

ALGORITMOS DE CENTROS PARA A PROGRAMAÇÃO CONVEXA NÃO
SUAVE

Marcos Augusto dos Santos

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:



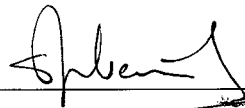
Prof. Paulo Roberto Oliveira, Dr. Ing. (orientador)



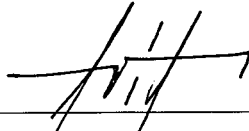
Prof. Nelson Maculan Filho, D. Sc.



Prof. Geraldo Robson Mateus, D.Sc.



Prof. Susana Sheimberg de Mackler, D. Sc.



Prof. Angel Ramon Sanchez Delgado, D. Sc.

RIO DE JANEIRO, RJ - BRASIL

SETEMBRO DE 1998

SANTOS, MARCOS AUGUSTO DOS

Métodos de Centros para a Programação Convexa Não Suave
[Rio de Janeiro] 1988

IX, 106 p., 29,7 cm (COPPE/UFRJ, D. Sc., Engenharia de Sistemas,
1998)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Métodos para resolução de de problemas de programação convexa
não diferenciável

I. COPPE/UFRJ II. Título (série)

“A good laboratory, like a good bank or a corporation or government, has to run like a computer. Almost everything is done flawlessly, by the book, and all numbers add up to the predicted sums. The days goes by. And then, if it is a lucky day, and a lucky laboratory, somebody makes a mistake; the wrong buffer, something in one of the blanks, a decimal misplaced in reading counts, the warm room off by a degree and half, a mouse out of his box, or just a misreading of the day’s protocol. Whatever, when the results come in, something is obviously screwed up, and the action can begin.

The misreading is not the important error; it opens the way. The next step is the crucial one. If the investigator can bring himself to say “But even so, look at that!” then the new finding what ever it is, is ready for snatching. What is needed, for progress to be made, is move based on the error.”

Lewis Thomas, *The Medusa and the Snail*

Aos jovens Josefino & Aída
À Liliane, Gabriel e Pedro
Ao professor Paulo Roberto

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D. Sc.)

ALGORITMOS DE CENTROS PARA A PROGRAMAÇÃO CONVEXA NÃO SUAVE

Marcos Augusto dos Santos

Setembro/1998

Orientador: Prof. Paulo Roberto Oliveira

Programa: Engenharia de Sistemas e Computação

Apresenta-se um novo algoritmo para a programação convexa não suave. As principais características são a estabilidade e o uso parcial do método de centros analíticos. Dada uma região convexa que contém, pelo menos, uma solução do problema, a estabilidade é alcançada utilizando pontos próximos à trajetória central de um polítopo que é construído a partir do polítopo original e de cortes gerados por subgradientes. O método provê uma forma de controlar o feixe de informações, incrementado em cada iteração, através de uma técnica “build-down” que o mantém dentro de limites razoáveis. A convergência é estabelecida e são apresentados testes com problemas típicos da literatura.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

CENTER ALGORITHMS FOR NONSMOOTH CONVEX PROGRAMMING

Marcos Augusto dos Santos

September/1998

Advisor: Prof. Paulo Roberto Oliveira

Department: System Engineering and Computing

We propose a new algorithm for minimizing nondifferentiable convex functions. The main features are the stability and the partial using of the analytic centers method. Given a convex polytope that contains, at least, one solution of the problem, we achieve the stability using points close to the central path of a polytope which is constructed from the original and the subgradients cuts. The evolution of the bundle of information, gained at each iteration, is controlled by a build-down strategy to keep a manageable size. The convergence is established and we present numerical tests with typical problems from literature.

ÍNDICE

1 - Introdução

2 - Algoritmos de otimização baseados na noção de centro

- 2.1 Métodos das Seções Centrais, Esferas e Elipsóides
- 2.2 Método dos Centros Analíticos
- 2.3 Métodos de Shor
 - 2.3.1 A escolha do passo
 - 2.3.2 A escolha do passo com transformação de espaços

3 - Um novo método para a programação convexa não suave

- 3.1 Introdução
 - 3.1.1 Planos secantes
 - 3.1.2 Estabilidade através da proximidade quadrática
 - 3.1.3 Estabilidade através do centro analítico
- 3.2 O método de centro analítico
- 3.3 Problema da reinicialização
 - 3.3.1 O caso quadrático
- 3.4 Algoritmo
 - 3.4.1 Obtenção de um ponto interior no sistema perturbado
 - 3.4.2 Atualização do parâmetro ρ
- 3.5 Gerência do feixe de informações
- 3.6 Convergência
 - 3.6.1 Convergência das sequências $\{\rho^k\}$ e $\{f_k^m\}$
 - 3.6.2 Complexidade para a redução de ρ^k para ρ^{k+1}
 - 3.6.3 Convergência de $\{\rho^k\}$ e $\{f_k^m\}$ para $f(x^*)$
 - 3.6.4 Complexidade por iteração
 - 3.6.5 Uma demonstração alternativa da convergência
- 3.7 Análise da complexidade
- 3.8 Experiência numérica

4 - Decomposição de problemas de programação linear de grande porte utilizando pontos interiores

- 4.1 Princípio de decomposição de Dantzig-Wolfe
- 4.2 Algoritmos
 - 4.2.1 O centro analítico e a função potencial
 - 4.2.2 Algoritmos para a decomposição de problemas de programação linear
- 4.3 Convergência

5 - Conclusões

Referências bibliográficas

Relação de quadros e figuras

- Quadro I - Algoritmo conceitual, baseado na noção de centro, para resolver problemas de otimização
- Quadro II - Algoritmo de relaxação
- Quadro III - Método dos centros analíticos para resolver problemas de programação linear
- Quadro IV - Algoritmo de Shor com dilatação de espaços
- Quadro V - Algoritmo para a programação convexa não suave que utiliza um método parcial de centros analíticos
- Quadro V.a- Reinicialização. Aplicação parcial do método de centros analíticos para recuperar a proximidade ao centro analítico da região perturbada.
- Quadro V.b - Obtenção do ponto próximo ao centro analítico de $S_{\rho^{k+1}}^l$
- Quadro VI - Códigos e autores
- Quadro VII - Valor da solução ótima, valor alcançado pelos algoritmos e dimensão (n) do problema
- Quadro VIII.a - Comparações com o código CA-1
- Quadro VIII.b - Numero de iterações
- Quadro IX - Algoritmo para a programação côncava não suave aplicado à Decomposição de Dantzig-Wolfe
- Quadro IX.a - Obtenção de um subgradiente, cálculo de $D(u_k)$ e determinação de d_{k+1} para a decomposição de Dantzig-Wolfe para o algoritmo do quadro IX
- Quadro X - Algoritmo que utiliza soluções parciais dos programas mestre e escravo
- Quadro X.a - Cálculo de uma solução parcial do problema escravo
- Quadro X.b - Obtenção de um ε -subgradiente, cálculo de $D_u(u_k)$ e determinação de \bar{d}_{l+1}
- Figura 1 - Ponto próximo ao centro analítico da região $S_{\rho^k}^l$
- Figura 2 - Introdução do corte $s_{l+1}^T x - r \leq d_{l+1}$, gerado por um subgradiente em (x^k, r^k)
- Figura 3 – Reinicialização a- relaxação do corte; b- aplicação parcial do método de

Figura 4 - Atualização do limite superior da solução ótima

Figura 5 – a. Aplicação de τ iterações do método de centros analíticos para obter (x^{k+1}, r^{k+1}) ; b. Nova região que contém a solução ótima

Capítulo 1

INTRODUÇÃO

A noção de centro tem tido um papel fundamental na teoria da otimização. Utilizando centros geométricos de polítopos, Levin obteve um algoritmo com convergência geométrica para minimização restrita de funções diferenciáveis, que independe da forma da função e das dimensões do problema (Levin, 1965). O algoritmo é conceitual, pois não se tem notícia de uma forma eficiente para calcular centros geométricos de polítopos gerais. Em (Khachian, 1979) foi estabelecida a polinomialidade de problemas de programação linear com coeficientes inteiros utilizando como pontos de teste centros de elipsóides de volumes decrescentes. Atualmente, há grande atividade na área de pontos interiores, uma classe de algoritmos explorada a partir dos trabalhos de (Karmakar, 1984). Nesta, o algoritmo de centros (Huard, 1967) vem desempenhando um papel importante, notadamente na resolução de problemas de programação convexa não suave (Kiwiel, 1996), (Nesterov, 1995), (Atkinson, Vaidya, 1995), (Bahn, du Merle, Goffin, Vial, 1995), (Goffin, Haurie, Vial, 1992).

Um revisor anônimo ao analisar o artigo “Maintaining closedness to the analytic center of a polytope by perturbing added hyperplanes” (Feijoo, Sanchez, Gonzaga, 1997)

sugeriu a aplicação do resultado principal na programação não diferenciável. Esta sugestão está cristalizada no capítulo 3 do presente trabalho, onde um novo algoritmo para a programação convexa não suave é apresentado. Antes, apresentamos uma versão mais geral e exata do resultado obtido por aqueles autores, com uma prova mais simples. As principais características do algoritmo para a programação não suave são a estabilidade e o uso parcial do método de centros analíticos. Dada uma região convexa que contém, pelo menos, uma solução do problema, a estabilidade é alcançada utilizando pontos próximos à trajetória central de um politopo que é construído a partir do politopo original e de cortes gerados por subgradientes. O método provê uma forma de controlar o feixe de informações, incrementado em cada iteração, através uma técnica “build-down”, que o mantém dentro de limites razoáveis. A convergência é estabelecida e são apresentados testes com problemas típicos da literatura. Este testes vêm sinalizando a efetividade do método proposto.

Este trabalho está organizado em cinco capítulos dos quais este é o primeiro. No que se segue tem-se a descrição de vários algoritmos baseados na noção de centro. Ele foi escrito com a intenção de dar uma idéia geométrica precisa da forma com que esta noção é aplicada em vários métodos de otimização. Neste sentido, ele é mais exploratório e visa identificar e/ou sugerir relações entre os diversos algoritmos. Um novo método para resolver problemas de programação convexa não suave é proposto no capítulo 3. O capítulo 4 discute o uso deste algoritmo no que é usualmente considerada a principal aplicação da programação convexa não diferenciável, os problemas de decomposição linear. São apresentados dois novos métodos para a decomposição de problemas de programação linear de grande porte, segundo o princípio de Dantzig-Wolfe. Em um deles, em cada iteração, o programa mestre é parcialmente resolvido e pressupõe que os subproblemas, originados da decomposição, sejam integralmente resolvidos. O outro resolve parcialmente, tanto o programa mestre, quanto os subproblemas. Finalizando, têm-se as Conclusões e as Referências Bibliográficas.

Capítulo 2

ALGORITMOS DE OTIMIZAÇÃO BASEADOS NA NOÇÃO DE CENTRO

Vários algoritmos utilizam, em cada iteração, um ponto considerado como centro, ou no sentido geométrico, ou em algum sentido analítico, de uma região convexa limitada onde está contida a solução. Ele é o ponto de teste para o qual estes algoritmos se voltam na ausência de informações sobre o elemento procurado. Um plano secante separa a região de viabilidade, mas não é a sua presença que permite caracterizar os métodos que vão ser apresentados; é muito mais a forma de escolher os centros. Tanto a velocidade quanto a convergência dependem do conceito de centro adotado.

Este capítulo foi escrito com a intenção de dar uma idéia dos métodos de otimização que utilizam a noção de centro em poliedros limitados, esferas e elipsóides. Optou-se por fazer uma apresentação geométrica e de caráter didático. Os métodos foram selecionados e organizados enquanto exemplos de utilização de um conceito para resolução de problemas de otimização; neste sentido, não existe a preocupação em expor com rigor aspectos teóricos.

Geometricamente, os métodos baseados na noção de centro realizam um corte em S^k , uma região convexa limitada que contém a solução ótima. Cada corte utiliza um hiperplano H^k ou “plano secante”, que separa S^k em dois volumes; naquele em que x^* , uma solução do problema, está contida, são realizadas operações, de acordo com o método que está sendo utilizado, para determinar o próximo centro ou ponto de teste. O algoritmo da biseção para busca unidimensional, é o exemplo mais simples ($n = 1$) de método baseado na noção de centro. O algoritmo conceitual encontra-se no quadro I.

algoritmo baseado na noção de centro

Obtenha x^0 , centro de S^0

$k := 0$

repita

se “ x^k é solução ótima”

então interrompa

fim se

determine o “plano secante” H^{k+1}

determine S^{k+1}

calcule o centro $x^{k+1} \in S^{k+1}$

$k := k + 1$

fim repita

fim algoritmo

Quadro I

Algoritmo conceitual, baseado na noção de centro,
para resolver problemas de otimização

Este capítulo está organizado em três seções. Na seção seguinte são detalhadas as formas de obter o centro, o plano secante e as regiões S^k nos métodos que utilizam uma noção geométrica de centro (Seções Centrais, Esferas e Elipsóides); o método dos

Centros Analíticos encontra-se descrito em 2. A seção 3 discute os métodos atribuídos a Shor que utilizam a noção de centro.

2.1 Métodos das Seções Centrais, Esferas e Elipsóides

O método das Seções Centrais (Levin, 1965) tem uma taxa de convergência considerada limite para os algoritmos baseados na noção de centro, segundo Judin e Nemirovskii, citados em (Polyak, 1978); nenhum algoritmo, utilizando as mesmas informações, obtém melhores resultados.

Seja o problema

$$\begin{aligned} & \text{Minimizar } f(x) & (1) \\ & x \in S \end{aligned}$$

onde $f(x)$ é diferenciável e S é um poliedro convexo. Na iteração k , x^k é o centro de gravidade de um poliedro S^k no qual está contida a solução ótima. Um hiperplano suporte ao epigrafo de $f(x)$ em x^k , representado por

$$H^k = \{x \in R^n : \langle \nabla f(x^k), x - x^k \rangle = 0\},$$

divide o espaço em dois sub-espacos; em um deles está contida a solução do problema.

Seja S^k a porção do politopo contida em $\langle \nabla f(x^k), x - x^k \rangle < 0$. A nova aproximação da solução é dada por:

$$x^{k+1} := \text{Centro de Gravidade } (S^k).$$

Usando o lema de Mitiagin (Mitiagin, 1968), Levin mostrou que a solução x^* de (1) é envolvida por politopos cujos volumes diminuem a uma razão $q < 1$:

$$\text{volume}(S^{k+1}) \leq q \cdot \text{volume}(S^k), \quad q = 1 - e^{-1}.$$

O volume de S^{k+1} tende a zero segundo uma progressão geométrica e não depende da forma da função diferenciável e nem das dimensões do problema.

Infelizmente, neste algoritmo dois procedimentos não triviais têm de ser usados - um para determinar as restrições não redundantes e um outro para calcular o centro de gravidade do politopo - que são da complexidade do problema original. Na prática, ele ainda é impossível de ser aplicado.

Os métodos de relaxação (Motzkin, Shoenberger, 1954), (Agmon, 1954), resolvem o problema da viabilidade linear:

determinar x :

$$a_i^T x \geq b_i, \quad i = 1, \dots, m, \quad (2)$$

onde $a_i \in R^n$ e $b_i \in R$. Lembremos que todo problema de programação linear pode ser colocado nesta forma através das condições de viabilidade primal, viabilidade dual e do teorema de dualidade forte (Oettli, 1972).

Em uma iteração k (ver quadro II) é verificada a viabilidade do ponto x^k . Se a restrição $a_i^T x \geq b_i$ é violada por x^k , é feita uma projeção ortogonal, em termos de norma euclideana, de x^k no hiperplano

$$H_i = \{x \in R^n : a_i^T x = b_i\}.$$

Usualmente, esta projeção é relaxada por um parâmetro λ e x^{k+1} calculado por:

$$x^{k+1} = x^k + \lambda(x_p^k - x^k), \quad (3)$$

onde

$$x_p^k = x^k + a_i^T \frac{(b_i - a_i^T x^k)}{\|a_i\|^2}. \quad (4)$$

Substituindo (4) em (3) obtém-se:

$$x^{k+1} = x^k + \lambda \frac{a_i^T (b_i - a_i^T x_k)}{\|a_i\|^2}.$$

algoritmo de Relaxação

escolha x^0 arbitrariamente

$k := 0$

repita

se “ x^k é viável”

então interrompa

fim se

$H^{k+1} :=$ hiperplano associado à restrição mais violada

calcule x_p^{k+1} , a projeção de x^k no hiperplano H^{k+1}

$x^{k+1} := x^k + \lambda(x_p^k - x^k)$, $\lambda \in (0,2)$

$k := k + 1$

fim repita

fim algoritmo

Quadro II

Os valores atribuídos a λ dão nome aos métodos. Se $\lambda = 1$, tem-se o método das projeções; $\lambda = 2$ tem-se o das reflexões. Já λ nos intervalos $(1,2)$, $(0,1)$ e $(2, +\infty)$ correspondem aos métodos das subreflexões, das subprojeções e das sobreflexões.

Em Motzkin e Shoenberg, λ é igual a dois. Já em Agmon, $\lambda \in (0,2)$. Neste último, se o conjunto de soluções de (2) é não vazio e, a cada iteração é escolhida a restrição mais violada, a taxa de convergência é linear. Além do mais, mostra-se que cada iteração fica mais próxima (a uma razão fixa) do conjunto de soluções viáveis que seu antecessor.

O algoritmo de Agmon, segundo uma interpretação de (Bland, GoldFarb, Todd, 1981), é um método baseado na noção de centro onde S^k é uma esfera. Suponha que, inicialmente, $x^0 \in S^0$ e que S^0 contenha alguma solução para o problema¹. Em uma iteração típica, escolhe-se para centro um ponto no segmento $(x^k, x^k + 2(x_p^k - x^k))$, onde x_p^k é a projeção de x^k no hiperplano associado à desigualdade mais violada. É sempre possível obter S^{k+1} tal que $\{x \in R^n : a_i^T x \geq b_i\} \cap S^k \subset S^{k+1}$.

O método pode consumir um número exponencial de iterações (em relação aos dados) para alcançar uma solução (Goffin, 1982).

Em resposta à dificuldade de obter centros de polítopos, (Judin, Nemirovskii, 1979) propôs o algoritmo dos Elipsóides, que foi utilizado por (Khachian, 1979) para resolver uma importante questão teórica da programação matemática. Foi a partir de seu trabalho que se pôde, definitivamente, classificar os problemas de programação linear como pertencentes à classe daqueles cuja solução pode ser obtida em tempo polinomial.

Em Khachian, o rito do centro é seguido construindo-se elipsóides S^k que contêm solução do sistema de inequações lineares. Na iteração $(k+1)$ o método testa se o centro de S^k satisfaz as restrições. Se isto ocorre, o método termina. Senão, alguma restrição violada por x^k , por exemplo, $a_i^T x \geq b_i$, é escolhida e o elipsóide de mínimo volume que contém o semi-elipsóide

$$\{x \in S^k : a_i^T x \leq a_i^T x^k\}$$

é construído. Algebricamente, pode-se representar S^k por

$$S^k = \{x \in R^n : (x - x^k) B_k^{-1} (x - x^k) \leq 1\}$$

onde x^k é o centro e B_k uma matriz simétrica definida positiva. Em função do cálculo de x^{k+1} , tem-se (ver apêndice B (Bland, Goldfarg, Todd, 1981)):

¹ No algoritmo, isto é desnecessário.

$$x^{k+1} = x^k - \tau \left(\frac{B_k a_i}{\sqrt{a_i^T B_k a_i}} \right) e$$

$$B_{k+1} = \delta \left(B_k - \frac{\sigma B_k a_i (B_k a_i)^T}{a_i^T B_k a_i} \right)$$

onde $\tau = \frac{1}{n+1}$, $\sigma = \frac{2}{n+1}$, $\delta = \frac{n^2}{n^2-1}$.

Embora este tenha sido o primeiro método reconhecido como polinomial para programação linear e tenha tido conseqüências teóricas importantes, na prática, ele mostrou-se surpreendentemente lento.

Atualmente, há grande atividade na área de pontos interiores, uma classe de algoritmos que alia a polinomialidade à eficiência (por exemplo, para programação linear, ver (Zang, 1996)). Um exemplo desta classe é visto a seguir.

2.2 Método dos Centros Analíticos

O centro analítico é o único máximo de uma função distância $d(x, S)$. Ele é fácil de ser computado e está associado a algoritmos de pontos interiores (Hertog, 1992) para resolver problemas de otimização.

Com o centro deixando de ser um conceito geométrico, além da facilidade de cálculo, também existe a vantagem de eliminar a necessidade de especificar o conjunto mínimo de desigualdades que representam o poliedro (em problemas onde a região de viabilidade é definida por inequações).

