

MITROIU MARIA ROXANA

**METODE DE PUNCT INTERIOR ÎN
OPTIMIZAREA MATEMATICĂ**

ISBN

978-973-0-39908-0

TÂRGU JIU, 2024

MITROIU MARIA ROXANA

**METODE DE PUNCT INTERIOR
ÎN OPTIMIZAREA MATEMATICĂ**



Târgu Jiu
2024

ISBN

978-973-0-39908-0

Cuprins

Introducere	4
Capitolul 1: Metode de punct interior de urmărire a traiectoriei în programarea liniară	5
1.1. Problema barieră.....	6
1.2. Multiplicatori Lagrange.....	8
1.3. Traiectoria centrală.....	10
1.4. Metoda primal-duală de urmărire a traiectoriei pentru programarea liniară.....	14
1.5. Metoda predictor-corector de urmărire a traiectoriei pentru programarea liniară.....	25
Capitolul 2. Metode de punct interior pentru programarea Matematică	33
2.1. Metode de punct interior pentru problema de programare pătratică.....	33
Capitolul 3. Algoritmul Karmarkar	40
3.1. Problema în forma standard Karmarkar.....	40
3.2. Minimizarea pe sferă.....	41
3.3. Evaluări de regularitate a domeniului de admisibilitate.....	41
3.4. Algoritmul Karmarkar.....	45
Concluzii	49
Bibliografie	50

INTRODUCERE

Lucrarea de față este consacrată prezentării principalelor aspecte teoretice ale metodelor de punct interior în programarea matematică. Aceasta își propune redarea principiilor care stau la baza metodelor de punct interior în programarea matematică, împreună cu capacitățile lor de a rezolva probleme reale de dimensiuni mari. Am ales această temă deoarece este incitantă prin actualitatea, extinderea și importanța ei, necesitând o muncă laborioasă de cercetare și documentare.

Lucrarea este structurată în trei capitole. Primul capitol tratează metodele de punct interior bazate pe urmărirea traiectoriei. Aceste metode au impus un concept foarte productiv, acela de traiectorie centrală, care a schimbat profund optica procedurilor de rezolvare a problemelor de programare matematică. Traectoria centrală este o curbă din domeniul de admisibilitate al problemei cu proprietăți foarte interesante, care, evidențiază geometric proprietățile metodelor barieră din programarea matematică neliniară. În acest capitol se vorbește despre metoda primal-duală și despre metoda predictor-corector de urmărire a traiectoriei pentru programarea liniară. Sunt prezentați algoritmul primal-dual și algoritmul predictor-corector. Aceștia urmăresc traiectoria centrală către optim, fiind deci algoritmi de urmărire a traiectoriei.

Cel de-al doilea capitol, „Metode de punct interior pentru programarea matematică”, are ca scop extinderea conceptelor proprii metodelor de punct interior la clasa problemelor de programare pătratică. Aceasta este o extensie foarte naturală, care pune în valoare rezultatele obținute în cadrul programării liniare.

În capitolul al treilea, este prezentată forma inițială a algoritmului Karmarkar, așa cum a fost publicat în 1984. Algoritmul lui Karmarkar este o metodă de punct interior care generează în interiorul domeniului de admisibilitate un șir de puncte ale căror costuri converg către costul optimal, obținându-se în final un punct suboptimal, care cu ajutorul unor transformări elementare este adus într-un punct optimal de pe frontiera domeniului de admisibilitate. Acesta a constituit punctul de plecare a unei noi direcții în programarea matematică.

Metodele de punct interior reprezintă o abordare serioasă a problemei programării matematice, capabilă de a furniza algoritmi deosebit de eficienți pentru rezolvarea problemelor de dimensiuni mari.

Capitolul 1. METODE DE PUNCT INTERIOR DE URMĂRIRE A TRAIECTORIEI ÎN PROGRAMAREA LINIARĂ

În acest capitol, se descriu o clasă de metode de punct interior utilizate în rezolvarea problemelor de programare liniară bazate pe conceptul de urmărire a traiectoriei centrale. Dintre toate metodele de punct interior, acestea par să fie cele mai simple și mai naturale. Proprietățile noțiunii de „traiectorie centrală” se particularizează într-un mod foarte interesant pentru clasa problemelor de programare liniară. Metodele de punct interior au readus în actualitate noțiunile de punct central și traiectorie centrală și sunt utilizate într-un mod magistral pentru rezolvarea problemelor de programare matematică.

Fie problema de programare liniară:

$$\begin{aligned} & \max c^T x, \\ & \text{referitor la:} \\ & Ax = b, x \geq 0, \end{aligned} \tag{1.1}$$

unde presupunem că $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ și $\text{rang}(A) = m$. Pentru evitarea cazurilor triviale, presupunem că domeniul de admisibilitate al problemei: $S = \{x \in \mathbb{R}^n / Ax = b, x \geq 0\}$ este mărginit și are un interior nevid, adică $S^0 = \{x \in \mathbb{R}^n / Ax = b, x > 0\} \neq \emptyset$.

Metoda lui Dantzig de rezolvare a problemei de programare liniară a fost dezvoltată de-a lungul a 40 de ani de cercetări teoretice și computaționale și este încă cel mai utilizat algoritm, fiind posibil să-și mențină multă vreme acest statut. Algoritmii simplex nu posedă însă o proprietate care în ultimii 15-20 de ani a devenit din ce în ce mai căutată, anume complexitatea polinomială.

În anul 1984, Karmarkar a proiectat un algoritm mult mai eficient decât algoritmul simplex, mai ales pentru rezolvarea problemelor de dimensiuni mari. Acest algoritm avea o complexitate polinomială mărginită de $O(n^{3.5}L)$ operații aritmetice ($L =$ lungimea intrării – numărul de biți necesari reprezentării datelor problemei). O variantă a algoritmului lui Karmarkar reproduce întocmai un algoritm datorat lui Dikin [1967], care acum este cunoscută ca metoda de scalare afină. Acest algoritm se comportă foarte bine, deoarece de-a lungul iterațiilor evită punctele aflate în vecinătatea frontierei domeniului de admisibilitate al problemei. Acest lucru se realizează apelând la funcția barieră logaritmică, care, pentru orice $x \in \mathbb{R}^n$, $x > 0$, este definită sub forma:

$$p(x) = \sum_{i=1}^n \log(x_i) \tag{1.2}$$

Această funcție scade nedefinit lângă frontiera domeniului S și poate deci fi utilizată ca o penalizare asociată acestor puncte. Combinând $p(\cdot)$ cu $c^T x$ din (1.1), obținem o funcție care ia valori foarte mici lângă frontiera lui S , forțând astfel orice algoritm de minimizare să evite aceste puncte.

Principala idee care se află în spatele tuturor metodelor de punct interior constă în realizarea simultană a două obiective:

- reducerea valorii funcției obiectiv,
- evitarea, pe cât posibil, a punctelor din vecinătatea frontierei domeniului de admisibilitate.

Metodele de punct interior de urmărire a traiectoriei implementează ideea de a ține iterațiile lângă traiectoria centrală. Pentru aceasta, ele apelează la o combinație a metodelor bazate pe condițiile de optimalitate ale cuplului primal-dual cu metode de penalizare.

1.1. PROBLEMA BARIERĂ

Fie problema de programare liniară cu restricții inegalități:

$$\begin{aligned} & \max c^T x, \\ & \text{referitor la:} \\ & Ax \leq b, x \geq 0. \end{aligned} \tag{1.3}$$

Problema duală corespunzătoare este:

$$\begin{aligned} & \min b^T y, \\ & \text{referitor la:} \\ & A^T y \geq c, y \geq 0. \end{aligned} \tag{1.4}$$

Vom introduce variabilele ecart, pentru a aduce aceste probleme la forma standard, cu restricții egalități:

$$\begin{aligned} & \max c^T x, \\ & \text{referitor la:} \\ & Ax + s = b, x, s \geq 0. \end{aligned} \tag{1.5}$$

și

$$\begin{aligned} & \min b^T y, \\ & \text{referitor la:} \\ & A^T y - z = c, y, z \geq 0. \end{aligned} \tag{1.6}$$

Pentru a elimina restricțiile inegalități din aceste probleme, vom aplica principiul penalizării. În (1.5), putem elimina restricția ca o variabilă oarecare, să zicem x_j , să fie nenegativă, prin adăugarea la funcția obiectiv a unui termen care ia valoarea minus infinit, când x_j este negativ și valoarea zero, în rest. Această reformulare a problemei nu este foarte folositoare, deoarece noua funcție obiectiv are discontinuități abrupte și nu ne permite utilizarea tehnicilor de studiu date de analiza matematică. Deci, vom înlocui această funcție discontinuă cu o altă funcție care este minus infinit, când x_j este negativ, dar ia o valoare finită, pentru x_j pozitiv și tinde la minus infinit, când x_j tinde la zero. Astfel, am netezit discontinuitatea și putem aplica așadar rezultatele clasice ale analizei matematice. Funcția logaritmică este, în mod evident, cea mai simplă funcție cu această proprietate. Pentru fiecare variabilă, vom introduce în funcția obiectiv un nou termen, care este logaritmul variabilei respective, ponderat cu o constantă pozitivă, obținând astfel problema:

$$\max c^T x + \mu \sum_{j=1}^n \log(x_j) + \mu \sum_{i=1}^m \log(s_i),$$

referitor la : (1.7)

$$Ax + s = b.$$

Această problemă nu pare prea diferită de problema originală, deși nu este echivalentă cu ea. Este evident că dacă parametrul μ , care este presupus pozitiv, devine din ce în ce mai mic, atunci problema (1.7) devine o aproximare din ce în ce mai bună a lui (1.5). Problema (1.7) se numește *problema barieră* asociată lui (1.5). Avem, de fapt, o familie de probleme, parametrizată după valorile lui μ . Fiecare dintre aceste probleme este neliniară. Funcția obiectiv neliniară se numește *funcția barieră*, sau, mai precis, *funcția barieră logaritmică*.

Este util să avem o imagine geometrică a acestei funcții barieră. Se știe că pentru probleme exprimate în forma standard, mulțimea soluțiilor admisibile formează un poliedru. Pe fiecare față a acestuia, o variabilă se anulează. Deci, funcția barieră este minus infinit pe fiecare față a poliedrului. Mai mult, ea este finită în interiorul poliedrului și tinde la minus infinit, când ne apropiem de frontiera acestuia. În figura 1.1 se observă câteva curbe de nivel constant ale funcției barieră pentru o problemă generică și pentru câteva valori ale parametrului μ . Pentru fiecare μ , maximul este atins într-un punct interior, și când μ tinde la zero, acest punct interior se deplasează din ce în ce mai aproape de soluția optimă a problemei de programare liniară originală, care este un vârf al poliedrului (fig. 1.1, a, b, c).

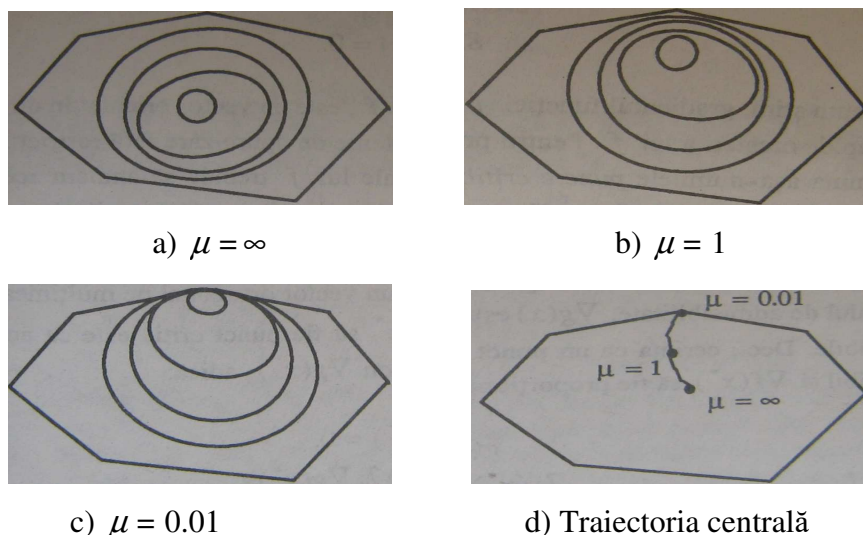


Fig. 1.1. Curbele de nivel constant ale funcției barieră pentru trei valori ale lui μ . Pentru fiecare valoare a lui μ se arată patru curbe de nivel constant.

În Figura 1.1. se observă că valoarea maximă a funcției barieră se atinge pe curba de nivel constant plasată în interiorul tuturor curbelor de nivel constant. Figura 1.1. arată traiectoria centrală. Privită ca o funcție de μ , mulțimea soluțiilor optime ale problemei barieră formează o traiectorie în interiorul poliedrului soluțiilor admisibile. Aceasta se numește *traiectorie centrală* – sau *drum central* (fig 1.1.d) și are un rol important în definirea și studierea algoritmilor de punct interior pentru programarea liniară și în general pentru programarea matematică.

1.2. MULTIPLICATORI LAGRANGE

În această secțiune, vom detalia condițiile privind existența unui maxim local a unei probleme de optimizare cu restricții liniare.

Pentru început, considerăm problema cu o singură restricție de egalitate:

$$\begin{aligned} & \max f(x), \\ & \text{referitor la:} \\ & g(x) = 0. \end{aligned} \tag{1.8}$$

După cum știm, gradientul funcției f , ∇f , este un vector orientat în direcția celei mai rapide creșteri ale lui f . În problemele de optimizare fără restricții, pentru a determina așa numitele *puncte critice* ale lui f , trebuie să anulăm acest vector. Punctul de maxim, dacă există, se află inclus în această mulțime a punctelor critice. În prezența restricției $g(x) = 0$, gradientul

trebuie să fie ortogonal pe mulțimea soluțiilor admisibile $\{x / g(x) = 0\}$. Se observă că, în fiecare punct x din domeniul de admisibilitate, $\nabla g(x)$ este un vector ortogonal pe mulțimea punctelor admisibile. Deci, cerința ca x^* să fie punct critic este ca acesta să fie admisibil și $\nabla f(x^*)$ să fie proporțional cu $\nabla g(x^*)$, adică:

$$\begin{aligned} g(x^*) &= 0, \\ \nabla f(x^*) &= \lambda \nabla g(x^*), \end{aligned}$$

unde λ este constanta de proporționalitate, numită *multiplicatorul Lagrange* asociat restricției $g(x) = 0$. λ poate fi orice număr real, pozitiv, negativ sau chiar zero.

