Matematinis Programavimas

A.Žilinskas

2005

Abstract

Matemathical Programming

A.Žilinskas

The textbook is oriented to rather a broad audience. The students of departments of Informatics, Economics, Engineering and Applied Mathematics are meant as the primary users of the text. The course has been developed and the textbook has been written as a part of TEMPUS project S_JEP 12382-97 (partner universities VMU, Copenhagen, Maastricht, KUT) [25]. The textbook contains theoretical material of the lectures. The instructions of practical work are included into WWW home page of the project. Some illustrative materials and examples of problem solutions, prepared by our students, are also available at the home page as well as advanced theoretical material (e.g. lectures by prof. J.Clausen) and references to software.

The course was prepared under strong influence of methodology of Project Based Learning (PBL). I was introduced into the ideas of PBL by D.Tempelaar and became a supporter of PBL mainly because of his enthusiasm. While preparing the course many similar courses of different universities were analised and discussed with the colleagues, especially from the partner universities. I would like to thank also my colleagues from Institute of Mathematics and Informatics J.Mockus, V.Šaltenis, G.Dzemyda, V.Tiešis whose critics helped me during many years. The indirect influence of the colleagues from institutions where I delivered lectures as a visiting professor or worked as a researcher (Dortmund university, RWTH Aachen, Abo Akademi, Danish Technical University) is acknowledged. I would like to thank many people, but let me mentioned here only few: J.Clausen, A.Kolen, K.Madsen, R.Mathar, H.Peters, H.-P.Schwefel, A.Törn.

ĮVADAS

1.1 Matematinio programavimo kurso apibūdinimas

Optimizavimo uždaviniai aktualūs įvairiose žmonių veiklos srityse. Inžinieriai, projektuojantys naujas technologijas, siekia maksimalaus efektyvumo. Verslininkai ir vadybininkai stengiasi darbą organizuoti taip, kad būtų gautas maksimalus pelnas. Šeimininkė, išėjusi apsipirkti, stengiasi pasirinkti maršrutą taip, kad kelionei sugaištų galimai mažiau laiko. Kasdieninėje veikloje dažnai optimizuojama intuityviai, nesinaudojant jokia teorija. Tačiau svarbiais ir sudėtingais atvejais stengiamasi uždavinį suformuluoti tiksliai ir jį spręsti tam skirtais metodais. Tradiciškai optimizavimo uždavinų esant ribojimams sprendimo teorija ir metodai vadinami **matematiniu programavimu.** Istoriškai pirmiausiai buvo sukurta tiesinio progrmavimo teorija, davusi pradžią ir termino *programavimas* paplitimui šiuo kontekstu. Šis terminas buvo pasirinktas todėl, kad sukurtieji metodai buvo skirti optimaliam ekonominių, transporto, karinių *programų* sudarymui. Terminas *programavimas* šiuo kontekstu atsirado netgi anksčiau negu kitu, gal būt dabar labiau įprastu kontekstu, reiškiančiu algoritmų rašymą, iš pradžių kompiuterio komandomis, vėliau - programavimo kalbomis.

Optimizavimo teorija ir optimizavimo metodų bei juos realizuojančios programinės įrangos kūrimas yra aktyviai vystoma mokslo ir technikos šaka. Pagrindinės mokslinės literatūros leidyklos publikuoja virš dešimt tarptautinių žurnalų ir knygų serijų, skirtų optimizacijos tematikai. Programinės įrangos rinkoje siūloma kelios dešimtys komerciškai platinamų optimizavimo paketų. Žymiai daugiau programinės įrangos sukurta nekomerciniais pagrindais. Kadangi optimizavimo metodai svarbūs ir teoriniu, ir taikomuoju požiūriais, daugelyje univeritetų skaitomi tokios tematikos kursai, kurie vadinami *Matematinis programavimas, Tiesinis/Netiesinis programavimas, Optimizavimo metodai* ir pan.

Įvairiuose universitetuose skaitomi kursai skiriasi tiek turiniu, tiek medžiagos atrinkimu, tiek ir jos pateikmo metodika. Tas ryšku jau vien pavarčius vadovėlius. Dažnai dėmesys sukoncentruojamas į tiesinį programavimą, netiesiniams uždaviniams skiriant santykinai mažą kurso dalį, pvz. [8], [21]. Toks medžiagos atrinkimas dažnai grindžiamas tuo, kad daugelis praktikoje, ypač ekonomikoje, kylančių uždavinių yra tiesiniai. Tačiau vis dažniau tenka nagrinėti netiesinius uždavinius, o tiesinių ir netiesinių uždavinių sprendimo metodų idėjos vis labiau persipina. Todėl šiame vadovėlyje stengtasi subalansuoti tiesiniams ir netiesiniams uždaviniams skirtos medžiagos apimtį.

Tokie vadovėliai, kaip [4], [26], labai formalūs: griežtai matematiškai formuluojamos teoremos, pateikiami detalūs įrodymai, beveik nėra iliustracijų ir praktinių uždavinų sprendimo pavyzdžių. Tokie vadovėliai skirti gan siaurai auditorijai, būtent matematikams besispecializuojantiems optimizacijos teorijos srityje.

Galima nurodyti priešingos orientacijos, t.y. labai praktiškų knygų, pvz. [9], [22], kurios apsiriboja metodų idėjų aprašymu ir jų pagrindu realizuotų algoritmų testavimo rezultatų aptarimu. Jos nesunkiai ir greitai skaitomos, jose pateikiamos rekomendacijos, kaip susirasti tinkamą metodą spręsti tipiniam uždaviniui. Tokios knygos skirtos vartotojui, kuris gali pasitenkinti gan paviršutiniškomis žiniomis. Paprastų receptų metodo pasirinkimui pakanka, jei iškylantys praktiniai uždaviniai nelabai sudėtingi, gerai atitinkantys pagrindžiančias metodus teorines prielaidas.

Tačiau informatikos, ekonomikos ir kitų sričių specialistai, gvildenantys šiuolaikines problemas, gan dažnai susiduria su sudėtingais optimizavimo uždaviniais, kurių sprendimas reikalauja kiek gilesnio, negu vien metodo idėja, optimizavimo teorijos suvokimo. Ketinantiems specializuotis optimizavimo metodų srityje nesunku išsirinkti studijoms keletą gerų knygų, pvz. [3], [2], [16], [11], [15], kuriose pavyko suderinti teoriją su praktika. Tačiau tokios studijos reikalauja daug laiko. Nors ir nerealu, kad taikomosios srities specialistas galėtų ilgam pasišvęsti optimizacijos teorijos studijoms, tačiau susidūręs su sudėtingesniu optimizavimo uždaviniu jis turi būti pasiruošęs ir pasiskaityti optimizacijos teorijos literatūrą, ir susivokti programinės įrangos aprašymuose, ir nesuklysti vertindamas programų paketų reklamą. Tokiam specialistui reikalingu teorijos ir metodų įvadu ir pretenduoja tapti šis vadovėlis.

Šalia standartinių matematinio programavimo metodų, vadovėlyje trumpai išdėstytas racionalaus išrinkimo teorijos bei optimizavimo su daugeliu kriterijų įvadas. Matematinio programavimo kursuose ši tematika dažniausiai nenagrinėjama. Tačiau ji tampriai susijusi su optimizavimo metodų taikymais ir jai reikia rasti vietos studijų programoje. Jei specialus *Optimalių sprendimų priėmimo/Daugiakriterių sprendimų* kursas nėra dėstomas, tai trumpą jų įvadą tikslingiausia pateikti metodologiškai artimame *Matematinio programavimo* kurse.

Šis vadovėlis sumanytas kaip optimizavimo teorijos ir praktikos vienovės projekcija į kvalifikuoto vartotojo interesus, atsižvelgiant į studijoms skiriamo laiko ribotumą, atitinkantį vieno semestro kursą bakalauro studijų programoje. Matyt, neįmanoma parašyti vadovėlį vienodai gerai tinkantį kelių specialybių studentams. Išdėstytai medžiagai studijuoti reikalingas pairuošimas neviršija matematinės analizės ir tiesinės algebros elementų. Ruošiant vadovėlį, buvo orientuojamasi į Vytauto Didžiojo ir Kauno Technologijos universitetų studentų poreikius. Autoriui teko dėstyti šį kursą ir silpniau matematiškai pasiruošusiai auditorijai, kuriai reikėdavo paaiškinti būtinus matricų teorijos rezultatus; be to, būdavo praleidžiami kai kurie įrodymai, o kursas papildomas praktiniais pavyzdžiais. Savarankiškai studijuojantį skaitytoją perspėsime, kad norint suprasti dėstomus metodus, reikėtų būti susipažinusiam su daugelio kintamųjų funkcijų skleidimu Teiloro eilute, vektorių tiesinio nepriklausomumo sąvoka, matricos tikrinių reikšmių bei vektorių sąvokomis.

Autoriaus nuomone, matematikams, besiruošiantiems dirbti taikymų srityje, taip pat būtų naudinga išklausyti matematinio programavimo kursą, orientuotą į praktinių uždavinių sprendimą. Vargu ar matematinio programavimo uždaviniai bei metodai yra tinkamiausia sritis matematinei erudicijai lavinti. Bet ši sritis, be abejonės, yra teorijos ir praktikos sąveikos sritis. Todėl vengėme ilgų ir painių įrodinėjimų, o stengėmės išryškinti metodų idėjas, paaiškinti, kada kokie metodai yra tinkamiausi. Gilesnėms teorijos studijoms iš literūros sąrašo galima pasirinkti, pavyzdžiui [2], [3], [11].

Vadovėlis parengtas vykdant TEMPUS projektą S_JEP 12382-97, koordinuojamą Vytauto Didžiojo universiteto ir vykdomą kartu su Kauno Technologijos universitetu ir dviem partneriais iš Europos Sąjungos valstybių: Kopenhagos ir Maastrichto universitetais [25]. Rašant vadovėlį buvo remtasi ilgamete mokslinių tyrimų optimizacijos srityje ir panašių disciplinų dėstymo partimi. Anksčiau publikuotos knygos [22], [23], nors formaliai ir neturėjo vadovėlių statuso, bet buvo vartojamos įvairius kursus klausiusių studentų. Pasiteisinusi dėstymo požiūriu medžiaga iš [22], [23] buvo perdirbta ir įtraukta į šį vadovėlį. Kadangi absoliuti šį kursą klausančiųjų studentų dauguma yra aktyvūs Interneto vartotojai, dalis kurso medžiagos yra patalpinta tinklalapyje, skirtame aukščiau minėtam projektui [25].

Šioje knygoje išdėstyta tik teorinė **Matematinio programavimo** kurso medžiaga. Tikimės, kad siūlomos savarankiško darbo užduotys ir kontroliniai klausimai padės geriau įsisavinti šios disciplinos teoriją. Tačiau Matematinio programavimo kursas neapsiriboja vien teorijos studijomis. Svarbu praktiškai išbandyti įvairius metodus, palyginti jų efektyvumą sprendžiant testinius uždavinius. Šio kurso praktikumui skirta medžiaga pateikta tinklalapyje [25]. Autoriaus nuomone, medžiagos padalinimas į teorinę dalį, atspausdintą tradiciniu būdu, ir praktinę Internetinę dalį, kuri turi būti įsisavinama dirbant kompiuteriu, atitinka racionalią studijų metodiką. Pastarosios paruošimui didelės įtakos turėjo pažintis su Maastrichto universitete plačiai taikoma Problem Based Learning metodika [19] ir diskusijos su šios metodikos entuziastu D.Tempellar'u. Tinklalapyje pateikti praktinių darbų aprašymai, kuriuos paruošė Dr. M.Tamošiūnaitė, sudėtingesnių situacijų iliustracijos, mūsų studentų paruošti kai kurių uždavinių sprendiniai, nuorodos į reikalingą programinę įrangą. Tinklalapyje taip pat patalpinta medžiaga, skirta gilesniam kai kurių temų nagrinėjimui, pavyzdžiui, projekto partnerio profesoriaus J.Clausen anglų kalba paruoštos paskaitos [6], [5].

Autoriaus pažiūroms į Matematinio Programavimo discipliną didelės įtakos turėjo ilgametis bendravimas su kolegomis iš Matematikos ir Informatikos Instituto J.Mockumi, V.Šalteniu, G.Dzemyda, V.Tiešiu. Atrenkant medžiagą šiam kursui buvo nagrinėti panašūs kituose universitetuose skaitomi kursai, ypač daug diskutuota su kolegomis iš projekto partnerių Kopenhagos ir Maastrichto universitetų, o taip pat iš Danijos Technikos, Dortmundo, Aacheno Technikos, Abo Akademi universitetų, kriuose buvo skaityti panašūs kursai einant vizituojančio profesoriaus pareigas ir/arba vykdyti moksliniai tyrimai artima šiam kursui tematika. Sunku būtų išvardinti visus kolegas, su kuriais vykusias diskusijas prisimindavau rašydamas šį vadovėlį, todėl norėčiau padėkoti bent jau tiems, su kuriais buvo diskutuotas konkrečiai šis projektas arba kurių įtaką jaučiau ryškiausiai, t.y. profesoriams J.Clausen, A.Kolen, K.Madsen, R.Mathar, H.Peters, H.-P.Schwefel, A.Törn. Pirmą vadovėlio versiją [24] išleido Vytauto Didžiojo Universiteto leidykla 1999 metais. Šis versija gerokai papildyta, ištaisytos redagavimo klaidos. Dėkoju recenzentui Doc. dr. A. Apiniui už vertingas patabas.

Iškilus terminologiniams klausimams buvo vadovaujamasi [12].

1.2 Uždavinio formulavimas

Matematinis uždavinio formulavimas labai svarbus optimizavimo etapas, nuo kurio priklauso ir jo išsprendžiamumas, ir gautų rezultatų adekvatumas praktiniam uždaviniui, kurį jis formalizuoja. Jei norima praktinį uždavinį spręsti kompiuteriu, reikia sudaryti uždavinio matematinį modelį ir jį atitinkančią kompiuterio programą. Optimizavimui skirtą kompiuterio programą galima pasirinkti iš daugelio komercinių ar laisvai platinamų optimizavimo paketų. Matematinio modelio sudarymas priklauso taikomosios srities specialistų kompetencijai. Nors pastarieji optimizavimo metodų paprastai nekuria, bet ir jiems būtinos įvadinės teorijos žinios, šiokia tokia metodų taikymo patirtis. Kadangi sudarant modelį beveik visada tenka rinktis kompromisą tarp modelio sudėtingumo ir adekvatumo, svarbu žinoti metodų, kuriais bus sprendžiamas uždavinys, galimybes ir bendras savybes. Sėkmingą uždavinio formulavimo ir jo sprendimo metodo suderinimą iliustruoja legenda apie Kartagenos įkūrimą.

Karalaitei Didonai buvo padovanota tiek žemės, kiek galima uždengti jaučio oda. Daug kas net nebandytų tokiame plote ir lūšnelės statytis, nekalbant jau apie miesto įkūrimą. Bet Didona rado genialų sprendinį. Ji sukarpė jaučio odą siauromis juostelėmis, jas surišo ir gautąja ilga juostele atsitvėrė sau teritoriją. Viena iš savo teritorijos ribų Didona pasirinko Viduržemio jūros pakrantę, kuri sudarė natūralią teritorijos ribą. Kita riba buvo pusapskritimiu išlenkta juostelė. Taip ji "uždengė", t.y. atsitvėrė, maksimalų plotą. Be to, Kartagena gavo išėjimą į jūrą. Didonos sprendinys buvo geras ne tik teritorijos maksimizavimo, bet ir kitu strategiškai svarbiu požiūriu, kuris net neatsispindi pradiniame uždavinio formulavime.

Pradinis uždavinio formulavimas apie teritoriją, kurią galima uždengti jaučio oda, daugeliui žmonių būtų suprantamas vienareikšmiai. Pasirodo, jis gali turėti ir kitokią interpretaciją. Ne tik šiuo mitologiniu atveju, bet ir nūdienos praktikoje labai svarbu, kad **uždavinio formulavimas būtų tikrai adekvatus siekiamiems tikslams**. Karalaitei Didonai rasti lemtingojo uždavinio sprendimo būdą, tikriausiai padėjo dievai. Matematiškai suformuluoti maksimalaus ploto uždavinį esant duotam perimetrui galima ir be matematinės analizės žinių, bet taip vadinamų izoperimetrinių uždavinių sprendimo metodai buvo sukurti kokiais dviem tūkstantmečiais vėliau. Kadangi karališka intuicija duota nedaugeliui, tai kuriamos teorijos ir metodai, padedantys parastiems mirtingiesiems spręsti sudėtingus uždavinius. Tačiau ir esant šiuolaikiniam mokslo lygiui **efektyvūs** sprendimo metodai žinomi toli gražu ne kiekvienam optimizacijos uždaviniui. Todėl formuluojant uždavinį reikia atsižvelgti į numatomų naudoti **optimizavimo metodų efektyvumą.**

Optimizavimo uždaviniais domėjosi, juos tyrinėjo ir metodus jiems kūrė tokie įžymūs mokslininkai kaip Newton, Lagrange, Euler ir daugelis kitų. Optimalumo principai svarbūs mokslo metodologijai. Nemažai gamtos dėsnių formuluojami kaip optimalumo sąlygos. Pavyzdžiui, šviesos spindulio kelias nevienalytėje aplinkoje nusakomas trumpiausiu sklidimo laiku. Optimalumo principais išreiškiami svarbūs ekonomikos dėsniai.

Toliau vartojamose formulėse ir matematiniuose reiškiniuose vektorius, matricas ir aibes žymėsime didžiosiomis raidėmis. Vektorių bei matricų komponentės bus žymimos atitinkamomis mažosiomis raidėmis su apatiniais indeksais. Vektorių ir matricų operacijose vektoriai suprantami kaip vektoriai-eilutės arba vektoriai-stulpeliai priklausomai nuo konteksto. Esant būtinybei skirti vektorius-eilutes ir vektorius-stulpelius, pastarieji bus žymimi su viršutiniu indeksu t, reiškiančiu transponavimą, pavyzdžiui, X - vektorius-eilutė, X^t - vektorius-stulpelis.

Optimizavimu vadinama paieška uždavinio formulavimu apibrėžtoje aibėje tokio elemento, kuriam kriterijaus reikšmė būtų minimali (arba maksimali). Dažnai nagrinėjamoji aibė yra n- matės Euklido erdvės poaibis. Tada galima kalbėti apie (parametrų) **vektoriaus su minimalia kriterijaus reikšme paiešką**. Parametrų vektoriaus $X \in R^n$ komponentės x_i, $i = 1, ..., n$, vadinamos *uždavinio kintamaisiais*, o kriterijų apibrėžianti funkcija $f(X)$ vadinama *tikslo funkcija.* Kintamieji turi patenkinti uždavinio formulavimą atitinkančias sąlygas. Pavyzdžiui, jei kintamasis x_i reiškia kažkokio objekto masę, jis negali būti neigiamas. Jei x_j reiškia temperatūrą Celsijaus skalėje, tai jis neturės prasmės būdamas mažesniu už -273.15. Be abejo, realiame uždavinyje temperatūros kitimo intervalas, apibrėžiamas technologinių uždavinio sąlygų, prasidės gerokai aukščiau absoliutaus nulio ir turės viršutinį rėžį. Kintamieji gali būti susieti lygybėmis ir nelygybėmis, išreikiančiomis gamtos dėsnius (pvz. energijos tvarumo), turimus resursus, standartų reikalavimus ir pan. Erdvės R^n poaibis A, kuriame patenkinamos uždavinyje apibrėžtos sąlygos kintamiesiems, vadinamas *leistinąja sritimi.* Leistinoji sritis paprastai apibrėžiama lygybėmis ir nelygybėmis *ribojimų funkcijų* reikšmėms, pvz.

$$A = \{X \in R^n : \; g_1(X) = 0, \; g_2(X) \geq 0\}.$$

Minimizavimo (analogiškai, maksimizavimo) uždavinys užrašomas

$$\min_{X \in A} f(X),$$

o jo sprendimas reiškia minimalios tikslo funkcijos reikšmės $f_* \leq f(X)$, $X \in A$ ir minimumo taško $X_* \in A$, $f(X_*) = f_*$ suradimą. Kad matematiškai suformuluotas uždavinys turėtų prasmę, reikia garantuoti sprendinio egzistavimą. Tam užtenka minimalių prielaidų, pvz. sprendinys egzistuoja, jei tikslo funkcija tolydinė ir leistinoji aibė aprėžta ir uždara. Sprendinio egzistavimui taip pat pakanka, kad leistinoji aibė būtų baigtinė.

Minimumo taško radimo uždavinys gali būti nekorektiškas, t.y. uždavinio sprendinys gali stipriai pasikeisti duomenims pasikeitus visai nežymiai. Pavyzdžiui, uždavinio

$$\min f(x), \; f(x) = \beta \cdot x, \; -10^6 \leq x \leq 10^6, \tag{1.1}$$

minimumo taškas yra $x_* = -10^6$, kai $\beta > 0$, ir $x_* = 10^6$, kai $\beta < 0$. Taigi, pakitus uždavinio parametrui β labai nedaug, sakysim, nuo -10^{-6} iki 10^{-6}, uždavinio sprendinys keičiasi maksimaliai galimu dydžiu $2 \cdot 10^6$.

Praktiškai priimtina ir teoriškai korektiška būtų tokia uždavinio formuluotė: rasti tikslo funkcijos reikšmę f_ε ir ją atitinkantį leistinosios aibės tašką, kuriame tikslo funkcijos reikšmė skiriasi nuo minimalios ne daugiau negu leistina tolerancija ε. Šiame vadovėlyje daroma prielaida, kad nagrinėjamieji minimizavimo uždaviniai yra korektiški. Tada minimumo ir minimumo taško suradimo uždaviniai yra ekvivalentūs.

Kaip pasinaudoti optimizavimo teorija, kai kokiame nors praktiniame uždavinyje iškyla optimizavimo būtinybė? Norint suformuluoti uždavinį matematiškai, reikia sudaryti taikomojo uždavinio *matematinį modelį:*

PIRMA, reikia išsiaškinti, kokį kriterijų tikslinga optimizauoti, kokie uždavinio parametrai gali būti keičiami ir kokioje srityje jie gali kisti, t.y. apibrėžti kintamųjų vektorių X, tikslo funkciją $f(X)$ ir leistinąją sritį A. Pasinaudojus optimizavimo metodu, bus gautas optimalių parametrų vektorius X_*, kuriuos ir reikės realizuoti praktiškai įgyvendinant uždavinio sprendinį.

ANTRA, reikia turėti pakankamai žinių apie taikomąją sritį, kad būtų galima sudaryti algoritmą tikslo funkcijos $f(X)$ ir ribojimus apibrėžiančių funkcijų $g_i(X)$ reikšmėms (daugeliu atveju ir išvestinėms) apskaičiuoti.

Remiantis matematinio modelio savybėmis ir optimizacijos teorijos rezultatais pasirenkamas tinkamiausias metodas suformuluotam optimizavimo uždaviniui spręsti.

1.3 Pavyzdžiai

Pradėkime nuo labai paprasto uždavinio. Kokia turėtų būti stačiakampio gretasienio formos dėžė, kad vienetiniam paviršiaus plotui jos tūris būtų maksimalus? Pažymėję dėžės ilgį, plotį ir aukštį x_1, x_2, x_3, tikslo funkciją galėsime užrašyti formule

$$f(X) = x_1 \, x_2 \, x_3,$$

ir leistinąją sritį trimis nelygybėmis ir viena lygybe

$$A = \{X \in R^3 : \, x_i \geq 0, \;\; 2x_1x_2 + 2\, x_1 \, x_3 + 2x_3 \, x_2 = 1\}.$$

Matematinis modelis čia labai paprastas: vos kelios analitinės formulės. Uždavinys tikrai turi sprendinį, nes tikslo funkcija yra tolydi, o leistinoji sritis aprėžta ir uždara. Uždavinio kintamųjų prasmė akivaizdi: gaminamos dėžės matavimai. Ar išsprendus uždavinį gautas rezultatas bus naudingas praktiniams taikymams? Taip, jei medžiagos taupymas yra pagrindinis tikslas: su duotu medžiagos, pvz. faneros, kiekiu galėtume pagaminti maksimalaus tūrio dėžę. Bet šis uždavinio formulavimas ignoruoja daugelį aplinkybių, kurios gali būti svarbios praktikoje. Pavyzdžiui, dažnai reikia ne vienos, o daugelio dėžių, fanerą reikia pjaustyti iš tam tikros formos lapų ir minimizuoti pjaustymo atliekas, dėžes reikia vežti tam tikro dydžio koteineriuose ir pan. Atsižvelgiant į konkretaus taikymo aplinkybes, tikriausiai reikėtų tikslinti matematinį modelį: atsižvelgti į daugiau ribojimų, ar net keisti tikslo funkciją. Daugelyje praktinių optimizavimo uždavinių adekvataus modelio sudarymas yra gan ilgas iteracinis procesas, kuriame dalyvauja suinteresuotų organizacijų specialistai ir vadybininkai [23].

Įvairiose technikos srityse galima rasti daug projektavimo pavyzdžių, kurių optimizacinis formulavimas laikytinas standartiniu. Vienas iš tokių pavyzdžių yra radioelektroninių filtrų projektavimas: duota dažnuminė charaktersitika $F_e(\varpi)$, $\varpi_- \leq \varpi \leq \varpi_*$, kurią reikia realizuoti pasirinktos struktūros R, L, C filtru. Pažymėję atitinkamų elementų elektrines charakteristikas R_i, L_i, C_i, ir sudarę elektrinės grandinės dažnuminės charakteristikos $F(\varpi, R_i, L_i, C_i)$ apskaičiavimo algoritmą, galime formuluoti optimizavimo uždavinį

$$\min_{R_i,L_i,C_i} \max_{\varpi_- \leq \varpi \leq \varpi_*} |F_e(\varpi) - F(\varpi, R_i, L_i, C_i)|,$$

apibrėžiant leistinąją sritį iš projekto specifikacijų išplaukiančiais reikalavimais projektuojamam filtrui. Kintamųjų reikšmės, surastos sprendžiant optimizavimo uždavinį, apibrėžia optimalaus filtro elementų elektrines charaktersitikas. Uždavinys būtų gerokai sudėtingesnis, jei filtro struktūra būtų iš anksto neapibrėžta, ir reikėtų ieškoti ne tik optimalių elementų charakteristikų, bet ir optimalios filtro struktūros.

Projektuojant įvairias statybines konstrukcijas svarbu ne tik užtikrinti pakankamą jų stiprumą (pvz., atsparumą gniuždymui, lenkimui ir pan.), bet ir minimizuoti medžiagų, ypač brangiųjų, sunaudojimą. Jei konstrukcijos kaina yra formuluojama kaip optimizavimo uždavinio tikslo funkcija, tai jos apskaičiavimo algoritmas yra paprastas. Tačiau gali būti problemų su duomenų neapibrėžtumu. Gal būt, iki bus pradėta gamyba pasikeis situacija statybinių medžiagų rinkoje, ir naudotinų medžiagų kainos bus nebe tokios, kokios yra projektavimo metu. Gali atsirasti naujos technologijos, ir rinkoje pasirodys medžiagos, kurios projektavimo metu apskritai nebuvo gaminamos. Uždavinio ribojimų formulavimui reikia žinoti konstrukcinių medžiagų atsparumo apskaičiavimo metodus. Nors algoritminiu požiūriu ši matematinio modelio dalis yra sudėtingesnė, bet ji gan stabili ir universali. Kartą sukurti, konstrukcijų atsparumo skaičiavimo algoritmai (ir juos realizuojantys programų paketai) gali būti naudingi įvairiems uždaviniams spręsti. Be minėtųjų aplinkybių gali tekti atsižvelgti į įvairius gamybos faktorius, nes kokia nauda iš labai geros konstrukcijos, kurios realiai nebus įmanoma pagaminti.

Pastarasis pavyzdys rodo, kad praktinio optimizavimo uždavinio matematinio modelio sudarymas ir suformuluoto matematinio optimizavimo uždavinio sprendimas gali būti sudėtingas, aukštos kvalifikacijos reikalaujantis uždavinys. Jo sprendimas gali pareikalauti iš gamintojo tam tikrų išlaidų. Tačiau įsotintos rinkos sąlygomis optimizuoti ir pačius gaminius, ir gamybos procesą, ir firmų vadybą tiesiog būtina, nes konkuruoti tenka su tais, kurie optimizuoja.

1.4 Uždavinių klasifikacija

Optimizavimo uždaviniai klasifikuojami priklausomai nuo tikslo funkcijos ir leistinosios srities savybių. Aukščiau pateiktame optimizacijos uždavinio formulavime buvo ieškoma optimalaus kintamųjų vektoriaus. Tokie uždaviniai vadinami *daugiamatės optimizacijos* uždaviniais, arba optimizacija n-matėje Euklidinėje erdvėje. Svarbus specialus atvejis $n = 1$ vadinamas *vienmate optimizacija*, arba vieno kintamojo funkcijų optimizacija.

Svarbi uždavinių klasė yra neapimta pateikto formulavimo. Tai optimizavimo teorijoje nagrinėjami uždaviniai, kuriuose ieškomasis objektas yra funkcija, pvz., valdymo signalo priklausomybė nuo laiko, užtvankos profilis ir pan. Paminėtas įvade izoperimetrinis uždavinys taip pat priklauso šiai klasei. Bendru atveju funkcija nėra apibrėžiama baigtiniu kintamųjų skaičiumi. Todėl ir uždaviniai formuluojami kaip *optimizacija funkcijų erdvėse*. Svarbus tokių uždavinių poklasis yra optimalaus valdymo uždaviniai. Šiame kurse tokie uždaviniai nenagrinėjami. Tačiau pažymėsime, kad jie dažniausiai sprendžiami pakeičiant pradinį uždavinį, suformuluotą funkcijų erdvėje, pagalbinių uždavinių su baigtiniu kintamųjų skaičiumi seka.

Jei leistinoji sritis baigtinė (atskirais atvejais gali būti begalinė suskaičiuojama), tai uždavinys vadinamas *diskretaus programavimo* (diskrečiosios optimizacijos) uždaviniu. Detaliau klasifikuojant išskiriami uždaviniai su sveikaskaitiniais ir Bulio kintamaisiais, o taip pat kombinatorinis programavimas. Pastaruoju atveju ieškoma optimalios apibrėžtų kintamųjų reikšmių kombinacijos, pvz. kuprinės uždavinyje ieškoma tokios dvimačių vektorių iš duotosios baigtinės aibės kombinacijos, kad jų pirmųjų komponenčių (sukraunamų į kuprinę daiktų kaina) suma būtų maksimali, esant antrųjų komponenčių (daiktų tūrio arba masės) sumos apribojimui.

Jei kintamieji nėra diskretūs, tai bendru atveju uždavinys vadinamas *netiesinio programavimo* uždaviniu, specialiai nepabrėžiant, kad leistinoji sritis yra kontinuali. Kalbant apie netiesinio programavimo uždavinius, paprastai, turima galvoje lokalaus minimumo paieška. Kai norima rasti globalų minimumą, kalbama apie *globalią optimizaciją.* Jei tikslo funkcija ir leistinoji aibė yra iškilos, uždavinys vadinamas *iškilaus programavimo* uždaviniu. Jei tikslo funkcija tiesinė, o leistinoji sritis apibrėžta tiesinėmis lygybėmis ir nelygybėmis, tai uždavinys vadinamas *tiesinio programavimo* uždaviniu.

Kai kuriose taikymo srityse, ypač ekonomikoje, tenka optimizuoti neapibrėžtumo sąlygomis. Tokiu atveju uždavinio matematiniame modelyje turi būti įvertinti neapibrėžtumai. Formuluojant tikimybių teorijos terminais gaunami *stochastinio programavimo* uždaviniai.

Šioje knygoje apie juos bus tik trumpai užsiminta.

Kitas svarbus optimizacijos uždavinio apibendrinimas, kuriam skirsime nemažai dėmesio, susijęs su būtinumu optimizuojant atsižvelgti į daugelį kriterijų. Tokius uždavinius nagrinėja *daugiakriterės optimizacijos* teorija.

1.5 Metodų efektyvumas

Tik labai specialiais atvejais optimizavimo uždavinį pasiseka išspręsti analitiškai, t.y. formule užrašyti sprendinį. Ne visada paprasta net patikrinti, ar duotasis taškas $X \in A$ yra nagrinėjamojo uždavinio minimumo tašku, ar ne. Minimumo Apibrėžtis bendru atveju nelabai kuo gali padėti, nes kontinualios leistinosios srities atveju neįmanoma palyginti nagrinėjamos tikslo funkcijos reikšmės su visomis kitomis. Kartais potencialių sprendinių aibė yra baigtinė, nors leistinoji sritis ir kontinuali. Taip yra, pavyzdžiui, tiesinio programavimo uždaviniuose, kuriuose leistinoji aibė yra daugiamatis briaunainis (poliedras), t.y. kontinuali. Tačiau nesunku įsitikinti, kad potencialių sprendinių aibė sutampa su poliedro viršūnių aibe ir todėl yra baigtinė. Surasti optimalų sprendinį baigtinėje potencialių sprendinių aibėje galima atlikus pilną aibės peržiūrą, t.y. baigtiniu žingsnių skaičiumi, nors patarasis gali būti ir labai didelis.

Kai optimizavimo uždavinys kokiu nors metodu išsprendžiamas baigtiniu žingsnių skaičiumi N, metodo efektyvumas gali būti vertinamas *algoritmų sudėtingumo teorijos* kriterijais. Įvertinama žingsnių skaičiaus N priklausomybė nuo n ir kitų uždavinio parametrų, ir šios priklausomybės tipas apsprendžia metodo sudėtingumo klasę. Kuo greičiau auga N , tuo sudėtingesnis metodas, arba tuo menkesnis jo efektyvumas.

Bendru atveju, baigtiniu žingsnių skaičiumi neįmanoma garantuoti sprendinio suradimo, todėl apsiribojama *apytiksliu uždavinio* *sprendimu*, kai apytiksliu sprendiniu vadinamas taškas apytiksliai patenkinantis (pakankamas arba būtinas) minimumo sąlygas. Uždavinys sprendžiamas iteratyviai: pradedama kokiu nors tašku X_0 ir konstruojama seka X_i, kurios nariai su indeksais $i \geq N$ apytiksliai patenkina minimumo sąlygas.

Jei baigtiniu žingsnių skaičiumi uždavinys neišsprendžiamas, reikalaujama, kad metodas bent jau konverguotų, t.y., kad metodo generuojamos sekos ribiniai taškai sutaptų su uždavinio sprendiniais. Bus paprasčiau įrodinėti matematinius teiginius, jei darysime prielaidą, kad minimumo taškas X_* lokaliai vienintelis, t.y. tam tikroje jo aplinkoje nėra kitų minimumo taškų. Kartais darysime dar stipresnę prielaidą, kad egzistuoja *vienintelis* lokalaus minimumo taškas, kuris yra ir globalaus minimumo tašku. Pažymėję paklaidą i-tajame žingsnyje

$$\begin{aligned} \varepsilon_i &= ||X_i - X_*||, \\ \epsilon_i &= f(X_i) - f(X_*), \end{aligned}$$

metodą vadinsime konverguojančiu į minimumo tašką, jei $\varepsilon_i \to 0$, ir konverguojančiu į minimumą, jei $\epsilon_i \to 0$. Jei paklaidos riba nebūtų lygi nuliui, tai vargu ar turėtume pakankamą pagrindą vadinti seką X_i generuojantį metodą minimizavimo metodu. Kita vertus, konvergavimas yra asimptotinė sekos savybė. Priminsime, kad sekos ε_i konvergavimas tereiškia tokio N_ε egzistavimą kiekvienam $\varepsilon > 0$, kad $\varepsilon_i \leq \varepsilon$, jei $i \geq N_\varepsilon$. Konvergavimo faktas nieko nesako apie N_ε dydį. Gal būt, N_ε toks didelis, kad per priimtiną laiką paklaida nesumažės iki norimo dydžio. Tada metodo konvergavimas neturės praktinės reikšmės.

Konverguojantys metodai lyginami jų konvergavimo greičio požiūriu. Jei pirmo metodo paklaida $\widetilde{\varepsilon_i}$, o antro $\widehat{\varepsilon_i}$, tai pirmas metodas konverguoja greičiau už antrą, kai

$$\lim_{i\to\infty} \frac{\widetilde{\varepsilon_i}}{\widehat{\varepsilon_i}} = 0.$$

Lyginant metodų paklaidas su standartinėmis sekomis, pvz. geometrine progresija, apibrėžiamos metodų klasės. Toliau bus apibrėžtos svarbiausios pirmo ir antro (konvergavimo) laipsnio metodų klasės. Kaip konvergavimo faktas, taip ir konvergavimo greitis yra asimptotinė savybė, kuri gali pasireikšti prie tokių didelių i, kai tas nebeturi praktinės reikšmės. Dėl to, kad konvergavimo tyrimų rezultatai yra asimptotinio pobūdžio, kai kurie taikymais besidominantys specialistai juos apskritai ignoruoja. Jų vertinimus lemia eksperimentai su testiniais uždaviniais bei anksčiau spręstų praktinių uždavinių rezultatai. Tačiau eksperimentinių tyrimų rezultatus sunku vienreikšmiškai interpretuoti ir apibendrinti. Gan dažnai eksperimento sąlygos būna nepilnai aprašytos. Nereti atvejai, kai vieno tyrinėtojo (pavyzdžiui, metodo autoriaus) rezultatai nepasitvirtina kitų tyrinėtojų eksperimentuose.

Kita vertus, kai kurių knygų, pvz. [4], [26], pagrindinis objektas yra grynai matematiniai konvergavimo aspektai. Neneigiant tokių tyrinėjimų matematinės vertės, tenka pastebėti, kad kartais teoremų sąlygos yra praktiškai nepatikrinamos arba labai retai pasitaikančios. Tokių atvejų nagrinėjimas vargu ar sudomins praktinių uždavinių sprendimu besidominančius optimizavimo metodų taikytojus. Tačiau, norint suprasti, prie kokių sąlygų vartotinas vienas ar kitas metodas, kaip parinkti metodo parametrus, būtina išnagrinėti pagrindines konvergavimo teoremas. Suvokti konvergavimo priežastis svarbu ne vien tam, kad būtų lengviau patikrinti konvergavimo sąlygas praktiniuose uždaviniuose. Tik jas suvokdamas metodo vartotojas gali išsiaiškinti, kodėl lėtai gerėja pasiektas minimumo įvertis, ką daryti, jei skaičiavimai sustoja nepasiekus tikėtos tikslo funkcijos reikšmės. Šioje knygoje pateikiama nedaug konvergavimo teoremų, manytume minimumas, be kurio optimizavimo metodų supratimas būtų pernelyg paviršutiniškas.

Gana dažnai konvergavimo terminas vartojamas netiksliai, pavyzdžiui, greitas funkcijos reikšmių mažėjimas minimizavimo proceso pradžioje vadinamas greitu konvergavimu. Gali būti, kad po kelių sėkmingų žingsnių tolimesnis progresas bus lėtas, o teorinis konvergavimo greitis mažesnis negu kitų metodų. Greitam grubiam priartėjimui prie sprendinio gali tikti toks metodas, kurio asimptotinis konvergavimas lėtas. Kita vertus, žinomi metodai, kurių asimptotinis konvergavimo greitis iš gero pradinio taško yra labai didelis, bet kurie gali apskritai nekonverguoti, jei pradinis taškas parinktas toli nuo minimumo taško. Tik tokie metodai, kurie pasižymi ir dideliu teoriniu konvergavimo greičiu ir praktiniams uždaviniams svarbiu greitu tikslo funkcijos mažėjimu pradinėje optimizavimo stadijoje, pripažįstami tinkamais plačiam vartotojų ratui, įtraukiami į komercinius optimizavimo paketus.

1.6 Iškilosios aibės ir funkcijos

Iškilumo savybė labai svarbi optimizavimo teorijoje. Aibė vadinama iškila, jei kartu su bet kuriais dviem aibės taškais tai aibei priklauso ir juos jungianti tiesės atkarpa. Formaliai tą galima užrašyti taip:

Apibrėžtis. $A \subset R^n$ yra iškila aibė, jei $\lambda X + (1-\lambda)Y \in A$ visiems $X, Y \in A$, ir $0 \leq \lambda \leq 1$.

Galima įrodyti, kad

∘ baigtinio skaičiaus iškilųjų aibių sankirta yra iškiloji aibė,

∘ jei C_1 ir C_2 iškilosios aibės, tai $\{X_1 + X_2 |\ X_1 \in C_1,\ X_2 \in C_2\}$ yra iškiloji aibė,

∘ tiesine transformacija gautos iškilosios aibės atvaizdas yra iškiloji aibė,

∘ jei dvi iškilosios aibės A, B, nesikerta, tai egzistuoja jas atskirianti hiperplokštuma, t.y. egzistuoja nenulinis vektorius C, toks kad $CX \geq c_0$, $X \in A$ ir $CX < c_0$, $X \in B$.

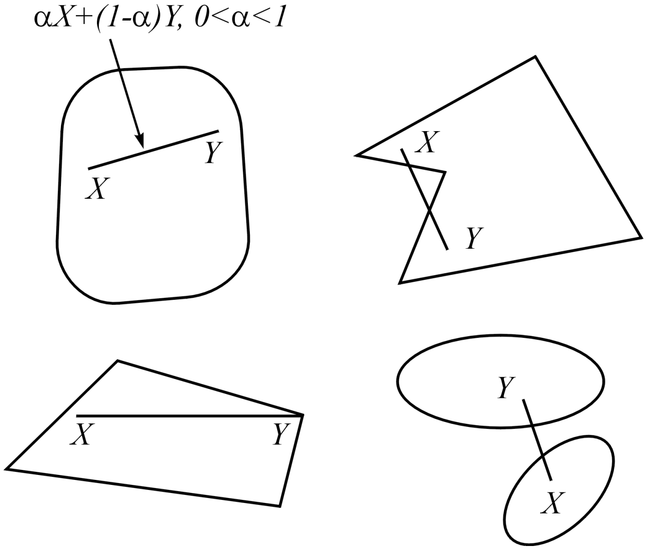

Figure 1.1: Iškilų ir neiškilų aibių pavyzdžiai

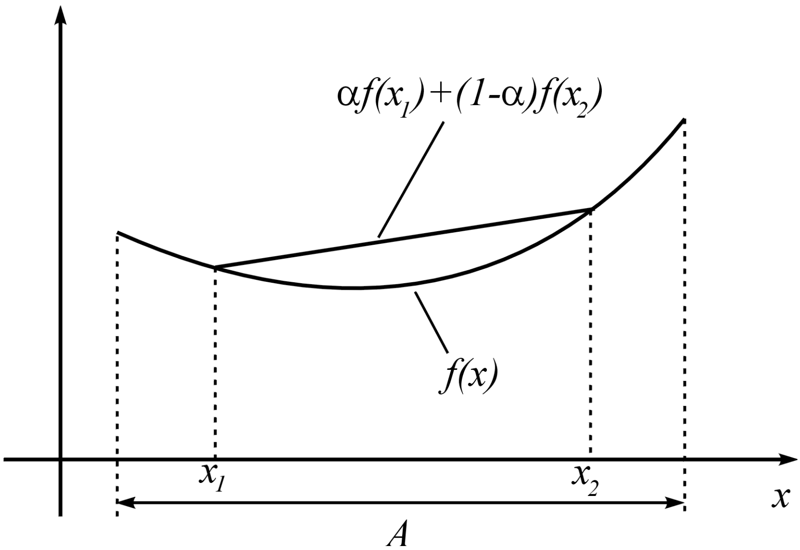

Figure 1.2: Vieno kintamojo iškilosios funkcijos pavyzdys

Apibrėžtis. Funkcija $f(X)$ apibrėžta iškilojoje aibėje A vadinama iškiląja, jei

$$f(\lambda X_1 + (1-\lambda)X_2) \leq \lambda f(X_1) + (1-\lambda)f(X_2), \qquad (1.2)$$

visiems $X_1, X_2 \in A$, ir $0 \leq \lambda \leq 1$.

Jei apibrėžime būtų parašytas griežtos nelygybės ženklas, tai funkcija būtų vadinama griežtai iškila.

Galima įrodyti, kad

∘ jei $f(X)$ iškiloji funkcija, tai jos, taip vadinamos, *lygio aibės* $\{X : f(X) \leq \alpha\}$, $\{X : f(X) < \alpha\}$ yra iškilos,

∘ tiesinė funkcija yra iškila,

∘ iškilųjų funcijų su teigiamais daugikliais suma yra iškiloji funkcija,

∘ jei $f_i(X)$, $X \in A, i \in I$, iškilosios funkcijos, tai $f(X) = \min_{i \in I} f_i(X)$ iškila,

∘ jei $f(X)$ iškila ir diferencijuojama, tai galioja nelygybė

$$f(X) \geq f(Y) + \nabla f(Y) \cdot (X - Y), \tag{1.3}$$

∘ jei $f(X)$ iškila ir du kartus diferencijuojama, tai jos antrųjų išvestinių matrica yra neneigiamai apibrėžta,

∘ iškilos funkcijos apibrėžtos iškiloje aibėje lokalus minimumas yra globalus; jei funkcija griežtai iškila, tai globalaus minimumo taškas vienintelis.

Įrodysime (1.3). Jei $f(X)$ diferencijuojama, tai diferencijuojama ir vieno kintamojo funkcija $\varphi(\lambda) = f(\lambda X + (1 - \lambda)Y)$ (žiūr. 3 užduotį 28 puslapyje). Nelygybė (1.2) reiškia, kad taške $\lambda = 0$, funkcijos $\varphi(\lambda)$ išvestinė nedidesnė už funkcijos $\lambda f(X) + (1 - \lambda) f(Y)$ išvestinę, t.y.

$$\varphi^{'}(0) = \nabla f(Y) \cdot (X - Y) \leq f(X) - f(Y). \tag{1.4}$$

Perkėlę $f(X)$ į kairiąją (1.4) nelygybės pusę, gausime (1.3).

Iš nelygybės (1.3) gauname svarbią optimizacijos teorijai išvadą. Jei kokiame nors taške Y iškilos diferencijuojamos funkcijos $f(X)$ gradientas lygus nuliui, tai (1.3) gali būti užrašyta šitaip

$$f(X) \geq f(Y),\ X \in A,$$

ir tas reiškia, kad Y yra funkcijos $f(\cdot)$ globalaus minimumo taškas.

Jei apibrėžime (1.2) nelygybės ženklą pakeistume į priešingą, tai gautume **įgaubtosios funkcijos** apibrėžimą.

Afininė funkcija

$$f(X) = a_0 + a_1 x_1 + ... + a_n x_n$$

yra ir iškila ir įgaubta.

Toliau bus sutinkamos funkcijos, kurių reikšmės vektorinės: funkcijos reikšmių vektorių - stulpelį sudaro funkcijų - komponenčių reikšmės.

$$F(X) = (f_1(X), ..., f_m(X))^t.$$

Jei kiekviena $F(X)$ komponentė $f_i(X)$ iškila (įgaubta) aukščiau duoto apibrėžimo prasme, tai ir funkcija $F(X)$ yra iškila (įgaubta). Afininė funkcija su vektorinėmis reikšmėmis užrašoma šitokia formule $F(X) = A \cdot X - B$, čia A yra $m \times n$ matrica.

Apibrėžtis. Funkcija $f(X)$, apibrėžta iškilojoje aibėje A, vadinama stipriai iškila su konstanta l, jei

$$f(\lambda X + (1-\lambda)Y) \leq$$
$$\lambda f(X) + (1-\lambda) f(Y) - \frac{1}{2} l\, \lambda (1-\lambda) ||X - Y||^2,$$

visiems $X, Y \in A$, ir $0 \leq \lambda \leq 1$.

Stipriai iškilai funkcijai nelygybė (1.3) gali būti pakeista nelygybe

$$f(X) \geq f(Y) + \nabla f(Y) \cdot (X - Y) + \frac{1}{2} l\ ||X - Y||^2.$$

1.7 Lipshitz'o sąlyga

Kuriant ir analizuojant optimizavimo metodus naudojami įvairūs optimizavimo uždavinių matematiniai modeliai. Pastarieji aprašomi funkcijų, turinčių norimas savybes, klasėmis. Optimizacijoje, kaip ir kitose disciplinose, plačiai vartojamos tiesinių bei kvadratinių funkcijų klasės. Kitose disciplinose gal rečiau vartojamos, bet optimizacijai būdingos iškilų funkcijų bei Lipshitz'o funkcijų klasės. Pirmąją ką tik apibrėžėme. Antrąją apibrėšime vieno kintamojo funkcijų atveju. Funkcijų, patenkinančių nelygybę

$$|f(x) - f(y)| \leq L\,|x - y|,\ L > 0, \tag{1.5}$$

klasė vadinama Lipshitz'o funkcijų klase; čia L - Lipshitz'o konstanta. (1.5) nelygybė aprėžia funkcijos reikšmių kitimo greitį. Todėl gretimuose taškuose funkcijos reikšmės negali smarkiai skirtis. Žinodami bent vieną funkcijos reikšmę, galime apskaičiuoti rėžius jos reikšmėms kituose taškuose, pvz.

$$f(x) - L\,|x - y| \leq f(y) \leq f(x) + L\,|x - y|.$$

Lipshitz'o funkcijos išvestinės absoliutiniu didumu ne didesnės už L. (1.5) nelygybė lengvai apibendrinama funkcijoms su vektoriniais argumentais bei reikšmėmis pakeičiant absoliutinio didumo ženklą normos ženklu.

1.8 Užduotys ir kontroliniai klausimai

1. Pakartokite (išnagrinėkite) matematinės analizės skyrių apie daugelio kintamųjų funkcijų diferencijavimą bei skleidimą Teiloro eilute. Išsiaiškinkite gradiento ir antrųjų išvestinių matricos savybes.

2. Pakartokite (išnagrinėkite) tiesinės algebros skyrių apie veiksmus su vektoriais ir matricomis. Išsiaiškinkite vektorių tiesinio nepriklausomumo sąvoką, matricos tikrinių reikšmių bei tikrinių vektorių prasmę, vektoriaus bei matricos normos prasmę.

3. Išreikškite funkcijos $f(X + tY)$ išvestinę pagal t per funkcijos $f(\cdot)$ gradientą.

4. Suformuluokite optimizavimo uždavinį iš savo klausytų kursų tematikos arba iš savo praktikos.

5. Remdamiesi turimomis matematikos žiniomis išspręskite dėžės tūrio optimizavimo uždavinį, suformuluotą 16 puslapyje.

6. Įrodykite pirmąsias dvi iškiliųjų aibių savybes (23 psl.) ir pirmąsias tris iškiliųjų funkcijų savybes (25 psl.).

OPTIMIZACIJA BE RIBOJIMŲ

2.1 Būtinos ir pakankamos minimumo sąlygos

Sprendžiamas optimizacijos uždavinys, kurio kintamiesiems neformuluojami jokie ribojimai, t.y. leistinoji sritis sutampa su visa n-mate Euklido erdve R^n

$$\min_{X \in R^n} f(X), \ \ X = (x_1, ..., x_n).$$

Toks uždavinys gali pasirodyti nelabai adekvatus realiems taikymams, nes bet kokiame praktiniame uždavinyje egzistuoja tam tikros kintamųjų ribos, išplaukiančios iš kintamųjų prasmės. Pavyzdžiui, masė negali būti neigiama, temperatūra negali išeiti iš tam tikrų ribų ir pan. Peržengus tas ribas, abstrakti uždavinio formuluotė tampa nebeadekvačia realiems uždaviniams nors matematiškai ir lieka korektiška. Visdėlto, tikslo funkcija kartais gali būti sudaryta taip, kad minimumas būtinai pasiekiamas nors ir *a priori* nežinomame, bet uždavinio sąlygas atitinkančiame taške. Tada nėra prasmės formuluoti ribojimus minimumo paieškai. Metodai uždaviniams be ribojimų ypač svarbūs kaip pagalbiniai, nes daugelis metodų uždaviniams su ribojimais sudaromi pakeičiant pradinį uždavinį seka uždavinių be ribojimų.

Pagal minimumo apibrėžimą patikrinti, ar taškas X yra minimumo tašku, įmanoma tik tada, kai leistinoji sritis A baigtinė, nes tik tada galima palyginti $f(X)$ su visomis kitomis funkcijos reikšmėmis $f(Y)$, $Y \in A$. Nagrinėjamuoju atveju to padaryti negalima iš principo, nes leitinoji sritis sutampa su R^n . Todėl reikia suformuluoti konstruktyvius (patikrinamus) požymius minimumo taškui atpažinti.

Apibrėžtis. Taškas X_* vadinamas funkcijos $f(X)$ lokalaus minimumo tašku, jei egzistuoja tokia X_* aplinka $S(X_*)$, kad $f(X_*) \leq f(X)$, $X \in S(X_*)$. Taškas X_* vadinamas funkcijos $f(X)$ globalaus minimumo tašku, jei $f(X_*) \leq f(X)$, $X \in R^n$.

Sakysim, kad funkcija $f(X)$ diferencijuojama. Vektorius, sudarytas iš funkcijos dalinių išvestinių, apskaičiuotų taške X vadinamas funkcijos gradientu ir žymimas

$$\nabla f(X) = \left(\frac{\partial f(X)}{\partial x_1}, ..., \frac{\partial f(X)}{\partial x_n} \right).$$

Būtina minimumo sąlyga. Jei X_* yra $f(X)$ lokalaus minimumo taškas, tai

$$\nabla f(X_*) = 0. \tag{2.1}$$

Ši sąlyga įrodoma matematinės analizės kurse remiantis funkcijos Teiloro eilutės pirmos eilės nariais. Trumpai šis įrodymas toks: sakysim taške X_* gardientas nelygus nuliui. Tada, žengus pakankamai trumpą žingsnį antigradiento kryptimi $X = X_* - \gamma \cdot \nabla f(X_*)$,

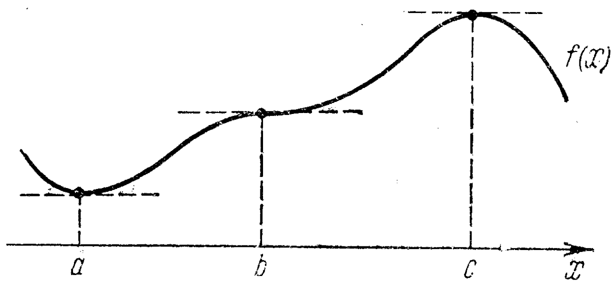

Figure 2.3: Būtina minimumo sąlyga patenkinama taškuose *a, b, c*

$\gamma > 0$, gaunama funkcijos $f(X)$ reikšmė mažesnė už $f(X_*)$:

$$f(X) =$$

$$= f(X_*) + \nabla f(X_*) \cdot (X - X_*) + o(||X - X_*||)$$

$$= f(X_*) - \gamma \left(||\nabla f(X_*)||^2 - \frac{o(\gamma)}{\gamma} ||\nabla f(X_*)|| \right)$$

$$< f(X_*).$$

Reiškia, taškas, kuriame gradientas nelygus nuliui negali būti lokalaus minimumo tašku. Ši sąlyga reiškia, kad funkcijos gradientas lygus nuliui ir globalaus minimumo taške, nes globalusis minimumas yra kartu ir lokalusis. Sąlyga (2.1) yra būtina, bet nepakankama. Ši sąlyga gali būti patenkinama ne tik lokalaus minimumo, bet ir kituose taškuose, pavyzdžiui, lokalaus maksimumo bei balno taškuose. Pavyzdžiui, 2.3 paveiksle būtinas minimumo sąlygas patenkina taškai a, b, c, bet tik taškas a yra lokalaus minimumo taškas.

Pakankamoms lokalaus minimumo sąlygoms suformuluoti pirmųjų tikslo funkcijos išvestinių neužtenka. Sakysim, kad $f(X)$ du kartus diferencijuojama. Jos antrųjų išvestinių (taške X) matricą, vadinamą Hesse matrica, pažymėsime $H(X)$

$$H(X) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} f(X) & \frac{\partial^2}{\partial x_1 \partial x_2} f(X) & \dots & \frac{\partial^2}{\partial x_1 \partial x_n} f(X) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(X) & \frac{\partial^2}{\partial x_2^2} f(X) & \dots & \frac{\partial^2}{\partial x_2 \partial x_n} f(X) \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(X) & \frac{\partial^2}{\partial x_n \partial x_2} f(X) & \dots & \frac{\partial^2}{\partial x_n^2} f(X) \end{pmatrix}.$$

Pakankamos minimumo sąlygos: X_* yra funkcijos $f(X)$ minimumo taškas, jei

1. taške X_* funkcijos gradientas lygus nuliui,
2. $H(X_*)$ teigiamai apibrėžta.

Pakankamos minimumo sąlygos įrodomos panašiai kaip ir būtinos. Remiamasi tikslo funkcijos skleidiniu Teiloro eilute iki antros eilės narių. Sakysim, taške X_* patenkintos pakankamos minimumo sąlygos ir iš jo žengiamas pakankamai trumpas žingsnis bet kuria kryptimi $X = X_* - \gamma \cdot S$. Tada patenkinama nelygybė

$$\begin{aligned} f(X) &= f(X_*) + \nabla f(X_*) \cdot (X - X_*) + \\ &+ \frac{1}{2}(X - X_*) \cdot H(X_*) \cdot (X - X_*) + o(||X - X_*||^2) \\ &= f(X_*) + \gamma^2 \left(\Delta^2 - \frac{o(\gamma^2)}{\gamma^2} ||S||^2 \right) > f(X_*), \end{aligned}$$

reiškianti, kad X_* yra lokalaus minimumo taškas. Geometriškai gradiento lygybės nuliui sąlyga reiškia, kad lokalaus minimumo taške liečiamoji hiperplokštuma ortogonali funkcijos reikšmių ašiai. Hesse funkcijos teigiamas apibrėžtumas reiškia, kad kvadratinė tikslo funkcijos aproksimacija lokalaus minimumo taške yra visomis kryptimis į begalybę augantis hiperparaboloidas. Tokiame taške vieno kintamojo funkcija $f(x)$ aproksimuojama parabole su teigiamu koeficientu

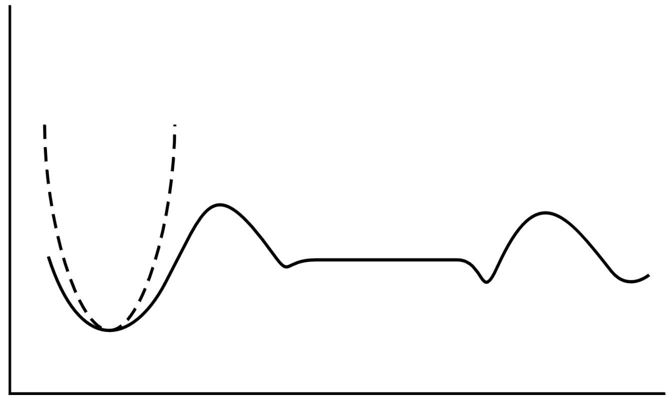

Figure 2.4: Įvairių atvejų, kai patenkinamos būtinos/pakankamos minimumo sąlygos, iliustracija

prie kvadratinio nario, kurios grafikas yra abiem kryptimis auganti parabolė. Situacija, kai funkcijos reikšmės kokiame nors intervale pastovios, vadinama "plato". Čia ir pirmoji ir antroji tikslo funkcijos išvestinės lygios nuliui. Grafiškai šie atvejai iliustruojami 2.4 paveikslu.

