-голямата ни интересува най-малката стойност на z върху P, то трябва успоредно да местим p наляво. Но колкото и далеч да отместим наляво правата p, сечението ѝ с допустимото множество P ще бъде непразно. Това означава, че в допустимото множество има точки, в които z приема произволно малки стойности. Следователно задачата за минимум на z върху P няма решение поради неограничено намаляване към $-\infty$ на стойностите на целевата функция z в точки от допустимото множество P.

Разгледаният метод за решаване може да се приложи към произволна линейна задача с две променливи и да послужи за намиране на нейно решение или за установяване на неограниченост на целевата ѝ функция върху допустимото ѝ множество. Той обаче е неуместен за линейни задачи с повече от две променливи, което налага за тяхното решаване да се търси по-систематичен подход.

Вектор $\mathbf{x}(x_1, \ldots, x_k) \in \mathbb{R}^k$ ще записваме в матрична форма като вектор-стълб, т.е. като $k \times 1$ матрица състояща се от неговите координати, или $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$. С $\mathbf{x}^{\mathrm{T}}$ ще означаваме транспонирания вектор на вектора $\mathbf{x}$, т.е. $1 \times k$ матрицата $\mathbf{x}^{\mathrm{T}} = [x_1, \ldots, x_k]$. Така скаларното произведение на векторите $\mathbf{c}$ и $\mathbf{x}$ в $\mathbb{R}^k$ можем да запишем като резултат от матрично умножение:

$$\mathbf{c}^{\mathrm{T}}\mathbf{x} = [c_1, \ldots, c_k] \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = \sum_{j=1}^{k} c_j x_j,$$

а целевата функция на линейната задача (L) можем да запишем като $z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}$.

Аналогично, всяко от m-те ограничения на (L) можем да запишем във вида

$$\mathbf{a}_i^{\mathrm{T}}\mathbf{x} \lesseqqgtr b_i, \quad i = 1, \ldots, m,$$

където $\mathbf{a}_i^{\mathrm{T}} = [a_{i1}, \ldots, a_{ik}]$.

§0.6. Канонична задача на линейното оптимиране

Линейната задача в общ вид (L) може да бъде за минимум или пък за максимум, а линейните ограничения могат да бъдат неравенства $\leqslant$ и $\geqslant$ и равенства. За да се улесни и уеднакви подхода към линейната задача в общ вид (L) тя се привежда в *каноничен вид.* Линейна задача, която е в каноничен вид се нарича *канонична задача.*

Дефиниция 0.1. Канонична задача е линейна задача за минимум с ограничения равенства с неотрицателни десни страни, върху всичките променливи на която са наложени условия за неотрицателност.

Линейна задача в общ вид (L) се привежда в съответната ѝ канонична задача (K) като се направят следните прости преобразувания:

- търсенето на $\max \mathbf{c}^{\mathrm{T}}\mathbf{x}$ се заменя с търсенето на $\min -\mathbf{c}^{\mathrm{T}}\mathbf{x}$, тъй като за произволно множество $X \subset \mathbb{R}^k$ и произволна функция $f : \mathbb{R}^k \to \mathbb{R}$ е изпълнено, че $\max\limits_{\mathbf{x}\in X} f(\mathbf{x}) = -\min\limits_{\mathbf{x}\in X} -f(\mathbf{x})$;

- ако $b_i < 0$ за някое i, двете страни на съответното ограничение се умножават с -1;

- всяка *свободна променлива* x_j, $j \notin J$, (променлива x_j, върху която в (L) няма наложено условие за неотрицателност) се заменя с разликата на двойка неотрицателни променливи x_j^+ и x_j^-:

$$x_j = x_j^+ - x_j^-, \quad x_j^+ \geqslant 0, \; x_j^- \geqslant 0;$$

- неравенство от вида

$$\mathbf{a}_i^{\mathrm{T}}\mathbf{x} \leqslant b_i$$

се превръща в равенство чрез добавяне към лявата му страна на нова неотрицателна променлива x_{k+i}, наречена *допълнителна променлива*

$$\mathbf{a}_i^{\mathrm{T}}\mathbf{x} + x_{k+i} = b_i, \quad x_{k+i} \geqslant 0.$$

- неравенство от вида
$$\mathbf{a}_i^{\mathrm{T}}\mathbf{x} \geqslant b_i$$
  се превръща в равенство чрез изваждане от лявата му страна на нова неотрицателна променлива x_{k+i}, наречена също *допълнителна променлива*:
$$\mathbf{a}_i^{\mathrm{T}}\mathbf{x} - x_{k+i} = b_i, \quad x_{k+i} \geqslant 0.$$

От казаното е ясно, че преобразуването на задачата (L) в съответната ѝ канонична задача (K) става за сметка на увеличаване на броя на променливите.

В матрична форма каноничната задача се записва така:

$$(K) \qquad \begin{aligned} \min \quad & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

където $\mathbf{c}$ е n-мерен вектор, който наричаме *вектор на целевата функция*, $\mathbf{x}$ е n-мерен вектор, който наричаме *вектор на променливите*, $\mathbf{b}(b_1, \ldots, b_m)$ е m-мерен вектор с неотрицателни координати, който наричаме *вектор на дясната страна*, а $m \times n$ матрицата $\mathbf{A} = \{a_{ij}\}_{i=1\div m}^{j=1\div n}$, чиито вектор-редове са векторите $\mathbf{a}_i^{\mathrm{T}}$ на линейните функции, описващи ограниченията, наричаме *матрица на ограниченията*.

Да отбележим, че неравенството $\mathbf{x} \geqslant \mathbf{0}$ трябва да се разбира като векторно неравенство и означава, че $x_j \geqslant 0$ за всяко $j = 1, \ldots, n$, т.е. $\mathbf{x} \geqslant \mathbf{0}$ е еквивалентно на система от n числови неравенства. Аналогично, равенството $\mathbf{A}\mathbf{x} = \mathbf{b}$ е векторно равенство и означава, че координатите на векторите $\mathbf{A}\mathbf{x}$ и $\mathbf{b}$ са съответно равни, т.е. $\mathbf{A}\mathbf{x} = \mathbf{b}$ е еквивалентно на система от m числови равенства.

Да отбележим също така, че координатите на вектора на целевата функция в каноничната задача (K), съответстващи на допълнителни променливи, са нули.

Изходната задача (L) и съответната ѝ канонична задача (K) са еквивалентни в смисъл, че на всяка допустима точка на (L) чрез подходящо преобразуване на координатите ѝ може да се съпостави допустима точка на (K) (и обратно) така че стойностите на целевите функции на двете задачи в тях да се различават евентуално само по знак.

Следователно, за да решим изходната задача (L) е достатъчно да решим съответната ѝ канонична задача (K).

По-нататък в курса ще работим предимно с канонични линейни задачи.

За да илюстрираме правилата за преобразуване на линейна задача в канонична, да приведем в каноничен вид задачата от Пример 0.1.

Тъй като x_1 е свободна по знак променлива, навсякъде (в целевата функция и в ограниченията) я представяме във вида $x_1 = x_1^+ - x_1^-$, където $x_1^+ \geqslant 0$ и $x_1^- \geqslant 0$.

От задача за максимум преминаваме към задача за минимум

$$\min z_K(\mathbf{x}) = -3x_1^+ + 3x_1^- + 2x_2.$$

Тъй като третото неравенство е с отрицателна дясна страна, го умножаваме с -1, за да получим

$$-x_1 - 2x_2 \leqslant 1.$$

Добавяме по една неотрицателна допълнителна променлива в лявата страна на трите неравенства, които сега са в посоката $\leqslant$.

Окончателно, каноничният вид на задачата от примера е

$$\begin{array}{l}
\min z_K(\mathbf{x}) = -3x_1^+ + 3x_1^- + 2x_2 \\
\;\; x_1^+ - \;\; x_1^- - 2x_2 + x_3 \qquad\qquad\quad = \;\; 1, \\
\;4x_1^+ - 4x_1^- + 3x_2 \qquad + x_4 \qquad\; = 15, \\
-x_1^+ + \;\; x_1^- - 2x_2 \qquad\qquad\quad + x_5 = \;\; 1, \\
x_1^+, x_1^-, x_2, x_3, x_4, x_5 \geqslant 0.
\end{array}$$

От изходната линейна задача с две променливи $\mathbf{x}_L(x_1, x_2)$ получихме съответната ѝ канонична задача с шест променливи $\mathbf{x}_K(x_1^+, x_1^-, x_2, x_3, x_4, x_5)$.

Ако векторът $\mathbf{x}_L(x_1, x_2)$ е допустим за изходната задача и

- $x_1 \geqslant 0$, то съответният вектор $\mathbf{x}_K(x_1, 0, x_2, 1 - x_1 + 2x_2, 15 - 4x_1 - 3x_2, 1 + x_1 + 2x_2)$ е допустим за каноничната задача;

- $x_1 < 0$, то съответният вектор $\mathbf{x}_K(0, -x_1, x_2, 1 - x_1 + 2x_2, 15 - 4x_1 - 3x_2, 1 + x_1 + 2x_2)$ е допустим за каноничната задача.

И в двата случая $z_K(\mathbf{x}_K) = -z(\mathbf{x}_L)$, т.е. стойностите на целевите функции на двете задачи в съответните допустими решения се различават само по знак.

Обратно, ако $\mathbf{x}_K(x_1^+, x_1^-, x_2, x_3, x_4, x_5)$ е допустим за каноничната задача, то съответният му $\mathbf{x}_L(x_1^+ - x_1^-, x_2)$ е допустим за изходната задача, като отново $z(\mathbf{x}_L) = -z_K(\mathbf{x}_K)$.

§1. Канонично многостенно множество. Върхове

В § 0.6 показахме как произволна линейна задача може да се сведе до съответна на нея канонична задача и казахме, че оттук нататък ще работим с канонична задача

$$(K) \qquad \begin{array}{ll} \min & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{array}$$

където $\mathbf{A}$ е $(m \times n)$ матрица, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m_+$ и $\mathbf{x} \in \mathbb{R}^n$ е n-мерен вектор на променливите.

Вектор $\mathbf{x}(x_1, \ldots, x_n)$ се нарича *допустима точка* или *допустимо решение* на задачата (K), ако координатите му удовлетворяват всички ограничения. Допустима точка се нарича *оптимално решение* на задачата (K), ако в нея целевата функция z приема най-малка стойност сред стойностите си във всички допустими точки. Множеството от всички допустими точки на задачата (K) означаваме с

$$M := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$$

и наричаме нейно *допустимо множество*.

За задачата (K) има точно три възможности:

- Допустимото ѝ множество M е празно. Това означава, че ограниченията на (K) са несъвместими. В този случай (K) няма решение и се нарича *несъвместима*.

- Допустимото множество M е непразно, а целевата функция $z(\mathbf{x})$ е неограничена отдолу върху M (т.е. $z(\mathbf{x}_n) \to -\infty$ за вектори $\mathbf{x}_n \in M$). Тогава (K) няма решение и се нарича *неограничена*.

- Допустимото множество M е непразно, а целевата функция $z(\mathbf{x})$ е ограничена отдолу върху M. В този случай ще докажем, че (K) има оптимално решение, т.е. съществува $\mathbf{x}^* \in M$, такова че $z(\mathbf{x}^*) = \min\{z(\mathbf{x}) : \mathbf{x} \in M\}$ и (K) се нарича *разрешима*.

Очевидно разрешимостта на каноничната задача (K) е свързана със свойствата на допустимото ѝ множество $M := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ и за начало ще разгледаме някои негови основни свойства.

Всяко множество от вида

$$M := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$$

ще наричаме *канонично множество*. От алгебрична гледна точка то се състои от неотрицателните решения $\mathbf{x} \geqslant \mathbf{0}$ на система линейни уравнения $\mathbf{A}\mathbf{x} = \mathbf{b}$. За да вникнем обаче в теорията на линейното оптимиране и за да разберем идеята на симплекс метода, използван за решаване на линейни задачи, от съществено значение е да изучим геометричните свойства на това множество.

§1.1. Изпъкнали множества. Някои свойства

Дефиниция 1.1. Нека $\mathbf{x}_1$ и $\mathbf{x}_2$ са две точки в $\mathbb{R}^n$. За всяко $\lambda \in [0,1]$ съответната точка

$$\mathbf{x}_\lambda = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$$

се нарича *изпъкнала комбинация* на точките $\mathbf{x}_1$ и $\mathbf{x}_2$ с *тегла* съответно λ и $1-\lambda$. Ако $\lambda \neq 0, 1$, то $\mathbf{x}_\lambda$ се нарича *истинска изпъкнала комбинация* на точките $\mathbf{x}_1$ и $\mathbf{x}_2$.

Очевидно, когато λ се мени от 0 до 1 съответните изпъкнали комбинации $\mathbf{x}_\lambda$ описват отсечката с краища $\mathbf{x}_2$ и $\mathbf{x}_1$.

Дефиниция 1.2. Множество $C \subseteq \mathbb{R}^n$ се нарича *изпъкнало*, когато за всеки две точки $\mathbf{x}_1$ и $\mathbf{x}_2$, принадлежащи на C, на C принадлежи и всяка изпъкнала комбинация на $\mathbf{x}_1$ и $\mathbf{x}_2$ (всяка точка от съединяващата ги отсечка).

С други думи, едно множество е изпъкнало, ако заедно с всеки две точки от множеството в него се съдържа изцяло и отсечката, която ги съединява (вж. фиг. 1.3). Празното и едноточковото множество по дефиниция са изпъкнали множества.

Фигура 1.3.

Дефиниция 1.3. Точка $\mathbf{x}$ от изпъкнало множество C се нарича *връх* на C ако $\mathbf{x}$ не може да се представи като истинска изпъкнала комбинация на две различни точки от C.

С други думи, $\mathbf{x} \in C$ е връх на C, ако $\mathbf{x}$ не е вътрешна точка за никоя отсечка с краища в C, т.е. представянето $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, където $0 < \lambda < 1$ и $\mathbf{x}_1, \mathbf{x}_2 \in C$, $\mathbf{x}_1 \neq \mathbf{x}_2$ е невъзможно.

Съгласно дефиницията, всеки (плътен) триъгълник има три върха и това са върховете му, а единичното евклидово кълбо $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leqslant 1\}$ има за върхове точките от единичната евклидова сфера $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$, , а правата и полуравнината са изпъкнали множества, които нямат върхове.

Като обобщение на Дефиниция 1.1 за k-точки получаваме

Дефиниция 1.4. *Изпъкнала комбинация* на k точки $\mathbf{x}_1, \ldots, \mathbf{x}_k$ се нарича точка от вида

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i, \quad \text{където } \lambda_i \geqslant 0 \text{ за всяко } i = 1, \ldots, k, \quad \text{и} \quad \sum_{i=1}^{k} \lambda_i = 1\,.$$

Да забележим, че в една изпъкнала комбинация теглата на участващите в нея точки са неотрицателни и сумата им е равна на 1.

Множеството от всички изпъкнали комбинации на k точки $\mathbf{x}_1, \ldots, \mathbf{x}_k$ се нарича тяхна *изпъкнала обвивка*, означава се с $\mathrm{co}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ и е най-малкото изпъкнало множество, което ги съдържа. Например, изпъкналата обвивка на две точки е отсечката, която ги съединява, изпъкналата обвивка на три точки в общо положение е триъгълникът с върхове в тях, на четири — образуваната от тях пирамида, и т. н.

Симплекс с размерност k се нарича изпъкналата обвивка $k+1$ точки $\mathrm{co}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k)$, такива че векторите $\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_k - \mathbf{x}_0$ са линейно независими.

Твърдение 1.1. *Ако $C_1, \ldots, C_k$ са изпъкнали множества в $\mathbb{R}^n$, то тяхното сечение $C := \bigcap_{i=1}^{k} C_i$ също е изпъкнало множество.*

Доказателство. Ако C е празното множество или е едноточково, то няма какво да доказваме. Да вземем две произволни точки $\mathbf{x}', \mathbf{x}'' \in C$ и произволно число $\lambda \in [0, 1]$ и да образуваме изпъкналата им комбинация $\lambda\mathbf{x}' + (1-\lambda)\mathbf{x}''$. Ако покажем, че тя принадлежи на C, твърдението ще бъде доказано. Да фиксираме индекс i в множеството $\{1, \ldots, k\}$. От това, че $\mathbf{x}', \mathbf{x}'' \in C_i$, което е изпъкнало, следва, че $\lambda\mathbf{x}' + (1-\lambda)\mathbf{x}'' \in C_i$. Тъй като това е вярно за всяко i в множеството $\{1, \ldots, k\}$ имаме, че $\lambda\mathbf{x}' + (1-\lambda)\mathbf{x}'' \in \bigcap_{i=1}^{k} C_i = C$. □

§1.2. Многостенно множество. Изпъкналост. Канонично многостенно множество

Дефиниция 1.5. Множество от точки $\mathbf{x} \in \mathbb{R}^n$, които са решения на някакво линейно уравнение $\mathbf{a}^{\mathrm{T}}\mathbf{x} = b$, където $\mathbf{a} \in \mathbb{R}^n$ е ненулев вектор, а $b \in \mathbb{R}$ е реално число, се нарича *хиперравнина* и се означава с $H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathrm{T}}\mathbf{x} = b\}$. Векторът $\mathbf{a}$ се нарича *нормален вектор* на H.

При $n = 2$ хиперравнината е права, при $n = 3$ хиперравнината е равнина и т. н.

Дефиниция 1.6. Множество от точки $\mathbf{x} \in \mathbb{R}^n$, които са решения на някакво линейно неравенство $\mathbf{a}^{\mathrm{T}}\mathbf{x} \geqslant b$, където $\mathbf{a} \in \mathbb{R}^n$ е ненулев вектор, а $b \in \mathbb{R}$ е реално число, се нарича *затворено полупространство* и се означава с $H^+ = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathrm{T}}\mathbf{x} \geqslant b\}$.

Всяка хиперравнина $H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathrm{T}}\mathbf{x} = b\}$ определя затворените полупространства $H^+ = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathrm{T}}\mathbf{x} \geqslant b\}$ и $H^- = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathrm{T}}\mathbf{x} \leqslant b\}$.

Дефиниция 1.7. *Многостенно множество* $P \subset \mathbb{R}^n$ наричаме множество, образувано от пресичането на краен брой затворени полупространства и хиперравнини. Ако многостенно множество е непразно и ограничено, то се нарича *многостен*.

Твърдение 1.2. *Многостенно множество $P \subset \mathbb{R}^n$ е изпъкнало.*

Доказателство. Лесно се доказва, че хиперравнините и затворените полупространства са изпъкнали множества. Като сечение на изпъкнали множества P е изпъкнало съгласно Твърдение 1.1. □

Следствие 1.1. *Канонично множество M е многостенно множество.*

Доказателство. Множеството M представлява сечение на m-те хиперравнини, определени с уравненията

$$\mathbf{a}_1^{\mathrm{T}}\mathbf{x} = b_1, \; \ldots, \; \mathbf{a}_m^{\mathrm{T}}\mathbf{x} = b_m$$

и на n-те затворените полупространства, определени с неравенствата

$$\mathbf{e}_1^{\mathrm{T}}\mathbf{x} \geqslant 0, \; \ldots, \; \mathbf{e}_n^{\mathrm{T}}\mathbf{x} \geqslant 0,$$

където $\mathbf{a}_i^{\mathrm{T}}$ е i-ят вектор-ред на матрицата $\mathbf{A}$, а $\mathbf{e}_i^{\mathrm{T}}$ е i-ят вектор-ред на единичната матрица от ред n. □

Канонично множество M ще наричаме още *канонично многостенно множество*.

Следствие 1.2. *Канонично множество M е изпъкнало множество.*

Доказателство. Следва непосредствено от Следствие 1.1 и Твърдение 1.2. □

§1.3. Как да разпознаем точките в канонично множество M, които не са негови върхове

От оптимизационна гледна точка върховете на канонично множество M са важни негови елементи, защото (както ще покажем по-късно в Теорема 5.2) ако каноничната задача (K) е разрешима, то сред решенията ѝ има такова, което е връх на допустимото ѝ множество M.

Нека е дадено канонично множество $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$. Нека е дадена и точка $\mathbf{x} \in M$.

Ще покажем, че ако съответните на положителните координати на $\mathbf{x}$ вектор-стълбове на $\mathbf{A}$ са линейно зависими, то от коефициентите на представянето на нулевия вектор като тяхна линейна комбинация може да се построи ненулев вектор $\mathbf{z}$, такъв че тръгвайки от $\mathbf{x}$ можем да се придвижим донякъде както по посока на $\mathbf{z}$ така и по посока на $-\mathbf{z}$ без да напускаме множеството M. Очевидно, в този случай $\mathbf{x}$ ще бъде във вътрешността на отсечка с краища в M и няма да бъде връх на M.

Лема 1.1. *Нека* $\mathbf{x} \in M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$. *Ако вектор-стълбовете на матрицата* $\mathbf{A}$*, съответстващи на положителните координати на* $\mathbf{x}$ *са линейно зависими, то съществува ненулев вектор* $\mathbf{z} \in \mathbb{R}^n$ *и числа* $t^- < 0 < t^+]$*, такива че*

1. $\mathbf{A}\mathbf{z} = \mathbf{0}$*;*
2. $z_i = 0$*, ако* $x_i = 0$*;*
3. *векторът* $\mathbf{x} + t\mathbf{z} \in M$ *за всяко* $t \in [t^-, t^+]$

и $\mathbf{x}$ *не е връх на* M*.*

Доказателство. Да означим вектор-стълбовете на матрицата $\mathbf{A}$ с $\mathbf{A}_1, \dots, \mathbf{A}_n$.

Да вземем произволно $\mathbf{x} \in M$ и да допуснем, че първите p координати на $\mathbf{x}$ са положителни, а последните $n - p$ са нули, т.е. $\mathbf{x}(x_1, \dots, x_p, 0, \dots, 0)$ като $x_j > 0$, $j = 1, \dots, p$ (винаги можем да го постигнем с преномериране на променливите).

От линейната зависимост на $\mathbf{A}_1, \dots, \mathbf{A}_p$, следва че съществуват числа $z_1, \dots z_p$ не всичките равни на нула и такива, че $\sum_{j=1}^{p} z_j \mathbf{A}_j = \mathbf{0}$. Нека означим $\mathbf{z}(z_1, \dots, z_p, 0, \dots 0) \in \mathbb{R}^n$. Да разгледаме векторите $\mathbf{x} + t\mathbf{z}$, където t е реално число.

Да забележим, че

$$\mathbf{A}\mathbf{z} = \sum_{j=1}^{n} z_j \mathbf{A}_j = \sum_{j=1}^{p} z_j \mathbf{A}_j = \mathbf{0}, \tag{1.1}$$

откъдето за всяко $t \in \mathbb{R}$ е изпълнено $\mathbf{A}(\mathbf{x} + t\mathbf{z}) = \mathbf{A}\mathbf{x} + t\mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x} = \mathbf{b}$.

Следователно $\mathbf{x} + t\mathbf{z} \in M$ точно тогава, когато $\mathbf{x} + t\mathbf{z} \geqslant \mathbf{0}$.

Векторът $\mathbf{x} + t\mathbf{z} \geqslant \mathbf{0}$ тогава и само тогава, когато $x_i + tz_i \geqslant 0$ за $i = 1, \dots, p$, т.е. $t \geqslant -\frac{x_i}{z_i}$ за всяко i, такова че $z_i > 0$ и $t \leqslant -\frac{x_i}{z_i}$ за всяко i, такова че $z_i < 0$.

Полагаме $t^- := \max\limits_{i:z_i>0} -\dfrac{x_i}{z_i}$, а ако няма i, такива че $z_i > 0$, полагаме $t^- := -\infty$. Аналогично, полагаме $t^+ := \min\limits_{i:z_i<0} -\dfrac{x_i}{z_i}$, а ако няма i, такива че $z_i < 0$, полагаме $t^+ := +\infty$. Тогава $t^- < 0 < t^+$ като поне едно от двете числа е крайно ($\mathbf{z} \neq \mathbf{0}$) и за всяко $t \in [t^-, t^+]$ е изпълнено $\mathbf{x} + t\mathbf{z} \in M$. □

Възниква логичният въпрос: след като от линейната зависимост на вектор-стълбовете на матрицата $\mathbf{A}$, съответстващи на положителните координати на вектор $\mathbf{x} \in M$ следва, че $\mathbf{x}$ не е връх на M можем ли да твърдим, че от линейната независимост на вектор-стълбовете на матрицата $\mathbf{A}$, съответстващи на положителните координати на вектор $\mathbf{x} \in M$ следва, че $\mathbf{x}$ е връх на M? Положителният отговор на този въпрос ни дава

§1.4. Алгебрична характеризация на върховете на M

Теорема 1.1 (алгебрична характеризация на върховете на M). *Точка $\mathbf{x} \in M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ е връх на M тогава и само тогава, когато стълбовете на матрицата $\mathbf{A}$, съответстващи на положителните координати на $\mathbf{x}$, са линейно независими.*

Доказателство. Да фиксираме $\mathbf{x} \in M$ като допуснем, че първите p координати на $\mathbf{x}$ са положителни, а последните $n-p$ са нули, т.е. $\mathbf{x}(x_1, \dots, x_p, 0, \dots, 0)$ като $x_j > 0$, $j = 1, \dots, p$.

Ако $p = 0$, т.е. $\mathbf{x} = \mathbf{0}$, то очевидно $\mathbf{x}$ е връх на M, защото ако допуснем че $\mathbf{0} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ за някое $\lambda \in (0,1)$ и някои $\mathbf{x}_1 \geqslant \mathbf{0}$ и $\mathbf{x}_2 \geqslant \mathbf{0}$ от M, то това води до $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{0}$.

Нека $p > 0$ и $\mathbf{x}$ е връх на M. Допускаме че векторите $\mathbf{A}_1, \dots, \mathbf{A}_p$ са линейно независими. Прилагаме Лема 1.1 и получаваме ненулев вектор $\mathbf{z} = (z_1, \dots, z_p, 0, \dots 0)$ и числа $t^- < 0 < t^+$, такива че $\mathbf{x} + t\mathbf{z} \in M$ за всяко $t \in [t^-, t^+]$. Да фиксираме $t > 0$, такова че t и $-t$ да бъдат в интервала $[t^-, t^+]$. Тогава точките $\mathbf{x}_1 := \mathbf{x} + t\mathbf{z}$ и $\mathbf{x}_2 := \mathbf{x} - t\mathbf{z}$ са в M, като $\mathbf{x}_1 \neq \mathbf{x}_2$, защото $t > 0$ и $\mathbf{z} \neq \mathbf{0}$. Като вземем средата на отсечката с краища $\mathbf{x}_1$ и $\mathbf{x}_2$, получаваме

$$\frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2 = \frac{1}{2}(\mathbf{x} + t\mathbf{z}) + \frac{1}{2}(\mathbf{x} - t\mathbf{z}) = \mathbf{x},$$

т.е. $\mathbf{x}$ е истинска изпъкнала комбинация на две различни точки $\mathbf{x}_1$ и $\mathbf{x}_2$ от M, което е в противоречие с това, че $\mathbf{x}$ е връх на M. Полученото противоречие показва, че ако точката $\mathbf{x}$ е връх та M, то стълбовете на $\mathbf{A}$, съответстващи на положителните ѝ координати са линейно независими.

Обратно, нека сега стълбовете $\mathbf{A}_1, \dots, \mathbf{A}_p$ са линейно независими. Да допуснем, че $\mathbf{x}$ не е връх на M. Това означава, че $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ за някои $\mathbf{x}_1$, $\mathbf{x}_2 \in M$, $\mathbf{x}_1 \neq \mathbf{x}_2$ и някое $0 < \lambda < 1$. Тъй като и двете точки $\mathbf{x}$ и $\mathbf{x}_1$ са в M, $\mathbf{A}(\mathbf{x} - \mathbf{x}_1) = \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Тъй като и двете числа λ и $1-\lambda$ са положителни, последните $n-p$ координати на $\mathbf{x}_1$ и $\mathbf{x}_2$ са нули, тъй като тези координати на $\mathbf{x}$ са равни на нула. Следователно последните $n-p$ координати на вектора $\mathbf{x} - \mathbf{x}_1$ са нули, а сред първите p има поне една ненулева (в противен случай $\mathbf{x} = \mathbf{x}_1 = \mathbf{x}_2$, което е невъзможно, тъй като $\mathbf{x}_1 \neq \mathbf{x}_2$). Тогава $\mathbf{0} = \mathbf{A}(\mathbf{x} - \mathbf{x}_1) = \sum_{j=1}^{n}(x_j - x_{1j})\mathbf{A}_j = \sum_{j=1}^{p}(x_j - x_{1j})\mathbf{A}_j$, откъдето следва, че стълбовете $\mathbf{A}_1, \dots, \mathbf{A}_p$ са линейно зависими. Полученото противоречие доказва теоремата. $\square$

§2. Канонично многостенно множество. Базисни допустими решения

Продължаваме разглежданията на канонично многостенно множество

$$M := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\},$$

които започнахме в предишния § 1.

