

IDEALIZAÇÃO DE PARÂMETROS
E
UM ALGORITMO PARAMETRIZADO DE BARREIRA LOGARÍTMICA

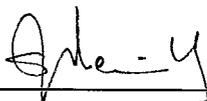
Bryon Richard Hall

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

Aprovado por



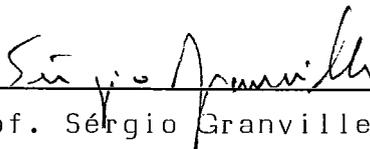
Prof. Paulo Roberto Oliveira, D.Sc.
(President)



Prof^a Susana Scheimberg de Makler, D.Sc.



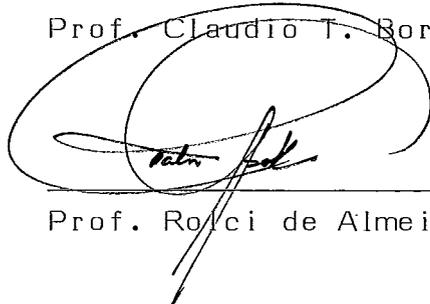
Prof. Clóvis C. Gonzaga, D.Sc.



Prof. Sérgio Granville, Ph.D.



Prof. Claudio T. Bornstein, Dr.Rer.Nat.



Prof. Ruy de Almeida Cipolatti, Ph.D.

HALL, BRYON RICHARD

Idealização de Parâmetros e um Algoritmo Parametrizado de Barreira Logarítmica [Rio de Janeiro] 1992.

p. 29.7 cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação, 1992)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Otimização Não-linear 2. Algoritmos de
Pontos Interiores 3. Método Quadrático
Seqüencial

I. COPPE/UFRJ II. Título (série)

DEDICATÓRIA

For my mother, Elsie A. Hall and in memory of my father, Elvin A Hall.

Agradeço áquele que possibilitou e inspirou a realização deste trabalho, meu orientador de tese Dr. Paulo Roberto Oliveira, e a todos aqueles que me proporcionaram apoio e cooperação. Entre outros menciono Marco Antônio Menezes, Clovis Gonzaga, Manoel Prado pelo apoio espiritual, meu filho Climbiê por momentos de alegria e diversão tão necessários e a minha esposa Marilena Silva.

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

IDEALIZAÇÃO DE PARÂMETROS
E
UM ALGORITMO PARAMETRIZADO DE BARREIRA LOGARÍTMICA

Bryon Richard Hall

JUNHO, 1992

Orientador: Prof. Paulo Roberto Oliveira

Programa: Engenharia de Sistemas e Computação

Dois métodos clássicos de otimização, o método da barreira logarítmica e o método quadrático seqüencial, são combinados no Algoritmo Parametrizado de Barreira Logarítmica com auxílio do conceito de autoconcordância. O algoritmo combina idealmente convergência global e convergência superlinear para problemas $(P) \min \{f(x) : g_i(x) \leq 0\}$, f, g_i convexas, de classe C^3 e (P) não-degenerado, se o conjunto ativo da solução é conhecida. O conceito de idealização de parâmetros é apresentado, onde um parâmetro que normalmente se comporta "passivamente" em algoritmos distintos é utilizado como um parâmetro de transição entre fases de um algoritmo híbrido. Os conceitos se estendem a variações do método seqüencial quadrático.

Algumas variações do algoritmo básico são implementadas e testadas em três problemas-teste.

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

IDEALIZATION OF PARAMETERS
AND
A PARAMETERIZED LOGARITHMIC BARRIER ALGORITHM

Bryon Richard Hall

JUNE, 1992

Thesis Supervisor: Paulo Roberto Oliveira

Department: System and Computational Engineering

Two classical optimization methods, the logarithmic barrier method and the sequential quadratic method, are combined in the Parametrized Logarithmic Barrier Algorithm with the aid of the concept of self-concordance. The algorithm combines ideally global convergence and superlinear convergence for problems $(P) \min \{f(x) \text{ s.t. } g_i(x) \leq 0\}$ with f, g_i convex, C^3 -class functions and (P) non-degenerate, if the active solution set is known. The concept of idealization of parameters is presented so that a normally "passive" parameter that acts as a limiting factor in real problems is used to optimize its use and guide the transition between phases in a hybrid algorithm. The concepts extend to variations of the sequential quadratic method.

A few variations of the basic algorithm are implemented and tested on three test problems.

INDICE

0. INTRODUÇÃO
 1. FUNÇÕES DE PENALIDADE E BARREIRA
 2. AUTOCONCORDÂNCIA
 3. MÉTODOS QUADRÁTICOS SEQÜENCIAIS
 4. R-IDEAL
 5. O ALGORITMO PARAMETRIZADO DE BARREIRA LOGARÍTMICA
 6. TESTES COMPUTACIONAIS
 7. CONCLUSÕES
- REFERÊNCIAS

Nos últimos anos, novos conceitos de otimização têm despertado muito interesse, primeiro pela possibilidade de formular algoritmos de complexidade polinomial e, segundo, pelo seu valor real enquanto bases de métodos novos e eficazes. Depois de Karmarkar [16], muitos novos resultados surgiram daquilo que se chama de "métodos de pontos interiores". Primeiro Gonzaga [10], Renegar [29] e outros reformularam e estenderam a proposta de Karmarkar em programação linear. Numa segunda fase, tentativas de generalizar os conceitos para programação convexa foram feitas por Kortanek Potra e Ye [17], Monteiro e Adler [22], Mehrotra e Sun [21] e outros.

Os métodos de barreira para problemas

$$(P) \quad \min f(x) \\ \text{s.a.} \quad g_i(x) \leq 0$$

f, g_i convexas, de classe C^2 , despertaram novo interesse pelo fato de alguns algoritmos de pontos interiores fazerem uso de funções-objetivo essencialmente idênticas às de métodos de barreira. Entre estes métodos, examinamos principalmente a barreira logarítmica. Esta reúne propriedades interessantes, como o trabalho de Gonzaga [11], Lootsma [19] e Den Hertog, Roos e Terlacky [4] vêm a confirmar. No contexto tradicional os métodos demonstraram problemas de convergência devido ao fato da matriz hessiana se tornar progressivamente pior condicionada quando a seqüência de iterações $\{x^k\}$ converge para uma solução x^* .

A barreira logarítmica foi restabelecida formalmente enquanto instrumento teórico por Nesterov e Nemirovsky [26]. A barreira logarítmica exhibe uma propriedade chamada

de autoconcordância pelos autores e por isto certos algoritmos fazendo uso da barreira logarítmica e a trajetória central (vide Jahrre [15]) para problemas convexos têm ϵ -convergência garantida em $O(p(L)\ln(\frac{1}{\epsilon}))$, onde $p(L)$ é um polinômio em L , o tamanho do problema em aritmética de ponto flutuante. O método ainda mantém a vantagem de ter convergência global garantida em condições gerais (Vide Lootsma, [19]).

Apesar de ressurgimento do método de barreira logarítmica, os métodos mais usados hoje continuam sendo os lagrangeanos, em particular os métodos quadráticos seqüenciais. (MQS) Pesquisadores como Han [13] e outros têm obtido bons resultados de convergência local e garantias de convergência global. Entretanto continua presente a potencialmente lenta convergência quando suficientemente afastado de x^* .

Frente à dificuldade de encontrar um único método que reúne todas as vantagens desejadas sem incorporar qualquer defeito, uma opção é de formular algoritmos híbridos nos quais vários procedimentos alternativos são disponíveis, sendo acionado aquele que for mais apropriado nas condições dadas. O tipo que visamos é de barreira logarítmica enquanto as iterações $\{x^k\}$ estiverem distante da região de convergência superlinear exibido pelo MQS. Ao entrar nesta região o MQS passa a ser executado. A dificuldade está em como monitorar a transição de uma fase à outra. Acreditamos que o conceito de autoconcordância pode servir para tal finalidade. Seu uso implica que consideraremos problemas (P) limitados a f, g_i convexas e de classe C^3 tais que a família $\{F_r(x) : r>0\}$ seja autoconcordante. Este conceito é discutido no capítulo 2. Antes disso resumimos diversos resultados que serão necessários no desenvolvimento posterior.

No capítulo 3 reabordamos os métodos quadráticos seqüenciais e uma combinação destes com a barreira logarítmica e no capítulo 4 a noção de idealização de parâmetros é desenvolvida. Em essência, ao parâmetro r da família de funções

$$F_r(x) := \frac{1}{r} f(x) - \sum_i \ln(-g_i(x))$$

é atribuído um valor que facilita a convergência das iterações $\{x^k\}$.

O algoritmo apresentado no capítulo 5 não pretende ser definitivo, mas ilustrativo do tipo de desenvolvimento possível com o uso "ativo" de um parâmetro no processo decisório de efetivar ou não uma dada iteração. No capítulo 6 serão relatados alguns exemplos numéricos, mais a título de exemplificação de que de comprovação prática.

0.1 Notação

Não existe uma notação inteiramente uniforme neste campo de estudos, sendo talvez a maior dificuldade tratar dos inúmeros índices. Procuramos usar a notação mais convencional e ainda ressaltamos os seguintes pormenores.

Índices serão padronizados: superíndices indicam vetores e ainda são empregados para as matrizes hessianas $\nabla^2 f$ e $\nabla^2 g_i$; sub-índices indicam escalares e outras matrizes. Assim, o gradiente da função $f: \mathbb{R}^n \Rightarrow \mathbb{R}$ nos pontos x , x^k , x^* e $x(r)^*$ é representado, respectivamente, por

$$\nabla f(x) \quad \nabla f^k \quad \nabla f^* \quad \nabla f(r)^*$$

Esta prática é usada também com outras funções, mas sempre que for necessário a expressão será escrita por extenso.

O multiplicador de Lagrange na Lagrangeana

$$L(x, \mu) = f(x) + \sum_i \mu_i g_i(x)$$

quando avaliado no ponto x^k é representado por μ^k . Escrevemos $L_{\mu}(x)$ para um representante da família $L(x, \mu)$, μ fixo

A aplicação $g: \mathbb{R}^n \Rightarrow \mathbb{R}^m$ representa m restrições $g_i: \mathbb{R}^n \Rightarrow \mathbb{R}$. A matriz jacobiana avaliada no ponto x^k é escrita ∇g_k .

Se atribuímos a métrica euclidiana ao espaço \mathbb{R}^n , este será indicado por E^n . Para um conjunto S , o fecho

de S é escrito \bar{S} e sua fronteira indicada por ∂S .

Usamos com frequência as funções de barreira

$$F_r(\mathbf{x}) = \frac{1}{r} f(\mathbf{x}) + F(\mathbf{x})$$

$$B_r(\mathbf{x}) = f(\mathbf{x}) + rF(\mathbf{x})$$

onde

$$F(\mathbf{x}) = - \sum_i \ln(-g_i(\mathbf{x}))$$

CAPÍTULO 1 FUNÇÕES DE PENALIDADE E BARREIRA

Os métodos de penalidade contam-se entre os primeiros usados na resolução de problemas não-lineares de minimização. Toda a classe de métodos consiste da definição de uma nova função $\Phi(x;\rho)$ a ser minimizada sem restrições. O parâmetro ρ possui um certo valor crítico $\bar{\rho}$ no qual Φ pode ser indefinido mas tal que

$$(1.1) \quad \lim_{\rho \Rightarrow \bar{\rho}} \underset{x}{\operatorname{argmin}} \Phi(x;\rho) = x^* \in S^*$$

sendo S^* conjunto-solução do problema inicial.

No caso de problemas não-lineares com condições de igualdade (PNLI), o método mais tradicional provém da função-objetivo

$$\Phi(x;\rho) = f(x) + \rho \sum_i |g_i(x)|^2$$

Pode ser provado que, para um valor de ρ suficientemente grande, já que $\bar{\rho}$ neste caso é igual a $+\infty$, o ponto minimal da função Φ é arbitrariamente próximo de x^* . Resultados análogos são válidos para outros métodos de penalidade.

As dificuldades inerentes ao uso deste método são bem documentadas (eg. Fletcher [6]) e se devem principalmente ao mau condicionamento da matriz hessiana de $\Phi(x;\rho)$ para valores de ρ próximos de $\bar{\rho}$. Este fato não invalida por inteiro os métodos mas motivou certo indeferimento dos mesmos por parte de usuários pragmáticos. Contudo, acreditamos que ainda há muito a ser aproveitado dos conceitos germinais.

diato de contornar em parte o mau condicionamento é por seqüenciação. ρ é gradativamente aproximado a $\bar{\rho}$ e $\Phi(x;\rho)$ é minimizado iterativamente. Denotamos por $x(\rho)^*$ o mínimo local de $\Phi(x;\rho)$ numa vizinhança de x^* . O algoritmo modelo é

- (0) Inicializar com $x^0 = x(\rho^0)^*$. $k = 1$
- (1) Conferir condições de convergência adequada .
- (2) Escolher ρ^k no intervalo aberto contido entre ρ^{k-1} e $\bar{\rho}$. Minimizar $\Phi(x;\rho^k)$ a partir de x^{k-1} .
- (3) Denotar a solução por x^k . $k:=k+1$. Voltar para (1).

Um problema não-linear com condições de desigualdade (PNLD) pode ser resolvido por outro método semelhante. Considere o problema

$$(P) \quad \begin{array}{ll} \min & f(x) \\ \text{s.a.} & g_i(x) \leq 0 \quad 1 \leq i \leq m \end{array}$$

onde f e cada g_i são funções de classe C^2 e tais que $S = \{x: g_i(x) < 0, \quad 1 \leq i \leq m\}$ seja não-vazio. Em particular, tomaremos f, g_i como funções convexas.

Sejam

$$\begin{aligned} E(x;r) &= \frac{1}{r} f(x) - \sum \frac{1}{g_i(x)} \\ (1.2) \quad F(x;r) &= \frac{1}{r} f(x) - \sum \ln(-g_i(x)) \end{aligned}$$

Dado certo $x^0 := x(r)^* \in S$ e supondo os passos bem definidos, o algoritmo-modelo proposto acima gera uma seqüência de pontos $\{x^k\}$ todos estritamente viáveis, usando por Φ , entre outras, qualquer uma das funções E ou F . Se o conjunto ativo for não-vazio, então para todo $r > 0$ fixo, escolhido um ponto arbitrário $\bar{x} \in \partial(\bar{S})$, tem-se

$$\lim_{x \rightarrow X} \Phi(x;r) = \infty$$

Esta característica motiva a distinção destes métodos de outros de penalidade, sendo eles conhecidos como métodos de barreira. Mais adiante detalharemos outra característica comum a todos, a de ser também progressivamente pior condicionado na medida em que $r \downarrow 0$ e $x(r)^* \rightarrow x^*$. Antes disto, consideramos outras propriedades do caso que usaremos neste trabalho - a barreira logarítmica (1.2).

Em princípio é possível a função logarítmica ser ilimitada por baixo, se S for conjunto ilimitado. Powell [28] apresentou o exemplo

$$\min \quad - \frac{1}{x^2 + 1} \quad \text{s.a.} \quad x \geq 1$$

sendo $F_r(x) := F(x;r)$ ilimitado inferiormente $\forall r > 0$. Repare entretanto que $f(x)$ é não-convexa neste exemplo. Fiacco e McCormick [5] mostram que se $f, g_i, (i = 1, \dots, m)$ são convexas, se o conjunto de soluções de (P) é compacto e se (P) tem a propriedade de complementaridade estrita então $F_r(x)$ terá ponto minimal $x(r)^* \forall r > 0$.

Denotamos por $\nabla g(x)$ a matriz jacobiana $m \times n$ das restrições $g_i(x) \leq 0$. Como é comum na literatura, chamamos de $Z(x)$ uma matriz cujas colunas formam uma base do espaço nulo de $\nabla g(x)$.

Proposição 1.1 Seja dado um problema viável (P), sendo $\nabla g(x^*)$ de posto máximo e toda restrição g_i suposta ativa em x^* . Suponha que existe um vetor μ^* tal que

- i) $\nabla f(x^*) = \nabla g(x^*)^T \mu^*$
- ii) $Z(x^*)^T \nabla^2 L(x^*, \mu^*) Z(x^*)$ é positiva definida
- iii) $\mu^* > 0$

Então existe um intervalo $(0, R)$ e uma trajetória T continuamente diferenciável de pontos $x(r)^*$ tal que

$$\lim_{r \downarrow 0} x(r)^* = x^*$$

Demo. Os detalhes se encontram em Fiacco e McCormick [5].

Em geral nem toda restrição g_i é ativa em x^* , mas esta pequena generalização não altera em nada a proposição: simplesmente destacamos as restrições que são ativas em x^* e a condição iii) se torna a de complementariedade estrita.

Os pontos de minimização irrestrita de $F_r(x)$, que em conjunto chamamos de trajetória central T , possuem propriedades interessantes. Se existir pelo menos uma restrição g_i ativa em x^* , por continuidade da função barreira, existe uma vizinhança relativa V de x^* em S tal que

$$-\sum_i \ln(-g_i(x)) > 0 \quad \forall x \in V$$

Isto é verdade mesmo quando $\{g_i\}$ contém restrições não-ativas em x^* . A proposição a seguir dependerá de existir $\bar{r} > 0$ tal que $x(r)^* \in V$ se $\bar{r} > r > 0$, o que é garantido pela proposição 1, mas que é freqüentemente satisfeito mesmo existindo certa g_i não-ativa em x^* .

Por definição de $x(r)^*$ tem-se $\forall r_1, r_2 > 0$

$$(1.3) \quad \frac{1}{r_1} f(x(r_1)^*) - \sum_i \ln(-g_i(x(r_1)^*)) \leq \frac{1}{r_2} f(x(r_2)^*) - \sum_i \ln(-g_i(x(r_2)^*))$$

$$(1.4) \quad \frac{1}{r_2} f(x(r_2)^*) - \sum_i \ln(-g_i(x(r_2)^*)) \leq \frac{1}{r_1} f(x(r_1)^*) - \sum_i \ln(-g_i(x(r_1)^*))$$

Somamos e rearranjamos (1.3) e (1.4), restando

$$\left(\frac{1}{r_1} - \frac{1}{r_2} \right) \left(f(x(r_1)^*) - f(x(r_2)^*) \right) \leq 0$$

ou seja

$$\frac{1}{r_1} - \frac{1}{r_2} < 0 \quad \text{implica que} \quad f(x(r_1)^*) \geq f(x(r_2)^*)$$

Este resultado, aplicado a qualquer uma das desigualdades (1.3) e (1.4), implica ainda que

$$-\sum_i \ln(-g_i(x(r_1)^*)) \leq -\sum_i \ln(-g_i(x(r_2)^*))$$

Em resumo, tanto $f(x(r)^*)$ como $\sum_i \ln(-g_i(x(r)^*))$ são funções não-decrescentes de r dentro da vizinhança V .

Repare que se \bar{S} for compacto, então existirá um mínimo absoluto de $-\sum \ln(-g_i(x))$, cujo argumento é chamado de centro analítico. Por outro lado, a substituição de restrições $g_i(x) \leq 0$ por múltiplos $\varepsilon g_i(x) \leq 0$ em nada altera o problema. Mas

$$-\sum_i \ln(-\varepsilon g_i(x)) = -m \cdot \ln \varepsilon - \sum_i \ln(-g_i(x))$$

e portanto é possível encontrar ε suficientemente pequeno de forma que $-\sum_i \ln(-\varepsilon g_i(x)) > 0 \quad \forall x \in S$. Daí decorre a seguinte proposição.

Proposição 1.2 : Seja $\{r_k\}$ seqüência estritamente decrescente e $\lim_{k \rightarrow \infty} r_k = 0$. Então existe uma vizinhança (relativa) V de x^* na qual $\{f(x(r_k)^*)\}$ e $\left\{ \sum_i \ln(-g_i(x(r_k)^*)) \right\}$ são não-decrescentes. Se \bar{S} for compacto, isso se verificará para todo r_k .

Se o conjunto ativo A de restrições $g_i(x)$ é conhecido, tem-se uma caracterização mais forte: cada seqüência $\{g_i(x^k)\}$ é crescente, possivelmente em outra vizinhança relativa V de x^* . Isso revela uma propriedade muito útil da

trajetória central T , que é sua pequena curvatura ou "relativa retilinearidade". dT/dr está contido na interseção de $m+1$ semi-espacos: $\{v \in E^n : v^T \nabla f(r)^* \geq 0\}$ e

$$\left\{ v \in E^n : v^T \nabla g_i(r)^* \leq 0 \right\} \quad \text{para } 1 \leq i \leq m$$

para todo r suficientemente pequeno. Esta propriedade viabilizará o uso de "passos longos" nos algoritmos que desenvolvemos nos capítulos 4 e 5.

O método de barreira logarítmica enfrenta dois problemas computacionais comuns a esta classe de métodos. Em primeiro lugar, é necessário um controle explícito da satisfação das restrições $g_i(x) \leq 0$ pois, com o crescente mau condicionamento do subproblema quando $x^k \rightarrow x^*$, é perfeitamente possível um passo discreto para fora de S tal que $\frac{1}{r} f(x) - \sum \ln(-g_i(x))$ seja avaliado como decrescente.

Relacionado à anterior complicação é o fato de muitos algoritmos clássicos de busca unidimensional somente funcionarem bem na medida em que os termos lineares quadráticos e cúbicos dominam na expressão polinomial da função-objetivo. $F_r(x)$ é singular ao longo da fronteira do conjunto S e pouco tratável por estes métodos.