Uma função contínua $d(x, S)$ para medir a distância de um ponto x , em um convexo S , à fronteira de S tem as seguintes propriedades (Huard, 1967):

- i) $d(x, S) = 0$ para todos os pontos de fronteira de S ;
- ii) $d(x, S) > 0$ para todos os pontos interiores a S ;
- iii) se \tilde{S} é o domínio obtido por uma translação de uma restrição de S , tal que \tilde{S} está contido em S , então $d(x, \tilde{S}) \leq d(x, S)$, para todo $x \in \tilde{S}$ e,
- iv) seja $a_0(x) \geq 0$ a restrição considerada em iii), sendo $a_0(x, \lambda)$ a translação parametrizada por λ , que determina a posição da restrição, e seja \tilde{S}_λ o domínio correspondente. Se $x(\lambda)$ é um ponto de \tilde{S}_λ , tem-se que:

se $a_0(x(\lambda), \lambda) \rightarrow 0$ quando $\lambda \rightarrow +\infty$, então

$$d(x(\lambda), \tilde{S}_\lambda) \rightarrow 0 \text{ quando } \lambda \rightarrow +\infty.$$

Dado um domínio convexo S e uma função $d(x, S)$ definida neste domínio, dá-se o nome de Centro Analítico ao ponto x_c ,

$$x_c : d(x_c, S) = \max\{d(x, S) : x \in S\}.$$

Seja o problema

$$\text{Minimizar } z = f(x) \tag{5}$$

sujeito a:

$$a_l(x) \geq 0, \quad l \in L$$

e ρ^k um limitante estritamente superior para a solução ótima de (5). Huard (Huard, 1967) propôs duas funções para calcular as distâncias:

$$a) \quad d(x, S) = [\rho^k - f(x)] \prod_l a_l(x) \text{ e}$$

$$b) d(x, S^k) = \min\{\rho^k - f(x), a_l(x) : l \in L\}.$$

Seja a função $d(x, S^k)$ definida segundo (a) e $i = 1, 2, \dots, m$. Fazendo

$$f_0(x) = \rho^k - f(x) \geq 0 \text{ e}$$

$$f_i(x) = a_i(x) \geq 0, \quad i = 1, 2, \dots, m,$$

tem-se

$$d(x, S^k) = f_0(x) \prod_{i=1}^m f_i(x),$$

É também possível considerar

$$e^{d(x, S^k)} = e^{f_0(x) \prod_{i=1}^m f_i(x)},$$

ou

$$d(x, S^k) = \text{Ln}(f_0(x)) + \sum_{i=1}^m \text{Ln}(f_i(x)).$$

Esta foi utilizada em (Sonnevend, 1985), atribuindo um peso q à função objetivo:

$$d(x, S^k) = q \text{Ln}(f_0(x)) + \sum_{i=1}^m \text{Ln}(f_i(x)).$$

Pode-se interpretar que a restrição

$$f_0(x) = \rho^k - f(x) \geq 0$$

é replicada q vezes. Com isto o centro analítico é deslocado em uma direção associada à $f(x)^2$.

Quanto à forma (b) da função distância, não se tem notícia de sua implementação.

A seguir tem-se, sucintamente, o método primal dos centros analíticos para a programação linear.

Seja o problema

$$\text{Min } c^T x : x \in S, \quad (6)$$

onde $S := \{x \in R^n : a_i^T x \leq b_i, i = 1, 2, \dots, m\}$, $c, a_i \in R^n$ e $b_i \in R$. Suponha que $\text{int}(S) \neq \emptyset$. Para $\rho > c^T x^*$, onde x^* é o valor ótimo de (6), $x \in \text{int}(S)$, $q \geq m + 1$ defina a função barreira logarítmica associada a S :

$$f_\rho(x) = -q \text{Ln}(\rho - c^T x) - \sum_{i=1}^m \text{Ln}(b_i - a_i^T x)$$

O método dos Centros Analíticos associa a (6) a seguinte família de problemas não lineares:

$$\text{Min } f_\rho(x) : x \in \text{int}(S). \quad (7)$$

Seja

$$x(\rho) := \arg \min \{f_\rho(x) : x \in \text{int}(S_\rho)\},$$

com S_ρ dado por

² A determinação do valor ótimo para q provém da análise da complexidade do método - ver (Den Hertog, 1992).

$$S_\rho = \{x \in S : \underbrace{c^T x \leq \rho, \dots, c^T x \leq \rho}_{q \text{ vezes}}\},$$

onde a restrição $c^T x \leq \rho$ é repetida q vezes e S_ρ diferente de vazio e limitado.

A curva contínua que une os pontos $x(\rho)$, é conhecida como trajetória central e cada $x(\rho)$ é o centro analítico de S_ρ .

O algoritmo está no quadro III. Dado ρ , um limitante superior do valor ótimo de (6), com iterações do tipo Newton-Raphson, procura-se um ponto na vizinhança de $x(\rho)$, centro analítico da região atual. Encontrado este ponto, o limitante superior é decrescido e as q restrições, relacionadas à função objetivo, são removidas. Outras q restrições, associadas ao novo limitante, são incorporadas, definindo uma nova região. Este procedimento é repetido até que seja alcançado um ponto próximo à trajetória central e suficientemente perto da solução ótima.

As iterações de Newton-Raphson para alcançar a vizinhança de centro analítico, são dadas por:

$$h := h_\rho(x) = -[H_\rho(x)]^{-1} g_\rho(x),$$

onde g_ρ é o gradiente de f_ρ , dado por

$$g_\rho(x) = \nabla f_\rho(x) = \left(\frac{q}{\rho - c^T x} c + \sum_{i=1}^m \frac{1}{b_i - a_i^T x} a_i \right)$$

e H_ρ é a Hessiana:

$$H_\rho(x) = \left(q \frac{cc^T}{(\rho - c^T x)^2} + \sum_{j=1}^m \frac{a_j a_j^T}{(b_j - a_j^T x)^2} \right).$$

algoritmo de centros analíticos

escolha $tol > 0$, $\theta, \varepsilon \in (0,1)$

escolha $\rho^0 : \rho^0 > c^T x^*$, $x^0 : x^0 \in \text{int}(S_{\rho^0})$ e $\delta_{\rho^0}(x^0) < \varepsilon$

$k := 0$

repita

se $(c^T x^k - \rho^k) \leq tol$

então interrompa

fim se

{ determine o "plano secante" $H^{k+1} := \{x \in R^n : c^T x = \rho^{k+1}\}$

$\rho^{k+1} := c^T x^k + \theta(\rho^k - c^T x^k)$

{ calcule o centro $x^{k+1} \in S^{k+1}$ }

$z := x^k$

repita

$h := \text{Newton_Raphson}(\rho^{k+1}, z)$

$v := \|h\|_{H_{\rho^{k+1}}}$

se $v \leq \varepsilon$

então interrompa

fim se

$\bar{\alpha} := \text{Busca_Linear}_{\alpha \geq 0} \{f_{\rho^{k+1}}(z + \alpha h) : z + \alpha h \in S\}$

$z := z + \bar{\alpha}h$

fim repita

$x^{k+1} := z$

$k := k + 1$

fim repita

fim algoritmo

Quadro III

Método dos centros analíticos para
resolver problemas de programação linear

A matriz $H = H_{\rho}(x)$ é simétrica e definida positiva e induz a norma

$$\|x\|_H = \sqrt{x^T H x}.$$

Dados ρ e $\bar{x} \in \text{int}(S_\rho)$, diz-se que \bar{x} está próximo à trajetória central se

$$\delta_\rho(\bar{x}) = \|h_\rho(\bar{x})\|_H < \varepsilon$$

para algum $\varepsilon \in (0, 1)$.

Conhecidos ρ^0 , um limitante superior para o valor da função objetivo e $x^0 \in \text{int}(S_{\rho^0})$, o método é inicializado com a subsequente aproximação do ponto inicial à trajetória central.

O método tem a complexidade aritmética da ordem de $O(n^{2.5} L)$ (Vaidya, 1990), onde L é o tamanho da entrada, isto é, o total de espaço necessário para armazenar os coeficientes do problema (ver (Khachian, 1979)).

2.3 Métodos de Shor

Os métodos oriundos da chamada escola russa, genericamente conhecidos como de Shor ou do subgradiente, utilizam o conceito de subgradiente para relaxar a necessidade de diferenciabilidade de $f(x)$. Suponha f convexa. Um subgradiente s_k é um elemento do conjunto subdiferencial $\partial f(x^k)$ definido como

$$\partial f(x^k) := \{s \in R^n : f(x) \geq f(x^k) + \langle s, x - x^k \rangle\}.$$

Desta definição tem-se, como consequência, um elemento básico dos algoritmos baseados na noção de centro: o hiperplano

$$H^k := \{x \in R^n : \langle s_k, x - x^k \rangle = 0\}$$

que separa S^{k+1} , uma região que contém a solução ótima. Não é por acaso que os métodos de melhor desempenho desta família, são aqueles que conseguem incorporar esta noção.

Os algoritmos de Shor não seguem o paradigma do decréscimo da função objetivo - uma direção $-s_k : s_k \in \partial f(x^k)$, nem sempre é de descida. No método

$$x^{k+1} := x^k - t^k \frac{s_k}{\|s_k\|}, \quad (8)$$

um passo t^k positivo não garante que $f(x^{k+1}) < f(x^k)$. Por outro lado, $-s_k$ é uma direção de descida para uma função $d(x)$ que mede a distância de x ao conjunto de valores ótimos de $f(x)$: um passo positivo convenientemente escolhido em (8) gera um ponto mais próximo da solução (veja teorema 1 a seguir), isto é, $d(x^{k+1}) < d(x^k)$.

Teorema 1

Seja o problema

$$\text{Min } f(x), \quad x \in R^n. \quad (\text{P})$$

Se x^* é a única solução de (P), então qualquer subgradiente de f em x é uma direção de descida para $d(x) = \|x - x^*\|$, para $x \neq x^*$.

Demonstração: (Shor, 1985)

■

Sem a busca linear para calcular o tamanho do passo, o recurso é pré-definir, de alguma forma, o tamanho do deslocamento. Mostramos a seguir algumas técnicas mais tradicionais. Posteriormente, na seção 2.3.2, apresentaremos uma generalização do método de subgradiente, que pode ser vista como um acelerador, os chamados métodos de subgradientes com dilatação do espaço.

Na discussão que se segue, supõe-se que não é explicitado todo o conjunto subdiferencial em x^k para escolher uma direção de descida.

2.3.1 A escolha do passo

Manter fixo o valor do passo em cada iteração não é uma boa estratégia. A seqüência gerada, $\{x^0, x^1, \dots, x^k\}$, fica no interior de uma bola, de raio proporcional ao passo, sem alcançar $x^* = \min f(x)$ (ver teorema 2, a seguir).

Teorema 2

Seja f uma função convexa de R^n em R . Se $t^k = \frac{t}{\|s^k\|}$ então a seqüência

$\{x^1, x^2, \dots, x^k\}$, gerada por

$$x^{k+1} := x^k - t^k \frac{s_k}{\|s_k\|},$$

é tal que existem $\varepsilon > 0$ e k onde:

$$\|x^k - x^*\| \leq t \frac{(1 + \varepsilon)}{2}$$

Demonstração: (Shor, 1985). ■

Variando o passo de modo que

$$t^0 + \dots + t^{k-1} + t^k \leq R,$$

com $R \neq \infty$ e $t^k \neq 0$, na iteração k tem-se:

$$\begin{aligned} \|x^0 - x^k\| &\leq \|x^0 - x^1\| + \|x^1 - x^2\| + \dots + \|x^{k-1} - x^k\| = \\ &= t^0 + \dots + t^{k-1} + t^k \leq R. \end{aligned}$$

Toda iteração fica no interior de uma esfera de raio R ; a solução só é alcançada se x^0 estiver suficientemente próximo da solução. Se a série for divergente, $\sum_{k=0}^{\infty} t^k = +\infty$, com

t^k tendendo para zero quando k vai para infinito, o método converge a partir de um x^0 arbitrário, conforme o

Teorema 3

Seja f uma função convexa de R^n em R . Se t^k satisfaz

$$i) \quad t^k \rightarrow 0 \text{ quando } k \rightarrow \infty \text{ e}$$

$$ii) \quad \sum_{k=0}^{\infty} t^k = +\infty,$$

então a sequência $\{x^1, x^2, \dots, x^k\}$, gerada por $x^{k+1} := x^k - t^k \frac{s_k}{\|s_k\|}$

é tal que

$$\lim (\inf (f(x^k))) = \min f(x) = f^*.$$

Demonstração: veja, por exemplo, (Minoux, 1983)

■

Observa-se, no entanto, que este último método é excessivamente lento (ver exemplos em (Minoux, 1983)).

Com a noção de centro, incorporada sob certas condições, são obtidos algoritmos mais eficientes, como é caso de um método visto a seguir, cuja série $\{t^k\}$ é convergente.

Seja f^* o valor ótimo de uma função convexa f e M o seu conjunto de valores mínimos:

$$M := \{x: f(x) = f^*\}.$$

A distância entre $x \notin M$ e o conjunto M é dada por

$$d(x) = \|P(x) - x\|,$$

onde $P(x)$ é a projeção de x em M . O índice de condicionamento χ (Goffin, 1980) de uma função convexa f é, por definição, o cosseno do maior ângulo existente entre $-s \in \partial f(x)$ e a direção $P(x) - x$, em um ponto $x \notin M$. Mais precisamente,

$$\chi := \inf_{x \notin M} \inf_{s \in \partial f(x)} \left\{ \frac{-s^T (P(x) - x)}{\|s\| \|P(x) - x\|} \right\}.$$

Considere S^0 uma esfera de raio R centrada em x^0 contendo o conjunto de soluções x^* , isto é, $\|x^0 - x^*\| \leq R$, para todo $x^* \in M$. Seja C o cone em x^0 definido por

$$C := \{x \in R^n : \langle x, -s^0 \rangle \leq \chi \|x\| \|s^0\|\},$$

onde $\chi = \cos \phi$ é o índice de condicionamento de f . Pela definição de χ , sabe-se

que $M \in S^0 \cap C$. Fazendo $x^1 = x^0 - t^0 \frac{s_0}{\|s_0\|}$,

$$t^0 := R \cos \phi = R\chi,$$

uma nova esfera centrada em x^1 e com raio $R \sin \phi = R\sqrt{1 - \chi^2}$ contém x^* . Depois da primeira iteração, podem ser obtidas esferas S^k , centradas em x^k cujos raios decrescem à razão $\sqrt{1 - \chi^2}$. A sequência $\{x^k\}$ corresponde ao método da série convergente onde $\{t^k\}$ é uma progressão geométrica com

$$t^k = t^0 \left(\sqrt{1 - \chi^2} \right)^k, \quad k = 1, 2, \dots$$

$$t^0 = R\chi$$

Na prática, não é conhecido χ ; isto impede obter o melhor t^k em cada iteração.

2.3.2 Algoritmos com transformação de espaços

O índice de condicionamento χ de f (Goffin, 1980) dá uma indicação do desempenho dos métodos do subgradiente. Com este índice próximo de zero, é possível que o ângulo entre o oposto a um subgradiente em um ponto x , e a direção de x à solução ótima, ser próximo de 90 graus, prejudicando a convergência. Este ângulo pode ser alterado com uma transformação de espaços.

Seja $B_k : R^n \rightarrow R^n$ uma transformação linear não singular e $x = B_k y$. A função $f(x)$ passa a ser

$$f(x) = f(B_k y) = \varphi(y).$$

Da definição de subgradiente, tem-se:

$$f(x) \geq f(x^k) + \langle s_k, x - x^k \rangle \Rightarrow$$

$$\varphi(y) \geq \varphi(y^k) + \langle s_k, B_k (y - y^k) \rangle,$$

onde $B_k^T s_k$ é um subgradiente para $\varphi(y)$.

O método do subgradiente com uma transformação linear $x = B_k y$, passa a ser:

$$y^{k+1} = y^k - t^k \frac{B_k^T s_k}{\|B_k^T s_k\|}.$$

Para mapear a transformação no espaço original, basta ver que

$$x^{k+1} = B_k y^{k+1} \Rightarrow$$

$$x^{k+1} = B_k \left(y^k - t^k \frac{B_k^T s_k}{\|B_k^T s_k\|} \right) \Rightarrow$$

$$x^{k+1} = x^k - t^k \frac{B_k B_k^T s_k}{\|B_k^T s_k\|}, \quad (13)$$

onde $B_k B_k^T$ é uma matriz simétrica definida positiva.

algoritmo de Shor com dilatação de espaços

escolha x^0

$k := 0$

$B_0 := I$

repita

se “ x^k é solução ótima”

então interrompa

fim se

escolha $s_k \in \partial f(x^k)$

escolha t^k

$$x^{k+1} = x^k - t^k \frac{B_k B_k^T s_k}{\|B_k^T s_k\|}$$

escolha α_k

$$B_{k+1} := B_k R \alpha_k (d_k)$$

$k := k + 1$

fim repita

fim algoritmo

Quadro IV

Algoritmo de Shor com dilatação de espaços

Os métodos de Shor que utilizam uma transformação especial, conhecida como dilatação de espaços, são descritos a seguir. Em particular, estamos interessados em uma instância destes métodos; aquela cujas operações correspondem a um método baseado na noção de centro.

Dada uma direção unitária $d \in R^n$, o operador de dilatação de espaços $R_\alpha(d)$, na direção d com coeficiente $\alpha > 1$ é definido como

$$R_\alpha(d) := x + (\alpha - 1)x_p,$$

onde $x_p = (x^T d)d$ é a projeção de x na direção d . De forma matricial, pode-se escrever $R_\alpha(d)$ como

$$R_\alpha(d) = [I + (\alpha - 1)dd^T],$$

onde I é a matriz identidade.

Seja $X = \langle d \rangle = \{\alpha d : \alpha \in R\}$. Então $R^n = X \oplus X^\perp$ e $x \in R^n$ pode ser unicamente escrito como $x = x_d + \bar{x}_d$, com $x_d \in X$ e $\bar{x}_d \in X^\perp$.

Portanto, $R_\alpha(d)x = \alpha x_d + \bar{x}_d$, isto é, $R_\alpha(d)$ multiplica a d -componente x_d por α e deixa o complemento ortogonal fixo.

O operador de dilatação de espaços tem as seguintes propriedades (Akgull, 1984) e (Shor, 1970):

$$i) \quad R_{\alpha\beta}(d) = R_\alpha(d) \cdot R_\beta(d);$$

$$ii) \quad R_\alpha(d).R_{1/\alpha}(d) = R_1(d) = I, \alpha \neq 0;$$

$$iii) \quad \|R_\alpha(d)x\|^2 = \|x\|^2 + (\alpha^2 - 1).(\langle x, d \rangle)^2 \quad e,$$

$$iv) \quad \det R_\alpha(d) = \alpha.$$

Com dilatação de espaços a matriz B_{k+1} é atualizada em cada iteração por

$$B_{k+1} := B_k R_{\alpha_k}(d_k) \quad (14)$$

No algoritmo do quadro IV, observa-se que em cada iteração é computado o produto $B_{k+1}B_{k+1}^T$, que resulta em uma matriz simétrica definida positiva. As fórmulas de Sokov (Sokov, 1974) levam este fato em conta e reescrevem x^{k+1} , sugerindo um método “quase Newton”, induzido pela métrica desta matriz.

Seja

$$H_{k+1} := B_{k+1}B_{k+1}^T,$$

substituindo B_{k+1} de (14), tem-se:

$$H_{k+1} = B_k R_{\alpha_k}(d_k) R_{\alpha_k}(d_k) B_k^T$$

que, pela propriedade i) fica

$$H_{k+1} = B_k R_{\alpha_k^2} B_k^T \Rightarrow$$

$$H_{k+1} = B_k [I + (\alpha_k^2 - 1) d_k d_k^T] B_k^T \Rightarrow$$

$$H_{k+1} = H_k + (\alpha_k^2 - 1)B_k d_k d_k^T B_k^T. \quad (15)$$

Fazendo

$$d_k = \frac{B_k^T s_k}{\|B_k^T s_k\|}$$

e substituindo-se em (15), tem-se:

$$H_{k+1} = H_k + \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} (\alpha_k^2 - 1).$$

Analogamente, chega-se à fórmula para atualização de x^{k+1} :

$$x_{k+1} = x_k - t_k \frac{H_k s_k}{\sqrt{s_k^T H_k s_k}}$$

Às diversas definições de t_k , α_k e s_k , correspondem diferentes métodos (ver (Minoux 1983)). Particularmente (Shor, 1977), fazendo

$$\alpha_k = \alpha = \sqrt{\frac{n-1}{n+1}}$$

e

$$t_k = t_{k-1} \frac{n}{\sqrt{n^2 - 1}},$$

com $t_0 = \frac{R}{n+1}$, onde R é o raio de uma esfera que contém x^0 e, pelo menos, uma solução ótima do problema, tem-se um algoritmo que pode ser interpretado como o método dos elipsóides para resolver problemas de programação convexa em R^n .

Capítulo 3

UM NOVO MÉTODO PARA A PROGRAMAÇÃO CONVEXA NÃO SUAVE

Uma possível classificação divide a otimização entre a linearidade e a não linearidade; entretanto, segundo (Rockafellar, 1992), o grande divisor de águas não se dá aí; ele está entre a convexidade e a não convexidade. Uma razão importante para isto é que quando funções convexas são minimizadas em conjuntos convexos, uma solução local também é global.

Ao lado da importância intrínseca da programação convexa, observa-se que vários casos particulares de otimização podem ser tratados diretamente, enquanto problemas de programação convexa e se beneficiarem de técnicas mais gerais. Particularmente, este é o caso da decomposição de problemas de programação linear de grande porte (capítulo 4).

Apresenta-se um novo algoritmo para resolver problemas de otimização convexa não suave. Dada uma função convexa f , não necessariamente diferenciável, de R^n em R , o problema da otimização convexa é encontrar x em R^n tal que $f(x) - f(x_{min}) \leq \delta$, para $\delta > 0$ arbitrariamente pequeno e x_{min} um mínimo de $f(\cdot)$.

Este capítulo está organizado da seguinte maneira. A seguir, o problema é explicitado juntamente com o modelo de planos secantes adotado para representá-lo. Detalhes do método de centros analíticos estão na seção 2; a seção 3 mostra como recuperar a proximidade à trajetória central quando a região que contém a solução é perturbada por um corte. O novo algoritmo está em 4; a seção 5 mostra como gerenciar o feixe de informações agregado pelo algoritmo em cada iteração. A convergência é estabelecida na seção 6 seguido da análise da complexidade em 7. A experiência computacional é apresentada no final do capítulo.

3.1 Introdução

3.1.1 Planos secantes

Seja f uma função convexa não necessariamente diferenciável em R^n e seja o seguinte problema:

$$\text{Minimizar } f(x) \text{ em } R^n \quad (1)$$

O subdiferencial de f em x substitui o gradiente na otimização não suave. É usual assumir que para cada x são conhecidos $f(x)$ e algum $s \in \partial f(x)$ (arbitrário), o chamado oráculo.