Вече знаем, че върховете на M се характеризират с линейната независимост на вектор-стълбовете на матрицата $\mathbf{A}$, съответстващи на положителните им координати.

Да припомним, че максималният брой линейно независими вектор-стълбове на $m \times n$ матрица $\mathbf{A}$ се нарича *ранг* на $\mathbf{A}$ и се означава с $r(\mathbf{A})$. Рангът на $\mathbf{A}$ също така е равен на максималния брой линейно независими вектор-редове на $\mathbf{A}$, както и на реда на най-големия ненулев минор на $\mathbf{A}$. Ако всичките m вектор-редове на $\mathbf{A}$ са линейно независими, казваме че $\mathbf{A}$ има пълен ранг по редове. Очевидно в този случай $r(\mathbf{A}) = m$ и $m \leqslant n$.

Когато $\mathbf{A}$ няма пълен ранг по редове, то

- или някои от уравненията в системата $\mathbf{A}\mathbf{x} = \mathbf{b}$ са *излишни* (в смисъл, че са линейно зависими от останалите уравнения);

- или системата линейни уравнения $\mathbf{A}\mathbf{x} = \mathbf{b}$ няма решение (и съответно M е празното множество).

Тези два случая зависят от съвместимостта на линейната зависимост на вектор-редовете на $\mathbf{A}$ и десните страни на уравненията.

Нека вектор-редът $\mathbf{a}_k^{\mathrm{T}}$ е в линейна зависимост от останалите вектор-редове на $\mathbf{A}$, т.е. за някои $\alpha_i \in \mathbb{R}$ не всички равни на нула е изпълнено $\mathbf{a}_k = \sum\limits_{1 \leqslant i \leqslant m, i \neq k} \alpha_i \mathbf{a}_i$.

Ако $b_k \neq \sum\limits_{1 \leqslant i \leqslant m, i \neq k} \alpha_i b_i$, то системата $\mathbf{A}\mathbf{x} = \mathbf{b}$ няма решение.

Ако $b_k = \sum\limits_{1 \leqslant i \leqslant m, i \neq k} \alpha_i b_i$, то k-то уравнение е в линейна зависимост от останалите, следователно е излишно и може да се отстрани от системата.

След като едно по едно бъдат отстранени всички излишни уравнения, се достига до система от уравнения, чиято матрица е с пълен ранг по редове.

§2.1. Базисни решения. Базисни допустими решения

Оттук нататък ще предполагаме, че матрицата $\mathbf{A}$, участваща в представянето на множеството M е с пълен ранг по редове. В този случай координатите на върховете на M се получават като *базисни решения* на линейната система $\mathbf{Ax} = \mathbf{b}$. Какво се разбира под базисно решение?

Тъй като $r(\mathbf{A}) = m$, то съществува поне една неособена подматрица от ред m на матрицата $\mathbf{A}$. Всяка такава подматрица се нарича *базисна матрица*.

Нека фиксираме базисна матрица $\mathbf{B}$. Означаваме с B множеството от индекси на стълбовете на $\mathbf{A}$, които са стълбове на $\mathbf{B}$. Ако предположим, че $\mathbf{B}$ се състои от първите m стълба на $\mathbf{A}$ (винаги можем да го постигнем чрез преномериране на променливите), тогава $B = \{1, \ldots, m\}$.

Матрицата $\mathbf{A}$ се разделя на две подматрици $\mathbf{A} = [\mathbf{B}|\mathbf{N}]$, където с $\mathbf{N}$ е означена подматрицата на $\mathbf{A}$, състояща се от останалите $n - m$ стълба. Съответно с N означаваме множеството от небазисните индекси, в случая $N = \{m + 1, \ldots, n\}$.

m-те променливи, чиито индекси са базисни (т.е. x_j за $j \in B$) наричаме *базисни променливи* или *базис*, а останалите $n - m$ променливи (т.е. x_j за $j \in N$) наричаме *небазисни*. Векторът на базисните променливи означаваме с $\mathbf{x}_B$, а векторът на небазисните променливи – с $\mathbf{x}_N$. Така произволен вектор $\mathbf{x} \in \mathbb{R}^n$ се представя във вида $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix}$.

Нека сега векторът $\mathbf{x}$ е решение на системата $\mathbf{Ax} = \mathbf{b}$. Тогава

$$\mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad [\mathbf{B}|\mathbf{N}] \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{Bx}_B + \mathbf{Nx}_N = \mathbf{b}.$$

Тъй като $\mathbf{B}$ е неособена матрица, съществува матрицата $\mathbf{B}^{-1}$, с която умножаваме отляво преобразуваната система, за да получим

$$\mathbf{B}^{-1}\,|\,\mathbf{Bx}_B + \mathbf{Nx}_N = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{x}_B + \mathbf{B}^{-1}\mathbf{Nx}_N = \mathbf{B}^{-1}\mathbf{b} \quad \Leftrightarrow$$

$$\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{Nx}_N. \tag{2.1}$$

Последното векторно равенство означава, че линейната система $\mathbf{Ax} = \mathbf{b}$ е разрешена относно базисните променливи $\mathbf{x}_B$, т.е. че посредством (2.1) базисните координати $\mathbf{x}_B$ на решение $\mathbf{x}$ на системата се изразяват чрез небазисните му координати $\mathbf{x}_N$.

Казано с други думи, за всеки зададен набор от $n - m$ числени стойности на небазисните променливи $\mathbf{x}_N$ чрез (2.1) се пресмятат m-те числени стойности на базисните променливи $\mathbf{x}_B$ и така се получават n-те координати на едно частно решение $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix}$ на системата. Ако зададем нулеви стойности на небазисните променливи, т.е. положим $\overline{\mathbf{x}}_N = \mathbf{0}$, от (2.1) за базисните променливи получаваме $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b}$.

По този начин на всяка базисна матрица $\mathbf{B}$ (или което е еквивалентно, на всеки базис B) по единствен начин се съпоставя решение на системата $\mathbf{Ax} = \mathbf{b}$ от вида $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix}$.

Дефиниция 2.1. Нека $\mathbf{B}$ е базисна матрица, съставена от m (линейно независими) стълба на $\mathbf{A}$. Векторът $\overline{\mathbf{x}}$, чиито небазисни координати са нули (т.е. $\overline{\mathbf{x}}_N = \mathbf{0}$), а базисните му координати са $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b}$ се нарича *базисно решение* с *базисна матрица* $\mathbf{B}$ (с *базис* B).

Терминът базис идва от това, че вектор-стълбовете на матрицата $\mathbf{B}$ образуват базис за линейното пространство, породено от вектор-стълбовете на $\mathbf{A}$, и на $\mathbf{Ax} = \mathbf{b}$ може да се гледа като на представяне на вектора $\mathbf{b}$ като линейна комбинация на вектор-стълбовете на $\mathbf{A}$, чиито коефициенти са координатите на решение $\mathbf{x}$ на системата.

Дефиниция 2.2. Ако базисно решение $\overline{\mathbf{x}}$ с базисна матрица $\mathbf{B}$ (с базис B) е неотрицателно, т.е. ако $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix} \geqslant \mathbf{0}$, то се нарича *базисно допустимо решение*, т.к. в този случай $\overline{\mathbf{x}}$ принадлежи на каноничното множество $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ и следователно е допустимо решение за канонична задача с допустимо множество M.

Геометричното понятие *връх* и алгебричното понятие *базисно допустимо решение* съответстват на един и същ обект в каноничното множество M, както ще се убедим от следната

Теорема 2.1. *Ако каноничното множество* $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ *е непразно, то за* $\overline{\mathbf{x}} \in M$ *следните твърдения са еквивалентни:*

(а) $\overline{\mathbf{x}}$ *е връх на* M*;*

(б) $\overline{\mathbf{x}}$ *е базисно допустимо решение.*

Доказателство. (а) $\Longrightarrow$ (б). От Теорема 1.1 следва линейната независимост на вектор-стълбовете на $\mathbf{A}$, съответстващи на положителните координати на $\overline{\mathbf{x}}$. Тъй като $r(\mathbf{A}) = m$ техният брой не надминава m и можем да ги допълним до базисна $m \times m$ матрица $\mathbf{B}$ на $\overline{\mathbf{x}}$ с някои от другите стълбове на $\mathbf{A}$.

(б) $\Longrightarrow$ (а). Нека $\overline{\mathbf{x}}$ е базисно допустимо решение с базисна матрица $\mathbf{B}$, т.е. $\overline{\mathbf{x}}_N = \mathbf{0}$, $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b} \geqslant \mathbf{0}$. Положителните координати на $\overline{\mathbf{x}}$ са сред базисните му координати, чиито индекси, разбира се, са индекси на стълбове на $\mathbf{B}$, а те са линейно независими. От Теорема 1.1 следва, че $\overline{\mathbf{x}}$ е връх на M. □

Теорема 2.1 позволява по-нататък да използваме термините връх и базисно допустимо решение като синоними.

§2.2. Изродени и неизродени базисни допустими решения

Тъй като $r(\mathbf{A}) = m$, връх в M може да има най-много m на брой положителни координати. Правим разлика между върхове с точно m положителни координати и върхове с по-малко от m положителни координати.

Дефиниция 2.3. Базисно допустимо решение $\overline{\mathbf{x}}$ с базис B се нарича *неизродено*, ако $\overline{\mathbf{x}}_B > \mathbf{0}$, и *изродено* в противен случай. Тогава сред базисните му координати има такива с нулева стойност, които се наричат *базисни нули*. Връх на M се нарича *неизроден*, ако съответното му базисно допустимо решение е неизродено и се нарича *изроден* в противен случай.

Знаем, че на всяка базисна матрица съответства точно едно базисно решение, което – ако е допустимо – е връх на M. Ако обаче връх на M има по-малко от m положителни координати, то той може да има няколко различни базиса (допълването до базисна матрица с някои от останалите вектор-стълбове на $\mathbf{A}$ може да става по различни начини).

Възможно е изроден връх $\overline{\mathbf{x}}$ да има огромен брой базиси: ако $\overline{\mathbf{x}}$ има $p < m$ положителни координати, то $\overline{\mathbf{x}}$ може да има

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

различни базиса. Координатите на върха $\overline{\mathbf{x}}$ като стойности са едни и същи, но множествата от координати, означени като базисни и небазисни са различни за всеки базис на $\overline{\mathbf{x}}$.

Типичен пример за изроденост се получава в задачата за назначения. Каноничното множество от допустими решения

$$M_k = \left\{ x_{ij} : \sum_{j=1}^{k} x_{ij}=1, i=1,\dots,k; \sum_{i=1}^{k} x_{ij}=1, j=1,\dots,k; x_{ij}\geqslant 0, i,j=1,\dots,k \right\}$$

на тази 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-те върха на M_8 има по 33 554 432 различни базиса!

От Теорема 2.1 следва и изключително важният факт, че върховете на канонично множество са краен брой.

Следствие 2.1. *Канонично множество* $M = \{\mathbf{x}\in\mathbb{R}^n : \mathbf{A}\mathbf{x}=\mathbf{b}, \mathbf{x}\geqslant\mathbf{0}\}$ *има само краен брой върхове.*

Доказателство. На всеки неизроден връх съответства една базисна матрица, а на всеки изроден е възможно да съответства и повече от една базисна матрица. Следователно

$$\text{(броят върхове)} \leqslant \text{(броят базисни матрици)}.$$

Всяка базисна матрица се състои от m линейно независими стълба, избрани сред n-те стълба на $\mathbf{A}$. Има краен брой начини, по които могат да бъдат избрани m от n обекта и той е $\binom{n}{m} = n!/m!(n-m)!$. Следователно,

$$\text{(броят базисни матрици)} \leqslant \binom{n}{m}$$

и върховете на M са най-много $\binom{n}{m}$ на брой. □

§3. Канонично многостенно множество. Посоки. Теорема за представяне на канонично многостенно множество

Нека $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ е канонично многостенно множество. Върховете на M са важни негови елементи, тъй като (както ще покажем в Следствие 3.2 по-долу), ако M е ограничено множество (т.е. M е многостен), то произволна точка от M се представя като изпъкнала комбинация на върховете на M. Това означава, че ограничено канонично множество може да бъде построено на базата на краен брой свои елементи (върховете му), като се вземат всичките им изпъкнали комбинации.

Когато каноничното множество е неограничено обаче, върховете му не са достатъчни за получаване на всичките му точки. В този случай за целта освен върховете на множеството са необходими и посоките в него.

§3.1. Посоки в канонично множество M. Алгебрична характеризация на посоките в M

Най-напред ще въведем термина посока за произволно множество.

Дефиниция 3.1. *Посока* в произволно множество $S \subset \mathbb{R}^n$ е ненулев вектор $\mathbf{d} \in \mathbb{R}^n$, такъв че за всяка точка $\mathbf{x}_0 \in S$ лъчът $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{x}_0 + \mu\mathbf{d},\ \mu \geqslant 0\}$ лежи изцяло в S.

От Дефиниция 3.1 е ясно, че ако $\mathbf{d}$ е посока в множество S, то $\nu\mathbf{d}$ за $\nu > 0$ също е посока в S.

Едно множество $K \subset \mathbb{R}^n$ се нарича *конус с връх в* $\mathbf{0}$, ако за всяко $\mathbf{d} \in K$ в K се съдържа и $\nu\mathbf{d}$ за произволно $\nu > 0$.

Следователно множеството от всички посоки в S е конус с връх в $\mathbf{0}$ и поради това се нарича *конус на посоките в* S. Ако S е изпъкнало множество, то лесно се доказва, че и конусът на посоките в S е изпъкнало множество, т.е. е *изпъкнал конус*. Тъй като M е изпъкнало множество (вж. Следствие 1.2), то посоките в M са изпъкнал конус с връх в $\mathbf{0}$. Това може да се покаже и директно като се използва характеризацията на посоките в M, дадена в Теорема 3.1 по-долу.

Посоки в многостенното множество P, изобразено на фиг. 0.1 например са направляващите вектори $\mathbf{d}_1(-2, 1)$ и $\mathbf{d}_2(-3, 4)$ на неограничените ръбове на многоъгълника. Изпъкналият конус на посоките на това множество е $K := \{\mu_1(-2, 1) + \mu_2(-3, 4)\ :\ \mu_1 \geqslant 0, \mu_2 \geqslant 0\}$.

Изглежда очевидно, а и в Следствие 3.1 по-долу ще го докажем, че едно канонично множество M е неограничено тогава и само тогава, когато в M има посока. Следователно за канонични задачи, в които допустимото множество е неограничено, ще ни бъде необходима алгебрична характеризация на посоките в M.

Теорема 3.1 (алгебрична характеризация на посоките в M). *Вектор $\mathbf{d} \neq \mathbf{0}$ е посока в M тогава и само тогава, когато*

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

Доказателство. Нека $\mathbf{d} \neq \mathbf{0}$ е посока в M. За фиксирана точка $\mathbf{x}_0 \in M$ имаме, че $\mathbf{x}_0 + \mu\mathbf{d} \in M$ за всяко положително μ. Това означава че за всяко $\mu > 0$ е изпълнено $\mathbf{A}(\mathbf{x}_0 + \mu\mathbf{d}) = \mathbf{b}$ и тъй като $\mathbf{Ax}_0 = \mathbf{b}$, то $\mu\mathbf{Ad} = \mathbf{0}$, откъдето $\mathbf{Ad} = \mathbf{0}$. Ако допуснем, че сред координатите на $\mathbf{d}$ има отрицателна, за големи μ съответната ѝ координата на вектора $\mathbf{x}_0 + \mu\mathbf{d}$ също ще бъде отрицателна, което противоречи на това, че $\mathbf{x}_0 + \mu\mathbf{d} \in M$ и следователно има само неотрицателни координати.

Обратно, ако $\mathbf{d} \neq \mathbf{0}$ удовлетворява условията на твърдението, то за произволна точка $\mathbf{x}_0 \in M$ имаме, че $\mathbf{A}(\mathbf{x}_0 + \mu\mathbf{d}) = \mathbf{Ax}_0 + \mu\mathbf{Ad} = \mathbf{Ax}_0 = \mathbf{b}$ и $\mathbf{x}_0 + \mu\mathbf{d} \geqslant \mathbf{0}$ за всяко неотрицателно μ. Съгласно Дефиниция 3.1, $\mathbf{d}$ е посока в M. □

Използваме алгебричните характеризации на върховете (Теорема 1.1) и на посоките (Теорема 3.1) в канонично множество M, за да докажем

§3.2. Теорема за представяне на канонично множество M

Да означим с $V := \{\overline{\mathbf{x}}_i : i \in I\}$ множеството от върховете на M. Да обърнем внимание на това, че индексното множество I е крайно множество, тъй като съгласно Следствие **??** M има краен брой върхове.

Теорема 3.2 (за представяне). *Всяка точка $\mathbf{x} \in M$ може да се представи във вида*

$$\mathbf{x} = \sum_{i \in I} \lambda_i \overline{\mathbf{x}}_i + \mathbf{d}\,, \tag{3.1}$$

където $\lambda_i \geqslant 0$ за всички $i \in I$, $\sum_{i \in I} \lambda_i = 1$ и $\mathbf{d}$ е посока в M или $\mathbf{d}$ е нулевият вектор.

Доказателство. Твърдението ще докажем с индукция по броя на положителните координати на точките в M.

Нека p_0 е най-малкият възможен брой положителни координати на точка в M.

Ще покажем, че всяка точка в M с p_0 положителни координати е връх на M.

Наистина, ако $p_0 = 0$, то нулевият вектор $\mathbf{0} \in M$. В този случай $\mathbf{0}$ е връх на M (вж. доказателството на Теорема 1.1).

Нека сега $p_0 > 0$ и да допуснем, че в M има точка $\mathbf{x}$ с p_0 положителни координати, която не е връх. Съгласно Теорема 1.1 вектор-стълбовете на $\mathbf{A}$, съответстващи на положителните ѝ координати са линейно зависими. Тогава, съгласно Лема 1.1 съществува вектор $\mathbf{z} \neq \mathbf{0}$, за който $z_i = 0$, ако $x_i = 0$ и числа $t^- < 0 < t^+$, такива че $\mathbf{x} + t\mathbf{z} \in M$ за всяко $t \in [t^-, t^+]$ като поне единият край на интервала е краен. Нека t^- е крайно. Точката $\mathbf{x}' = \mathbf{x} + t^-\mathbf{z}$ е точка от M, която обаче има поне една нулева координата повече от $\mathbf{x}$ (наистина, ако $t^- = -x_r/z_r$, то $x_r \neq 0$, докато $x'_r = x_r + t^- z_r = x_r + \dfrac{-x_r}{z_r} z_r = 0$) а това е невъзможно съгласно дефиницията на p_0.

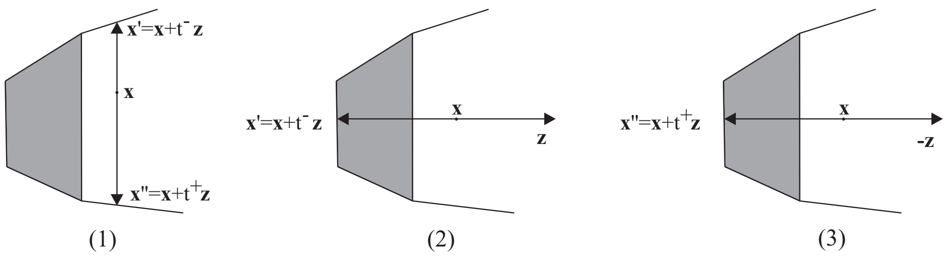
Представянето (3.1) очевидно е в сила за точки, които са върхове на M, тъй като ако $\mathbf{x}$ е връх, то $\mathbf{x} = \overline{\mathbf{x}}_i$ за някое $i \in I$.

Следователно точките в M с минимален брой положителни координати имат представянето (3.1).

Да предположим, че точките в M с по-малко от p положителни координати имат представянето (3.1).

Да разгледаме сега точка $\mathbf{x} \in M$, която има точно p положителни координати.

Ако $\mathbf{x}$ е връх, казахме, че за него представянето е в сила. Ако $\mathbf{x}$ не е връх, то съгласно Теорема 1.1 вектор-стълбовете на $\mathbf{A}$, съответстващи на положителните му координати са линейно зависими. Тогава, съгласно Лема 1.1 съществува вектор $\mathbf{z} \neq \mathbf{0}$, за който $z_i = 0$ ако $x_i = 0$, $\mathbf{Az} = \mathbf{0}$ и числа $t^- < 0 < t^+$, такива че $\mathbf{x} + t\mathbf{z} \in M$ за всяко $t \in [t^-, t^+]$. За вектора $\mathbf{z}$ са възможни три случая, илюстрирани на фиг. 3.4.



Фигура 3.4.

Случай 1. Векторът $\mathbf{z}$ има положителни и отрицателни координати. В този случай t^- и t^+ приемат крайни стойности. Точките $\mathbf{x}' = \mathbf{x} + t^-\mathbf{z}$ и $\mathbf{x}'' = \mathbf{x} + t^+\mathbf{z}$ са точки от M, които лежат върху правата през $\mathbf{x}$, определена от $\mathbf{z}$, и имат поне една нулева координата повече от $\mathbf{x}$. Следователно, $\mathbf{x}'$ и $\mathbf{x}''$ са точки от M с по-малко от p положителни координати. Съгласно индукционното предположение те имат желаното представяне

$$\mathbf{x}' = \sum_{i \in I} \lambda'_i \overline{\mathbf{x}}_i + \mathbf{d}', \text{ където } \lambda'_i \geqslant 0,\ i \in I,\ \sum_{i \in I} \lambda'_i = 1,\ \mathbf{d}' \geqslant \mathbf{0},\ \mathbf{Ad}' = \mathbf{0};$$

$$\mathbf{x}'' = \sum_{i \in I} \lambda''_i \overline{\mathbf{x}}_i + \mathbf{d}'', \text{ където } \lambda''_i \geqslant 0,\ i \in I,\ \sum_{i \in I} \lambda''_i = 1,\ \mathbf{d}'' \geqslant \mathbf{0},\ \mathbf{Ad}'' = \mathbf{0}.$$

Точката $\mathbf{x}$ лежи на отсечката с краища $\mathbf{x}'$ и $\mathbf{x}''$ и следователно е тяхна изпъкнала комбинация, т.е. съществува $\lambda \in (0, 1)$, такова че

$$\mathbf{x} = \lambda \mathbf{x}' + (1 - \lambda)\mathbf{x}''.$$

Стойността на λ намираме от

$$\mathbf{x} = \lambda\mathbf{x}' + (1-\lambda)\mathbf{x}'' = \lambda(\mathbf{x} + t^-\mathbf{z}) + (1-\lambda)(\mathbf{x} + t^+\mathbf{z})$$
$$= \mathbf{x} + (\lambda t^- + (1-\lambda)t^+)\mathbf{z}.$$

Получаваме $(\lambda t^- + (1-\lambda)t^+)\mathbf{z} = \mathbf{0}$, но $\mathbf{z}$ е ненулев вектор и следователно $\lambda t^- + (1-\lambda)t^+ = 0$. Оттук получаваме $\lambda = t^+/(t^+ - t^-)$.

Като отразим това, че точките $\mathbf{x}'$ и $\mathbf{x}''$ имат желаното представяне,

$$\mathbf{x} = \lambda\mathbf{x}' + (1-\lambda)\mathbf{x}'' = \lambda\left(\sum_{i\in I}\lambda_i'\overline{\mathbf{x}}_i + \mathbf{d}'\right) + (1-\lambda)\left(\sum_{i\in I}\lambda_i''\overline{\mathbf{x}}_i + \mathbf{d}''\right)$$
$$= \sum_{i\in I}\left(\lambda\lambda_i' + (1-\lambda)\lambda_i''\right)\overline{\mathbf{x}}_i + \lambda\mathbf{d}' + (1-\lambda)\mathbf{d}''$$

и като положим

$$\lambda_i := \lambda\lambda_i' + (1-\lambda)\lambda_i'' \quad \text{за всяко } i \in I \ \text{ и } \quad \mathbf{d} := \lambda\mathbf{d}' + (1-\lambda)\mathbf{d}''$$

получаваме, че

$$\mathbf{x} = \sum_{i\in I}\lambda_i\overline{\mathbf{x}}_i + \mathbf{d},$$

където

$$\lambda_i \geqslant 0 \text{ за } i \in I,$$
$$\sum_{i\in I}\lambda_i = \sum_{i\in I}(\lambda\lambda_i' + (1-\lambda)\lambda_i'') = \lambda\sum_{i\in I}\lambda_i' + (1-\lambda)\sum_{i\in I}\lambda_i'' = \lambda + (1-\lambda) = 1,$$
$$\mathbf{d} \geqslant \mathbf{0} \text{ и } \mathbf{A}\mathbf{d} = \mathbf{A}(\lambda\mathbf{d}' + (1-\lambda)\mathbf{d}'') = \lambda\mathbf{A}\mathbf{d}' + (1-\lambda)\mathbf{A}\mathbf{d}'' = \mathbf{0}.$$

От последното е ясно, че $\mathbf{d}$ е нулевият вектор или посока в M и следователно $\mathbf{x}$ има желаното представяне.

Случай 2. Векторът $\mathbf{z} \geqslant \mathbf{0}$. В този случай само t^- е крайно. Дефинираме $\mathbf{x}'$, както в случай 1. Точката $\mathbf{x}$ записваме като

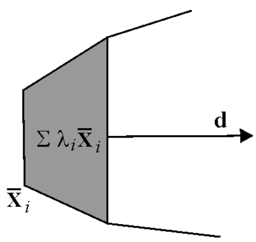
$$\mathbf{x} = \mathbf{x}' + (-t^-)\mathbf{z}, \quad \text{където} \ \ t^- < 0.$$

Тъй като според индукционното предположение $\mathbf{x}'$ има желаното представяне и тъй като $\mathbf{z}$ е посока в M (понеже $\mathbf{z} \geqslant \mathbf{0}$ и $\mathbf{A}\mathbf{z} = \mathbf{0}$), то очевидно $\mathbf{x}$ има представянето (3.1).

Случай 3. Векторът $\mathbf{z} \leqslant \mathbf{0}$. Доказателството е аналогично на това в случай 2, като заменим $\mathbf{x}'$, t^- и $\mathbf{z}$ съответно с $\mathbf{x}''$, t^+ и $-\mathbf{z}$. □

От Теоремата за представяне е ясно, че за представянето на произволен елемент от канонично множество M, са необходими както изпъкналите комбинации на върховете на M, така и посоките в M, ако има такива (вж. фиг. 3.5).

Отговор на въпроса кога в M има посока дава



Фигура 3.5.

Следствие 3.1. *Непразно канонично множество M е неограничено тогава и само тогава, когато в M има посока.*

Доказателство. Нека в M има посока $\mathbf{d} \neq \mathbf{0}$. Тъй като $M \neq \emptyset$, в M има поне една точка $\mathbf{x}_0$. Съгласно дефиницията за посока, в M се съдържа и лъчът $\{\mathbf{x}_0 + \mu\mathbf{d},\ \mu \geqslant 0\}$, който е неограничено множество. Следователно M е неограничено множество.

Обратно, нека M е неограничено. Да допуснем, че в M няма посока. От Теоремата за представяне имаме, че произволен елемент $\mathbf{x} \in M$ се представя във вида

$$\mathbf{x} = \sum_{i\in I} \lambda_i \overline{\mathbf{x}}_i, \quad \text{където } \lambda_i \geqslant 0,\ i \in I \quad \text{и } \sum_{i\in I} \lambda_i = 1.$$

Да означим $k := \max_{i\in I} \|\overline{\mathbf{x}}_i\|$. Като максимум на краен брой числа константата k е добре дефинирана и от представянето на $\mathbf{x}$ имаме, че

$$\|\mathbf{x}\| = \left\| \sum_{i\in I} \lambda_i \overline{\mathbf{x}}_i \right\| \leqslant \sum_{i\in I} \lambda_i \|\overline{\mathbf{x}}_i\| \leqslant k \sum_{i\in I} \lambda_i = k,$$

което означава, че нормата на произволен вектор в M не надминава k, т.е. M се съдържа в кълбо с център $\mathbf{0}$ и радиус k, и следователно M е ограничено множество. Полученото противоречие доказва, че ако M е неограничено, то в M има посока. $\square$

Следствие 3.2. *Ако M е непразно и ограничено канонично множество (т.е. M е многостен), то всяко $\mathbf{x} \in M$ се представя като изпъкнала комбинация на върховете на M.*

Доказателство. От Теоремата за представяне в представянето (3.1) на произволен елемент $\mathbf{x}$ на M ще имаме $\mathbf{d} = \mathbf{0}$, тъй като M е ограничено и според Следствие 3.1 в M няма посока. $\square$

§4. Основни теореми на линейното оптимиране

Разглеждаме каноничната задача на линейното оптимиране

$$
(K) \qquad \begin{array}{ll} \min & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{array}
$$

където $\mathbf{A}$ е $m \times n$ матрица, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m_+$ и $\mathbf{x} \in \mathbb{R}^n$ е n-мерен вектор на променливите.

