§1. Кратък увод в Изследване на операциите

§1.1. Какво се разбира под Изследване на операциите?

Под **операции (Operations)** се разбират дейностите, които осъществява дадена организация или система за постигане на своите цели.

Под **изследване (Research)** се разбира процесът на анализ на организацията или системата чрез използване на научни методи.

Следователно Изследване на операциите (или още Operations Research) е научна дисциплина за анализ на управлението на организации и системи.

Или, както пише Хамди Таха в увода на неговата станала класическа книга „Увод в изследване на операциите“, „*Изследване на операциите* (*ИО*) *е научен подход за вземане на решения за управление на дадена система в условия на ограничения от технически, икономически или друг характер върху системата.*“.

Под ИО почти винаги се разбира използването на математически методи за моделиране на системата и за анализ на нейните характеристики. Наистина, математическите модели и методи заемат централно място в ИО, но решаването на задачи на организационното управление далеч не винаги се свежда до построяване на математическия модел и до реализирането на съответните пресмятания. Това се обуславя от факта, че в хода на вземане на управленски решения често е необходимо да се отчитат фактори, съществени за правилното решаване на поставената задача, които обаче не могат непосредствено да бъдат включени в математическия модел. Такъв фактор например е човешкото поведение. Наличието на поведенчески елемент в моделираната система често води до това, че използваният за изработване на управленско решение математически модел се оказва твърде груб и поради това неспособен да даде правилно решение на поставената управленска задача.

В тази връзка като пример най-често се привежда станалият христоматиен за ИО „*асансьорен проблем*“. Същността на проблема е много проста. Служителите в административния офис на една голяма фирма се оплаквали от твърде дългото чакане на асансьора. Бил предприет опит възникналият проблем да бъде решен чрез математическите методи на теорията за масовото обслужване. Решението по ред причини се оказало неудачно. По-нататъшното изследване на проблема показало, че претенциите на служителите към режима на работа на асансьора са неоснователни, тъй като на практика времето за чакане на асансьора било напълно приемливо. Тогава възникнала идеята, позволила решаването на „асансьорния проблем“, а именно: било предложено на всеки етаж редом с асансьорните врати да се поставят големи огледала. Когато това било направено, жалбите веднага секнали. В очакване на асансьора хората се оглеждали в огледалата. При това времето за чакане минавало неусетно за тях и те нямали никакви основания да се оплакват от бавната работа на асансьора.

Приведеният пример показва необходимостта от правилна преценка на възможностите за математическо описание на изследваните процеси и за това, че в сферата на организационното управление далеч не всичко се поддава на формализация и следователно на адекватно отразяване в математическия модел. Интересно е да се отбележи, че това обстоятелство е било взето предвид още в първото използване в практиката на методите на ИО. Първото приложение на ИО било направено с военни цели, т.е. първите изследвани операции били военните операции. Малко преди и по време на Втората световна война Генералният щаб на британската армия възлага на екип от изявени в различни области учени да анализира някои военни операции: разполагане на радарите, управление на конвоите, подобряване на ефективността на бомбардировките, на антиподводничарските акции, на операциите по минирането и т.н. Всички тези проблеми са варианти на добре изучената днес задача за оптимално разпределение на ограничени ресурси (най-често военна техника), но сред включените в екипа специалисти е имало и психолози.

ИО е сравнително нова, но динамично развиваща се научна дисциплина. Методите на ИО се прилагат успешно при решаване на проблеми в сферата на промишленото производство, военното дело, селското стопанство, транспорта, финансите, икономиката, планирането и др. Използването на методи на ИО за решаване на проблеми от областта финансите и икономиката се нарича още Management Science. Напоследък термините Operations Research и Management Science се използват като синоними.

Тъй като ИО се използва за решаване на задачи на организационното управление, да видим каква е

§1.2. Структурата на задачата за организационно управление

При постановката на задачата за организационното управление преди всичко е важно да се определят три неща:

1. *целта*, преследвана от субекта на управлението.

   Под *цел* разбираме този краен резултат, който е необходимо да се получи по пътя на избор и реализация на някакви управляващи въздействия върху управляваната система. В икономическата сфера целта най-често е да се максимизира печалбата или да се минимизират разходите. Когато целта е определена, за *оптимален* се счита такъв начин на действие, който води до постигането на тази цел.

2. стойностите на кои характеристики на изследваната система могат да се променят (*управляеми променливи*).

При проучването на обекта на управление трябва да се намерят тези негови характеристики, чиито стойности могат да се променят (да се управляват). С други думи, трябва да се определи цялото множество от т. нар. *управляеми променливи*, които често ще наричаме само *променливи*. В случаите, когато априорната информация за идентификация на променливите не е достатъчна, за тяхното изясняване се налага да се привлекат специалисти от различни области, които в една или друга степен да вземат участие в процеса на анализ на системата. Например най-подходящото решение на „асансьорния проблем“ е предложено от този член на екипа от анализатори, който е имал повече познания в областта на психологията, отколкото в областта на формалните методи на ИО.

3. стойностите на кои характеристики на изследваната система не се променят (*параметри* на системата).

   Идентификацията на *неуправляемите променливи* (параметри), т.е. променливите, изменението на чиито стойности не зависи от решенията на субекта на управление, също е важен момент при вземане на управленски решения. Например, за да се вземе решение относно обема и структурата на продукцията, която смята да реализира на пазара, фирменото ръководство трябва да има предвид такива параметри като пазарните цени на отделните изделия. Ясно е, че в реални ситуации игнорирането на някои параметри може да доведе до построяването на неадекватен модел и следователно до вземането на погрешни решения.

След като сме изяснили структурата на задачата за управление, да преминем към

§1.3. Изкуството да се моделира

В ИО централната роля е отредена на математическото моделиране. За построяването на математически модел е необходимо да се има точна представа за целта на изследваната система и да се разполага с информация за ограниченията върху системата, които определят областта от допустими стойности на управляемите променливи. Анализът на модела трябва да доведе до намирането на най-добро управленско въздействие върху обекта на управление, като при това се изпълняват и всички ограничения.

Сложността на реалните системи силно затруднява представянето на целта и ограниченията в аналитичен вид. Въпреки големият брой на променливи и ограничения, които на пръв поглед е необходимо да се отчетат при анализа на реалната ситуация, само една част от тях се оказва съществена за описание на поведението на изследваната система. Поради това е уместно първо да се направи опростено описание на реалната система и вече на негова база да се строи математически модел. При това най-важното е да се идентифицират доминиращите променливи, параметри и ограничения. Опростеният образ на реалната система се отличава от оригиналната система по това, че в него намират отражение само доминиращите фактори (променливи, ограничения и параметри), които определят основната линия на поведение на реалната система. Математическият модел от своя страна представя най-съществените за описанието на опростения образ на системата съотношения във вид на целева функция и съвкупност от функционални ограничения.

Правила, определящи прехода от реалната система към математическия модел, не съществуват. Свеждането на множеството фактори, управляващи поведението на системата, към относително неголям брой доминиращи фактори и преходът от опростения образ на оригиналната система към модела, е изкуство, а не наука. Степента на адекватност на построения модел зависи преди всичко от способностите на членовете на екипа от анализатори.

Въпреки че строги предписания за това как следва да се разработва даден модел е невъзможно да се формулират, добре е да се има представа за възможните типове модели, за тяхната структура и характеристики.

§1.4. Една класификация на моделите на ИО

Най-важният тип модели на ИО са *математическите модели.* В основата на тяхното построяване стои допускането на това, че всички променливи, параметри и ограничения на системата, както и целта на управление, са количествено измерими.

Нека с x_j, $j = 1, \dots, k$ означим променливите на задачата. Нека X бъде множеството от всички k-мерни вектори (точки) $\mathbf{x}(x_1, \dots, x_k) \in \mathbb{R}^k$, чиито координати $(x_1, \dots, x_k)$ удовлетворяват всички ограничения, наложени на системата. Множеството X се нарича *допустимо множество.* Ако условията за функциониране на управляваната система се описват с m на брой ограничения, то след изясняване на функционалната зависимост между параметрите и променливите на системата, допустимото множеството X може да се представи като множество от решения на система от m функционални равенства и/или неравенства, а целта на управление да се опише посредством функция (наречена *целева функция*) на променливите $x_1, \dots, x_k$ на системата.

Следователно, математическият модел се заключава в свеждането на задачата за организационно управление до следната математическа задача: Да се намери оптималната стойност на целевата 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, $i = 1, \ldots, m$, са функции на k променливи, а b_i, $i = 1, \ldots, m$, са реални числа. Знакът $\lesseqgtr$ 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 (т.е. точка $\mathbf{x} \in X$) се нарича *допустимо решение. Оптимално решение* е такова допустимо решение, в което целевата функция приема най-добрата си стойност (оптимум) сред стойностите си във всички допустими решения. 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.

От своя страна математическите модели могат да се разделят на *детерминистични* и *стохастични* модели.

При *детерминистичните модели* стойностите на параметрите на системата са известни константи. С такива модели работят линейното оптимиране, целочисленото оптимиране, оптимирането върху графи, теорията на оптималното управление и др.

При *стохастичните модели* параметрите на системата приемат определени стойности с някаква вероятност. С такива модели работят теорията случайните процеси, теорията на масовото обслужване, теорията за полезността, теорията на игрите, динамичното оптимиране и др.

Настоящият курс е посветен основно на детерминистичните математически модели и методите за тяхното решаване.

Въпреки че в курса се разглеждат предимно математически модели, добре е да се знае, че освен математическите модели в ИО се използват и *имитационни* и *евристични* модели.

Имитационните модели „възпроизвеждат“ поведението на системата в продължение на някакъв период от време. Това се постига чрез определяне на поредността на събитията, които във времето дават важна информация за поведението на системата. След като тези събития са определени, характеристиките на системата се регистрират само в моментите на реализация на тези събития. Информация за характеристиките на системата в тези моменти се събира във вид на статистически данни от наблюденията на тези събития. Тази информация се обновява всеки път при настъпване на някое от интересуващите анализатора събития.

Тъй като при построяването на имитационен модел не е необходимо използването на математически функции, свързващи в явен вид променливите и параметрите на системата, тези модели по правило позволяват да се имитира поведението на много сложни системи, за които построяването на математически модели и получаването чрез тях на решение е невъзможно. Освен това гъвкавосттар присъща на имитационните модели, позволява да се получи по-точно представяне на системата. Основният недостатък на имитационното моделиране се състои в това, че неговата реализация е еквивалентна на провеждането на множество експерименти, а това неизбежно обуславя наличието на експериментални грешки. При имитационното моделиране обикновено възникват големи трудности, свързани с разработката на експеримента (статистически аспект), натрупването и съхраняването на резултатите от наблюденията и необходимата проверка на статистическите изводи. Имитационното моделиране е по-неудобен инструмент за изследване от математическите модели, които позволяват да се получи решение на поставената в общ вид задача.

Видяхме, че математическият модел на ИО търси оптимално решение на поставения проблем. Съответната му математическа задача понякога може да се окаже толкова сложна, че намирането на нейно точно решение да бъде невъзможно. Дори когато съществуването на оптимално решение е теоретически доказано, необходимите за неговото получаване пресмятания може да се окажат огромен брой, а необходимото за тяхното извършване време — твърде дълго. В такива случаи за получаване на рационално (приближено) решение могат да се използват *евристични методи*, базирани на интуитивно или емпирично избрани правила, които позволяват да се подобри текущото решение. По същество евристичните методи представляват процедури за търсене на разумен преход от едно допустимо решение към друго допустимо решение с цел подобряване на текущата стойност на целевата функция на модела. Когато по-нататъшно приближение към оптимума е невъзможно да се получи, най-доброто от получените до момента решения се приема като приближено решение на задачата.

§1.5. Изчислителни аспекти на ИО

Повечето алгоритми, разработени за решаване на оптимизационни задачи, не позволяват тяхното решение да се получи директно в аналитична форма. Като правило решението се намира чрез последователност от повтарящи се изчислителни процедури (*итерации*). Основната особеност на итерационния процес се състои в това, че на всяка стъпка съществува възможност за получаване на решение, което е по-близко до оптимума, отколкото текущото решение. В този случай се казва, че решението клони итеративно към оптимална точка.

Итерационният характер на изчислителния процес, използван за решаване на задачи с помощта на методи на ИО, обуславя използването на компютри за числените пресмятания. Освен това размерността на някой оптимизационни задачи е толкова голяма, че опитите да се получи тяхно решение, като се пресмята на ръка, са безсмислени.

Въпреки, че успехите в използването в практиката на методите на ИО се дължат на мощните изчислителни машини, остават ред задачи, за които не съществуват ефективни числени методи за решаване. Основната причина за възникващите трудности се състои в това, че даже в случая, когато сходимостта на алгоритъма е теоретически доказана и се използват мощни компютри за пресмятанията, времето за извършване на пресмятанията може да бъде много дълго. Друг отрицателен момент е характерният за компютърните пресмятания недостатък — грешките от закръгляване. Влиянието на грешките от закръгляване неизбежно се проявява при решаване на задачи с целочислени стойности на променливите, понеже в компютъра аритметичните действия се изпълняват с плаваща запетая. При реализиране на итерационен изчислителен процес влиянието на натрупването на грешките от закръгляване в даден момент може да стане съществено. Този проблем напоследък не е толкова актуален поради постигнатото вече високо технологично ниво на компютрите, което позволява постигането на по-голяма точност при пресмятанията.

Възможните трудности при осъществяване на изчислителния процес при решаване на оптимизационни задачи винаги трябва да се имат предвид. Затруднения от този род могат да заставят анализатора да опрости модела и да го приведе в съответствие с наличните възможности за числени пресмятания. За съжаление такива опростявания по правило увеличават разрива между оригиналната система и модела, което води до получаване на по-неточно решение на изходната задача.

§1.6. Етапи на ИО

Вече се убедихме, че е добре в процеса по ИО специалистът в областта на математическото моделиране да бъде подпомогнат от други специалисти, компетентни в различните аспекти на изследваната проблемна ситуация. Следователно в състава на екипа от анализатори е добре да влизат и представители на възложителя, носещи пряка отговорност за функционирането на изследваната система, а също така за проверката и практическото използване на разработените препоръки.

Работата, изпълнявана от анализаторския екип в процеса на операционното изследване, се състои от следните етапи:

1. идентифициране на проблема;
2. построяване на модела;
3. решаване на поставената задача с помощта на модела;
4. проверка на адекватността на модела;
5. прилагане на получените резултати в практиката.

Въпреки че тази последователност не е задължителна, тя може да се счита за общоприета. С изключение на третия етап, при който на базата на разработения модел решението се получава като се използват апробирани формализирани методи, всички останали етапи се изпълняват без строго регламентирани правила. Това се обуславя от факта, че изборът на една или друга процедура за реализиране на всеки етап зависи от конкретния проблем, от условията за функциониране на системата и от опита на екипа.

На *първия етап* задачата е да се идентифицира проблемът. Тук могат да се отделят следните основни стъпки:

- формулировка на задачата;
- намиране на присъщите на изследваната система параметри и променливи, цел и ограничения.

Вторият етап е свързан с построяването на модела. На този етап екипът, след като отчете особеностите на поставената задача, трябва да избере най-подходящия за адекватно описание на изследваната система модел. При построяването на такъв модел трябва да се установят количествените съотношения, необходими за изразяване на целевата функция и ограниченията във вид на функции от управляемите променливи. Ако разработеният модел съответствува на някой общ клас математически модели на ИО (например на модела на линейното оптимиране), то за получаване на решението е удобно да се използва съответният математически модел. Ако математическите съотношения, използвани в модела, са твърде сложни и не позволяват да се получи аналитично решение на задачата, по-подходящ за изследване може да се окаже имитационен модел. В някои случаи може да се наложи съвместно използване на математически, имитационни и евристични модели.

На *третия етап* се осъществява решаването на формулираната задача. При използване на математически модели решението се получава с помощта на апробирани оптимизационни методи, които водят до оптимално решение на задачата. В случая на прилагане на имитационни или евристични модели понятието за оптималност става по-неопределено и полученото решение само приблизително удовлетворява критерия за оптималност за функциониране на системата.

На третия етап освен намирането на оптимално решение винаги, когато това е възможно, трябва да бъде получена допълнителна информация за възможното изменение на решението при изменение на параметрите на системата. Тази част от изследването обикновено се нарича *анализ за чувствителност* на модела. Той е много полезен, когато някои параметри на изследваната система не подлежат на точна оценка. В такава ситуация е важно да се изследват възможните изменения на оптимума в зависимост от изменението в някакви интервали на съответните параметри на системата.

Четвъртият етап се състои в проверка на адекватността на модела. Моделът може да се счита за адекватен, ако е способен да даде достатъчно надеждно предсказание за поведението на системата. Общият метод за проверка за адекватност на модела се състои в съпоставяне на получените резултати с характеристиките на системата, които тя е имала при определени условия в миналото. Ако при аналогични входни параметри моделът достатъчно точно възпроизвежда поведението на системата, той се счита за адекватен. Тъй като построяването на модела обикновено става като се използват щателно проверени ретроспективни данни, благоприятният изход от едно такова сравнение е предопределен.

Този способ за оценка на адекватност на модела е непригоден при разработване на нови системи, тъй като липсват необходимите за проверка на модела данни. В някои случаи е допустима паралелна разработка на имитационен модел, предназначен за генериране на данни, необходими за проверката на основния математически модел.