Generalizarea la cazul mai multor restricții este imediată. Într-adevăr, pentru problema:

$$\begin{aligned} & \max f(x), \\ & \text{referitor la:} \\ & g_i(x) = 0, \quad i = 1, \dots, m \end{aligned} \tag{1.9}$$

domeniul de admisibilitate este intersecția a m hipersuprafețe. În acest caz, $\nabla f(x^*)$ trebuie să se afle în spațiul generat de gradientii restricțiilor active în x^* . Prin urmare, un punct critic trebuie să verifice următorul sistem:

$$\begin{aligned} g(x^*) &= 0, \\ \nabla f(x^*) &= \sum_{i=1}^m \lambda_i \nabla g_i(x^*) \end{aligned} \tag{1.10}$$

Ecuatiile (1.10) au fost obținute utilizând argumente pur geometrice, dar există o cale algebrică care conduce exact la aceleași rezultate. Pentru aceasta, vom utiliza funcția Lagrange:

$$L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i g_i(x) \tag{1.11}$$

pentru care vom căuta punctele ei critice în privința variabilelor x și λ . Deoarece aceasta este o problemă de optimizare fără restricții, punctele critice sunt obținute prin anularea derivatelor de ordin întâi:

$$\begin{aligned} \frac{\partial L}{\partial x_j} &= \frac{\partial f}{\partial x_j} - \sum_{i=1}^m \lambda_i \frac{\partial g_i}{\partial x_j} = 0, \quad j = 1, \dots, n, \\ \frac{\partial L}{\partial \lambda_i} &= -g_i = 0, \quad i = 1, \dots, m. \end{aligned} \tag{1.12}$$

Utilizând o notație matriceală, vedem că aceste ecuații sunt exact cele din (1.10) obținute din argumente geometrice. Faptul că o soluție a condițiilor de optimalitate de ordinul unu reprezintă într-adevăr un maxim global în care suntem interesați poate fi o problemă dificilă. Totuși, dacă restricțiile problemei sunt liniare, atunci este necesar să investigăm hessianul funcției de optimizat. Următoarea teoremă este relevantă în acest sens:

Teorema 1.1. Dacă restricțiile sunt liniare, un punct critic x^* este un maxim local restricționat pentru problema (1.9) dacă:

$$y^T \nabla^2 f(x^*) y < 0, \quad (1.13)$$

pentru orice vector y care satisface:

$$y^T \nabla g_i(x^*) = 0, \quad i = 1, \dots, m. \quad (1.14)$$

Demonstrație. Dezvoltând în serie Taylor funcția f în jurul lui x^* , obținem:

$$f(x^* + y) = f(x^*) + \nabla f(x^*)^T y + \frac{1}{2} y^T \nabla^2 f(x^*) y + o(\|y\|^2).$$

Vectorul y reprezintă deplasarea din punctul x^* . Dar, singurele deplasări relevante sunt cele care se află în domeniul de admisibilitate. Deci, fie y o direcție care verifică (1.14). Atunci, din (1.10) și (1.14) rezultă că $\nabla f(x^*)^T y = 0$, adică:

$$f(x^* + y) = f(x^*) + \frac{1}{2} y^T \nabla^2 f(x^*) y + o(\|y\|^2).$$

Din (1.13) rezultă imediat că x^* este un punct de maxim al problemei.

(q.e.d)

1.3. TRAIECTORIA CENTRALĂ

În această secțiune, vom utiliza rezultatele date de multiplicatorii Lagrange pentru a studia soluția dată de problema barieră. Vom arăta că, pentru orice valoare a parametrului barieră μ , problema barieră are soluție. Vom arăta că dacă μ tinde la zero, atunci soluția problemei barieră tinde la soluția problemei originale de programare liniară. În cursul acestor dezvoltări ne vom întâlni cu traiectoria centrală a problemei duale. Ecuțiile care definesc traiectoriile centrale primale și duale joacă un rol foarte important în definirea algoritmilor de punct interior, conducându-ne natural la noțiunea de traiectorie centrală primal-duală.

În continuare, considerăm problema barieră:

$$\max c^T x + \mu \sum_{j=1}^n \log(x_j) + \mu \sum_{i=1}^m \log(s_i),$$

referitor la :

$$Ax + s = b.$$

(1.15)

Aceasta este o problemă de optimizare cu restricții egalități. În vederea obținerii unei soluții a acestei probleme, putem aplica tehnica multiplicatorilor lui Lagrange, așa cum am procedat în secțiunea anterioară. Lagrangeanul asociat problemei barieră este:

$$L(x, s, y) = c^T x + \mu \sum_{j=1}^n \log(x_j) + \mu \sum_{i=1}^m \log(s_i) + y^T(b - Ax - s). \quad (1.16)$$

Condițiile necesare de ordinul întâi pe care trebuie să le satisfacă un punct staționar al relației (1.16) sunt următoarele:

$$\begin{aligned} \frac{\partial L}{\partial x_j} &= c_j + \frac{\mu}{x_j} - \sum_{i=1}^m y_i a_{ij} = 0, & j &= 1, \dots, n, \\ \frac{\partial L}{\partial s_i} &= \frac{\mu}{s_i} - y_i = 0 & i &= 1, \dots, m, \\ \frac{\partial L}{\partial y_i} &= b_i - \sum_{j=1}^n a_{ij} x_j - s_i = 0 & i &= 1, \dots, m. \end{aligned} \quad (1.17)$$

Introducând vectorul $z = \mu X^{-1} e$, condițiile de optimalitate (1.17) se rescriu sub forma:

$$\begin{aligned} Ax + s &= b, \\ A^T y - z &= c, \\ XZe &= \mu e, \\ YSe &= \mu e, \end{aligned} \quad (1.18)$$

unde X , Z , Y și S sunt matrice diagonale ale căror elemente sunt componentele vectorilor x , y , z , respectiv s . Se observă că prima relație din (1.18) reprezintă primal admisibilitatea, iar a doua admisibilitatea duală. De asemenea, ultimele două relații leagă variabilele primale și duale de variabilele ecart primale și respectiv duale. Dacă în (1.18) luăm $\mu = 0$, atunci vedem că ultimele două relații sunt chiar condițiile de complementaritate, motiv pentru care acestea sunt numite condițiile de μ -complementaritate.

Condițiile de optimalitate de ordinul unu (1.18) formează un sistem de $2n + 2m$ ecuații, cu același număr de necunoscute. Este evident că, dacă aceste ecuații erau liniare, atunci, utilizând de exemplu metoda de eliminare a lui Gauss, am fi putut rezolva ușor acest sistem, problema programării liniare reducându-se astfel la rezolvarea unui sistem de ecuații algebrice liniare. În acest sistem, singurele neliniarități sunt date de condițiile de μ -complementaritate care sunt simple înmulțiri ale variabilelor, de tipul $x_j z_j$ și respectiv $y_i s_i$. Aceste neliniarități fac ca subiectul programării liniare să nu fie trivial.

Se impune acum *existența și unicitatea* unei soluții pentru sistemul (1.18). Teorema următoare se referă la existența punctelor critice pentru acest sistem:

Teorema 1.2. Sistemul corespunzător condițiilor necesare de optimalitate de ordinul unu are un punct critic (o soluție) dacă și numai dacă domeniile de admisibilitate a problemei primale, respectiv a celei duale au un interior nevid.

Demonstrație. Implicația directă este evidentă și mai puțin importantă. Demonstrăm acum implicația inversă. Să presupunem că atât domeniul de admisibilitate al problemei primale cât și al celei duale au fiecare un interior nevid. Acest lucru înseamnă că există un punct admisibil primal (x^0, s^0) cu $x^0 > 0$, $s^0 > 0$, precum și un punct admisibil dual (y^0, z^0) cu $y^0 > 0$, $z^0 > 0$. Pentru orice punct admisibil primal (x, s) dat, considerăm expresia $z^{0T}x + y^{0T}s$. Din (1.18) obținem:

$$\begin{aligned} z^{0T}x + y^{0T}s &= (A^T y^0 - c)^T x + y^{0T}(b - Ax) \\ &= b^T y^0 - c^T x. \end{aligned}$$

Rezolvând această ecuație în privința funcției obiectiv $c^T x$, obținem:

$$c^T x = z^{0T}x - y^{0T}s + b^T y^0.$$

Deci, funcția barieră definită în (1.15) poate fi rescrisă sub forma:

$$\begin{aligned} f(x, s) &= c^T x + \mu \sum_{j=1}^n \log(x_j) + \mu \sum_{i=1}^m \log(s_i) \\ &= \sum_{j=1}^n (-z_j^0 x_j + \mu \log(x_j)) + \sum_{i=1}^m (-y_i^0 s_i + \mu \log(s_i)) + b^T y^0. \end{aligned}$$

Se observă că ultimul termen este o constantă. De asemenea, fiecare termen al celor două sume din relația de mai sus este o funcție de o singură variabilă. Toate aceste funcții au forma generală:

$$h(\xi) = -a\xi + \mu \log(\xi), \quad 0 < \xi < \infty,$$

unde $a > 0$. Aceste funcții au un maxim unic în punctul a/μ și acest maxim tinde la $-\infty$ când ξ tinde la $+\infty$. Pentru orice constantă c , mulțimea

$$\{(x, s) \in \mathbf{R}^n / f(x, s) \geq c\}$$

este mărginită. Fie acum $f^0 = f(x^0, s^0)$, și considerăm mulțimea:

$$P^0 = \{(x, s) / Ax + s = b, x \geq 0, s \geq 0, f(x, s) \geq f^0\}.$$

Este evident că P^0 este diferită de mulțimea vidă, deoarece conține punctul (x^0, s^0) . Mai mult, această mulțime este închisă și mărginită. În continuare, pentru a arăta că mulțimea P^0 este închisă, o exprimăm ca intersecția a trei mulțimi și anume:

$$\{(x, s) / Ax + s = b\} \cap \{(x, s) / x \geq 0, s \geq 0\} \cap \{(x, s) / f(x, s) \geq f^0\}.$$

Este evident că primele două mulțimi sunt închise. A treia mulțime este închisă pentru că este imaginea inversă a unei mulțimi închise $[f^0, \infty]$ prin funcția f , care este o funcție continuă. Deoarece intersecția a trei mulțimi închise este închisă, avem că P^0 este de asemenea închisă.

În spațiile euclidiene, o mulțime mărginită și închisă este un compact și cum Teorema lui Weierstrass ne asigură că o funcție continuă pe un compact nevid își atinge valoarea maximă, rezultă că există un punct în acest compact în care funcția ia valoarea maximă. Considerând acest

rezultat în cazul funcției f și a mulțimii P^0 , avem că f își atinge maximul pe P^0 , ceea ce implică faptul că își atinge maximul pe mulțimea $\{(x, s) / x > 0, s > 0\}$, deoarece P^0 , prin definiție, era acea submulțime a acestui domeniu pe care f ia valorile maxime.

(q.e.d)

Pentru a arăta unicitatea soluției sistemului (1.18), vom utiliza informațiile date de derivatele de ordinul doi ale funcției barieră:

$$f(x, s) = c^T x + \mu \sum_{j=1}^n \log(x_j) + \mu \sum_{i=1}^m \log(s_i).$$

Derivatele de ordinul întâi sunt următoarele:

$$\frac{\partial f}{\partial x_j} = c_j + \frac{\mu}{x_j}, \quad j = 1, \dots, n,$$

$$\frac{\partial f}{\partial s_i} = \frac{\mu}{s_i}, \quad i = 1, \dots, m.$$

Derivatele de ordinul doi sunt:

$$\frac{\partial^2 f}{\partial x_j^2} = -\frac{\mu}{x_j^2}, \quad j = 1, \dots, n,$$

$$\frac{\partial^2 f}{\partial s_i^2} = -\frac{\mu}{s_i^2}, \quad i = 1, \dots, m.$$

toate derivatele mixte fiind nule. Așadar, hessianul funcției barieră este o matrice diagonală cu elemente strict negative. Teorema 1.1 arată că sistemul (1.18) poate avea cel mult un punct critic, și dacă acesta există, el este un maxim global.

Concluzionând dezvoltările anterioare, vedem că dacă domeniul de admisibilitate primal și cel dual au fiecare un interior nevid, atunci pentru orice $\mu > 0$, sistemul (1.18) are o soluție unică $(x_\mu, s_\mu, y_\mu, z_\mu)$. Drumul definit de $\{(x_\mu, s_\mu, y_\mu, z_\mu) / \mu > 0\}$ se numește *traiectorie centrală primal-duală*. Aceasta are un rol foarte important în metodele de punct interior. Pe traiectoria centrală suntem centrați atât din punctul de vedere al hiperplanelor care definesc polipolul primal cât și al celor care definesc polipolul dual. Mai mult, pentru $\mu \rightarrow 0$ traiectoria centrală deplasează soluțiile din (1.18) către soluția optimă a problemei primale și a celei duale.

1.4. METODA PRIMAL-DUALĂ DE URMĂRIRE A TRAIECTORIEI PENTRU PROGRAMAREA LINIARĂ

Să considerăm problema de programare liniară:

$$\begin{aligned} & \min c^T x, \\ \text{referitor la:} & \\ & Ax = b, \\ & 0 \leq x \leq u, \end{aligned} \tag{1.19}$$

unde, câteva, sau toate marginile superioare, componente ale vectorului u , pot fi $+\infty$. Introducând variabilele ecart, problema devine:

$$\begin{aligned} & \min c^T x, \\ \text{referitor la:} & \\ & Ax = b, \\ & x + s = u, \\ & x \geq 0, s \geq 0. \end{aligned} \tag{1.20}$$

Eliminând restricțiile inegalități, prin intermediul funcției barieră logaritmice, din (1.20) obținem șirul de probleme:

$$\begin{aligned} & \max c^T x - \mu \sum_{j=1}^n \log(x_j) - \mu \sum_{j=1}^n \log(s_j), \\ \text{referitor la :} & \\ & Ax = b, \\ & x + s = u, \end{aligned} \tag{1.21}$$

parametrizat după valorile parametrului barieră $\mu > 0$. Utilizând formalismul dat de funcția Lagrange,

$$\begin{aligned} L(x, s, y, w, \mu) = & c^T x - \mu \sum_{j=1}^n \log(x_j) - \mu \sum_{j=1}^n \log(s_j) \\ & - y^T (Ax - b) - w^T (u - x - s), \end{aligned}$$

condițiile necesare de ordinul întâi pe care trebuie să le satisfacă un punct staționar al acestei funcții sunt:

$$Ax = b, \tag{1.22, a}$$

$$x + s = u,$$

$$A^T y + z - w = c, \tag{1.22, b}$$

$$XZe = \mu e,$$

$$SWe = \mu e, \quad (1.22, c)$$

unde X, Z, S, W sunt matrice diagonale cu elemente x_j, z_j, s_j și respectiv $w_j, j = 1, \dots, n$, iar $z \in \mathbb{R}^n$ este vectorul variabilelor duale corespunzător ecarturilor. Condițiile (1.22, a) mențin admisibilitatea primală, condiția (1.22, b) menține admisibilitatea duală, iar (1.22, c) furnizează relațiile de μ -complementaritate: $x_j z_j = \mu, s_j w_j = \mu$ pentru $j = 1, \dots, n$. De asemenea, pentru sistemul (1.22) se poate demonstra o teoremă care să asigure existența unui punct critic. Mai mult, hessianul funcției barieră din (1.21) este o matrice diagonală cu elemente strict pozitive, ceea ce asigură cu sistemul (1.22) poate avea cel mult un punct critic și, dacă acesta există, el este un minim global. Condiția esențială este ca domeniul de admisibilitate al problemei (1.19) să aibă un interior nevid.

Luând $\mu > 0$, sistemul (1.22) se poate rezolva în privința variabilelor x, s, y, z , și w . Cum acești vectori depind de alegerea parametrului barieră μ , vom obține o familie de soluții care definește traiectoria centrală a problemei (1.19).