Ще докажем две теореми, които са в основата на разработването на алгоритми (и в частност на симплекс алгоритъма) за нейното решаване. Тези теореми разкриват значението, което върховете на допустимото множество на задачата $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ имат за тези методи.

Теорема 4.1. *Ако M е непразно множество, то M има поне един връх.*

Доказателство. Тъй като $M \neq \emptyset$, то в M има поне една точка $\mathbf{x}$. Тя има представянето (3.1), доказано в Теорема 3.2. Ако допуснем, че множеството $V = \{\overline{\mathbf{x}}_i, i \in I\}$ от върховете на M е празно, то от Теоремата за представяне $\mathbf{x} = \mathbf{d}$, където $\mathbf{d} \geqslant \mathbf{0}$, $\mathbf{A}\mathbf{d} = \mathbf{0}$. Тъй като $\mathbf{x}$ е в M, то $\mathbf{b} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{d} = \mathbf{0}$, т.е. $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{0}, \mathbf{x} \geqslant \mathbf{0}\}$ и следователно нулевият вектор $\mathbf{0} \in M$. Но $\mathbf{0}$ е връх на M, откъдето получаваме противоречие с допускането, че множеството от върховете на M е празно. $\square$

Теорема 4.2. *Ако M е непразно множество, то или функцията $z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}$ е неограничена отдолу върху M, или минималната ѝ стойност върху M се достига в някой от върховете на M.*

Доказателство. Има два възможни случая:

Случай 1. В M има посока $\mathbf{d}$, която сключва тъп ъгъл с вектора на целевата функция $\mathbf{c}$, т.е. такава, че $\mathbf{c}^{\mathrm{T}}\mathbf{d} < 0$.

В този случай M е неограничено множество и стойностите на z намаляват неограничено върху M. Наистина, ако вземем произволна точка $\mathbf{x}$ от непразното множество M, то по дефиницията за посока всички точки от лъча $\{\mathbf{x} + \mu\mathbf{d} \ : \ \mu \geqslant 0\}$ са в M. Стойностите на z върху точките от този допустим лъч са $z(\mathbf{x} + \mu\mathbf{d}) = \mathbf{c}^{\mathrm{T}}(\mathbf{x} + \mu\mathbf{d}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} + \mu\mathbf{c}^{\mathrm{T}}\mathbf{d}$ и очевидно клонят към $-\infty$, когато $\mu \to +\infty$, поради това, че $\mathbf{c}^{\mathrm{T}}\mathbf{d} < 0$.

Случай 2. За всички посоки $\mathbf{d}$ в M е изпълнено, че $\mathbf{c}^{\mathrm{T}}\mathbf{d} \geqslant 0$.

Да вземем произволна точка $\mathbf{x} \in M$. Според Теоремата за представяне съществуват числа $\lambda_i \geqslant 0$ за $i \in I$, такива че $\sum\limits_{i \in I} \lambda_i = 1$ и вектор $\mathbf{d}$, който е посока в M или $\mathbf{d} = \mathbf{0}$, такива че

$$
\mathbf{x} = \sum_{i \in I} \lambda_i \overline{\mathbf{x}}_i + \mathbf{d},
$$

където $V = \{\overline{\mathbf{x}}_i, i \in I\}$ е множеството от върховете на M.

За стойността на z в $\mathbf{x}$ имаме

$$\begin{aligned} z(\mathbf{x}) &= \mathbf{c}^{\mathrm{T}}\left(\sum_{i\in I}\lambda_i\overline{\mathbf{x}}_i + \mathbf{d}\right) = \sum_{i\in I}\lambda_i\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}_i + \mathbf{c}^{\mathrm{T}}\mathbf{d} \\ &\geqslant \sum_{i\in I}\lambda_i\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}_i \geqslant \sum_{i\in I}\lambda_i \min_{i\in I}\{\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}_i\} = \min_{i\in I}\{\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}_i\}\sum_{i\in I}\lambda_i \\ &= \min_{i\in I}\{\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}_i\} = \mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}_{i_0} = z(\overline{\mathbf{x}}_{i_0}), \end{aligned}$$

което означава, че минимумът на z се достига във върха $\overline{\mathbf{x}}_{i_0}$ на M. $\square$

Този резултат стои в основата на решаването на линейни задачи. Той показва, че ако в M има посока $\mathbf{d}$, която сключва тъп ъгъл с вектора на целевата функция $\mathbf{c}$, задачата (K) е неограничена, а в противен случай като кандидати за оптимално решение е достатъчно да бъдат разгледани само върховете на M.

§5. Симплекс метод. Геометрична мотивация. Описание на стъпките на симплексната итерация

От Теорема 4.2 е ясно, че като кандидати за оптимално решение на каноничната задача

$$(K) \qquad \begin{aligned} \min \quad & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

където $\mathbf{A}$ е $m \times n$ матрица, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m_+$ и $\mathbf{x} \in \mathbb{R}^n$ е n-мерен вектор на променливите е достатъчно да разгледаме само върховете на допустимото ѝ множество $M = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$. За големи стойности на m и n намирането на всички върхове на M е непрактично, тъй като M може да има $\binom{n}{m}$ на брой върхове, а $\binom{n}{m}$ расте много бързо при нарастването на n и m. Следователно за решаването на (K) е необходим по-систематичен подход. Такъв подход прилага симплекс методът, разработен от Джордж Данциг през 1947 г. На практика симплекс методът се оказва толкова успешен, че и до момента е един от най-известните и най-широко използваните методи за компютърно реализиране на числените пресмятания при решаване на линейни задачи.

За начало ще изложим идеята на симплекс метода, която има ясна геометрична мотивация.

§5.1. Описание на симплекс метода (СМ)

Симплекс методът има две фази.

Първата фаза (фаза I) или установява, че допустимото множество M е празно, или намира връх на M (Теорема 4.1 гарантира съществуването на поне един връх, ако M е непразно множество). Намереният връх служи за начало на втората фаза на метода и поради това се нарича още *начален връх*.

Втората фаза (фаза II) е същинската част на метода. Намирайки се в текущия връх $\overline{\mathbf{x}}$, методът търси ръб на M, излизащ от $\overline{\mathbf{x}}$, по който целевата функция z намалява.

Ако СМ установи, че по всички ръбове на M, излизащи от $\overline{\mathbf{x}}$ целевата функция z нараства, то $\overline{\mathbf{x}}$ е оптимално решение и следва край на алгоритъма.

Ако СМ намери неограничен ръб на M, излизащ от $\overline{\mathbf{x}}$, по който целевата функция z намалява, то задачата (K) е неограничена.

Ако СМ намери ограничен ръб на M, излизащ от $\overline{\mathbf{x}}$, по който целевата функция z намалява, той отива в другия му край, който е връх $\overline{\mathbf{x}}'$ на M, в който $z(\overline{\mathbf{x}}') < z(\overline{\mathbf{x}})$.

По този начин, започвайки от началния връх, методът се движи от връх на връх по ръбовете на M, като движението се извършва само по ръбове, по които целевата функция z намалява. Следователно ако методът напусне даден връх, той никога не се връща в него и след краен брой итерации (тъй като според Следствие 2.1 M има краен брой върхове) или се достига до връх на M, който е оптимално решение на задачата (K), или се намира неограничен ръб в допустимото множество M, по който целевата функция z намалява неограничено.

По-нататък целта ни е да намерим съответната алгебрична форма на горното просто геометрично описание на симплекс метода.

Най-напред ще разгледаме същинската фаза на СМ — фаза II, която намира оптимално решение, като предположим, че фаза I вече е намерила начален връх. Както ще се убедим в § **??**, симплекс методът успешно може да бъде приложен и за решаването на задачата на фаза I – намирането на началния връх.

Предполагаме че редовете на матрицата $\mathbf{A}$ са линейно независими, т.е., че $r(\mathbf{A}) = m$ и че $m < n$.

За да опишем една итерация на симплекс метода, да предположим, че $\overline{\mathbf{x}}$ е текущият връх. Нека базисната му матрица $\mathbf{B}$ се състои от първите m вектор-стълба на $\mathbf{A}$ (нещо което винаги бихме постигнали чрез преномериране на променливите), т.е. базисът е $B = \{1, \dots, m\}$, а множеството от небазисните индекси е $N = \{m+1, \dots, n\}$. С направените означения, върхът $\overline{\mathbf{x}}$ на M е $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix} \geqslant \mathbf{0}$.

§5.2. Базисно представяне на задачата (K) спрямо базиса B

Нека $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix} \geqslant \mathbf{0}$ е връх на M с базис B. Задачата (K) се представя в *базисен вид спрямо базиса* B по следния начин:

1. Изразяваме базисните координати $\mathbf{x}_B$ на решение $\mathbf{x}$ на системата $\mathbf{Ax} = \mathbf{b}$ чрез небазисните му координати $\mathbf{x}_N$ (както в § 2.1)

$$\mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad [\mathbf{B}|\mathbf{N}] \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{Bx}_B + \mathbf{Nx}_N = \mathbf{b} \quad \Leftrightarrow$$

$$\Leftrightarrow \mathbf{B}^{-1}|\mathbf{Bx}_B + \mathbf{Nx}_N = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{x}_B + \mathbf{B}^{-1}\mathbf{Nx}_N = \mathbf{B}^{-1}\mathbf{b} \quad \Leftrightarrow$$
$$\mathbf{x}_B = \overline{\mathbf{x}}_B - \mathbf{B}^{-1}\mathbf{Nx}_N.$$

Така за произволно решение $\mathbf{x}$ на системата $\mathbf{Ax} = \mathbf{b}$ получаваме представянето

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{x}}_B - \mathbf{B}^{-1}\mathbf{Nx}_N \\ \overline{\mathbf{x}}_N + \mathbf{Ix}_N \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} + \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{bmatrix} \mathbf{x}_N =$$

$$\overline{\mathbf{x}} + \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{bmatrix} \mathbf{x}_N,$$

като с $\mathbf{I}$ е означена единичната матрица от ред $n-m$.

Нека за всеки небазисен индекс $j \in N$ (т.е. за $j > m$) означим с $\mathbf{d}_j$ съответния стълб на матрицата $\begin{bmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{bmatrix}$, т.е.

$$\mathbf{d}_j := \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_j \\ \mathbf{e}_j \end{bmatrix}, \tag{5.1}$$

където $\mathbf{A}_j$ е j-ият стълб на матрицата $\mathbf{A}$, а $\mathbf{e}_j$ е $(n-m)$-мерен вектор, чиято $(j-m)$-та координата е 1, а останалите му координати са нули. След въведеното означение $\begin{bmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{bmatrix} \mathbf{x}_N = \sum_{j\in\mathbf{N}} x_j \mathbf{d}_j$ и

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j\in N} x_j \mathbf{d}_j. \tag{5.2}$$

2. Като използваме представянето (5.2) на решение $\mathbf{x}$ на системата, за стойността на целевата функция z в $\mathbf{x}$ получаваме

$$z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} = \mathbf{c}^{\mathrm{T}}\left(\overline{\mathbf{x}} + \sum_{j\in N} x_j \mathbf{d}_j\right) = \mathbf{c}^{\mathrm{T}}(\overline{\mathbf{x}}) + \sum_{j\in N} x_j \mathbf{c}^{\mathrm{T}}\mathbf{d}_j.$$

Нека за всеки небазисен индекс $j \in N$ означим с $\overline{c}_j$ скаларното произведение $\mathbf{c}^{\mathrm{T}}\mathbf{d}_j$. Имаме

$$\overline{c}_j = \mathbf{c}^{\mathrm{T}}\mathbf{d}_j = [\mathbf{c}_B^{\mathrm{T}}, \mathbf{c}_N^{\mathrm{T}}] \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_j \\ \mathbf{e}_j \end{bmatrix} = c_j - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_j.$$

Ако използваме същата формула, за да пресметнем $\overline{c}_i$ за базисен индекс $i \in B$, получаваме

$$\overline{c}_i = c_i - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_i = c_i - \mathbf{c}_B^{\mathrm{T}}\mathbf{e}_i = c_i - c_i = 0.$$

За всеки индекс $j \in \{1,\dots,n\}$ така полученото число $\overline{c}_j$ се нарича *относителна оценка* или още *редуцирана цена* на променливата x_j спрямо базиса B. Векторът $\overline{\mathbf{c}}^{\mathrm{T}} = [\overline{\mathbf{c}}_B^{\mathrm{T}}, \overline{\mathbf{c}}_N^{\mathrm{T}}] = [\mathbf{0}^{\mathrm{T}}, \overline{\mathbf{c}}_N^{\mathrm{T}}]$, чиито координати са относителните оценки на променливите спрямо базиса B, се нарича *вектор на относителните оценки спрямо базиса* B. Да забележим, че относителните оценки на базисните променливи са нули.

След направеното полагане имаме

$$z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j\in N} \overline{c}_j x_j.$$

3. Като отразим направените в 1 и 2 преобразувания и полагания в задачата (K), получаваме еквивалентната на нея задача

$$(K_B) \qquad \begin{aligned} &\min z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j\in N} \overline{c}_j x_j \\ &\mathbf{x} = \overline{\mathbf{x}} + \sum_{j\in N} x_j \mathbf{d}_j \\ &\mathbf{x} \geqslant \mathbf{0}, \end{aligned}$$

   в която решенията на системата и стойностите на целевата функция в тях са изразени само посредством небазисните променливи x_j, $j \in N$.

Представянето на задачата (K) във вида (K_B) наричаме *базисно представяне спрямо базиса* B.

§5.3. Какво представляват $\mathbf{d}_j$ и $\overline{c}_j$ в базисния вид на (K)

Казахме че, намирайки се в текущия връх $\overline{\mathbf{x}}$, симплекс методът решава по кой от ръбовете на допустимото множество M, излизащи от $\overline{\mathbf{x}}$, да тръгне. При направените предположения $M \neq \emptyset$, $r(\mathbf{A}) = m$, от всеки връх на M излизат точно $n-m$ ръба (размерността на M е $n-m$).

Върхът $\overline{\mathbf{x}}$ на M се намира в сечението на m-те хиперравнини, определени с уравненията

$$\mathbf{a}_1^{\mathrm{T}}\mathbf{x} = b_1, \ldots, \mathbf{a}_m^{\mathrm{T}}\mathbf{x} = b_m, \tag{5.3}$$

и $(n-m)$-те хиперравнини, определени с уравненията

$$\mathbf{e}_{m+1}^{\mathrm{T}}\mathbf{x} = 0, \ldots, \mathbf{e}_n^{\mathrm{T}}\mathbf{x} = 0, \tag{5.4}$$

които се удовлетворяват в $\overline{\mathbf{x}}$ поради това, че $\overline{x}_j = 0$ за $j \in N$, т.е. $(n-m)$-те му небазисни координати са равни на нула.

Да разгледаме матрицата

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

чиито вектор-редове са нормалните вектори на тези n хиперравнини. Тъй като $\mathbf{B}$ е неособена матрица, редовете на матрицата $\mathbf{M}$ са линейно независими и следователно $\mathbf{M}$ също е неособена.

Следователно, върхът $\overline{\mathbf{x}}$ е единствената точка в сечението на n-те линейно независими хиперравнини (5.3) и (5.4) и $\overline{\mathbf{x}} \in M$.

Всеки ръб на M, излизащ от $\overline{\mathbf{x}}$ се получава като сечение на M с m-те хиперравнини (5.3) и $n-m-1$ от хиперравнините (5.4) – (общо $n-1$ хиперравнини). Сечението на $(n-1)$-те линейно независими хиперравнини е права през $\overline{\mathbf{x}}$, а сечението на правата с изпъкналото множество M може да е точката $\overline{\mathbf{x}}$, отсечка или лъч (не е възможно сечението да е цялата права, защото в този случай $\overline{\mathbf{x}}$ няма да бъде връх). Ако сечението е само точката $\overline{\mathbf{x}}$, имаме *фиктивен ръб*, в противен случай ръбът е *действителен* като ако е отсечка е *ограничен ръб*, а ако е лъч е *неограничен ръб*, излизащ от $\overline{\mathbf{x}}$.

Базисът B определя $n-m$ ръба, излизащи от $\overline{\mathbf{x}}$. Тези ръбове се представят във вида

$$\{\overline{\mathbf{x}}+t\mathbf{d}_j, t\geqslant 0\}\subset M, \quad \mathbf{d}_j=\begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_j \\ \mathbf{e}_j \end{bmatrix},\ j\in N,$$

т.е. векторите $\mathbf{d}_j$, $j\in N$ от базисния вид на задачата (K_B) са направляващи вектори по ръбовете, излизащи от $\overline{\mathbf{x}}$. За да се убедим в това да пресметнем за $q\in N$

$$\mathbf{M}\mathbf{d}_q=\begin{bmatrix} \mathbf{B} & \mathbf{N} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_q \\ \mathbf{e}_q \end{bmatrix}=\begin{bmatrix} -\mathbf{B}\mathbf{B}^{-1}\mathbf{A}_q+\mathbf{A}_q \\ \mathbf{e}_q \end{bmatrix}=\begin{bmatrix} \mathbf{0} \\ \mathbf{e}_q \end{bmatrix}, \tag{5.5}$$

което означава, че векторът $\mathbf{d}_q$ е ортогонален на нормалните вектори на хиперравнините (5.3) и (5.4) с изключение на хиперравнината $\mathbf{e}_q^{\mathrm{T}}\mathbf{x}=0$, т.е. векторът $\mathbf{d}_q$ лежи на правата, която е сечение на тези $n-1$ линейно независими хиперравнини.

Остава да видим кои са допустимите точки от тази права, т.е. точките от нея, които са в M. Да фиксираме $q\in\mathbf{N}$ и да разгледаме точките от вида

$$\mathbf{x}(t)=\overline{\mathbf{x}}+t\mathbf{d}_q, \tag{5.6}$$

където t е реален параметър. Съгласно (5.5) $\mathbf{A}\mathbf{d}_q=\mathbf{0}$ и

$$\mathbf{A}\mathbf{x}(t)=\mathbf{A}(\overline{\mathbf{x}}+t\mathbf{d}_q)=\mathbf{A}\overline{\mathbf{x}}+t\mathbf{A}\mathbf{d}_q=\mathbf{A}\overline{\mathbf{x}}=\mathbf{b} \tag{5.7}$$

т.е. за произволно реално число t векторът $\mathbf{x}(t)$ е решение на системата $\mathbf{A}\mathbf{x}=\mathbf{b}$. Следователно, $\mathbf{x}(t)\in M$ тогава и само тогава, когато $\mathbf{x}(t)\geqslant\mathbf{0}$. Да положим

$$\mathbf{w}_q:=\mathbf{B}^{-1}\mathbf{A}_q \tag{5.8}$$

и да изразим координатите на

$$\mathbf{x}(t)=\overline{\mathbf{x}}+t\mathbf{d}_q=\begin{bmatrix} \overline{\mathbf{x}}_B \\ \mathbf{0} \end{bmatrix}+t\begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_q \\ \mathbf{e}_q \end{bmatrix}=\begin{bmatrix} \overline{\mathbf{x}}_B-t\mathbf{B}^{-1}\mathbf{A}_q \\ t\mathbf{e}_q \end{bmatrix}=\begin{bmatrix} \overline{\mathbf{x}}_B-t\mathbf{w}_q \\ t\mathbf{e}_q \end{bmatrix}. \tag{5.9}$$

Векторът $\mathbf{x}(t)\geqslant\mathbf{0}$ ако $\overline{\mathbf{x}}_B-t\mathbf{w}_q\geqslant\mathbf{0}$ и ако $t\mathbf{e}_q\geqslant\mathbf{0}$. Второто векторно неравенство води до $t\geqslant 0$ поради $\mathbf{e}_q\geqslant\mathbf{0}$. Първото векторно неравенство е еквивалентно на системата от m числови неравенства $\overline{x}_i-tw_{iq}\geqslant 0$, $i\in B$, където с w_{iq} е означена i-та координата на вектора $\mathbf{w}_q$. Ако $w_{iq}\leqslant 0$, i-то неравенство очевидно е в сила, а ако $w_{iq}>0$ трябва $t\leqslant\frac{\overline{x}_i}{w_{iq}}$.

Ако векторът $\mathbf{w}_q$ има положителни координати от *теста за минимално отношение* полагаме

$$\overline{t}:=\min\left\{\frac{\overline{x}_i}{w_{iq}}:\ w_{iq}>0, i\in B\right\},$$

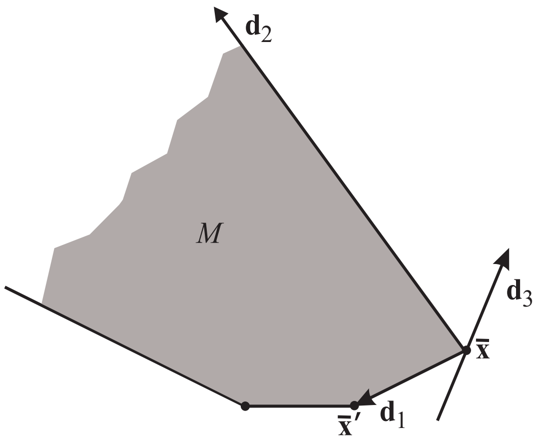
а ако векторът $\mathbf{w}_q$ няма положителни координати полагаме $\bar{t} = +\infty$. За всяко $t \in [0, \bar{t}]$ имаме, че $\mathbf{x}(t) \geqslant \mathbf{0}$, или $\mathbf{x}(t) \in M$.

Ако $\mathbf{w}_q \leqslant \mathbf{0}$, то $\bar{t} = +\infty$ и ръбът с направление $\mathbf{d}_q$ е неограничен. В този случай $\mathbf{d}_q \geqslant \mathbf{0}$, $\mathbf{A}\mathbf{d}_q = \mathbf{0}$ и $\mathbf{d}_q$ е посока в M.

Ако $\mathbf{w}_q$ има положителна координата, има две възможности: или $\bar{t} > 0$ и ръбът с направление $\mathbf{d}_q$ е ограничен, като другият му край е $\mathbf{x}(\bar{t})$, или $\bar{t} = 0$, което означава че $\overline{\mathbf{x}} = \mathbf{x}(0)$ е единствената допустима точка, т.е. ръбът с направление $\mathbf{d}_q$ е фиктивен.

Да отбележим, че ако $\overline{\mathbf{x}}$ е неизродено базисно допустимо решение (т.е. $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b} > 0$), то всеки от излизащите от него ръбове е действителен. Наистина, в този случай по всяко направление $\mathbf{d}_q$ или $\bar{t} = +\infty$ и имаме неограничен ръб или $\bar{t} > 0$ като минимум на краен брой положителни числа (напомняме, че $\overline{x}_i > 0$ за всяко $i \in B$) и имаме ограничен ръб.

Ако $\overline{\mathbf{x}}$ е изроден връх обаче, възможно е някой от векторите $\mathbf{d}_q$ да определя фиктивен ръб: $\mathbf{d}_q$ определя фиктивен ръб, излизащ от $\overline{\mathbf{x}}$, ако за някой базисен индекс $i \in B$ базисната координата $\overline{x}_i = 0$ като същевременно $w_{iq} > 0$: в този случай $\bar{t} = 0$ и $\overline{\mathbf{x}} = \mathbf{x}(0)$ е единствената допустима точка от вида $\overline{\mathbf{x}} + t\mathbf{d}_q$. В случай на изроден връх за намиране на $(n-m)$-те действителни ръба, излизащи от него трябва да се използват всичките му базиси.



Фигура 5.6.

На фигура 5.6 векторът $\mathbf{d}_1$ определя действителен ограничен ръб на M, излизащ от върха $\overline{\mathbf{x}}$, другият край на който е върхът $\overline{\mathbf{x}}'$, векторът $\mathbf{d}_2$ определя действителен неограничен ръб на M, излизащ от $\overline{\mathbf{x}}$, докато векторът $\mathbf{d}_3$ определя фиктивен ръб.

Както вече казахме, симплекс методът търси сред излизащите от $\overline{\mathbf{x}}$ ръбове такъв, по който целевата функция z намалява. От вида (5.6) на $\mathbf{x}(t)$ е ясно, че движението по ръба с направление $\mathbf{d}_q$ е еквивалентно на увеличаване на стойностите на небазисната променлива x_q, докато стойностите на всички останали небазисни променливи остават фиксирани на нула.

Как се изменя при това движение стойността на целевата функция z? От базисния вид (K_B) на задачата спрямо базиса B на $\overline{\mathbf{x}}$ следва, че за векторите от ръба с направление $\mathbf{d}_q$ ще имаме

$$z(\mathbf{x}(t)) = \mathbf{c}^{\mathrm{T}}\mathbf{x}(t) = \mathbf{c}^{\mathrm{T}}(\overline{\mathbf{x}} + t\mathbf{d}_q) = \mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} + t\mathbf{c}^{\mathrm{T}}\mathbf{d}_q = z(\overline{\mathbf{x}}) + t\overline{c}_q\,,$$

откъдето е ясно как би се изменила целевата функция, ако симплекс методът тръгне по ръба $\mathbf{d}_q$ (т.е. започне да увеличава небазисната променлива x_q):

- ако $\overline{c}_q < 0$, градиентът $\mathbf{c}$ на целевата функция z сключва тъп ъгъл с направлението $\mathbf{d}_q$ на ръба и z ще намалява при движение по този ръб;
- ако $\overline{c}_q > 0$, $\mathbf{c}$ сключва остър ъгъл с $\mathbf{d}_q$ и z ще нараства при движение по този ръб;
- ако $\overline{c}_q = 0$, $\mathbf{c}$ е ортогонален на $\mathbf{d}_q$ и стойностите на z не се изменят при движение по този ръб.

Това изяснява ролята на относителните оценки: относителната оценка $\overline{c}_q$ оценява небазисната променлива x_q спрямо базиса B — дали, намирайки се в базисното допустимо решение $\overline{\mathbf{x}}$ с базис B, увеличаването на небазисната променлива x_q ще доведе до намаляване или до увеличаване на целевата функция.

§5.4. Критерий за оптималност на текущото базисно допустимо решение $\overline{\mathbf{x}}$

От казаното дотук и от базисния вид на задачата (K_B) получаваме

Теорема 5.1. (Критерий за оптималност). *Ако за всички $j \in N$ относителните оценки $\overline{c}_j \geqslant 0$, то базисното допустимо решение $\overline{\mathbf{x}}$ с базис B е оптимално решение на каноничната задача (K).*

Доказателство. За произволна допустима точка $\mathbf{x} \in M$ от базисния вид на задачата (K_B) имаме, че стойността на целевата функция z в $\mathbf{x}$ е

$$z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j\in N} \overline{c}_j x_j. \tag{5.10}$$

Тъй като $x_j \geqslant 0$ за $j \in N$, ако всички относителни оценки $\overline{c}_j \geqslant 0$ за $j \in N$, то сумата в (5.10) ще има неотрицателна стойност и следователно $z(\mathbf{x}) \geqslant z(\overline{\mathbf{x}})$ за произволно $\mathbf{x} \in M$, което означава, че $\overline{\mathbf{x}}$ е оптимално решение на (K). □

Вярно ли е обратното твърдение? Не. Ако $\overline{\mathbf{x}}$ е неизродено оптимално базисно допустимо решение, то тогава $\overline{c}_j \geqslant 0$ за всеки индекс $j \in N$, но ако $\overline{\mathbf{x}}$ е изродено оптимално базисно допустимо решение е възможно $\overline{c}_q < 0$ за някое $q \in N$ — ако ръбът с направление $\mathbf{d}_q$ е фиктивен ръб през $\overline{\mathbf{x}}$.

Непосредствени следствия от (5.10) са

Следствие 5.1. *Ако за всички $j \in N$ относителните оценки $\overline{c}_j > 0$, то базисното допустимо решение $\overline{\mathbf{x}}$ с базис B е единствено оптимално решение на каноничната задача (K).*

Следствие 5.2. *Ако $\overline{\mathbf{x}}$ е оптимално базисно допустимо решение на (K) с небазисни относителни оценки $\overline{c}_{j_1} = \overline{c}_{j_2} = \cdots = \overline{c}_{j_k} = 0$, то всяка допустима точка $\mathbf{x} \in M$ от вида*

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{i=1}^{k} x_{j_i} \mathbf{d}_{j_i} \tag{5.11}$$

също е оптимално решение на (K).

Да обърнем внимание на това че ако някои небазисни променливи имат нулеви относителни оценки то това не означава, че $\overline{\mathbf{x}}$ не е единствено оптимално решение. Наистина, ако $\overline{\mathbf{x}}$ е изродено то може да бъде единствената допустима точка от вида (5.11) поради фиктивност на ръбовете с направления $\mathbf{d}_{j_i}$, участващи в (5.11).