Griežtai kalbant, apytikslus minimumo sąlygų patenkinimas taške X dar nereiškia nei X artumo lokalaus minimumo taškui, nei $f(X)$ artumo lokaliam minimumui. Pavyzdžiui, taip gali būti, kai kažkokiame leistinosios srities poaibyje tikslo funkcijos reikšmės beveik pastovios, bet už "plato" poaibio ribos mažėja, ir minimumas pasiekiamas taške gana daug nutolosiame nuo nagrinėjamojo taško. Todėl reikalingas tam tikras atsargumas, jei pagal apytikslų būtinų sąlygų patenkinimą sprendžiama, ar tiriamasis taškas yra lokalaus minimumo taškas.

2.2 Konvergavimas

Optimizavimo uždavinius per baigtinį laiką (iteracijų skaičių) pavyksta išspręsti tik atskirais atvejais, pavyzdžiui, jei sprendžiamas tiesinio programavimo uždavinys. Kitiems atvejams kuriami iteraciniai metodai, generuojantys begalines taškų sekas X_i. Be abejo, siekiama, kad generuojamos begalinės sekos taškai artėtų į sprendinį. Jei vykdant metodą būtų gautas taškas X_j, kuriame patenkinamos pakankamos minimumo sąlygos, be abejo, minimizavimą nutrauktume, nes X_j būtų uždavinio sprendinys. Tačiau tokie atvejai yra retenybė. Bendru atveju tikimasi per priimtiną laiką su priimtinu tikslumu priartėti prie sprendinio. Metodas vadinamas konverguojančiu, jei jo generuojamų taškų sekos ribiniai taškai sutampa su uždavinio sprendiniais. Kad būtų paprasčiau, apsiribokime vieno ribinio taško atveju ir konverguojančia taškų X_i seka

$$\lim_{i\to\infty} ||X_i - X_*|| = 0. \tag{2.2}$$

Konvergavimas yra teoriškai būtina optimizavimo metodo savybė. Bet vien tik konvergavimio faktas nepakankamai charakterizuoja metodą, nes konvergavimas gali būti labai lėtas. Konvergavimo greitis įvertinamas lyginant metodų paklaidas su standartinėmis sekomis. Jei metodo paklaida $\varepsilon_i = ||X_i - X_*||$ artėja į nulį greičiau negu geometrinės progresijos narys, t.y. egzistuoja tokie $c > 0$ ir $\beta \in (0, 1)$, kad

$$\varepsilon_i \leq c \cdot \beta^i,$$

tai sakoma, kad metodas konverguoja *geometrinės progresijos* (tiesiniu) greičiu. Jei santykis $\varepsilon_{i+1}/\varepsilon_i$ artėja į nulį, tai sakoma, kad konvergavimo greitis virštiesinis. Tiksliau virštiesinis greitis apibūdinamas *konvergavimo laipsniu:* jei egzistuoja tokie $c > 0$, $p > 1$, ir $\beta \in (0, 1)$, kad

$$\varepsilon_i \leq c \cdot \beta^{(p^i)},$$

tai sakoma, kad metodas konverguoja virštiesiškai, o konvergavimo laipsnis ne žemesnis negu p. Svarbus atskiras atvejis, kai $p = 2$, vadinamas kvadratiniu konvergavimu.

2.3 Nusileidimo metodų apibūdinimas

Nusileidimo metodai pagrįsti informacija apie pirmąsias ir antrąsias dalines tikslo funkcijos $f(X)$ išvestines. Tokia informacija naudinga krypčiai link lokalaus minimumo parinkti. Kaip žinoma iš matematinės analizės kurso, gradientas nukreiptas funkcijos greičiausio augimo kryptimi. Todėl priešinga – antigradiento kryptis yra reikšminga ieškant minimumo. Pirmosios funkcijos išvestinės (gradientas) apibrėžia tiesinę funkcijos aproksimaciją. Pastaroji gali būti pakankamai tiksli tik tam tikroje, dažnai nedidelėje, taško, kuriame apskaičiuotos išvestinės, aplinkoje. Todėl antigradiento kryptimi daromas žingsnis neturi būti ilgas, nes priešingu atveju bus išeita iš aproksimacijos adekvatumo srities. Teoriškai ir antrųjų išvestinių informacija apie funkcijos kitimą patikima tik taško, kuriame jos apskaičiuotos, aplinkoje. Tačiau praktiškai tokia informacija dažnai leidžia prognozuoti funkcijos kitimo pobūdį didesnėje srityje. Naudojantis antrosiomis išvestinėmis dažnai galima nustatyti kryptį, kuria galima žengti gan ilgą žingsnį link minimumo, t.y. informacija apie antrąsias išvestines leidžia greičiau artėti prie ieškomojo minimumo taško.

Nagrinėdami nusileidimo metodus remsimės prielaida, kad išvestinių reikšmės yra žinomos. Praktiškai jas tenka apskaičiuoti arba pagal tikslias formules, arba kitais būdais. Dažnai išvestinės įvertinamos skaitmeniniu būdu. Pastaruoju metu pasiūlyti automatinio diferencijavimo metodai, kurie pagrįsti tikslo funkcijos apskaičiavimo programos, parašytos tam tikra programavimo kalba, analize.

Nusileidimo metodai yra iteraciniai. Pradedama nuo vartotojo parinkto taško X_0 ir generuojama taškų seka

$$X_{i+1} = X_i + \gamma_i\, S_i\,, \tag{2.3}$$

čia S_i yra krypties vektorius (dažniausiai vienetinio ilgio), o γ_i žingsnio daugiklis (reiškiantis žingsnio ilgį, jei $||S_i|| = 1$). Realizuojant nusileidimo idėją pageidautina, kad kryptis rodytų link minimumo taško, o žengiamas žingsnis maksimaliai priartintų prie minimumo taško. Tačiau, automatiškai tą užtikrinti ne visada parasta. Pavyzdžiui, norint užtikrinti metodo konvergavimą, reikia, kad žingsnio ilgis tam tikru greičiu trumpėtų. Daugelio metodų konvergavimo tyrimas gana sudėtingas. Todėl įrodymai bus pateikti tik keletui atvejų, kuriems nereikia ilgų išvedžiojimų ar specialių matematikos žinių.

Nusileidimo metodai pagrįsti lokalia informacija apie tikslo funkcijos kitimą. Todėl daugiaekstremių tikslo funkcijų atveju negalima garantuoti, kad šiais metodais surastas sprendinys bus globalusis minimumas. Daugiaekstremiu atveju gali būti randami į vairūs lokalūs minimumai, priklausomai nuo to, kokio lokalaus minimumo pritraukimo zonoje parinktas pradinis taškas.

2.4 Gradientinis nusileidimas

Šiuo metodu, matyt, paprasčiausiai realizuojama nusileidimo metodo (2.3) idėja. Nusileidimo krypties vektorius čia sutampa su antigradientu $S_i = -\nabla f(X_i)$, o žingsnio daugiklis γ pastovus. Artėjant prie minimumo taško, šio metodo žingsnio ilgis trumpėja, nes mažėja gradiento norma. Tačiau žingsnio daugiklis negali būti pernelyg didelis, nes vietoj norimo nuoseklaus artėjimo link minimumo taško nuo pat pradžių gali prasidėti nereguliarus šokinėjimas ilgėjančiais šuoliais.

1 teorema. Sakysim, $f(X)$, $X \in R^n$, diferencijuojama ir aprėžta iš apačios

$f(X) \geq f_{-} > -\infty$,
o jos gradientas patenkina Lipshitz'o sąlygą
$||\nabla f(X) - \nabla f(Z)|| \leq L \cdot ||X - Z||$.
Jei žingsnio daugiklis patenkina sąlygą $0 < \gamma < 2/L$, tai gradientinio nusileidimo metodas generuoja taškų seką X_i, kuri monotoniškai mažėja tikslo funkcijos reikšmių prasme, t.y. $f(X_{i+1}) < f(X_i)$, ir konverguoja gradiento prasme, t.y.
$\lim_{i\to\infty} \nabla f(X_i) = 0$.

Įrodymas. Iš matematinės analizės kurso žinoma šitokia lygybė

$$f(X+Y) = f(X) + \int_0^1 (\nabla f(X+tY) \cdot Y)dt, \tag{2.4}$$

kurią galima užrašyti ir taip

$$\begin{aligned} f(X+Y) &= f(X) + (\nabla f(X) \cdot Y) \\ &+ \int_0^1 ((\nabla f(X+tY) - \nabla f(X)) \cdot Y)dt. \end{aligned}$$

Įstatę šioje lygybėje $X = X_i$ ir $Y = -\gamma \cdot \nabla f(X_i)$, gausime

$$f(X_{i+1}) = f(X_i) - \gamma||\nabla f(X_i)||^2 -$$

$$-\gamma \int_0^1 ((\nabla f(X_i - t\gamma\nabla f(X_i)) - \nabla f(X_i)) \cdot \nabla f(X_i))dt \leq$$

$$\leq f(X_i) - \gamma||\nabla f(X_i)||^2 +$$

$$+\gamma \int_0^1 ||(\nabla f(X_i - t\gamma\nabla f(X_i)) - \nabla f(X_i)|| \cdot ||\nabla f(X_i))||dt \leq$$

$$
\begin{aligned}
&\leq \quad f(X_i) - \gamma||\nabla f(X_i)||^2 + L\gamma^2||\nabla f(X_i)||^2 \int_0^1 t\ dt = \\
&= \quad f(X_i) - \gamma \cdot (1 - L\gamma/2) \cdot ||\nabla f(X_i)||^2.
\end{aligned}
$$

Pagal teoremos prielaidas $0<\gamma < 2/L$, todėl $f(X_{i+1}) < f(X_i)$. Susumuokime nelygybes, imdami $i = 0, ..., N$, ir suprastindami $f(X_i)$, įeinančius į sumą kairėje ir dešinėje nelygybės pusėje

$$f(X_{N+1}) \leq f(X_0) - \gamma \cdot (1 - L\gamma/2) \sum_0^N ||\nabla f(X_i)||^2,$$

$$\sum_0^N ||\nabla f(X_i)||^2 \leq \frac{1}{\gamma \cdot (1 - L\gamma/2)} (f(X_0) - f_-).$$

Paskutinioji nelygybė įrodo, kad gradientų normų kvadratų eilutė konverguoja. Būtina konvergavimo sąlyga yra eilutės narių konvergavimas į 0. Todėl

$$\lim_{i\to\infty} \nabla f(X_i) = 0.$$

Sakykim, kad aibė

$$\{X : f(X) \leq f(X_0)\} \tag{2.5}$$

aprėžta. Tada iš taškų X_i sekos galima išrinkti konverguojančią į stacionarų tašką posekę. Esant papildomoms prielaidoms galima įrodyti, kad pati X_i seka konverguoja į stacionarų tašką. Toliau daroma prielaida, kad sąlyga (2.5) yra patenkinama.

Funkcijos reikšmių ir gradiento normos mažėjimas paprastai reiškia konvergavimą į lokalųjį minimumą. Tačiau galima sukonstruoti specialių pavyzdžių, kai žengus žingsnį gradientiniu metodu gaunamas taškas X_{i+1}, kuriame $\nabla f(X_{i+1}) = 0$, bet X_{i+1} yra ne lokalaus minimumo, o lokalaus maksimumo arba balno taškas. Tiksliai pataikyti į tašką, kuriame gradientas lygus nuliui, galima nebent specialiai sukonstruotame pavyzdyje. Tačiau ir praktiškai pasitaikančiose situacijose apytiksli gradiento lygybė nuliui nebūtinai reiškia einamojo taško artumą lokaliojo minimumo taškui. Kitaip sakant, skirtinguose uždaviniuose artumo nuliui mąstai gali labai skirtis.

Šia teorema įrodyta, kad gradientinio nusileidimo metodas konverguoja, bet nieko neteigiama apie konvergavimo greitį. Bendruoju atveju konvergavimo greičio įverčiai nežinomi. Detalesniam tyrimui tenka susiaurinti nagrinėjamų funkcijų klasę. Paprastai daromos prielaidos apie tikslo funkcijos iškilumą. Panašaus pobūdžio yra ir prielaidos apie tikslo funkcijos antrųjų išvestinių matricos tikrines reikšmes, pvz. $0 < \lambda_0 \leq \lambda_i \leq \lambda_*$, čia λ_i, $i = 1, ..., n$, yra matricos $H(X)$ tikrinės reikšmės. Pastaroji prielaida reiškia, kad funkcija $f(X)$ yra ne tik iškili, bet ir visuose taškuose jos kitimo pagreitis negali būti nei mažesnis už λ_0, nei didesnis už λ_*. Tokios vieno kintamojo funkcijos grafikas yra lokaliai "įspraustas" tarp dviejų parabolių su koeficientais $\lambda_0/2$ ir $\lambda_*/2$ prie kvadratinių narių. Kai funkcija yra iškili, jos lokalusis minimumas yra kartu ir globalusis.

2 teorema. Sakysim, $f(X)$ du kartus diferencijuojama, ir jos antrųjų išvestinių matricos $H(X)$ tikrinės reikšmės λ_i, $i = 1, ..., n$, patenkina nelygybę $0 < \lambda_0 \leq \lambda_i \leq \lambda_*$ visiems X. Tada gradientinio nusileidimo metodas su žingsnio daugikliu γ, patenkinančiu nelygybę $0 < \gamma < 2/\lambda_*$, konverguoja geometrinės progresijos greičiu į minimumo tašką X_*

$$||X_i - X_*|| \leq q^i \, ||X_0 - X_*|| \, ,$$

$$q = \max\left(|1 - \gamma\, \lambda_0|, \, |1 - \gamma\, \lambda_*|\right) < 1.$$

Metodas konverguoja greičiausiai, kai $\gamma = \gamma^* = 2/(\lambda_0 + \lambda_*)$, nes tada $q = q^* = \frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0}$ yra minimalus.

Įrodymas. Kadangi $f(X)$ du kartus diferencijuojama, tai jos pirmosios išvestinės diferencijuojamos, ir visoms išvestinėms galioja lygybės panašios į (2.4); užrašę jas vektorine forma, gausime tokią gradiento išraišką

$$\nabla f(X+Y) = \nabla f(X) + \int_0^1 H(X+tY)\cdot Y dt,$$

į kurią įstatę $X = X_*$, $Y = X_i - X_*$, gradientą $\nabla f(X_i)$ išreikšime šitokia formule

$$\begin{aligned} \nabla f(X_i) &= \\ &= \int_0^1 H(X_* + t(X_i - X_*))\cdot (X_i - X_*)dt = \\ &= A_i(X_i - X_*). \end{aligned} \tag{2.6}$$

Remdamiesi (2.6), galime įvertinti

$$\begin{aligned} ||X_{i+1} - X_*|| &= ||X_i - X_* - \gamma\nabla f(X_i)|| = \\ ||(I-\gamma A_i)(X_i - X_*)|| &\leq ||(I-\gamma A_i)|| \cdot ||(X_i - X_*)||. \end{aligned}$$

Įvertindami prielaidas apie matricos $H(X)$ tikrines reikšmes, galime teigti, kad $\lambda_0 I \leq A_i \leq \lambda_* I$, čia $I-$ vienetinė matrica. Kadangi $||(I - \gamma A_i)|| = \max(|1-\gamma\lambda_0|, |1-\gamma\lambda_*|)$, tai pažymėję $q = \max(|1-\gamma\lambda_0|, |1-\gamma\lambda_*|)$, gausime nelygybę

$$||X_{i+1} - X_*|| \leq q||(X_i - X_*)|| \leq q^{i+1}||(X_0 - X_*)||,$$

įrodančią, kad metodas konverguoja geometrinės progresijos, kurios vardiklis lygus q, greičiu. Konvergavimas greičiausias, kai q yra minimalus. Nubraižius funkcijos

$$q = \max(|1-\gamma\lambda_0|, |1-\gamma\lambda_*|) \tag{2.7}$$

grafiką (žiūr. 5 užduotį 77) , nesunku įsitikinti, kad minimali q reikšmė lygi

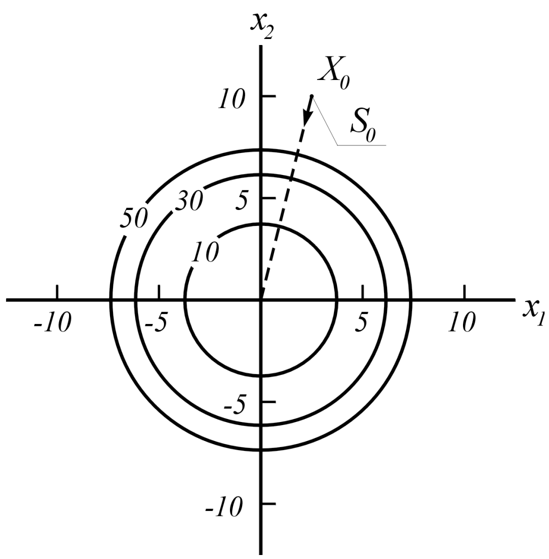

Figure 2.5: Palankus atvejis gradientinio nusileidimo algoritmui

$$q_{\min} = \frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0} \ ,$$

ir minimumas pasiekiamas, kai $\gamma = \gamma^* = 2/(\lambda_0 + \lambda_*)$.

Gautasis konvergavimo greičio įvertis rodo, kad lygybės $\lambda_* = \lambda_0$ atveju tikslus sprendinys gaunamas jau pirmojoje iteracijoje. Šiuo specialiu atveju tikslo funkcija yra sukimosi hiperparaboloidas. Bet kuriame taške tokios funkcijos antigradiento kryptis rodo į minimumo tašką. Parinkę žingsnio daugiklį γ^* taip, kaip rekomenduojama teoremos formulavime, jau pirmuoju žingsniu pateksime tiesiai į minimumo tašką. Dvimatis šios situacijos atvejis iliustruojamas 2.5 paveiksle.

Kita vertus, teoremos rezultatas rodo, kad esant stipriai besiskiriančioms λ_* ir λ_0 reikšmėms, q yra artimas vienetui. Kadangi santykis tarp maksimalios ir minimalios tikrinių reikšmių praktiniuose uždaviniuose būna 10^6 ir daugiau, tai tokiais atvejais gradientinio metodo konvergavimas būtų akivaizdžiai per lėtas.

Kaip bus parodyta toliau, yra metodų, turinčių geresnį teorinį konvergavimo greitį. Praktiniu požiūriu gradientinio nusileidimo metodo taikymas susijęs su žingsnio daugiklio parinkimo sunkumais, nes įvertinti Lipshitz'o konstantą gradientui, o juo labiau rėžius tikrinėms matricos $H(X)$ reikšmėms įmanoma tik labai retais atvejais. Šis metodas paprastas teorinio nagrinėjimo požiūriu, gerai iliustruoja pagrindines nusileidimo metodų savybes. Nors išnagrinėtasis metodas praktiškai beveik nenaudojama, bet apibendrinant gradientinio nusileidimo idėją sudaryti kiti, efektyvesni nusileidimo metodai.

2.5 Greičiausias nusileidimas

Paprasčiausioje gradientinio nusileidimo versijoje žingsnio daugiklis yra pastovus. Tačiau jį parinkti pagal teorines rekomendacijas sunku. Žingsnio ilgio parinkimui buvo sugalvota įvairių euristinių rekomendacijų, pavyzdžiui, žengus sėkmingą žingsnį, pakartoti tokio pat ilgio žingsnį ta pačia kryptimi ir pan. [9]. Daugelis tokių rekomendacijų išreiškia *minimizavimo nusileidimo kryptimi* apytikslią realizaciją. Idėja minimizuoti antigradiento kryptimi atrodo labai patraukliai. Jei antigradientas nukreiptas funkcijos greičiausio mažėjimo kryptimi, tai kodėl maksimaliai neišnaudojus geros krypties: atlikime šia kryptimi vienmatę minimizaciją! Įdomu pažymėti, kad tokią idėją dar 1847 metais realizavo Augustin Cauchy. Reikia pažymėti, kad tiksliai realizuoti vienmatės minimizacijos bendruoju atveju negalėsime, o apytikslus skaitmeninis metodas pareikalaus skaičiavimų laiko. Tačiau norėdami ištirti potencialias galimybes panagrinėkime abstrakčią *greičiausio nusileidimo metodo* versiją:

$$X_{i+1} = X_i + \gamma_i \, S_i \,, \tag{2.8}$$

čia $S_i = \frac{-\nabla f(X_i)}{||\nabla f(X_i)||}$, $\gamma_i = \arg\min_{\gamma \geq 0} f(X_i + \gamma \cdot \, S_i)$, t.y. žingsnis žengiamas iki funkcijos minimumo antigradiento kryptimi.

3 teorema. Sakysim, $f(X)$, $X \in R^n$, tolygiai diferencijuojama ir aprėžta iš apačios

$f(X) \geq f_- > -\infty.$

Tada greičiausio nusileidimo metodas (2.8) konverguoja gradiento prasme, t.y.

$\lim_{i \to \infty} \nabla f(X_i) = 0.$

Šios teoremos sąlygos bendresnės negu anksčiau nagrinėtu (gradientinio nusileidimo metodo su pastoviu žingsnio daugikliu) atveju. Tačiau vykdant (2.8) metodą, atliekamas minimizavimas antigradiento kryptimi, todėl (2.8) konvergavimas įrodomas netgi paprasčiau. Manydami, kad konvergavimo faktas abejonių nekelia, įrodymo nepateiksime. Teoremoje kalbama, kad ribinis metodo sugeneruotos sekos taškas yra stacionarus. Su tam tikromis išlygomis, paminėtomis aptariant gradientinio metodo konvergavimą, galima teigti, kad sugeneruotos sekos ribinis taškas yra lokalaus minimumo taškas.

Įdomu palyginti gradientinio ir greičiausio nusileidimo metodų konvergavimo greičius. Tokio palyginimo rezultatai svarbūs nusileidimo krypties ir žingsnio ilgio parinkimo racionalumui įvertinti. Gradientinio nusileidimo metodo konvergavimo greitis buvo įvertintas iškilių funkcijų, "įspraudžiamų" tarp dviejų teigiamai apibrėžtų kvadratinių funkcijų, klasei. Greičiausio nusileidimo metodo konvergavimas nagrinėjamas siauresnei teigiamai apibrėžtų kvadratinių funkcijų klasei. Sakysim, kad

$$f(X) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j - \sum_{i=1}^{n} b_i x_i = \\ = \frac{1}{2} X \cdot A \cdot X - B \cdot X, \tag{2.9}$$

čia matrica A yra teigiamai apibrėžta, t.y. visos jos tikrinės reikšmės yra teigiamos. Žinoma, kad (2.9) gradientas lygus

$$\nabla f(X) = A \cdot X - B, \tag{2.10}$$

o minimumo taškas gali būti užrašytas analitine formule $X_* = A^{-1}B$. Apskaičiavimams pagal šią formulę užtenka baigtinio operacijų skaičiaus, nors apversti matricą ne visada lengva.

Panagrinėsime (2.9) minimumo taško paiešką greičiausio nusileidimo metodu. Bet koks vienmatis tokios funkcijos pjūvis (t.y. kai X priklauso kokiai nors n-matės erdvės tiesei $X(\gamma)$, $-\infty < \gamma < \infty$) yra kvadratinė funkcija

$$\varphi(\gamma) = f(X(\gamma)) = c_2\gamma^2 + c_1\gamma + c_0,\ c_2 > 0,$$

kurios minimumo taškas

$$\gamma_* = -\frac{c_1}{2c_2}\,.$$

Tiesės, einančios per tašką X_i antigradiento kryptimi atveju

$$X = X_i - \gamma\,\nabla f(X_i)$$

(2.9) pjūvis yra kvadratinė kintamojo γ funkcija, kurios koeficientai išreiškiami šitokiomis formulėmis

$$\begin{aligned}
c_1 &= \sum_{k=1}^{n}\sum_{j=1}^{n} a_{kj}\frac{\partial f(X_i)}{\partial x_k}x_{ij} - \sum_{k=1}^{n} b_k\frac{\partial f(X_i)}{\partial x_k} = \\
&= \sum_{k=1}^{n}\left(\sum_{j=1}^{n} a_{kj}x_{ij} - b_k\right)\frac{\partial f(X_i)}{\partial x_k} = \sum_{k=1}^{n}\left(\frac{\partial f(X_i)}{\partial x_k}\right)^2 \\
&= ||\nabla f(X_i)||^2,
\end{aligned}$$

$$\begin{aligned}
c_2 &= \frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n} a_{kj}\frac{\partial f(X_i)}{\partial x_k}\cdot\frac{\partial f(X_i)}{\partial x_j} = \\
& \nabla f(X_i)\cdot A\cdot\nabla f(X_i),
\end{aligned}$$

o minimumo taškas šitaip

$$\gamma_i = \frac{||\nabla f(X_i)||^2}{\nabla f(X_i)\cdot A\cdot\nabla f(X_i)}. \qquad (2.11)$$

4 teorema. Greičiausio nusileidimo metodas (2.8) teigiamai apibrėžtos kvadratinės tikslo funkcijos atveju konverguoja geometrinės progresijos greičiu

$$||X_i - X_*|| \le \sqrt{\frac{2}{\lambda_0}(f(X_0) - f(X_*))}\cdot\left(\frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0}\right)^i, \qquad (2.12)$$

kur λ_* didžiausia, o λ_0 mažiausia matricos A tikrinės reikšmės.

Įrodymas. Atsižvelgdami į tikslo funkcijos (2.9) ir žingsnio daugiklio (2.11) išraiškas, gauname šitokią lygybę

$$\begin{aligned}
f(X_{i+1}) &= \left(\frac{1}{2}(X_i - \gamma_i \nabla f(X_i)) \cdot A - B\right)(X_i - \gamma_i \nabla f(X_i)) = \\
&= f(X_i) - \gamma_i \nabla f(X_i) \cdot (AX_i - B) \cdot \nabla f(X_i) + \\
&\quad + \frac{1}{2}\gamma_i^2 \nabla f(X_i) \cdot A \cdot \nabla f(X_i) \\
&= f(X_i) - \gamma_i \nabla f(X_i) \cdot \nabla f(X_i) + \\
&\quad + \frac{1}{2}\gamma_i^2 \nabla f(X_i) \cdot A \cdot \nabla f(X_i) \\
&= f(X_i) - \frac{1}{2}\frac{||\nabla f(X_i)||^4}{\nabla f(X_i) \cdot A \cdot \nabla f(X_i)}\,.
\end{aligned}$$

Kadangi

$$2(f(X_i) - f(X_*)) = X_i A X_i - 2BX_i + X_* A X_* = \tag{2.13}$$

$$= X_i A X_i - 2X_* A X_i + X_* A X_* = (X_i - X_*)A(X_i - X_*) =$$

$$= \left(A^{-1}(AX_i - B)\right)(AX_i - B) = (A^{-1}\nabla f(X_i)) \cdot \nabla f(X_i), \tag{2.14}$$

tai

$$\frac{f(X_{i+1}) - f(X_*)}{f(X_i) - f(X_*)} = 1 -$$

$$-\frac{||\nabla f(X_i)||^4}{(\nabla f(X_i) \cdot A \cdot \nabla f(X_i)) \cdot (\nabla f(X_i) \cdot A^{-1} \cdot \nabla f(X_i))}\,. \tag{2.15}$$

Pasinaudodami Kantorovičiaus nelygybe

$$(XAX) \cdot (XA^{-1}X) \leq \frac{(\lambda_0 + \lambda_*)^2 ||X||^4}{4\lambda_0\lambda_*}, \tag{2.16}$$

(2.15) galime užrašyti taip

$$\frac{f(X_{i+1}) - f(X_*)}{f(X_i) - f(X_*)} \leq \left(\frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0} \right)^2 .$$

Iteratyviai pritaikę pastarąją formulę i kartų, gausime šitokią nelygybę

$$f(X_i) - f(X_*) \leq (f(X_0) - f(X_*)) \left(\frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0} \right)^{2i} . \tag{2.17}$$

Kadangi

$$\begin{aligned} 2(f(X_i) - f(X_*)) &= (X_i - X_*)A(X_i - X_*) \geq \\ &\lambda_0 ||X_i - X_*||^2, \end{aligned}$$

tai (2.17) gali būti užrašyta šitokia forma

$$||X_i - X_*|| \leq \sqrt{\frac{2}{\lambda_0}(f(X_0) - f(X_*))} \cdot \left(\frac{\lambda_* - \lambda_0}{\lambda_* + \lambda_0} \right)^i ,$$

ką ir reikėjo įrodyti.

Gautasis rezultatas nenuteikia optimistiškai: labai palankiu greičiausio nusileidimo metodui atveju, kai minimumo taškas antigradiento kryptimi apskaičiuojamas analitiškai, šio metodo konvergavimo greitis yra toks pat, kaip ir paprastesnio gradientinio nusileidimo metodo konvergavimo greitis platesnei funkcijų klasei. Iš gautų rezultatų darytina išvada, kad metodų, kurių nusileidimo kryptis sutampa su antigradiento kryptimi, konvergavimas ne greitesnis už tiesinį.

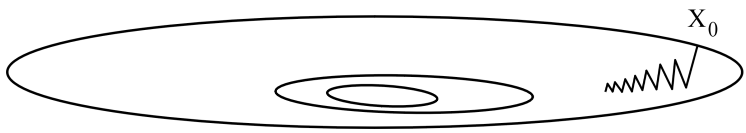

Figure 2.6: Lėto greičiausio nusileidimo metodo konvergavimo iliustracija

Palyginę abiejų metodų konvergavimo greičio įverčius matome, kad jiems palankus simetriškos funkcijos atvejis. Bet tokia funkcija iš esmės tepriklauso nuo vieno kintamojo, ir šis atvejis praktiniu požiūriu nelabai įdomus. Esant stipriai nesimetrinėms funkcijoms, vaizdžiai apibūdinamoms "gilaus griovio" įvaizdžiu, abu metodai konverguoja lėtai. 2.6 pav. iliustruoja lėtą "zigzaginį" konvergavimą dvimačiu atveju.

Greičiausio nusileidimo metodas paprastas tiek teorinio nagrinėjimo, tiek programinės realizacijos požiūriais. Jį nagrinėjant galima išsiaiškinti daug svarbių optimizavimo metodų savybių. Tačiau jis praktiškai beveik nenaudojamas, kadangi yra pernelyg lėtas.

2.6 Niutono metodas

Pagrindinė gradientinių metodų idėja gan paprasta. Iš turimo taško žengiama ta kryptimi, kuria funkcija greičiausiai mažėja, t.y. tiesiogiai realizuojamas tikslas minimizuoti $f(X)$. Tačiau, remdamiesi būtinosiomis minimumo sąlygomis, galime uždavinį spręsti netiesiogiai, pakeisdami jį lygčių sistemos $\nabla f(X) = 0$ sprendimu, ir pasinaudodami šių uždavinių sprendimo teorija bei sukaupta patirtimi. Niutono metodas yra vienas iš klasikinių metodų lygčių sistemoms spręsti. Jo esmė ta, kad kairiosiose lygčių sistemos pusėse esančios funkcijos ištiesinamos taške X_i ir, išsprendus tiesinę lygčių sistemą, turėtasis taškas pakeičiamas tiesinės lygčių sistemos sprendiniu X_{i+1}. Norėdami realizuoti šią idėją ir ištiesinti gradiento komponentes sudarančias funkcijas, pasinaudosime tikslo funkcijos $f(X)$ Teiloro eilutės nariais iki antrosios eilės:

$$\nabla f(X) \approx \nabla f(X_i) + H(X_i) \cdot (X - X_i). \tag{2.18}$$

Prilyginę (2.18) nuliui, gausime metodą aprašančią iteracinę formulę

$$X_{i+1} = X_i - \nabla f(X_i) \cdot H^{-1}(X_i). \tag{2.19}$$

Kadangi formule (2.19) išreiškiamas tikslo funkcijos $f(X)$ kvadratinės aproksimacijos minimumo taškas, tai Niutono metodą galima interpretuoti ir šitaip. Taške X_i tikslo funkcija aproksimuojama Teiloro eilute, įvertinant ne aukštesnės negu antros eilės narius. Apskaičiuojamas aproksimuojančiosios kvadratinės funkcijos minimumo taškas X_{i+1} ir juo pakeičiamas X_i. Metodas (2.19) apibrėžtas naudojant pirmąsias ir antrąsias išvestines, bet nenaudojant tikslo funkcijos reikšmių. Dviejų kintamųjų funkcijos atveju Niutono metodas iliustruojamas 2.7 pav. Jei tikslo funkcija būtų kvadratinė, jos minimumas būtų gautas vienu žingsniu. Bendru atveju formulė (2.19) taikoma iteratyviai.

Įdomi dar viena Niutono metodo interpretacija. Vektoriaus daugyba iš matricos geometriškai reiškia tiesinę erdvės transformaciją, t.y. vektoriaus pasukimo ir ištempimo (suspaudimo) operacijas. Todėl vektorius $\nabla f(X_i) \cdot H^{-1}(X_i)$ gali būti interpretuojamas kaip gradiento vektorius tiesiškai transformuotoje erdvėje, o Niutono metodo žingsnio kryptis, kaip antigradiento kryptis matrica $H(X_i)$ apibrėžtoje metrikoje.

Sakysime, kad egzistuoja taškas X_*, kuriame patenkintos pakankamos minimumo sąlygos ir taško X_* $\bar{\delta}-$aplinkoje $H(X)$ tolydi.

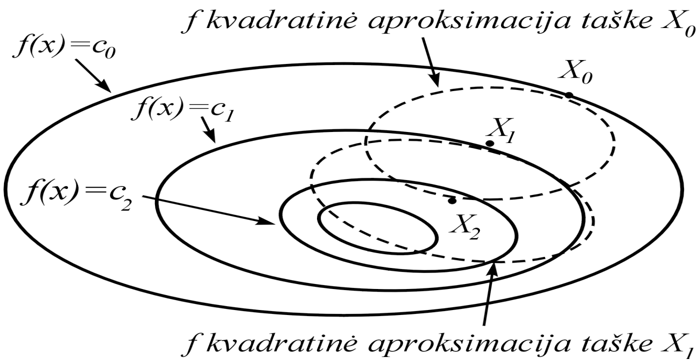

Figure 2.7: Niutono algoritmo iliustracija ; $c_0 > c_1 > c_2$

5 teorema. Jei pradinis taškas X_0 parenkamas pakankamai artimoje minimumo taško aplinkoje, tai (2.19) metodo generuojama seka konverguoja į minimumo tašką greičiau negu tiesiškai.

Jei be to $H(X)$ patenkina Lipshitz'o sąlygą

$$||H(X) - H(Z)|| \leq L \cdot ||X - Z||, \tag{2.20}$$

tai Niutono metodas konverguoja kvadratiniu greičiu.

Įrodymas. Sakysim $\delta > 0$ ir $M > 0$ tokie, kad matrica $H(X)$ neišsigimusi ir $||H^{-1}(X)|| \leq M$, visiems X, $||X - X_*|| \leq \delta$. Pagal metodo apibrėžimą

$$||X_{i+1} - X_*|| = ||X_i - X_* - \nabla f(X_i) \cdot H^{-1}(X_i)|| =$$
$$= ||H^{-1}(X_i) \cdot (H(X_i) \cdot (X_i - X_*) - \nabla f(X_i))||.$$

Pastarojoje formulėje gradientą $\nabla f(X_i)$, pakeitę jį išreikiančia (2.6) formule, ir atlikę keletą akivaizdžių pertvarkymų, gausime nelygybę

$$||X_{i+1} - X_*|| = ||H^{-1}(X_i) \cdot [H(X_i) -$$

$$- \int_0^1 H(X_* + t(X_i - X_*))dt] \cdot (X_i - X_*)|| =$$

$$\begin{aligned} = \quad & ||H^{-1}(X_i) \cdot \left(\int_0^1 (H(X_i) - H(X_* + t(X_i - X_*)))dt \right) \cdot \\ & \cdot (X_i - X_*)|| \leq \end{aligned}$$

$$\begin{aligned} \leq \quad & M \left(\int_0^1 ||H(X_i) - H(X_* + t(X_i - X_*))||dt \right) \cdot \\ & \cdot ||X_i - X_*||. \end{aligned} \tag{2.21}$$

Kadangi $H(X)$ tolydi, galima parinkti tokį mažą δ, kad pirmąją iteraciją atitinkantis pointegrinis skirtumas būtų norimai mažas. Pratęsdami pagal indukciją, gausime nelygybių seką

$$\begin{aligned} ||X_{i+1} - X_*|| &\leq q_i ||X_i - X_*||, \\ q_i &\to 0, \text{ kai } ||X_i - X_*|| \to 0, \end{aligned}$$

reiškiančią, kad $||X_i - X_*|| \to 0$ virštiesiniu greičiu.

Remdamiesi (2.20) ir (2.21) gausime šitokią nelygybę

$$\begin{aligned} ||X_{i+1} - X_*|| &\leq M \left(\int_0^1 Lt ||X_i - X_*|| dt \right) ||X_i - X_*|| \\ &= \frac{LM}{2} ||X_i - X_*||^2, \end{aligned}$$

kuri įrodo, kad Niutono metodo konvergavimo greitis yra kvadratinis:

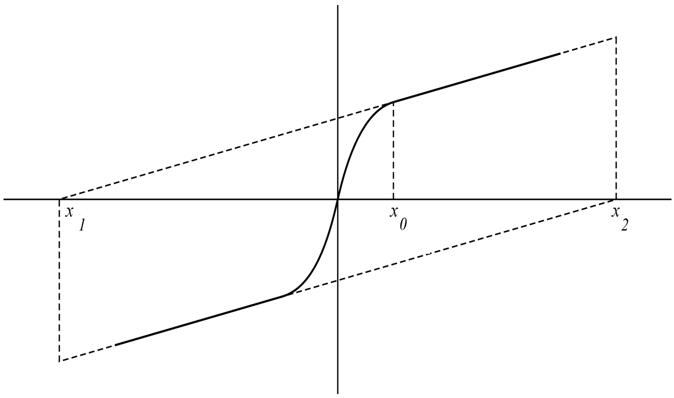

Figure 2.8: Niutono metodo užsiciklinimo pavyzdys

$$||X_i - X_*|| \leq \frac{2}{L \cdot M} \left(\frac{L \cdot M}{2} ||X_0 - X_*|| \right)^{(2^i)}. \quad (2.22)$$

Jei patenkintos teoremos sąlygos, tai Niutono metodas konverguoja labai greitai. Iš pirmo žvilgsnio tas tikrai patrauklu. Tačiau teoremos sąlygas galima užtikrinti tik parinkus tokį pradinį tašką, kuris būtų pakankamai artimas sprendiniui. Nelygybė (2.22) konkrečiau išreiškia pradinio ir minimumo taškų artumo matą: skliausteliuose esančio reiškinio reikšmė turi būti mažesnė už 1. Iš blogo pradinio taško Niutono metodas gali apskritai nekonverguoti. Tokią situaciją iliustruoja 2.8 paveikslas, kuriame pavaizduotos funkcijos $f'(x)$ atveju šaknies $f'(x) = 0$ paieška Niutono metodu užsiciklina.

Šis pavyzdys rodo, kad Niutono metodas gali nekonverguoti net iškilos tikslo funkcijos atveju. Ypač neaiškus metodo konvergavimo klausimas, jei metodo vykdymo metu gaunamuose taškuose antrųjų išvestinių matrica nėra teigiamai apibrėžta. Pavyzdžiui, jei kokiame nors taške X_j matrica $H(X_j)$ būtų neigiamai apibrėžta, tai iš to taško būtų žengiamas žingsnis į kvadratinės aprosimacijos maksimumą.

Siekiant susilpninti griežtas konvergavimo sąlygas, buvo pasiūlytos įvairios Niutono metodo modifikacijos. Viena iš jų tokia – modifikuojama kvadratinė tikslo funkcijos aproksimacija taip, kad antros eilės išvestinių matrica būtų visada teigiamai apibrėžta. Kitoje modifikacijoje kvadratinė aproksimacija vartojama tik mažėjimo krypčiai einamojo taško aplinkoje nustatyti; po to vienmatė minimizacija atliekama nustatytąja kryptimi panašiai, kaip greičiausio nusileidimo metode. Tokiu būdu praplečiama konvergavimo sritis, tačiau nebeužtikrinamas kvadratinis konvergavimo greitis.

Vertinant praktinio panaudojimo požiūriu, Niutono metodo privalumas (didelis konvergavimo greitis) dažnai nustelbiamas jo trūkumų: siauros konvergavimo srities ir realizavimo sunkumų, sąlygojamų antrųjų išvestinių skaičiavimu bei jų matricos apvertimu. Niutono metodas ypač vertingas tokiais atvejais, kai būtinas labai didelis sprendinio tikslumas ir turimas geras pradinis taškas, pavyzdžiui, surastas kitu metodu, turinčiu platesnę konvergavimo sritį.

2.7 Patikimos srities metodai

Anksčiau nagrinėti metodai pagrįsti tiesine (gradientinis ir greičiausio nusileidimo metodai) arba kvadratine (Niutono metodas) aproksimacija. Tokios aproksimacijos gerai aprašo tikslo funkciją tam tikroje srityje, kuri paprastai sutampa su einamojo taško aplinka. Bet už tos aplinkos ribų aproksimuojančioji ir aproksimuojamoji funkcijos gali labai skirtis. Patikimos srities (trust region) metoduose aproksimacija naudojama tik tokioje srityje, kurioje aproksimacija yra tikrai priimtina. Šiame skyrelyje paaiškinsime tik pačią patikimos srities metodų idėją, ją panaudodami Niutono metodo modifikavimui. Detaliau su šiais metodais, kuriuos nagrinėti gan sudėtinga, galima susipažinti, pavyzdžiui, pasiskaičius [3] knygą.

Niutono metodo $i+1-$ masis žingsnis žengiamas į $f(X)$ kvadratinės aproksimacijos $f^i(X)$ minimumo tašką, čia $f^i(X)$ yra $f(X)$ Teiloro eilutės taško X_i aplinkoje du pirmieji nariai

$$f^i(X) = f(X_i) + \nabla f(X_i) \cdot (X - X_i) +$$

$$+\frac{1}{2}(X - X_i) \cdot H(X_i) \cdot (X - X_i). \tag{2.23}$$

Teoriškai $f^i(X)$ yra gera $f(X)$ aproksimacija taško X_i aplinkoje. Jei, savo ruožtu, X_i yra minimumo taško X_* aplinkoje, tai tikėtina, kad $f^i(X)$ minimumo taškas yra arti X_*. Tuo pagrįstas Niutono metodo greitas konvergavimas iš gero pradinio taško. Jei X_i yra toli nuo $f(X)$ minimumo taško X_*, tai $f^i(X)$ gali ne tik kad labai skirtis nuo $f(X)$ taško X_* aplinkoje, bet ir $f^i(X)$ minimumo taškas gali būti labai nutolęs nuo X_*. Kaip jau anksčiau minėjome, iš blogo pradinio taško Niutono metodas gali apskritai nekonverguoti.

Atsižvelgdami į tai, kad $f^i(X)$ yra gera $f(X)$ aproksimacija taško X_i aplinkoje, galime modifikuoti Niutono metodą šitaip

$$X_{i+1} = \arg \min_{||X-X_i|| \le \delta} f^i(X). \tag{2.24}$$

Svarbu pažymėti, kad net tuo atveju, kai $H(X_i)$ nėra teigiamai apibrėžta, (2.24) sumažina tikslo funkcijos reikšmę. Naujasis taškas (2.24) apskaičiuojamas sprendžiant kvadratinės funkcijos minimizavimo uždavinį esant ribojimui $||X - X_i|| \le \delta$. Minimizavimo su ribojimais metodai bus nagrinėjami kitame skyriuje. Kadangi (2.24) uždavinyje tėra vienintelis ribojimas $||X - X_i|| \le \delta$, tai jį galima palyginti lengvai įvertinti.

Žengiant žingsnį patikimoje srityje, galima užtikrinti monotonišką tikslo funkcijos mažėjimą ir metodo konvergavimą iš bet kokio pradinio taško. Priartėjus prie minimumo taško tiek, kad ribojimas

$||X - X_i|| \leq \delta$ nebeturi reikšmės, metodas, aprašomas (2.23), nebesiskiria nuo Niutono metodo. Tuo užtikrinama ir plati metodo konvergavimo sritis, ir didelis konvergavimo greitis.

2.8 Kintamos metrikos metodai

Niutono metodas buvo modifikuojamas įvairiai. Viena iš idėjų pasirodė ypač sėkminga ir ja remiantis buvo sukurti metodai, vadinami kintamos metrikos arba kvaziniutono metodais. Juose antrųjų išvestinių matrica nėra naudojama, o atvirkštinė antrųjų išvestinių matrica aproksimuojama remiantis tik informacija apie pirmąsias funkcijos išvestines. Vienas iš žymiausių šios klasės metodų, DFP (vadinamas jo autorių Davidon-Fletcher-Powell vardu), aprašomas šitaip:

1. Inicializacija. Parinkti pradinį tašką X_0. Suteikti reikšmę sustojimo parametrui ε. Suteikti reikšmes $n \times n$ matricos D_1 elementams taip, kad matrica būtų simetrinė ir teigiamai apibrėžta, pvz. vienetinė. $k = 0$, $j = 1$. $Y_1 = X_0$.

2. Minimizacija nusileidimo kryptimi. Jei

$$||\nabla f(Y_j)|| < \varepsilon,$$

sustoti. Priešingu atveju tęsti: apskaičiuoti nusileidimo krypties vektorių $S_j = -\nabla f(Y_j) \cdot D_j$; rasti

$$\gamma_j = \arg\min_{\gamma > 0} f(Y_j + \gamma S_j), \tag{2.25}$$

ir pereiti į naują tašką

$$\begin{aligned} Y_{j+1} &= Y_j + \gamma_j \cdot S_j, \\ j &= j + 1, \end{aligned}$$

jei $j \leq n+1$, vykdyti 3 žingsnį, priešingu atveju $k = k+1$, $j = 1$, $X_k = Y_1 = Y_{n+1}$, kartoti **2** žingsnį**.**

3.Matricos perskaičiavimas.

$$D_{j+1} = D_j + \frac{P_j^t \cdot P_j}{P_j \cdot Q_j^t} - \frac{D_j \cdot Q_j^t \cdot Q_j \cdot D_j}{Q_j \cdot D_j \cdot Q_j^t},$$

čia P_j ir Q_j yra vektoriai - eilutės

$$\begin{aligned} P_j &= \gamma_j S_j = Y_{j+1} - Y_j, \\ Q_j &= \nabla f(Y_{j+1}) - \nabla f(Y_j), \end{aligned}$$

o viršutinis indeksas t reiškia transponavimą,

$j = j+1$, vykdyti **2.**

Atkreipsime dėmesį į tai, kad kas n vienmačių (2.25) minimizacijų matricai suteikiama pradinė reikšmė. Tas daroma siekiant išvengti paklaidų susikaupimo.

DFP yra vienas iš daugelio šios klasės metodų. Visų jų vienos iteracijos atlikimo laikas proporcingas n^2. Kvadratinės funkcijos atveju šie metodai išsprendžia uždavinį per n žingsnių. Šiuo požiūriu jų efektyvumas toks pat kaip ir Niutono metodo, kadangi Niutono metodo iteracijos atlikimo laikas proporcingas n^3. Kintamos metrikos metodai konverguoja iš bet kokio pradinio taško. Konvergavimo greitis yra virštiesinis $n-$žingsnių iteracijų skaičiaus atžvilgiu. Viena n-žingsnė iteracija apima n vienmačių minimizacijų. Kontamos metrikos metodai yra efektyvūs ir plačiai vartojami.

2.9 Jungtinių krypčių metodai

Jungtinių krypčių (conjugate directions) metodai remiasi kvadratine tikslo funkcijos aproksimacija. Atskiros jų versijos buvo pasiūlytos apie 1950 metus. Metodų pavadinimas kilo iš to, kad kvadratinės funkcijos atveju šių metodų generuojamos nusileidimo kryptys yra jungtinės kvadratinę funkciją aprašančios matricos atžvilgiu. Kintamos metrikos metodai yra atskiras šios klasės metodų atvejis, kadangi kvadratinės tikslo funkcijos atveju (2.25) vienmatės minimizacijos vykdomos jungtinėmis kryptimis. Paprastai, kai kalbama apie jungtinių krypčių metodus, turimi galvoje metodai, kurių vienai iteracijai atlikti reikia laiko proporcingo n. Kitais atvejais jie vadinami siauresnės klasės vardu, pvz. kintamos metrikos metodais.

Sakysim, tikslo funkcija yra kvadratinė, apibrėžiama (2.9) formule. Žinoma, kad ortogonalia Euklido ervės transformacija galima kintamuosius pakeisti taip, kad (2.9) naujais kintamaisiais bus užrašoma taip

$$f(V) = \sum_{i=1}^{n} \alpha_i v_i^2.$$

Akivaizdu, kad naujojoje koordinačių sistemoje kvadratinės funkcijos minimumas iš bet kokio pradinio taško randamas atlikus n vienačių minimizavimų išilgai naujų koordinačių ašių. Minimas koordinačių transformacijos uždavinys ekvivalentus matricos A, įeinančios į (2.9) formulę, tikrinių vektorių apskaičiavimo uždaviniui, kuris skaičiavimų sudėtingumo požiūriu nėra lengvas. Bet nebūtina atlikti koordinačių transformaciją išreikšta forma. Tą patį efektą gautume, jei pradinėje koordinačių sistemoje atliktume minimizavimą naujosios sitemos koordinačių vektorių kryptimis. Teisingas dar bendresnis teiginys: minimumo tašką galima pasiekti atlikus n vienmačių minimizacijų kryptimis atitinkančiomis matricos A jungtines kryptis. Pastarasias galima apskaičiuoti truputį modifikavus greičiausio nusileidimo metodą. Viena iš jungtinių krypčių metodo versijų yra Flecher - Reeves jungtinių gradientų metodas, kuris aprašomas šitaip:

$$
\begin{aligned}
S_0 &= -\nabla f(X_0),\ i=0, \\
X_{i+1} &= X_i + \gamma_i \cdot S_i,\ \gamma_i = \arg\min_{\gamma>0} f(Y_i+\gamma S_i), \\
i &= i+1, \\
\beta_i &= \frac{\nabla f(X_i)\cdot \nabla f(X_i)}{\nabla f(X_{i-1})\cdot \nabla f(X_{i-1})}, \\
S_i &= -\nabla f(X_i) + \beta_i \cdot S_{i-1}.
\end{aligned} \tag{2.26}
$$

Paeiliui einančias n vienmačių minimizacijų pavadinkime n-žingsne iteracija. Galima įrodyti, kad šiuo metodu kvadratinės funkcijos minimumas randamas per vieną n-žingsnę iteraciją. Bendruoju (nekvadratinės tikslo funkcijos) atveju konvergavimo greitis yra virštiesinis n-žingsnių iteracijų skaičiaus atžvilgiu . Panašiomis savybėmis pasižymi ir kiti jungtinių krypčių metodai. Jie praktinio efektyvumo požiūriu nusileidžia kintamos metrikos metodams ir yra jautresni išvestinių įvertinimo bei vienmačio minimizavimo paklaidoms. Tačiau jie naudotini tada, kai kintamųjų yra tiek daug, kad matricinėms operacijoms reikalingas n^2 proporcingas laikas yra per ilgas, o minėtos paklaidos yra paneigtinos. Programinė (2.26) metodo realizacija ne ką sudėtingesnė už greičiausio nusileidimo metodo realizaciją.

Toliau įrodysime, kad (2.26) metodu kvadratinės funkcijos minimumas randamas per baigtinį iteracijų skaičių. Teorija besidominčiam skaitytojui ši medžiaga turėtų būti įdomi, nes panašus matematinis aparatas vartojamas ir kintamos metrikos metodų analizei. Silpnesnio matematinio pasiruošimo skaitytojas gali praleisti šią skyrelio dalį be žalos toliau dėstomos medžiagos supratimui.

Sakysim A yra $n\times n$ teigiamai apibrėžta matrica. Nenuliniai vektoriai $S_1,...,S_k$, $k\le n$, vadinami A-jungtiniais, jei jie patenkina lygybę

$$S_i A S_j = 0,\ \ i\neq j.$$

Nesungu įrodyti, kad A-jungtiniai vektoriai $S_1, ..., S_k$, yra tiesiškai nepriklausomi. Sakysim priešingai $S_k = \sum_{i=1}^{k-1} \alpha_i S_i$. Tada turėtų galioti lygybė

$$S_k A S_k = S_k A \sum_{i=1}^{k-1} \alpha_i S_i = \sum_{i=1}^{k-1} \alpha_i S_k A S_i = 0,$$

kuri neteisinga, nes A - teigiamai apibrėžta matrica.

Sakysim $f(X) = \frac{1}{2} XAX - BX$, ir $S_0, ..., S_{n-1}$ yra n A-jungtiniai vektoriai. Nuosekliai atliekant n vienmačių minimizacijų tų vektorių kryptimis gaunama seka

$$\begin{aligned} X_{i+1} &= X_i + \gamma_i S_i, \\ \gamma_i &= \arg\min_r f(X_i + rS_i). \end{aligned}$$

Lema. X_{j+1} yra funkcijos $f(X)$ minimumo taškas vektorių $S_0, ..., S_j$ generuojamoje tiesinėje daugdaroje T_j, čia

$$T_j = \{X : X = X_0 + V,\ V = \sum_{k=0}^{j} \alpha_k S_k.$$

.

Įrodymas. Kadangi X_{j+1} yra $f(X)$ minimumo taškas kryptimi S_j, tai galioja lygybė

$$\left.\frac{\partial f(X_j + \gamma S_j)}{\partial \gamma}\right|_{\gamma = \gamma_j} = \nabla f(X_{j+1}) \cdot S_j = 0. \tag{2.27}$$

Todėl visiems $i = 0, ..., j-1$ galioja lygybė

$$\nabla f(X_{j+1}) \cdot S_i = (X_{j+1} A - B) \cdot S_i =$$

$$= \left(X_{i+1} + \sum_{k=i+1}^{j} \gamma_k S_k \right) AS_i - BS_i =$$

$$= X_{i+1} AS_i - BS_i = \nabla f(X_{i+1}) \cdot S_i = 0,$$

kurios įrodyme buvo remtasi S_i ir S_k, $k = i+1, ..., j$ jungtinumu. Kadangi funkcijos išvestinė vektoriaus S_k, $k = 0, ..., j$, kryptimi lygi funkcijos gradiento ir vektoriaus S_k skaliarinei sandaugai, tai

$$\left. \frac{\partial f(X_{j+1} + \gamma S_k)}{\partial \gamma} \right|_{\gamma=0} = \nabla f(X_{j+1}) \cdot S_i = 0, \tag{2.28}$$

Paskutinioji lygybė reiškia, kad funkcijos $f(X)$ minimumas aibėje T_j pasiekiamas taške X_{j+1}.

Išvada.

$$X_n = \arg \min_{X \in R^n} f(X).$$

Bet kokių tiesiškai nepriklausomų vektorių aibę $Z_0, ..., Z_k$ galima pakeisti A jungtinių vektorių aibe $S_0, ..., S_k$ taip, kad vektorių $Z_0, ..., Z_k$ ir $S_0, ..., S_k$ generuojami poerdviai sutaptų. Toks pakeitimas gali būti atliktas Gram-Schmidt prosedūra:

$$S_0 = Z_0,\ S_{i+1} = Z_{i+1} + \sum_{l=0}^{i} c_{i+1,l}\, S_l,$$

čia koeficientai $c_{i+1,l}$ apskaičiuojami remiantis S_j A-jungtinumu:

$$\begin{aligned} S_{i+1} AS_j &= Z_{i+1} AS_j + \sum_{l=0}^{i} c_{i+1,l}\, S_l AS_j = 0, \\ c_{i+1,j} &= -\frac{Z_{i+1} AS_j}{S_j AS_j},\ j = 0, ..., i. \end{aligned}$$

Gram-Schmidt prosedūra gali būti naudojama A-jungtinių vektorių sudarymai ir iš tiesiškai priklausomų vektorių aibės. Tokiu atveju kai kuriems j bus gauta $S_j = 0$, o nenulinių $A-$jungtinių vektorių generuojamas poerdvis sutaps su $Z_0, ..., Z_k$ generuojamu poerdviu.

6 teorema. Vektoriai, apibrėžti (2.26), yra A-jungtiniai, o vienmačių minimizacijų skaičius, reikalingas $f(X)$ minimumui surasti, yra ne didesnis už n.

Įrodymas. Sakysim $Z_i = -\nabla f(X_i) = -AX_i - B$, čia taškai X_i, apibrėžiami pagal (2.26). Parodysim, kad (2.26) apibrėžiami vektoriai yra A-jungtiniai. Visų pirma pastebėsime, kad

$$\nabla f(X_1) = AX_1 - B,\ X_1 - X_0 = -\gamma_1 \nabla f(X_0),$$

$$\nabla f(X_1) - \nabla f(X_0) = A(X_1 - X_0) = -\gamma_1 A \nabla f(X_0),$$

čia γ_1 pirmojo žingsnio daugiklis. Su vektoriais Z_0, Z_1 pritaikysime Gram-Schmidt metodą:

$$S_1 = -\nabla f(X_1) + \frac{\nabla f(X_1) A \nabla f(X_0)}{\nabla f(X_0) A \nabla f(X_0)} \nabla f(X_0). \tag{2.29}$$

Kadangi galioja tokios lygybės

$$\nabla f(X_0) A \nabla f(X_0) = -\frac{1}{\gamma_1}(X_1 - X_0) A \nabla f(X_0) =$$

$$\begin{aligned} &= -\frac{1}{\gamma_1}(\nabla f(X_1) - \nabla f(X_0)) \cdot A^{-1}A \cdot \nabla f(X_0) = \\ &= \frac{1}{\gamma_1}||\nabla f(X_0)||^2, \end{aligned}$$

$$\begin{aligned}\nabla f(X_1)A\nabla f(X_0) &= \nabla f(X_1)A(-\frac{1}{\gamma_1}(X_1 - X_0)) = \\ &= -\frac{1}{\gamma_1}||\nabla f(X_1)||^2,\end{aligned}$$

tai (2.29) gali būti užrašyta šitokia forma

$$S_1 = -\nabla f(X_1) - \frac{||\nabla f(X_1)||^2}{||\nabla f(X_0)||^2}\nabla f(X_0),$$

kuri sutampa su (2.26). Pratęsdami pagal indukciją ir įvertindami lygybes

$$\begin{aligned}S_iAS_i &= \frac{1}{\gamma_{i+1}}(X_{i+1} - X_i)AS_i = \\ &= \frac{1}{\gamma_{i+1}}(\nabla f(X_{i+1}) - \nabla f(X_i))S_i = \\ &= \frac{1}{\gamma_{i+1}}||\nabla f(X_i)||^2,\end{aligned}$$

$$\begin{aligned}\nabla f(X_{i+1})AS_i &= \nabla f(X_{i+1})A\frac{1}{\gamma_{i+1}}(X_{i+1} - X_i) = \\ &= \frac{1}{\gamma_{i+1}}\nabla f(X_{i+1})(\nabla f(X_{i+1}) - \nabla f(X_i)) = \\ &= \frac{1}{\gamma_{i+1}}||\nabla f(X_{i+1})||^2,\end{aligned}$$

įsitikinsime, kad Gram-Schmidt metodu gaunami jungtiniai vektoriai sutampa su (2.26) metode naudojamais vienmatės minimizacijos krypčių vektoriais. Todėl (2.26) metodo krypčių vektoriai yra A-jungtiniai. Kvadratinės funkcijos minimumo taškui surasti reikės ne daugiau negu n vienmačių minimizacijų, kaip tas išplaukia iš lemos išvados.