Prin aplicarea unei iterații a metodei Newton, din condițiile de optimalitate (1.22) se obține algoritmul corespunzător metodei primal-duale de urmărire a traiectoriei. Să presupunem că dispunem de o estimăție (x, s, y, z, w) a variabilelor, care satisface următoarele condiții: $x > 0, s > 0, z > 0, w > 0$ și pentru care $x + s = u$, adică marginile superioare sunt în mod explicit satisfăcute. Această ipoteză de explicitare a variabilelor problemei nu restrânge cu nimic generalitatea metodei. Astfel, fixând o valoare finită și pozitivă pentru parametrul barieră μ și aplicând un pas al metodei Newton, din (1.22) obținem sistemul liniar:

$$\begin{aligned} A\Delta x &= b - Ax, \\ \Delta x + \Delta s &= 0, \\ A^T \Delta y + \Delta z - \Delta w &= c - A^T y - z + w, \\ Z\Delta x + X\Delta z &= \mu e - XZe, \\ W\Delta s + S\Delta w &= \mu e - SWe, \end{aligned} \quad (1.23)$$

unde $\Delta x, \Delta s, \Delta y, \Delta z$ și Δw sunt deplasările din punctul curent, care trebuie determinate ca soluții ale sistemului (1.23) cu care vom calcula noua estimăție a optimului problemei.

Notând:

$$\begin{aligned} \Theta &= (X^{-1}Z + S^{-1}W)^{-1} \\ \rho(\mu) &= \mu(S^{-1} - X^{-1})e - (W - Z)e, \end{aligned} \quad (1.24)$$

atunci soluția sistemului în deplasări (1.23) se poate explicita sub forma:

$$\begin{aligned} \Delta y &= (A\Theta A^T)^{-1}[(b - Ax) + A\Theta((c - A^T y - z + w) + \rho(\mu))], \\ \Delta x &= \Theta[A^T \Delta y - \rho(\mu) - (c - A^T y - z + w)], \end{aligned}$$

$$\Delta z = \mu X^{-1}e - Ze - X^{-1}Z\Delta x, \quad (1.25)$$

$$\Delta w = \mu S^{-1}e - We + S^{-1}W\Delta x,$$

$$\Delta s = -\Delta x.$$

O nouă estimăție $(x^+, s^+, y^+, z^+, w^+)$ a variabilelor este definită sub forma:

$$\begin{aligned} x^+ &= x + \alpha_p \Delta x, & y^+ &= y + \alpha_D \Delta y, \\ s^+ &= s + \alpha_p \Delta s, & z^+ &= z + \alpha_D \Delta z, \\ w^+ &= w + \alpha_D \Delta w, \end{aligned} \quad (1.26)$$

unde α_p și α_D sunt lungimile pașilor de deplasare din spațiul primal, respectiv dual, alese astfel

încât să se mențină pozitivitatea variabilelor $x, s, z,$ și w , adică $\alpha_p = \tau \hat{\alpha}_p$ și $\alpha_D = \tau \hat{\alpha}_D$, unde:

$$\begin{aligned} \hat{\alpha}_p &= \min \left\{ \min_j \left\{ \frac{x_j}{-\Delta x_j}, \Delta x_j < 0 \right\}, \min_j \left\{ \frac{s_j}{-\Delta s_j}, \Delta s_j < 0 \right\} \right\}, \\ \hat{\alpha}_D &= \min \left\{ \min_j \left\{ \frac{z_j}{-\Delta z_j}, \Delta z_j < 0 \right\}, \min_j \left\{ \frac{w_j}{-\Delta w_j}, \Delta w_j < 0 \right\} \right\}, \end{aligned} \quad (1.27)$$

și $\tau \in (0,1)$ este factorul de menținere a iterației curente în interiorul domeniului de admisibilitate. De obicei, $\tau = 0.9995$.

La fiecare iterație, parametrul barieră μ este redus și algoritmul continuă până când „saltul relativ de dualitate” satisface o condiție de oprire a iterațiilor:

$$\frac{|c^T x - b^T y + u^T w|}{1 + |b^T y - u^T w|} < \varepsilon, \quad (1.28)$$

unde $\varepsilon > 0$ este o toleranță prespecificată de oprire a iterațiilor. În continuare, putem prezenta următorul algoritm:

Algoritm PD (prima-dual)

(Descriere)

Pasul 1. Se consideră o aproximație inițială $t_0 = [x_0 \ s_0 \ y_0 \ z_0 \ w_0]^T$, unde $x_0 > 0$, $s_0 > 0$, $z_0 > 0$ și $w_0 > 0$, pentru care $x_0 + s_0 = u$. Se pune parametrul $k = 0$,

Pasul 2. Se testează criteriul (1.28) de oprire a iterațiilor,

Pasul 3. Se alege o valoare $\mu > 0$ a parametrului barieră și se rezolvă sistemul (1.23) obținându-se soluția $\Delta t = [\Delta x \ \Delta s \ \Delta y \ \Delta z \ \Delta w]^T$,

Pasul 4. Se alege $\tau_k \in (0, 1)$ și se pune $\alpha_p = \tau_k \hat{\alpha}_p$ și $\alpha_D = \tau_k \hat{\alpha}_D$, unde $\hat{\alpha}_p$ și $\hat{\alpha}_D$ sunt dați de (1.27),

Pasul 5. Se calculează noua aproximație τ_{k+1} ca în (1.26), se pune $k = k+1$ și se continuă cu pasul 2.

În scopul reactualizării parametrului μ , s-au propus mai multe formule. Parametrul μ are rolul unui parametru de admisibilitate, de „centrare”, obligând variabilele primale și pe cele duale de a nu se anula. Într-adevăr, dacă $\mu = 0$ și presupunem că componenta a i -a a lui x este nulă, adică $x_i = 0$ și aceeași componentă a lui z este nenulă, adică $z_i \neq 0$, atunci din (1.23) rezultă că la fiecare iterație $\Delta x_i = 0$.

Se observă că dacă estimăția curentă a soluției optime este primal și dual admisibilă, atunci Δt dat de algoritmul PD este o „direcție admisibilă”, adică verifică:

$$\begin{aligned} A\Delta x &= 0, \\ \Delta x + \Delta s &= 0, \\ A^T\Delta y + \Delta z - \Delta w &= 0. \end{aligned} \tag{1.29}$$

Propoziția 1.1. Pentru orice $(\Delta x, \Delta s, \Delta y, \Delta z, \Delta w)$ care verifică (1.29) are loc relația:
 $\Delta x^T\Delta z + \Delta s^T\Delta w = 0$.

Demonstrație. Luând $\Delta z = \Delta w - A^T\Delta y$ și $\Delta x = -\Delta s$ în (1.29), prin calcul direct se stabilește concluzia propoziției.

În afară de admisibilitate, singura condiție pentru optimalitate este ca saltul de dualitate să fie nul. Când soluția este admisibilă, complementaritatea $x^Tz + s^Tw$ este egală cu saltul de dualitate $c^Tx - b^Ty + u^Tw$. Pentru a reduce complementaritatea, continuăm iterațiile dintr-un punct admisibil. O alegere corespunzătoare a lui μ conduce la reducerea complementarității cu o rată liniară. Principala observație este că $x^T\Delta z + z^T\Delta x = n\mu - x^Tz$ și $s^T\Delta w + w^T\Delta s = n\mu - s^Tw$.

Teorema 1.3. Dacă se consideră lungimea comună α a pasului de deplasare, în spațiile primal și dual, atunci complementaritatea $g = x^Tz + s^Tw$ descrește liniar la fiecare iterație a algoritmului primal-dual de urmărire a traiectoriei.

Demonstrație. Prin calcul direct obținem:

$$\begin{aligned} g^+ &= (x + \alpha\Delta x)^T(z + \alpha\Delta z) + (s + \alpha\Delta s)^T(w + \alpha\Delta w) \\ &= x^Tz + s^Tw + \alpha[x^T\Delta z + s^T\Delta w + \Delta x^Tz + \Delta s^Tw] + \alpha^2[\Delta x^T\Delta z + \Delta s^T\Delta w] \\ &= x^Tz + s^Tw - \alpha[x^Tz + s^Tw - 2n\mu] + \alpha^2[\Delta x^T\Delta z + \Delta s^T\Delta w], \end{aligned}$$

Din Propoziția 1.1, se obține:

$$\begin{aligned} (x + \alpha\Delta x)^T(z + \alpha\Delta z) + (s + \alpha\Delta s)^T(w + \alpha\Delta w) &= \\ &= x^Tz + s^Tw - \alpha[x^Tz + s^Tw - 2n\mu], \end{aligned}$$

ceea ce demonstrează teorema.

(q.e.d)

În cazul în care lungimile pașilor în spațiile primal și dual sunt diferite, analiza este mai complicată. Un calcul simplu arată că dacă $\alpha_p \neq \alpha_D$, atunci saltul de dualitate g^+ conține un termen suplimentar care este fie variația funcției obiectiv primale, fie variația funcției obiectiv duale cu semn schimbat. Într-adevăr, dacă $\alpha_p < \alpha_D$, atunci:

$$\begin{aligned} & (x + \alpha_p \Delta x)^T(z + \alpha_D \Delta z) + (s + \alpha_p \Delta s)^T(w + \alpha_D \Delta w) = \\ & = x^T z + s^T w + \alpha_p [\Delta x^T z + \Delta s^T w] + \alpha_D [x^T \Delta z + s^T \Delta w] + \alpha_p \alpha_D [\Delta x^T \Delta z + \Delta s^T \Delta w] = \\ & = x^T z + s^T w - \alpha_p [x^T z + s^T w - 2n \mu] + (\alpha_D - \alpha_p)[x^T \Delta z + s^T \Delta w] + \\ & \quad \alpha_p \alpha_D [\Delta x^T \Delta z + \Delta s^T \Delta w]. \end{aligned}$$

Utilizând propoziția 1.1 pentru a elimina ultimul termen și observând că:

$$x^T \Delta z + s^T \Delta w = x^T [\Delta w - A^T \Delta y] + (u - x)^T \Delta w = -b^T \Delta y + u^T \Delta w,$$

rezultă că:

$$\begin{aligned} & (x + \alpha_p \Delta x)^T(z + \alpha_D \Delta z) + (s + \alpha_p \Delta s)^T(w + \alpha_D \Delta w) = \\ & = x^T z + s^T w - \alpha_p [x^T z + s^T w - 2n \mu] - (\alpha_D - \alpha_p)[b^T \Delta y - u^T \Delta w]. \end{aligned}$$

Analog, dacă $\alpha_D < \alpha_p$, atunci:

$$\begin{aligned} & (x + \alpha_p \Delta x)^T(z + \alpha_D \Delta z) + (s + \alpha_p \Delta s)^T(w + \alpha_D \Delta w) = \\ & = x^T z + s^T w - \alpha_D [x^T z + s^T w - 2n \mu] + (\alpha_p - \alpha_D)c^T \Delta x, \end{aligned}$$

deoarece $\Delta x^T z + \Delta s^T w = c^T \Delta x$.

Teorema 1.3, împreună cu analiza de mai înainte recomandă, pentru parametrul barieră μ , următoarea estimatie:

$$\mu = \sigma \frac{x^T z + s^T w}{2n}, \quad (1.30)$$

unde $\sigma \in (0, 1)$ este un parametru de modificare al lui μ . Cu această alegere a lui μ , următorul exemplu ilustrează funcționarea algoritmului PD.

Exemplul 1.1 (Algoritmul primal-dual de urmărire a traiectoriei pentru programarea liniară). Să considerăm programul liniar:

$$\min -19x_1 - 13x_2 - 12x_3 - 17x_4,$$

referitor la:

$$3x_1 + 2x_2 + x_3 + 2x_4 + x_5 = 225,$$

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 117,$$

$$4x_1 + 3x_2 + 3x_3 + 4x_4 + x_7 = 420,$$

$$0 \leq x_i \leq 50, \quad i = 1, \dots, 7.$$

Fie punctul inițial $x_0 = [25 \dots 25]^T$, $s_0 = [25 \dots 25]^T$, $y_0 = [333]^T$, $z_0 = [5 \dots 5]^T$ și

$w_0 = [2 \dots 2]^T$. Redau în tabelele 1.1 – 1.3 câteva rezultate numerice, reprezentând rezultatele furnizate de algoritmul PD [Andrei N., 1999].

În tabelul 1.1 este ilustrată evoluția variabilelor primale de-a lungul iterațiilor algoritmului PD. Se vede imediat convergența rapidă a algoritmului către soluția optimă. Evoluția variabilelor duale este redată în tabelul 1.2. Se observă că atât parametrul barieră cât și saltul de dualitate tind rapid la zero. Așa cum recomandă metoda Newton, lungimea pasului tinde la 1.

Tabelul 1.1. Evoluția variabilelor primale ($\sigma_k = 0.001$, $\tau_k = 0.99995$, $\varepsilon = 10^{-18}$)

Nr.	x(1)	x(2)	x(3)	x(4)	x(5)	x(6)	x(7)
0	25,0000	25,0000	25,0000	25,0000	25,0000	25,0000	25,0000
1	25,1073	22,0934	29,1793	30,1969	15,9177	13,9511	25,1181
2	25,9200	16,5043	32,5849	36,4371	8,7717	6,8083	16,2445
3	27,7064	7,1330	36,2178	43,8599	3,6776	2,4303	16,2445
4	29,7667	4,1925	37,0848	44,0836	2,0623	2,0083	0.8628E-4
5	32,2821	2,2266	38,4710	42,1314	0,9664	1,9336	0.4060E-3
6	37,7660	0,4281	46,3010	32,1743	0,1960	0,3395	0.2178E-3
7	38,9341	0.4799E-1	47,9648	30,0542	0.2806E-1	0.1698E-4	0.6357E-4
8	38,9993	0.2400E-5	48,0004	30,0002	0.9494E-3	0.1518E-4	0.7824E-5
9	39,0000	0.5207E-6	48,0000	30,0000	0.2883E-6	0.4199E-6	0.1801E-6
10	39,0000	0.1874E-9	48,0000	30,0000	0.761E-10	0.2510E-9	0.5836E-10
11	39,0000	0.112E-12	48,0000	30,0000	0.636E-13	0.8607E-13	0.3924E-13
12	39,0000	0.453E-16	48,0000	30,0000	0.183E-16	0.6084E-16	0.1409E-16

Tabelul 1.2. Evoluția variabilelor duale ($\sigma_k = 0.001$, $\tau_k = 0.99995$, $\varepsilon = 10^{-18}$)

Nr. iter	y(1)	y(2)	y(3)
0	3,0000	3,0000	3,0000
1	-1,2199	-1,7706	1,3561
2	-2,2739	-2,8921	-0,6076
3	-2,5151	-3,2680	-1,5530
4	-2,2060	-1,8932	-2,4233
5	-1,9545	-0,6984	-3,0410
6	-1,9800	-0,7447	-3,0623
7	-1,9956	-0,8619	-3,0350

8	-1,9998	-0,9928	-3,0018
9	-2,0000	-0,9999	-3,0000
10	-2,0000	-1,0000	-3,0000
11	-2,0000	-1,0000	-3,0000
12	-2,0000	-1,0000	-3,0000

Tabelul 1.3. Caracteristicile procesului de optimizare

$$(\sigma_k = 0.001, \tau_k = 0.99995, \varepsilon = 10^{-18})$$

Nr. iter	J_P	J_D	μ	g	α
1	-1525,000	1586,000	0.8750E-1	1225,000	0,5589
2	-1627,755	-167,7183	0.3509E-1	491,3702	0,6443
3	-1717,489	-1204,260	0.1276E-1	178,7729	0,7228
4	-1799,387	-1658,325	0.3601E-2	50,4220	0,6084
5	-1814,513	-1761,171	0.1493E-2	20,9034	0,6705
6	-1820,193	-1803,210	0.5206E-3	7,2890	0,7982
7	-1825,696	-1822,259	0.1945E-3	1,4630	0,8521
8	-1826,874	-1826,364	0.1533E-4	0,2147	0,9648
9	-1826,997	-1826,979	0.5240E-6	0.7337E-2	0,9999
10	-1827,000	-1827,000	0.5158E-9	0.7221E-5	1,0006
11	-1827,000	-1827,000	0.1906E-12	0.2669E-8	1,0004
12	-1827,000	-1827,000	0.1112E-15	0.1557E-11	1,0005

J_P - valoarea funcției obiectiv primale;

J_D - valoarea funcției obiectiv duale;

g - saltul de dualitate;

α - $\min(\alpha_P, \alpha_D)$, lungimea pasului.