§5.5. Критерий за неограниченост на целевата функция

Нека сега за текущото $\overline{\mathbf{x}}$ с базис B не е в сила критерият за оптималност, т.е. нека съществува небазисен индекс $q \in N$, за който $\overline{c}_q < 0$. Това означава, че при движение по ръба с направление $\mathbf{d}_q$ целевата функция намалява. Въпреки че симплекс методът може да тръгне по произволен такъв ръб на намаляване, обичайното правило[1] е да се избере ръбът, съответстващ на най-малката относителна оценка $\overline{c}_q$.

След като симплекс методът е избрал ръб на намаляване с направление $\mathbf{d}_q$, той се придвижва от $\overline{\mathbf{x}}$ по този ръб, т.е. започва да увеличава t в (5.6).

Има две възможности:

- $\mathbf{w}_q \leqslant \mathbf{0}$. Тогава $\overline{t} = \infty$ и ръбът с направление $\mathbf{d}_q$ е неограничен ръб на намаляване на z.

Получаваме

Теорема 5.2. (критерий за неограниченост на целевата функция). *Ако за някой индекс $q \in N$, $\overline{c}_q < 0$ и $\mathbf{w}_q \leqslant \mathbf{0}$, то задачата (K) е неограничена.*

[1] Нарича се правило на Бил

Доказателство. От $\mathbf{w}_q \leqslant \mathbf{0}$ следва $\mathbf{d}_q \geqslant \mathbf{0}$, а от (5.5) имаме $\mathbf{A}\mathbf{d}_q = \mathbf{0}$. От Теорема 4.1, която характеризира посоките в M, следва че ненулевият (тъй като $\mathbf{e}_q \neq \mathbf{0}$) вектор $\mathbf{d}_q$ е посока в M. Тъй като $\mathbf{c}^{\mathrm{T}}\mathbf{d}_q = \overline{c}_q < 0$, векторът $\mathbf{c}$ на целевата функция z сключва тъп ъгъл с тази посока. От Теорема 4.2 следва, че z намалява неограничено по ръба $\{\overline{\mathbf{x}} + t\mathbf{d}_q,\ t \geqslant 0\}$. □

- $\mathbf{w}_q \nleqslant \mathbf{0}$. Тогава $\bar{t}$ е крайно като при $\bar{t} > 0$ ръбът с направление $\mathbf{d}_q$ е ограничен ръб на M, а при $\bar{t} = 0$ е фиктивен ръб. И в двата случая обаче се прави

§5.6. Преход към съседно базисно допустимо решение $\overline{\mathbf{x}}'$

Тъй като $\mathbf{w}_q \nleqslant \mathbf{0}$, най-голямата стъпка, която можем да направим по ръба с направление $\mathbf{d}_q$ без да нарушим допустимостта намираме от *теста за минимално отношение*

$$\bar{t} := \min \left\{ \frac{\overline{x}_i}{w_{iq}} :\ w_{iq} > 0, i \in B \right\}, \tag{5.12}$$

т.е. $\mathbf{x}(t) \in M$ за стойности на $t \in [0, \bar{t}]$ и $\mathbf{x}(t) \notin M$ за стойности на $t > \bar{t}$. Както ще докажем в Лема 5.1, точката $\overline{\mathbf{x}}' := \overline{\mathbf{x}} + \bar{t}\mathbf{d}_q$ е базисно допустимо решение, чийто базис се различава от базиса на $\overline{\mathbf{x}}$ само по един индекс. Две базисни допустими решения, чиито базиси се различават само по един индекс се наричат *съседни*, както и съответните им върхове, ако са различни. Да отбележим, че ако $\bar{t} > 0$, то $\overline{\mathbf{x}}'$ е връх на M, свързан с върха $\overline{\mathbf{x}}$ с ограничения ръб $\mathbf{d}_q$, а ако $\bar{t} = 0$, то $\overline{\mathbf{x}}'$ съвпада с $\overline{\mathbf{x}}$, но има базис, различен от базиса на $\overline{\mathbf{x}}$.

Нека в теста за минимално отношение (5.12) имаме

$$\bar{t} := \min \left\{ \frac{\overline{x}_i}{w_{iq}}, \quad w_{iq} > 0, i \in B \right\} = \frac{\overline{x}_p}{w_{pq}}.$$

Координатата w_{pq} на вектора $\mathbf{w}_q$, при която се достига минимума в това отношение, винаги е положително число. Числото w_{pq} се нарича още *ключово число.*

Координатите на точката $\overline{\mathbf{x}}'$, която се получава в края на стъпката с дължина $\bar{t}$ по ръба $\mathbf{d}_q$, съгласно (5.9) и дефиницията на $\bar{t}$, са

$$\begin{aligned} &\overline{x}'_j = 0, \quad j \in N, j \neq q\,, \\ &\overline{x}'_q = \bar{t} = \overline{x}_p / w_{pq}\,, \\ &\overline{x}'_i = \overline{x}_i - \bar{t} w_{iq} = \overline{x}_i - \overline{x}_p w_{iq}/w_{pq}\,, \quad i \in B\,. \end{aligned} \tag{5.13}$$

Да забележим, че ако $\overline{\mathbf{x}}$ е неизродено ($\overline{x}_i > 0$, $i \in B$), то $\overline{x}_p > 0$ (базисна координата), а $\overline{x}_q = 0$ (небазисна координата). Обратно, в $\overline{\mathbf{x}}'$ имаме, че координатата $\overline{x}'_p = \overline{x}_p - \overline{x}_p w_{pq}/w_{pq} = 0$, докато координатата $\overline{x}'_q = \bar{t} > 0$.

Това идва да подскаже, че така полученото $\overline{\mathbf{x}}'$ е базисно допустимо решение съседно на $\overline{\mathbf{x}}$: неговият базис B' се отличава от базиса B на $\overline{\mathbf{x}}$ по това, че в него не участва променливата x_p, но участва променливата x_q. Това ще докажем в следващата

Лема 5.1. *Точката $\overline{\mathbf{x}}'$ с координати* (5.13) *е друго базисно допустимо решение с базис B', който се различава от базиса B на $\overline{\mathbf{x}}$ по това че съдържа индекса q, а не съдържа индекса p, т.е. $B' = B \setminus \{p\} \cup \{q\}$, а базисната му матрица $\mathbf{B}'$ се различава от базисната матрица $\mathbf{B}$ на $\overline{\mathbf{x}}$, по това, че един от нейните стълбове – стълбът $\mathbf{A}_p$ – е заменен със стълба $\mathbf{A}_q$, т.е.*

$$\mathbf{B}' = \mathbf{B} + (\mathbf{A}_q - \mathbf{A}_p)\mathbf{e}_p^{\mathrm{T}} .$$

Доказателство. Точката $\overline{\mathbf{x}}'$ принадлежи на M съгласно дефиницията на $\bar{t}$. Да забележим, че стълбовете на матрицата $\mathbf{A}$, съответстващи на положителните координати на $\overline{\mathbf{x}}'$, са стълбове на матрицата $\mathbf{B}'$. Като вземем предвид Теорема 2.1 и Теорема 1.1, достатъчно е да покажем, че стълбовете на матрицата $\mathbf{B}'$ са линейно независими.

Да допуснем противното, т.е. че стълбовете $\mathbf{A}_i$, $i \in B'$ са линейно зависими. Това означава, че съществуват числа β_i, $i \in B'$ не всичките равни на нула и такива, че $\displaystyle\sum_{i\in B'} \beta_i \mathbf{A}_i = \mathbf{0}$ или като вземем предвид, че $B' = B \setminus \{p\} \cup \{q\}$

$$\sum_{i\in B, i\neq p} \beta_i \mathbf{A}_i + \beta_q \mathbf{A}_q = \mathbf{0}. \tag{5.14}$$

От $\mathbf{A}_q = \mathbf{B}\mathbf{B}^{-1}\mathbf{A}_q = \mathbf{B}\mathbf{w}_q$ имаме, че $\mathbf{A}_q$ се представя като линейна комбинация на стълбовете на $\mathbf{B}$ като коефициентите на линейната комбинация са координатите на вектора $\mathbf{w}_q$ или $\mathbf{A}_q = \displaystyle\sum_{i\in B} w_{iq}\mathbf{A}_i$. Като заместим това представяне на вектора $\mathbf{A}_q$ в (5.14), получаваме

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

Последното е линейна комбинация на векторите $\mathbf{A}_i$, $i \in B$, но те са линейно независими. Следователно всички коефициенти в тази линейна комбинация са равни на нула. От $\beta_q w_{pq} = 0$ и $w_{pq} > 0$ (понеже w_{pq} е ключово число!) следва, че $\beta_q = 0$, което, заместено в останалите коефициенти води до $\beta_i = 0$ за всяко $i \in B, i \neq p$. Получаваме, че всички коефициенти β_i, $i \in B'$, са нули, а с това и търсеното противоречие. □

При така настъпилата смяна на базиса B с базиса B', казваме, че променливата x_q *влиза* в базиса на мястото на променливата x_p, която *излиза* от базиса.

Като обобщим горните разсъждения, получаваме

Теорема 5.3. *Ако* $\overline{c}_q < 0$ *и* $\mathbf{w}_q$ *има положителна координата, то* $\overline{\mathbf{x}}'$ *с координати* (5.13) *е различно от* $\overline{\mathbf{x}}$ *базисно допустимо решение, в което стойността на целевата функция* $z(\overline{\mathbf{x}}') = z(\overline{\mathbf{x}}) + \bar{t}\overline{c}_q$ *е по-малка от стойността* $z(\overline{\mathbf{x}})$*, когато* $\bar{t}$*, определено с* (5.12)*, е положително число* (*в частност, когато* $\overline{\mathbf{x}}$ *е неизродено*).

За да завършим итерацията на симплекс метода, остава само да заменим базисното допустимо решение $\overline{\mathbf{x}}$ с базисното допустимо решение $\overline{\mathbf{x}}'$, да заменим базиса B с B' и да заменим базисната матрица $\mathbf{B}$ с базисната матрица $\mathbf{B}'$.

§6. Алгоритъм на симплекс метода. Методи за намиране на начално базисно допустимо решение

В § 5 описахме подробно интерацията на симплекс метода за решаване на канонична задача

$$(K)\qquad \begin{aligned} \min\quad & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

където $\mathbf{A}$ е $m \times n$ матрица, $\mathbf{b} \in \mathbb{R}^m_+$, $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n$. Сега ще дадем

§6.1. Алгоритъм на симплекс метода

Фаза I:

(0) Или се установява, че допустимото множество е празното и тогава КРАЙ – задачата (K) е несъвместима; или се намира начално базисно допустимо решение $\overline{\mathbf{x}}$ с базисна матрица $\mathbf{B} = [\mathbf{A}_{j_1}, \ldots, \mathbf{A}_{j_m}]$. Множеството от индексите на базисните променливи се означава с $B = \{j_1, \ldots, j_m\}$ (т.е. x_{j_i} е i-та базисна променлива, $i = 1, \ldots, m$), а множеството от индексите на небазисните променливи се означава с N.

Фаза II:

(1) Пресмятат се относителните оценки на небазисните променливи

$$\overline{c}_j = c_j - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_j \quad \text{за всяко } j \in N.$$

(2) Проверка на критерия за оптималност: Ако $\overline{c}_j \geqslant 0$, за всяко $j \in N$, то КРАЙ – текущото решение $\overline{\mathbf{x}}$ е оптимално, като оптималната стойност на функцията е $z(\overline{\mathbf{x}}) = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b}$. В противен случай се преминава към стъпка (3).

(3) Избор на небазисна променлива x_q за влизане в базиса. Тук е възможно да има нееднозначност — избира се ръб на намаляване на z, като се избере индекс

$$q \in \{j \in N \;:\; \overline{c}_j < 0\}\,.$$

(4) Проверка на критерия за неограниченост на целевата функция: Намират се координатите на вектора

$$\mathbf{w}_q = \mathbf{B}^{-1}\mathbf{A}_q.$$

Ако $\mathbf{w}_q \leqslant \mathbf{0}$, то КРАЙ – в допустимото множество M има неограничен ръб с направление $\mathbf{d}_q = \begin{bmatrix} -\mathbf{w}_q \\ \mathbf{e}_q \end{bmatrix}$, по който $z \to -\infty$. В противен случай се преминава към стъпка (5).

(5) Избор на базисна променлива x_{j_p} за излизане от базиса: Пресмята се

$$\bar{t} = \min_{1 \leqslant i \leqslant m} \left\{ \frac{\overline{x}_{j_i}}{w_{iq}} \ : \ w_{iq} > 0 \right\}$$

и се избира базисна променлива x_{j_p}, за която $\bar{t} = \frac{\overline{x}_{j_p}}{w_{pq}}$ за излизане от базиса. Ако минимумът се достига за повече от един базисен индекс, изборът на излизаща от базиса променлива е нееднозначен.

(6) Обновява се текущото решение, базиса B и базисната матрица $\mathbf{B}$: Присвоява се

$$\overline{x}_q \Leftarrow \bar{t} = \overline{x}_{j_p}/w_{pq} \, ,$$
$$\overline{x}_{j_i} \Leftarrow \overline{x}_{j_i} - \bar{t} w_{iq}, \quad 1 \leqslant i \leqslant m \, ,$$
$$B \Leftarrow B \setminus \{j_p\} \cup \{q\} \, ,$$
$$\mathbf{B} \Leftarrow \mathbf{B} + (\mathbf{A}_q - \mathbf{A}_{j_p})\mathbf{e}_p^{\mathrm{T}} \, ,$$
$$j_p \Leftarrow q$$

и се преминава към стъпка (1).

При описанието на итерацията на симплекс метода в § 5 (фаза II), предположихме че във фаза I вече е намерено начално базисно допустимо решение (това е стъпка (0) на алгоритъма). Да видим как симплекс методът, който намира оптимално решение във фаза II, може успешно да се приложи и за решаването на задачата на фаза I — намирането на начално базисно допустимо решение (начален връх).

§6.2. Методи за намиране на начален връх

За реализиране на фаза I на симплекс метода ще разгледаме два метода.

При *двуетапния метод* на първия етап се решава спомагателна канонична задача (I). От нейното оптимално базисно допустимо решение се получава базисно допустимо решение на (K), което на втория етап симплекс методът използва като начален връх за решаване на (K).

Спомагателната канонична задача (I) се формулира по следния начин:

- всяко от ограниченията на задачата (K) се преобразува като към лявата му страна се добавя *изкуствена неотрицателна променлива* $y_i \geqslant 0$, при което то добива вида $\mathbf{a}_i^{\mathrm{T}}\mathbf{x} + y_i = b_i$;

- върху новото множество от ограничения се минимизира сумата на изкуствените променливи.

Така се получава следната канонична задача

$$(I) \qquad \begin{aligned} \min \quad & \xi(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{m} y_i \\ & \mathbf{A}\mathbf{x} + \mathbf{I}\mathbf{y} = \mathbf{b} \quad (\mathbf{b} \geqslant \mathbf{0}), \\ & \mathbf{x} \geqslant \mathbf{0}, \quad \mathbf{y} \geqslant \mathbf{0}, \end{aligned}$$

където $\mathbf{y}(y_1, \dots, y_m) \in \mathbb{R}^m$ е векторът на изкуствените променливи, а с $\mathbf{I}$ е означена единичната матрица от ред m.

Задачата (I) има очевидно базисно допустимо решение $(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = (\mathbf{0}, \mathbf{b})$. Неговият базис се състои само от изкуствените променливи $\mathbf{y}$, а базисната му матрица е $\mathbf{I}$. В частност допустимото множество на (I) не е празно. Целевата функция ξ приема неотрицателни стойности върху него. Следователно, задачата (I) е разрешима.

И така, за каноничната задача (I) прилагаме симплекс метода като използваме $(\mathbf{0}, \mathbf{b})$ за начален връх, за да намерим нейно оптимално базисно допустимо решение $(\overline{\mathbf{x}}^*, \overline{\mathbf{y}}^*)$. Да означим с ξ^* оптималната стойност на ξ, т.е. $\xi^* := \xi(\overline{\mathbf{x}}^*, \overline{\mathbf{y}}^*)$.

Възможни са три случая:

1. $\xi^* > 0$, което означава, че в оптималния базис има изкуствена променлива y_i, която има положителна стойност $\overline{y}_i^* > 0$ в оптималното решение. В този случай допустимото множество на задачата (K) е празно. Наистина, ако допустимото множество M на задачата (K) не е празно, то $\xi^* = 0$: като вземем $\mathbf{x} \in M$ и положим $\mathbf{y} = \mathbf{0}$, съответният вектор $(\mathbf{x}, \mathbf{0})$ е допустим за задачата (I) и в него $\xi(\mathbf{x}, \mathbf{0}) = 0$.

2. $\xi^* = 0$ и всички изкуствени променливи y_i са небазисни за $(\overline{\mathbf{x}}^*, \overline{\mathbf{y}}^*)$. В този случай $\overline{\mathbf{x}}^*$ е базисно допустимо решение на задачата (K), което фаза II на СМ ще използва за начално базисно допустимо решение при нейното решаване.

3. $\xi^* = 0$, но някои от изкуствените променливи y_i са останали в базиса с нулева стойност (базисни нули) в оптималното решение $(\overline{\mathbf{x}}^*, \overline{\mathbf{y}}^*)$, откъдето следва, че то е изродено. В този случай всяка изкуствена променлива, която е останала в базиса, може или да се елиминира заедно с излишното уравнение, с което е асоциирана, или да се замени в базиса от някоя небазисна $\mathbf{x}$-променлива.

   По точно казано, нека $\overline{y}_i^* = 0$ и y_i е k-та базисна променлива. Ако

   - $\mathbf{e}_k^{\mathrm{T}} \mathbf{B}^{-1} \mathbf{A}_j = 0$ за всички небазисни стълбове $\mathbf{A}_j$, то това означава, че след елементарни преобразувания k-ият ред на матрицата $\mathbf{A}$ се е трансформирал в нулевия вектор, а k-то уравнение на системата се е трансформирало в тъждеството $\mathbf{0}^{\mathrm{T}}\mathbf{x} = 0$. Следователно в системата линейни уравнения $\mathbf{A}\mathbf{x} = \mathbf{b}$ k-то е излишно и k-ият ред може да се отстрани от матрицата $\mathbf{A}$, а k-ият ред и k-ият стълб да се отстранят от базисната матрица на $(\overline{\mathbf{x}}^*, \overline{\mathbf{y}}^*)$ заедно с k-та базисна променлива y_i.

- $\mathbf{e}_k^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_j \neq 0$ за някой небазисен стълб $\mathbf{A}_j$, то y_i може да се замени в базиса от небазисната променлива x_j. Тъй като при тази замяна $\overline{t}$=0 (поради $\overline{y}_i^* = 0$), променя се само базисът на решението.

След като изкуствените променливи се заменят последователно в базиса с $\mathbf{x}$-променливи, получава се базисно допустимо решение на задачата (I) с базис, състоящ се само от $\mathbf{x}$-променливи, от което се получава базисно допустимо решение за задачата (K), както в случай 2.

По-често за намиране на начално базисно допустимо решение се използва т.нар. *M-метод*. Той използва същите идеи, както двуетапния метод, но обединява в едно неговите два етапа. Разглежда се т.нар. *M-задача*

$$(M) \qquad \begin{array}{ll} \min & z_M(\mathbf{x},\mathbf{y}) = \displaystyle\sum_{j=1}^{n} c_j x_j + M \sum_{i=1}^{m} y_i \\ & \mathbf{A}\mathbf{x} + \mathbf{y} = \mathbf{b} \quad (\mathbf{b} \geqslant \mathbf{0}), \\ & \mathbf{x} \geqslant \mathbf{0}, \quad \mathbf{y} \geqslant \mathbf{0}, \end{array}$$

където $M > 0$ е достатъчно голямо число.

Следващите две теореми показват как каноничната задача (K) се решава посредством решаване на съответната ѝ (M)-задача.

Теорема 6.1. *Ако задачата (K) е разрешима, то съществува число $M_0 > 0$, такова че за всяко $M \geqslant M_0$ съответната (M)-задача е разрешима и $\mathbf{y}$-координатите на всички нейни оптимални базисни допустими решения са нули.*

Доказателство. Тъй като в доказателството се използват основни резултати от двойнствеността в линейното оптимиране, на които е посветен § 8, то ще бъде направено в § 8. □

Теорема 6.2. *Ако $(\mathbf{x}^*, \mathbf{0})$ е оптимално решение на задачата (M), то $\mathbf{x}^*$ е оптимално решение на задачата (K).*

Доказателство. Нека $\mathbf{x}$ е произволна допустима точка за задачата (K). Тогава $(\mathbf{x}, \mathbf{0})$ е допустима точка за (M) и $z_M(\mathbf{x}^*, \mathbf{0}) \leqslant z_M(\mathbf{x}, \mathbf{0})$. Но $z_M(\mathbf{x}, \mathbf{0}) = z(\mathbf{x})$ за всяко $\mathbf{x}$, откъдето $z(\mathbf{x}^*) \leqslant z(\mathbf{x})$ и $\mathbf{x}^*$ е оптимално решение на (K). □

И така, (M)-задачата има очевидно начално базисно допустимо решение $(\mathbf{0}, \mathbf{b})$. Симплекс методът, приложен за решаване на (M)-задачата приключва по един от следните два начина:

1. Установява, че (M)-задачата е неограничена. Тогава от Теорема 6.1 следва, че каноничната задача (K) няма решение. За изясняване на това дали неразрешимостта на (K) може да се дължи на празно допустимо множество или на неограниченост на целевата функция са необходими допълнителни изследвания;

2. Намира е оптимално базисно допустимо решение $(\overline{\mathbf{x}}^*, \overline{\mathbf{y}}^*)$ на (M)-задачата. Ако $\overline{\mathbf{y}}^* \neq \mathbf{0}$ то от доказателството на Теорема 6.1 приведено в § 8 става ясно, че в този случай задачата (K) е с празно допустимо множество. Ако $\overline{\mathbf{y}}^* = \mathbf{0}$, то $\overline{\mathbf{x}}^*$ е оптимално решение на задачата (K) съгласно Теорема 6.2.

§7. Изроденост и зацикляне на симплекс алгоритъма. Правило на Бленд за избягване на зацикляното

Вече се запознахме със симплекс алгоритъма за решаване каноничната задача

$$(K) \qquad \begin{array}{ll} \min & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{array}$$

където $\mathbf{A}$ е $m \times n$ матрица, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m_+$ и $\mathbf{x} \in \mathbb{R}^n$ е n-мерен вектор на променливите.

От Теорема 5.3 следва, че ако всички върхове на допустимото множество $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ са неизродени, то стойността на целевата функция z ще намалява на всяка итерация и че напускайки даден връх, симплекс методът няма да се върне пак в него. Тъй като M има краен брой върхове, това означава, че симплекс алгоритъмът е краен. Зацикляне на алгоритъма може да се получи само в изроден връх.

Ако текущият връх $\overline{\mathbf{x}}$ с базис B е изроден, за влизане в базиса е определена променлива x_q, чиято $\overline{c}_q < 0$ и $\overline{\mathbf{x}}$ има базисна нула $\overline{x}_p = 0$, такава че $w_{pq} > 0$, то тогава $\overline{t} = 0$. Като резултат предприетата стъпка ще бъде с нулева дължина и в края на итерацията върхът $\overline{\mathbf{x}}'$ ще съвпадне с върха $\overline{\mathbf{x}}$, но ще има **различен базис** $B' = B \setminus \{p\} \cup \{q\}$.

Тъй като текущият връх $\overline{\mathbf{x}}$ и стойността на целевата функция в него $z(\overline{\mathbf{x}})$ не се променят, теоретично е възможно симплекс методът да зацикли — да преминава безкрайно през редица от базиси на един и същи изроден връх, и да не го напуска. На практика това не представлява проблем, тъй като съществуват правила за избор на влизащата в базиса променлива и на излизащата от базиса променлива, които предотвратяват зацикляното.

Пример за такова правило е **правилото на Бленд**, което гласи: **на всяка итерация от променливите, които са кандидати за влизане в базиса и от променливите, които са кандидати за излизане от базиса, винаги се избират тези с най-малък индекс**. Като кандидат за влизане в базиса се разглежда всяка небазисна променлива x_q, такава че $\overline{c}_q < 0$, а като кандидат за излизане от базиса се разглежда всяка базисна променлива x_p, такава че $\overline{t} = \dfrac{\overline{x}_p}{w_{pq}}$, където $\overline{t}$ се определя от теста за минимално отношение (5.12). Да забележим, че ако минимумът се достига за повече от един базисен индекс, то следващото базисно допустимо решение $\overline{\mathbf{x}}'$ ще бъде изродено и изборът на излизаща от базиса променлива ще бъде нееднозначен.

Ще докажем, че при спазване на правилото на Бленд симплекс методът приключва след краен брой итерации.

Теорема 7.1. *Симплекс методът приключва след краен брой итерации, ако на всяка итерация от кандидатите за влизане и от кандидатите за излизане от базиса се избират променливите 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 покажем, че това води до противоречие.

И така, да допуснем, че въпреки спазването на правилото на Бленд се получава цикъл и че зациклянето е в изродения връх $\overline{\mathbf{x}}$.

Нека $\{B_1, \dots, B_k\}$ е редицата от базиси на $\overline{\mathbf{x}}$, през които зациклят методът, т.е. симплекс методът генерира следната безкрайна редица от базиси на $\overline{\mathbf{x}}$

$$B_1, \dots, B_k, B_1, \dots, B_k, B_1, \dots. \tag{7.1}$$

Ще казваме, че дадена променлива е *непостоянна*, ако тя участва в един и не участва в друг от тези базиси.

Нека x_p е непостоянната променлива с най-голям индекс, нека B е базиса, от който тя излиза и нека x_q е променливата, влизаща в базиса на мястото на x_p. Тъй като променливата x_q влиза в базиса B, а от него излиза променливата x_p, то $q \in N$ (с N както обикновено са означени небазисните за базиса B индекси), а $p \in B$ и следователно променливата x_q е непостоянна променлива (не участва в базиса B, но участва в следващия го базис). Оттук $q < p$.

От базисния вид (K_B) на задачата спрямо базиса B за решение $\mathbf{x}$ на системата $\mathbf{A}\mathbf{x} = \mathbf{b}$ имаме

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j \in N} x_j \mathbf{d}_j \quad \text{и} \quad z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j \in N} \overline{c}_j x_j. \tag{7.2}$$

Нека сега означим с B^* този базис от базисите в (7.1), в който променливата x_p влиза. Очевидно B^* е различен от B. С N^* да означим множеството от небазисните спрямо базиса B^* индекси, т.е. $N^* := \{1, \dots, n\} \setminus B^*$.

От базисния вид (K_{B*}) на задачата спрямо базиса B^* за решение $\mathbf{x}$ на системата $\mathbf{A}\mathbf{x} = \mathbf{b}$ имаме

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j \in N^*} x_j \mathbf{d}_j^* \quad \text{и} \quad z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j \in N^*} \overline{c}_j^* x_j = z(\overline{\mathbf{x}}) + \sum_{j=1}^{n} \overline{c}_j^* x_j, \tag{7.3}$$

като последното равенство в представянето на $z(\mathbf{x})$ следва от това, че относителните оценки на базисните спрямо базиса B^* променливи са нули, т.е. $\overline{c}_j^* = 0$ за $j \in B^*$ и включването им в сумата не я променя.

Знаем, че за произволно зададен набор от стойности на променливите x_j, $j \in N$, след като от (7.2) се пресметнат стойностите на променливите x_i, $i \in B$ се получава частно решение $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix}$ на системата $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Нека $\mathbf{x}$ е частното решение на системата, което се получава като зададем $x_q := \alpha$ за произволно число $\alpha \in \mathbb{R}$ и $x_j := 0$ за $j \in N$, $j \neq q$. От (7.2), като вземем предвид, че $\mathbf{d}_q = \begin{bmatrix} -\mathbf{w}_q \\ \mathbf{e}_q \end{bmatrix}$, получаваме $x_i = \overline{x}_i - \alpha w_{iq}$ за $i \in B$.

Стойността на целевата функция z в така полученото частно решение $\mathbf{x}$ според (7.2) е $z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \overline{c}_q\alpha$, а според (7.3) е $z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \overline{c}_q^*\alpha + \sum\limits_{i\in\mathbf{B}} \overline{c}_i^*(\overline{x}_i - \alpha w_{iq})$. Като приравним тези два израза получаваме

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

Тъй като това равенство е в сила за всяко $\alpha \in \mathbb{R}$, то коефициентът пред α (а също така и дясната страна на равенството) е нула:

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

Тъй като променливата x_q влиза в базиса B, това означава, че относителната ѝ оценка спрямо този базис е $\overline{c}_q < 0$. Тъй като променливата x_q **не** влиза в базиса B^* (в този базис влиза променливата x_p) и тъй като $q < p$, то по правилото на Бленд за избор на променлива за влизане в базиса B^* имаме, че $\overline{c}_q^* \geqslant 0$. Следователно

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

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

$$\overline{c}_r^* w_{rq} > 0. \tag{7.4}$$

В частност $\overline{c}_r^* \neq 0$. Тъй като относителните оценки на базисните променливи са нули, то r е небазисен индекс, т.е. $r \in N^*$. Получаваме, че променливата x_r е непостоянна, понеже участва в базиса B и не участва в базиса B^*. Следователно $r \leqslant p$. Нещо повече, $r < p$, тъй като $\overline{c}_p^* w_{pq} < 0$ (относителната оценка $\overline{c}_p^* < 0$, понеже x_p влиза в базиса B^*, а $w_{pq} > 0$, понеже x_p излиза от базиса B, за да влезе на нейно място x_q, откъдето w_{pq} е ключовото число, а то винаги е положително).