Por definição tem-se

$$\begin{aligned} \nabla F_r(x(r)^*) &= \frac{1}{r} \nabla f(r)^* - \sum_i \frac{\nabla g_i(r)^*}{g_i(r)^*} = 0 \\ \Rightarrow \nabla f(r)^* - \sum_i \frac{r}{g_i(r)^*} \nabla g_i(r)^* &= 0 \end{aligned}$$

Se g_i não for ativa em x^* , na medida em que r é reduzido para zero, o coeficiente $-\frac{r}{g_i(x)}$ tende para zero; se g_i for ativa em x^* , tende para o coeficiente μ_i^* , caso $\nabla g(x^*)$ seja de posto máximo. Convém definir

$$(1.5) \quad \mu_i(r) = - \frac{r}{g_i(r)^*}$$

Para discutir o mau condicionamento do problema, utilizamos outra função-objetivo:

$$B_r(x) = f(x) - r \sum_i \ln(-g_i)$$

cuja hessiana no ponto $x(r)^*$ é dada por

$$(1.6) \quad \nabla^2 f(r)^* = \sum_i \frac{r}{g_i(r)^*} \nabla^2 g_i(r)^* + \\ + \sum_i \frac{r}{g_i^2(r)^*} \nabla g_i(r)^* (\nabla g_i(r)^*)^T$$

Quando r tende para 0, os primeiros dois termos de (1.6) convergem para

$$(1.7) \quad \nabla^2 f^* = \sum \mu_i^* \nabla^2 g_i^*$$

(1.7) é uma matriz positiva definida e limitada. No terceiro termo de (1.6) tem-se no limite

$$\sum_i \nabla g_i^* \nabla g_i^{T*} \mu_i \frac{1}{g_i}$$

que é uma matriz de posto menor ou igual a m com autovalores $\{v_1, \dots, v_m\}$ que tendem para infinito e autovetores no espaço-imagem da aplicação

$$\nabla g(x)^T: \mathbb{R}^m \rightarrow \mathbb{R}^n$$

No limite $r \downarrow 0$ tem-se

$$\frac{\mu_i^*}{\mu_j^*} = \lim_{r \downarrow 0} \frac{r/g_i}{r/g_j}$$

e portanto os autovalores v_i têm entre si a relação

$$\frac{v_i}{v_j} = \lim_{r \downarrow 0} \frac{r/g_i^2}{r/g_j^2} = \left(\frac{\mu_i^*}{\mu_j^*} \right)^2$$

Entretanto, não usaremos o método barreira-logarítmica aqui, mas um algoritmo que evita este mau condicionamento e cuja hessiana converge para $\nabla^2 L(x^*, \mu^*)$.

Proposição 1.3 : Considere o problema (P) tendo $\{g_i\}$ como conjunto de restrições ativas em x^* . Seja $\nabla g(x^*)$ de posto máximo. Então a trajetória central T não incide em x^* assintoticamente a qualquer hiperfície $\{x \in \mathbb{R}^n: g_i(x) = 0\}, 1 \leq i \leq m$.

Demo. Pela proposição 1, existe uma trajetória central definida num intervalo $(0, \bar{r})$. A expansão de Taylor de sua expressão em torno de x^* é dada por

$$(1.8) \quad x(r)^* = x^* + ry + O(r^2)$$

Em $x = x(r)^* \in T$ tem-se

$$(1.9) \quad r \nabla f(x) - \nabla g(x)^T \left\{ \frac{r^2}{g_i(x)} \right\} = 0$$

No limite $r \downarrow 0$ sabe-se que

$$(1.10) \quad \nabla f^* = \lim_{r \downarrow 0} (\nabla g(r)^*)^T \left\{ \frac{r}{g_i(r)^*} \right\} = - (\nabla g^*)^T \mu^*$$

Derivamos (1.9) em relação a r , obtendo

$$(1.11) \quad \begin{aligned} \nabla f(x) + r \nabla^2 f(x) y - \sum_i H_i \left(\frac{r^2}{g_i(x)} \right) y - 2 \nabla g(x)^T \left\{ \frac{r}{g_i(x)} \right\} \\ + \nabla g(x)^T \left[\text{Diag} \frac{r^2}{g_i^2(x)} \right] \nabla g(x) y = 0 \end{aligned}$$

Substituindo (1.10) em (1.11),

$$\begin{aligned} - (\nabla g^*)^T \mu^* + \lim_{r \downarrow 0} r \nabla^2 f^* y - \\ - \lim_{r \downarrow 0} \sum H_i \left(\frac{r^2}{g_i(r)^*} \right) \cdot y + 2 \nabla g(x^*)^T \mu^* + \\ + \nabla g(x^*)^T \left[\text{Diag} \mu_i^* \right]^T \left[\text{Diag} \mu_i^* \right] \nabla g(x^*) y = 0 \end{aligned}$$

o que implica

$$(\nabla g^*)^T [\text{Diag } \mu_i^*]^T e + (\nabla g^*)^T [\text{Diag } \mu_i^*]^T [\text{Diag } \mu_i^*] \nabla g^* y = 0$$

onde $e = (1, 1, \dots, 1)^T$. Sendo $\nabla g(x^*)$ de posto máximo, decorre que

$$(1.12) \quad - [\text{Diag } \mu_i^*] \nabla g^* y = e$$

De (1.12) segue

$$(1.13) \quad \nabla g^* y = - \left\{ \frac{1}{\mu^*} \right\}$$

Considerando o significado de $\nabla g(x^*) y$ e a hipótese $\mu^* > 0$, tem-se em conclusão a proposição.

■

Supondo conhecido o conjunto de restrições ativas em x^* tem-se

$$(1.14) \quad \nabla f(x(r)^*) = - \sum_i \mu_i(r) \nabla g_i(x(r)^*)$$

O produto escalar dos dois lados de (1.14) com o vetor $x(r)^* - x^*$ e uso da expansão de Taylor de primeira ordem levam a

$$f(x(r)^*) - f(x^*) = - \sum_i \mu_i(r) g_i(x(r)^*) + O(r^2)$$

No caso do problema (P) ser convexo, pode ser mostrado que

$$(1.15) \quad f(x(r)^*) - f(x^*) \leq - \sum_i \mu_i(r) g_i(x(r)^*) \\ \approx mr$$

já que $\mu_i(r) = - \frac{r}{g_i(x(r)^*)}$.

(1.15) pode ser usado como critério de convergência ou fa-

zer parte de um tal critério, conforme o caso.

Vale a pena ressaltar a natureza muito proveitosa da trajetória central na maioria dos casos encontrados. As propriedades da função logarítmica faz com que a trajetória central seja imune a comportamento como de se desdobrar repetidamente. Exemplos de comportamento mais atípico podem ser construídos apenas ao usar funções f, g_1 , que não admitem boas aproximações em expressões polinomiais de ordem pequena, e para valores maiores de r .

2.1 Funções Autoconcordantes

Os pesquisadores da CIE Yu. E. Nesterov e A. S. Nemirovsky enunciaram um conceito que ajuda significativamente na compreensão *in loco* dos algoritmos que se baseiam na função-barreira. Este conceito, autoconcordância de funções e famílias de funções, é essencialmente geométrico em conteúdo, e serve na detecção de diversas regiões em S com suas respectivas garantias de taxa de convergência ao longo da trajetória central. Procedemos para as definições básicas. Para maiores detalhes, consulte [26].

Def. Seja S um subconjunto aberto, convexo e não-vazio de E^n . A função $f:S \Rightarrow R$ é dita autoconcordante sobre S com parâmetro $a > 0$ ($f \in A_a(S, E^n)$) se f é convexa, de classe C^3 e se para todo $x \in S$ e todo $h \in E^n$ tem-se

$$(2.1) \quad |D^3f(x)[h,h,h]| \leq 2 a^{-1/2} (D^2f(x)[h,h])^{3/2}$$

Geometricamente, entendemos autoconcordância como um limite imposto nas terceiras derivadas unidirecionais, visando garantir que a aproximação de f por uma função quadrática seja proveitosa.

Numa vizinhança de um ponto mínimo local degenerado encontramos exemplos de não-autoconcordância: considere $\{x^{2k}\}$ em $S = (-\varepsilon_k, \varepsilon_k)$, $k \in \mathbb{N} - \{1\}$. No contexto de função-barreira, examinemos para $x > 0$

$$f(x) = x^{-k} \quad k \in \mathbb{N}$$

$$g(x) = -\ln x$$

A condição (2.1) se traduz respectivamente em

$$(2.2) \quad -k(k+1)(k+2)x^{-k-3} \leq 2 a^{-1/2} (k(k+1)x^{-k-2})^{3/2}$$

$$(2.3) \quad 2 x^{-3} \leq 2 a^{-1/2} x^{-3}$$

Enquanto (2.3) se satisfaz para todo $x > 0$ com $a = 1$, a desigualdade (2.2) não é válida fora de certo intervalo $(0, b(a,k)]$ para todo $a > 0$ e $k \geq 1$, ié. para todo f convexa.

Quanto maior for o parâmetro a , melhor será a aproximação (global) de f por uma função quadrática. Algumas estimativas do valor de a serão discutidas no capítulo 6.

O conceito de autoconcordância se aplica através da definição de uma função associada.

$$(2.4) \quad \lambda(f, x) = \inf \left\{ \lambda : |\nabla f(x)^T h| \leq \lambda a^{1/2} (h^T \nabla^2 f(x) h)^{1/2} \right. \\ \left. \forall h \in E^n \right\}$$

com $\lambda(f, x) = \infty$ no caso de não existir nenhum valor λ que torne válida a desigualdade.

Suponhamos f função estritamente convexa. Se definirmos a aproximação quadrática de f em torno do ponto x por

$$Q_x(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2 f(x) (y - x)$$

$Q_x(y)$ atinge mínimo num ponto $y = x - \nabla^2 f(x)^{-1} \nabla f(x)$ o que implica que

$$Q_x(x) - \inf \{ Q_x(y) : y \in \mathbb{R}^n \} = \frac{1}{2} \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$$

Definimos o vetor $h = - \nabla^2 f(x)^{-1} \nabla f(x)$ e lembramos que

$$a \lambda^2 \geq \frac{|\nabla f(x)^T v|^2}{v^T \nabla^2 f(x) v} \quad \forall v \in E^n \\ \geq \frac{|\nabla f(x)^T h|^2}{h^T \nabla^2 f(x) h}$$

$$= \frac{|\nabla f^T(x) \nabla^2 f^{-1} \nabla f(x)|^2}{\nabla f(x)^T \nabla^2 f^{-1} \nabla^2 f \nabla^2 f^{-1} \nabla f(x)} = \nabla f(x)^T \nabla^2 f^{-1} \nabla f(x)$$

Considerando o significado do λ usado, temos demonstrado

Proposição 2.1: Dado $x \in S$ onde $f: S \rightarrow \mathbb{R}$ é função λ -autoconcordante, considere a aproximação quadrática $Q_x(y)$ de f . Tem-se

$$\frac{\lambda}{2} \geq Q_x(x) - \inf \{ Q_x(y) : y \in \mathbb{R}^n \}$$

ou seja, $\lambda^2/2$ é supremo do decréscimo esperado no valor de f por um passo no algoritmo de Newton.

$\lambda(f, x) \downarrow 0$ quando $x \rightarrow x^*$ e seu valor possui um significado absoluto se supusermos λ determinado por

$\lambda = \inf_x \{ \lambda(x) : x \in S \}$ para certa função λ com

$$\lambda(x) = \sup \{ \alpha > 0 : \alpha \text{ satisfaz (2.1) em } x \in S \text{ no papel de } \lambda \}$$

Repare que, ao tratar de funções-barreira $F_r(x)$, r fixo, sempre se tem $\lambda(F_r, x) < \infty$. Nemirovsky e Nesterov mostram que isto implica na existência de um ponto $x(r)^*$ onde F_r tem valor mínimo e $\nabla F_r(x(r)^*) = 0$. Resumimos a seguir os resultados principais sobre funções autoconcordantes em S .

1) Se $\lambda(f, x) < \infty$ para algum $x \in S$, então $\lambda(f, x)$ é finita e contínua em todo S .

2) Seja $\{y \in S : f(y) \leq f(x)\}$ fechado em \mathbb{R}^n para certo $x \in S$ fixo e seja $\lambda(f, x) < \infty$

2.1) se $\lambda(f, x) < 1$ então f tem ponto de mínimo no conjunto S .

2.2) existe $\lambda' \in [\lambda_*, 1)$, onde $\lambda_* = 2 - \sqrt{3} = 0,2679\dots$ com as seguintes características. Supo-

nhamos que a partir do ponto x seja aplicado o algoritmo de Newton na busca de $\min f(x)$, definido por

$$(2.5) \quad x^{i+1} = x^i + c(\lambda(f, x^i)) p^i$$

onde p^i é passo de Newton, $x^0 = x$ e c é o comprimento relativo do passo definido por

$$(2.6) \quad c(\lambda) = \begin{cases} (1 + \lambda)^{-1} & \text{se } \lambda > \lambda' \\ (1 - \lambda)\lambda^{-1}(3 - \lambda)^{-1} & \text{" } \lambda' \geq \lambda \geq \lambda_* \\ 1 & \text{" } \lambda < \lambda_* \end{cases}$$

Estes passos são bem-definidos (i.é. $x^{i+1} \in S$). Designando por λ_i o valor de $\lambda(f, x^i)$, tem-se respectivamente nos três casos de (2.6) as seguintes relações:

$$(2.7) \quad \begin{aligned} f(x^{i+1}) &\leq f(x^i) - a \{ \lambda_i - \ln(1 + \lambda_i) \} \\ &\leq f(x^i) - a \{ \lambda' - \ln(1 + \lambda') \} \end{aligned}$$

$$(2.8) \quad \begin{cases} \lambda_{i+1} \leq (6\lambda_i - \lambda_i^2 - 1)/4 < \lambda_i \\ 1 - \lambda_{i+1} \geq (1 - \lambda_i)(5 - \lambda_i)/4 \geq \\ \geq (1 - \lambda_i)(5 - \lambda')/4 \end{cases}$$

$$(2.9) \quad \lambda_{i+1} \leq \lambda_i^2 (1 - \lambda_i)^{-2} < \frac{1}{2} \lambda_i$$

Repare que no primeiro caso ($\lambda > \lambda'$), garante-se apenas a redução de \underline{f} , enquanto no terceiro caso ($\lambda < \lambda_*$), tem-se assegurada uma forte redução do valor λ_i . Em interpretação mais sintética, verifica-se o uso de $\lambda(f, x)$ enquanto parâmetro que mede o possível erro num passo puro de Newton, erro este induzido pela derivada terceira $D^3f(x)$. A novidade está no uso apenas das derivadas terceiras unidirecionais enquanto expressão de cotas superiores do erro em questão. Assim, o algoritmo de Newton se divide em três partes:

1) $(\lambda > \lambda')$ o número de iterações nesta fase é menor ou igual a

$$\frac{f(x) - \min_S f(x)}{a (\lambda' - \ln(1 + \lambda'))} + 1$$

2) $(\lambda' \geq \lambda \geq \lambda_*)$ nesta fase intermediária, λ_i decresce de forma progressivamente mais acelerada. O número de iterações é menor ou igual a

$$\frac{\ln(1 - \lambda_*) - \ln(1 - \lambda')}{\ln((5 - \lambda')/4)} + 1$$

3) $(\lambda < \lambda_*)$ a fase final é de convergência quadrática. Repare que nesta fase não existe dependência de λ' .

2.2 Famílias Autoconcordantes

Ao se tratar de uma família de funções $\{F_r: S \Rightarrow \mathbb{R}, r > 0\}$, precisamos levar em conta a variação dos parâmetros envolvidos na definição de autoconcordância. Nesterov e Nemirovsky definem-na por

$$\mathfrak{F} = \{S_r, F_r, E^n\}_{r \in (0, \bar{r})}$$

sendo cada conjunto S_r aberto, convexo e não-vazio e cada função F_r convexa e de classe C^3 . A definição da condição de autoconcordância exige a associação a \mathfrak{F} de um conjunto de seis parâmetros $\alpha(r), \gamma(r), \mu(r), \xi(r), \eta(r)$ e K . Cada função F_r é $\alpha(r)$ -autoconcordante. Uma métrica ρ_K do conjunto $(0, \bar{r})$ é definida, onde $K \in (0, \lambda_*)$. O resultado principal é o seguinte.

Teorema. Seja $\mathfrak{F} = \{S_r, F_r, E^n\}$ uma família autoconcordante. Suponha que para $(r, x) \in (0, \bar{r}) \times S_r$ tem-se

$$\lambda(F_r, x) < K$$

e que

$$\rho_K(r, r') \leq K^{-1}(K - \lambda(F_r, x))$$

para certo $r' \in (0, \bar{r})$. Então podemos concluir que

$$(r', x) \in (0, \bar{r}) \times S_r,$$

e que

$$\lambda(F_{r'}, x) \leq K.$$

2.3 Função-Barreira Logarítmica

Nemirovsky e Nesterov desenvolvem o conceito de autoconcordância visando o enquadramento de algoritmos "clássicos" de barreira, com potencial de aceleração. Resumimos brevemente esta abordagem.

Uma função f se diz fortemente autoconcordante em S ($f \in A_a^+(S, E^n)$) se f é a -autoconcordante em S e se todo conjunto $\{x \in S : f(x) \leq t\}$ é fechado.

Se $F \in A_1^+(S, E^n)$ é uma função definida sobre S tal que $\sup\{\lambda(F, x) : x \in S\} =: \nu^{1/2}(F)$ então F é dita uma barreira ν -autoconcordante ($F \in \mathcal{B}(\bar{S}, \nu)$). Dizemos da função $f: X \ni \bar{S} \Rightarrow \mathbb{R}$ que ela é β -compatível com $F \in \mathcal{B}(\bar{S}, \nu)$ se f é convexa, de classe C^3 , inferiormente semi-contínua no conjunto S e se

(2.10)

$$| D^3 f(x)[h, h, h] | \leq \beta (3D^2 f(x)[h, h]) \cdot (3D^2 F(x)[h, h])^{1/2}$$

Notacionalmente escrevemos $f \in \mathcal{A}(F, \beta)$. A seguinte proposição detalha quatro propriedades importantes de funções β -compatíveis com barreiras ν -autoconcordantes.

Proposição 2.2 Sejam $f \in \mathfrak{A}(F, \beta)$, $F \in \mathfrak{B}(\bar{S}, \nu)$

i) Seja $T: E \rightarrow \mathbb{R}^n$ uma transformação afim de forma que $T(E) \cap S \neq \emptyset$. Definimos

$$\begin{aligned} S^+ &= T^{-1}(S) \\ F^+(y) &= F(Ty): S^+ \rightarrow \mathbb{R} \\ f^+(y) &= f(Ty): S^+ \rightarrow \mathbb{R} \end{aligned}$$

Então $F^+ \in \mathfrak{B}(\bar{S}^+, \nu)$ e $f^+ \in \mathfrak{A}(F^+, \beta)$, isso é, ν e β são preservados.

ii) Sejam $f_i \in \mathfrak{A}(F, \beta_i)$, $\rho_i \geq 0$, $i = 1, 2$. Então $\rho_1 f_1 + \rho_2 f_2 \in \mathfrak{A}(F, \max(\beta_1, \beta_2))$

iii) Sejam S_i conjuntos abertos não-vazios em \mathbb{R}^n , $1 \leq i \leq m$, tais que

$$S^+ = \bigcap_{i=1}^m S_i \neq \emptyset.$$

Sejam $F_i \in \mathfrak{B}(S_i, \nu_i)$, $1 \leq i \leq m$ e $F^+ = \sum_{i=1}^m F_i: S^+ \rightarrow \mathbb{R}$

Então $F^+ \in \mathfrak{B}(S, \sum_{i=1}^m \nu_i)$ e $\mathfrak{A}(F_i, \beta) \subset \mathfrak{A}(F^+, \beta) \quad \forall i$.

iv) Seja $x \in S$, $h \in E^n$ e $q_x(h) = \sup \{t : x \pm th \in S\}$

$$\begin{aligned} \text{Então } (h^T \nabla F(x) h)^{-1/2} &\leq q_x(h) \leq \\ &\leq (1 + 3\nu) (h^T \nabla F(x) h)^{-1/2}. \end{aligned}$$

Repare que a chamada condição (relativa) de Lipschitz se assemelha à β -compatibilidade. Den Herzog, Roos e Terla-ky [4] definem esta condição pela existência de $L > 0$ tal que

$$|D^3 f_j(h, h, p)| \leq L(p^T H p)^{1/2} D^2 f_j(h, h)$$

para todo $h, p \in \mathbb{R}^n$ e todo i , $0 \leq j \leq m$. f_j corresponde a g_j , $1 \leq j \leq m$, e f_0 a f , na nossa formulação. H é a matriz hessiana de $\hat{F}_r(x)$, onde

$$\hat{F}_r(x) := -\frac{1}{r} b^T x - \sum_{i=1}^m \ln(-g_i(x))$$

com a função-objetivo, se não-linear, incorporada ao conjunto $\{g_i(x)\}$.

Nesterov e Nemirovsky enquadram uma versão do método clássico de barreira logarítmica na linguagem de autoconcordância. Com isso, eles mostram que o algoritmo toma lugar em até três etapas. Na primeira, preliminar, uma aproximação x^0 do centro analítico $x(F)$ é obtida, sob a condição

$$\lambda(F, x^0) \leq \lambda_2$$

onde λ_2 é uma dada constante menor que λ_* e que satisfaz outras condições relacionadas à função-barreira F .

A seguir, uma etapa de minimização da função-objetivo f é iniciada. Para tal, a função F_r é minimizada aproximadamente por uma seqüência de valores de $r \downarrow 0$. No caso particular de $f \in \mathcal{A}(F, 0)$ — o que significa que f é uma forma quadrática — são obtidos resultados sobre o número de iterações necessário para obter uma ϵ -aproximação de x^* .

Prosseguimos com resultados mais próximos ao uso dos conceitos de barreiras ν -autoconcordantes nos capítulos 4 e 5.

Proposição 2.3 Sejam convexas e de classe C^3 as funções $g_i: S \Rightarrow \mathbb{R}$, $1 \leq i \leq m$, onde $S = \{x \in \mathbb{R}^n : g_i(x) < 0 \ \forall i\}$ e sejam fortemente autoconcordantes sobre S , com parâmetros $a_i = 1$, as funções $F_i(x) = -\ln(-g_i(x))$, $1 \leq i \leq m$.

Então

$$F(x) \doteq - \sum_i^m \ln(-g_i(x))$$

é barreira ν -autoconcordante com $0 \leq \nu \leq m$.