Заключителният етап е свързан с реализацията на получените и проверени резултати. На този етап е необходимо да се оформят крайните резултати от изследването във вид на точни инструкции по прилагане на резултатите, които трябва да са съставени така, че лесно да се възприемат от лицата, обезпечаващи управлението на системата.

§1.7. Кратко описание, цели и източници на курса по ИО

Целта на курса по ИО е да запознае студентите с различни типове модели на ИО и с математически методи за тяхното решаване като се изтъкнат възможностите и недостатъците, свойствени на конкретните методи. Опасно заблуждение е мнението, че не е необходимо да се изучава математическия апарат на ИО, понеже компютърът може сам да „осигури“ решение на задачата. Следва винаги да се помни, че компютърът решава формулираната от анализатора задача. Ако той не е осведомен за възможностите на използвания модел, това неизбежно ще се отрази на качеството на полученото решение.

Целта на курса е да даде на студентите представа за математическия апарат на ИО и с помощта на прости примери да покаже сферата на приложение на методите на ИО. След като получи знания за математическите основи на ИО студентът би могъл се занимава с практическо изследване на реални задачи и сам да повишава нивото на своята подготовка в дадена конкретна област на ИО, като изучава съответните специализирани публикации.

За начало в курса подробно ще бъдат разгледани основите на линейното оптимиране (ЛО). Такава е традицията, спазвана във всички книги по ИО, която се обуславя от това, че ЛО е послужило за основа на разработката на много други математически методи на ИО, като например целочисленото оптимиране. По-нататък в курса ще бъдат разгледани целочисленото оптимиране, някои графови модели и др.

Полезни за доброто и бързо овладяване на преподавания материал са вече придобитите знания в областта на линейната алгебра и теория на графите.

Лекциите върху преподавания материал в първата част на курса 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.

§2. Задача на линейното оптимиране. Каноничен вид

В §1.4 видяхме, че математическият модел води до формулирането задача на математическото оптимиране (оптимизационна задача) от следния вид:

$$\begin{aligned}
&\text{да се намери}\\
&\min/\max\ f(x_1,\dots,x_k)\quad \text{(целева функция)}\\
&\text{при}\\
&g_i(x_1,\dots,x_k) \lesseqqgtr b_i,\ i=1,\dots,m\quad \text{(ограничения)},
\end{aligned}$$

където $f, g_i : \mathbb{R}^k \to \mathbb{R}$, $i = 1,\dots,m$ 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 $\geqslant$.

С други думи решаването на оптимизационната задача се състои в това да се намери оптималната стойност (минимум или максимум) на някаква функция $f(\mathbf{x})$ за стойности на аргумента $\mathbf{x}$, които принадлежат на някакво подмножество X на пространството $\mathbb{R}^k$, състоящо се от решенията на m-те функционални равенства и неравенства.

Основен дял на математическото оптимиране е *линейното оптимиране.*

§2.1. Задача на линейното оптимиране

Задача на линейното оптимиране е такава задача на математическото оптимиране, в която целевата функция f, както и всички функции g_i, описващи ограниченията, са *линейни функции.*

Следователно общият вид на задача на линейното оптимиране е следният

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

Теоретичните основи на линейното оптимиране се полагат с изучаването на линейните неравенства, което може да се проследи назад във времето до една работа на Фурие от 1826 г. По-късно много математици доказват редица частни случаи на най-важния резултат на линейното оптимиране, който и ние ще докажем — т. нар. *силна теорема за двойнственост.* Приложната страна на теорията започва да се разработва през 1939 г. от руския математик Канторович, който пръв забелязва практическата важност на някои класове линейни оптимизационни задачи и пръв дава алгоритъм за тяхното решаване. За съжаление години наред работите на Канторович остават непознати на Запад и незабелязани на Изток. Силен тласък в развитието на приложната страна на линейното оптимиране дават работите на Данциг, който през 1947 г. разработва симплекс метода за решаване на линейни задачи, които възникват в планирането на 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 Купман показва, че линейното оптимиране представлява подходящ модел за анализ на класически икономически теории. През 1975 г. Шведската Кралска Академия на Науките присъжда Нобеловата награда за икономика на Канторович и Купман „за техния принос в теорията за оптималното разпределение на ресурси“. Както се вижда, въпросната Академия разглежда работата на Данциг като твърде математическа за присъждане на наградата по икономика (а знаем, че не се присъжда Нобелова награда за математика).

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.

По-нататък 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.

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

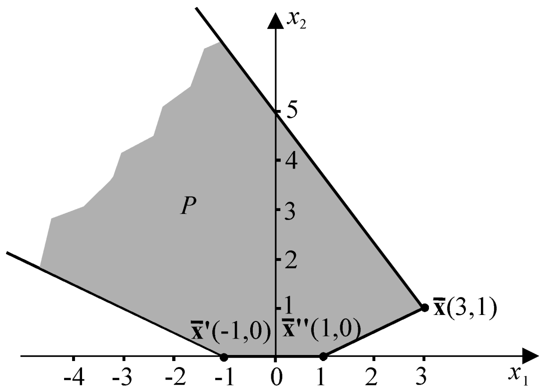
Пример 2.1. Да се намери

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

при ограничения

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

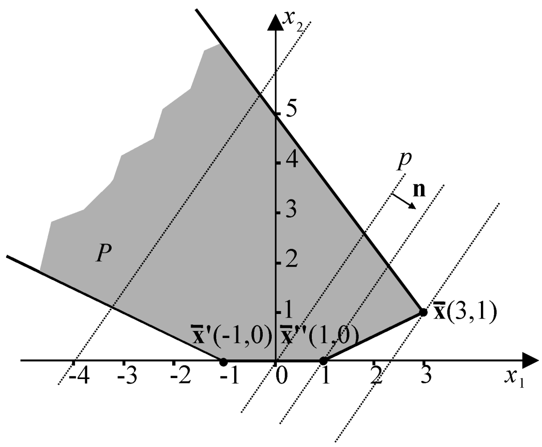
Нека P е допустимото множество на задачата. То се състои от всички вектори $\mathbf{x}(x_1, x_2)$, чиито координати удовлетворяват и четирите линейни неравенства. От геометрична гледна точка, множеството от точки, чиито координати са решения на линейно неравенство, е *полуравнина*. Следователно допустимото 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. 2.1.



Фигура 2.1.

Интересуваме се от най-голямата стойност на линейната функция $z(\mathbf{x}) = 3x_1 - 2x_2$, която тя може да приеме за точките $\mathbf{x}(x_1, x_2) \in P$. 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. 2.2.

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 ще бъде празно. Следователно максималната стойност на z



Фигура 2.2.

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 $x_2 = 1$.

Ако при същите данни на задачата вместо най-голямата ни интересува най-малката стойност на z върху допустимото множество, то трябва успоредно да местим p наляво. Но колкото и далеч да отместим наляво правата p, сечението ѝ с допустимото множество ще бъде непразно. Това означава, че в допустимото множество ще съществуват точки, за които z ще получава произволно малки стойности. Следователно задачата за минимум няма да има решение поради неограничено намаляване към $-\infty$ на целевата функция z върху 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 изразим 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 $\mathbf{a}_i^{\mathrm{T}} = [a_{i1}, \ldots, a_{ik}]$.

§2.2. Канонична задача

Линейната задача в общ вид (L) може да бъде за минимум или пък за максимум, а линейните функции, описващи ограниченията, могат да бъдат свързани с 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*.

Каноничната задача е линейна задача за минимум с ограничения равенства, чиито десни части b_i са неотрицателни числа и върху всичките ѝ променливи x_j 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, $j \in J$, (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. Двете задачи са еквивалентни в смисъл, че на всяко допустимо решение на (L) можем да съпоставим допустимо решение на (K) и обратно като подходящо преобразуваме координатите им, а съответните стойности на целевите функции на двете задачи в тях се различават евентуално само по знак.

В матрична форма каноничната задача може да се запише 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*, $\mathbf{b}(b_1, \dots, b_m)$ е m-мерен вектор с неотрицателни координати, който наричаме *вектор на дясната част*, 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 $j = 1, \dots, n$, 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.

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

За да илюстрираме правилата за преобразуване на линейна задача в каноничен вид, да приведем в каноничен вид задачата от Пример 2.1.

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 минаваме към задача за минимум

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

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

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

Добавяме по една неотрицателна допълнителна променлива в лявата част на всяко от трите неравенства, които вече са в посоката $\leqslant$.

Окончателно, каноничният вид на задачата от примера е

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

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

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

— ако $x_1 \geqslant 0$, то съответният $\mathbf{x}_K(x_1, 0, x_2, 0, 0, 0)$ е допустим за каноничната задача;

— ако $x_1 < 0$, то съответният $\mathbf{x}_K(0, -x_1, x_2, 0, 0, 0)$ е допустим за каноничната задача.

И в двата случая $z(\mathbf{x}_L) = -z_K(\mathbf{x}_K)$, т.е. стойностите на целевите функции 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 $z_K(\mathbf{x}_K) = -z(\mathbf{x}_L)$.

§3. Канонично многостенно множество. Върхове и базисни допустими решения

В § 2.2 показахме как произволна линейна задача (L) може да се сведе до съответна на нея канонична задача и казахме, че оттук нататък ще работим с 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, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m_+$ и $\mathbf{x} \in \mathbb{R}^n$ е n-мерен вектор на променливите.

За произволна съвкупност от числени стойности на променливите $(x_1, \ldots, x_n)$ казваме, че е *решение* на задачата (K). Едно решение $\mathbf{x}(x_1, \ldots, x_n)$ се нарича *допустимо*, ако удовлетворява всички ограничения. Допустимо решение се нарича *оптимално*, ако в него целевата функция z приема най-малка стойност измежду стойностите си в допустими решения. Множеството от всички допустими решения

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

на задачата (K) се нарича нейно *допустимо множество*.

За задачата (K) има точно три възможности:

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

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

- Допустимото множество M е непразно, а целевата функция $z(\mathbf{x})$ е ограничена отдолу върху M. Тогава (K) се нарича *разрешима*. Както ще се убедим от Теорема 5.2 в § 5, в този случай (K) има оптимално решение, т.е. съществува $\mathbf{x}^* \in M$, такова че $z(\mathbf{x}^*) = \min\{z(\mathbf{x}) : \mathbf{x} \in M\}$.

Очевидно допустимото множество $M := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ на каноничната задача (K) до голяма степен определя това дали тя е разрешима или не. Следователно е целесъобразно да изучим свойствата на това множество, което поради очевидната му връзка с каноничната задача 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 симплекс метода, използван за решаване на линейни задачи, от съществено значение е да изучим геометричните свойства на това множество.

За тази цел първо ще ни бъдат нужни няколко дефиниции.

Дефиниция 3.1. Нека $\mathbf{x}_1$ и $\mathbf{x}_2$ са две точки в $\mathbb{R}^n$. За всяко $\lambda \in [0,1]$ съответната точка

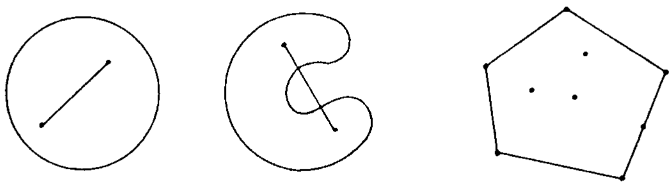
$$\mathbf{x}_\lambda = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$$

се нарича *изпъкнала комбинация* на точките $\mathbf{x}_1$ и $\mathbf{x}_2$ с тегла 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 като меним λ от 0 до 1 съответните изпъкнали комбинации $\mathbf{x}_\lambda$ ще опишат отсечката с краища $\mathbf{x}_2$ и $\mathbf{x}_1$.

Дефиниция 3.2. 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 съединителната 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. 3.3). Празното и едноточковото множество са изпъкнали множества.



Фигура 3.3.

Като обобщение на Дефиниция 3.1 за k-точки имаме

Дефиниция 3.3. *Изпъкнала комбинация* на k точки $\mathbf{x}_1, \dots, \mathbf{x}_k$ се нарича точка от вида

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

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 е най-малкото изпъкнало множество, което ги съдържа. Множеството от всички изпъкнали комбинации на две точки е отсечката, която ги свързва, на три точки в общо положение е триъгълникът с върхове в тях, на четири — образуваната от тях пирамида, и т. н.

Дефиниция 3.4. Точка $\mathbf{x}$ от изпъкнало множество C се нарича *екстремна точка* (*връх*) на C, когато $\mathbf{x}$ 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 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.

Например, крайните точки на триъгълника са върховете му, а крайните точки на единичното кълбо $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leqslant 1\}$ са точките от единичната сфера $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$.

Твърдение 3.1. *Ако $C_1, \ldots, C_k$ са изпъкнали множества в $\mathbb{R}^n$, то тяхното сечение $C := \bigcap\limits_{i=1}^{k} C_i$ 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 $\lambda\mathbf{x}' + (1-\lambda)\mathbf{x}'' \in \bigcap\limits_{i=1}^{k} C_i = C$. $\square$

Дефиниция 3.5. Множество от точки $\mathbf{x} \in \mathbb{R}^n$, които са решения на някакво линейно уравнение $\mathbf{a}^{\mathrm{T}}\mathbf{x} = b$, където $\mathbf{a}^{\mathrm{T}} \in \mathbb{R}^n$ 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 3.6. 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 $\mathbf{a}^{\mathrm{T}} \in \mathbb{R}^n$ 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 $H^- = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathrm{T}}\mathbf{x} \leqslant b\}$ и $H^+ = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathrm{T}}\mathbf{x} \geqslant b\}$.

Дефиниция 3.7. *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 3.2. *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 3.1. □

Следствие 3.1. *Каноничното множество M е многостенно множество.*

Доказателство. Множеството M представлява сечение на m-те хиперравнини, определени от 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 от 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. □

Множеството M ще наричаме още *канонично многостенно множество.*

Следствие 3.2. *Каноничното множество M е изпъкнало множество.*

Доказателство. От Следствие 3.1 и Твърдение 3.2. □

§3.1. Алгебрична характеризация на върховете на канонично многостенно множество M

Върховете на каноничното множество M са важни геометрични обекти, защото както ще покажем по-късно в Теорема 5.2, ако каноничната задача (K) е разрешима, то тя има оптимално решение, което е връх на каноничното множество M от всичките ѝ допустими решения.

За работа с тези геометрични обекти — върховете на каноничното множество M, ни е необходимо да намерим тяхна алгебрична характеризация. Първо ще докажем

Теорема 3.1. *Точка $\mathbf{x} \in M$ е връх на $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ тогава и само тогава, когато стълбовете на матрицата $\mathbf{A}$, съответстващи на положителните координати на $\mathbf{x}$, са линейно независими.*

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

Нека фиксираме $\overline{\mathbf{x}} \in M$. Да допуснем, че първите p координати на $\overline{\mathbf{x}}$ са положителни, а последните $n-p$ са нули, т.е. $\overline{\mathbf{x}}(\overline{x}_1, \ldots, \overline{x}_p, 0, \ldots, 0)$ като $\overline{x}_j > 0$, $j = 1, \ldots, p$ (винаги можем да го постигнем с преномериране на променливите).

Нека $\overline{\mathbf{x}}$ е връх на M. Трябва да докажем, че векторите $\mathbf{A}_1, \ldots, \mathbf{A}_p$ са линейно независими. Допускаме противното.

От линейната зависимост на $\mathbf{A}_1, \ldots, \mathbf{A}_p$, съществуват числа $z_1, \ldots z_p$ не всичките равни на нула и такива, че $\sum_{j=1}^{p} z_j \mathbf{A}_j = \mathbf{0}$. Да разгледаме векторите $\overline{\mathbf{x}} + \theta \mathbf{z}$, където $\mathbf{z} = (z_1, \ldots, z_p, 0, \ldots 0)$, а θ е реално число. Да забележим, че

$$\mathbf{A}\mathbf{z} = \sum_{j=1}^{n} z_j \mathbf{A}_j = \sum_{j=1}^{p} z_j \mathbf{A}_j = \mathbf{0}. \tag{3.1}$$

Имаме, че $\mathbf{A}(\overline{\mathbf{x}} + \theta \mathbf{z}) = \mathbf{A}\overline{\mathbf{x}} + \theta \mathbf{A}\mathbf{z} = \mathbf{A}\overline{\mathbf{x}} = \mathbf{b}$.
Векторът $\overline{\mathbf{x}} + \theta \mathbf{z} \geqslant \mathbf{0}$, ако $\overline{x}_i + \theta z_i \geqslant 0$ за всяко $i = 1, \ldots, p$, което е еквивалентно на това $\theta \geqslant -\frac{\overline{x}_i}{z_i}$ за всяко i, такова че $z_i > 0$ и $\theta \leqslant -\frac{\overline{x}_i}{z_i}$ за всяко i, такова че $z_i < 0$. Ако означим $\theta'' := \max_{i: z_i > 0} -\dfrac{\overline{x}_i}{z_i}$ или $\theta'' = -\infty$, ако няма i, такива че $z_i > 0$ и означим $\theta' := \min_{i: z_i < 0} -\dfrac{\overline{x}_i}{z_i}$ или $\theta' = +\infty$, ако няма i, такива че $z_i < 0$, то $\theta'' < 0 < \theta'$.
За всяко $\theta \in [\theta'', \theta']$ точката $\overline{\mathbf{x}} + \theta \mathbf{z} \in M$.

