



1150068589



IMECC

T/UNICAMP M523a

Aceleração quase-Newton para problemas de minimização com restrições

Luziane Ferreira de Mendonça

Profa. Dra. Véra Lucia da Rocha Lopes
Orientadora

Prof. Dr. José Mario Martínez
Co-Orientador

Tese apresentada ao Instituto de
Matemática, Estatística e Computação
Científica, UNICAMP, como requisito
parcial para a obtenção do título de
Doutor em Matemática Aplicada.

IMECC - UNICAMP
Abril de 2006

**I.M.E.C.C.
BIBLIOTECA**

Aceleração quase-Newton para problemas de minimização com restrições

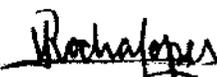
Luziane Ferreira de Mendonça

Este exemplar corresponde à redação final da Tese devidamente corrigida e defendida por Luziane Ferreira de Mendonça e aprovada pela Comissão Julgadora.

Banca Examinadora

Profa. Dra. Véra Lucia da Rocha Lopes
Prof. Dr. Lúcio Tunes dos Santos
Prof. Dr. Roberto Andreani
Prof. Dr. Angel Ramon Sanchez Delgado
Prof. Dr. Ernesto G. Birgin

Campinas, 05 de Abril de 2006


Profa. Dra. Véra Lucia da Rocha Lopes


Prof. Dr. José Mario Martínez

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para a obtenção do título de Doutor em Matemática Aplicada.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecária: Maria Júlia Milani Rodrigues – CRB8a / 2116

Mendonça, Luziane Ferreira de

M523a Aceleração quase-Newton para problemas de minimização com restrições / -- Campinas, [S.P. :s.n.], 2006.

Orientadores : Véra Lucia da Rocha Lopes; José Mario Martínez
Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

I. Sistemas não-lineares. 2. Métodos numéricos. 3. Otimização. 4. Equações de Lagrange. I. Lopes, Véra Lucia da Rocha. II. Martínez, José Mario. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Título em inglês: Quasi-Newton acceleration for constrained minimization problems

Palavras-chave em inglês (Keywords): 1. Nonlinear systems. 2. Numerical methods. 3. Optimization. 4. Lagrange equations.

Área de concentração: Otimização – Métodos numéricos

Titulação: Doutora em Matemática Aplicada

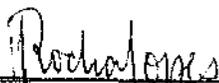
Banca examinadora: Profa. Dra. Véra Lucia da Rocha Lopes (IMECC-UNICAMP)
Prof. Dr. Lúcio Tunes dos Santos (IMECC-UNICAMP)
Prof. Dr. Roberto Andreani (IMECC-UNICAMP)
Prof. Dr. Angel Ramón Sanchez Delgado (ICE-UFRRJ)
Prof. Dr. Ernesto Julián Goldberg Birgin (IME-USP)

Data da defesa: 05/04/2006

UNIDADE IMECC
Nº CHAMADA 7/UNICAMP
M523a IM/H 2445
V EX
TOMBO BC/ 68529
PROC 16.123-06
C D X
PREÇO 11,00
DATA 24/05/06
Nº CPD 13181D-379955

Tese de Doutorado defendida em 05 de abril de 2006 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



Prof. (a). Dr. (a). VERA LÚCIA DA ROCHA LOPES



Prof. (a). Dr. (a). LÚCIO TUNES DOS SANTOS



Prof. (a). Dr. (a). ROBERTO ANDREANI



Prof. (a). Dr. (a). ERNESTO JULIÁN GOLDBERG BIRGIN



Prof. (a). Dr. (a). ANGEL RAMÓN SANCHEZ DELGADO

*“And love is not the easy thing
The only baggage you can bring...
And the love is not the easy thing
The only baggage you can bring...
Is all that you can't leave behind.”*

Walk on - Bono Vox

Agradecimentos

Aos meus pais, Antonina e José; especialmente à minha mãe, pelo apoio incondicional e suporte emocional para as minhas decisões, hoje e sempre;

Ao meu irmão, Helder, por ser do jeito especial que é. Pelo companheirismo e a disposição de ouvir e compartilhar, estando sempre presente apesar da distância.

À Véra, por repartir comigo um trecho de sua vida, transformando uma orientação em uma forte amizade, que irá perdurar por muitos e muitos anos; ao professor Martínez, pela proposta de trabalho, por sua co-orientação e paciência;

Aos amigos, responsáveis por vários momentos de alegria e cumplicidade impescífveis, tornando esta jornada mais leve. A todos vocês que compartilharam um pedaço de suas vidas, muito obrigada. Em particular, agradeço a Alexandre, Cícero, Clair, Duvier, Ederson, Elianc, Flávia, Glauber, Heloísa, Irene, Jair, Julia, Juliana, Larissa, Laura, Lilian, Lucelina, Luciana, Luizinho, Marcos Eduardo, Marina, Maristela, Mayra, Momoe, Rodrigo Ferroni, Rodrigo Lima, Rodrigo Morante, Sebastián, Suzana, Wesley, Marcão, Seu Zé e D. Ivone;

A alguns professores que, sempre dispostos a ajudar, transformaram-se em amigos ao longo do tempo; em especial, agradeço à Cheti, Lúcio, Marcia, Rosana Pérez, Sandra, Silvia e Valéria. Aos funcionários mais amigos que já conheci: Cidinha, Tânia, Ednaldo, Fátima e Rafael.

Aos demais membros da Banca Examinadora, pelos comentários que engrandeceram o conteúdo desta Tese;

Ao CNPQ e à UNICAMP, pelo suporte financeiro impescindível para a realização deste trabalho.

Resumo

Sistemas de Otimalidade (ou Sistemas KKT) são sistemas formados pelas condições primais-duais estacionárias para a solução de problemas de otimização. Sob hipóteses adequadas (condições de qualificação), os minimizadores locais de um problema de minimização satisfarão as equações e inequações KKT; entretanto, infelizmente, muitos outros pontos estacionários (incluindo maximizadores) também são soluções desse sistema não linear. Por essa razão, os métodos destinados à resolução de problemas de programação não-linear fazem uso constante da estrutura de minimização, e o uso simples de métodos destinados à resolução de sistemas não-lineares podem gerar soluções espúrias. Todavia, caso o método destinado à resolução do sistema KKT tenha um ponto inicial situado na região de atração para um minimizador, esse método pode vir a ser muito eficiente. Neste trabalho, os métodos quase-Newton para a resolução de sistemas não-lineares são usados como aceleradores de algoritmos de programação não-linear (Lagrangiano Aumentado) com restrições de igualdade, desigualdade e caixa. Utilizamos como acelerador o método simétrico inverso de correção de posto um (ISR1), o qual realiza reinícios periódicos e faz uso das estruturas esparsas das matrizes para armazenamento. São demonstrados resultados de convergência e são realizados vários experimentos numéricos que comprovam a eficiência desta estratégia para problemas de minimização com restrições de igualdade, e indicam outros caminhos para problemas de minimização com restrições gerais (igualdade, desigualdade e caixa).