Demo. Considere a função $F_i(x)$. Tem-se

$$\lambda(F_i, x) = \max_{h \in S^{n-1}} \frac{\nabla g_i^T h}{[(\nabla g_i^T h)^2 - g_i^T \nabla^2 g_i h]^{1/2}}$$

(todas as funções avaliadas em $x \in S$). Já que $g_i(x) < 0$ e $\nabla^2 g_i(x)$ é positiva semi-definida, decorre que $F_i: S \Rightarrow \mathbb{R}$ é barreira ν -autoconcordante com $0 \leq \nu \leq 1$.

A proposição 2.2, (iii), completa a demonstração. ■

2.4 λ -Razão e Regiões de Confiança

Os resultados de Nemirovsky e Nesterov sugerem o uso da família $\{ \lambda(F_r, x) : S \Rightarrow \mathbb{R}^+ \}_r$ como meio de definição de uma "região de confiança". Considere um dado par (f, F) , $f \in \mathcal{A}(F, \beta)$, $F \in \mathcal{B}(\bar{S}, \nu)$. $T \subset \bar{S}$ é a trajetória central e

$$C = \left\{ (x, r) : \lambda(F_r, x) < \lambda_* \right\}$$

Uma vez que seja dado o conjunto C , os valores $a(x)$ e $\lambda(F_r, x)$, $x \notin C$, não desempenham mais nenhum papel no decorrer dos algoritmos que desenvolvemos mais adiante. Isso contribui com certa simplificação; assim dividiremos o trabalho em duas etapas. Na primeira, partindo de $y \in S$ arbitrário, encontramos um ponto x^0 próximo de T . Na segunda, todas as iterações (x^k, r) são contidas em C .

Pela definição (2.4), dado $x \in S$, $r \in (0, \bar{r})$, existe $h \in E^n$, h unitário, tal que

$$(2.11) \quad \lambda(F_r, x) = \frac{|\nabla F_r(x)^T h|}{(a h^T \nabla^2 F_r(x) h)^{1/2}} \leq \frac{\|\nabla F_r(x)\|}{(a h^T \nabla^2 F_r(x) h)^{1/2}}$$

Seja v_1 o menor autovalor da hessiana $\nabla^2 F_r(x(r)^*)$ e v_1 um autovetor unitário correspondente. A autoconcordância de $F_r(x)$ limita a variação dos autovalores de $\nabla^2 F_r(x)$ numa vizinhança $N_1(\varepsilon)$ de $x(r)^*$, de forma que escrevemos

$$(2.12) \quad \lambda(F_r, x) \leq \frac{\|\nabla F_r(x)\|}{(a v_1^T \nabla^2 F_r(x(r)^*) v_1)^{1/2}} + \varepsilon \approx \frac{\|\nabla F_r(x)\|}{(a v_1)^{1/2}}$$

para $x \in N_1(\varepsilon)$.

Já que

$\nabla F_r(x) = \nabla^2 F_r(x(r)^*)(x - x(r)^*) + O(\|x - x(r)^*\|^2)$
 tem-se $\|\nabla F_r(x)\| = k(r)\|x - x(r)^*\| + \delta$ numa vizinhança $N_2(\delta)$ de $x(r)^*$. Definimos

$$(2.13) \quad R(r) := (a v_1)^{1/2}$$

Chamamos esta razão $R(r)$ da λ -razão em r do par (f, F) . Por definição, $R(r) > 0$ e mede a relação entre $\lambda(F_r, x)$ e $\|\nabla F_r(x)\|$ para $x \in N_1(\varepsilon) \cap N_2(\delta)$

Em $x = x(r)^*$ tem-se

$$(2.14) \quad \frac{1}{r} \nabla f(x) = \sum_i \frac{\nabla g_i(x)}{g_i(x)}$$

Partindo de

$$(2.15) \quad r \cdot \nabla^2 F_r(x) = \nabla^2 f(x) + \sum_i r \left(\frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i^2(x)} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right)$$

já que $\lim_{r \downarrow 0} -\frac{r}{g_i} = \mu_i^*$, tem-se para r pequeno

$$r \nabla^2 F_r(\mathbf{x}) \approx \nabla^2 f(\mathbf{x}) + \sum_i \left(r \frac{\nabla g_i(\mathbf{x}) \nabla g_i(\mathbf{x})^T}{g_i^2(\mathbf{x})} + \mu_i^* \nabla^2 g_i(\mathbf{x}) \right)$$

(2.14) implica que

$$(2.16) \quad \frac{1}{r^2} \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T = \sum_i \frac{\nabla g_i(\mathbf{x}) \nabla g_i(\mathbf{x})^T}{g_i^2(\mathbf{x})} + \sum_{\substack{i,j \\ i \neq j}} \frac{\nabla g_i(\mathbf{x}) \nabla g_j(\mathbf{x})^T}{g_i(\mathbf{x}) g_j(\mathbf{x})}$$

de onde decorre, para r pequeno

$$\begin{aligned} r^2 \nabla^2 F_r(\mathbf{x}) &\approx r \nabla^2 f(\mathbf{x}) + r \sum_i \mu_i^* \nabla^2 g_i(\mathbf{x}) + \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T - \\ &\quad - r^2 \sum_{\substack{i,j \\ i \neq j}} \frac{\nabla g_i(\mathbf{x}) \nabla g_j(\mathbf{x})^T}{g_i(\mathbf{x}) g_j(\mathbf{x})} \\ &\approx \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T - r^2 \sum_{\substack{i,j \\ i \neq j}} \frac{\nabla g_i(\mathbf{x}) \nabla g_j(\mathbf{x})^T}{g_i(\mathbf{x}) g_j(\mathbf{x})} \end{aligned}$$

ou seja

$$(2.17) \quad \nabla^2 F_r(\mathbf{x}) \approx \frac{1}{r^2} \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T - \frac{1}{r^2} \sum_{\substack{i,j \\ i \neq j}} \mu_i^* \mu_j^* \nabla g_i(\mathbf{x}) \nabla g_j(\mathbf{x})^T$$

e

$$\begin{aligned} (a h^T \nabla^2 F_r(\mathbf{x}) h)^{1/2} &\approx \\ &\approx \frac{a^{1/2}}{r} \left[h^T \left(\nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T - \sum_{\substack{i,j \\ i \neq j}} \mu_i^* \mu_j^* \nabla g_i \nabla g_j^T \right) h \right]^{1/2} \end{aligned}$$

Posto que o conteúdo dos parênteses do termo à direita possui limite finito quando $\mathbf{x} = \mathbf{x}(r)^* \Rightarrow \mathbf{x}^*$, e lembrando que $\mathbf{x}(r)^* - \mathbf{x}^*$ varia linearmente com r numa

vizinhança de x^* , resulta a seguinte proposição.

Proposição 2.4 Dado o par (f, F) , $f \in \mathcal{U}(F, \beta)$, $F \in \mathcal{B}(S, \nu)$, com $a = \inf\{a(x) : x \in C \subset S\} > 0$, a λ -razão $R(r)$ é diretamente proporcional a r^{-1} para valores de r suficientemente pequenos.

Examinemos o comportamento de $\lambda(F_r, x)$ para um certo ponto fixo $x(\bar{r})^*$ e variação de r . Para $r' := \alpha\bar{r}$ suficientemente próximo de \bar{r}

$$\frac{\| \nabla F_{r'}(x(\bar{r})^*) \|}{R(\bar{r})} \geq \lambda(F_{r'}, x(\bar{r})^*)$$

Por outro lado

$$\nabla F_r(x(\bar{r})^*) = \frac{1}{r} \nabla f(x(\bar{r})^*) + \nabla F(x(\bar{r})^*) = 0$$

$$\begin{aligned} \nabla F_{r'}(x(\bar{r})^*) &= \frac{1}{r'} \nabla f(x(\bar{r})^*) + \nabla F(x(\bar{r})^*) \\ &= \frac{1 - \alpha}{\alpha\bar{r}} \nabla f(x(\bar{r})^*) + \frac{1}{\bar{r}} \nabla f(x(\bar{r})^*) + \nabla F(x(\bar{r})^*) \\ &= \frac{1 - \alpha}{\alpha\bar{r}} \nabla f(x(\bar{r})^*) \end{aligned}$$

\Rightarrow

$$(2.18) \quad \lambda(F_{r'}, x(\bar{r})^*) \leq \frac{1 - \alpha}{\alpha\bar{r}R} \| \nabla f(x(\bar{r})^*) \|$$

Para valores pequenos de r onde $rR \approx k$ (constante)

$$\lambda(F_{r'}, x(r)^*) \leq \frac{1 - \alpha}{\alpha} \frac{1}{k} \| \nabla f(x(r)^*) \|$$

o que mostra que $\lambda(F_{\alpha r}, x(r)^*)$ varia linearmente em $\bar{\alpha} := 1 - \alpha$ numa vizinhança de $x(r)^*$.

Normalmente, α terá valores (iniciais) no intervalo $(0.5, 1)$ na segunda etapa dos algoritmos.

Os resultados expostos permitem que se enuncie um conceito idealizado de algoritmo. Suponhamos dado $x^k \in S$ tal que para dado $r \in (0, \bar{r})$

$$\left\| \frac{1}{r} \nabla f^k - \sum_i \frac{\nabla g_i^k}{g_i^k} \right\| \leq \frac{\lambda_*}{2} R(r) := \frac{\lambda_*}{2} R$$

Da desigualdade triangular tem-se

$$(2.19) \quad \left\| \frac{1}{\alpha r} \nabla f^k - \sum_i \frac{\nabla g_i^k}{g_i^k} \right\| \leq \frac{\lambda_*}{2} R + \frac{1-\alpha}{\alpha r} \|\nabla f^k\|$$

Escolhemos um valor de $\alpha \in (0, 1)$ tal que

$$(2.20) \quad \left\| \frac{1}{\alpha r} \nabla f^k - \sum_i \frac{\nabla g_i^k}{g_i^k} \right\| \leq \lambda_* R$$

implicando em

$$(2.21) \quad \left\| \frac{1}{\alpha r} \nabla f^{k+1} - \sum_i \frac{\nabla g_i^{k+1}}{g_i^{k+1}} \right\| < \frac{\lambda_*}{2} R$$

A desigualdade (2.21) se verifica pela aplicação de (2.9) à desigualdade (2.20).

Assim garante-se a existência de uma seqüência $\{x^k\}$ convergente para x^* e contida em C . Obviamente este resultado não é, em si, de alguma forma novo. A novidade reside, ao contrário, na possibilidade de "monitorar" a seqüência usando o valor do gradiente $\|\nabla F_r(x)\|$. Por outro lado o conjunto C e o monitoramento da seqüência são independentes de como são calculadas as iterações $\{x^k\}$, sendo aplicáveis mesmo num processo algorítmico acelerado, como será discutido em capítulo 4.

3.1 O Método Clássico

Podemos distinguir duas classes de métodos para a resolução de

$$(P) \quad \begin{aligned} &\min f(x) \\ &\text{s.a. } g_i(x) \leq 0 \end{aligned}$$

no caso de f, g_i serem convexas e de classe C^2 . Uma abordagem penaliza o não-cumprimento das restrições; outra trata de procurar um ponto (x^*, μ^*) que satisfaz as condições Karush-Kuhn-Tucker. Os métodos quadráticos seqüenciais pertencem a esta segunda classe. Este tipo de método foi proposto primeiro por Wilson [31] e é conhecido também como método de métrica variável.

Suponhamos ∇g de posto máximo, que todas as restrições g_i sejam ativas na solução x^* e que $\mu^* > 0$. A resolução de (P) equivale encontrar uma raiz $(x^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m$ do sistema

$$(3.1) \quad \begin{aligned} \nabla f(x) + \nabla g(x)^T \mu &= 0 \\ g(x) &= 0 \end{aligned}$$

Dado um ponto (x^k, μ^k) , o método de Newton para a resolução de (3.1) consiste de resolver o sistema

$$(3.2) \quad \nabla^2 L(x^k, \mu^k) \begin{bmatrix} \Delta x^k \\ \Delta \mu^k \end{bmatrix} = -\nabla L(x^k, \mu^k)$$

onde $L(x, \mu) = f(x) + \sum \mu_i g_i(x)$. (3.2) pode ser reescrito como

$$(3.3) \quad \begin{cases} H_k \Delta x^k + \nabla g_k^T \Delta \mu^k = -\nabla f^k - \nabla g_k^T \mu^k \\ \nabla g_k \Delta x^k = -g^k \end{cases}$$

onde H_k é a matriz hessiana da lagrangeana avaliada em (x^k, μ^k) . Já que $\Delta\mu^k = \mu^{k+1} - \mu^k$, (3.3) pode ser escrito da forma

$$(3.4) \quad \begin{cases} \nabla f^k + H_k p^k + \nabla g_k^T \mu^{k+1} = 0 \\ \nabla g_k p^k + g^k = 0 \end{cases}$$

onde Δx^k é substituído pelo mais familiar passo p^k . O sistema (3.4) consiste das condições KKT do problema

$$(3.5) \quad \begin{aligned} \min \quad & p^T \nabla f^k + \frac{1}{2} p^T H_k p \\ \text{s.a.} \quad & \nabla g_k p = -g^k \end{aligned}$$

Ou seja, ao resolver o problema (3.5), obtemos um passo p^k e uma nova aproximação μ^{k+1} do multiplicador μ^* .

A convergência do processo é garantida localmente pelo teorema sobre a convergência do método de Newton. (Vide Bertsekas [2])

Teorema 3.1 Seja dado o sistema (3.1) com f, g de classe C^2 com solução (x^*, μ^*) e a matriz $\nabla^2 L(x^*, \mu^*)$ inversível. Existe $\delta > 0$ tal que se, $(x^0, \mu^0) \in B_\delta(x^*, \mu^*)$, então a seqüência $\{(x^k, \mu^k)\}$ gerada pela resolução reiterada de (3.2) (e portanto de (3.5)) é bem definida, converge para (x^*, μ^*) e satisfaz $(x^k, \mu^k) \in B_\delta(x^*, \mu^*)$ para todo k . A não ser que $(x^k, \mu^k) = (x^*, \mu^*)$ para todo $k \geq k^*$, a seqüência $\{(x^k, \mu^k)\}$ exhibe convergência Q-superlinear. Se também verifica-se que $\nabla L: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ possui constante de Lipshitz $M > 0$ em $B_\delta(x^*, \mu^*)$ e se

$$\| \nabla^2 L(x, \mu)^{-1} \| \leq M < \infty$$

$\forall (x, \mu) \in B_\delta(x^*, \mu^*)$, então a seqüência $\{(x^k, \mu^k)\}$ exhibe convergência Q-superlinear de ordem pelo menos dois.

Na sua forma "clássica" este método pode trazer problemas, pois a região de convergência Q-superlinear pode

ser muito reduzida e a de convergência em si pouco maior. Outra dificuldade é a possível inconsistência da restrição $\nabla g_k p = -g^k$ em certos pontos. Em geral estes problemas têm sido contornados eficientemente pelo uso de funções de mérito (vide Han [13]) Proporemos um método de natureza distinta da destes no capítulo 4.

A grande vantagem do uso dos métodos MQS é a transformação de um problema geral não-linear (função-objetivo e restrições) num problema quadrático com restrições lineares. Observe que grande número de problemas práticos contém algumas restrições lineares e que há grande economia computacional em reduzir todas as restrições a esta forma. Por outro lado a resolução de problemas quadráticos com restrições lineares é relativamente fácil com software empregado hoje.

Resumimos a técnica utilizada na resolução de um subproblema da forma (3.5). Seja Z_k uma matriz $n \times (n-m)$ cujas colunas geram o subespaço ortogonal àquele gerado pelas linhas de ∇g_k . Todo vetor p possui expressão única com componentes destes dois subespaços. Escrevemos a solução de (3.5), p^k , na forma

$$(3.6) \quad p^k = \nabla g_k^T p_y + Z_k p_z$$

Por definição, $\nabla g_k p^k = -g^k$ e portanto

$$\begin{aligned} \nabla g_k \left(\nabla g_k^T p_y + Z_k p_z \right) &= \left(\nabla g_k \nabla g_k^T \right) p_y \\ &= -g^k \end{aligned}$$

A equação $(\nabla g_k \nabla g_k^T) p_y = -g^k$ determina unicamente p_y . Substituímos (3.6) na função-objetivo de (3.5), com o valor de p_y já determinado, obtendo um novo problema, reduzido, de minimização irrestrita:

$$(3.7) \quad \left((\nabla f^k)^T Z_k + p_y^T \nabla g_k H_k Z_k \right) p_z + \frac{1}{2} p_z^T Z_k^T H_k Z_k p_z$$

Igualamos o gradiente da função (3.7) a zero e resolvemos

$$Z_k^T H_k Z_k p_z = - Z_k^T \left(\nabla f^k + H_k \nabla g_k^T p_y \right)$$

O multiplicador μ^{k+1} é computado da relação

$$(3.8) \quad \nabla f^k + H_k p^k = - \nabla g_k^T \mu^{k+1}$$

É comum referir-se a $Z_k^T \nabla f^k$ como o gradiente projetado e a $Z_k^T H_k Z_k$ como a hessiana projetada da função lagrangeana. Obviamente, a resolução do problema de cima depende da matriz $Z_k^T H_k Z_k$ ser positiva definida. Já que Z_k é de posto máximo, isso decorre de H_k ser positiva definida. Como

$$H_k = \nabla^2 f^k + \sum_i \mu_i^k \nabla^2 g_i^k$$

f, g_i são convexas e $\mu^k \geq 0$, neste caso H_k é positiva definida para todo $x^k \in S$.

3.2 MQS-Deslocado

Considere um subproblema da forma

$$(3.9) \quad \begin{aligned} \min \quad & p^T \nabla f^k + \frac{1}{2} p^T H_k p \\ \text{s.a.} \quad & \nabla g_k p = - g^k - \varepsilon^k \quad (\varepsilon^k \geq 0) \end{aligned}$$

(3.9) é uma iteração MQS, a partir de (x^k, μ^k) , na resolução do problema

$$(3.10) \quad \begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & g_i(x) \leq - \varepsilon_i^k \end{aligned}$$

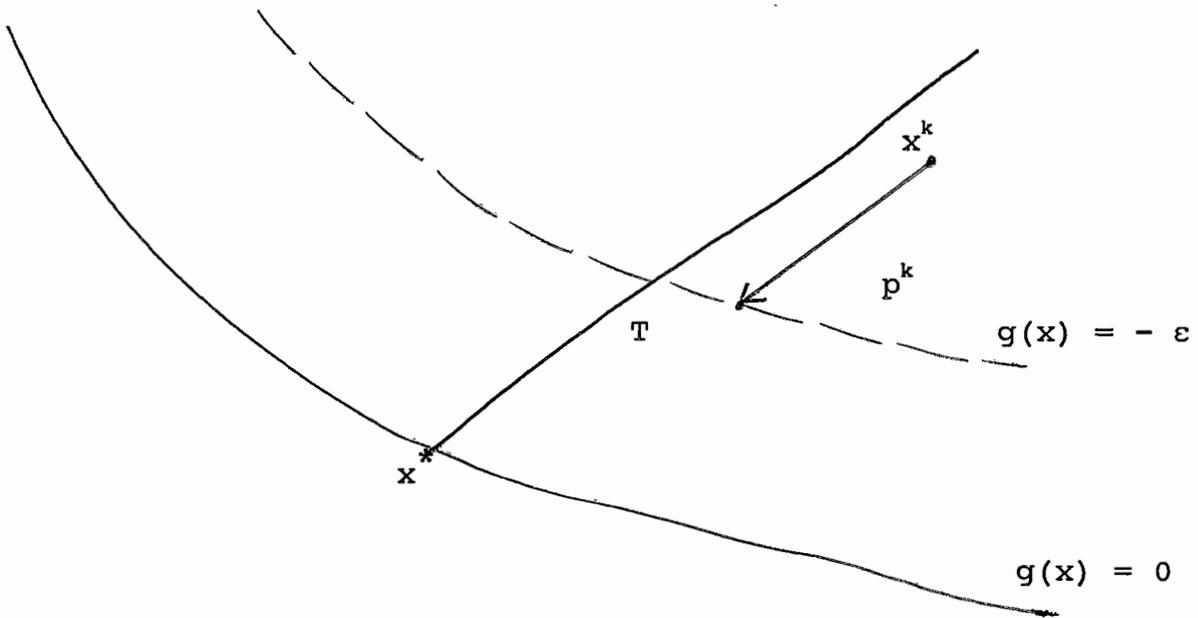


Figura 1

A resolução de (3.9) providencia um passo p^k em direção a x^* caso $-g_k > \varepsilon_k$ (vide Fig. 1). Outro modo de entender o significado da resolução de (3.9) decorre da seguinte proposição.

Proposição 3.1 Seja dado o problema (P) e correspondente trajetória central T. Seja $x^k \in T$, $x^k = \operatorname{argmin} \{ F_r(x) : r = r_k \} = x(r_k)^*$, $r_k > 0$. Seja $\varepsilon^k = -\frac{r'}{r_k} g^k$. Para valores de $r' \in (r_k - \tau, r_k + \tau)$, τ suficientemente pequeno, verifica-se que existe a solução x^{k+1} do problema (3.9) e

$$(3.11) \quad \frac{1}{r'} \nabla f^{k+1} - \sum_i \frac{\nabla g_i^{k+1}}{g_i(x^{k+1})} = O(\tau)$$

Demo. Sabemos que

$$(3.12) \quad \nabla f^{k+1} = \nabla f^k + \nabla^2 f^k p + O(\|p\|^2)$$

$$\nabla g_i^{k+1} = \nabla g_i^k + \nabla^2 g_i(x^k) p + O(\|p\|^2) \quad 1 \leq i \leq m$$

Por outro lado, o problema (3.9) implica que

$$(3.13) \quad g^{k+1} = \frac{r'}{r_k} g^k + O(\|p\|^2)$$

Substituímos (3.12) e (3.13) no termo à esquerda de (3.11), obtendo

$$\begin{aligned} \frac{1}{r'} (\nabla f^k + \nabla^2 f^k p) - \frac{r_k}{r'} \sum_i \frac{(\nabla g_i^k + \nabla^2 g_i(x^k) p)}{g_i(x^k)} &= \\ &= \frac{1}{r'} \left(\nabla f^k - r_k \sum_i \frac{\nabla g_i^k}{g_i(x^k)} \right) + \frac{1}{r'} \left(\nabla^2 f^k - \right. \\ &\quad \left. - r_k \sum_i \frac{\nabla^2 g_i(x^k)}{g_i(x^k)} \right) p \end{aligned}$$

Já que $\nabla f^k - r_k \sum_i \frac{\nabla g_i^k}{g_i(x^k)} = 0$, tem-se que para p suficientemente pequeno, (3.11) se verifica, com o erro diretamente proporcional a $\|p\|$. O tamanho do passo p , entretanto, é diretamente proporcional a τ , completando a demonstração.