Да фиксираме $\theta > 0$, такова че θ и $-\theta$ са в интервала $[\theta'', \theta']$. Тогава точките $\mathbf{x}_1 := \overline{\mathbf{x}} + \theta \mathbf{z}$ и $\mathbf{x}_2 := \overline{\mathbf{x}} - \theta \mathbf{z}$ са в M, като $\mathbf{x}_1 \neq \mathbf{x}_2$, защото $\theta > 0$ и $\mathbf{z} \neq \mathbf{0}$. Като вземем средата на отсечката с краища $\mathbf{x}_1$ и $\mathbf{x}_2$, получаваме

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

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

Нека сега стълбовете $\mathbf{A}_1, \ldots, \mathbf{A}_p$ са линейно независими. Трябва да покажем, че $\overline{\mathbf{x}}$ е връх на M.

Да допуснем, че $\overline{\mathbf{x}}$ не е връх на M. Това означава, че $\overline{\mathbf{x}} = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$ 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 $\overline{\mathbf{x}}$ и $\mathbf{x}_1$ са в M, $\mathbf{A}(\overline{\mathbf{x}} - \mathbf{x}_1) = \mathbf{A}\overline{\mathbf{x}} - \mathbf{A}\mathbf{x}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Тъй като и двете числа λ и $1 - \lambda$ са строго положителни, последните $n - p$ координати на $\mathbf{x}_1$ и $\mathbf{x}_2$ са нули, тъй като тези координати на $\overline{\mathbf{x}}$ са равни на нула. Следователно последните $n - p$ координати на вектора $\overline{\mathbf{x}} - \mathbf{x}_1$ са нули, а сред първите p има поне една ненулева (в противен случай $\overline{\mathbf{x}} = \mathbf{x}_1 = \mathbf{x}_2$, което е невъзможно, тъй като $\mathbf{x}_1 \neq \mathbf{x}_2$). Тогава $\mathbf{0} = \mathbf{A}(\overline{\mathbf{x}} - \mathbf{x}_1) = \sum_{j=1}^{n} (\overline{x}_j - x_{1j})\mathbf{A}_j = \sum_{j=1}^{p} (\overline{x}_j - x_{1j})\mathbf{A}_j$, откъдето следва, че стълбовете $\mathbf{A}_1, \ldots, \mathbf{A}_p$ са линейно зависими.

Полученото противоречие показва, че ако вектор-стълбовете на $\mathbf{A}$, съответстващи на положителните координати на $\overline{\mathbf{x}}$, са линейно независими, то $\overline{\mathbf{x}}$ е връх на M. $\square$

§3.2. Базисни решения. Базисни допустими решения. Изродени и неизродени базисни допустими решения

Да разгледаме линейната система $\mathbf{Ax} = \mathbf{b}$, която е система от m уравнения с n неизвестни.

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

Когато $\mathbf{A}$ няма пълен ранг по редове, то или системата линейни уравнения $\mathbf{Ax} = \mathbf{b}$ няма решение (и в този случай M е празното множество), или някои от уравненията в системата $\mathbf{Ax} = \mathbf{b}$ са *излишни* в смисъл, че са линейно зависими от останалите уравнения. Във втория случай излишните уравнения се отстраняват едно по едно, докато се достигне до система от уравнения, чиято матрица има пълен ранг по редове.

Без ограничение на общността, оттук нататък ще предполагаме, че матрицата $\mathbf{A}$ има пълен ранг по редове. В този случай, характеризация на върховете на M може да се даде посредством т. нар. *базисни решения* на линейната система $\mathbf{Ax} = \mathbf{b}$.

Нека $m \times m$ матрицата $\mathbf{B}$ е неособена подматрица на матрицата $\mathbf{A}$ (тъй като $r(\mathbf{A}) = m$, то съществува поне една такава подматрица). Да предположим, че $\mathbf{B}$ се състои от първите m стълба на $\mathbf{A}$ (винаги можем да го постигнем чрез преномериране на променливите). 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 $n - m$ невключени в $\mathbf{B}$ стълбове на $\mathbf{A}$.

Множеството от n-те променливи $x_1, \ldots, x_n$ разделяме на две подмножества — m-те променливи, чиито индекси са индекси на стълбове на $\mathbf{A}$, включени в $\mathbf{B}$, означаваме с $\mathbf{x_B}$ и наричаме *базисни променливи*, а $n - m$-те променливи, чиито индекси са индекси на стълбове в $\mathbf{N}$, означаваме с $\mathbf{x_N}$ и наричаме *небазисни променливи*. Така координатите на произволен вектор $\mathbf{x} \in \mathbb{R}^n$ представяме 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

$$\mathbf{B}^{-1} \,|\, \mathbf{Bx_B} + \mathbf{Nx_N} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{x_B} + \mathbf{B}^{-1}\mathbf{Nx_N} = \mathbf{B}^{-1}\mathbf{b} \quad \Leftrightarrow$$

$$\mathbf{x_B} = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{N}\mathbf{x_N}. \tag{3.2}$$

Последното векторно равенство означава, че линейната система $\mathbf{Ax} = \mathbf{b}$ е разрешена относно базисните променливи $\mathbf{x_B}$, т.е. че посредством (3.2) 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 набор от числени стойности на небазисните променливи $\mathbf{x_N}$ от (3.2) могат да бъдат пресметнати числените стойности на базисните променливи $\mathbf{x_B}$ и така да се получи едно частно решение $\mathbf{x} = \begin{bmatrix} \mathbf{x_B} \\ \mathbf{x_N} \end{bmatrix}$ на системата. Ако зададем нулеви кординати на вектора на небазисните променливи $\mathbf{x_N}$, т.е. положим $\overline{\mathbf{x}}_\mathbf{N} = \mathbf{0}$, от (3.2) за координатите на вектора на базисните променливи ще получим единствено решение $\overline{\mathbf{x}}_\mathbf{B} = \mathbf{B}^{-1}\mathbf{b}$.

По този начин на всяка неособена подматрица $\mathbf{B}$ на матрицата $\mathbf{A}$ по единствен начин се съпоставя решение на системата $\mathbf{Ax} = \mathbf{b}$ от вида $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_\mathbf{B} \\ \overline{\mathbf{x}}_\mathbf{N} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix}$.

Дефиниция 3.8. Нека $\mathbf{B}$ е неособена $m \times m$ матрица, съставена от m (линейно независими) стълба на $\mathbf{A}$. Единственият вектор $\overline{\mathbf{x}}$, чиито небазисни координати са нули (т.е. $\overline{\mathbf{x}}_\mathbf{N} = \mathbf{0}$), а базисните му координати са $\overline{\mathbf{x}}_\mathbf{B} = \mathbf{B}^{-1}\mathbf{b}$ се нарича *базисно решение* с *базисна матрица* $\mathbf{B}$ (с *базис* $\mathbf{B}$).

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 3.9. Ако базисно решение $\overline{\mathbf{x}}$ с базисна матрица $\mathbf{B}$ е неотрицателно, т.е. ако $\overline{\mathbf{x}} = \begin{bmatrix} \overline{\mathbf{x}}_\mathbf{B} \\ \overline{\mathbf{x}}_\mathbf{N} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix} \geqslant \mathbf{0}$ то се нарича *базисно допустимо решение*, т.к. в този случай $\overline{\mathbf{x}} \in M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ и следователно е допустимо решение за каноничната задача (K).

Върховете на каноничното множество M са базисни допустими решения, както ще се убедим от следната

Теорема 3.2. *Ако каноничното множество* $M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ *е непразно, 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 3.1 получаваме линейната независимост на 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 $\mathbf{B}$ 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. $\overline{\mathbf{x}}_{\mathbf{N}} = \mathbf{0}$, $\overline{\mathbf{x}}_{\mathbf{B}} = \mathbf{B}^{-1}\mathbf{b} \geqslant \mathbf{0}$. 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 3.1 следва, че $\overline{\mathbf{x}}$ е връх на M. □

Благодарение на доказания резултат по-нататък ще използваме термините връх и базисно допустимо решение като синоними. От него следва и важният факт, че върховете на канонично множество са краен брой.

Следствие 3.3. *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. От Теорема 3.2 следва, че всеки връх на M е и базисно допустимо решение, т.е. на всеки връх съответства базисна матрица, състояща се от m линейно независими базисни 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, върховете на каноничното множество M са най-много $\binom{n}{m}$ на брой. □

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 3.10. Базисно допустимо решение $\overline{\mathbf{x}}$ с базис $\mathbf{B}$ се нарича *неизродено*, ако $\overline{\mathbf{x}}_{\mathbf{B}} > \mathbf{0}$, 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 допустимо решение, следователно точно един връх. Ако обаче върхът има по-малко от m положителни координати, то на него могат да съответстват няколко различни базисни матрици (допълването до базис да стане с различни вектор-стълбове на $\mathbf{A}$).

По-точно казано, ако базисното допустимо решение $\overline{\mathbf{x}}$ с базис $\mathbf{B}$ е неизродено, то лежи в сечението на m-те хиперравнини, определени с уравненията

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{3.3}$$

и на $(n-m)$-те хипер-равнини, определени с уравненията

$$\mathbf{e}_j^{\mathrm{T}}\mathbf{x} = 0, \quad j \in \mathbf{N}, \tag{3.4}$$

които съответстват на това, че $n-m$-те небазисни координати на $\overline{\mathbf{x}}$ 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. Върхът $\overline{\mathbf{x}}$ се определя по единствен начин от пресичането на n-те линейно независими хиперравнини (3.3) и (3.4).

Ако базисното допустимо решение $\overline{\mathbf{x}}$ с базис $\mathbf{B}$ е изродено, то някои от координатите на вектора $\overline{\mathbf{x}}_{\mathbf{B}}$ (базисните му координати) също са равни на нула. Следователно повече от $n-m$ неравенства от вида $x_j \geqslant 0$ се удовлетворяват във върха $\overline{\mathbf{x}}$ като равенства, което означава, че координатите на върха $\overline{\mathbf{x}}$ удовлетворяват повече от n на брой равенства. Следователно възможно е върхът $\overline{\mathbf{x}}$ да се определи по различни начини като сечение на n линейно независими хиперравнини, т.е. представянето му може да не бъде единствено, както беше в неизродения случай, и поради тази причина на изроден връх могат да съответстват няколко различни базисни матрици.

Да наблегнем още веднъж на факта, че докато на една базисна матрица $\mathbf{B}$ може да съответства точно един връх $\overline{\mathbf{x}}$ (точно едно базисно решение, което е връх, ако е допустимо), то на изроден връх е възможно да съответстват различни базисни матрици. На изроден връх $\overline{\mathbf{x}}$ е възможно да съответства огромен брой базисни матрици. В действителност, ако $\overline{\mathbf{x}}$ има $p < m$ положителни координати, то е възможно $\overline{\mathbf{x}}$ да има

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

различни базиса. Координатите на върхът $\overline{\mathbf{x}}$ са едни и същи, но той има множество различни базиси — множествата от координати, означени като базисни и небазисни са различни.

Типичен пример за изроденост се получава в задачата за назначения. Може да се покаже, че каноничното множество от допустими решения

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

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!

§4. Канонично многостенно множество. Посоки. Теорема за представяне

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 геометрични обекти, тъй като (както ще покажем в Следствие 4.2 по-долу), ако M е ограничено множество, т.е. M е многостен, произволна точка от M може да се представи като изпъкнала комбинация на върховете на M. Това означава, че ограничено многостенно множество може да бъде построено на базата на краен брой свои елементи (върховете му), като се вземат всичките им изпъкнали комбинации.

Когато M е неограничено множество обаче, върховете на M не са достатъчни за представяне на произволна точка от M. В този случай освен върховете на M за целта са необходими и посоките в M.

§4.1. Посоки в канонично множество M. Алгебрична характеризация

Най-напред ще въведем термина посока за произволно множество.

Дефиниция 4.1. *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 4.1 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 нулата 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 3.2), то посоките в M са изпъкнал конус с връх в $\mathbf{0}$. Това може да се докаже и директно като се използва характеризацията на посоките в M, получена в Теорема 4.1 по-долу.

Като пример можем да разгледаме многостенното множество P, изобразено на фиг. 2.1. Посоки в него 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 4.1 по-долу ще го докажем строго, 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 4.1. *Вектор* $\mathbf{d} \neq \mathbf{0}$ *е посока в* M *тогава и само тогава, когато*

$$\mathbf{A}\mathbf{d} = \mathbf{0} \qquad \textit{и} \qquad \mathbf{d} \geqslant \mathbf{0}\,.$$

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 $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, то $\mu\mathbf{A}\mathbf{d} = \mathbf{0}$, откъдето $\mathbf{A}\mathbf{d} = \mathbf{0}$. 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 $\mathbf{A}(\mathbf{x}_0+\mu\mathbf{d}) = \mathbf{A}\mathbf{x}_0+\mu\mathbf{A}\mathbf{d} = \mathbf{A}\mathbf{x}_0 = \mathbf{b}$ и $\mathbf{x}_0 + \mu\mathbf{d} \geqslant \mathbf{0}$ за всяко неотрицателно μ. Съгласно дефиницията $\mathbf{d}$ е посока в M. $\square$

След като вече разполагаме с характеризация на върховете и посоките в канонично множество M, можем да докажем

§4.2. 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 3.3 M има краен брой върхове.

Теорема 4.2 (Теорема за представяне). *Всяка точка* $\mathbf{x} \in M$ *може да се представи във вида*

$$\mathbf{x} = \sum_{i\in I} \lambda_i \overline{\mathbf{x}}_i + \mathbf{d}\,, \tag{4.1}$$

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. Представянето (4.1) 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 $i \in I$.

Теоремата ще докажем с индукция по броя p на положителните координати на точка $\mathbf{x} \in M$.

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

Ако $p_0 = 0$, то нулевият вектор $\mathbf{0}$ принадлежи на M. Той очевидно е връх на M и за него представянето е в сила.

Нека $p_0 > 0$. Всяка точка в M с p_0 положителни координати е връх на M. Наистина, да допуснем, че в M има точка $\overline{\mathbf{x}}$ с p_0 положителни координати, която не е връх. Следователно вектор-стълбовете на $\mathbf{A}$, съответстващи на положителните ѝ координати са линейно зависими. Продължаваме, както в заградената в каре част от доказателството на

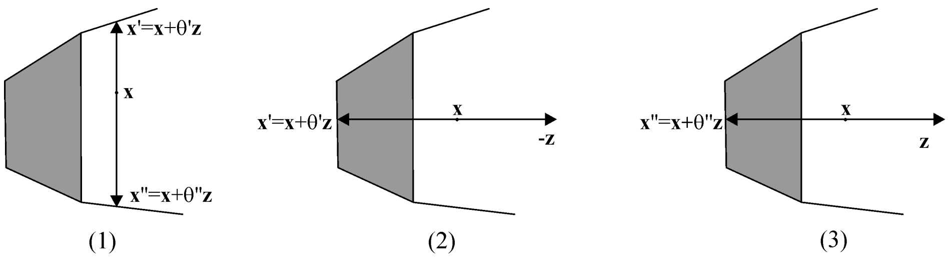
Теорема 3.1 в § 3.1: съществува вектор $\mathbf{z} \neq \mathbf{0}$, такъв че $z_i = 0$ ако $\overline{x}_i = 0$ и такъв че $\mathbf{Az} = \mathbf{0}$. Определяме съответните θ', θ'', като забелязваме, че поне едно от тях не е безкрайност, понеже $\mathbf{z} \neq \mathbf{0}$. Нека θ' е крайно. Да разгледаме точката $\mathbf{x}' = \overline{\mathbf{x}} + \theta'\mathbf{z}$, която е такава, че $\mathbf{Ax}' = \mathbf{b}$ и $\mathbf{x}' \geqslant \mathbf{0}$, т.е. $\mathbf{x}' \in M$. Освен това $\mathbf{x}'$ е точка от M, която има поне още една нулева координата в сравнение с $\overline{\mathbf{x}}$ (ако $\theta' = -\overline{x}_r / z_r$, то $\overline{x}_r \neq 0$, докато $x'_r = 0$). Това е в противоречие с дефиницията на p_0.

Следователно точките в M с минимален брой положителни координати имат представянето (4.1).

Да предположим, че точките в M с по-малко от p положителни координати имат представянето (4.1).

Да разгледаме 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 вектор-стълбовете на $\mathbf{A}$, съответстващи на положителните му координати, са линейно зависими и съществува вектор $\mathbf{z} \neq \mathbf{0}$, такъв че $z_i = 0$, ако $x_i = 0$, за който $\mathbf{Az} = \mathbf{0}$. Вижте отново заградената в каре част от доказателството на Теорема 3.1 в § 3.1. За вектора $\mathbf{z}$ са възможни три случая, илюстрирани на фиг. 4.4.



Фигура 4.4.