Abstract

Optimality (or KKT) systems arise as primal-dual stationarity conditions for constrained optimization problems. Under suitable constraint qualifications, local minimizers satisfy KKT equations but, unfortunately, many other stationary points (including, perhaps, maximizers) may solve these nonlinear systems too. For this reason, nonlinear-programming solvers make strong use of the minimization structure and the naive use of nonlinear-system solvers in optimization may lead to spurious solutions. Nevertheless, in the basin of attraction of a minimizer, nonlinear-system solvers may be quite efficient. In this work quasi-Newton methods for solving nonlinear systems are used as accelerators of nonlinear-programming (augmented Lagrangian) algorithms. A periodically-restarted memoryless symmetric rank-one (SR1) correction method is introduced for that purpose. Convergence results are given. For problems with only equality constraints, numerical experiments that confirm that the acceleration is effective are presented. A bunch of problems with equalities, inequalities and box constraints is tested and several comments and suggestions for further work are presented.

Conteúdo

1	Introdução	1
2	Método Lagrangiano Aumentado	4
2.1	O método	4
2.2	Algoritmo - Lagrangiano Aumentado (ALGENCAN)	7
2.2.1	Sobre o vetor σ_k	10
2.3	Análise teórica	11
2.4	Conclusão	15
3	Método Inverso Simétrico de Correção de Posto Um - ISR1	16
3.1	Os métodos quase-Newton	17
3.2	O método ISR1	18
3.3	Método direto (SR1) e método inverso (ISR1)	20
3.4	ISR1 em sistemas lineares	21
3.5	Algoritmo padrão	26

3.5.1	Detalhamento do algoritmo	28
3.6	Resultados de convergência	29
3.7	Conclusão	36
4	Método Lagrangiano Aumentado Acelerado	38
4.1	Introdução	39
4.2	O algoritmo proposto	40
4.3	Análise Teórica	42
4.4	Exemplo numérico	43
5	Método Lagrangiano Aumentado Acelerado - Problemas com Restrições de Igualdade	46
5.1	Introdução	46
5.2	O método Lagrangiano Aumentado	47
5.3	Análise teórica	49
5.4	Algoritmos para resolução dos subproblemas	52
5.5	O método Lagrangiano Aumentado Acelerado	52
5.6	Testes numéricos	55
5.6.1	Perfil de desempenho	55
5.6.2	Problemas de pequeno porte	57
5.6.3	Problemas de grande porte	60
5.6.4	Uma observação sobre a estruturação do algoritmo Lagrangiano Aumentado	66

5.7	Conclusões	68
6	Método Lagrangiano Aumentado Acelerado - Problemas com Restrições de Desigualdade	69
6.1	Acelerador	70
6.1.1	Detalhes da implementação	73
6.2	Resultados numéricos	75
6.2.1	Problemas de pequeno porte	75
6.2.2	Problemas de esferas-rígidas	76
6.2.3	Problema do passo da montanha	79
6.3	Conclusões	84
7	Conclusão	85

Lista de Figuras

4.1	Esquema descritivo do Lagrangiano Aumentado Acelerado.	41
4.2	Exemplo Numérico - Aceleração do método Lagrangiano Aumentado. . .	45
5.1	<i>Exemplo do método CPR</i> - Estrutura esparsa.	54
5.2	<i>Perfil de Desempenho com Milonga</i> - Problemas de Pequeno Porte. . . .	60
5.3	<i>Perfil de Desempenho com Milonga</i> - Tempo de execução.	62
5.4	<i>Perfil de Desempenho com Milonga</i> - Avaliações de Função.	63
5.5	<i>Perfil de Desempenho com Milonga</i> - Iterações Internas.	63
5.6	<i>Perfil de Desempenho com Gencan</i> - Tempo de execução.	64
5.7	<i>Perfil de Desempenho com Gencan</i> - Avaliações de função.	65
5.8	<i>Perfil de Desempenho</i> - Algoritmo proposto direto e invertido.	67
6.1	<i>Perfis de Desempenho: Avaliações de função e tempo de execução.</i> . . .	76
6.2	Ilustração da configuração inicial e da configuração ótima - Esfera Rígida.	77
6.3	<i>Perfis de Desempenho: Avaliações de função e tempo de execução</i> - $N \leq 1000$	78

6.4	<i>Perfis de Desempenho:</i> Avaliações de função e tempo de execução - $N \geq 1000$	78
6.5	Ilustração do caminho ótimo - Passo da Montanha.	79
6.6	Superfície utilizada nos teste numéricos - Passo da Montanha.	81
6.7	<i>Perfis de Desempenho:</i> Avaliações de função e tempo de execução - $N \leq 2000$	82
6.8	<i>Perfis de Desempenho:</i> Avaliações de função e tempo de execução - $N \geq 2000$	82
6.9	Solução do problema do Passo da Montanha - Superfície suave.	83
6.10	<i>Perfis de Desempenho:</i> Avaliações de função e tempo de execução - Superfície suave.	84

Capítulo 1

Introdução

O objetivo deste trabalho é resolver problemas do tipo

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & h(x) = 0 \\ & g(x) \leq 0 \\ & l \leq x \leq u \end{aligned}$$

onde $x \in \mathbb{R}^n$ é o vetor de variáveis a ser determinado, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$ são as restrições de igualdade e desigualdade (respectivamente), e $l, u \in \mathbb{R}^n$ são os vetores que contêm os limites superiores e inferiores (respectivamente) do vetor de variáveis.

Os métodos do tipo Lagrangiano Aumentado [4] são algoritmos clássicos destinados à resolução desse tipo de problema. Tais algoritmos resolvem uma seqüência de subproblemas, os quais são obtidos através da adição das restrições à função objetivo, utilizando funções de penalização. O acréscimo provocado na nova função objetivo (devido à essa adição) pode ser controlado através de um parâmetro de penalização (em geral, denotado por ρ). Tal método possui seu desempenho intrinsecamente ligado aos valores adotados para ρ : valores pequenos de ρ sugerem que a solução obtida na resolução do subproblema pode não ser factível; por outro lado, um valor alto para ρ pode acarretar instabilidade numérica durante a implementação computacional. Portanto, em geral, é comum utilizarmos um valor inicial pequeno para tal parâmetro, e irmos aumentando-o conforme percebemos que a factibilidade não está sendo cumprida.

Pode-se ter uma outra visão do problema de minimização observando-se que as condições de otimalidade desse problema podem ser expressas como um sistema de equações não-lineares. Logo, pode-se tentar encontrar a solução de um problema de minimização resolvendo o sistema não-linear associado a ele.

Os métodos mais comuns utilizados para resolver sistemas não-lineares são o método de Newton e os métodos quase-Newton [19], [24]. Em princípio, utilizar o sistema não-linear para resolver o problema de minimização não é um bom caminho, uma vez que a diferença entre minimizadores e maximizadores não é detectada. Entretanto, se a aproximação inicial é suficientemente próxima de um minimizador, então a resolução do sistema não-linear se apresenta como uma boa alternativa.

Neste trabalho, pretende-se realizar um estudo sobre como concatenar um método quase-Newton com um método tipo Lagrangiano Aumentado. A idéia principal consiste em utilizar o método Lagrangiano Aumentado como o responsável por fornecer um bom ponto inicial para o método quase-Newton, no sentido de que este último convirja para um minimizador.

Para tanto, inicialmente é dada uma visão geral sobre os métodos Lagrangiano Aumentado destinado à resolução de problemas de minimização com restrições de igualdade, desigualdade e caixa. Apresentamos o algoritmo que descreve a implementação do método Lagrangiano Aumentado escolhido, assim como uma visão geral dos resultados teóricos existentes na literatura recente sobre tal método. Esse capítulo não apresenta contribuição original. Porém, é de fundamental importância para a compreensão e a fundamentação teórica dos demais capítulos.