Това, че $r < p$ води до $\overline{c}_r^* \geqslant 0$, защото в противен случай съгласно критерия за избор на променлива с най-малък индекс за влизане в базиса B^* влизаща в базиса B^* щеше да бъде променливата x_r, а не x_p и от (7.4) следва

$$w_{rq} > 0. \tag{7.5}$$

Тъй като всеки от базисите в редицата (7.1) е базис на един и същи връх $\overline{\mathbf{x}}$, то всяка непостоянна променлива е с нулева стойност в $\overline{\mathbf{x}}$, т.е. имаме базисна нула $\overline{x}_i = 0$, ако i е непостоянна променлива. В частност

$$\overline{x}_r = 0. \tag{7.6}$$

От (7.5) и (7.6) следва, че променливата x_r е била кандидат за излизане от базиса B и понеже $r < p$, то по правилото на Бленд тя е трябвало да бъде избрана за излизане от базиса B, а не променливата x_p. Получихме търсеното противоречие и с това доказахме теоремата. $\square$

Задачи, допустимото множество на които има изродени върхове, са често срещани. За това, че симплекс методът няколко итерации е бил в изроден връх може да се съди по това, че целевата функция приема една и съща стойност неколкократно (преди да бъде намерен действителен ръб на намаляване), а след това стойностите ѝ отново започват да намаляват.

§8. Приложни реализации на симплекс метода

В § 5 и § 6 се запознахме със симплекс метода за решаване на канонична задача

$$(K)\qquad \begin{aligned} \min \quad & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

където $\mathbf{A}$ е $m \times n$ матрица, $\mathbf{b} \in \mathbb{R}^m_+$, $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n$.

Сега ще се спрем на това как на практика се извършват пресмятанията, необходими на симплекс метода.

§8.1. Таблична форма на симплекс метода

Един от начините да се организират пресмятанията, които се правят при итерацията на симплекс метода е данните за текущото базисно решение да се вложат в една голяма матрица, наречена *симплексна таблица*, която може да бъде генерирана директно от данните на задачата.

При зададени матрица $\mathbf{A}$, дясна част $\mathbf{b}$ и вектор на целевата функция $\mathbf{c}$ изходната таблица е просто следната по-голяма матрица:

$$T'' = \left[\begin{array}{c:c} \mathbf{A} & \mathbf{b} \\ \hdashline \mathbf{c}^{\mathrm{T}} & 0 \end{array}\right].$$

Таблицата T'' е с $m+1$ реда и $n+1$ стълба.

Нека $\overline{\mathbf{x}}$ е базисно решение с базисна матрица $\mathbf{B}$. Ако е необходимо, преномерираме променливите така, че $x_1, \dots, x_m$ да бъдат базисните променливи. Тогава базисната матрица $\mathbf{B}$ на $\overline{\mathbf{x}}$ се състои от първите m стълба на $\mathbf{A}$, а последните ѝ $n-m$ стълба образуват подматрицата $\mathbf{N}$ с размерност $m \times (n-m)$.

След евентуалното преномериране на променливите в таблицата имаме, че първите стълбове са базисните, т.е.

$$T''(\overline{\mathbf{x}}) = \left[\begin{array}{c:c:c} \mathbf{B} & \mathbf{N} & \mathbf{b} \\ \hdashline \mathbf{c}_B^{\mathrm{T}} & \mathbf{c}_N^{\mathrm{T}} & 0 \end{array}\right].$$

Изразяването на базисните координати $\mathbf{x}_B$ на решение $\mathbf{x}$ на системата $\mathbf{A}\mathbf{x} = \mathbf{b}$ чрез небазисните му координати $\mathbf{x}_N$ се постига чрез елементарни Гаусови преобразувания на горната част на таблицата

$$T'(\overline{\mathbf{x}}) = \left[\begin{array}{c:c:c} \mathbf{I} & \mathbf{B}^{-1}\mathbf{N} & \mathbf{B}^{-1}\mathbf{b} \\ \hdashline \mathbf{c}_B^{\mathrm{T}} & \mathbf{c}_N^{\mathrm{T}} & 0 \end{array}\right].$$

Матрицата, която преобразува $\mathbf{B}$ в $\mathbf{I}$, е матрицата $\mathbf{B}^{-1}$ и направените елементарни преобразувания са еквивалентни на умножение отляво с матрицата $\mathbf{B}^{-1}$ на горната част на таблицата.

За да завършим, остава да заместим така изразените базисни променливи в целевата функция z, при което тя става функция само на небазисните променливи. За целта умножаваме с $\mathbf{c}_{\mathbf{B}}^{\mathrm{T}}$ горната част на таблицата и я вадим от долната:

$$T(\overline{\mathbf{x}}) = \left[\begin{array}{c:c:c} \mathbf{I} & \mathbf{B}^{-1}\mathbf{N} & \mathbf{B}^{-1}\mathbf{b} \\ \hdashline \mathbf{0} & \mathbf{c}_N^{\mathrm{T}} - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{N} & -\mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} \end{array}\right] = \left[\begin{array}{c:c:c} \mathbf{I} & \mathbf{B}^{-1}\mathbf{N} & \mathbf{B}^{-1}\mathbf{b} \\ \hdashline \mathbf{0} & \overline{\mathbf{c}}_N^{\mathrm{T}} & -z(\overline{\mathbf{x}}) \end{array}\right].$$

Това е окончателният вид на симплексната таблица за базисното решение $\overline{\mathbf{x}}$.

Ако $\overline{\mathbf{x}}$ е текущото решение за симплексната итерация, то е базисно допустимо решение, т.е. $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b} \geqslant \mathbf{0}$ и горните преобразувания са еквивалентни на привеждането на задачата (K) в базисен вид спрямо базиса B на $\overline{\mathbf{x}}$.

Векторът от базисните координати на $\overline{\mathbf{x}}$, който е $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b}$ се намира горе вдясно на $T(\overline{\mathbf{x}})$, а небазисните му координати, разбира се, са $\overline{\mathbf{x}}_N = \mathbf{0}$. Стойността на целевата функция z в $\overline{\mathbf{x}}$ е $z(\overline{\mathbf{x}}) = \mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b}$ и се намира долу вдясно на таблицата с обратен знак.

Симплексната таблица $T(\overline{\mathbf{x}})$ съдържа цялата информация, необходима на симплексната итерация.

От нея може да се прецени дали базисното допустимо решение $\overline{\mathbf{x}}$ е оптимално или не — координатите на вектора на относителните оценки на небазисните променливи $\overline{\mathbf{c}}_N^{\mathrm{T}} = \mathbf{c}_N^{\mathrm{T}} - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{N}$, се намират долу в средата на таблицата.

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

Ако сред координатите на вектора $\overline{\mathbf{c}}_N$ има отрицателни, то $\overline{\mathbf{x}}$ не удовлетворява критерия за оптималност и стойността на целевата функция още може да се намали. Всяка отрицателна координата $\overline{c}_j < 0$ на вектора $\overline{\mathbf{c}}_N$ съответства на ръб $\mathbf{d}_j$, по който целевата функция намалява.

Избира се отрицателна координата $\overline{c}_q$ на $\overline{\mathbf{c}}_N$ (като се спазва правилото на Бил или правилото на Бленд). Това означава, че в базиса ще влезе небазисната променлива x_q.

За да се определи коя от базисните променливи ще излезе от базиса, е необходимо да се намери $\overline{t}$ от теста за минимално отношение:

$$\overline{t} = \min\left\{\frac{\overline{x}_i}{w_{iq}} = \frac{(\mathbf{B}^{-1}\mathbf{b})_i}{(\mathbf{B}^{-1}\mathbf{A}_q)_i} : w_{iq} > 0, 1 \leqslant i \leqslant m\right\} = \frac{\overline{x}_p}{w_{pq}} = \frac{(\mathbf{B}^{-1}\mathbf{b})_p}{(\mathbf{B}^{-1}\mathbf{A}_q)_p}. \tag{8.1}$$

В него участват координатите на вектора $\mathbf{w}_q := \mathbf{B}^{-1}\mathbf{A}_q$ (той се намира в q-ия стълб на таблицата $T(\overline{\mathbf{x}})$ — стълбът в горната част на таблицата, който стои над координатата $\overline{c}_q$ на вектора $\overline{\mathbf{c}}_N$) и на вектора $\mathbf{B}^{-1}\mathbf{b}$ (той, както вече отбелязахме, са намира горе вдясно на таблицата).

Минимумът в (8.1) се взима само по положителните координати на вектора $\mathbf{w}_q$. Ако в стълба $\mathbf{w}_q$ няма положителни координати, то от текущия връх $\overline{\mathbf{x}}$ излиза неограничен ръб $\mathbf{d}_q = \begin{bmatrix} -\mathbf{w}_q \\ \mathbf{e}_q \end{bmatrix}$, по който целевата функция z намалява неограничено.

В противен случай се преминава към съседно базисно допустимо решение $\overline{\mathbf{x}}'$, чиято базисна матрица $\mathbf{B}'$ се получава, като p-ия стълб на матрицата $\mathbf{B}$ се замени със стълба $\mathbf{A}_q$.

За да се получи симплексната таблица $T(\overline{\mathbf{x}}')$ на новото базисно допустимо решение $\overline{\mathbf{x}}'$, се използва таблицата $T(\overline{\mathbf{x}})$ на предходното $\overline{\mathbf{x}}$, а не началната таблица T''. Преобразуванието, което трансформира $T(\overline{\mathbf{x}})$ в $T(\overline{\mathbf{x}}')$, се нарича *pivot* или *завъртане.* То осъществява стъпка (6) на алгоритъма и е еквивалентно привеждане на задачата (K) в базисен вид спрямо новия базис B'.

Да означим с t_{ij} елемента, който се намира в i-ия ред и j-ия стълб на $T(\overline{\mathbf{x}})$ за $i = 1, \dots m+1$, $j = 1, \dots, n+1$.

Да означим с t'_{ij} елемента, който ще се намира в i-ия ред и j-ия стълб на $T(\overline{\mathbf{x}}')$ за $i = 1, \dots m+1$, $j = 1, \dots, n+1$.

Завъртането (pivot) прави следното:

1. $t'_{pj} = \dfrac{t_{pj}}{t_{pq}}, \quad j = 1, \dots, n+1,$

2. $t'_{ij} = t_{ij} - t_{iq}\dfrac{t_{pj}}{t_{pq}}, \quad j = 1, \dots, n+1, \quad i = 1, \dots, m+1, i \neq p.$

Така 1. и 2. за $1 \leqslant i \leqslant m$ и $1 \leqslant j \leqslant n$ изразяват новите базисни променливи $\mathbf{x}_{B'}$ чрез новите небазисни променливи $\mathbf{x}_{N'}$;

1. и 2. за $1 \leqslant i \leqslant m$ и $j = n+1$ пресмятат новите базисни координати $\overline{\mathbf{x}}_{B'}$ от старите $\overline{\mathbf{x}}_B$;

2. за $i = m+1$ и $1 \leqslant j \leqslant n$ изразява z като функция само на новите небазисни променливи $\mathbf{x}_{N'}$, т.е. така новите относителни оценки $\overline{\mathbf{c}}'$ се получават от старите $\overline{\mathbf{c}}$.

По този начин на всяка итерация се създава симплексна таблица, съдържаща цялата необходима на симплексната итерация информация — за оптималност, за неограниченост, за край на алгоритъма.

Именно табличната форма на симплекс метода е формата в която той е бил описан при своето създаване. Въпреки че тази форма е илюстративна и приемлива за учебни примери с малко променливи, тя не е подходяща за решаване на задачи с голяма размерност, както и на задачи, в които матрицата $\mathbf{A}$ има някаква структура (например има много нулеви коефициенти), а такива често възникват в практиката. Това е така, понеже завъртането на таблицата обикновено разрушава структурата на матрицата $\mathbf{A}$. Друг недостатък на завъртането е, че при него на всяка итерация се генерират всичките $n - m$ стълба на матрицата $\mathbf{B}^{-1}\mathbf{N}$, докато само един от тях — стълбът $\mathbf{B}^{-1}\mathbf{A}_q$ е необходим на теста за минимално отношение (8.1).

§8.2. Модифициран симплекс метод

За отстраняване на недостатъците на табличната форма на симплекс метода на по-късен етап е разработен т.нар. *модифициран симплекс метод.*

Когато симплекс методът се намира в текущия връх $\overline{\mathbf{x}}$ с базисна матрица $\mathbf{B}$, необходимо е да бъдат направени пресмятанията на стъпки (1), (4) и (5) от алгоритъма. Пресмятането на относителните оценки на стъпка (1) се извършва на два етапа: първо да се намират координатите на вектора

$$\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$$

и след това те да се използват за пресмятане на относителните оценки $\overline{c}_j = c_j - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{A}_j$, $j \in N$. Векторът $\boldsymbol{\pi}$ се нарича *вектор на симплексните множители.*

Следователно, за целите на симплекс метода е необходимо да бъдат намерени координатите на векторите $\boldsymbol{\pi}^{\mathrm{T}}$, $\mathbf{w}_q$ и $\overline{\mathbf{x}}_B$. Те се пресмятат съответно така:

$$\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}, \quad \mathbf{w}_q = \mathbf{B}^{-1}\mathbf{A}_q \quad \text{и} \quad \overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b}, \tag{8.2}$$

и следователно за намирането им е достатъчно да се знае обратната на базисната матрица $\mathbf{B}$ — матрицата $\mathbf{B}^{-1}$.

При първоначалните реализации на модифицирания симплекс метод матрицата $\mathbf{B}^{-1}$ се е съхранявала и поддържала в явен вид, като се е обновявала при всяка смяна на базиса.

В последствие се отчита фактът, че най-бързо и лесно става решаването на система линейни уравнения, чиято матрица е *елементарна матрица*. Една матрица се нарича *елементарна*, ако се различава от единичната само по елементите на един от своите стълбове. Реализира се идеята на всяка итерация матрицата $\mathbf{B}^{-1}$ да бъде представяна и поддържана във вид на произведение от елементарни матрици.

Ако $\mathbf{B}$ е базисната матрица на текущото решение, следващата базисна матрица $\mathbf{B}'$ (която се получава като p-ия стълб на $\mathbf{B}$ се замени със стълба $\mathbf{A}_q$) може да се получи след умножение отдясно на $\mathbf{B}$ с елементарна матрица, която ще означим с $\mathbf{E}$, т.е.

$$\mathbf{B}' = \mathbf{BE},$$

където

$$\mathbf{E} := \begin{array}{c} \begin{bmatrix} 1 & & w_{1q} & & \\ & \ddots & \vdots & & \\ & & w_{pq} & & \\ & & \vdots & \ddots & \\ & & w_{mq} & & 1 \end{bmatrix} \\ \begin{array}{c} \uparrow \\ \text{стълб } p \end{array} \end{array}. \tag{8.3}$$

Лесно се проверява, че умножаването на $\mathbf{B}$ отдясно с $\mathbf{E}$ оставя всички стълбове на $\mathbf{B}$ непроменени с изключение на p-ия, който се трансформира в $\mathbf{B}\mathbf{w}_q = \mathbf{A}_q$, както е необходимо. С други думи казано, това умножение заменя p-ия стълб на $\mathbf{B}$ с вектора $\mathbf{A}_q$.

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

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

където обратната на матрицата $\mathbf{E}$ от (8.3) е елементарната матрица

$$\mathbf{E}^{-1} = \mathbf{I} - \frac{\mathbf{e}_p^{\mathrm{T}}(\mathbf{w}_q - \mathbf{e}_p)}{w_{pq}} = \begin{array}{c} \begin{bmatrix} 1 & & -w_{1q}/w_{pq} & & \\ & \ddots & \vdots & & \\ & & 1/w_{pq} & & \\ & & \vdots & \ddots & \\ & & -w_{mq}/w_{pq} & & 1 \end{bmatrix} \\ \uparrow \\ \text{стълб } p \end{array}. \tag{8.4}$$

В § 6 видяхме, че след въвеждане на изкуствени променливи базисната матрица на началното базисно допустимо решение е единичната матрица $\mathbf{I}$. Тогава на k-та итерация обратната на матрицата $\mathbf{B}_k$ ще се представя като произведение на k елементарни матрици

$$\mathbf{B}_k^{-1} = \mathbf{E}_k^{-1}\mathbf{E}_{k-1}^{-1}\cdots\mathbf{E}_1^{-1}\mathbf{I}, \tag{8.5}$$

където всяка от матриците $\mathbf{E}_i^{-1}$ е от вида (8.4). Обновяването на обратната на базисната матрица става просто чрез добавяне на нова елементарна матрица в произведението. Ясно е, че вместо да се помни цялата елементарна матрица $\mathbf{E}_i^{-1}$, достатъчно е да се помни само стълба на $\mathbf{E}_i^{-1}$, който се различава от съответния стълб на единичната матрица, както и неговото място в $\mathbf{E}_i^{-1}$, което пести памет.

След натрупването на $2m$ на брой елементарни матрици в произведението е препоръчително текущата матрица $\mathbf{B}^{-1}$ да се преизчислява, а използваните до момента елементарни матрици – да се изтриват. Тъй като е известно кои са стълбовете на $\mathbf{A}$, включени в $\mathbf{B}$, то матрицата $\mathbf{B}^{-1}$ се получава чрез последователна замяна на стълбовете на $\mathbf{I}$ със съответните базисни стълбове като всяка замяна съответства на умножение на $\mathbf{I}$ с елементарна матрица от вида $\mathbf{E}^{-1}$. Така след преизчисляването си $\mathbf{B}^{-1}$ вече е представена като произведение на не повече от m елементарни матрици.

По описания начин се поддържа представянето на $\mathbf{B}^{-1}$ като произведение на не повече от $2m$ елементарни матрици, което спестява изчислително време, пести памет и намалява грешките от закръгляване.

Още по-съвременен подход е да се гледа на (8.2) като на три системи линейни уравнения с една и съща матрица на коефициентите $\mathbf{B}$

$$\boldsymbol{\pi}^{\mathrm{T}}\mathbf{B} = \mathbf{c}_B^{\mathrm{T}}, \quad \mathbf{B}\mathbf{w}_q = \mathbf{A}_q, \quad \mathbf{B}\overline{\mathbf{x}}_B = \mathbf{b} \tag{8.6}$$

и за тяхното решаване да се използуват числените методи на линейната алгебра, като се направи стандартното директно разлагане на матрицата $\mathbf{B} = \mathbf{LU}$, където $\mathbf{L}$ е долна триъгълна, а $\mathbf{U}$ е горна триъгълна матрица. Това позволява лесното и бързо решаване на трите системи. Възможно е елементите на разлагането $\mathbf{L}'$ и $\mathbf{U}'$ на следващата базисна матрица $\mathbf{B}' = \mathbf{L}'\mathbf{U}'$ да не се пресмятат чрез директно разлагане, а да се получат от предходните $\mathbf{L}$ и $\mathbf{U}$ чрез подходящо умножение с елементарни матрици.

§9. Двойнственост в линейното оптимиране. Двойка спрегнати задачи. Теореми за двойнственост

Двойнствеността в линейното оптимиране е способ за изследване на линейни задачи с помощта на спомагателни, тясно свързани с тях линейни задачи.

Ще покажем как с всяка линейна задача може да се асоциира друга линейна задача, наречена нейна двойнствена и как поведението на едната от тях (разрешима, неограничена) влияе на поведението на другата.

§9.1. Права задача

Нека е дадена линейна задача в общ вид (L) (вж. § 0.5). Произволна такава задача може да бъде преобразувана в задача за минимум (като се вземе предвид, че $\max\limits_{x\in X} f(x) = -\min\limits_{x\in X}[-f(x)]$ за произволна функция f и произволно множество $X \subset \mathbb{R}^n$), а всички неравенствата в множеството от ограничения могат да бъдат обърнати в посоката $\geqslant$ (ако неравенството е от вида $\leqslant$ то двете му страни се умножават с -1 без да се държи сметка за знака на дясната му страна!).

Така получената задача се нарича *права задача*:

$$(P)\qquad \begin{aligned} \min z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}& \\ \mathbf{a}_i^{\mathrm{T}}\mathbf{x} \geqslant b_i,&\quad i\in I,\\ \mathbf{a}_i^{\mathrm{T}}\mathbf{x} = b_i,&\quad i\in \overline{I},\\ x_j \geqslant 0,&\quad j\in J,\\ x_j \lesseqgtr 0,&\quad j\in \overline{J}. \end{aligned}$$

Тук I е индексно множество, подмножество на множеството $\{1,\ldots,m\}$ от индекси на ограниченията, като $\overline{I} := \{1,\ldots,m\}\setminus I$, а J е индексно множество, подмножество на множеството $\{1,\ldots,n\}$ от индекси на променливите, като $\overline{J} := \{1,\ldots,n\}\setminus J$.

Знакът $\lesseqgtr$ използваме, за да означим това, че върху съответната променлива не е наложено условие за неотрицателност и тя е свободна по знак.

Ако означим с $\mathbf{A}$ матрицата, чиито вектор-редове са векторите $\mathbf{a}_i^{\mathrm{T}}$, $i = 1,\ldots,m$, получаваме $m\times n$ матрицата на ограниченията на задачата, чиито вектор-стълбове, както и преди, ще означаваме с $\mathbf{A}_j$, $j = 1,\ldots,n$. Векторът на целевата функция $\mathbf{c}(c_1,\ldots,c_n)$ и векторът на променливите $\mathbf{x}(x_1,\ldots,x_n)\in\mathbb{R}^n$, а векторът на дясната страна е $\mathbf{b}(b_1,\ldots,b_m)\in\mathbb{R}^m$.

§9.2. Правила за писане на двойнствената задача

От казаното по-горе е очевидно, че на произволна задача на линейното оптимиране може да се гледа като на права задача. На всяка права линейна задача (P) се съпоставя друга линейна задача (DP), която се нарича *двойнствена задача* на задачата (P). Правилото, по което се прави това съпоставяне се нарича *спрягане*. Първо ще напишем съответната на (P) двойнствена задача (DP), а след това ще обясним по-подробно връзката между тях.

$$
(P)\quad
\begin{aligned}
&\min z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}\\
&\mathbf{a}_i^{\mathrm{T}}\mathbf{x} \geqslant b_i, \quad i \in I,\\
&\mathbf{a}_i^{\mathrm{T}}\mathbf{x} = b_i, \quad i \in \overline{I},\\
&\quad x_j \geqslant 0, \quad j \in J,\\
&\quad x_j \lesseqgtr 0, \quad j \in \overline{J},
\end{aligned}
\qquad \Rightarrow \qquad
(DP)\quad
\begin{aligned}
&\max v(\mathbf{y}) = \mathbf{b}^{\mathrm{T}}\mathbf{y}\\
&\quad y_i \geqslant 0, \quad i \in I,\\
&\quad y_i \lesseqgtr 0, \quad i \in \overline{I},\\
&\mathbf{A}_j^{\mathrm{T}}\mathbf{y} \leqslant c_j, \quad j \in J,\\
&\mathbf{A}_j^{\mathrm{T}}\mathbf{y} = c_j, \quad j \in \overline{J}.
\end{aligned}
$$

На всяко от ограниченията $i \in \{1, \dots, m\}$ се съпоставя двойнствена променлива y_i, като ако i-то ограничение е неравенство, то в (DP) на променливата y_i е наложено условие за неотрицателност, а ако е равенство, то y_i е свободна по знак.

Останалите ограничения на (DP) се получават като за всяко $j \in \{1, \dots, n\}$ векторът на двойнствените променливи $\mathbf{y}(y_1, \dots, y_m)$ се умножи със съответния вектор стълб $\mathbf{A}_j$ на матрицата на ограниченията $\mathbf{A}$ (стълбът пред променливата x_j) като полученото скаларно произведение не трябва да надминава c_j (коефициента в целевата функция пред x_j), ако върху x_j в (P) е наложено условие за неотрицателност и трябва да е равно на c_j, ако променливата x_j в (P) е свободна по знак.

Накрая, (DP) е задача за максимум, а целевата ѝ функция е скаларното произведение на вектора на двойнствените променливи $\mathbf{y}$ и вектора $\mathbf{b}$ в дясната страна на ограниченията на (P).

§9.3. Двойка спрегнати задачи

Видяхме как на всяка права задача (P) чрез спрягане се съпоставя двойнствена задача (DP). Разбира се, така получената двойнствената задача (DP) можем да преработим в права задача, като я преобразуваме в задача за минимум и представим неравенствата в посоката $\geqslant$. Ако спрегнем получената права задача по горното правило, получената двойнствена задача ще съвпадне с изходната права задача (P), което ще докажем в

Лема 9.1. *Двойнствената задача на задачата (DP) е правата задача (P).*

Доказателство. Трябва да покажем, че $(DDP) \equiv (P)$. За целта първо преработваме двойнствената задача (DP) така, че да я направим права задача

$$
(DP)\quad
\begin{array}{rl}
\max v(\mathbf{y}) = \mathbf{b}^{\mathrm{T}}\mathbf{y} & \\
y_i \geqslant 0, & i \in I,\\
y_i \lesseqgtr 0, & i \in \overline{I},\\
\mathbf{A}_j^{\mathrm{T}}\mathbf{y} \leqslant c_j, & j \in J,\\
\mathbf{A}_j^{\mathrm{T}}\mathbf{y} = c_j, & j \in \overline{J},
\end{array}
\quad \sim \quad (DP)\quad
\begin{array}{rl}
-\min[-\mathbf{b}^{\mathrm{T}}\mathbf{y}] & \\
y_i \geqslant \ \ 0, & i \in I,\\
y_i \lesseqgtr \ \ 0, & i \in \overline{I},\\
-\mathbf{A}_j^{\mathrm{T}}\mathbf{y} \geqslant -c_j, & j \in J,\\
-\mathbf{A}_j^{\mathrm{T}}\mathbf{y} = -c_j, & j \in \overline{J}.
\end{array}
$$

На така получената права задача по правилото за спрягане пишем съответната двойнствена и опростяваме

$$
(DDP)\quad
\begin{array}{rl}
-\max[-\mathbf{c}^{\mathrm{T}}\mathbf{x}] & \\
-\mathbf{a}_i^{\mathrm{T}}\mathbf{x} \leqslant -b_i, & i \in I,\\
-\mathbf{a}_i^{\mathrm{T}}\mathbf{x} = -b_i, & i \in \overline{I},\\
x_j \geqslant 0, & j \in J,\\
x_j \lesseqgtr 0, & j \in \overline{J},
\end{array}
\quad \sim \quad (P)\quad
\begin{array}{rl}
\min \mathbf{c}^{\mathrm{T}}\mathbf{x} & \\
\mathbf{a}_i^{\mathrm{T}}\mathbf{x} \geqslant b_i, & i \in I,\\
\mathbf{a}_i^{\mathrm{T}}\mathbf{x} = b_i, & i \in \overline{I},\\
x_j \geqslant 0, & j \in J,\\
x_j \lesseqgtr 0, & j \in \overline{J},
\end{array}
$$

за да получим правата задача. □

От лемата е ясно, че всяка от задачите (P) и (DP) се получава от другата чрез спрягане. Затова те още се наричат *двойка спрегнати задачи.*

Нека е дадена задача на линейното оптимиране в общ вид (L). Двойнствената на задачата (L) е двойнствената задача на съответната ѝ права задача (P). На задачата (L) съпоставяме и съответната ѝ канонична задача (K). Очевидно (K) също е права задача (задача за минимум, чиито ограничения са равенства и върху всичките променливи на която е наложено условие за неотрицателност). Двойнствената задача (DK) на задачата (K) се различава от (DP) евентуално само по знаците на някои от променливите. Наистина, да означим с $\boldsymbol{\pi} \in \mathbb{R}^m$ векторът на променливите за задачата (DK). Ако задачата (P) е такава, че векторът в дясната ѝ страна $\mathbf{b}$ е с неотрицателни координати (за което очевидно е достатъчно $b_i \geqslant 0$, за всяко $i \in I$), то съответната ѝ канонична задача (K) също има за двойнствена (DP), т.е. $(DK) \equiv (DP)$ (Проверете!) и тогава $\pi_i = y_i$ за всяко $i = 1, \dots, m$. Ако обаче в задачата (P) има $b_i < 0$ за някое $i \in I$, то при канонизиране i-то ограничение се умножава с -1, което води до това че $\pi_i = -y_i$, т.е. съответните на i-то ограничение двойнствени променливи се различават по знак.