Analiza algoritmului

În continuare, considerăm aceeași lungime α a pasului de deplasare, atât în spațiul primal, cât și în cel dual, adică la fiecare iterație a algoritmului, $\alpha_P = \alpha_D = \alpha$.

În scopul determinării progresului efectuat de algoritmul PD la fiecare iterație, avem nevoie de niște criterii care să cuantifice evoluția algoritmului. Din teoria dualității, cunoaștem că o soluție primal-duală este optimă, dacă verifică următoarele trei criterii:

- a) admisibilitatea primală;
- b) admisibilitatea duală;
- c) complementaritatea.

Din (1.23), se observă că pentru admisibilitatea primală putem utiliza vectorul infeasibilității primale $r_p = b - Ax$, pentru admisibilitatea duală putem utiliza vectorul infeasibilității duale $r_D = c - A^T y - z + w$, iar pentru complementaritate vectorul $g = x^T z + s^T w$, saltul de dualitate.

Din (1.26) se observă că la fiecare iterație, evoluția primelor două criterii este următoarea:

$$\begin{aligned} r_p^+ &= b - Ax^+ = b - A(x + \alpha \Delta x) = b - Ax - \alpha A \Delta x = (I - \alpha) r_p, \\ r_D^+ &= c - A^T y^+ - z^+ + w^+ = c - A^T (y + \alpha \Delta y) - (z + \alpha \Delta z) + (w + \alpha \Delta w) = \\ &= r_D - \alpha (A^T \Delta y - \Delta z + \Delta w) = (I - \alpha) r_D. \end{aligned}$$

Cum $\alpha \in (0, 1)$, la fiecare iterație, algoritmul PD produce o reducere a infeasibilității primale și duale, și aceasta este cu atât mai pronunțată cu cât α este mai aproape de unu, așa cum recomandă metoda Newton pentru rezolvarea sistemului neliniar care definește traiectoria centrală. Vom obține:

$$\begin{aligned} g^+ &= x^{+T} z^+ + s^{+T} w^+ = (x + \alpha \Delta x)^T (z + \alpha \Delta z) + (s + \alpha \Delta s)^T (w + \alpha \Delta w) = \\ &= x^T z + s^T w + \alpha [x^T \Delta z + \Delta x^T z + s^T \Delta w + \Delta s^T w] + \alpha^2 [\Delta x^T \Delta z + \Delta s^T \Delta w]. \end{aligned}$$

Dar,

$$\begin{aligned} x^T \Delta z + \Delta x^T z &= e^T (X \Delta z + Z \Delta x) = e^T (\mu - X Z e) = n \mu - x^T z, \\ s^T \Delta w + \Delta s^T w &= e^T (S \Delta w + W \Delta s) = e^T (\mu e - S W e) = n \mu - s^T w, \end{aligned}$$

și

$$\begin{aligned} \Delta x^T \Delta z + \Delta s^T \Delta w &= \Delta x^T \Delta z - \Delta x^T \Delta w = \Delta x^T (\Delta z - \Delta w) = \Delta x^T (r_D - A^T \Delta y) \\ &= \Delta x^T r_D - (A \Delta x)^T \Delta y = \Delta x^T r_D - r_p^T \Delta y. \end{aligned}$$

Înlocuind aceste expresii în relația de definiție a lui g^+ , obținem:

$$g^+ = g + \alpha (2n\mu - g) + \alpha^2 (\Delta x^T r_D - r_p^T \Delta y).$$

Dar, $2n\mu = \sigma g$, deci

$$g^+ = (1 - (1 - \sigma)\alpha)g + \alpha^2 (\Delta x^T r_D - r_p^T \Delta y).$$

În acest moment, abandonăm egalitățile și lucrăm cu estimări. Pentru orice doi vectori $a, b \in \mathbb{R}^n$ avem:

$$|a^T b| = \left| \sum_j a_j b_j \right| \leq \sum_j |a_j| |b_j| \leq (\max_j |a_j|) (\sum_j |b_j|) = \|a\|_\infty \|b\|_1.$$

Cu aceasta obținem:

$$\begin{aligned} |\Delta x^T r_D| &\leq \|r_D\|_1 \|\Delta x\|_\infty, \\ |r_P^T \Delta y| &\leq \|r_P\|_1 \|\Delta y\|_\infty, \end{aligned}$$

Deci,

$$g^+ \leq (1 - (1 - \sigma)\alpha)g + \alpha[\|r_P\|_1 \|\alpha\Delta y\|_\infty + \|r_D\|_1 \|\alpha\Delta x\|_\infty].$$

Dar, $\|\alpha\Delta x\|_\infty \leq \|x\|_\infty$ și $\|\alpha\Delta y\|_\infty \leq \|y\|_\infty$. Dacă presupunem că $\|x\|_\infty$ și $\|y\|_\infty$ sunt mărginite de un număr real suficient de mare M , atunci o estimăție a complementarității la următoarea iterație este dată de:

$$g^+ \leq (1 - (1 - \sigma)\alpha)g + M\|r_P\|_1 + M\|r_D\|_1.$$

În continuare, să vedem progresul înregistrat de algoritmul primal-dual după un număr de iterații.

Fie r_P^k , r_D^k , g^k și α^k valorile acestor cantități la iterația k .

Teorema 1.4. Presupunem că există numerele reale $\bar{\alpha} > 0$ și $M < \infty$, precum și un întreg K astfel încât, pentru orice $k \leq K$:

$$\begin{aligned} \alpha^k &\geq \bar{\alpha}, \\ \|x^k\|_\infty &\leq M, \\ \|y^k\|_\infty &\leq M. \end{aligned}$$

Atunci există o constantă $\bar{M} < \infty$ astfel încât:

$$\begin{aligned} \|r_P^k\|_1 &\leq (1 - \bar{\alpha})^k \|r_P^0\|_1, \\ \|r_D^k\|_1 &\leq (1 - \bar{\alpha})^k \|r_D^0\|_1, \\ g^k &\leq (1 - \bar{\alpha})^k \bar{M}, \end{aligned}$$

pentru orice $k \leq K$, unde $\bar{\alpha} = (1 - \sigma)\bar{\alpha}$.

Demonstrație. Din relațiile de mai sus, notând faptul că pentru fiecare iterație α^k este mărginit, pentru infeasibilitatea primală și duală obținem imediat:

$$\begin{aligned} \|r_P^k\|_1 &\leq (1 - \bar{\alpha})^k \|r_P^{k-1}\|_1 \leq \dots \leq (1 - \bar{\alpha})^k \|r_P^0\|_1, \\ \|r_D^k\|_1 &\leq (1 - \bar{\alpha})^k \|r_D^{k-1}\|_1 \leq \dots \leq (1 - \bar{\alpha})^k \|r_D^0\|_1. \end{aligned}$$

Pentru g^k obținem:

$$\begin{aligned} g^k &\leq (1 - (1 - \sigma)\bar{\alpha})g^{k-1} + M(1 - \bar{\alpha})^{k-1} (\|r_P^0\|_1 + \|r_D^0\|_1) \\ &= (1 - \bar{\alpha})g^{k-1} + \bar{M}(1 - \bar{\alpha})^{k-1}, \end{aligned}$$

unde $\bar{M} = M(\|r_p^0\| + \|r_D^0\|)$. Deoarece o inegalitate similară leagă g^{k-1} de g^{k-2} , putem scrie:

$$\begin{aligned} g^k &\leq (1 - \underline{\alpha})[(1 - \underline{\alpha})g^{k-2} + \bar{M}(1 - \bar{\alpha})^{k-2}] + \bar{M}(1 - \bar{\alpha})^{k-1} \\ &= (1 - \underline{\alpha})^2 g^{k-2} + \bar{M}(1 - \bar{\alpha})^{k-1} \left[\frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} + 1 \right]. \end{aligned}$$

Continuând în acest fel, obținem:

$$\begin{aligned} g^k &\leq (1 - \underline{\alpha})^2 [(1 - \underline{\alpha})g^{k-3} + \bar{M}(1 - \bar{\alpha})^{k-3}] + \bar{M}(1 - \bar{\alpha})^{k-1} \left[\frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} + 1 \right] \\ &= (1 - \underline{\alpha})^3 g^{k-3} + \bar{M}(1 - \bar{\alpha})^{k-1} \left[\left(\frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} \right)^2 + \frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} + 1 \right] \leq \dots \leq \\ &(1 - \underline{\alpha})^k g^0 + \bar{M}(1 - \bar{\alpha})^{k-1} \left[\left(\frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} \right)^{k-1} + \dots + \frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} + 1 \right]. \end{aligned}$$

Dar,

$$(1 - \bar{\alpha})^{k-1} \left[\left(\frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} \right)^{k-1} + \dots + \frac{1 - \underline{\alpha}}{1 - \bar{\alpha}} + 1 \right] = \frac{(1 - \underline{\alpha})^k - (1 - \bar{\alpha})^k}{\bar{\alpha} - \underline{\alpha}} \leq \frac{(1 - \underline{\alpha})^k}{\sigma \alpha}.$$

Deci,

$$g^k \leq (1 - \underline{\alpha})^k \left[g^0 + \frac{\bar{M}}{\sigma \alpha} \right].$$

Notând expresia din paranteză cu \bar{M} , obținem concluzia teoremei.

(q.e.d)

Teorema 1.4 furnizează numai un rezultat parțial asupra convergenței algoritmului primal-dual, deoarece aceasta depinde de presupunerea că la fiecare iterație lungimile pașilor rămân mărginite față de zero.

Dacă estimăția curentă a soluției este admisibilă, atunci complementaritatea $g = x^T z + s^T w$ este egală cu saltul de dualitate $c^T x - b^T y + u^T w$. Cazul în care soluția curentă nu este admisibilă ne permite să completăm algoritmul PD prin precizarea criteriilor de oprire a iterațiilor. Ținând cont de definiția infeasibilității duale, obținem:

$$g = x^T z + s^T w = x^T (c - A^T y + w - r_D) + (u - x)^T w.$$

După unele calcule algebrice, rezultă:

$$g = (c^T x - b^T y + u^T w) + r_p^T y - r_D^T x \quad (1.31)$$

Mai departe vom utiliza inegalitatea lui Hölder pentru a mărgini termenii din (1.31), cu scopul de a obține o estimare a saltului de dualitate:

$$\begin{aligned} |c^T x - b^T y + u^T w| &\leq g + |r_p^T y| + |r_D^T x| \leq \\ &\leq g + \|r_p\|_1 \|y\|_\infty + \|r_D\|_1 \|x\|_\infty \end{aligned} \quad (1.32)$$

Acum, dacă g , $\|r_p\|_1$ și $\|r_D\|_1$ sunt mici (și $\|x\|_\infty$ și $\|y\|_\infty$ nu sunt prea mari), atunci saltul de dualitate va fi mic. Această estimare ne arată că saltul de dualitate devine mic de îndată ce soluțiile primală și duală devin „aproape admisibile”.

Algoritmul PD poate fi completat în sensul următor. Fie $\varepsilon > 0$ o toleranță suficient de mică și $M < \infty$ o toleranță suficient de mare. Dacă $\|x\|_\infty > M$, atunci algoritmul se oprește declarând problema primală nemărginită. Dacă $\|y\|_\infty > M$, atunci din nou algoritmul se oprește declarând problema duală nemărginită. În final, dacă $g < \varepsilon$, $\|r_p\|_1 < \varepsilon$ și $\|r_D\|_1 < \varepsilon$, atunci algoritmul se oprește declarând soluția curentă optimă (cel puțin în toleranțele specificate).

Observația 1.1. Membru drept al sistemului (1.23) se poate scrie ca o sumă de doi termeni, dintre care unul depinde de parametrul barieră μ . Astfel, pasul 3 al algoritmului PD se descompune în următorii subpași:

Pasul 3.1. Se rezolvă sistemul liniar:

$$\begin{aligned} A\Delta x &= b - Ax, \\ \Delta x + \Delta s &= 0, \\ A^T\Delta y + \Delta z - \Delta w &= c - A^T y - z + w, \\ Z\Delta x + X\Delta z &= -XZe, \\ W\Delta s + S\Delta w &= -SWe, \end{aligned} \quad (1.33)$$

Obținându-se soluția $\Delta t_a = [\Delta x_a \ \Delta s_a \ \Delta y_a \ \Delta z_a \ \Delta w_a]^T$.

Pasul 3.2. Se alege o valoare $\mu > 0$ a parametrului barieră și se rezolvă sistemul liniar:

$$\begin{aligned} A\Delta x &= 0, \\ \Delta x + \Delta s &= 0, \\ A^T\Delta y + \Delta z - \Delta w &= 0, \\ Z\Delta x + X\Delta z &= \mu e, \\ W\Delta s + S\Delta w &= \mu e, \end{aligned} \quad (1.34)$$

obținându-se soluția $\Delta t_c = [\Delta x_c \ \Delta s_c \ \Delta y_c \ \Delta z_c \ \Delta w_c]^T$.

Pasul 3.3. Se consideră: $\Delta t = \Delta t_a + \Delta t_c$.

Pasul Newton Δt_a , definit în pasul 3.1 de mai sus poate fi direcționat către frontiera cadranelui pozitiv, implicând astfel o lungime foarte mică a pasului de deplasare α . Pasul de centrare Δt_c definit în pasul 3.2 are rolul de a remedia această situație. Astfel, pare imediată alegerea parametrului μ_k ca o funcție de pasul Newton Δt_a . Din punct de vedere matematic, conceptele de *polinomialitate* și *rată de convergență* sunt incompatibile. Polinomialitatea are sens numai pentru algoritmi care se termină într-un număr finit de pași, în timp ce rata de convergență este definită numai pentru algoritmi care se efectuează în număr infinit de pași. Prin algoritm de punct interior polinomial se înțelege că datele problemei sunt întregi (sau raționale) și că acesta se termină într-un număr finit de pași, iar prin algoritm de punct interior liniar convergent se înțelege că datele problemei sunt date de relația [Andrei, 1999]:

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \beta < 1.$$

Cu aceste precizări se poate discuta simultan atât polinomialitatea cât și rata de convergență a unui algoritm.