Propomos um método que resolve problemas de minimização com restrições de igualdade, desigualdade e caixa, o qual é uma combinação de um algoritmo quase-Newton (método inverso de correção de posto um - ISR1), implementado com uma busca linear não-monótona, com um método Lagrangiano Aumentado [31].

Para tanto, no Capítulo 2, expomos o método Lagrangiano Aumentado adotado em nossas implementações, bem como um resumo dos resultados de convergência presentes na literatura recente. No Capítulo 3, falamos sobre o método quase-Newton adotado. Apresentamos o método ISR1: sua fórmula de atualização das matrizes de aproximação, uma comparação entre o método direto e o inverso, e algumas propriedades que o ISR1 possui quando destinado à resolução de sistemas lineares. Em particular, na Seção 3.4 são apresentadas contribuições teóricas originais, uma vez que parte desses mesmo

resultados existem apenas no caso em que a matriz do sistema linear é definida positiva.

Embora existam resultados de convergência para o método ISR1 quando implementado sem busca linear, utilizamos um algoritmo que faz uso de uma busca linear não-monótona (baseada em [32], na qual permitimos que o valor da norma infinito da função do sistema no novo iterando possa ser superior que o seu valor no iterando anterior), a qual nos permite obter resultados de convergência global e local superlinear para o método. Esses resultados de convergência estendem-se para todos os métodos quase-Newton secantes que satisfazem uma propriedade particular (Weak Bounded Deterioration - WBD) [10] (ver Definição 3.6.1).

A matriz Jacobiana associada à função do sistema de otimalidade possui estrutura esparsa por blocos. Todas as matrizes utilizadas durante a execução do método ISR1 são armazenadas e manipuladas de forma esparsa, com o intuito de reduzir o tempo computacional e a quantidade requerida de memória.

A contribuição diretriz da Tese encontra-se exposta no Capítulo 4, onde propomos o novo algoritmo (proposta de aceleração) de forma detalhada, e as justificativas para as idéias de aceleração.

No capítulo seguinte são expostos os resultados teóricos existentes na literatura recente sobre o método Lagrangiano Aumentado quando destinado a resolução de problemas de minimização com apenas restrições de igualdade. É também exposto o método Lagrangiano Aumentado Acelerado e todas as particularidades de sua implementação para esse tipo de problema. Por fim, são apresentados e analisados vários testes numéricos para problemas de pequeno e grande porte, com o intuito de avaliar o desempenho do método proposto.

No Capítulo 4, propomos uma técnica de aceleração e apresentaremos os resultados numéricos obtidos na resolução de problemas de minimização com restrições de igualdade. No Capítulo 5 propomos uma aceleração semelhante na resolução de problemas de minimização com restrições. Os resultados numéricos obtidos são expostos e comentados.

O Capítulo 6 segue a mesma estrutura que o capítulo anterior, destinando-se agora à resolução de problemas de minimização com restrições de igualdade, desigualdade e caixa. Está presente uma descrição detalhada do algoritmo proposto assim como os perfis de desempenho dos métodos obtidos para vários problemas-teste. Finalizando este trabalho, o Capítulo 7 apresenta as conclusões e propostas para trabalhos futuros.

Capítulo 2

Método Lagrangiano Aumentando

Neste capítulo expomos o método Lagrangiano Aumentado utilizado em nossas implementações. Apresentamos o algoritmo e resultados teóricos para o mesmo, quando destinado a problemas de minimização com restrições de igualdade, e também problemas de minimização com restrições gerais (igualdade, desigualdade e caixa). Os resultados teóricos encontram-se na literatura recente [1], [3], e são expostos neste capítulo por se tratarem de justificativa fundamental para a convergência do algoritmo acelerado. A proposta do algoritmo acelerado, acompanhada de uma análise teórica e uma implementação computacional criteriosa de tal algoritmo, é o tema desta Tese.

2.1 O método

Os métodos tipo Lagrangiano Aumentado para problemas de minimização consistem, em linhas gerais, na eliminação de algumas (ou todas as) restrições por meio da adição à função objetivo, de uma função de penalização que depende das restrições, do vetor de multiplicadores de Lagrange e do parâmetro de penalização.

Tais métodos resolvem uma seqüência de *subproblemas* com restrições simples, onde os vetores de multiplicadores de Lagrange e os parâmetros de penalização são fixos durante a resolução dos subproblemas, e atualizados a cada iteração externa [4], [35].

Em geral, as restrições que não são penalizadas são escolhidas por constituírem um

subproblema resultante de fácil resolução.

Considere o problema de minimização

$$\begin{aligned}
 \min \quad & f(x) \\
 \text{s.a} \quad & h(x) = 0 \\
 & g(x) \leq 0 \\
 & l \leq x \leq u
 \end{aligned} \tag{2.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$ representa as restrições de igualdade, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$ representa as restrições de desigualdade, $l \in \mathbb{R}^n$ é o limite inferior e $u \in \mathbb{R}^n$ é o limite superior da variável $x \in \mathbb{R}^n$.

No caso de problemas de minimização como (2.1), as restrições de desigualdade permitem a utilização de uma grande quantidade de métodos do tipo Lagrangiano Aumentado. Um estudo recente de Birgin, Castillo e Martínez [5] concluiu que, dentre um conjunto de 65 métodos do tipo Lagrangiano Aumentado utilizados na resolução de 194 problemas da coleção CUTE [15] que possuem restrições de desigualdade, a função de penalização PHR (Powell-Hestenes-Rockafellar) [26],[38],[40] obteve o melhor desempenho tanto em questão de eficiência quanto em robustez, o que motivou o seu uso em nossas implementações.

A função Lagrangiana utilizada na resolução do problema original (2.1) é dada por

$$\mathcal{L}(x, \mu, \rho, \lambda) = f(x) + \lambda^T h(x) + \frac{\rho}{2} \|h(x)\|_2^2 + P(x, \rho, \mu) \tag{2.2}$$

onde P é a função de penalização PHR:

$$P(x, \rho, \mu) = \frac{1}{2\rho} \sum_{i=1}^{m_2} ([\max\{0, \mu_i + \rho g_i(x)\}]^2 - \mu_i^2), \tag{2.3}$$

referente à penalização das restrições de desigualdade. Conseqüentemente, o gradiente da função Lagrangiana (em relação a x) é dado por

$$\nabla_x \mathcal{L}(\cdot) = \nabla f(x) + J_h(x)^T (\lambda + \rho h(x)) + J_g(x)^T \max\{0, \mu + \rho g(x)\} \tag{2.4}$$

Uma condição de qualificação é uma propriedade que um minimizador local deve possuir para que satisfaça as condições KKT. Dentre as mais conhecidas, temos a Independência

Linear das Restrições Ativas, a condição de Mangasarian-Fromovitz e a condição de Dependência Linear Positiva Constante (CPLD) (ver [3], [41]). Se x_* é minimizador local do problema (2.1) e se uma condição de qualificação [3] é válida em x_* , então existem vetores de multiplicadores de Lagrange $\lambda_*, \mu_*, r_*, t_*$, que verificam o seguinte sistema de equações e inequações:

$$\left\{ \begin{array}{l} \nabla f(x) + J_h(x)^T \lambda + J_g(x)^T \mu - r + t = 0 \\ \text{diag}(g(x))\mu = 0 \\ \text{diag}(l - x)r = 0 \\ \text{diag}(x - u)t = 0 \\ h(x) = 0 \\ g(x) \leq 0 \\ l \leq x \leq u \\ \mu \geq 0, r \leq 0, t \geq 0. \end{array} \right. \quad (2.5)$$

Esse sistema é conhecido como as condições KKT [24] associadas ao problema (2.1), e o ponto x_* como um ponto KKT. As igualdades $\text{diag}(g(x))\mu = 0$, $\text{diag}(l - x)r = 0$, $\text{diag}(x - u)t = 0$ são denominadas *condições de complementaridade*. Essas condições significam que as restrições inativas em x devem possuir o respectivo multiplicador de Lagrange nulo.