No passo k de um algoritmo de planos secantes (Kelley, 1960) para resolver o problema, tem-se as iterações x^1, x^2, \dots, x^k geradas, juntamente com os valores $f(x^1), f(x^2), \dots, f(x^k)$ e os subgradientes $s_1 \in \partial f(x^1), \dots, s_k \in \partial f(x^k)$. A aproximação de f por um modelo de planos secantes, associado a este feixe de informações, é a seguinte função seccionalmente afim:

$$x \in R^n \rightarrow f_k^\Delta(x) = \max\{f(x^i) + \langle s_i, x - x^i \rangle, i = 1, 2, \dots, k\}$$

e a próxima iteração x^{k+1} é obtida resolvendo o problema:

$$\begin{aligned} & \text{Minimizar } f_k^\Delta(x) & (2) \\ & x \in S \end{aligned}$$

onde S é um conjunto convexo compacto que contém, pelo menos, um mínimo de f . O problema (2) é equivalente a:

$$\begin{aligned} & \text{Minimizar } r & (3) \\ & \text{sujeito a:} \\ & r \geq f(x^i) + \langle s_i, x - x^i \rangle, \quad i = 1, 2, \dots, k \\ & x \in S \\ & r \in R \end{aligned}$$

Em cada passo o algoritmo computa x^{k+1} e $s_{k+1} \in \partial f(x^{k+1})$. Se

$$f(x^{k+1}) \leq f_k^\Delta(x^k) + \varepsilon,$$

para algum $\varepsilon > 0$, então uma ε -solução foi encontrada e o algoritmo termina. Caso contrário, k é substituído por $k + 1$ e é novamente resolvido o problema relaxado (3), depois de ser acrescido um novo plano secante.

Infelizmente, a convergência é extremamente lenta e o método é instável.

3.1.2 Estabilidade através da proximidade quadrática (feixes)

A metodologia de feixes, veja, por exemplos (Hiriart-Urruty, Lemaréchal, 1993), usa a idéia de planos secantes agregada a um procedimento de estabilização. Na iteração k tem-se um modelo $f_k^\Delta(\cdot)$, que representa f e um centro de estabilização x^k . A próxima

iteração y^{k+1} é obtida buscando os compromissos de diminuir o modelo, $f_k^\Delta(y^{k+1}) < f_k^\Delta(x^k)$, e de permanecer próximo do centro de estabilidade; isto é,

$$\|y^{k+1} - x^k\| < \tau$$

para algum τ a ser fixado.

O centro de estabilidade deve aproximar-se iterativamente, do mínimo de f , (veja (Hirriat-Urruty, Lemaréchal, 1993)).

Há vários algoritmos conceitualmente equivalentes que seguem este esquema; por exemplo, planos secantes com região de confiança. Neles, a sequência de iterações é definida por:

$$y^{k+1} := \arg \min \{ f_k^\Delta(y) : \frac{1}{2} \|y - x^k\| < \frac{\tau^2}{2} \};$$

equivalente a determinar a solução do problema:

Minimizar r

(P_τ^k)

sujeito a:

$$r \geq f(x^i) + \langle s_i, x - x^i \rangle, \quad i = 1, 2, \dots, k$$

$$\frac{1}{2} \|x - x^k\|^2 \leq \frac{\tau^2}{2}$$

onde $\tau \in R$ é dado. Se o modelo f_k^Δ é uma boa aproximação de f em $B(x^k, \tau)$, uma região de confiança em torno do centro de estabilidade, ocorre um decréscimo suficiente de f em y^{k+1} , o passo é dito sério e o centro de estabilidade é atualizado,

$$x^{k+1} := y^{k+1},$$

podendo, ainda assim, haver enriquecimento do modelo através do corte gerado pelo novo ponto. Caso contrário, o centro permanece inalterado e o modelo é enriquecido pelo acréscimo de um corte gerado pelo subgradiente em y^{k+1} . Neste caso, diz-se que o passo é nulo.

3.1.3 Estabilidade através do centro analítico

Evidências experimentais (Sapatnekar, Rao, Vaidya, 1991), (Goffin, Haurie, Vial, 1992), (Bahn, Merle, Goffin, Vial, 1995), têm mostrado a efetividade do centro analítico como ponto de teste em algoritmos de planos secantes. Como foi visto no capítulo 2, a idéia do plano secante está intimamente relacionada com os algoritmos que usam algum tipo de noção de centro, como os métodos das seções centrais (Levin, 1965), dos elipsóides (Khachiyan, 1976) e o de centro analítico (Sonnevend, 1985). A característica comum destes métodos é o uso, em cada iteração, de um ponto considerado como centro (ou ponto de teste), ou no sentido geométrico, ou em algum sentido analítico, de uma região convexa que contém o ótimo. Estas regiões podem ser esferas, elipsóides ou poliedros limitados.

De um modo geral, os métodos que utilizam centros analíticos para resolver o problema (1), podem ser postos em uma escala na qual, em um extremo, estão algoritmos cuja única informação do comportamento da função é um subgradiente arbitrário, computado na iteração corrente para gerar um corte que separa a região que contém a solução ótima - ver, por exemplo, (Nesterov, 1995). Em outro extremo, situa-se o algoritmo proposto neste capítulo. A função objetivo é minimizada por meio de uma função modelo, construída a partir de hiperplanos de suporte obtidos em cada iteração, e um limite superior do valor da solução ótima é monotonamente decrescido. Desta forma, nosso algoritmo se destaca pelo fato de se utilizar efetivamente do método primal de centros analíticos, em cada subproblema de programação linear. Outros algoritmos, situados em uma faixa intermediária nesta escala, não utilizam o conceito de trajetória central; ver (Kiwel, 1996) e (Goffin, Haurie, Vial, 1992).

(Nesterov, 1995) propôs e analisou a complexidade do método de centros analíticos aplicado a (3) com S dado por uma bola centrada no ponto inicial, com raio suficientemente grande que contenha alguma solução. Os passos do algoritmo básico são:

$$1. x^0 \in R^n, \|x^* - x^0\| \leq R, \tau > 0$$

$$2. F_0(x) = \frac{\tau}{2R^2} \|x - x^0\|^2$$

$$3. F_{k+1}(x) = F_k(x) + \frac{1}{2R^2} \|x - x^0\|^2 - Ln \langle g_k, x^k - x \rangle,$$

$$\text{onde } g_k \in \partial f(x^k) \text{ e } x^{k+1} = \arg \min_{x \in R^n} F_{k+1}(x)$$

Nesterov utiliza a hipótese de limitação uniforme dos subgradientes para obter a convergência deste algoritmo. Obtém também, para uma certa variante, a complexidade de $O(N^2n + N^{1.5}n^2)$ operações aritméticas, necessárias para a execução de N iterações do algoritmo, n sendo a dimensão do espaço de variáveis, e o fator constante dependente de determinados parâmetros que definem o processo.

(Kiwel, 1996) considera o problema de obter $f_* = \{\min f(z), 0 \leq z \leq e\}$, f convexa, e sendo o vetor de $\mathbf{1}$, com $f_* \geq 0$ e $f(\frac{e}{2}) = 1$. O autor mostra que nos casos em que a caixa é arbitrária (e limitada), e com um mínimo não verificando a positividade, uma simples transformação leva à sua forma canônica.

O algoritmo parte do hiperplano $[0,1]^n$, com $y \in R^n$, $\zeta \in R$, de modo a chegar ε -solução tal que $f(z) \leq \zeta$, $\zeta \leq f_* + \varepsilon$. Através de sucessivos cortes no hiperplano, e respectivo cálculo de centro analítico, com a função potencial usual, o algoritmo tem basicamente dois procedimentos, em uma dada iteração k , sendo (z^k, ζ_k) o centro atual:

1. Se $f(z^k) < \zeta_k$, atualize a cota superior para o valor $f(z^k)$, passando à iteração seguinte;

2. $f(z^k) \geq \zeta_k$, mantenha a cota e acrescente um novo corte, através do subgradiente obtido em z^k , passando à iteração seguinte.

Kiwiel demonstra a convergência de $\min_{j < k} f(z^j)$ para f_* (algoritmo não é, evidentemente, de descida). Mostra também que nos casos em que f é seccionalmente afim, o decaimento acima é exponencial.

(Goffin, Haurie, Vial, 1992) consideram o problema de minimizar funções seccionalmente afins,

$$\min_x \max_{i \in I} \{d_i - \sum_{j=1}^n g_{i,j} x_j\}$$

$$l \leq x \leq h$$

onde I pode ser infinito. Se $\text{card}(I) = m$, este problema se reescreve como

$$\text{minimizar } r \tag{D}$$

s. a:

$$re + Gx \geq d$$

$$l \leq x \leq h$$

$$(x, r) \in R^{n+1},$$

onde $G = (g_{ij})$, $e \in R^m$ é o vetor de $\mathbf{1}$. O dual de (D) é

$$\text{maximizar } d^T \zeta + e^T \eta - h^T \xi \tag{P}$$

s. a:

$$G^T \zeta + \theta - \xi$$

$$e^T \zeta = 1$$

$$\zeta, \eta, \xi \geq 0,$$

que possui o formato padrão (primal) de Karmarkar.

O algoritmo é iterado sobre funções e politopos (conjuntos de localização) que aproximam o problema original, e que são definidos por

$$f_K(x) = \max_{i \in K} \{d_i - \sum_{j=1}^n g_{ij} x_j\},$$

$$F_K(z_u) = \{(x, r): r \leq z_u, r \geq d_i - \sum_{j=1}^n g_{ij} x_j, i \in K, l \leq u \leq h\}.$$

Através destes, se definem naturalmente (D_K) e (P_K) , similares a (D) e (P) . Os passos básicos são os seguintes:

1. Determinação de um ponto central (x_c, r_c) .

Isto é feito através do uso da versão de (Goffin, Vial, 1989) do algoritmo de Karmarkar aplicada a (P_K) . Como subproduto, o algoritmo provê uma cota inferior para o problema relaxado $\min\{f_K(x): l \leq x \leq h\}$, assim como para o problema original. Desta forma, z_l é atualizado por $z_l = \max\{z_l, z\}$.

2. No ponto x_c se calcula

$$f(x_c) = \max_{i \in I} \{d_i - \sum_{j=1}^n g_{ij} x_j^c\} = d_k - \sum_{j=1}^n g_{kj} x_j^c$$

$$K = K \cup \{k\}$$

$f(x_c)$ é uma cota superior para o problema original. Desta forma, z_u é atualizado por $z_u = \max\{z_u, f(x_c)\}$.

3. Se $z_u - z_l < \varepsilon$, para $\varepsilon > 0$ dado, parar.

A opção pela resolução da forma (P) faz com que entre duas iterações se gere uma nova coluna (eventualmente idêntica à anterior). Como consequência é necessário obter um ponto interior após cada geração de coluna. Os autores se utilizam de um procedimento devido a (Mitchell, Todd, 1992), em que a direção de Karmarkar é modificada de forma a incluir a coluna.

A convergência é assegurada porque o número de formas lineares é finito.

Os autores testaram o algoritmo em problemas seccionalmente afins e não afins, obtendo bons resultados.

No novo algoritmo para a programação convexa não diferenciável apresentado neste capítulo, é implementado o método de Centros (Huard, 1967) em um contexto de planos secantes. Resgata-se o conceito de trajetória central, existe uma seqüência monótona estritamente decrescente para monitorar o desenvolvimento das iterações e é um método primal. Além disto, utiliza um novo mecanismo para a reinicialização.

Outro aspecto importante, geralmente tratado com dificuldade pelos algoritmos, refere-se à gerência das informações que vão sendo acumuladas em cada iteração. É usual sugerir uma estratégia para limitar o crescimento do feixe. Normalmente, são utilizadas técnicas heurísticas. No nosso caso, propomos um critério preciso; são descartadas aquelas restrições, introduzidas em algum momento da execução do algoritmo, que causam impacto desprezível no cálculo do centro analítico. Finalmente, outra característica diferencial é a flexibilidade quanto a forma da região inicial que contém uma solução ótima. Ela pode ser expressa por variáveis canalizadas, por um politopo, ou por regiões quadráticas limitadas como esferas e elipsóides.

3.2 O método do centro analítico

Para uma região fixa S , cada iteração x^k do método de centros analíticos aplicado à (1) está próximo do centro analítico de uma região associada com a solução do problema (2). Seja $d_i := s_i^T x^i - f(x^i)$, $i = 1, 2, \dots, l$ e

$$S^l := S \times R \cap \{(x, r) \in R^{n+1} : s_i^T x - r \leq d_i, i = 1, 2, \dots, l\}$$

Então (3) pode ser escrito como

$$\begin{aligned} & \text{Minimizar } r & (4) \\ & \text{sujeito a:} \\ & (x, r) \in S^l \end{aligned}$$

e o método de centro analítico associa a (4) uma família de problemas não lineares, dada por

$$\text{Minimizar } f_{\rho^k}^l(x, r)$$

sujeito a:

$$(x, r) \in \text{int}(S_{\rho^k}^l)$$

onde $\rho^k \in R$ é dado, $\rho^k \geq r^*$ (valor ótimo),

$$f_{\rho^k}^l(x, r) := -q \text{Ln}(\rho^k - r) - \sum_{i=1}^l \text{Ln}(d_i - s_i^T x + r) - f_s$$

q é um inteiro positivo, f_s é a função barreira logarítmica associada a S e

$$S_{\rho^k}^l := S^l \cap \{(x, r) \in R^{n+1} : \underbrace{r \leq \rho^k, \dots, r \leq \rho^k}_{q \text{ vezes}}\}.$$

O conjunto inicial que tomamos no algoritmo é dado por

$$V^0 = S \times R \cap \{(x, r) \in R^{n+1} : \underbrace{\rho^0 \geq r, \dots, \rho^0 \geq r}_{q \text{ vezes}}, r \geq s_i^T (x - x^0) + f(x^0)\}$$

onde x^0 é o ponto de partida. Observe que se $\|x\| \leq R$, para $x \in S$ e L é a constante de Lipschitz de f em S , temos:

$$r \geq -2RL + f(x^0).$$

Consequentemente V^0 é um compacto.

O método do centro analítico (ver o capítulo anterior), a partir de um dado limite superior ρ^k para a solução ótima, busca a vizinhança de

$$(x(\rho^k), r(\rho^k)) = \arg \min \{f_{\rho^k}^l(x, r) : (x, r) \in \text{int}(S_{\rho^k}^l)\};$$

o centro analítico de $S_{\rho^k}^l$. A seguir, este limite é diminuído, o que significa que q cópias da restrição associada à função objetivo são deslocadas, e, novamente, é procurado um ponto próximo do novo centro.

Seja a região S em R^n dada por

$$S := \{x \in R^n : a_i^T x \leq b_i, i = 1, 2, \dots, m\},$$

onde $a_i \in R^n$, $b_i \in R$ e é assumido que o interior de S , indicado por $\text{int}(S)$, é não

vazio e limitado. Então $f_S := \sum_{j=1}^m \text{Ln}(b_j - a_j^T x)$,

$$f_{\rho^k}^l(x, r) := -q \text{Ln}(\rho^k - r) - \sum_{i=1}^l \text{Ln}(d_i - s_i^T x + r) - \sum_{j=1}^m \text{Ln}(b_j - a_j^T x)$$

e $q \geq m + l + 1$.

Necessitaremos das derivadas de primeira e segunda ordens de $f_{\rho^k}^l$, que são dadas por

$$g_{\rho^k}^l(x, r) = \nabla f_{\rho^k}^l(x, r) = \begin{pmatrix} \sum_{j=1}^m \frac{1}{b_j - a_j^T x} a_j + \sum_{j=1}^l \frac{1}{d_j - s_j^T x + r} s_j \\ \frac{q}{\rho^k - r} - \sum_{j=1}^l \frac{1}{d_j - s_j^T x + r} \end{pmatrix}$$

e

$$H_{\rho^k}^l(x, r) = \begin{pmatrix} \sum_{j=1}^m \frac{a_j a_j^T}{(b_j - a_j^T x)^2} + \sum_{j=1}^l \frac{s_j s_j^T}{(d_j - s_j^T x + r)^2} & - \sum_{j=1}^l \frac{s_j}{(d_j - s_j^T x + r)^2} \\ - \sum_{j=1}^l \frac{s_j^T}{(d_j - s_j^T x + r)^2} & \mu_{\rho^k}^l(x, r) \end{pmatrix},$$

onde $\mu_{\rho^k}^l(x, r) = \frac{q}{(\rho^k - r)^2} + \sum_{j=1}^l \frac{1}{(d_j - s_j^T x + r)^2}$.

É fato conhecido nas técnicas de pontos interiores que um dos fatores que garantem a polinomialidade em programação linear e quadrática (e convexa, no caso autoconcordante), veja por exemplo (Den Hertog, 1992), é o uso de pontos próximos à trajetória central. Para a efetivação deste conceito, seja a matriz $H = H_{\rho^k}^l(x, r)$ definida positiva para $(x, r) \in \text{int}(S_{\rho^k}^l)$; ela induz a norma

$$\|(x, r)\|_H = \sqrt{(x, r)^T H(x, r)}$$

para $(x, r) \in R^{n+1}$.

A proximidade de $(x, r) \in \text{int}(S_{\rho^k}^l)$ a $(x(\rho^k), r(\rho^k))$ é definida pela norma (induzida por H) da direção de Newton para $f_{\rho^k}^l(\cdot, \cdot)$ em (x, r) dada por:

$$\delta_{\rho^k}^l(x, r) = \|H^{-1} g\|_H,$$

onde $g = g_{\rho^k}^l(x, r)$; o ponto (x, r) é considerado próximo de $(x(\rho^k), r(\rho^k))$ se

$$\delta = \delta_{\rho^k}^l(x, r) < \varepsilon,$$

para algum $\varepsilon \in (0,1)$. É claro, $\delta_{\rho^k}^l(x, r) = 0$, se e somente se, $(x, r) = (x(\rho^k), r(\rho^k))$.

Se a região S for uma esfera em R^n ;

$$S := \{x \in R^n : \frac{1}{2} \|x\|^2 < \frac{\tau^2}{2}\};$$

o mesmo procedimento pode ser adotado, através das derivadas primeira e segunda, que levarão à definição de proximidade.

3.3 O problema da reinicialização

Uma das dificuldades para aplicar o método de centro analítico em um ambiente de planos secantes representado por (3), está na recuperação de um ponto interior, próximo à trajetória central, quando uma nova desigualdade é incorporada ao feixe.

Este problema tem sido atacado por diversos pesquisadores. Através do dual, técnicas de geração de colunas com métodos de pontos interiores foram propostas por (Mitchell, Todd, 1992) e (Goffin, Vial, 1990), para algoritmos projetivos. (Ye, 1992) propôs um método de redução potencial (não-interior) para o problema da viabilidade linear que permite geração de coluna. Em cada iteração uma desigualdade violada no centro corrente é adicionada ao sistema (em uma posição deslocada), até que um ponto viável seja encontrado. (Dantzig, Ye, 1990) propuseram um esquema de “build-up” para o algoritmo afim-escala dual. Este método difere do afim-escala “padrão” na escolha do elipsóide que gera a direção de descida, que é construído a partir de um conjunto de restrições duais “promissoras”. (Tone, 1991) propôs uma estratégia de conjuntos ativos para o método de redução potencial de (Ye, 1989) e (Ye, 1991). Nela a direção de busca é também construída de um subconjunto de restrições que têm folgas duais pequenas na iteração corrente. (Den Hertog, Ross, Terklay, 1992) propuseram uma estratégia de

“build-up and down” para o método da barreira logarítmica em programação linear. Ela parte de um pequeno subconjunto das restrições e segue a correspondente trajetória central até que a iteração esteja próxima (ou viole) uma das demais restrições; neste ponto a restrição é adicionada ao sistema corrente. Por outro lado, uma restrição é eliminada se a correspondente folga do correspondente centro é “grande”. (Den Hertog, 1992) analisa o efeito de deslocamento, adição e eliminação de uma restrição na posição do centro, a distância à trajetória e na variação na função barreira.

Nós optamos por utilizar o trabalho desenvolvido por (Feijoo, Sanchez, Gonzaga, 1997). Apresentamos a seguir uma versão simplificada (e mais flexível) dos resultados do artigo citado, seguido de sua extensão ao caso quadrático. No caso linear os resultados são exatos.

Sejam $S_{\rho^k}^l$, $H_{\rho^k}^l$ e $g_{\rho^k}^l$ como anteriormente. Assume-se que (\bar{x}, \bar{r}) é conhecido e está próximo do centro analítico de $S_{\rho^k}^l$. Uma nova desigualdade linear, $s^T x - r \leq d$, é acrescentada ao conjunto daquelas que definem $S_{\rho^k}^l$; isto é,

$$S_{\rho^k}^{l+1} = S_{\rho^k}^l \cap \{(x, r) \in R^{n+1} : s^T x - r \leq d\},$$

onde se supõe que $\text{int}(S_{\rho^k}^{l+1}) \neq \emptyset$. O problema é encontrar um ponto no interior de $S_{\rho^k}^{l+1}$.