■

Concluimos com uma observação. Em primeiro lugar, se g_i é restrição ativa em x^* , a norma do gradiente $\|\nabla g_i^k\|$ é não-decrescente para $\{x^k: x^k = x(r_k)^*, r_k \downarrow 0\}$. Isto implica que o hiperplano $(x - x^k)^T \nabla g_i^k = -g_i(x^k)$ (correspondente à restrição $g_i(x) \leq 0$ no passo MQS) em geral não intercepta a região viável S .

Formalmente isso é mostrado considerando um ponto genérico \bar{x} que pertença à interseção dos hiperplanos

$$\left\{ x: (x - x^k)^T \nabla g_i^k = -g_i^k \right\}, \quad 1 \leq i \leq m$$

Tem-se

$$\begin{aligned}
g_i(\bar{x}) &= g_i^k + (\bar{x} - x^k)^T \nabla g_i^k + \frac{1}{2} (\bar{x} - x^k)^T \nabla^2 g_i^k (\bar{x} - x^k) + \\
&\quad + O(\|\bar{x} - x^k\|^3) \\
&= \frac{1}{2} (\bar{x} - x^k)^T \nabla^2 g_i^k (\bar{x} - x^k) + O(\|\bar{x} - x^k\|^3) \\
&> 0 \quad \forall i, 1 \leq i \leq m
\end{aligned}$$

Por este motivo, um passo do algoritmo MQS "puro" tende a gerar a iteração x^{k+1} fora de S . O MQS-deslocado reduz este passo, de forma que x^{k+1} continue viável e mais próximo de x^* de que seria o caso da iteração MQS.

3.3 Convergência Superlinear do MQS

Dado um problema (P), f, g_i convexas de classe C^3 , o teorema 3.1 garante a existência de uma região de convergência superlinear em torno de x^* . O conceito de autoconcordância serve para a detecção desta região. Supomos inicialmente que uma aproximação muito boa do valor do multiplicador μ^* seja conhecida. Em $x^k \in S$ tem-se

$$(3.14) \quad \lambda(L_{\mu^*}, x^k) = \max_{h \in S^{n-1}} \frac{|h^T \nabla L_{\mu^*}(x^k)|}{(ah^T \nabla^2 L_{\mu^*}(x^k)h)^{1/2}} \leq \frac{\|\nabla L_{\mu^*}(x^k)\|}{(a\zeta_1^*)^{1/2}}$$

onde ζ_1^* é o menor autovalor da hessiana $\nabla^2 L_{\mu^*}$ e a é o parâmetro de autoconcordância da função $L_{\mu^*}(x)$ - e que de fato precisa ser conhecido apenas numa vizinhança convexa de x^* que inclui x^k .

3.4 Sobre Restrições de Igualdade e o Conjunto Ativo

Para maior clareza de exposição nos capítulos 4 e 5, adiantamos algumas observações sobre um assunto que será abordado no final do capítulo 5. Consideramos nesta seção idealizações de algoritmos de barreira logarítmica (ABL) e do método quadrática seqüencial deslocado (MQSD). Em geral, dado um problema (P), não é conhecido de antemão o conjunto ativo A na solução x^* . Assim, torna-se necessária a inclusão, anterior à aplicação do MQSD, de um "processo" capaz de indicar A corretamente.

Seja indicado um incorreto conjunto ativo A' . Se $A \subset A'$, o ABL continua convergindo para a solução x^* , deixando as restrições excedentes não-ativas. O MQSD leva a uma resposta errada, ao exigir $g_i(x^*) = 0 \quad \forall i$. Se, ao contrário, A não está contido em A' , então tanto o ABL como o MQSD levam a respostas erradas, acompanhado de violação das restrições $g_i \in A \setminus A'$.

Estes mesmos problemas existem para o algoritmo que desenvolvemos neste trabalho. Há necessidade de determinação do conjunto ativo A .

Um problema comum é do tipo

$$(P') \quad \begin{array}{ll} \min & f(x) \\ \text{s.a.} & g_i(x) \leq 0 \quad 1 \leq i \leq m_1 \\ & h_j(x) = 0 \quad 1 \leq j \leq m_2 \end{array}$$

com f, g_i, h_j funções convexas de classe C^2 . Nestas condições é tentador simplesmente incluir $h_j(x) \leq 0$ como restrições supostas ativas em x^* . De fato o MQSD leva a uma resposta correta caso seja acertado o conjunto ativo A na sua totalidade. O ABL leva a uma resposta incorreta sempre que existe $\bar{x}^* \in S, h_j(\bar{x}^*) < 0$ para algum j e $f(\bar{x}^*) < f(x^*)$. Estas observações nos levam a omitir o

problema geral (P') de consideração.

Entretanto há um subcaso importante que pode ser incluído: h_j função linear $\forall j$. Já que neste caso tanto h_j como $-h_j$ é convexa, podemos incluir uma das duas como restrição $h_j(x) \leq 0$ suposta ativa em x^* e garantir que o ABL e o MQSD convirjam para x^* .

4.1 Introdução

A proposição 3.1 demonstrou a existência de uma correspondência fraca entre o algoritmo da barreira logarítmica e o método quadrático seqüencial deslocado, indicada abaixo.

$$\begin{array}{ccc}
 \text{ABL} & & \text{MQSD} \\
 \text{Em } x^k \approx x(r)^*, \text{ resolve} & \longleftrightarrow & \min p^T \nabla f^k + \frac{1}{2} p^T H_k p \\
 \min F_{\alpha r}(x) & & \text{s.a. } \nabla g_k p = (\alpha - 1)g^k
 \end{array}$$

O método MQS-deslocado tem sido pouco explorado, talvez por motivo desta correspondência ser fraca.

Exemplificamos. Seja $\alpha \in (0, 1)$ o fator de redução e dados o problema com $n > 1$

$$\begin{array}{ll}
 \min & x^n \\
 \text{s.a.} & x - a \geq 0 \quad (a > 0)
 \end{array}$$

e o ponto $x^0 > a$. Assim

$$\begin{aligned}
 F_r(x) &= \frac{1}{r} x^n - \ln(x - a) \\
 (4.1) \quad \nabla F_r(x) &= \frac{1}{r} n x^{n-1} - \frac{1}{x - a}
 \end{aligned}$$

(4.1) mostra que há uma correspondência biunívoca entre os conjuntos $\{x(r)^* : r > 0\}$ e (a, ∞) dada por

$$r(x) = n(x - a)x^{n-1}$$

O subproblema MQS-deslocado é

$$\begin{array}{ll}
 \min & n x^{n-1} p + \frac{1}{2} n(n-1) x^{n-2} p^2 \\
 \text{s.a.} & p = (\alpha - 1)(x - a)
 \end{array}$$

o que implica que o passo é sempre fixo de forma que

$$\frac{(x_{k+1} - a)}{(x_k - a)} = \alpha$$

Repare agora que, partindo de x_0 , com $r_0 \doteq r(x_0) = n(x_0 - a)x_0^{n-1}$, alvejamos x_1 com $r(x_1) = \alpha r(x_0)$. Escrevemos

$$(4.2) \quad r_{\text{previsto}}(x_1) \doteq r_{\text{pr}}(x_1) = \alpha r_0$$

O ponto x_1 , entretanto, é tal que

$$\begin{aligned} r(x_1) &= n(x_1 - a)x_1^{n-1} \\ &= n\alpha(x_0 - a)(a - \alpha a + \alpha x_0)^{n-1} \end{aligned}$$

e

$$\frac{r(x_1)}{r_{\text{pr}}(x_1)} = \frac{x_1^{n-1}}{x_0^{n-1}}$$

Em geral, na k -ésima iteração, $r_{\text{pr}}(x_k) = \alpha^k r_0$ e

$$(4.3) \quad \frac{r(x_k)}{r_{\text{pr}}(x_k)} = \frac{x_k^{n-1}}{x_0^{n-1}}$$

A expressão (4.3) implica que, exceto no caso $n = 1$ (linear), o MQS-deslocado terá um erro de estimativa grande e crescente com k do "valor verdadeiro" de $r(x_k)$. Este fato motiva a convergência apenas linear neste exemplo. Repare que quanto maior o expoente n , maior será a superestimação de $r(x_k)$; isso torna $r_{\text{pr}}(x_k)$ essencialmente inútil.

4.2 R-Ideal: Conceito e Propriedades

Num problema geral (P) definimos para $x \in S$

$$(4.4a) \quad r_{id}(F_r, x) \doteq \operatorname{argmin}_r \{ \|\nabla F_r(x)\|^2 \}$$

$$(4.4b) \quad r_{id}(B_r, x) \doteq \operatorname{argmin}_r \{ \|\nabla B_r(x)\|^2 \}$$

(leia-se r ideal de x) As seqüências $\{ r_{id}(F_r, x^k) \}$ e $\{ r_{id}(B_r, x^k) \}$ convergem para o mesmo valor $\operatorname{arg} \{ x(r)^* \}$ para $\{ x^k \}$ convergente para $x(r)^* \in T$. Após indicar qual das duas versões $r_{id}(F_r, x)$, $r_{id}(B_r, x)$ está sendo usada em dado momento, indicaremos o valor por $r_{id}(x)$.

Proposição 4.1 Numa vizinhança da trajetória central T , $r_{id}(F_r, x)$ e $r_{id}(B_r, x)$ são funções bem definidas, estritamente convexas e estritamente positivas.

Demo. Consideremos primeiro o caso de $r_{id}(x) = r_{id}(F_r, x)$. Para $x = x(r)^* \in T$ tem-se $r_{id}(x) = r > 0$.

Definimos

$$U = \{ x: r_{id}(x) > 0 \} \cap S$$

Tem-se $T \setminus x^* \subset U$. Dágora em diante r_{id} será considerado como função definida apenas em U . Fixo $x \in U$, $\|\nabla F_r(x)\|^2$ é função quadrática de r , com

$$(4.5) \quad \frac{\partial}{\partial r} (\|\nabla F_r(x)\|^2) = -\frac{2}{r^3} \nabla f^T \nabla f + \frac{2}{r^2} \nabla f^T \sum_i \frac{\nabla g_i}{g_i}$$

Se no ponto $x \in U$ tem-se $r = r_{id}(x)$ então

$$(4.6a) \quad r = \frac{\nabla f^T \nabla f}{\nabla f^T \sum_i \frac{\nabla g_i}{g_i}}$$

e numa vizinhança de $x(r)^*$ a expressão (4.6) é bem definida. Examinemos a segunda derivada.

$$\begin{aligned} \frac{\partial^2}{\partial r^2} (\|\nabla F_r(x)\|^2) &= \frac{1}{r^4} \left[6 \nabla f^T \nabla f - 4r \nabla f^T \sum_i \frac{\nabla g_i}{g_i} \right] \\ &= \frac{1}{r^4} \left[6 (\nabla f^T \nabla f - r \nabla f^T \sum_i \frac{\nabla g_i}{g_i}) + 2r \nabla f^T \sum_i \frac{\nabla g_i}{g_i} \right] \end{aligned}$$

Para $r_{id}(x)$ em (4.5)

$$\frac{\partial^2}{\partial r^2} (\|\nabla F_r(x)\|^2) = \frac{1}{r^4} 2r \nabla f^T \sum_i \frac{\nabla g_i}{g_i}$$

Perto de $x(r)^*$ esta segunda derivada é maior que zero e cresce rapidamente quando $r \downarrow 0$.

No caso de $r_{id}(x) \doteq r_{id}(B_r, x)$ os mesmos resultados seguem ainda mais facilmente, pois neste caso

$$\frac{\partial}{\partial r} (\|\nabla B_r(x)\|^2) = -2 \nabla f(x)^T \sum_i \frac{\nabla g_i}{g_i} + 2r \sum_{i,j} \frac{\nabla g_i^T \nabla g_j}{g_i g_j}$$

e em $r = r_{id}(x)$

$$(4.6b) \quad r = \frac{\nabla f(x)^T \sum_i \nabla g_i / g_i}{\sum_{i,j} \frac{\nabla g_i^T \nabla g_j}{g_i g_j}}$$

é bem definido. A segunda derivada

$$\frac{\partial^2}{\partial r^2} (\|\nabla B_r(x)\|^2) = 2 \sum_{i,j} \frac{\nabla g_i^T \nabla g_j}{g_i g_j} > 0$$

garante a convexidade estrita de $B_r(x)$.

■

Experiências com o conceito r_{id} mostram que U é

comumente quase igual a S numa vizinhança V de x^* com

$$V = \left\{ x \in S : r_{id}(x) < \infty \right\}$$

tanto para $r_{id}(F_r, x)$ como para $r_{id}(B_r, x)$.

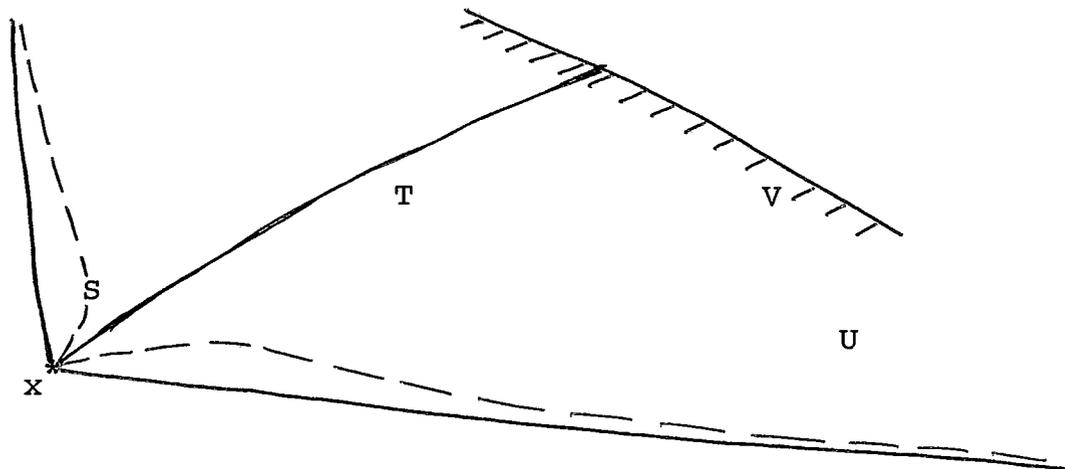


Figura 2

No caso de $r_{id}(B_r, x)$,

$$(4.7) \quad \lim_{x \Rightarrow \partial S} r_{id}(x) = 0$$

e a situação é ilustrada na figura 2.

No caso $r_{id}(F_r, x)$, o limite é indeterminado quando x se aproxima à fronteira de S , porém em geral é finito e maior que zero. Ao determinar $r_{id}(x)$ próximo de T usamos sempre $r_{id}(F_r, x)$. Na fase preliminar do algoritmo de seção 5.1, x às vezes distante de T , usamos $r_{id}(B_r, x)$

O cálculo de $r_{id}(x)$ para $x \in U$ é relativamente simples por ser o mínimo de uma função estritamente convexa de uma variável. Além disso, tem-se uma estimativa inicial de seu valor: $r_{pr}(x)$.

Proposição 4.2 Sejam dados o problema (P) e $x^0 = x(r_0)^*$, $r_0 > 0$. Para dado $\alpha \in (0, 1)$ seja efetivada uma iteração de MQSD ou ABL, obtendo x^1 . Então para α suficientemente grande

$$(4.8) \quad \frac{r_{id}(x^1)}{\alpha r_0} < 1$$

Demo. Para α suficientemente grande temos $r_{id}(x^1) > 0$ bem definido. Provamos que $r_{id}(x^1) < \alpha r_0$. As iterações pelo MQSD implicam em passos p^k que satisfazem $\nabla g_k p^k = -(1 - \alpha)g^k$, ou seja, tal que seja "previsto" $g_i(x^k + p^k) = \alpha g_i(x^k) \forall i$. Sendo g_i convexa e crescente ao longo de T para $r < \bar{r}$, tem-se

$$(4.9a) \quad g_i(x^k + p^k) > \alpha g_i(x^k) \quad \forall i$$

Para α suficientemente grande $x^{k+1} := x^k + p^k$ é próximo de T e portanto (4.9a) implica que $r_{id}(x^{k+1}) < \alpha r_{id}(x^k)$, pois as superfícies $V_r = \{x \in U : r_{id}(x) = r\}$ são transversais a T.

As iterações pelo ABL para resolução do problema (P), não-degenerado, são localmente equivalentes à geração de uma seqüência $\{y^k\}$, soluções aproximadas de

$$(4.9b) \quad \min (F_r(y) \doteq \frac{1}{r} f(y) - m \ln(-y))$$

por passos de Newton ou Quasi-Newton, para valores sucessivos decrescentes de r . Formalmente, efetuamos uma mudança de variáveis $\Phi: x \rightarrow y$, diferenciável, e com $\det(\Phi'(x)) \neq 0 \forall x \in U$, com

$$\begin{aligned} \Phi(T) &= \{(y_1, 0, 0, \dots, 0) : y_1 \leq 0\} \\ &\doteq \{(y_1, 0, 0) \in \mathbb{R}^- \times \mathbb{R}^{m-1} \times \mathbb{R}^{n-m}\} \end{aligned}$$

Já que T não incide assintoticamente em nenhum hiperplano tangente $\{v \in \mathbb{R}^n : v^T \nabla g_i^* = 0\}$, existe aplicação Φ tal que todas as imagens $\{\Phi(x^k)\}$ são arbitrariamen-

te próximos do eixo y_1 negativo, para α suficientemente próximo de 1. T ser ao longo do eixo y_1 implica $\Phi^*(F_r): \mathbb{R}^n \rightarrow \mathbb{R}$ pode ser descrito por (4.9b).

Consideremos $y^0 = y(r_0)^*$ e $y^1 = y^0 + h^0$,

$$h^0 = - \left[\frac{d^2 F_{\alpha r_0}}{dy^2} (y^0) \right]^{-1} \frac{dF_{\alpha r_0}}{dy} (y^0)$$

(Vide Figura 3, onde y^1 e $y(\alpha r_0)^*$ são ilustrados como se esta proposição fosse correta) Vê-se que (4.8) equivale provar

$$\frac{dF_{\alpha r_0}}{dy} (y^1) > 0$$

Por Taylor

$$\begin{aligned} \frac{dF_{\alpha r_0}}{dy} (y^1) &= \frac{dF_{\alpha r_0}}{dy} \Big|_{y^0} + \frac{d^2 F_{\alpha r_0}}{dy^2} \Big|_{y^0} (y^1 - y^0) + \\ &+ \frac{1}{2} \frac{d^3 F_{\alpha r_0}}{dy^3} \Big|_{y^0} (y^1 - y^0)^2 + O(y^1 - y^0)^3 \end{aligned}$$

Basta mostrar

$$(4.10) \quad \frac{d^3 F_{\alpha r_0}}{dy^3} (y^0) = \frac{1}{\alpha r_0} \frac{d^3 f}{dy^3} (y^0) - \frac{2}{(y^0)^3} > 0$$

Por hipótese $\left| \frac{1}{\alpha r_0} \frac{d^3 f}{dy^3} (y^0) \right| \leq k \quad \forall y$. Para y suficientemente próximo de zero, $y < 0$, a desigualdade (4.10) é satisfeita.

■

Observação: (4.8) implica que $r_{id}(x^1)$ é, além de "ideal" enquanto quantificação do próximo subproblema, ainda tem a vantagem de ser menor de que $r_{pr} = \alpha r_0$. Isto, de fato, constitui a grande vantagem oferecida pelo uso do conceito

r_{id} . Resta observar que, na prática, este fenômeno é constatado para quase todo $\alpha \in (0, 1)$, em todas as iterações.

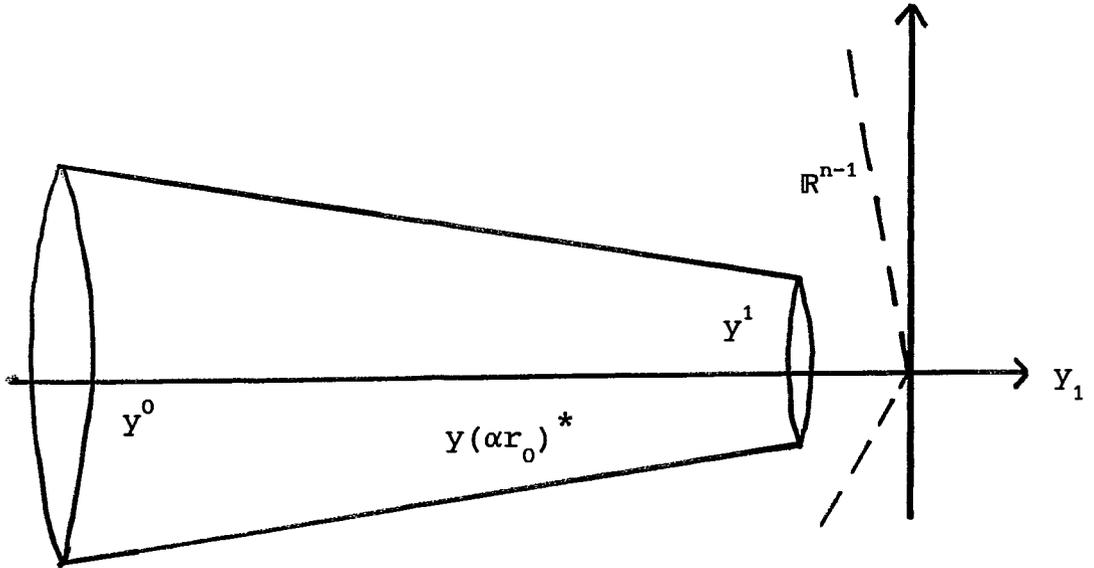


Figura 3

4.3 Elementos para um Algoritmo Parametrizado de Barreira Logarítmica

Na seção 2.4 foi enunciado um algoritmo que agora reenunciamos em termos de MQSD.