Caso, em (2.2), a função de penalização PHR também seja utilizada para penalizar as restrições de caixa, o gradiente (2.4) da função Lagrangiano Aumentado seria reescrito como

$$\begin{aligned} \nabla_x \mathcal{L}(\cdot) = & \nabla f(x) + J_h(x)^T (\lambda + \rho h(x)) + J_g(x)^T \max\{0, \mu + \rho g(x)\} \\ & - \max\{0, r + \rho(l - x)\} + \max\{0, t + \rho(x - u)\}, \end{aligned}$$

e por comparação direta com (2.5), podemos verificar que uma boa estimativa para os valores de λ^* , μ^* , r^* e t^* é dada por

$$\begin{aligned} \lambda^* & \approx \lambda + \rho h(x) \\ \mu^* & \approx \max\{0, \mu + \rho g(x)\} \\ r^* & \approx \max\{0, r + \rho(l - x)\} \\ t^* & \approx \max\{0, t + \rho(x - u)\}. \end{aligned} \quad (2.6)$$

Essa comparação determinou a escolha das fórmulas de atualização das estimativas dos vetores de multiplicadores de Lagrange (λ_k e μ_k) utilizadas no Passo 4 do Algoritmo 2.2.1. As atualizações de r_k e t_k são utilizadas no Algoritmo 4.2.1. Tal escolha de atualização é clássica, e uma das mais utilizadas durante a implementação de métodos Lagrangiano Aumentado.

2.2 Algoritmo - Lagrangiano Aumentado (ALGENCAN)

Algoritmo 2.2.1. - Lagrangiano Aumentado

Considere o problema original de minimização (2.1). Assuma que x_0 é um ponto factível, $\varepsilon > 0$, $\gamma > 1$, $-\infty < \lambda_{min} < \lambda_{max} < \infty$, $-\infty < \mu_{min} < \mu_{max} < \infty$, $\tau \in (0, 1)$, $0 < \rho_1 \in \mathbb{R}$, $\lambda_1 \in \mathbb{R}^{m_1}$, $\lambda_1 \in [\lambda_{min}, \lambda_{max}]^{m_1}$ e $\mu_1 \in \mathbb{R}^{m_2}$, $\mu_1 \in [\mu_{min}, \mu_{max}]^{m_2}$. Seja $\{\varepsilon_k\}_{k \in \mathbb{N}}$ uma seqüência de números positivos que converge para zero. Tome inicialmente o contador de iterações externas igual a 1 ($k = 1$).

Passo 1 - Resolver o subproblema

$$\begin{aligned} \min \quad & f(x) + \lambda_k^T h(x) + \frac{\rho_k}{2} \|h(x)\|_2^2 + \\ & \frac{1}{2\rho_k} \sum_{i=1}^{m_2} ([\max\{0, [\mu_k]_i + \rho_k g_i(x)\}]^2 - [\mu_k]_i^2) \end{aligned} \quad (2.7)$$

$$\text{s.a. } l \leq x \leq u$$

com λ_k , μ_k e ρ_k fixos, utilizando x_{k-1} como ponto inicial e atribuindo a x_k a solução encontrada.

Passo 2 - Cálculo do vetor σ_k

$$\sigma_k = \max \left\{ g(x_k), -\frac{\mu_k}{\rho_k} \right\}. \quad (2.8)$$

Passo 3 - Critério de Parada

Seja \mathcal{P} o projetor na caixa $l \leq x \leq u$. Se

$$\max \{ \|h(x_k)\|_\infty, \|\sigma_k\|_\infty, \|\mathcal{P}[x_k - \nabla_x \mathcal{L}(x_k, \mu_k, \rho_k, \lambda_k)] - x_k\|_\infty \} \leq \varepsilon$$

tome $x^* = x_k$ e pare a execução do algoritmo.

Passo 4 - Atualização dos vetores de multiplicadores

Faça $\bar{\lambda}_{k+1} = \lambda_k + \rho_k h(x_k)$;

calcule λ_{k+1} com sendo a projeção de $\bar{\lambda}_{k+1}$ em $[\lambda_{\min}, \lambda_{\max}]^{m1}$.

Faça $\bar{\mu}_{k+1} = \max \{0, \mu_k + \rho_k g(x)\}$;

calcule μ_{k+1} com sendo a projeção de $\bar{\mu}_{k+1}$ em $[\mu_{\min}, \mu_{\max}]^{m2}$.

Passo 5 - Atualização do parâmetro de penalização

Se $\max \{ \|h(x_k)\|_\infty, \|\sigma_k\|_\infty \} \leq \tau \max \{ \|h(x_{k-1})\|_\infty, \|\sigma_{k-1}\|_\infty \}$, tome $\rho_{k+1} = \rho_k$.

Caso contrário, $\rho_{k+1} = \gamma \rho_k$.

Passo 6 - Atualização do contador

Se $k + 1$ é inferior ao limite máximo de iterações externas, faça $k = k + 1$ e volte para o Passo 1.

Caso contrário, a execução do algoritmo é interrompida e é declarado que o algoritmo não converge.

Na implementação do Algoritmo 2.2.1, utilizamos a rotina ALGENCAN (Augmented Lagrangian algorithm using GENCAN) [6], disponível na página da internet do projeto *Tango* (www.ime.usp.br/~egbirgin/tango)¹, usando os valores padrões sugeridos pelos autores para os parâmetros. ALGENCAN é uma rotina destinada à resolução de problemas de programação de grande porte, e utiliza a subrotina GENCAN para resolver o subproblema do Passo 1 do Algoritmo 2.2.1.

¹TANGO (Trustable Algorithms for Nonlinear General Optimization) é um conjunto de rotinas em Fortran voltadas para a resolução de problemas de otimização, desenvolvido no DMA - IMECC - Unicamp e DCC - IME - USP, sob a coordenação do Prof. J. M. Martínez.

GENCAN (Generalização de QUACAN - Quadráticas Canalizadas) [6] foi desenvolvido por E. G. Birgin e J. M. Martínez; essa rotina computacional utiliza uma busca linear, a qual combina “backtracking” e extrapolação, durante a minimização interna às faces. GENCAN associa a teoria de conjunto-ativo com gradiente espectral projetado [7], [9] para a resolução dos subproblemas de minimização com restrições de caixa. A técnica do Gradiente Espectral Projetado [7], [9] permite, durante a execução, que restrições possam ser adicionadas ou retiradas em cada iteração, fazendo com o que o método seja aconselhável para problemas de grande porte.

A direção de descida é determinada por meio de uma aproximação do tipo Newton-truncada, e é obtida durante a implementação por meio do método dos Gradientes Conjugados.