Será mostrado que existe um parâmetro de relaxação γ tal que (\bar{x}, \bar{r}) continua próximo do centro analítico da região

$$S_{\gamma} := S_{\rho^k}^l \cap \{(x, r) \in R^{n+1} : \underbrace{s^T x - r \leq \gamma, \dots, s^T x - r \leq \gamma}_{p \text{ vezes}}\},$$

onde p é um inteiro positivo e $p \geq m + l$. Conhecido o valor de γ , pode-se associar a seguinte função barreira a S_{γ} :

$$P_\gamma(x, r) := -p \text{Ln}(\gamma - s^T x + r) + P(x, r),$$

onde $P(x, r)$ é a função barreira logarítmica relativa a $S_{\rho^k}^l$. Como $\text{int}(S_{\rho^k}^{l+1}) \neq \emptyset$, a aplicação parcial do método de centros analíticos para minimizar $P_\gamma(x, r)$ fornece o ponto desejado. Em outras palavras, seja o seguinte problema

$$\text{Minimizar } z = s^T x - r \quad (\text{P})$$

sujeito a

$$(x, r) \in S_{\rho^k}^l$$

e seja a seguinte família de problemas não lineares, utilizada pelo método de centros analíticos aplicado ao problema (P) :

$$\text{Minimizar } P_{\gamma^i}(x, r) \quad (P_{\gamma^i})$$

sujeito a

$$(x, r) \in S_{\rho^k}^l.$$

Para $\gamma^0 = \gamma$, (\bar{x}, \bar{r}) é um ponto próximo à trajetória central associada ao problema (P). Seja $(x^*, r^*) = \arg \min \{s^T x - r : (x, r) \in S_{\rho^k}^l\}$; por hipótese,

$$S_{\rho^k}^{l+1} \neq \emptyset \Rightarrow s^T x^* - r^* \leq d$$

e o método de centros analíticos permite, ou concluir que $z^* = d$, ou concluir que $z^* < d$. Se o último caso ocorre, o método é parcialmente aplicado e obtém-se um ponto pertencente ao $\text{int}(S_{\rho^k}^{l+1})$; caso contrário, o ponto resultante está na fronteira de $S_{\rho^k}^{l+1}$ e $\text{int}(S_{\rho^k}^{l+1}) = \emptyset$.

Teorema 1 (veja (Feijoo, Sanchez, Gonzaga, 1997))

Sejam

$$\theta = (s, -1)^T H^l(s, -1),$$

$$\alpha = (s, -1)^T H^l g,$$

$$\varepsilon^2 < 1 - \delta^2,$$

e

$$\beta = \sqrt{p^2 \alpha^2 - \varepsilon^2 p[\alpha^2 - \theta(p - \varepsilon^2)]}.$$

onde $\delta < 1$ é a proximidade de (x, r) ao centro analítico do problema perturbado. Então para

$$\gamma \geq (s^T, -1)(x^T, r)^T + \frac{1}{\varepsilon^2}(p\alpha + \beta),$$

$$\text{se } \alpha^2 - \theta(p - \varepsilon^2) < 0;$$

$$\gamma \geq (s^T, -1)(x^T, r)^T + \frac{1}{\varepsilon^2}(p\alpha + \beta) \text{ ou}$$

$$0 \leq \gamma \leq (s, -1)^T(x, r) + \frac{1}{\varepsilon^2}(p\alpha - \beta),$$

$$\text{se } \alpha^2 - \theta(p - \varepsilon^2) > 0 \text{ e } \alpha > 0;$$

$$\gamma \geq (s, -1)^T(x, r),$$

$$\text{se } \alpha^2 - \theta(p - \varepsilon^2) \geq 0 \text{ e } \alpha < 0$$

e

$$\gamma \geq (s, -1)^T(x, r) + \frac{2p\alpha}{\varepsilon^2}$$

$$\text{se } \alpha^2 - \theta(p - \varepsilon^2) = 0 \text{ e } \alpha > 0,$$

a proximidade de (x, r) ao centro da região perturbada é ε .

Demonstração

Seja $P_\gamma(x, r) = -p \ln(\gamma - s^T x + r) + P(x, r)$ e defina

$$\omega = (s, -1)^T,$$

$$\begin{aligned}\eta &= \gamma - s^T x + r \quad \text{e} \\ \mu &= \frac{p^{\frac{1}{2}}}{\eta} \omega.\end{aligned}\tag{6}$$

Da definição de $P\gamma$, temos

$$\begin{aligned}\nabla P\gamma &= g + p^{\frac{1}{2}} \mu \\ \nabla^2 P\gamma &= H + \mu\mu^T\end{aligned}$$

onde

$$g = \nabla P \quad \text{e} \quad H = \nabla^2 P.$$

Estamos interessados no cálculo de

$$\left[(g + p^{\frac{1}{2}}\mu)^T (H + \mu\mu^T)^{-1} (g + p^{\frac{1}{2}}\mu) - (g^T H^{-1} g) \right]^{\frac{1}{2}}.$$

Utilizando a identidade de Sherman-Morrisson, temos

$$\begin{aligned}(g + p^{\frac{1}{2}}\mu)^T \left(H^{-1} - \frac{H^{-1} \mu\mu^T H^{-1}}{1 + \mu^T H^{-1} \mu} \right) (g + p^{\frac{1}{2}}\mu) &= \\ = g^T H^{-1} g + 2p^{\frac{1}{2}} g^T H^{-1} \mu + p\mu^T H^{-1} \mu - \\ \frac{1}{1 + \mu^T H^{-1} \mu} \left(g^T H^{-1} \mu\mu^T H^{-1} g + 2p^{\frac{1}{2}} g^T H^{-1} \mu\mu^T H^{-1} \mu + p\mu^T H^{-1} \mu\mu^T H^{-1} \mu \right).\end{aligned}$$

Substituindo μ por (6), $\theta = \omega^T H^{-1} \omega$, $\alpha = \omega^T H^{-1} g$, temos, desconsiderando a primeira parcela,

$$\Delta = 2\frac{p}{\eta} \alpha + \frac{p^2}{\eta^2} \theta - \frac{1}{1 + \frac{p}{\eta^2} \theta} \left(\frac{p}{\eta^2} \alpha^2 + 2\frac{p^2}{\eta^3} \alpha\theta + \frac{p^3}{\eta^4} \theta^2 \right).\tag{7}$$

Supondo que $g^T H^{-1} g = \delta^2$, $\delta \in (0,1)$, podemos escolher $\varepsilon \in (0,1)$ tal que

$$\delta^2 + \varepsilon^2 < 1, \quad \text{e} \quad \Delta \leq \varepsilon^2.$$

Deste modo, se terá o novo sistema com proximidade $\sqrt{\delta^2 + \varepsilon^2}$. De (7) temos:

$$\varepsilon^2 \eta^2 - 2p\alpha\eta + p[\alpha^2 - \theta(p - \varepsilon^2)] \geq 0. \quad (8)$$

Como $4p^2\alpha^2 - 4\varepsilon^2 p[\alpha^2 - \theta(p - \varepsilon^2)] > 0$, designemos

$$\beta = \sqrt{p^2\alpha^2 - \varepsilon^2 p[\alpha^2 - \theta(p - \varepsilon^2)]},$$

para obter as soluções positivas, que verificam (8), dadas por

$$\eta \geq \frac{1}{\varepsilon^2}(p\alpha + \beta), \text{ se } \alpha^2 - \theta(p - \varepsilon^2) < 0;$$

$$0 \leq \eta \leq \frac{1}{\varepsilon^2}(p\alpha - \beta), \text{ ou } \eta \geq \frac{1}{\varepsilon^2}(p\alpha + \beta),$$

$$\text{se } \alpha^2 - \theta(p - \varepsilon^2) > 0 \text{ e } \alpha > 0;$$

$$\eta \geq 0, \text{ se } \alpha^2 - \theta(p - \varepsilon^2) \geq 0 \text{ e } \alpha < 0;$$

$$\eta \geq \frac{2p\alpha}{\varepsilon^2}, \text{ se } \alpha^2 - \theta(p - \varepsilon^2) = 0 \text{ e } \alpha > 0.$$

Chegamos enfim ao resultado do teorema, pela substituição da expressão de η .

■

Observação:

Embora teoricamente a nossa demonstração seja equivalente à do artigo citado de Feijoo et alii, o resultado acima possui a vantagem computacional de deixar livre a escolha da proximidade final.

3.3.1 O caso quadrático

O teorema 2 a seguir, obtém resultados para perturbação em regiões quadráticas. Eles não são utilizados pelo algoritmo apresentado neste capítulo. Métodos de centros

analíticos que imitem feixes (Sanchez, Oliveira, Santos, 1996), por exemplo, podem ser beneficiados com este resultado.

Suponha S uma região quadrática limitada em R^{n+1} de interior não vazio e seja, como antes, conhecido $v \in \text{int}(S)$, v próximo de v_c , centro analítico de S e que se tenha um problema de reinicialização associado a

$$q(x) = \frac{1}{2} x^T Q x + a^T x - b \leq 0,$$

onde $Q \in R^{(n+1) \times (n+1)}$ é uma matriz simétrica semi-definida positiva, $a \in R^{n+1}$ e $b \in R$. O problema então é determinar um parâmetro de relaxação γ tal que a inclusão da restrição

$$q(x) = \frac{1}{2} x^T Q x + a^T x - b \leq \gamma$$

não altera significativamente a proximidade de v ao centro analítico v_c .

Seja $P_\gamma(x)$ a nova função barreira dada por

$$P_\gamma(x) = -P(x) - l \ln(\gamma - q(x)).$$

O gradiente e o hessiano de $P_\gamma(\cdot)$ em v são:

$$g_\gamma(v) = \nabla P_\gamma(v) = \frac{l}{(\gamma - q(v))} (a + Qv) + g(v) \text{ e}$$

$$H_\gamma(v) = \nabla^2 P_\gamma(v) = \frac{l}{(\gamma - q(v))} Q + \frac{l}{(\gamma - q(v))^2} (a + Qv)(a + Qv)^T + H(v).$$

A matriz H_γ é simétrica e definida positiva. A proximidade de v ao novo centro $x_{c,\gamma}$ é dada por

$$\delta_\gamma = \delta_\gamma(v) = \|H_\gamma^{-1}(v)g_\gamma(v)\|_{H_\gamma(v)} = \sqrt{g_\gamma(v)H_\gamma^{-1}(v)g_\gamma(v)}.$$

Seja $\beta = \beta(\gamma) = \frac{1}{(\gamma - q(v))} > 0$ e $d = Qv + a$. Então

$$g_\gamma = l\beta d + g \quad \text{e}$$

$$H_\gamma = l\beta Q + l\beta^2 dd^T + H.$$

Note que para cada γ , H_γ é definida positiva. Além disto,

$$H + l\beta^2 d d^T \prec H_\gamma,$$

então

$$H_\gamma^{-1} \prec (H + l\beta^2 d d^T)^{-1} \quad (\text{corolário 7.7.4 de (Horn, 1985)})$$

e

$$\delta_\gamma^2 = g_\gamma^T H_\gamma^{-1} g_\gamma \leq g_\gamma^T (H + \beta^2 d d^T)^{-1} g_\gamma$$

O resultado a seguir é obtido de forma similar ao teorema 1.

Teorema 2

Sejam

$$\theta = d^T H^{-1} d,$$

$$\alpha = d^T H^{-1} g,$$

$$\varepsilon^2 < 1 - \delta^2,$$

$$\zeta = l(1 - \varepsilon^2)(\alpha^2 + \varepsilon^2\theta)$$

e

$$\eta = l(\theta\varepsilon^2 + \alpha^2 - l\theta),$$

onde $\delta < 1$ é a proximidade de ν ao centro analítico do problema não perturbado.

Então, para

$$\gamma \geq q(\nu) + \frac{\eta}{l\alpha - \sqrt{\zeta}}, \quad \text{se } \eta < 0;$$

$$\gamma \geq q(\nu) + \frac{\eta}{l\alpha - \sqrt{\zeta}} \quad \text{ou} \quad \gamma \leq q(\nu) + \frac{\eta}{l\alpha + \sqrt{\zeta}},$$

$$\text{se } \eta > 0 \text{ e } \alpha > 0;$$

$$\gamma > q(\nu), \quad \text{se } \eta > 0 \text{ e } \alpha < 0;$$

$$\gamma \geq q(\nu) + \frac{\varepsilon^2}{2l\alpha}, \quad \text{se } \eta = 0,$$

a proximidade de ν ao centro da região perturbada é ε .

3.4 Algoritmo

No algoritmo conceitual do quadro I (seção 2.1), a região S^k é refinada por um único corte; aquele associado ao plano secante H^{k+1} . Em um contexto no qual uma nova restrição, além daquela associada a H^{k+1} , é adicionada ao problema em cada iteração, é necessário recuperar o centro antes de dar continuidade ao algoritmo. O método parcial de centros analíticos, descrito a seguir, antes de proceder à perturbação, utiliza o procedimento de reinicialização (seção anterior) para recuperar a proximidade ao centro analítico.

3.4.1 Obtenção de um ponto interior no sistema perturbado

Com (x^k, r^k) , próximo do centro analítico de $S_{\rho^k}^l$ (figura 1), o algoritmo do quadro V computa um subgradiente arbitrário s_{l+1} em x^k , $f(x^k)$ e $d_{k+1} = s_{k+1}^T x^k - f(x^k)$, elementos para gerar o corte $s_{k+1}^T x - r \leq d_{k+1}$ (figura 2). Se o interior da região

$$S_{\rho^k} \cap \{(x, r) \in R^{n+1} : s_{k+1}^T x - r \leq d_{k+1}\}$$

é não vazio, o procedimento de reinicialização (figuras 3.a e 3.b) aplica parcialmente o método de centro analítico ao problema (P), encontrando (x_c^k, r_c^k) próximo à trajetória central, e incorpora o corte ao feixe. Caso contrário (interior da região perturbada vazio), o algoritmo ignora a restrição gerada pelo subgradiente¹.

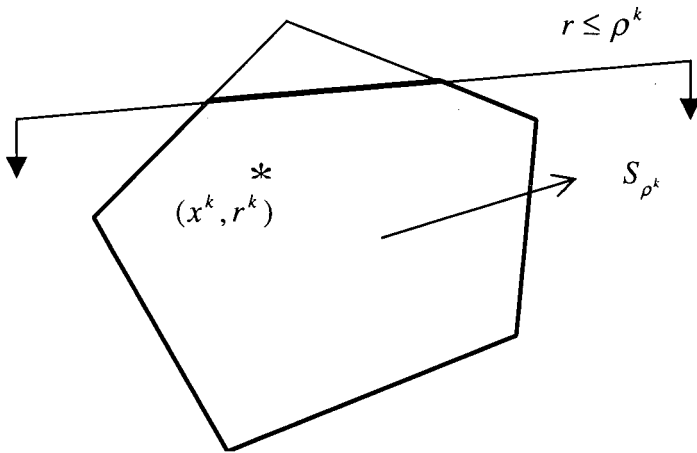


Figura 1

Ponto próximo ao centro analítico de S_{ρ^k}

¹ Este fato, para manter a legibilidade, não é mostrado no algoritmo do quadro V.

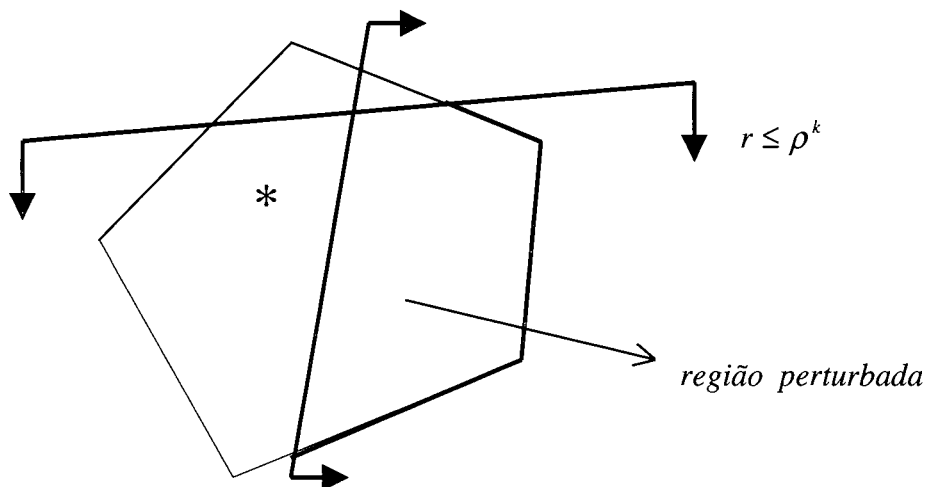


Figura 2

Introdução do corte $s_{k+1}^T x - r \leq d_{k+1}$, gerado por um subgradiente em (x^k, r^k)

3.4.2 Atualização do parâmetro ρ

A seguir, é atualizado ρ^k . Seja $f_m^{k+1} := \min\{f(x^k), f_m^k\}$. Utilizamos como novo valor do limite superior ρ^{k+1} o seguinte:

$$\rho^{k+1} := \rho^k - \theta(\rho^k - f_m^k),$$

onde $\theta \in (0,1)$ - ver figura 4. De posse deste limite, o algoritmo aplica um número finito τ ($\tau \geq 1$) de iterações do método de centros analíticos ao problema (3) seção 3.1, começando com $\rho_p^0 = \rho^k$, $(x_p^0, r_p^0) = (x_c^k, r_c^k)$ e termina quando $\rho_p^\tau = \rho^{k+1}$, obtendo (x^{k+1}, r^{k+1}) - ver figuras 5.a e 5.b. Se $\rho^{k+1} > r^k$, então $\tau = 1$; isto é $\rho_p^1 = \rho^{k+1}$; caso $r^k > \rho^{k+1}$, um número $\tau > 1$ de iterações é aplicado até que $\rho_p^\tau = \rho^{k+1}$ (ver quadro V.b).

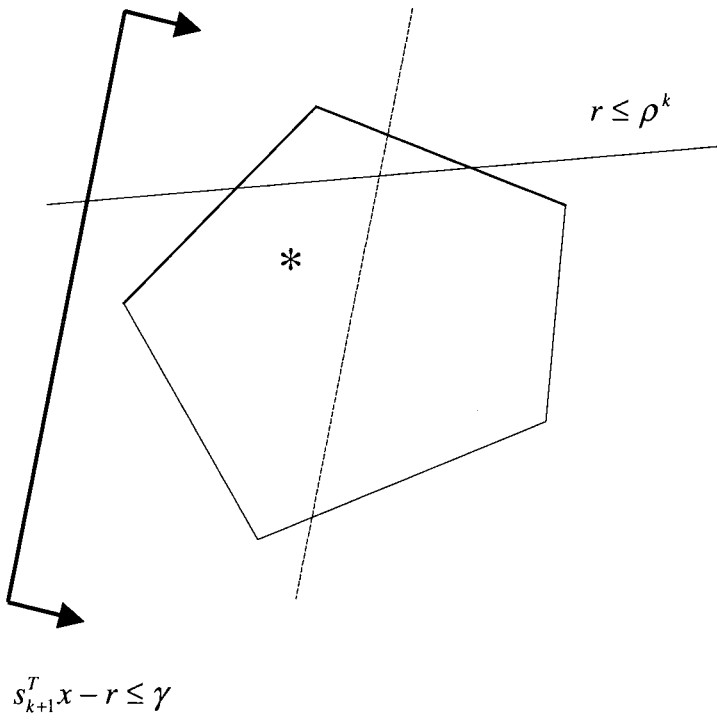


Figura 3a

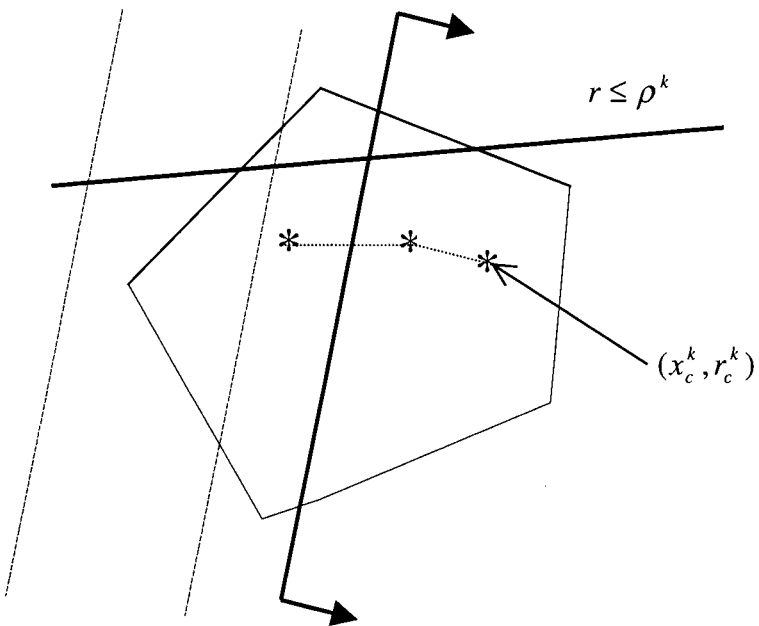


Figura 3b

Figura 3 - Reinicialização

a- relaxação do corte; **b-** aplicação parcial do método de centros analíticos para obter um ponto interior à nova região

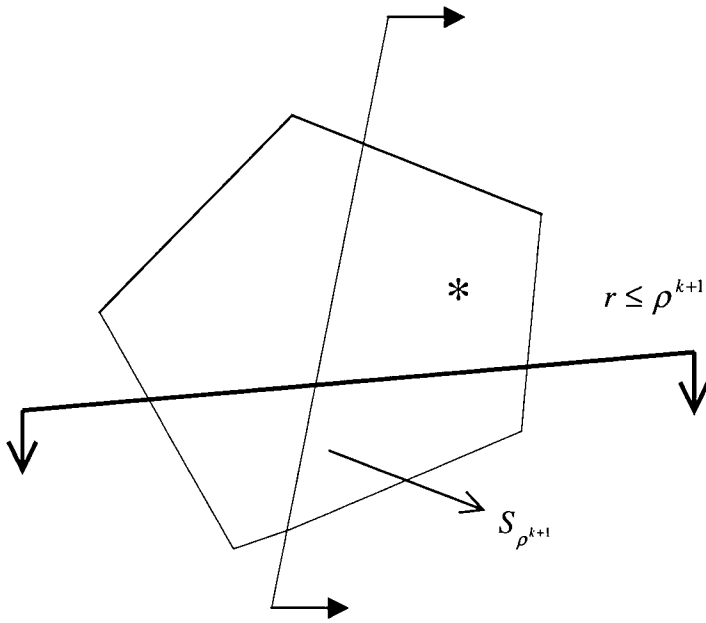


Figura 4

Atualização do limite superior da solução ótima

O algoritmo do quadro V parte de uma região fixa S , seja ela um poliedro limitado ou uma esfera (a representação de S incide somente nas rotinas de nível mais baixo).