1.5. METODA PREDICTOR-CORECTOR DE URMĂRIRE A TRAIECTORIEI PENTRU PROGRAMAREA LINIARĂ

Algoritmul corespunzător acestei metode, pentru rezolvarea problemei (1.19) utilizează aceeași funcție barieră logaritmică pentru a obține condițiile necesare de optimalitate de ordinul unu date de sistemul (1.22). În cazul acesta, presupunând că dispunem de o estimăție curentă $t = [x \ s \ y \ z \ w]^T$ a variabilelor care satisfac condiția: $x > 0$, $s > 0$, $z > 0$ și $w > 0$, în loc de a aplica o iterație a metodei Newton pentru a determina Δt , vom substitui direc noul punc $t + \Delta t$ în (1.22), obținând sistemul:

$$\begin{aligned} A(x + \Delta x) &= b, \\ (x + \Delta x) + (s + \Delta s) &= u, \\ A^T(y + \Delta y) + (z + \Delta z) - (w + \Delta w) &= c, \\ (X + \Delta X)(Z + \Delta Z)e &= \mu e, \\ (S + \Delta S)(W + \Delta W)e &= \mu e, \end{aligned} \tag{1.35}$$

unde ΔX , ΔZ , ΔS , și ΔW sunt matrice diagonale cu elemente Δx_i , Δz_i , Δs_i și respectiv, Δw_i ,

($i = 1, \dots, n$). Sistemul (1.35) devine:

$$\begin{aligned}
A\Delta x &= b - Ax, \\
\Delta x + \Delta s &= u - x - s, \\
A^T\Delta y + \Delta z - \Delta w &= c - A^T y - z + w, \\
Z\Delta x + X\Delta z &= \mu e - XZe - \Delta X\Delta Ze, \\
W\Delta s + S\Delta w &= \mu e - SWe - \Delta S\Delta We.
\end{aligned} \tag{1.36}$$

Se observă că membrul stâng al sistemului (1.36) este identic cu cel din (1.23), dar termenii liberi sunt diferiți. Mai întâi, ecuația care implică marginile superioare asupra variabilelor din (1.36) arată că condițiile de mărginire sunt tratate la fel ca și condițiile de admisibilitate primală și duală. Apoi, prezența termenilor neliniari $\Delta X\Delta Ze$ și $\Delta S\Delta We$, situați în membrul drept în (1.36), constituie diferența majoră dintre (1.23) și (1.36).

Ideea metodei predictor-corector de urmărire a traiectoriei, expusă pentru prima dată de Mehrota [1992], care s-a inspirat din teoria rezolvării ecuațiilor diferențiale, este de a rezolva sistemul (1.36) în două etape.

În prima etapă, „*de predicție*”, se consideră $\mu = 0$ și se ignoră termenii pătratici $\Delta X\Delta Ze$ și $\Delta S\Delta We$, rezolvându-se sistemul algebric liniar:

$$\begin{aligned}
A\Delta x &= b - Ax, \\
\Delta x + \Delta s &= u - x - s, \\
A^T\Delta y + \Delta z - \Delta w &= c - A^T y - z + w, \\
Z\Delta x + X\Delta z &= -XZe, \\
W\Delta s + S\Delta w &= -SWe.
\end{aligned} \tag{1.37}$$

obținându-se soluția $\Delta t_a = [\Delta x_a \ \Delta s_a \ \Delta y_a \ \Delta z_a \ \Delta w_a]^T$. Observăm că, în afara termenului liber din a doua ecuație, care se referă la condițiile de mărginire a variabilelor, sistemul (1.37) este identic cu cel din (1.33). Δt_a este *direcția primal-duală afină*, fiind responsabilă în principal cu minimizarea funcției obiectiv.

În a doua etapă, etapa „*de corecție*”, direcția afină este utilizată pentru:

1. Aproximarea termenilor neliniari din membrul drept din (1.36).
2. Calculul unei noi estimări a parametrului barieră μ .

Pentru estimarea lui μ , Mehrota propune executarea testului rapoartelor standard asupra variabilelor primale și duale, pentru a determina lungimea pasului care ar rezulta dacă se utilizează direcția primal-duală afină definită de (1.37). Astfel, calculăm:

$$\bar{\delta}_p = \min \left\{ \min_j \left\{ \frac{x_j}{-\Delta x_{aj}}, \Delta x_{aj} < 0 \right\}, \min_j \left\{ \frac{s_j}{-\Delta s_{aj}}, \Delta s_{aj} < 0 \right\} \right\}, \quad (1.38)$$

$$\bar{\delta}_D = \min \left\{ \min_j \left\{ \frac{z_j}{-\Delta z_{aj}}, \Delta z_{aj} < 0 \right\}, \min_j \left\{ \frac{w_j}{-\Delta w_{aj}}, \Delta w_{aj} < 0 \right\} \right\}$$

și considerăm: $\delta_p = \tau \bar{\delta}_p$ și $\delta_D = \tau \bar{\delta}_D$, unde $\tau \in (0,1)$ este un parametru (de obicei $\tau = 0,99995$) de atenuare a blocării (anulării unei variabile).

Valoarea complementarității care ar rezulta dacă am considera un pas în direcția afină Δt_a este:

$$\bar{g} = (x + \delta_p \Delta x_a)^T (z + \delta_D \Delta z_a) + (s + \delta_p \Delta s_a)(w + \delta_D \Delta w_a) \quad (1.39)$$

Astfel, estimația lui Mehrota [1990], generalizată de Lustig, Marsten și Shanno [1992] pentru a include și margini simple asupra variabilelor este:

$$\mu = \left(\frac{\bar{g}}{x^T z + s^T w} \right)^2 \left(\frac{\bar{g}}{n} \right), \quad (1.40)$$

care, după cum se vede, furnizează o valoare mică pentru μ , când direcția afină îmbunătățește valoarea complementarității și o valoare mare, în caz contrar. Această comportare a formulei (1.40) de definire a lui μ este foarte convenabilă, deoarece o îmbunătățire nesemnificativă a complementarității în direcția afină indică necesitatea unei centrări mai pronunțate, adică a unei valori mai mari pentru μ .

Cu această valoare a parametrului barieră μ se rezolvă următorul sistem algebric liniar:

$$\begin{aligned} A\Delta x &= 0, \\ \Delta x + \Delta s &= 0, \\ A^T \Delta y + \Delta z - \Delta w &= 0, \\ Z\Delta x + X\Delta z &= \mu e - \Delta X_a \Delta Z_a e, \\ W\Delta s + S\Delta w &= \mu e - \Delta S_a \Delta W_a e, \end{aligned} \quad (1.41)$$

care furnizează așa-numita *direcție de centrare* $\Delta t_c = [\Delta x_c \ \Delta s_c \ \Delta y_c \ \Delta z_c \ \Delta w_c]^T$. Astfel, direcția afină Δt_a este combinată cu o direcție de centrare Δt_c , pentru a se obține o nouă estimație a variabilelor, ca în:

Algoritm PC (predictor-corector)

(Descriere)

Pasul 1. Se consideră o aproximație inițială $t_0 = [x_0 \ s_0 \ y_0 \ z_0 \ w_0]^T$, unde $x_0 > 0$, $s_0 > 0$, $z_0 > 0$ și $w_0 > 0$. Se pune parametrul $k = 0$,

Pasul 2. Se testează criteriul (1.28) de oprire a iterațiilor,

Pasul 3. Se rezolvă sistemul algebric liniar (1.37) în privința direcției afine Δt_a ,

Pasul 4. Se alege $\sigma_k \in [0, 1)$ și se calculează estimăția parametrului barieră μ conform relației (1.40). Se pune $\mu = \sigma_k \mu$,

Pasul 5. Se rezolvă sistemul algebric liniar (1.41) în privința direcției de centrare Δt_c ,

$$\begin{aligned} \text{Pasul 6. Se pune } \Delta x &= \Delta x_a + \Delta x_c, & \Delta y &= \Delta y_a + \Delta y_c, \\ \Delta s &= \Delta s_a + \Delta s_c, & \Delta z &= \Delta z_a + \Delta z_c, \\ & & \Delta w &= \Delta w_a + \Delta w_c \end{aligned}$$

Pasul 7. Se alege $\tau_k \in (0, 1)$ și se pune $\alpha_P = \tau_k \bar{\alpha}_P$ și $\alpha_D = \tau_k \bar{\alpha}_D$, unde $\bar{\alpha}_P$ și $\bar{\alpha}_D$ sunt dați de (1.27).

Pasul 8. Se consideră noua estimăție τ_{k+1} a variabilelor, unde:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_P \Delta x, & y_{k+1} &= y_k + \alpha_D \Delta y, \\ s_{k+1} &= s_k + \alpha_P \Delta s, & z_{k+1} &= z_k + \alpha_D \Delta z, \\ & & w_{k+1} &= w_k + \alpha_D \Delta w, \end{aligned}$$

se pune $k = k + 1$ și se continuă cu pasul 2.

Astfel, o dată ce matrice $A\Theta A^T$ a fost factorizată, calculele suplimentare implicate de algoritmul PC constă în aplicarea unei substituții, pentru determinarea direcției de centrare, precum și calculele rapoartelor (1.38), pentru obținerea unei estimății a lui μ .

Avantajul important oferit de această metodă constă în utilizarea informației de ordinul doi asupra traiectoriei centrale.

Exemplul 1.2. Se consideră același program liniar din exemplul (1.1). Algoritmul predictor-corector, inițializat în același punct, furnizează rezultatele din tabelele 1.4-1.6 [Andrei, 1999]:

Tabelul 1.4. Evoluția variabilelor primale ($\sigma_k = 0.001$, $\tau_k = 0.99995$, $\varepsilon = 10^{-18}$)

Nr.	x(1)	x(2)	x(3)	x(4)	x(5)	x(6)	x(7)
0	25,0000	25,0000	25,0000	25,0000	25,0000	25,0000	25,0000
1	25,8621	18,6251	31,3569	30,1969	10,7024	9,2016	18,5969
2	29,7838	4,2163	36,6760	36,4371	1,9361	2,1962	0.3169E-3
3	33,5339	1,2938	39,1842	43,8599	0,5027	1,9830	0.1288E-4
4	38,8970	0.8756E-1	47,9452	44,0836	0.4136E-1	0.1185E-3	0.1697E-5

5	38,9997	0.1749E-3	47,9997	42,1314	0.4780E4	0.1882E-4	0.5136E-5
6	39,0000	0.114E-10	48,0000	30,0000	0.4363E-8	0.1717E-7	0.5849E-8
7	39,0000	0.338E-11	48,0000	30,0000	0.924E-12	0.363E-12	0.9920E-13
8	39,0000	0.403E-14	48,0000	30,0000	0.843E-16	0.322E-15	0.1131E-15
9	39,0000	0.653E-19	48,0000	30,0000	0.178E-19	0.702E-20	0.1918E-20

Tabelul 1.5. Evoluția variabilelor duale ($\sigma_k = 0.001$, $\tau_k = 0.99995$, $\varepsilon = 10^{-18}$)

Nr. iter	y(1)	y(2)	y(3)
0	3,0000	3,0000	3,0000
1	-0,7950	-0,9372	-1,0499
2	-2,2001	-2,3522	-2,2475
3	-1,9316	-0,6876	-3,0680
4	-1,9876	-0,8262	-3,0456
5	-1,9999	-0,9999	-3,0000
6	-2,0000	-1,0000	-3,0000
7	-2,0000	-1,0000	-3,0000
8	-2,0000	-1,0000	-3,0000
9	-2,0000	-1,0000	-3,0000

Tabelul 1.6. Caracteristicile procesului de optimizare

($\sigma_k = 0.001$, $\tau_k = 0.99995$, $\varepsilon = 10^{-18}$)

Nr. iter	J_P	J_D	μ	g	α
1	-1525,000	-974,967	0.5139E-2	1225,000	0,7377
2	-1627,755	-1759,462	0.1006E-3	427,270	0,9169
3	-1717,489	-1809,888	0.3943E-4	40,773	0,7768
4	-1799,387	-1825,698	0.1614E-5	8,825	0,9124
5	-1827,000	-1826,999	0.1566E-4	0,767	0,9990
6	-1827,000	-1827,000	0.1924E-7	0.9429E-3	1,0000
7	-1827,000	-1827,000	0.3381E-11	0.1656E-6	1,0000
8	-1827,000	-1827,000	0.3720E-15	0.1822E-10	1,0000
9	-1827,000	-1827,000	0.6537E-19	0.3203E-14	1,0000

J_P - valoarea funcției obiectiv primale;

J_D - valoarea funcției obiectiv duale;

g - saltul de dualitate;

α - $\min(\alpha_p, \alpha_D)$, lungimea pasului.

Analiza convergenței

Notăm $t = [x \ s \ y \ z \ w]^T$ și considerăm următoarea transformare:

$$F(t) = \begin{bmatrix} Ax - b \\ x + s - u \\ A^T y + z - w - c \\ XZe \\ SWe \end{bmatrix} \quad (1.42)$$

Fie totodată vectorul $\hat{e} = [0 \dots 0 \ 1 \dots 1]^T \in \mathbb{R}^{4n+m}$, unde numărul componentelor egale cu 1 din \hat{e} este egal cu $2n$.

Este clar că metoda Newton „pură” poate fi rescrisă sub forma:

$$N(t) = t - F'(t)^{-1} F(t). \quad (1.43)$$

Pe de altă parte, metoda predictor-corector se scrie sub forma:

$$\bar{N}(t) = t - \alpha F'(t)^{-1} [F(t) + F(N(t)) - \mu \hat{e}] \quad (1.44)$$

Deci,

$$\begin{aligned} \bar{N}(t) - t^* &= t - t^* - F'(t)^{-1} [F(t) + F(N(t))] + \\ &+ (1 - \alpha) F'(t)^{-1} [F(t) + F(N(t))] + \alpha \mu F'(t)^{-1} \hat{e} = \\ &= F'(t)^{-1} [F'(t)(N(t)) - F(N(t)) - F'(t)t^*] + \\ &+ (1 - \alpha) F'(t)^{-1} [F(t) + F(N(t))] + \alpha \mu F'(t)^{-1} \hat{e} = \\ &= F'(t)^{-1} [F(N(t)) + F(t^*) - F'(t^*)(N(t) - t^*)] + \\ &+ F'(t)^{-1} [F'(t^*) - F'(t) - F'(t^*)](N(t) - t^*) + \\ &+ (1 - \alpha) F'(t)^{-1} [F(t) + F(N(t))] + \alpha \mu F'(t)^{-1} \hat{e}. \end{aligned} \quad (1.45)$$

Acum, local, adică în vecinătatea soluției optime t^* , din analiza standard a metodei Newton [Andrei, 1999], rezultă că:

$$\|N(t) - t^*\| = O(\|t - t^*\|^2) \quad (1.46)$$

Dar, $\|F(t)\| = O(\|t - t^*\|)$ și $\|F(N(t))\| = O(\|t - t^*\|^2)$. Deci, din (1.45) obținem imediat:

$$\|\bar{N}(t) - t^*\| = O(\|t - t^*\|^4) + O(\|t - t^*\|^3) + |1 - \alpha| O(\|t - t^*\|) + \mu O(1) \quad (1.47)$$

care se simplifică la:

$$\|\bar{N}(t) - t^*\| = O(\|t - t^*\|^3) + |1 - \alpha| O(\|t - t^*\|) + \mu O(1) \quad (1.48)$$

Evident, alegând μ corespunzător, termenul $\mu O(1)$ poate fi făcut la mărimea lui $O(\|t - t^*\|^3)$.

Totul depinde acum de $|1 - \alpha| O(\|t - t^*\|)$. Deci, trebuie să analizăm cantitatea $1 - \alpha$. Pentru convergența cubică a metodei predictor-corrector trebuie ca mărimea $1 - \alpha$ să fie de mărimea $O(\|t - t^*\|^2)$.