Observe que o vetor σ_k é utilizado tanto no Passo 3 (critério de parada) quanto no Passo 5 (atualização das estimativas dos multiplicadores) do algoritmo anterior. Intuitivamente, podemos pensar que quando o valor da norma infinito de σ_k é bem pequeno (digamos, menor que $\varepsilon > 0$ suficientemente pequeno), podemos garantir que, para cada $i = 1, \dots, m_2$,

$$|[\sigma_k]_i| < \varepsilon;$$

nesse caso, existem duas possibilidades:

- Se $[g(x)]_i < -\frac{[\mu_k]_i}{\rho_k} \Rightarrow [g(x)]_i < \varepsilon \quad \text{e} \quad -\varepsilon \rho_k < [\mu_k]_i < \varepsilon \rho_k$
- Se $-\frac{[\mu_k]_i}{\rho_k} < [g(x)]_i \Rightarrow -\varepsilon < [g(x)]_i < \varepsilon \quad \text{e} \quad [\mu_k]_i > -\varepsilon \rho_k$

Podemos concluir que (com uma ε - tolerância), se o primeiro caso for verificado, temos que as restrições de desigualdade são satisfeitas, com valor nulo para o respectivo multiplicador de Lagrange. Se o segundo caso for verificado, temos que a restrição g_i é ativa em x , e que o seu respectivo multiplicador de Lagrange é positivo. Então, uma norma suficientemente pequena para o vetor σ_k indica que o ponto x em questão é factível, e que são válidas as hipóteses de complementaridade.

Em [4], Bertsekas demonstra como chegar à expressão (2.8) para o valor do vetor σ_k . Na seção a seguir, reproduzimos parte dessa demonstração, a fim de facilitar a compreensão do texto.

2.2.1 Sobre o vetor σ_k

Ao longo desta seção, usaremos a seguinte notação:

- dado $u \in \mathbb{R}^n$, u^2 é um vetor onde a i -ésima componente é dada por u_i^2 .
- dados $u, v \in \mathbb{R}^n$, $w = \max\{u, v\}$ é um vetor onde a i -ésima componente é dada por $\max\{u_i, v_i\}$.

O vetor σ_k tem sua origem através de uma reformulação do problema (2.1). Uma vez que a teoria sobre métodos tipo Lagrangiano Aumentado destinados à resolução de problemas de minimização com apenas restrições de igualdade está bem consolidada, considere o problema estendido associado a (2.1):

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & h(x) = 0 \\ & g_i(x) + z^2 = 0 \\ & x \in \Omega \end{aligned} \tag{2.9}$$

onde $\Omega = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ e z_1, \dots, z_{m_2} são variáveis de folga.

Para o problema (2.9), o método Lagrangiano Aumentado resolve o seguinte problema de minimização:

$$\begin{aligned} \min_{x, z} \quad & f(x) + \lambda^T h(x) + \mu^T [g(x) + z^2] + \frac{\rho}{2} [\|h(x)\|_2^2 + \|g(x) + z^2\|_2^2] \\ \text{s.a} \quad & x \in \Omega \end{aligned} \tag{2.10}$$

para valores fixos de ρ , λ e μ .

Conforme descrito em [4], pode-se minimizar apenas para z , considerando um vetor x fixo factível. Considerando apenas a minimização em z , podemos observar que o problema (2.10) é equivalente a

$$\min_z \quad \mu^T [g(x) + z^2] + \frac{\rho}{2} \|g(x) + z^2\|_2^2. \tag{2.11}$$

Derivando a função objetivo e igualando a zero, temos:

$$\mu + \rho g(x) + \rho z^2 = 0$$

e então podemos concluir que o valor de z ótimo é dado por

$$z^* = \sqrt{\max\left\{0, -\frac{\mu}{\rho} - g(x)\right\}}.$$

Assim, substituindo o valor encontrado em (2.10), temos

$$\begin{aligned} \min_x \quad & f(x) + \lambda^T h(x) + \frac{\rho}{2} \|h(x)\|_2^2 + \frac{1}{2\rho} \sum_{i=1}^{m_2} ([\max\{0, \mu_i + \rho g_i(x)\}]^2 - \mu_i^2) \\ \text{s.a.} \quad & x \in \Omega \end{aligned} \quad (2.12)$$

cuja função objetivo é conhecida como a função Lagrangiano Aumentado de Powell-Hestenes-Rockafeller (PHR).

Representaremos por σ o valor das restrições de desigualdade acrescidas das variáveis de folga,

$$\sigma_k = g(x_k) + z_k^2.$$

Substituindo o valor de z_k ,

$$\sigma_k = g(x_k) + \max \left\{ 0, -\frac{\mu_k}{\rho_k} - g(x_k) \right\} = \max \left\{ g(x_k), -\frac{\mu_k}{\rho_k} \right\}$$

de onde temos a fórmula para o cálculo de σ_k no Passo 2 do Algoritmo 2.2.1.

2.3 Análise teórica

Andreani, Birgin, Martínez e Schuverdt [1] realizaram um estudo recente sobre algoritmos Lagrangiano Aumentado para problemas de minimização com restrições gerais, utilizando como hipótese a Dependência Linear Positiva Constante (CPLD) sobre todas as restrições, que é uma hipótese mais fraca que a utilizada nos trabalhos anteriores.

Esse estudo fornece os seguintes resultados teóricos para o Algoritmo 2.2.1, quando os subproblemas são resolvidos de forma inexata:

- se um ponto limite satisfaz a condição CPLD, então ou ele é um ponto KKT da soma dos quadrados das infactibilidades, ou ele é factível;
- pontos factíveis são pontos KKT; a demonstração desse resultado segue através da hipótese de que tais pontos satisfazem a condição CPLD;

Para facilitar a redação da definição a seguir, considere a função $G : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2+2n}$ definida por blocos da seguinte forma

$$G(x) = \begin{pmatrix} g(x) \\ l - x \\ x - u \end{pmatrix}.$$

Então as restrições do problema (2.1) podem ser reescritas como

$$h(x) = 0; \quad G(x) \leq 0.$$

Definição 2.3.1. CPLD para restrições de igualdade e desigualdade.

Dizemos que um ponto x satisfaz a Condição de Dependência Linear Positiva Constante (CPLD) em relação às restrições de igualdade, desigualdade e caixa se existem

- $I_h \subset \{1, \dots, m_1\}$
- $I_G \subset I_A = \{i = 1, \dots, m_2 + 2n \mid G_i(x) = 0\},$

tais que, se a igualdade

$$\sum_{i \in I_h} \lambda_i \nabla h_i(x) + \sum_{i \in I_G} \mu_i \nabla G_i(x) = 0$$

é satisfeita com $\sum_{i \in I_h} |\lambda_i| + \sum_{i \in I_G} \mu_i > 0$, então os gradientes

$$\{\nabla h_i(z)\}_{i \in I_h}, \{\nabla G_i(z)\}_{i \in I_G}$$

são linearmente dependentes para todo z em uma vizinhança de x .

O teorema a seguir é o primeiro resultado apresentado em [1], o qual diz que se x_* (ponto limite da seqüência de pontos $\{x_k\}$ gerada pelo Algoritmo 2.2.1) não é factível, então ele é um minimizador local da infactibilidade das restrições penalizadas, sujeito às restrições de caixa.