От току що казаното е ясно, че ако намерим решение на (DK), от него лесно получаваме решение на (DP) и обратно. Това позволява да продължим разглежданията по-долу само на двойката спрегнати задачи, състояща се от канонична задача и съответната ѝ двойнствена.

Съответната на каноничната задача

$$
(K)\qquad
\begin{array}{ll}
\min & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}\\
 & \mathbf{A}\mathbf{x} = \mathbf{b}\,,\\
 & \mathbf{x} \geqslant \mathbf{0}\,.
\end{array}
$$

двойнствена задача е

$$(DK)\qquad \begin{aligned} \max \quad & v(\boldsymbol{\pi}) = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ & \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}, \\ & \boldsymbol{\pi} \lesseqqgtr 0. \end{aligned}$$

§9.4. Теореми за двойнственост

За връзката между двойката спрегнати задачи (K) и (DK) ще докажем няколко основни резултата.

Теорема 9.1 (слаба теорема за двойнственост). *Ако* $\mathbf{x}$ *е допустима точка за* (K) *и* $\boldsymbol{\pi}$ *е допустима точка за* (DK)*, то* $\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}^{\mathrm{T}}\mathbf{x}$.

Доказателство. Ако $\mathbf{x}$ е допустима точка за (K), имаме $\mathbf{A}\mathbf{x} = \mathbf{b}$ и $\mathbf{x} \geqslant \mathbf{0}$. Ако $\boldsymbol{\pi}$ е допустима точка за двойнствената задача (DK), то $\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}$. Следователно $\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} = (\mathbf{A}\mathbf{x})^{\mathrm{T}}\boldsymbol{\pi} = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{x}^{\mathrm{T}}\mathbf{c} = \mathbf{c}^{\mathrm{T}}\mathbf{x}$. □

Слабата теорема за двойнственост всъщност твърди, че:

I. ако (DK) е съвместима и $\boldsymbol{\pi}$ е допустима точка за (DK), то $v(\boldsymbol{\pi})$ е долна граница за стойността на целевата функция z на (K) в произволна нейна допустима точка $\mathbf{x}$;

II. ако (K) е съвместима и $\mathbf{x}$ е допустима точка за (K), то $z(\mathbf{x})$ е горна граница за стойността на целевата функция v на (DK) в произволна нейна допустима точка $\boldsymbol{\pi}$.

Като следствие от Теорема 9.1 получаваме

Следствие 9.1. *Нека* $\mathbf{x}^*$ *е допустима точка за* (K)*, а* $\boldsymbol{\pi}^*$ *е допустима точка за* (DK)*. Ако* $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}^*$*, то* $\mathbf{x}^*$ *и* $\boldsymbol{\pi}^*$ *са оптимални решения съответно на* (K) *и на* (DK).

Доказателство. Според слабата теорема за двойнственост за произволна допустима точка $\mathbf{x}$ на (K) и за допустимото $\boldsymbol{\pi}^*$ за (DK) имаме, че $\mathbf{c}^{\mathrm{T}}\mathbf{x} \geqslant \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}^* = \mathbf{c}^{\mathrm{T}}\mathbf{x}^*$ и следователно $\mathbf{x}^*$ е оптимално за (K).

Използвайки същата теорема, за произволна допустима точка $\boldsymbol{\pi}$ на (DK) и за допустимото $\mathbf{x}^*$ за (K) имаме, че $\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}^{\mathrm{T}}\mathbf{x}^* = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}^*$ и следователно $\boldsymbol{\pi}^*$ е оптимално за (DK). □

Съществуват ли обаче допустими точки $\mathbf{x}^*$ за (K) и $\boldsymbol{\pi}^*$ за (DK), които да удовлетворяват условието на това следствие?

Отговор на този въпрос дава:

Теорема 9.2 (силна теорема за двойнственост). (а) *Ако едната от двойката спрегнати задачи* (K) *и* (DK) *е разрешима, то разрешима е и другата задача, като* $\min \mathbf{c}^{\mathrm{T}}\mathbf{x} = \max \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}$.

(б) *Ако едната от двойката спрегнати задачи* (K) *и* (DK) *е неограничена, то другата задача е несъвместима.*

Доказателство. (a) Ще допуснем, че е разрешима задачата (K) и ще покажем, че е разрешима задачата (DK). Обратното се доказва аналогично.

И така, нека $\overline{\mathbf{x}}$ е текущото базисно допустимо решение с базис B за каноничната задача (K), решавана със симплекс метода, т.е.

$$\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix} \geqslant \mathbf{0}.$$

Да разгледаме съответния на $\overline{\mathbf{x}}$ вектор на симплексните множители $\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$ (вж. § 8.2). Имаме, че

$$\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} = [\mathbf{c}_B^{\mathrm{T}}, \mathbf{c}_N^{\mathrm{T}}] \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = [\mathbf{c}_B^{\mathrm{T}}, \mathbf{c}_N^{\mathrm{T}}] \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} = \boldsymbol{\pi}^{\mathrm{T}}\mathbf{b} = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}\,.$$

Това означава, че на всяка симплексна итерация за текущия връх $\overline{\mathbf{x}}$ и за съответния му вектор на симплексните множители $\boldsymbol{\pi}$ е в сила равенството

$$z(\overline{\mathbf{x}}) = \mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} = v(\boldsymbol{\pi}).$$

Нека сега $\overline{\mathbf{x}}^*$ с базис B е полученото със симплекс метода оптимално базисно допустимо решение за (K). Съответният му вектор на симплексните множители е $\boldsymbol{\pi}^* = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$. Тъй като $\overline{\mathbf{x}}^*$ е оптимално, то за него е в сила критерият за оптималност

$$\overline{\mathbf{c}}_N^{\mathrm{T}} = \mathbf{c}_N^{\mathrm{T}} - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{N} \geqslant \mathbf{0}$$

или критерият за оптималност на $\overline{\mathbf{x}}^*$ е еквивалентен на

$$\mathbf{c}_N^{\mathrm{T}} \geqslant \boldsymbol{\pi}^{*\mathrm{T}}\mathbf{N}.$$

Тъй като очевидно $\mathbf{c}_B^{\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{B} = \boldsymbol{\pi}^{*\mathrm{T}}\mathbf{B}$, получаваме

$$\mathbf{c}^{\mathrm{T}} = [\mathbf{c}_B^{\mathrm{T}}, \mathbf{c}_N^{\mathrm{T}}] \geqslant [\boldsymbol{\pi}^{*\mathrm{T}}\mathbf{B}, \boldsymbol{\pi}^{*\mathrm{T}}\mathbf{N}] = \boldsymbol{\pi}^{*\mathrm{T}}[\mathbf{B}|\mathbf{N}] = \boldsymbol{\pi}^{*\mathrm{T}}\mathbf{A},$$

или $\mathbf{c} \geqslant \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi}^*$, т.е. получаваме че $\boldsymbol{\pi}^*$, съответстващ на оптималното $\overline{\mathbf{x}}^*$, е допустим за задачата (DK).

Да обобщим: ако $\overline{\mathbf{x}}$ с базис B е текущото базисно допустимо решение за симплексната итерация, съответният му вектор на симплексните множители $\boldsymbol{\pi}$ удовлетворява условието $\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}$, но само в случая, когато всички относителни оценки спрямо базиса B са неотрицателни, векторът $\boldsymbol{\pi}$ е допустим за двойнствената задача. Казано с други думи на всяка итерация симплекс алгоритъмът поддържа допустимостта на решението $\overline{\mathbf{x}}$ за задачата (K) и условието $\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}$, като се стреми да постигне допустимост на $\boldsymbol{\pi}$ за задачата (DK).

Показахме, че ако $\overline{\mathbf{x}}^*$ е оптимално базисно допустимо решение на (K) с базис B, то $\boldsymbol{\pi}^{*\mathrm{T}} := \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$ е допустимо решение за двойнствената задача (DK), което удовлетворява $\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}^* = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}^*$. В този случай от Следствие 9.1 имаме, че $\boldsymbol{\pi}^*$ е оптимално решение за (DK), с което доказателството на (a) е приключено.

(б) следва директно от Теорема 9.1 с допускане на противното. □

Да отбележим, че обратното на Теорема 9.2(б) твърдение в общия случай не е вярно, т.е. ако едната от двойката спрегнати задачи е несъвместима, то от това **не следва** че другата задача е неограничена. Възможно е и двете задачи да бъдат несъвместими, както се вижда от следния

Пример. В двойката спрегнати задачи

$$(K)\quad \begin{array}{ll} \min & z(\mathbf{x}) = -x_1 - x_2 \\ & x_1 - x_2 = 1, \\ & x_1 - x_2 = 0, \\ & x_1 \geqslant 0\,,\ x_2 \geqslant 0, \end{array} \qquad (DK)\quad \begin{array}{ll} \max & v(\boldsymbol{\pi}) = \pi_1 \\ & \pi_1 + \pi_2 \leqslant -1, \\ & -\pi_1 - \pi_2 \leqslant -1, \\ & \pi_1 \lesseqgtr 0,\ \pi_2 \lesseqgtr 0 \end{array}$$

и двете са несъвместими.

Както вече казахме, вече имаме достатъчно средства да направим доказателството на Теорема 6.1 от § 6.

Теорема 6.1. *Ако задачата (K) е разрешима, то съществува число $M_0 > 0$, такова че за всяко $M \geqslant M_0$ съответната (M)-задача е разрешима и $\mathbf{y}$-координатите на всички нейни оптимални базисни допустими решения са нули.*

Доказателство. Нека $\mathbf{x}$ е допустима точка за (K). Тогава $(\mathbf{x}, \mathbf{0})$ е допустима точка за всяка (M)-задача

$$(M)\qquad \begin{array}{ll} \min & z_M(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{n} c_j x_j + M \sum_{i=1}^{m} y_i \\ & \mathbf{A}\mathbf{x} + \mathbf{y} = \mathbf{b} \quad (\mathbf{b} \geqslant \mathbf{0}), \\ & \mathbf{x} \geqslant \mathbf{0}, \quad \mathbf{y} \geqslant \mathbf{0}, \end{array}$$

(допустимото множество на (M)-задачите не зависи от числото M) и следователно M-задачата не е несъвместима.

Тъй като (K) е разрешима, от Силната теорема за двойнственост следва, че задачата (DK) е разрешима. Нека $\boldsymbol{\pi}$ е произволна допустима точка за задачата $(DK)\ \begin{array}{l} \max \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}. \end{array}$ Да положим $\overline{M} := \max\limits_{1 \leqslant i \leqslant m} \pi_i$. Тогава за $M \geqslant \overline{M}$ $\boldsymbol{\pi}$ ще бъде допустима точка и за задачата

$$(DM)\quad \begin{array}{l} \max \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c} \\ \pi_i \leqslant M,\ i = 1, \dots, m. \end{array}$$

Следователно за $M \geqslant \overline{M}$ задачата (DM) е с непразно допустимо множество. Съгласно Слабата теорема за двойнственост, съответната (M)-задача не е неограничена.

Следователно, за $M \geqslant \overline{M}$ съответната (M)-задача е разрешима.

За доказателството на втората част на теоремата е достатъчно да предположим, че (K) има непразно допустимо множество.

Нека $\mathbf{x}$ е произволна допустима точка за (K).

Нека $(\overline{\mathbf{x}}_1, \overline{\mathbf{y}}_1)$ е базисно допустимо решение за (M), такова че $\overline{\mathbf{y}}_1 \neq \mathbf{0}$. Тогава съществува $M_1 \geqslant \overline{M}$ такова че $z_{M_1}(\overline{\mathbf{x}}_1, \overline{\mathbf{y}}_1) = \mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}_1 + M_1(\sum_{i=1}^{m} \overline{y}_i) >$

$\mathbf{c}^{\mathrm{T}}\mathbf{x} = z_{M_1}(\mathbf{x}, \mathbf{0})$ и следователно $(\overline{\mathbf{x}}_1, \overline{\mathbf{y}}_1)$ няма да бъде оптимално за (M)-задача с $M \geqslant M_1$.

Тъй като базисните допустими решения $(\overline{\mathbf{x}}, \overline{\mathbf{y}})$ (и в частност тези, за които $\overline{\mathbf{y}} \neq \mathbf{0}$) са краен брой, то ще съществува $\widetilde{M}$ такова че за всяко $M \geqslant \widetilde{M}$ оптималните базисни допустими решения за (M)-задачата (ако съществуват) ще бъдат от вида $(\overline{\mathbf{x}}, \mathbf{0})$.

Полагаме $M_0 := \max\{\overline{M}, \widetilde{M}\}$ и доказателството приключва. $\square$

§10. Двойнственост в линейното оптимиране. Лема на Фаркаш. Условия за допълнителност. Икономическа интерпретация на двойнствените променливи

С теоремите за двойнственост, които доказахме в § 9 може да бъде доказан и един важен резултат за системи линейни уравнения и неравенства, известен като

§10.1. Лема на Фаркаш

Този резултат е много добре известен и с широка сфера на приложение, а ние ще го използваме съществено в курса по Математическо оптимиране—2.

Теорема 10.1 (Лема на Фаркаш). *Системата*

$$\text{(I)} \qquad \mathbf{A}\mathbf{x} = \mathbf{b}\,, \quad \mathbf{x} \geqslant \mathbf{0}\,,$$

има решение, тогава и само тогава, когато системата

$$\text{(II)} \qquad \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{0}, \quad \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} > 0$$

няма решение.

Доказателство. Да разгледаме следната двойка спрегнати задачи

$$\begin{array}{llcll} (P) & \min\ \mathbf{0}^{\mathrm{T}}\mathbf{x} & \qquad & (D) & \max\ \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, & & & \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{0}\,. \\ & \mathbf{x} \geqslant \mathbf{0}\,, & & & \end{array}$$

Нека (I) има решение и нека $\mathbf{x}$ е нейно решение. Да допуснем, че (II) също има решение и да вземем нейно решение $\boldsymbol{\pi}$. Тогава

$$0 < \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} = (\mathbf{A}\mathbf{x})^{\mathrm{T}}\boldsymbol{\pi} = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant 0,$$

което води до противоречие.

Нека сега (II) няма решение, т.е. за всяко $\boldsymbol{\pi}$, такова че $\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{0}$ имаме, че $\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \leqslant 0$. Това означава, че задачата (D) има непразно допустимо множество ($\boldsymbol{\pi} = \mathbf{0}$ е нейна допустима точка) и върху него целевата ѝ функция е ограничена отгоре от нула. Следователно (D) е разрешима. От Силната теорема за двойнственост, (P) също е разрешима, в частност системата (I) има решение. ■

Лемата на Фаркаш, доказана през 1902 г. стои в основата на развитието на линейното оптимиране и често се използва, за да бъдат доказани с нейна помощ теоремите за двойнственост, вместо както ние постъпихме. Геометрично Лемата на Фаркаш твърди, че е вярно точно едно от следните твърдения:

(I) векторът $\mathbf{b}$ е положителна линейна комбинация на вектор-стълбовете на $\mathbf{A}$; или

(II) съществува вектор $\boldsymbol{\pi}$, който сключва неостри ъгли с всички вектор-стълбове на $\mathbf{A}$ и сключва остър ъгъл с $\mathbf{b}$.

§10.2. Условия за допълнителност

Ще покажем, че когато линейна оптимизационна задача има икономическа интерпретация, то икономическа интерпретация имат както променливите на нейната двойнствена задача, така и самата двойнствена задача.

За целта ще изясним по-подробно връзката между променливите и ограниченията на следната двойка спрегнати задачи:

$$(P') \quad \begin{aligned} &\min \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ &\mathbf{A}\mathbf{x} \geqslant \mathbf{b}\,, \\ &\mathbf{x} \geqslant \mathbf{0}\,. \end{aligned} \qquad\qquad (D') \quad \begin{aligned} &\max \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ &\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}, \\ &\boldsymbol{\pi} \geqslant \mathbf{0}. \end{aligned}$$

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 една теорема, която характеризира оптималните решения на тази двойка спрегнати задачи.

Теорема 10.2 (Условия за допълнителност). *Нека* $\mathbf{x}$ *е допустима точка за* (P')*, а* $\boldsymbol{\pi}$ *е допустима точка за* (D')*. Необходими и достатъчни условия* $\mathbf{x}$ *и* $\boldsymbol{\pi}$ *да бъдат оптимални решения на* (P') *и* (D') *са*

$$\mathbf{x}^{\mathrm{T}}(\mathbf{c} - \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi}) = 0 \tag{10.1}$$

и

$$\boldsymbol{\pi}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0\,. \tag{10.2}$$

Доказателство. За допустимото $\mathbf{x}$ за (P') и за допустимото $\boldsymbol{\pi}$ за (D') имаме

$$\mathbf{s} \equiv \mathbf{A}\mathbf{x} - \mathbf{b} \geqslant \mathbf{0}, \quad \mathbf{x} \geqslant \mathbf{0}, \quad \text{и} \quad \mathbf{w} \equiv \mathbf{c} - \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \geqslant \mathbf{0},\ \boldsymbol{\pi} \geqslant \mathbf{0}\,. \tag{10.3}$$

Следователно,

$$0 \leqslant (\mathbf{c} - \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi})^{\mathrm{T}}\mathbf{x} = \mathbf{c}^{\mathrm{T}}\mathbf{x} - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{A}\mathbf{x} \quad \text{или} \quad \boldsymbol{\pi}^{\mathrm{T}}\mathbf{A}\mathbf{x} \leqslant \mathbf{c}^{\mathrm{T}}\mathbf{x},$$

$$0 \leqslant \boldsymbol{\pi}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) = \boldsymbol{\pi}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{b} \quad \text{или} \quad \boldsymbol{\pi}^{\mathrm{T}}\mathbf{b} \leqslant \boldsymbol{\pi}^{\mathrm{T}}\mathbf{A}\mathbf{x},$$

откъдето

$$\boldsymbol{\pi}^{\mathrm{T}}\mathbf{b} \leqslant \boldsymbol{\pi}^{\mathrm{T}}\mathbf{A}\mathbf{x} \leqslant \mathbf{c}^{\mathrm{T}}\mathbf{x}. \tag{10.4}$$

Ако условията (10.1) и (10.2) са в сила, то неравенствата в (10.4) се превръщат в равенства и оптималността на $\mathbf{x}$ и $\boldsymbol{\pi}$ идва от Следствие 9.1.

Обратно, ако $\mathbf{x}$ и $\boldsymbol{\pi}$ са оптимални решения на (P') и (D'), от Силната теорема за двойнственост следва, че $\mathbf{c}^{\mathrm{T}}\mathbf{x} = \boldsymbol{\pi}^{\mathrm{T}}\mathbf{b}$, откъдето в (10.4) имаме равенства вместо неравенства, което означава че са изпълнени условията (10.1) и (10.2). ■

За двойката спрегнати задачи (P') и (D'), поради неотрицателността на променливите $\mathbf{x}$ и $\boldsymbol{\pi}$, както и на съответните допълнителните променливи $\mathbf{s}$ и $\mathbf{w}$ (вж. (10.3)) условията за допълнителност (10.1) и (10.2) могат да се изкажат в следната полезна форма.

$$\begin{aligned} &w_j \equiv (\mathbf{c} - \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi})_j = 0 \quad \text{или } x_j = 0\,, \quad \text{за } j = 1, \dots, n\,, \\ &s_i \equiv (\mathbf{A}\mathbf{x} - \mathbf{b})_i = 0 \quad \text{или } \pi_i = 0\,, \quad \text{за } i = 1, \dots, m\,. \end{aligned} \tag{10.5}$$

От условията за допълнителност в този им вид, е ясно, че допустими точки $\mathbf{x}$ за правата (P') и $\boldsymbol{\pi}$ за двойнствената задача (D') са техни оптимални решения, тогава и само тогава, когато

(1) ако допълнителна променлива $w_j > 0$, то съответната ѝ променлива $x_j = 0$ за всяко j; ако допълнителна променлива $s_i > 0$, то съответната ѝ променлива $\pi_i = 0$ за всяко i;

(2) ако променлива $x_j > 0$, то съответната ѝ допълнителна променлива $w_j = 0$ за всяко j; ако променлива $\pi_i > 0$, то съответната ѝ допълнителна променлива $s_i = 0$ за всяко i.

Да отбележим, че за двойката спрегнати задачи (K) и (DK), която разглеждахме в § 9 от значение е само условието (10.2), тъй като (10.1) е вярно за всяко допустимо решение $\mathbf{x}$ на задачата (K).

§10.3. Икономическа интерпретация на двойнствените променливи

Често в линейните задачи, описващи икономически модели, десните страни на ограниченията представляват количества от определени ресурси.

За да конкретизираме разглежданията, да разгледаме задачата за оптимално разпределение на ресурси, наречена още задача за смесено производство: дадена фирма е в състояние да произведе n на брой различни стоки използвайки за целта m на брои различни ресурси. Нека означим с c_j печалбата (приходите 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 мениджъра на фирмата е да състави план на производството (т.е. да определи какво количество от всяка стока да бъде произведено), така че общата печалба да бъде максимална.

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

$$(D') \qquad \max z(\mathbf{x}) = \sum_{j=1}^{n} c_j x_j$$
$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i\,, \quad i = 1, 2, \ldots, m\,,$$
$$x_j \geqslant 0\,, \quad j = 1, 2, \ldots, n\,.$$

Записана в матричен вид това е задача от вида (D'), който разглеждахме по-горе и затова я означаваме по същия начин. Координатите b_i на вектора $\mathbf{b}$ в дясната страна на (D') са количествата налични ресурси i, $i = 1, \ldots, m$. Да предположим, че

$$\mathbf{x}^* = \begin{bmatrix} \mathbf{x}_B^* \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix} \geqslant \mathbf{0}$$

е оптимално базисно допустимо решение на (D'). Ако предположим, че $\mathbf{x}^*$ е неизродено, т.е. $\mathbf{x}_B^* > \mathbf{0}$, то малка промяна $\Delta\mathbf{b}$ на вектора $\mathbf{b}$ не би променила оптималния базис $\mathbf{B}$ (тъй като относителните оценки $\overline{\mathbf{c}}_N = \mathbf{c}_N - \mathbf{c}_B\mathbf{B}^{-1}\mathbf{N}$ не се влияят от изменение на $\mathbf{b}$, то малко изменение в дясната част $\mathbf{b}$ няма да повлияе на оптималността, а само на допустимостта на решението). Следователно, ако векторът $\mathbf{b}$ се замени с вектора $\mathbf{b}+\Delta\mathbf{b}$, новото оптимално решение на (D') за малки по норма $\Delta\mathbf{b}$ ще бъде

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

а оптималната стойност на целевата функция ще се измени с

$$\Delta z = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\Delta\mathbf{b} = \boldsymbol{\pi}^{*\mathrm{T}}\Delta\mathbf{b},$$

където $\boldsymbol{\pi}^{*\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$. От доказателството на Теорема 4.2 знаем, че $\boldsymbol{\pi}^*$ е оптимално решение на двойнствената на (D') задача, която е (P').

Оттук е ясно, че на оптималната стойност π_i^* на двойнствената променлива π_i може да се гледа като на *неявна цена* или *цена в сянка на i-ия ресурс*, тъй като π_i^* е увеличението на оптималната стойност на целевата функция, което се получава при увеличение с единица на количеството на този ресурс. Тази икономическа интерпретация е много полезна понеже тя показва кои са ценните за даденото производство ресурси.

Казваме, че ресурсът i е дефицитен, ако i-то неравенство се изпълнява като равенство от оптималното решение (т.е. $\sum_{j=1}^{n} a_{ij}x_j^* = b_i$ или $s_i^* = 0$) и недефицитен ако i-то неравенство се изпълнява като строго неравенство от оптималното решение (т.е. $\sum_{j=1}^{n} a_{ij}x_j^* < b_i$ или $s_i^* > 0$).

Ако i е недефицитен ресурс, то от условията за допълнителност следва, че неявната му цена π_i^* е нула и малко изменение на количеството b_i на този ресурс няма да доведе до изменение на оптималната стойност на z. Разбира се смисъл от увеличаване на количеството на недефицитен ресурс няма, а е разумно количеството му да бъде намалено с разликата $s_i^* = b_i - \sum_{j=1}^n a_{ij} x_j^*$.

Ако i е дефицитен ресурс, то от това, че неявната му цена π_i^* е неотрицателна следва, че увеличаване на количеството му (т.е. $\Delta b_i > 0$) би довело до подобряване на оптималната стойност на целевата функция, а намаляване на количеството му (т.е. $\Delta b_i < 0$) би довело до влошаване на оптималната стойност на целевата функция.

Следователно, за подобряване на оптималната стойност на целевата функция е необходимо количествата на дефицитните ресурси да се увеличават, като очевидно, до най-голямо подобрение би се стигнало при увеличаването на дефицитния ресурс с най-голяма неявна цена.

Така определените неявни цени показват, че ценни за производството са дефицитните ресурси. Те дават количествено изражение на относителната ценност за дефицитния ресурс за даденото производство – колкото по-голяма е неявната му цена, толкова по-ценнен е той за производството.

Неявните цени на ресурсите са полезни и в случая, когато при вече намерено оптимално решение трябва да се вземе решение дали да се въведе ново производство.

Като пример ще разгледаме задачата за диета: 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 дневния 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:

$$\begin{aligned} (P') \qquad \min \quad & \sum_{j=1}^n c_j x_j \\ & \sum_{j=1}^n a_{ij} x_j \geqslant b_i \,, \quad i = 1, 2, \ldots, m \,, \\ & x_j \geqslant 0 \,, \quad j = 1, 2, \ldots, n \end{aligned}$$

която в матричен вид е точно от вида (P') и затова я означаваме по същия начин. Очевидно нейната двойнствена е от вида (D').

Да предположим, че след като сме решили (P') и сме получили оптималната диета $\mathbf{x}^*$, на пазара се е появила нова храна $k = n + 1$ – очевидно невключена в нея. 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 ѝ цена бъде c_k. Тъй като оптималната стойност π_i^* на i-та двойнствена променлива се интерпретира като неявна цена на единица от i-та хранителна съставка, то хранителните вещества, осигурявани от единица от новата храна сумарно имат следната неявна цена $\sum_{i=1}^m \pi_i^* a_{ik}$. Следователно, ако $c_k < \sum_{i=1}^m \pi_i^* a_{ik}$, то е изгодно новата храна да бъде закупена и включена в диетата, а ако $c_k \geqslant \sum_{i=1}^m \pi_i^* a_{ik}$ то включването на новата храна в диетата не е оправдано. Защо? Очевидно това, което пресметнахме по-горе е относителната оценка на новата променлива (храна) $\bar{c}_k = c_k - \sum_{i=1}^m \pi_i^* a_{ik}$ спрямо базиса на оптималното до момента решение $\mathbf{x}^*$. В първия случай $\bar{c}_k < 0$ и $\boldsymbol{\pi}^*$ не е допустимо за новата двойнствена задача, понеже не удовлетворява новото двойнствено ограничение. Прилагането на симплекс метода за ре-оптимизация води до незабавно избиране на новата храна за влизане в базиса (включване в диетата) тъй като тя има отрицателна относителна оценка. Във втория случай имаме, че $\bar{c}_k \geqslant 0$, т.е. новата променлива е с неотрицателна относителна оценка и намерената оптимална диета си остава оптимална ($\boldsymbol{\pi}^*$ остава допустимо за новата двойнствена задача, т.к. удовлетворява и новото двойнствено ограничение).

§10.4. Икономическа интерпретация на двойнствената задача

За да се убедим, че често не само двойнствените променливи, а и самата двойнствена задача може да има икономическа интерпретация нека да разгледаме отново задачата за диета (P'). Следователно, нейната двойнствена е във формата (D'). Можем да разгледаме задачата (D') като задача на конкурент на бакалина, от който този, който прави диетата, купува храните. 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 $\pi_1, \dots, \pi_m$, които да са неорицателни и да са конкурентни на цените c_j на бакалина, т.е. да удовлетворяват

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

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

§11. Двойнствен симплекс метод

Ще разгледаме вариант на симплекс метода, който е разработен за справяне със случаите, когато при вече намерено оптимално решение на канонична задача

$$(K) \qquad \begin{aligned} &\min z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ &\mathbf{A}\mathbf{x} = \mathbf{b} \\ &\mathbf{x} \geqslant \mathbf{0}, \end{aligned}$$

където $\mathbf{A}$ е $(m \times n)$ матрица, $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m_+$ се налага да бъде въведено ново ограничение. Ако намереното оптимално решение не удовлетворява новото ограничение се налага реоптимизация. Интелигентен начин на справяне с тази ситуация предлага така нареченият двойнствен симплекс метод, създаден от Лемке през 1954 г.