Presupunând că soluția curentă satisface strict relațiile de complementaritate și că t^* este un vârf nedegenerat pentru problema fără restricții simple de mărginire a variabilelor, Zhang, Tapia și Dennis [1992], pentru metoda primal-duală, au obținut următoarea evaluare:

$$1 - \alpha_k = \frac{1 - \tau_k + \sigma_k \theta_k}{1 - \sigma_k \theta_k} + O(x_k^T z_k) \quad (1.49)$$

unde τ_k și σ_k sunt definiți ca în algoritmul PC, iar $\theta_k \in [1/n, 1]$. Se observă că

$$O(x_k^T z_k) = O(\|t - t^*\|), \text{ deoarece pentru } \tau_k \text{ admisibil } x_k^T z_k = \|F(t_k)\|_1.$$

Din punctul de vedere al cantității $1 - \alpha_k$, metodele de punct interior primal-duală și predictor-corrector de urmărire a traiectoriei sunt identice, adică acestea pot fi privite ca *perturbații ale metodei Newton* aplicate sistemului $F(t) = 0$ din (1.42). În primul caz, perturbația aplicată membrului drept este definită de $\mu \hat{e}$, în timp ce în al doilea caz, perturbația este $\mu \hat{e} - F(t - F'(t)^{-1} F(t))$. Se observă că aceste două perturbații diferă printr-un termen de mărimea lui $O(\|t - t^*\|^2)$. Deci, metoda predictor-corrector de urmărire a traiectoriei, chiar pentru probleme nedegenerate, nu poate fi cubic convergentă. Totuși, alegând $\alpha_k = 1$ (în vecinătatea soluției optime) și considerând $\mu_k = O((x_k^T z_k)^3)$, din (1.48) rezultă că este posibil să obținem convergența cubică. Toate aceste observații pot fi ansamblate în:

Teorema 1.5. Fie $\{t_k\}$ șirul de iterații generat de metoda predictor-corrector de urmărire a traiectoriei cu t_0 strict admisibil, aplicată problemei $\min\{c^T x / Ax = b, x \geq 0\}$. Presupunând că:

- 1) relațiile de complementaritate sunt strict satisfăcute,
- 2) t^* este un vârf nedegenerat,
- 3) șirul $\{t_k\}$ converge la t^* .

Dacă alegem σ_k și τ_k care să satisfacă:

$$0 \leq \sigma_k \leq \min\{\sigma, c_1(x_k^T z_k)\}, \quad (1.50)$$

și

$$\max\{\tau, 1 - c_2(x_k^T z_k)\} \leq \tau_k < 1 \quad (1.51)$$

unde $\sigma_k \in (0,1)$, $\tau \in (0,1)$ și $c_1, c_2 > 0$, atunci convergența metodei predictor-corector este pătratică, adică există o constantă $\gamma_2 > 0$, astfel încât pentru k suficient de mare:

$$\|t_{k+1} - t^*\| \leq \gamma_2 \|t_k - t^*\|^2 \quad (1.52)$$

Pe de altă parte, dacă în locul lui (1.50) se alege:

$$0 \leq \sigma_k \leq \min\{\sigma, c_1(x_k^T z_k)^2\} \quad (1.53)$$

și în locul lui (1.51), pentru k suficient de mare, se consideră $\alpha_k = 1$, atunci convergența este cubică, adică există o constantă γ_3 astfel încât pentru k suficient de mare:

$$\|t_{k+1} - t^*\| \leq \gamma_3 \|t_k - t^*\|^3 \quad (1.54)$$

Capitolul 2. METODE DE PUNCT INTERIOR PENTRU PROGRAMAREA MATEMATICĂ

2.1. METODE DE PUNCT INTERIOR PENTRU PROBLEMA DE PROGRAMARE PĂTRATICĂ

Problemele de programare matematică pătratică sunt importante deoarece formează o punte de legătură între problemele de programare liniară și cele de programare neliniară, dar și pentru că, în jurul unui punct curent, orice problemă de programare neliniară, cu proprietăți de netezime suficient de rezonabile, poate fi destul de bine aproximată ca o problemă de programare pătratică. Acest lucru constituie fundamentul pentru foarte multe metode și algoritmi de optimizare. De asemenea, acestea sunt importante deoarece, în jurul unui punct de funcționare foarte multe fenomene pot fi modelate ca probleme de programare pătratică. Majoritatea legilor de mecanică, teoria microscopică a câmpului electromagnetic, chimie, biologie, economie, finanțe, etc. sunt exprimate ca relații matematice în care neliniaritățile sunt de ordinul doi, astfel încât prezența modelelor de optimizare pătratică este foarte probabilă, fiind o bună aproximație a mișcării din diferite domenii de activitate.

De obicei, problemele de programare pătratică sunt formulate ca probleme de minimizare:

$$\min \frac{1}{2} x^T Q x + c^T x$$

referitor la (2.1)

$$Ax \geq b, x \geq 0,$$

unde $x \in \mathbb{R}^n$, A este o matrice $m \times n$ -dimensională, iar Q este o matrice $n \times n$ -dimensională.

Cum problema duală are o poziție specială în metodele de punct interior pentru programarea liniară, în cele ce urmează vom prezenta o tehnică de construcție a problemei duale a lui (2.1). Modul cel mai direct de construcție a dualei se bazează pe legătura care există între dualitate și condițiile de optimalitate de ordinul unu pentru problema barieră asociată. Pentru a scrie problema barieră asociată lui (2.1), să considerăm variabilele ecart $s \in \mathbb{R}^m$ în restricțiile din (2.1), cu care scriem următoarea problemă barieră:

$$\min \frac{1}{2} x^T Q x + c^T x - \mu \sum_{j=1}^n \log x_j - \mu \sum_{i=1}^m \log s_i,$$

referitor la

(2.2)

$$Ax - s = b,$$

Pentru această problemă, introducem funcția Lagrange:

$$L(x, s, y) = \frac{1}{2} x^T Q x + c^T x - \mu \sum_{j=1}^n \log x_j - \mu \sum_{i=1}^m \log s_i + y^T (b - Ax + s).$$

Condițiile de optimalitate de primul ordin ale problemei barieră (2.2) se obțin prin anularea derivatelor funcției Lagrange în raport cu toate variabilele x , s și y :

$$\frac{\partial L}{\partial x} = Qx + c - \mu X^{-1} e - A^T y = 0,$$

$$\frac{\partial L}{\partial s} = -\mu S^{-1} e + y = 0,$$

$$\frac{\partial L}{\partial y} = b - Ax + s = 0.$$

La fel ca în cazul programării liniare, introducând vectorul $z = \mu X^{-1} e$, atunci condițiile de optimalitate de ordinul unu se pot scrie sub formă compactă:

$$Ax - s = b,$$

$$A^T y + z - Qx = c,$$

$$XZe = \mu e,$$

$$YSe = \mu e,$$

unde X , Z , Y și S sunt matrice diagonale a căror elemente sunt componentele vectorilor x , y , z și respectiv s . Din ultimele două condiții de mai sus se observă că problema duală implică vectorul n -dimensional al variabilelor z , care sunt complementare variabilelor primale x , precum și vectorul m -dimensional al variabilelor y , care sunt complementare variabilelor ecart primale s . Deci, variabilele y și z sunt nenegative în problema duală. De asemenea, pentru a stabili legătura între condițiile de optimalitate și problema duală, trebuie ca a doua relație din condițiile de mai sus să fie restricția problemei duale. Astfel, restricțiile problemei duale sunt:

$$A^T y + z - Qx = c,$$

$$y, z \geq 0.$$

Restricțiile problemei duale implică vectorul n -dimensional x , care pare să aparțină problemei primale. Variabilele x care apar în duală, în principiu, nu au nici o legătură cu variabilele x care apar în problema primală, cu excepția faptului că, în punctul de optim ele sunt egale.

Vom utiliza teorema slabă de dualitate pentru a scrie funcția obiectiv a problemei duale. În această teoremă presupunem că dispunem de o soluție admisibilă primală (x, s) și de o soluție admisibilă duală (x, y, z) . Atunci

$$y^T(Ax) = y^T(b + s) = b^T y + y^T s$$

și

$$(A^T y)^T x = (c - z + Qx)^T x = c^T x - z^T x + x^T Qx.$$

Acum, deoarece $y^T(Ax) = (A^T y)^T x$, rezultă că:

$$\begin{aligned} 0 &\leq y^T s + z^T x = c^T x + x^T Qx - b^T y = \\ &= \left(c^T x + \frac{1}{2} x^T Qx \right) - \left(b^T y - \frac{1}{2} x^T Qx \right). \end{aligned}$$

Din această inegalitate rezultă clar că funcția obiectiv a problemei duale este $b^T y - \frac{1}{2} x^T Qx$.

Deci, problema duală a lui (2.1) este:

$$\max b^T y - \frac{1}{2} x^T Qx,$$

referitor la (2.3)

$$\begin{aligned} A^T y + z - Qx &= c, \\ y, z &\geq 0. \end{aligned}$$

La fel ca în cazul programării liniare, și pentru programarea pătratică legătura fundamentală dintre problema primală (2.1) și cea duală (2.3) este dată de teorema ecarturilor complementare.

Teorema 2.1. Pentru problema de programare pătratică convexă, dată o soluție (x^*, s^*) care este admisibilă pentru problema primală și o soluție (x^*, y^*, z^*) care este admisibilă pentru problema duală, dacă acestea satisfac relația:

$$y^{*T} s^* + z^{*T} x^* = 0,$$

atunci (x^*, s^*) este soluție optimă pentru problema primală și (x^*, y^*, z^*) este soluția optimă pentru problema duală.

Demonstrație. Într-adevăr, fie (x, s) o soluție primal admisibilă și (\hat{x}, y, z) o soluție dual admisibilă. Vom utiliza teorema slabă de dualitate și faptul că $y^T Ax$ se poate scrie în două moduri, adică $y^T(Ax) = (A^T y)^T x$. Din condițiile de optimalitate de mai sus obținem:

$$y^T(Ax) = y^T(b + s) = b^T y + y^T s,$$

și respectiv:

$$(A^T y)^T x = (c - z + Q\hat{x})^T x = c^T x - z^T x + x^T Qx.$$

Comparând aceste două relații și ținând cont de faptul că atât $y^T s$ cât și $z^T x$ sunt nenegative, rezultă că:

$$0 \leq y^T s + z^T x = c^T x + \hat{x}^T Q x - b^T y.$$

Fie deci (x, s) o soluție primal admisibilă. Atunci teorema de dualitate slabă aplicată lui (x, s) din punctul de vedere al primalei și lui (x^*, y^*, z^*) , din punctul de vedere al dualei arată că:

$$c^T x + x^{*T} Q x - b^T y^* \geq 0.$$

Dar, în punctul de optim (x^*, s^*) , inegalitatea de mai sus e chiar egalitate:

$$c^T x^* + x^{*T} Q x^* - b^T y^* = 0.$$

Combinând aceste două relații de mai sus, obținem:

$$c^T x^* + x^{*T} Q x^* \leq c^T x + x^{*T} Q x.$$

Mai departe, trebuie să arătăm că funcția obiectiv primală evaluată în x^* nu este mai mare decât valoarea ei în x , adică:

$$c^T x^* + \frac{1}{2} x^{*T} Q x^* \leq c^T x + \frac{1}{2} x^T Q x.$$

Acest lucru se arată ușor. De exemplu, dacă plecăm de la membrul stâng al inegalității de mai sus obținem:

$$\begin{aligned} c^T x^* + \frac{1}{2} x^{*T} Q x^* &= c^T x^* + x^{*T} Q x^* - \frac{1}{2} x^{*T} Q x^* \\ &\leq c^T x + x^{*T} Q x - \frac{1}{2} x^{*T} Q x^* = \\ &= c^T x + \frac{1}{2} x^T Q x - \frac{1}{2} x^T Q x + x^{*T} Q x - \frac{1}{2} x^{*T} Q x^* \\ &= c^T x + \frac{1}{2} x^T Q x - \frac{1}{2} (x - x^*)^T Q (x - x^*). \end{aligned}$$

Dar, din ipotezele teoremei, matricea Q este pozitiv semidefinită, astfel încât din inegalitatea de mai sus, rezultă:

$$c^T x^* + \frac{1}{2} x^{*T} Q x^* \leq c^T x + \frac{1}{2} x^T Q x.$$

Cum x este un punct primal admisibil, rezultă că x^* (împreună cu s^*) constituie o soluție optimă pentru problema primală. O analiză similară arată că y^* (împreună cu x^* și z^*) formează o soluție optimă pentru problema duală.

(q.e.d)

După modul în care sunt rezolvate condițiile de optimalitate asociate problemei barieră, pentru problema de programare pătratică, la fel ca în cazul programelor liniare, se pot dezvolta metode de punct interior de urmărire a traiectoriei: *primal-duale*, *predictor-corector*, sau *predictor-corector de ordin superior*.

În cele ce urmează, vom dezvolta metoda primal-duală de punct interior pentru problema de programare pătratică (2.1). Într-adevăr, condițiile de optimalitate de ordinul unu sunt:

$$\begin{aligned} A^T y + z - Qx &= c, \\ Ax - s &= b, \\ XZe &= \mu e, \\ YSe &= \mu e. \end{aligned}$$

La fel ca în cazul programării liniare, vom înlocui punctul (x, s, y, z) cu $(x + \Delta x, s + \Delta s, y + \Delta y, z + \Delta z)$ pentru a obține următorul sistem algebric neliniar, a cărui soluție formează direcția de deplasare $(\Delta x, \Delta s, \Delta y, \Delta z)$:

$$\begin{aligned} A^T \Delta y + \Delta z - Q \Delta x &= c - A^T y - z + Qx \equiv r_D, \\ A \Delta x - \Delta s &= b - Ax + s \equiv r_p, \\ Z \Delta x + X \Delta z + \Delta X \Delta z &= \mu e - XZe, \\ S \Delta y + Y \Delta s + \Delta Y \Delta s &= \mu e - YSe, \end{aligned}$$

unde r_p și r_D sunt reziduurile primale și respectiv cele duale. Dacă eliminăm termenii neliniari (pătratici), obținem următorul sistem algebric liniar pentru direcția $(\Delta x, \Delta s, \Delta y, \Delta z)$:

$$\begin{aligned} A^T \Delta y + \Delta z - Q \Delta x &= r_D, \\ A \Delta x - \Delta s &= r_p, \\ Z \Delta x + X \Delta z &= \mu e - XZe, \\ S \Delta y + Y \Delta s &= \mu e - YSe. \end{aligned}$$

Prin rezolvarea acestui sistem, se obține direcția de deplasare căutată. La fel ca în cazul programării liniare, putem să reducem acest sistem în sensul eliminării variabilelor Δz și Δs . Din ultimele două ecuații ale sistemului de mai sus se obține imediat:

$$\begin{aligned} \Delta z &= X^{-1}(\mu e - XZe - Z \Delta x), \\ \Delta s &= Y^{-1}(\mu e - YSe - S \Delta y), \end{aligned} \tag{2.4}$$

Cu acestea, variabilele Δx și Δy determină soluții ale sistemului KKT redus:

$$\begin{bmatrix} -(X^{-1}Z + Q) & A^T \\ A & Y^{-1}S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} c - A^T y - \mu X^{-1}e + Qx \\ b - Ax + \mu Y^{-1}e \end{bmatrix}. \tag{2.5}$$

Astfel, din (2.4) și (2.5) se poate determina direcția de deplasare din punctul curent considerat. Lungimea pasului se determină exact ca în cazul liniar, prin calculul deplasării maxime admisibile penalizată cu un factor $\tau \in (0,1)$.