2.10 Paieškos metodai

Nusileidimo metodai pagrįsti tam tikromis prielaidomis apie tikslo funkciją ir remiasi informacija apie tikslo funkcijos išvestines. Tačiau tokios prielaidos ne visada adekvačios sprendžiamam uždaviniui. Jei tikslo funkcija nėra glotni, tai jos hiperpaviršiaus nereguliarumai gali užgošti esmines funkcijos reikšmių mažėjimo kryptis: vaizdžiai kalbant, penelyg pasikliaudami lokalia informacija (išvestinėmis), galime įstrigti artimiausioje baloje vietoj judėję kalno šlaitu žemyn.

Paieškos metodai realizuoja įvairias labai skirtingas idėjas, todėl juos nelengva apibrėžti kaip metodų klasę. Kai kurie metodai modeliuoja gamtos reiškinius, gyvūnų elgseną ar net žmonių euristines sprendinio paieškos strategijas. Gal būt, labiausiai charakteringa šių metodų savybė ta, kad jiems pagrįsti nėra sudaromi kiekvieno žingsnio lokalūs tikslo funkcijos modeliai.

Vienas iš metodų, priskiriamų mokslo klasikų palikimui, yra vadinamas **Gauss-Seidel** vardu. Šio metodo kiekvienoje iteracijoje ieškoma tikslo funkcijos reikšmės vieno iš kintamųjų atžvilgiu kitiems esant fiksuotiems. Dažnai šis metodas vadinams pakoordinatine minimizacija. Jo idėja labai paprasta: iš pradinio taško judama pagal pirmąją koordinatę x_1 kol pasiekiamas vienmatis minimumas pagal x_1. Iš gautojo taško judama pagal antrąją koordinatę kol pasiekiamas minimumas pagal x_2, ir tt. Formaliai metodą galima aprašyti šitaip

$$\begin{aligned} X_{k+1} &= (x_{k1}, x_{k2}, ..., x_{ki}, x_{k+1\ i+1}, x_{k\ i+2}, ..., x_{kn}), \\ x_{k+1\ i+1} &= \arg\min_t f(x_{k1}, x_{k2}, ..., x_{ki}, t, x_{k\ i+2}, ..., x_{kn}). \end{aligned}$$

Galima įrodyti, kad metodas konverguoja į globalų minimumą, jei tikslo funkcija iškila, o vieno kintamojo funkcija, gaunama iš $f(X)$ fiksavus $n-1$ kintamojo reikšmes, griežtai iškila. Tačiau metodas gali būti efektyvus ir sudėtingesnių tikslo funkcijų atveju. Metodas, pavyzdžiui, efektyvus daugiaekstremėms funkcijoms, jei

$$f(X) = \sum_{i=1}^{n} f_i(x_i),$$

ir turimas efektyvus vienmatis globalinės minimizacijos metodas funkcijoms $f_i(\cdot)$.

Daugelyje praktinių uždavinių pasiteisino taip vadinamas **deformuojamo simplekso metodas** [9]. Atkreipsime dėmesį į tai, kad šis metodas vadinamas panašiu vardu kaip ir garsus tiesinio programavimo uždavinių sprendimo metodas. Bet tuo jų panašumas ir baigiasi. Vis dėlto, vengiant skirtingus metodus vadinti panašiais vardais, nagrinėjamasis metodas dažnai vadinamas jo autorių Nelder - Meed vardu.

Metodo idėją paaiškinsime specialiu taisyklingo simplekso atveju. Sudaromas taisyklingas $n+1$ viršūnės daugiakampis (simpleksas), kurio centras yra pasirinktasis pradinis taškas. Kraštinės ilgį rekomenduotina parinkti taip, kad jis būtų artimas nuspėjamam pradinio taško atstumui iki minimumo taško. Simplekso viršūnėse apskaičiuojamos tikslo funkcijos reikšmės. Naujasis taškas yra viršūnės, kurioje tikslo funkcijos reikšmė didžiausia, atspindys prieš ją esančios sienos atžvilgiu. 2.9 paveiksle pateiktas dviejų kintamųjų pavyzdys; čia pradinis simpleksas yra trikampis kurio 1,2,3 viršūnėse apskaičiuotos tikslo funkcijos reikšmės. "Blogiausia" iš jų yra 1. Naujasis 4 taškas gaunamas kaip 1 viršūnės atspindys trikampio kraštinės, jungiančios 2 ir 3 viršūnes, atžvilgiu. Tikslo funkcijos reikšmė 4 taške geresnė negu 1 taške, taigi žingsnis sėkmingas. Naująjį simpleksą sudaro 2,3,4 viršūnės. "Blogiausia" 2 viršūnė atspindima kraštinės 3-4 atžvilgiu. 5 viršūnė yra blogiausia naujajame simplekse 3,4,5. Jos atspindys sug-

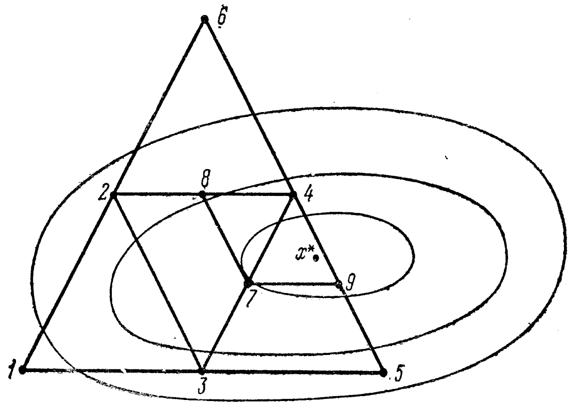

Figure 2.9: Simplekso algoritmo iliustracija

rąžina į simpleksą 2,3,4. Todėl jame galima pabandyti atspindėti 3 viršūnę. Bet ir šis žingsnis nesėkmingas. Kai visų viršūnių atspindžiai nepagerina tikslo funkcijos, simpleksas mažinamas, pavyzdžiui, dalinant kraštinę pusiau. Iš 2,3,4 tokiu būdu gaunamas 4,7,8. Pastaroji operacija dažnai vadinama redukcija. Po to vėl bandomi viršūnių atspindžiai. Sustojama, kai simplekso kraštinė tampa trumpesnė už pasirinktą toleranciją ε. Atskiri metodo žingsniai pagrindžiami euristiniais samprotavimais. Galima pasiūlyti įvairių modifikacijų, kurios atrodys ne mažiau pagrįstos. Pavyzdžiui, gavus simpleksą 3,4,5, kurio visų viršūnių atspindžiai nesėkmingi, redukuoti jį, o ne grįžti į 2,3,4. Galimos ir kitos modifikacijos, kuriose neišlaikomas simplekso taisyklingumas.

Praktiniuose uždaviniuose gan dažnai kintamųjų masteliai yra labai skirtingi. Kitaip sakant, tokio pat dydžio skirtingų kintamųjų pokyčiai duoda labai skirtingus tikslo funkcijos reikšmių pokyčius. Taisyklingas simpleksas nėra tokia geometrinė figūra, kuri gerai tiktų prie tokio nesimetriško tikslo funkcijos hiperpaviršiaus. Todėl buvo pasiūlyta paieškos procese keisti simpleksą taip, kad jis greičiau judėtų tikslo funkcijos hiperpaviršiumi link minimumo taško, prisitaikydamas prie hiperpaviršiaus savybių.

Deformuojamo simplekso metode naudojamos tempimo ir suspaudimo operacijos. Tempiama ta kryptimi, kuria vyksta gerėjimas. Jei atspindys toks sėkmingas, kad naujoji funkcijos reikšmė geresnė net už geriausią iki tol žinotą, galima daryti išvadą, kad užčiuopta labai gera judėjimo kryptis. Todėl ta kryptimi žengiama toliau, ištempiant ta kryptimi gautąjį simpleksą. Jei atspindys nesėkmingas, tas nebūtinai reiškia, kad kryptis bloga. Gali būti, kad žingsnis per ilgas. Todėl nebūtina atsisakyti šios tikėtinai geros krypties, o pabandyti simpleksą, gautą atspindint blogiausią viršūnę, suspausti atlikto žingsnio kryptimi. Tiek tempimo, tiek suspaudimo operacijos priklauso nuo empiriškai parinktų parametrų [9]. Patyręs vartotojas gali atsisakyti siūlomų parametrų reikšmių ir jas parinkti pagal savo uždavinio ypatybes.

Literatūroje aprašyta nemažai atvejų, kai deformuojamo simplekso metodu buvo sėkmingai išspręsti uždaviniai, kuriems dėl vienokių ar kitokių priežasčių netiko nusileidimo metodai.

Nemažą paieškos metodų poklasę sudaro **atsitiktinės paieškos metodai**. Vienas iš paprasčiausių metodų yra atsitiktinis taškų generavimas su tolygiu skirstiniu leistinojoje aibėje. Sugeneruotuose taškuose apskaičiuojamos tikslo funkcijos reikšmės ir išrenkamas taškas su mažiausia tikslo funkcijos reikšme. Ypač lengvai metodas realizuojamas, kai leistinoji sritis yra hiperstačiakampis. Metodo konvergavimas (tikimybine prasme) įrodomas esant minimalioms prielaidoms apie tikslo funkciją. Tačiau metodo teorinis konvergavimo greitis mažas, o praktiškai metodo efektyvumas dažniausiai nepakankamas. Tas ir natūralu, kadangi metodas *pasyvus,* t.y. einamasis taškas parenkamas nepriklausomai nuo anksčiau apskaičiuotų tikslo funkcijos reikšmių. Tolygi atsitiktinė paieška dažniausiai naudojama kombinuojant ją su kitais metodais.

Tam kad nustatytume greičiausio funkcijos kitimo kryptį, turime apskaičiuoti funkcijos gradientą. Tas reikalauja tam tikrų skaičiavimo laiko sąnaudų. Kartais tokios sąnaudos gali neatsipirkti ta prasme, kad žingsnis "pigiau" nustatyta, nors ir "prastesne", kryptimi gali būti santykinai "naudingesnis". Čia mes apsiribosime šiuo neformaliu paaiškinimu, tikėdamiesi, kad kabutėse parašyti žodžiai skaitytojui iš esmės suprantami. Pasirinkę tam tikrą žingsnio ilgį ρ, asitiktinai su tolygiu skirstiniu ant sferos $S(X_i, \rho)$ paviršiaus generuojame tašką, ir jame apskaičiuojame tikslo funkcijos reikšmę. Jei gautoji funkcijos reikšmė mažesnė, naujas taškas tampa X_{i+1}. Jei pagerėjimo nėra, $X_{i+1} = X_i$. Tikimybė gauti geresnę funkcijos reikšmę apytiksliai lygi 0.5, kai funkcija nagrinėjamojoje srityje $S(X_i, \rho)$ artima tiesinei. Tokia prielaida priimtina, kai ρ pakankamai mažas ir X_i pakankami toli nuo minimumo taško. Buvo pasiūlytos įvairios metodo modifikacijos, kuriomis buvo siekiama padidinti sėkmingo žingsnio tikimybę esant kitokioms prielaidoms apie funkciją, pvz., kai einamasis taškas yra "griovio" dugne.

Paminėsime vieną iš pastarojo metodo modifikacijų, kuri panaši į gradientinį nusileidimą tiek, kad net vadinama **stochastinio gradiento metodu**. Generuokime ant sferos ne vieną, o N taškų. Visuose sugeneruotuose taškuose apskaičiuokime tikslo funkcijos reikšmes ir išrinkime geriausią. Geriausias taškas tampa X_{i+1}. Kai sferos radiusas pakankamai mažas, o N pakankamai didelis, $X_{i+1} - X_i$ kryptis yra artima tikslo funkcijos antigradiento taške X_i krypčiai.

Pastaruoju metu ypač popoliarūs taip vadinami **evoliuciniai metodai** [17], kuriuose yra daug atsitiktinės paieškos elementų. Šiuose metoduose modeliuojamas populiacijų dauginimasis ir natūralioji individų atranka. Suprantama, kad tokie modeliai tik labai paviršutiniškai modeliuoja gyvosios gamtos evoliucinius procesus, bet jų pagrindu buvo sukurti įdomūs ir daugeliu atvejų pasiteisinę optimizavimo metodai. Kai evoliuciniai metodai skirti diskrečiosios optimizacijos uždaviniams spręsti, juos paprastai vadina genetiniais algoritmais, o kai skirti uždaviniams su tolydžiais kintamaisiais, tai juos vadina evoliucinėmis strategijomis.

Evoliucinės strategijos inicializacijai reikalinga pradinė populiacija. Ja gali būti atsitiktinai su tolygiu skirstiniu leistinojoje srityje sugeneruoti taškai. Juose apskaičiuojamos tikslo funkcijos reikšmės, kurios apsprendžia "individų" prisitaikymą prie aplinkos. Geriau prisitaikę "individai" turi didesnę tikimybę palikti palikuonis, kurie yra kažkiek panašūs į savo "tėvus". Modeliuojant mutacijas, prie apskaičiuoto (palikuonio) vektoriaus pridedamas atsitiktinis (mutacijų) vektorius. Tėvų ir vaikų populiacijoje modeliuojamas išmirimas, kuris priklauso nuo individo amžiaus ir prisitaikymo prie aplinkos. Gaunama nauja karta, kurioje vėl modeliuojamas reprodukcijos procesas.

Šitokiu būdu generuojami taškai labiau koncentruojasi leistinosios srities poaibiuose, kuriuose tikslo funkcijos reikšmės mažesnės. Visas taškų klasteris juda link geresnių tikslo funkcijos reikšmių. Čia aprašytos tik pačios bendriausios evoliucinių metodų idėjos. Jų realizacijų, besiskiriančių selekcijos, reprodukcijos, mutacijos, išmirimo modeliais, yra gana daug.

Paieškos metodai pasižymi nelokaliu ieškojimo būdu, t.y. jų randamas sprendinys yra nebūtinai artimiausias pradiniam taškui lokalusis minimumas. Šiuo požiūriu paieškos metodai skiriasi nuo nusileidimo metodų. Tačiau ne visi paieškos metodai orientuoti globalaus minimumo paieškai. Tam skirta ištisa globalios minimizacijos metodų klasė.

2.11 Globali optimizacija

Globalaus minimumo apibrėžimas nėra konstruktyvus. Nėra žinomos ir globalaus minimumo sąlygos, kurios palengvintų paiešką. Kadangi globalus minimumas yra ir lokalus, tai jis gali būti surastas lokalaus nusileidimo metodu iš pradinio taško, priklausančio globalaus minimumo pritraukimo sričiai. Dažnai bandoma tokį pradinį tašką surasti atsitiktinės paieškos metodu. Jei taškai generuojami atsitiktinai su tolygiu skirstiniu leistinojoje srityje, tai vieno atsitiktinio pataikymo į tokią sritį tikimybė lygi $p = G/V$, čia G ir V yra globalaus minimumo pritraukimo srities ir leistinosios srities hipertūriai. Tikimybė, kad bent vienas iš N bandymų pataikys į globalaus minimumo pritraukimo sritį lygi

$$p_N = 1 - (1 - p)^N.$$

Parenkant N pakankamai didelį (tuo didesnį, kuo p mažesnis), p_N teoriškai gali būti padarytas pakankamai artimas vienetui. Tačiau didesnio kintamųjų skaičiaus uždaviniuose p gali būti labai mažas, pavyzdžiui, $p < 10^{-10}$, jei dešimties kintamųjų atveju pagal kiekvieną kintamąjį ieškomoji sritis sudarytų mažiau negu 10% leistinosios srities. Tada N būtų nepriimtinai didelis. Todėl daugeliui praktinių uždavinių šis paprastas ir universalus metodas yra nepakankamai efektyvus.

Panašiai vertininta ir kita paprasta globalaus minimumo paieškos idėja vadinama atsitiktine lokalių minimumų peržiūra, kai išrenkamas geriausias iš lokalių minimumų, gautų lokalaus minimizavimo algoritmu iš atsitiktinai sugeneruotų pradinių taškų. Lokali minimizacija dažnai vykdoma ne nusileidimo, o paieškos metodais. Realizacija labai paprasta, užtenka turėti lokalaus minimizavimo metodą, kuris kombinuojamas su pradinių taškų generavimu. Efektyvumo požiūriu ši paprasta idėja ne visada patraukli. Viena - kelių lokalių nusileidimų rezultatu gali būti tas pats lokalaus minimumo taškas, ypač tada, kai tas lokalusis minimumas turi didelę pritraukimo sritį. Kitas probleminis atvejis - smulkus tikslo funkcijos hiperpaviršiaus bangavimas, dėl kurio lokalaus nusileidimo metodai gali nedaug tepajudėti iš sugeneruotų pradinių taškų. Kaip ir aukščiau išanalizuotu atveju, tikimybė surasti globalų minimumą per praktiškai priimtiną laiką gali būti labai maža. Tačiau šie trūkumai nėra esminiai, jei tikslo funkcija turi tik keletą lokalių minimumų su panašaus dydžio pritraukimo sritimis. Tada šio paprasto metodo panaudojimas tikrai pasiteisina. Nemažai sudėtingesnių metodų gali būti interpretuojami kaip jo patobulinimai.

Atsitiktinis taškų generavimas naudojamas kuriant globalios paieškos metodus, modeliuojančius gyvos ir negyvos gamtos procesus. Daug dėmesio mokslinėje literatūroje buvo skirta **modeliuojamo atkaitinimo** (angliškai *simulated annealing)* metodui, kuris vis dėlto, dažniausiai naudojamas diskrečiosios optimizacijos uždaviniams spręsti [1].

Klasterinis metodas pradedamas panašiai kaip ir lokalių minimumų perrinkimas: sugeneruojama pradinių taškų aibė. Lokalūs nusileidimai pradedami iš visų pardinių taškų padarant keletą žingsnių link minimumo. Gauti taškai analizuojami klasterizavimo metodu, t.y. ieškoma taškų sankaupų (klasterių). Jei aptinkami klasteriai, tai daroma išvada, kad keli nusileidimo procesai juda link to paties minimumo. Iš klasterio išrenkamas taškas su geriausia tikslo funkcijos reikšme. Lokalus nusileidimas tęsiamas tik iš išrinktojo taško, o kiti lokalaus nusileidimo procesai nutraukiami. Po klaster - analizės gali būti sustabdytas tam tikras skaičius lokalaus nusileidimo procesų. Sugeneruojamas tam tikras skaičius naujų pradinių taškų. Keletas nusileidimo procesų žingsnių įvykdoma iš taškų, išlikusių po klasterizavimo, ir naujai sugeneruotų taškų. Vykdoma gautų taškų klaster - analizė ir tt. Metodo sustojimo sąlyga apibrėžiama euristiškai: sustojama, kai kelios iš eilės klaster - analizės nebeaptinka naujų klasterių. Lyginant su atsitiktine lokalių minimumų peržiūra šis metodas efektyvesnis, nes lokalūs nusileidimai į tą patį lokalųjį minimumą gan greitai aptinkami ir sustabdomi (išskyrus vieną). Kad būtų išvengta įstrigimų nereikšminguose lokaliuose minimumuose, naudojami ne nusileidimo, bet paieškos metodai. Teoriškai tirti šį metodą sunku dėl to, kad jis konstruojamas remiantis euristinėmis idėjomis, o ne matematinėmis prielaidomis apie tikslo funkciją. Tačiau eksperimentinis jo palyginimas su kitais metodais sprendžiant testinius uždavinius parodė aukštą šio metodo efektyvumą.

Klasterizacijos metodas ne išimtis - globaliojoje optimizacijoje labai daug euristikos. Vis dėlto, yra ir matematinėmis prielaidomis pagrįstų metodų. Panagrinėkime, atrodytų paprastą, kvadratinės tikslo funkcijos atvejį. Net šiuo atveju galimas daugiaekstremiškumas. Pavyzdžiui, kvadratinė neigiamai apibrėžta funkcija gali turėti daugybę lokalių minimumų ant leistinosios srities kraštų. Net paprasčiausiu vieno kintamojo atveju $f(x) = -x^2$, $-1 \leq x \leq 2$, uždavinys daugiaeksremis: intervalo galai yra lokalių minimumų taškai. Potencialių lokalaus minimumo taškų skaičius auga labai greitai, augant erdvės dimensijai n. Kadangi tikslo funkcija šiuo atveju aprašoma matematine formule, tai palyginti paprastai galima įvertinti galimus funkcijos reikšmių pokyčius ir atmesti tuos leistinosios srities poaibius, kuriuose tikrai negali būti minimumo. Panašiai sudaromi metodai uždaviniams su įgaubtomis (iškilomis į viršų) tikslo funkcijomis. Tokio pobūdžio uždaviniai nagrinėjami ir jų sprendimo metodai sudaromi tradiciniais matematiniais metodais [10].

Bendresniems uždaviniams tradicinis modelis yra **Lipshitz'o funkcijų klasė**, t.y. klasė funkcijų, patenkinančių nelygybę

$$|f(X_1) - f(X_2)| \leq L \cdot ||X_1 - X_2||,$$

čia L - Lipshitz'o konstanta, kurią reikia įvertinti konkrečiam (sprendžiamam) uždaviniui. Jei leistinajai sričiai priklausančiuose taškuose $X_i, i = 1, ..., N$, apskaičiuotos tikslo funkcijos reikšmės $y_i = f(X_i)$, tai galima apibrėžti apatinį funkcijos reikšmių rėžį

$$\varphi(X) = \max_{1 \leq i \leq N} f(X_i) - L \cdot ||X_i - X||.$$

Tikslo funkcijos globalusis minimumas negali būti pasiekiamas tuose leistinosios srities poaibiuose, kuriuose

$$\varphi(X) \geq f^+ = \min_{1 \leq i \leq N} f(X_i).$$

Pastaroji išvada iliustruota vieno kintamojo funkcijos atveju 2.10 paveiksle. Tie poaibiai, kuriuose globalaus minimumo nėra, gali būti toliau nebenagrinėjami, o tikslo funkcijos skaičiavimas planuojamas likusioje leistinosios srities dalyje; 2.10 paveiksle ši sritis pavaizduota užbrūkšniuotais pointervalais. Intuityviai natūralu (toks metodas turi ir matematinį pagrindimą) skaičiuoti tikslo funkcijos reikšmę tame taške, kuriame galėtų būti pasiekiama minimum minimorum $f(X)$ reikšmė

$$\begin{aligned} f^- &= \min_{X \in A} \varphi(X), \\ X_{i+1} &= \arg \min_{X \in A} \varphi(X). \end{aligned} \tag{2.30}$$

Vieno kintamojo atveju tokį metodą palyginti nesunku realizuoti, kaip pavaizduota 2.10 paveiksle.

Tačiau bendru atveju tokio metodo algoritminis realizavimas bei praktinis panaudojimas susijęs su šitokiais sunkumais. Praktinių uždavinių tikslo funkcijos tik išimtinais atvejais yra vieno kintamojo. Tačiau $\varphi(X)$ minimizavimas daugiamatėje erdvėje yra labai sudėtingas daugiaekstremis uždavinys, gal būt, ne paprastesnis už tą, kurio sprendimui bus naudojamas sukurtasis metodas. Pastarojo efektyvumas priklauso nuo konstantos L įverčio. Jei L per mažas, gali būti atmestas poaibis, kuriam priklauso globalaus minimumo taškas. Jei L labai didelis, vienu funkcijos reikšmės apskaičiavimu atmetama tik maža leistinosios srities dalis.

Globalioji optimizacija žymiai sudėtingesnė už lokaliąją. Nepaisant praktinio globaliosios optimizacijos uždavinių aktualumo,

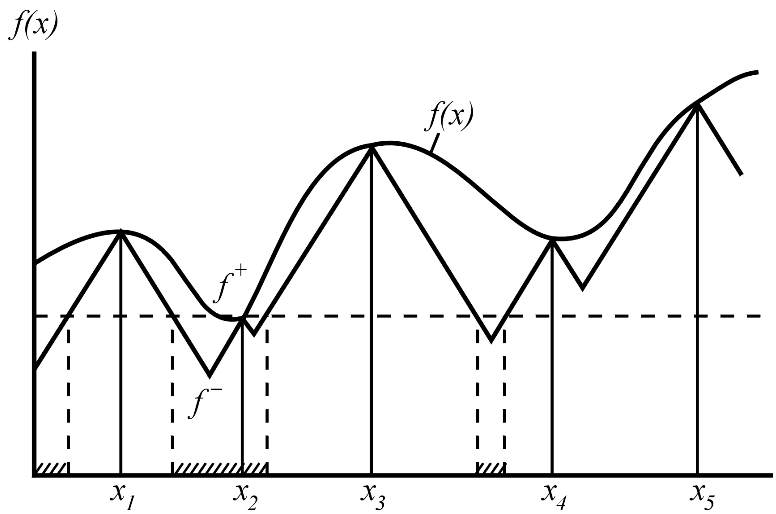

Figure 2.10: Globalios optimizacijos metodo pagrįsto apatiniu funkcijos reikšmių rėžiu, priklausančiu nuo Lipshitz'o konstantos, iliustracija

plačiai paplitusiuose komerciniuose optimizavimo paketuose nėra tam skirtų metodų. Su įvairiomis idėjomis, teorijomis ir metodais galima susipažinti pasiskaičius [20] knygą. Lietuvoje aktyviai dirbama statistiniais modeliais pagrįstoje globaliosios optimizacijos srityje. Bajesinė globalinės optimizacijos teorija ir iš jos išplaukiantys metodais pateikti šios teorijos pradininko J.Mockaus monografijoje [14].

2.12 Užduotys ir kontroliniai klausimai

1. Įrodykite (2.10).
2. Raskite dviejų kintamųjų funkcijų

$$\begin{aligned} f(X) &= x_1^2 + \sin^2 x_2, \\ f(X) &= (x_1 - 2)^2 \cdot \cos^3(x_2), \end{aligned}$$

lokaliųjų minimumų taškus.

3. Šviesa nehomogeninėje aplinkoje sklinda tokiu keliu, kuris nueinamas per trumpiausią laiką. Išnagrinėkite šviesos sklidimą plokštumoje. Sakysim vienoje pusplokštumėje šviesos sklidimo greitis lygus v_1, o kitoje v_2. Šviesos šaltinis ir stebėjimo taškas yra skirtingose pusplokštumėse. Remdamiesi šviesos sklidimo laiko minimumo sąlygomis išveskite formulę, išreikiančią Snellius dėsnį.

4. Išnagrinėkite vienmačio gradientinio metodo su pastoviu žingsnio daugikliu

$$x_{k+1} = x_k - f'(x_k)$$

konvergavimą, kai

$$f(x) = \begin{cases} \frac{3(1-x)^2}{4} - 2(1-x), \ x > 1, \\ \frac{3(1+x)^2}{4} - 2(1+x), \ x < -1, \\ x^2 - 1, \ -1 \leq x \leq 1, \end{cases}$$

o pradinis taškas lygus $x_0 = 2$.

5. Nubraižykite funkcijos (2.7) grafiką.

6. Į kokį tašką bus pereita, atlikus vieną greičiausio nusileidimo metodo iteraciją, kai $f(X)$ yra kvadratinė dviejų kintamųjų funkcija (2.9), ir

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 10 \end{pmatrix}, \ B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ X_i = \begin{pmatrix} 5 \\ 5 \end{pmatrix}.$$

7. Sakysim, Q yra simetrinė matrica. Įrodykite, kad tikriniai Q vektoriai, atitinkantys skirtingas tikrines reikšmes yra Q jungtiniai.

8. Kokiu atveju simplekso metodo žingsnio kryptis artima antigradiento krypčiai.

9. Panagrinėkite metodo (2.30) dvimatę versiją.

NETIESINIS PROGRAMAVIMAS

3.1 Minimumo sąlygos

Kiekvienas inžinierius žino, kiek daug visokių sąlygų bei ribojimų būtina įvertinti projektuojant. Techniniuose reikalavimuose paprastai būna nurodyta daug ribinių projektuojamojo objekto charakteristikų reikšmių, reikalavimai konstrukcinėms medžiagoms ir pan. Pavyzdžiui, masė ir kaina turi būti mažesnės, o patikimumas didesnis už tam tikras nustatytas reikšmes. Taip pat būtina atsižvelgti į projektuojamojo objekto charakteristikas siejančius gamtos dėsnius. Todėl techniniai optimizavimo uždaviniai dažniausiai būna su ribojimais. Panašiai formuluojami ir ekonominio planavimo uždaviniai, nes tenka atsižvelgti ir į turimas atsargas, ir į biudžeto ribotumą, ir į reikalavimus terminams, ir tt. Formuluojant uždavinį matematiškai, **ribojimais** vadinamos leistinąją sritį apibrėžiančios lygybės ir nelygybės, pavyzdžiui,

$$
\begin{aligned}
\min \quad f(X),\ X &= (x_1, ..., x_n), \\
g_i(X) &\leq 0,\ i = 1, ..., k, \\
g_i(X) &= 0,\ i = k+1, ..., m,
\end{aligned}
\tag{3.1}
$$

reiškia funkcijos $f(X)$ minimizavimą leistinoje srityje, apibrėžtoje funkcijų $g_i(X)$ neteigiamumo arba lygybės nuliui reikalavimais. Patenkinantis ribojimus taškas X vadinamas (optimizavimo uždavinio) **leistinu sprendiniu**. Minimumo taškas vadinamas **optimaliu sprendiniu**, arba trumpiau, **sprendiniu**. Jei bent viena iš (3.1) funkcijų yra netiesinė, toks optimizavimo uždavinys vadinamas netiesinio programavimo uždaviniu, jei jos visos iškilos, - iškilaus programavimo uždaviniu. Ribojimai klasifikuojami į tiesinius ir netiesinius, žiūrint kokios funkcijos $g_i(X)$. Pirmieji k ribojimų vadinami nelygybiniais, kiti - lygybiniais. Uždavinių su ribojimais lokalaus optimizavimo metodai sudaromi iš metodų be ribojimų, juos modifikuojant taip, kad būtų įvertinama ne tik funkcijos mažėjimo kryptis, bet ir ribojimai.

Antrajame skyriuje pateikta būtina minimumo sąlyga uždaviniams be ribojimų, išreikšta funkcijos gradiento lygybe nuliui. Esant ribojimams, funkcijos gradientas gali būti nelygus nuliui visoje leistinoje srityje, kaip parodyta 3.11 paveiksle. Tada minimumas pasiekiamas taške X_*, esančiame ant leistinosios srities krašto. Būtiną minimumo sąlygą galima paaiškinti geometriškai: jei taške X_* pasiekiamas lokalusis minimumas, tai leistinoje srityje jokia kryptimi negalima sumažinti tikslo funkcijos reikšmės. Todėl tikslo funkcijos gradientas taške X_* turi būti ortogonalus leistinosios srities liečiamajai hiperplokštumai.

Nelygybiniai ribojimai, kurie taške X patenkinami kaip lygybės, vadinami aktyviais; jų indeksų aibę pažymėsime $I^*(X)$, ir $I(X) = I^*(X) \cup \{k+1, ..., m\}$.

6 teorema. Jei X_* yra lokalaus minimumo taškas, tai egzis-

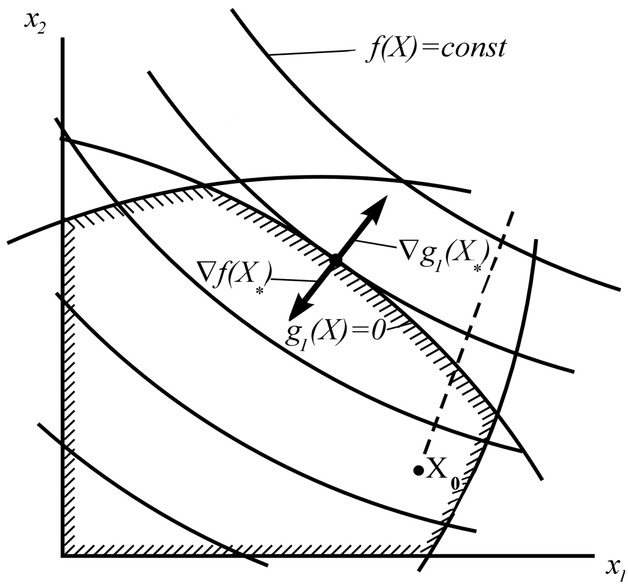

Figure 3.11: Būtinų minimumo sąlygų iliustracija

tuoja tokie visi kartu nelygūs nuliui daugikliai $\lambda_i, i = 0, ..., m$, kad $\lambda_i \geq 0,\ i = 1, ..., k$,

$$\lambda_0\ \nabla f(X_*)\ +\ \sum_{i=1}^{m} \lambda_i\ \nabla g_i(X_*)\ =\ 0,$$

$$\lambda_i\ g_i(X_*)\ =\ 0, \quad i = 1, ..., k. \tag{3.2}$$

Įrodymas. Sakysim, kad $\nabla g_i(X_*)$, $i = k+1, ..., m$, tiesiškai nepriklausomi. Priešingu atveju teoremos teiginys įrodomas trivialiai, paėmus $\lambda_i = 0,\ i \leq k$.

Sudarysime dvi aibes $\Phi, \Psi \subset R^{m+1}$:

$$\Phi = \{\Theta : \Theta \in R^{m+1},\ \theta_0 = \nabla f(X_*) \cdot S,$$

$$\theta_i = \nabla g_i(X_*) \cdot S,\ i \in I(X_*),\ S \in R^n\},$$

$$\Psi = \{\Theta : \Theta \in R^{m+1},\ \theta_0 < 0,\ \theta_i < 0,$$

$$i \in I^*(X_*),\ \theta_i = 0,\ i = k+1, ..., m\}.$$

Įrodysime, kad šios aibės nesikerta. Sakysim, kad yra priešingai. Tada turėtų egzistuoti toks $S \in R^n$, kad būtų patenkintos šitokios lygybės ir nelygybės

$$\begin{array}{rcl} \nabla f(X_*) \cdot S & < & 0, \\ \nabla g_i(X_*) \cdot S & < & 0,\ i \in I^*(X_*), \\ \nabla g_i(X_*) \cdot S & = & 0,\ i = k+1, ..., m. \end{array} \tag{3.3}$$

Įrodysime, kad tada kryptimi S, rodančia į leistinosios srities vidų, tikslo funkcijos reišmės mažėja. Kadangi S yra paviršiaus $g_i(X) = 0, i = k+1, ..., m$, liečiamoji kryptis taške X_*, tai egzistuoja

$$X(\delta) = X_* + \delta S + o(\delta),$$

patenkinantis ribojimus

$$g_i(X(\delta)) = 0,\ i\ = k+1, ..., m.$$

Iš (3.3) išplaukia, kad pakankamai mažam δ

$$f(X(\delta)) = f(X_*) + \delta \cdot \nabla f(X_*) \cdot S + o(\delta) < f(X_*),$$

ir panašiai įrodoma, kad $g_i(X(\delta)) < 0,\ i \in I^*(X_*)$. Bet tada $X(\delta)$ patenkintų visus ribojimus, ir galiotų nelygybė $f(X(\delta)) < f(X_*)$, reiškianti, kad X_* nėra lokalaus minimumo tašku. Taigi, prielaida, kad aibės Φ ir Ψ kertasi, neteisinga.

Galima įsitikinti, kad aibės Φ ir Ψ iškilos. Nesikertančios iškilos aibės gali būti atskirtos hiperplokštuma. Kadangi abiem aibėm Φ ir Ψ priklauso kiek norima mažos taško $0 \in R^{m+1}$ aplinkos taškų, tai 0 turi priklausyti skiriančiai hiperplokštumai. Tas reiškia, kad egzistuoja ne visi lygūs nuliui $\lambda_i, i = 0, ..., m$, tokie, kad

$$\sum_{i=0}^{m} \lambda_i \theta_i \geq 0,\ \ \Theta \in \Phi,\ \text{ir} \sum_{i=0}^{m} \lambda_i \theta_i < 0,\ \ \Theta \in \Psi.$$

Pirmoji nelygybė gali būtų užrašyta šitaip

$$\lambda_0 \nabla f(X_*) \cdot S + \sum_{i \in I(X_*)} \lambda_i \nabla g_i(X_*) \cdot S + \sum_{i \notin I(X_*)} \lambda_i \theta_i \geq 0, \tag{3.4}$$

čia S bet koks R^n erdvės vektorius. Kadangi $\theta_i,\ i \notin I(X_*)$, gali būti tiek teigiami, tiek neigiami, bei kiek norima dideli absoliutiniu didumu, tai nelygybė (3.4) negalėtų būti patenkinama visiems $\theta_i,\ i \notin I(X_*)$, jei nebūtų patenkinta lygybė $\lambda_i = 0,\ i \notin I(X_*)$. Kadangi (3.4) turi būti patenkinama ir su S, ir su $-S$, tai nelygybės ženklas $\geq$ turi būti pakeistas lygybės ženklu. Todėl iš (3.4) išplaukia tokios lygybės

$$\begin{aligned} \lambda_0 \, \nabla f(X_*) \; + \; \sum_{i=1}^{m} \lambda_i \, \nabla g_i(X_*) \quad &= \quad 0, \\ \lambda_i \quad &= \quad 0, \, i \notin I(X_*). \end{aligned}$$

Kadangi visiems $\theta \in \Psi$ teisinga nelygybė $\theta_i < 0$, kai $i = 0$ ir $i \in I^*(X_*)$, bei galioja nelygybė

$$\lambda_0 \theta_* + \sum_{i \in I^*(X_*)} \lambda_i \theta_i < 0,$$

tai $\lambda_0 \geq 0$ ir $\lambda_i \geq 0$, $i \in I^*(X_*)$. Gautą rezultatą, atsižvelgdami į tai, kad $g_i(X_*) = 0$, $i \in I(X_*)$, galima užrašyti vadinamąja **F.John** forma, kuri pateikta teoremos formulavime (3.2).

Svarbus šios teoremos atvejis $\lambda_0 \neq 0$ įrodomas esant papildomai *reguliarumo prielaidai,* pavyzdžiui, $\nabla g_i(X_*)$, $i \in I(X_*)$, yra tiesiškai nepriklausomi. Tada teoremos formulavime galima rašyti $\lambda_0 = 1$, o gautos būtinos minimumo sąlygos vadinamos **Karush-Kuhn-Tucker** sąlygomis

$$\begin{aligned} \nabla f(X_*) \; + \; \sum_{i=1}^{m} \lambda_i \, \nabla g_i(X_*) \; &= \; 0, \\ \lambda_i \; g_i(X_*) \; &= \; 0, \quad i = 1, ..., k. \end{aligned} \tag{3.5}$$

Mokslinėje literatūroje vartojama: santrumpa KKT sąlygos. Reguliarumo sąlygos, pavyzdžiui, yra nepatenkinamos uždavinio

$$\min \, (x_1 - 10)^2 + x_2^2,$$

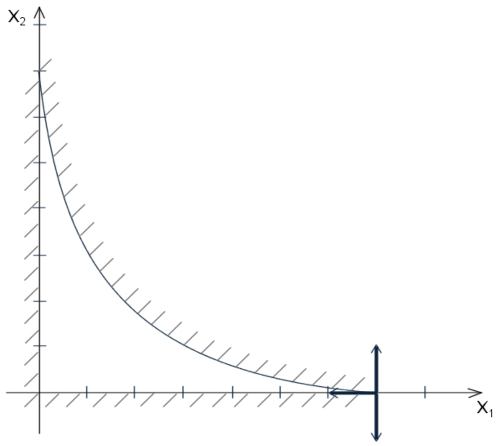

Figure 3.12: Nepatenkintų reguliarumo sąlygų pavyzdys

$$\begin{array}{rcl} x_2 & \leq & 7 - \sqrt{-x_1^2 + 14x_1}, \\ 0 & \leq & x_1 \leq 7, \\ 0 & \leq & x_2 \leq 7, \end{array}$$

minimumo taške (0, 7), nes abiejų aktyvių ribojimų gradientai yra statmeni abscisų ašiai, kuria yra nukreiptas tikslo funkcijos gradientas (žiūr., 3.12 paveikslą).

Daugelis netiesinio programavimo metodų susiję su Lagrange'o funkcija

$$L(X, \Lambda) = f(X) + \sum \lambda_i \cdot g_i(X).$$

Jei Lagrange'o funkcija diferencijuojama, tai ką tik įrodytas minimumo sąlygas (priėmus reguliarumo prielaidą) galima užrašyti naudojantis Lagrange'o funkcija

$$\begin{aligned} \nabla_X L(X,\Lambda) &= 0, \\ \lambda_i &\geq 0, \; i \in I^*(X_*), \\ \lambda_i \; g_i(X_*) &= 0, \;\; i = 1,...,k. \end{aligned}$$

Lagrange'o funkcijos terminais galima suformuluoti būtinas minimumo sąlygas, kurios nereikalauja, kad funkcijos $f(X)$, $g_i(X)$, $i = 1,...,m$, būtų diferencijuojamos. Tarkime, (3.1) uždavinys turi **tik nelygybinius** ribojimus, visos šio uždavinio **funkcijos iškilos** ir ribojimai tenkina reguliarumo sąlygą: egzistuoja toks taškas X, kuriame (3.1) nelygybės tenkinamos griežtai: $g_i(X) > 0, i = 1,...,k$. Tada lokalaus minimumo taškas X_* lemia Lagrange'o funkcijos balno tašką, t. y. egzistuoja tokie $\lambda_{0\,i} \geq 0$, $i = 1,...,k$, kad

$$L(X_*,\Lambda) \leq L(X_*,\Lambda_0) \leq L(X,\Lambda_0) \tag{3.6}$$

su visais $\lambda_i \geq 0$, $X \in R^n$. Tas rodo, kad lokalaus minimumo su sudėtingais ribojimais ieškojimas gali būti pakeistas Lagrange'o funkcijos balno taško ieškojimu esant paprastiems ribojimams $\lambda_i \geq 0$. Šios būtinos minimumo sąlygos vadinamos (Kuhn – Tucker) teorema apie balno tašką.

Uždavinių su ribojimais būtinos minimumo sąlygos tiriamos jau seniai. Vienas iš svarbiausių dar ir šiandien yra Lagrange'o pasiūlytas principas - perkelti ribojimus į tikslo funkciją. Kuhn – Tucker sąlygos išvestos remiasi ta pačia idėja. Atrodytų, pakankamos minimumo sąlygos svarbesnės už būtinas. Tačiau jas galima suformuluoti tik tuo atveju, kai visos (3.1) funkcijos tenkina griežtus reikalavimus, kurių praktiniuose uždaviniuose dažniausiai neįmanoma patikrinti. Be to, jos gerokai sudėtingesnės. Todėl stengiamasi suformuluoti praktiškai patikrinamas būtinas sąlygas, kaip galima artimesnes pakankamoms. Vienas iš svarbiausių teorinių netiesinio programavimo metodų pagrindimų yra įrodymas, kad pagal tuos metodus generuojamų taškų seka artėja prie taško, tenkinančio būtinas minimumo sąlygas.

3.2 Dualumas

Silpnesnio matematinio pasiruošimo skaitytojams šis paragrafas gali būti sunkiau suprantamas, bet jis nėra būtinas toliau dėstomai medžiagai studijuoti. Prie dualumo problemos bus dar sugrįžta, nagrinėjant paprastesnį tiesinio programavimo atvejį.

Nagrinėsime minimizavimo uždavinį, kurį vadinsime **pirminiu**

$$\mathbf{P}: \qquad \min f(X), \tag{3.7}$$

$$\begin{aligned} g_i(X) &\leq 0, \ i = 1, ..., k, \\ h_i(X) &= 0, \ i = 1, ..., m, \end{aligned}$$

čia lygybinių ir nelygybinių ribojimų funkcijas pažymėjome skirtingomis raidėmis, kadangi apie skirtingų tipų ribojimus bus daromos skirtingos prielaidos. Be to, darysim prielaidą, kad **P** uždavinio kintamųjų vektorius priklauso paprastos struktūros aibei $X \in A_0$, kuri nėra aprašoma ribojimų funkcijomis. Pavyzdžiui, $A_0 = \{X : x_i \geq 0, \ i = 1, ..., n\}$; dažnai A_0 tiesiog sutampa su R^n. Leistina aibė, kaip paprastai, žymima A.

P uždavinio Lagrange'o funkcija užrašoma šitaip

$$L(X, U, V) = f(X) + \sum_{i=1}^{k} u_i g(X) + \sum_{i=1}^{m} v_i h(X) =$$

$$= f(X) + UG(X) + VH(X).$$

Dualiu vadinamas maksimizavimo uždavinys

$$\mathbf{D}: \qquad \max \Theta(U,V), \tag{3.8}$$

$$U \geq 0,\ V \in R^m,$$

kurio tikslo funkcija apibrėžiama kaip Lagrange'o funkcijos minimalios reikšmės X atžvilgiu priklausomybė nuo U ir V

$$\Theta(U,V) = \inf_{X \in A_0} L(X,U,V);$$

čia parašytas simbolis inf apibrėžia analogišką minimui reikšmę, kai pastarasis neegzistuoja, pavyzdžiui, jei galima rasti norimai mažą funkcijos reikšmę, tai minimumas neegzistuoja, o infimumas lygus -∞.

Geometrinę dualaus uždavinio prasmę paaiškinsime pavyzdžiu

$$\min\ f(X),\ g(X) \leq 0,\ X \in A_0. \tag{3.9}$$

Pažymėkime $y = g(X)$, $z = f(X)$, ir Γ – vektoriaus (y,z) reikšmių aibę, kai $X \in A_0$. **P** uždavinio (3.9) minimumas atitinka aibės Γ tašką ant z ašies su minimalia z reikšme $f(\overline{X})$, kaip tas pavaizduota 3.13 paveiksle. **P** uždavinį (3.9) atitinka **D** uždavinys $\max_{u \geq 0} \Theta(u)$, kurio tikslo funkcija yra

$$\begin{aligned} \Theta(u) &= \min_{X \in A_0} f(X) + u \cdot g(X) = \\ &= \min\,(z + u \cdot y). \end{aligned}$$

Esant fiksuotai $u \geq 0$ reikšmei $\Theta(u)$ priklauso nuo dviejų parametrų z ir y, o pastaroji priklausomybė geometriškai reiškia tiesę

$$z = \Theta(u) - u \cdot y, \tag{3.10}$$

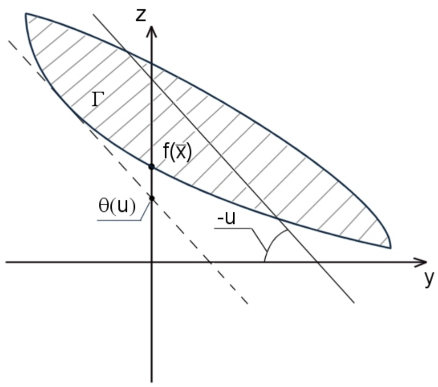

Figure 3.13: Dualaus uždavinio iliustracija

kuri yra žemiausiai z atžvilgiu einanti tiesė, turinti bent vieną bendrą tašką su Γ, t.y. aibės Γ apatinė lietėja su krypties koeficientu $-u$. $\Theta(u)$ reikšmė lygi z reikšmei, gaunamai susikirtus tiesei (3.10) su z ašimi. Dualus uždavinys: rasti $u \geq 0$ reikšmę $\overline{u}$, maksimizuojančią $\Theta(u)$. Geometriškai -$\overline{u}$ reikštų krypties koeficientą lietėjos, einančios per tašką $y = 0$, $z = f(\overline{X})$, vaizduojantį optimalų **P** uždavinio sprendinį.

7 teorema. Sakysim X yra leistinas **P** uždavinio (3.7) sprendinys, o (U, V) – leistinas **D** uždavinio (3.8) sprendinys. Tada galioja nelygybė

$$f(X) \geq \Theta(U, V). \tag{3.11}$$

Įrodymas. Pagal apibrėžimą

$$\begin{aligned}\Theta(U,V) &= \inf_{X\in A_0} f(X)+UG(X)+VH(X) \leq \\ &\leq f(X)+UG(X)+VH(X),\ \forall\, X\in A_0.\end{aligned}$$

Kadangi leistiniems sprendiniams galioja nelygybės $U \geq 0$, $G(X)\leq 0$, $H(X)=0$, tai $\Theta(U,V)\leq f(X)$, $\forall\, X\in A$, $U\geq 0$.

1 išvada.

$$\inf_{X\in A} f(X) \geq \sup_{U\geq 0,\, V\in R^m} \Theta(U,V). \tag{3.12}$$

2 išvada. Jei $f(\overline{X})=\Theta(\overline{U},\overline{V})$, $\overline{X}\in A$, $\overline{U}\geq 0$, tai $\overline{X}$ ir $(\overline{U},\overline{V})$ yra **P** ir **D** uždavinių sprendiniai.

3 išvada. Jei, $\inf_{X\in A} f(X)=-\infty$, tai ir $\Theta(U,V)=-\infty$, $\forall\, U\geq 0$. Jei $\sup_{U\geq 0}\Theta(U,V)=\infty$, tai **P** neturi leistinų sprendinių.

3.13 paveikslo pavyzdyje pirminio ir dualaus uždavinių optimalios reikšmės sutampa. Įrodytoji teorema teigia, kad bendru atveju galioja nelygybė (3.11) ir iš jos išplaukianti nelygybė (3.12), t.y. **P** uždavinio minimumas ne mažesnis už **D** uždavinio maksimumą. Kai jie nesutampa (pvz., 3.14 paveiksle), skirtumas vadinamas dualumo tarpu. Toliau pateikiamas pavyzdys [2], detaliau paaiškinantis dualumo tarpo sąvoką.

Pavyzdys. P: $\min f(X)=\min(-2x_1+x_2)$, $h(X)=x_1+x_2-3=0$, $A_0=\{(0,0),\,(0,4),\,(4,4),\,(4,0),(1,2),\,(2,1)\}$. Nesunku patikrinti, kad lygybinį ribojimą patenkina du aibės A_0 taškai, o minimumo taškas yra $\overline{X}=(2{,}1)$; $f(\overline{X})=-3$.

D: $\max\Theta(v)$, čia

$$\Theta(v)=\min_{X\in A_0}(-2x_1+x_2+v\cdot(x_1+x_2-3)).$$

Nesunku įsitikinti, kad

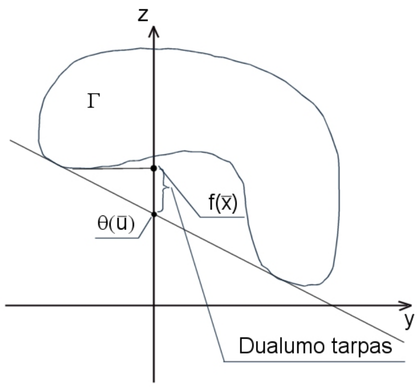

Figure 3.14: Dualumo tarpo pavyzdys

$$\Theta(v) = \begin{cases} -4+5v, & v \le -1, \\ -8+v, & -1 < v \le 2, \\ -3v, & v > 2, \end{cases}$$

todėl

$$\max \Theta(v) = \Theta(2) = -6.$$

Šiuo atveju dualumo tarpas lygus 3. Funkcijos $\Theta(v)$ grafikas pateiktas 3.15 paveiksle, o **P** ir **D** uždavinių optimalūs sprendiniai pavaizduoti 3.16 paveiksle.

Lema. Sakysim A_0 – netuščia aibė, $e(X)$, $g_i(X)$ – iškilos funkcijos, o $H(X)$ –afininė funkcija (t.y., $H(X) = TX - B$, čia T yra $m \times n$ matrica, $B \in R^m$).

Jei nelygybių sistema

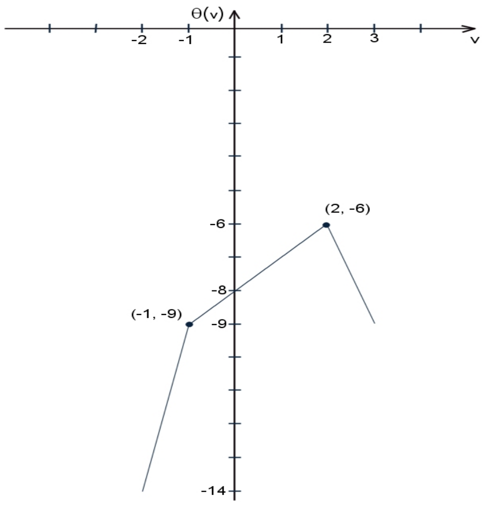

Figure 3.15: Dualaus uždavinio tikslo funkcijos grafikas

$$e(X) < 0,\ G(X) \leq 0,\ H(X) = 0,\ X \in A_0 \tag{3.13}$$

neturi sprendinio, tai $\forall\, X \in A_0$ nelygybių sistema

$$u_0 \cdot e(X) + U \cdot G(X) + V \cdot H(X) \geq 0 \tag{3.14}$$

turi sprendinį $(u_0, U, V) \neq 0$, $(u_0, U) \geq 0$.

Atvirkščiai, jei (3.14) turi sprendinį ir $u_0 > 0$, tai (3.13) neturi sprendinio.

Įrodymas. Remiantis tuo, kad funkcijos $e(X)$, $g_i(X)$ – iškilos, o $H(X)$ –afininė, nesunku įrodyti, kad aibė

$$\begin{aligned} \Psi &= \{(p, Q, R) : \exists\ X \in A_0, \\ p &> e(X),\ Q \geq G(X),\ R = H(X)\}, \end{aligned} \tag{3.15}$$

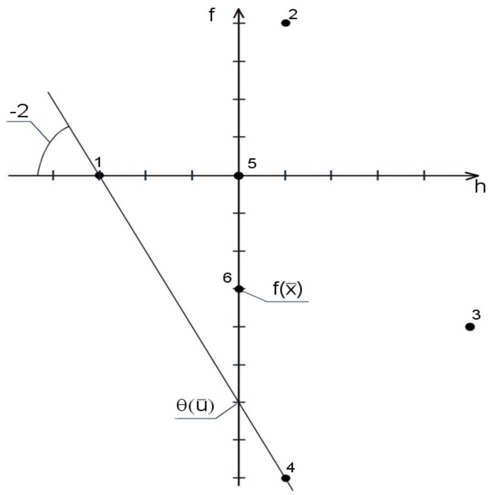

Figure 3.16: Dualumo tarpas lygus 3

yra iškila. Sakysim (3.13) neturi sprendinio. Tada $(0, 0, 0) \notin \Psi$, čia antrasis nulis reiškia k-matį nulinį vektorių, o trečiasis nulis m-matį nulinį vektorių. Kadangi egzistuoja hiperplokštuma, atskirianti iškilą aibę ir jai nepriklausantį tašką, tai egzistuoja toks nenulinis vektorius (u_0, U, V), kad visiems ribiniams aibės Ψ taškams (p, Q, R) patenkinama nelygybė

$$u_0 p + UQ + VR \geq 0. \tag{3.16}$$

Imkime kokį nors $X \in A_0$. Pagal aibės Ψ apibrėžimą p ir Q gali būti paimti kiek norima dideli. Todėl nelygybė (3.16) gali būti užtikrinta visiems aibės Ψ taškams tik tada, kai $u_0 \geq 0\,, U \geq 0$. Kadangi bet kuriam $X \in A_0$ $(e(X), G(X), H(X))$ yra ribinis aibės Ψ taškas, tai nelygybių sistema

$$u_0 \cdot e(X) + U \cdot G(X) + V \cdot H(X) \geq 0$$

patenkinta visiems $X \in A_0$, t.y. (u_0, U, V) yra teoremos sąlygas patenkinantis (3.14) sprendinys.

Sakysim (3.14) turi sprendinį $(u_0, U, V) \neq 0$, ir $u_0 > 0$, $U \geq 0$. Jei $X \in A_0$ toks, kad $G(X) \leq 0$, $H(X) = 0$, tai iš (3.14) išplaukia, kad $u_0 \cdot e(X) \geq 0$. Kadangi $u_0 > 0$, tai turi būti patenkinama nelygybė $e(X) \geq 0$, kas prieštarauja pirmajai sistemos (3.13) nelygybei. Todėl (3.13) neturi sprendinio.

8 teorema. Sakysim A_0 – netuščia iškila aibė, $f(\cdot)$, $g_i(\cdot)$ – iškilos funkcijos, o $H(X)$ –afininė funkcija. Be to, sakysim, kad patenkinama tokia ribojimų reguliarumo prielaida: egzistuoja $\widehat{X} \in A_0$ toks, kad $G(\widehat{X}) < 0$, $H(\widehat{X}) = 0$, o $0 \in R^m$ yra vidinis aibės $\Gamma = \{\gamma : \gamma = H(X),\ X \in A_0\} \subseteq R^m$ taškas.

Tada

$$\inf_{X \in A} f(X) = \sup_{U \geq 0} \Theta(U, V). \tag{3.17}$$

Jei inf baigtinis, tai $\exists\ \overline{U} \geq 0, \overline{V}$, tokie kad

$$\sup_{U \geq 0} \Theta(U, V) = \Theta(\overline{U}, \overline{V}).$$

Jei $\exists\ \overline{X}$, toks kad $\inf_{X \in A} f(X) = f(\overline{X})$, tai $\overline{U} \cdot G(\overline{X}) = 0$.

Įrodymas. Pažymėkime $f_- = \inf_{X \in A} f(X)$. Jei $f_- - \infty$, tai įrodymas trivialus, nes 1 išvada teigia, kad $\sup_{U \geq 0} \Theta(U, V) = -\infty$.

Sakysim $f_- > -\infty$. Panagrinėkime nelygybių sistemą

$$f(X) - f_- < 0,\ G(X) \leq 0,\ H(X) = 0.$$

Pagal f_- apibrėžimą pastaroji nelygybių sistema neturi sprendinio. Todėl iš Lemos teiginio išplaukia, kad egzistuoja $(u_0, U, V) \neq 0$, $(u_0, U) \geq 0$, tokie kad

$$u_0 \cdot (f(X) - f_-) + U \cdot G(X) + V \cdot H(X) \geq 0,$$

$$\forall X \in A_0. \tag{3.18}$$

Įrodysim, kad $u_0 > 0$. Sakysim priešingai $u_0 = 0$. Pagal teoremos prielaidas egzistuoja $\widehat{X}$, toks kad $G(\widehat{X}) < 0$, $H(\widehat{X}) = 0$. Įstačius $\widehat{X}$ į (3.18), gaunamos tokios išvados: $U \cdot G(X) \geq 0$, $U = 0$. Todėl turi būti teisinga nelygybė $V \cdot H(X) \geq 0$. Kadangi $0 \in \Gamma$, galima parinkti $X \in A_0$ taip, kad $H(X) = -\epsilon V, \epsilon > 0$. Tada turi galioti nelygybė $-\epsilon||V|| \geq 0$, ir $||V|| = 0$. Gautoji išvada $(u_0, U, V) = 0$ prieštarauja anksčiau suformuluotam rezultatui. Taigi, prielaida $u_0 = 0$ neteisinga, ir $u_0 > 0$. Padalinę (3.18) iš u_0 ir pažymėję atitinkamus koeficientus $\overline{U}, \overline{V}$ gausime nelygybę

$$f(X) + \overline{U} \cdot G(X) + \overline{V} \cdot H(X) \geq f_-, \forall X \in A_0.$$

Todėl turi galioti

$$\inf_{X \in A_0} (f(X) + \overline{U} \cdot G(X) + \overline{V} \cdot H(X)) \geq f_-. \tag{3.19}$$

Bet dešinioji pastarosios nelygybės pusė sutampa su $\Theta(\overline{U}, \overline{V})$, o 1 išvada teigia, kad $\Theta(\overline{U}, \overline{V}) \leq f_-$. Dvi paskutiniosios nelygybės įrodo (3.17).