Дефиниция 11.1. *Базисно решение $\overline{\mathbf{x}}$ за задачата (K) с базис B, $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix}$ се нарича псевдо оптимално, ако относителните оценки на небазисните променливи са неотрицателни (т.е. ако $\overline{\mathbf{c}}_N = \mathbf{c}_N - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{N} \geqslant \mathbf{0}$).*

С други думи казано, базисно решение $\overline{\mathbf{x}}$ е псевдо оптимално, ако съответният му вектор на симплексните множители $\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$ е допустим за двойнствената задача (DK)

$$\begin{aligned} &\max\ v(\boldsymbol{\pi}) = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ &\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}. \end{aligned}$$

Да отбележим, че е възможно едно псевдо оптимално базисно решение $\overline{\mathbf{x}}$ да не бъде оптимално решение на (K), тъй като $\overline{\mathbf{x}}$ може да не е допустимо (т.е. може да не удовлетворява условието $\overline{\mathbf{x}} \geqslant \mathbf{0}$) и, че ако псевдо оптимално решение $\overline{\mathbf{x}}$ е и допустимо, то е оптимално решение на (K) (вж. теоремите за двойнственост).

Да вземем псевдо оптимално базисно решение $\overline{\mathbf{x}}$, което не е допустимо за (K), т.е. $\overline{\mathbf{x}} \not\geqslant \mathbf{0}$. Такова решение например се получава при добавяне на ново $(m+1)$-во ограничение неравенство към ограниченията на (K). Нека $\mathbf{x}^*$ е оптималното базисно допустимо решение на (K). Да положим $k = m+1$ и да добавим ограничението $\displaystyle\sum_{j=1}^{n} a_{kj}x_j \leqslant b_k$. Ако $\mathbf{x}^*$ удовлетворява новото ограничение, няма какво повече да правим, тъй като то си остава оптимално. Ако обаче новото ограничение не се удовлетворява от $\mathbf{x}^*$, т.е. ако $\displaystyle\sum_{j=1}^{n} a_{kj}x_j^* > b_k$, то превръщаме новото ограничение в равенство $\displaystyle\sum_{j=1}^{n} a_{kj}x_j + s_k = b_k$ с добавяне на неотрицателна допълнителна променлива $s_k \geqslant 0$. Ясно е, че оптималният базис за изходната задача (K) и новата допълнителна променлива образуват базис за новата задача. За този базис е изпълнен критерият за оптималност (небазисните променливи и относителните им оценки са същите), но базисното решение, което му съответства $(\mathbf{x}^*, s_k^*)$ е недопустимо тъй като в него стойността на новата допълнителна променлива е отрицателна $(s_k^* = b_k - \sum_{j=1}^{n} a_{kj}x_j^* < 0)$.

Двойнственият симплекс метод е създаден, за да разрешава точно такава ситуация. Също както симплекс метода (който вече ще наричаме *прав симплекс метод*) той решава задачата (K) 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.

§11.1. Описание на стъпките на итерацията на двойнствения симплекс метод

За да опишем итерацията на двойнствения симплекс метод, да допуснем, че решаваме каноничната задача (K) и че базисът B на текущото базисно решение $\overline{\mathbf{x}}$ се състои от първите m променливи, т.е. $\mathbf{B} = [\mathbf{A}_1, \dots \mathbf{A}_m]$ и $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b}$. Нека $\overline{\mathbf{x}}$ е псевдо оптимално, т.е. $\overline{\mathbf{c}}_N^{\mathrm{T}} = \mathbf{c}_N^{\mathrm{T}} - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{N} \geqslant \mathbf{0}$. Ако $\overline{\mathbf{x}}_B \ngeqslant \mathbf{0}$, векторът $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \mathbf{0} \end{bmatrix}$ съответства на псевдо оптимален, но недопустим връх на многостенното множество от допустими решения на задачата (K).

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 извеждане на базисната променлива x_p от базиса. Изборът на това коя небазисна променлива x_q да влезе в базиса на мястото на x_p се прави така, че да се поддържа псевдо оптималност на текущото базисно решение, т.е. да се поддържа критерият за оптималност.

От базисния вид на задачата (K_B) спрямо базиса B на $\overline{\mathbf{x}}$ за произволно решение $\mathbf{x}$ на системата $\mathbf{A}\mathbf{x} = \mathbf{b}$ имаме, че е от вида

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j \in N} x_j \mathbf{d}_j, \tag{11.1}$$

където $\mathbf{d}_j = \begin{bmatrix} -\mathbf{w}_j \\ \mathbf{e}_j \end{bmatrix}$ за $\mathbf{w}_j = \mathbf{B}^{-1}\mathbf{A}_j$, а стойността на целевата функция в него се изразява като

$$z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j \in N} \overline{c}_j x_j. \tag{11.2}$$

Ако изберем небазисната променлива x_q за влизане в базиса на мястото на x_p, то новият базис ще бъде $B' = B \backslash \{p\} \cup \{q\}$, а новите небазисни индекси ще бъдат $N' = N \setminus \{q\} \cup \{p\}$. Нека означим базисното решение с базис B' с $\overline{\mathbf{x}}'$.

За да изразим стойностите на целевата функция z в произволно решение $\mathbf{x}$ на системата $\mathbf{Ax} = \mathbf{b}$ спрямо новите небазисни променливи е достатъчно да изразим променливата x_q от (11.1) чрез променливите с индекси в N' и да заместим в (11.2).

И така, от съответното на променливата x_p уравнение в (11.1), имаме че

$$x_p = \overline{x}_p - \sum_{j\in N} w_{pj}x_j = \overline{x}_p - \sum_{j\in N\setminus\{q\}} w_{pj}x_j - w_{pq}x_q$$

и за $w_{pq} \neq 0$ можем да изразим променливата x_q като

$$x_q = \frac{\overline{x}_p}{w_{pq}} - \frac{x_p}{w_{pq}} - \sum_{j\in N} \frac{w_{pj}}{w_{pq}} x_j,$$

след което я заместваме в (11.2), за да получим

$$z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j\in N\setminus q} \overline{c}_j x_j + \overline{c}_q \left(\frac{\overline{x}_p}{w_{pq}} - \frac{x_p}{w_{pq}} - \sum_{j\in N} \frac{w_{pj}}{w_{pq}} x_j \right) =$$

$$\left\{ z(\overline{\mathbf{x}}) + \overline{c}_q \frac{\overline{x}_p}{w_{pq}} \right\} + \sum_{j\in N\setminus q} \left(\overline{c}_j - \overline{c}_q \frac{w_{pj}}{w_{pq}} \right) x_j - \frac{\overline{c}_q}{w_{pq}} x_p.$$

Тъй като $\overline{\mathbf{x}}'$, разбира се, е решение на системата, то от последното равенство за $\mathbf{x} = \overline{\mathbf{x}}'$ получаваме

$$z(\overline{\mathbf{x}}') = z(\overline{\mathbf{x}}) + \overline{c}_q \frac{\overline{x}_p}{w_{pq}} + \sum_{j\in N\setminus q} \left(\overline{c}_j - \overline{c}_q \frac{w_{pj}}{w_{pq}} \right) \overline{x}'_j - \frac{\overline{c}_q}{w_{pq}} \overline{x}'_p = z(\overline{\mathbf{x}}) + \overline{c}_q \frac{\overline{x}_p}{w_{pq}}$$

(като помним, че $\overline{x}'_j = 0$ за $j \in N \setminus \{q\} \cup \{p\} = N'$, тъй като това са небазисни променливи за $\overline{\mathbf{x}}'$). Следователно изразът във фигурните скоби при изразяването на $z(\mathbf{x})$ е точно $z(\overline{\mathbf{x}}')$ и заествайки го там получаваме

$$z(\mathbf{x}) = z(\overline{\mathbf{x}}') + \sum_{j\in N\setminus q} \left(\overline{c}_j - \overline{c}_q \frac{w_{pj}}{w_{pq}} \right) x_j - \frac{\overline{c}_q}{w_{pq}} x_p. \tag{11.3}$$

От друга страна, от базисния вид на задачата $(K_{B'})$ спрямо базиса B' на $\overline{\mathbf{x}}'$ за произволно решение $\mathbf{x}$ на системата $\mathbf{Ax} = \mathbf{b}$, имаме че

$$z(\mathbf{x}) = z(\overline{\mathbf{x}}') + \sum_{j\in N'} \overline{c}'_j x_j, \tag{11.4}$$

където $\overline{c}'_j$ са относителните оценки спрямо новите небазисни променливи $j \in N' = N \setminus \{q\} \cup \{p\}$.

От (11.3) и (11.4) получаваме че относителните оценки спрямо новите небазисни променливи са

$$\overline{c}'_p = -\frac{\overline{c}_q}{w_{pq}}$$
$$\overline{c}'_j = \overline{c}_j - \overline{c}_q \frac{w_{pj}}{w_{pq}}, \quad j \in N \setminus q.$$

От рекурентната връзка между относителните оценки следва, че за да бъдат новите относителни оценки на небазисни променливи неотрицателни (т.е. $\overline{\mathbf{x}}'$ да бъде псевдо оптимално) индексът $q \in N$ трябва да се избере така, че

$$0 \leqslant -\frac{\overline{c}_q}{w_{pq}}$$

т.е. w_{pq} трябва да бъде отрицателно число и освен това

$$-\frac{\overline{c}_q}{w_{pq}} \leqslant -\frac{\overline{c}_j}{w_{pj}}, \qquad \text{за всяко } w_{pj} < 0,\ j \in N.$$

С други думи казано, q се избира като този небазисен индекс, за който

$$\frac{\overline{c}_q}{w_{pq}} = \max\left\{ \frac{\overline{c}_j}{w_{pj}} :\ w_{pj} < 0, j \in N \right\}.$$

Какво се случва, ако такъв избор не може да бъде направен, т.е. ако $w_{pj} \geqslant 0$ за всяко $j \in N$? В такъв случай допустимото множество на задачата (K) е празното. Наистина, след като означим вектора $\mathbf{u}_p^{\mathrm{T}} := \mathbf{e}_p^{\mathrm{T}}\mathbf{B}^{-1}$, забелязваме че $0 \leqslant w_{pj} = (\mathbf{B}^{-1}\mathbf{A}_j)_p = \mathbf{e}_p^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_j = \mathbf{u}_p\mathbf{A}_j$ за всеки небазисен индекс $j \in N$, т.е. $\mathbf{u}_p^{\mathrm{T}}\mathbf{N} \geqslant \mathbf{0}$. Същевременно $\mathbf{u}_p^{\mathrm{T}}\mathbf{B} = \mathbf{e}_p^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{B} = \mathbf{e}_p^{\mathrm{T}} \geqslant \mathbf{0}$, откъдето $\mathbf{u}_p^{\mathrm{T}}\mathbf{A} = \mathbf{u}_p^{\mathrm{T}}[\mathbf{B} \mid \mathbf{N}] = [\mathbf{u}_p^{\mathrm{T}}\mathbf{B}, \mathbf{u}_p^{\mathrm{T}}\mathbf{N}] \geqslant \mathbf{0}$ и $\mathbf{u}_p^{\mathrm{T}}\mathbf{A}$ е вектор с неотрицателни координати. Тогава уравнението $\mathbf{A}\mathbf{x} = \mathbf{b}$ не може да има неотрицателно решение $\mathbf{x} \geqslant \mathbf{0}$, тъй като в такъв случай $0 \leqslant \mathbf{u}_p^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{u}_p^{\mathrm{T}}\mathbf{b} = \mathbf{e}_p^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} = \overline{x}_p < 0$ и имаме противоречие. Следователно каноничната задача (K) е с празно допустимо множество, когато $w_{pj} \geqslant 0$ за всяко $j \in N$.

§11.2. Алгоритъм на двойнствения симплекс метод

(0) Намира се псевдо оптимално базисно решение $\overline{\mathbf{x}}$ на каноничната задача (K) с базисна матрица $\mathbf{B} = [\mathbf{A}_{j_1}, \ldots, \mathbf{A}_{j_m}]$. Намират се базисните координати $\overline{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b}$.

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

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

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

(3) Проверка за празно допустимо множество: Намираме $\mathbf{u}_p$ като решим системата $\mathbf{B}^{\mathrm{T}}\mathbf{u}_p = \mathbf{e}_p$ относно $\mathbf{u}_p$ и след това пресмятаме $w_{pj} = \mathbf{u}_p^{\mathrm{T}}\mathbf{A}_j$, за всяко $j\ inN$. Ако $w_{pj} \geqslant 0$ за всяко $j \in N$, то КРАЙ; задачата (K) е с празно допустимо множество.

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

$$\frac{\bar{c}_q}{w_{pq}} = \max\left\{\frac{\bar{c}_j}{w_{pj}} :\ w_{pj} < 0, j \in N\right\},$$

ако максимумът се достига за няколко небазисни индекса.

(5) Обновяване на базисното решение, на базиса и на базисната матрица: намираме $\mathbf{w}_q$ като решим

$$\mathbf{B}\mathbf{w}_q = \mathbf{A}_q\,.$$

Полагаме

$$\begin{gathered}
\bar{x}_q \leftarrow \bar{t} = \bar{x}_{j_p}/w_{pq}\,,\\
\bar{x}_{j_i} \leftarrow \bar{x}_{j_i} - \bar{t}w_{iq}\,, \quad \text{за } 1 \leqslant i \leqslant m,\ i \neq p\\
\mathbf{B} \leftarrow \mathbf{B} + (\mathbf{A}_q - \mathbf{A}_{j_p})\mathbf{e}_p^{\mathrm{T}}\,,\\
B \leftarrow B \cup \{q\} \setminus \{j_p\}\,,\\
j_p \leftarrow q\,,
\end{gathered}$$

и отиваме на стъпка (1).

Ако задачата (K) 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 (K), или ще покажем, че допустимото ѝ множество е празно. В случай на изроденост на (K) можем да прибегнем до алгоритъма на Бленд, избягващ зациклянето.

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

$$(K) \quad \begin{array}{l} \min \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \mathbf{A}\mathbf{x} + \mathbf{u} = \mathbf{b} \\ \mathbf{x} \geqslant \mathbf{0},\ \mathbf{u} \geqslant \mathbf{0} \end{array} \quad \text{и да образуваме двойнствената } (DK) \quad \begin{array}{l} \max \mathbf{b}^{\mathrm{T}}\mathbf{y} \\ \mathbf{A}^{\mathrm{T}}\mathbf{y} \leqslant \mathbf{c} \\ \mathbf{y} \leqslant \mathbf{0}. \end{array}$$

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

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

За двойката спрегнати канонични задачи

$$\begin{array}{llcll} (K) & \min \mathbf{c}^{\mathrm{T}}\mathbf{x} & \text{и} & (DK) & \min \mathbf{b}^{\mathrm{T}}\mathbf{z} \\ & \mathbf{A}\mathbf{x} + \mathbf{u} = \mathbf{b} & & & -\mathbf{A}^{\mathrm{T}}\mathbf{z} + \mathbf{s} = \mathbf{c} \\ & \mathbf{x} \geqslant \mathbf{0}, \mathbf{u} \geqslant \mathbf{0} & & & \mathbf{z} \geqslant \mathbf{0},\ \mathbf{s} \geqslant \mathbf{0} \end{array}$$

имаме базисно псевдо оптимално решение $\overline{\mathbf{x}}^{\mathrm{T}} = (\mathbf{0}, \mathbf{b})$ с базисни променливи $\mathbf{u}$ за (K) и базисно допустимо решение $\overline{\mathbf{z}}^{\mathrm{T}} = (\mathbf{0}, \mathbf{c})$ с базисни променливи $\mathbf{s}$ за (DK) съответно. Симплексните таблици за съответните базисни решения са

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

Да разгледаме преобразувание чрез правия симплекс метод на симплексната таблица $T(\overline{\mathbf{z}})$: от реда с относителните оценки трябва да изберем отрицателна относителна оценка, т.е. $b_p < 0$, а след това променливата, излизаща от базиса трябва да изберем по теста за минимално отношение

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

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

§12. Анализ за чувствителност. Параметрично оптимиране

Разглеждаме, както обикновено, каноничната задача

$$(K) \qquad \begin{aligned} &\min z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ &\mathbf{A}\mathbf{x} = \mathbf{b} \\ &\mathbf{x} \geqslant \mathbf{0}, \end{aligned}$$

където $\mathbf{A}$ е $(m \times n)$ матрица, $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m_+$.

Във въпрос 11 видяхме как можем да получим оптимално решение на (K) след добавяне на ново ограничение, използвайки оптималното решение на задачата (K) и двойнствения симплекс метод. Във въпрос 10 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 $\mathbf{b}$ или във вектора на целевата функция $\mathbf{c}$ влияят на вече полученото оптимално решение. Такива изследвания се наричат още *анализ за чувствителност* или *следоптимален анализ*.

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

§12.1. Изменения във вектора на дясната част b

Да разгледаме фамилията канонични задачи с параметър в дясната част:

$$(K_t) \qquad \begin{aligned} &\min z(t) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ &\mathbf{A}\mathbf{x} = \mathbf{b} + t\mathbf{d} \\ &\mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

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

Твърдение 12.1. *Множеството от стойности на параметъра t, за които задачата (K_t) е съвместима е интервал $[\alpha, \beta]$ (с крайни или безкрайни граници).*

Доказателство. Множеството $P := \{(\mathbf{x}, t) | \mathbf{A}\mathbf{x} - t\mathbf{d} = \mathbf{b};\ \mathbf{x} \geqslant \mathbf{0}\}$ е многостенно множество в $\mathbb{R}^{n+1}$ и следователно е изпъкнало и затворено.

Проекцията на изпъкнало множество върху всяка от координатните оси е изпъкнало множество (докажете!). Изпъкнало затворено множество върху реалната права е затворен интервал. Означаваме го с $[\alpha, \beta]$ и отбелязваме че числата α и β могат да приемат крайни или безкрайни стойности. ■

Твърдение 12.2. *Ако за една стойност $t_0 \in [\alpha, \beta]$, задачата (K_{t_0}) е разрешима, то за всяко $t \in [\alpha, \beta]$ задачата (K_t) е разрешима.*

Доказателство. Тъй като задачата (K_{t_0}) е разрешима, от Силната теорема за двойнственост следва, че разрешима е и нейната двойнствена задача (D_{t_0}) $\begin{array}{l} \min v(t_0) = (\mathbf{b} + t_0\mathbf{d})^{\mathrm{T}}\boldsymbol{\pi} \\ \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c} \end{array}$. Очевидно множеството от ограничения на двойнствената задача (D_t) $\begin{array}{l} \min v(t) = (\mathbf{b} + t\mathbf{d})^{\mathrm{T}}\boldsymbol{\pi} \\ \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c} \end{array}$ не зависи от стойностите на параметъра t. Следователно, за всяко $t \in [\alpha, \beta]$ двойнствената задача (D_t) ще има непразно допустимо множество. Тъй като за всяко $t \in [\alpha, \beta]$ задачата (K_t) има допустимо решение, то съгласно слабата теорема за двойнственост целевата функция на (D_t) е ограничена отгоре върху допустимото ѝ множество. Тъй като установихме, че за всяко $t \in [\alpha, \beta]$ допустимото множество на (D_t) не е празното и целевата ѝ функция е ограничена отгоре върху него, то тя има крайно оптимално решение. Сега от Силната теорема за двойнственост имаме, че за всяко $t \in [\alpha, \beta]$ задачата (K_t) има крайно оптимално решение. ■

Ако $\mathbf{B}$ е базисната матрица на оптимално базисно допустимо решение $\overline{\mathbf{x}}$ за някаква стойност на $t = t_0$, то краищата на интервала $I_B = [\underline{t}, \overline{t}]$, в който базисното решение $\overline{\mathbf{x}}(t) = \begin{bmatrix} \overline{\mathbf{x}}_B(t) \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}(\mathbf{b} + t\mathbf{d}) \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{b}}^{\mathrm{T}} + t\overline{\mathbf{d}}^{\mathrm{T}} \\ \mathbf{0} \end{bmatrix}$, където $\overline{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b}$ и $\overline{\mathbf{d}} = \mathbf{B}^{-1}\mathbf{d}$ остава допустимо (и следователно оптимално) се определят при решаването на системата линейни относно t неравенства $\overline{\mathbf{b}}^{\mathrm{T}} + t\overline{\mathbf{d}}^{\mathrm{T}} \geqslant \mathbf{0}$ като

$$\underline{t} = \max_{i \in B} \left\{ -\frac{\overline{b}_i}{\overline{d}_i} : \overline{d}_i > 0 \right\}, \quad \text{или } \underline{t} = -\infty, \text{ ако } \overline{\mathbf{d}} \leqslant \mathbf{0}$$

и

$$\overline{t} = \min_{i \in B} \left\{ -\frac{\overline{b}_i}{\overline{d}_i} : \overline{d}_i < 0 \right\}, \quad \text{или } \overline{t} = \infty, \text{ ако } \overline{\mathbf{d}} \geqslant \mathbf{0}.$$

В интервала I_B координатите на оптималното решение $\overline{\mathbf{x}}(t)$ зависят линейно от t, но базисът B, базисната матрица $\mathbf{B}$ и оптималното решение на двойнствената задача $\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$ остават едни и същи.

Ако $\overline{\mathbf{x}}(t)$ е оптимално базисно допустимо решение на (K_t) за t в интервала $[\underline{t}, \overline{t}]$, то или съществува съседно на него базисно допустимо решение, което е оптимално решение на (K_t) за стойности на t в някакъв интервал $[\underline{\underline{t}}, \underline{t}]$, такъв че $-\infty \leqslant \underline{\underline{t}}$, или (K_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 от базиса B излиза променливата x_i, за която $\underline{t} = -\bar{b}_i/\bar{d}_i$, а $\underline{\underline{t}}$ се определя като се използва новия базис. Ако в симплексната итерация се достигне до празно допустимо множество, то (K_t) е с празно допустимо множество за всяко $t < \underline{t}$. По аналогичен начин се разсъждава за $t \geqslant \bar{t}$.

Твърдение 12.3. *Функцията $z^*(t)$ е изпъкнала, на части линейна функция на t в интервала $[\alpha, \beta]$.*

Доказателство. Нека I_B е интервалът на изменение на t, в който базисната матрица на оптималното решение $\overline{\mathbf{x}}$ е матрицата $\mathbf{B}$. Следователно,

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

е линейна функция на t в 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 $\mathbf{x}'$ е оптимално решение на $(K_{t'})$, а $\mathbf{x}''$ е оптимално решение на $(K_{t''})$. Образуваме изпъкналата комбинация на двата вектора $\mathbf{x} := \lambda\mathbf{x}' + (1-\lambda)\mathbf{x}''$ за $0 \leqslant \lambda \leqslant 1$.

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

§12.2. Изменения във вектора на целевата функция c

Да разгледаме фамилия от канонични задачи с параметър във вектора на целевата функция:

$$(K_t) \qquad \begin{aligned} &\min z(t) = (\mathbf{c} + t\mathbf{d})^{\mathrm{T}}\mathbf{x} \\ &\mathbf{A}\mathbf{x} = \mathbf{b} \\ &\mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

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

Твърдение 12.4. *Множеството от стойности на параметъра t, за които задачата (K_t) е разрешима е интервал $[\alpha, \beta]$.*

Доказателство. От силната теорема за двойнственост (K_t) е разрешима $\iff$ двойнствената ѝ задача $(D_t) \quad \begin{aligned} &\min v(t) = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ &\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c} + t\mathbf{d} \end{aligned}$ е разрешима $\iff$ (D_t) има непразно допустимо множество и целевата ѝ функция е ограничена отгоре върху него. Забелязваме, че ако допустимото множество на (K_t) (което не зависи от t!) не е празното, то от Слабата теорема за двойнственост целевата функция на (D_t) ще бъде ограничена отгоре върху допустимото ѝ множество. Следователно, необходимо е да намерим тези t, за които (D_t) има непразно допустимо множество. Множеството $\{(\boldsymbol{\pi}, t) | \mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} - t\mathbf{d} \leqslant \mathbf{c}\}$ е затворено многостенно множество в $\mathbb{R}^{m+1}$. Следователно, проекцията му върху оста на t е затворен интервал $[\alpha, \beta]$ с крайни или безкрайни граници. ■

Да допуснем, че при $t = t_0$, оптимално базисно допустимо решение на задачата (K_{t_0}) е $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_B \\ \overline{\mathbf{x}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix}$, с базисна матрица $\mathbf{B}$. Забелязваме, че координатите му не зависят от t и искаме да определим краищата на интервала $I_B = [\underline{t}, \overline{t}]$, в който $\overline{\mathbf{x}}$ остава оптимално. За целта нека означим със $\mathbf{c}_B$ и $\mathbf{d}_B$ базисните части, а със $\mathbf{c}_N$ и $\mathbf{d}_N$ небазисните части на векторите $\mathbf{c}$ и $\mathbf{d}$ съответно

Базисът B е оптимален тогава и само тогава, когато небазисните спрямо него относителни оценки са неотрицателни, т.е. $(\mathbf{c}_N + t\mathbf{d}_N)^{\mathrm{T}} - (\mathbf{c}_B + t\mathbf{d}_B)^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{N} \geqslant \mathbf{0}$. Като положим $\overline{\mathbf{c}}_N^{\mathrm{T}} = \mathbf{c}_N^{\mathrm{T}} - \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{N}$ и $\overline{\mathbf{d}}_N^{\mathrm{T}} = \mathbf{d}_N^{\mathrm{T}} - \mathbf{d}_B^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{N}$ това означава $t\overline{\mathbf{d}}_N^{\mathrm{T}} \geqslant -\overline{\mathbf{c}}_N^{\mathrm{T}}$. От тази система линейни относно t неравенства определяме

$$\underline{t} = \max_{j \in N} \left\{ -\frac{\overline{c}_j}{\overline{d}_j} : \overline{d}_j > 0 \right\}, \quad \text{или } \underline{t} = -\infty, \text{ ако } \overline{\mathbf{d}} \leqslant \mathbf{0}$$

и

$$\overline{t} = \min_{j \in N} \left\{ -\frac{\overline{c}_j}{\overline{d}_j} : \overline{d}_j < 0 \right\}, \quad \text{или } \overline{t} = \infty, \text{ ако } \overline{\mathbf{d}} \geqslant \mathbf{0}.$$

Ако $t_0 = 0$ и изберем $\mathbf{d} = \mathbf{e}_j$, то $[c_j + \underline{t}, c_j + \overline{t}]$ е *интервалът 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 $\overline{\mathbf{x}}$ е оптимално базисно допустимо решение на (K_t) за t в интервала $[\underline{t}, \overline{t}]$, то или съществува съседно на $\overline{\mathbf{x}}$ базисно допустимо решение, което е оптимално за стойности на t в някакъв интервал $[\underline{\underline{t}}, \underline{t}]$, такъв че $-\infty \leqslant \underline{\underline{t}}$, или $z^*(t) = -\infty$ (т.е. целевата функция на задачата (K_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 в базиса влиза променливата x_j, за която $\underline{t} = -\overline{c}_j/\overline{d}_j$, а $\underline{\underline{t}}$ се определя като се използва новия базис. Ако в симплексната итерация се достигне до неограничен ръб, то $z^*(t) = -\infty$ за всяко $t < \underline{t}$. По аналогичен начин се разсъждава за $t \geqslant \overline{t}$.

Твърдение 12.5. *Функцията $z^*(t)$ е вдлъбната, на части линейна функция на t в интервала $[\alpha, \beta]$.*

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

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

т.е. z^* е линейна функция на t в интервала I_B.

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

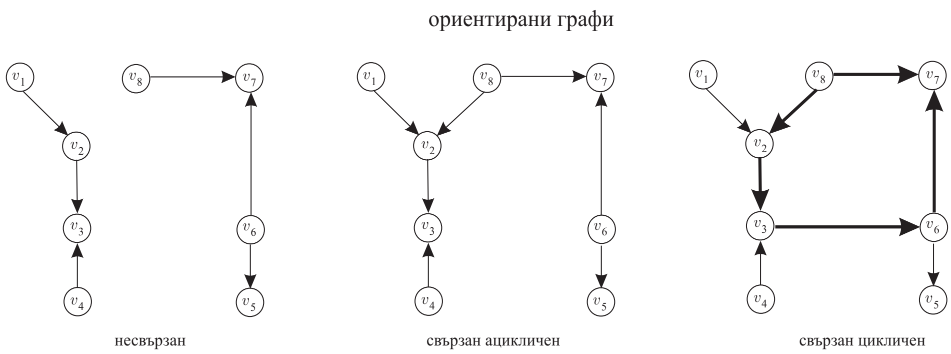
§13. Графи. Покриващи дървета. Мрежова задача. Характеризация на базисните решения

Много линейни задачи могат да се разглеждат като задачи за минимизиране на разходите за транспортиране на ресурс, разположен в някои от възлите на мрежа, така че да се задоволят потребностите от този ресурс 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.

§13.1. Графи

Математически мрежите най-удачно се моделират с графи. За целта ще припомним някои основни дефиниции и резултати от теория на графите.

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 $i = t(e)$, $j = h(e)$.



Фигура 13.7.

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 i_{k-1} и i_k има дъга, която може да е от вида (i_{k-1}, i_k) (*права* дъга), или от вида (i_k, i_{k-1}) (*обратна* дъга) за $1 \leqslant k \leqslant l$. Пътищата могат да се разглеждат като последователност 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 *покриващ* (*обхващащ*) подграф на G, ако $W \equiv V$.