Definim astfel o nouă estimăție a variabilelor (x^+, s^+, y^+, z^+) sub forma:

$$\begin{aligned}x^+ &= x + \alpha_p \Delta x, & y^+ &= y + \alpha_D \Delta y, \\s^+ &= s + \alpha_p \Delta s, & z^+ &= z + \alpha_D \Delta z,\end{aligned}$$

unde α_p și α_D sunt lungimile pașilor de deplasare din spațiul primal, respectiv dual, alese astfel încât să se mențină pozitivitatea variabilelor x , s , y și z , adică $\alpha_p = \tau \hat{\alpha}_p$ și $\alpha_D = \tau \hat{\alpha}_D$, unde:

$$\begin{aligned}\hat{\alpha}_p &= \min \left\{ \min_j \left\{ \frac{x_j}{-\Delta x_j}, \Delta x_j < 0 \right\}, \min_j \left\{ \frac{s_j}{-\Delta s_j}, \Delta s_j < 0 \right\} \right\}, \\ \hat{\alpha}_D &= \min \left\{ \min_j \left\{ \frac{z_j}{-\Delta z_j}, \Delta z_j < 0 \right\}, \min_j \left\{ \frac{y_j}{-\Delta y_j}, \Delta y_j < 0 \right\} \right\},\end{aligned}\tag{2.6}$$

și $\tau \in (0,1)$ este factorul de menținere a iterației curente în interiorul domeniului de admisibilitate. De obicei, $\tau = 0.9995$.

Astfel se poate prezenta următorul algoritm de punct interior, primal-dual de urmărire a traiectoriei pentru programarea pătratică.

Algoritm QPD

(Descriere)

Pasul 1. Se consideră o aproximație inițială $t_0 = [x_0 \ s_0 \ y_0 \ z_0]^T$, unde $x_0 > 0$, $s_0 > 0$, $y_0 > 0$ și $z_0 > 0$. Se pune parametrul $k = 0$,

Pasul 2. Se testează un criteriu de oprire a iterațiilor,

Pasul 3. Se calculează reziduurile primale și duale:

$$r_p = b - A x_k + s_k, \quad r_D = c - A^T y_k - z_k + Q x_k.$$

Pasul 4. Se calculează saltul de dualitate: $\gamma = z^T x + y^T s$,

Pasul 5. Se determină valoarea parametrului barieră: $\mu = \sigma \frac{\gamma}{n+m}$, unde $\sigma \in (0,1)$ este un parametru de modificare a lui μ ,

Pasul 6. Se rezolvă sistemul KKT redus:

$$\begin{bmatrix} -(X^{-1}Z+Q) & A^T \\ A & Y^{-1}S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} c - A^T y - \mu X^{-1}e + Qx \\ b - Ax + \mu Y^{-1}e \end{bmatrix},$$

Pasul 7. Se determină direcțiile:

$$\Delta z = X^{-1}(\mu e - XZe - Z\Delta x),$$

$$\Delta s = Y^{-1}(\mu e - YSe - S\Delta y),$$

Pasul 8. Se alege $\tau_k \in (0, 1)$ și se pune $\alpha_p = \tau_k \hat{\alpha}_p$ și $\alpha_D = \tau_k \hat{\alpha}_D$, unde $\hat{\alpha}_p$ și $\hat{\alpha}_D$ sunt dați de (2.6),

Pasul 9. Se calculează noua aproximație $t_{k+1} = \begin{bmatrix} x_{k+1} & s_{k+1} & y_{k+1} & z_{k+1} \end{bmatrix}^T$

unde $x_{k+1} = x_k + \alpha \Delta x$, $s_{k+1} = s_k + \alpha \Delta s$, $y_{k+1} = y_k + \alpha \Delta y$, $z_{k+1} = z_k + \alpha \Delta z$, iar $\alpha = \min(\alpha_p, \alpha_D)$. Se pune $k = k + 1$ și se continuă cu pasul 2.

La implementarea practică a metodelor de punct interior, prezența liniilor sau a coloanelor dense în matricea restricțiilor face să preferăm rezolvarea sistemului KKT redus, ceea ce implică utilizarea unui software care poate trata sisteme de ecuații algebrice liniare nedefinite. La fel ca în cazul programării liniare, continuăm reducerea sistemului (2.5) în scopul obținerii ecuațiilor normale. Astfel, avem două posibilități:

(1) Într-adevăr, din (2.5) putem obține imediat *ecuațiile normale în forma primală*, prin rezolvarea primei ecuații în privința lui Δx :

$$\Delta x = -(X^{-1}Z + Q)^{-1} [c - A^T y - \mu X^{-1} e + Qx - A^T \Delta y].$$

Atunci, din a doua ecuație din (2.5) deducem imediat *ecuațiile normale (în forma primală)*:

$$\left[A(X^{-1}Z + Q)^{-1} A^T + Y^{-1} S \right] \Delta y = b - Ax + \mu Y^{-1} e +$$

$$A(X^{-1}Z + Q)^{-1} (c - A^T y - \mu X^{-1} e + Qx).$$

După cum vom vedea, matricea coeficienților ecuațiilor normale în forma primală are tiparul zero/nezero dat de matricea $A(D + Q)^{-1} A^T$, unde D este o matrice diagonală. Dacă Q este matrice diagonală, ne aflăm în situația programării liniare, iar în caz contrar structura de matrice rară a lui $A(D + Q)^{-1} A^T$ este total ruinată.

(2) Tot din (2.5) putem obține imediat *ecuațiile normale în forma duală*, prin rezolvarea celei de-a doua ecuații în privința lui Δy :

$$\Delta y = YS^{-1} [b - Ax + \mu X^{-1} e - A \Delta x],$$

și apoi eliminarea lui din prima ecuație, pentru a obține:

$$\begin{aligned} -(X^{-1}Z + Q - A^T YS^{-1} A) \Delta x &= c - A^T y - \mu X^{-1} e + Qx - \\ &- A^T YS^{-1} (b - Ax + \mu Y^{-1} e). \end{aligned}$$

În această situație, tiparul zero/nezero al matricei coeficienților ecuațiilor normale în forma duală este dat de matricea $A^T A + Q$, matrice care poate fi mult mai rară decât $A(D + Q)^{-1} A^T$. Totuși, în acest caz avem de rezolvat un sistem n -dimensional, față de cazul ecuațiilor normale în forma primală, în care dimensiunea sistemului este m (în general mai mică decât n).

Capitolul 3. ALGORITMUL KARMARKAR

În anul 1984, Karmarkar a propus un algoritm foarte important pentru rezolvarea problemelor de programare liniară. Acesta a fost primul algoritm de programare liniară în timp polinomial care să concureze în mod viabil cu algoritmul simplex în rezolvarea problemelor de dimensiuni mari, provenite din lumea reală. În timp ce algoritmul simplex generează puncte extreme pe frontiera domeniului de admisibilitate, algoritmul lui Karmarkar este o metodă de punct interior. Acest algoritm generează în interiorul domeniului de admisibilitate un șir de puncte ale căror costuri converg către costul optimal, obținându-se în final un punct suboptimal, care cu ajutorul unor transformări elementare este adus într-un punct optimal de pe frontiera domeniului de admisibilitate.

În cele ce urmează, vom face prezentarea algoritmului proiectiv al lui Karmarkar, bazându-ne pe articolul lui inițial.

3.1. PROBLEMA ÎN FORMA STANDARD KARMARKAR

Problemele de programare liniară rezolvate de algoritmul Karmarkar au o formă specială, numită *forma standard Karmarkar*. Datele acestor probleme sunt reprezentate de o matrice $A \in \mathbb{R}^{m \times n}$ (matricea coeficienților restricțiilor) și de un vector $c \in \mathbb{R}^n$ (vectorul coeficienților funcției obiectiv). Notăm cu $e \in \mathbb{R}^n$ vectorul care are toate componentele egale cu 1.

Definiție 3.1. Simplexul standard $(n - 1)$ -dimensional este mulțimea:

$$\Delta = \left\{ x \in \mathbb{R}^n / x \geq 0, e^T \cdot x = \sum_{j=1}^n x_j = 1 \right\}$$

Vectorul $a_0 = \frac{1}{n} e$ se numește centrul simplexului.

Notăm spațiul liniar $\Omega = \{x \in \mathbb{R}^n / A \cdot x = 0\}$ și deci, domeniul de admisibilitate Π a problemei în forma standard este:

$$\Pi = \Omega \cap \Delta$$

Putem presupune că toate cele $m+1$ restricții care definesc domeniul Π sunt liniar independente.

Forma standard Karmarkar presupune că:

- centrul simplexului este un punct admisibil: $a_0 = \frac{1}{n}e \in \Pi$;

- costul $c^T \cdot x$ este nenegativ, pentru orice $x \in \Pi$:

$$\forall x \in \Pi \Rightarrow c^T \cdot x \geq 0.$$

Cu ipotezele și notațiile de mai sus, problema de programare liniară în forma standard Karmarkar se scrie astfel:

$$\inf \{c^T \cdot x / x \in \Omega \cap \Delta\} .$$

3.2. MINIMIZAREA PE SFERĂ

Fie $z \in \square^n$ și $r \in \square$, $r > 0$. Sfera de centru z și rază r este mulțimea:

$$S(z, r) = \{x \in \square^n / \|x - z\| \leq r\}$$

Intersecția unui spațiu afin de dimensiune nenulă cu o sferă a cărui centru aparține acestui spațiu, este o sferă de aceeași rază, dar de dimensiune mai mică cu o unitate.

Considerăm $z \in \square^n$ cu proprietatea: $A \cdot z = b$ și

$$S' = \{x \in \square^n / A \cdot x = b\} \cap S(z, r),$$

adică S' este intersecția unui spațiu afin cu o sferă având centrul în acel spațiu.

Minimizarea funcției liniare $c^T \cdot x$ peste S' se realizează astfel: dacă notăm cu c_p proiecția lui c pe spațiul afin $\{x \in \square^n / A \cdot x = b\}$, $-c_p$ este direcția din acest spațiu, de-a lungul căreia funcția obiectiv descrește cel mai rapid. Prin urmare, plecând din centrul z al sferei și deplasându-ne în direcția $-c_p$ cu un pas egal cu lungimea razei r , determinăm un punct de pe suprafața sferei $S(z, r)$, care este punctul de cost minim căutat. [Bătătorescu, 2002]

3.3. EVALUĂRI DE REGULARITATE A DOMENIULUI DE ADMISIBILITATE

Presupunem că deținem un punct admisibil z al problemei:

$$\inf \{c^T \cdot x / A \cdot x = b\}$$

Cum putem să minimizăm $c^T \cdot x$ peste o sferă de dimensiune mai mică, obținută ca intersecție dintre un spațiu afin și o sferă centrată în el, încercăm să înscriem în interiorul regiunii admisibile o sferă de dimensiune mai mică S , centrată în z și să minimizăm $c^T \cdot x$ peste S . Dacă sfera S nu este cu mult mai mică decât regiunea admisibilă, minimumul peste S ar putea să nu fie cu mult mai mare decât minimumul problemei de la care am plecat.

Fie $P \subseteq \mathbb{R}^n$ un poliedru și $z \in P$. Fie $E \subseteq P$ o mulțime convexă închisă cu $z \in E$. Pentru a obține o evaluare a aproximării soluției optime, vom scala mulțimea E printr-un factor v suficient de mare, având centrul scalării în z , astfel încât noua mulțime E' să conțină mulțimea P :

$$E' = \{z + v(x - z) / x \in E\}$$

$$E \subseteq P \subseteq E'$$

Mulțimea E' are aceeași formă ca E . Fiecărui punct $x \in E$ îi corespunde un punct $x' = z + v(x - z) \in E'$, și reciproc.

Considerăm $c^T \cdot x_E$ valoarea minimă a lui $c^T \cdot x$ peste E . Deoarece funcția $c^T \cdot x$ este liniară, luând $x_{E'} = z + v(x_E - z)$, valoarea minimă a lui $c^T \cdot x$ peste E' va fi $c^T \cdot x_{E'}$. Avem:

$$z - x_{E'} = v(z - x_E)$$

de unde obținem:

$$c^T \cdot (z - x_{E'}) = v[c^T \cdot (z - x_E)]$$

Valoarea care ne interesează este $c^T \cdot x_P$, minimumul lui $c^T \cdot x$ peste poliedrul P . Deoarece $P \subseteq E'$, avem $c^T \cdot x_P \geq c^T \cdot x_{E'}$ și astfel:

$$c^T \cdot (z - x_P) \leq c^T \cdot (z - x_{E'}) = v[c^T \cdot (z - x_E)]$$

Deci, putem scrie:

$$\begin{aligned} (v-1)(c^T \cdot z) - v(c^T \cdot x_E) &\geq -(c^T \cdot x_P) \\ (v-1)(c^T \cdot z) - (v-1)(c^T \cdot x_P) &\geq v(c^T \cdot x_E) - v(c^T \cdot x_P) \\ (v-1)[(c^T \cdot z) - (c^T \cdot x_P)] &\geq v[(c^T \cdot x_E) - (c^T \cdot x_P)] \end{aligned}$$

Deoarece $c^T \cdot z > c^T \cdot x_P$, avem:

$$\frac{c^T \cdot (x_E - x_P)}{c^T \cdot (z - x_P)} \leq 1 - \frac{1}{v}$$

Cu cât este mai mic v , cu atât mai mare este progresul realizat în funcția obiectiv prin trecerea de la z la x_E . Cu cât P va avea o formă mai apropiată de cea a unei sfere centrate în z , cu atât v va avea o valoare mai mică și astfel, el poate fi folosit pentru a exprima regularitatea poliedrului.

Algoritmul Karmarkar are drept punct de plecare pe $z = a_0$ și generează o succesiune de puncte admisibile a căror costuri converg către 0. Acest procedeu poate fi descris astfel: să considerăm că am determinat un punct $a \in \Pi$ pentru care $c^T \cdot a > 0$ și vrem să generăm următorul punct b . Pentru aceasta, va trebui să înscriem în Π o sferă S centrată în a și să minimizăm $c^T \cdot x$ peste S . Cum este puțin probabil ca minimumul lui $c^T \cdot x$ peste S să aproximeze bine adevăratul minim al lui $c^T \cdot x$ peste Π , deoarece în general Π nu are o formă sferică cu centrul în a , Karmarkar a avut ideea de a transforma problema dată printr-o aplicație T_a , astfel încât să se obțină o nouă problemă în care imaginea punctului curent a prin T_a să fie centrul simplexului Δ , a_0 . Regiunea admisibilă Π' a noii probleme va fi dată la rândul ei tot de intersecția simplexului Δ cu un spațiu afin, la fel ca și Π , iar mulțimea Δ este regulată și are drept centru a_0 . Aceasta înseamnă că dacă în noua problemă minimizăm peste o sferă înscrisă în Π' , vom obține un punct b' al cărui cost va fi semnificativ mai mic decât cel al lui a_0 . Dacă aplicăm lui b' transformarea inversă T_a^{-1} , se obține un nou punct b din Π . Deoarece costul lui b' în problema transformată este mult mai mic decât cel al lui a_0 , în problema originală costul lui b va fi și el mai mic decât cel al lui a . Prin aplicarea repetată a acestui procedeu, se obține un șir de puncte ale căror costuri vor converge către 0.

Notăm cu $S(a_0, r)$ cea mai mare sferă $(n - 1)$ -dimensională centrată în a_0 , care este inclusă în Δ (sfera înscrisă), și cu $S(a_0, R)$ cea mai mică sferă $(n - 1)$ -dimensională centrată în a_0 , care îl conține pe Δ (sfera circumscrisă).

Lema 3.1. Razele sferelor $S(a_0, r)$ și $S(a_0, R)$ sunt date de următoarele formule:

$$r = \frac{1}{\sqrt{n(n-1)}} \text{ și } R = \sqrt{\frac{n-1}{n}} .$$

Demonstrație. Sfera înscrisă.