Jei egzistuoja $\overline{X} \in A$, toks kad $f(\overline{X}) = f_-$, tai iš (3.19) išplaukia $\overline{U} \cdot G(\overline{X}) = 0$.

Apibrėžtis. Vektorius $(\overline{X}, \overline{U}, \overline{V})$ vadinamas Lagrange'o funkcijos balno tašku, jei $\overline{X} \in A_0, \overline{U} \geq 0$, ir $\forall X \in A_0$, (U, V), $U \geq 0$, patenkintos nelygybės

$$L(\overline{X}, U, V) \leq L(\overline{X}, \overline{U}, \overline{V}) \leq L(X, \overline{U}, \overline{V}). \tag{3.20}$$

9 teorema. Vektorius $(\overline{X}, \overline{U}, \overline{V})$, $\overline{X} \in A_0, \overline{U} \geq 0$, yra Lagrange'o funkcijos balno taškas tada ir tik tada, kai

a) $L(\overline{X},\overline{U},\overline{V}) = \min_{X\in A_0} L(X,\overline{U},\overline{V})$,
b) $G(\overline{X}) \leq 0$, $H(\overline{X}) = 0$,
c) $\overline{U}\cdot G(\overline{X}) = 0$.

Įrodymas. Sakysim $(\overline{X},\overline{U},\overline{V})$, $\overline{X} \in A_0, \overline{U} \geq 0$, yra Lagrange'o funkcijos balno taškas. Lygybė a) turi galioti pagal balno taško apibrėžimo (3.20) kairiąją nelygybę. Dešinioji nelygybė, kurios išskleista forma yra šitokia

$$f(\overline{X}) + \overline{U}\cdot G(\overline{X}) + \overline{V}\cdot H(\overline{X}) \geq f(\overline{X}) + U\cdot G(\overline{X}) + V\cdot H(\overline{X}),$$

galėtų būti pažeista, jei negaliotų bent viena iš sąlygų $G(\overline{X}) \leq 0$, $H(\overline{X}) = 0$, nes U ir V gali būti parinkti norimai dideli absoliutaus dydžio prasme. Paėmę $U0$, gauname nelygybę $\overline{U}\cdot G(\overline{X}) \geq 0$, o įvertinę $G(\overline{X}) \leq 0$, $\overline{U} \geq 0$, gauname lygybę $\overline{U}\cdot G(\overline{X}) = 0$. Todėl galioja a), b), c), ir pirmoji teoremos dalis įrodyta.

Sakysim priešingai, žinomas $(\overline{X},\overline{U},\overline{V})$, $\overline{X} \in A_0, \overline{U} \geq 0$, toks, kad galioja a), b), c). Dėl a) turi galioti dešinioji (3.20) nelygybė

$$L(\overline{X},\overline{U},\overline{V}) \leq L(X,\overline{U},\overline{V}),\ \forall X \in A_0.$$

Remiantis b) ir c), akivaizdu

$$L(\overline{X},\overline{U},\overline{V}) = f(\overline{X}) + \overline{U}\cdot G(\overline{X}) + \overline{V}\cdot H(\overline{X}) =$$

$$= f(\overline{X}) \geq f(\overline{X}) + U\cdot G(\overline{X}) + V\cdot H(\overline{X}) =$$

$$= L(\overline{X},U,V),$$

t.y. , galioja ir kairioji (3.20) nelygybė.

10 teorema. Vektorius $(\overline{X},\overline{U},\overline{V})$, $\overline{X} \in A_0, \overline{U} \geq 0$, yra Lagrange'o funkcijos balno taškas tada ir tik tada, kai optimali **P** reikšmė $f(\overline{X})$ sutampa su optimalia **D** reikšme $\Theta(\overline{U},\overline{V})$.

Įrodymas. Sakysim $(\overline{X}, \overline{U}, \overline{V})$, $\overline{X} \in A_0, \overline{U} \geq 0$, yra Lagrange'o funkcijos balno taškas. Todėl galioja tik ką įrodytos teoremos b) teiginys, ir $\overline{X}$ yra leistinas P uždavinio sprendinys. Kadangi $\overline{U} \geq 0$, tai $(\overline{U}, \overline{V})$ yra leistinas D uždavinio sprendinys. Remiantis a), b) ir c) galima padaryti tokias išvadas

$$\begin{aligned}
\Theta(\overline{U}, \overline{V}) &= L(\overline{X}, \overline{U}, \overline{V}) = \\
&= f(\overline{X}) + \overline{U} \cdot G(\overline{X}) + \overline{V} \cdot H(\overline{X}) = f(\overline{X}).
\end{aligned}$$

7 teoremos 2 išvada teigia, kad tokiu atveju $\overline{X}$ yra P uždavinio optimalus sprendinys, o $(\overline{U}, \overline{V})$ yra optimalus D uždavinio sprendinys.

Sakysim atvirkščiai, $\overline{X}$ yra P uždavinio optimalus sprendinys, $(\overline{U}, \overline{V})$ yra optimalus D uždavinio sprendinys ir $\Theta(\overline{U}, \overline{V}) = f(\overline{X})$. Todėl $\overline{X} \in A_0$, $\overline{U} \geq 0$, $G(\overline{X}) \leq 0$, $H(\overline{X}) = 0$. Kadangi $\overline{X}$ ir $(\overline{U}, \overline{V})$ yra leistini P ir D uždavinių sprendiniai, tai atitinkamos tikslo funkcijų reikšmės turi patenkinti (3.11):

$$\Theta(\overline{U}, \overline{V}) = \min_{X \in A_0} f(X) + \overline{U} \cdot G(X) + \overline{V} \cdot H(X) \leq \tag{3.21}$$

$$\leq f(\overline{X}) + \overline{U} \cdot G(\overline{X}) + \overline{V} \cdot H(\overline{X}) = f(\overline{X}) + \overline{U} \cdot G(\overline{X}) \leq$$

$$\leq f(\overline{X}).$$

Bet pagal prielaidą turi galioti $\Theta(\overline{U}, \overline{V}) = f(\overline{X})$. Todėl tarp kraštinių (b) narių galioja lygybė, ir visas (3.21) nelygybes galima pakeisti lygybėmis. Tą padarius, akivaizdu: $\overline{U} \cdot G(\overline{X}) = 0$, ir

$$L(\overline{X}, \overline{U}, \overline{V}) = f(\overline{X}) = \min_{X \in A_0} L(X, \overline{U}, \overline{V}).$$

Įrodėme, kad galioja 9 teoremos a), b), c) sąlygos, vadinasi, $(\overline{X}, \overline{U}, \overline{V})$ yra Lagrange'o funkcijos balno taškas.

Galima parodyti, kad esant patenkintoms 8 teoremos prielaidoms, iš P uždavinio sprendinio $\overline{X}$ egzistavimo išplaukia tokio vektoriaus $(\overline{U}, \overline{V})$, $\overline{U} \geq 0$, egzistavimas, kad $(\overline{X}, \overline{U}, \overline{V})$ yra Lagrange'o funkcijos balno taškas. Papildomai galiojant diferencijuojamumo prielaidai galima įrodyti, kad balno taškas patekina Karush-Kuhn-Tucker sąlygas ir atvirkščiai, taškas, patenkinantis šias sąlygas, yra balno taškas.

Šiame paragrafe palietėme tik pagrindinius dualumo teorijos rezultatus, liečiančius būtinas ir pakankamas optimalių sprendinių sąlygas. Dualumo teorijos taikymai algoritmų konstravimui ir analizei aprašyti, pavyzdžiui [3].

3.3 Baudos ir barjerų metodai

Paprastumo dėlei tarkime, kad tikslo funkcija apibrėžta su visais $X \in R^n$, ir žinomas leistinosios srities vidaus taškas X_1. Pradiniu tašku parinkus X_1, lokalųjį minimumą, jei toks egzistuotų srities viduje, galima būtų rasti naudojant metodą, skirtą uždaviniams be ribojimų. Tačiau techniniuose bei ekonominiuose optimizavimo uždaviniuose minimumas dažniausiai pasiekiamas leistinosios srities ribiniame taške, kaip pavyzdžiui pavaizduota 3.11 paveiksle. Pagal ieškojimo be ribojimų metodą gauta trajektorija, prasidedanti taške X_1, šiuo atveju išeina už leistinosios srities, pažeisdama ribojimą $g_1(X) < 0$. Trajektorijos dalis iki susikirtimo su ribojimu galėtų būti gauta ir pagal metodą be ribojimų. Trajektorijos tąsa turi kompromisiškai suderinti funkcijos mažėjimo kryptį ir ribojimus. Bendrą funkcijos reikšmės sumažėjimą ir ribojimų pažeidimą įvertina vadinamoji **baudos funkcija**

$$B(X, r) = f(X) + \frac{1}{r} b(X) \tag{3.22}$$

čia $b(X) = 0$, jei ribojimai nepažeisti: $g_i(X) \leq 0,\ i = 1, ..., k,\ g_i(X) = 0,\ i = k+1, ..., m$, ir $b(X) > 0$, jei ribojimai pažeisti; $r > 0$. Antrasis baudos funkcijos dėmuo rodo, kad ribojimų pažeidimas ekvivalentus tam tikram tikslo funkcijos reikšmės padidėjimui. Dažnai naudojamos kvadratinė ir absoliuti baudos:

$$b(X) = \sum_{i=1}^{k}(\max(0,\ g_i(X))^2 + \sum_{i=k+1}^{m}(g_i(X))^2, \tag{3.23}$$

$$b(X) = \sum_{i=1}^{k}\max(0,\ g_i(X) + \sum_{i=k+1}^{m}|g_i(X)|. \tag{3.24}$$

Baudos funkcija $B(X, r)$ su pakankamai mažu r staigiai didėja už leistinosios srities ribų. Parametrui r artėjant prie nulio, $B(X, r)$ minimumo taškas artėja prie (3.1) uždavinio sprendinio. Todėl (3.1) uždavinys keičiamas seka uždavinių su diferencijuojama funkcija $b(X)$, pavyzdžiui (3.23), apibrėžiama mažėjančia parametro r seka. Kiekvieno paskesnio uždavinio pradinis taškas yra anstesnio uždavinio sprendinys. Paprastai, mažinant r, griežtinamos ir metodo be ribojimų sustojimo sąlygos.

Galima parodyti, kad egzistuoja toks pakankamai mažas r_0, kad, $r \leq r_0$ absoliučios baudos funkcijos minimumas sutamps su sprendiniu. Todėl uždavinį (3.1) galima pakeisti vienu minimizavimo uždaviniu, bet su nediferencijuojama, pavyzdžiui absoliutinės baudos, funkcija.

Diferencijuojamą baudos funkciją, kai parametro r reikšmės mažos, sunku minimizuoti, nes baudos funkcijos hesianas blogai apibrėžtas, t.y. minimumo taškas yra gilaus griovio dugne. Tam tikrų sunkumų kyla ir tuo atveju, kai baudos funkcija nediferencijuojama.

Vartojant baudos metodus, generuojama seka taškų, artėjančių prie sprendinio iš leistinosios srities išorės. Kartais tikslo funkcija būna apibrėžta tik leistinoje srityje. Tada optimizavimo metodo generuojami taškai turi likti srities viduje viso minimizavimo poceso metu.

Barjerų funkcija skiriasi nuo baudos tuo, kad ji ima staigiai didėti artėjant prie ribojimų ir dar jų nepažeidus. Metodu be ribojimų gauta tokios funkcijos minimizavimo trajektorija ima krypti išilgai (3.1) uždavinio ribojimų, jų dar nepasiekus. Aišku, barjerų metodas vartotinas tik tuo atveju, kai ribojimai nelygybiniai ir leistinosios srities hipertūris nelygus nuliui. Plačiausiai vartojama lagaritminė **barjerų funkcija**

$$B(X,r) = f(X) - r\sum_{1=1}^{k} ln(-g_i(X)). \tag{3.25}$$

Vartojant barjerų metodą, uždavinys su ribojimais keičiamas uždavinių be ribojimų seka su $r \to 0$. Analogija tarp barjerų ir baudos metodų akivaizdi. Dėl šios analogijos abu metodai dažnai vadinami baudos metodais, tik vienas išorinės baudos, o kitas vidinės baudos metodu. Sustojimo sąlyga nustatoma pagal mažą minimumo taško pokytį keliems gretimiems uždaviniams be ribojimų, t. y. su keliomis mažėjančiomis parametro r reikšmėmis.

Baudos metodas gana plačiai vartojamas techniniams optimizavimo uždaviniams spręsti. Jis efektyvus tada, kai baudos funkcija natūraliai išplaukia iš techninio uždavinio formuluotės. Baudos metodas, kartais to nesuvokiant, vartojamas daugiakriteriam optimizavimo uždaviniui pakeisti vienkriteriu. Uždavinio formulavimo etape inžinieriai išskiria vieną kriterijų kaip tikslo funkciją, kitus kriterijus pakeisdami ribojimais. Po to programuotojai, vartodami baudos metodą, ribojimus perkelia į tikslo funkciją. Taip visi kriterijai pakeičiami vienu. Kadangi toka pakeitimas nevienintelis ir toli gražu ne visada geriausias, reikėtų jį pagrįsti techninio uždavinio analize.

Griežtai matematiškai baudos metodą pirmą kartą pavartojo žymus amerikiečių matematikas R. Courant 1943 m. judėjimo ribotoje srityje analizei. Baudos metodas gana plačiai taikomas tiek teoriniuose tyrimuose, tiek konstruojant algoritmus. Beje, jis gan gerai tinka daugiaekstremių uždavinių su ribojimais apytiksliam globalinio minimumo įvertinimui. Lokalinio minimizavimo uždviniams spręsti baudos ir barjerų metodai buvo plačiai vartojami septintajame dešimtmetyje. Vieną iš populiariausių to laikotarpio ESM programų SUMT sukūrė amerikiečių specialistai A.Fiacco ir G.McCormick. Pagrindinis šių metodų trūkumas yra tas, kad baudos ir barjerų funkcijų reljefe susidaro gilūs sudėtingos formos grioviai, kuriuose nusileidimo metodų žingsniai negali būti ilgi. Šiuo metu lokalinio minimizavimo uždaviniams su diferencijuojamomis tikslo ir ribojimų funkcijomis spręsti sukurti efektyvūs metodai, kuriuose netiesiogiai panaudojamos baudos funkcijos, kombinuojant jas su Lagrange'o funkcijomis. Pastaruoju metu baudos funkcijų idėja pergyvena savo renesansą ryšium su vidinio taško metodų teorija.

Optimizavimo programų paketuose verta turėti baudos funkcija pagrįstą metodą, kuriame pagalbinis uždavinys (baudos funkcijos minimizavimas) būtų sprendžiamas metodu, nejautriu įvairiems uždavinio nereguliarumams. Tokiu pavyzdžiui galėtų būti FLEXI metodas [9], kuriame baudos funkcija minimizuojama deformuojamo simplekso metodu. Šis metodas vartotinas tada, kai tikslo ar ribojimų funkcijos apskaičiuojamos su reikšmingomis paklaidomis, gali turėti nediferencijuojamumo poaibius ar kitokius nukrypimus nuo funkcijų glotnumo prielaidos.

3.4 Leistinųjų krypčių metodai

Šios klasės metodų pagrindinė idėja - rasti kryptį, nepažeidžiančią ribojimų, kuria tikslo funkcija mažėtų. Nenulinis vektorius D vadinamas leistina kryptimi taške X, jei egzistuoja toks r_0, kad $X + rd$ tenkina ribojimus su visais $0 < r < r_0$. Vektorius D vadinamas leistina nusileidimo kryptimi, jei egzistuoja toks r_0 kad $f(X + rD) < f(X)$ ir $X + rD$ tenkina ribojimus su visais $0 < r < r_0$. Pasirinkus leistiną nusileidimo kryptį, žingsnis paprastai nustatomas, minimizuojant tikslo funkciją šia kryptimi, t.y. sprendžiant vienmatį minimizavimo uždavinį. Įvairūs metodai skiriasi krypties ir žingsnio parinkimu. Paprastai šis metodas taikomas tada, kai ribojimai tiesiniai arba netiesiniai nelygybiniai.

Dėl paprastumo nagrinėkime uždavinį su tiesiniais ribojimais:

$$\min f(X),\ AX + B \leq\ 0,\ CX + E = 0, \tag{3.26}$$

kur, A ir C - $k \times n$ ir $(m-k) \times n$ dimensijų matricos, o B ir E - k ir $m-k$ - dimensijų vektoriai, 0 - nulinis atitinkamo dimensijos vektorius, pavyzdžiui, uždavinio

$$\begin{aligned} f(X) &= (x_1-7)^2 + (x_2-5)^2, \\ A &= \begin{pmatrix} -1 & 1 \\ 2 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix},\ B = \begin{pmatrix} -3 \\ -15 \\ 0 \\ 0 \end{pmatrix}, \end{aligned} \tag{3.27}$$

tikslo funkcija yra kvadratinė, o keturi ribojimai yra tiesiniai nelygybiniai.

Suformuluotą uždavinį spręsime leistinųjų krypčių metodų pagrindėjo Zoutendijk'o metodu. Sakykime X_k - taškas leistinoje srityje. Jei visi nelygybiniai ribojimai tenkinami kaip griežtos nelygybės, tai taškas X_k yra vidinis leistinosios srities taškas. Iš jo galima žengti metodo be ribojimų žingsnį iki susikirtimo su leistinosios srities riba. Jei taškas X_k yra srities riboje, tai ribojimai, tenkinami kaip lygybiniai, vadinami aktyviais ribojimais. Leistinąjai kryp\v{c}iai iš taško, esančio ant srities krašto, nustatyti sprendžiamas tiesinio programavimo uždavinys:

$$\begin{aligned} &\min_D \ \nabla f(X_k)\, D, \\ A_1 D &\leq 0,\ C\cdot D = 0, \\ -1 &\leq d_i \leq 1,\ i = 1, ..., n, \end{aligned} \tag{3.28}$$

kur A_1 yra matricos A dalis, atitinkanti aktyvius ribojimus. Geometrinė uždavinio prasmė šitokia: ieškoma kryptis, tenkinanti aktyvius ribojimus ir sudaranti minimalų kampą su antigradiento kryptini. Sakykime, (3.27) pavyzdyje $X_k = (4, 7)$. Tada pirmieji du ribojimai yra aktyvūs, o tikslo funkcijos gradientas taške X_k lygus $(-6,\ 4)$. Taigi, leistinąjai krypčiai nustatyti reikia spręsti tiesinio programavimo uždavinį:

$$\min\,(-6d_1 + 4d_2),$$

$$\begin{aligned} -d_1 + d_2 &\leq 0, \\ 2d_1 + d_2 &\leq 0, \end{aligned} \tag{3.29}$$

$$-1 \leq d_1,\ d_2\ \leq\ 1.$$

Lengva įsitikinti, kad pagalbinio tiesinio programavimo uždavinio (3.29) sprendinys yra $D_k = (1/2, -1)$, o minimali tikslo funkcijos reikšmė lygi -7. Šiuo atveju tikslo funkcijos minimumas kryptimi D_k sutampa su (3.27) uždavinio sprendiniu (šeštasis pratimas). Jei taip nebūtų, rastame taške būtų galima nustatyti naują leistinąją kryptį ir tęsti iteracinį procesą.

Šį metodą galima modifikuoti uždaviniams su netiesiniais ribojimais. Aktyvieji ribojimai taške X_k aproksimuojami tiesiniais, t. y. $g_i(X) \approx\ g_i(X_k)+\nabla g_i(X_k)(X_k-X)$. Kaip ir anksčiau, leistinoji kryptis nustatoma aproksimuotam uždaviniui. Minimizuojant tikslo funkciją šia kryptimi, gaunamas naujas taškas X_{k+1} ir t. t. Šį metodą galima apibendrinti, įvertinant tikslo funkcijos, išskleistos Taylor'o eilute, aukštesnės eilės narius. Tada ieškoma krypties, tenkinančios ištiestintus ribojimus ir mažiausiai besiskiriančios nuo krypties, rodančios į aproksimuotos kvadratine funkcija tikslo funkcijos minimumo tašką. Suprantama, aproksimuojančioji kvadratinė funkcija turi būti teigiamai apibrėžta.

Šios klasės metodai dažniausia naudojami uždaviniams su tiesiniais ribojimais spręsti. Jei ribojimai labai netiesiniai, metodai neefektyvūs dėl trumpo žingsnio kiekvienoje iteracijoje.

3.5 Gradientų projekcijų ir redukcijų metodai

Rosen'o metodas - vienas iš plačiausiai naudojamų gradientų **projekcijų metodų**. Juo realizuojama ortogonali gradiento projekcija į ištiestintus aktyvius ribojimus. Dėl paprastumo tarkime, kad ribojimai tiesiniai. Iš pradinio taško pagal minimizavimo be ribojimų metodą einama iki trajektorijos sankirtos su leistinosios srities riba. Gali atsitikti ir taip, kad antigradientas kai kurių iš ribojimų, kurie tenkinami kaip lygybės, atžvilgiu nukreiptas į leistinosios srities vidų. Tada tokie ribojimai išbraukiami iš aktyvių ribojimų sąrašo. Antigradientas projektuojamas į aktyvių ribojimų sankirtą. Projekcija apskaičiuojama dauginant antigradientą iš matricos, sudarytos iš aktyvių ribojimų (tiesinių nelygybių) koeficientų. Šia kryptimi ieškoma tikslo funkcijos minimumo. Nauja iteracija pradedama aktyvių ribojimų sąrašo sudarymu ir t. t.

Rosen'o metodas, apibendrintas uždaviniams su netiesiniais ribojimais, artimas leistinųjų krypčių metodui ir remiasi tiesine ribojimų aproksimacija. Kadangi einant liestine, galima gerokai nutolti nuo leistinosios srities, tai daromas tik nedidelis žingsnelis ir gautas taškas koreguojamas, projektuojant jį į netiesinius ribojimus. Projekcija apskaičiuojama iteraciniais metodais, kurie artimi optimizavimo be ribojimų metodams. Todėl aišku, kad žengti didelį žingsnį liestinės kryptimi gali būti neracionalu, kad nebūtų per daug nutolta nuo leistinosios srities.

Projekcijų metodai plačiai taikomi uždaviniams su tiesiniais ribojimais spręsti. Jie ypač efektyvūs (palyginti su kitais), kai yra daug kintamųjjų. Remiantis išdėstytomis idėjomis, kintamosios metrikos metodai be ribojimų nesunkiai modifikuojami uždaviniams su tiesiniais ribojimais. Tereikia modifikuoti hesiano perskaičiavimo formules taip, kad būtų įvertinami aktyvių ribojimų aibės pasikeitimai, t. y. jos sumažėjimas arba papildymas naujais ribojimais.

Redukcijos metodų idėja ta, kad dalis kintamųjų traktuojami kaip priklausomi ir gali būti išreikšti nepriklausomais kintamaisiais sprendžiant aktyviųjų ribojimų lygtis. Aktyviųjų tiesinių ribojimų lygtys sprendžiamos gana lengvai, bet eliminuoti priklausomus kintamuosius netiesinių ribojimų atveju, sprendžiant (netiesines) lygtis, praktiškai neįmanoma. Tačiau palyginti paprastai galima gauti priklausomų kintamųjų diferencialines charakteristikas. Panagrinėkime dviejų kintamųjų uždavinį su vienu lygybiniu ribojimu

$$g(x_1, x_2) = 0. \tag{3.30}$$

Tikslo ir ribojimo funkcijos diferencialai lygūs:

$$\partial f(X) = \frac{\partial f(X)}{\partial x_1} dx_1 + \frac{\partial f(X)}{\partial x_2} dx_2, \tag{3.31}$$

$$\partial g(X) = \frac{\partial g(X)}{\partial x_1} dx_1 + \frac{\partial g(X)}{\partial x_2} dx_2. \tag{3.32}$$

Kadangi iš (3.30) išplaukia, kad $\partial g(X) = 0$, tai iš (3.32) gauname:

$$dx_2 = -\left(\frac{\partial g(X)}{\partial x_1}\Big/\frac{\partial g(X)}{\partial x_2}\right) dx_1. \tag{3.33}$$

(3.33) įrašę į (3.31), gausime:

$$\partial f(X) = \left(\frac{\partial f(X)}{\partial x_1} - \left(\frac{\partial f(X)}{\partial x_2} \cdot \frac{\partial g(X)}{\partial x_1}\Big/\frac{\partial g(X)}{\partial x_2}\right)\right) dx_1.$$

Iš čia išplaukia, kad funkcijos $f(X)$ (redukuota) išvestinė nepriklausomo kintamojo x_1 atžvilgiu lygi

$$\frac{df(X)}{dx_1} = \frac{\partial f(X)}{\partial x_1} - \frac{\partial f(X)}{\partial x_2}\left(\frac{\partial g(X)}{\partial x_2}\right)^{-1}\frac{\partial g(X)}{\partial x_1}.$$

Lokalaus minimumo taško galima ieškoti kaip lygties

$$\frac{df(X)}{dx_1} = 0$$

sprendinio. Pavyzdžiui, uždavinio su tikslo funkcija $f(X) = x_1 + x_2$ ir ribojimu $g(X) = x_1^2 + x_2^2 - 1 = 0$ redukuota išvestinė lygi

$$\frac{df(X)}{dx_1} = 1 - \frac{x_1}{x_2}.$$

Prilyginę pastarąją išvestinę nuliui ir įvertindami ribojimą gausime minimizavimo uždavinio sprendinį $x_1 = x_2 = -1/\sqrt{2}$. Analitiškai išspręsti išvestinės lygybės nuliui lygtį galima tik specialiais atvejais. Tačiau redukuota išvestinė gali būti panaudota ir iteraciniam algoritmui sudaryti. Sakysim, žinomas leistinas sprrendinys X. Įvertinę redukuotos išvestinės ženklą ir parinkę žingsnį pagal kintamąjį $\triangle x_1$, žingsnį pagal kintamąjį $\triangle x_2$ nustatysime pagal (3.33). Šiuo atveju svarbus tik redukuotos išvestinės ženklas. Naujas taškas gaunamas, kaip įprasta, iš taško X žengus žingsnį ($\triangle x_1$, $\triangle x_2$). Gautasis taškas dažniausiai nepatenkina ribojimų ir todėl turi būti koreguojamas. Korekcija paprastai vykdoma sprendžiant lygtį $g(X) = 0$. Todėl šio metodo žingsnis negali būti ilgas.

Daugelio kintamųjų atveju metodas sudaromas analogiškai. Randamas tikslo funkcijos gradientas nepriklausomų kintamųjų atžvilgiu, įvertinant $m < n$ lygybinių ribojimų. Jis vadinamas redukuotu gradientu:

$$\begin{aligned} \nabla_R f(X) \quad = \quad & \nabla_{X_1} f(X) - \\ & \nabla_{X_2} f(X) \cdot \left(\frac{\partial g(X)}{\partial X_2}\right)^{-1} \cdot \frac{\partial g(X)}{\partial X_1}; \end{aligned} \tag{3.34}$$

čia $X_1 = (x_1, ..., x_{n-m})$ - nepriklausomų kintamųjų vektorius, $X_2 = (x_{n-m+1}, ..., x_n)$ - priklausomų kintamųjų vektorius, o gradientų vektoriai suprantami kaip vektoriai-eilutės. Dalinių išvestinių vektoriai šių kintamųjų atžvilgiu žymimi gradiento simboliu su atitinkamu indeksu, o vektorinės ribojimų funkcijos $g = (g_1, ..., g_m)$ išvestinių matricos žymimos $\frac{\partial g(X)}{\partial X_1}$ ir $\frac{\partial g(X)}{\partial X_2}$

$$\frac{\partial g(X)}{\partial X_1} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & ... & \frac{\partial g_1}{\partial x_{n-m}} \\ ... & ... & ... \\ \frac{\partial g_m}{\partial x_1} & ... & \frac{\partial g_m}{\partial x_{n-m}} \end{pmatrix},$$

$$\frac{\partial g(X)}{\partial X_2} = \begin{pmatrix} \frac{\partial g_1}{\partial x_{n-m+1}} & ... & \frac{\partial g_1}{\partial x_n} \\ ... & ... & ... \\ \frac{\partial g_m}{\partial x_{n-m+1}} & ... & \frac{\partial g_m}{\partial x_n} \end{pmatrix}.$$

Žingsnio kryptis $(\Delta X_1, \Delta X_2)$ parenkama šitaip:

$$\begin{aligned}\triangle X_1 &= -\nabla_R f(X), \\ \triangle X_2 &= -\nabla_R f(X) \cdot \left(\left(\frac{\partial g(X)}{\partial X_2} \right)^{-1} \cdot \frac{\partial g(X)}{\partial X_1} \right)^t .\end{aligned} \tag{3.35}$$

Kai ribojimai tiesiniai, (3.35) kryptis neišveda iš leistinosios srities. Jei ribojimai netiesiniai, tai jie dažniausiai pažeidžiami. Todėl, žengus žingsnį kryptimi (3.35), reikia koreguoti gautąjį tašką $X + \gamma(\triangle X_1, \triangle X_2)$. Paprastai Niutono metodu sprendžiama lygčių sistema $g(X) = 0$ kintamųjų X_2 atžvilgiu, esant fiksuotam $X_1 + \gamma \triangle X_1$. Taigi, kaip ir vartojant projekcijų metodą, gaunamas trumpas žingsnis, nors skaičiavimai ir komplikuoti. Tiek gradientų projekcijų, tiek gradientų redukcijų metodai (ypač pastaruoju metu pasiūlyti apibendrinti metodai) efektyvūs uždaviniams su tiesiniais ribojimais ir dideliu kintamųjų skaičiumi. Vienas iš pagrindinių apibendrinimų yra kintamos metrikos idėjų panaudojimas krypties parinkimui. Literatūroje paskelbta rezultatų, rodančių redukuotų gradientų tipo metodų efektyvumą sprendžiant kai kuriuos netiesinius uždavinius. Tiksliai apibrėžti jų klasę gana sunku, bet vaizdžiai tariant, galima ją nusakyti kaip klasę uždavinių su artimais tiesiniams ribojimams.

3.6 Lagrange'o funkcijų metodai

Nagrinėdami būtinas minimumo sąlygas 1 skyrelyje, pabrėžėme Lagrange'o funkcijos reikšmę netiesinio programavimo uždaviniams analizuoti. Jos pagrindu galima konstruoti ir minimizavimo metodus. Kaip jau buvo minėta, Lagrange'o funkcijos balno taškas nusako iškilo programavimo uždavinio sprendinį. Todėl metodas balno taškui ieškoti kartu yra ir metodas iškilojo programavimo uždaviniui spręsti. Priminsime, kad Lagrange'o funkcija vadinama

$$L(X, \Lambda) = f(X) + \sum_{i=1}^{m} \lambda_i g_i(X), \ \ \lambda_i \geq 0.$$

Reikia rasti tašką (X_*, Λ_*), tenkinantį nelygybes

$$L(X_*,\, \Lambda) \leq L(X_*,\, \Lambda_*) \leq L(X,\, \Lambda_*),\, \lambda_{*i} \geq 0.$$

Kadangi pastarasis uždavinys yra su paprastais (neneigiamumo) ribojimais, tai jį galima spręsti gradientiniu metodu

$$\begin{aligned} X_{j+1} &= X_j - \gamma_j \sum_{i=1}^{m} \mu_{ji} \, \nabla g_i(X_j), \\ \mu_{j+1} &= (\mu_j + \gamma_j \, g(X_j))_+ \end{aligned} \tag{3.36}$$

čia ženklas $(\cdot)_+$ reiškia vektoriaus projekciją į teigiamą kvadrantą: t.y. aibę $\{\mu\colon\ \mu_i\ \geq 0, \mathrm{i} = 1, ..., m\ \}$, γ_j - žingsnio ilgio daugiklis. Tačiau kaip ir paprasti gradientiniai minimizavimo metodai, (3.36) nelabai efektyvus, o kartais juo apskritai nerandamas tikslus sprendinys. Antra vertus, lokalaus minimumo taškas išreiškiamas Lagrange'o funkcijos balno tašku tik esant gan griežtoms prielaidoms (iškilumas, reguliarumas). Bendru atveju tegalima teigti, kad Lagrange'o funkcija yra stacionari minimumo taške: jei X_* - lokalaus minimumo taškas, ir aktyviųjų ribojimų gradientai $\nabla g_i(X_*)$, $i \in I_*$, tiesiškai nepriklausomi, tai egzistuoja toks Λ_*, kad

$$\nabla f(X_*) + \sum_{i=1}^{m} \lambda_{*i} \, \nabla g_i(X_*) = 0,$$

ir X_* yra funkcijos $L(X, \Lambda_*)$ stacionarus taškas. Metodu (3.36) pradinio uždavinio minimumo taškas nebūtinai bus rastas. Taip gali atsitikti, jei Lagrange'o funkcijos stacionarus taškas bus ne minimumo, o perlinkio arba maksimumo taškas. Iš čia išplaukia natūralus noras "pataisyti" Lagrange'o funkciją taip, kad stacionarus taškas virstų minimumo tašku. Tą galima lengvai padaryti pridėjus prie $L(X,\Lambda)$ narį, kuris pats ir jo gradientas būtų lygus nuliui stacionariame taške, o tolstant nuo jo, pakankamai greitai didėtų. Viena iš paprasčiausių ir plačiausiai paplitusių modifikuotų Lagrange'o funkcijų yra ši:

$$L_I(X,\Lambda,r) = f(X) + \sum_{i=1}^{m} \lambda_i\, g_i(X) + \frac{1}{r}\sum_{i\in I} g_i^2(X), \tag{3.37}$$

čia I - aktyviųjų ribojimų indeksų aibė. Nesunku įrodyti, kad egzistuoja baigtinis r_1, užtikrinantis $L_I(X,\Lambda,r)$ hesiano teigiamą apibrėžtumą su visais $r < r_1$. Taigi X_* yra modifikuotos Lagrange'o funkcijos minimumo taškas. Ši išvada teisinga tik esant tam tikroms Λ ir r reikšmėms, kurios iš anksto nežinomos. Todėl modifikuotų Lagrange'o funkcijų pagrindu kuriamuose metoduose turi būti apskaičiuoti priimtini šių reikšmių įverčiai. Baudos daugiklio r įtaką nesunku suprasti, prisiminus baudos metodų aptarimą. Parinkę per mažą r reikšmę, turėsime staigiai kylantį funkcijos reljefo šlaitą tuoj už leistinosios srities ribos. Anksčiau buvo pabrėžta, kad tokią funkciją minimizuoti sunku, ypač kai srities riba kreiva. Kai r reikšmė per didelė, gali nepavykti transformuoti Lagrange'o funkcijos stacionaraus taško į (3.37) minimumo tašką. Jei būtų žinoma daugiklių vektoriaus reikšmė Λ_*, tai netiesinio programavimo uždavinį galėtume išspręsti, vieną kartą minimizuodami modifikuotą Lagrange'o funkciją (be ribojimų). Tačiau sprendžiant realius uždavinius Λ_* būna nežinomas, ir jį tenka vertinti naudojantis informacija apie $f(X)$ ir $g_i(X)$, $i = 1, ..., m$, gauta ankstesnėse iteracijose. Panagrinėsime specialų minimizavimo uždavinio atvejį, kai nėra nelygybinių ribojimų

$$\min_{G(X)=0} f(X),\ G(X) = (g_1(X), ..., g_m(X))^t.$$

Šiuo atveju modifikuota Lagranžo funkcija lygi

$$L(X, \Lambda, r) = f(X) + \sum_{i=1}^{m} \lambda_i \, g_i(X) + \frac{1}{2r} \sum_{i=1}^{m} g_i^2(X),$$

o jos gradientas išreiškiamas formule

$$\nabla_X L(X, \Lambda, r) = \nabla f(X) + \sum_{i=1}^{m} (\lambda_i \, + \frac{1}{r} g_i(X)) \; \nabla g_i(X). \qquad (3.38)$$

Algoritmo vykdymui pradėti reikia pasirinkti pradines Λ ir r reikšmes. Turint einamąsias Lagrange'o ir baudos daugiklių reikšmes, Λ_k , r_k, sprendžiamas minimizavimo be ribojimų uždavinys

$$\min \; L(X, \, \Lambda_k \, , r_k).$$

Išsprendus šį uždavinį, gaunamas naujas minimumo taško įvertis X_k. Remiantis modifikuotos Lagranžo funkcijos gradiento (3.38) lygybe nuliui taške X_k Lagrange'o daugikliai perskaičiuojami pagal formulę

$$\Lambda_{k+1} = \Lambda_k + \frac{1}{r_k} G(X_k).$$

Baudos daugiklis sumažinamas padauginant jį iš koeficiento: $r_{k+1} = \tau \, r_k$, $0 < \tau < 1$. Įrodoma, kad esant pakankamai bendroms prielaidoms, toks metodas konverguoja į lokalųjį minimumą.

Jei uždavinyje be lygybinių yra ir nelygybiniai ribojimai, pastaruosius galima pakeisti lygybiniais uždavinio kintamuosius papildant fiktyviaisiais kintamaisiais. Kitokia metodo versija, kurioje einamasis taškas nustatomas (3.37) minimizavimu. Šiuo atveju reikšmę turi aktyviųjų ribojimų aibės nustatymo būdas. Aktyvieji ribojimai dažniausiai prognozuojami kiekvienam minimizavimo (be ribojimų) kintamųjų vektoriaus X atžvilgiu ciklui. Įvairūs modifikuotų Lagrange'o funkcijų metodai skiriasi baudos ir Lagrange'o daugiklių perskaičiavimo, aktyvinių ribojimų aibės nustatymo, minimizavimo kintamųjų vektoriaus X atžvilgiu būdais. Praktinis metodo efektyvumas priklauso ir nuo programinės realizacijos. Pabrėšime, kad Lagrange'o funkcijomis pagrįstų metodų efektyvumo laidas - ribojimų įvertinimas Lagrange'o daugikliais, kai baudos narys tik "pataiso" Lagrange'o funkciją; todėl baudos daugiklį (3.37) formulėje daug lengviau įvertinti negu tiesiogiai taikomos baudos metoduose (3.24), (3.23), (3.25). Lagrange'o daugiklių įverčiai perskaičiuojami iteraciniais metodais, kurių iteracijos eina pakaitomis su (3.37) funkcijos minimizavimo X atžvilgiu iteracijomis (kintamos metrikos metodu).

Efektyvūs netiesinio programavimo metodai sukurti remiantis panašia idėja. Nors bendru atveju Lagrange'o funkcijos stacionarus taškas X atžvilgiu nebūtinai yra minimumo tašku, bet tam tikru būdu parinkę X kitimo aibę, galime garantuoti, kad šioje aibėje stacionarus taškas būtų minimumo tašku. Sakykime, X_*-lokalaus minimumo taškas, o λ_{*i}, $i \in I$ tokie, kad

$$\nabla f(X_*) + \sum_{i \in I} \lambda_{*i} \nabla g_i(X_*) = 0,$$

čia I - aktyviųjų taške X_* ribojimų indeksų aibė. Tada X_* yra funkcijos $L(X, \Lambda_*)$ minimumo taškas aibėje, kurią apibrėžia ortogonalūs aktyvinių ribojimų gradientams vektoriai. Taigi, žinodami Λ_*, galėtume rasti X_* vienąkart minimizuodami Lagrange'o funkciją aibėje, aprašytoje tiesiniais ribojimais. Kadangi Λ_* nežinomas, tai jį reikia įvertinti naudojantis informacija apie tikslo ir ribojimų funkcijas, gaunama minimizavimo eigoje. Metodą sudaro viena kitą keičiančios iteracijos: Lagrange'o daugiklių įverčių tikslinimo ir Lagrange'o funkcijos minimizavimo X atžvilgiu. Tačiau tiksliai minimizuoti Lagrange'o funkciją esant netiksliems Λ_* įverčiams dažniausiai neverta Todėl vietoj šio minimizavimo uždavinio formuluojamas kvadratinio programavimo uždavinys, kurio tikslo funkcija yra kvadratine funkcija aproksimuota Lagrange'o funkcija. Ji konstruojama naudojantis hesiano perskaičiavimo formulėmis, kaip ir kintamos metrikos metoduose. Kvadratinio programavimo uždaviniams spręsti yra labai efektyvių algoritmų, todėl ši iteracija nekelia problemų.

Modifikuotomis Lagrange'o funkcijomis pagrįstų metodų efektyvumą lemia Lagrange'o daugiklių įverčių ir hesiano perskaičiavimo metodai. Modifikuotų Lagrange'o funkcijų metodai yra vieni iš efektyviausių netiesinio programavimo uždaviniams su diferencijuojamomis tikslo ir ribojimų funkcijomis spręsti, kai kintamųjų skaičius nelabai didelis (iki kelių šimtų). Šie metodai ypač efektyvūs uždaviniams su tiesiniais ribojimais.

3.7 Metodų aptarimas

Paprasčiausi yra uždaviniai su diferencijuojamomis tikslo funkcjomis ir tiesiniais ribjimais. Jiems spręsti sukurta daug efektyvių metodų ir juos realizuojančios programinės įrangos. Uždaviniai su tiesiniais ribojimais netgi paprastesni už uždavinius be ribojimų ta prasme, kad dėl ribojimų sumažėja leistinosios aibės dimensija. Uždaviniams su tiesiniais ribojimais yra efektyvūs pojekcijų ir redukcijų metodai, nes šių metodų operacijos realizuojamos nesudėtingomis matricų daugybos operacijomis. Lagrange'o funkcijų metodai taip pat tinka minimizavimo uždaviniams su tiesiniais ribojimais. Tačiau naudoti šiems uždaviniams kompiuterių programas, skirtas bendro pobūdžio netiesiniams uždaviniams, ne visada tikslinga. Pavyzdžiui Powell'o arba Schittkowski'o programos labai efektyvios kreipinių į tikslo ir ribojimų funkcijų papogrames skaičiaus atžvilgiu. Tačiau jos užima labai daug operatyvios kompiuterio atminties, todėl jomis galima spręsti mažesnio kintamųjų skaičiaus uždavinius negu gradientų projekcijų ar redukcijų metodais. Jei ribojimai tiesiniai, o tikslo funkcija nediferencijuojama, tai pastarieji metodai gali būti nesunkiai apibendrinti, imant vietoj gradiento kvazigradientą. Tais atvejais, kai tikslo funkcija daugiaekstremė arba jai būdingi kitokie nereguliarumai, retai pavyksta pasinaudoti tiesinių ribojimų specifika.

Lokalaus minimizavimo su netiesiniais ribojimais uždaviniams, kai tikslo ir ribojimų funkcijos diferencijuojamos, o kintamųjų nelabai daug (iki kelių šimtų), efektyviausi yra Lagrange'o funkcijų metodai. Jie visapusiškai ištyrinėti, juos realizuojančios kompiuterių programos eksplotuojamos ilgus metus. Todėl programų konstrukcija yra išbaigta, sukurti gana tobuli pagalbinių uždavinių sprendimo metodai. Tačiau šie metodai nelabai tinka tais atvejais, kai tikslo ar ribojimų funkcijos nediferencijuojamos. Jei skaičiavimo paklaidos pernelyg didelės, tai skaitmeniškai vertinant dalines išvestines, galima gauti beprasmius rezultatus. Kartais funkcijos reikšmės skaičiuojamos iteraciniais metodais, kurių tikslumas nesunkiai reguliuojamas. Pavyzdžiui, iteracinis netiesinių diferencialinių lygčių sprendimo procesas nutraukiamas sprendinio tikslumą lemiantiems parametrams pasiekus tam tikras reikšmes. Tokiu atveju pradiniame minimizavimo etape racionalu imti didesnį nepriklausomo kintamojo pokytį skirtuminėse formulėse ir leisti didesnę funkcijos skaičiavimo paklaidą. Artėjant prie minimumo, funkcijos skaičiavimo tikslumą verta didinti, o diferencijavimo žingsnį mažinti. Panagrinėkime paprasčiausią vieno kintamojo atvejį. Gautos išvados tinka ir dalinėms išvestinėms, t. y. gradientui įvertinti. Naudojantis Taylor'o formule, funkcijos reikšmės prieaugį galima išreikšti šitaip:

$$f(x+h) - f(x) = f'(x) \cdot h + \frac{1}{2} f''(z) * h^2, x \leq \ z \leq \ x+h.$$

Aproksimuodami išvestinės reikšmę baigtiniu skirtumu

$$f'(x,h) \approx \frac{f(x+h) - f(x)}{h} \tag{3.39}$$

"ištiesinimo" paklaidą galime įvertinti dydžiu

$$\max{}_{x \leq\ z \leq\ x+h} f''(z)\ h^2/2 = f''\ \cdot h^2/2.$$

Kadangi funkcijos reikšmės skaičiuojamos su paklaida, kurią pažymėsime ε, tai bendras (3.39) paklaidos įvertis lygus

$$\frac{1}{2} f''\, h^2 + \frac{2\varepsilon}{h}.$$

Ši paprasta formulė atspindi abu pagrindinius paklaidų šaltinius: funkcijos "ištiesinimo" paklaida didėja proporcingai h^2, o funkcijos skaičiavimo paklaidos įtaka didėja proporcingai h^{-1}. Minimizavę paklaidos įvertį h atžvilgiu, gausime optimalų žingsnio ilgį $h_0 = 2\sqrt{\epsilon / f''}$. Tiesiogiai pasinaudoti šia formule neįmanoma, nes nežinomas antrosios išvestinės maksimumas f''. Padarę papildomą prielaidą apie funkcijos ir jos išvestinės kitimą, galėtume parodyti, kad optimalus žingsnio ilgis proporcingas $\sqrt{\varepsilon_0}$, kur ε_0 – santykinė funkcijos skaičiavimo paklaida. Tada santykinė išvestinės įverčio paklaida taip pat proporcinga $\sqrt{\varepsilon_0}$. Grubiai kalbant, reikšminių ženklų skaičius išvestinės įvertyje ne didesnis už pusę reikšminių ženklų skaičiaus funkcijos reikšmės įvertyje. Artėjant prie minimumo, ε_0 gali neleistinai išaugti net tada, kai parenkamas optimalus diferencijavimo žingsnis. Todėl gali būti tikslinga pereiti prie tikslesnės, pavyzdžiui, centrinio skirtumo formulės. Žinomos adaptyvios paprogramės, realizuojančios optimalius skaitmeninio diferencijavimo metodus. Tačiau jos yra sudėtingos, naudoja daug funkcijos reikšmių. Todėl paprastai apsiribojama vienpusio arba centrinio skirtumo formulėmis, kuriose žingsnis parenkamas vadovaujantis analogiškais anksčiau pateiktiems samprotavimams. Optimaliai parinkti ribojimų funkcijų diferencijavimo formulę ir jos parametrus sunkiau, nes svarbus ne pats gradientas, o tikslo funkcijos gradiento (arba kokios nors kitos nusileidimo krypties) projekcija į aibę, apibrėžiamą ribojimų funkcijų gradientais. Kadangi šiuo klausimu nėra netgi teorinių rekomendacijų, tai vadovaujamasi tais pačiais samprotavimais, kaip ir tikslo funkcijos atveju.

Svarbi efektyvaus netiesinio programavimo metodų panaudojimo sąlyga yra tinkamas tikslo ir ribojimų funkcijų bei nepriklausomų kintamųjų mastelio parinkimas. Teoriškai metodai dažniausiai invariantiški mastelių atžvilgiu, tačiau garantuoti, kad šią savybę turės ir jų realizacijos, negalima. Viena iš priežasčių ta, kad kompiuterio programos vykdo veiksmus ne tik su santykiniais bet ir su absoliučiais dydžiais. Pavyzdžiui, dabar paplitusiu Pentium tipo kompiuteriu minimizuojant ta pačia gradientinio tipo metodą realizuojančia kompiuterio programa funkcijas $f(X)$ ir $10^{-20}f(X)$, vargu ar galima tikėtis to paties atsakymo (suprantama, atsižvelgiant į funkcijos reikšmės mastelį) jau vien dėl to, kad antrajai funkcijai daugelis taškų bus laikoma optimaliais dėl mažesnės už kompiuterio programoje užfiksuotas ribines gradiento normos reikšmes.

Beveik visiems algoritmams nepageidautina, kad funkcijos variacija būtų artima nuliui. Taip pat nepageidautinos didelio absoliutaus didumo funkcijos reikšmės. Kuriant taikomojo uždavinio matematinį modelį, rekomenduotina funkcijos reikšmėms parinkti tokį mastelį, kad jos minimumas būtų apie vienetą, o maksimalios reikšmės, pakeltos kubu, būtų ne didesnės už maksimalų kompiuterio skaičių. Sudėtingiau parinkti nepriklausomų kintamųjų ir ribojimų funkcijų mastelius. Gerai apibrėžta laikoma tokia funkcija, kurios nepriklausomų kintamųjų vienodi pokyčiai sukelia tos pačios eilės funkcijos prieaugius. Specialioje literatūroje galima rasti samprotavimų, kaip tiesiškai transformuoti nepriklausomus kintamuosius, kad būtų optimaliai suderinti masteliai. Kadangi praktiškai juos realizuoti ne taip paprasta, dažniausiai apsiribojama tik natūralių uždavinio nepriklausomų kintamųjų mastelio parinkimu. Ribojimų funkcijoms šalia gero apibrėžtumo keliamas papildomas reikalavimas - jų įtaka turi būti apylygė, t.y. vienodo dydžio nepriklausomų kintamųjų pokyčiai turėtų iššaukti tos pačios eilės ribojimų funkcijų pokyčius. Geriausi netiesinio programavimo algoritmai automatiškai parenka priimtinus mastelius, tačiau vartotojas, formuluodamas uždavinį, turėtų atsižvelgti į šias paprastas rekomendacijas. Mastelių parinkimo neįmanoma visiškai automatizuoti, o formuluojant uždavinį įvertinti šiuos samprotavimus dažniausiai ne taip jau sunku.

Uždaviniams su ribojimais, kaip ir be jų, viena iš pagrindinių rekomendacijų - panaudoti kuo išsamesnę informaciją apie funkcijų išvestines. Kai tikslo ir netiesinės ribojimų funkcijos diferencijuojamos, efektyviausi yra Lagrange'o funkcijų metodai. Tą patvirtina gausūs ekspermentinių tyrimų rezultatai paskelbti literatūroje. Tačiau šie metodai netinka uždaviniams, kurių tikslo funkcija neapibrėžta už leistinosios srities ribų. Tokiu atveju vartotinas barjerų metodas.

Daug sudėtingiau parinkti efektyvų metodą, kai funkcijų reikšmes tegalima apskaičiuoti duotuoju algoritmu, kurio paklaidos tokios didelės, kad skaitmeniškai diferencijuoti beprasmiška. Tada ribojimai įvertinami baudos arba barjerų metodais, o uždavinys be ribojimų sprendžiamas pagal vieną iš tiesioginio ieškojimo algoritmų. Kuriant tokius metodus, daug ką tenka spręsti euristiškai. Todėl sunku apibrėžti jų racionalaus panaudojimo sritį. Kai kurios rekomendacijos, kada vartoti tiesioginio ieškojimo ir baudos metodus, pateiktos anksčiau. Plačiai naudojamas metodas - FLEXI, kuris pagrįstas baudos ir deformuojamo simplekso metodais. Tačiau uždaviniams su diferencijuojamomis tikslo funkcijomis, Lagrange'o funkcijomis pagtįsti metodai gerokai efektyvesni negu FLEXI. Tai reikia turėti galvoje ruošiant optimizavimo uždavinį spręsti kompiuteriu.

Daugiaekstremiai uždaviniai su bendro pobūdžio netiesiniais ribojimais baudos metodu dažniausiai pakeičiami uždaviniais su paprastais ribojimais, kuriems skirta dauguma globalinio minimizavimo metodų. Šis būdas grindžiamas tuo, kad kuriant globaliuosius metodus, apie tikslo funkciją paprastai daromos tik labai bendros priekaidos. Taigi natūralu perkelti į tikslo funkciją visus uždavinio sunkumus, nes pagrindinės prielaidos lieka patenkintos. Kadangi baudos daugiklis parenkamas gana laisvai, o globalaus minimizavimo metodais minimumas nustatomas labai apytiksliai, tai tokiu metodu gautą sprendinį verta patikslinti lokalaus netiesinio programavimo metodu, kuris parenkamas atsižvelgiant į tikslo funkcijos ir ribojimų savybes. Kai kuriuose, dažniausiai atsitiktinio globalaus ieškojimo, metoduose ribojimai vertinami taip: minimizuojama hiperstačiakampyje, aprašytame apie leistinąją sritį. Jei taškas patenka į leistinąją sritį, imama tikslo funkcijos reikšmė, jei ne - koks nors didelis skaičius. Tačiau šis metodas visiškai netinka dažnai pasitaikantiems uždaviniams, kurių leistinosios srities hipertūris yra labai mažas palyginti su aprašyto hiperstačiakampio, t. y. tikimybė pataikyti į leistinąją sritį artima nuliui. Beje šį metodą galima interpretuoti kaip baudos metodą su trūkia baudos funkcija.

Kartais kyla nesusipratimų dėl gauto optimizavimo uždavinio sprendinio tikslumo. Dažnas nepatyręs vartotojas, nurodęs baigimo sąlygų parametrą lygų ε, tikisi, kad rastas sprendinys skiriasi nuo tikrojo ne daugiau ε. Tokia nuomonė susiformuoja naudojantis paprastesnių uždavinių sprendimo algoritmais, kurie garantuoja randamo sprendinio tikslumą. Pavyzdžiui, pagal standartines sinuso arba kvadratinės šaknies radimo paprogrames funkcijos reikšmė apskaičiuojama mašininiu tikslumu. Netiesinio programavimo metodai gali garantuoti randamo sprendinio tikslumą tik labai specialiais atvejais. Praktiškai baigimo sąlygos apibrėžiamos keliais būdais. Gradientinio tipo metodų viena iš baigimo sąlygų yra būtinų minimumo sąlygos patenkinimas metodo vartotojo užduotu tikslumu. Tačiau tokias sąlygas (ypač neiškiliųjų funkcijų atveju) gali tenkinti taškai, esantys gana toli nuo lokalaus minimumo taško. Jeigu tikslo ir ribojimų funkcijoms nekeliami labai griežti reikalavimai, negalima garantuoti, kad gradiento projekcija į ribojimus monotoniškai mažės. Todėl sustojimo sąlyga gali būti tenkinama ne tik lokalaus minimumo aplinkoje, bet ir nutolusiuose nuo jo "spąstuose". Kitos, ypač tiesioginio ieškojimo metodų, sustojimo sąlygos išreiškia minimizavimo proceso sulėtėjimą, t.y. menką rastojo taško bei funkcijos reikšmės kitimą keliose iš eilės einančiose įteracijose. Tai būdinga lokalaus minimumo aplinkai, bet vėlgi ne tik jai. Todėl gautą sprendinį verta patyrinėti. Galima pabandyti iš gauto arba kito taško minimizuoti tuo pačiu metodu su kitais parametrais arba kitu metodu. Dažnai sprendinio analizei būna naudinga tikslo ir ribojimų funkcijų pjūvius per gautąjį tašką pavaizduoti grafikai.

Galima pabrėžti, kad netiesinio programavimo uždaviniai sudėtingi ir nėra universalių metodų jiems spręsti. Parenkant metodą, reikia atidžiai palyginti uždavinio savybes ir prielaidas, kuriomis pagrįstas tas ar kitas metodas. Nors matematinė optimizavimo teorija gerai išvystyta, bet sėkmingam sudėtingų praktikos uždavinių sprendimui reikia tam tikros patirties.

3.8 Istorinės pastabos

Trumpai paliesime tik minimumo sąlygų temą. Ne todėl, kad metodų kūrimas ir tyrimas būtų mažiau svarbūs. Reikalas tas, kad metodų kūrimas aktyviai tebesitęsia ir todėl sunku objektyviai įvertinti gautų rezultatų svarbą bei aktualumą. Minimumo sąlygų teorija jau gan išbaigta, jos rezultatų vertinimas pateikiamas apžvalgose, dalyko istorijai skirtuose straipsniuose ir netgi vadovėliuose.

Paprasčiausios būtinos minimumo sąlygos diferencijuojamai vieno kintamojo funkcijai išreiškiamos išvestinės lygybe nuliui. Šios sąlygos dažnai vadinamos Fermat (1601-1665) teorema, nors tuo metu, kai kūrė P.Fermat, išvestinės sąvoka dar nebuvo griežtai suformuluota.

Būtinas lokalaus minimumo sąlygas uždaviniams su lygybiniais ribojimais suformulavo žymus matematikas ir mechanikas J.Lagrange, gyvenęs 1763-1813. Vėliau Lagrange'o metodas buvo panaudotas įvairiems uždaviniams spręsti, bet nagrinėjamais atvejais ribojimai būdavo lygybės tipo.

Pirmuosius rezultatus apie būtinas minimumo sąlygas uždaviniams su nelygybės tipo ribojimais gavo W.Karush savo magistriniame darbe apgintame Čikagos universitete 1939 metais. Pažymėtina, kad tuo metu Čikagos universitete buvo aktyviai dirbama variacinio skaičiavimo srityje.

Kitas, taip pat likęs nepaskelbtu darbas apie būtinas minimumo są lygas uždaviniams su nelygybiniais ribojimais buvo parašytas F.John'o, kuris nagrinėjo uždavinius apie elipsoidų tūrius. F.John pateikė straipsnį 1939 metais solidžiam matematikos žurnalui *Duke Mathematical Journal,* kurio redakcija straipsnį atmetė. Darbas buvo paskelbtas 1948 metais knygoje, skirtoje žymaus amerikiečių matematiko R.Courant, kuris domėjosi optimizacija, šešiasdešimtmečiui. Gal būt, geometrinis uždavinys, kurį nagrinėdamas F.John išvedė būtinas minimumo sąlygas, ir nebuvo labai svarbus. Tačiau abu, W.Karush'o ir F.John'o, nepublikuotieji darbai dabar laikomi optimizacijos teorijos klasika.

Po to, kai 1951 metais H.Kuhn ir A.Tucker paskelbė garsųjį straipsnį apie būtinas minimumo sąlygas, ši tematika buvo pripažinta solidžia ir aktualia matematikos kryptimi. Beje, šiame straipsnyje buvo pirmą kartą pavartotas terminas *netiesinis programavimas.* Šiuo metu moksliniai straipsniai ir knygos, kuriuose remiamasi minimumo sąlygomis, skaičiuojami šimtais. Retas rezultatas cituojamas taip dažnai, kaip Kuhn - Tucker sąlygos.