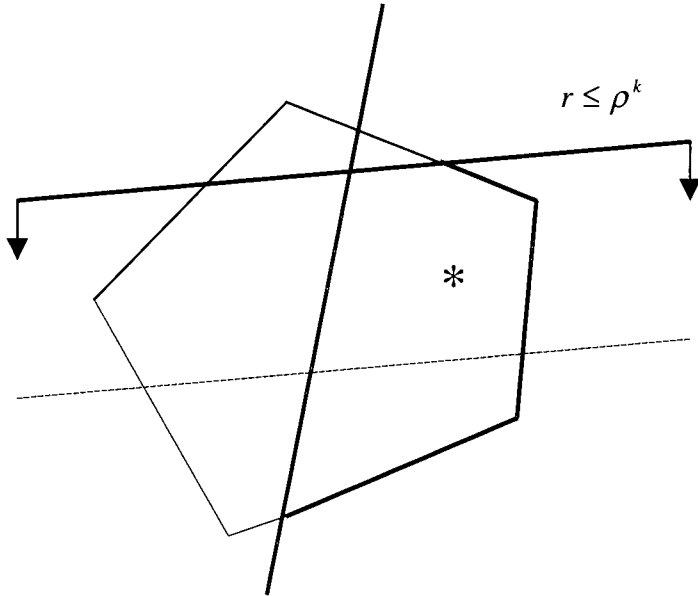


Figura 5a

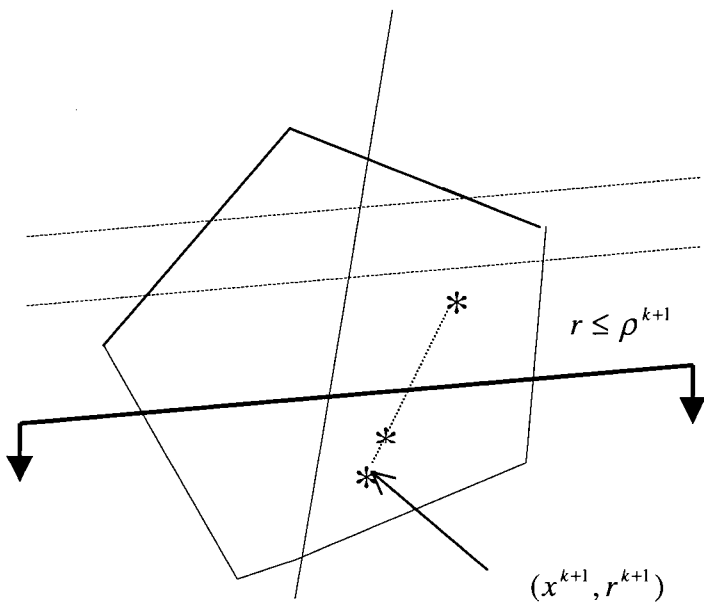


Figura 5b

Figura 5

- a- Aplicação de τ iterações do método de centros analíticos para obter (x^{k+1}, r^{k+1}) ;
 b- Nova região que contém a solução ótima

algoritmo para Programação Convexa Não Suave

1. Obtenha $(x^0, r^0) \in \text{int}(S_{\rho^0}) : \delta_{\rho^0} \leq 0,1$
2. Escolha $f_m^0 = f(x^0)$, $\rho^0 > f_m^0$, $\theta \in (0,1)$ e $\varepsilon > 0$, $\eta > 0$

- $k := 0$
- repita
3. se $(\rho^k - r^k) < \frac{72}{91}\varepsilon$
então interrompa
fim se
4. Compute $f(x^k)$, $s_{k+1} \in \partial f(x^k)$ e $d_{k+1} = s_{k+1}^T x^k - f(x^k)$
5. Compute γ { ver Teorema 1 }
6. Reduza γ até d_{k+1} { Aplicação do método de C.A.;
ver Quadro V.a }
{ Incorpora o corte ao feixe }
7. $S_{\rho^k} := S_{\rho^k} \cap \{(x, r) \in R^{n+1} : s_{k+1}^T x - r \leq d_{k+1}\}$
8. $f_m^k := \min\{f(x^k), f_m^k\}$
9. $\rho^{k+1} := \rho^k - \theta(\rho^k - f_m^k)$
10. Obtenha (x^{k+1}, r^{k+1}) , próximo ao Centro Analítico de $S_{\rho^{k+1}}$
{ ver Quadro V.b }

11. $k := k + 1$
- fim repita
- fim algoritmo

Quadro V

Algoritmo para programação convexa não suave que utiliza um método parcial de centros analíticos

ref. “obtenha (x^{k+1}, r^{k+1}) , próximo ao centro analítico de $S_{\rho^{k+1}}$ ”

$i := 0$

$\rho_c^i := \gamma$

$(x_c^i, r_c^i) := (x^k, r^k)$

repita

se $(\rho_c^i = d_{k+1})$

então interrompa

fim se

se $\rho_c^i < \rho^{k+1}$

então $\rho_p^{i+1} = \rho^{k+1}$

$p := 1$

senão $\rho_c^{i+1} := \rho_c^i + \theta(\rho_c^i - r_c^i)$

fim se

$(x_c^{i+1}, r_c^{i+1}) :=$ ponto próximo ao *Centro Analítico de*

$$S_{\rho_c^{i+1}} := S_{\rho^i} \cap \{(x, r) \in R^{n+1} : \underbrace{s_{k+1}^T x - r \leq \rho_c^{i+1}, \dots, s_{k+1}^T x - r \leq \rho_c^{i+1}}_{p \text{ vezes}}\}$$

$i := i + 1$

fim repita

$(x_c^k, r_c^k) := (x_c^i, r_c^i)$

fim ref.

Quadro V.a- Reinicialização

Aplicação parcial do método de centros analíticos para recuperar a proximidade ao centro analítico da região perturbada.

ref. “obtenha (x^{k+1}, r^{k+1}) , próximo ao centro analítico de $S_{\rho^{k+1}}$ ”

$\tau := 0$

$\rho_p^\tau := \rho^k$

$(x_p^\tau, r_p^\tau) := (x_c^k, r_c^k)$, ponto próximo à trajetória central, obtido pela redução de

γ até d_{k+1}

repita

se $(\rho_p^\tau = \rho^{k+1})$

então interrompa

fim se

se $(r^\tau < \rho^{k+1})$

então $\rho_p^{\tau+1} = \rho^{k+1}$

$q := 1$

senão $\rho_p^{\tau+1} := \rho_p^\tau + \theta(\rho_p^\tau - r_p^\tau)$

fim se

$(x^{k+1}, r^{k+1}) :=$ ponto próximo ao *Centro Analítico de*

$$S_{\rho_p^{\tau+1}} := S_{\rho^k} \cap \{(x, r) \in R^{n+1} : \underbrace{r \leq \rho_p^{\tau+1}, \dots, \rho_p^{\tau+1}}_{q \text{ vezes}}\}$$

$\tau := \tau + 1$

fim repita

$(x^{k+1}, r^{k+1}) := (x_p^\tau, r_p^\tau)$

fim ref.

Quadro V.b

Obtenção do ponto próximo ao centro analítico de $S_{\rho^{k+1}}$

3.5 Gerência do feixe de informações

Implementações do método de planos secantes têm de lidar com o crescimento do feixe de informações proveniente das restrições adicionadas em cada iteração. O aumento de tamanho dos subproblemas tornam rapidamente a sua solução difícil, senão impossível, de ser obtida (Eaves, Zangwill, 1971). O método proposto na seção 3.4 admite uma forma natural de gerenciar o tamanho do feixe. Já que se sabe como relaxar uma restrição, tal que a sua inclusão no feixe não perturbe significativamente a proximidade à trajetória central (seção 3.3), isto pode ser aplicado na análise do efeito do descarte de alguma restrição introduzida anteriormente.

Não mostramos em detalhe, porque não foi utilizado nas implementações, porém se trata apenas de utilizar a identidade de Sherman-Morrisson em

$$(g_k - u_i)^T (H_{\rho^k} + u_i u_i^T)^{-1} (g_k - u_i),$$

$$u_i = (d_i - s_i^T x + r)^{-1} \begin{pmatrix} -s_i \\ 1 \end{pmatrix},$$

onde u_i é alguma restrição que se deseja avaliar.

3.6 Convergência

Nas proposições que se seguem, k é a iteração e s_{k+1} é um subgradiente arbitrário de f em x^k . Supõe-se conhecida uma região compacta S convexa que contém a solução ótima do problema e que $f(x) \leq K \neq \infty, \forall x \in S$.

Nesta seção obtemos a convergência do algoritmo, mais precisamente, mostraremos que $\rho^k \rightarrow f(x^*)$ e $f_m^k \rightarrow f(x^*)$. Será mostrado que o teste de parada produz uma ε -solução no sentido em que $f_m^k = \min f(x^i) \leq f(x^*) + \varepsilon$. Calcularemos também a complexidade (polinomial) por iteração do algoritmo. Trata-se de estimar o número

máximo de iterações de centro analítico, aplicado em problemas de programação linear. A análise se baseia sobretudo em resultados teóricos conhecidos do método de centros analíticos (C. A.). Em particular, usaremos o excelente texto de (Den Hertog, 1992) como referência para o desenvolvimento. Em acréscimo, apresentaremos uma demonstração da convergência que prescinde da análise de complexidade.

3.6.1 Convergência das seqüências $\{\rho^k\}$ e $\{f_m^k\}$

Nos próximos lemas mostramos que $\{\rho^k\}$ e $\{f_m^k\}$ possuem um limite comum.

Lema 1

Tem-se, para todo $k \geq 0$, $\rho^k > f_m^k$ e $\{\rho^k\}$ é uma seqüência monótona estritamente decrescente. Mais precisamente,

$$\begin{aligned}\rho^k - f_m^k &\geq (1 - \theta)^k (\rho^0 - f_m^0) \\ \rho^k - \rho^{k+1} &\geq \theta(1 - \theta)^k (\rho^0 - f_m^0).\end{aligned}$$

Demonstração: da fórmula de atualização do ρ (passo 9) tem-se

$$\rho^{k+1} - f_m^{k+1} = \rho^k - f_m^{k+1} - \theta(\rho^k - f_m^k) \geq \rho^k - f_m^k - \theta(\rho^k - f_m^k), \quad (9)$$

onde a desigualdade decorre da definição de $f_m^{k+1} = \min_{i \leq k+1} f(x^i)$, que faz de $\{f_m^k\}$ uma seqüência não-crescente. De (9) se deduz que

$$\begin{aligned}\rho^{k+1} - f_m^{k+1} &\geq (1 - \theta)(\rho^k - f_m^k) \\ &\geq (1 - \theta)^{k+1} (\rho^0 - f_m^0),\end{aligned}$$

que é positivo devido à condição de inicialização no passo 2 do algoritmo. Temos assim o primeiro resultado.

Retornando à fórmula de ρ^{k+1} (passo 9), reescrevemos

$$\rho^k - \rho^{k+1} = \theta(\rho^k - f_m^k) \geq \theta(1 - \theta)^k(\rho^0 - f_m^0),$$

a desigualdade sendo devida à primeira parte.

Lema 2

As seqüências $\{\rho^k\}$ e $\{f_m^k\}$ convergem para o mesmo limite

$$\rho^* \geq f(x^*) = \min_{x \in S} f(x).$$

Demonstração: supomos as seqüências infinitas. Devido ao lema anterior, e da definição de f_m^k , temos, para todo $k \geq 0$,

$$\rho^k \geq f_m^k \geq f(x^*).$$

Esta limitação, junto à monotonia das seqüências $\{\rho^k\}$ e $\{f_m^k\}$ garantem, respectivamente, limites ρ^* e f_m^* , tais que

$$\rho^* \geq f_m^* \geq f(x^*).$$

Por outro lado, levando ao limite ambos os lados de

$$\rho^{k+1} = \rho^k - \theta(\rho^k - f_m^k),$$

obtemos

$$\rho^* = \rho^* - \theta(\rho^* - f_m^*),$$

de que se conclui, por ser $\theta \neq 1$, que

$$\rho^* = f_m^*.$$

3.6.2 Complexidade para a redução de ρ^k para ρ^{k+1}

Ao fim da iteração k do algoritmo, temos o ponto (x_k, r^k) próximo do C.A. do politopo S_{ρ^k} . A incorporação de uma restrição correspondente ao subgradiente $s_{k+1} \in \bar{\partial}f(x_k)$ é realizada através de um passo iterativo de C.A., que leva ao ponto (x_c^k, r_c^k) , próximo ao C.A. do politopo

$$V := \{(x, r) \in S, s_i^T x - r \leq d_i, i = 1, \dots, k+1, \underbrace{r \leq \rho^k, \dots, r \leq \rho^k}_{q \text{ vezes}}\},$$

para q convenientemente escolhido. O problema linear em pauta é, portanto,

$$\text{Min } t \tag{11}$$

$$t \geq s_i^T y - d_i, \quad i = 1, \dots, k+1$$

$$y \in S, \quad t \in R,$$

onde o ponto inicial é $(y^0, t^0) = (x_c^k, r_c^k)$ e $\rho^k > r_c^k$.

O lema apresentado a seguir, sem demonstração, é adaptação à nossa notação, dos lemas 3.2 e 3.6 do texto citado de Den Hertog.

Lema 3

Sejam (y_{k+1}^*, r_{k+1}^*) solução ótima para (11), e z^j a cota superior na j -ésima iteração de C. A. Seja (y^j, t^j) o iterado ao fim de j iterações. Defina m_{k+1} o número de restrições de (11). Suponha $q \geq 2\sqrt{m_{k+1}}$, e a proximidade de cada iteração igual a $\frac{1}{2}$. Tem-se

$$z^{j-1} - r_{k+1}^* \leq \left(1 + \frac{m_{k+1}}{q}\right) \left(1 + \frac{2\sqrt{m_{k+1}}}{q}\right) (z^{j-1} - t^{j-1}). \quad (12)$$

O resultado que se segue é um resultado intermediário na demonstração do teorema principal de complexidade do método de C. A. (teorema 3.1 de Den Hertog). Definamos a atualização das cotas de C. A. aplicado ao problema (11) por

$$z^j = z^{j-1} - \beta(z^{j-1} - t^{j-1}),$$

para algum $0 < \beta < 1$.

Lema 4

Com as mesmas hipóteses do lema anterior, se tem

$$z^j \leq t^j + \left(1 - \frac{\beta}{2\left(1 + \frac{m_{k+1}}{q}\right)}\right)^j (\rho^k - r_{k+1}^*). \quad (13)$$

Demonstração: temos

$$\begin{aligned} \frac{z^j - r_{k+1}^*}{z^{j-1} - r_{k+1}^*} &= \frac{z^{j-1} - \beta(z^{j-1} - t^{j-1}) - r_{k+1}^*}{z^{j-1} - r_{k+1}^*} = \\ &= 1 - \beta \frac{z^{j-1} - t^{j-1}}{z^{j-1} - r_{k+1}^*}. \end{aligned}$$

Usamos a estimativa (12) para obter

$$\frac{z^j - r_{k+1}^*}{z^{j-1} - r_{k+1}^*} \leq 1 - \frac{\beta}{2(1 + \frac{m_{k+1}}{q})} := 1 - \beta'.$$

Assim, após j iterações de C. A. temos

$$\begin{aligned} z^j - t^j &\leq z^j - r_{k+1}^* \\ &\leq (1 - \beta')(z^{j-1} - r_{k+1}^*) \\ &\leq (1 - \beta')^j (z^0 - r_{k+1}^*) = (1 - \beta')^j (\rho^k - r_{k+1}^*), \end{aligned}$$

conforme queríamos demonstrar.

A seguir, apresentamos o teorema 3.1 de Den Hertog.

Lema 5

Após um máximo de

$$\frac{2}{\beta} \left(1 + \frac{m_{k+1}}{q}\right) Ln \frac{4(1 + \frac{m_{k+1}}{q})(\rho^k - r_{k+1}^*)}{\varepsilon} \quad (14)$$

iterações de C. A., o algoritmo de centros termina com uma ε -solução do problema (11).

Podemos agora apresentar o resultado principal desta seção:

Proposição 1

Sob as hipóteses do lema 3, após um máximo de

$$\frac{2}{\beta} \left(1 + \frac{m_{k+1}}{q}\right) \text{Ln} \frac{8 \left(1 + \frac{m_{k+1}}{q}\right) (\rho^k - r_{k+1}^*)}{\rho^k - \theta(\rho^k - f_m^k) - r_{k+1}^*} \quad (15)$$

iterações de C. A., o passo 10 do algoritmo é realizado.

Demonstração: desejamos que j seja tal que

$$z^j \leq \rho^{k+1} = \rho^k - \theta(\rho^k - f_m^k).$$

De acordo com (13) isto certamente ocorre se

$$t^j + (1 - \beta')^j (\rho^k - r_{k+1}^*) \leq \rho^k - \theta(\rho^k - f_m^k). \quad (16)$$

A análise do pior caso implica em que devemos supor

$$\rho^{k+1} \leq r_c^k. \quad (17)$$

Por outro lado, como o lado esquerdo de (16) converge para r_{k+1}^* , precisamos mostrar que

$$\rho^{k+1} > r_{k+1}^*, \quad (18)$$

de modo que seja garantida a solução para (16), com j suficientemente grande. Ora, r_{k+1}^* é solução ótima da aproximação de planos secantes do problema convexo original; por ter menos restrições, é verdade que

$$r_{k+1}^* \leq f(x^*).$$

Além disso, do lema 1 se tem, para todo i ,

$$\rho^i > f_m^i \geq f(x^*),$$

logo

$$\rho^{k+1} > f(x^*) \geq r_{k+1}^*,$$

que é (18).

Agora propomos resolver (16), considerando j tal que verifique simultaneamente

$$t^j - r_{k+1}^* \leq \frac{1}{2}(\rho^{k+1} - r_{k+1}^*) \quad (19)$$

e

$$(1 - \beta^i)^j (\rho^k - r_{k+1}^*) \leq \frac{1}{2}(\rho^{k+1} - r_{k+1}^*). \quad (20)$$

Usamos o lema 5 com $\varepsilon = \frac{(\rho^{k+1} - r_{k+1}^*)}{2}$ para obter (19),

$$j > \frac{2}{\beta} \left(1 + \frac{m_{k+1}}{q}\right) Ln \frac{8(1 + \frac{m_{k+1}}{q})(\rho^k - r_{k+1}^*)}{\rho^{k+1} - r_{k+1}^*}. \quad (21)$$

Por outro lado (20) implica em

$$j > \frac{2}{\beta} \left(1 + \frac{m_{k+1}}{q}\right) Ln \frac{2(\rho^k - r_{k+1}^*)}{\rho^{k+1} - r_{k+1}^*},$$

cota que claramente menor do que a dada em (21), que é o resultado desejado.

3.6.3 Convergência de $\{\rho^k\}$ e $\{f_m^k\}$ para $f(x^*)$

A partir de agora suporemos que

$$q = Cm_{k+1}, \quad (22)$$

para algum $C > 0$. Além disto, chamaremos n_k o número de iterações de C.A., necessárias para passar da cota ρ^k à cota ρ^{k+1} . Sabemos, da proposição 1, que

$$n_k \leq \frac{4}{\beta} Ln \frac{16(\rho^k - r_{k+1}^*)}{\rho^{k+1} - r_{k+1}^*}. \quad (23)$$

Podemos então escrever que

$$\rho^{k+1} \geq z^{n_k}.$$

A fim de simplificar a análise, suporemos que $\rho^{k+1} = z^{n_k}$.

Teorema 3

Escolha, em (22), $C = 12$, e $1 > \beta \geq 19/42$. Então as seqüências $\{\rho^k\}$ e $\{f_m^k\}$ convergem para

$$\rho^* = f(x^*). \quad (24)$$

Além disto, o algoritmo termina com uma solução ε -ótima, no seguinte sentido:

$$f_m^{\bar{k}} \leq f(x^*) + \varepsilon, \quad (25)$$

\bar{k} sendo a iteração final.