Demonstrăm mai întâi că, dacă $y \in \square^n$ satisface condițiile: $y_i \geq \frac{1}{n}$ pentru $i < n$, $y_n = 0$

și $y_1 + y_2 + \dots + y_n \geq 1$, atunci

$$\|y - a_0\|^2 \geq r^2 = \frac{1}{n(n-1)} .$$

Fie $z_i = y_i - \frac{1}{n}$ pentru $i = 1, 2, \dots, n-1$. Atunci

$$\|y - a_0\|^2 = \sum_{i=1}^{n-1} \left(y_i - \frac{1}{n} \right)^2 + \frac{1}{n^2} = \sum_{i=1}^{n-1} z_i^2 + \frac{1}{n^2}$$

Avem evident $z_i \geq 0$ pentru orice i , iar

$$\sum_{i=1}^{n-1} z_i \geq 1 - \frac{n-1}{n} = \frac{1}{n}.$$

Deoarece $z_i \geq 0$ și $\sum_{i=1}^{n-1} z_i \geq \frac{1}{n}$, minimumul expresiei $\sum_{i=1}^{n-1} z_i^2 + \frac{1}{n^2}$ este obținut dacă $\sum_{i=1}^{n-1} z_i = \frac{1}{n}$ și toți z_i

sunt identici, adică $z_i = \frac{1}{n(n-1)}$ pentru $i = 1, 2, \dots, n-1$. În acest caz,

$$\sum_{i=1}^{n-1} z_i^2 + \frac{1}{n^2} = \frac{1}{n^2(n-1)} + \frac{1}{n^2} = \frac{1}{n(n-1)} = r^2$$

și afirmația este demonstrată.

Distanța de la a_0 la punctul $\left(0, \frac{1}{n-1}, \frac{1}{n-1}, \dots, \frac{1}{n-1} \right)^T$ de pe frontiera lui Δ este:

$$\sqrt{\frac{1}{n^2} + (n-1) \left(\frac{1}{n-1} - \frac{1}{n} \right)^2} = \sqrt{\frac{1}{n^2} + \frac{1}{(n-1)n^2}} = r$$

astfel încât raza oricărei sfere $(n-1)$ -dimensionale incluse în Δ nu poate fi mai mare decât r .

Demonstrăm că orice punct $x \in \square^n$ aparține lui Δ dacă satisface condițiile $e^T \cdot x = 1$ și $\|x - a_0\| \leq r$. Presupunem prin absurd că x satisface condițiile cerute, dar că există o anumită componentă a sa, să zicem ultima, care este negativă. Definim vectorul $y \in \square^n$ astfel:

$$y_i = \begin{cases} x_i & \text{daca } i < n \text{ si } x_i \geq \frac{1}{n} \\ \frac{2}{n} - x_i & \text{daca } i < n \text{ si } x_i < \frac{1}{n} \\ 0 & \text{daca } i = n \end{cases}$$

Este ușor de verificat că $y_i \geq \frac{1}{n}$ pentru $i < n$, $y_n = 0$, $e^T \cdot y \geq 1$ și $\|y - a_0\| \leq \|x - a_0\| \leq r$.

Dar aceasta contrazice afirmația demonstrată mai sus și astfel prima formulă este demonstrată.

Sfera circumscrisă

Distanța de la a_0 la orice punct extremal al lui Δ este:

$$\sqrt{\left(1 - \frac{1}{n} \right)^2 + \frac{n-1}{n^2}} = \sqrt{(n-1)n} = R.$$

Astfel, orice sferă sferă $(n - 1)$ -dimensională care include pe Δ trebuie să aibe raza cel puțin R . Pentru a completa demonstrația, arătăm că distanța de la orice punct din Δ la a_0 este cel mult R .

Dacă α și β sunt numere reale nenegative, este ușor de verificat că

$$\left(\alpha - \frac{1}{n}\right)^2 + \left(\beta - \frac{1}{n}\right)^2 \leq \left[\left(\alpha + \beta\right) - \frac{1}{n}\right]^2 + \left(0 - \frac{1}{n}\right)^2$$

Fie $x \in \Delta$. Pentru $i = 2, 3, \dots, n$, putem înlocui succesiv componentele x_1 și x_i ale lui x prin $x_1 + x_i$ și 0, respectiv. Procedând astfel și ținând cont de relația de mai sus, obținem puncte din Δ a căror distanță față de a_0 este cel puțin egală sau mai mare decât cea a punctului anterior. Deoarece distanța acestuia față de a_0 este R , nici un punct din Δ nu se va afla față de a_0 la o distanță mai mare decât R .

(q.e.d)

3.4. ALGORITMUL KARMARKAR

Fie L mărimea problemei de programare liniară în forma standard Karmarkar, care este numărul:

$$L = (m + 1)n + \lceil \log_2 |P| \rceil + n \lceil \log_2 n \rceil$$

unde, m, n sunt dimensiunile matricei A , iar P este produsul tuturor coeficienților nenuli din problemă.

Fie $\alpha \in \left(0, \frac{1}{2}\right)$ un număr real oarecare și fie

$$\delta = \alpha - \frac{\alpha^2}{1 - \alpha} > 0.$$

În cele ce urmează, vom arăta că algoritmul Karmarkar rezolvă în timp polinomial orice problemă în forma standard.

Pentru a descrie mai ușor acest algoritm, vom utiliza o funcție Φ , care generează șirul de puncte din interiorul domeniului Π , adică $x^{(k+1)} = \Phi(x^{(k)})$. Considerăm că funcția Φ are următoarele proprietăți:

- dacă $a \in \Pi$, $a > 0$ și $c^T \cdot a > 0$, atunci $\Phi(a) \in \Pi$ și $\Phi(a) > 0$,
- dacă problema are costul optim egal cu 0, atunci ori $c^T \cdot \Phi(a) = 0$,
ori $f(\Phi(a)) \leq f(a) - \delta$.

Algoritmul lui Karmarkar.

Iterația 0. Se consideră $x^{(0)} = a_0 = \frac{1}{n}e$ și fie $K = \left\lceil \frac{2nL}{\delta} \right\rceil$ (numărul maxim de iterații).

Dacă $c^T \cdot a_0 = 0$, atunci o soluție optimă este a_0 . STOP.

Dacă $c^T \cdot a_0 > 0$, atunci se efectuează iterațiile următoare, pentru $k = \overline{1, K}$:

Iterația k.

- Cu ajutorul funcției Φ se generează $x^{(k)} = \Phi(x^{(k-1)})$.

- Dacă $c^T \cdot x^{(k)} = 0$, atunci soluția optimă este $x^{(k)}$. STOP.

- Dacă $f(x^{(k)}) > f(x^{(k-1)}) - \delta$, atunci problema are costul optim pozitiv. STOP.

Iterația K + 1. Găsește un punct external $v \in \Pi$ al cărui cost nu este mai mare decât $c^T \cdot x^{(K)}$. În acest caz, v este o soluție de bază optimă, adică: $c^T \cdot v = 0$.

Teorema 3.1. Presupunem că:

- pentru orice $a \in \Pi$, $a > 0$ și $c^T \cdot a > 0$, avem $\Phi(a) \in \Pi$ și $\Phi(a) > 0$;

- dacă costul minim al problemei este 0, atunci ori $c^T \cdot \Phi(a) = 0$, ori $f(\Phi(a)) \leq f(a) - \delta$.

În aceste condiții, algoritmul Karmarkar rezolvă corect problemele de programare liniară în forma standard Karmarkar.

Demonstrație. Singura situație în care algoritmul poate semnaliza un cost optim pozitiv este la iterația k , dacă costul $c^T \cdot x^{(k)} > 0$ și $f(x^{(k)}) > f(x^{(k-1)}) - \delta$. În acest caz, datorită ipotezei făcute asupra lui Φ , știm că problema are un cost optim pozitiv.

Dacă condiția $c^T \cdot x^{(k)} = 0$ este verificată, atunci evident $x^{(k)}$ este o soluție optimă a problemei.

Să presupunem că pentru orice $k = 1, 2, \dots, K$, avem $f(x^{(k)}) \leq f(x^{(k-1)}) - \delta$. Atunci

$$f(x^{(K)}) \leq f(x^{(0)}) - \left(\frac{2nL}{\delta} \right) \delta = f(x^{(0)}) - 2nL$$

și

$$c^T \cdot x^{(K)} \leq e \frac{f(x^{(K)})}{n} \leq e \frac{f(x^{(0)}) - 2nL}{n}.$$

Deoarece $x^{(0)} = a_0$ deducem că:

$$f(x^{(0)}) = \ln \left[\frac{(c^T \cdot a_0)^n}{\left(\frac{1}{n}\right)^n} \right] = n \ln(nc^T \cdot a_0) = n \ln(c_1 + c_2 + \dots + c_n)$$

Dar, se poate arăta ușor că $\ln(c_1 + c_2 + \dots + c_n) \leq L$, astfel încât avem $f(x^{(0)}) \leq nL$. Rezultă

$$\frac{f(x^{(0)}) - 2nL}{n} \leq -L. \text{ Prin urmare,}$$

$$c^T \cdot x^{(K)} \leq e^{-L} < 2^{-L}.$$

În acest caz, fie $v \in \Pi$ punctul extremal obținut de iterația $K+1$. Conform Lemei 3.1, soluția de bază v este un vector cu toate componentele raționale, având numitorul comun pozitiv și mai mic ca 2^L . Astfel, dacă $c^T \cdot v > 0$, atunci $c^T \cdot v > 2^{-L}$. Pe de altă parte, $0 \leq c^T \cdot v \leq c^T \cdot x^{(K)} < 2^{-L}$, ceea ce implică $c^T \cdot v = 0$.

În concluzie, algoritmul Karmarkar determină un punct de cost 0, dacă există unul, sau semnalează că un astfel de punct nu există.

(q.e.d)

În cele ce urmează, ne vom ocupa de construcția unei funcții Φ cu proprietățile cerute de algoritm.

Caracteristica principală a lui Φ este de a duce printr-o transformare proiectivă, un punct a din interiorul lui Π , în centrul $a_0 \in \Pi'$ a simplexului Δ , apoi de a găsi un punct $b' \in \Pi'$, astfel încât $f'(b') \leq f'(a_0) - \delta$ și în final, prin aplicarea transformării inverse, de a determina punctul $b \in \Pi$.

Așa cum am arătat, domeniul transformat Π' are o bună regularitate cu centrul în a_0 , ceea ce justifică alegerea lui b' ca soluție optimă în minimizarea lui $c^T \cdot y$ pe sfera $S'(a_0, r) \subseteq \Pi'$. Poate deci să apară următoarea problemă: aflându-se pe frontiera sferei, anumite componente ale lui b' pot fi zero și în acest caz, evaluarea progresului nu mai este posibilă prin intermediul funcției potențial f' , care este definită pentru vectori cu toate componentele strict pozitive.

Pentru a evita acest lucru, trebuie să ne asigurăm că b' rămâne strict în interiorul domeniului Π' , astfel că vom efectua minimizarea pe sfera $S'\left(a_0, \frac{a}{n}\right)$, pentru care raza este

$$\frac{a}{n} < \frac{a}{\sqrt{n(n-1)}} < ar < \frac{r}{2}.$$

Cu această alegere a lui b' vom putea demonstra că $f'(b') \leq f'(a_0) - \delta$. Introducem următoarele notații:

$$\begin{aligned}\Sigma &= \{x \in \mathbb{R}^n / e^T \cdot x = 1\} \\ \bar{\Omega} &= \Omega' \cap \Sigma \\ B &= \begin{pmatrix} A \cdot D \\ e^T \end{pmatrix}\end{aligned}$$

Evident, $a_0 \in \bar{\Omega}$, iar B este o matrice de dimensiune $(m+1) \times n$, obținută prin adăugarea unei linii la cele ale lui $A \cdot D$. De asemenea, putem scrie

$$\bar{\Omega} = \{x \in \mathbb{R}^n / B \cdot x = e^{(m+1)}\}$$

unde $e^{(m+1)}$ este vectorul unitar de dimensiune $m+1$, care are pe 1 pe ultima poziție.

Modul în care funcția Φ asociază punctului $a \in \Pi$, vectorul $b = \Phi(a) \in \Pi$, este următorul:

Funcția Φ .

1. Fie $a \in \Pi$, $a > 0$, și matricea diagonală $D = \text{diag}(a)$.

Se determină vectorul $c' = D \cdot c$ și matricea $B = \begin{pmatrix} A \cdot D \\ e^T \end{pmatrix}$.

2. Se calculează c_p , proiecția lui c' pe spațiul nul al lui B :

$$c_p = c' - B^T \cdot (B \cdot B^T)^{-1} \cdot B \cdot c'.$$

și se determină vectorul

$$\bar{c}_p = \begin{cases} \frac{c_p}{\|c_p\|} & \text{daca } c_p \neq 0 \\ 0, & \text{daca } c_p = 0 \end{cases}$$

3. Se calculează vectorul

$$b' = a_0 - \frac{\alpha}{n} \bar{c}_p$$

care se obține plecând din a_0 cu un pas de lungime $\frac{\alpha}{n}$ în direcția $-\bar{c}_p$ (optimizarea pe sferă).

4. Se calculează

$$b = T_a^{-1}(b') = \frac{D \cdot b'}{e^T \cdot D \cdot b'}.$$

CONCLUZII

Metodele de punct interior rămân o zonă activă și rodnică de cercetare, deși ritmul frenetic care caracteriza acest domeniu a încetinit în ultimii ani. Acestea continuă să se dezvolte, deși competiția cu metoda simplex este strânsă. Aplicații la programarea pătratică arată promisiuni considerabile, din cauza capacității superioare ale metodelor de punct interior de a exploata structura problemei în mod eficient. Influența asupra teoriei de programare neliniară și practică nu a fost încă stabilită, chiar dacă cercetări importante au fost deja consacrate acestui subiect.

Utilizarea abordării punct-interior în metodele de descompunere pare promițătoare, deși nu au fost efectuate studii comparative riguroase cu abordări alternative.

Principala idee care se află în spatele tuturor metodelor de punct interior constă în realizarea simultană a două obiective:

- reducerea valorii funcției obiectiv,
- evitarea, pe cât posibil, a punctelor din vecinătatea frontierei domeniului de admisibilitate.

Metodele de punct interior de urmărire a traiectoriei implementează ideea de a ține iterațiile lângă traiectoria centrală. Pentru aceasta, ele apelează la o combinație a metodelor bazate pe condițiile de optimalitate ale cuplului primal-dual cu metode de penalizare.

BIBLIOGRAFIE

- [1] **Andrei, N.**, *Programarea Matematică, Metode de punct interior*. Editura Tehnică, București, 1999.
- [2] **Bătătorescu, A.**, *Metode de optimizare liniară*. Editura Universității din București, București, 2002.
- [3] **Dantzig, G.B.**, *Linear Programming and Extensions*. Princeton University Press, New Jersey, 1963.
- [4] **Karmar, N.K.**, *A new polynomial time algorithm for linear programming*. *Combinatorica*, vol.4, 1984, pp.337-395.
- [5] **Karmarkar, N., Lagarias, J.C, Slutsman, L., Wang, P.**, *Power series variants of Karmarkar-types algorithms*. *AT&T Tech.J.*, vol.68, pp.20-36, 1989.
- [6] **Mehrotra, S.**, *On finding a vertex solution using interior points methods*. *Linear Algebra and its Applications*, vol.152, 1990.
- [7] **Vanderbei, R.J.**, *Linear Programming: Foundations and Extensions*. Kluwer Academic Publisher, Boston, 1996.
- [8] **Zhang, Y., Tapia, R.A., Dennis, J.E.**, *On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms*. *SIAM Journal on Optimisation*, vol.2, 1992.