3.9 Užduotys ir kontroliniai klausimai

1. Remdamiesi būtinomis minimumo sąlygomis raskite uždavinio

$$\min\,(x_1-2)^2+(x_2-2)^2,$$

$$\begin{aligned} -x_1 - x_2 &\leq 0, \\ -x_1 &\leq 0, \\ -x_2 &\leq 0, \end{aligned}$$

sprendinį.

2. Suformuluokite dualų uždavinį, kai

P: $\min(x_1^2 + x_2^2)$, esant ribojimui $-x_1 - x_2 + 4 \leq 0$, ir $A_0 = \{X : X \geq 0\}$. Išspręskite **P** ir **D** ir pavaizduokite sprendimą grafiškai.

3. Nubraižykite 1 pratimo baudos funkcijos (3.22), (3.23), (3.24) lygio linijas su $r = 0.5$, $r = 0.1$.

4. Nubraižykite 1 pratimo barjerų funkcijos (3.25) lygio linijas su $r = 0.5$, $r = 0.1$.

5. Apskaičiuokite 1 pratimo tikslo funkcijos gradiento taške (0, 1) projekcijas į pirmąjį ir antrąjį ribojimus.

6. Nubraižykite 1 pratimo Lagrange'o funkcijos lygio linijas, kai Lagrange'o daugikliai optimalūs: $\Lambda = \Lambda_0$.

7. Pateikite uždavinio (3.27) ir jo sprendimo eigos leistinųjų krypčių metodu grafinę iliustraciją: tikslo funkcijos lygio linijas, leistinąją sritį, gradientą taške (4, 7) ir leistinąją kryptį tame taške.

–

TIESINIS PROGRAMAVIMAS

4.1 Uždavinio formulavimas

Tiesiniu programavimu (TP) vadinamas optimizavimo uždavinių su tiesine tikslo funkcija ir tiesinėmis lygybėmis bei nelygybėmis apibrėžta leistinąja sritimi sprendimas. Vaizdus TP uždavinio atvejis yra pašarų raciono optimalaus sudarymo uždavinys. Sakysim, turime n pašarų rūšių, kurių m maistinių savybių nusako matrica A, t.y. a_{ij} reiškia i-tosios pašaro rūšies vienete esančių j-tojo tipo maistinių vienetų. Pavyzdžiui, pirmoji matricos eilutė reiškia kiekvienos pašarų rūšies vieneto kaloringumą, antroji - baltymų kiekį ir pan. Žinomos šėrimo normos reikalauja, kad gyvulys su pašarais gautų nemažiau negu tam tikrą kiekį b_i kiekvieno iš maistinių vienetų

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &\geq b_1, \\
a_{22}x_1 + a_{22}x_2 + ... + a_{2n}x_n &\geq b_2, \\
& ..... \\
a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &\geq b_m,
\end{aligned}$$

čia x_i reiškia i-tosios rūšies pašarų kiekį gyvulio racione. Kadangi įvairūs racionai gali patenkinti būtinus maistingumo reikalavimus, stengiamasi parinkti pigiausią, t.y. minimizuoti funkciją $f(X) = c_1x_1 + c_1x_1 + ... + c_1x_1$, kur c_i reiškia i-tosios pašaro rūšies vieneto kainą. Kadangi pašarų kiekis racione negali būti neigiamas, turi būti patenkinti ribojimai $x_i \geq 0$, kurie užrašomi atskirai, nes sudarant algoritmus gali būti įvertinti paprasčiau negu bendro pobūdžio tiesiniai ribojimai.

Kitas charakteringas TP uždavinys susijęs su pervežimų optimizavimu. Sakysim, nagrinėjamas produktas sandėliuojamas n sandėlių, kurių talpa a_i, $i = 1, ..., n$ vienetų. Produktas turi būti pristatomas į m vartojimo (ar pardavimo) punktų, kurių poreikiai yra b_j, $j = 1, ..., m$ vienetų. Reikia sudaryti pervežimų planą taip, kad visi vartotojų poreikiai būtų patenkinti ir bendros transporto išlaidos

$$\sum_{i,j} c_{ij}x_{ij}$$

būtų minimalios, kur x_{ij} yra produkto kiekis pervežamas iš i-tojo sandėlio j-tajam vartotojui, o c_{ij} produkto vieneto pervežimo kaina tokiu marštutu. Minėtos pervežimo sąlygos išreiškiamos šiomis lygybėmis ir nelygybėmis: $x_{ij} \geq 0$, $\sum_{i=1}^{n} x_{ij} = b_j$, $\sum_{j=1}^{m} x_{ij} \leq a_i$.

Bendruoju atveju TP uždavinys matriciniais žymėjimais gali būti užrašytas taip:

$$\min CX,$$

$$
\begin{aligned}
AX &= B, \\
A'X &\geq B', \\
x_1 &\geq 0, ..., x_r \geq 0,
\end{aligned}
\tag{4.1}
$$

čia kintamasis $X \in R^n$ yra n-matis vektorius, A - $l \times n$ matrica, A' $-$ $(m-l) \times n$ matrica, $\begin{pmatrix} B \\ B' \end{pmatrix} \in R^m$, X pirmosios r komponenčių turi būti neneigiamos, o kitų X komponenčių ženklas gali būti bet koks. TP uždaviniams būdingas kai kurių kintamųjų neneigiamumo reikalavimas; todėl natūralu ir kitas tiesines nelygybes užrašyti $\geq$ forma. Čia užrašyta tiesinio programavimo uždavinio *bendroji forma.* Dažnai patogiau nagrinėti atskirus bendrosios formos atvejus.

Tiesinio programavimo uždavinio *standartinė forma* yra tokia

$$\min CX, \tag{4.2}$$

$$AX = B, \ X \geq 0.$$

Tiesinio programavimo uždavinio *kanoninė forma* yra tokia

$$\min CX, \tag{4.3}$$

$$AX \geq B, \ X \geq 0.$$

Akivaizdu, kad (4.2) ir (4.3) yra daliniai (4.1) atvejai. Tačiau bet kurį TP uždavinį galima užrašyti ir standartine ir kanonine formomis. Užrašykime TP bedrąjį uždavinį standartine forma, kuri labai patogi simplekso algoritmo analizei. Nelygybinius ribojimus $DX \geq H$ galima pakeisti lygybiniais $DX - H - S = 0$ ir neneigiamumo $S \geq 0$ ribojimais. Keičiant bendrąjį uždavinį standartiniu, uždavinio kintamųjų skaičius padidėja tiek, kiek bendrajame uždavinyje yra nelygybinių ribojimų. Dėl uždavinio performulavimo apibrėžti kintamieji S dažnai vadinami *fiktyviaisiais.* Užrašykime bendrąjį uždavinį kanonine forma. Lygybinius ribojimus $DX = H$ galima pakeisti nelygybių pora: $DX \geq H$, $-DX \geq -H$. Keičiant bendrąjį uždavinį kanoniniu ribojimų skaičius dvigubai padidėja, nes lygybinis ribojimas pakeičiamas dviem nelygybiniais. Iš šių paprastų samprotavimų išplaukia visų trijų tiesinio programivimo uždavinio formų ekvivalentumas.

Kadangi TP uždavinio tikslo funkcija yra akivaizdžiai tolydi, tai esant aprėžtai leistinajai sričiai minimumo taškas visada egzistuoja. Kita vertus, jei leistinoji sritis neaprėžta, galima parinkti tokius tikslo funkcijos koeficientus, kad tikslo funkcijos reikšmės leistinojoje srityje neaprėžtai mažėtų. Pastarasis atvejis vadinamas *neaprėžtu TP uždaviniu.*

Tiesinio programavimo uždavinys yra specialus netiesinio programavimo uždavinio atvejis. Todėl jį galima spręsti ankstesniuose knygos skyriuose nagrinėtais metodais. Tačiau tikslo ir ribojimų funkcijų tiesiškumo specifika gali būti panaudota efektyvesniems algoritmams sukurti.

4.2 Geometrinė interpretacija

Tiesinėmis lygybėmis ir nelygybėmis apibrėžta sritis yra baigtinio skaičiaus puserdvių sankirta, vadinama poliedru (daugiasieniu). Aprėžtas poliedras vadinamas politopu. Galima parodyti, kad politopas yra baigtinio skaičiaus taškų (viršūnių) iškilus apvalkalas. Dvimačiu atveju leistinoji sritis yra daugiakampis, pavyzdžiui, parodytas 4.12 paveiksle.

Kadangi tikslo funkcija tiesinė, tai jos lygio linijos yra hiperplokštumos. Dvimačiu atveju lygio linijos yra tiesės, pavyzdžiui parodytos

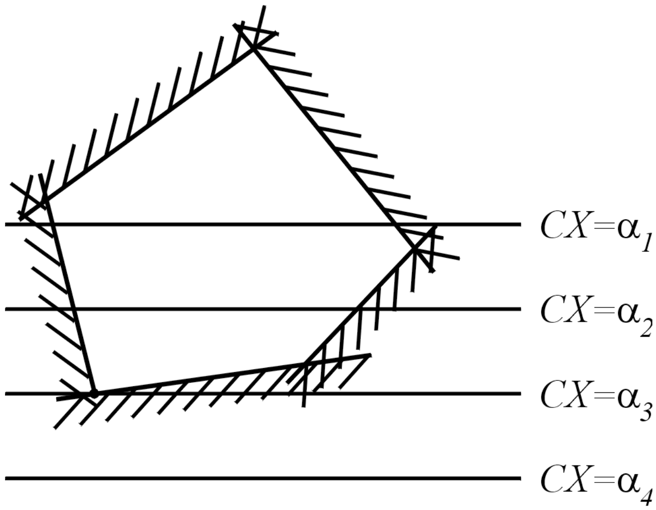

Figure 4.12: Geometrinė TP uždavinio iliustracija

4.12 paveiksle:

$$\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4.$$

Priminsime, kad būtinos minimumo sąlygos geometriškai interpretuojamos kaip tikslo funkcijos lygio linijos lietimasis su leistinąja sritimi. Tiesei liečiantis su daugiakampiu bent viena iš daugiakampio viršūnių yra lietimosi tašku. Todėl viena iš daugiakampio (bendru atveju poliedro) viršūnių yra tiesinio programavimo uždavinio sprendiniu. Paveiksle pavaizduotu atveju minimali tikslo funkcijos reikšmė lygi α_3.

4.3 Baziniai sprendiniai

Iš geometrinės tiesinio programavimo uždavinio interpretacijos išplaukia, kad leistinosios srities viršūnės labai svarbios uždavinio sprendimo algoritmui sudaryti, kadangi minimumo galima ieškoti peržiūrint viršūnes tikslo funkcijos mažėjimo kryptimi. Norint šiuo pagrindu kurti optimizavimo algoritmus, reikia, visų pirma, mokėti apskaičiuoti viršūnes. Geometrinį srities viršūnės vaizdinį atitinka algoritmiškai apskaičiuojami leistinieji baziniai sprendiniai.

Apibrėžtis. $V \in P \subset R^n$ yra poliedro P viršūnė, jei V nėra išreiškiama kaip jokių P taškų W, Y griežtai iškili kombinacija.

Galima įrodyti, kad šis apibrėžimas yra ekvivalentus geometriškai vaizdesniam apibrėžimui, kurio viršūnė apibrėžiama kaip n hiperplokštumų (poliedro sienų) susikirtimas.

Apibrėžtis. Sakysim A yra matrica, kurios dimensija lygi $m \times n$ ir rangas lygus m. Bet kurie m tiesiškai nepriklausomų A stulpelių sudaro bazę, t.y. matricos A stulpelių erdvės R^m bazę. Kad būtų trumpiau, dažnai baze vadinama bazinių stulpelių indeksų aibė.

TP uždavinį, užrašytą standartine forma, kurio matricos rangas lygus m, sutrumpintai žymėsime TPS.

Apibrėžtis. Sakysim $J = \{i_1, i_2, ..., i_m\}$ yra tiesinio programavimo uždavinio, užrašyto standartine forma, matricos A bazė, ir $i_1 < i_2 < ... < i_m$. Vektorius X vadinamas TPS *baziniu sprendiniu,* jei $x_{j_i} = z_i$, $Z = H^{-1}B$, $j_i \in J$, $x_{j_i} = 0$, $j_i \notin J$, čia matrica H yra sudaryta iš matricos A bazinių stulpelių. Kadangi matrica H neišsigimusi, tai bazinis sprendinys visada egzistuoja.

Apibrėžtis. Bazinis sprendinys vadinamas *leistinuoju baziniu sprendiniu* (LBS), jei $x_{j_i} \geq 0$, $j_i \in J$.

Iš leistino bazinio sprendinio apibrėžimo matyti, kad tai yra leistinosios srities taškas. LBS nepriklauso nuo tikslo funkcijos.

Pavyzdžiui, TPS

$$\min\,(x_1 + 2x_2 + 3x_3 + 4x_4),$$

$$\begin{aligned} 2x_1 + x_2 + 5x_3 + 2x_4 &= 10, \\ x_1 + x_2 - x_3 - x_4 &= 4, \\ x_i &\geq 0, \end{aligned} \tag{4.4}$$

bazę sudaro antras ir ketvirtas stulpeliai. Todėl šiuo atveju dviejų lygčių sistemos sprendinys $Z = H^{-1}B$ randamas taip

$$Z = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 10 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} 10 \\ 4 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}.$$

Šią bazę atitinkantis LBS yra keturmatis vektorius, kurio antra ir ketvirta komponentės lygios atitinkamoms vektoriaus Z komponentėms, o kitos komponentės lygios nuliui, t.y. $X = (0, 6, 0, 2)$. Bazė vienareikšmiai apibrėžia LBS, tačiau tą patį LBS gali apibrėžti kelios bazės. Tokiu atveju, LBS vadinamas išsigimusiu, o jo nulinių komponenčių skaičius yra didesnis negu $n - m$. Jeigu matricos A rangas lygus m, tai uždavinys turi bazinių sprendinių. Jei leistinoji aibė netuščia, tai uždavinys turi ir leistinųjų bazinių sprendinių.

11 teorema. TPS leistinosios srities taškas X yra srities viršūnė tada ir tik tada, kai jis yra leistinasis bazinis sprendinys.

Įrodymas. Sakysim X yra leistinosios srities viršūnė. Pažymėsime A_j matricos A j-tąjį stulpelį, ir $I = \{i : x_i > 0\}$.

Įrodymą pradėsime priimdami papildomą prielaidą, kad pasirinktajam X matricos A stulpeliai A_j, $j \in I$, tiesiškai nepriklausomi. Matricos A rangas lygus m ir $|I| \leq m$, todėl aibę I galime papildyti iki m elementinės aibės J, taip, kad A_j, $j \in J$, būtų tiesiškai nepriklausomi. Kadangi X yra leistinosios srities taškas, J yra matricos A bazė ir $x_i = 0$, $i \notin J$, tai X yra TPS leistinasis bazinis sprendinys.

Jeigu atsisakytume papildomos prielaidos apie tiesinę nepriklausomybę, tai X būtų tokia leistinosios srities viršūnė, kurios A_j, $j \in I$, būtų tiesiškai priklausomi. Tada egzistuotų ne visi lygūs nuliui koeficientai z_j, $j \in I$, tokie, kad

$$\sum_{j \in I} z_j A_j = 0.$$

Galėtume sudaryti vektorių Z, kurio komponentės būtų pastarieji koeficientai z_j, $j \in I$, o komponentės su indeksais $i \notin I$ būtų lygios 0. Akivaizdu, kad galiotų lygybė $AZ = 0$. Pažymėkime

$$\tau = \min_{z_j \neq 0} \frac{x_j}{|z_j|}.$$

Kadangi z_j gali būti nelygus nuliui tik kai $j \in I$, o šitokiems indeksans $x_j > 0$, tai $\tau > 0$. Sudarykime vektorius

$$X^+ = X + \tau Z,\ X^- = X - \tau Z.$$

Abu vektoriai patenkina TPS lygybinius ribojimus: $AX^\pm = AX \pm \tau(AZ) = B$. Kai $j \notin I$, $x_j^\pm = x_j \geq 0$. Kai $j \in I$, tai $\tau|z_j| \leq \frac{x_j}{|z_j|} \cdot |z_j| = x_j$, todėl teisinga nelygybė $x_j^\pm = x_j \pm \tau z_j \geq 0$. Įrodėme, kad abu vektoriai $X^\pm$ yra leistinieji sprendiniai nelygūs X, nes $\tau > 0$, $Z \neq 0$. Bet tada X negalėtų būti leistinosios srities viršūne, nes $X = \frac{1}{2}X^+ + \frac{1}{2}X^-$. Reiškia, prielaidos, kad A_j, $j \in I$, tiesiškai nepriklausomi, atsisakymas lemia prieštarą. Tuo ir baigiame įrodymą, kad leistinosios srities viršūnė yra leistinas bazinis sprendinys.

Sakysim X yra bazinis sprendinys. Jeigu priimtume prielaidą, kad X nėra leistinosios srities viršūnė, tai egzistuotų leistinieji sprendiniai $Y \neq X$, $Z \neq X$ tokie, kad $X = \lambda Y + (1 - \lambda)Z$, $1 < \lambda < 1$, Kadangi X yra vienintelis leistinasis vektorius, kurio komponentės $x_i = 0$, $i \notin I$, čia I- bazė, atitinkanti X, tai bent vienam $j \notin I$ galioja nelygybė $y_j > 0$. Kadangi $z_j \geq 0$, tai teisinga nelygybė

$$\lambda y_j + (1 - \lambda) z_j > 0,$$

reiškianti, kad $X \neq \lambda Y + (1-\lambda)Z$. Gauta prieštara įrodo, kad leistinasis bazinis sprendinys yra leistinosios srities viršūnė.

Išvada. Leistinosios srities viršūnių skaičius ne didesnis negu m elementų išrinkimo iš n elementų aibės kombinacijų skaičius C_n^m; dažnai tas skaičius žymimas $\begin{pmatrix} n \\ m \end{pmatrix}$.

12 teorema. Sakysim W yra TPS leistinosios srities taškas. Arba egzistuoja leistinosios srities viršūnė V, tokia kad $CV \leq CW$, arba TPS neaprėžtas.

Įrodymas. Teoremos įrodymas gan ilgas. Todėl iš pradžių pateiksime įrodymo idėją. Nagrinėjamas koks nors leistinosios srities taškas W. Ieškosime kito leistinosios srities taško, kuriame kaip lygybė patenkinamas dar bent vienas ribojimas ir tikslo funkcijos reikšmė mažesnė už CW. Jei surastajame taške kaip lygybės patenkinami n ribojimai, tai surastasis taškas yra leistinosios srities viršūnė. Priešingu atveju įrodysime, kad leistinojoje srityje galima rasti taškų, kuriuose tikslo funkcijos reikšmės norimai mažos.

Kadangi W leistinosios srities taškas, tai $AW = B$. Sunumeruokime visus ribojimus pradėdami nuo kintamųjų neneigiamumo: $1, ..., n+m$. Iš ribojimų, kurie taške W patenkinami kaip lygybės, aibės išrinkime didžiausią tiesiškai nepriklausomų ribojimų aibę ir jų indeksų aibę pažymėkime I_0; $r = |I_0|$, $I = I_0 \cap \{1, ..., n\}$.

Jei $r = n$, tai W yra n hiperplokštumų susikirtimas, t.y. leistinosios srities viršūnė. Šiuo atveju teoremos teiginys trivialus.

1 teiginys. Kadangi buvo išrinkta didžiausia galima aibė I_0, tai ribojimas $x_j \geq 0$, $j \notin I_0$, patenkinamas taške W kaip lygybė, yra tiesiškai priklausomas nuo ribojimų, kurių indeksai sudaro aibę I_0.

Pažymėsime

$$\Omega = \{X : x_j = 0,\ j \in I,\ AX = 0\}$$

ir pastebėsime, kad Ω yra $n-r$ matavimų poerdvis. Raskime, pvz.

Gauso eliminavimo metodu, nenulinį Ω vektorių U. Galima U parinkti taip, kad $CU \leq 0$; jei surastajam vektoriui U galiotų priešinga nelygybė, galėtume pakeisti jo komponenčių ženklus ir turėtume vektorių su norima savybe.

2 teiginys. Jei $w_j = 0$, tai ir $u_j = 0$.

Įrodymas. Jei $j \in I$, tai $u_j = 0$ pagal Ω apibrėžimą.

Sakysim, $w_j = 0$, $j \notin I$. Remdamiesi 1 teiginiu, galime išreikšti j-tąjį aktyvų taške W ribojimą, tiesine ribojimų, kurių indeksai priklauso I, kombinacija. Vietoj w_i šioje tiesinėje kombinacijoje įrašę u_i gausime $u_j = 0$, kadangi dešiniosios lygybių pusės Ω apibrėžime lygios 0.

Bet kokiems skaliarams λ teisinga lygybė

$$A(W + \lambda U) = AW + \lambda \cdot 0 = B.$$

Reiškia, $W + \lambda U$ patenkina lygybinius ribojimus. Anksčiau įrodyta (2 teiginys), kad $u_j = 0$, kai $w_j = 0$, todėl ir $w_j + \lambda u_j = 0$, kai $w_j = 0$. Įrodysime, kad

◦ arba galima parinkti norimai didelį λ nepažeidžiant ribojimų ir taip gauti norimai mažą tikslo funkcijos reikšmę,

◦ arba galima parinkti tokią λ reikšmę, kad taške $W + \lambda U$ būtų patenkinta dar bent vienu lygybiniu ribojimu daugiau negu taške W ir $C(W + \lambda U) \leq CW$. Naujasis lygybe tapęs neneigiamumo ribojimas tiesiškai nepriklausomas nuo ribojimų su indeksais iš I_0.

Pažymėkime $J = \{j : u_j < 0\}$. Galimi trys atvejai.

1 atvejis. $CU < 0$, $J = \emptyset$.

Šiuo atveju visi $u_j \geq 0$. Todėl $W + \lambda U$ priklauso leistinąjai sričiai prie bet kokio $\lambda \geq 0$, ir

$$C(W + \lambda U) = CW + \lambda CU.$$

Kadangi $CU < 0$, tai parinkdami kiek norima didelį λ gausime tašką leistinojoje srityje su norimai maža tikslo funkcijos reikšme.

2 atvejis. $J \neq \emptyset$.

Pažymėkime

$$s = \min_{j \in J} \frac{w_j}{-u_j} \,.$$

Kadangi visiems $j \in J$ galioja nelygybės $w_j > 0,\ u_j < 0,$ tai $s > 0$; priminsime, kad $u_j = 0$, jei $w_j = 0$. Parinkime $l \in J$ taip, kad $w_l/(-u_l) = s$ ir pažymėkime $Z = W + sU$. Visiems $j \in J$ galioja

$$z_j = w_j + su_j \geq w_j + \frac{w_j}{-u_j} u_j = 0,$$

o visiems $j \notin J$ tokia pat nelygybė galioja pagal pagal J apibrėžimą, nes W yra leistinosios srities taškas. Todėl Z yra leistinosios srities taškas. Be to, $z_l = w_l + su_l = 0$, kai $w_l > 0$. Reiškia, taške Z kaip lygybė patenkintas l-tasis ribojimas, kuris taške W buvo griežtas, ir galioja nelygybė

$$CZ = CW + sCU \leq CW.$$

3 atvejis. $CU = 0,\ J = \emptyset$.

Visiems j galioja nelygybė $u_j \geq 0$, ir bent vienam j nelygybė griežta, kadangi vektorius U yra nenulinis. Todėl teisinga nelygybė $s > 0$, čia

$$s = \min_{u_j > 0} \frac{w_j}{u_j} \,.$$

Parinkime $l \in J$ taip, kad $w_l/u_l = s$ ir pažymėkime $Z = W - sU$. Akivaizdu, kad Z patenkina lygybinius ribojimus $AZ = B$, ir $z_l = 0$.

Sakysim j toks, kad $w_j > 0,\ u_j > 0$. Kadangi $s \leq w_j/u_j$, tai

$$z_j \geq w_j - (w_j/u_j)u_j = 0.$$

Tiems j, kuriems galioja $w_j \geq 0,\ u_j = 0$, nelygybė $z_j \geq 0$ akivaizdžiai patenkinama. Reiškia, Z yra leistinosios srities taškas. Jame kaip lygybė patenkinamas l-tasis ribojimas, kuris taške W buvo griežtas, ir galioja lygybė

$$CZ = CW - sCU = CW.$$

Išvada. Jei TPS aprėžtas, t.y. egzistuoja (kad ir labai mažas) skaičius $w > -\infty$, toks, kad visiems leistiniems sprendiniams X galioja nelygybė $CX \geq w$, tai TPS uždavinys yra baigtinis. Tas reiškia, kad sprendiniui surasti užtenka atlikti baigtinį žingsnių skaičių. Sakysim, Y_1 koks nors leistinosios srities taškas. Remiantis 12 teorema galima surasti leistinosios srities viršūnę V_1 tokią, kad $CV_1 \leq CY_1$; tam užtenka patikrinti leistinosios srities viršūnes. Arba V_1 yra uždavinio sprendinys, arba V_1 aplinkoje galima rasti leistinosios srities tašką Y_2 tokį, kad $CY_2 < CV_1$. Remiantis 12 teorema galima surasti leistinosios srities viršūnę V_2 tokią, kad $CV_2 \leq CY_2$ ir tt. Kadangi kiekviename žingsnyje randama viršūnė su griežtai mažesne tikslo funkcijos reikšme, tai po baigtinio žingsnių skaičiaus bus surasta optimali viršūnė.

4.4 Atraminis bazės keitimas

Geometrinė idėja, kaip tiesinio programavimo uždavinio sprendinio ieškoti pereinat iš vienos leistinosios srities viršūnės į kitą, algoritmiškai gali būti realizuota bazės keitimu. Panagrinėsime pavyzdį (4.4), kurio bazę $\{2,\ 4\}$ atitinka leistinas bazinis sprendinys $X = (0, 6, 0, 2)$, t.y.

$$x_2A_2 + x_4A_4 = 6\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 2 \\ -1 \end{pmatrix} = B = \begin{pmatrix} 10 \\ 4 \end{pmatrix}.$$

Pabandykime padidinti nebazinę komponentę x_3 nuo nulio iki kokios nors reikšmės λ taip, kad lygybiniai ribojimai (4.4) liktų patenkinti. Siekdami gauti bazinį sprendinį, atitinkantį gretimą leistinosios srities viršūnę, turime bazėje pakeisti tik vieną stulpelį. Kadangi nebazinis A_3 įkeliamas į bazę, kažkuris iš bazinių stulpelių turės būti iš bazės iškeltas. Kiti nebaziniai stulpeliai (šiuo atveju A_1) lieka nebaziniais. Pažymėję naujas kintamųjų, atitinkančių bazines komponentes, reikšmes su štrichais, gausime tokią lygtį

$$x_2' A_2 + x_4' A_4 + \lambda A_3 = B, \tag{4.5}$$

Nebazinis stulpelis visada gali būti išreikštas kaip tiesinė bazinių stulpelių kombinacija, kurios koeficientus bendru atveju galima apskaičiuoti sprendžiant tiesinių lygčių sistemą. Nagrinėjamam pavyzdžiui lengva matyti, kad $A_3 = A_2 + 2A_4$. Įstatčius šią išraišką į (4.5), gaunama tokia lygtis

$$(x_2' + \lambda)A_2 + (x_4' + 2\lambda)A_4 = B. \tag{4.6}$$

Lygybiniai ribojimai (4.4) tikrai būtų patenkinti, jei lygtyje (4.6) daugikliai prie bazinių stulpelių būtų lygūs bazinėms X komponentėms: $x_2 = x_2' + \lambda$, $x_4 = x_4' + 2\lambda$. Gautasis vektorius bus naujas leistinas bazinis sprendinys, jei $x_2' \geq 0$, $x_4' \geq 0$, $\lambda > 0$, ir vienas iš x_i' taps lygiu nuliui. Kadangi $x_2 = 6$ ir $x_4 = 2$, tai nesunku rasti tinkamą reikšmę $\lambda = 1$. Lygybė $x_4' = 0$ reiškia, kad ketvirtas stulpelis iškeliamas iš bazės. Nauja bazė yra $\{2, 3\}$, o bazinis sprendinys lygus (0,5,1,0).

Aprašyto bazės keitimo esmė yra nebazinės komponentės didinimas nuo nulio iki tol, kol kuris nors kitas kintamasis pasieks (atsirems į) ribojimą. Kai reikės terminą trumpinti, tokį keitimą vadinsime *ABAKS* (Atraminis BAzės KeitimaS). Angliškai toks procesas vadinamas *pivoting.*

Sakysim, žinoma bazė, kurią atitinka leistinasis bazinis sprendinys. Algoritmą patogiau aprašyti vartojant ne tikrus bazinių stulpelių indeksus, o sąlyginius, atitinkančius *sutvarkytą bazę* $\mathcal{B} = \{\mathcal{B}(1), ..., \mathcal{B}(m)\}$; nagrinėtuoju atveju

$\mathcal{B} = \{\mathcal{B}(1), \mathcal{B}(2)\} = \{2, 4\}$. Nebaziniai stulpeliai gali būti užrašyti kaip tiesinės bazinių stulpelių kombinacijos

$$A_j = \sum_{i=1}^{m} x_{ij} A_{\mathcal{B}(i)} \ , \tag{4.7}$$

čia x_{ij} žymi koeficientą prie $\mathcal{B}(i)-$tojo bazinio stulpelio išreiškiant A_j tiesine bazinių stulpelių kombinacija. Nagrinėtu atveju buvo $x_{13} = 6$, $x_{23} = 2$. Abaksu įnešant nebazinį $j-$tąjį stulpelį į bazę, bazinio sprendinio X_0 $j-$toji komponentė didinama nuo 0 iki λ patenkinant lygybinius ribojimus. Sudėję lygybinių ribojimų sąlygą vektoriui X_0 ir (4.7) padaugintą iš λ, gausime lygybę

$$\sum_{i=1}^{m} x_{i0} A_{\mathcal{B}(i)} + \lambda(A_j - \sum_{i=1}^{m} x_{ij} A_{\mathcal{B}(i)}) = B,$$

kurią truputį pertvarkę

$$\sum_{i=1}^{m} (x_{i0} - x_{ij}\lambda) A_{\mathcal{B}(i)} + \lambda A_j = B \tag{4.8}$$

galime panaudoti naujo bazinio vektoriaus apskaičiavimui. Sudarykime vektorių, kurio j-toji komponentė lygi $x'_{j0} = \lambda$, $\mathcal{B}(i)$-toji komponentė lygi $x'_{i0} = x_{i0} - x_{ij}\lambda$, o kitos komponentės lygios 0. Lygybė (4.8) garantuoja, kad TPS lygybiniai ribojimai patenkinti. Belieka užtikrinti naujojo vektoriaus komponenčių neneigiamumą parenkant

$$\lambda = \min_{i,\, x_{ij} > 0} \frac{x_{i0}}{x_{ij}}. \tag{4.9}$$

Naujoji bazė, atitinkanti sudarytą bazinį vektorių X'_0, yra $\mathcal{B}'(i) = \mathcal{B}(i)$, $i \neq k$, $\mathcal{B}'(k) = j$, kur k yra minimumo indeksas (4.9). Apskaičiuojant λ pagal (4.9) galimi tokie atvejai:

- nėra teigiamų x_{ij},
- gauta λ reikšmė lygi 0.

Pirmuoju atveju leistinoji TPS sritis neaprėžta, ir norimai didinant λ gautas taškas nepažeidžia ribojimų. Antruoju atveju gautas bazinis sprendinys yra išsigimęs.

4.5 Simplekso algoritmas

Tiesinio programavimo uždavinio sprendimo idėja buvo anksčiau pateikta geometriškai: pereiti iš vienos leistinosios srities viršūnės į kaimyninę viršūnę su mažesne tikslo funkcijos reikšme. Perėjimas į kaimyninę viršūnę gali būti realizuotas abaksu. Pasirinkus nebazinį stulpelį, kurs įkeliamas į bazę, abaksas automatiškai nustato iš bazės iškeliamą stulpelį. Tikslo funkcijos reikšmės mažėjimą užtikrinantis į bazę įkeliamo stulpelio parinkimas ir bazės keitimas abaksu sudaro *simplekso algoritmo* pagrindą. Atrodytų, kad įkeliamą į bazę stulpelį tikslinga parinkti taip, kad tikslo funkcijos reikšmė gautajame abaksu taške būtų minimali. Tačiau toks parinkimas, nors ir optimalus vieno žingsnio požiūriu, nebūtinai bus geras viso sprendimo proceso požiūriu. Perėjimas į kaimyninę viršūnę, kurioje tikslo funkcijos reikšmė minimali, gali būti nelabai racionalus todėl, kad naujoji viršūnė gali būti neperspektyvi tolimesnės paieškos požiūriu. Kitaip sakant, mažesnis pagerinimas šiame žingsnyje gali atsipirkti geresne vėliau daromų žingsnių perspektyva. Įvertinti kelių algoritmo žingsnių rezultatą labai sudėtinga; žymiai paprasčiau nustatyti, ar tikslo funkcijos reikšmė sumažės, negu įvertinti, kokio dydžio bus sumažėjimas.

Atraminį bazės keitimą patogu formalizuoti lentelės perskaičiavimo forma. Kad būtų suprantamiau, pradėkime jau nagrinėtu pavyzdžiu. Užrašykime TPS lygybinius ribojimus (4.4) lentele

	x1	x2	x3	x4
10	2	1	5	2
4	1	1	-1	-1

.

Anksčiau nagrinėdami šį pavyzdį, bazę tiesiog atspėjome. Bazės Apibrėžtis nusako ir metodą jai surasti: tereikia rasti m nepriklausomų matricos A stulpelių. Iš tiesinės algebros kurso žinoma, kad tiesinės operacijos su uždavinio matricos eilutėmis nekeičia lygčių sistemos sprendinio. Todėl galima tiesiškai nepriklausomų stulpelių visus, išskyrus vieną, elementus pakeisti nuliais. Šiuo metodu ne tik surandami tiesiškai nepriklausomi stulpeliai, bet ir jų sudaroma matrica pakeičiama vienetine. Tas patogu todėl, kad nulinio stulpelio elementai tada tiesiog lygūs nenuliniams bazinio sprendinio elementams. Pavyzdžiui, pasirinkę pirmo stulpelio antros eilutės elementą (patogus skaičiavimui, nes lygus 1) galime to paties stulpelio pirmos eilutės elementą pakeisti 0 (padauginę antrąją eilutę iš 2 ir atėmę iš pirmosios eilutės):

	x1	x2	x3	x4
2	0	-1	7	4
4	1	1	-1	-1

.

.

Pridėję pirmąją eilutę prie antrosios ir po to pirmąją eilutę padauginę iš -1 gausime lentelę

	x1	x2	x3	x4
-2	0	1	-7	-4
6	1	0	6	3

,

iš kurios akivaizdu, kad baziniais stulpeliai yra pirmasis ir antrasis $\mathcal{B}(1) = 2$, $\mathcal{B}(2) = 1$. Šią bazę atitinkantis bazinis sprendinys $(6, -2\,, 0\,, 0)$ nėra leistinas, nes jo antroji komponentė yra neigiama. Taigi, nors bazę nesunku surasti metodu panašiu į Gauso eliminavimo metodą tiesinių lygčių sistemoms spręsti, bet toks metodas negarantuoja, kad surastą bazę atitinkantis bazinis sprendinys bus leistinas. Bendras metodas leistinajam baziniam sprendiniui surasti pateiktas skyrelyje "Startas". Čia, nesigilindami į šį klausimą, tarsim, kad bazės paieška buvo sėkmingesnė, ir buvo surasta bazė $\mathcal{B} = \{2, 4\}$ su leistinuoju baziniu sprendiniu $X_0 = (0, 6, 0, 2)$:

	x1	x2	x3	x4
6	4/3	1	1	0
2	1/3	0	2	1

.

Kai baziniai stulpeliai yra vienetiniai, labai lengva nebazinius stulpelius užrašyti kaip tiesines bazinių stulpelių kombinacijas ir išspręsti lygčių sistemą su perskaičiuotu (4.4) dešiniųjų pusių vektoriumi β. Nagrinėjamuoju atveju, pavyzdžiui, galima išreikšti $A_3 = \sum_{i=1}^{m} x_{i3} A_{\mathcal{B}(i)}$, $x_{13} = 1$, $x_{23} = 2$:

$$\begin{pmatrix} 5 \\ -1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2 \cdot \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Lygčių sistemos

$$\sum_{i=1}^{m} x_{i0} A_{\mathcal{B}(i)} = \beta,\ \beta = (6,\ 2),$$

sprendinys lygus $x_{10} = 6$, $x_{20} = 2$. Iš bazės iškeliamas stulpelis nustatomas pagal (4.9)

$$\lambda = \min_{i,\, x_{ij} > 0} \frac{x_{i0}}{x_{i3}} = \min\left(\frac{6}{1}, \frac{2}{2}\right) = 1, \quad k = 2. \tag{4.10}$$

Todėl, įkeliant į bazę trečią stulpelį, iš bazės iškeliamas $\mathcal{B}(2)$, t.y. ketvirtasis stulpelis. Naujai gautą bazę vėl patogu sudaryti iš vienetinių stulpelių. Todėl lentelę perskaičiuojame padalindami antrąją eilutę iš 2 ir atimdami gautą rezultatą iš pirmosios eilutės:

	x1	x2	x3	x4
5	7/6	1	0	-1/2
1	1/6	0	1	1/2

.

Iki šiol nagrinėjome bazės keitimą apskritai neatsižvelgdami į tikslo funkciją. Tačiau toks keitimas atliekamas siekiant sumažinti tikslo funkcijos reikšmę. Perėjimas į kitą bazinį sprendinį realizuojamas dviem žingsniais: įkeliamo į bazę stulpelio išrinkimo ir abakso. Pastarasis yra vienareikšmiškai apibrėžtas įkeliamo į bazę stulpelio išrinkimu. Todėl stulpelis turėtų būti išrenkamas atsižvelgiant į išrinkimo sąlygotą tikslo funkcijos reikšmės pokytį. Panagrinėsime, kaip gali būti įvertintas toks pokytis. Tikslo funkcijos reikšmė bazinio sprendinio X_0 taške lygi

$$F = \sum_{i=1,\ i\neq k}^{m} c_{\mathcal{B}(i)} x_{i0} + c_{\mathcal{B}(k)} x_{k0},$$

čia atskirai parašytas dėmuo, atitinkantis iškeliamą iš bazės stulpelį. Naujajame baziniame sprendinyje, gautame įkėlus į bazę $j-$tąjį stulpelį ir iškėlus $k-$tąjį, tikslo funkcijos reikšmė lygi

$$F' = \sum_{i\neq k} c_{\mathcal{B}(i)}(x_{i0} - \lambda x_{ij}) + c_j \lambda.$$

Tikslo funkcijos reikšmių skirtumas, atsižvelgiant į lygybę $x_{k0} = \frac{x_{k0}}{x_{kj}} x_{kj} = \lambda x_{kj}$, išreiškiamas šitaip

$$F' - F = \lambda(c_j - \sum_{i=1}^{m} c_{\mathcal{B}(i)} x_{ij}). \qquad (4.11)$$

Akivaizdu, kad **bazės pakeitimas sumažina tikslo funkcijos reikšmę, jei (4.11) skliausteliuose esantis reiškinys neigiamas**, t.y. tokiu atveju $j-$tasis stulpelis gali būti išrinktas įkelti į bazę.

Veiksmai, kuriuos reikia atlikti apskaičiuojant (4.11), labai panašūs į veiksmus, atliekamus realizuojant atraminį bazės keitimą. Todėl kiekvieno stulpelio įkėlimo į bazę kriterijų galima apskaičiuoti atliekant veiksmus su ABAKS lentelės eilutėmis. Panagrinėkime (4.4) pavyzdį. Nulinė pradinės lentelės eilutė, kurioje buvo užrašyta nereikšminga informacija (x1, x2, x3, x4), pakeičiama tikslo funkcijos koeficientų eilute:

0	1	2	3	4
10	2	1	5	2
4	1	1	-1	-1

.

Suradus bazę, iš nulinės eilutės atimamos lentelės eilutės padaugintos iš $c_{\mathcal{B}(i)}$. Nagrinėjamuoju atveju iš nulinės eilutės būtų atimta pirmoji eilutė padauginta iš 2 ir antroji padauginta iš 4:

-20	-3	0	-7	0
6	4/3	1	1	0
2	1/3	0	2	1

.

Pastebėsime, kad nulinės eilutės ir nuliniame stulpelyje gaunama su priešingu ženklu tikslo funkcijos reikšmė, atitinkanti einamąjį bazinį sprendinį. Abiejų nebazinių stulpelių kriterijų reikšmės neigiamos, todėl, įkeliant į bazę bet kurį iš jų, galima sumažinti tikslo funkcijos reikšmę. Pasirinkus trečiąjį stulpelį, iškeliamo iš bazės stulpelio numeris nustatomas šitaip

$$\lambda = \min_{i,x_{ij}>0} \frac{x_{i0}}{x_{i3}} = \min\left(\frac{6}{1}, \frac{2}{2}\right) = 1, \quad k = 2$$

Atlikę tuos pačius veiksmus, kaip ir aukščiau, gausime naują bazę $\mathcal{B} = \{2,\ 3\}$ su leistinuoju baziniu sprendiniu $X_0 = (0,\ 5,\ 1,\ 0)$. Prie nulinės eilutės pridėjus antrąją eilutę padaugintą iš 7/2, nulinės eilutės trečiajame stulpelyje gaunamas nulinis elementas:

-13	-11/6	0	0	7/2
5	7/6	1	0	-1/2
1	1/6	0	1	1/2

,

o nulinės eilutės nulinio stulpelio elemento reikšmė tampa lygi einamojo bazinio sprendinio tikslo funkcijos reikšmei su priešingu ženklu, t.y. šiame baziniame sprendinyje tikslo funkcijos reikšmė yra lygi 13. Iš gautosios lentelės taip pat matosi, kad tikslo funkcijos reikšmę galima sumažinti įkeliant į bazę pirmąjį stulpelį. Tada iš bazės reikėtų iškelti antrąjį stulpelį, nes

$$\begin{aligned} \lambda &= \min_{i,x_{ij}>0} \frac{x_{i0}}{x_{i1}} = \min\left(\frac{30}{7}, 6\right) = \frac{30}{7}, \\ k &= 1,\ \mathcal{B}(1) = 2. \end{aligned}$$

Atlikus atraminį bazės keitimą, gaunamas rezultatas

-36/7	0	11/7	0	19/7
30/7	1	6/7	0	-3/7
2/7	0	-1/7	1	4/7

,

reiškiantis, kad tolimesnis bazės keitimas nebegali pagerinti tikslo funkcijos reikšmės, nes visų nebazinių stulpelių kriterijų reikšmės teigiamos. Šio uždavinio optimalus bazinis sprendinys ir optimali tikslo fukcijos reikšmė yra šitokie

$$X_* = (30/7,\ 0,\ 2/7,\ 0),\ CX_* = 36/7.$$

Bendruoju atveju sąlygą

$$c_j - \sum_{i=1}^{m} c_{\mathcal{B}(i)} x_{ij} > 0$$

matricine forma galima užrašyti šitaip

$$C - C_{\mathcal{B}}\mathcal{X} \geq 0,$$

čia $C_{\mathcal{B}} = (c_{\mathcal{B}(1)}, ..., c_{\mathcal{B}(m)})$, ir $\mathcal{X}$ yra koeficientų x_{ij}, apibrėžtų lygybe (4.7), matrica $\mathcal{X} = \mathcal{H}_{\mathcal{B}}^{-1}A$, $\mathcal{H}_{\mathcal{B}}$ – bazę $\mathcal{B}$ atitinkanti stulpelių matrica.

13 teorema. Jei $C - C_{\mathcal{B}}\mathcal{X} \geq 0$, tai nagrinėjamasis bazinis sprendinys yra optimalus.

Įrodymas. Sakysim, Y yra leistinosios srities vektorius (nebūtinai bazinis): $AY = B$, $Y \geq 0$. Pažymėkime $Z = C_{\mathcal{B}}\mathcal{X}$. Pagal teoremos prielaidą visiems j galioja nelygybė $c_j \geq z_j$. Todėl teisinga šitokia nelygybė

$$\begin{aligned} CY &\geq ZY = (C_{\mathcal{B}}\mathcal{H}_{\mathcal{B}}^{-1}A)Y \\ &= C_{\mathcal{B}}\mathcal{H}_{\mathcal{B}}^{-1}(AY) = C_{\mathcal{B}}\mathcal{H}_{\mathcal{B}}^{-1}B. \end{aligned} \tag{4.12}$$

Tikslo funkcijos reikšmė nagrinėjamajame baziniame sprendinyje X_0 lygi

$$\sum_{i=1}^{m} c_{\mathcal{B}(i)} x_{i0} = \sum_{i=1}^{m} c_{\mathcal{B}(i)} (\mathcal{H}_{\mathcal{B}}^{-1}B)_i = C_{\mathcal{B}}\mathcal{H}_{\mathcal{B}}^{-1}B. \tag{4.13}$$

Palyginę (4.12) ir (4.13) matome, kad $CX_0 \leq CY$, t.y. nagrinėjamajame baziniame sprendinyje tikslo funkcijos reikšmė ne blogesnė negu bet kokiame kitame leistinosios srities taške. Reiškia, X_0 yra optimalus bazinis sprendinys.

4.6 Startas

Simplekso algoritmu galime rasti optimalų sprendinį, bet pradėti turime nuo kokio nors leistinojo bazinio sprendinio. Kaip jį sužinoti? Sakysim, sprendžiamas uždavinys

$$\min CX,$$

$$AX \leq B,\ X \geq 0,$$

ir $B \geq 0$. Tada papildę uždavinį fiktyviais neneigiamais kintamaisiais, kurių vektorių pažymėsime X', ir papildę matricą A vienetine matrica I, pakeičiame pradinį uždavinį šitokiu

$$\min CX,$$

$$(A, I)\begin{pmatrix} X \\ X' \end{pmatrix} = B,\ X \geq 0,\ X' \geq 0.$$

Pastarasis uždavinys ekvivalentus pradiniam ir turi akivaizdų leistiną bazinį sprendinį $X' = B$, nuo kurio ir galima pradėti tikslo funkcijos minimizavimą simplekso algoritmu.

Jei leistinasis bazinis sprendinys nėra toks akivaizdus, tenka spręsti dviejų fazių uždavinį. Papildydami uždavinį neneigiamais fiktyviais kintamaisiais, visus nelygybinius ribojimus galime pakeisti šitokiais lygybiniais

$$\begin{aligned}
\sum_i a_{ji}x_i &\geq b_j > 0 &&\to \sum_i a_{ji}x_i - x'_j = b_j, \\
\sum_i a_{ji}x_i &\geq b_j,\ b_j < 0 &&\to -\sum_i a_{ji}x_i + x'_j = -b_j, \\
\sum_i a_{ji}x_i &\leq b_j < 0 &&\to -\sum_i a_{ji}x_i - x'_j = -b_j,
\end{aligned}$$

su neneigiama dešiniąja puse. Gautojo TPS uždavinio kintamųjų (tikrų ir fiktyvių) vektorių pažymėsime X, ribojimų matricą, pašalinę tiesiškai priklausomas eilutes, pažymėsime Γ, o jos rangą m. Uždavinys

$$\min CX,$$

$$\begin{aligned} \Gamma X &= \mathrm{abs}(B), \\ X &\geq 0, \end{aligned} \tag{4.14}$$

ekvivalentus pradiniam; čia abs(B) reiškia vektorių, sudarytą iš vektoriaus B elementų absoliutinių didumų. Pastarąjį uždavinį galime pakeisti (papildydami jį m fiktyviais neneigiamais kintamaisiais, sudarančiais vektorių X'') kitu TPS uždavniu

$$\min \sum_{i=1}^{m} x_i'',$$

$$(\Gamma, I) \cdot (X, X'')^t = \mathrm{abs}(B),\ X \geq 0,\ X'' \geq 0,$$

kurio leistinasis bazinis sprendinys akivaizdus $X'' = \mathrm{abs}(B)$. Išsprendę pastarąjį uždavinį simplekso metodu, galime gauti optimalų bazinį sprendinį atitinkantį nulinę optimalią kainą. Tada gautasis sprendinys yra (4.14) uždavinio leistinuoju baziniu sprendiniu. Todėl pradėdami šiuo sprendiniu galime spręsti (4.14) uždavinį simplekso algoritmu ir gauti pradinio uždavinio sprendinį. Tačiau gali atsitikti ir taip, kad (4.14) uždavinio optimali reikšmė nebus lygi nuliui. Tas reikštų, kad pradinis uždavinys neturi leistino sprendinio.

Jei pradinio uždavinio leistinas bazinis sprendinys nėra akivaizdus, tenka spręsti tiesinio programavimo uždavinį du kartus. Išsprendus pirmosios fazės uždavinį gaunamas leistinasis bazinis pradinio uždavinio sprendinys. Antroji fazė - pradinio uždavinio sprendimas simplekso metodu pradedant leistinuoju baziniu sprendiniu, surastu sprendžiant pirmos fazės uždavinį.

4.7 Dualumas TP

Bendras dualumo atvejis išnagrinėtas ankstesniame skyriuje. Todėl nesunku performuluoti gautus rezultatus TP terminais. Kita vertus, dualus uždavinys TP atveju formuluojamas paprasčiau ir yra vaizdesnis negu bendru atveju. Gal būt, silpnesnio matematinio pasiruošimo skaitytojas praleido paragrafą, skirtą dualumui netiesiniame programavime, dėl tame pagrarafe vartoto šiek tiek sudėtingesnio matematinio aparato. Todėl paaiškinsime dualumo idėją paprastu TP pavyzdžiu.

Panagrinėkime **pirminį** TP uždavinį

$$\begin{array}{rcl} \min\,(x_1 + 2x_2 - 7x_3),\; x_i & \geq & 0, \\ 3x_1 - x_2 + 5x_3 & = & 6, \\ 2x_1 + 0x_2 - x_3 & = & 12, \end{array}$$

kuris yra užrašytas standartine forma. Sakysim, X yra leistinas šio uždavinio sprendinys, o $Y = (y_1, y_2)$ naujų kintamųjų vektorius. Susumavę ribojimų funkcijas padaugintas iš y_i gausime funkciją

$$\begin{array}{rcl} \Theta(Y) & = & (3x_1 - x_2 + 5x_3)\, y_1 + (2x_1 - x_3)\, y_2 = \\ & = & 6\, y_1 + 12\, y_2 = \\ & = & (3y_1 + 2y_2)x_1 - y_1 x_2 + (5y_1 - y_2)x_3. \end{array}$$

Palyginkime pirminio TP uždavinio tikslo funkciją su $\Theta(Y)$. Jei skliausteliuose esantys daugikliai mažesni už tikslo funkcijos koeficientus, tai $\Theta(Y)$ reikšmė mažesnė už pirminio uždavinio tikslo funkcijos reikšmę bet kuriam leistinam pirminio uždavinio sprendiniui. Todėl galioja nelygybių sistema

$$\begin{aligned}
\max\Theta(Y) &\leq \min\,(x_1+2x_2-7x_3),\\
3y_1+2y_2 &\leq 1,\\
-y_1 &\leq 2,\\
5y_1-y_2 &\leq -7,
\end{aligned}$$

reiškianti, kad TP uždavinio

$$\max\,(3y_1+2y_2), \tag{4.15}$$

$$\begin{aligned}
3y_1+2y_2 &\leq 1,\\
-y_1 &\leq 2,\\
5y_1-y_2 &\leq -7,
\end{aligned}$$

optimali reikšmė yra pirminio uždavinio optimalios reikšmės apatinis rėžis. TP uždavinys (4.15) vadinamas **dualiu** pirminiam uždaviniui. TP dualaus uždavinio kintamųjų vektorius paprastai žymimas Y, tuo pabrėžiant simetriškumą tarp pirminio ir dualaus uždavinio. Galima parodyti, kad jei bent vienas iš uždavinių turi optimalų sprendinį, tai ir kitas turi optimalų sprendinį, o jų optimalios reikšmės sutampa.

Panašiai kaip išnagrinėtame pavyzdyje pirminiam TP uždaviniui, užrašytam bendra forma,

$$\min\, CX,$$

$$\begin{aligned}
A'X &= B',\\
A''X &\geq B'', \\
x_1 &\geq 0,...,x_r\geq 0,
\end{aligned} \tag{4.16}$$

formuluojamas dualus TP uždavinys

$$\max BY,$$

$$\begin{array}{rcl} (A^t)'Y & \leq & C', \\ (A^t)''Y & = & C'', \\ y_{l+1} & \geq & 0, ..., y_m \geq 0, \end{array} \tag{4.17}$$

čia $C = (C', C'')$, $B = (B', B'')$, matricą $(A^t)'$ sudaro r transponuotos matricos $A = \begin{pmatrix} A' \\ A'' \end{pmatrix}$ eilučių, o $(A^t)''$ sudaro likusios $n - r$ transponuotos matricos A eilutės. Panašiai vektorius C padalintas į C' ir C'', o matrica A ir vektorius B padalinti į l ir $m - l$ eilučių ir elementų. Dualaus uždavinio tikslo funkcijos koeficientai yra (4.16) ribojimų dešiniosios pusės vektoriaus elementai. Dualaus uždavinio ribojimų dešiniosios pusės vektorius yra (4.16) tikslo funkcijos koeficientų vektorius. Iki šiol transponavimo operacijos nežymėdavome, kadangi ji lengvai būdavo suprantama iš konteksto. Tačiau (4.17) ribojimų matrica užrašyta naudojant transponavimo ženklą, kad būtų pabrėžtas šitoks santykis tarp pirminio ir dualaus uždavinio: *dualiojo uždavinio ribojimų matrica gaunama transponuojant pirminio uždavinio ribojimų matricą.* Neneigiamumo ribojimus kintamiesiems x_i, $i = 1, ..., r$, uždavinyje (4.16) atitinka r nelygybinių $\leq$ tipo ribojinų uždavinyje (4.17). Lygybinius ribojimus uždavinyje (4.16) atitinka neribojami kintamieji y_i, $i = 1, ..., l$, o likusius nelygybinius $\geq$ ribojimus atitinka neneigiamumo ribojimai kintamiesiems y_i, $i = l + 1, ..., m$.

Dualumo tyrimai sudaro svarbią tiesinio programavimo teorijos dalį. Dualių uždavinių analize pagrįsti ir kai kurie TP uždavinių sprendimo algoritmai. TP dualumas detaliau išdėstytas prof. J.Clausen paskaitoje, kuri pateikta tinklalapyje [6]. Čia pagrindinius

TP dualumo teorijos teiginius išvesime iš bendrų dualumo rezultatų, pateiktų trečiajame šios knygos skyriuje.

Nagrinėsime TP uždavinį, užrašytą standartine forma

$$\min CX,$$

$$AX = B,\ X \geq 0.$$

Pažymėję Lagrange'o daugiklių vektorių Y, Lagrange'o funkciją galėsime užrašyti šitaip

$$L(X,Y) = CX + Y(B - AX),\ X \geq 0,$$

o dualaus uždavinio tikslo funkciją – šitaip

$$\begin{aligned} \Theta(Y) &= \inf_{X\geq 0} CX + Y(B - AX) = \\ &= \inf_{X\geq 0} \left(\sum_{j=1}^{n} \left[c_j - \sum_{i=1}^{m} y_i a_{ij} \right] x_j + \sum_{i=1}^{m} y_i b_i \right). \end{aligned} \tag{4.18}$$

Jei bent vienam j galiotų nelygybė

$$c_j - \sum_{i=1}^{m} y_i a_{ij} < 0,$$

tai, didėjant x_j Lagrange$'$o funkcijos reikšmė neaprėžtai mažėtų, t.y. $\Theta(Y) = -\infty$. Jei nelygybės

$$c_j - \sum_{i=1}^{m} y_i a_{ij} \geq 0 \tag{4.19}$$

patenkintos visiems j, tai (4.18) minimumas pasiekiamas, kai $X = 0$, ir

$$\Theta(Y) = \sum_{i=1}^{m} y_i b_i = BY. \tag{4.20}$$

Užrašius ribojimus (4.19) matricine forma, (4.20) maksimizavimo uždavinys atitiks (4.17) formą:

$$\max\, BY, \tag{4.21}$$

$$A^t Y \leq C.$$

Remdamiesi bendru dualaus uždavinio apibrėžimu, suformuluokime dualų uždavinį TP uždaviniui (4.21), kuris ekvivalentus $\min\,(-BY)$, $A^tY \;\geq\; C$. Pastarojo uždavinio Lagrange'o funkcijos daugiklius pažymėję $x_j \geq 0$, $j = 1, ..., n$, dualaus uždavinio tikslo funkciją galėsime išreikšti šitaip

$$\begin{aligned} R(X) &= \inf_{Y \in R^m} -BY + X(A^tY - C) = \\ &= \inf_{Y \in R^m} \left(\sum_{i=1}^{m} \left[\sum_{j=1}^{n} a_{ij} x_j - b_i \right] y_i - \sum_{j=1}^{n} c_j x_j \right), \end{aligned}$$

ir padaryti išvadą: $R(X) = -CX$, jei $AX = B$, $R(X) = -\infty$, jei $AX \neq B$. Todėl (4.21) dualus uždavinys yra šitoks

$$\max\,(-CX) = \min\, CX,$$

$$AX = B,\ X \geq 0,$$

t.y. dualaus uždavinio dualus uždavinys sutampa su pirminiu uždaviniu. Pastarasis teiginys TP uždaviniui užrašytam bendra forma (4.16) gali būti suformuluotas taip: atsižvelgiant į operacijų max ir min, nelygybių $\geq$ ir $\leq$, bei nelygybinių ir kintamųjų neneigiamumo ribojimų simetriškumą, galima teigti, kad pirminis uždavinys yra dualaus uždavinio dualus uždavinys.

Performuluojant bendros dualumo teorijos išvadas tiesiniam programavimui nesunku įrodyti tokius teiginius:

1. Jei X ir Y leistinieji pirminio ir dualaus uždavinio sprendiniai, tai $CX \geq BY$.

2. Jei vienas iš (4.16), (4.17) uždavinių turi optimalų sprendinį, tai ir kitas turi optimalų sprendinį, ir abiejų uždavinių optimalios tikslo funkcijos reikšmės sutampa.

3. Jei vienas iš (4.16), (4.17) uždavinių neturi leistinų sprendinių, tai kitas neturi leistinų sprendinių arba yra neaprėžtas.

4. X^* ir Y^* yra pirminio ir dualaus uždavinių sprendiniais tada ir tik tada, kai

$$\begin{aligned} (A \cdot \ X^* - B) \cdot Y^* &= 0, \\ (A^t \cdot \ Y^* - C) \cdot X^* &= 0. \end{aligned} \tag{4.22}$$

Svarbi dualaus uždavinio sprendinio ekonominė interpretacija. Sakysim, firma gamina n produktų, kurių kiekius $x_1, ..., x_n$, norima nustatyti taip, kad pardavus pagamintus produktus gautas pelnas būtų maksimalus. Jei $i-$tojo produkto kainą pažymėsime c_i, tai reikia spręsti maksimizavimo uždavinį max CX. Kiekvieno produkto vieneto gamybai reikalingų m resursų poreikiai apibrėžiami matrica A, kurios elementas a_{ij} reiškia $i-$tojo produkto vienetui pagaminti reikalingo $j-$tojo resurso kiekį. Gamybos apimtį riboja turimų resursų kiekis, nusakomas vektoriumi B. Pelnas maksimizuojamas esant ribojimams $AX \leq B, X \geq 0$. Tokį tiesinio programavimo uždavinį galėtume spręsti, pavyzdžiui, simplekso algoritmu, pakeitę uždavinį ekvivalenčiu standartiniu TP uždaviniu.