Случай 1. Векторът $\mathbf{z}$ има положителни и отрицателни координати. В този случай θ' и θ'' приемат крайни стойности. Точките $\mathbf{x}' = \mathbf{x} + \theta'\mathbf{z}$ и $\mathbf{x}'' = \mathbf{x} + \theta''\mathbf{z}$ са точки от M, които лежат върху правата през $\mathbf{x}$, определена от $\mathbf{z}$, и имат поне още една нулева координата в сравнение с $\mathbf{x}$. Точките $\mathbf{x}'$ и $\mathbf{x}''$ са точки от M и имат по-малко от p положителни координати. Според 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

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

Получаваме $(\lambda\theta'+(1-\lambda)\theta'')\mathbf{z} = \mathbf{0}$, но $\mathbf{z}$ е ненулев вектор и следователно $\lambda\theta' + (1-\lambda)\theta'' = 0$. Оттук намираме стойността на $\lambda = -\theta''/(\theta'-\theta'')$.

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

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

Като положим

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

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 $\mathbf{z} \leqslant \mathbf{0}$. В този случай само θ' е крайно. Дефинираме $\mathbf{x}'$, както в случай 1. Точката $\mathbf{x}$ записваме като

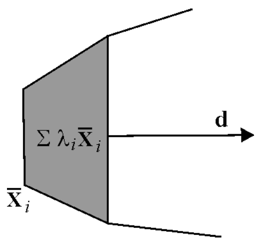
$$\mathbf{x} = \mathbf{x}' + \theta'(-\mathbf{z}), \quad \text{където} \quad \theta' > 0.$$

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 $\mathbf{A}(-\mathbf{z}) = \mathbf{0}$ от (3.1)), то очевидно $\mathbf{x}$ има представянето (4.1).

Случай 3. Векторът $\mathbf{z} \geqslant \mathbf{0}$. Доказателството е аналогично на доказателството в случай 2, като заменим $\mathbf{x}'$, θ' и $-\mathbf{z}$ съответно с $\mathbf{x}''$, $-\theta''$ 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. 4.5).

Отговор на въпроса кога в M има посока дава



Фигура 4.5.

Следствие 4.1. *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 $\mathbf{x}_0$. По 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, или M е ограничено множество. Полученото противоречие се дължи на допускането, че в M няма посока. Следователно ако M е неограничено, в M има посока. □

Следствие 4.2. *Ако M е непразно и ограничено 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 (4.1) на произволен елемент $\mathbf{x}$ на M ще имаме $\mathbf{d} = \mathbf{0}$, тъй като M е ограничено и според Следствие 4.1 в M няма посока. □

§5. 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 5.1. *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 (4.1), доказано в Теорема 4.2. 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 $\mathbf{0}$ очевидно е връх на M, откъдето получаваме противоречие с допускането, че множеството от върховете на M е празно. □

Теорема 5.2. *Ако M е непразно множество, то или $z(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}$ е неограничена отдолу върху M или минималната ѝ стойност за $\mathbf{x} \in M$ се достига във връх 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 $z \to -\infty$ по посоката $\mathbf{d}$. 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 този 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 клонят към $-\infty$, когато $\mu \to +\infty$, поради това, че $\mathbf{c}^{\mathrm{T}}\mathbf{d} < 0$.

Случай 2. В M няма посока $\mathbf{d}$, такава че $\mathbf{c}^{\mathrm{T}}\mathbf{d} < 0$. С други думи, 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

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

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. $\square$

Тази теорема стои в основата на решаването на линейни задачи. Тя 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 се разгледат само върховете на M.

§6. Симплекс метод

От втората основна теорема на линейното оптимиране — Теорема 5.2 е ясно, че за да решим каноничната задача на линейното оптимиране

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

където $\mathbf{A}$ е $m \times n$ матрица, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$ и $\mathbf{b} \in \mathbb{R}^m_+$, е достатъчно като кандидати за оптимално решение 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 обаче да се намерят всички върхове на M е непрактично, тъй като M може да има $\binom{n}{m} = \dfrac{n!}{m!(n-m)!}$ на брой върхове, а $\binom{n}{m}$ расте много бързо. Следователно за решаването на (K) е необходим по-систематичен подход. Такъв подход предлага 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 численните пресмятания при решаване на линейни задачи.

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

§6.1. Описание на симплекс метода (СМ)

Симплекс методът има две фази.

Първата фаза (фаза I) или намира връх на M (Теорема 5.1 гарантира съществуването на поне един връх, ако M е непразно множество), или установява, че допустимото множество M е празно. Намереният връх служи за начало на метода и поради това се нарича още *начален връх*.

Втората фаза (фаза II) е същинската част на метода. Намирайки се в текущия връх, методът намира ръб на M, излизащ от него и сключващ тъп ъгъл с вектора $\mathbf{c}$ на целевата функция z.

Ако този ръб е неограничен, той определя посока в M, по която z клони към $-\infty$ и задачата (K) е неограничена.

Ако ръбът е ограничен, методът отива в другия му край, който е връх на M, в който обаче z приема по-малка стойност.

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 е строго намаляваща. Следователно ако методът напусне даден връх, той никога не се връща в него и след краен брой итерации (тъй като според Следствие 3.3 M има краен брой върхове) или се достига до връх на M, който е оптимално решение на задачата (K), или се намира посока в M, по която целевата функция z намалява неограничено.

По-нататък задачата ни е да облечем горното просто геометрично описание на симплекс метода в алгебрична форма.

Най-напред ще разгледаме същинската фаза на СМ — фаза II, която намира оптимално решение, като предполага, че фаза I вече е намерила начален връх. Както ще се убедим в § 7.2, симплекс методът успешно може да бъде приложен и за решаването на задачата на фаза I – намирането на начален връх.

Ще предполагаме, че редовете на матрицата $\mathbf{A}$ са линейно независими, т.е., че $r(\mathbf{A}) = m$ и че $m < n$, така че задачата не е тривиална. Направените предположения за $\mathbf{A}$ гарантират, че от стълбовете ѝ може да се избере базисна матрица $\mathbf{B}$ и следователно има поне едно базисно решение на системата $\mathbf{Ax} = \mathbf{b}$.

За да опишем една итерация на симплекс метода, да предположим, че $\overline{\mathbf{x}}$ е текущ връх за метода. Нека $\overline{\mathbf{x}}$ е такова базисно допустимо решение, чиято базисна матрица $\mathbf{B}$ се състои от първите m вектор-стълба на $\mathbf{A}$ (нещо което винаги бихме постигнали чрез преномериране на променливите). Това означава, че първите m координати на произволен вектор $\mathbf{x} \in \mathbb{R}^n$ считаме за базисни, а останалите му $n - m$ координати — за небазисни, а матрицата $\mathbf{A}$ е представена като $\mathbf{A} = [\mathbf{B}|\mathbf{N}]$. Тогава върхът $\overline{\mathbf{x}}$ на M има представянето

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

С $\mathbf{B}$ освен базисната матрица ще означаваме и множеството от индексите на базисните променливи, т.е. $\mathbf{B} = \{1, \ldots, m\}$, а с $\mathbf{N}$ ще означаваме както небазисната част от матрицата $\mathbf{A}$, така и множеството от индексите на небазисните променливи, т.е. $\mathbf{N} = \{m+1, \ldots, n\}$.

§6.2. Привеждане на задачата (K) в базисен вид спрямо базиса B

След като базисната матрица $\mathbf{B}$ е фиксирана, задачата (K) можем да приведем в *базисен вид спрямо базиса* $\mathbf{B}$. Това става по следния начин:

1. Базисните координати $\mathbf{x_B}$ на решение $\mathbf{x}$ на системата $\mathbf{Ax} = \mathbf{b}$ изразяваме чрез небазисните му координати $\mathbf{x_N}$ (вж. също § 3.2)

$$\mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad [\mathbf{B}|\mathbf{N}] \begin{bmatrix} \mathbf{x_B} \\ \mathbf{x_N} \end{bmatrix} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{Bx_B} + \mathbf{Nx_N} = \mathbf{b} \quad \Leftrightarrow$$

$$\Leftrightarrow \mathbf{B}^{-1}|\mathbf{Bx_B} + \mathbf{Nx_N} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{x_B} + \mathbf{B}^{-1}\mathbf{Nx_N} = \mathbf{B}^{-1}\mathbf{b} \quad \Leftrightarrow$$
$$\mathbf{x_B} = \overline{\mathbf{x}}_\mathbf{B} - \mathbf{B}^{-1}\mathbf{Nx_N}.$$

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

$$\mathbf{x} = \begin{bmatrix} \mathbf{x_B} \\ \mathbf{x_N} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{x}}_\mathbf{B} - \mathbf{B}^{-1}\mathbf{N}\mathbf{x_N} \\ \overline{\mathbf{x}}_\mathbf{N} + \mathbf{I}\mathbf{x_N} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{x}}_\mathbf{B} \\ \overline{\mathbf{x}}_\mathbf{N} \end{bmatrix} + \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{bmatrix} \mathbf{x_N} =$$

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

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

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

$$\boldsymbol{\eta}_j := \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_j \\ \mathbf{e}_j \end{bmatrix}, \tag{6.1}$$

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 $\begin{bmatrix} -\mathbf{B}^{-1}\mathbf{N} \\ \mathbf{I} \end{bmatrix} \mathbf{x_N} = \sum_{j \in \mathbf{N}} x_j \boldsymbol{\eta}_j$ и

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j \in \mathbf{N}} x_j \boldsymbol{\eta}_j. \tag{6.2}$$

2. Като използваме представянето (6.2) на решение $\mathbf{x}$ на системата, за стойността на целевата функция z в $\mathbf{x}$ получаваме

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

Нека за всеки небазисен индекс $j \in \mathbf{N}$ означим с $\overline{c}_j$ скаларното произведение $\mathbf{c}^{\mathrm{T}}\boldsymbol{\eta}_j$ на вектора на целевата функция $\mathbf{c}$ и вектора $\boldsymbol{\eta}_j$. Имаме

$$\overline{c}_j = \mathbf{c}^{\mathrm{T}}\boldsymbol{\eta}_j = [\mathbf{c}_\mathbf{B}^{\mathrm{T}}, \mathbf{c}_\mathbf{N}^{\mathrm{T}}] \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_j \\ \mathbf{e}_j \end{bmatrix} = c_j - \mathbf{c}_\mathbf{B}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_j.$$

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

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

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 $\mathbf{B}$. Векторът $\overline{\mathbf{c}}^{\mathrm{T}} = [\overline{\mathbf{c}}_\mathbf{B}^{\mathrm{T}}, \overline{\mathbf{c}}_\mathbf{N}^{\mathrm{T}}] = [\mathbf{0}^{\mathrm{T}}, \overline{\mathbf{c}}_\mathbf{N}^{\mathrm{T}}]$, чиито координати са относителните оценки на променливите спрямо базиса

$\mathbf{B}$, се нарича *вектор на относителните оценки спрямо базиса* $\mathbf{B}$. Да забележим, че относителните оценки на базисните променливи са нули.

След направеното полагане имаме

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

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

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

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

Задачата $(K_{\mathbf{B}})$ наричаме *базисен вид на каноничната задача* (K) *спрямо базиса* $\mathbf{B}$.

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

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. Направляващи вектори по ръбовете, излизащи от върха $\overline{\mathbf{x}}$, са векторите $\boldsymbol{\eta}_j$ за $j \in \mathbf{N}$ (вж. дефиницията им в (6.1) и мястото им в базисния вид $(K_{\mathbf{B}})$ на задачата).

Да допуснем, че $\overline{\mathbf{x}}$ е неизродено базисно допустимо решение (т.е. $\overline{\mathbf{x}}_{\mathbf{B}} = \mathbf{B}^{-1}\mathbf{b} > 0$). В този случай за всяко $q \in \mathbf{N}$ векторът $\boldsymbol{\eta}_q$ определя *допустимо направление* за множеството M, т.е. за достатъчно малки положителни θ, точките от вида

$$\mathbf{x}(\theta) = \overline{\mathbf{x}} + \theta \boldsymbol{\eta}_q \tag{6.3}$$

остават в M. Да проверим това твърдение. Да фиксираме произволно $q \in \mathbf{N}$ и да забележим, че тъй като

$$\mathbf{A}\boldsymbol{\eta}_q = [\mathbf{B}|\mathbf{N}] \begin{bmatrix} -\mathbf{B}^{-1}\mathbf{A}_q \\ \mathbf{e}_q \end{bmatrix} = -\mathbf{B}\mathbf{B}^{-1}\mathbf{A}_q + \mathbf{A}_q = -\mathbf{A}_q + \mathbf{A}_q = \mathbf{0}, \tag{6.4}$$

то

$$\mathbf{A}\mathbf{x}(\theta) = \mathbf{A}(\overline{\mathbf{x}} + \theta\boldsymbol{\eta}_q) = \mathbf{A}\overline{\mathbf{x}} + \theta\mathbf{A}\boldsymbol{\eta}_q = \mathbf{A}\overline{\mathbf{x}} = \mathbf{b} \tag{6.5}$$

т.е. за произволно реално число θ векторът $\mathbf{x}(\theta)$ е решение на системата $\mathbf{Ax} = \mathbf{b}$. Остава само да установим, че координатите на $\mathbf{x}(\theta)$ са неотрицателни за малки положителни θ. Но тъй като по допускане $\overline{\mathbf{x}}$ е неизродено и всичките му базисни координати $\overline{\mathbf{x}}_{\mathbf{B}} = \mathbf{B}^{-1}\mathbf{b} > \mathbf{0}$, то

$$\mathbf{x}(\theta)=\overline{\mathbf{x}}+\theta\boldsymbol{\eta}_q=\begin{bmatrix}\overline{\mathbf{x}}_{\mathbf{B}}\\ \mathbf{0}\end{bmatrix}+\theta\begin{bmatrix}-\mathbf{B}^{-1}\mathbf{A}_q\\ \mathbf{e}_q\end{bmatrix}=\begin{bmatrix}\overline{\mathbf{x}}_{\mathbf{B}}-\theta\mathbf{B}^{-1}\mathbf{A}_q\\ \theta\mathbf{e}_q\end{bmatrix}=\begin{bmatrix}\overline{\mathbf{x}}_{\mathbf{B}}-\theta\mathbf{w}_q\\ \theta\mathbf{e}_q\end{bmatrix}, \tag{6.6}$$

където за последното равенство сме положили $\mathbf{w}_q := \mathbf{B}^{-1}\mathbf{A}_q$. Векторът $\mathbf{x}(\theta) \geqslant \mathbf{0}$ ако $\overline{\mathbf{x}}_{\mathbf{B}}-\theta\mathbf{w}_q \geqslant \mathbf{0}$ и ако $\theta\mathbf{e}_q \geqslant \mathbf{0}$. Второто векторно неравенство води до $\theta \geqslant 0$ поради $\mathbf{e}_q \geqslant \mathbf{0}$. Първото векторно неравенство е еквивалентно на системата от m числови неравенства $\overline{x}_i - \theta w_{iq} \geqslant 0$, $i \in \mathbf{B}$, където с w_{iq} е означена i-та координата на вектора $\mathbf{w}_q$. Ако $w_{iq} \leqslant 0$ неравенството е очевидно е в сила, а ако $w_{iq} > 0$ трябва $\theta \leqslant \frac{\overline{x}_i}{w_{iq}}$. Ако означим $\overline{\theta} := \min\{\dfrac{\overline{x}_i}{w_{iq}},\ w_{iq} > 0\}$, то $\overline{\theta} > 0$ като минимум на краен брой положителни числа (напомняме, че $\overline{x}_i > 0$ за всяко $i \in \mathbf{B}$) и за всяко $\theta \in [0, \overline{\theta}]$ имаме, че $\mathbf{x}(\theta) \geqslant \mathbf{0}$, което е достатъчно за $\mathbf{x}(\theta) \in M$.

Както вече казахме, симплекс методът търси сред ръбовете, излизащи от $\overline{\mathbf{x}}$ такъв, по който целевата функция z намалява. От вида (6.6) на $\mathbf{x}(\theta)$ е ясно, че движението по ръба с направляващ вектор $\boldsymbol{\eta}_q$ за някое $q \in \mathbf{N}$ 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 $(K_{\mathbf{B}})$ на задачата спрямо базиса $\mathbf{B}$ на $\overline{\mathbf{x}}$ следва, че при движение по ръба с направляващ вектор $\boldsymbol{\eta}_q$ ще имаме

$$z(\mathbf{x}(\theta)) = \mathbf{c}^{\mathrm{T}}\mathbf{x}(\theta) = \mathbf{c}^{\mathrm{T}}(\overline{\mathbf{x}} + \theta\boldsymbol{\eta}_q) = \mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} + \theta\mathbf{c}^{\mathrm{T}}\boldsymbol{\eta}_q = z(\overline{\mathbf{x}}) + \theta\overline{c}_q\,,$$

откъдето е ясно какво би било изменението на целевата функция, ако симплекс методът тръгне по ръба $\boldsymbol{\eta}_q$ (т.е. започне да увеличава небазисната променлива x_q):

- ако $\overline{c}_q < 0$, градиентът на целевата функция $\mathbf{c}$ сключва тъп ъгъл с направлението $\boldsymbol{\eta}_q$ на ръба и z ще намалява при движение по този ръб;

- ако $\overline{c}_q > 0$, градиентът на целевата функция $\mathbf{c}$ сключва остър ъгъл с направлението $\boldsymbol{\eta}_q$ на ръба и z ще нараства при движение по този ръб;

- ако $\overline{c}_q = 0$, градиентът на целевата функция $\mathbf{c}$ е ортогонален на направлението $\boldsymbol{\eta}_q$ на ръба и стойностите на z не се изменят при движение по този ръб.