Dado $(x^k, r_k) \in U$, onde $r_k = r_{id}(x^k)$, tal que

$$\|\nabla F_{r_k}(x^k)\| \leq \frac{\lambda^*}{2} R(r_k) \doteq \frac{\lambda^*}{2} R$$

tomamos $\alpha \in (0, 1)$ tal que

$$\frac{1 - \alpha}{\alpha r_k} \|\nabla f^k\| \leq \frac{\lambda^*}{2} R$$

Obtemos de (2.19)

$$\| \nabla F_{\alpha r_k}(x^k) \| \leq \lambda_* R$$

o que implica, por definição, que $\lambda(F_{\alpha r_k}, x^k) < \lambda_*$. De (2.9) é garantido um passo p^k tal que em $x^{k+1} = x^k + p^k$ tem-se

$$\| \nabla F_{\alpha r_k}(x^{k+1}) \| \leq \frac{\lambda_*}{2} R(\alpha r_k) \leq \frac{\lambda_*}{2} R$$

para α suficientemente grande. De (4.8) tem-se

$$(4.11) \quad \| \nabla F_{r_{k+1}}(x^{k+1}) \| < \frac{\lambda_*}{2} R$$

e concluímos um passo no algoritmo.

Já que na prática (4.11) pode ser uma desigualdade "forte", devido à diferença entre r_{k+1} e αr_k , temos seguramente a possibilidade de dar um passo mais longo de que aquele apresentado acima. Ao usar a norma do gradiente $\| \nabla F_{r_k}(x^k) \|$ como medida indireta de $\lambda(F_{r_k}, x^k)$, reduzimos em muito o trabalho computacional da avaliação do valor do passo dado.

Seguiremos a tática de não "desperdiçar" o passo p^k uma vez efetuada sua computação, sem aproveitá-lo o máximo possível na determinação de um passo longo. Como será indicado nos testes computacionais (vide cap. 6), há considerável potencial nesta idéia simples, devido ao fato da trajetória central T ser relativamente retilínea em regiões razoavelmente grandes. Este algoritmo será formalizado em seção 5.2 e constitui a etapa principal do APBL.

5.1 Fase 1: Preliminar

Já aludimos a um algoritmo parametrizado barreira logarítmica em seção 4.3. A etapa apresentada exige $x^0 \in S$ próximo da trajetória central T . Desenvolvemos a seguir uma fase preliminar do algoritmo, que visa encontrar x^0 nestas condições a partir de $y \in S$ arbitrário.

Nesterov e Nemirovsky [26] encontram dificuldades semelhantes. A etapa preliminar destes autores consiste de supor \bar{S} compacto. Existe um centro analítico único e Nesterov e Nemirovsky propõem uma família

$$(5.1) \quad F_r^{(1)}(x) = \left\{ r(x - x^k)^T \sum_i \frac{\nabla g_i^k}{g_i^k} - \sum_i \ln(-g_i(x)) : r \geq 0 \right\}$$

que, ao reduzir r para 0, converge em mínimo para o centro analítico. r inicial é tomado como 1 ; $r_k = \Lambda^{-k}$ onde Λ é uma constante para a qual é demonstrada convergência da etapa preliminar do algoritmo. Repare que para $F_r^{(1)}$ não faz sentido falar em r ideal, pois a "trajetória central" na etapa preliminar inclui todo o conjunto S , devido ao fato de $F_r^{(1)}$ depender de $n+1$ parâmetros (r, x^k) e não de um parâmetro só como no caso de F_r .

Nesterov e Nemirovsky propõem que os sucessivos mínimos sejam encontrados usando método de Newton ou Quase-Newton. Vale a pena frisar que Λ formalmente permitido em [26] é freqüentemente muito próximo de 1 e, conseqüentemente, de convergência lenta. Além disso, a combinação da etapa preliminar citada com a etapa da seção 4.3 implicaria, em geral, em inicialmente incorporar uma restrição $g_{m+1}(x) \leq 0$ no problema (P) para tornar S um conjunto

limitado e, na etapa seguinte, descartar esta restrição artificial.

Quanto à extensão do método quadrático seqüencial deslocado à primeira etapa do algoritmo, na seção 3.2 foi mostrada a fraca equivalência entre MQSD e o ABL. Isso faz com que o MQSD seja inteiramente inadequado para a localização de x^0 próximo de T . A restrição

$$(5.2) \quad \nabla g_k p = (\alpha - 1)g^k$$

tenta reduzir os valores de $g_i(x)$, linearmente, por um fator fixo. Se $y^0 \in S$ for próximo de ∂S , (5.2) tende a manter y^1 tão ou mais próximo, dificultando, em muito, progresso em direção a T .

Há ainda outra dificuldade na aplicação do MQSD à fase preliminar do algoritmo: a quase impossibilidade de usar r ideal em substituição a r e α fixos. (4.6a) mostra que

$$\lim_{x \rightarrow \partial S} r_{id}(F_r, x) \quad \text{indeterminado}$$

e de (4.6b)

$$\lim_{x \rightarrow \partial S} r_{id}(B_r, x) = 0$$

Nos dois casos o uso de r_{id} é pouco eficiente, pois seu valor significa pouco quanto à proximidade entre y^0 e $x(r_{id})^*$, e por isto a estimativa do multiplicador de Lagrange μ^0 perde seu valor.

Na ausência de um método geral pela localização de $x(r)^* \in T$, propomos duas alternativas a serem usadas em casos diferentes. Se f, g_i são convexas um passo \bar{p}^k de Newton ou Quase-Newton é direção de descida de F_r e B_r . Em casos mais gerais o método de descida máxima com salvaguardas é apropriado. Sabe-se que T normalmente se encontra longe de ∂S para valores de r suficientemente grandes, e portanto qualquer passo p^k deve ter monitoramento de forma

que

$$(5.3) \quad p^k = \gamma \bar{p}^k$$

$$\gamma \quad \text{tal que} \quad g_i(y^k + p^k) \leq \varepsilon g_i(y^k) \quad \forall i$$

Isso é, o ponto y^{k+1} está contido no conjunto $S_\delta := \{ x : g_i(x) \leq \delta_i \quad \forall i \}$, onde $\delta_i := \varepsilon g_i(y^k)$. S_δ é um conjunto que contém os pontos $x \in T$ em que $r_{id}(x) \geq \delta' > 0$ (Vide Fig. 4)

Quanto à função-objetivo da fase 1, testamos duas alternativas: $B_r(x)$, r fixo e

$$(5.4) \quad B_{r_k}(x) : r_k = \max \{ r_{id}(y^{k-1}), C \}$$

para prefixado $C > 0$, sendo o propósito de (5.4) usar como função-objetivo a melhor alternativa possível — mas evitando valores de r_k excessivamente pequenos

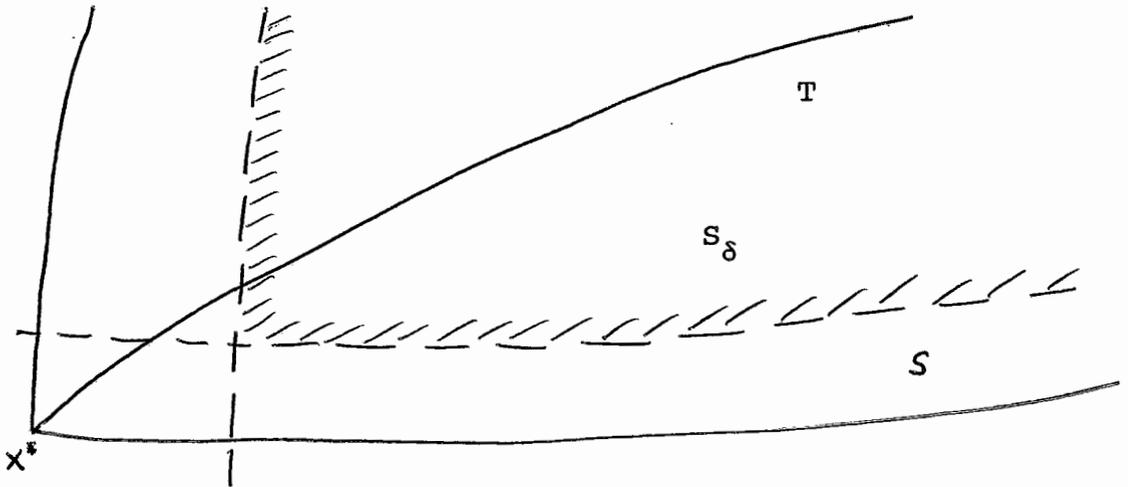


Figura 4

Resumimos a Fase 1 (Preliminar)

$$y^0 \in S. \quad k := 0.$$

(5.5) Escolha r

Calcule cota superior $\lambda \geq \lambda(F_r, y^k)$

se $\lambda < \lambda_*/2$, ponha $x^0 = y^k$
 Calcule o passo \bar{p}^k de Newton ou Quase-Newton para
 $\min B_r(x)$
 Seja $y^{k+1} = y^k + p^k$ onde $p^k = \gamma \bar{p}^k$ é calculado via busca unidimensional
 $k := k+1$. Vá para (5.5)

5.2 Algoritmo Parametrizado de Barreira Logarítmica (APBL)

(Versões 1a e 1b)

Continuamos o algoritmo com a fase 2

$x^0 \in U$, $\varepsilon > 0$, $k := 0$. $\lambda(F_{r_0}, x^0) < \lambda_*/2$

Fixe $\alpha \in (0, 1)$ inicial. (vide observacso i)

(5.6) Estime $R(r) = (av_1)^{1/2}$

(5.7) Calcule o passo \bar{p}^k do MQSD (em 3.9)

Calcule $r := r_{id}(x^k + \bar{p}^k)$

(5.8) Estime o valor de $\bar{\lambda} = \lambda(F_r, x^k + \bar{p}^k)$

(5.9) se $\bar{\lambda} > \lambda_*/2$, majore α e vá para (5.7)

(5.10) se $\bar{\lambda} \leq \lambda_*/2$, teste para ε -convergência.

caso não haja convergência: $j := 0$

(5.11) Ponha $x_{(j)}^{k+1} = x^k + \xi_j \bar{p}^k$

Calcule $r^j := r_{id}(x_{(j)}^{k+1})$. Novamente, há duas possibilidades:

se $\lambda(F_{r^j}, x_{(j)}^{k+1}) \leq \frac{\lambda_*}{2}$, ponha $j := j + 1$ e vá para (5.11)

se $\lambda(F_{r^j}, x_{(r)}^{k+1}) > \frac{\lambda^*}{2}$, ponha $\bar{j} := j - 1$ e continue

Com o passo $\xi_{\bar{j}} \bar{p}^k$ determinado, há duas possibilidades.

Versão 1a: $x^{k+1} := x^k + \xi_{\bar{j}} \bar{p}^k$. $k := k + 1$ e vá para (5.6)

Versão 1b: Resolva o novo sistema

$$\begin{aligned} \min \quad & p^T \nabla f^k + \frac{1}{2} p^T H_k p \\ \text{s.a.} \quad & \nabla g_k^T p = (\xi_j (1 - \alpha) - 1) g^k \end{aligned}$$

para p^k e tome $x^{k+1} = x^k + p^k$. Calcule $r := r_{id}(x^{k+1})$. Teste para ε -convergência e averigüe que $\lambda(F_r, x^{k+1}) < \frac{\lambda^*}{2}$. $k := k+1$ e vá para (5.6)

Observações:

i) O parâmetro de inicialização $\alpha \in (0,1)$ é uma garantia da convergência do subproblema MQSD. Se (P) convexo, o menor valor possível de α é estritamente menor que 1 e tem limite igual a zero quando $\{x^k\} \rightarrow x^*$. Este valor depende do problema e não pode ser calculado anteriormente à solução do problema, no caso geral. Usamos valores heurísticos nos casos apresentados no Cap. 6. Gonzaga [11] obteve redução linear de r para programação linear, mas há pouca informação para o caso geral.

ii) O valor de $\bar{\lambda}$ (5.8) é facilmente estimada pela relação

$$\lambda(F_r, x^k + \bar{p}^k) \leq \frac{\|\nabla F_r(x^k + \bar{p}^k)\|}{R(r)}$$

e normalmente será menor que $\lambda^*/2$, a não ser que α seja

excessivamente pequeno. α inicial é dado um valor mediano e será majorado ou não conforme seu desempenho seja falho ou promissor, respectivamente. As estimativas de $R(r)$ e r_{id} são, em nosso caso, feitas em simples subrotinas. Para detalhes, consulte o início do Cap. 6.

iii) Este algoritmo tem convergência garantida mas na prática é lenta. A versão 2 do algoritmo inclui uma terceira fase que assegura convergência super-linear.

iv) A seqüência $\{\xi_j\}_{j \in \mathbb{N}}$ é crescente e modulada de forma apropriada a detectar a possibilidade de um passo longo. Nos testes computacionais do capítulo 6 usaremos $\xi_j = 2^{j/2}$.

Lembremos que $\mu^k = -\frac{r}{g^k}$ é estimativa do multiplicador de Lagrange. Há vários meios de estimar uma cota superior para $|\mu_i^k - \mu_i^*|$, $1 \leq i \leq m$. No decorrer do algoritmo, uma vez que se tem certeza desta cota ser pequena, torna-se atrativa a alternativa de substituir o problema das fases iniciais por

$$(5.12) \quad \min L_{\mu^k}(x)$$

Propomos como fase 3 do APBL (Versões 2a e 2b) a substituição de (5.10) por

$$\text{Ponha } \bar{\mu} = -\frac{r}{g(x^k + \bar{p}^k)} \quad \text{e estime } \hat{\lambda} := \lambda(L_{\bar{\mu}}, x^k + \bar{p}^k)$$

$$\text{onde } L_{\bar{\mu}}(x) = f(x) + \sum_i \bar{\mu}_i g_i(x). \quad \text{Se } \hat{\lambda} < \lambda_*, \text{ vá à fase 3.}$$

Fase 3

$$\begin{aligned} \min \quad & p^T v f^k + \frac{1}{2} p^T H_k p \\ \text{s.a.} \quad & \nabla g_k^T p = -(1 - \alpha) g^k \end{aligned}$$

onde agora α tem um valor fixo e pequeno.

5.3 Condições de Convergência do APBL

A proposição que segue formaliza as diversas avaliações feitas ao longo das últimas seções.

Proposição 5.1 Seja dado um problema (P) com f, g_i convexas e de classe C^3 com f β -compatível com a barreira m -autoconcordante $F(x) = -\sum_i \ln(-g_i(x))$. Suponha que existe um conjunto não-vazio $S = \{x: g_i(x) < 0 \ \forall i\}$ e seja a solução única de (P) um ponto x^* com multiplicador μ^* tal que $\mu_i^* > 0 \ \forall i$. Então para qualquer $\varepsilon > 0$ e qualquer $x^0 \in S$ o algoritmo APBL produzirá uma seqüência finita de pontos $\{x^k\} \subset S$ com as seguintes características.

i) Numa primeira fase levará a um ponto x^{k_1} tal que para certo $r > 0$, tem-se $\lambda(F_r, x^{k_1}) < \frac{\lambda^*}{2}$.

ii) Numa segunda fase levará a um ponto x^{k_2} numa sucessão de iterações para as quais $\lambda(F_{r_1}, x^{k_1+1}) < \frac{\lambda^*}{2}$ e onde $\{r_1\}$ é uma sucessão de reais positivos decrescente por pelo menos um fator linear α e tal que $\lambda(L_{\mu^*}, x^{k_2}) < \frac{\lambda^*}{2}$.

iii) Numa terceira fase levará a um ponto x^{k_3} numa sucessão de iterações onde

$$\lambda(L_{\mu^*}, x^{k_2+1}) < \frac{1}{2} \lambda(L_{\mu^*}, x^{k_2})$$

e tal que

$$\lambda(L_{\mu^*}, x^{k_3}) < \varepsilon$$

Demo. A afirmação i) decorre como corolário do teorema 3.1 e a iii) de (2.9). ii) é comprovado ao observar que há garantia de $r_{id}(x^{k_1+1}) := r_1 \leq r_{pr}(x^{k_1+1}) = \alpha^1 r_{k_1}$ onde $\alpha < 1$ é o fator de redução do parâmetro r . Já que a

sucessão $\{ x_1^{k+1} \}$ tem convergência linear para x^* , está garantida a ocorrência de um ponto x_2^k com $\lambda(L_{\mu}^*, x_2^k) < \lambda_*$

■

Observação. A proposição 5.1 garante muito pouco, menos que de fato ocorre com a aplicação do APBL. Isso decorre da natureza parcialmente heurística do algoritmo. Por este motivo não tentamos colocar cotas superiores no número de iterações.

Um algoritmo pode ser teoricamente muito correto mas ineficiente na resolução de problemas reais. No caso do APBL confronta-se com um número de parâmetros que, conforme os valores realmente estabelecidos, podem implicar que o algoritmo tem pouco mérito prático.

Foi implementado uma versão bastante extensa (em torno de 2500 linhas de FORTRAN 77) visando testar diversos aspectos do APBL. Três problemas-teste foram usados:

$$\begin{aligned}
 \text{Problema 1} \quad & \min \quad x_1^2 + 3x_2^2 + 0.1x_3^4 \\
 & \text{s.a.} \quad 2.025 - x_1 - 0.5x_2 - 2.55x_3 \leq 0 \\
 & \quad \quad 0.25 \quad \quad \quad - x_2 + x_3^2 \leq 0
 \end{aligned}$$

$$\begin{aligned}
 \text{sendo} \quad x^* &= (0.5, 0.5, 0.5) & \mu^* &= (1, 2.5) \\
 f(x^*) &= 1.00625
 \end{aligned}$$

Cinco pontos iniciais distintos foram usados. A tabela A-1 no Apêndice contém informação a respeito destes pontos.

$$\text{Problema 2} \quad (\text{n}^\circ 43 \text{ em } [14], \text{ Rosen-Suzuki})$$

$$\begin{aligned}
 \min \quad & x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\
 \text{s.a.} \quad & -8 + x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 \leq 0 \\
 (6.1) \quad & -5 + 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 \leq 0 \\
 & -10 + x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 \leq 0
 \end{aligned}$$

$$\begin{aligned}
 \text{sendo} \quad x^* &= (0, 1, 2, -1) & \mu^* &= (1, 2, 0) \\
 f(x^*) &= -44
 \end{aligned}$$

$$\text{Ponto inicial: } x^0 = (0, 0, 0, 0)$$

Problema 3 (Não-convexo)

$$\begin{aligned} \min \quad & x_1^2 x_2^2 \\ \text{s.a.} \quad & (x_1 - 3.0)^2 + (x_2 - \rho)^2 - 1 \leq 0 \end{aligned}$$

ρ é um parâmetro, i.é. o problema 3 é uma família de problemas. Tem-se resultados distintos para cada valor do parâmetro usado.

Caso 3.1	$\rho = 2.5$ $\mu^* = 12.048$	$x^* = (2.4325, 1.6764)$ $f^* = 16.63330$
Caso 3.2	$\rho = 1.5$ $\mu^* = 4.1738$	$x^* = (2.8199, 0.5164)$ $f^* = 2.12011$
Caso 3.3	$\rho = 1.1$ $\mu^* = 0.8852$	$x^* = (2.9661, 0.10057)$ $f^* = 0.088999$

Observação 1. O valor do parâmetro \underline{a} de autoconcordância é de difícil avaliação e constitui a nosso ver um assunto ainda a ser tratado satisfatoriamente. Uma subrotina em nossa implementação faz uma estimativa de uma cota inferior de seu valor. No problema 3 o parâmetro \underline{a} não existe formalmente, mas é usado $a(U)$, onde U é uma região em torno dos pontos da seqüência $\{x^k\}$. Para maiores informações sobre \underline{a} , consulte a seção 6.5.

Observação 2. O problema 2 é um problema-teste clássico. Por ser quadrático e convexas as funções f, g_1 , o verdadeiro problema é claramente a identificação do conjunto ativo A . Testamos as duas alternativas de previamente i) conhecer A e ii) não conhecer A .

O algoritmo parametrizado em suas várias versões exigiu programação relativamente extensa. Tendo como prioritário a obtenção de diversos tipos de informação sobre o

APBL, abrimos mão de medir o performance em tempo de CPU. Diversas subrotinas heurísticas foram usadas na implementação atual que poderiam afetar negativamente os resultados computacionais. Em particular, mencionamos duas destas.

i) Subrotina RDBU. Razão Dourada Busca Unidimensional. Após estabelecer um intervalo conveniente, o método da razão dourada é usada até reduzir a 10^{-4} seu comprimento relativo. Este método foi escolhido por sua transparência.

ii) Subrotina XLAMBDA. A obtenção de estimativas dos valores de $\lambda(F_r, x^k)$ e $R(r)$ não encontraram precedentes. Foi utilizado um método relativamente "caro" de maximização da função

$$\Lambda(h) = \frac{|h^T \nabla F_r^k|}{(ah^T \nabla^2 F_r^k h)^{1/2}}$$

sobre a esfera unitária $S^{n-1} = \{h \in \mathbb{R}^n : \|h\|_2 = 1\}$ por maximização unidimensional em direções $\nabla \Lambda(h^k)$ até os passos dados serem pequenos.

A idealização do parâmetro r traz em seu bojo a possibilidade de introduzir um viés sistemático nos dados do problema em cada iteração. Tentamos nos certificar quanto à ocorrência deste fenômeno. Um parâmetro chamado MULT determina se $R(r)$ é para ser avaliado no ponto x^k , próximo de $x(r_{id})^*$, ou num ponto mais afastado de T . Também foram calculados os valores de $\lambda(F_r, x)$ e $R(r)$ em pontos escolhidos e não ocorrentes em seqüências de iterações.