Sakysim, optimaliai suplanavus ir pradėjus gamybą atsiranda investicinių lėšų, kurias būtų galima investuoti į pradėtą gamybą.

Kokių resursų verta pirkti? Kokią kainą už juos verta mokėti? Jeigu bus nupirkta papildomai resursų, kurių kiekių vektorių pažymėsime T, tai uždavinio ribojimai bus užrašomi nelygybe $AX \leq B + T$. Norint sužinoti maksimalų pelną, kurį galima gauti su papildomais resursais, tektų spręsti TP uždavinį iš naujo. Jei T variantų ne vienas, tai norint juos palyginti tektų spręsti tiek naujų TP uždavinių, kiek yra variantų. Tiesa, jei T pakankamai mažas lyginant su B, tai modifikuoto uždavinio optimalią bazę sudaro tie patys kintamieji. Pasinaudojus tokia informacija modifikuotų uždavinių maksimalius pelnus galima greitai apskaičiuioti. Bet yra ir kitas, geresnis variantų palyginimo būdas. Išsprendus dualų uždavinį, jo optimalų sprendinį galima panaudoti modifikuotų uždavinių maksimaliam pelnui įvertinti. Kadangi esant pakankamai mažam T dualiame uždavinyje keičiasi tik tikslo funkcijos koeficientai, o optimali bazė išlieka, tai naujo dualaus uždavinio maksimalus pelnas lygus $(B+T)Y^*$, čia Y^* - dualaus uždavinio su resursų vektoriumi B sprendinys. Iš dualumo teorijos žinoma optimalių reikšmių lygybė: $CX^* = BY^*$, čia X^*- pradinio uždavinio optimalus sprendinys. Nupirkus papildomai T resursų, pajamų prieaugis būtų TY^*. Todėl Y^* interpretuotinas kaip resursų pelningumų (marginal value) vektorius, nusakantis maksimalias mokėtinas kainas.

Jei, sakysim, $y_i^* = 3$, tai papildomai nupirktas $i-$tojo resurso kiekis t_i už kainą γ_i duos papildomą pelną $t_i(\ y_i^* - \ \gamma_i)$. Šiuo požiūriu neapsimoka mokėti už $i-$tąjį resursą brangiau negu $y_i^* = 3$.

Jei $y_i^* = 0$, tai turimas $i-$tojo resurso kiekis b_i pirminio uždavinio optimaliame sprendinyje yra perteklinis, t.y. $\sum_{j=1}^{n} a_{ij} x_j^* < b_i$. Todėl papildomai pirkti $i-$tojo resurso apskritai neapsimoka. Pabrėšime, kad taip nustatytos maksimaliai mokėtinos kainos turi prasmę tik to uždavinio požiūriu.

4.8 Sudėtingumas

Tiesinio programavimo uždavinio, užrašyto standartine forma, sprendinį galima rasti simplekso algoritmu. Sakysim, leistinas bazinis sprendinys žinomas, kadangi jį galima rasti išsprendus kitą panašų uždavinį. Kadangi simplekso algoritmas kaip potencialius sprendinius tikrina poliedro viršūnes, tai tikrinimų skaičius yra baigtinis, blogiausiu atveju sutampantis su poliedro viršūnių skaičiumi. Jei TPS apibrėžtas $n-$matėje erdvėje su m ribojimų, tai leistinų bazinių sprendinių (viršūnių) skaičius ne didesnis negu C_n^m. Sakysim, $n = 2m$. Tada, naudojantis Stirling'o formule viršūnių skaičių galima aproksimuoti $C_n^m \backsim 2^{2m}(m/e)^m$, kas yra astronominiškai didelis skaičius, kai m siekia vos kelias dešimts. Net ir labai paprastos struktūros leistinosios srities, pvz. hiperkubo, viršūnių skaičius yra labai didelis; kai $n = 100$, tai viršūnių skaičius lygus $2^{100} \sim 10^{30}$.

Egzistuoja tiesinio programavimo uždavinių pavyzdžiai, kai simplekso algoritmas, eidamas nuo pradinio iki optimalaus sprendinio, patikrina visas poliedro viršūnes. Taigi, augant uždavinio dydžiui, blogiausiu atveju uždavinio sprendimo simlekso algoritmu laikas auga eksponentiškai. Jei praktiniai uždaviniai būtų panašūs į blogiausio atvejo uždavinius, simplekso algoritmas būtų visai nenaudingas, nes per priimtiną laiką juo nebūtų įmanoma išspręsti bent kiek didesnio uždavinio. Laimei, praktiniai uždaviniai palankesni simplekso algoritmui: daugeliui praktinių uždavinių sprendimo laikas auga ne greičiau negu $2m$. Vis dėlto, kyla svarbus teorinis klausimas, ar egzistuoja algoritmai, kurie blogiausio atvejo požiūriu paprastesni negu simplekso algoritmas.

Simplekso algoritmo sudėtingumo įvertis gaunamas įvertinant poliedro viršūnių skaičių ir įrodant, kad egzistuoja uždaviniai, kuriuos sprendžiant patikrinamos visos poliedro viršūnės. Gautasis sudėtingumo įvertis išreiškiamas šitaip $\tau(m) = \Omega(2^{2m})$; čia $\tau(\cdot)$ yra uždavinio sprendimo laikas, o $\Omega(\cdot)$ reiškia funkciją, augančią ne lėčiau negu jos argumentas. Ne visiems algoritmams šitokia paprasta idėja tinka. Bendru atveju reikia apibrėžti uždavinio dydžio matą ν ir įvertinti uždavinio sprendimo laiko τ priklausomybę nuo uždavinio dydžio. Sprendimo laikas paprastai suprantamas kaip reikalingų uždaviniui išspręsti kompiuteriu vykdomų operacijų skaičius. Realiai kompiuteriu atliekami veiksmai su dvejetainiais kodais. Kiekvienas dvejetainis kodas gali būti interpretuojamas kaip dvejetainis skaičius. Algoritmų sudėtingumo tyrimuose dažniausiai abstrahuojamasi nuo kitų skaičiavimo kompiuteriu ypatybių ir nagrinėjami uždaviniai, kurių duomenys yra sveikieji skaičiai. TPS atveju tas reikštų prielaidą, kad visi matricos A bei vektorių B, C elementai yra sveikieji skaičiai; TPS uždavinį su sveikaisiais koeficientais sutrumpintai žymėsime TPSS. Prielaida apie koeficientų sveikaskaitiškumą praktinių taikymų požiūriu nėra varžanti, nes tinkamai parinkus mastelį uždavinio duomenis norimu tikslumu galima išreikšti sveikaisias skaičiais. Natūralus uždavinio dydžio matas yra bitų skaičius reikalingas uždavinio duomenims užrašyti dvejetaine sistema. TPSS duomenis sudaro $m \times n$ dydžio matrica A ir žymiai mažiau elementų turintys vektoriai B ir C. Todėl TPSS dydis gali būti įvertintas šitaip

$$\nu \geq m \times n + \lceil \log_2 |P| \rceil,$$

čia pirmasis dėmuo įvertina skirtukų skaičių uždavinio koeficientams vienam nuo kito atskirti nestruktūrizuotoje atmintyje, P yra visų nelygių nuliui sveikųjų skaičių, sudarančių TPS duomenis, sandauga; $\lceil \cdot \rceil$ reiškia apvalinimą iki didesnio sveikojo skaičiaus.

Toliau pateiksime polinominio sudėtingumo TPSS algoritmų pavyzdžius. Polinominis sudėtingumas reiškia, kad skaičiavimo laikas gali būti aprėžtas polinomine uždavinio dydžio funkcija

$$\tau(\nu) = O(\nu^N),$$

čia $O(\cdot)$ reiškia funkciją, kuri auga ne greičiau negu jos argumentas, N priklauso nuo uždavinio duomenų. Kadangi eksponentinė funkcija auga greičiau už polinominę, tai pakankamai dideliems uždaviniams polinominio sudėtingumo algoritmai bus greitesnis negu simplekso algoritmas. Priminsime, kad algoritmai lyginami blogiausio atvejo požiūriu.

TPSS algoritmų sudėtingumui analizuoti patogesnis uždavinio dydžio matas

$$L = m \times n + \lceil \log_2 |P| \rceil + n \, \lceil \log_2 |n| \rceil.$$

Kadangi L patenkina nelygybę $L \leq \nu^2$, tai iš algoritmo polinominio sudėtingumo L atžvilgiu išplaukia jo polinominis sudėtingumas ν atžvilgiu: $L^M \leq \nu^{2M}$.

Simplekso algoritmu optimalaus sprendinio ieškoma tikrinant leistinosios srities viršūnes. Toliau nagrinėjami metodai generuoja taškų, nesutampančių su srites viršūnėmis, sekas. Tokios sekos asimptotiškai konverguoja į optimalią viršūnę. Begalinės sekos nutraukimo sąlygai apibrėžti ir tiksliam TPSS uždavinio (kurio dydis lygus L) sprendiniui apskaičiuoti pagal gautą apytikslų sprendinį svarbios toliau suformuluotos bazinių sprendinių savybės.

Lema. Jei X yra bazinis sprendinys, tai egzistuoja sveikieji skaičiai $D_0, D_1, ..., D_n$, tokie, kad $x_i = D_i/D_0$, $D_0 \neq 0$, $D_i \leq 2^L$, $|x_i| < 2^L$.

Įrodymas. Sakysim H yra $m \times m$ matrica, sudaryta iš matricos A bazinių stulpelių, atitinkančių bazinį sprendinį X. Pažymėkime matricos H determinantą D_0. Kadangi D_0 yra lygus $m!$ dėmenų sumai, kurių kiekvienas lygus m matricos H elementų sandaugai, tai

$$|D_0| \leq m! \cdot |P| \leq n! \cdot |P| \leq 2^{n \lceil \log_2 n \rceil + \log_2 P} < 2^L.$$

Kiekvienas bazinis X elementas pagal Kramerio taisyklę lygus D_i/D_0, D_i determinantas matricos, gautos pakeitus matricoje H

vieną stulpelį stulpeliu B. Panašiai, kaip D_0 atveju, parodoma, kad $|D_i| < 2^L$. Nebaziniai X elementai lygūs nuliui. Todėl visus X elementus galima užrašyti taip, kaip teigiama Lemos teiginyje, ir $|x_i| < 2^L$.

Lema. Sakysim X ir Y du baziniai sprendiniai ir patenkinta nelygybė

$$CX - CY \leq 2^{-2L}. \tag{4.23}$$

Tada $CX = CY$.

Įrodymas. Sakysim teiginys neteisingas, t.y. $CX \neq CY$. Kadangi iš anksčiau įrodytos lemos išplaukia, kad $CX = D'/D$, $CY = E'/E$, ir visi dešiniosiose lygybių pusėse esantys sveikieji skaičiai absoliutiniu dydžiu neviršija 2^L, tai teisinga nelygybė

$$|CX - CY| = \left|\frac{D'E - E'D}{DE}\right| \geq \frac{1}{|DE|} > 2^{-2L},$$

kuri prieštarauja (4.23). Gautas prieštaravimas įrodo Lemos teiginį.

4.9 Elipsoidų metodas

Elipsoidai naudojami konstuojant optimizavimo metodus įvairioms uždavinių klasėms. Čia trumpai aptariamas elipsoidų metodas tiesinio programavimo uždaviniams. Šį metodą 1979 m. paskelbė L.Chačijanas, remdamasis visai kitomis negu simplekso algoritmo idėjomis. Šio metodo žingsnių skaičius, augant uždavinio dydžiui, blogiausiu atveju auga polinomo greičiu, t.y. metodas priklauso polinominio sudėtingumo klasei. Šiuo požiūriu elipsoidų metodas žymiai pranašesnis už eksponentinio sudėtingumo klasei priklausantį simplekso algoritmą.

Nagrinėsime tiesinio programavimo uždavinį standartine forma

$$\min CX,$$

$$AX = B, \; X \geq 0, \tag{4.24}$$

darydami prielaidą, kad A, B ir C elementai yra sveikieji skaičiai; tokį uždavinį sutrumpintai žymėsime TPSS.

14 teorema. TPSS polinominio sudėtingumo algoritmas egzistuoja tada ir tik tada, kai egzistuoja polinominio sudėtingumo algoritmas tiesinių nelygybių su sveikaisiais koeficientais sistemoms spręsti.

Įrodymas. Sakysim, žinomas polinominio sudėtingumo TPSS algoritmas. Duotajai nelygybių sistemai galima užrašyti TP uždavinį, kurio ribojimais būtų duotosios nelygybės, o tikslo funkcijos koeficientų vektorius būtų nulinis $C = (0, ..., 0)$. Gautą uždavinį pakeitę jam ekvivalenčiu TPSS uždaviniu ir jį išsprendę turimu algoritmu arba gausime duotosios tiesinių nelygybių sistemos leistnąjį sprendinį, arba įrodysime, kad toks sprendinys neegzistuoja. Kadangi minimi pakeitimai atliekami per tiesinį laiką, tai bendras uždavinio sprendimo laikas polinominis, t.y. sprendimo laikas gali būti aprėžtas polinomu, kurio argumentas yra nelygybių sistemos dydis.

Sakysim, žinomas polinominio sudėtingumo algoritmas tiesinių nelygybių su sveikaisiais koeficientais sistemoms spręsti. Šiuo algoritmu galima surasti kokį nors nelygybių sistemos (kiekvieną lygtį galima pakeisti dviem nelygybėmis), kuria apibrėžti TPSS ribojimai sprendinį, arba įsitikinti, kad ribojimai nesuderinti, t.y. sistema neturi sprendinio. Pirmuoju atveju leistinų sprendinių aibė netuščia, ir TPSS sprendimą, remiantis dualumo teorijos rezultatais, galima pakeisti šitokios lygčių ir nelygybių sistemos su sveikasiais koeficientais

$$\begin{aligned} AX &= B, \; X \geq 0, \\ A^t Y &\leq C, \, CX - BY = 0, \end{aligned} \tag{4.25}$$

sprendimu. Iš dualumo teorijos žinoma, kad tiesinių nelygybių (4.25) sistema turi sprendinį, jei TPSS aprėžtas. Naudodamiesi algoritmu tiesinių nelygybių sistemoms spręsti per polinominį laiką apskaičiuosime (4.25) sprendinį (X_*, Y_*), arba įsitikinsime, kad (4.25) neturi sprendinio. Kadangi (4.25) sistemos dydis nedidesnis už trigubą TPSS (4.24) dydį, tai TPSS sprendinys X_* bus surastas per polinominį laiką; arba bus nustatyta, kad TPSS sprendiniai neaprėžti.

Įrodėme, kad polinominio sudėtingumo algoritmas TPSS uždaviniui (4.24) egzistuoja tada ir tik tada, kai toks algoritmas egzistuoja tiesinių nelygybių sistemai (4.25) spręsti; priminsime, kad lygybė yra ekvivalenti dviem nelygybėms. Pastarosios sistemos nelygybės yra negriežtos. Kaip pamatysime vėliau, lengviau sudaryti polinominio sudėtingumo algoritmą griežtų nelygybių sistemoms. Todėl svarbu, kad egzistuotų polinominio sudėtingumo algoritmas negriežtos nelygybių sistemos sprendiniui atstatyti pagal artimos pastarajai griežtų nelygybių sistemos sprendinį.

15 teorema. Nelygybių su sveikaisiais koeficientais sistema $AX \leq B$, kurios dydis L, turi sprendinį tada ir tik tada, kai $AX < B'$ tuti sprendinį, čia $b'_i = b_i + \varepsilon$, $\varepsilon = 2^{-2L}$. Egzistuoja polinominio sudėtingumo L atžvilgiu algoritmas sistemos $AX \leq B$ sprendiniui apskaičiuoti remiantis sistemos $AX < B'$ sprendiniu.

Šios teoremos neįrodinėsime. Ja remiantis galima teigti, kad egzistuoja polinominio sudėtingumo TPSS algoritmas, jei polinominio sudėtingumo algoritmas egzistuoja griežtų tiesinių nelygybių sistemoms su sveikaisiais koeficientais spręsti. Iš tikrųjų, remiantis dualumo teorija, TPSS per tiesinį laiką gali būti pakeistas ekvivalenčiu negriežtų tiesinių nelygybių su sveikaisiais koeficientais sistemos $\tilde{A}Z \leq \widetilde{B}$ sprendimo uždaviniu, kurio dydis $£$ ne daugiau kaip tris kartus didesnis už TPSS dydį. Sudarykime nelygybių sistemą $\tilde{A}'Z < \widetilde{B}'$, $\tilde{A}' = \tilde{A} \cdot 2^{2L}$, $\widetilde{B}' = \widetilde{B} \cdot 2^{2L} + I$, I - vienetinis vektorius. Pastarosios sistemos dydis ne didesnis negu $2£^2$. Jei pastarajai sistemai spręsti žinomas polinominio sudėtingumo algoritmas, tai jos sprendimo laikas aprėžtas ir polinomu L atžvilgiu. Suradę nelygybių sistemos $\tilde{A}'Z < \widetilde{B}'$ sprendinį, kaip teigia 15 teorema, sistemos $\tilde{A}Z \leq \widetilde{B}$ sprendinį, galima atstatyti (tuo pačiu surasti ir TPSS sprendinį) per polinominį laiką. Kadangi polinominų įverčių kompozicija duoda polinominį įvertį, tai ir bendras TPSS uždavinio sprendimo algoritmas yra polinominio sudėtingumo.

Elipsoidų algoritmas skirtas griežtų tiesinių nelygybių sistemoms spręsti. Pažymėkime nagrinėjamąją nelygybių sistemą $AX < B$. Inicializavimo žingsnyje sukuriamas pradinis elipsoidas Ξ_0, kuris apima nelygybių sistemos sprendinių aibę T arba (jei, pavyzdžiui, aibė neaprėžta) jos dalį.

$k-$tajame algoritmo žingsnyje tikrinama, ar hiperelipsoido Ξ_{k-1} centras ϑ_k patenkina nelygybių sistemą. Jei taip, tai uždavinys išspręstas. Jei ne, nustatoma nelygybė,

$$\sum_i a_{ji}x_i < b_j,$$

kuri yra nepatenkinama. Per elipsoido centrą pravesta hiperplokštuma lygiagreti hiperplokštumai

$$\sum_i a_{ji}x_i = b_j$$

dalija hiperelipsoidą pusiau, ir T priklauso hiperelipsoido pusei, kuri patenkina nelygybę

$$\sum_i a_{ji}x_i < \sum_i a_{ji}\vartheta_{ki}.$$

Akcentuosime svarbią hiperelipsoidų savybę: nesunku sukonstruoti hiperelipsoidą, kuris apima padalinto hiperelipsoido pusę ir kurio hipertūris reikšmingai mažesnis už padalinto hiperelipsoido tūrį. Remiantis šia savybe sudaromas hiperelipsoidas Ξ_k. $k := k + 1$, ir žingsnis kartojamas.

Galima parodyti, kad po baigtinio žingsnių skaičiaus bus gautas nelygybių sistemos sprendinys. Algoritmas be papildomos sustojimo

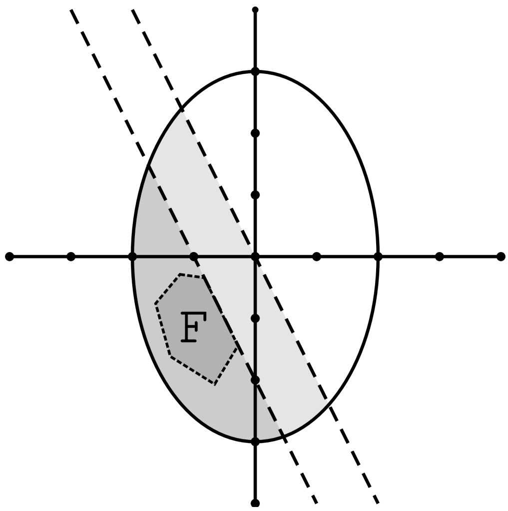

Figure 4.13: Elipsoido centras nepatenkina nelygybių sistemos. Nubrėžta tiesė, atitinkanti nepatenkintą nelygybę, ir jai lygiagreti tiesė, einanti per elipsoido centrą

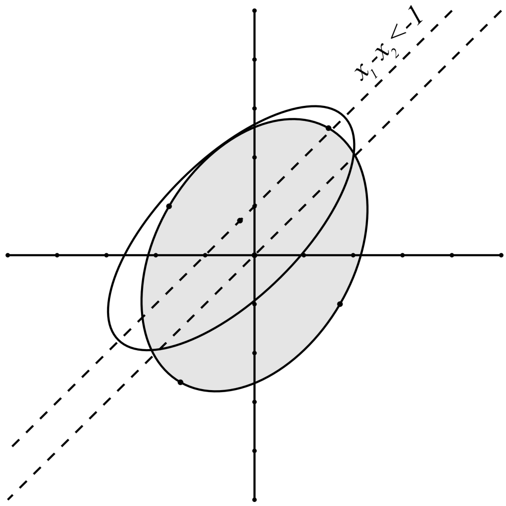

Figure 4.14: Naujas elipsoidas apima atkirstąją pusę

sąlygos būtų nekorektiškas, nes tuščios aibės T atveju būtų vykdomas begalinis ciklas. Tačiau, remiantis informacija apie uždavinio duomenis, teoriškai galima įvertinti minimalų hiperelipsoido tūrį netuščios leistinosios srities atveju. Jei vykdant algoritmą gaunamas mažesnio tūrio hiperelipsoidas, tai algoritmas stabdomas, ir šitoks sustojimas reiškia, kad nelygybių sistema neturi sprendinio.

Sakysim, pradinis hiperelipsoidas yra sfera su centru koordinačių pradžioje ir radiusu $n \cdot 2^L$. Tada ne vėliau kaip po $K = 16 \cdot n \cdot (n + 1) \cdot L$ žingsnių algoritmas sustos. Tas reiškia, kad blogiausio atvejo sprendimo laikas yra polinominis uždavinio dydžio atžvilgiu. Toliau pateikiama elipsoidų algoritmo schema. k- žingsnyje hiperelipsoidą aprašanti matrica pažysmėta H_k :

$$\Xi_k = \{X : (X - \vartheta_k)^t \cdot H_k^{-1} \cdot (X - \vartheta_k) \leq 1.$$

Elipsoidų algoritmas:
1. $k = 0,\ K = 16 \cdot n \cdot (n+1) \cdot L,\ \vartheta_0 = (0, ..., 0)^t,$

$$H_0 = n^2\ 2^{2L} \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{pmatrix}.$$

2. **For** $k = 0$ **to** K **do**
If $A\vartheta_k < B$ **then** Print (ϑ_k) **and Stop**
else
Rasti stulpelį A_i tokį, kad $A_i^t \vartheta_k \geq b_i$,
$D = A_i$,
$\vartheta_{k+1} = \vartheta_k - \frac{1}{n+1} \cdot \frac{H_k \cdot D}{\sqrt{D^t \cdot H_k \cdot D}}$,
$H_{k+1} = \frac{n^2}{n^2-1} \left(H_k - \frac{2}{n+1} \frac{(H_k \cdot D)(H_k \cdot D)^t}{D^t \cdot H_k \cdot D} \right).$
3. Print ('neturi sprendinio") **and Stop**.

Nepaisant to, kad elipsoidų algoritmas (blogiausio atvejo) sudėtingumo kriterijaus požiūriu yra žymiai geresnis už simplekso algoritmą, jis pasirodė kaip prastas pastarojo konkurentas sprendžiant praktinius uždavinius. Viena iš pagrindinių priežasčių ta, kad simplekso algoritmo žingsnis realizuojamas žymiai paprastesniais skaičiavimais negu elipsoidų algoritmo žingsnis. Be to, praktiniuose uždaviniuose simplekso algoritmo vykdomų žingsnių skaičius daug - daug mažesnis negu blogiausiu atveju, o elipsoidų algoritmui tokia savybė (kurią sunku net tiksliai suformuluoti) nėra būdinga.

4.10 Vidinio taško metodai

Simplekso algoritmas dominavo tiesinio programavimo uždavinių sprendimo srityje nuo jo išradimo iki elipsoidų metodo paskelbimo. Jis buvo detaliai teoriškai ištirtas, ištobulinta jį realizuojanti programinė įranga. Simplekso algoritmu buvo sėkmingai sprendžiami praktiniai uždaviniai. Atskiros kitokių idėjų realizacijos pritraukdavo nepalyginamai mažiau intelektualiųjų jėgų ir negalėjo konkuruoti su dominuojančiąja kryptimi. Situacija pasikeitė po elipsoidų metodų paskelbimo. Nors elipsoidų metodas ir realiai nekonkuravo su simplekso algoritmu praktinių TP uždavinių sprendimo srityje, tačiau jo sukūrimas parodė, kad skirtingos negu simplekso algoritmo idėjos gali būti perspektyvios, kad kelias efektyvių TP metodų kūrimui nėra vienintelis. Praėjus keliems metams po elipsoidų metodo paskelbimo, N.Karmarkar pasiūlė kitokią idėją, pradėjusią naują, taip vadinamų vidinio taško metodų, kryptį. Buvo įrodyta, kad šis, kaip ir elipsoidų, metodas priklauso polinominio sudėtingumo metodų klasei. Be to, Karmarkar'o metodo sudėtingumą aprašančio polinomo laipsnis žemesnis negu elipsoidų metodo sudėtingumą aprašančio polinomo laipsnis. Ypač svarbu, kad šis metodas sėkmingai konkuravo su simplekso algoritmu sprendžiant praktinius uždavinius.

Karmarkar'o algoritmas tiesiogiai tinka specialiam TPS atvejui, kai visų, išskyrus vieną, ribojimų dešiniosios pusės lygios nuliui. Išskirtinis ribojimas yra toks

$$x_1 + x_2 + ... + x_n = 1.$$

Be to, reikalaujama, kad leistinajai sričiai priklausytų taškas $V_0 = \frac{I}{n}$, čia I - vienetinis vektorius.

Apibrėžtis. Politopas

$$K = \{X \in R^n : X \geq 0, \sum x_i = 1\},$$

vadinamas standartiniu $n-1$ matavimo simpleksu, o taškas $V = \frac{I}{n}$ jo centru.

Apibrėžtis. TP uždavinys vadinamas Karmarkar'o standartiniu uždaviniu (TPKS), jei, leistinoji sritis yra apibrėžta šitaip

$$G = K \cap P, \;\; P = \{X : \; AX = 0\},$$

čia A yra sveikųjų skaičių matrica, $V \in G$, tikslo funkcijos koeficientai C yra sveikieji skaičiai ir tikslo funkcijos reikšmės CX neneigiamos, kai $X \in G$. Reikia rasti srities G tašką, kuriame tikslo funkcijos reikšmė lygi nuliui arba nustatyti, kad toks taškas neegzistuoja.

Pabrėšime: TPKS formuluojamas taip, kad nereikalaujama rasti optimalaus sprendinio, jei minimali tikslo funkcijos reikšmė nelygi nuliui. TPKS yra specialus TPS atvejis. Galima įrodyti, kad TPS per polinimininį laiką gali būti pakeistas TPKS. Todėl nesiaurindami bendrumo spręsime TPKS, kadangi jo struktūra labiau atitinka nagrinėjamąjį algoritmą. Paminėsime keletą svarbių Karmarkar'o algoritmo paaiškinimui faktų.

Sakysim $X_0 \in G$. Įbrėžkime į G sferą $S(X_0, r)$ su centru X_0 ir radiusu r. Nesunku rasti minimizavimo uždavinio $\min_{X \in S(X_0,r)} CX$ minimumo tašką (8 pratimas). Tačiau surastasis minimumo taškas gali labai skirtis nuo uždavinio $\min_{X \in G} CX$ minimumo taško. Tas priklauso nuo srities apvalainumo taško X_0 atžvilgiu: aibę G galėtume vadinti apvalaina taško X_0 atžvilgiu, jei įbrėžtos į G ir apibrėžtos apie G sferų radiusai mažai skirtųsi. Jei aibė apvalaina taško X_0 atžvilgiu, tai abiejų minėtų uždavinių minimumo taškai skiriasi nelabai daug, ir priešingai: jei G nėra apvalaina, tai jie gali labai skirtis. Pavyzdžiui, trikampis kurio kraštinių ilgiai 1000, 1000 ir 1 tikrai nelaikytinas apvalainu. Standartinis ($n-1$ matavimo) simpleksas laikytinas apvailainu, nes apie jį apibrėžtos ir į jį įbrėžtos sferų radiusų santykis nėra labai didelis (9 pratimas)

$$r = \frac{1}{\sqrt{n(n-1)}}, \;\; R = \sqrt{\frac{n-1}{n}}, \;\; \frac{R}{r} = n - 1.$$

Projektyvinė transformacija $R^n \longmapsto R^n$ apibrėžiama šitaip

$$X \longmapsto \frac{HX + D}{FX + c},$$

čia H yra $n \times n$ matrica, D, F - n mačiai vektoriai ir $c-$ konstanta, patenkinanys reikalavimą, kad matrica $\begin{pmatrix} E & D \\ F^t & c \end{pmatrix}$ būtų neišsigimusi; ši transformacija apibrėžta tik tokiems X, kuriems $FX + c \neq 0$. Jei $\alpha_i \neq 0$, tai

$$P_\alpha(X) = \frac{\Gamma X}{I^t \Gamma X},\ \Gamma^{-1} = \begin{pmatrix} \alpha_1 & 0 & ... & 0 \\ 0 & \alpha_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & 0 & \alpha_n \end{pmatrix},$$

apibrėžia specialų projektyvinės transformacijos atvejį $K \longmapsto K$. $P_\alpha(X)$ atvaizduoja standartinį simpleksą standartiniame simplekse, kadangi $I^t \cdot P_\alpha(X) = 1$, ir visos atvaizdo komponentės neneigiamos. Nesunku patikrinti (10 pratimas), kad atvirkštinė transformacija apibrėžiama šitaip

$$P_\alpha^{-1}(X) = \frac{\Gamma^{-1} X}{I^t \Gamma^{-1} X}\ , \tag{4.26}$$

ir $P_\alpha^{-1}(X) \in K$, $X \in K$.

Jei $X' = P_\alpha(X)$, tai $x_j = \frac{\alpha_j x_j'}{\sum \alpha_i x_i'}$, ir lygybė $AX = \sum A_j x_j = 0$ patenkinama tada ir tik tada, kai $\frac{1}{\sum \alpha_i x_i'} \sum_j (A_j \alpha_j) x_j' = 0$, arba $\sum_j (A_j \alpha_j) x_j' = 0$. Kadangi $A\Gamma^{-1}$ yra $m \times n$ matrica, kurios $j-$ tasis stulpelis yra $(A_j \alpha_j)$, tai lygybė $AX = 0$ patenkinama tada ir tik tada, kai $A\Gamma^{-1} X' = 0$.

Teisinga tokia transformacijos $P_\alpha(X)$ savybė, apibendrinanti prieš tai įrodytąsias: jei $\alpha \in G$, tai transformacija $P_\alpha(X)$ atvaizduoja standartinio simplekso ir tiesinio poerdvio sankirtą $G = K \cap \{X :\ AX = 0\}$ į standartinio simplekso ir tiesinio poerdvio sankirtą $G' = K \cap \{X' :$

$A\Gamma^{-1}X' = 0\}$; $P_\alpha(\alpha) = V$. Kadangi leistinosios srities atvaizdui G' galioja tokie santykiai

$$S(V,r) \cap \{A\Gamma^{-1}X' = 0\} \subset G' \subset S(V,R) \cap \{A\Gamma^{-1}X' = 0\},$$

čia $S(V,\cdot)$ pažymėtos įbrėžta ir apibrėžta sferos, tai galima teigti, kad transformacija $P_\alpha(X)$ išlaiko leistinosios srities apvailainumą. Santykinė paklaida, gaunama minimizaciją leistinojoje srityje pakeičiant minimizacija įbrėžtoje sferoje, ne didesnė už

$$\frac{R-r}{R} = 1 - \frac{r}{R} = 1 - \frac{1}{n-1}.$$

Remiantis atvaizdavimo $P_\alpha(X)$ savybėmis galima sudaryti tokį algoritmą: leistinojoje srityje randamas X_1 - tikslo funkcijos minimumo taškas įbrėžtojoje sferoje. Atliekama erdvės transformacija $P_{X_1}(X)$. Naujojoje leistinoje srityje X_1 yra atvaizduotas į srities centrą V. Apskaičiuojamas taškas X_2' atliekant transformuotos tikslo funkcijos minimizavimą sferoje, įbrėžtoje į naująją sritį, kuri yra to paties tipo kaip ir pradinė sritis, ir tt. Atliekant atvirkštinį atvaizdavimą, galima atstatyti einamojo taško X_k' pirmvaizdį pradinėje leistinojoje srityje.

Tačiau ši graži schema turi vieną nepažymėtą trūkumą: transformuota tikslo funkcija nebėra tiesinė:

$$CX = \frac{\sum_j (c_j \alpha_j) x_j'}{\sum_i \alpha_i x_i'},\ X' \in G'.$$

Žinomi metodai negarantuoja netiesinės funkcijos minimimumo taško radimo per baigtinį žingsnių skaičių. Jeigu nepavyktų šio pagalbinio uždavinio pakeisti kitu, sprendžiamu per baigtinį žingsnių skaičių, tai polinominio sudėtingumo algoritmo sudarymas taptų neįmanomas. Tačiau pasirodė, kad užtenka minimizuoti einamosios tikslo funkcijos skaitiklį. Nors atskirai paimtos iteracijos rezultatas gali ir pabloginti tikslo funkcijos reikšmę, bet ilga iteracijų seka užtikrina tikslo funkcijos mažėjimą. Minimizavimo proceso progresui vertinti buvo pasiūlyta naudotis potencijalo funkcija

$$f(X) = \ln\left(\frac{(CX)^n}{x_1 \cdot x_2 \cdot \ldots \cdot x_n}\right).$$

Akivaizdu, kad $CX \to 0$, jei $f(X) \to -\infty$.

Apibrėšime funkciją $\Phi(X),\ X \in G$ šitaip:

1. Atliekama transformacija $P_X(\cdot)$,

2. Randamas taškas X' minimizuojant $C'Y$ sferoje $S(V, \frac{\beta}{n})$, čia $C' = \Gamma^{-1}C$.

3. $\Phi(X) = P_X^{-1}(X')$.

Įrodytas šitoks faktas: jei atvaizdavimu $\Phi(\cdot)$ potencijalas sumažėja mažiau negu δ, tai Karmarkar'o uždavinio minimali tikslo funkcijos reikšmė yra teigiama.

Karmarkar'o algritmas:

Pasirenkama parametro β reikšmė $0 < \beta < \frac{1}{2}$ ir apskaičiuojamas algoritmo parametras $\delta = \beta - \frac{\beta^2}{1-\beta}$.

1. $X_0 = (\frac{1}{n}, ..., \frac{1}{n})^t$, $K = \lceil 2nL/\delta \rceil$.
   If $CX_0 = 0$ **then** Print (X_0) **and Stop.**
2. **for** $k = 1$ **to** K **do**
   $X_k = \Phi(X_{k-1})$
   If $CX_k = 0$ **then** Print (X_k) **and Stop.**
   If $f(X_k) > f(X_{k-1}) - \delta$ **then** Print("min >0")
   and Stop.
3. Rasti leistinosios aibės viršūnę $X : CX \leq CX_k$.
   Print (X) **and Stop.**

Tikslo funkcijos reikšmė apskaičiuotame taške X turi būti lygi 0. Karmarkar'o algoritmo vienos iteracijos sudėtingumas yra $O(n^3)$. Kadangi iteracijų skaičius yra $O(nL)$, tai bendras sudėtingumas $O(n^4L)$ yra polinominis.

Karmarkar'o algoritmo reikšmė optimizacijos mokslo kryptčiai žymiai didesnė negu atskiro, kad ir labai efektyvaus, algoritmo sukūrimas. Buvo įrodyta, kad sudėtingų TP praktinių uždavinių sprendimui gali būti sėkmingai taikomos netiesinių funkcijų optimizavimo idėjos. Prasidėjęs vidinio taško metodų tyrimas ir jų sėkmigas taikymas gerokai išplėtė optimizacijos metodų taikymo sritį. Toliau, nesigilindami į sudėtingumo vertinimą, pateiksime vieno vidinio taško metodo, vadinamo **afininės skalės keitimo** metodu, aprašymą. Jo idėja panaši į gradiento projekcijos metodo idėją. Reikia rasti TPS

$$\min CX,$$

$$AX = B, \ X \geq 0,$$

sprendinį. Sakysim, žinomas priklausantis leistinajai sričiai pradinis taškas X_0. Tikslo funkcijos antigradiento vektoriaus $-C$ projekcija į aibę

$$\{X : AX = B\}$$

yra leistina nusileidimo kryptis. Tikslo funkciją galima mažinti einant šia kryptimi iki leistinosios srities ribos, t.y. iki viena iš vektoriaus X komponenčių taps lygi 0. Toliau iš tokio taško tektų eiti leistinosios srities riba, apibrėžiama aktyviaisiais nelygybiniais ribojimais. Aktyviųjų ribojimų įvertinamas projektuojant nusileidimo kryptį į leistinąją sritį kelia metodo realizavimo sunkumų. Esminė šio metodo idėja ta, kad neinama iki leistinosios srities ribos, bet naujuoju pasirenkamas vidinis srities taškas.

Kadangi tikslo funkcija tiesinė, tai gradientas visur tas pats. Taigi, iš naujojo taško vėl tektų eiti ta pačia kryptimi. Geresnei krypčiai gauti reikia įvertinti artimiausią leistinosios srities ribą. Nuo to priklauso metodo efektyvumas.

Galime prie tikslo funkcijos pridėti pagalbinį narį, kuris įvertintų nelygybinius ribojimus. Paprasčiausia, bet sudėtingesnė už tiesinę, funkcija yra kvadratinė. Suformuluosime pagalbinį minimizavimo uždavinį su kvadratine tikslo funkcija, kurios matrica H_k bus parenkama taip, kad įvertintų einamojo taško artumą leistinosios srities ribai

$$\min \left(C(X - X_k) + \frac{1}{2\upsilon_k}(X - X_k)H_k(X - X_k)\right),$$

$$AX = B, \ X \geq 0.$$

Šio uždavinio minimumo taškas lygus

$$\begin{aligned} \overline{X_k} &= X_k - \upsilon_k(H_k)^{-1}(C - A\Theta_k), \\ \Theta_k &= \left(A\,(H_k)^{-1}A\right)^{-1} A\,(H_k)^{-1}C. \end{aligned} \tag{4.27}$$

Vienas iš galimų matricos H_k parinkimo variantų yra $H_k = \Omega_k^{-2}$, čia Ω_k yra diagonali matrica, sudaryta iš (teigiamų) vektoriaus X_k komponentų. Kai X_k yra vidinis leistinosios srities taškas, o υ_k pakankamai mažas, tai ir $\overline{X_k}$ yra vidinis leistinosios srities taškas. Naująjį tašką rasime minimizuodami tikslo funkciją ant spindulio, einančio iš taško X_k link $\overline{X_k}$

$$X_{k+1} = X_k + \gamma_k(\overline{X_k} - X_k),$$

čia γ_k yra žingsnio daugiklis, nustatytas sprendžiant vienmatės minimizacijos uždavinį. Vėlgi, minimizuojant einama ne iki pat srities ribos, bet sustojama šiek tiek anksčiau.

Priminsime, kad Niutono metodo žingsnis gali būti interpretuojamas kaip žingsnis antigradiento, apskaičiuoto kitoje metrikoje, kryptimi. Taip ir (4.27) metodo žingsnio kryptis $-T_kC$ yra antigradiento kryptis erdvėje, gautoje atlikus R^n afininę transformaciją apibrėžtą matrica T_k. Pastaroji matrica išreiškiama per ribojimų matricą A bei teigiamai apibrėžtą matricą H_k, kurią galima parinkti taip, kad metodas turėtų norimas savybes. Galima įrodyti, kad šiuo metodu su pasirinkta aukščiau matrica H_k TP uždavinio sprendinys gaunamas po baigtinio žingsnių skaičiaus.

Dėl naudojamos metodo apibrėžime afininės transformacijos anglų kalba šitoks metodas vadinamas *affine scaling method.* Jis buvo pasiūlytas dar 1967 metais I.Dikino. Tačiau tuo metu nebuvo įrodytas metodo polinominis sudėtingumas.

Pastaruoju metu vidinio taško metodų tyrimas buvo viena iš aktyviausių krypčių optimizacijoje. Buvo paskelbti ir ištirti metodai, sudarantys didelę klasę, kurią prireikė toliau detalizuoti. Specialiojoje literatūroje dažnai sutinkami terminai "pirminis - dualus", "logaritminis barjeras", "potencijalo mažinimas", "afininis skalės keitimas" beveik vienareikšmiškai parodo, kad straipsnyje nagrinėjami arba taikomi vidinio taško metodai. Reikia pažymėti, kad sėkmingas vidinio taško metodų taikymas davė impulsą ir simplekso tipo algoritmų tobulinimui.

4.11 Istorinės pastabos

Algoritmo tiesinio programavimo uždaviniams spręsti išradėjas ir šios mokslo srities krikštatėvis G.Danzig teigia, kad *"Tiesinį programavimą galima vertinti kaip revoliucinį išradimą, įgalinantį žmoniją praktinėse sudėtingose situacijose nusibrėžti aiškius tikslus ir detaliai pagrįsti geriausius žingsnius tiems tikslams pasiekti".* Pats G.Danzig susidūrė su sudėtingomis praktinėmis situacijomis, kai Antrojo Pasaulinio karo metais dirbo Pentagone planavimo reikalų patarėju. Po karo jam buvo pasiūlyta išanalizuoti planavimo automatizavimo galimybes. Reikėjo trumpais terminais ruošti iki tol neregėtos apimtimis tiekimo ir logistikos planus. To laiko mechanizacija reiškė analoginių ir/arba perforacinių skaičiavimo mašinų panaudojimą. Jų galimybių ribotumą gerai matė planavimo ekspertai. Sužinoję apie kompiuterio kūrimo idėjas, jie prisidėjo prie idėjos realizavimo įtikinėdami Pentagono vadovus finansuoti kompiuterio kūrimą. Taigi, galvodamas apie uždavinį, G.Danzig turėjo metodo realizavimo perspektyvą.

G.Danzig matematikos daktaro laipsnį gavo Berkeley universitete. Jo gabumai pasireiškė anksti. Pasakojama, kad po paskaitos jis greit nusirašė nuo lentos namų darbo sąlygą ir turėjo šiek tiek sunkumų spręsdamas tą uždavinį. Bet kai atnešė profesoriui J.Neyman rezultatus, pasirodė, kad nuo lentos buvo nusirašyta ne sąlyga namų darbams, o tema daktaro disertacijai. Savo moksliniuose tyrimuose G.Danzig įgavo patirties klasikinėje optimizacijos teorijoje. Matyt, gilus Lagrange'o daugiklių metodo suvokimas suvaidino teigiamą vaidmenį nagrinėjant tiesinio programavimo uždavinius. Be to, G.Danzig turėjo galimybę konsultuotis su tokiais mokslo autoritetais kaip J. von Neumann, J.Neyman, H.Kuhn ir kitais. Gilios matematikos žinios, orientacija į praktinį pritaikymą ir kompiuterio perspektyvos suvokimas buvo G.Danzig'o sulydyti į vieną iš įžymiausių algoritmų – *simplekso algoritmą* tiesinio programavimo uždaviniams spręsti. 1947 laikomi šio algoritmo sukūrimo metais. Vos tik sukurtas jis buvo panaudotas optimaliam karinių pervežimų planavimui. Pirmas komercinis panaudojimas datuojamas 1952 metais, kai benzino gamyboje tiesinio programavimo metodais buvo pradėti optimizuoti rektifikuojamieji naftos mišiniai. Pirmoji komercinė kompiuterio programa, realizuojanti simplekso algoritmą, buvo sukurta mokslinių tyrimų kompanijoje Rand Corporation 1954 metais.

Kitas tiesinio programavimo teorijos pirminis šaltinis – L.Kantorovičiaus darbas, skirtas iš pirmo žvilgsnio nuobodžiam uždaviniui – faneros gamybos mašinų optimalaus darbo planui sudaryti. L.Kantorovičiaus matematinis talentas paireiškė labai anksti. Jis baigė universitetą 18 metų, o būdamas dvidešimties, jis su tokiais garsiais bendraautoriais kaip V.Smirnovas ir V.Krylovas paskelbė pirmąją savo knygą *Variacins skaičiavimas*. Būdamas 22 metų jis tapo Leningrado universiteto profesoriumi, pagarsėjusiu rezultatais funkcijų analizės ir skaičiavimų matematikos srityse. 1937 metais į dvidešimt penkerių metų Leningrado universiteto profesorių ir prestižiškiausio matematikos V.Steklovo vardo instituto bendradarbį L.Kantorovičių kreipėsi faneros gamybos įmonė su prašymu padėti išspręsti sunkų gamybos planavimo uždavinį. Turimos aštuonios skirtingos faneros gamybos mašinos, kurios kiekviena gali gaminti penkių tipų produkciją. Tačiau mašinų pajėgumai gaminti tam tikro tipo produkciją yra skirtingi. Gamybos planas reikalavo, kad tarp išleidžiamų produkcijos rūšių būtų išlaikytos reikalingos proporcijos. Reikia paskirstyti darbus turimoms mašinoms taip, kad būtų išlaikytos reikalaujamos proporcijos, ir bendras produkcijos išleidimas būtų maksimalus. Visų galimų variantų perrinkti neįmanoma, nes reikėtų išspręst apie milijardą tiesinių lygčių sistemų su dvylika kintamųjų. Su panašiais uždaviniais matematikai, dirbę taikymų srityje, susidurdavo jau nuo senų laikų. Pavyzdžiui, žinomas prancūzų matematikas G.Monge panašų uždavinį suformulavo stengdamasis minimizuoti transporto išlaidas, susijusias su karinių įtvirtinimų statyba. Tačiau G.Monge'ui nepavyko išspręsti uždavinio. Matematikos tradicija orientavosi į lygčių sprendimą. Buvo žinomi dar Lagrange'o pasiūlyti metodai optimizavimo uždaviniams su lygybiniais ribojimais spręsti. Tačiau naujajam uždaviniui žinomi metodai netiko. Juos benagrinėdamas, L.Kantorovičius 1938 metais sukūrė naują metodą, kuris toli pranoko pradinio gamybinio uždavinio sprendimo poreikius. Remiantis sukurta teorija buvo pradėta optimaliai planuoti didelių sistemų darbą, svarbių projektų (ypač karinių) vykdymą ir pan., o taip pat nagrinėti tokias fundamentalias ekonomikos kategorijas kaip kainas. L.Kantorovičiaus rezultatai apie faneros gamybos planavimo uždavinio sprendimą buvo paskelbti Leningrado universiteto 1939 metais išleistoje brošiūroje. Apibendrinantis matematinis straipsnis buvo paskelbtas 1942 metais žurnale *Doklady Akademiji Nauk SSSR.* Kaip jau minėta aukščiau, keturiasdešimtaisiais metais JAV nepriklausomai atrasti tiesinio programavimo metodai buvo detaliau išvystyti ir iš karto pritaikyti ypatingos svarbos uždaviniams spręsti. Tačiau tas faktas nemenkina L.Kantorovičiaus pionieriško atradimo reikšmės. Gal būt, jo rezultatai pralenkė laiką ta prasme, kad tuometinėje SSSR nebuvo techninės įrangos (kompiuterių) sukurtiems metodams realizuoti. Be to, optimalaus planavimo metodai neatitiko vyraujančių ideologizuotos ekonomikos tendencijų. Kiek vėliau, kai susilpnėjo ideologijos diktatūra mokslui, L.Kantorovičiaus autoritetas buvo pripažintas ir sovietinių ekonomistų. Be daugelio jo mokslinių rezultatų pripažinimo ženklų buvusioje SSSR, jis buvo pripažintas tarptautinės mokslinės visuomenės kaip vienas iš žymiausių šio šimtmečio matematikų ir ekonomikos teoretikų. Jis pelnė aukščiausią mokslinį pripažinimą – Nobelio premiją (1975 metų apdovanojimas ekonomikos srityje).

Nuo pat simplekso algoritmo išradimo jis iš karto pradėtas plačiai taikyti ir visapusiškai tyrinėti. Vargu ar yra kitas algoritmas susilaukęs panašaus mokslininkų dėmesio ir taip dažnai vartojamas, nors kaip teigia algoritmo išradėjas G.Danzig, simplekso algoritmo idėja iš pradžių pasirodė neperspektyvi. Tokio pesimizmo priežastis mūsų jau buvo aukščiau pažymėta: iki bus surastas sprendinys nepalankiu atveju gali tekti pežiūrėti labai daug viršūnių. Laimei algoritmas buvo išbandytas. Jo efektyvumas pranoko visus lūkesčius. Negatyvi simplekso algoritmo praktinio efektyvumo pasekmė ta, kad buvo nuslopintas dėmesys kitoms galimoms idėjoms.

Situacija pasikeitė, kai 1979 metais L.Chačijanas įrodė elipsoidų metodo polinominį sudėtingumą. Ne taip dažnai didieji dienraščiai pirmajame puslapyje rašo apie matematinius rezultatus. New York Times L.Chačijano metodą pavadino ”matematiniu rusų sputniku”. Nors pats elipsoidų metodas ir netapo efektyvia priemone tiesinio programavimo uždaviniams spręsti, bet ledai buvo pralaužti. Buvo pradėtas intensyvus kitokių idėjų tyrimas. Didelę reikšmę turėjo 1984 metais paskelbtas N.Karmarkar'o metodas. Šiuo metu vidinio taško metodai sėkmingai konkuruoja su simplekso metodu įvairiose taikymų srityse.

4.12 Užduotys ir kontroliniai klausimai

1. Užrašykite tiesinio programavimo uždavinį

$$\max\ 2x_1 - 3x_2 + 5x_{4,}$$

$$\begin{aligned} -x_1 + x_2 - x_3 + x_4 &= 8 \\ 2x_1 + 4x_2 &= 10 \\ -x_3 - x_4 &\leq 3 \\ x_1 &\geq 0 \\ x_3 &\geq 0 \end{aligned}$$

standartine ir kanonine formomis.

2. Paaiškinkite bazinio neleistino sprendinio geometrinę prasmę.

3. Kokiais atvejais tiesinio programavimo uždavinys neturi sprendinio?

4. Ar pavyzdžio (4.4) pirmasis ir trečiasis stulpeliai sudaro bazę? Jei taip, atlikite atraminį šios bazės keitimą.

5. Kokiu atveju simplekso algoritmas "užsiciklina"?

6. Užrašykite tiesinio programavimo uždavino (4.4) dualųjį uždavinį.

7. Kokiu pagrindu tiesinio programavimo uždavinį galima pakeisti tiesinių lygybių ir nelygybių sistemos sprendimo uždaviniu? Atlikite tokį pakeitimą (4.4) uždaviniui.

8. Afininio poerdvio ir sferos, kurios centras prikklauso poerdviui, sankirta yra mažesnės dimensijos sfera. Karmarkar'o metodo realizacija numato tiesinės funkcijos minimizaciją taip apibrėžtoje srityje. Išspręskite šitokio uždavinio pavyzdį

$$\min\ x_1 + 2x_2 + 3x_3,$$

$$(x_1 - 1)^2 + (x_2 - 2)^2 + (x_3 - 3)^2 \leq 1,$$

čia uždavinio formulavime minimas afininis poerdvis sutampa su R^3.

9. Raskite įbrėžtos ir apibrėžtos apie $(n-1)$ dimensijos standartinį simpleksą sferų radiusus.

10. Įrodykite, kad (4.26) formule apibrėžtas atvaizdavimas yra atvirkštinis atvaizdavimui $P_\alpha(X)$.

RACIONALUS IŠRINKIMAS

5.1 Ar visada galime optimizuoti?

Nagrinėdami optimizavimo uždavinių sprendimo metodus, kaip savaime suprantmą dalyką priimdavome, kad optimizavimo uždaviniai yra adekvatus matematinis praktinių uždavinių modelis. Iš tiesų, daugelyje žmonių praktinės veiklos sričių optimizavimo uždaviniai iškyla labai natūraliai. Kita vertus, yra nemažai atvejų, kai optimizacijos tikslingumas intuityviai aiškus, bet optimizavimo uždavinio formulavimas nėra trivialus. Sakysim, investuojant kapitalą, norima gauti galimai didesnį pelną. Bet dažnai niekas negali garantuoti, koks bus investicijos rezultatas. Kartais projektas visiškai sužlunga, projektą vykdanti firma bankrutuoja. Kaip pasirinki investavimo būdą, jei jo rezultatai neaiškūs? Daugelis ne tik komercinės, bet ir inžinerinės veiklos rezultatų paaiškėja vėliau, o nuspręsti reikia dabar, atsižvelgiant į daugelį kriterijų, esant rezultatų neapibrėžtumui bei interesų konfliktams. Tokiose sudėtingose situacijose dažniausiai sprendžia aukšto rango vadybininkai, remdamiesi išmoktais vadybos metodais, patirtimi, intuicija. Bet ar visada jų sprendimai logiški? Gal būt, žmogaus intuicijai ir euristiniam sugebėjimams gali padėti teorija panaši į optimizaciją, puikiai pasiteisinusią paprastesniais atvejais. Kad būtų konkrečiau, nagrinėkime inžinerinį projektavimą [23]. Bet ne techninius jo aspektus, braižymą ir pan., o projektinio varianto išrinkimą, kaip esminį "*pakitimų žmogų supančioje dirbtinėje aplinkoje pradžios*" [23] faktorių. Projektavimo pavyzdys geras tuo, kad jam būdingi aukščiau minėti sunkumai, ir ši veikla turi būti racionali. Čia negalima pasikliauti atsitiktinumu ir rizikuoti, "viską pastatant ant kortos".

5.2 Išrinkimo uždavinių savybės

Projektavimo procesas susideda iš kelių, iteratyviai kartojamų etapų [23]. Kiekviename etape turi būti išrinktas variantas, kurio specifikacijas apibrėžia prieš tai buvusio etapo rezultatas. Kadangi visas projektavimo procesas yra iteracinio pobūdžio, tai ir išrinkimo uždaviniai sprendžiami patikslinus specifikacijas kelis kartus. Vien sveiku protu ir euristiniais metodais dažnai sunku surasti bent kelis variantus, patenkinančius duotas specifikacijas. Todėl abejotina, kad bus išrinktas geriausias variantas. Jei varianto, patenkinančio specifikacijas, visai nepavyksta surasti, tai dažnai nelengva įtikinti vadovus ir kitus projektuotojus, kad toks variantas neegzistuoja. Projektinių variantų išrinkimas reikalauja daug jėgų ir laiko, todėl stengiamasi jį, kiek įmanoma, automatizuoti naudojant matematinius metodus.

Silpniausia prielaida, įgalinani matematiškai nagrinėti išrinkimą (anglų kalba *choice*), yra dviejų variantų palyginamumas. Jei palyginus bet kokius du variantus galima pasakyti, kuris iš jų bent jau ne blogesnis už kitą, tai sakoma, kad variantų aibėje apibrėžtas preferencijos santykis. Savo struktūra matematinis išrinkimo uždavinio modelis turi įvertinti bendriausias preferencijos santykio savybes. Apibrėžti preferencijos santykį dažnai nelengva, dėl tokių aplinkybių:

1. *Kriterijų gausa.* Visas objektas (taip pat ir jo mazgai bei detalės) apibūdinamas ne vienu, o keliais kriterijais, pavyzdžiui, kaina, patikimumu, naudingo veiksmo koeficientu, patvarumu, objekto mase, energijos sąnaudomis ir pan. Kartais reikia atsižvelgti į sunkiai kiekybiškai įvertinamus kriterijus, pavyzdžiui, pastato fasado atitikimą architektūrinei aplinkai, gamybos atliekų įtaką gamtai ir pan. Kai kurie kriterijai yra prieštaringi. Kai kurie kriterijai apibrėžti projekto specifikacijomis. Kai kuriuos siekiama optimizuoti.

2. *Neapibrėžtumas.* Sudėtingesniame uždavinyje tiksliai numatyti projektinio varianto charakteristikų neįmanoma. Projektuojant visiškai naujus objektus, kai kurių duomenų paprasčiausiai nėra. Kitais atvejais, duomenis surinkti sunku arba brangu, pavyzdžiui, reikia atlikti tyrinėjimus. Sudėtingi objektai gali būti projektuojami net kelis metus, o patys objektai naudojami dešimtmečiais. Bėgant laikui pasikeičia vertinimo kriterijai, kainos, atsiranda naujos medžiagos ir įrenginiai, t. y. pasikeičia dirbtinė žmogų supanti aplinka ir visuomenės poreikiai; tam kartais gali turėti įtakos gamtiniai faktoriai, pvz. sausros, potvyniai. Projektuotojai dažnai nežino, kokie bus atskirų įmonių, organizacijų arba įtakingų žmonių sprendimai. Todėl ne visada galima tiksliai numatyti varianto išrinkimo pasekmes.

3. *Rizika.* Neigiamos pasirinkto sprendimo pasekmės paprastai paaiškėja įdiegus projektą. Jos gali būti labai sunkios, pavyzdžiui, katastrofa su žmonių aukomis. Tačiau ir mažesnes nesėkmes lydi nemalonumai. Vengiant rizikos, atsižvelgiama į įprastus, anksčiau pasiteisinusius sprendimus. Tačiau tokie sprendimai pernelyg konservatyvūs. Sunku įsivaizduoti sudėtingą projektą vykdytą be jokios rizikos.

4. *Konfliktai.* Didelę reikšmę lyginant variantus gali turėti faktoriai, kuriuos sunku įtraukti į matematinį projektuojamojo objekto modelį. Pavyzdžiui, projektavimo, gamybos arba mokslo įstaigų atstovai su vienu arba kitu projektiniu variantu gali sieti savo savanaudiškus tikslus. Įtaką projektuotojui gali daryti įvairios potencialių projektuojamojo objekto vartotojų grupės. Jei projektuotojų kolektyvo narių tikslai yra nevienodi, tai išrenkant variantą gali kilti konfliktai.

Užbaigdami šį įvadinį skyrelį, aptarsime dar vieną terminą. Be termino "išrinkimas" anglų kalboje naudojamas ir šiek tiek bendresnis terminas *decision making*, kuris lietuviškoje literatūroje pažodžiui verčiamas kaip "sprendimo priėmimas". Mūsų vartojamas terminas "projektuotojas" bendru atveju atitinka terminą "decision maker".