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

От казаното дотук и от базисния вид на задачата $(K_{\mathbf{B}})$ получаваме

Теорема 6.1. (Критерий за оптималност). *Ако за всички $j \in \mathbf{N}$ относителните оценки $\overline{c}_j \geqslant 0$, то базисното допустимо решение $\overline{\mathbf{x}}$ е оптимално решение на каноничната задача (K).*

Доказателство. За произволно допустимо $\mathbf{x} \in M$ от базисния вид на задачата $(K_{\mathbf{B}})$ имаме, че $\mathbf{x} = \overline{\mathbf{x}} + \sum_{j \in \mathbf{N}} x_j \boldsymbol{\eta}_j$, като $x_j \geqslant 0$ за $j \in \mathbf{N}$, а стойността на целевата функция z в $\mathbf{x}$ е

$$z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j \in \mathbf{N}} \overline{c}_j x_j. \tag{6.7}$$

Тъй като $x_j \geqslant 0$ за $j \in \mathbf{N}$, ако всички относителни оценки $\overline{c}_j \geqslant 0$ за $j \in \mathbf{N}$, то от (6.7) следва, че $z(\mathbf{x}) \geqslant z(\overline{\mathbf{x}})$ за всяко $\mathbf{x} \in M$ и следователно $\overline{\mathbf{x}}$ е оптимално решение на (K). □

По-горе видяхме, че ако базисното допустимо решение $\overline{\mathbf{x}}$ е неизродено ($\overline{\mathbf{x}}_{\mathbf{B}} > 0$), то всички вектори $\boldsymbol{\eta}_j$ за $j \in \mathbf{N}$ определят допустими за M направления. Ако $\overline{\mathbf{x}}$ е изродено обаче, възможно е някои от векторите $\boldsymbol{\eta}_j$ да определят недопустими за M направления. Ако за всеки небазисен индекс $j \in \mathbf{N}$ означим $\mathbf{w}_j := \mathbf{B}^{-1}\mathbf{A}_j$, то от казаното за допустимите направления в началото на параграфа е ясно, че $\boldsymbol{\eta}_j$ ще определя недопустимо за M направление в $\overline{\mathbf{x}}$, ако за някой базисен индекс $i \in \mathbf{B}$ базисната координата $\overline{x}_i = 0$ като същевременно $w_{ij} > 0$, където с w_{ij} е означена i-та координата на вектора $\boldsymbol{\eta}_j$. В този случай $\overline{\theta} = 0$ и $\overline{\mathbf{x}} = \mathbf{x}(0)$ е единствената допустима точка от вида $\overline{\mathbf{x}} + \theta\boldsymbol{\eta}_j$.

Ако $\overline{\mathbf{x}}$ е неизродено, то вярна е и обратната на Теорема 6.1, т.е. ако $\overline{\mathbf{x}}$ е оптимално базисно допустимо решение, което е неизродено, то $\overline{c}_j \geqslant 0$ за всеки индекс $j \in \mathbf{N}$. Ако оптималното базисно допустимо решение $\overline{\mathbf{x}}$ е изродено обаче, възможно е то да е оптимално дори и в случай, че относителна оценка на някоя небазисна променлива $\overline{c}_j$ е отрицателна — ако съответният ѝ вектор $\boldsymbol{\eta}_j$ определя недопустимо за M направление.

Непосредствени следствия от (6.7) са

Следствие 6.1. *Ако за всички $j \in \mathbf{N}$ относителните оценки $\overline{c}_j > 0$, то базисното допустимо решение $\overline{\mathbf{x}}$ е единствено оптимално решение на каноничната задача (K).*

Следствие 6.2. *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 всяко допустимо решение $\mathbf{x} \in M$ от вида*

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{i=1}^{k} x_{j_i} \boldsymbol{\eta}_{j_i} \tag{6.8}$$

също е оптимално решение на (K).

Ако оптимално базисно допустимо решение $\overline{\mathbf{x}}$ е изродено и относителните оценки на някои от небазисните му променливи са нули, Следствие 6.2 не води до това, че $\overline{\mathbf{x}}$ не е единствено оптимално решение. Това е така, защото изроденото $\overline{\mathbf{x}}$ може да е единствената допустима точка от вида (6.8) поради недопустимост на направленията, определени от $\boldsymbol{\eta}_{j_i}$ в (6.8).

§6.4. Критерий за неограниченост на целевата функция

Нека сега за текущото $\overline{\mathbf{x}}$ с базис $\mathbf{B}$ не е в сила критерият за оптималност, т.е. нека съществува небазисен индекс $q \in \mathbf{N}$, за който $\overline{c}_q < 0$. Това означава, че при движение по ръба с направляващ вектор $\boldsymbol{\eta}_q$ 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 отрицателна относителна оценка $\overline{c}_q$.

След като симплекс методът е избрал ръб на намаляване $\boldsymbol{\eta}_q$ (с относителна оценка $\overline{c}_q < 0$), той се придвижва от $\overline{\mathbf{x}}$ по този ръб, т.е. започва да увеличава θ в (6.3).

За вектора $\mathbf{w}_q$, който съществено участва в описанието на направлението $\boldsymbol{\eta}_q$ имаме две възможности — или $\mathbf{w}_q \leqslant \mathbf{0}$, или $\mathbf{w}_q$ има положителна координата.

- Ако $\mathbf{w}_q \leqslant \mathbf{0}$, то $\mathbf{x}(\theta) = \overline{\mathbf{x}} + \theta\boldsymbol{\eta}_q$ ще бъде такова, че $\mathbf{x}(\theta) \geqslant \mathbf{0}$ за всяко $\theta \geqslant 0$ и ръбът с направляващ вектор $\boldsymbol{\eta}_q$ ще бъде неограничен ръб на M.

Получаваме

Теорема 6.2. (Критерий за неограниченост на целевата функция). *Ако за някой индекс $q \in \mathbf{N}$, $\overline{c}_q < 0$ и $\mathbf{w}_q \leqslant \mathbf{0}$, то задачата (K) е неограничена.*

Доказателство. Това, че $\mathbf{w}_q \leqslant \mathbf{0}$ влече $\boldsymbol{\eta}_q \geqslant \mathbf{0}$, а от (6.4) имаме $\mathbf{A}\boldsymbol{\eta}_q = \mathbf{0}$. От Теорема 4.1, която характеризира посоките на M следва, че ненулевият (тъй като $\mathbf{e}_q \neq \mathbf{0}$) вектор $\boldsymbol{\eta}_q$ е посока в M. Тъй като $\mathbf{c}^{\mathrm{T}}\boldsymbol{\eta}_q = \overline{c}_q < 0$, векторът $\mathbf{c}$ на целевата функция сключва тъп ъгъл с тази посока. От Теорема 5.2 следва, че z намалява неограничено по ръба $\{\overline{\mathbf{x}} + \theta\boldsymbol{\eta}_q,\ \theta \geqslant 0\}$. □

- Ако $\mathbf{w}_q \nleqslant \mathbf{0}$, то движението по ръба с направляващ вектор $\boldsymbol{\eta}_q$ не може да продължава неограничено, т.е. θ не може да нараства неограничено без това да наруши допустимостта на точките $\mathbf{x}(\theta)$. В този случай ръбът с направляващ вектор $\boldsymbol{\eta}_q$ е ограничен ръб на M. При движение по него правим

[1] Нарича се правило на Бил

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

Тъй като $\mathbf{w}_q \nleqslant \mathbf{0}$, ще съществува стойност $\overline{\theta} \geqslant 0$, такава че $\mathbf{x}(\theta) \in M$ за стойности на $\theta \in [0, \overline{\theta}]$ и $\mathbf{x}(\theta) \notin M$ за стойности на $\theta > \overline{\theta}$. В този случай ръбът с направляващ вектор η_q е ограничен ръб на M — отсечка с край в $\overline{\mathbf{x}}$, другият край на която $\overline{\mathbf{x}}' := \overline{\mathbf{x}} + \overline{\theta}\boldsymbol{\eta}_q$ се оказва друго базисно допустимо решение, както ще докажем в Лема 6.1 по-долу. Базисът на $\overline{\mathbf{x}}'$ 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*. Аналогично два върха на M, на които съответстват съседни базисни допустими решения са свързани с ръб на M и се наричат *съседни* върхове.

За да намерим координатите на съседния връх $\overline{\mathbf{x}}'$ е достатъчно да определим $\overline{\theta}$ — най-голямата стъпка, която можем да направим по ръба $\boldsymbol{\eta}_q$ без да нарушим допустимостта. В началото на § 6.3 видяхме, че $\overline{\theta}$ се определя от т.нар. *тест за минимално отношение*:

$$\overline{\theta} := \min\left\{\frac{\overline{x}_i}{w_{iq}}, \quad w_{iq} > 0\right\} = \frac{\overline{x}_p}{w_{pq}}. \tag{6.9}$$

Да отбележим, че ако $\overline{\mathbf{x}}$ е неизродено, то $\overline{\theta} > 0$. 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 $\overline{\theta}$ по ръба $\boldsymbol{\eta}_q$, съгласно (6.6) и избора на $\overline{\theta}$, са

$$\begin{aligned}
&\overline{x}'_j = 0, \quad j \in \mathbf{N}, j \neq q\,,\\
&\overline{x}'_q = \theta = \overline{x}_p / w_{pq}\,,\\
&\overline{x}'_i = \overline{x}_i - \theta w_{iq} = \overline{x}_i - \overline{x}_p w_{iq}/w_{pq}\,, \quad i \in \mathbf{B}\,.
\end{aligned} \tag{6.10}$$

Да забележим, че $\overline{x}'_p = 0$, т.е. че базисната за $\overline{\mathbf{x}}$ координата x_p приема нулева стойност в новото $\overline{\mathbf{x}}'$.

Ако текущото $\overline{\mathbf{x}}$ е неизродено ($\overline{x}_i > 0$, $i \in \mathbf{B}$), то $\overline{x}'_q = \overline{\theta} > 0$.

Следователно ако $\overline{\mathbf{x}}$ е неизродено, то базисната за неговия базис $\mathbf{B}$ променлива x_p приема нулева стойност в $\overline{\mathbf{x}}'$, а небазисната за $\mathbf{B}$ променлива x_q приема положителна стойност в $\overline{\mathbf{x}}'$.

Това идва да подскаже, че така получената точка $\overline{\mathbf{x}}'$ всъщност е базисно допустимо решение, чийто базис се отличава от базиса на $\overline{\mathbf{x}}$ по това, че за него променливата x_p е небазисна, а променливата x_q е базисна. Това ще докажем в следващата

Лема 6.1. *Точката $\overline{\mathbf{x}}'$ с координати (6.10) е друго базисно допустимо решение с базис $\mathbf{B}'$, който се различава от базиса $\mathbf{B}$ на $\overline{\mathbf{x}}$ по това че съдържа индекса q, а не съдържа индекса p, т.е. $\mathbf{B}' = \mathbf{B} \setminus \{p\} \cup \{q\}$, а базисната му матрица $\mathbf{B}'$ се различава от базисната матрица $\mathbf{B}$ на*

$\overline{\mathbf{x}}$, *по това, че един от нейните стълбове* $\mathbf{A}_p$ *е заменен със стълба* $\mathbf{A}_q$, *т.е.*

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

Доказателство. Точката $\overline{\mathbf{x}}'$ очевидно принадлежи на M, тъй като $\mathbf{A}\overline{\mathbf{x}}' = \mathbf{A}(\overline{\mathbf{x}} + \overline{\theta}\boldsymbol{\eta}_q) = \mathbf{A}\overline{\mathbf{x}} + \overline{\theta}\mathbf{A}\boldsymbol{\eta}_q = \mathbf{A}\overline{\mathbf{x}} = \mathbf{b}$, а $\overline{\mathbf{x}}' \geqslant \mathbf{0}$ следва от избора на $\overline{\theta}$. 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 3.1 и Теорема 3.2, достатъчно е да покажем, че стълбовете на матрицата $\mathbf{B}'$ са линейно независими.

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

$$\sum_{i\in\mathbf{B}, i\neq p} \beta_i \mathbf{A}_i + \beta_q \mathbf{A}_q = \mathbf{0}. \tag{6.11}$$

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 $\mathbf{A}_q = \sum\limits_{i\in\mathbf{B}} w_{iq}\mathbf{A}_i$. Като заместим това представяне на вектора $\mathbf{A}_q$ в (6.11), получаваме

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

Последното е линейна комбинация на векторите $\mathbf{A}_i$, $i \in \mathbf{B}$, но те са линейно независими, тъй като са вектор-стълбовете на базисната матрица $\mathbf{B}$. Следователно всички коефициенти на 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 $i \in \mathbf{B}, i \neq p$. Получаваме, че всички коефициенти β_i, $i \in \mathbf{B}'$, са нули, а с това и търсеното противоречие. □

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

Ако $\overline{\mathbf{x}}$ е изродено, възможно е определеното от теста за минимално отношение $\overline{\theta} = 0$. Това ще се случи, ако базисната координата $\overline{x}_p = 0$ и същевременно $w_{pq} > 0$. В този случай стъпката по ръба ще бъде с нулева дължина, т.е. методът няма да напусне изродения връх $\overline{\mathbf{x}}$, но ще смени неговия базис.

Да забележим, че ако минимумът в теста за минимално отношение (6.9) 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.

Като обобщим горните разсъждения, получаваме

Теорема 6.3. *Ако* $\overline{c}_q < 0$ *и* $\mathbf{w}_q$ *има положителна координата, то* $\overline{\mathbf{x}}'$ *с координати* (6.10) *е различно от* $\overline{\mathbf{x}}$ *базисно допустимо решение, в което стойността на целевата функция* $z(\overline{\mathbf{x}}') = z(\overline{\mathbf{x}}) + \overline{\theta}\overline{c}_q$ *е строго по-малка от стойността* $z(\overline{\mathbf{x}})$*, когато* $\overline{\theta}$*, определено с* (6.9)*, е положително число* (*в частност, когато* $\overline{\mathbf{x}}$ *е неизродено*)*.*

За да завършим итерацията на симплекс метода, остава само да заменим върха $\overline{\mathbf{x}}$ с върха $\overline{\mathbf{x}}'$ и да заменим базисната матрица $\mathbf{B}$ с базисната матрица $\mathbf{B}'$ като направим x_q базисна променлива, а x_p — небазисна променлива.

§6.6. Крайност на симплекс метода. Изроденост и зацикляне. Правило на Бленд за избягване на зациклянето

От Теорема 6.3 следва, че ако всички върхове на M са неизродени, то стойността на целевата функция 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 текущото базисно допустимо решение $\overline{\mathbf{x}}$ е изродено, 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 $\overline{\theta} = 0$. Като резултат предприетата стъпка ще бъде с нулева дължина и в края на итерацията върхът $\overline{\mathbf{x}}$ няма да се промени, т.е. при пресмятане с $\overline{\theta} = 0$ координатите на върха $\overline{\mathbf{x}}$ не се променят, но **се сменя неговия базис** — променливата x_q влиза в базиса на мястото на x_p.

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: **сред променливите, които са кандидати за влизане и излизане от базиса, винаги се избират тези с най-малък индекс**. Като кандидат за влизане в базиса се разглежда всяка небазисна променлива x_q с $\overline{c}_q < 0$, а като кандидат за излизане от базиса се разглежда всяка базисна променлива x_p, при която се достига минимумът $\overline{\theta}$ в теста за минимално отношение.

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

Теорема 6.4. *Симплекс методът приключва след краен брой итерации, ако от кандидатите за влизане и излизане от базиса винаги 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 докажем, че това води до противоречие.

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

Нека $\{\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_k\}$ е редицата от базиси на $\overline{\mathbf{x}}$, през които зацикля методът, т.е. симплекс методът генерира следната безкрайна редица от базиси на $\overline{\mathbf{x}}$

$$\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_k, \mathbf{B}_0, \mathbf{B}_1, \ldots \tag{6.12}$$

Както обикновено с $\mathbf{B}$ означаваме базисната матрица, както и множеството от базисни индекси, а с $\mathbf{N}$ означаваме множеството от съответните небазисни индекси, т.е. $\mathbf{N} := \{1, 2, \ldots, n\} \setminus \mathbf{B}$, както и съответната небазисна част на матрицата $\mathbf{A}$.

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 $\mathbf{B}$ е базиса, от който тя излиза и нека x_q е променливата, влизаща в базиса $\mathbf{B}$ на мястото на x_p. Тъй като променливата x_q влиза в базиса $\mathbf{B}$, а от него излиза променливата x_p, то $q \in \mathbf{N}$, а $p \in \mathbf{B}$ и следователно променливата x_q е непостоянна променлива (не участва в базиса $\mathbf{B}$, но участва в следващия го базис). Оттук $q < p$.

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

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j \in \mathbf{N}} \boldsymbol{\eta}_j x_j \quad \text{и} \quad z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j \in \mathbf{N}} \overline{c}_j x_j. \tag{6.13}$$

Нека сега означим с $\mathbf{B}^*$ този базис сред базисите в (6.12), в който променливата x_p влиза. Очевидно $\mathbf{B}^*$ е различен от $\mathbf{B}$. Да означим с $\mathbf{N}^*$ множеството от небазисните спрямо базиса $\mathbf{B}^*$ индекси, т.е. $\mathbf{N}^* := \{1, \ldots, n\} \setminus \mathbf{B}^*$.