Demonstração: apliquemos o lema 3 com $j - 1 = n_k$. Obtemos

$$\rho^{k+1} - r_{k+1}^* \leq \left(1 + \frac{1}{12}\right) \left(1 + \frac{2}{12\sqrt{m_{k+1}}}\right) (\rho^{k+1} - r^{k+1}). \quad (26)$$

Analogamente, o lema 4 fornece

$$\rho^{k+1} - r_{k+1} \leq \left(1 - \frac{\beta}{2\left(1 + \frac{1}{12}\right)}\right)^{n_k} (\rho^k - r_{k+1}^*). \quad (27)$$

De (26) e (27) deduzimos que

$$\rho^{k+1} - r_{k+1}^* \leq \frac{91}{72} \left(1 - \frac{6\beta}{13}\right)^{n_k} (\rho^k - r_{k+1}^*),$$

ou, substituindo ρ^{k+1} por sua expressão,

$$\rho^{k+1} - \theta(\rho^k - f_m^k) - r_{k+1}^* \leq \frac{91}{72} \left(1 - \frac{6\beta}{13}\right)^{n_k} (\rho^k - r_{k+1}^*),$$

isto é:

$$\left(1 - \theta - \frac{91}{72} \left(1 - \frac{6\beta}{13}\right)^{n_k}\right) \rho^k + \theta f_m^k \leq \left(1 - \frac{91}{72} \left(1 - \frac{6\beta}{13}\right)^{n_k}\right) r_{k+1}^*. \quad (28)$$

Destaquemos a seqüência

$$\alpha_k := 1 - \frac{91}{72} \left(1 - \frac{6\beta}{13}\right)^{n_k}.$$

Temos, devido à hipótese sobre β que,

$$\frac{1}{72} \leq \alpha_k < 1, \quad (29)$$

de que se conclui pela existência de uma subsequência convergente, que denominamos de $\{\alpha_{k_j}\}$. Considere agora as seqüências $\{\rho^{k_j}\}$, $\{f_m^{k_j}\}$ e $\{r_{(k+1)_j}^*\}$. Devido ao lema 2,

$$\lim_{j \rightarrow \infty} \rho_{k_j} = \lim_{j \rightarrow \infty} f_m^{k_j} = \rho^*. \quad (30)$$

Por outro lado, sendo $\{r_{k+1}^*\}$ a seqüência das soluções do método de planos secantes aplicado a f (em um compacto), sabemos que

$$\lim_{j \rightarrow \infty} r_{(k+1)_j}^* = \lim_{k \rightarrow \infty} r_{k+1}^* = f(x^*). \quad (31)$$

Passemos agora ao limite em (28), levando em consideração (29), (30) e (31). Seja $\alpha_k^* > 0$ o ponto de acumulação relativo à subsequência $\{\alpha_{k_j}\}$. Temos

$$(\alpha_k^* - \theta)\rho^* + \theta\rho^* \leq \alpha_k^* f(x^*),$$

ou,

$$\rho^* \leq f(x^*).$$

Como, pelo lema 2, $\rho^* \geq f(x^*)$, temos (24).

Se o algoritmo parar em \bar{k} , $\rho^{\bar{k}} - r_{\bar{k}}^* \leq \frac{72\varepsilon}{91}$, logo, segundo (26)

$$\frac{72}{91}(\rho^{\bar{k}} - r_{\bar{k}}^*) \leq \frac{72}{91}\varepsilon,$$

e

$$f_m^{\bar{k}} - r_{\bar{k}}^* < \rho^{\bar{k}} - r_{\bar{k}}^* \leq \varepsilon,$$

$$f_m^k \leq r_k^* + \varepsilon \leq f(x^*) + \varepsilon,$$

conforme desejado.

3.6.4 Complexidade por iteração

Nosso objetivo é estimar o máximo de iterações de C.A., por iteração do algoritmo. Lembramos que na seção 3.6.2 obtivemos a complexidade relativa ao passo 10 do algoritmo, Falta, portanto, apenas, estimar a complexidade do passo 6, que corresponde ao processo de incorporação de restrição. Fica desde logo claro que a estrutura modular do algoritmo permite que se utilize qualquer outro método de incorporação, eventualmente de custo inferior ao que usamos.

Complexidade do algoritmo de incorporação de restrição

A metodologia a ser seguida é similar à de seção 3.6.2. Supomos que (x^k, r^k) é um centro

analítico (aproximado) do conjunto S_{ρ^k} , dado por

$$S_{\rho^k} = \{(x, r) \in R^{n+1} : (x, r) \in S, s_i^T x - r \leq d_i, i = 1, \dots, k, \underbrace{r \leq \rho^k, \dots, r \leq \rho^k}_{q \text{ vezes}}\}$$

Em x^k é determinado um novo corte caracterizado pelo subgradiente $s_{k+1} \in \partial f(x^k)$:

$$s_{k+1}^T x - r \leq d_{k+1} = s_{k+1}^T x^k - f(x^k) \quad (32)$$

Através do teorema 1 calcula-se uma relaxação, cujo valor denotaremos por γ^k , tal que (x^k, r^k) seja um C.A. aproximado de S_{ρ^k} interceptado por

$$\{(x, r) : s_{k+1}^T x - r \leq \gamma^k\} \quad (33)$$

Por simples inspeção dos resultados do Teorema 1, podemos supor que γ^k é escolhido de modo a sempre satisfazer

$$\gamma^k \geq d_{k+1}. \quad (34)$$

Evidentemente que só haverá necessidade de implementar o processo de recuperação do centro analítico se $\gamma^k > d_{k+1}$, que é o que suporemos a partir de agora.

Considere, como na seção 3.4, o seguinte problema auxiliar:

$$\begin{aligned} \text{Min } & s_{k+1}^T y - t \\ & (y, t) \in S_{\rho^k} \end{aligned} \quad (35)$$

É claro que o objetivo é que a cota superior γ^k para a função custo de (35) seja diminuída até o valor d_{k+1} . O método de C. A. é utilizado tendo como ponto de partida (x^k, r^k) . Temos resultados análogos aos lemas 3, 4 e 5. Abaixo, apenas enunciamos os que serão utilizados posteriormente, relativos à estimativa para a cota z^j , e à complexidade.

Lema 6

Sejam z_k^* o custo ótimo de (35) e z^j a cota superior na j -ésima iteração. Suponha $q \geq 2\sqrt{m_k}$, e a proximidade de cada iteração igual a $\frac{1}{2}$. Então

$$z^j \leq s_{k+1}^T y^j - t^j + \left(1 - \frac{\xi}{2(1 + \frac{m_k}{q})}\right)^j (\gamma^k - z_k^*), \quad (36)$$

onde $0 < \xi < 1$ é o parâmetro de atualização das cotas: $z^j = z^{j-1} - \xi(z^{j-1} - t^{j-1})$.

Lema 7

Após um máximo de

$$\frac{2}{\xi} \left(1 + \frac{m_k}{q}\right) \text{Ln} \left(\frac{4 \left(1 + \frac{m_k}{q}\right) (\gamma^k - z_k^*)}{\varepsilon} \right) \quad (37)$$

iterações de C. A., o algoritmo de centros termina com uma ε -solução do problema (35).

Podemos finalmente apresentar a

Proposição 2

Sob as hipóteses do lema 6, após um máximo de

$$\frac{2}{\xi} \left(1 + \frac{m_k}{q}\right) \text{Ln} \left(\frac{8 \left(1 + \frac{m_k}{q}\right) (\gamma^* - z_k^*)}{d_{k+1} - z_k^*} \right) \quad (38)$$

iterações de C. A., o passo 6 é realizado.

Demonstração: desejamos que j seja tal que $z^j \leq d_{k+1} = s_{k+1}^T x^k - f(x^k)$. Face ao resultado do lema 6, isto certamente ocorre se

$$s_{k+1}^T y^j - t^j + \left(1 - \frac{\xi}{2 \left(1 + \frac{m_k}{q}\right)}\right)^j (\gamma^k - z_k^*) \leq s_{k+1}^T x^k - f(x^k). \quad (39)$$

hipótese de que o corte gera um politopo de interior não vazio faz com que a desigualdade acima se resolva para j suficientemente grande. De fato temos, por um lado,

$$d_{k+1} = s_{k+1}^T x^k - f(x^k) < z_k^*,$$

e, por outro, as seqüências $\{s_{k+1}^T x^j - t^j\}_j$ e $\{(1 - \frac{\xi}{2(1 + \frac{m_k}{q})})^j (\gamma^j - z_k^*)\}_j$, ambas não-

crescentes, convergindo (monotonamente) para, respectivamente, z_k^* e 0. Deste modo, podemos resolver (39), obtendo j tal que verifique, simultaneamente,

$$s_{k+1}^T y^j - t^j - z_k^* \leq \frac{1}{2}(d_{k+1} - z_k^*) \quad (40)$$

e

$$(1 - \frac{\xi}{2(1 + \frac{m_k}{q})})^j (\gamma^k - z_k^*) \leq \frac{1}{2}(d_{k+1} - z_k^*). \quad (41)$$

Usamos o lema 7 para resolver (40) com $\varepsilon = \frac{1}{2}(d_{k+1} - z_k^*)$, obtendo

$$j > \frac{2}{\xi} (1 + \frac{m_k}{q}) \text{Ln} \left(\frac{8(1 + \frac{m_k}{q})(\gamma^k - z_k^*)}{d_{k+1} - z_k^*} \right). \quad (42)$$

De (41) temos a cota

$$j > \frac{2}{\xi} (1 + \frac{m_k}{q}) \text{Ln} \left(\frac{2(1 + \frac{m_k}{q})(\gamma^k - z_k^*)}{d_{k+1} - z_k^*} \right),$$

inferior à anterior, resultando assim em (38).

Complexidade por iteração

Nas proposições 1 e 2, escolhemos, conforme o teorema 3, $q = 12m_k$,

$$\text{supomos } 1 > \beta \geq \frac{19}{42},$$

$0 < \xi < 1$, obtendo, por iteração, um máximo de

$$\frac{13}{6\beta} \text{Ln} \frac{\frac{26}{3}(\rho^k - r_{k+1}^*)}{(\rho^{k+1} - r_{k+1}^*)} + \frac{13}{6\xi} \text{Ln} \frac{\frac{26}{3}(\gamma^k - z_k^*)}{(d_{k+1} - z_k^*)}$$

passos do algoritmo de centro analítico.

Comentários

A análise de convergência utilizou dois ingredientes básicos, um de teoria de algoritmos de planos secantes - a convergência da seqüência r_k^* para $f(x^*)$, e a teoria do algoritmo de centros, com suas estimativas de decréscimo da cota superior, que nos forneceram cotas para a quantidade de iterações de C. A., por iteração do algoritmo. Como desejávamos obter estimativa de complexidade por iteração, iniciamos por esta abordagem.

Na próxima seção, apresentamos uma demonstração alternativa, que prescinde da complexidade entre ρ^k e ρ^{k+1} . Ainda a metodologia se utilizará de propriedade de ambos os algoritmos de planos secantes e C. A. em programação linear, conforme veremos.

3.6.5 Uma demonstração alternativa de convergência

O lema que se segue é um resultado geral do método de planos secantes, fazendo parte do teorema geral de convergência. Ele espelha a proximidade do modelo à função para um número suficiente de iterações.

Lema 8

Suponha que a seqüência $\{x^k\}$ do algoritmo seja infinita. Então, dado $\varepsilon > 0$, existe $K(\varepsilon) > 0 \in N$ tal que, para todo $k \geq K(\varepsilon)$

$$f(x^k) - r^k \leq \varepsilon.$$

Demonstração:

Por absurdo, supor $f(x^k) - r^k > \varepsilon$, para $k \geq \bar{k}$, $\forall \bar{k} > 0$, para algum $\varepsilon > 0$.

Então, $r^k < f(x^k) - \varepsilon \Rightarrow f(x_i) + s_i^T(x^k - x_i) < f(x^k) - \varepsilon$, $i = 1, 2, \dots, k-1$.

Portanto, $-\varepsilon > f(x_i) - f(x^k) + s_i^T(x^k - x_i)$. Seja L a constante de Lipschitzian para f no compacto S ; tem-se que

$$\varepsilon < 2L\|x^k - x_i\|, \text{ para } i < k = 1, 2, \dots$$

Isto é um absurdo, pois $\{x^k\} \subset S$. Logo, $f(x^k) - r^k \leq \varepsilon$ e $k \geq \bar{k}$.

Lema 9

Se $f(x^k) - r^k \leq \frac{\varepsilon}{2}$, então existe $\bar{k}(\varepsilon) \in N$ tal que, para todo $k \geq \bar{k}(\varepsilon)$,

$$\rho^k - r^k \leq \varepsilon.$$

Demonstração: por hipótese, para $k \geq \bar{k}(\varepsilon)$,

$$\rho^k - r^k \leq \rho^k - f(x^k) + \frac{\varepsilon}{2} \leq \rho^k - f_m^k + \frac{\varepsilon}{2},$$

onde foi utilizada a definição de f_m^k . Podemos escrever

$$\rho^k - r^k \leq \rho^k - \rho^* + \rho^* - f_m^k + \frac{\varepsilon}{2} \leq \rho^k - \rho^* + \frac{\varepsilon}{2}.$$

Como $\rho^k \rightarrow \rho^*$, existe $\bar{k}(\varepsilon)$, que podemos escolher superior a $k(\varepsilon)$, tal que, para $k \geq \bar{k}(\varepsilon)$,

$$\rho^k - r^k \leq \varepsilon,$$

como queríamos demonstrar.

Um resultado particular da teoria do algoritmo de C. A. é dado pela fórmula (26), da demonstração do teorema 3 (lemas 3.2 e 3.6 de Den Hertog):

$$\rho^k - r_k^* \leq \frac{91}{72}(\rho^k - r^k). \quad (44)$$

Teorema 3'

As seqüências $\{\rho^k\}$ e $\{f_m^k\}$ convergem para $f(x^*)$ e o algoritmo termina com uma solução ε -ótima relativa a f_m^k :

$$f_m^k \leq f(x^*) + \varepsilon.$$

Demonstração: para k suficientemente grande, temos, de (44) e do lema 9

$$\rho^k - r_k^* \leq \frac{91}{72}(\rho^k - r^k) \leq \varepsilon.$$

Sabemos que $\rho^k \rightarrow \rho^*$ e $r_k^* \rightarrow f(x^*)$, logo

$$\rho^* \leq f(x^*) + \varepsilon.$$

Este resultado foi, na verdade, demonstrado para todo $\varepsilon > 0$. Como $\rho^* \geq f(x^*)$, temos a igualdade. A segunda parte, é como no teorema 3.

3.8 Experiência numérica

O algoritmo foi implementado em Matlab[®] em um microcomputador sem recursos especiais. Para verificar a confiabilidade do software, alguns experimentos numéricos com problemas clássicos de programação convexa não suave retirados de (Lemaréchal, Mifflin, 1977), (Shor, 1985), (Schram, 1990), (Kiwel, 1990) e em referências citadas nestes artigos, são reportados. Os resultados são comparados com os obtidos por quatro outros métodos para otimização não suave. O quadro VI informa códigos, autores e métodos; o método apresentado neste capítulo corresponde ao códigos CA-1. Ele implementa o algoritmo básico do quadro V.

Código	Autor	Método
<i>MIFCI</i>	<i>C. Lemaréchal</i>	<i>ε-steepest descent</i>
<i>BT</i>	<i>H. Schramm</i>	<i>região de confiança</i>
<i>PB</i>	<i>M. M. Makeda</i>	<i>feixes proximais</i>
<i>GHV</i>	<i>Goffin, Haurie, Vial</i>	<i>Centro analítico</i>
<i>CA-1</i>	<i>M. A. dos Santos</i>	<i>Alg. centro analítico</i>

Quadro VI
Códigos e autores

Os resultados estão sumarizados nos quadros VII e VIII. As quatro primeiras colunas do quadro VIII.b são de (Makela, Neittaanmaki, 1992), onde a seguinte convenção é adotada:

it: número de iterações,

nf: número de avaliações da função e subgradiente (oráculos) e

*f**: valor final da função objetivo.

O quadro VIII ajuda a dar uma idéia mais precisa do desempenho do método. Nele estão arrolados o número total de iterações de centros analíticos despendidas.

A utilização do resultado do teorema 1 para a reinicialização, na forma em que foi implementada, mostrou ser ineficiente, quando comparada com os resultados de (Goffin, Haurie, Vial, 1992).

Finalmente, cumpre acrescentar que no código produzido, não foram implementadas estratégias para a atualização da inversa da matriz H , como em (Karmakar, 1979); sempre que necessário, a matriz H foi invertida e não simplesmente atualizada. Isto se deve ao fato do código ser um protótipo, que utiliza uma ferramenta típica de prototipagem (MatLab).

A limitada experiência computacional com o algoritmo do quadro V tem mostrado que o método trabalha, pelo menos, tão bem quanto os outros. A seguir são apresentados os problemas usados como teste.

Problemas Teste

CB2 (Charalambous/Bandler) (Charalambous, Conn, 1978)

Dimensão 2,

Função objetivo $f(x) = \max \{ x_1^2 + x_2^4, \quad (2 - x_1)^2 + (2 - x_2)^2, \quad 2e^{-x_1 + x_2} \},$

Ponto ótimo $x^* = (1,1393, 0,8994),$

Valor ótimo $f(x^*) = 1,9522$,

Ponto inicial $x^0 = (-1.5, 2)$.

CB3 (Charalambous/Bandler) (Charalambous, Conn, 1978)

Dimensão 2,

Função objetivo $f(x) = \max \{x_1^4 + x_2^2,$
 $(2 - x_1)^2 + (2 - x_2)^2,$
 $2e^{-x_1+x_2} \},$

Ponto ótimo $x^* = (1, 1)$,

Valor ótimo $f(x^*) = 2$,

Ponto inicial $x^0 = (2, 2)$.

Dem (Demyanov/Malozemov) (Dem'yanov, 1986)

Dimensão 2,

Função objetivo $f(x) = \max \{5x_1 + x_2,$
 $-5x_1 + x_2,$
 $x_1^2 + x_2^2 + 4x_2 \},$

Ponto ótimo $x^* = (0, -3)$,

Valor ótimo $f(x^*) = -3$,

Ponto inicial $x^0 = (1, 1)$.

QL (Makela, Neittaanmaki, 1992)

Dimensão 2,

Função objetivo $f(x) = \max \{x_1^2 + x_2^2,$
 $x_1^2 + x_2^2 + 10(-4x_1 - x_2 + 4),$
 $x_1^2 + x_2^2 + 10(-x_1 - 2x_2 + 6) \},$

Ponto ótimo $x^* = (1, 2, 2, 4)$,

Valor ótimo $f(x^*) = -7.2$,

Ponto inicial $x^0 = (-1, 5)$.

LQ (Makela, Neittaanmaki, 1992)

Dimensão 2,
 Função objetivo $f(x) = \max \{ -x_1 - x_2, -x_1 - x_2 + (x_1^2 + x_2^2 - 1) \}$,
 Ponto ótimo $x^* = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$,
 Valor ótimo $f(x^*) = -\sqrt{2}$,
 Ponto inicial $x^0 = (-0,5, -0,5)$.

Mifflin 1 (Lemaréchal, Mifflin, 1977)

Dimensão 2,
 Função objetivo $f(x) = -x_1 + 20 \max \{ x_1^2 + x_2^2 - 1, 0 \}$,
 Ponto ótimo $x^* = (1, 0)$,
 Valor ótimo $f(x^*) = -1$,
 Ponto inicial $x^0 = (0,8, 0,6)$.

Mifflin 2 (Lemaréchal, Mifflin, 1977)

Dimensão 2,
 Função objetivo $f(x) = -x_1 + 2(x_1^2 + x_2^2 - 1) + 1,75|x_1^2 + x_2^2 - 1|$
 Ponto ótimo $x^* = (1, 0)$,
 Valor ótimo $f(x^*) = -1$,
 Ponto inicial $x^0 = (-1, -1)$.

Rosen (Makela, Neittaanmaki, 1992)

Dimensão 4,
 Função objetivo $f(x) = \max \{ f_1(x), f_1(x) + 10 f_2(x), f_1(x) + 10 f_3(x), f_1(x) + 10 f_4(x) \}$,

where

$$\begin{aligned} f_1(x) &= x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4, \\ f_2(x) &= x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8, \\ f_3(x) &= x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10, \end{aligned}$$

$$f_4(x) = x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5,$$

Ponto ótimo $x^* = (0, 1, 2, -1)$,

Valor ótimo $f(x^*) = -44$,

Ponto inicial $x^0 = (0, 0, 0, 1)$.

Shor (Shor, 1985)

Dimensão 5,

Função objetivo $f(x) = \{ \max_{1 \leq i \leq 10} \{ b_i \sum_{j=1}^5 (x_j - a_{i,j})^2 \} \}$,

Valor ótimo $f(x^*) = 22,60016$,

Ponto inicial $x^0 = (0, 0, 0, 0, 1)$.

Maxquad (Lemaréchal, Mifflin, 1977)

Dimensão 10,

Função objetivo $f(x) = \max_{1 \leq i \leq 5} \{ x^T A_i x - b_i^T x \}$,

Valor ótimo $f(x^*) = 0,8414084$,

Ponto inicial 1 $x^0 = (1, \dots, 1)$,

Ponto inicial 2 $x^0 = (0, \dots, 0)$,

Maxq (Makela, Neittaanmaki, 1992)

Dimensão 20,

Função objetivo $f(x) = \max_{1 \leq i \leq 20} \{ x_i^2 \}$,

Valor ótimo $f(x^*) = 0$,

Ponto inicial $x_i^0 = i, i = 1, \dots, 10$,

$x_i^0 = -i, i = 11, \dots, 20$.

Maxl (Makela, Neittaanmaki, 1992)

Dimensão 20,

Função objetivo $f(x) = \max_{1 \leq i \leq 20} \{ |x_i| \}$,

Valor ótimo $f(x^*) = 0$,

Ponto inicial $x_i^0 = i, i = 1, \dots, 10$,

$x_i^0 = -i, i = 11, \dots, 20$.

Goffin (Makela, Neittaanmaki, 1992)

Dimensão 50,

Função objetivo $f(x) = 50 \max_{1 \leq i \leq 50} \{x_i\} - \sum_{i=1}^{50} x_i$,Valor ótimo $f(x^*) = 0$,Ponto inicial $x_i^0 = i - 25.5, i = 1, \dots, 50$.**TR48** (Kiwel, 1987)

Dimensão 48,

Função objetivo $f(x) = \sum_{j=1}^{48} d_j \max_{1 \leq i \leq 48} \{x_i - a_{i,j}\} - \sum_{i=1}^{48} s_i x_i$,Valor ótimo $f(x^*) = -638565$,Ponto inicial $x_i^0 = 0, i = 1, \dots, 48$.

Legendas dos Quadros

Uma iteração completa do algoritmo é composta de

1. Incorporação de uma restrição: c^k iterações de C. A. e N_p^k passos de Newton e
2. Melhoria da cota ρ^k : p^k iterações de C. A. e N_c^k passos de Newton.