5.3 Naudingumo funkcija

Du projekto variantus (alternatyvas) lengviausia palyginti, jei svarbus tik vienas kriterijus. Pavyzdžiui, gamyklos projektų variantai gali būti lyginami pagal numatomą pelną. Variantas X_1 yra ne blogesnis negu variantas X_2, jei $f(X_1) \geq f(X_2)$, $X_i \in A$, $i = 1, 2$; čia A - variantų, tenkinančių projekto specifikacijas, aibė, $f(\cdot)$ - pelnas.

Preferencijos santykis tarp variantų (alternatyvų) žymimas ženklu $\succsim$. Paprastai sakoma, kad X_2 nedominuoja X_1, jei $X_1 \succsim X_2$. Jei $X_2 \succsim X_1$ ir $X_1 \succsim X_2$, tai alternatyvos vadinamos ekvivalenčiomis ir rašomos šitaip: $X_1 \sim X_2$. Jei $X_1 \succsim X_2$, bet jos nėra ekvivalenčios, tai sakoma, kad X_1 dominuoja X_2 ir rašoma šitaip: $X_1 \succ X_2$. Funkcija $f(\cdot)$ aukščiau minėtame pavyzdyje apibrėžia preferencijos santykį. Šis atitikimas nėra abipus vienareikšmis, nes ir kita funkcija $F(X) = \varphi(f(X))$ apibrėžia lygiai tą patį preferencijos santykį, čia $\varphi(\cdot)$- griežtai didėjanti funkcija. Lygindami variantus ir pagal pelną, ir pagal kokį nors kitą kriterijų, kuris yra monotoniškai didėjanti pelno funkcija, gausime tokį patį rezultatą.

Racionaliai išrinkti projekto variantą iš visų galimų variantų aibės galima tik tuomet, kai apibrėžtasis preferencijos santykis pasižymi tam tikromis racionalumo savybėmis. Sakykime, kad toks santykis apibrėžtas. Tačiau išrinkimo uždavinys būtų žymiai aiškesnis, jei, kaip anksčiau nagrinėtame pavyzdyje, žinotume analogišką pelnui funkciją, išreiškiančią preferencijos santykį. Tuomet išrinkimo uždavinį galėtume pakeisti tos funkcijos optimizavimo uždaviniu. Kyla klausimas, kokiomis savybėmis turi pasižymėti preferencijos santykis, kad egzistuotų jį reprezentuojanti funkcija.

Sakykime, kad duotas baigtinis projekto variantų skaičius $A = \{X_1, ..., X_m\}$, o peferencijos santykis yra toks, kad tarp nagrinėjamų variantų nėra ekvivalenčių. Ši prielaida nemažina bendrumo, nes ekvivalenčius variantus galima reprezentuoti vienu iš jų. Sakykime, kad preferencijos santykis tranzityvus, t.y. iš $X_1 \succ X_2$ ir $X_2 \succ X_3$ išplaukia, kad $X_1 \succ X_3$. Daugeliui žmonių ši prielaida priimtina. Tačiau, sprendžiant konkrečius uždavinius, dažnai paaiškėja, kad realūs lyginimai gali būti netranzityvūs. Todėl panagrinėkime logines netranzityvumo pasekmes.

Sakykime, kad $X_1 \succ X_2$, $X_2 \succ X_3$ ir $X_3 \succ X_1$. Kadangi nė viena alternatyva šiuo atveju neturi aiškių privalumų, apsistokime ties X_1. Tačiau X_1 dominuoja X_2, todėl sutiktume pasirinkti X_2 vietoj X_1 tik gaudami už tai kažkokį užmokestį a_1, pvz. sutaupydami statybos išlaidų. Kadangi X_2 dominuoja X_3, tai pakeisdami X_2 į X_3 gautume dar užmokestį a_2. Tačiau iš paskutiniosios alternatyvų poros palyginimo matome, kad X_3 dominuoja X_1, todėl pasiėmę užmokestį a_3 už X_3 pakeitimą į X_1, sugrįžtume prie pradinio varianto X_1 ir dar gauname pelną $a_1 + a_2 + a_3$. Tačiau tokios "pelno pompos", kuri iš nieko - pompuotų pelną, logiškai mąstant, negali būti. Vadinasi, dėl preferencijos netranzityvumo gali atsirasti prieštaravimų. Detaliau tranzityvumo savybės aptariamos specialioje literatūroje [7], [13].

Jei alternatyvų skaičius baigtinis, ir preferencijos santykis tranzityvus, tai alternatyvas galima sunumeruoti taip, kad geresnė alternatyva turėtų didesnį numerį

$$X_K \succ X_{K-1} \succ ... \succ X_1.$$

Apibrėžkime funkciją, kurios reikšmė lygi alternatyvos numeriui $U(X_i) = i$, $i = 1, ..., K$. Iš funkcijos apibrėžimo akivaizdu, kad $X_i \succ X_j$ tada ir tik tada, kai $U(X_i) > U(X_j)$, t.y. $U(\cdot)$ reprezentuoja duotąjį preferencijos santykį. Tokia funkcija vadinama **naudingumo funkcija.**

Tuomet, kai alternatyvų skaičius yra begalinis, preferencijos tranzityvumo nepakanka, kad egzistuotų naudingumo funkcija [7]. Neretai tenka nagrinėti uždavinius su kontinualia alternatyvų aibe. Dažniausiai tai n - matės erdvės poaibiai, kai kiekviena vektoriaus koordinatė atitinka vieną iš kriterijų. Nesileisdami į matematinius išvedžiojimus ir prielaidų aptarimą, suformuluosime bendrą išvadą: jei preferencijos santykis patenkina natūralias ir logiškas prielaidas, tai egzistuoja jį reprezentuojanti naudingumo funkcija. Taigi racionalaus projektinio varianto išrinkimo uždavinį iš esmės galima pakeisti naudingumo funkcijos optimizavimo uždaviniu. Tačiau praktiškai tuo pasinaudoti galima tik tuomet, kai žinomas konstruktyvus naudingumo funkcijos apskaičiavimo algoritmas. Atkreipsime dėmesį į tai, kad preferencijos santykis naudingumo teorijoje yra pirminis. Naudingumo funkcija jį tik aprašo. Bet kuri kita funkcija, gauta monotoniškai transformavus naudingumo funkciją, taip pat reprezentuoja nagrinėjamąjį preferencijos santykį ir todėl taip pat yra naudingumo funkcija.

Kaip sudaryti naudingumo funkciją, gana plačiai aprašyta literatūroje. Iš pradžių teoriniais samprotavimais pagrindžiama, kokiai funkcijų klasei konkrečiu atveju turi priklausyti naudingumo funkcija. Po to, naudojantis ekspertine informacija (atsižvelgiant į tai, kaip projektinius variantus palygina specialistai) identifikuojama konkreti funkcija. Dažnai antrasis etapas reiškia nežinomų parametrų įvertinimą.

Išnagrinėkime paprasčiausią pavyzdį, kai projekto variantas apibūdinamas vieninteliu kriterijum - pelnu. Kadangi variantą galima sutapatinti su jį atitinkančiu pelnu, tai alternatyvų aibė yra neneigiamų skaičių intervalas. Aišku, kad didėjant pelnui, naudingu-

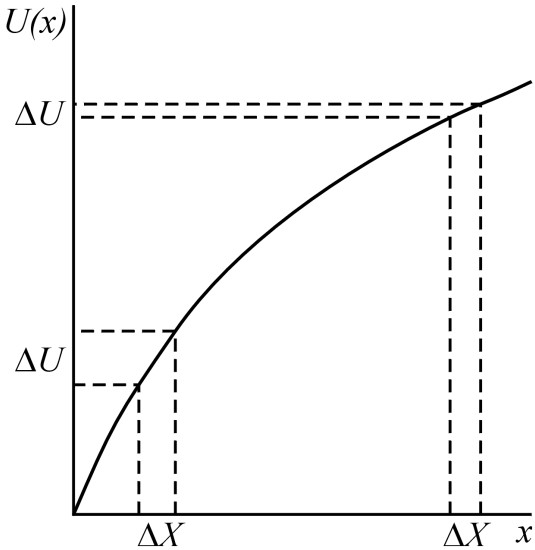

Figure 5.15: Naudingumo funkcijos augimo greitis mažėja, augant x.

mas didėja, todėl $U(\cdot)$ yra didėjanti funkcija. Tačiau pelno prieaugiui, sakykime, dešimčiai litų, suteikiama nevienoda reikšmė, nes ji priklauso nuo to, koks buvo pradinis pelnas: 100 ar 1000 litų. Antruoju atveju prieaugis paprastai vertinamas kaip žymiai mažiau naudingas. Todėl funkcijos augimo greitis didėjant pelnui mažėja (5.15 pav.).

Tokią naudingumo funkciją galima aproksimuoti eksponentine funkcija

$$U(x) = a + b\exp(cx),$$

čia koeficientai b ir c neigiami.

Kaip jau buvo minėta anksčiau, $\varphi(U(x))$ (čia $\varphi(\cdot)$ griežtai didėjanti) yra naudingumo funkcija, reprezentuojanti preferencijos santykį. Pastarasis dažnai pasižymi papildomomis savybėmis, iš kurių galima spręsti, kodėl viena alternatyva geresnė už kitą. Pavyzdžiui, preferencija gali būti apibrėžta ne tik pačioms alternatyvoms, bet ir jų poroms. Jei $(X_1, X_2) \succsim (X_3, X_4)$ reikštų, kad perėjimas iš X_1 į X_2 yra ne blogesnis negu perėjimas iš X_3 į X_4 ir preferencija pasižymėtų racionalumo savybėmis, tai naudingumo funkcijai patenkintų nelygybę $U(X_2) - U(X_1) \geq U(X_4) - U(X_3)$. Galima įrodyti, kad šitokia savybė invariantiška ne bet kokiai monotoniškai didėjančiai funkcijai $\varphi(\cdot)$, o tik tiesinei funkcijai $\varphi(x) = ax + b$, $a > 0$. Šis atvejis atitinka įprastą matavimo vienetų pasirinkimo laisvę. Kaip žinia, fizikos dėsniai nepriklauso nuo to, kaip matuojama masė (gramais, kilogramais, svarais), atstumas (metrais, kilometrais, pėdomis, coliais) ir kiti fiziniai dydžiai. Be to, šis specialus naudingumo funkcijos atvejis svarbus tuomet, kai lyginamas vidutinis naudingumas esant neapibrėžtumui, o taip pat sprendžiant daugiakriterio optimizavimo uždavinius.

Panagrinėkime atvejį, kai yra daug kriterijų. Jei kiekvieną alternatyvą pilnai charakterizuoja kriterijų vektorius $F = F(X)$ (čia $F(\cdot) = (f_1(\cdot), ..., f_m(\cdot)))$, tai galime nagrinėti naudingumo funkciją apibrėžtą kriterijų vektoriams, o ne pačioms alternatyvoms, t.y. $U(F)$. Kartais bendrą naudingumo funkciją lengviau sudaryti netiesiogiai, t.y. ne kriterijams $(f_1, ..., f_m)$, o jų naudingumui. Pirmiausia sukonstruojamos atskirų kriterijų naudingumo funkcijos $z_i = \nu_i(f_i)$, $i = 1, ..., m$. Po to, sudaroma kompozicijos funkcija $u(Z(F)) = U(F)$. Plačiau šis klausimas nagrinėjamas kitame knygos skyriuje. Čia pažymėsime tik vieną plačiai naudojamą kompoziciją - kriterijų tiesinę kombinaciją $u(Z) = a_1 z_1 + a_2 z_2 + ... + a_m z_m$. Šiuo atveju, naudojantis ekspertine informacija, tereikia įvertinti kriterijų lyginamuosius svorius

5.4 Neapibrėžtumas

Kai kurios projektuojamų objektų charakteristikos nustatomos visiškai tiksliai, pavyzdžiui, kiek ir kokios rūšies statybinių medžiagų reikia numatomam siuvimo cecho pastatui. Tačiau tiksliai nustatyti, kiek jis kainuos, sunkiau, nes keičiasi ir statybinių medžiagų, ir darbo jėgos kainos. Dar sunkiau įvertinti viso cecho kainą. Kol bus užbaigta statyba, gali atsirasti naujų mašinų ir technologinių procesų, todėl daug kas gali keistis net paskutiniojoje projekto realizavimo stadijoje. Yra pavyzdžių, kai reali statinio kaina kelis kartus viršijo projektinę. Būsimo cecho pelnas priklauso nuo daugelio neapibrėžtų ir kintančių faktorių, todėl gali net kilti abejonių, ar prasminga jį kažkaip vertinti. Tačiau pelno kriterijus labai svarbus lyginant kelis konkuruojančius projektus.

Projektuotojai įpratę intuityviai vertinti įvairius kriterijus neapibrėžtumo sąlygomis, nes jų darbas pradeda keisti ateities dirbtinę aplinką. Ateitis susijusi su dabartimi dialektiškai sąveikaujančiais būtinumo ir atsitiktinumo ryšiais. Kuo sudėtingesnis projektuojamasis objektas, kuo tolimesnėje ateityje jis bus realizuotas, tuo svarbesnė antroji komponentė, t. y. tuo didesnis neapibrėžtumas vertinant projektą. Norėdami aptarti matematinius metodus, turime išsiaiškinti, kokie kiekybiniai neapibrėžtumo matai naudotini, kokie modeliai ir metodai taikytini lyginant projektinius variantus neapibrėžtumo sąlygomis.

Ilgą laiką objektai ir procesai, kurių nebuvo galima tiksliai apibrėžti, mokslo metodais apskritai nebuvo tyrinėjami. Tačiau šiuolaikiniai projektavimo ir valdymo uždaviniai būtent tokie. Pavyzdžiui, lyginant projektuojamus objektus pagal pelno kriterijų U, turime atsižvelgti į tai, kad šis priklauso ne tik nuo projektuotojo parenkamų parametrų $X = (x_1, ..., x_n)$, bet ir nuo nevaldomų faktorių $Y = (y_1, ..., y_m)$, kurių reikšmės projektuojant nežinomos, t. y. $U = U(X, Y)$. Jei nelygybė $U(X_1, Y) \geq U(X_2, Y)$ teisinga visoms Y

reikšmėms, tai pirmasis variantas yra, be abejonių, geresnis už antrąjį. Jei $U(X_1, Y) \geq U(X_2, Y)$ tik kai kurioms Y reikšmėms, tai atsakymas nevienareikšmis. Vargu ar galima pagrįsti vienokį ar kitokį išrinkimo principą, jei neapibrėžtumas neformalizuojamas kokiomis nors racionalumo prielaidomis. Nagrinėtas pelno kriterijus bendru atveju gali būti suprantamas kaip apibendrinantis daugelį kriterijų **naudingumas**.

5.5 Garantuoto rezultato principas

Atsargūs žmonės, prieš pradėdami dirbti, dažnai pagalvoja, koks bus rezultatas nepalankiai susiklosčius aplinkybėms. Vadovaudamiesi šiuo principu, anksčiau pateiktame pavyzdyje projektinius variantus turėtume lyginti pagal jų minimalų pelną Y atžvilgiu

$$\min_{Y \in B} U(X_1, Y) \text{ ir } \min_{Y \in B} U(X_2, Y),$$

čia B - nežinomų faktorių vektoriaus Y galimų reikšmių aibė. Tuomet optimalų variantą gautume maksimizuodami minimaliai galimą pelną $\min_{Y \in B} U(X, Y)$

$$\max_{X \in A} \min_{Y \in B} U(X, Y).$$

Šiuo atveju neapibrėžtumo modelis – nežinomų faktorių vektoriaus galimų reikšmių aibė B. Kadangi garantuoto rezultato principas naudojamas analizuojant nepalankiausias sąlygas, o minimumas dažniausiai pasiekiamas esant ribinėms Y reikšmėms, labai svarbu objektyviai apibrėžti aibės Y ribą, atitinkančią nepalankiausias sąlygas. Pavyzdžiui, projektuojant cecho pastatą, reikia atsižvelgti į nepalankiausias klimatines, seismines sąlygas, prasčiausią medžiagų kokybę ir kt. Gali kilti klausimas, kokia turi būti minimali temperatūra apibrėžiant aibę B cecho statybos pavyzdyje. Ilgaamžės praktikos suformuotos statybinės normos orientuojamos į pasitaikančias ekstremalias reikšmes bei vidutinę žiemos temperatūrą, bet ne į žemiausią fiziškai įmanomą temperatūrą. Nagrinėtame pavyzdyje garantuoto rezultato principo panaudojimas yra ginčytinas. Tačiau jis visiškai natūralus, jei faktorių vektorių Y valdo priešiški projektuotojui veiksniai. Detaliau antagonistinių interesų susidūrimą analizuoja lošimų teoriją, kurioje šis principas išvedamas iš aksiomų, apibrėžiančių konflikto sprendinį.

Garantuoto rezultato reikia siekti ir tuomet, kai pažeidus projektuojamų objektų, pavyzdžiui, galingų energetikos objektų, transporto sistemų, darbo režimą, galimos sunkios, katastrofiškos pasekmės. Projektuotojas priverstas tarsi, persikūnyti į savo priešininką ir ieškoti tokių sąlygų Y, kurioms susidarius gali kilti katastrofos grėsmė. Pavyzdžiui, projektuojant atominę elektrinę, reikia įsitikinti, ar reaktorius nesusprogs esant visoms galimoms nevaldomų faktorių vektoriaus Y reikšmėms, t. y. ir tokioms, kurios labai mažai tikėtinos.

Išskyrėme būdingiausius atvejus, kai reikėtų taikyti garantuoto rezultato principą. Praktiškai jis taikomas dažniau, o kartais ir absoliutinamas kaip vienintelis objektyvus, nepriklausantis nuo subjektyvių faktorių.

Jei žinome aibę B, tai galime suformuluoti teiginį, skambantį kaip matematikos teorema: visiems $Y \in B$ maksimizavimo $\max_{X\in A} U(X,Y)$ rezultatas ne blogesnis už garantuotą, t.y.

$$\max_{X\in A} \min_{Y\in B} U(X,Y).$$

Tačiau matematikos teoremose, kuriuose yra teiginys "*visiems* $Y \in B$ *galioja*", abstrakti aibė B natūraliai sudaro pačios matematinės problemos dalį, pvz. diferencijuojamųjų funkcijų aibė, natūrinių skaičių aibė ir kt. Nevaldomų faktorių aibės B abibrėžimas projektavimo uždaviniuose dažniausiai priklauso nuo subjektyvių nežinomų faktorių įverčių, pvz., būsimų muito mokesčių bei žaliavų kainų, kvalifikuotų darbuotojų samdymo galimybių ir pan. Todėl pripažindami, kad garantuoto rezultato principas atitinka matematikos tradicijas, negalime nekreipti dėmesio į tai, kad jis vienpusiškai orientuotas į ekstremalias sąlygas.

Projektuojant plataus vartojimo gaminius, svarbiau atsižvelgti ne į ekstremalias, bet į labiausiai tikėtinas, vidutines sąlygas, kurios apsprendžia, ar gaminys bus ekonomiškai priimtas. Reziumuosime tokiu palyginimu: garantuoto rezultato principas, be abejo, racionalus projektuojant povandeninio laivo radio aparatūrą, bet vargu ar šis principas tiks projektuojant buitinį radijo imtuvą.

5.6 Vidutinio naudingumo principas

Masinio naudojimo gaminiai gali būti eksploatuojami esant įvairioms nevaldomų faktorių vektoriaus Y reikšmėms. Kuo dažnesnės, labiau tikėtinos sąlygos, tuo didesnė jų reikšmė projektuojant gaminį. Sakykime, kad Y-atsitiktinis. Jei žinomas vektoriaus Y tikimybinis tankis $p(Y)$, tai vidutinį projektinio varianto X naudingumą galima išreikšti formule

$$u(X) = \int U(X,Y)\, p(Y)\, dY. \tag{5.1}$$

Vidutinis naudingumas – tai bendras įvairiomis sąlygomis eksploatuojamų gaminių naudingumas. Atskiru atveju, kai $p(Y)$ išreiškia eksploatavimo sąlygų dažnius, $U(X,Y)$ – projekto varianto X pelną, gaunamą esant sąlygomis Y, tuomet $u(X)$ lygus vidutiniam pelnui. Todėl, siekiant maksimizuoti pelną, konstruktyvius parametrus natūralu parinkti maksimizuojant $u(X)$.

Vidutinio naudingumo principo racionalumas sprendžiant panašius uždavinius nekelia abejonių. Tačiau gali būti sunku parinkti tikimybių skirstinį. Remiantis tradicinais tikimybių teorijos vadovėlias reikėtų ištirti vektoriaus Y realizacijų įvairių reikšmių poaibių dažnius. Jei šie stabilūs, tai Y gali būti laikomas stochastiniu, o jo skirstinį (atskiru atveju tikimybinį tankį) galima įvertinti statistiniais metodais naudojant duomenis. Klasikinis šitoks pavyzdys, ištyrę kaip dažnai iškrinta atitinkamos lošimo kauliuko kombinacijos ir įsitikinę jų stabilumu, ateityje įvairių kombinacijų tikimybes priimtume lygiomis gautiems pagal eksperimento rezultatus dažniams.

Deja, sprendžiant realius uždavinius, neapibrėžtumai nepalyginamai sudėtingesni. Vienareikšmiškai nustatyti ateityje vyksiančių reiškinių pobūdį pagal praeityje stebėtus atsitiktinius reiškinius ne visada įmanoma. Pagrįsti vidutinio naudingumo principą dar sudėtingiau tuomet, kai projekto realizacija vienintelė. Pavyzdžiui, nagrinėtame pavyzdyje Y – tai technologinių įrenginių, numatomų pirkti užsienio rinkoje, kainos, o $U(X,Y)$ – projektuojamojo cecho pelnas. Išrenkant projekto variantą X, kainų vektorius nežinomas. Nors prielaida, kad X yra atsitiktinis dydis atrodo pilnai priimtina, bet ar šiuo atveju tikrai pagrįstas vidutinio pelno maksimizavimas. Tiesiogiai remtis klasikinės tikimybių teorijos išvadomis tuomet, kai kalbama apie pavienį atsitiktinį įvykį, kažin ar korektiška.

Tikimybių teorija nagrinėja masinius atsitiktinius reiškinius, bet panagrinėkime vienkartinio dalyvavimo loterijoje pavyzdį. Kodėl žmogus pirkdamas vieną bilietą, turėtų galvoti apie **vidutinį** išlošimą? Tuo vadovaujasi tie, kurie pavienį įvykį interpretuoja kaip vieną iš daugelio analogiškų įvykių realizaciją ir yra linkę vienodose situacijose vienodai elgtis. Tokie racionalūs žmonės, tikriausiai, retai dalyvauja loterijose ir azartiniuose lošimuose. Tačiau projektuotojas **privalo** išrinkti vieną iš galimų variantų, t.y. jis negali išvengti dalyvavimo "loterijoje". Todėl, būdamas racionaliu, jis norėtų rinktis tą variantą, kurio vidutinis naudingumas didžiausias. Nagrinėtame pavyzdyje projektuotojas, matyt, kreiptųsi į užsakovą, kad pastarasis užsakytų rinkos tyrimo specialistams atlikti įreginių kainų prognozę, t.y. įvertinti kainų skirstinio tankį $p(Y)$, kuriam šiuo atveju labiau tiktų subjektyvių tikimybių, o ne dažnumų interpretacija. Turėdamas tokį įvertį projektuotojas galėtų išrinkti variantą maksimizuodamas vidutinį pelną, kuris apibrėžiamas formule (5.1).

5.7 Subjektyvios tikimybės

Ateities dirbtinės aplinkos modelis sudaromas remiantis esminiais, nekintamais gamtos ir visuomenės dėsniais. Kai kurie iš jų formuluojami remiantis tikimybių teorija. Šią modelio dalį apima filosofinė būtinumo kategorija. Jos priešybė – modelio neapibrėžtumai. Jei pastaruosius ir galima matematiškai aprašyti tai gautasis modelis vis tiek bus subjektyvus. Kad būtų paprasčiau, panagrinėkime vieną modelio parametrą, pvz. vieno iš technologinės įrangos, skirtos projektuojamam siuvimo cechui, varianto kainą pasaulinėje rinkoje po penkerių metų. Įvertinant bendras politines ir ekonomines tendencijas, matyt, galima įvertinti ir bendrą kainų kitimo tendenciją. Taip pat reikia atsižvelgti į specifines siuvimo pramonės ir mašinų gamybos ypatybes ir kt. Jei ekspertas įvertins šią kainą 0.5 milijono litų, tai ekspertizės užsakovas tikrai neturės pretenzijų dėl 10 tūkstančių litų paklaidos. Tačiau, ar negali kaina pakilti iki 0.7 milijono? Ar ji gali viršyti milijoną? Išreiškiamas vienu skaičiumi ekspertinis kainos įvertis kažin ar pakankamas, nors jis ir atitinka labiausiai tikėtiną reikšmę. Visas reikšmių intervalas yra daugiau ar mažiau tikėtinas, ir į tai reikia atsižvelgti išrenkant variantą.

Kasdieniniame gyvenime be kiekybinių vertinimų, pavyzdžiui, skaičiais išreiškiamų ilgio, svorio, laiko matų sutinkami ir mažiau tikslūs teiginiai: "*labai garsiai kalba*", "*gražiai piešia*", "*intensyviai treniruodamasis pasieks gerų sporto rezultatų*". Ar galima tokiems vertinimams suteikti tikslesnę prasmę, t.y. išmatuoti juos naudojantis kokia nors skale? Tam pirmiausia būtina, kad vertinamos savybės realizacijos būtų palyginamos. Šis reikalavimas išreiškiamas binariuoju palyginimo santykiu, paprastai žymimu $\succ$. Pavyzdžiui, teiginį " *a kalba garsiau už b*" galima užrašyti šitaip: $a \succ b$. Jei binarusis palyginimo santykis tenkina tam tikrus racionalumo reikalavimus, tai nagrinėjamą savybę galima išmatuoti. Žmonės lengvai palygina garso stiprumą. Turbūt todėl, jau seniai sukurta skalė ir specialūs prietaisai garso stiprumui matuoti. Palyginti piešinius sudėtingiau. Nors stojančiųjų į dailės mokyklas piešinių kokybę vertinimo komisija įvertina balais, t. y. išmatuoja, bet kiek dėl to būna ginčų! Kaip palyginti savybės išraiškas, minimas trečiajame teiginyje, apskritai neaišku.

Šioje knygoje nepateiksime tikslaus palyginimo santykio racionalumo apibrėžimo ir jo detaliau neaptarsime. Apie tai rašoma specialiuose matavimo teorijos veikaluose. Gal būt, skaitytojui bus įdomu, kad palyginimo santykio reprezentavimo problemos atsiradimą lėmė psichologijos, pedagogikos ir kai kurių kitų "netikslių" mokslų poreikiai. Pažymėsime, kad matuoti psichologines savybes sudėtingiau negu fizines, nes pirmąsias sunkiau apibrėžti, izoliuoti, tiesiogiai stebėti, realizuoti reikalingu intensyvumu. Tokios problemos būdingos ir matuojant subjektyvią tikimybę.

Anksčiau aprašytame pavyzdyje ekspertas tikriausiai pradės lyginti dviejų galimų kainų intervalų tikėtinumus. Kas labiau tikėtina, ar kaina tarp 0.4 ir 0.5 milijono, ar kaina didesnė 0.7 milijono? Panagrinėsime bendrą vieno faktoriaus atvejį. Raidėmis A, B, C pažymėsime uždarus intervalus, o žyme $\succ$ - tikėtinumo santykį. Užrašas $A \succ B$ reiškia, kad A labiau tikėtinas negu B. Aptarsime keletą racionalumo sąlygų.

1. **Pilnumas.** Esant bet kuriems dviem A ir B, arba $A \succ B$, arba $B \succ A$, arba $A \backsim B$; čia ekvivalentumo sąryšiu $\backsim$ paneigiamos pirmosios dvi alternatyvos.

2. Tranzityvumas. Jei $A \succ B$ ir $B \succ C$, tai $A \succ C$.

3. Adityvumas. Sakykime, kad C turi tik po vieną bendrą tašką su A ir B. $A \succ B$ tuomet ir tik tuomet, kai $A \cup C \succ B \cup C$.

4. Tolydumas. Jei $A \succ B$, tai intervalo A pradžioje ir gale galima atskirti tokius pointervalius A_- bei A_+, kad $A_- \backsim A_+ \backsim B$.

Šios sąlygos, galima sakyti, neatsiejamos nuo tikėtinumo santykio. Detaliai jos aptariamos specialioje literatūroje [13]. Prie šių sąlygų pridėję dar keletą, gauname išvadą [20]: egzistuoja vienintelis tikimybinis tankis $p(s)$, suderintas su tikėtinumo santykiu: $A \succ B$ tada ir tik tada, kai

$$\int_A p(s)ds > \int_B p(s)ds.$$

Remiantis šiuo rezultatu, neapibrėžtumo faktorius interpretuotinas kaip atsitiktinis dydis su tikimybiniu tankiu $p(\cdot)$. Kita vertus, tai reiškia, kad eksperto informacija apie neapibrėžtumą gali būti išreikšta tikimybiniu skirstiniu. Neapibrėžtumo išreiškimas standartiniu būdu, t. y. tikimybių skirstiniu, pasižymi daugeliu privalumų, pavyzdžiui, kur kas aiškesni tampa įverčių, kuriuos pateikia įvairūs ekspertai, panašumai ir skirtumai. Žinomi metodai, skirti daugelio ekspertų informacijai apie tą patį faktorių integruoti. Todėl lengviau atpažinti klaidas ir suderinti ekspertų informaciją apie nagrinėjamus faktorius. Subjektyvius modelius galima "adaptuoti" įvertinant stebėjimų duomenis. Tikimybinį subjektyvaus neapibrėžtumo modelį galima naudoti vidutiniam projektinių variantų naudingumui apskaičiuoti. Matematiniai metodai lieka tie patys, skiriasi tik gaunamų rezultatų interpretacija.

5.8 Savage'o teorija

Racionalaus išrinkimo neapibrėžtumo sąlygomis uždavinys buvo nagrinėjamas keliais etapais. Pirmiausia paaiškinome, kad esant apibrėžtumui racionalaus išrinkimo uždavinys pakeičiamas naudingumo funkcijos $U(X)$ maksimizavimo uždaviniu. Jei išrinkimas atliekamas neapibrėžtumo sąlygomis, tai sudaromas tikimybinis neapibrėžtumo modelis, kuris atskiru atveju yra tikimybinis tankis $p(s)$. Tuomet naudingumo funkcija taip pat priklauso nuo neapibrėžto faktoriaus s: $U(X) = U(X, s)$, ir kiekvieną variantą X galima apibūdinti vidutiniu naudingumu

$$u(X) = \int U(X, Y)\, p(Y)\, dY.$$

Jei šis kriterijus priimtinas, tai išrinkimo neapibrėžtumo sąlygomis uždavinį galima pakeisti vidutinio naudingumo maksimizavimo uždaviniu [7]. Projektuotojai, tur būt, šias teorines išvadas vertintų pirmiausiai atsižvelgdami į tai, ar sudėtinga sprendžiant konkretų uždavinį parinkti naudingumo funkciją ir tikimybinį tankį.

Norėtume labai trumpai aptarti Savage'o teorijos pagrindines idėjas. Šioje teorijoje subjektyvių tikimybių ir naudingumo funkcijos egzistavimas įrodomas bei vidutinio naudingumo kriterijus pagrindžiamasremiantis kiek kitokiomis prielaidomis: lyginant neapibrėžtų rezultatų situacijas.

Sakykim, kad $\mathbf{S} = \{s\}$ yra situacijas apibrėžiančių būvių aibė (states of the world), $\mathbf{C} = \{c\}$ - pasekmių aibė, $\mathbf{D} = \{d(\cdot)\}$ sprendimo funkcijų aibė, kurios argumentai yra iš aibės $\mathbf{S}$, o reikšmės priklauso aibei $\mathbf{C}$. Vaizdžiu "avantiūros" (gamble) atveju $\mathbf{S}$ yra neapibrėžtumo sąlygomis įvykstantys įvykiai, $\mathbf{C}$ - galimų išlošimų aibė. Projektuotojas, nežinodamas koks būvis bus realizuotas, turi pasirinkti funkciją $d(\cdot) \in \mathbf{D}$. Čia pasirinkimas neapibrėžtumo ir rizikos sąlygomis vadinamas "avantiūra". Preferencijos santykis tarp $\mathbf{D}$ elementų atitinka preferencijas tarp "avantiūrų". Jei šis santykis patenkina labai bendras racionalumo prielaidas, tai egzistuoja tokia naudingumo funkcija $U(\cdot)$ ir toks tikimybinis skirstinys (atskiru atveju tikimybinis tankis $p(\cdot)$), kad $d(\cdot) \succ g(\cdot)$ tada ir tik tada, kai

$$\int U(d(s))\, p(s)\, ds > \int U(g(s))\, p(s)\, ds\,.$$

Reziumuojame: jei "avantiūras" galima racionaliai palyginti, tai egzistuoja toks tikimybinis neapibrėžtumo modelis ir tokia naudingumo funkcija, kad "avantiūras" galima lyginti pagal jų vidutinį naudingumą.

5.9 Neryškios aibės

Labai artima neapibrėžtumui yra neryškumo (neaiškumo) sąvoka. Šiuo terminu (fuzziness) pavadinta ir nauja mokslo kryptis. Neryškių aibių teorija pradedant L. Zadeh straipsniu 1965 metais labai intensyviai vystėsi. Neryški aibė $A \in R^1$ apibrėžiama priklausomumo funkcija $M_A(x)$, kurios reikšmės priklauso intervalui [0, 1] (5.16 paveikslas). Kiekvienam argumentui x funkcijos $M_A(x)$ reikšmė apibrėžia priklausomumo aibei A laipsnį. Neryški aibė atitinka tradicinį aibės apibrėžimą, jei priklausomumo funkcijai įgauna tik dvi reikšmes: 0 ir 1. Tuomet $M_B(x)$ yra stačiakampė (5.16 paveikslas) ir nurodo, kad x arba priklauso aibei B, arba jai nepriklauso, t. y. jokių tarpinių priklausomumo laipsnių nėra.

Tie projektų parametrai, kurių tikslios reikšmės nežinomos, gali būti pakeisti neryškiomis aibėmis. Todėl neryškių aibių teorija gali būti panaudota išrenkant projektinius variantus. Panagrinėkime, kokie skirtumai tarp tradicinio tikimybinio modelio ir modelio, sudaryto neryškių aibių metodu. Jei visos aibės apibrėžtos baigtiniame intervale, tai priklausomumo funkcijos gali būti sunormuotos taip, kad jų integralai būtų lygūs 1. Tuomet priklausomumo funkcija atitiktų tikimybinį tankį. Tačiau formaliai tikimybinio tankio sąvoka siauresnė, nes priklausomumo funkcijos integralas gali ir nebūti baigtinis (pvz., $M_A(x) = 1$, $-\infty \le x \le \infty$). Neryškios aibės nuo tikimybinių

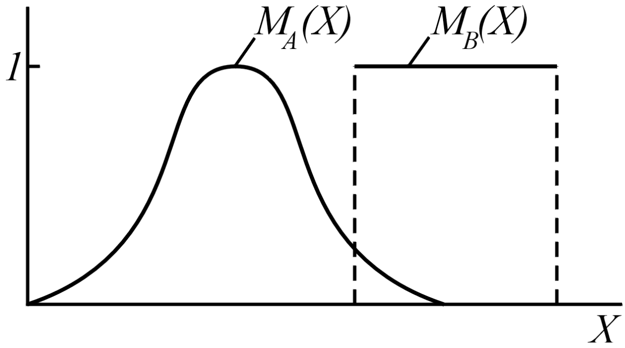

Figure 5.16: Priklausomumo funkcijų pavyzdžiai

modelių skiriasi ir tuo, kad priklausomumo funkcijoms apibrėžiamos tokios operacijos, kurios neturi analogų tikimybių teorijoje.

Yra nemažai darbų, apibendrinančių anksčiau nagrinėtą vidutinio naudingumo principą, kuriuose naudingumas bei tikimybės išreikšiami neryškiomis aibėmis. Tuo pabrėžiama, kad nei naudingumo, nei tikimybės negalima tiksliai žinoti. Neryškių aibių teoriją taip pat galima sėkmingai taikyti netiksliems samprotavimams modeliuoti. Tačiau negalima teigti, kad neryškios aibės visiškai pakeičia labiau tradicinius tikimybinius neapibrėžtumo modelius. Pasirenkant metodą konkrečiam uždaviniui spręsti, reikia kritiškai palyginti įvairias galimybes, nes ne visada naujesnis metodas yra automatiškai geresnis.

5.10 Psichologinės preferencijos santykio savybės

Įdomu palyginti realias žmonių preferencijų savybes su anksčiau išdėstytomis teorinėmis. Literatūroje šis klausimas gana plačiai aptartas, o šiame skyrelyje pateiksime kai kuriuos apibendrintus psichologinių tyrimų rezultatus [7]. Eksperimentai, kuriais remiantis daromos išvados, yra abstraktaus pobūdžio, jų sąlygos suprantamos plačiam potencialių psichologinių eksperimento dalyvių ir skaitytojų ratui. Būtų labai įdomu, jei skaitytojas analogiškus tyrimus atliktų konkrečiomis projektinių variantų lyginimo sąlygomis.

Jau anksčiau aptarėme, kad alternatyvas neapibrėžtumo sąlygomis reikėtų lyginti pagal jų vidutinį naudingumą. Tačiau panagrinėkime tokį pavyzdį. Sakykime, kad realizuodami kokį nors sumanymą, jūs privalote pasirinkti vieną iš dviejų alternatyvų:

1) garantuojančią 80 litų pelną;
2) su tikimybe 0,85, kad bus 100 litų pelnas;
   tikimybė, kad jokio pelno nebus, lygi 0,15.

Kurią alternatyvą jūs, skaitytojau, pasirinktumėt? Teoriškai racionalaus išrinkimo būdas rekomenduoja antrąją alternatyvą, nes jos vidutinis pelnas $100 \times 0.85 + 0 \times 0.15 = 85$ litai (pirmosios 80 litų). Tačiau daugelis psichologinio eksperimento, aprašyto literatūroje, dalyvių pasirinko pirmąją.

Vienas iš paaiškinimų toks: pasirinkus antrąją alternatyvą, yra tikimybė nieko negauti. Žmogus, matyt, įvertina tą nemalonią psichologinę būseną, kurion pakliūtų, jei pasirinkęs antrąją alternatyvą nieko negautų. Jei apsiribotume šiuo paaiškinimu, tai alternatyvas turėtume apibūdinti ne vien pelnu, bet ir psichologinėmis pasekmėmis, t. y. vektoriniu dydžiu, kurio antrąją komponentę nežinia kaip matuoti.

Eksperimento rezultatus galima paaiškinti ir kitaip, neatmetant racionalaus išrinkimo prielaidos.

Sakykime, pelno naudingumas yra netiesinė pelno funkcija: $y = U(x)$. Jei $U(\cdot)$ yra įgaubta funkcija (kaip 5.17 paveiksle), tai santykis tarp vidutinio abiejų alternatyvų naudingumo gali būti būtent toks, kokį išreiškė psichologinio eksperimento dalyviai. Kodėl piniginio pelno naudingumas turėtų būti įgaubta didėjanti funkcija, jau anksčiau buvo minėta.

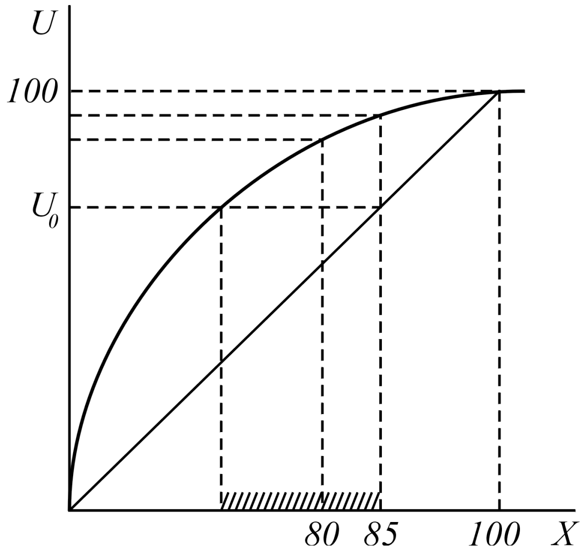

Figure 5.17: Loterijų palyginimas naudingumo atžvilgiu

Jei alternatyvos apibūdinamos kriterijumi, kurio naudingumas yra įgaubta funkcija, tai racionaliam išrinkimui būdingas rizikos vengimas. Ši savybė išplaukia iš matematinio įgaubtos funkcijos apibrėžimo:

$$U(x_0) > U_0 = p_1 \times U(x_1) + p_2 \times U(x_2),$$

čia $p_1 + p_2 = 1$, $x_0 = p_1 x_1 + p_2 x_2$, t. y. garantuoto pelno alternatyva x_0 dominuoja loterijas su tikimybe p_1, duodančias didesnį negu x_0 pelną, bet su tikimybe p_2, duodančias mažesnį negu x_0 pelną. Kadangi funkcija tolydi, tai yra nemažai x reikšmių, pavyzdžiui užbrūkšniuotas intervalas 5.17 paveiksle, kurioms esant garantuoto pelno (mažesnio už vidutinį loterijos pelną) alternatyvos dominuoja nagrinėjamąsias loterijas.

Analogiškai galima išnagrinėti pavyzdį, kai tenka išrinkti alternatyvą nuostolių atveju. Sakykime, vėl galimos dvi alternatyvos :

1) garantuojanti 80 litų nuostolius;
2) su tikimybe 0.85, kad bus 100 litų nuostoliai
(tikimybė, kad nuostolių nebus, lygi 0.15).

Racionalaus išrikimo būdas rekomenduoja pirmąją alternatyvą. Tačiau daugelis pasirenka antrąją. Psichologiškai sunkiau susitaikyti su garantuotais nuostoliais, negu bandyti jų išvengti, nors antruoju atveju galima turėti didesnius nuostolius. Šiuo atveju naudingumo funkcija yra iškili, t.y. racionaliam išrinkimui būdingas polinkis į riziką.

Bendru atveju naudingumo funkcija turi S raidės pavidalą. Tačiau ji nėra simetriška neutralaus taško atžvilgiu. Panagrinėkime paprastą pavyzdį. Pasiūlymo dalyvauti loterijoje, kurioje tikimybė, kad bus išlošta 100 $ ir tikimybė, kad bus pralošta 100 $, lygios 0.5, atsisakė beveik visi eksperimento dalyviai. Dalyvavimas loterijoje būtų psichologiškai priimtinas tuomet, kai esant vienodoms alternatyvų tikimybėms nuostoliai būtų gerokai mažesni už pelną. Nuostolių ir pelno atveju naudingumo funkcija pavaizduota 5.18 paveiksle. Abi šakas galima aproksimuoti eksponentinėmis funkcijomis, kurių koeficientai skirtingi.

Išrinkimui būdingas polinkis į riziką arba jos vengimas priklauso nuo neutralaus taško padėties. Tačiau nuostolių ir pelno sąvokos kartais subjektyvios, todėl reikia atidžiai išanalizuoti alternatyvų pasekmes ir pagrįsti neutralaus taško padėtį. Literatūroje aprašytas psichologinis eksperimentas, kuriame dalyvavusiems gydytojams buvo užduodamas toks klausimas: *JAV laukiama retos Azijoje prasidėjusios ligos epidemija. Jei nebus imtasi priemonių, tai vertinant pagal kitų valstybių duomenis gali mirti 600 žmonių. Atlikti tyrimai rekomenduoja du kovos su šia liga būdus. Panaudojus būdą A, bus išgelbėta 200 žmonių. Panaudojus būdą B, tikimybė, kad bus išgelbėta 600 žmonių, lygi 1/3, bet tikimybė, kad nebus niekas niekas išgelbėtas, lygi 2/3. Kurį būdą jūs rekomenduotumėte?* Dauguma rekomendavo vengti

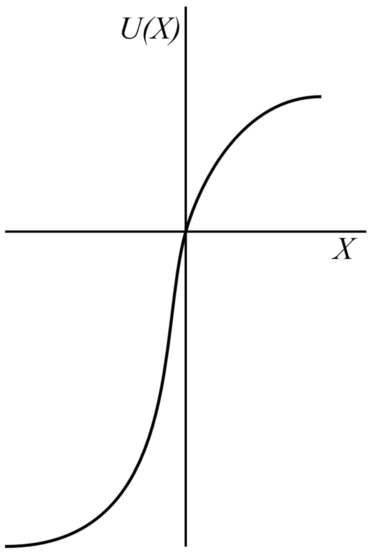

Figure 5.18: Naudingumo funkcija nesimetriška neutralaus taško atžvilgiu

rizikos ir pasirinkti pirmąjį būdą. Iš uždavinio formuluotės išplaukia, kad neutralus taškas atitinka 600 žmonių netektį. Vadinasi, kiekvienas išgelbėtasis - tai pelnas. Taigi gydytojų rekomendacija atitinka teoriną.

Kitiems gydytojams buvo pateiktas šitoks klausimas: *Ligai gydyti pasiūlyti du būdai - C ir D. kitų valstybių duomenimis, jei bus panaudotas C būdas, mirs 400 žmonių, jei D, tai tikimybė, kad niekas nemirs, lygi 1/3, o tikimybė, kad mirs 600 žmonių, lygi 2/3. Kurį būdą jūs rekomenduotumėte?*

Dauguma rekomendavo antrąjį rizikingą būdą. Matyt, uždavinys suformuluotas taip, kad susidaro neteisingas įspūdis, jog neutralus taškas atitinka nulį. Vadinasi, kiekviena mirtis yra nuostoliai, taigi gydytojų rekomendacija vėl atitinka teoriją. Abiem formuluotėmis, tik kitais žodžiais, nusakoma ta pati objektyvi situacija. Tačiau antrąja formuluote tiesiai nepasakoma, kas įvyktų, jei nebūtų panaudotas joks gydymo būdas. Tai vienintelis, bet esminis abiejų formuluočių skirtumas.

Daugelyje gyvenimiškų situacijų žmonės, išrinkdami vieną arba kitą alternatyvą, vadovaujasi ne vien racionalumu, bet ir emocijomis. Kai kuriais atvejais pasirinkimas būna toks, kad aplinkiniai jį tegali įvertinti kaip neapgalvotą. Šioje knygoje, neturėdami galimybės plačiau nagrinėti psichologinių klausimų, pažymėsime tik keletą įdomesnių momentų. 300 litų premija už gerą darbą gali žmogų nuliūdinti ir netgi įžeisti, jei jis tikėjosi gauti žymiai daugiau. Tačiau net ir pusė šios sumos gali sukelti didelį džiaugsmą, jei ji yra "išmušta" beveik neįtikėtinomis sąlygomis. Ta pati 300 litų premija bus priimama vienaip, jei ji didžiausia iš gautų kolektyve, ir visai kitaip, jei ji mažiausia. Todėl įdomu, kaip nusivylimo koncepciją reikėtų suderinti su išdėstytu racionalaus išrinkimo modeliu. O štai dar vienas įdomus klausimas: ar atsiperka analizė, be kurios neįmanomas racionalus išrinkimas. Jį galima suformuluoti ir kitaip: kiek kainuoja galvojimas?

Sakykime, kad projektuotojas, dirbdamas savo darbą, stengiasi būti racionalus. Tuomet išrenkant variantus neapibrėžtumo sąlygomis reikia remtis vidutinio naudingumo įvertinimu. Vidutinis naudingumas apibrėžiamas naudingumo funkcija ir tikimybiniu skirstiniu. Kad būtų galima išreikšti individo naudingumo funkciją, pirmiausia reikia nustatyti jos formą, t.y., ar ji atitinka rizikos vengimą, ar polinkį į riziką. Sakysim, nagrinėjamas tik vienas kriterijus. Jei garantuotas variantas $(x_1 + x_2)/\ 2$ dominuoja loteriją, kurioje su tikimybėmis 0.5 realizuojamos kriterijaus reikšmės x_1 ir x_2, tai naudingumo funkcija gali būti aproksimuota eksponentine

$$U(x) = a + b\, e^{-cx},$$

čia konstantos a, b, c $(b < 0, c > 0)$ įvertinamos remiantis ekspertine informacija. Pavyzdžiui, jei variantas x_3 ekvivalentus loterijai, kurioje su tikimybėmis 0.5 realizuodami x_1 ir x_2, tai naudojantis lygybe

$$U(x_3) = 0.5\, U(x_1) + 0.5\, U(x_2)$$

galima apskaičiuoti konstantą c. Praktiškai stengiamasi gauti kuo daugiau ekspertinės informacijos ir po to konstantas įvertinti naudojantis bendrais metodais, pavyzdžiui, mažiausiųjų kvadratų metodu.

Dažnai naudingumas (su pakankamu tikslumu) yra tiesinė kriterijaus x funkcija. Tuomet naudingumą galime sutapatinti su x. Kriterijų, kai jų yra ne vienas, kompozicijos problema nagrinėjama kitame skyriuje.

Analogiškai įvertinami subjektyvūs tikimybiniai skirstiniai. Reikia pažymėti, kad žmonės subjektyviai vertindami tikimybes pagal statistinių stebėjimų duomenis linkę pervertinti labai mažas tikimybes ir sumažinti artimų vienetui tikimybių įverčius. Visi mažų tikimybių įvykiai traktuojami kaip beveik negalimi, tačiau kokybiškai besiskiriantys nuo garantuotai neįmanomo įvykio. Panašiai vertinami ir beveik garantuotai įvyksiantys įvykiai. Artimų nuliui tikimybių įvertinimas yra labai sudėtinga problema taip pat ir klasikinėje statistikoje. Išanalizavus stambių avarijų priežastis, paaiškėja, kad avarijos įvyksta būtent tuomet, kai sutampa praktiškai neįtikėtini įvykiai, kurių statistikos surinkti neįmanoma. Subjektyviais tokių tikimybių įverčiais pasikliauti taip pat negalima. Todėl, projektuojant unikalius objektus, kurių (nors ir labai mažai tikėtini) sutrikimai gali sukelti katastrofiškas pasekmes, variantas išrinkamas remiantis garantuoto rezultato principu.

5.11 Stochastinė optimizacija

Remiantis racionalaus išrinkimo teorija, projektiniai variantai lyginami pagal vidutinio naudingumo kriterijų. Jei tokių variantų yra tik keletas, tai visų jų vidutinį naudingumą įmanoma apskaičiuoti net ir tuomet, kai neapibrėžtų faktorių daug. Bet kaip pasinaudoti teorijos išvadomis, jei konstruktyvių parametrų vektorius X yra daugiamatis ir priklauso sudėtingos struktūros aibei A, o nevaldomų faktorių vektorius Y taip pat yra daugiamatis? Funkciją $u(X)$ maksimizuoti ankstesniuose knygos skyriuose nagrinėtais metodais dažniausiai nepavyktų. Sunkumai kyla todėl, kad daugiamačiam integralui (5.1) apskaičiuoti reikia daug kompiuterio laiko, nes kitaip $u(X)$ reikšmės būtų gaunamos su didelėmis paklaidomis. Todėl (5.1) tipo integralams maksimizuoti sukurti specialūs metodai naudojantys tik pointegrinės funkcijos reikšmes. Pavyzdžiui, sudaroma atsitiktinė trajektorija, kurios žingsniai vidutiniškai nukreipti $U(X, Y)$ didėjimo kryptimi. Jei žingsnio ilgis ir atsitiktinė judesio kryptis patenkina tam tikras sąlygas, tai atsitiktinė trajektorija konverguoja (tikimybine prasme) į funkcijos $u(X)$ lokalaus maksimumo tašką.

Yra ir kitų idėjų, kuriomis pagrindžiami metodai, iš esmės besiskiriantys nuo ką tik aprašytojo. Reikia pažymėti, kad stochastinės optimizacijos uždavinius spręsti nėra lengva net ir turint tam skirtus programinės įrangos paketus.

5.12 Užduotys ir kontroliniai klausimai

1. Ar pagrįstai postuluojama, kad preferencijos santykis yra tranzityvus? Sukonstruokite pavyzdį su netranzityviu preferencijos santykiu ir jį išanalizuokite.

2. Kokiomis savybėmis pasižymi naudingumo funkcija?

3. Kokius žinote neapibrėžtumo modelius? Kokiu savo būsimų pajamų neapibrėžtumo modeliu remtumėtės, jei reikėtų apsispręsti dėl skolinimosi iš banko buto pirkimui?

4. Kokioms aplinkybėms esant sprendimą priimtumėt vadovaudamiesi vidutinio naudingumo principu, ir kokioms - minimakso principu?

5. Kokia naudingumo funkcija atitinka sprendimus su polinkiu į riziką ir kokia atitinka sprendimus vengiančius rizikos?

6. Paaiškinkite neutralaus rizikos funkcijos taško prasmę.

DAUGIAKRITERIS OPTIMIZAVIMAS

6.1 Intuityvusis išrinkimas

Daugelio kriterijų uždaviniai iškyla įvairiose sprendimų priėmimo situacijose. Panagrinėkime projektavimą. Kiekvienas projektavimo etapas baigiamas vieno arba kelių variantų išrinkimu tolesniam darbui. Pavyzdžiui, ekspertams pateikiami penki galimi preliminarūs siuvimo cecho varijantai X_i, $i = 1, ..., 5$, besiskiriantys kriterijų vektoriaus $F(X) = (f_1(X), ..., f_5(X))$ reikšmėmis. Jas palyginę, ekspertai savo išvadas dažnai formuoja, pavyzdžiui, šitaip: varijantas X_1 nepriimtinas dėl per didelės kriterijaus f_1 (kainos) reikšmės, o X_2 - dėl per mažo f_2 (patikimumo). Variantas X_3 atmetamas, nes beveik pagal visus kriterijus blogesnis už X_1 , tik pasižymi mažu f_5 (triukšmo lygiu). Rekomenduojama dirbti su variantais X_1 ir X_5 , tačiau reikia sumažinti varianto X_5 triukšmo lygį ir pagerinti kitus sanitarinius kriterijus bei sumažinti varijanto X_1 kainą.

Be abejo, ekspertai, visapusiškai išnagrinėję kelis variantus, priims racionalų sprendimą. Tačiau ar tie keli variantai atspindi visas galimybes? Kita vertus, žmogus gali neaprėpti visų galimų variantų, jei jų yra keli šimtai, ypač tuomet, kai kiekvienas varijantas apibūdinamas dešimtimis kriterijų. Euristiniais metodais vargu ar pavyktų ištirti didelę variantų aibę ir, atmetus aiškiai nepriimtinus, išrinkti ekspertizei tam tikrą potencialiai geriausių variantų skaičių. Todėl ekspertai efektyviai gali dirbti tik naudodami matematinius metodus. Minėjome, kad racionalaus išrinkimo uždavinį teoriškai galima pakeisti naudingumo funkcijos optimizavimo uždaviniu. Šiame skyriuje panagrinėsime, kaip tą padaryti praktiškai. Jei sudaryti naudingumo funkciją dėl įvairių priežasčių sunku, tai galime pasinaudoti alternatyviais metodais, kuriuos taip pat aptarsime.

6.2 Apibrėžtys

Projekuotojo keičiamų parametrų vektorius $X = (x_1, ..., x_n)$ parenkamas tam tikroje srityje $A \in R^n$, vadinamojoje leistinąja sritimi arba leistinąja aibe. Pavyzdžiui, kai kuriems kintamiesiems apibrėžiamos viršutinės, kai kuriems - apatinės, kitiems - abi ribos. Plačiau leistinosios srities sąvoka aptarta skyrelyje, skirtame optimizavimo uždavinių formulavimui. Pavyzdžiui, leistinoji sritis 6.19 paveiksle apibrėžta trimis paprastaisiais, vienu netiesiniu ir dviem tiesiniais ribojimais.

Kiekvienam vektoriui X apibrėžtas kriterijų vektorius $(f_1, ..., f_m) = F(X)$. Vektoriui X kintant aibėje A vektorius F perbėga kriterijų reikšmių aibę Φ. Jei kriterijus būtų tik vienas (tolydinis X atžvilgiu), tai Φ sutaptų su intervalu, kurio pradžią ir galą nustato kriterijaus funkcijos (vieno kriterijaus uždaviniuose vadinamos tikslo funkcija) minimumas ir maksimumas. Jei yra daug kriterijų, tai jų maksimumų taškai paprastai nesutampa. Pavyzdžiui,

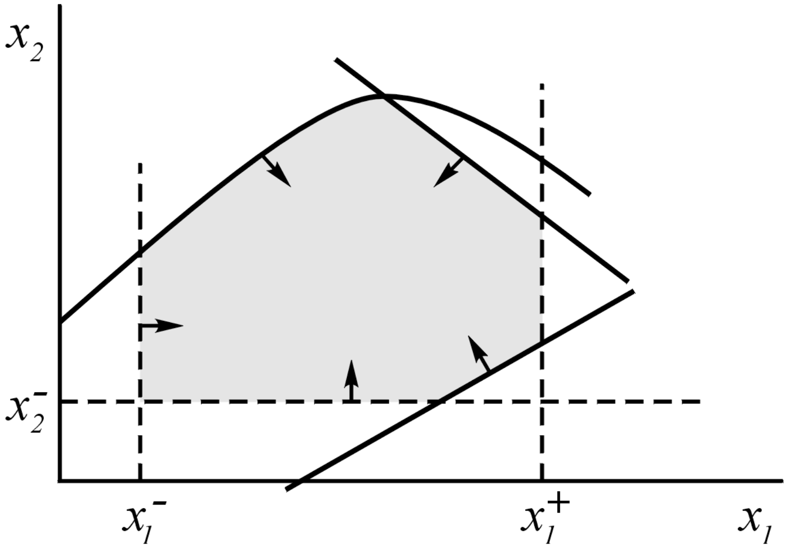

Figure 6.19: Leistinosios aibės pavyzdys

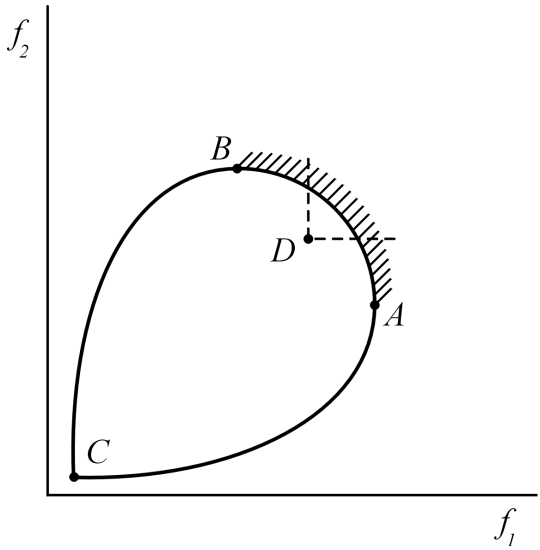

Figure 6.20: Kriterijų reikšmių aibės pavyzdys

kriterijų reikšmių aibėje, pavaizduotoje 6.20 paveiksle, pirmojo kriterijaus maksimumo taškas A nesutampa su antrojo kriterijaus maksimumo tašku B. Tačiau sutampa abiejų kriterijų minimumai taške C. Jei spręsdami šį uždavinį surastume bet kurio kriterijaus minimimą, tuomet išspręstume ir dvikriterį minimizmo uždavinį.