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

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j \in \mathbf{N}^*} \boldsymbol{\eta}_j^* x_j \quad \text{и} \quad z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \sum_{j \in \mathbf{N}^*} \overline{c}_j^* x_j = z(\overline{\mathbf{x}}) + \sum_{j=1}^{n} \overline{c}_j^* x_j, \tag{6.14}$$

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

От (6.13) е ясно, че за произволно зададен набор от стойности на променливите $\mathbf{x}_{\mathbf{N}}$ могат да се пресметнат стойностите на променливите

$\mathbf{x_B}$ и да така да се получи частно решение $\mathbf{x} = \begin{bmatrix} \mathbf{x_B} \\ \mathbf{x_N} \end{bmatrix}$ на системата $\mathbf{Ax} = \mathbf{b}$.

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 $j \in \mathbf{N}$, $j \neq q$. От (6.13), като вземем предвид, че $\boldsymbol{\eta}_q = \begin{bmatrix} -\mathbf{w}_q \\ \mathbf{e}_q \end{bmatrix}$, получаваме $x_i = \overline{x}_i + \alpha\eta_{iq} = \overline{x}_i - \alpha w_{iq}$ за $i \in \mathbf{B}$.

Стойността на целевата функция z в така полученото частно решение $\mathbf{x}$ според (6.13) е $z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \overline{c}_q\alpha$, а според (6.14) е $z(\mathbf{x}) = z(\overline{\mathbf{x}}) + \overline{c}_q^*\alpha + \sum_{i\in\mathbf{B}} \overline{c}_i^*(\overline{x}_i - \alpha w_{iq})$. Като приравним тези два израза получаваме

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

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

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

Тъй като променливата x_q влиза в базиса $\mathbf{B}$, това означава, че относителната ѝ оценка спрямо този базис е $\overline{c}_q < 0$. Тъй като променливата x_q **не** влиза в базиса $\mathbf{B}^*$ (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 $\mathbf{B}^*$ имаме, че $\overline{c}_q^* \geqslant 0$. Следователно

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

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

$$\overline{c}_r^* w_{rq} > 0. \tag{6.15}$$

В частност $\overline{c}_r^* \neq 0$. Тъй като относителните оценки на базисните променливи са нули, то r е небазисен индекс, т.е. $r \in \mathbf{N}^*$. Получаваме, че променливата x_r е непостоянна, понеже участва в базиса $\mathbf{B}$ и не участва в базиса $\mathbf{B}^*$. Следователно $r \leqslant p$. В действителност $r < p$, тъй като $\overline{c}_p^* w_{pq} < 0$ (относителната оценка $\overline{c}_p^* < 0$, понеже x_p влиза в базиса $\mathbf{B}^*$, а $w_{pq} > 0$, понеже x_p излиза от базиса $\mathbf{B}$, 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 $\mathbf{B}^*$ влизаща в базаса $\mathbf{B}^*$ щеше да бъде променливата x_r, а не x_p. Следователно от (6.15) следва

$$w_{rq} > 0. \tag{6.16}$$

Тъй като всеки от базисите в редицата (6.12) поражда един и същ връх $\overline{\mathbf{x}}$, то всяка непостоянна променлива е с нулева стойност във всеки от тези базиси, т.е. имаме базисна нула $\overline{x}_i = 0$, ако i е непостоянна променлива. В частност

$$\overline{x}_r = 0. \tag{6.17}$$

От (6.16) и (6.17) следва, че променливата x_r е била кандидат за излизане от базиса $\mathbf{B}$ 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. □

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

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

§7. Алгоритъм и приложни реализации на симплекс метода

В § 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, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$ и $\mathbf{b} \in \mathbb{R}^m_+$. Сега ще дадем

§7.1. Алгоритъм на симплекс метода

Фаза I:

(0) Намира се начално базисно допустимо решение $\overline{\mathbf{x}}$ с базисна матрица $\mathbf{B} = [\mathbf{A}_{j_1}, \ldots, \mathbf{A}_{j_m}]$. Нека $\mathbf{B} = \{j_1, \ldots, j_m\}$ означава индексното множество от базисни променливи, т.е. x_{j_i} е i-та базисна променлива, $i = 1, \ldots, m$.

Фаза II:

(1) Пресмятат се относителните оценки на небазисните променливи

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

(2) Проверка на критерия за оптималност: Ако $\overline{c}_j \geqslant 0$, за всяко $j \notin \mathbf{B}$, то КРАЙ – текущото решение $\overline{\mathbf{x}}$ е оптимално, като оптималната стойност на функцията е $z(\overline{\mathbf{x}}) = \mathbf{c}_{\mathbf{B}}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b}$. 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, като се избере индекс

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

(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 направляващ вектор $\boldsymbol{\eta}_q = \begin{bmatrix} -\mathbf{w}_q \\ \mathbf{e}_q \end{bmatrix}$, 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
$$\overline{\theta} = \frac{\overline{x}_{j_p}}{w_{pq}} = \min_{1\leqslant i\leqslant m} \left\{ \frac{\overline{x}_{j_i}}{w_{iq}} \ :\ w_{iq} > 0 \right\} .$$
Ако минимумът се достига за повече от един индекс, изборът на излизаща от базиса променлива е нееднозначен.

(6) Обновява се текущото решение, базиса и базисната матрица $\mathbf{B}$: Присвоява се
$$\begin{aligned}
&\overline{x}_q \Leftarrow \theta = \overline{x}_{j_p}/w_{pq}\,,\\
&\overline{x}_{j_i} \Leftarrow \overline{x}_{j_i} - \theta w_{iq}, \quad 1 \leqslant i \leqslant m\,,\\
&\mathbf{B} \Leftarrow \mathbf{B} + (\mathbf{A}_q - \mathbf{A}_{j_p})\mathbf{e}_p^{\mathrm{T}}\,,\\
&\mathbf{B} \Leftarrow \mathbf{B} \setminus \{j_p\} \cup \{q\}\,,\\
&j_p \Leftarrow q
\end{aligned}$$
и се преминава към стъпка (1).

За описанието на фаза II на симплекс метода, което направихме в § 6, предполагахме, че на фаза I вече е намерено начално базисно допустимо решение (това е стъпка (0) на алгоритъма). Да видим как симплекс методът, който намира оптимално решение на фаза II, може успешно да се приложи и за решаването на проблема на фаза I — намирането на начален връх.

§7.2. Изкуствени променливи и фаза I на симплекс метода

Задачата за намиране на базисно допустимо решение за канонична задача (K) може да се формулира като изкуствена минимизационна задача (I) по следния начин:

- ограниченията на задачата (K) се преобразуват така, че за всяко $i = 1, \ldots, m$, i-то уравнение $\mathbf{a}_i^{\mathrm{T}}\mathbf{x} = b_i$ на системата се преобразува в уравнението $\mathbf{a}_i^{\mathrm{T}}\mathbf{x} + y_i = b_i$, като се добавя *изкуствена неотрицателна променлива* $y_i \geqslant 0$;
- 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 $\mathbf{y}(y_1, \ldots, y_m) \in \mathbb{R}^m$ е векторът на изкуствените променливи, а с $\mathbf{I}$ е означена единичната матрица от ред m.

Целта при решаване на каноничната задача (I) е изкуствените променливи да бъдат сведени до нулева стойност. Защо? Защото, ако допустимото множество M на първоначалната задача (K) не е празно, то като вземем $\mathbf{x} \in M$ и положим $\mathbf{y} = \mathbf{0}$, съответният вектор $(\mathbf{x}, \mathbf{0})$ ще бъде не само допустим за задачата (I), а ще бъде и нейно оптимално решение, понеже в него целевата функция ξ ще приема оптималната си стойност 0.

И така, да пристъпим към решаване на каноничната задача (I). Тя има очевидно базисно допустимо решение $(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = (\mathbf{0}, \mathbf{b})$, чийто базис се състои само от изкуствените променливи $\mathbf{y}$ и чиято базисна матрица е $\mathbf{I}$. Следователно можем да приложим фаза II на симплекс метода за задачата (I), като използваме $(\mathbf{0}, \mathbf{b})$ за начален връх.

Задачата (I) е разрешима тъй като допустимото ѝ множество не е празно ($(\mathbf{0}, \mathbf{b})$ е допустима точка) и целевата функция ξ е ограничена отдолу от нула. Приложен за (I), симплекс методът намира 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 е еквивалентно на това, че в оптималния базис има изкуствена променлива y_i с положителна стойност $\overline{y}_i^* > 0$. Това означава, че допустимото множество на задачата (K) е празно, понеже както обяснихме по-горе, в противен случай ξ^* ще бъде нула, а случаят не е такъв.

2. $\xi^* = 0$ и всички изкуствени променливи $\mathbf{y}$ 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 ще използва за начално базисно допустимо решение при нейното решаване.

3. $\xi^* = 0$, но някои от изкуствените променливи $\mathbf{y}$ са останали в базиса с нулева стойност (базисни нули) и оптималното решение $(\overline{\mathbf{x}}^*, \overline{\mathbf{y}}^*)$ е изродено. В този случай всяка изкуствена $\mathbf{y}$-променлива, която е останала в базиса, може или да се елиминира заедно с излишното уравнение, с което е асоциирана, или да се замени с 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 $\mathbf{A}_j$, $j \in \mathbf{N}$, то това означава, че след елементарни преобразувания k-ия ред на матрицата $\mathbf{A}$ се е трансформирал в нулевия вектор, а k-то уравнение на системата се е трансформирало в тъждеството $\mathbf{0}^{\mathrm{T}} \mathbf{x} = \mathbf{0}$. 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 някое $j \in \mathbf{N}$, то y_i може да се замени в базиса от небазисната променлива x_j. Тъй като при тази замяна $\overline{\theta}$=0 поради $\overline{y}^* = 0$, то се променя само базисът на решението.

След като в базиса изкуствените $\mathbf{y}$-променливи се заменят последователно 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 използваният подход е да се комбинират решаваната на фаза I задача (I) и решаваната на фаза II задача (K) в една задача, наречена *M-задача*

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

където числото M се избира толкова голямо, че всички изкуствени $\mathbf{y}$-променливи от даден момент нататък да бъдат сведени до нула, в случай че изходната задача има допустимо решение.

Сега ще се спрем на някои приложни реализации на симплекс метода.

§7.3. Таблична форма на симплекс метода

Един от начините да се организира итерацията на симплекс метода е да се вложат данните за текущото базисно решение 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 $x_1, \ldots, x_m$ 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.

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

Изразяването на базисните координати $\mathbf{x}_{\mathbf{B}}$ на решение $\mathbf{x}$ на системата $\mathbf{Ax} = \mathbf{b}$ чрез небазисните му координати $\mathbf{x}_{\mathbf{N}}$ се постига чрез елементарни Гаусови преобразувания на горната част на таблицата

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

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:

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

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. $\overline{\mathbf{x}}_{\mathbf{B}} = \mathbf{B}^{-1}\mathbf{b} \geqslant \mathbf{0}$ и горните преобразувания са еквивалентни на привеждането на задачата (K) в базисен вид спрямо базиса $\mathbf{B}$ на $\overline{\mathbf{x}}$.

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

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

Ако сред координатите на вектора $\overline{\mathbf{c}}_{\mathbf{N}}$ има отрицателни, то $\overline{\mathbf{x}}$ не удовлетворява критерия за оптималност и стойността на целевата функция още може да се намали. Всяка отрицателна координата $c_j < 0$

на вектора $\overline{\mathbf{c}}_{\mathbf{N}}$ съответства на ръб $\boldsymbol{\eta}_j$, по който целевата функция намалява. Обичайното правило е да се избере най-малката отрицателна координата на $\overline{\mathbf{c}}_{\mathbf{N}}$.

Нека най-малката отрицателна относителна оценка е $\overline{c}_q$. 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 $\overline{\theta}$ от теста за минимално отношение:

$$\overline{\theta}=\min\left\{\frac{\overline{x}_i}{w_{iq}}=\frac{(\mathbf{B}^{-1}\mathbf{b})_i}{(\mathbf{B}^{-1}\mathbf{A}_q)_i}:w_{iq}>0,1\leqslant i\leqslant m\right\}=\frac{\overline{x}_p}{w_{pq}}=\frac{(\mathbf{B}^{-1}\mathbf{b})_p}{(\mathbf{B}^{-1}\mathbf{A}_q)_p}\,. \tag{7.1}$$

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 $\overline{\mathbf{c}}_{\mathbf{N}}$) и на вектора $\mathbf{B}^{-1}\mathbf{b}$ (той, както вече отбелязахме, са намира горе вдясно на таблицата).

Минимумът в (7.1) 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 $\boldsymbol{\eta}_q = \begin{bmatrix} -\mathbf{w}_q \\ \mathbf{e}_q \end{bmatrix}$, 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 $\mathbf{B}'$.

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

Завъртането (pivot) прави следното:

1. дели p-ия ред на $T(\overline{\mathbf{x}})$ на $t_{pq} = w_{pq}$, и
2. за $1 \leqslant i \leqslant m + 1$, $i \neq p$ вади от i-ия ред на $T(\overline{\mathbf{x}})$ произведението на t_{iq} с новия p-ти ред,

при което в новата таблица $T(\overline{x}')$ q-ият стълб става p-тият единичен вектор $\mathbf{e}_p$.

По-точно, ако означим с t'_{ij} елемента, който се намира в i-ия ред и j-ия стълб на $T(\overline{\mathbf{x}}')$, при завъртането се получава

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

1. и 2. за $1 \leqslant i \leqslant m$ и $1 \leqslant j \leqslant n$ са изразяването на новите базисни променливи $\mathbf{x_{B'}}$ чрез новите небазисни променливи $\mathbf{x_{N'}}$;

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

2. за $i = m+1$ и $1 \leqslant j \leqslant n$ е заместването на така изразените $\mathbf{x_{B'}}$ в z и изразяването ѝ като функция само на новите небазисни променливи $\mathbf{x_{N'}}$, т.е. на получаването на новите относителни оценки $\overline{\mathbf{c}}'$ от старите $\overline{\mathbf{c}}$.

Така на всяка итерация се създава симплексна таблица за текущото решение, 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 $\mathbf{A}$ е с 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 (7.1).

§7.4. Модифициран симплекс метод. Реализации.

За отстраняване на недостатъците на табличната форма на симплекс метода на по-късен етап е разработен т.нар. *модифициран симплекс метод*. Нека да разгледаме алгоритъма на симплекс метода, за да установим кои са наистина необходимите за него пресмятания.

Намирайки се в текущия връх $\overline{\mathbf{x}}$ с базисна матрица $\mathbf{B}$, е необходимо да се извършат пресмятанията на стъпки (1), (4) и (5) от алгоритъма. Пресмятането на относителните оценки на стъпка (1) може да стане на два етапа: първо да се намерят координатите на вектора

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

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

Следователно трябва да се решат относно $\boldsymbol{\pi}^{\mathrm{T}}$, $\mathbf{w}_q$ и $\overline{\mathbf{x}}_{\mathbf{B}}$ следните три системи линейни уравнения

$$\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_{\mathbf{B}}^{\mathrm{T}}\mathbf{B}^{-1}, \quad \mathbf{w}_q = \mathbf{B}^{-1}\mathbf{A}_q \quad \text{и} \quad \overline{\mathbf{x}}_{\mathbf{B}} = \mathbf{B}^{-1}\mathbf{b}, \tag{7.2}$$

които имат една и съща матрицата от коефициенти и тя е 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 взима предвид, 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 стълбовете си. Реализира се идеята на всяка итерация матрицата $\mathbf{B}^{-1}$ да бъде представяна и поддържана във вид на произведение на 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.

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

където

$$\mathbf{E} := \begin{array}{c} \begin{bmatrix} 1 & & w_{1q} & & \\ & \ddots & \vdots & & \\ & & w_{pq} & & \\ & & \vdots & \ddots & \\ & & w_{mq} & & 1 \end{bmatrix} \\ \uparrow \\ \text{стълб } p \end{array}. \tag{7.3}$$

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 (7.3) е елементарната матрица

$$\mathbf{E}^{-1} = \mathbf{I} - \frac{\mathbf{e}_p^{\mathrm{T}}(\mathbf{w}_q - \mathbf{e}_p)}{w_{pq}} = \begin{array}{c} \begin{bmatrix} 1 & & -w_{1q}/w_{pq} & & \\ & \ddots & \vdots & & \\ & & 1/w_{pq} & & \\ & & \vdots & \ddots & \\ & & -w_{mq}/w_{pq} & & 1 \end{bmatrix} \\ \uparrow \\ \text{стълб } p \end{array}. \tag{7.4}$$

В § 7.2 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 елементарни матрици

$$\mathbf{B}_k^{-1} = \mathbf{E}_k^{-1}\mathbf{E}_{k-1}^{-1}\cdots\mathbf{E}_1^{-1}\mathbf{I}, \tag{7.5}$$

където всяка от матриците $\mathbf{E}_i^{-1}$ е от вида (7.4). Обновяването на обратната на базисната матрица става просто чрез добавяне на нова елементарна матрица в произведението. Ясно е, че вместо да се пази 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 $2m$ матрици, което спестява изчислително време, пести памет и намалява грешките от закръгляване.