6.1 Uso de r -ideal na fase 1

Passamos a descrever os diversos testes realizados. $r_{id}(B_r, x^0)$ foi avaliado em diversos pontos para problemas 1 e 2. Usamos como alternativa mais viável para a primeira fase:

$$\min B_r(x) \quad r = \max \{r_{id}(x), C\}$$

com $C = 10$ para evitar uma redução indevida do parâmetro r , e ainda $C = 100$ no caso de um ponto "difícil". Para os pontos iniciais escolhidos temos os seguintes resultados, onde N representa o número de iterações até atingir $\|\nabla B_r(x^k)\| \leq 0.01/r_{id}$. O critério para esta escolha foi de simplesmente estabelecer $\lambda(F_r, x^0) \ll \lambda^*/_2$ para iniciar a segunda fase. A tabela 6.1 contém informações a respeito de problema 1. As tabelas 6.2 e 6.3 se referem ao problema 2, sendo respectivamente o caso i) de conhecer previamente o conjunto ativo A , e ii) de ignorar o conjunto A .

x^0	$r_{id}(B_r, x^0)$	$\ \nabla B_r(x^0)\ $	N	$\ \nabla B_r(x^k)\ $	$r_{id}(x^k)$
(5, 5, 2)	0.3770	31.5361	4	0.0521	10.0
(1, 5, 1)	54.61	15.3126	2	0.0047	48.24
(-5, 10, 1)	0.1296	60.7328	4	0.0021	10.0
(-5, 5, 2)	6.809	28.0871	4	0.0866	10.0
(5, 17, -4)	-0.8561	110.2251	19	0.0216	10.0
(5, 17, -4)	-0.8561	110.2251	7	0.1721	100.0

Tabela 6.1

x^0	$r_{id}(B_r, x^0)$	$\ \nabla B_r(x^0)\ $	N	$\ \nabla B_r(x^k)\ $	$r_{id}(x^k)$
(0, 0, 0, 0)	11.7413	21.6962	2	0.0454	10.28
(0, 2, 0, 0)	1.0000	22.2236	3	0.0075	10.00
(0, 1, -.2, .1)	12.1305	22.2282	2	0.0433	10.71
(.5, 0, .3, .2)	4.3075	20.3006	3	0.0032	10.00
(-.5, .5, -.5, -.2)	9.3434	26.0018	3	0.0039	10.00

Tabela 6.2

x^0	$r_{id}(B_r, x^0)$	$\ \nabla B_r(x^0) \ $	N	$\ \nabla B_r(x^k) \ $	$r_{id}(x^k)$
(0, 0, 0, 0)	12.6426	21.5147	2	0.0173	11.74
(0, 2.0, 0, 0)	0.2727	22.5543	3	0.0603	10.00
(0, 1, -.2, .1)	5.5123	22.9302	3	0.0005	10.00
(.5, 0, .3, .2)	4.3075	20.0526	3	0.0021	10.00
(-.5, .5, -.5, -.2)	5.7675	26.2103	3	0.0013	10.00

Tabela 6.3

É evidente que o processo de idealização na fase 1 providencia informação que freqüentemente não será utilizada, como por exemplo ocorre ao revelar que $r_{id}(B_r, x^0)$ é negativo. Mas a informação é muito facilmente obtida e servirá nestes casos apenas de desqualificar r_{id} enquanto parâmetro confiável. De todo modo, métodos heurísticos servem bem para aproximação inicial a T, e a ocorrência de casos como o do ponto (5, 17, -4) no problema 1, de convergência lenta para um ponto próximo de T, nos parece vinculado mais a limitações específicas da implementação da fase 1 de que a uma dificuldade inerente ao ponto e ao problema.

6.2 APBL sem passos longos

O uso de $r_{id}(F_r, x)$ permite medir melhor o progresso ao longo da trajetória central T mesmo sem usá-lo na detecção de possíveis passos longos. Comparamos aqui a norma $\| \nabla F_r(x^k) \|$, r prefixado e r-ideal. Já que a determinação do conjunto ativo constitui uma questão à parte, supomo-lo conhecido no caso do problema 2. Neste sentido escolheu-se pontos iniciais próximos de T e os valores de diversos

parâmetros.

Todas as buscas unidimensionais foram realizadas com precisão relativa $\varepsilon = 10^{-4}$ em todos os problemas-testes. No problema 1, x^0 foi estabelecido com $r_{id}(x^0) = 50$ e $\lambda(F_{50}, x^0) \leq 0.001$. No problema 2 usou-se um ponto x^0 obtido da execução da fase 1 a partir do ponto de referência $(0, 0, 0, 0)$. Uma redução uniforme DRED = 0.95 foi executada em cada iteração. Os resultados para os problemas 1 e 2 estão nas tabelas A-2 e A-4.

Os três casos do problema 3 são apresentados nas tabelas A-6, A-7 e A-8. Aqui o fator de redução DRED de \underline{r} utilizado foi 0.90.

A determinação de uma aproximação de μ^* é necessária para a passagem à terceira fase do algoritmo. Uma estimativa fácil é

$$(\mu^k)^* = - \frac{r_{id}}{g^k}$$

Previsões lineares extrapolativas podendo $\mu_k^*(r) = c_k r + d_k$ e computando c_k e d_k das últimas duas iterações também são de cálculo muito fácil. As tabelas A-3 e A-5 mostram estas estimativas para Problema 1 e 2 respectivamente.

As tabelas mostram a crescente disparidade entre $r_{id}(x^k)$ e \underline{r} reduzido por um parâmetro fixo. A razão entre as normas dos gradientes de $F_{r_{id}}$ e $F_{r_{pr}}$ cresce linearmente ao longo da Fase 2. Isso ocorre de maneira marcante no último caso do Problema 3.

6.3 APBL Versões 1a e 1b

Todas as versões do APBL foram testados nos três problemas-testes com vários valores dos parâmetros.

Testamos as versões 1a e 1b do APBL no problema 1 de forma ilustrar vários aspectos dos algoritmos. A partir do

ponto $y^0 = (5, 17, -4)$ iniciou-se aproximação de T, terminando a Fase 1 no ponto $x^0 = (2.4095, 2.0953, 0.6317)$ em ambos os casos, com r-ideal igual a 15.00. Os fatores de redução DRED usados foram 0.90 e 0.95 respectivamente. Impediu-se que $\lambda(F_r, x^k)$ ultrapassasse $\lambda_*/2$ neste e nos outros problemas-testes.

No problema 2 as versões 1a e 1b do APBL foram testados a partir de $y^0 = (0, 0, 0, 0)$ como especificado em [14]. A primeira fase em ambas as versões termina no ponto $x^0 = (-0.1711, 0.8235, 1.3190, -0.2966)$ com r-ideal igual a 10.28 . A partir deste ponto inicial de Fase 2 as duas versões foram testados com fator de redução DRED igual a .84. Na tabela 6-4 tem-se os resultados da versão 1a e na tabela 6-5 os da versão 1b. Indica-se por Extens. o múltiplo do passo padrão que é dado e por Reaprox. o passo de reaproximação à trajetória central T.

k	r-ideal	$\ \nabla F_r(k)\ $	$g_1(x^k)$	$g_2(x^k)$	$\ x^k - x^*\ $
0	10.27958	.00442	-4.84386	-3.39270	1.009430
1	6.63172	.12877	-3.66030	-2.58321	.720394
2	4.57013	.16208	-2.79065	-1.98075	.520695
3	3.62649	.19437	-2.33565	-1.65841	.428593
4	2.91161	.22874	-1.95637	-1.38950	.353927
5	2.90026	.00299	-2.04084	-1.35314	.385793
6	1.82979	.12683	-1.36976	-.91251	.251983
7	1.18700	.24762	-.92333	-.61739	.163525
8	1.04094	.33193	-.81840	-.54706	.147660
9	.79090	.42698	-.63166	-.42262	.110545
10	.72378	.50111	-.58101	-.38868	.103071
11	.66304	.55550	-.53446	-.35755	.094641
12	.60759	.61487	-.49164	-.32892	.086927
13	.60407	.00955	-.52806	-.30972	.103518
14	.40452	.24795	-.35799	-.21024	.070808
15	.27137	.86731	-.24162	-.14215	.046229
16	.25879	.52863	-.23130	-.13571	.044744
17	.17486	.95952	-.15707	-.09223	.030088
18	.11832	1.75592	-.10671	-.06264	.022038
19	.09024	2.53860	-.08180	-.04769	.023585
20	.08571	2.11008	-.07805	-.04523	.015179

Tabela 6-4 (primeira parte)

k	$f(x^k)$	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
0	-28.76203	1.414	2.218	.00350	1.262105
1	-33.28791	1.414	1.568	.08754	1.470951
2	-36.21352	1.000	1.374	.11109	1.458978
3	-37.64540	1.000	1.290	.11929	1.629347
4	-38.78429	reaprox.	1.232	.14039	1.629347
5	-38.76236	4.000	-	.00144	2.067615
6	-40.59352	4.000	1.153	.05152	2.461881
7	-41.74762	1.414	1.098	.09174	2.699156
8	-42.01421	2.828	1.088	.09672	3.431913
9	-42.47964	1.000	1.062	.11493	3.715250
10	-42.60523	1.000	1.061	.11757	4.262060
11	-42.71972	1.000	1.049	.12255	4.533028
12	-42.82460	reaprox.	1.045	.13564	4.533028
13	-42.82285	8.000	-	.00201	4.746076
14	-43.20805	8.000	1.032	.03627	6.836783
15	-43.46505	1.000	1.029	.15873	5.463932
16	-43.49173	8.000	1.158	.05584	9.466462
17	-43.65593	8.000	1.014	.07246	13.242066
18	-43.76674	5.657	1.010	.09321	18.838836
19	-43.82004	1.000	1.049	.21114	12.023076
20	-43.83086	reaprox.	1.255	.17550	12.023076

Tabela 6-4 (segunda parte)

Posição Final: .000811 .998783 1.993243 -.983531

Versão 1b

k	r-ideal	$\ \nabla F_r(k)\ $	$g_1(x^k)$	$g_2(x^k)$	$\ x^k - x^*\ $
0	10.27958	.00442	-4.84386	-3.39270	1.009430
1	4.87299	.26456	-2.90895	-2.08658	.558633
2	4.84408	.00155	-3.00677	-2.04638	.581778
3	2.46174	.18130	-1.73218	-1.19634	.316351
4	1.45461	.32369	-1.08898	-.75455	.193513
5	1.44872	.00409	-1.14648	-.72534	.216526
6	.91141	.17322	-.74489	-.47301	.138078
7	.59081	.39855	-.49266	-.31346	.090250
8	.44776	.59348	-.37689	-.23992	.068646
9	.34105	.84444	-.28910	-.18411	.052459
10	.33935	.01070	-.31051	-.17360	.062867
11	.22682	.24990	-.20888	-.11693	.042082
12	.15255	.61509	-.14107	-.07904	.028303
13	.10297	1.14608	-.09550	-.05353	.019105
14	.06967	1.92315	-.06474	-.03630	.012927
15	.04721	3.06339	-.04393	-.02464	.008761
16	.03203	4.73865	-.02983	-.01674	.005944
17	.02175	7.20122	-.02027	-.01137	.004036
18	.01477	10.82182	-.01377	-.00773	.002741
19	.01004	16.14569	-.00936	-.00525	.001863
20	.00682	23.97439	-.00636	-.00357	.001266

k	$f(x^k)$	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
0	-28.76203	1.000	-	.00350	1.262105
1	-35.74538	reaprox.	3.287	.20962	1.262105
2	-35.74088	4.000	-	.00099	1.569867
3	-39.51763	1.000	1.537	.08753	2.071312
4	-41.26565	reaprox.	1.278	.15627	2.071312
5	-41.25905	8.000	-	.00172	2.377562
6	-42.25007	8.000	1.159	.04719	3.670686
7	-42.85503	5.657	1.099	.07991	4.987607
8	-43.12852	1.000	1.070	.09644	6.153558
9	-43.33405	reaprox.	1.489	.13723	6.153558
10	-43.33194	16.000	-	.00152	7.045133
11	-43.55264	16.000	1.036	.02385	10.475978
12	-43.69875	16.000	1.023	.04138	14.864549
13	-43.79648	16.000	1.016	.05380	21.301959
14	-43.86222	16.000	1.011	.06253	30.757643
15	-43.90659	16.000	1.007	.06860	44.653053
16	-43.93660	16.000	1.005	.07281	65.084520
17	-43.95695	16.000	1.003	.07570	95.126886
18	-43.97075	16.000	1.002	.07769	139.296605
19	-43.98013	16.000	1.002	.07904	204.262004
20	-43.98649	16.000	1.001	.07995	299.865579

Tabela 6-5

Posição Final: -.000016 .999915 1.999613 -.999236

A partir das tabelas de cima e do Apêndice vê-se a realização de extensões às vezes consideráveis. Com os valores do fator de redução $DRED = 0.92$, $DRED = 0.96$ e $DRED = 0.98$ houve extensões linearmente equivalentes a 0.68 (previsto) e que de fato resultaram em reduções de até 0.51.

A versão 1b parece à primeira vista ser superior à versão 1a, mas deve lembrar que cada iteração de 1b equivale duas iterações de 1a. Neste sentido, a versão 1b não parece demonstrar nenhuma vantagem em relação a 1a.

A proporcionalidade entre $R(r)$ e r_{id}^{-1} é revalidada e isto poderia proporcionar meios de tornar menos oneroso o cálculo de alguns parâmetros. Diversas reaproximações a T após passos excessivamente longos se tornaram necessárias, e apenas a fixação de valores conservativos de parâmetros impediria isso. Mesmo assim, passos longos foram dados após

reaproximações, não servindo de contra-indicação ao método.

Problema 2 Caso do Conjunto Ativo não ser conhecido.

Existem diversos meios de identificar o conjunto ativo de um problema (P) e não aprofundaremos o assunto aqui. Apenas ressaltamos a relativa facilidade deste trabalho ao usar uma das versões do APBL. A equivalência de ordem 0 entre o ABL e o MQS-deslocado depende do acerto na escolha de conjunto ativo. Assim, a correspondência entre os passos é perdida no caso de erro na determinação de A.

Exemplificamos abaixo, tomando inicialmente todo o conjunto (6.1) do problema 2 como conjunto supostamente ativo. Obviamente, a Fase 1 não é afetada e a Fase 2 é iniciada em x^0 próximo de $x(11.78)^*$.

Os resultados são claros. Apesar de usarmos um fator de redução de 0.92, maior de que o valor de 0.84 usado no caso anterior de conhecer A, o valor limite de $\lambda(F_r, x^k)$ é logo ligeiramente violado e uma reaproximação a T é efetivada. Nas vinte iterações tabeladas, seis reaproximações e revalorizações do fator de redução DRED são realizadas, tornando DRED = 0.99875 sem melhorar muito sua precisão na determinação de pontos sucessivos que acompanham a trajetória central.

Observe que o valor de $g_3(x^k)$ diminui ao invés de aumentar, por cinco vezes. Extrapolações lineares de valores $g_1(x)$ fazendo uso de r_{id} ou de ∇g_1 indicam que, das três restrições, $g_3(x)$ é muito menos provavelmente ativo em x^* de que $g_1(x)$ ou $g_2(x)$. Ao modificar o conjunto ativo a convergência se torna imediatamente vigorosa, podendo fixar DRED = 0.84 e obtendo longos passos. Utiliza-se

$$x^0 = (-.06554662 \quad .55942341 \quad 1.13168937 \quad -.05112715)$$

Problema 3

Apresentamos os três casos ($\rho = 2.5, 1.5$ e 1.1) do problema 3 na versão 1a do algoritmo. Como este caso é ilustrativo, porém não formalmente sob inspeção, omitimos testagem extensiva. Em particular, omitimos a versão 1b. Esta versão foi testada e se comporta de forma muito parecida com a versão 1a. O valor artificial $a = 1.0$ do parâmetro \underline{a} de autoconcordância foi usado. A convexidade local de F_r próximo de T (i.é. com valor apropriado de \underline{r}) torna factível o uso de um parâmetro de autoconcordância, mas nenhuma tentativa de estimar o valor foi realizada.

Os pontos iniciais usados são próximos dos pontos $x(10)^* \in T$ dos respectivos valores de ρ :

$$\begin{array}{ll} \rho = 2.5 & x^0 = (2.88, 2.00) \\ \rho = 1.5 & x^0 = (2.88, 1.00) \\ \rho = 1.1 & x^0 = (2.88, 0.60) \end{array}$$

Em todos os três casos foram usados como critério de convergência da subrotina de busca unidimensional $\varepsilon = 10^{-3}$. Um controle da grandeza de $\lambda(F_r, x^k)$ foi mantido como nos outros problemas. Passos longos de até quatro vezes a magnitude do passo padrão foram permitidos. Já que o fator de redução do parâmetro \underline{r} usado foi 0.9, isso significa que reduções "previstas" de até 0.6 eram contempladas. A entrada Red R das tabelas indica por quanto foi ultrapassado a redução "prevista", e Extens. indica o múltiplo efetivado do passo padrão.

Caso $\rho = 1.5$

k	r-ideal	$\ \nabla F_r(k)\ $	$f(x^k)$	$g(x^k)$	$\ x^k - x^*\ $
0	10.00000	.00056	7.48075	-.68659	.45431
1	2.51131	.11749	4.05720	-.35013	.21362
2	1.46786	.27193	3.33706	-.24038	.12849
3	1.03039	.31172	3.03002	-.18759	.12112
4	.75538	.38673	2.79582	-.14406	.07511
5	.54732	.46257	2.63198	-.11184	.07725
6	.40816	.59636	2.50285	-.08520	.04568
7	.29777	.73830	2.40918	-.06524	.05223
8	.22092	.98650	2.33408	-.04874	.03092
9	.17894	.45921	2.29478	-.04049	.02744
10	.12249	1.21202	2.24071	-.02796	.02001
11	.08989	1.57264	2.21046	-.02099	.02525
12	.07412	.80113	2.19333	-.01726	.00986
13	.05770	1.00625	2.17760	-.01358	.01240
14	.04507	1.29175	2.16494	-.01060	.00790
15	.03478	1.64326	2.15498	-.00825	.01018
16	.02682	2.14358	2.14696	-.00635	.00731
17	.02033	2.78309	2.14064	-.00485	.00895
18	.01528	3.73158	2.13553	-.00363	.00728
19	.01258	1.84623	2.13271	-.00300	.00401
20	.00990	2.35245	2.13002	-.00236	.00297

k	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
0	4.00000	-	.000244	2.287608
1	2.82843	1.872	.053453	2.198046
2	2.00000	1.469	.100506	2.705605
3	2.00000	1.490	.103910	2.999880
4	2.00000	1.334	.110416	3.502524
5	2.00000	1.377	.118678	3.897648
6	2.00000	1.271	.130652	4.564492
7	2.00000	1.352	.144333	5.115267
8	1.41421	1.290	.163178	6.045533
9	2.82843	1.344	.070243	6.537434
10	2.00000	1.115	.152157	7.965589
11	1.41421	1.331	.173523	9.063023
12	2.00000	1.240	.079340	10.097444
13	2.00000	1.108	.088910	11.317519
14	2.00000	1.095	.100209	12.890496
15	2.00000	1.142	.113133	14.525070
16	2.00000	1.144	.128651	16.662057
17	2.00000	1.210	.146833	18.954110
18	1.41421	1.241	.169376	22.031308
19	2.00000	1.252	.076517	24.128257
20	-	1.065	.097498	-

Tabela 6-6

6.4 APBL Versões 2a e 2b

O resultado de maior valor prático seria a implementação eficiente do APBL em sua versão mais completa. Este valor seria consagrado ao mostrar a resolução efetiva de problemas difíceis e mais gerais de que os da estreita classe de problemas convexos. Aqui levamos à conclusão nossos exemplos apenas para tornar concretas algumas observações, e a resolução "eficiente" pelo APBL, Versão 2a e 2b, nada tem de surpreendente. As tabelas A-14 e A-15 exemplificam o processo para problemas 1 e 2 respectivamente. A convergência no problema 2 é tão rápida que logo na nona iteração é alterado o subproblema para

$$\begin{aligned} \min \quad & p^T \nabla f^k + \frac{1}{2} p^T H_k p \\ \text{s.a.} \quad & \nabla g_k p = -0.8 g^k \end{aligned}$$

No problema 1 a passagem da segunda à terceira fase do algoritmo é adiada para a 16^a iteração. Nenhuma tentativa de obter estimativas de ordem superior dos multiplicadores μ^k foi realizada, ié foi aplicado a versão mais simples possível do algoritmo MQS-deslocado. O fator 0.8 foi escolhido sem muita deliberação; a sua escolha se basearia em considerações inerentes aos problemas-objetos a serem resolvidos. Nos casos destes dois problemas, um passo unitário de Newton pode ser tomado e convergência quadrática obtida, por serem problemas relativamente fáceis.

A versão 2b é a ilustrada nas tabelas, sem mostrar diferença significativa da versão 2a além das já mencionadas acima.