Daugiakriterių uždavinių teorijoje paprastai nagrinėjamas visų kriterijų **maksimizavimas**. Tokiu atveju nelygybė tarp kriterinių įverčių vektorių atrodo panašiai kaip preferencijos santykio ženklas. Toliau laikysimės šios tradicijos, nors vieno kriterijaus optimizavimo uždaviniuose tikslo funkciją minimizavome. 6.20 paveiksle iliustruojamas dvikriterio maksimizavimo uždavinys gali atrodyti prieštaringas. Jei kaip sprendinį imsime tašką A, tai gerokai nukrypsime nuo maksimalios antrojo kriterijaus riekšmės, ir atvikščiai. Todėl reikia ieškoti kompromisinio varijanto. Kita vertus, D taip pat negali būti sprendiniu, nes visi taškai pažymėtame sektoriuje yra geresni uš tašką D abiejų kriterijų atšvilgiu. Taškai, iš kurių negalima pajudėti pagerinant abu kriterijus, sudaro aibės Φ šiaurės - rytų ribą, kuri pavyzdžiui užbrūkšniuota 6.20 paveiksle. Toks Φ poaibis vadinamas **Pareto arba kompromisų aibe**. Jei abu kriterijai lygiaverčiai, tai nė vieno iš šio poaibio taškų negalime išskirti, nes kiekvieną iš jų vienodai tinkamas kaip kompromisinis variantas. Todėl bendru atveju nepagerinamą kriterijų vektorių galime apibrėšti šitaip: F nepagerinamas, jei kūgio, sudaryto iš pustiesių išeinančių iš taško F koordinačių kryptimis, sankirta su aibe Φ susideda iš veinintelio taško F. Pavyzdžiui, 6.20 paveiksle kiekvienas užbrūkšniuotas kreivės taškas nepagerinamas.

Projektuotoją domina, kaip surasti variantą, atitinkantį nepagerinamą kriterijų vektorių. Taškas X^* vadinamas **efektyviuoju sprendiniu** (optimaliu Pareto prasme), jei nelygybė $F(X) \geq F(X^*)$ ekvivalenti lygybei $X = X^*$, t.y. neegzistuoja toks $X \in A$, kuriam $f_i(X) \geq f_i(X^*)$, $i = 1, ..., m$ ir bent vienam j, $f_j(X) > f_j(X^*)$. Efektyvaus sprendinio kriterijų vektorius nepagerinamas, ir atvirkščiai: kiekvieną nepagerinamą kriterijų (Pareto) vektorių atitinka efektyvus sprendinys.

6.3 Kriterijų kompozicija

Praeitame skyriuje buvo parodyta, kad racionalus išrinkimas gali būti pakeistas naudingumo funkcijos maksimizavimu. Tačiau ir nežinant šio rezultato atrodo gan natūralu pakeisti daugiakriterį uždavinį vienakriteriniu, sudarytu susumavus su svoriais visus uždavinio kriterijus

$$f(X) = \sum_{i=1}^{m} a_i\ f_i(X),\ \sum_{i=1}^{m} a_i = 1,\ a_i \geq 0. \tag{6.1}$$

Kokiu pagrindu gautąją funkciją $f(X)$ galima laikyti pradindinio uždavinio naudingumo funkcja? Ar teoriškai pagrįstas daugiakriterio uždavinio keitimas funkcijos $f(X)$ maksimizavimo uždaviniu?

Paprastai, preferencijos santykį (psichologiniu) eksperimentu pavyksta nustatyti tik nedideliam alternatyvų skaičiui. Todėl galima tikėtis parinkti svorius (6.1) taip, kad funkcijos $f(X)$ reikšmių santykiai atitiks eksperimentinius duomenis. Kita vertus, kartais kriterijų svorius galima nustatyti atsižvelgiant į kriterijų svarbą. Panagrinėsime svertinės kriterijų sumos savybes:

16 teorema. Jei aibė A yra iškili, funkcijos $f_i(X)$- įgaubtos (iškilos į viršų) ir X^0- efektyvaus sprendinio taškas, tai egzistuoja tokie $a_i^0 \geq 0$, $\sum_{i=1}^{m} a_i^0 = 1$, kad

$$\max_{X \in A} \sum_{i=1}^{m} a_i^0 \ f_i(X) = \sum_{i=1}^{m} a_i^0 \ f_i(X^0). \tag{6.2}$$

Šis S.Karlin'o rezultatas teoriškai pagrindžia svertinės svorių kompozicijos taikymą.

17 teorema. Jei $a_i \geq 0$, $i = 1, ..., m$, bent vienas $a_j > 0$ ir

$$\sum_{i=1}^{m} a_i \ f_i(X^*) = \max_{X \in A} \sum_{i=1}^{m} a_i \ f_i(X) \tag{6.3}$$

tai X^* yra efektyvus sprendinys.

Iš šių teoremų matome, kad maksimizuojant svertinę kriterijų sumą visada gaunamas efektyvusis sprendinys. Atitinkamai parinkdami svorius, galime gauti bet kurį efektyvųjį sprendinį. Gali susidaryti įspūdis, kad naudodamiesi šia kriterijų kompozicija galėsime išspręsti visas daugiakriterio optimizavimo problemas. Tačiau iškilumo riekalavimas Karlin'o teoremoje susiaurina jos pritaikymo sritį, nes kriterijų iškilumas projektavimo uždaviniuose gali būti abejotinas ir tik atskirais atvejais įrodomas. Jei iškilumo sąlyga ir priimtina, tai dar lieka neaišku, kokia yra priklausomybė tarp parenkamų svorių ir juos atitinkančių Pareto aibės taškų. Jei apie kriterijus turime papildomos informcijos, tai kaip ją įvertinti parenkant svorius? Pavyzdžiui, reikia gauti N tolygiai pasiskirsčiusių Pareto aibėje taškų. Kaip parinkti svorius? Parenkant svorius asitiktinai su tolygiu pasiskirstymu, juos atitinką kriterijų vektoriai gali būti pasiskirstę Pareto aibėje labai netolygiai. (6.21) paveiksle pavaizduotas gana tipiškas pavyzdys: atskiruose Pareto aibės poaibiuose taškai išsidėstę tankiai, o kai kur jų visai nėra. Todėl lieka neaišku, kokia yra Pareto aibės forma. Naudodamiesi turimais duomenimis negalime atsakyti, ar egzistuoja tokie kriterijų vektoriai, kad $f_i > f_i^*$. Ekstrapoliuodami taip, kaip parodyta paveiksle brūkšnine linija, galėtume tikėtis, kad tinkamai parinkus svorius, gausime geresnį už (f_1^*, f_2^*) kriterijų vektorių. Tačiau gali pasirodyti, kad tikroji riba yra tokia, kokią vaizduoja ištisinė linija, ir nėra vektoriaus su komponenčių reišmėmis f_i^*.

Svertinė kriterijų kompozicija nėra panacėja, bet ji dažnai naudojama kuriant daugiakriterio optimizavimo metodus. Bendras kriterijus, gautas atlikus svertinės kriterijų sumos kompoziciją, pasižymi tuo, kad sumažėjus vienam kriterijui, tai kompensuojama, padidinus kitą. Kitu atveju, tai gali būti lyg ir atskirų pelno komponenčių suma: svarbu maksimizuoti bendrą pelną ir visai nesvarbu, kurios komponentės sąskaita jis gaunamas.

Esti uždavinių, kuriuose variantas nepriimtinas, jei bent vieno kriterijaus reikšmė lygi "nuliui". Čia nulis parašytas kabutėse todėl, kad natūraliame mastelyje nepriimtina kriterijaus reikšmė gali ir nebūti lygi nuliui. Tačiau, pakeitus skalės ataskaitos tašką, šią reikšmę galima laikyti nuline. Antra vertus, net ir labai geros kelių kriterijų reikšmės gali būti nuvertintos, kai vienas iš kriterijų artimas nuliui. Šiuo atveju variantas priimtinas tik tuomet, kai visų kriterijų reikšmės bent jau "neblogos" (anksčiau aprašytu svertinės sumos atveju variantas galėtų būti priimtinas ir tuomet, kai vieno kriterijaus reikšmė "labai bloga", t.y. nulinė). Tokias savybes turi kompozicija, išreiškiama kriterijų

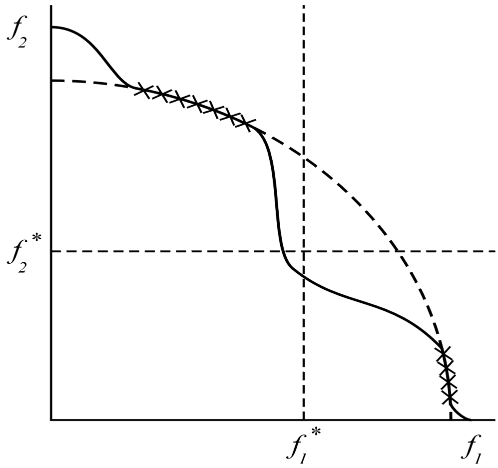

Figure 6.21: Optimizuojant svertinę svorių kompoziciją, taškai Pareto aibėje išsidėsto netolygiai

sandauga:

$$f(X) = f_1(X) \cdot f_2(X) \cdot \ldots \cdot f_m(X). \tag{6.4}$$

Jei funkciją $f(X)$ logaritmuosime, tai gausime atskirų kriterijų logaritmų sumą. Todėl šios kompozicijos matematinės savybės yra analogiškos svertinės sumos savybėms.

Pateiksime dar vieną kriterijų kompzicijos pavyzdį

$$f(X) = \min_i a_i \cdot f_i(X), \tag{6.5}$$

kuria išreiškiamas "silpniausios grandies stiprinimo" principas. Juo dažnai naudojasi projektuotojai. Šiuo atveju, net nereikalaujant kriterijų iškilumo prielaidos, galima įrodyti teoremą, panašią į Karlin'o teoremą.

18 teorema. Jei X^* yra efektyvaus sprendinio taškas ir $f_i(X^*) > 0$ visiems $1 \le i \le m$, tai egzistuoja tokie $a_i^* > 0$, $\sum_{i=1}^m a_i^* = 1$, kad

$$\max_{X \in A} \min_i a_i^* \cdot f_i(X) = \min_i a_i^* \cdot f_i(X^*). \tag{6.6}$$

Jei X nepriklauso efetyviųjų sprendinių aibei, tai esant bet kokiems $a_i > 0$, $\sum_{i=1}^m a_i = 1$, taške X funkcijos $\min_i a_i \cdot f_i(X)$ maksimumas nepasiekiamas.

Teoriniu požiūriu šis (J.Germejerio) rezultatas yra labai bendras. Tačiau praktiškai parinkti svorius (6.6) kompozicijai gali būti gana sudėtinga.

Kartais, sprendžiant daugiakriterį uždavinį, formuluojamas idealus tikslas, t. y. kriterijų vektorius F^0, kurio gal ir neįmanoma pasiekti, bet į kurį reikia orientuotis. Pavyzdžiui, žinome, kad gaminys, kurio parametrai F^0, rinkoje turi paklausą. Norint jį nukonkuruoti, reikia pasiūlyti gaminį, kurio parametrai būtų beveik tokie patys, bet jo gamybai reikėtų mažiau sąnaudų. Taigi, siekiame priartėti prie užsibrėžto tikslo. Todėl toks daugiakriterio uždavinio pakeitimas vienakriteriniu vadinamas tikslo programavimu. Šis terminas yra vertinys iš anglų kalbos **goal programming**. Tikslo programavimo uždavinio formulavimui be tikslo vektoriaus F^0 reikia dar apibrėšti atstumo tarp kriterijų vektorių funkciją $r(F^0, F)$. Tuomet daugiakriterį uždavinį galima pakeisti minimizavimo uždaviniu

$$\min_{X \in A} r(F^0, F(X)).$$

Kadangi atstumo funkcija dažniausiai apibrėžiama šitaip :

$$\begin{aligned} r_1(F,G) &= \sum_{i=1}^{m} c_i|f_i - g_i|, \\ r_2(F,G) &= \sum_{i=1}^{m} c_i(f_i - g_i)^2, \end{aligned}$$

tai ir vėl kyla klausimas: kaip parinkti atskirų kriterijų svorius? Dažnai neaišku, kurią atstumo formulę pasirinkti. Be to, tikslą ne visada galima apibrėžti vienareikšmiškai.

Kriterijų kompozicijos optimizavimas pagrindžiamas įvairiais euristiniais samprotavimais ir matematinėmis kompozicijos funkcijų savybėmis. Viena iš pagrindinių šio metodo problemų – kaip informaciją apie kriterijus išreikšti jų svoriais. Apibrėžus kompozicijos funkciją, uždavinys sprendžiamas įprastais optimizavimo metodais.

6.4 Nuosekli kriterijų analizė

Administracinių sprendimų, kurie priimami esant sudėtingoms situacijoms (kai yra daug kriterijų ir tikslų) tyrimai parodė, kad žmonės alternatyvas išrenka nuosekliai analizuodami atskirus kriterijus. Pirmiausia atmetamos alternatyvos, netinkančios pagal svarbiausiąjį kriterijų, po to - pagal antrąjį ir tt. Tinkamiausi variantai išrenkami nuosekliai atmetant netinkamus atskirų kriterijų požiūriu.

Panašūs rezultatai buvo gauti atliekant psichologinius eksperimentus. Pavyzdžiui, dalyviai turėjo išrinkti iš daugelio daiktų vieną, reikalingą eksperimento vedėjui. Spręsdamas šį uždavinį, eksperimento dalyvis užduodavo vedėjui klausimus apie reikalingo daikto požymius. Daugelis klausimų buvo formuojami taip, kad būtų galima atmesti visus daiktus, neturinčius reikalingo daikto požymio.

Ranžuojant (surikiuojant) kriterijus, gauname **leksikografinės optimizacijos uždavinį**: variantas X_1 geresnis už X_2 , jei $f_1(X_1) > f_1(X_2)$. Jei $f_1(X_1) = f_1(X_2)$, tai lyginama pagal antrąjį kriterijų ir t. t. Naudojant leksikografinį išrinkimą, kai kuriuose čempionatuose išaiškinami čempionai, pavyzdžiui:

- čempionu tampa komanda, surinkusi maksimalų taškų skaičių (pregalė - 1 taškas, lygiosios - 1/2 taško, pralaimėjimas - 0 taškų);

- jei kelios komandos surenka vienodą taškų skaičių, tai čempionu tampa ta komanda, kuri surinko daugiausia taškų tarpusavio susitikimuose;

- jei pagal pirmuosius du kriterijus kelios komandos yra lygiavertės, tai čempionu tampa ta, kurios įmuštų ir praleistų įvarčių santykis jų tarpusavio rungtynėse geriausias;

- jei kelios komandos lyiavertės pagal tris ankstesnius krtiterijus, tai čempionas išaiškinamas atsižvelgiant į įvarčių santykį visose čempionato rungtynėse.

Kad kelios komandos būtų lygiavertės pagal kelis kriterijus nors ir mažai tikėtina, bet pasitaiko. Pavyzdžiui, 1969 m. ledo ritulio pasaulio čempionate trys komandos (TSRS, Čekoslovakija ir Švedija) surinko veinodą taškų skaičių apskritai ir tarpusvio rungtynėse. TSRS komanda tapo čempione tik palyginus komandas pagal trečiąjį kriterijų. Dabar dvi iš šių trijų valstybių nebeegzistuoja, o geriausi įvairių tautybių žaidėjai žaidžia Šiaurės Amerikoje. Klubų vadovų vykdomas žaidėjų pasirinkimas galėtų iliustruoti kitą daugiakriterio išrinkimo uždavinį, bet jo formulavimą paliksime skaitytojui.

Leksikografinis optimizavimo uždavinys gali būti sprendžiamas nuosekliai maksimizuojant kriterijus siaurėjančiuose leistinos aibės poaibiuose $A_i \subset A$:

pirmasis poaibis:

$$A_1 = \{X : X \in A,\ f_1(X) = f_1^*\},\ \max_{X \in A} f_1(X) = f_1^*,$$

antrasis poaibis:

$$A_2 = \{X : X \in A_1, \ f_2(X) = f_2^*\}, \ \max_{X \in A_1} f_2(X) = f_2^*,$$

m-asis poaibis:

$$A_m = \{X : X \in A, \ f_m(X) = f_m^*\}, \ \max_{X \in A_{m-1}} f_m(X) = f_m^*.$$

$m-$ojo maksimizavimo uždavinio maksimumo taškas ir bus sprendiniu. Jei kuriame nors žingsnyje aibę sudaro veinintelis taškas, tai jis ir yra leksikografinio ušdavinio sprendinys. Leksikogrfiniai uždaviniai, matyt, natūraliausi tuomet, kai leistinoji (arba kriterijų reikšmių) aibė yra diskreti, nes jei A kontinuali, o $F(X)$ tolydinė, tai sudaryti poaibius A_i sudėtinga.

Aukščiau nurodytas būdas, skirtas leksikografiniam uždaviniui spręsti, susideda iš m etapų, kuriuose turi būti surastas funkcijos maksimumas aibėje, apibrėžtoje lygybiniu ribojimu. Leksikografiniai uždaviniai su begalinėmis kriterijų reikšmių aibėmis įdomūs tuo, kad ne visada egzistuoja naudingumo funkcija, atitinkanti leksikografinį preferencijos santykį. Todėl leksikografinis uždavinys ne visada gali būti pakeistas vienu optimizavimo (naudingumo funkcijos maksimizavimo) uždaviniu. Taigi, naudingumo teorijos prielaidos nėra universalios, o leksikografinis uždavinys iliustruoja atvejį, kai gali neegzistuoti uždavinį atitinkanti kriterijų kompozicija.

Praktiniuose uždaviniuose krterijų ranžavimas, ypač svarbiausiojo išrinkimas, dažnai natūraliai išplaukia iš paties uždavinio. Pavyzdžiui, kaina masinės produkcijos prekėms arba masė aviaciniams įrenginiams dažnai yra svarbesni už kitus kriterijus. Tačiau toks griežtas kriterijų ranžavimas (surikiavimas), kaip leksikografinės optimizacijos atveju, ne visada reikalingas. Surikiuotiems kriterijams lieka tam tikra tolerancija, t.y. gali būti pasirinktas variantas, kuris nežymiai blogesnis už kitą pagal svarbiausiąjį kriterijų, nes jis žymiai geresnis pagal antrąjį kriterijų.

Parinkę pagrindiniu maksimizuojamu kriterijumi, pavyzdžiui $f_1(X)$, o kitus apibrėžę kaip ribojimus, gausime įprastą matematinio programavimo uždavinį

$$\max_{X\in B} f_1(X),\ B = \{X : X \in A,\ f_i(X) \geq \phi_i,\ i = 2, ..., m\},$$

čia ribojimų lygiai ϕ_i gali būti apibrėžiami ptojektavimo tikslais, žinomų prototipų parametrais ir kt.

Nesant natūralių, uždavinio sąlygomis reglamentuojamų ribojimų lygių, juos parinkti gali būti sunku. Jei nors vienas ribojimų lygis bus per aukštas ($\phi_i > \max_{X\in A} f_i(X)$, gausime tuščią aibę B. Kita vertus, ϕ_i negali būti žemesnis negu i-ojo kriterijaus priimtina riba. Ypaš sunku parinkti ϕ_i tuomet, kai kriterijai prieštaringi, pavyzdšiui, kaina ir patikimumas. Kartais lygių parinkimą galima pagrįsti analizuojant atskirų kriterijų maksimumus ir "vidutinių" vektorių F komponenčių reikšmes. Sprendžiant daugiakriterį uždvinį šiuo metodu, tenka daug kartų spręsti optimizvimo ušdavinius, besiskiriančius tik ribojimų nelygybių dešniosiomis pusėmis.

Leksikografinio uždavinio sprendimo ir daugiakriterio uždavinio pakeitimo įprastu matematinio programavimo uždaviniu metodai gali būti tam tikra prasme apjungti į vieną. Daugelyje praktinių uždavinių kriterijai gali būti surikiuoti (nors ir ne taip griežtai kaip leksikografiniu atveju). Sakykime, kad kriterijus tuo svarbesnis, kuo mažesnis jo numeris. Kartais lengviau suformuluoti ne ribojimus kriterijų reikšmėms, o nuolaidas d_i, kurios leidžiamas nuo maksimalių kriterijų reikšmių, siekiant pagerinti mažiau svarbių kriterijų reikšmes. Tuomet apskaičiuotas maksimalias svarbiausiųjų kriterijų reikšmes galima naudoti leistinajai sričiai apibrėžti žemesniojo rango kriterijų maksimizavimo uždaviniuose. Taikant tokį metodą, reikia nuosekliai išspręsti šiuos optimizavimo uždavinius:

$$
\begin{aligned}
f_1^* &= \max_{X \in A} f_1(X), \\
A_1 &= \{X : X \in A,\ f_1(X) \geq f_1^* - d_1\}, \\
f_2^* &= \max_{X \in A_1} f_2(X), \\
A_2 &= \{X : X \in A_1,\ f_2(X) \geq f_2^* - d_2\}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
f_m^* &= \max_{X \in A_{m-1}} f_m(X).
\end{aligned}
$$

Daugiakriterio uždadvinio sprendiniu yra m-ojo maksimizavimo uždavinio maksimumo taškas. Prieš spendžiant uždavinį nustatyti nuolaidų dydžius yra nelengva. Jei visi $d_i = 0$, gaunamas leksikografinis uždavinys. Jei d_i yra tokie dideli, kad papildomi ribojimai nebesusiaurina leistinos aibės, tai gautasis sprendinys sutampa su žemiausiojo rango kriterijaus maksimumo tašku. Sprendžiant uždavinius nuolaidų metodu, tenka eksperimentuoti su įvairiais d_i nuolat konsultuojantis su ekspertais. Dažnai tam padeda grafikai, pavyzdžiui, optimalios antrojo kriterijaus maksimalios reikšmės priklausomybė nuo pirmojo kriterijaus nuolaidos pavaizduota paveiksle (6.22):

$$f_2^* = f_2^*(d_1).$$

Suprantama, kad norint sudaryti tokius grafikus, reikia daug kartų spręsti maksimizavimo uždavinius. Nuosekliųjų nuolaidų metodas lengvai suprantamas ir mėgstamas vartotojų. Jį teoriškai pagrindžia tokia teorema:

19 teorema. Kiekvienam kriterijų vektoriui F^*, priklausančiam Pareto aibei, galima parinkti tokias nuolaidas d_i, kad nuosekliųjų nuolaidų metodu gautas sprendinys $(f_1^*, f_2^*, ..., f_m^*)$ sutaps su F^*.

Bendru atveju nuosekliųjų nuolaidų metodu gautas sprendinys gali ir nepriklausyti efektyviųjų sprendinių aibei. Tačiau galima įrodyti,

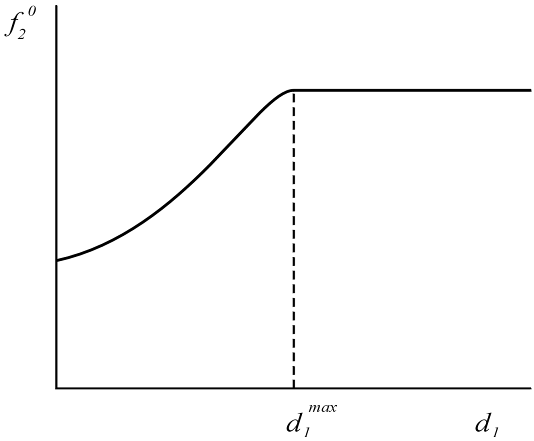

Figure 6.22: Antrojo kriterijaus maksimalios reikšmės priklausomybė nuo pirmojo kriterijaus nuolaidos

kad jei A yra aprėžta uždara aibė, o visos kriterijų funkcijos - tolydinės, tai nuoseklių̨jų nuolaidų metodu gaunamų sprendinių aibėje yra bent vienas efektyvus sprendinys. Nuoseklių̨jų nuolaidų metodą geriausia naudoti tuomet, kai sprendžiant daugiakriterius uždavinius kriterijai yra natūraliai suranžuoti pagal svarbą taip, kad analizuojant i-ąjį kriterijų reikia atsižvelgti į $i+1$-ąjį, bet nebūtinai į $i+2$-ąjį.

6.5 Pareto aibės aproksimacija

Aukščiau išdėstyti daugiakriterės optimizacijos teorijos faktai rodo, kad vienintelė sprendinio ypatybė, tiesiogiai išplaukianti iš daugiakriterio uždavinio formulavimo, yra ta, kad sprendinys priklauso Pareto aibei. Jei apie kriterijus turime papildomos informacijos, tai galima išskirti arba Pareto aibės dalį, arba rasti kurį nors vieną jos tašką. Pavyzdžiui, priskyrus kriterijams svorius ir suradus jų svertinės sumos maksimumą, gaunamas vienas iš Pareto aibės taškų. Dažnai asmenys, sprendžiantys daugiakriterius uždavinius, pvz. prjektuotojai, sutinka, kad kriterijams galima racionaliai priskirti svorius, tačiau praktiškai jie tą daro nenoriai ir tik su didelėmis išlygomis. Panašūs sunkumai kyla ir naudojant kitus kriterijų palyginimo metodus. Žmonių turimai informacijai išgauti reikalingi specialūs metodai, įvertinantys psichologinius faktorius. Kita vertus, reikia padėti uždavinio sprendėjui geriau susivokti turimoje situacijoje, geriau suprasti uždavinio ypatybes. Tam visų pirma reikia kuo daugiau informacijos apie kriterinių įverčių aibę Φ, t.y. apie galimas kriterijų reikšmes ir jų sąryšius. Tik turėdamamas tokią informaciją, sprendėjas gali racionaliai derinti išorinius (pvz., projektinius) reikalavimus, konkretaus atvejo ypatybes ir vartojamo metodo galimybes.

Jei, pavyzdžiui, bent du kriterijai prieštarauja vienas kitam, tai nepagerinamų vektorių (Pareto) aibė sutampa su aibe A. Norint kuo nuodugniau ištirti Φ, reikėtų sudaryti tolygiai pasiskirsčiusių taškų tinklelį $X_i, i = 1, ..., N$, aibėje A ir analizuoti gaunamas tuose taškuose kriterijų $f_j(X_i)$ reikšmes. Tolygaus ieškojimo metodas gerai tinka tuomet, kai reikia ne itin tiksliai įvertinti globalinius ekstremumus. Suradus kriterijų ekstremumus, galima nurodyti aibės Φ ekstremalius taškus. Atmetę dominuojamus kriterijų vektorius (t.y. tuos F_j - kuriems galime surasti $F_k \geq F_j$), gaunama Pareto aibės aproksimacija. Ji pavaizduota 6.23 paveiksle taškais, sujungtais brūkšnine linija. Juos atitinką taškai aibėje A sudaro efektyviųjų sprendinių aibės aproksimaciją.

Taškų X_i, $i = 1, ..., N$, aibėje A tolygumą galime išreikšti kriterijumi

$$d(X_i,\, i = 1, ..., N) = \max_{X \in A} \min_{1 \leq i \leq N} ||X - X_i||,$$

kuris lygus didžiausios "skylės" radiusui tinklelyje $X_i, i = 1, ..., N$.

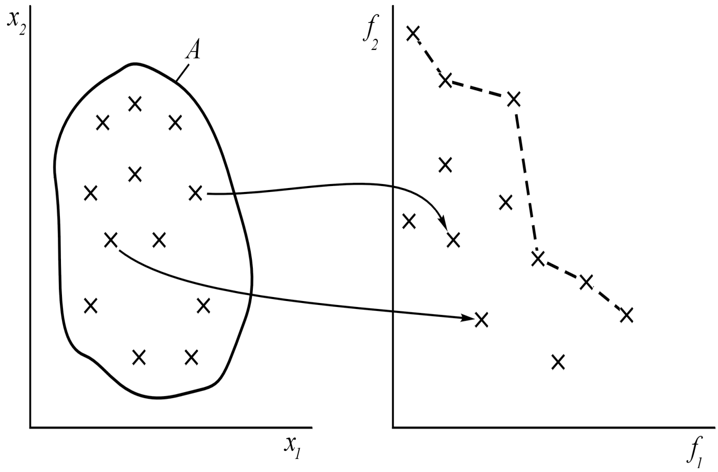

Figure 6.23: Pareto aibės aproksimacija

Bendru atveju, norint sudaryti tolygų tinklelį, reikia spręsti sudėtingą optimizavimo uždavinį:

$$\min_{X_i \in A,\ i=1,...,N} d(X_i,\, i = 1, ..., N).$$

Jei aibė A apibrėžta paprastais ribojimais (stačiakampis gretasienis daugiamatėje erdvėje), tai tolygų tinklelį pagrįstai galima pakeisti tinkleliu, kuris yra apytikriai tolygus ir paprastai sudaromas, pvz. LP_τ [22]. Kai aibė A yra sudėtinga, bet žinomas stačiakampis gretasienis $S \supset A$, kurio hipertūris nėra daug kartų didesnis už A hipertūrį, patartina naudoti tą patį metodą. Tuomet iš taškų $X_i \in S$ galima atrinkti tuos, kurie priklauso aibei A. Atrinktųjų taškų pasiskirstymas aibėje A apytiksliai tolygus, o jų dalis lygi aibių A ir S hipertūrių santykiui. Bendru atveju, sugeneruoti tolygiai pasiskirsčiusius taškus aibėje A, kai A apibrėžiama netiesiniais (ypač lygybiniais) ribojimais, nėra paprasta.

Gana dažnai stačiakampio gretasienio taškai generuojami naudojant pseudo atsitiktinių skaičių generatorių (specialią kompiuterio paprogramę). Toks būdas paprastesnis ir greitesnis už tolygių tinklelių panaudojimą, bet taškai išsidėsto ne itin tolygiai.

Net jei kriterijai ir neprieštarauja vienas kitam, efektyviųjų sprendinių aibės E hipertūris gali sudaryti labai mažą aibės A hipertūrio dalį. Todėl tolygaus tinklelio taškai gali nepatekti į aibės E aplinką, o aibių E (efektyviųjų sprendinių) ir Φ (Pareto) aproksimacijos bus nelabai tikslios. Šiuo atveju, siekiant padidinti tikslumą, reikėtų tolygų tinklelį pakeisti netolygiu, kurio taškai labiau koncentruotųsi efektyvių sprendinių aibės aplinkoje. Tačiau tokie metodai sudėtingesni, pagrįsti kriterijų funkcijų matematinėmis svybėmis, ir jų šioje knygoje nenagrinėsime.

Analitinį būdą efektyvių sprendinių ir Pareto aibėms apskaičiuoti iliustruosime pavyzdžiu. Sakykime, kad leistinoji aibė $A \subset R^2$ yra apibrėžta paprastais ribojimais $A = \{X : 0 \leq x_1, x_2 \leq 1\}$, du kriterijai - formulėmis

$$\begin{aligned} f_1(X) &= 2 - x_1^2 - x_2^2, \\ f_2(X) &= 2 - 2(x_1 - 0.8)^2 - (x_2 - 0.7)^2. \end{aligned}$$

6.24 paveiksle kriterijų funkcijos ir yra pavaizduotos lygio linijomis. Apskaičiuoti kriterijų reikšmių aibės ribas net šiuo atveju nėra taip paprasta. Kiekvieną tašką (f_1, f_2) srityje esančioje žemiau brūkninės linijos, atitinka du taškai (x_1, x_2). Pareto aibę sudaro kriterijų reikšmių aibės riba, pažymėta kreive, jungiančia kriterijų maksimalias reikšmes atitinkančius taškus, kurie pavaizduoti žvaigšdutėmis.

Aišku, kad (x_1, x_2) plokštumoje abiejų kriterijų maksimumų taškai įeina į efektyvių sprendinių aibę. Bet kuris efektyvusis sprendinys pasižymi tuo, kad jokia kryptimi, einančia iš to taško ir nevedančia iš leistinosios srities, negalima pagerinti iš karto abiejų kriterijų. Todėl

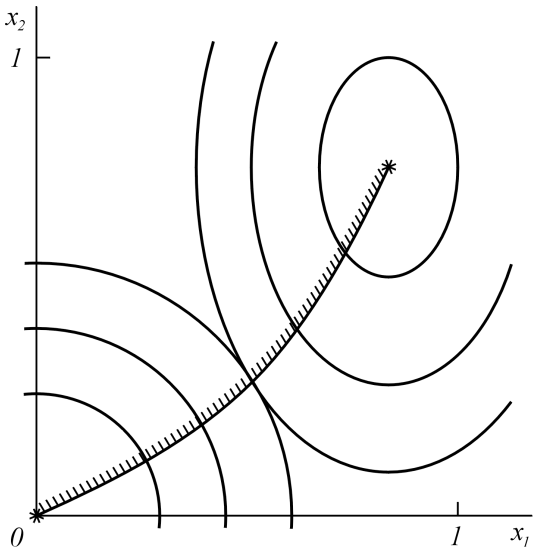

Figure 6.24: Kriterijų funkcijų lygio linijos ir efektyviųjų sprendinių aibė (pavaizduota užbrūkšniuota linija)

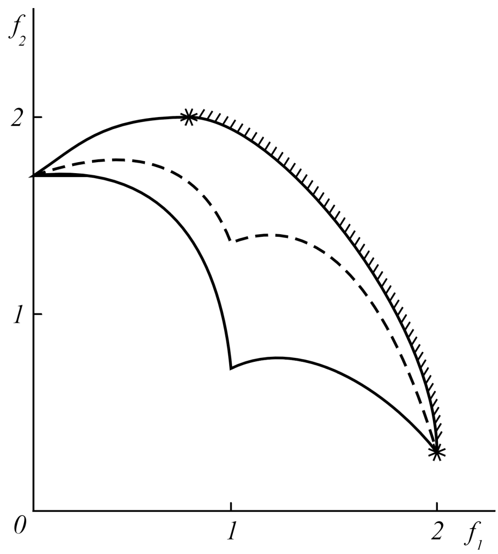

Figure 6.25: Kriterijų reikšmių aibė

būtiną efektyvaus sprendinio sąlygą galima suformuluoti šitaip: jei abi kriterijų funkcijos diferencijuojamos ir vidinis leistinosios srities taškas X yra efektyvus, tai kriterijų gradientai taške X yra priešingų krypčių

$$\lambda_1 \cdot \nabla f_1(X) + \lambda_2 \cdot \nabla f_2(X) = 0, \ \lambda_i \geq 0,$$

Apskaičiavę gradientų komponentes ir pakeitę vektorinę lygtį lygčių sistema komponentėms apskaičiuoti, gausime

$$\begin{aligned} -2x_1 - 4\lambda(x_1 - 0.8) &= 0, \\ -2x_2 - 2\lambda(x_2 - 0.7) &= 0. \end{aligned}$$

Iš pirmosios lygties išplaukia, kad

$$\lambda = -\frac{x_1}{2\,(x_1 - 0.8)}\ , \ \ \lambda \geq 0.$$

Įstatę λ į antrąją lygtį ir gausime

$$-x_2 + \frac{x_1(x_2 - 0.7)}{2(x_1 - 0.8)} = 0, \ 0 \leq x_1 \leq 0.8.$$

Šia lygtimi apibrėžiama efektyvių sprendinių aibė. Ją pertvarkę gausime kreivės lygtį

$$-2x_1x_2 + 1.6x_2 + x_1x_2 - 0.7x_1 = 0,$$

$$x_2(1.6 - x_1) = 0.7x_1,$$

$$x_2 = \frac{0.7x_1}{1.6 - x_1}\ , \ \ 0 \leq x_1 \leq 0.8.$$

Efektyvių sprendinių aibę atitinkanti kreivė prasideda taške (0, 0) ir monotoniškai auga, kol pasiekia tašką (0.8, 0.7). Tolimesnė aprašomos šia lygtimi kreivės tąsa jau nebeatitinka sąlygų, keliamų efektyvių sprendinių aibei. Kai $x_1 > 0.8$, tuomet parametro λ reikšmė yra neigiama. Todėl abiejų kriterijų gradientų kryptys sutampa. 6.24 paveiksle efektyvių sprendinių aibė pavaizduota užbrūkšniuota kreive.

6.6 Dialoginiai metodai

Keičiant daugiakriterio optimizavimo uždavinį vienkriteriu, naudojama papildoma informacija apie kriterijus, kriterijų svorius, minimalias leistinas kriterijų vertes, tikslo vektorius ir kt. Tačiau gautasis vienkriteris uždavinys gali pasirodyti neišsprendžiamas. Pavyzdžiui, keičiant daugiakriterį uždavinį vienkriteriu matematinio programavimo uždaviniu ir nustačius pernelyg dideles reikšmes ribojimų nelygybių dešiniosiose pusėse leistinoji aibė gali pasirodyti tuščia. Tas reikštų, kad pareikalauta sprendinio su pernelyg didelėmis kriterijų reikšmėmis. Kita vertus, gautasis sprendinys gali nepatenkinti vartotojo. Kai kada Pareto aibei aproksimuoti labiau naudotinas tolygus tinklelis, kai kada - netolygus. Išskirtinę reikšmę gali turėti vienas arba kitas Pareto aibės poaibis. Todėl svarbu, kad tuoj pat, gavus atsakymą, būtų galima performuluoti uždavinį ir vėl jį spręsti kitais metodais arba su kitais duomenimis. Kad vartotojui būtų lengviau vykdyti kompiuteriu daugiakriterės optimizacijos algoritmus, kuriamos specialios dialoginės programų sistemos. Daugiakriterio uždavinio sprendėjui interaktyvaus sprendimo galimybė ypaš svarbi tuomet, kai reikia įvertinti gautą kriterijų vektorių atsižvelgiant į neformalią vartotojo informaciją apie uždavinį.

Vienu iš pirmųjų dialoginių daugiakriterio optimizavimo metodų buvo gradientinis metodas naudingumo funkcijai maksimizuoti. Tačiau vietoje naudingumo funkcijos reikšmių skaičiavimo taške X_k buvo klausiama eksperto nuomonės apie kriterijų, apskaičiuotų tame taške, santykinį reikšmingumą. Sakykime, kad egzistuoja, nors ir nežinoma, naudingumo funkcija $U(f_1(X), ..., f_m(X))$. Leistinoji sritis A yra iškila aibė. Visos uždavinio funkcijos įgaubtos (iškilos į viršų) ir tolygiai diferencijuojamos. Standartinis lokalaus maksimizavimo metodas aprašomas šitaip:

1. Inicializacija: parenkamas pradinis taškas X_0, $k = 0$.

2. Leistinos krypties parinkimas: vektorius S_k apskaičiuojamas taip, kad jo projekcija į funkcijos $U(f_1(X), ..., f_m(X))$ gradiento vektorių taške X_k būtų maksimali:

$$S_k = \arg \max_{S,\ X_k+S\in A} (\nabla_X\ U(f_1(X_k), ..., f_m(X_k)),\ S).$$

3. Žingsnio prinkimas: apskaičiuojamas žingsnio ilgis

$$t_k = \arg \max_{0\leq t\leq 1} U(f_1(X_k + t\, S_k), ..., f_m(X_k + t\, S_k)).$$

4. Naujo sprendinio apskaičiavimas: $X_{k+1} = X_k + t\, S_k$, $k = k + 1$.

5. Sugrįžimas į **2.**

Tačiau, nežinodami funkcijos $U(\cdot)$, negalime įvykdyti nei 2, nei 3 etapo. Taigi, kyla klausimas, kaip formalią informaciją apie $U(\cdot)$ reikšmes pakeisti informacija gaunama iš vartotojo. Funkcijos $U(\cdot)$ gradientą taške X_k galima išreikšti formule

$$\nabla_X\ U(f_1(X_k), ..., f_m(X_k) = \sum_{i=1}^{m} \frac{\partial U_k}{\partial f_i} \cdot \nabla_X\ f_i(X_k),$$

čia $\frac{\partial U_k}{\partial f_i}$ - dalinė funkcijos $U(\cdot)$ išvestinė pagal f_i, apskaičiuota taške X_k. Būtent šios dalinės išvestinės ir yra nežinomos. Kadangi naudingumo funkcija yra monotoniškai didėjanti, tai visos dalinės išvestinės yra teigiamos. Todėl vektorių S_k galima apskaičiuoti šitaip:

$$S_k = \arg \max_{\substack{S \\ X_k+S\in A}} (\nabla_X\ U(f_1(X_k), ..., f_m(X_k))\, / \frac{\partial U_k}{\partial f_1},\ S) =$$

$$\arg \max_{\substack{S \\ X_k+S \in A}} (\sum_{i=1}^{m} a_i^k \cdot \nabla_X f_i(X_k),\, S),$$

čia $a_i^k = \frac{\partial U_k}{\partial f_i} / \frac{\partial U_k}{\partial f_1}$.

Koeficientu a_i^k išreiškiamas i-ojo kriterijaus santykinis reikšmingumas, t.y., kiek kartų naudingumo prieaugis bus didesnis, padidėjus kriterijui f_i, negu tiek pat padidėjus kriterijui f_1 (kriterijų prieaugiai turi būti maži). Remdamasis šia interpretacija, ekspertas dažniai gali įvertinti koeficentus a_i^k.

Koeficientus įvertinti galima ir kitaip. Sakykime, kad pajudėjus iš taško X_k, kriterijaus f_1 reikšmė padidėja Δ_1. Visi kiti kriterijai, išskyrus i-ąjį, nekinta. Koks turi būti i-ojo kriterijaus neigiamas prieaugis -Δ_i, kad naująjį tašką galėtume laikyti ekvivalenčiu X_k? Parinkę Δ_i, patenkinančius suformuluotą sąlygą, a_i^k galėsime apskaičiuoti pagal formulę $a_i^k = \Delta_1/\Delta_i$.

Nežinodami funkcijos $U(\cdot)$, formaliai negalime įvykdyti ir 3-ojo etapo. Tačiau, pateikę ekspertui visų kriterijų priklausomybės nuo t grafikus $f_i(X_k + t\, S_k)$, galime tikėtis, kad jis parinks tinkamą t_k reikšmę.

Šis metodas demonstruoja principinę interaktyvaus daugiakriterio uždavinio sprendimo galimybę, nors praktiškai ir nelabai vartojamas. Kuriant interaktyvius metodus reikia atsižvelgti į psichologinius faktorius. Pavyzdžiui, ekspertai labiau linkę duoti kokybinius dviejų alternatyvų palyginimus negu kiekybinius įvertinimus. Vartojant interaktyvią gradientinio metodo versiją sunku atsižvelgti į priimtiną ekspertui sprendimo laiką: gradientinio metodo žingsniai dažnai būna trumpi, ir todėl sprendinio ieškojimas gali pernelyg užsitęsti.

Vartotojams psichologiniu požiūriu priimtinesnis deformuojamo simplekso metodo dialoginė modifikacija, skirta daugiakriteriams uždaviniams spręsti. Aibėje A sudaromas simpleksas, kurio viršūnėse $X_1, ..., X_{k+1}$ apskaičiuojami kriterijų vektoriai $F_1, ..., F_{k+1}$. Projektuotojui siūloma palyginti jų poras. Išrenkamas blogiausias taškas ir jis atspindimas simplekso svorio centro atžvilgiu erdvėje R^n. Naująjį simpleksą sudaro senojo simplekso viršūnės išskyrus tą, kuri buvo įvertinta kaip blogiausia. Pastaroji pakeičiama jos atspindžiu. Naujojoje viršūnėje apskaičiuojamas kriterijų vektorius F, kurį reikia palyginti su kai kuriais kriterijų vektoriais, atitinkančiais senojo simplekso viršūnes. Simplekso deformacijas (tempiamą, suspaudimą ar redukciją), kaip ir originaliame vienakriteriame metode, apsprendžia preferencijos santykis tarp atitinkančių viršūnes kriterijų vektorių. Tai yra neesminis apibendrinimas lyginant su vienkriteriu deformuojamojo simplekso metodu, kuriame viršūnės palyginamos pagal jose apskaičiuotas tikslo funkcijos reikšmes. Ekspertui, sprendžiančiam šį uždavinį, tereikia palyginti kriterijų vektoruis. Aptariant deformuojamojo simplekso metodo savybes vienkriterio optimizavimo atveju, buvo pažymėta, kad jis nejautrus tikslo funkcijos nereguliarumams bei skaičiavimo paklaidoms. Todėl ir esant kriterijų gausai galima tikėtis, kad, vartototojui nežymiai suklydus lyginant altrnatyvas, nebus nepataisomų pasekmių.

Dažnai uždavinio charakteristikų nepavyksta suklasifikuoti į ribojimus ir kriterijus. Siekiant praktinio uždavinio charakteristiką formalizuoti kaip matematinio programavimo uždavinio ribojimą, gali būti nelengva apibrėžti ribojimų dešiniąsias puses. Panašūs sunkumai gali iškilti bandant apibrėžti leistinas kriterijų nuolaidas. Tada galima rinktis patį bendriausią, taip vadinamą, parametrų erdvės zondavimo metodą, kuris panašus į pasyvų ieškojimą su tolygiu taškų išdėstymu vieno kriterijaus optimizavimo uždaviniuose. Leistinoji sritis apibrėžiama parastaisiais ribojimais ir tik neabejotinai ribojimų prasmę turinčiomis funkcijomis $g_i(X) \geq 0$. Jei funkciją $g_i(X)$ galima interpretuoti ir kaip kriterijų, ir kaip ribojimą, tai ji priskiriama prie kriterijų. Parametrų erdvės zondavimo metodas [18] susideda iš trijų toliau aprašytų dalių. Antroji ir trečioji pakaitomis kartojamos tol, kol bus pasiektas priimtinas sprendinys.

1. Sudaroma bandymų lentelė. Leistinoje srityje A (pseudo) atsitiktinai tolygiai generuojami taškai $X_1, ..., X_N$. Visuose tškuose apskaičiuojamos m kriterijų reikšmės

$$f_1(X_i), ..., f_m(X_i), i = 1, ..., N,$$

kurios išdėstomos mažėjimo tvarka (žiūr. lentelę), nurodant taško numerį:

$$f_k(X_{k1}) \geq f_k(X_{k2} \geq ... \geq f_k(X_{kN}),\ k = 1, ..., m.$$

Iš tinklelio savybių išplaukia, kad

$$f_k(X_{k1}) \to \max_{X \in A} f_k(X)\,, \quad f_k(X_{kN}) \to \min_{X \in A} f_k(X),$$

o funkcijos reikšmių intervalų empyriniai dažniai artėja prie tų intervalų tikimybių. Todėl iš lentelės duomenų galima spręsti apie minimalią ir maksimalią kriterijaus reikšmes bei kriterijų reikšmių tikimybių skirstinį.

2. Nustatomi kriterijų reikšmių slenksčiai. Bandymų lentelė pateikiama projektuotojui. Žemiau pateiktas konkretaus uždavinio sprendimo lentelės pavyzdys [18]. Išanalizavęs gautąsias kriterijų reikšmes, jis nustato kiekvieno kriterijaus slenkstį f_i^*, žemiau kurio kriterijaus reikšmės laikytinos nepriimtinomis.

3. Tikrinama, ar aibė, kurioje kriterijai viršija slenksčius, yra netuščia. Tas realizuojama parastu algoritmu: atrenkami taškai, kuriuose slenkstį viršija pirmajasis kriterijus $f_1(X_i) \geq f_1^*$, po to antrajasis ir tt. Jei nė vieno tokio taško nėra, tai galima daryti išvadą: arba aibė tuščia, arba nepavyko jos aptikti. Jei padidinus N, taško, tenkinančio ribojimus, vis tiek nepavyksta rasti, tai manoma, kad aibė tuščia arba jos hipertūtis labai mažas. Vartotojui pranešama, kad jo reikalavimai kriterijams per griežti (nesuderinami) ir siūloma sumažinti slenksčių reikšmes. Suradus priimtiną tašką, jis tikslinamas: tinklelui padengiama palyginti nedidelė rastojo taško aplinka ir apskaičiuotos kriterijų reikšmės, priklausančios Pareto aibei, pateikiamos vartotojui; iš jų išrenkamas uždavinio sprendinys.

Gali atrodyti, jog uždavinys labai paprastas. Vartotojas tik analizuoja vieną iš lentelės stulpelių. Jam nereikia ieškoti kompromiso tarp kriterijų, kaip to reikalauja nuosekliųjų nuolaidų metodas. Aišku, vartotojui norėtųsi, kad slenksčių reikšmės būtų artimos maksimalioms kriterijų reikšmėms. Todėl kompromisas tarp kriterijų pasiekiamas slenksčių nustatymu. Atkreipsime dėmesį į tai, kad parinkus pernelyg dideles slenksčių reikšmes būtų gauta tuščia aibė $\{X : X \in A,\ f_i(X) \geq f_i^*\ , i = 1, ..., m\ \}$, ir nebeliktų iš ko išrinkti sprendinį.

L e n t e l ė. Parametrų erdvės zondavimas, $N = 512$

$f_1(X_i)$	i	$f_2(X_i)$	i	...	$f_m(X_i)$	i
-1.46	416	-3.13	416		-1.31	384
-1.53	348	-3.36	348		-1.41	374
-1.54	112	-3.46	112		-1.47	260
-1.62	280	-3.52	448		-1.50	248
-1.65	448	-4.02	302		-1.54	353
...	...	...	...	...	...	...
-9.82	407	-32.8	89		-17.0	46
-10.0	71	-34.2	407		-17.5	507

Dažnai kyla klausimas, kokia konkrečiu atveju turi būti parinkta N reikšmė? Metodo autoriai nurodo, kad dauguma praktinių uždavinių buvo sėkmingai išspręsti su $N = 128$, arba $N = 256$, arba netgi $N = 30$ [18]. Jei kriterijų funkcijoms apskaičiuoti reikia labai daug kompiuterio laiko, tai pasirenkamos mažesnės N reikšmės.

Tolygiai padengti leistinąją sritį tinkleliu gan paprasta tuomet, kai aibė A - stačiakampis gretasienis; nesudėtinga, kai A yra įrašyta į stačiakampį gretasienį, kurio hipertūris artimas A hipertūriui;

labai sunku, kai A aprašoma netiesiniais lygybiniais ribojimais. Nerekomenduojama naudoti šį metodą sprendžiant uždavinius su prieštaringais kriterijais. Parametrų erdvės zondavimo metodas artimas nuosekliųjų nuolaidų metodui, kurio dialoginius variantus naudoja nemažai projektuotojų.

Sudarant daugelį dialoginių algoritmų aukščiau aprašyti matematiniai metodai naudojami arba kaip galimi sprendimo metodo variantai, arba kaip pagalbinės priemonės.

6.7 Metodų aptarimas

Skaitytojas, baigdamas skaityti skyrelį, aišku, norėtų gauti konkrečias rekomendacijas, kaip pasirinkti tinkamą metodą. Deja, daugiakriterių optimizavimo uždavinių klasifikacijos labai sąlyginės. Antra vertus, vartotojų informacija apie kriterijus labai įvairi, ir kiekvieno sudėtingo praktinio uždavinio sprendimas, galima sakyti, yra menas. Matyt, todėl net storose knygose, skirtose daugiakriterei optimizacijai, nėra vienareikšmių rekomendacijų, kaip pasirinkti metodą konkretiems uždaviniams spręsti.

Kad projektuotojas galėtų pagrįsti savo pasirinktą variantą, jis siekia sužinoti kriterijų kitimo ribas, tipines kriterijų reikšmių kombinacijas. Todėl, kai kriterijai yra neprieštaringi, pasiteisina parametrų erdvės zondavimo metodas. Jį vartoja nemažai mašinų projektuotojų.

Dažnai pasiteisina daugiakriterių uždavinių keitimas vieno kriterijaus optimizavimo uždavinių seka: tam tikras kriterijus pasirenkamas tikslo funkcija, visi kiti keičiami ribojimais. Taip buvo išspręsta nemažai įvairių sričių optimalaus projektavimo uždavinių. Nors šis metodas ir nėra labai gerai teoriškai pagrįstas, bet, remiantis sukaupta patirtimi, galima efektyviai panaudoti įprastus optimizavimo metodus. Parinkdami kriterijams leidžiamas tolerancijas, gauname nuosekliųjų nuolaidų metodą.

Jei kriterijai gali būti interpretuojami kaip pelnas arba nuostoliai, tai tikėtina surasti bendrą jų matavimo skalę, įgalinančią naudotis svertinės sumos kompozicija.

Norint pasirinkti tinkamą metodą uždaviniui spręsti, reikėtų išanalizuoti metodų ir kompiuterių programų aprašymus, praktinių uždavinių sprendimo rezultatus. Mokslinio tyrimo darbuose įdomu patyrinėti įvairius metodus, palyginti jais gaunamus rezultatus. Bet praktikoje iškylantys uždaviniai dažnai turi būti išspręsti iki griežtai nustatytų terminų. Todėl projektuotojui nebėra kada eksperimentuoti bandant įvairius metodus. Pagrindinis šios knygos tikslas – supažindinti su pagrindiniais metodais, jų racionalaus taikymo sritimis. Susipažinęs su šiame skyriuje aptartais metodais, skaitytojas galės lengviau orientuotis reklaminėse įvairių daugiakriterių uždavinių sprendimo metodų anotacijose bei objektyviau vertinti praktinių uždavinių sprendimo rezultatų publikacijas.

Pasirenkant metodą, rekomenduotume atreipti dėmesį į šiuos aspektus:

1. *Adektvatumas.* Daugiakriteriai uždaviniai paprastai sprendžiami dialoginiais metodais. Dialoge su kompiuteriu turi dalyvauti kompetentingas specialistas arba jų grupė. Svarbu užtikrinti, kad tai tikrai būtų tas dialogas, kuris numatytas metode. Reikėtų detaliai apgalvoti, ar tie klausimai, į kuriuos turės atsakyti ekspertai, yra būdingi jų veiklai, ar dialogas padės giliau suprasti uždavinį. Metodą galima vadinti tinkamu, jei jį naudodamas vartotojas dirba tikslingai, turima informacija naudojama racionaliai.

2. *Naudojimo paprastumas.* Nepatyrusiam vartotojui yra sukurti specialūs dialoginių metodų variantai. Vėliau, įgijęs patirties, vartotojas gali dirbti naudodamas efektyvesnius metodo variantus, skirtus patyrusiam vairuotojui arba "meistrui". To paties metodo variantų realizacjos skirtingoms vartotojų katrgorijoms, gali gerokai skirtis. Pasirenkant programinę įrangą reikėtų atsižvelgti į tai, kokiomis prielaidomis apie vartotojo žinias daugiakriterės optimizacijos sri-

tyje buvo vadovautasi kuriant daugiakriterės optimizacijos taikomųjų programų paketą. Nemažą reikšmę turi ir tai, per kiek laiko galima gauti programinę įrangą, kokia jos kaina.

3. *Efektyvumas.* Metodų kūrėjų požiūriu tai pati svarbiausia metodo savybė, įrodoma teoriškai arba nustatoma eksperimentiškai. Ar metodas praktiškai efektyvus, galima sužinoti tik tada, kai jis yra pakankamai plačiai vartojamas, t. y. pakankamai paprastas ir priimtinas daugeliui vartotojų. Metodas laikomas praktiškai efektyviu, jei juo randami geresni variantai, negu vartojant kitus metodus.

6.8 Užduotys ir kontroliniai klausimai

1. Ar daugiakriterio uždavinio atskirų kriterijų funkcijų maksimumų taškai priklauso efektyvių sprendinių aibei? Kodėl?

2. Ar galima svertinės kriterijų sumos kompozicijos svorius parinkti taip, kad ši kompozicija atitiktų "silpniausios grandies stiprinimo" principą?

3. Dvikriteriam uždaviniui

$$
\begin{aligned}
f_1(x) &= x_1 + x_2, \\
f_2(x) &= 1 - (x_1 - 2)^2 - (x_2 - 0.5)^2, \\
x &= (x_1, x_2),\ 0 \le x_1, x_2 \le 1,
\end{aligned}
$$

grafiškai pavaizduokite kriterijų reikšmių, Pareto ir efektyvių sprendinių aibes.

4. Išanalizuokite trijų kriterijų leksikografinį uždavinį su kontinualia leistinąja sritimi.

5. Suformuluokite daugiakriterį uždavinį, atitinkantį savo savaitės biudžeto planavimą.

Bibliography

[1] E. Aarts and J.Korst. *Simulated Annealing and Boltzman Machines*. J.Wiley and Sons, 1989.

[2] M. Bazaraa, H.Sherali, and C.Shetti. *Nonlinear Programming, Theory and Algorithms*. John Wiley and Sons, 1993.

[3] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.

[4] J. Cea. *Optimisation Theorie et Algorithmes*. Dunod, 1971. (yra vertimas į rusų kalbą).

[5] J. Clausen. Branch and bound algorithms – principles and examples. *"http://vaidila.vdu.lt/tempus"*,, 1998.

[6] J. Clausen. Teaching duality in linear programming – the multiplier approach. *"http://vaidila.vdu.lt/tempus"*, 1998.

[7] P. Fishburn. *Utility Theory for Decision Making*. J.Wiley, 1970. (yra vertimas į rusų kalbą).

[8] F. Hillier and G.Lieberman. *Introduction to Mathematical Programming*. McGraw-Hill, 1991.

[9] D. Himmelblau. *Applied Nonlinear Programming*. McGrw-Hill, 1972. (yra vertimas į rusų kalbą).

[10] R. Horst, P.Pardalos, and N.Thoai. *Introduction to Global Optimization.* Kluver, 1995.

[11] C. Kelley. *Iterative Methods for Optimization.* SIAM, 1999.

[12] J. Kubilius(red.). *Matematikos terminų žodynas.* Mokslo ir enciklopedijų leidykla, 1994.

[13] R. Luce and H.Raiffa. *Games and Decisions.* J.Wiley, 1957. (yra vertimas į rusų kalbą).

[14] J. Mockus. *Bayesian Approach to Global Optimization.* Kluver, 1989.

[15] J. Nocedal and S.Wright. *Numerical Optimization.* Springer, 1999.

[16] B. Poliak. *Vvedenije v optimizaciju.* Nauka, 1983. (Rusų kalba).

[17] H-P. Schwefel. *Evolution and Optimum Seeking.* J.Wiley and Sons, 1995.

[18] I. Sobol and R.Statnikov. Lp - poisk v zadačach optimalnogo konstruirovanija. *Problemy slučainogo poiska*, 1:117–135, 1972. (Rusų kalba).

[19] D. Tempelaar, F.Wiedersheim-Paul, and E.Gunnarsson. *Educational Innovations in Economics and Business II.* Kluver Academic Publishers, 1998.

[20] A. Törn and A.Žilinskas. *Global Optimization.* Springer, 1989.

[21] V. Čiočys ir R.Jasilionis. *Matematinis programavimas.* Mokslas, 1990.

[22] V. Šaltenis ir A.Žilinskas. *Techninių optimizavimo uždavinių sprendimas*. Mokslas, 1986.

[23] A. Žilinskas. *Naujieji projektavimo metodai*. Mokslas, 1990.

[24] A. Žilinskas. *Matematinis programavimas*. Kaunas, VDU leidykla, 1999.

[25] A. Žilinskas et al. Modelling of economics and business systems, s-jep 12382-97. Home page, Vytautas Magnus University: Kaunas, http://vaidila.vdu.lt/tempus, 1998.

[26] F. Vasiljev. *Čislenyje metody rešenija ekstremalnych zadač*. Nauka, 1988. (Rusų kalba).