Още по-нов и съвременен подход е системите (7.2) да бъдат разгледани като три системи с една и съща матрица на коефициентите $\mathbf{B}$

$$\boldsymbol{\pi}\mathbf{B} = \mathbf{c}_{\mathbf{B}}, \quad \mathbf{B}\mathbf{w}_q = \mathbf{A}_q, \quad \mathbf{B}\mathbf{x}_{\mathbf{B}} = \mathbf{b} \tag{7.6}$$

и за тяхното решаване да се използуват числените методи на линейната алгебра, като се направи стандартното директно разлагане на матрицата $\mathbf{B} = \mathbf{L}\mathbf{U}$, 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.

§8. Двойнственост в линейното оптимиране

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, която е тясно свързана с критерия за оптималност на решението на изходната задача.

§8.1. Права задача

Нека е дадена линейна задача в общ вид (L) (вж. § 2). 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 част!).

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

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

§8.2. 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 на това съпоставяне се нарича *спрягане*. Първо ще напишем съответната на (P) двойнствена задача (DP), а след това ще обясним по-подробно връзката между тях.

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

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

Останалите ограничения на (DP) се получават като за всяко $j \in \{1, \dots, n\}$ векторът на двойнствените променливи $\boldsymbol{\pi}(\pi_1, \dots, \pi_m)$ се умножи със съответния вектор стълб $\mathbf{A}_j$ на матрицата на ограниченията $\mathbf{A}$ на (P) (стълбът пред променливата x_j) като полученото скаларно произведение трябва да не 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 $\boldsymbol{\pi}$ и вектора дясна част $\mathbf{b}$ на (P).

§8.3. 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 ѝ 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 8.1. *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

$$
(DP)\quad \begin{aligned} &\max v(\boldsymbol{\pi}) = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \\ &\pi_i \geqslant 0, \quad i \in I, \\ &\pi_i \lesseqgtr 0, \quad i \in \overline{I}, \\ &\mathbf{A}_j^{\mathrm{T}}\boldsymbol{\pi} \leqslant c_j, \quad j \in J, \\ &\mathbf{A}_j^{\mathrm{T}}\boldsymbol{\pi} = c_j, \quad j \in \overline{J}, \end{aligned} \qquad \sim \qquad (DP)\quad \begin{aligned} &-\min[-\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}] \\ &\pi_i \geqslant \;\; 0, \quad i \in I, \\ &\pi_i \lesseqgtr \;\; 0, \quad i \in \overline{I}, \\ &-\mathbf{A}_j^{\mathrm{T}}\boldsymbol{\pi} \geqslant -c_j, \quad j \in J, \\ &-\mathbf{A}_j^{\mathrm{T}}\boldsymbol{\pi} = -c_j, \quad j \in \overline{J}. \end{aligned}
$$

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 приведем задачата (P) в съответната ѝ канонична задача (K). Очевидно (K) също е права задача (задача за минимум, чиито ограничения са равенства с неотрицателна дясна част, върху всичките променливи на която е наложено условие за неотрицателност). Ако на каноничната задача (K) се напише двойнствената ѝ задача (DK), тя ще се различава от (DP) евентуално само по знаците на някои от променливите. Нека $\mathbf{y}\in\mathbb{R}^m$ е векторът на променливите за задачата (DK). Ако задачата (P) е такава, че векторът ѝ дясна част $\mathbf{b}$ 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 $y_i=\pi_i$ 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 $y_i=-\pi_i$, т.е. съответните променливи се различават по знак.

От току що казаното е ясно, че ако разполагаме с решение на (DK), можем да получим от него решение на (DP) и обратно. Следователно можем да се концентрираме върху каноничните задачи на линейното оптимиране и съответните им двойнствени задачи.

Да разгледаме сега канонична задача на линейното оптимиране

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

Съответната ѝ двойнствена задача е

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

§8.4. Теореми за двойнственост

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

Теорема 8.1 (Слаба теорема за двойнственост). *Ако* $\mathbf{x}$ *е допустимо решение за правата задача* (K) *и* $\boldsymbol{\pi}$ *е допустимо решение за двойнствената задача* (DK)*, то* $\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}^{\mathrm{T}}\mathbf{x}$.

Доказателство. От това, че $\mathbf{x}$ е допустимо за (K), имаме $\mathbf{A}\mathbf{x} = \mathbf{b}$ и $\mathbf{x} \geqslant \mathbf{0}$. За произволно $\boldsymbol{\pi} \in \mathbb{R}^m$ получаваме $\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} = (\mathbf{A}\mathbf{x})^{\mathrm{T}}\boldsymbol{\pi} = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{T}\boldsymbol{\pi}$. Ако $\boldsymbol{\pi}$ е допустимо за двойнствената задача (DK), то $\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}$ и от $\mathbf{x} \geqslant \mathbf{0}$ следва, че $\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{x}^{\mathrm{T}}\mathbf{c} = \mathbf{c}^{\mathrm{T}}\mathbf{x}$, т.е. $\mathbf{b}^{\mathrm{T}}\boldsymbol{\pi} \leqslant \mathbf{c}^{\mathrm{T}}\mathbf{x}$. $\square$

Слабата теорема за двойнственост всъщност твърди, че:

I. стойността на целевата функция z в произволно допустимо решение $\mathbf{x}$ за правата задача представлява горна граница за стойността на целевата функция v на двойнствената задача в произволно нейно допустимо решение $\boldsymbol{\pi}$ (включително и оптималното);

II. стойността на целевата функция v в произволно допустимо решение $\boldsymbol{\pi}$ за двойнствената задача представлява долна граница за стойността на целевата функция z на правата задача в произволно нейно допустимо решение $\mathbf{x}$ (включително и оптималното).

Като следствие от теоремата получаваме

Следствие 8.1. *Нека* $\mathbf{x}^*$ *е допустимо решение за правата задача* (K)*, а* $\boldsymbol{\pi}^*$ *е допустимо решение за двойнствената задача* (DK)*. Ако* $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}^*$*, то* $\mathbf{x}^*$ *и* $\boldsymbol{\pi}^*$ *са оптимални решения на съответните задачи.*

Доказателство. Според слабата теорема за двойнственост за всяко допустимо решение $\mathbf{x}$ на (K) 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 всяко допустимо решение $\boldsymbol{\pi}$ на (DK) 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). $\square$

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

Отговор на този въпрос дава:

Теорема 8.2 (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. (а) Тъй като всяка от задачите (K) и (DK) може да бъде разглеждана като права задача, ще допуснем, че е разрешима задачата (K) и ще покажем, че е разрешима задачата (DK).

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

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

Да разгледаме съответния на $\overline{\mathbf{x}}$ вектор на симплексните множители $\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{c}_{\mathbf{B}}^{\mathrm{T}}\mathbf{B}^{-1}$ (вж. § 7.4). Имаме, че

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

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

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

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

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

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

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

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

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

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 $\mathbf{B}$ е текущото базисно допустимо решение за симплексната итерация, неговият 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 $\mathbf{B}$ са неотрицателни, той 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 $\mathbf{B}$, то $\boldsymbol{\pi}^{*\mathrm{T}} := \mathbf{c}_{\mathbf{B}}^{\mathrm{T}}\mathbf{B}^{-1}$ е допустимо решение за двойнствената задача (DK), което удовлетворява $\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}}^* = \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}^*$.

От Следствие 8.1 получаваме, че $\boldsymbol{\pi}^*$ е оптимално решение за (DK), с което доказателството на (а) е приключено.

(б) Следва непосредствено от Теорема 8.1 с допускане на противното. □

Да отбележим, че обратното на Теорема 8.2(б) в общия случай 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 несъвестими, както се вижда от следния

Пример. В двойката спрегнати задачи

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

и двете са несъвместими.

§9. Класическа транспортна задача

§9.1. Икономическа постановка

Даден продукт се произвежда в пунктове $A_1, \ldots, A_m$ в количества съответно $a_1, \ldots, a_m$, а се потребява в пунктове $B_1, \ldots, B_n$ съответно в количества $b_1, \ldots, b_n$, като сумарното производство е равно на сумарното потребление, т.е.

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{9.1}$$

Транспортните разходи за превоза на единица продукт от пункта A_i до пункта B_j са съответно c_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Очевидно $c_{ij} \geqslant 0$.

Задачата е пунктовете на потребление $B_1, \ldots, B_n$ да се снабдят, така че техните потребности да бъдат задоволени и общите транспортни разходи да бъдат минимални.

Условието (9.1) се нарича още *условие за баланс*. Без ограничение на общността считаме, че $a_i > 0$, $i = 1, \ldots, m$ и $b_j > 0$, $j = 1, \ldots, n$.

§9.2. Математически модел

Да означим с x_{ij} количеството продукт, с което пунктът A_i, $i = 1, \ldots, m$, ще снабди пункта B_j, $j = 1, \ldots, n$. Като означим вектора на променливите с $\mathbf{x}(x_{11}, \ldots, x_{1n}, x_{21}, \ldots x_{2n}, \ldots, x_{mn})$, задачата е

$$(TP) \qquad \begin{aligned} &\min z(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\ &\sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1, \ldots, m, \\ &\sum_{i=1}^{m} x_{ij} = b_j, \quad j = 1, \ldots, n, \\ &x_{ij} \geqslant 0, \qquad i = 1, \ldots, m, \ j = 1, \ldots, n. \end{aligned}$$

Следователно *транспортната задача* (TP) е задача на линейното оптимиране в каноничен вид. Броят на променливите е mn, а броят на ограниченията от тип равенство е $m + n$.

Ако в (TP) означим с $\mathbf{c}$ вектора на целевата функция, с $\tilde{\mathbf{b}}(a_1, \ldots, a_m, b_1, \ldots, b_n)$ — вектора на дясната част, а с $\mathbf{A}$ — съответната $(m + n) \times mn$ матрица на ограниченията, можем да запишем задачата в следния матричен вид:

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

§9.3. (TP) като частен случай на линейна задача

Нека да разпишем $(m+n)$-те ограничения равенства на транспортната задача:

$$
\begin{array}{lllll}
x_{11}+\cdots+x_{1n} & & & & = a_1,\\
 & x_{21}+\cdots+x_{2n} & & & = a_2,\\
\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\
 & & & x_{m1}+\cdots+x_{mn} & = a_m,\\
x_{11} & +x_{21} & & +x_{m1} & = b_1,\\
x_{12} & +x_{22} & & +x_{m2} & = b_2,\\
\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\
x_{1n} & +x_{2n} & & +x_{mn} & = b_n.
\end{array}
$$

Коефициентите пред неизвестните са нули или единици. Всяка променлива x_{ij} участва с коефициент 1 само в две уравнения: i-то и $(m+j)$-то. Това означава, че матрицата на ограниченията $\mathbf{A}$ на (TP) съдържа само 0 и 1, а вектор-стълбовете ѝ $\mathbf{A}_{ij}$ имат точно по две координати различни от нула и това са i-та и $(m+j)$-та.

§9.4. Свойства на транспортната задача

Теорема 9.1. *Условието за баланс (9.1) е необходимо и достатъчно условие (TP) да бъде разрешима.*

Доказателство. Необходимостта е очевидна: ако $\mathbf{x}^*$ е оптимално решение на (TP), то тъй като $\mathbf{x}^*$ е допустимо, то удовлетворява всички ограничения. Като сумираме първите m ограничения за $\mathbf{x}^*$ получаваме

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^* = \sum_{i=1}^{m} a_i, \tag{9.2}$$

а като сумираме последните n ограничения за $\mathbf{x}^*$ получаваме

$$\sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij}^* = \sum_{j=1}^{n} b_j, \tag{9.3}$$

От това, че левите страни на равенствата (9.2) и (9.3) са равни следва, че и десните им страни са равни, което е (9.1).

Обратно, ако е изпълнено (9.1) и означим $s := \sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, то векторът $\mathbf{x}$ с координати $x_{ij} := \dfrac{a_i b_j}{s}$ е допустим за (TP) и тя следователно е съвместима. Понеже целевата ѝ функция z е ограничена отдолу върху допустимото ѝ множество $(z \geqslant 0)$, то транспортната задача е разрешима. $\square$

Тъй като за стойности на m и n, поне едната от които не е единица (и това е нормалният случай), е изпълнено $m+n \leqslant mn$, то очевидно $r(\mathbf{A}) \leqslant m+n$.

Ако умножим първите m вектор-редове на $\mathbf{A}$ с 1, а останалите ѝ n вектор-редове с -1 и ги съберем, ще получим нулевия вектор, което означава, че вектор-редовете на $\mathbf{A}$ са линейно зависими. Следователно $r(\mathbf{A}) \leqslant m+n-1$. Ще докажем

Теорема 9.2. *$r(\mathbf{A}) = m+n-1$.*

Доказателство. Тъй като детерминантата

$$\det \left\| \frac{\mathbf{A}_{1n}\mathbf{A}_{2n}\ldots\mathbf{A}_{mn}\mathbf{A}_{11}\mathbf{A}_{12}\ldots\mathbf{A}_{1\,n-1}}{\text{без последната координата}} \right\| =$$

$$= \det \left\| \begin{array}{cccccc} 1 & 0 & 0\ldots 0 & 1 & 1\ldots 1 \\ 0 & 1 & 0\ldots 0 & 0 & 0\ldots 0 \\ 0 & 0 & 1\ldots 0 & 0 & 0\ldots 0 \\ \multicolumn{6}{c}{\dotfill} \\ 0 & 0 & 0\ldots 1 & 0 & 0\ldots 0 \\ 0 & 0 & 0\ldots 0 & 1 & 0\ldots 0 \\ \multicolumn{6}{c}{\dotfill} \\ 0 & 0 & 0\ldots 0 & 0 & 0\ldots 1 \end{array} \right\| = 1$$

(за да се убедите, можете да развиете детерминантата по първия стълб), рангът на $\mathbf{A}$ е точно $m+n-1$. □

Следователно всяко базисно допустимо решение $\overline{\mathbf{x}}$ на (TP) ще има базис от $m+n-1$ променливи.

§9.5. Транспортна таблица. Основни понятия

Транспортната таблица (ТТ) се състои от редове, стълбове и клетки. Всяка клетка се определя с една наредена двойка числа (i,j) — номерата на реда и стълба в ТТ, в които тя се намира. С клетката (i,j) се асоциират променливата x_{ij}, транспортните разходи c_{ij} и съответният вектор-стълб $\mathbf{A}_{ij}$ на матрицата на ограниченията $\mathbf{A}$.

Базисни клетки наричаме ония клетки, за които съответните променливи x_{ij} образуват базис на транспортната задача, разглеждана като задача на линейното оптимиране. Съгласно Теорема 9.2 техният брой е $m+n-1$.

Затворена начупена линия наричаме линия, образувана от отсечки по такъв начин, че всеки край на дадена отсечка от линията е край на точно още една отсечка от нея. Краищата на отсечките се наричат *върхове* на начупената линия.

Цикъл наричаме всяка затворена начупена линия с върхове в клетки на ТТ, която удовлетворява следните две условия:

а) всяка отсечка лежи изцяло в ред или в стълб на ТТ;

б) двете отсечки, които излизат от всеки връх на цикъла, лежат едната в ред, а другата в стълб на ТТ.

Съвкупност от клетки на ТТ, която съдържа поне един цикъл се нарича *циклична*. В противен случай се нарича *ациклична*. Най-общо един произволен цикъл γ може да се опише по следния начин

$$\gamma : (i_1, j_1)(i_2, j_1)(i_2, j_2) \dots (i_s, j_s)(i_1, j_s).$$

§9.6. Свойства на базисните допустими решения на класическата транспортна задача

Теорема 9.3. *Съвкупност от вектор-стълбове $\{\mathbf{A}_{ij}\}$ на матрицата $\mathbf{A}$ е линейно независима тогава и само тогава, когато съответната ѝ съвкупност от клетки в ТТ е ациклична.*

Доказателство. Нека съвкупността от вектор-стълбове $\mathbf{A}_{ij}$ на матрицата $\mathbf{A}$ е линейно независима.

Да допуснем, че съответният им набор от клетки в ТТ е цикличен, т.е. съдържа цикъл

$$\gamma : (i_1, j_1)(i_2, j_1)(i_2, j_2) \dots (i_s, j_s)(i_1, j_s).$$

Тогава

$$\mathbf{A}_{i_1 j_1} - \mathbf{A}_{i_2 j_1} + \mathbf{A}_{i_2 j_2} + \dots + \mathbf{A}_{i_s j_s} - \mathbf{A}_{i_1 j_s} = \mathbf{0}.$$

Защо това е така? Освен нулеви координати, векторът $\mathbf{A}_{i_1 j_1}$ има 1 само на i_1 и $m+j_1$ място. С изваждане от него на вектора $\mathbf{A}_{i_2 j_1}$ се елиминира 1 на $m+j_1$ място. Координатата $m+j_1$ става равна на 0, а координатата i_2 става -1. С добавяне към разликата на вектора $\mathbf{A}_{i_2 j_2}$ се елиминира 1 на i_2 място, т.е. i_2 координата става нула, но се появява 1 на $m + j_2$ място, и т. н. Накрая, с изваждането на вектора $\mathbf{A}_{i_1 j_s}$ се елиминират 1 на $m + j_s$ място, появила се от добавянето на предходния вектор $\mathbf{A}_{i_s j_s}$ и 1 на i_1 място, появила се от първия вектор $\mathbf{A}_{i_1 j_1}$. Получаваме, че всички координати са нули.