Teorema 2.3.1. *Seja $\{x_k\}$ uma seqüência gerada pelo Algoritmo 2.2.1. Seja x_* um ponto limite de $\{x_k\}$. Então, se a seqüência de parâmetros de penalização $\{\rho_k\}$ é limitada, o ponto limite x_* é factível. Caso contrário, ao menos uma das seguintes possibilidades é verdadeira:*

- x_* é um ponto KKT do problema

$$\begin{aligned} \min \quad & \frac{1}{2} \left[\sum_{i=1}^{m_1} [h(x)]_i^2 + \sum_{i=1}^{m_2} \max \{0, [g(x)]_i\}^2 \right] \\ \text{s.a} \quad & l \leq x \leq u \end{aligned}$$

- x_* não satisfaz a restrição de qualificação CPLD, levando em conta o conjunto de restrições $\Omega = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$.

Demonstração. Ver [1], Theorem 3.1. ■

A exigência de que um ponto x satisfaça a restrição de qualificação CPLD em relação ao conjunto de restrições $\Omega = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ é sempre satisfeita, uma vez que os gradientes de cada uma dessas restrições são dados pelos vetores $-e_i$, $i = 1, \dots, n$, no caso da limitação inferior de x e e_i , $i = 1, \dots, n$, no caso da limitação superior de x .

Assim, as idéias do Teorema 2.3.1 podem ser resumidas da seguinte forma: se a seqüência de parâmetros de penalização $\{\rho_k\}$ é limitada, então ou o ponto limite x_* é factível ou x_* é um ponto KKT do problema de minimização das normas das infactibilidades.

O Teorema 2.3.2 a seguir diz que pontos limites factíveis são pontos estacionários (KKT) do problema original, supondo que a condição de qualificação CPLD é válida.

Teorema 2.3.2. *Seja $\{x_k\}_{k \in \mathbb{N}}$ uma seqüência gerada pelo Algoritmo 2.2.1. Assuma que x_* é um ponto limite factível que satisfaz a condição de qualificação CPLD em relação a todas as restrições. Então, x_* é um ponto KKT do problema original. Além disso, se x_* satisfaz a condição de qualificação de Mangasarian-Fromovitz [39] com $\{x_k\}_{k \in \mathbb{K}}$ uma subseqüência que converge para x_* , o conjunto*

$$\{\|\lambda_{k+1}\|, \|\mu_{k+1}\|\}_{k \in \mathbb{K}}$$

é limitado.

Demonstração. Ver [1], Theorem 3.2. ■

O próximo resultado corresponde à limitação da seqüência de parâmetros de penalização gerados pelo Algoritmo 2.2.1. Esse resultado é obtido através da associação da seqüência $\{x_k\}$ gerada pelo Algoritmo 2.2.1 quando aplicado ao problema original, e a seqüência de pontos gerada pelo mesmo algoritmo quando aplicado a problemas que possuem apenas restrições de igualdade:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & c(x) = 0, \end{aligned} \tag{2.13}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $c : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$.

Para tanto, considere as seguintes hipóteses:

- H0. A seqüência $\{x_k\}$ é gerada pelo Algoritmo 2.2.1 e

$$\lim_{k \rightarrow \infty} x_k = x_*.$$

- H1.

1. O ponto x_* é factível;
2. Os gradientes das restrições em x_* são linearmente independentes;
3. As funções f e c possuem derivadas segunda contínuas em uma vizinhança de x_* ;
4. É válida a condição suficiente de segunda ordem para minimizadores locais, isto é, se $\lambda_* \in \mathbb{R}^{m_1}$ é o vetor de multiplicadores de Lagrange associado à x_* , então:

$$\langle z, \nabla_{xx}^2 L(x_*, \lambda_*) z \rangle > 0$$

para todo $z \neq 0$ tal que $\nabla c(x_*)^T z = 0$, sendo L a função Lagrangiana aumentada associada ao problema (2.13).

- H2. Para todo $i = 1, \dots, m_1$,

$$[\lambda_*]_i \in ([\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i).$$

Teorema 2.3.3. *Suponha que as Hipóteses $H0$, $H1$ e $H2$ são satisfeitas pela seqüência gerada pelo Algoritmo 2.2.1, aplicado ao problema (2.13). Além disso, suponha que existe uma seqüência $\eta_k \rightarrow 0$ tal que*

$$\varepsilon_k \leq \eta_k \|c(x_k)\|_\infty$$

para todo $k \in \mathbb{N}$, onde a seqüência $\{\varepsilon_k\}$ é formada pelos limites superiores dos testes de parada, para cada iteração k .

Então, a seqüência dos parâmetros de penalidade $\{\rho_k\}$ é limitada.

Demonstração. Ver [1], Theorem 4.2. ■

2.4 Conclusão

Neste capítulo descrevemos o método Lagrangiano Aumentado que será utilizado em nossas implementações numéricas. Apresentamos detalhadamente o seu algoritmo, assim como um resumo de sua análise teórica, apresentada em [4] e mais recentemente em [1].

A implementação de algoritmos tipo Lagrangiano Aumentado é naturalmente associada a estratégias de modificação de um ou mais parâmetros de penalização. Se esses parâmetros crescem muito durante o processo, o problema de minimização associado torna-se muito difícil de resolver. Por outro lado, um aumento muito pequeno torna a convergência lenta.

Esses dois fatos mostram a necessidade e sugerem um estudo de uma proposta de aceleração para a convergência de métodos do tipo Lagrangiano Aumentado, que seja cuidadoso o suficiente para não perder a estabilidade dos algoritmos a serem utilizados.

Capítulo 3

Método Inverso Simétrico de Correção de Posto Um - ISR1

Antes de passar à apresentação da nossa proposta de um novo algoritmo que visa acelerar a convergência do método Lagrangiano Aumentado, vamos nos dedicar, neste capítulo, ao método quase-Newton ISR1, que escolhemos como método acelerador para o algoritmo proposto.

O vetor solução (associado ao correspondente vetor de multiplicadores de Lagrange) de um problema de minimização satisfaz um conjunto de equações (e/ou inequações), as quais são denotadas por condições de otimalidade (ou condições KKT) desse problema.

Uma outra forma de encontrarmos as soluções de problemas de programação não linear seria resolver o sistema formado por suas condições de otimalidade¹, utilizando métodos numéricos. Para tanto, neste capítulo realizamos uma breve descrição dos métodos quase-Newton (em particular, descrevemos o método Inverso de Correção de Posto Um - ISR1), destinados à resolução de sistemas não-lineares.

Fazemos a seguir um estudo detalhado do método ISR1: sua aplicação na resolução de sistemas lineares, algoritmo e resultados de convergência obtidos, quando utilizada uma busca linear não-monótona.

¹Essa forma de resolução será melhor explorada no Capítulo 4.

3.1 Os métodos quase-Newton

A resolução de um sistema de equações não lineares é uma tarefa necessária durante a resolução de problemas das mais diversas áreas (física, engenharia, economia e outras ciências).

Dada a função não-linear $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $F = (f_1, \dots, f_n)^T$, que possui derivadas contínuas, consideremos o problema de resolver

$$F(x) = 0. \quad (3.1)$$

Denotaremos a matriz de derivadas parciais de F (a matriz jacobiana ou simplesmente o Jacobiano) por $J(x)$.

Existem vários métodos propostos para a resolução de sistemas não lineares. O método de Newton é o mais conhecido, e serve de base para a obtenção de outros métodos eficientes; nesse método iterativo, a seqüência de aproximações x_k é gerada por:

$$x_{k+1} = x_k - J(x_k)^{-1}F(x_k). \quad (3.2)$$

Sob certas condições, o método de Newton tem convergência quadrática local.