6.5 Sobre o Parâmetro \hat{a}

O valor de $a(x)$, cuja definição repetimos aqui por conveniência

$$a(x;r) = \underset{a}{\operatorname{argmin}} \left\{ |D^3 F_r(x)[h,h,h]| \leq 2a^{-1/2} (D^2 F_r(x)[h,h])^{3/2} \right\}$$

foi calculado em vários pontos de S nos casos dos problemas 1 e 2, para diferentes valores de \underline{r} em F_r . Os resultados estão nas tabelas 6.8 e 6.9, respectivamente. Os pontos escolhidos são de dois tipos: pontos selecionados próximo da trajetória central T (\underline{a} e \underline{b}) e pontos "representativos", ié. aleatórios e bem-distribuídos por S (\underline{c} - \underline{g}). Para problema 1, estes pontos já foram usados e são listados na tabela A-1. Para problema 2, os pontos em questão são dados abaixo.

a)	(-0.260, 0.330, 0.114, 0.470)	$r_{id} = 100.0$
b)	(0.047, 0.868, 1.940, -0.736)	$r_{id} = 1.0$
c)	(0, 0, 0, 0)	
d)	(0, 2, 0, 0)	
e)	(0, 1, -0.2, 0.1)	
f)	(0.5, 0, 0.3, 0.2)	
g)	(-0.5, 0.5, -0.5, -0.2)	

Ponto	r		
	1.00	10.00	100.0
a	7.0413	1.8584	1.2844
b	2.1529	1.2167	1.0774
c	1.5122	1.0560	1.0111
d	6.0950	5.1728	1.8043
e	1.2277	1.0414	1.0166
f	2.4828	1.2118	1.0200
g	1.6883	1.0921	1.0313
min	1.2277	1.0414	1.0166

Tabela 6-8

Ponto	r		
	1.00	10.00	100.0
a	765.05	26.811	10.963
b	1.6575	1.4005	1.376
c	95.496	6.274	3.691
d	8.619	2.2167	1.857
e	99.023	6.455	3.745
f	668.70	24.297	11.485
g	9.280	2.1815	1.8002
min	1.6575	1.4005	1.376

Tabela 6-9

Os valores encontrados motivaram o uso de $\underline{a} = 1.0$ em ambos os problemas. Observa-se uma relativa uniformidade de valores e em ambos os casos valores crescentes de $a(x)$ na medida em que \underline{r} é reduzido. Em termos gerais, isso deve ser tributado ao fato de, com redução de \underline{r} , o termo $F(x)$ em $F_r(x) = \frac{1}{r}f(x) + F(x)$, fonte do mau condicionamento de $F_r(x)$, é progressivamente mais dominado numa vizinhança de x^k pelo outro termo. Pontos mais distantes da barreira logarítmica têm também valores mais altos de $a(x)$, r fixo.

7.1 Extensões do APBL

Abordamos algumas extensões do APBL. Quanto à aplicabilidade do APBL, do seu alcance, já nos referimos em seção 3.4 à inclusão de restrições lineares de igualdade $Ax = b$. De maior significado é o fato da trajetória central T da função barreira logarítmica ser bem definida em qualquer domínio convexo se a função-objetivo $f(x)$ for semi-convexa numa vizinhança de T , ou seja, com os autovalores $\{ \mu(x) \}$ não-negativos da matriz hessiana $\nabla^2 F_r(x)$ em todo ponto x de S . Portanto, o problema (P) pode ser estendido para incluir $g_i(x)$ quasiconvexas assim que for estendido o conceito de autoconcordância para este caso. De fato, seja problema (P')

$$\begin{array}{ll}
 \min & f(x) \\
 (P') & \text{s.a. } g_i(x) \leq 0 \quad 1 \leq i \leq m_1 \\
 & Ax = b
 \end{array}$$

onde existe uma região não-vazia $S = \{x: g_i(x) < 0 \quad \forall i\}$, com solução (x^*, μ^*) não-degenerada e seja V uma vizinhança relativa de x^* onde o MQS demonstra convergência super-linear. O algoritmo APBL pode ser estendido a (P') se for possível obter um conceito análogo ao de autoconcordância. Lembremos que este conceito é essencialmente local e pode ser definido para uma função $F(x)$ dentro de uma região onde F é convexa.

Neste sentido há a opção de usar o parâmetro a localmente e estimá-lo por métodos iterativos. A determinação exata das derivadas terceiras é muito dispendiosa, mas elas podem ser estimadas por métodos de diferença finita a

partir da seqüência de matrizes hessianas $\{\nabla^2 F_r(x^k)\}$. Estas matrizes, por sua vez, podem ser ou estimadas por métodos Quasi-Newton ou calculadas exatamente.

Todos os exemplos examinados demonstram suave variação de $a(x^k)$ ao longo do algoritmo, decrescente em direção à trajetória T (exceto em casos extremos com passos muito longos) e novamente decrescente ao longo de T em direção a x^* (sem exceção). Por outro lado, usando funções $g_i(x)$ que são de classe C^3 mas não de classe C^4 , é fácil fabricar exemplos nos quais $a(x)$ varia de forma arbitrariamente rápida por variação de $x \in S$ em torno de subvariedades de codimensão maior ou igual a um.

A extensão mais promissora parece ser o uso de $\lambda(L_\mu, x^k)$ mencionado na seção 3.3. A extensão não é imediata pois o parâmetro μ varia conforme o valor de μ e a família $\{L_\mu\}$ não é geralmente convexa em μ . Enquanto as iterações $\{x^k\}$ da Fase 2 do APBL estiver convergindo linearmente para x^* , as estimativas $\{\mu^*\}^k$ do multiplicador μ^* também convergem linearmente.

Uma vez que se determina como calcular com precisão conhecida o valor de $\lambda(L_{\mu^*}, x^k)$ um passo do seguinte tipo parece atrativo:

$$\begin{aligned} \text{Em } x^k, \text{ resolve-se } & p^T \nabla f^k + \frac{1}{2} p^T H_k p \\ \text{s.a. } & \nabla g_k p = -g^k \end{aligned}$$

e teste αp^k tal que $\alpha = \operatorname{argsup}_\alpha \left\{ g_i(x^k + \alpha p^k) \leq -\epsilon \quad \forall i \right\}$

7.2 Questões em aberto

Todo algoritmo parcialmente heurístico deixa uma miríade de questões em aberto. A primeira destas talvez seja o uso do parâmetro r -ideal na primeira fase do algoritmo. Não temos evidência segura para atestar nossa opinião de que o uso é sempre benéfico ié. existência de um valor C tal que se $r_{id}(B_r, x^0) \geq C$ então a solução de $\text{argmin}_x \{ F_r(x) \}$ será encontrada mais rapidamente usando $r := r_{id}(x^0)$ de que mediante a prefixação de \underline{r} . A necessidade de se aproximar inicialmente à trajetória central T , no caso do APBL, deixa dúvidas quanto à comparação geral dele com algoritmos MQS clássicos, como o de Han [13].

Uma condição possivelmente muito crítica - a de determinar o mais cedo possível o conjunto ativo em x^* , isso é, antes de iniciar a Fase 2 do APBL - ainda não possui classificação quantitativa. Por este e outros motivos citados nos capítulos 4 e 5, não podemos por enquanto assegurar convergência polinomial a não ser em casos triviais que nada demonstram para o caso geral.

Finalmente, frisamos a necessidade de estender e quantificar o conceito de autoconcordância. Sua determinação prática é não-trivial, apesar de não ter apresentado dificuldade em nenhum caso examinado por nós. Não há garantia ainda de não existirem problemas (P) para os quais \underline{a} ou até $\{a(x): x \text{ próximo de } T\}$ tais que \underline{a} seja muito pequeno, e sendo também este fato de averiguação não-trivial. Problemas que detêm o mesmo papel em programação não-linear como aquele demonstrado pelo problema Klee-Minty em relação ao simplex, são mais numerosos que este, porém simultaneamente mais "anormais" que o Klee-Minty. Ainda não há como concluir muito a respeito de sua relevância na programação não-linear.

7.3 Conclusões Gerais

Consideramos demonstrada a superioridade do APBL em relação ao ABL clássico. Mesmo com extensões como as propostas por Lasdon, Fox e Ratner [18] que contornam parcialmente o mau condicionamento da matriz hessiana de B_r , um algoritmo que providencia a passagem a um algoritmo (fase de algoritmo) de convergência superlinear num momento adequado é intrinsecamente superior a um algoritmo "puro" de barreira logarítmica.

A crítica pode ser levantada a respeito do fato do APBL ser parcialmente heurístico. Consideramos que, descontada a perda de elegância e simplicidade em relação a algoritmos "puros" como os de Nesterov e Nemirovsky, o ganho em eficiência e aplicabilidade serve como justificativa mais de que suficiente. Além disso, o APBL representa uma tentativa de "otimizar" a integração de algoritmos distintos. O uso de parâmetros idealizados possibilita a passagem de um método "puro" a outro mais indicado sob condições de garantia de convergência linear e superlinear nas fases distintas do algoritmo.

Acreditamos ter demonstrado ser definitiva a reabilitação do uso da barreira logarítmica iniciada principal e rigorosamente por Nesterov e Nemirovsky em [26]. Pelas propriedades analíticas desta família de funções, há motivo de acreditar que seu uso é quase sempre efetivo na resolução de problemas bastante gerais que demonstram boas condições de convergência superlinear apenas numa vizinhança restrita de (x^*, μ^*) .

REFERENCIAS:

- [1] Bazarra, M.S. e Shetty, C.M. (1979). *Non-Linear Programming*, John Wiley & Sons.
- [2] Bertsekas, D.P. (1982). *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press.
- [3] Den Hertog, P., Roos, C., Terlaky, T. (1990). *A Potential Reduction Method for a Class of Smooth Convex Programming Problems*, Report No. 90-01, Faculty of Mathematics and Informatics/Computer Science, Delft University of Technology, Delft, Holanda.
- [4] Den Hertog, P., Roos, C., Terlaky, T. (1990). *On the Classical Logarithmic Barrier Function Method for a Class of Smooth Convex Programming Problems*, Report No. 90-28, Faculty of Mathematics and Informatics/Computer Science, Delft University of Technology, Delft, Holanda.
- [5] Fiacco, A.V. e McCormick, G.P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons.
- [6] Fletcher, R. (1980). *Practical Methods of Optimization*, John Wiley & Sons.
- [7] Fritsch, R. (1955). *The Logarithmic Potential Method of Convex Programming*, Memorandum of the University Institute of Economics, Noruega.
- [8] Gill, P.E., Murray, W., Wright, M.H. (1981). *Practical Optimization*, Academic Press.
- [9] Gill, P.E., Murray, W., Saunders, M.A., Wright, M.H. (1989). *Constrained Nonlinear Programming* em Nemhauser et al (eds), *Handbook in OR and MS*, Vol. 1, Elsevier Science Publishers/ North-Holland.

- [10] Gonzaga, C.C. (1987). *An algorithm for solving linear programming problems in $O(n^3L)$ operations*. Tech. Report, Dept. of Electrical Engineering and Computer Sciences, Berkley, EUA.
- [11] Gonzaga, C.C. (1988). *Polynomial Affine Algorithms for Linear Programming*, internal report, Programa de Eng. de Sistemas e Computação, COPPE/UFRJ, Rio de Janeiro, Brasil.
- [12] Hall, B.R. (1991). *Self-Concordance and a Parametrized Logarithmic Barrier Algorithm*, documento apresentado no XXX-TIMS, Rio de Janeiro.
- [13] Han, S.P. (1977). *A Globally Convergent Method for Nonlinear Programming*, *Jornal of Optimization Theory & Applications*, Vol. 22, No. 3, pp. 297-309.
- [14] Hock, W. Schittkowski, K. (1980). *Test Examples for Nonlinear Programming Codes*, Springer-Verlag.
- [15] Jahre, F. (1989). *The Method of Analytic Centers for Smooth Convex Programming*, Dissertação, Institut für Angewandte Mathematic und Statistik, Universität Würzburg, Würzburg, Alemanha.
- [16] Karmarkar, N. (1984). *A new polynomial-algorithm for linear programming*, *Combinatorica* 4, pp. 373-395.
- [17] Kortanek, K.O., Potra, F., Ye, Y. (1989). *On some efficient interior point methods for nonlinear convex programming*, Working Paper Series No. 89-24, University of Iowa, Iowa City, EUA.
- [18] Lasdon, L.S., Fox, R.L. e Ratner, M. (1973). *An efficient one-dimensional search procedure for barrier functions*, *Mathematical Programming* 4, pp. 279-296.
- [19] Lootsma, F.A. (1969). *Hessian matrices of penalty functions for solving constrained optimization problems*, *Philips Research Reports* 24, pp. 322-331.

- [20] Lootsma, F.A. (1970). *Boundry properties of penalty functions for constrained optimization problems*, Philips Research Reports Suppl. 3.
- [21] Mehrotra, S., Sun, J. (1987). *An algorithm for convex quadratic programming that requires $O(n^{3.5}L)$ arithmetic operations*, Technical Report 87-24, Northwestern University, Illinois, EUA.
- [22] Monteiro, R.C., Adler, I. (1989). *An $O(n^3L)$ primal-dual interior point algorithm for convex programming*, Mathematical Programming 44, pp. 27-42.
- [23] Murray, W. (1971). *Analytic expressions for the eigenvalues and eigenvectors of the hessian matrices of barrier and penalty functions*, J. Optimization and Applications 7, pp. 189-196.
- [24] Murray, W. e Wright, M.H. (1978). *Projected Lagrangian methods based on the trajectories of penalty and barrier functions*, Report SOL-78-23, Dept. Operations Research, Stanford University, EUA.
- [25] Nemirovsky A.S. e Yudin, D.B. (1979). *Problem Complexity and Method Efficiency in Optimization*, traduzido por E.R. Dawson (1983), John Wiley & Sons.
- [26] Nesterov, Yu.E. e Nemirovsky, A.S. (1989). *Self-Concordant Functions and Polinomial Time Methods in Convex Programming*, Moscou.
- [27] Ortega, J.M., Rheinboldt, W.C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press.
- [28] Powell, M.J.D. (1972). *Problems relating to unconstrained optimization*, em Murray, W. (ed) *Numerical Methods for Unconstrained Optimization*, Academic Press, pp. 29-55.

- [29] Renegar, J. (1988). *A polynomial time algorithm based on Newton's method for linear programming*, Mathematical Programming 40, pp. 59-93.
- [30] Sonnevend, G., Stoer, J., Zhao, G. (1989). *On the Complexity of following the Central Path of Linear Programs by Linear Extrapolation*, comunicação dos autores Universität Würzburg, Würzburg, Alemanha.
- [31] Wilson, R.B. (1963). *A Simplicial Algorithm for Concave Programming*, Tese PhD, Harvard University, EUA.
- [32] Wright, M.H. (1976). *Numerical Methods for Nonlinearly Constrained Optimization*, Tese PhD, Stanford University, EUA.
- [33] Vaidya, P.M. (1987). *An Algorithm for Linear Programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ Arithmetic Operations*, preprint, AT&T Bell Laboratories, Murray Hill, EUA.

APÊNDICE

$x^0 = (5, 5, 2)$	$g(x^0) = (-10.575, -0.75)$
$x^0 = (1, 5, 1)$	$g(x^0) = (- 4.025, -3.75)$
$x^0 = (-5, 10, 1)$	$g(x^0) = (- 0.525, -8.75)$
$x^0 = (-5, 5, 2)$	$g(x^0) = (- 0.575, -0.75)$
$x^0 = (5, 17, -4)$	$g(x^0) = (- 1.275, -0.75)$

Tabela A-1

Problema 1

Iter.	r_{id}	$\ \nabla F_{r_{id}}\ $	r_{pr}	$\ \nabla F_{r_{pr}}\ $	$\ x^k - x^*\ $
0	50.0000	.00103	50.0000	.00103	4.930456
1	41.1850	.00684	45.0000	.04418	4.425550
2	33.9412	.01424	40.5000	.09374	3.973013
3	28.0125	.02307	36.4500	.14896	3.566869
4	23.1552	.03348	32.8050	.21046	3.202359
5	19.1726	.04566	29.5245	.27897	2.875232
6	15.9032	.05982	26.5720	.35525	2.581665
7	13.2166	.07618	23.9148	.44018	2.318222
8	11.0056	.09498	21.5234	.53473	2.081813
9	9.1838	.11649	19.3710	.63996	1.869662
10	7.6805	.14101	17.4339	.75706	1.679273
11	6.4383	.16884	15.6905	.88736	1.508407
12	5.4099	.20035	14.1215	1.03232	1.355053
13	4.5570	.23592	12.7093	1.19356	1.217405
14	3.8484	.27597	11.4384	1.37290	1.093846
15	3.2585	.32098	10.2946	1.57234	0.982922
16	2.7665	.37148	9.2651	1.79410	0.883332
17	2.3551	.42803	8.3386	2.04067	0.793907
18	2.0103	.49129	7.5047	2.31479	0.713601
19	1.7207	.56196	6.7543	2.61953	0.641475
26	.8230	1.05758	3.9883	4.73515	0.376991
31	.4185	1.90376	2.3551	8.32140	0.221921
36	.2235	3.34039	1.3906	14.39680	0.130787
41	.1238	5.77505	0.8212	24.68752	0.077136

Tabela A-2

k	$g_1(x^k)$	$g_2(x^k)$	μ_1^*	μ_2^*	$\bar{\mu}_1^*$	$\bar{\mu}_2^*$
0	-5.7508	-2.9194	8.694	17.127	-	-
1	-5.1757	-2.6274	7.957	15.675	4.514	10.343
2	-4.6581	-2.3646	7.286	14.354	4.142	8.164
3	-4.1923	-2.1281	6.682	13.163	3.828	7.536
4	-3.7731	-1.9152	6.137	12.090	3.539	6.974
5	-3.3958	-1.7236	5.650	11.124	3.305	6.474
6	-3.0562	-1.5512	5.204	10.252	3.035	6.010
7	-2.7506	-1.3961	4.805	9.467	2.842	5.605
8	-2.4755	-1.2564	4.446	8.760	2.659	5.241
9	-2.2280	-1.1307	4.122	8.122	2.489	4.906
14	-1.3156	-0.66757	2.925	5.765	1.882	3.714
19	-0.86316	-0.43794	1.993	3.929	1.380	2.728
24	-0.50969	-0.25857	1.862	3.670	1.352	2.664
29	-0.30096	-0.15267	1.585	3.124	1.260	2.484
34	-0.17772	-0.09014	1.421	2.802	1.214	2.292
39	-0.10494	-0.04790	1.325	2.658	1.198	1.989

Tabela A-3

μ é a estimativa $-r_{id}/g$; $\bar{\mu}$ é a extrapolação linear.

Problema 2

k	r-ideal	$\ \nabla_{r_{id}} F_r(k)\ $	r_{prev}	$\ \nabla_{r_{pr}} F_r(k)\ $	$\ x^k - x^*\ $
0	10.27958	.00442	10.27958	.00442	1.009430
1	9.40335	.01943	9.76560	.07334	.945616
2	8.63444	.03769	9.27732	.14734	.885780
3	7.95158	.05237	8.81345	.22258	.830386
4	7.34049	.06453	8.37278	.30003	.779293
5	6.79071	.07506	7.95414	.38034	.732158
6	6.29398	.08460	7.55644	.46391	.688585
7	5.84367	.09355	7.17861	.55088	.648197
8	5.43415	.10219	6.81968	.64208	.610667
9	5.06067	.11070	6.47870	.73721	.575713
10	4.71914	.11923	6.15476	.83671	.543095
11	4.40593	.12787	5.84702	.94086	.512605
12	4.11823	.13671	5.55467	1.04995	.484062
13	3.85327	.14580	5.27694	1.16427	.457308
14	3.60880	.15519	5.01309	1.28411	.432203
15	3.38285	.16495	4.76244	1.40982	.408622
16	3.17360	.17511	4.52431	1.54171	.386453
17	2.97962	.18572	4.29810	1.68015	.365595
18	2.79947	.19680	4.08319	1.82551	.345957
19	2.63191	.20841	3.87903	1.97817	.327456
20	2.47591	.22058	3.68508	2.13852	.310015
25	1.83853	.29115	2.85145	3.07168	.236478
30	1.37940	.38180	2.20639	4.27120	.18105
35	1.04277	.49873	1.70727	5.81690	.13899
39	.83709	.61637	1.39058	7.37037	.11265

Tabela A-4

k	$g_1(x^k)$	$g_2(x^k)$	μ_1	μ_2	$\bar{\mu}_1$	$\bar{\mu}_2$
0	-4.8439	-3.3927	2.122	3.030	-	-
1	-4.5975	-3.2210	2.045	2.919	1.219	1.728
2	-4.3639	-3.0583	1.979	2.823	1.238	1.745
3	-4.1426	-2.9040	1.919	2.738	1.104	1.748
4	-3.9328	-2.7577	1.866	2.662	1.229	1.749
5	-3.7339	-2.6188	1.819	2.593	1.238	1.741
6	-3.5433	-2.4869	1.776	2.531	1.231	1.745
7	-3.3664	-2.3617	1.735	2.474	1.203	1.734
8	-3.1966	-2.2429	1.700	2.423	1.236	1.746
9	-3.0356	-2.1301	1.667	2.376	1.220	1.739
10	-2.8827	-2.0230	1.637	2.333	1.222	1.739
20	-1.7217	-1.2087	1.438	2.048	1.216	1.732
30	-1.0295	-0.7229	1.340	1.908	1.216	1.731
40	-0.6484	-0.4554	1.291	1.838	1.217	1.731

Tabela A-5

μ é a estimativa $-r_{id}/g$; $\bar{\mu}$ é a extrapolação linear.