No quadros a seguir, adotamos a convenção $c = \sum_{k \geq 0} c^k$, $N_r = \sum_{k \geq 0} N_r^k$, $p = \sum_{k \geq 0} p^k$ e

$$N_c = \sum_{k \geq 0} N_c^k.$$

Algoritmo		M1FC1	BT	PB	CA-1
Problema (n)	Solução ótima	f^*	f^*	f^*	f^*
CB2 (2)	1,952225	1.952253	1.952225	2.000000	1.952225
CB3 (2)	2,000000	2.001415	2.000000	-3.000000	2.000003
DEM (2)	-3,000000	-3.000000	-3.000000	7.200001	-2.999999
QL (2)	7,200000	7.200018	7.200009	-1.414213	7.200000
LQ (2)	-1,414214	-1.141420	-1.414214	-1.1414213	-1.414207
Miffilin1 (2)	-1,000000	-0.999967	-1.000000	-1.000000	-0.999999
Miffilin2 (2)	-1,000000	-0.999993	-1.000000	-43.99999	-0.999995
Rosen (4)	-44,000000	-43.99998	-43.99998	-43.99999	-43.99995
Shor (5)	22,600160	22.60018	22.60016	22.60016	22.60019
Maxquad1 (10)	-0,841408	-0.841359	-0.841408	-0.841408	-0.841413
Maxquad2 (10)	-0,841408	-0.841359	-0.841408	-0.841408	-0.841202
Maxq (20)	0,000000	0.000000	0.000000	0.000000	0.000002
Maxl (20)	0,000000	0.000000	0.000000	0.000000	0.000000
TR48 (48)	-638565	-633625.5	-638565.0	-638565.0	-638565.9
Goffin (50)	0,000000	46.2	0.000000	0.000000	0.000000

Quadro VII

Valor da solução ótima, valor alcançado pelos algoritmos e dimensão (n) do problema¹

¹ Não dispomos do valor da solução alcançada por GHV

Método	M1FC1		BT		PB		GHV		CA-1
Testes	<i>it</i>	<i>nf</i>	<i>it</i>	<i>nf</i>	<i>it</i>	<i>nf</i>	<i>it</i>	<i>nf</i>	<i>it</i>
CB2	11	31	13	16	15	16			10
CB3	12	44	13	21	15	16			11
DEM	10	33	9	13	7	8			11
QL	12	30	12	17	17	18			10
LQ	16	52	10	11	14	15			8
Mifflin1	143	281	49	74	22	23			10
Mifflin2	30	71	6	13	16	17			14
Rosen	22	61	22	32	40	41			19
Shor	21	69	29	30	26	27	29	29	39
Maxquad1	29	69	45	56	41	42	87	87	79
Maxquad2	20	54	45	49	33	34			106
Maxq	144	207	125	128	158	159			170
Maxl	138	213	74	84	34	35			7
Goffin	72	194	50	53	51	52			1
TR48	163	284	165	179	152	153	237	237	186

Quadro VIII.a
 Comparações com o código CA-1

Método	CA-1						GHV
	c	N_r	p	N_c	$c + p$	$N_r + N_c$	
CB2	10	66	21	37	31	103	
CB3	11	53	22	43	33	96	
DEM	11	11	41	80	52	91	
QL	11	51	21	49	32	100	
LQ	8	39	12	26	20	65	
Mifflin1	10	41	16	36	26	77	
Mifflin2	14	96	18	37	32	103	
Rosen	19	129	19	38	38	167	
Shor	39	302	39	70	78	372	68
Maxquad1	81	638	79	79	160	717	182
Maxquad2	106	618	106	128	212	746	
Maxq	291	401	170	321	461	722	
Maxl	7	7	35	68	42	65	
Goffin	1	1	8	16	9	17	
TR48	412	513	360	911	772	1424	515

Quadro VIII.b
Número de iterações

Capítulo 4

DECOMPOSIÇÃO DE PROBLEMAS DE PROGRAMAÇÃO LINEAR

O princípio de Decomposição de Dantzig-Wolfe (Dantzig, Wolfe, 1960) deu origem a uma classe de algoritmos na qual, durante algum tempo, foram depositadas as esperanças de resolver eficientemente problemas de programação linear de grande porte. Ao lado da atraente idéia da divisão do problema em partes, ele traz, menos aparente, um esquema de descentralização e distribuição. Infelizmente, as esperanças não se confirmaram. A impressão, depois de todo o esforço de pesquisa empreendido nas décadas de 60 e 70, é de que o método carece de boas qualidades de convergência; muito do ganho obtido com o trabalho em subproblemas menores é perdido com o elevado número de iterações entre o problema mestre e as partes.

Depois de décadas de domínio do método Simplex, a programação matemática sofreu o impacto do trabalho de Khachian (Khachian, 1979), que vislumbrou, teoricamente, a possibilidade de resolver em tempo polinomial, problemas de programação linear. Antes disto, Klee-Minty (Klee-Minty, 1972) divulgou uma classe de problemas que demonstrava a exponencialidade do método Simplex. Posteriormente, a partir do

trabalho de (Karmakar, 1984), vários autores obtiveram métodos, genericamente conhecidos como pontos interiores, que reduzem os limites obtidos por Khachian.

A principal aplicação da programação convexa é na decomposição de problemas de grande porte, conforme afirmado por (Hiriart-Urruty, Lemarechal, 1993). Alguns autores (Kim, Nazareth, 1991), (Den Hertog, 1992) conjecturam que o uso de pontos interiores para resolver, ou o problema mestre, ou os subproblemas originados da decomposição, possa dar nova vida aos métodos tradicionais.

O objetivo deste capítulo é aplicar um método de pontos interiores à decomposição de problemas de programação linear, consolidando e estendendo o uso da metodologia desenvolvida no capítulo 3. São apresentados algoritmos para resolver o problema originado da aplicação do princípio de decomposição de Dantzig-Wolfe a problemas de programação linear de grande porte, utilizando o método do centro analítico.

4.1 Princípio de decomposição de Dantzig-Wolfe

A decomposição consiste em reduzir os elementos de um modelo a partes mais simples que interagem, hierarquicamente, entre si. Um problema é decomposto por partições que, basicamente, podem ser feitas de dois modos; ou dividem-se as restrições, ou dividem-se as variáveis. A partição de restrições (Dantzig, Wolfe, 1960), induz a decomposição via preços, onde a avaliação de uma decisão x^k , tomada em um nível hierárquico inferior, é feita segundo preços imputados aos recursos partilhados. Na partição de variáveis (Benders, 1962), os níveis hierárquicos inferiores estabelecem cotas para a solução ótima, que são reavaliadas, iterativamente, pela partição situada em um nível hierárquico mais alto. Estas duas estratégias estão intimamente relacionadas; no caso da programação linear, existe uma relação de dualidade entre ambas. Métodos mistos geralmente são sobreposições destes dois princípios.

O método de Dantzig-Wolfe divide as restrições e, utilizando a função dual associada, obtém critérios para avaliar o desempenho dos subproblemas resultantes.

Seja o seguinte problema de programação linear:

$$\text{Minimizar } c^T x \quad (1)$$

sujeito a:

$$Ax = b,$$

$$Dx \geq d,$$

$$x \geq 0,$$

onde c , x têm n componentes; b e d são vetores de dimensões m_1 e m_2 ; A e D são matrizes de dimensões (m_1, n) e (m_2, n) .

Sem perda de generalidade, (1) pode ser escrito como

$$\text{Minimizar } c^T x \quad (2)$$

sujeito a:

$$Ax = b,$$

$$x \in X,$$

onde $X := \{x: Dx = d, x \geq 0\}$. No que se segue, supõe-se que X seja compacto.

A função Lagrangeana associada, com u sendo um vetor de dimensão apropriada, é:

$$L(x, u) := c^T x - u^T (Ax - b)$$

ou,

$$L(x, u) := u^T b + (c^T - u^T A)x.$$

A função dual $D(u)$ associada é:

$$D(u) := u^T b + \min\{(c^T - u^T A)x : x \in X\}.$$

Para cada u fixo, resolve-se o seguinte subproblema para obter $D(u)$:

$$\text{Minimizar } (c^T - u^T A)x : x \in X \quad (3)$$

O problema de otimização

$$\text{Max } D(u) : u \in R^m \quad (4)$$

É, como se sabe, o problema dual de (1). Neste capítulo a estratégia é resolver diretamente o problema dual.

Sabe-se que a função

$$D(u) := u^T b + (c^T - u^T A)x^*(u), \quad (5)$$

onde $x^*(u)$ é a solução do problema (3), é concava e não necessariamente diferenciável em u e que

$$s = (b - A x^*(u))$$

é um subgradiente de D no ponto u . De fato, temos

$$D(v) = v^T b + (c^T - v^T A)x^*(v) \leq v^T b + (c^T - v^T A)x^*(u), \text{ para todo } u, v$$

ou,

$$D(v) \leq v^T b + (c^T - v^T A)x^*(u) + u^T A[x^*(u) - x^*(u)] + u^T (b - b) \Rightarrow$$

$$D(v) \leq u^T b + (c^T - u^T A)x^*(u) + s^T (v - u) = D(u) + s^T (v - u).$$

Uma solução $x_\varepsilon(u) \in X$ do problema (3), que atende um dado critério de aproximação, gera um ε -subgradiente de D em u (lema 1) e produz cotas para o valor de $D(u)$.

4.2 Algoritmos

Nesta seção estão propostos dois algoritmos para resolver o problema de programação côncava associado ao princípio de decomposição de Dantzig-Wolfe. O primeiro deles é uma aplicação direta do algoritmo básico do quadro V do capítulo 3; onde somente o problema mestre é parcialmente resolvido em cada iteração. Já o outro, utiliza soluções parciais, tanto do problema mestre, quanto do subproblema escravo em cada iteração¹. Neste caso, são empregados ε -subgradientes no lugar de subgradientes para obter os cortes que separam a solução ótima.

4.2.1 O centro analítico e a função potencial

A seguir apresentamos resultados retirados de (Goffin, Haurie, Vial, 1992). Eles relacionam o centro analítico de um polítopo associado a um problema de programação linear ao mínimo de uma função potencial associada a um polítopo do problema dual, conforme estabelecido por (Ross, Vial, 1988).

Para o problema de programação linear

$$\text{Minimizar } \rho = \gamma^T x \quad (6)$$

sujeito a:

$$Dx - s = \delta,$$

$$s \geq 0;$$

onde $D \in R^{m_2 \times n}$, $\gamma, x \in R^n$ e $s, \delta \in R^{m_2}$; o problema dual é:

$$\text{Maximizar } \delta^T y \quad (7)$$

sujeito a:

$$D^T y = \gamma,$$

¹ Em Dantzig-Wolfe tanto o problema mestre quanto o subproblema, são completamente resolvidos em cada iteração.

$$y \geq 0.$$

Considere $\bar{\rho}$ uma cota superior da solução ótima de (6) e seja

$$S_{\bar{\rho}} := \{x : \underbrace{\bar{\rho} \geq \gamma^T x, \dots, \bar{\rho} \geq \gamma^T x}_q, Dx \geq \delta\}.$$

O centro analítico de $S_{\bar{\rho}}$ é o mínimo da função barreira logarítmica

$$\varphi_{\bar{\rho}}(x, \bar{\rho}) := -q \text{Ln}(\bar{\rho} - \gamma^T x) - \sum_{i=1}^{m_2} \text{Ln}(s_i), \quad (8)$$

$$s. a: Dx - s = \delta, \quad s > 0.$$

Dada a estrita convexidade de (8), as condições de otimalidade são também suficientes.

Denominando y o multiplicador relativo às igualdades, temos:

$$\gamma - D^T y \left(\frac{\bar{\rho} - \gamma^T x}{q} \right) = 0, \quad (9)$$

$$S^{-1} e - y = 0$$

e

$$Dx - s = \delta.$$

Por conveniência, definimos a mudança de variável dada por

$$z = \frac{y(\bar{\rho} - \gamma^T x)}{q}.$$

O sistema (9) se reescreve como:

$$D^T z = \gamma, \quad (a) \quad (9')$$

$$Dx - s = \delta \quad (b)$$

e

$$Sz = \frac{\rho - \gamma^T x}{q} e. \quad (c)$$

Agora, seja a seguinte função potencial associada ao politopo dual:

$$\psi(z, \bar{\rho}) = (m_2 + q) \text{Ln}(\bar{\rho} - \delta^T z) - \sum_{i=1}^{m_2} \text{Ln}(z_i). \quad (10)$$

Sabe-se que ψ possui um único ponto de mínimo no conjunto $\{y: D^T z = \gamma, z \geq 0\}$, caracterizado por :

$$-\frac{m_2 + q}{\bar{\rho} - \delta^T z} \delta - Z^{-1} e + Dv = 0 \quad (11)$$

$$D^T z = \gamma,$$

onde denotamos por v o multiplicador associado às igualdades. A seguir, mostraremos que (9) e (11) são equivalentes.

Se fizermos a mudança de variável

$$x = v \frac{\bar{\rho} - \delta^T z}{m_2 + q} \quad \text{e} \quad s = \frac{\bar{\rho} - \delta^T z}{m_2 + q} Z^{-1} e,$$

a primeira das condições (11) se reescreve

$$-\frac{m_2 + q}{\bar{\rho} - \delta^T z} \delta - \frac{m_2 + q}{\bar{\rho} - \delta^T z} s + \frac{m_2 + q}{\bar{\rho} - \delta^T z} Dx = 0,$$

que é a segunda igualdade em (9'). Além disto,

$$Sz = \frac{\bar{\rho} - \delta^T z}{m_2 + q} e. \quad (11')$$

Observa-se a similaridade (9') com (9'a), (9'b) e (11'). De (9'c) temos

$$z^T s = e^T SZe = m_2 \frac{\bar{\rho} - \gamma^T x}{q},$$

enquanto de (9'a) e (9'b),

$$z^T s = z^T D x - z^T \delta = \gamma^T x - z^T \delta.$$

Assim

$$\gamma^T x - z^T \delta = m_2 \frac{\bar{\rho} - \delta^T z}{m_2 + q}.$$

Por substituição direta, temos, com (11'),

$$m_2 \frac{\bar{\rho} - \gamma^T x}{q} = m_2 \frac{\bar{\rho} - \delta^T z}{m_2 + q},$$

ou,

$$\delta^T z = \frac{m_2 + q}{q} \gamma^T x - \frac{m_2}{q} \bar{\rho},$$

o que mostra que a solução de (9') é a mesma que de (9'a), (9'b) e (11').

Logo, um ponto que minimiza a função barreira logarítmica (8) permite computar o mínimo da função potencial (10), através de (11'), associado ao problema dual do problema de programação linear. Este relacionamento será explorado para computar uma cota inferior para a solução ótima do problema escravo. No que se segue, supõe-se que o problema escravo está sendo resolvido pelo método dos centros analíticos (ver quadro XIII.b)

4.2.2 Algoritmos para decomposição de problemas de programação linear

Decomposição de Dantzig-Wolfe

No método de planos secantes de (Kelley, 1960), uma função côncava $D(u)$ é sucessivamente aproximada por um conjunto de formas lineares dadas por :

$$D(u) \approx D^\Delta(u) := \min\{D(u_i) + s_i^T(u - u_i), i = 1, 2, \dots, l, u \in S\},$$

onde S é um conjunto compacto que contém $u^* = \arg \max D(u)$. Em cada iteração, u_{k+1} é computado por

$$u_{k+1} := \arg \max[\min\{D(u_i) + s_i^T(u - u_i), i = 1, 2, \dots, l, u \in S\}],$$

que é equivalente a

$$(u_{k+1}, r^{k+1}) := \arg \max\{r : r \leq D(u_i) + s_i^T(u - u_i), i = 1, 2, \dots, l, u \in S, r \in R\}.(12)$$

O problema (12) é conhecido como problema mestre e o subproblema (3), como problema escravo.

Resolvendo-se integralmente o problema escravo em cada iteração, não há dificuldade em empregar os resultados do capítulo anterior. No quadro IX encontra-se a adaptação do algoritmo do quadro V para a decomposição de Dantzig-Wolfe.

algoritmo para Decomposição de Dantzig-Wolfe

Escolha $\rho^0 \leq r^*$, $\rho^0 < D_{\max}^0 = D(u_0)$, $\theta \in (0,1)$ e $\varepsilon > 0$

Escolha $(u_0, r^0) \in \text{int}(S_{\rho^0}^0) : \delta_{\rho^0}^0 \leq 0,1$

$k := 0$

repita

se $(r^k - \rho^k) < \varepsilon$

então interrompa

fim se

Compute $D(u_k)$,

$s_{k+1} \in \partial D(u_k)$ e $d_{k+1} = s_{k+1}^T u_k - D(u_k)$ { ver quadro IX.a}

Compute γ

Reduza γ até d_{k+1}

{ Incorpora o corte ao feixe }

$S_{\rho^k} := S_{\rho^k} \cap \{(x, r) \in R^{n+1} : s_{k+1}^T u - r \leq d_{k+1}\}$

$D_{\max}^{k+1} := \max\{D(u_k), D_{\max}^k\}$

$\rho^{k+1} := D_{\max}^{k+1} - \theta(D_{\max}^{k+1} - \rho^k)$

Obtenha (u_{k+1}, r^{k+1}) , próximo ao Centro Analítico de $S_{\rho^{k+1}}$

$k := k + 1$

fim repita

fim algoritmo

Quadro IX

Algoritmo para programação côncava não suave aplicado à
Decomposição de Dantzig-Wolfe

Decomposição de Dantzig-Wolfe com problema escravo aproximado

Seja agora, um algoritmo que utiliza soluções sub-ótimas do problema escravo.

Veremos que estas soluções geram ε -subgradientes e e cotas, para cada iterado u_i :

$$D_l(u_i) \leq D(u_i) \leq D_u(u_i).$$

refinamento “Compute $D(u_k)$, $s_{k+1} \in \partial D(u_k)$ e $d_{k+1} = s_{k+1}^T u_k - D(u_k)$ ”

$$x^*(u^k) := \arg \min \{ (c^T - u_k^T A)x : x \in X \}$$

$$D(u_k) := u_k^T b + (c^T - u_k^T A)x^*(u_k)$$

$$s_{k+1} := [b - Ax^*(u_k)]$$

$$d_{k+1} := s_{k+1}^T u_k - D(u_k)$$

fim refinamento

Quadro IX.a

Obtenção de um subgradiente, cálculo de $D(u_k)$ e determinação de d_{k+1} para o algoritmo do quadro IX

Na iteração k é conhecido (u_k, r^k) , próximo do centro analítico de S_{ρ^k} ,

$$S_{\rho^k} := S \cap \{ (u, r) \in R^{m+1} : e_i^T u - r \geq d_i, i = 1, 2, \dots, k \} \cap \\ \cap \{ (u, r) \in R^{m+1} : \underbrace{r \geq \rho^k, \dots, r \geq \rho^k}_{q \text{ vezes}} \},$$

onde $\bar{d}_i = e_i^T u_i - D_u(u_i)$, associada ao modelo de planos secantes

$$\text{Maximizar } r: e_i^T u - r \geq \bar{d}_i, i = 1, 2, \dots, k \text{ e } u \in S.$$

O algoritmo (quadro X) usa uma solução subótima $x_x(u_k)$ do problema escravo (quadro X.a) tal que

$$(c^T - u_k^T A)x_x(u_k) < r^k - u_k^T b. \quad (13)$$

Esta solução, quando existente, produz um semi-espaço que separa a iteração corrente (lema 3).

algoritmo para Decomposição de Dantzig-Wolfe (com soluções sub-ótimas)

Escolha $\rho^0 \leq r^*$, $\rho^0 < D_{\max}^0 = D(u_0)$, $D_{\max}^0 \leq D(u^*)$, $\theta \in (0,1)$ e $\varepsilon > 0$

Escolha $(u_0, r^0) \in \text{int}(S_{\rho^0})$: $\delta_{\rho^0} \leq 0,1$

$k := 0$

repita

se $(r^k - \rho^k) < \varepsilon$

então interrompa

fim se

Compute $x_x(u_k)$ { ver quadro X.a }

Compute $D_u(u_k)$, $e_{l+1} \in \partial_\varepsilon D(u_k)$ e $\bar{d}_{k+1} = e_{k+1}^T u_k - D_u(u_k)$
 { ver quadro X.b }

Compute γ

Reduza γ até d_{k+1}

{ Incorpora o corte ao feixe }

$S_{\rho^k} := S_{\rho^k} \cap \{(x, r) \in R^{n+1} : e_{k+1}^T u - r \geq \bar{d}_{k+1}\}$

Compute $D_l(u_k)$

$D_{\max}^{k+1} := \max\{D_l(u_k), D_{\max}^k\}$

$\rho^{k+1} := D_{\max}^{k+1} - \theta(D_{\max}^{k+1} - \rho^k)$

Obtenha (u_{k+1}, r^{k+1}) , próximo ao Centro Analítico de $S_{\rho^{k+1}}$

$k := k + 1$

fim repita

fim algoritmo

Quadro X

Algoritmo que utiliza soluções parciais dos programas mestre e escravo

refinamento "Compute $x_x(u_k)$ "

{ Dado $x^0 : x^0 \in X$, próximo ao centro analítico de X }

Compute γ' : x^0 continue próximo ao centro analítico da região

$$X \cap \{x \in R^n : \underbrace{(c^T - u_k^T A)x \leq \gamma', \dots, (c^T - u_k^T A)x \leq \gamma'}_{l' \text{ vezes}}\}$$

$k := 0$

repita

se $x^k = \arg \min \{(c^T - u_k^T A)x : x \in X\}$ ou

$$(c^T - u_k^T A)x^k < r^k - u_k^T b$$

então interrompa

fim se

$x^{k+1} :=$ próxima iteração do método de centros analíticos aplicado ao problema escravo

$k := k + 1$

fim repita

$$x_x(u_k) := x^k$$

fim refinamento

Quadro X.a

Cálculo de uma solução parcial do problema escravo

Para obter esta solução sub-ótima, supõe-se conhecida uma solução inicial x^0 , próxima do centro analítico da região X , que será utilizada em todas as iterações do algoritmo. Um procedimento de reinicialização, nos moldes daquele visto no capítulo anterior, é então aplicado, até obter (13). Esta solução, quando existente², fornece um ε -subgradiente e_{k+1} de D em u_k e permite obter \bar{d}_{k+1} (ver quadro XIII.b).