Os métodos quase-Newton buscam, através de uma aproximação B_k para $J(x_k)$, evitar o cálculo da matriz jacobiana em cada iteração. A seqüência de aproximações é gerada por

$$x_{k+1} = x_k - B_k^{-1}F(x_k), \quad (3.3)$$

onde a matriz B_{k+1} é obtida a partir de B_k utilizando fórmulas de recorrência baseadas em iterações anteriores. Sob certas condições, os métodos quase-Newton têm convergência local superlinear.

Uma classe de métodos quase-Newton das mais bem sucedidas é a dos métodos secantes. Nestes, utiliza-se como critério de escolha para atualizar a matriz B_k , aproximação da matriz jacobiana, uma matriz que satisfaça a equação secante:

$$B_{k+1}s_k = F(x_{k+1}) - F(x_k) = y_k \quad (3.4)$$

onde $s_k = x_{k+1} - x_k$.

Os métodos quase-Newton inversos são métodos iterativos que atualizam, a cada iteração, a matriz H_k que aproxima a inversa da matriz Jacobiana $J(x)$ no ponto x_k . Neste caso, a classe de métodos secantes refere-se aos métodos cujas matrizes satisfazem a seguinte equação secante inversa:

$$H_{k+1}y_k = x_{k+1} - x_k = s_k. \quad (3.5)$$

Neste capítulo trataremos do método de correção de posto um simétrico inverso (ISR1 - Inverse Symmetric Rank-One Updating Method) [16], [33]. Nesse método, atualiza-se uma aproximação da matriz inversa (H_k) adicionando uma matriz de posto um, mantendo sua simetria; em sua implementação, realizamos reinícios periódicos, nos quais realizamos uma iteração do tipo Newton. Apresentamos também uma análise teórica e resultados de convergência para métodos quase-Newton secantes (de uma forma geral), que utilizem uma busca linear não-monótona. Muitos trabalhos apresentam resultados de convergência para métodos quase-Newton [11], [12], [21], [30] implementados ou não com busca linear; entretanto, até o presente momento, nenhum trabalho realiza tal análise teórica utilizando uma busca linear não-monótona.

3.2 O método ISR1

Utilizando um método quase-Newton inverso para obter a solução de $F(x) = 0$, dado um ponto inicial x_0 , podemos escrever cada novo iterando como sendo

$$x_{k+1} = x_k - H_k F(x_k), \quad (3.6)$$

onde $H_k \approx J(x_k)^{-1}$. Caso $J(x_k)$ seja simétrica (e, conseqüentemente, $J(x_k)^{-1}$), é natural desejar que a aproximação H_k também seja simétrica, o que é o caso do método ISR1.

No método ISR1, a atualização das matrizes H_k é feita pela seguinte correção de posto um:

$$H_{k+1} = H_k + \beta v_k v_k^T, \quad (3.7)$$

onde β pode assumir os valores ± 1 .

Sejam $s_k = x_{k+1} - x_k$ e $y_k = F(x_{k+1}) - F(x_k)$. Exigindo que H_{k+1} satisfaça a equação

secante $H_{k+1}y_k = s_k$, temos

$$\begin{aligned} H_{k+1}y_k &= s_k \\ (H_k + \beta v_k v_k^T) y_k &= s_k \\ (\beta v_k^T y_k) v_k &= s_k - H_k y_k \end{aligned} \quad (3.8)$$

Logo, v é um múltiplo de $s_k - H_k y_k$; então, existe $\delta \in \mathbb{R}$ tal que

$$v = \delta (s_k - H_k y_k). \quad (3.9)$$

Substituindo a igualdade (3.9) em (3.8), obtemos

$$s_k - H_k y_k = \beta \delta^2 \left[(s_k - H_k y_k)^T y_k \right] (s_k - H_k y_k).$$

Como β pode assumir apenas os valores ± 1 , e δ^2 é não-negativo, concluímos que

$$\beta = \text{sign} \left[y_k^T (s_k - H_k y_k) \right] \quad \delta = \pm \left| y_k^T (s_k - H_k y_k) \right|^{-1/2}.$$

Logo, a fórmula de atualização do método de correção de posto um simétrico inverso será dada por

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)}. \quad (3.10)$$

Devido à expressão da atualização da matriz H_k , verifica-se facilmente que, dada uma aproximação inicial H_0 simétrica, todas as demais aproximações H_k ($k \geq 1$) também o serão.

Devido à fórmula de atualização (3.10), são necessárias algumas observações à respeito do seu denominador:

1. Se $s_k = H_k y_k$, então a única fórmula de atualização simétrica que satisfaz a última equação secante é

$$H_{k+1} = H_k;$$

2. Se $s_k \neq H_k y_k$, mas $(s_k - H_k y_k)^T y_k = 0$, então não existe uma fórmula de atualização simétrica que satisfaça a última equação secante. Durante a implementação numérica, essa situação é evitada por meio de uma salvaguarda (ver (3.23));
3. Se $(s_k - H_k y_k)^T y_k \neq 0$, a fórmula (3.10) é utilizada para atualização.

3.3 Método direto (SR1) e método inverso (ISR1)

De forma análoga à descrita na seção anterior, é possível deduzir a forma direta do Método de Atualização Simétrica de Posto Um (SR1):

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}. \quad (3.11)$$

Suponha, a princípio, que as fórmulas de atualizações (3.10) e (3.11) possuem denominadores não nulos. Valores muito pequenos dos denominadores são evitados por meio de uma salvaguarda na implementação do método. Nesta seção, será demonstrado que a matriz H_{k+1} , cuja fórmula de atualização é dada por (3.10) é exatamente a matriz inversa de B_k , com a fórmula de atualização dada em (3.11).

Para tanto, definamos $w_k = y_k - B_k s_k$ e $r_k = \frac{w_k}{w_k^T s_k}$. Então, de (3.11), temos que

$$B_{k+1} = B_k + r_k w_k^T. \quad (3.12)$$

Aplicando a fórmula de Sherman-Morrison [36] para o cálculo da matriz inversa, obtemos

$$B_{k+1}^{-1} = B_k^{-1} - \frac{B_k^{-1} r_k w_k^T B_k^{-1}}{1 + w_k^T B_k^{-1} r_k}. \quad (3.13)$$

Chamando $H_k = B_k^{-1}$ a matriz inversa de B_k , e substituindo o valor de r_k na expressão acima, obtemos

$$H_{k+1} = H_k - \frac{H_k w_k w_k^T H_k}{w_k^T s_k + w_k^T H_k w_k}. \quad (3.14)$$

Supondo B_0 simétrica (e, conseqüentemente, H_0 simétrica), temos que

$$B_k = B_k^T \quad \text{e} \quad H_k = H_k^T, \quad \forall k.$$

Assim,

$$H_{k+1} = H_k - \frac{(H_k w_k)(H_k w_k)^T}{w_k^T (s_k + H_k w_k)}. \quad (3.15)$$

Como $H_k w_k = -(s_k - H_k y_k)$, de (3.15),

$$\begin{aligned} H_{k+1} &= H_k - \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{w_k^T (s_k - s_k + H_k y_k)} \\ &= H_k - \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(H_k w_k)^T y_k} \\ &= H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k} \end{aligned} \quad (3.16)$$

Portanto, no caso do Método de Correção de Posto Um Simétrico, H_{k+1} é igual a B_{k+1}^{-1} , como queríamos demonstrar.