§13.2. Мрежа и мрежова задача

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

$$\sum_{i \in V} b_i = 0, \tag{13.1}$$

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

Целта е да се преразпредели ресурсът 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. 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 е даден вектор с цените за превоз на единица от ресурса по всяка от дъгите $c = (c_e)_{e \in E}$.

Целта ни е да минимизираме общите транспортни разходи за преразпределяне на ресурса за задоволяване на потребностите във върховете.

Както вече казахме, ограниченията върху променливите идват от това че трябва да бъдат удовлетворени потребностите във всеки връх – да бъде доставено недостигащото количество или да бъде извозено излишното количество. И така, да фиксираме връх $k \in V$. Общото количество ресурс, което влиза във върха k е

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

а общото количество ресурс, което излиза от върха k е

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

Ако k е връх, в който имаме излишък, то трябва

$$\mathrm{Out}(k) - \mathrm{In}(k) = b_k,$$

а ако k е връх, в който имаме недостиг, то е необходимо

$$\mathrm{In}(k) - \mathrm{Out}(k) = |b_k| = -b_k.$$

И в двата случая получаваме

$$\mathrm{Out}(k) - \mathrm{In}(k) = b_k,$$

т.е.

$$\sum_{e \in E : t(e)=k} x_e - \sum_{e \in E : h(e)=k} x_e = b_k.$$

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 канонична линейна задача

$$(\hat{T}) \qquad \begin{aligned} &\min \sum_{e \in E} c_e x_e \\ &\sum_{e : t(e)=i} x_e - \sum_{e : h(e)=i} x_e = b_i\,, \quad i \in V, \\ &x_e \geqslant 0\,, \quad e \in E\,. \end{aligned}$$

Мрежа се нарича ориентиран свързан граф $G = (V, E)$, с чиито върхове са асоциирани числа b_i, $i \in V$ удовлетворяващи условието за баланс (13.1). Вектор $\mathbf{x} = (x_e)_{e \in E}$, чиито координати удовлетворяват $\sum_{e:t(e)=i} x_e - \sum_{e:h(e)=i} x_e = b_i$ за всяко $i \in V$ и $x_e \geqslant 0$ за всяко $e \in E$ се нарича *поток* в мрежата. С други думи, горната задача може да се разглежда като задача за намиране на поток в мрежа, който да задоволява определени потребности във възлите на мрежата и да бъде с минимална цена. Поради това тя се нарича *мрежова задача.*

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

За целта, да означим с $\hat{\mathbf{A}}$ *матрицата на инцидентност* между върховете и дъгите на G, т.е. $\hat{\mathbf{A}}$ има толкова редове, колкото са върховете на G и толкова стълбове, колкото са дъгите на G като елементът $\hat{a}_{ie}$, който съответства на върха i и дъгата e е

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

В матричен вид записваме задачата $(\hat{T})$ като

$$(\hat{T}) \qquad \begin{aligned} &\min\ \mathbf{c}^{\mathrm{T}} \mathbf{x} \\ &\hat{\mathbf{A}} \mathbf{x} = \hat{\mathbf{b}}\,, \\ &\mathbf{x} \geqslant \mathbf{0}\,. \end{aligned}$$

Да забележим, че в задачата $(\hat{T})$ матрицата на ограниченията $\hat{\mathbf{A}}$ е с линейно зависими редове. Наистина, във всеки стълб на $\hat{\mathbf{A}}$ има точно една $+1$ и точно една -1, докато останалите координати са нули. Следователно, сумата от вектор-редовете на $\hat{\mathbf{A}}$ е нулевият вектор, т.е. те са линейно зависими: ако означим с $\hat{\mathbf{a}}_i^{\mathrm{T}}$ редът на $\hat{\mathbf{A}}$, съответстващ на върха $i \in V$, имаме $\displaystyle\sum_{i\in V} \hat{\mathbf{a}}_i = \mathbf{0}$.

Сега вече е ясно и защо се налага условието за баланс (13.1) върху координатите на $\hat{\mathbf{b}}$. Причината е, че това условие е необходимо условие за това задачата $(\hat{T})$ да бъде с непразно допустимо множество. Наистина, ако $\hat{\mathbf{x}}$ е произволна допустима точка за $(\hat{T})$, то за всяко $i \in V$ ще имаме $\hat{\mathbf{a}}_i^{\mathrm{T}}\hat{\mathbf{x}} = b_i$. Сумираме и получаваме $\displaystyle\sum_{i\in V} b_i = \left(\sum_{i\in V} \hat{\mathbf{a}}_i^{\mathrm{T}}\right)\hat{\mathbf{x}} = \mathbf{0}^{\mathrm{T}}\hat{\mathbf{x}} = 0$, което е условието за баланс.

От линейната зависимост на вектор-редовете на $\hat{\mathbf{A}}$ и от условието за баланс следва, че сред ограниченията на задачата $(\hat{T})$ има излишни и те трябва да бъдат премахнати. Както ще се убедим, достатъчно е да премахнем само едно, произволно избрано ограничение.

Нека фиксираме произволен връх r във V, който ще наричаме *корен*. Нека означим с $\mathbf{A}$ матрицата, която се получава след като от матрицата $\hat{\mathbf{A}}$ бъде изтрит реда, съответстващ на r, а с $\mathbf{b}$ означим вектора, който се получава от вектора $\hat{\mathbf{b}}$ след изтриване на координатата b_r. Да разгледаме задачата

$$(T) \qquad \begin{aligned} &\min\ \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ &\mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ &\mathbf{x} \geqslant \mathbf{0}, \end{aligned}$$

която има едно ограничение по-малко от задачата $(\hat{T})$. Задачите (T) и $(\hat{T})$ са еквивалентни, тъй като допустимите им множества съвпадат. Наистина, ако $\hat{\mathbf{x}}$ е допустима точка за $(\hat{T})$, тя очевидно е допустима и за (T), а ако $\mathbf{x}$ е допустима точка за (T) от $\displaystyle\sum_{i\in V} \hat{\mathbf{a}}_i = \mathbf{0}$ имаме, че $\displaystyle\hat{\mathbf{a}}_r = -\sum_{i\in V, i\neq r} \hat{\mathbf{a}}_i$. Оттук, $\displaystyle\hat{\mathbf{a}}_r^{\mathrm{T}}\mathbf{x} = -\sum_{i\in V, i\neq r} \hat{\mathbf{a}}_i^{\mathrm{T}}\mathbf{x} = -\sum_{i\in V, i\neq r} b_i = b_r$ като второто равенство следва от това, че $\mathbf{x}$ е допустимо за (T), а последното – от условието за баланс (13.1) и следователно $\mathbf{x}$ е допустима точка за $(\hat{T})$.

§13.3. Покриващи дървета и базисни решения

Ще използваме следния резултат, който е добре известен от теория на графите и който ще оставим без доказателство.

Лема 13.1. *Нека $H = (V, F)$ е покриващ подграф на ориентирания свързан граф $G = (V, E)$ с $|V| = m$. 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 от $E \setminus F$ *създава цикъл.*

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

Ще използваме горния резултат, за да покажем, че матрицата $\mathbf{A}$ е с линейно независими редове, т.е че $r(\mathbf{A}) = m - 1$, където $m = |V|$. В същото време ще дадем характеризация на неособените подматрици на $\mathbf{A}$ от ред $m - 1$ (базисните матрици).

Теорема 13.1. *Нека* $G = (V, E)$ *е ориентиран свързан граф с* $|V| = m$. *Нека* $\hat{\mathbf{A}}$ *е неговата матрица на инцидентност,* $r \in V$ *е произволен връх и нека* $\mathbf{A}$ *е получена от* $\hat{\mathbf{A}}$ *след отстраняването на реда с индекс* r.

Тогава $r(\mathbf{A}) = m - 1$ *и ако* $\mathbf{B}$ *е* $(m-1) \times (m-1)$ *подматрица на* $\mathbf{A}$*, то* $\mathbf{B}$ *е неособена, тогава и само тогава, когато нейните стълбове съответстват на дъги на покриващо дърво на* G.

Доказателство. $\Longleftarrow$) Според Лема 13.1(iv) ориентираният свързан граф G притежава покриващо дърво, което се получава чрез премахване на дъги от G дотогава, докато полученият подграф стане минимално свързан, а според Лема 13.1(i) той ще бъде с $m - 1$ дъги. Следователно съществува поне една $(m-1) \times (m-1)$ подматрица на $\mathbf{A}$, стълбовете на която съответстват на покриващо дърво за G. Да означим коя да е такава подматрица с $\mathbf{B}$. Това, че $\mathbf{B}$ е неособена директно следва от

Лема 13.2. *Нека* $H = (V, F)$ *е покриващо дърво за* $G(V, E)$ *и нека* $\mathbf{B}$ *е съответната му подматрица на* $\mathbf{A}$. *Тогава редовете и стълбовете на* $\mathbf{B}$ *могат да се пренаредят така, че да образуват горна триъгълна матрица с ненулеви елементи по диагонала. В частност,* $\mathbf{B}$ *е неособена.*

Доказателство на Лема 13.2. С индукция по $|V| = k$.

За $k = 2$, $\mathbf{B}$ е 1×1 матрица, т.е. число, което е ± 1.

Нека твърдението е вярно за граф $G(V, E)$ с $|V| < m$ и да разгледаме случая $|V| = m$.

Нека $H = (V, F)$ е покриващо дърво за $G(V, E)$ с $|V| = m$. За всеки връх $i \in V$ дефинираме *степен* на i относно H, която е равна на броя на влизащите и излизащите от върха i дъги, които са в F, т.е. на броя на дъгите $e \in F$, такива че $t(e) = i$ или $h(e) = i$. Сумата от степените на всички върхове относно H е $2|F| = 2(m-1)$, тъй като началото и края на всяка дъга от F 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 от F). Оттук следва, че има поне два върха със степен 1 относно H, 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 от графа H 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 13.1(ii) следва, че H' е покриващо дърво за графа с $m-1$ върха

$$G' = (V \setminus \{i\}, E \setminus \{f \in E : t(f) = i \text{ или } h(f) = i\}).$$

Според индукционната хипотеза, можем да наредим върховете и дъгите на H' така, че съответната им матрица $\mathbf{B}'$ да бъде горно триъгълна с ненулеви елементи по диагонала. Като добавим реда за върха i като последен ред и стълба за дъгата e като последен стълб, получаваме

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}' & \mathbf{u} \\ \mathbf{0} & \pm 1 \end{pmatrix} \tag{13.2}$$

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

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

$$\sum_{e \in P} \mathbf{A}_e - \sum_{e \in Q} \mathbf{A}_e = \mathbf{0}\,.$$

И така, наборът от дъгите е цикличен и съгласно Лема 13.1(ii) това не могат да са дъги на покриващо дърво. Полученото противоречие доказва теоремата. ■

Показахме, че в задачата (T) матрицата $\mathbf{A}$ е с линейно независими редове. Задачата е канонична и има базисно решение (съответстващо на покриващо дърво). Намирането на базисните координати $\overline{\mathbf{x}}_B$ на базисно решение $\overline{\mathbf{x}}$ с базисна матрица $\mathbf{B}$ става чрез решаване на системата $\mathbf{B}\overline{\mathbf{x}}_B = \mathbf{b}$, а небазисните му координати, разбира се, са нули. Тъй като базисната матрица $\mathbf{B}$ може да се пренареди във във вида (13.2), то решаването на $\mathbf{B}\overline{\mathbf{x}}_B = \mathbf{b}$ относно $\overline{\mathbf{x}}_B$ става без да е необходимо да извършваме деления (т.к. диагоналните елементи на $\mathbf{B}$ са ± 1) и без да е необходимо да извършваме умножения (т.к. извън диагоналните ненулеви елементи на $\mathbf{B}$ са ± 1). С други думи, системата $\mathbf{B}\overline{\mathbf{x}}_B = \mathbf{b}$ се решава чрез последователност от събирания и изваждания. Следователно, ако векторът $\mathbf{b}$ е целочислен, базисното решение $\overline{\mathbf{x}}$ също ще бъде целочислено.

Следствие 13.1. *Ако b_i, $i \in V$ са цели числа, то координатите на всяко базисно решение $\mathbf{x}$ на (T) са цели числа, като $x_e \neq 0$ само за $e \in F$ за множеството от дъги F на някое покриващо дърво H на G.*

§14. Мрежов симплекс метод

Ще разгледаме какви са спецификите на симплекс метода, когато той се прилага за решаване на мрежова задача

$$(T) \qquad \begin{aligned} &\min\ \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ &\mathbf{A}\mathbf{x} = \mathbf{b} \\ &\mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

т.е. канонична линейна задача, чиято матрица на ограниченията $\mathbf{A}$ е получена от матрицата на инцидентност $\hat{\mathbf{A}}$ на ориентиран свързан граф $G = (V, E)$ с $|V| = m$ чрез изтриване на реда, съответстващ на връх $r \in V$ – корен, а векторът $\mathbf{b}$ е получен от вектора $\hat{\mathbf{b}} = (b_i)_{i\in V}$, удовлетворяващ условието за баланс $\sum_{i\in V} b_i = 0$ чрез премахване на координатата b_r, $\mathbf{c} = (c_e)_{e\in E}$, $\mathbf{x} = (x_e)_{e\in E}$.

Адаптирането на симплекс метода 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 текущото базисно допустимо решение е $\overline{\mathbf{x}}$.

Основната идея е пресмятанията да се осъществяват не с помощта на базисната матрица $\mathbf{B}$ на базисното допустимо решение $\overline{\mathbf{x}}$, а да се използва съответното ѝ покриващо дърво $H = (V, F)$ на $G = (V, E)$.

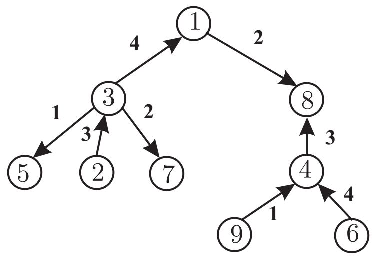
За всеки връх $i \in V$ построяваме единствения път $P_{i,r}$ в H от i до корена r (такъв път винаги съществува, т.к. H е свързан, а ако допуснем, че съществува и друг път ще получим цикъл в H, което е в противоречие с това, че H е ацикличен). За всеки връх $i \in V$ пазим неговият предшественик $p(i)$ – върхът следващ i в пътя $P_{i,r}$ като ако дъгата свързваща i и $p(i)$ в този път е права дъга, пишем $+p(i)$, а ако е обратна дъга пишем $-p(i)$. Пазим и $d(i)$ – дължината на пътя $P_{i,r}$.

Накрая нареждаме последователно върховете започвайки от корена r така, че върхът $p(i)$ да се среща винаги преди върха i. Такова нареждане се нарича *преднаредба* на дървото. Съществуват множество преднаредби. Фиксира се една от тях и върхът, който следва върха i в нея се означава с $s(i)$.

За пълното описание на дървото H са достатъчни трите m-мерни вектора $\mathbf{d} = (d(i))_{i\in V}$, $\mathbf{p} = (p(i))_{i\in V}$, $\mathbf{s} = (s(i))_{i\in V}$.

За илюстрация на казаното, да разгледаме следния

Пример 14.1. Нека $V = \{1, 2, \dots, 9\}$ с корен $r = 1$, като H е покриващото дърво, показано на Фигура 14.8 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 $\mathbf{p}$, $\mathbf{d}$ и $\mathbf{s}$ с дължина $|V| = 9$, дадени в Таблица 14.1.



Фигура 14.8.

Таблица 14.1

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

Да отбележим, че вместо да съхраняваме $(m-1)^2$ числа – елементите на базисна матрица $\mathbf{B}$ достатъчно е да съхраняваме $3m$ числа, описващи съответното ѝ покриващо дърво H.

Как се използват векторите $\mathbf{d}$, $\mathbf{p}$, $\mathbf{s}$ в хода на мрежовия симплекс метод?

Фиксираната преднаредба дава реда, в който се пресмятат координатите на вектора на симплексните множители $\boldsymbol{\pi}$. Тъй като $\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_B^{\mathrm{T}}\mathbf{B}^{-1}$, то $\boldsymbol{\pi}$ е решение на системата $\mathbf{B}^{\mathrm{T}}\boldsymbol{\pi} = \mathbf{c}_B$, която в случая на мрежова задача изглежда така:

$$\pi_i - \pi_j = c_f, \quad f = (i,j) \in F.$$

Това е линейна система с $m-1$ уравнения и m неизвестни. За да я решим полагаме $\pi_r = 0$ и пресмятаме останалите координати на $\boldsymbol{\pi}$ в реда, предписан от преднаредбата.

Ако за влизане в базиса е избрана небазисна променлива x_e с отрицателна относителна оценка, то е необходимо да се намерят координатите на вектора $\mathbf{w} = \mathbf{B}^{-1}\mathbf{A}_e$ или – което е еквивалентно – трябва да се реши системата $\mathbf{B}\mathbf{w} = \mathbf{A}_e$. За целта да разгледаме дъгата $e = (u,v)$. Тя принадлежи на $E \setminus F$ тъй като x_e е небазисна променлива. Тъй като H е свързан граф, от u до v съществува път $P_{u,v}$ от дъги в F. Добавяме към този път дъгата e и получаваме цикъл $P_{u,v} \cup \{e\}$. От доказателството на посоката $\Longrightarrow$ на Теорема 13.1 е ясно как да да използваме този цикъл, за да намерим координатите на вектора $\mathbf{w}$: ако означим с P множеството от правите дъги в цикъла, а с Q множеството от обратните дъги в цикъла от това че $\displaystyle\sum_{e\in P}\mathbf{A}_e - \sum_{e\in Q}\mathbf{A}_e = \mathbf{0}$ е очевидно, че $w_f = +1$, ако $f \in P$; $w_f = -1$ ако $f \in Q$; и $w_f = 0$ в противен случай (ако f е дъга от F, която не участва в цикъла).

§14.1. Алгоритъм на мрежовия симплекс метод

Нека $\overline{\mathbf{x}}$ е текущото базисно допустимо решение с покриващо дърво $H(V, F)$.

(1) Решаваме системата

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

като положим $\pi_r \equiv 0$ и пресмятаме останалите π-та като използваме преднаредбата $\mathbf{s}$, фиксирана в представянето на H: така $\pi_{p(i)}$ е вече известно, когато пресмятаме π_i.

Редът на пресмятане в нашия пример е следния: π_3, π_5, π_2, π_7, π_8, π_4, π_9, и накрая π_6.

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

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

(4) Проверка за неограниченост на целевата функция: намираме пътя $P_{u,v}$ от u до v в H, като използваме векторите $\mathbf{p}$ и $\mathbf{d}$. С негова помощ намираме координатите на вектора $\mathbf{w}$. Ако в пътя $P_{u,v}$ няма прави дъги ($w_f = -1$ за всяка дъга f от пътя), то тогава имаме неограничен ръб в допустимото множество на задачата, по който целевата функция намалява неограничено; КРАЙ. В противен случай отиваме на стъпка (5).

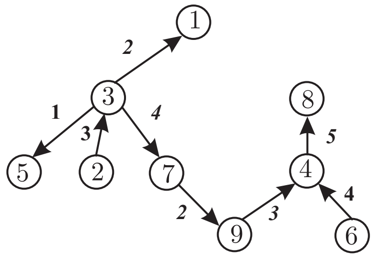
В нашия пример, да предположим, че $e = (u, v) = (7, 9)$. Тъй като $d(9) > d(7)$, намираме $p(9) = +4$; следователно $(9, 4)$ е обратна дъга в пътя. Сега $d(7) = d(4)$, но $7 \neq 4$. Така че намираме $p(7) = -3$, $p(4) = +8$, откъдето $(3, 7)$ и $(4, 8)$ са обратни дъги в пътя. Тъй като $3 \neq 8$, 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 $P_{7,9}e$ от върха 7 до върха 9 в H.

(5) Определяне на напускаща дървото дъга $f \in F$: това е правата дъга ($w_f = +1$) в пътя $P_{u,v}$ с най-малък поток по нея $\overline{x}_f$, която можем да определим като при намирането на пътя $P_{u,v}e$ на стъпка (4) сравняваме потока по всяка новопоявила се в него права дъга с текущия минимален поток по правите му дъги.

За нашия пример $f = (1, 8)$ и $\overline{x}_f = 2$.

(6) Обновяване на покриващото дърво (на базиса)

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



Фигура 14.9.

Таблица 14.2

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 $\overline{x}$:

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

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

Отиваме на стъпка (1).

За нашия пример новото дърво е показано на Фигура 14.9, представянето му е дадено в Таблица 14.2, а новото базисно допустимо решение (потокът по него) е $\overline{x}_{79} \leftarrow 2$, $\overline{x}_{37} \leftarrow 4$, $\overline{x}_{31} \leftarrow 2$, $\overline{x}_{18} \leftarrow 0$, $\overline{x}_{48} \leftarrow 5$ и $\overline{x}_{94} \leftarrow 3$, а останалите не се променят.

Да отбележим, че обновяването на представянето на дървото се състои в обновяването на векторите **p**, **d** и **s**. Съществуват правила на прехода, но те са сравнително сложни и тук няма да се спираме на тях.

§15. Генериране на стълб. Задача за едномерен разкрой

Ще разгледаме как симплекс методът може да се прилага при решаване на линейни задачи с голяма размерност. Нека е дадена канонична задача

$$(K) \qquad \begin{aligned} \min \quad & z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ & \mathbf{x} \geqslant \mathbf{0}\,, \end{aligned}$$

в която $m\times n$ матрицата на ограниченията $\mathbf{A}$ има огромен брой стълбове n. Ако структурата на вектор-стълбовете на $\mathbf{A}$ е известна, то съхраняването на този огромен брой стълбове е неоправдано, тъй като те могат да се генерират в хода на симплекс алгоритъма едва тогава, когато това стане необходимо – оттук идва и наименованието „генериране на стълб“ на този подход.

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

§15.1. Задача за едномерен разкрой

Да предположим, че фирма за производство на хартия разполага с широки ролки хартия с ширина W. Клиентите ѝ обаче търсят ролки хартия с по-малка ширина. Да предположим, че за да задоволи това търсене, фирмата трябва да произведе b_i броя ролки с ширина w_i, $i = 1, 2, \dots, m$. По-тесните ролки се получават чрез разрязване на широка ролка по различни *начини.* Например, нека широките ролки са с ширина $W = 70$ см, а се търсят ролки с ширини $w_1 = 17$ см и $w_2 = 15$ см. Широка ролка може да бъде разрязана на четири ролки с ширина w_1 като се бракува ширина от 2 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 и т.н. Всички такива начини на разкрояване образуват матрицата $\mathbf{A} = \{a_{ij}\}$, в която a_{ij} е броят ролки с ширина w_i за $i = 1, 2, \dots, m$, които се получават при j-ия начин на разкрояване за $j = 1, 2, \dots, n$. Стълбът $\mathbf{A}_j$ на матрицата $\mathbf{A}$ отговаря на j-ия начин на разкрояване.

Начините на разрязване от нашия пример се описват със стълбовете $\begin{bmatrix} 4 \\ 0 \end{bmatrix}$ и $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Допишете стълбовете, които отговарят на останалите възможни начини на разкрояване от примера.

Задачата за разкрой се състои в това да се намери минималният брой широки ролки, които трябва да се нарежат, за да се задоволи търсенето на b_i тесни ролки с ширина w_i за всяко i.

Следователно задачата за разкрой е задача (K) в която всяка координата x_j на вектора на променливите $\mathbf{x}$ означава броя широки ролки, които ще бъдат разкроени по j-ия начин; в матрицата на ограниченията $\mathbf{A} = \{a_{ij}\}$, a_{ij} е броят ролки с ширина w_i за $i = 1, 2, \ldots, m$, които се получават при j-ия начин на разкрояване за $j = 1, 2, \ldots, n$, а координатите b_i на вектора дясна част $\mathbf{b}$ са количествата b_i на ролките с ширина w_i за $i = 1, 2, \ldots, m$, които се търсят; накрая координатите c_j на вектора на целевата функция $\mathbf{c}$ са $c_j = 1$ за всяко j.

Получаваме задачата за едномерен разкрой:

$$(P) \qquad \begin{aligned} &\min \sum_{j=1}^{n} x_j \\ &\sum_{j=1}^{n} a_{ij} x_j = b_i\,, \quad i = 1, \ldots, m \\ &x_j \geqslant 0 \quad j = 1, \ldots, n. \end{aligned}$$

В действителност, бихме искали да решим съответната *целочислена задача*, в която x_j – броят големи ролки, нарязани по j-ия начин, е цяло число. Въпреки това, решаването на линейната нецелочислена задача (P) след закръгляване често дава достатъчно добро приближено решение на целочислената, или поне такъв е случаят, когато количествата b_i са разумно големи.

§15.2. Генериране на стълб в задачата за едномерен разкрой

Решаването на (P) само по себе си 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 създава и съхранява матрицата от коефициенти $\mathbf{A}$ е нецелесъобразно. Въпреки това, както показват Гилмор и Гомори през 1961–1963 г., задачата (P) може ефективно да бъде решена със симплекс алгоритъма като стълбовете на $\mathbf{A}$ се генерират в хода на алгоритъма тогава, когато това стане необходимо.

Намирането на начално базисно допустимо решение е лесно. Наистина, нека начинът j се състои в разрязване на широка ролка само на тесни ролки от един вид – тези с ширина w_j. Тогава първите m стълба на $\mathbf{A}$ ще бъдат

$$\mathbf{A}_j = \begin{bmatrix} 0 \\ \vdots \\ [W/w_j] \\ \vdots \\ 0 \end{bmatrix} \leftarrow j\text{-та координата},$$

(като $[\lambda]$ означава най-голямото естествено число, което не надминава λ) и очевидно тези стълбове $[\mathbf{A}_1, \ldots \mathbf{A}_m]$ образуват диагонална базисна подматрица на $\mathbf{A}$.

Да предположим, че $\overline{\mathbf{x}}$ с базис $\mathbf{B}$ е текущото базисно допустимо решение за симплексната итерация.

Пресмятаме симплексните множители $\boldsymbol{\pi}$ като $\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{e}^{\mathrm{T}}\mathbf{B}^{-1}$ (да напомним, че всичките $c_j = 1$, следователно $\mathbf{c}_B = \mathbf{e}$, където с $\mathbf{e}$ е означен векторът с размерност m, всичките координати на който са равни на единица).

Следващата стъпка е да се пресметнат относителните оценки

$$\overline{c}_j = 1 - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{A}_j \tag{15.1}$$

за всеки небазисен индекс 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 $\overline{c}_q < 0$. Това изглежда трудна задача, тъй като не знаем стълбовете $\mathbf{A}_j$ на матрицата $\mathbf{A}$, които са необходими за пресмятанията. Въпреки това, благодарение на структурата на задачата, можем да направим тази стъпка неявно.

Вектор $\mathbf{a} = (a_1, a_2, \ldots, a_m)^{\mathrm{T}} \in \mathbf{Z}_+^m$ (всяка от координатите a_i е естествено число) ще бъде стълб на $\mathbf{A}$ тогава и само тогава, когато съответства на допустим начин на разкрояване, т.е. тогава и само тогава, когато $\mathbf{w}^{\mathrm{T}}\mathbf{a} \leqslant W$, където $\mathbf{w} = (w_1, w_2, \ldots, w_m)^{\mathrm{T}}$.

Следователно това, което ни интересува е дали относителната оценка $1 - \boldsymbol{\pi}^{\mathrm{T}}\mathbf{a}$ е неотрицателна за всеки такъв вектор $\mathbf{a}$. Така стигаме до спомагателната задача

$$(SP) \qquad \begin{array}{ll} \max & \boldsymbol{\pi}^{\mathrm{T}}\mathbf{a} \\ & \mathbf{w}^{\mathrm{T}}\mathbf{a} \leqslant W\,, \\ & \mathbf{a} \in \mathbf{Z}_+^m\,. \end{array}$$

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

И така, решаваме спомагателната задача (SP). Ако оптималната стойност на (SP) е по-малка или равна на 1, тогава всички относителни оценки ще бъдат неотрицателни и текущото базисно допустимо решение ще бъде оптимално. Ако оптималната стойност на (SP) е по-голяма от 1, оптималното решение $\mathbf{a}^*$ на (SP) ще бъде стълб на матрицата $\mathbf{A}$, който съответства на небазисна променлива (на начин на разкрояване) с отрицателна относителна оценка, която се избира за влизане в базиса.

Итерацията на 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 стълбове на матрицата $\mathbf{A}$, а генерираме само тези от тях, които са необходими на симплексната итерация като решения на спомагателната задача (SP).

Остава да се спрем накратко на това как задачата (SP) може да бъде решена 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 $v = w_{\min}, w_{\min} + 1, \ldots, W$. Оптималното решение $\mathbf{a}^*$, за което се получава оптималната стойност $f(W)$ се получава чрез връщане назад, като се пази максимизиращият индекс i 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.