² Se não for possível atingir este limite, tem-se o subproblema escravo inteiramente resolvido, possibilitando produzir um subgradiente.

Se $\text{int}(S_{\rho^k}^{k+1})$ é não vazio, o procedimento de reinicialização aplica parcialmente o método de centro analítico a um problema auxiliar (ver seção 3.4.1), e encontra um ponto interior (u_c^k, r_c^k) , próximo à trajetória central associada à região perturbada (caso contrário, $\text{int}(S_{\rho^k}^{k+1}) = \emptyset$ e o algoritmo ignora a restrição gerada pelo ε -subgradiente³). A seguir, é atualizado D_{\max}^{k+1} , o melhor valor conhecido para a solução do problema.

refinamento “Compute $D_u(u_k)$, $e_{k+1} \in \partial_\varepsilon D(u_k)$ e $\bar{d}_{k+1} = e_{k+1}^T u_k - D_u(u_k)$ ”

$$D_u(u_k) := u_k^T b + (c^T - u_k^T A)x_k(u_k)$$

$$e_{k+1} := [b - Ax_k(u_k)]$$

$$d_{k+1} := e_{k+1}^T u_k - D_u(u_k)$$

fim refinamento

Quadro X.b

Obtenção de um ε -subgradiente, cálculo de $D_u(u_k)$ e determinação de \bar{d}_{k+1}

Seja $z_k(\pi)$ o valor da função objetivo do problema dual associado ao problema escravo (3), avaliado no ponto π que minimiza a função potencial (10). Do teorema da dualidade fraca, tem-se que

$$(c^T - u_k^T A)x^* \geq z_k(\pi) \Rightarrow$$

$$u_k^T b + (c^T - u_k^T A)x^* \geq u_k^T b + z_k(\pi) \Rightarrow$$

$$D(u_k) \geq u_k^T b + z_k(\pi) .$$

³ A ocorrência deste fato não é mostrada no algoritmo.

Seja $D_l(u_k) = u_k^T b + z_k(\pi)$, uma cota inferior para D . A fórmula de atualização de D_{\max}^{k+1} passa a ser

$$D_{\max}^{k+1} := \max\{D_l(u_k), D_{\max}^k\}.$$

O novo valor do limite inferior ρ^{k+1} é obtido de

$$\rho^{k+1} := \rho^k - \theta(D_{\max}^{k+1} - \rho^k).$$

De posse deste limite, o algoritmo aplica $\tau(\tau \geq 1)$ iterações do método de centros analíticos ao problema (12), começando com $\rho_p^0 = \rho^k$, $(u_p^0, r_p^0) = (u_c^k, r_c^k)$ e termina com $\rho_p^\tau = \rho^{k+1}$ e (u_{k+1}, r^{k+1}) .

O teste de otimalidade é

$$(r^k - \rho^k) < \varepsilon,$$

onde ε é uma constante positiva que indica a tolerância. Quando este teste é satisfeito, é garantido que uma ε -solução do problema de programação convexa foi encontrada

4.3 Convergência

Esta seção estende o resultado de convergência do algoritmo básico tratado no capítulo 3, para o algoritmo do quadro X.

O algoritmo básico (quadro V) não utiliza nenhuma propriedade de D , a não ser, é claro, a convexidade. É óbvio, a partir das demonstrações, que a convergência é mantida mesmo que e não seja um subgradiente. É apenas necessário que seja produzido um

corte que separa a solução ótima do resto; qualquer hiperplano capaz de separar (u^*, r^*) em S_{ρ^k} pode ser usado.

A principal diferença entre os dois algoritmos está na resolução do problema escravo em cada iteração. O segundo algoritmo utiliza uma solução sub-ótima, que não dá origem a um subgradiente; o que é possível obter é um ε -subgradiente.

Um ε -subgradiente de uma função côncava $f: R^{m_1} \rightarrow R$ em $u \in R^{m_1}$ é um elemento do conjunto ε -subdiferencial definido como:

$$\partial_\varepsilon D(u) := \{e \in R^{m_1} : D(w) \leq D(u) + \langle e, w - u \rangle + \varepsilon, \forall w \in R^{m_1}\},$$

onde $\varepsilon \geq 0$. O lema 1 mostra que uma solução parcial $x_k(u_k)$ que atende ao critério (13), produz um ε -subgradiente para D em u . Com esta solução obtém-se D_u , o Lagrangeano de D em $(u_k, x_k(u_k))$. O lema 2 mostra que com D_u e $e \in \bar{\partial}_\varepsilon D(u_k)$ tem-se um corte válido:

$$(u^*, r^*) \in \{(u, r) \in R^{m_1+1} : e^T u - r \geq e^T u_k - D_u(u_k)\}.$$

A seguir os lemas são demonstrados.

Lema 1

Seja $x \in X$ e u tais que $(c^T - u^T A)x < r - u^T b$. Então

$$s = (b - Ax) \in \bar{\partial}_\varepsilon D(u).$$

Demonstração

$$D(w) = w^T b + \min\{(c^T - w^T A)x : x \in X\} \Rightarrow$$

$$D(w) \leq w^T b + (c^T - w^T A)x \Rightarrow$$

$$D(w) \leq w^T b + (c^T - w^T A)x - u^T b + u^T b - u^T Ax + u^T Ax \Rightarrow$$

$$D(w) \leq (w - u)^T (b - Ax) + u^T b + (c^T - u^T A)x.$$

Por hipótese,

$$(c^T - u^T A)x < r - u^T b \Rightarrow$$

$$D(w) \leq (w - u)^T (b - Ax) + r.$$

Mas $r \leq \max D^\Delta(u) = D(u) + \varepsilon$, para algum $\varepsilon \geq 0$.

Então

$$D(w) \leq (w - u)^T (b - Ax) + D(u) + \varepsilon$$

$$\Rightarrow s = (b - Ax) \in \bar{o}_\varepsilon D(u)$$

Lema 2

Seja $e \in \bar{o}_\varepsilon D(u_k)$ e $D_u(u_k) := u_k^T b + (c^T - u_k^T A)x$, $x \in X$.

Então

$$(u^*, r^*) \in \{(u, r) \in R^{m+1} : e^T u - r \geq e^T u_k - D_u(u_k)\}.$$

Demonstração

Imediata, pois $e \in \bar{o}_\varepsilon D(u_k)$ e $D_u(u_k) \geq D(u_k)$.

Lema 3

Seja $x_k(u_k)$ uma solução subótima do problema $\text{Min}\{(c^T - u_k^T A)x : x \in X\}$ tal que

$$(c^T - u_k^T A)x_k(u_k) < r^k - u_k^T b$$

Então para $e_{k+1} \in \bar{o}_\varepsilon f(u_k)$, se tem

$$e_{k+1}^T u_k - r^k < \bar{d}_{k+1}.$$

Demonstração

Por hipótese,

$$(c^T - u_k^T A)x_k(u_k) < r^k - u_k^T b \Rightarrow$$

$$-r^k < -(c^T - u_k^T A)x_k(u_k) - u_k^T b \Rightarrow$$

$$e_{k+1}^T u_k - r^k < e_{k+1}^T u_k \underbrace{-(c^T - u_k^T A)x_k(u_k) - u_k^T b}_{-D_u(u_k)} \Rightarrow$$

$$e_{k+1}^T u_k - r^k < e_{k+1}^T u_k - D_u(u_k) = \bar{d}_{k+1}.$$

■

Capítulo 5

CONCLUSÕES

A noção de centro, seja ela um conceito geométrico ou analítico, permeia a teoria da otimização. Levin, Khachian e Huard dentre outros, obtiveram resultados memoráveis utilizando aquele que é “um ponto para onde convergem as coisas, como para uma natural posição de repouso”¹.

Vários algoritmos desta classe foram agrupados no capítulo 2. O objetivo deste capítulo foi sobretudo didático, não de uma revisão completa. Não vimos, por exemplo, as relações como as do método de Shor com os algoritmos dos elipsóides e das esferas. Não apresentamos também algoritmos híbridos (Goffin, 1979), como por exemplo, algoritmos que utilizassem Khachian conjugado com o método das esferas. Nestes, uma iteração de Agmon poderia vir a ser uma opção aos elipsóides de mínimo volume calculado em cada iteração do método dos elipsóides.

Uma idéia que pode ser efetiva na prática, é aplicar um dos algoritmos de centro, como por exemplo, o algoritmo dos elipsóides, para reduzir o volume de uma esfera inicial que contém a solução ótima, antes de aplicar os algoritmos dos capítulos 3 e 4. Isto é particularmente interessante, quando o problema é originado da decomposição.

¹ Dicionário Aurélio básico da língua portuguesa, Editora Nova Fronteira, 1988.

Estabelecer limites para as variáveis primais, normalmente, não é difícil; o conhecimento do significado das variáveis de decisão torna esta tarefa trivial na maioria dos problemas. O mesmo não pode ser dito quanto às variáveis duais. A tendência é de que sejam fornecidos limites artificialmente exagerados. Esta idéia lembra o procedimento para refinar intervalos de raízes de uma equação, onde o algoritmo da bissecção é usado antes de aplicar um método mais sofisticado.

No capítulo 3 foi proposto um novo método, baseado na noção de centro, para resolver o problema da programação convexa não diferenciável. A experiência computacional, no momento limitada, tem sinalizado a sua efetividade.

O algoritmo tem, pelo menos, uma característica importante. Ele emprega uma seqüência monótona estritamente decrescente de cotas para alcançar a solução e, conseqüentemente, segue trajetórias centrais associadas ao mínimo da função modelo f^Δ . Isto difere, por exemplo, do trabalho de (Nesterov, 1995), onde a convergência é estabelecida de forma semelhante aos métodos do subgradiente. Também, são desnecessários conceitos como “passo sério” e “passo nulo” dos métodos de feixe (Hiriart-Urruty, Lemaréchal, 1993).

Quanto à complexidade, não há como não dar provimento ao fato de que a estimativa apresentada é exacerbada; ela é um mero indicativo da polinomialidade do método por iteração. As duas fontes de complexidade aritmética têm um aspecto em comum; o método de centros analíticos é aplicado para reduzir uma cota da função objetivo (γ e ρ^k) para um valor conhecido (d_{l+1} e ρ^{k+1}).

Para uma eficiente implementação de um algoritmo de planos secantes no contexto em que uma nova restrição é adicionada em cada iteração, tem de ser usada alguma estratégia “build-down”. Este tipo de procedimento está relacionado com a gerência do feixe. Nos métodos convencionais de programação convexa não suave, o instante de descartar alguma restrição, introduzida em uma iteração anterior, é determinado com critérios sem uma justificativa precisa, de forma heurística. Embora não tenhamos

desenvolvido detalhadamente, propusemos uma metodologia que poderá ser útil em trabalho futuro.

No capítulo 4 foram apresentados novos algoritmos para resolver o problema resultante da aplicação do princípio de decomposição de Dantzig-Wolfe em problemas de programação linear de grande porte. Para avaliar a efetividade do algoritmo, estão sendo implementados os problemas de Khun-Quandt (Chvatal, 1980)). Estes problemas têm a seguinte forma:

$$\text{Min } -x$$

s. a:

$$Ax \leq b$$

$$x \geq 0,$$

onde A é uma matriz $m \times n$ densa e com elementos inteiros, gerados aleatoriamente em $(1, 100)$. Os elementos do vetor b estão no intervalo $(5000, 10000)$. Estes valores garantem que o problema é limitado e a origem é uma solução viável. A decomposição é realizada relaxando-se 20 por cento das restrições iniciais; as 80 por cento restantes são consideradas nos subproblemas. Em breve esperamos publicar os resultados.

A decomposição de problemas de grande porte é dirigida pela estrutura; os problemas fáceis são candidatos a serem os problemas escravos. Se o segundo algoritmo apresentado no capítulo 4, realmente exigir na maioria das vezes, soluções parciais dos problemas escravos, perde o sentido o critério de modularização baseado na estrutura. Um novo critério, baseado na efetividade do corte em um ambiente de planos secantes, talvez possa ser formulado.

REFERÊNCIAS BIBLIOGRÁFICAS

AGMON, S. 1954, “The relaxation method for linear inequalities, *Canadian Journal of Mathematics*, n. 6, pp. 382-392.

AKGUL, M., “*Topics in relaxation and ellipsoidal methods*”, Pitman Publishing Limited, 1984.

ATKINSON, D. S., VAIDYA, P.M., 1995, “A cutting plane algorithm for convex programming that uses analytic centers”, *Mathematical Programming*, n. 69, pp. 1-43.

BAHN, O., MERLE O. du, GOFFIN, J. L., VIAL, J. P., 1995, “A cutting plane method from analytic centers for stochastic programming”, *Mathematical Programming*, n. 69, pp. 45-73.

BENDERS, J., F., 1962, “Partitioning procedures for solving mixed variables programming problems”, *Numerische Mathematik*, n. 4, pp. 238-252.

BLAND, R. G., GOLDFARB, D., TODD, M. J., 1981, “The ellipsoid method: a survey”, *Operations Research*, n. 29, pp. 1039-1090.

CHARALAMBOUS, C., CONN, A. R., 1978, “An efficient method to solve the minimax problem directly”, *SIAM Journal on Numerical Analysis*, n. 15, pp. 162-187.

CHAVTAL, V., 1980, *Linear Programming*, Freeman, New York.

DANTZIG, G. B., WOLFE, P., 1960, “The decomposition algorithm for linear programs” *Operations Research*, n. 8, pp. 101-111.

DANTZIG, G. B., YE, Y., 1990, “A Build-up interior method for linear programming: affine scaling form”, *Technical Report SOL 90-4*, System Optimization Laboratory, Dept. of Operations Research, Stanford, USA.

- DEM'YANOV, V. F., MALOZEMOV, V. N., 1986, "*Introduction to minimax*", Wiley, New York.
- DEN HERTOOG, D., 1992, "Interior point approach to linear, quadratic and convex programming: algorithms and complexity", *Ph. D. Thesis*, Delft University of Technology, Delft, The Netherlands.
- DEN HERTOOG, D., ROOS, C. and TERKLAY, T., 1992, "A Build-up variant of the path following method for LP", *Operations Research Letters*, n. 12, pp., 181-186.
- EAVES, B. CURTIS, ZANGWILL, W. I., 1971, "Generalized cutting plane algorithms", *Siam J. Control*, n. 4, pp. 529 - 542.
- FEIJOO B., SÁNCHEZ, A., GONZAGA C. C., 1997,, "Maintaining closedness to the analytic center of a polytope by perturbing added hyperplanes", *Journal of Applied Mathematics and Optimization*, 2, pp. 22-32.
- GOFFIN, J. L., 1979, "Convergence of cyclic Shor-Khachian method for systems of linear inequalities", *Working Paper 79-54*, Faculty of Management, McGill University, Montreal, Quebec.
- GOFFIN, J. L., VIAL, J. P., 1989, , "On the computation of weighted analytic centers and dual ellipsoids with the projective algorithm", Manuscript, Department d'Économie Commerciale et Industrielle, University of Geneve, Italy.
- GOFFIN, J., 1980, "The relaxation method for solving systems of linear inequalities" *Management Science* 5, pp. 388-414.
- GOFFIN, J. L., 1982, "On the nonpolynomiality of the relaxation method for systems of linear inequalities", *Mathematical Programming* 22, pp. 93-103.

GOFFIN, J. L., VIAL, J.P., 1990, "Cutting planes and column generation techniques with projective algorithm", *Journal of Optimization Theory and Applications*, n. 65, pp. 409-429.

GOFFIN, J. L., HAURIE, A., VIAL, J.P., 1992, "Decomposition and nondifferentiable optimization with the projective algorithm", *Management Science*, n. 38, pp. 284-302.

HIRIART-URRUTY, J. B., LEMARÉCHAL, C., 1993, "*Convex Analysis and Algorithms II*", Springer-Verlag Berlin Heidelberg.

HORN, R., JOHNSON C., R., "Matrix Analysis", Cambridge University Press, 1985.

HUARD, P., "Resolution of mathematical Programming problems with nonlinear constraints by the method of centres, 1967, *Nonlinear Programming*, J. Abadie, North Holland Publishing Co., Amsterdam, Holland, pp. 207-219.

JUDIN, D. B., NEMIROVSKII, A. S., 1979, "Efficient methods for solving large scale convex programming problems", *Ekonomika i Matematicheskie Metody*, n. 15, pp. 135-152.

KARMAKAR, N., 1984, "A New Polynomial-time Algorithm for Linear Programming", *Combinatorica*, n. 4, pp. 373-395.

KHACHIAN, L. G., 1979, "A polynomial algorithm in linear programming", *Soviet Math. Doklady* n. 20, pp. 191-194.

KELLEY, J. E., 1960, "The cutting plane method for solving convex programming", *J. Soc. Indus. Appl. Math.*, n. 4, pp. 703-712.

KIM, K., NAZARETH, J., L., 1991, "The decomposition principle and algorithms for linear programming", *Linear Algebra and Applications*, n. 152, pp. 119-133.

KIWIEL, K. C., 1996, "Complexity of some cutting plane methods that use analytic centers", *Mathematical Programming*, n. 74, pp. 47-54.

KLEE, V., Minty, G. L., 1972, "How good is the simplex algorithm?", *Inequalities III*, Academic Press, New York, pp. 159-175.

LEMARÉCHAL, C., MIFFLIN R., (Eds.), 1988, "Nonsmooth Optimization", In: *Proceedings of a IIASA Workshop*, Pergamon Press, Oxford.

LEVIN, A. Ju, 1965, "An algorithm for the minimization of convex functions, *Soviet Mathematics Doklady* n. 6, pp. 286-290.

MAKELA, M. M., NEITTAANMAK, P. I., 1992, "*Nonsmooth Optimization*", World Scientific Publishing Co. Pte. Ltd..

Matlab, High-Performance Numeric Computation and Visualization Software, 1994, The MathWorks Inc., Massassuchets, USA.

MITIAGIN, B. S., "Two inwqualities for volumes of convex bodies", 1968, *Mathematical Notes*, n. 5, pp., 61-65.

MITCHEL, J. E., TODD, M. J., 1992, "Solving combinatorial Optimization problems using Karmakar's algorithm", *Mathematical Programming*, n. 56, pp. 245-284.

MOTZKIN, T, SHOENBERG, I. J., 1954, "The relaxation method for linear inequalities, *Canadian Journal of Mathematics*" , n. 6, 393-404.

NESTEROV, Yu., 1995, "Complexity estimates of some cutting plane methods based on the analytic barrier", *Mathematical Programming*, n. 69, pp. 149-176.

POLYAK, B. T., 1978, "Subgradient methods: a survey of soviet research", *Nonsmooth Optimization (eds. Lemarechal et Mifflin) - Pergamanon Press*, pp. 5 - 30.

ROCKAFELLAR, R. T., 1992, "Lagrange multipliers and optimality", *1992 SIAM Annual Meeting*, Los Angeles.

TONE, K., 1991, "An active-set strategy in interior point methods for linear programming", *Working Paper*, Graduate School, Saita University, Uzawa, Japan.

ROOS, C., VIAL, J. P., 1988, "Analytic Centers in Linear Programming", *Report 88-68*, Delft University of Tecnology, Delft, Netherlands.

SANCHEZ, A., OLIVEIRA, P. R., SANTOS, M. A. dos, 1996, "Recovering of the analytic center in quadratic regions and applications, *Technical Report ES-496/96*, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

SAPATNEKAR, S. S., RAO, V. B., VAIDYA, P. M., 1991, "A convex optimization approach to transistor sizing for CMOS circuits", *Proceedings of the IEEE International Conference on Computer Aided Design*, IEEE Computer Society Press, Los Alamitos, CA, pp. 482-485.

SCHRAMM H., ZOWE J., 1990, "A version of the bundle idea for minimizing a nonsmooth function: conceptual idea convergence analysis, numerical results", *DFG-report 206*, Universitat Bayreuther.

SHOR, N. Z., 1970, "Utilization of the operation of space dilatation in the minimization of convex function", *Cybernetics*, n. 6, pp. 7-15.

SHOR, N. Z., 1977, "Cut-off method with space extension in convex programming problems", *Cybernetics*, n. 13, pp. 94-96.

SHOR, N. Z., 1985, "Minimization Methods for non-differentiable functions", Springer, Verlag, Berlin.

SOKOV, N. A., 1974, "Note on minimization methods employing space stretching", *Cybernetics* n. 10, pp. 689-692.

SONNEVEND, G., 1985, "An 'analytical centre' for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming, *Lecture Notes in Control and Information Sciences* 84 (Springer Verlag, New York) pp. 866-876.

VAIDYA, P. M., 1990, "An Algorithm for Linear Programming which requires $O((m + n)n^2 + (m+n)^{1.5} n)L$ arithmetic operations", *Mathematical Programming* n. 47, pp. 175-201.

YE, Y., 1989, "Eliminating columns and rows in potential reduction and path following algorithms for linear programming", *Working Paper 89-7*, Department of Management Sciences, University of Iowa, USA.

YE, Y., 1991, "An $O(n^3)L$ potential reduction algorithm for linear programming", *Mathematical Programming*, n. 50, pp. 239-258.

YE, Y., 1992A potential reduction algorithm allowing column generation", *SIAM Journal of Optimization*, n. 2, pp. 7-29.

ZHANG, Y., 1996, Solving large-scale linear programs by interior-point methods under Matlab environment, *Technical Report TR96-01*, February, 1996, University of Maryland, Department of Mathematics and Statistics, Baltimore County - USA,.