Следователно имаме линейна зависимост на част от вектор-стълбовете , а оттам и на цялата съвкупност $\mathbf{A}_{ij}$, а това води до противоречие.

Обратно, нека векторите $\mathbf{A}_{ij}$ са линейно зависими. Да означим с β_{ij} съответните коефициенти в линейната комбинация, която изразява нулевия вектор чрез $\mathbf{A}_{ij}$. Нека $\beta_{i_1 j_1} > 0$. Това означава, че ще съществува $\beta_{i_2 j_1} < 0$ (за да се компенсира положителната координата $\beta_{i_1 j_1}$ на вектора $\beta_{i_1 j_1}\mathbf{A}_{i_1 j_1}$ в сумата). Аналогично трябва да съществува $\beta_{i_2 j_2} > 0$ (за да се компенсира отрицателната координата $\beta_{i_2 j_1}$ на вектора $\beta_{i_2 j_1}\mathbf{A}_{i_2 j_1}$ в сумата) и т. н. Като проследим зависимостта на индексите, виждаме че можем да построим цикъл

$$\gamma : (i_1, j_1)(i_2, j_1)(i_2, j_2) \dots (i_s, j_s)(i_1, j_s)$$

в ТТ. □

Следствие 9.1. *Ако* $\overline{\mathbf{x}}$ *е базисно допустимо решение за* (TP)*, то съвкупността от клетки в ТТ, съответстваща на базисните му координати, е ациклична.*

Доказателство. Знаем, че ако $\overline{\mathbf{x}}$ е базисно допустимо решение, то вектор-стълбовете на $\mathbf{A}$, съответстващи на положителните му координати, са линейно независими (Теорема 3.1) и остава да приложим Теорема 9.3. □

Следователно, ако $\overline{\mathbf{x}}$ е базисно допустимо решение (връх) за (TP) и клетките на ТТ, съответстващи на базисните му променливи, запълним с базисните му координати а клетките, съответстващи на небазисните му координати (които, разбира се, са нули), оставим празни, то пълните клетки ще бъдат $m+n-1$ на брой и ще образуват ациклична съвкупност. Получаваме ТТ на базисното допустимо решение $\overline{\mathbf{x}}$, която означаваме с $T(\overline{\mathbf{x}})$.

Теорема 9.4. *Нека* $\overline{\mathbf{x}}$ *е връх. За произволна празна клетка на* $T(\overline{\mathbf{x}})$ *съществува точно един цикъл, който я свързва с базисните (пълните) клетки.*

Доказателство. В $T(\overline{\mathbf{x}})$ базисните (пълните) клетки са $m+n-1$ на брой и образуват ациклчна съвкупност H. Нека (k,l) е празна клетка. Да я присъединим към H. Получаваме нова съвкупност $H' = H \cup (k,l)$, която е циклична, тъй като се състои от $m+n$ клетки (всеки $m+n$ стълба на $\mathbf{A}$ са линейно зависими съгласно Теорема 9.2, а на линейно зависима съвкупност от вектори съответства циклична съвкупност от клетки, съгласно Теорема 9.2). При това всеки цикъл в H' съдържа (k,l), тъй като в противен случай H би била циклична съвкупност.

За да установим единствеността, да допуснем, че съществуват два различни цикъла, свързващи (k,l) с H:

$$(k,l)(k,j_1)(i_1,j_1)\ldots(i_p,j_p)(i_p,l)$$

и

$$(k,l)(k,s_1)(r_1,s_1)\ldots(r_q,s_q)(r_q,l).$$

Всички клетки в тях без (k,l) са базисни. Махаме (k,l) и построяваме цикъл от базисни клетки

$$(i_p,l)(i_p,j_p)\ldots(i_1,j_1)(k,j_1)(k,s_1)(r_1,s_1)\ldots(r_q,s_q)(r_q,l),$$

което е противоречие с това, че $\overline{\mathbf{x}}$ е базисно допустимо решение и базисните му клетки образуват ациклична съвкупност. □

§9.7. Разпределителен метод за решаване на (TP)

Преминаване от едно базисно допустимо решение $\overline{\mathbf{x}}$ към съседно на него базисно допустимо решение $\overline{\mathbf{x}}'$.

Нека $\overline{\mathbf{x}}$ е базисно допустимо решение за (TP) с транспортна таблица $T(\overline{\mathbf{x}})$. Нека (k,l) е празна клетка в $T(\overline{\mathbf{x}})$. Според Теорема 9.4 съществува единствен цикъл, който я свързва с базисните (пълните) клетки на $T(\overline{\mathbf{x}})$. Нека това е цикълът

$$\gamma_{kl}: \begin{array}{cccccc} (k,l) & (k,j_1) & (i_1,j_1) \ldots & (i_s,j_s) & (i_s,l). \\ + & - & + \;\;\ldots & + & - \end{array}$$

Да отбележим, че в него всички клетки освен (k,l) са базисни за $\overline{\mathbf{x}}$.

На клетките на цикъла γ_{kl} присвояваме алтернативно знаците $+$ и $-$, като започнем с $+$ от клетката (k,l). Тогава клетките на цикъла γ_{kl} се разделят на две равномощни съвкупности γ_{kl}^+ и γ_{kl}^-. Въвеждаме числото

$$\overline{c}_{kl} := \sum_{(i,j)\in\gamma_{kl}^+} c_{ij} - \sum_{(i,j)\in\gamma_{kl}^-} c_{ij}.$$

Да означим

$$\overline{\theta} := \min_{(i,j)\in\gamma_{kl}^-} \overline{x}'_{ij}.$$

Напомняме, че $\overline{x}_{ij}$ са координатите на $\overline{\mathbf{x}}$ и тъй като в съвкупността γ_{kl}^- участват само базисни клетки, то координатите $\overline{x}_{ij}$, участващи в определянето на $\overline{\theta}$ са базисни за $\overline{\mathbf{x}}$.

Координатите $\overline{x}'_{ij}$ на съседното базисно допустимо решение $\overline{\mathbf{x}}'$ намираме по следния начин

$$\begin{array}{ll} \overline{x}'_{ij} = \overline{x}_{ij} + \overline{\theta}, & (i,j) \in \gamma_{kl}^+, \\ \overline{x}'_{ij} = \overline{x}_{ij} - \overline{\theta}, & (i,j) \in \gamma_{kl}^-, \\ \overline{x}'_{ij} = \overline{x}_{ij}, & (i,j) \notin \gamma_{kl}. \end{array}$$

В ТТ, съответстваща на новото $\overline{\mathbf{x}}'$ се изпразва **точно една** от старите базисни клетки (клетка, принадлежаща на γ_{kl}^-, за чийто индекс се достига $\overline{\theta}$), а клетката (k,l) се запълва с количество $\overline{\theta}$, понеже $(k,l) \in \gamma_{kl}^+$ и следователно

$$\overline{x}'_{kl} = \overline{x}_{kl} + \theta = 0 + \overline{\theta} = \overline{\theta}.$$

Ако минимумът при определянето на $\overline{\theta}$ се достигне за повече от една клетка $(i,j) \in \gamma_{kl}^-$, то само една клетка (i_p,j_p) от тях се оставя празна (излиза от базиса), а останалите се запълват с нули (базисни нули!) в новото $\overline{\mathbf{x}}'$.

Новополученият вектор $\overline{\mathbf{x}}'$ остава допустим за транспортната задача, тъй като разпределянето на количеството $\overline{\theta}$ (откъдето идва и името на метода — разпределителен метод) се извършва последователно в рамките на един ред или стълб и така се осигурява изпълнението на ограниченията равенства (т.е. $\overline{\mathbf{x}}'$ удовлетворява $\mathbf{A}\overline{\mathbf{x}}' = \tilde{\mathbf{b}}$), а от избора на $\overline{\theta}$ следва, че $\overline{\mathbf{x}}' \geqslant \mathbf{0}$.

Нещо повече, $\overline{\mathbf{x}}'$ е базисно допустимо решение, тъй като съвкупността от пълните клетки H' на $\overline{\mathbf{x}}'$ се различава от съвкупността от пълните клетки H на $\overline{\mathbf{x}}$ само по това, че съдържа клетката (k,l), а не съдържа една от клетките на γ_{kl}^- — клетката (i_p, j_p). Следователно в H' с изключение на (k,l) всички останали клетки принадлежат на ациклична съвкупност. Тогава, ако в H' има цикъл, той непременно съдържа (k,l). Така получаваме, че съществуват поне два различни цикъла — този в H' (който не съдържа клетката (i_p, j_p)) и γ_{kl} (който съдържа клетката (i_p, j_p)), които свързват (k,l) с базисните клетки H на $\overline{\mathbf{x}}$, което е противоречие с Теорема 9.4. □

Сравняване на $z(\overline{\mathbf{x}})$ и $z(\overline{\mathbf{x}}')$.

Ако означим

$$S := \sum_{(i,j)\notin\gamma_{kl}} c_{ij}\overline{x}_{ij}, \quad S^+ := \sum_{(i,j)\in\gamma_{kl}^+} c_{ij}\overline{x}_{ij}, \quad S^- := \sum_{(i,j)\in\gamma_{kl}^-} c_{ij}\overline{x}_{ij},$$

то

$$z(\overline{\mathbf{x}}') = \sum_{(i,j)} c_{ij}\overline{x}'_{ij} = \sum_{(i,j)\notin\gamma_{kl}} c_{ij}\overline{x}_{ij} + \sum_{(i,j)\in\gamma_{kl}^+} c_{ij}(\overline{x}_{ij}+\overline{\theta}) + \sum_{(i,j)\in\gamma_{kl}^-} c_{ij}(\overline{x}_{ij}-\overline{\theta}) =$$

$$= S + S^+ + S^- + \overline{\theta}\left(\sum_{(i,j)\in\gamma_{k,l}^+} c_{ij} - \sum_{(i,j)\in\gamma_{kl}^-} c_{ij}\right) = z(\overline{\mathbf{x}}) + \overline{\theta}\overline{c}_{kl}.$$

Критерий за оптималност на $\overline{\mathbf{x}}$.

Ако за всички празни (небазисни) за $T(\overline{\mathbf{x}})$ клетки (k,l) е изпълнено, че $\overline{c}_{kl} \geqslant 0$, то базисното допустимо решение $\overline{\mathbf{x}}$ е оптимално. В противен случай може да се премине към „не по-лошо“ съседно базисно допустимо решение $\overline{\mathbf{x}}'$. Числата $\overline{c}_{kl}$, които се пресмятат за празните (небазисните) клетки (k,l), са относителните оценки на небазисните променливи x_{kl}.

§9.8. Алгоритъм на разпределителния метод за решаване на (TP)

(0) Намира се начално базисно допустимо решение $\overline{\mathbf{x}}$.

(1) За всяка празна (небазисна) клетка (i,j) в таблицата $T(\overline{\mathbf{x}})$ се пресмятат относителните оценки $\overline{c}_{ij}$. Ако всички празни клетки имат $\overline{c}_{ij} \geqslant 0$, то КРАЙ, $\overline{\mathbf{x}}$ е оптимално. В противен случай се преминава към стъпка (2).

(2) Избира се празна клетка (k,l), такава че $\overline{c}_{kl} < 0$ и се преминава към съседно на $\overline{\mathbf{x}}$ базисно допустимо решение $\overline{\mathbf{x}}'$ (или към нов базис на $\overline{\mathbf{x}}$, ако $\overline{\mathbf{x}}$ е изродено). Сменяме $\overline{\mathbf{x}}$ с $\mathbf{x}'$ и се връщаме на стъпка (1).

§9.9. Крайност на алгоритъма. Зацикляне. Избягване на зациклянето

Алгоритъмът на разпределителния метод е краен поради това, че имаме само краен брой базисни допустими решения. Зацикляне може да се получи само в изроден връх. Необходимо и достатъчно условие задачата да бъде изродена, е да съществува частичен баланс, т. е. баланс между производството и потреблението на непълни групи производители и потребители. Зацикляне в малки учебни примери на практика не се получава, а в големи задачи зациклянето може да се избегне, като предварително задачата се преобразува в неизродена чрез разрушаване на частичните баланси и преминаване към т.нар. ε-задача

$$
(TP_\varepsilon) \qquad \begin{aligned}
&\min z(\mathbf{x}) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}\\
&\sum_{j=1}^{n} x_{ij} = a'_i, \quad i = 1,\dots,m,\\
&\sum_{i=1}^{m} x_{ij} = b'_j, \quad j = 1,\dots,n,\\
&x_{ij} \geqslant 0, \qquad\quad i = 1,\dots,m,\ j = 1,\dots,n,
\end{aligned}
$$

където $a'_i = a_i + \varepsilon$, $i = 1,\dots m$, $b'_j = b_j$, $j = 1,\dots,n-1$, $b'_n = b_n + m\varepsilon$.

§9.10. Методи за намиране на начално базисно допустимо решение за $(\boldsymbol{TP})$

Метод на северозападния ъгъл.

Започва се от северозападната клетка на ТТ — клетката $(1,1)$. На базисната променлива x_{11} се присвоява стойност $\overline{x}_{11} := \min\{a_1, b_1\}$.

1) Ако $a_1 < b_1$, то $b_1 \Leftarrow b_1 - a_1$ и се зачертава 1-ия ред (реда на a_1) от ТТ.

2) Ако $b_1 < a_1$, то $a_1 \Leftarrow a_1 - b_1$ и се зачертава 1-ия стълб (стълба на b_1) от ТТ.

3) Ако $a_1 = b_1$, то или $b_1 \Leftarrow 0$ и се зачертава 1-ия ред от ТТ, или $a_1 \Leftarrow 0$ и се зачертава 1-ия стълб от ТТ.

Така размерността на ТТ се намалява с единица. За новата таблица процедурата се повтаря и така до пълно изчерпване.

В третия случай на следващата итерация ще се получи нулева стойност на съответната базисна променлива, т. е. началното базисно допустимо решение ще бъде изродено.

Запълването на клетките става от крайната горе вляво (северозападната клетка) до крайната долу вдясно, като се върви алтернативно в редове или стълбове, при което очевидно се получава ациклична съвкупност от $m + n - 1$ клетки. Следователно, стига се до базисно допустимо решение.

Метод на минималния елемент.

Започва се от клетка (i_0, j_0) на ТТ с минимални транспортни разходи, т. е. от клетка, за която $c_{i_0 j_0} = \min c_{ij}$. На базисната променлива $x_{i_0 j_0}$ се присвоява стойност $\overline{x}_{i_0 j_0} := \min\{a_{i_0}, b_{j_0}\}$.

1) Ако $a_{i_0} < b_{j_0}$, то $b_{j_0} \Leftarrow b_{j_0} - a_{i_0}$ и се зачертава ред i_0 (реда на a_{i_0}) от ТТ.

2) Ако $b_{j_0} < a_{i_0}$, то $a_{i_0} \Leftarrow a_{i_0} - b_{j_0}$ и се зачертава стълб j_0 (стълба на b_{j_0}) от ТТ.

3) Ако $a_{i_0} = b_{j_0}$, то или $b_{j_0} \Leftarrow 0$ и се зачертава ред i_0 от ТТ, или $a_{i_0} \Leftarrow 0$ и се зачертава стълб j_0 от ТТ.

Така размерността на ТТ се намалява с единица. За новата таблица процедурата се повтаря и така до изчерпване.

В третия случай в някоя от следващите итерации ще получим нулева стойност на базисна променлива и началното базисното допустимо решение ще бъде изродено.

Запълването на клетките става така, че на всяка итерация се елиминира ред или стълб и не е възможно да се получи цикъл. Следователно получава се базисно допустимо решение.

Поради факта, че методът на минималния елемент отчита транспортните разходи, а методът на северозападния ъгъл не, вероятността първият да намери начално базисно допустимо решение, което е по-близо до оптималното, е по-голяма.

§9.11. Целочисленост на решението на (TP)

Нека $\overline{\mathbf{x}}$ е базисно решение на (TP) с базисна матрица $\mathbf{B}$. Понеже коефициентите на базисната матрица $\mathbf{B}$ са само нули и единици, решаването на системата $\mathbf{B}\overline{\mathbf{x}}_{\mathbf{B}} = \tilde{\mathbf{b}}$ става без да е необходимо да се извършват деления (дели се на 1) и без да е необходимо да се извършват умножения (умножава се с 1). С други думи, гаусовата елиминация при решаване на системата $\mathbf{B}\overline{\mathbf{x}}_{\mathbf{B}} = \tilde{\mathbf{b}}$ се свежда до проста последователност от събирания и изваждания. Следователно, ако векторът $\tilde{\mathbf{b}}$ с координати $(a_1, \ldots, a_m, b_1, \ldots, b_n)$ е целочислен, базисните кординати $\overline{\mathbf{x}}_{\mathbf{B}}$ също ще бъдат целочислени, тъй като ще се получават чрез събиране и изваждане на целите числа a_i, b_j.

Следователно, ако всички a_i и b_j са цели числа, то всяко базисно решение $\overline{\mathbf{x}}$ ще има целочислени координати.