Problema 3 ($\rho = 2.5$)

Iteração	r_{id}	$\ \nabla F_{r_{id}}\ $	r_{pr}	$\ \nabla F_{r_{pr}}\ $	$\ x^k - x^*\ $
0	10.000000	.000127	10.000000	.000127	.27126
1	8.346680	.000666	9.000000	.264818	.23935
2	7.049508	.000644	8.100000	.545751	.21185
3	6.009483	.000564	7.290000	.847342	.18798
4	5.161766	.000492	6.561000	1.173894	.16714
5	4.461259	.000433	5.904900	1.529682	.14887
6	3.875882	.000382	5.314410	1.919095	.13278
7	3.382114	.000338	4.782970	2.346832	.11857
8	2.962306	.000301	4.304670	2.817838	.10599
9	2.602956	.000268	3.874200	3.337559	.09482
10	2.293559	.000239	3.486780	3.911924	.08490
11	2.025823	.000213	3.138110	4.547343	.07606
12	1.793115	.000191	2.824300	5.251031	.06819
13	1.590071	.000171	2.541870	6.030873	.06116
14	1.412310	.000153	2.287680	6.895420	.05488
15	1.256217	.000137	2.058910	7.854551	.04927
20	0.710938	.000079	1.215770	14.484550	.02887
25	0.409655	.000048	0.717900	19.532718	.01703
30	0.238476	.000028	0.423910	34.268292	.01011

Tabela A-6

Problema 3 ($\rho = 1.5$)

Iteração	r_{id}	$\ \nabla F_{r_{id}}\ $	r_{pr}	$\ \nabla F_{r_{pr}}\ $	$\ x^k - x^*\ $
0	10.000000	.000457	10.00000	.000457	.45435
1	7.448944	.003934	9.00000	.348723	.39326
2	5.788368	.003721	8.10000	.695913	.34313
3	4.624092	.002924	7.29000	1.055273	.30120
4	3.767774	.002293	6.56100	1.435719	.26559
5	3.116292	.001843	5.90490	1.844285	.23498
6	2.608041	.001513	5.31441	2.287249	.20845
7	2.203665	.001261	4.78297	2.770669	.18531
8	1.876733	.001064	4.30467	3.300680	.16501
9	1.608967	.000906	3.87420	3.883708	.14715
10	1.387242	.000777	3.48678	4.526618	.13137
11	1.201904	.000683	3.13811	5.236835	.11741
12	1.045758	.000594	2.82430	6.022440	.10501
13	0.913285	.000519	2.54187	6.892320	.09400
14	0.800165	.000454	2.28768	7.856251	.08419
15	0.703064	.000398	2.05891	8.922584	.07545
20	0.380461	.000214	1.21577	16.307994	.04388
25	0.213534	.000121	0.71790	24.509056	.02568
30	0.122390	.000068	0.42391	42.693052	.01509

Tabela A-7

Problema 3 ($\rho = 1.1$)

Iteração	r_{id}	$\ \nabla F_{r_{id}}\ $	r_{pr}	$\ \nabla F_{r_{pr}}\ $	$\ x^k - x^*\ $
0	10.000000	.000952	10.00000	.000952	.56223
1	6.375596	.007149	9.00000	.445098	.47545
2	4.467332	.006753	8.10000	.859106	.40913
3	3.287939	.004898	7.29000	1.275839	.35572
4	2.496261	.003492	6.56100	1.711205	.31148
5	1.936961	.002595	5.90490	2.175594	.27412
6	1.527943	.002009	5.31441	2.677268	.24215
7	1.221120	.001603	4.78297	3.223689	.21453
8	.986468	.001307	4.30467	3.822144	.19050
9	.804220	.001083	3.87420	4.480087	.16948
10	.660884	.000909	3.48678	5.205396	.15100
11	.546918	.000769	3.13811	6.006529	.13472
12	.455482	.000656	2.82430	6.892663	.12032
13	.381529	.000562	2.54187	7.873856	.10756
14	.321284	.000485	2.28768	8.961173	.09623
15	.271883	.000421	2.05891	10.166801	.08615
20	.125646	.000217	1.21577	18.495907	.04994
25	.063056	.000118	0.71790	28.809474	.02918
30	.033537	.000067	0.42391	50.065905	.01713

Tabela A-8

Versão 1a

k	r-ideal	$\ \nabla F_r(k)\ $	$f(x^k)$	$g_1(x^k)$	$g_2(x^k)$	$\ x^k - x^*\ $
1	15.0000	.04653	18.99271	-3.04313	-1.44621	2.491731
2	8.1395	.05112	11.57125	-2.18241	-1.03389	1.818395
3	4.8230	.09248	7.41442	-1.56513	-.74013	1.216941
4	3.9746	.08192	6.46892	-1.40862	-.66439	1.137505
5	2.3639	.18778	4.43415	-1.01020	-.47614	.813296
6	1.4503	.32358	3.21422	-.72447	-.34137	.575926
7	.9065	.56255	2.46068	-.51956	-.24456	.419669
8	.5905	.79376	1.98419	-.37261	-.17538	.285898
9	.3801	1.37479	1.67382	-.26722	-.12511	.234219
10	.2637	1.54545	1.47165	-.19164	-.08864	.135956
11	.2253	1.71698	1.41125	-.17247	-.07916	.134743
12	.2247	.02802	1.41125	-.18624	-.07634	.138923
13	.1892	2.38626	1.34858	-.14613	-.06834	.115069
14	.1593	2.82074	1.29708	-.12547	-.05868	.098907
15	.1345	3.32029	1.25368	-.10772	-.05038	.084592
16	.1137	3.91636	1.21701	-.09249	-.04325	.073147
17	.0965	4.58044	1.18598	-.07941	-.03714	.061795
18	.0818	5.41715	1.15965	-.06818	-.03188	.054932
19	.0698	6.24990	1.13730	-.05854	-.02737	.043945
20	.0590	7.48875	1.11831	-.05026	-.02347	.044650
25	.0410	11.05573	1.08412	-.03512	-.01651	.027505
30	.0280	16.15528	1.05994	-.02434	-.01143	.023617
35	.0202	21.22526	1.04497	-.01770	-.00826	.013655

k	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
1	2.828	-	.030066	1.547488
2	2.828	1.6170	.032531	1.571509
3	1.000	1.4406	.098940	.934738
4	2.828	1.7591	.027648	2.963054
5	2.828	1.4327	.047144	3.983021
6	2.828	1.3664	.061707	5.243816
7	2.828	1.3256	.077303	7.277162
8	2.828	1.2324	.098771	8.036332
9	2.828	1.2599	.106542	12.903745
10	1.000	1.0824	.224015	6.898878
11	reapprox	-	.248877	6.898878
12	2.828	-	.001437	19.497748
13	2.828	1.1176	.100599	23.720601
14	2.828	1.1169	.099033	28.482779
15	2.828	1.1001	.104064	31.906285
16	2.828	1.0916	.102384	38.251665
17	2.828	1.0722	.106518	43.001376
18	2.828	1.0772	.105637	51.280986
19	2.828	1.0411	.106943	58.441228
20	2.828	1.0882	.109986	68.088022
25	2.828	1.0336	.111264	99.364631
30	2.828	1.0951	.115330	140.078087
35	2.828	1.0145	.107788	196.916113

Tabela A-9

Posição Final: (0.5177370 0.5029644 0.4972724)

Problema 1		Versão 1b				
k	r-ideal	$\ \nabla F_r(k)\ $	$f(x^k)$	$g_1(x^k)$	$g_2(x^k)$	$\ x^k - x^*\ $
1	15.0000	.04653	18.99271	-3.04313	-1.44621	2.491731
2	10.4333	.00860	14.15211	-2.50636	-1.19043	2.056014
3	7.6517	.01753	10.89724	-2.09004	-.99250	1.707892
4	5.6954	.04881	8.56344	-1.75097	-.83135	1.425163
5	4.2979	.08720	6.86251	-1.47310	-.69932	1.194714
6	3.2880	.13342	5.60657	-1.24466	-.59079	1.006234
7	2.5494	.18822	4.66697	-1.05620	-.50128	.851472
8	2.0026	.25238	3.95473	-.90011	-.42715	.723849
9	1.5927	.32672	3.40786	-.77032	-.36552	.618129
10	1.2815	.41217	2.98263	-.66190	-.31405	.530130
11	1.0424	.50972	2.64808	-.57097	-.27089	.456547
12	.8565	.62050	2.38180	-.49434	-.23452	.394710
13	.7102	.74580	2.16751	-.42947	-.20373	.342479
14	.5939	.88698	1.99337	-.37431	-.17756	.298167
15	.5004	1.04562	1.85052	-.32721	-.15521	.260402
16	.4245	1.22349	1.73231	-.28682	-.13605	.228072
17	.3624	1.42259	1.63370	-.25205	-.11955	.200276
18	.3111	1.64510	1.55085	-.22201	-.10530	.176290
19	.2683	1.89353	1.48076	-.19595	-.09294	.155511
20	.2325	2.17061	1.42113	-.17328	-.08218	.137451
25	.1196	4.10622	1.22685	-.09588	-.04547	.075921
30	.0652	7.40205	1.12887	-.05447	-.02583	.043094
35	.0367	12.99392	1.07615	-.03144	-.01491	.024860

k	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
1	2.000	-	.030066	1.547488
2	2.000	3.0445	.004576	1.878389
3	2.000	2.6661	.007941	2.207522
4	2.000	2.5566	.019673	2.481173
5	2.000	2.4537	.030409	2.867443
6	2.000	2.3497	.040373	3.304695
7	2.000	2.2463	.049503	3.802254
8	2.000	2.1449	.057790	4.367097
9	2.000	2.0470	.065247	5.007391
10	2.000	1.9540	.071914	5.731401
11	2.000	1.8658	.077847	6.547684
12	2.000	1.7833	.083119	7.465149
13	2.000	1.7080	.087778	8.496353
14	2.000	1.6379	.091900	9.651565
15	2.000	1.5739	.095544	10.943780
16	2.000	1.5161	.098758	12.388755
17	2.000	1.4637	.101600	14.001934
18	2.000	1.4161	.104115	15.800790
19	2.000	1.3738	.106335	17.807249
20	2.000	1.3355	.108296	20.043186
25	2.000	1.1949	.111526	36.818625
30	2.000	1.1138	.115446	64.116791
35	2.000	1.0665	.117683	110.414184

Tabela A-10

Posição Final: (0.5106097 0.5097900 0.5011028)

Problema 2 Caso $m = 3$

$$x^0 = (-.06554662 \quad .55942341 \quad 1.13168937 \quad -.05112715)$$

k	r-ideal	$g_1(x^k)$	$g_2(x^k)$	$g_3(x^k)$	$\ x^k - x^*\ $
0	11.7381	-5.84157	-4.03712	-7.96717	1.36115
1	9.5385	-5.34046	-3.68853	-7.27582	1.19156
2	9.5129	-5.32134	-3.62788	-7.53158	1.22191
3	7.9466	-4.87087	-3.31949	-6.88867	1.07614
4	7.3602	-4.67182	-3.18408	-6.60611	1.01505
5	7.3258	-4.65918	-3.10888	-6.94057	1.05875
6	6.2759	-4.26936	-2.84816	-6.35659	.93734
7	6.0691	-4.18314	-2.79081	-6.22797	.90980
8	5.8711	-4.09877	-2.73458	-6.10218	.88469
9	5.8475	-4.09539	-2.67078	-6.40771	.92972
10	5.0897	-3.75521	-2.44844	-5.87347	.82608
11	4.8767	-3.64725	-2.37886	-5.70393	.78721
12	4.8002	-3.61060	-2.35488	-5.64664	.78011
13	4.7828	-3.61335	-2.30008	-5.93077	.82509
14	4.2077	-3.31470	-2.10922	-5.43960	.73594
15	4.0978	-3.24597	-2.06683	-5.32604	.70058
16	4.0595	-3.22904	-2.05606	-5.29827	.70528
17	4.0482	-3.23535	-2.01285	-5.54283	.74612
18	3.5864	-2.96880	-1.84592	-5.08588	.66806
19	3.5309	-2.91966	-1.81783	-5.00101	.62599
20	3.4964	-2.90866	-1.81127	-4.98239	.64233
21	3.4878	-2.91793	-1.77482	-5.20595	.68130

k	$f(x^k)$	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
0	-23.71144	1.000	.000	.00010	1.251377
1	-26.48191	reaprox	-	.13855	1.251377
2	-26.41892	2.000	-	.00076	1.425507
3	-28.66919	1.000	2.058	.11511	1.469021
4	-29.56687	reaprox	-	.16781	1.469021
5	-29.49820	4.000	-	.00130	1.494067
6	-31.23467	1.000	1.791	.10720	1.582174
7	-31.59221	1.000	1.648	.13005	1.596173
8	-31.93442	reaprox	-	.15437	1.596173
9	-31.85977	8.000	-	.00113	1.567169
10	-33.24769	2.828	1.620	.10278	1.702970
11	-33.65898	1.000	1.480	.13534	1.693636
12	-33.79713	reaprox	-	.14849	1.693636
13	-33.72205	16.000	-	.00103	1.678212
14	-34.85843	4.000	1.503	.10050	1.834891
15	-35.10082	1.000	1.307	.12957	1.722228
16	-35.16474	reaprox	-	.13651	1.722228
17	-35.09351	32.000	-	.00105	1.918577
18	-36.05621	5.657	1.426	.09984	1.961880
19	-36.20390	1.000	1.095	.15366	1.666631
20	-36.25874	reaprox	-	.13937	1.666631
21	-36.19060	64.000	-	.00138	2.035303

Tabela A-11

Posição Final: (0.050338 0.785840 1.699318 -0.559182)

Caso $\rho = 2.5$

k	r-ideal	$\ \nabla F_r(k)\ $	$f(x^k)$	$g(x^k)$	$\ x^k - x^*\ $
0	10.00000	.00057	24.06733	-.46576	.27127
1	4.25288	.01535	20.29968	-.26309	.14370
2	2.18518	.06356	18.64420	-.15377	.08048
3	1.19807	.31238	17.78200	-.09065	.05465
4	.91853	.39050	17.52005	-.07068	.03876
5	.70059	.48512	17.31892	-.05505	.03725
6	.53553	.62149	17.15915	-.04245	.02657
7	.40493	.79099	17.03533	-.03257	.02745
8	.30463	1.04299	16.93634	-.02460	.02067
9	.22470	1.37390	16.85919	-.01834	.02230
10	.18434	.69375	16.81655	-.01505	.00989
11	.12580	1.83128	16.76023	-.01035	.01588
12	.10329	.92669	16.73635	-.00849	.00658
13	.07063	2.45457	16.70471	-.00583	.01158
14	.05798	1.24350	16.69127	-.00479	.00458
15	.03967	3.30238	16.67346	-.00328	.00858
16	.03255	1.67395	16.66588	-.00269	.00329
17	.02226	4.45545	16.65586	-.00185	.00643
18	.01825	2.25993	16.65158	-.00151	.00241
19	.01247	6.02709	16.64595	-.00103	.00486
20	.01021	3.05942	16.64353	-.00085	.00178

k	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
0	4.00000	-	.000291	1.951239
1	4.00000	1.437	.006500	2.362095
2	4.00000	1.215	.020858	3.047169
3	2.00000	1.129	.079729	3.918052
4	2.00000	1.167	.087614	4.457004
5	2.00000	1.186	.096955	5.003545
6	2.00000	1.178	.108468	5.729739
7	2.00000	1.219	.121964	6.485434
8	2.00000	1.238	.138886	7.509641
9	1.41421	1.312	.159266	8.626423
10	2.82843	1.270	.072536	9.564186
11	1.41421	1.123	.159507	11.480860
12	2.82843	1.265	.072870	12.717048
13	1.41421	1.118	.160536	15.289830
14	2.82843	1.266	.073460	16.927606
15	1.41421	1.117	.162035	20.380700
16	2.82843	1.270	.074207	22.557814
17	1.41421	1.117	.163847	27.192812
18	2.82843	1.276	.075087	30.097637
19	1.41421	1.119	.165934	36.322227
20	-	1.283	.084230	-

Tabela A-12

Caso $\rho = 1.1$

k	r-ideal	$\ \nabla F_r(k)\ $	$f(x^k)$	$g(x^k)$	$\ x^k - x^*\ $
0	10.00000	.00267	3.65137	-.79544	.56195
1	1.10394	.25535	.76668	-.35125	.22576
2	.68033	.26509	.52961	-.26851	.14488
3	.43401	.27515	.39551	-.21039	.13097
4	.29333	.30782	.30449	-.16343	.08537
5	.20010	.33946	.24589	-.12857	.08074
6	.14206	.39666	.20375	-.10031	.05162
7	.10098	.45619	.17476	-.07886	.05182
8	.07395	.55090	.15313	-.06149	.03202
9	.05394	.65494	.13762	-.04816	.03476
10	.04022	.81306	.12574	-.03736	.02102
11	.02972	.99476	.11697	-.02903	.02476
12	.02228	1.26794	.11013	-.02228	.01550
13	.01646	1.59720	.10499	-.01706	.01904
14	.01223	2.10025	.10093	-.01283	.01328
15	.00997	1.00039	.09873	-.01066	.00906
16	.00683	2.62364	.09575	-.00739	.00942
17	.00562	1.26769	.09455	-.00615	.00593
18	.00389	3.32428	.09287	-.00428	.00689
19	.00322	1.61995	.09219	-.00357	.00400
20	.00224	4.24595	.09123	-.00249	.00513

k	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
0	4.0000	.000	.001755	1.521744
1	2.0000	2.224	.111445	2.291235
2	2.0000	1.919	.099037	2.676662
3	2.0000	1.810	.092083	2.988039
4	2.0000	1.621	.089376	3.444141
5	2.0000	1.589	.088274	3.845570
6	2.0000	1.450	.089953	4.409693
7	2.0000	1.446	.092559	4.928701
8	2.0000	1.338	.097601	5.644419
9	2.0000	1.353	.103589	6.322470
10	2.0000	1.272	.112117	7.251867
11	2.0000	1.306	.121968	8.155902
12	2.0000	1.251	.134894	9.399486
13	2.0000	1.307	.149982	10.649247
14	1.4142	1.286	.169440	12.395230
15	2.8284	1.305	.074068	13.506338
16	1.4142	1.113	.160792	16.316932
17	2.8284	1.248	.071269	17.787405
18	1.4142	1.092	.155176	21.422645
19	2.8284	1.213	.069336	23.363905
20	-	1.080	.181731	23.363905

Tabela A-13

Problema 1 Versão 2b

$$x^0 = (2.409551, 2.095300, 0.631738)$$

k	r-ideal	$\ \nabla F_r(x^k)\ $	$f(x^k)$	$g_1(x^k)$	$g_2(x^k)$	$\ x^k - x^*\ $
1	15.0000	.04653	18.99271	-3.04313	-1.44621	2.491731
2	10.4333	.00860	14.15211	-2.50636	-1.19043	2.056014
3	7.6517	.01753	10.89724	-2.09004	-.99250	1.707892
4	5.6954	.04881	8.56344	-1.75097	-.83135	1.425163
5	4.2979	.08720	6.86251	-1.47310	-.69932	1.194714
6	3.2880	.13342	5.60657	-1.24466	-.59079	1.006234
7	2.5494	.18822	4.66697	-1.05620	-.50128	.851472
8	2.0026	.25238	3.95473	-.90011	-.42715	.723849
9	1.5927	.32672	3.40786	-.77032	-.36552	.618129
10	1.2815	.41217	2.98263	-.66190	-.31405	.530130
11	1.0424	.50972	2.64808	-.57097	-.27089	.456547
12	.8565	.62050	2.38180	-.49434	-.23452	.394710
13	.7102	.74580	2.16751	-.42947	-.20373	.342479
14	.5939	.88698	1.99337	-.37431	-.17756	.298167
15	.5004	1.04562	1.85052	-.32721	-.15521	.260402
16	.4245	1.22349	1.73231	-.28682	-.13605	.228072
17	.0686	6.77621	1.13518	-.05736	-.02705	.044703
18	.0131	34.54978	1.03138	-.01147	-.00540	.009034
19	.0026	173.42657	1.01125	-.00229	-.00108	.001812
20	.0005	867.81085	1.00725	-.00046	-.00022	.000362

k	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
1	2.000	.0000	.030066	1.547488
2	2.000	3.0445	.004576	1.878389
3	2.000	2.6661	.007941	2.207522
4	2.000	2.5566	.019673	2.481173
5	2.000	2.4537	.030409	2.867443
6	2.000	2.3497	.040373	3.304695
7	2.000	2.2463	.049503	3.802254
8	2.000	2.1449	.057790	4.367097
9	2.000	2.0470	.065247	5.007391
10	2.000	1.9540	.071914	5.731401
11	2.000	1.8658	.077847	6.547684
12	2.000	1.7833	.083119	7.465149
13	2.000	1.7080	.087778	8.496353
14	2.000	1.6379	.091900	9.651565
15	2.000	1.5739	.095544	10.943780
16	-	1.5161	.098758	12.388755
17	-	16.7677	.114649	59.103918
18	-	16.1928	.114482	301.793482
19	-	16.0399	.115211	1505.293276
20	-	16.0080	.115354	7523.025603

Tabela A-14

Posição Final: (0.5000107 0.5000097 0.5000011)

Problema 2 Versão 2b

$$x^0 = (-.17114630 \quad .82345745 \quad 1.31899470 \quad -.29663456)$$

k	r-ideal	$\ \nabla F_r(x^k)\ $	$g_1(x^k)$	$g_2(x^k)$	$\ x^k - x^*\ $
0	10.27958	.00442	-4.84386	-3.39270	1.009430
1	4.87299	.26456	-2.90895	-2.08658	.558633
2	4.84408	.00155	-3.00677	-2.04638	.581778
3	2.46174	.18130	-1.73218	-1.19634	.316351
4	1.45461	.32369	-1.08898	-.75455	.193513
5	1.44872	.00409	-1.14648	-.72534	.216526
6	.91141	.17322	-.74489	-.47301	.138078
7	.59081	.39855	-.49266	-.31346	.090250
8	.44776	.59348	-.37689	-.23992	.068646
9	.08451	4.33526	-.07229	-.04677	.013078
10	.01672	22.66453	-.01435	-.00931	.002577
11	.00334	114.28788	-.00287	-.00186	.000514
12	.00067	572.39941	-.00057	-.00037	.000103
13	.00013	2862.95898	-.00011	-.00007	.000021
14	.00003	14315.77832	-.00002	-.00001	.000004
15	.00001	71582.00000	.00000	.00000	.000001

k	$f(x^k)$	Extens.	Red R	$\lambda(F_r, x^k)$	R(r)
0	-28.76203	1.000	-	.00350	1.262105
1	-35.74538	reaprox	-	.20962	1.262105
2	-35.74088	4.000	-	.00099	1.569867
3	-39.51763	1.000	1.537	.08753	2.071312
4	-41.26565	reaprox	-	.15627	2.071312
5	-41.25905	8.000	-	.00172	2.377562
6	-42.25007	8.000	1.159	.04719	3.670686
7	-42.85503	5.657	1.099	.07991	4.987607
8	-43.12852	1.000	1.070	.09644	6.153558
9	-43.83363	1.000	20.281	.16384	26.460810
10	-43.96701	1.000	20.054	.17851	126.966745
11	-43.99342	1.000	20.011	.18148	629.767460
12	-43.99868	1.000	20.002	.18209	3143.419445
13	-43.99974	1.000	20.000	.18222	15711.626749
14	-43.99995	1.000	20.000	.18224	78552.901931
15	-43.99999	1.000	20.001	.18226	392751.778207

Posição Final: -.0000001 1.0000000 1.9999999 -.9999998

Tabela A-15