3.4 ISR1 em sistemas lineares

Nesta seção, estaremos interessados em encontrar a solução do sistema linear $Ax = b$, onde A é uma matriz simétrica. Seja $q(x) = Ax - b$; então resolver tal sistema é equivalente a encontrar a solução de $q(x) = 0$. Suponha A inversível, o que faz com que $q(x)$ tenha solução única $x_* = A^{-1}b$.

A seguir são demonstrados alguns resultados teóricos originais, especialmente obtidos quando o método ISR1 é usado para resolver tais problemas. Observamos que existe resultado semelhante ao teorema a seguir para matrizes simétricas definidas positivas [29].

Mostramos inicialmente que, para a resolução desse tipo de problema, a matriz H_{j+1} satisfaz todas as $j + 1$ equações secantes anteriores.

Teorema 3.4.1. *Dada uma matriz inicial H_0 simétrica, para $j = 0, \dots, n - 1$, temos que*

$$s_i = H_{j+1} y_i \quad (3.17)$$

para todo $i \leq j$.

Demonstração. A prova é feita por indução. Para $j = 0$, o único valor possível para i é 0, fazendo com que a relação (3.17) seja válida (por se tratar da última equação secante).

Suponha que a relação seja válida para H_j , com $i \leq j - 1$. Para a matriz H_{j+1} , (3.17) é verdadeira quanto $i = j$ (novamente, por se tratar da última equação secante). Para $i < j$, utilizando a hipótese de indução,

$$\begin{aligned} H_{j+1} y_i &= \left[H_j + \frac{(s_j - H_j y_j)(s_j - H_j y_j)^T}{(s_j - H_j y_j)^T y_j} \right] y_i \\ &= H_j y_i + \left[\frac{(s_j - H_j y_j)}{y_j^T (s_j - H_j y_j)} \right] (s_j^T y_i - y_j^T H_j^T y_i) \\ &= s_i + \left[\frac{(s_j - H_j y_j)}{y_j^T (s_j - H_j y_j)} \right] (s_j^T y_i - y_j^T H_j y_i) \end{aligned}$$

pois H_j é simétrica (visto que H_0 o é). Como

$$y_i = q(x_{i+1}) - q(x_i) = Ax_{i+1} - b - Ax_i + b = As_i, \quad (3.18)$$

usando a hipótese de indução,

$$s_j^T y_i - y_j^T H_j y_i = s_j^T y_i - y_j^T s_i = s_j^T As_i - (As_j)^T s_i = 0;$$

então,

$$H_{j+1} y_i = s_i, \quad (3.19)$$

o que encerra a demonstração. ■

Assim, temos que

$$\left\{ \begin{array}{l} H_n y_0 = s_0 \\ H_n y_1 = s_1 \\ \vdots \\ H_n y_{n-1} = s_{n-1}, \end{array} \right. \quad (3.20)$$

o que determina univocamente a matriz H_n , caso os vetores y_0, y_1, \dots, y_{n-1} sejam linearmente independentes.

Provamos ainda que, no caso da resolução de sistemas lineares, a relação entre dependência e independência linear dos vetores s_i e y_i , com $i = 1, \dots, n-1$, é dada pelo lema a seguir.

Lema 3.4.1.

Os vetores s_0, s_1, \dots, s_{n-1} são linearmente independentes se e somente se os vetores y_0, y_1, \dots, y_{n-1} são linearmente independentes.

Demonstração. Suponha que s_0, s_1, \dots, s_{n-1} sejam linearmente independentes, e sejam $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ tais que

$$\alpha_0 y_0 + \dots + \alpha_{n-1} y_{n-1} = 0.$$

Multiplicando por H_n em ambos os lados, e usando (3.19),

$$\alpha_0 H_n y_0 + \dots + \alpha_{n-1} H_n y_{n-1} = 0$$

$$\alpha_0 s_0 + \dots + \alpha_{n-1} s_{n-1} = 0,$$

e então os coeficientes α_k , $k = 0, \dots, n-1$, são todos nulos, provando que y_0, y_1, \dots, y_{n-1} são linearmente independentes.

Suponha agora que y_0, y_1, \dots, y_{n-1} são linearmente independentes; sejam $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ tais que

$$\alpha_0 s_0 + \dots + \alpha_{n-1} s_{n-1} = 0;$$

multiplicando por A em ambos os lados,

$$\alpha_0 A s_0 + \dots + \alpha_{n-1} A s_{n-1} = 0.$$

Agora, por (3.18),

$$\alpha_0 y_0 + \dots + \alpha_{n-1} y_{n-1} = 0,$$

donde os coeficientes α_k , $k = 0, \dots, n-1$, são todos nulos, provando que s_0, s_1, \dots, s_{n-1} são linearmente independentes. ■

No lema anterior, a relação direta entre dependência (independência) linear dos vetores s_i e y_i é válida desde que o conjunto de índices dos vetores linearmente dependentes (independentes) s_i seja o mesmo que para os vetores y_i .

Nas observações a seguir analisaremos em que casos temos garantia (ou não) da convergência do método para a solução do sistema linear. Mais precisamente, a Observação 3.4.1 garante que, se as direções quase-Newton s_i são linearmente independentes, o método ISR1 converge para a solução do sistema de dimensão n em no máximo n iterações. A Observação 3.4.2 tenta esclarecer quais são os possíveis resultados quando os vetores s_i gerados pelo método não são linearmente independentes.

Observação 3.4.1. *Suponha que os vetores y_i , $i = 0, \dots, n-1$, sejam linearmente independentes. Utilizando (3.18) e (3.20), temos que a matriz H_n (escrita por colunas) é dada por*

$$H_n = [s_0 \mid \dots \mid s_{n-1}] [As_0 \mid \dots \mid As_{n-1}]^{-1};$$

como s_0, s_1, \dots, s_{n-1} também são linearmente independentes (Lema 3.4.1), podemos escrever H_n como

$$\begin{aligned} H_n &= [s_0 \mid \dots \mid s_{n-1}] [s_0 \mid \dots \mid s_{n-1}]^{-1} A^{-1} \\ H_n &= A^{-1}. \end{aligned}$$

Então, o ponto x_{n+1} será dado por

$$\begin{aligned} x_{n+1} &= x_n - H_n q(x_n) \\ &= x_n - A^{-1} (Ax_n - b) \\ &= x_n - A^{-1} Ax_n + A^{-1} b \\ &= A^{-1} b; \end{aligned}$$

portanto

$$Ax_{n+1} = b,$$

fazendo com que x_{n+1} seja a solução do sistema original.

Observação 3.4.2. Consideremos agora que s_0, s_1, \dots, s_{n-1} sejam vetores linearmente dependentes, e seja s_j o primeiro vetor (a partir de s_0) que possa ser escrito como combinação linear de alguns vetores anteriores (com coeficientes não nulos):

$$s_j = \sum_{i \in D} \alpha_i s_i, \quad D \subset \{0, \dots, j-1\},$$

onde D corresponde ao conjunto de índices dos vetores s_i que formam tal combinação linear.

Como

$$s_j = \sum_{i \in D} \alpha_i s_i \rightarrow As_j = \sum_{i \in D} \alpha_i As_i \rightarrow y_j = \sum_{i \in D} \alpha_i y_i,$$

do Teorema 3.4.1 temos que

$$s_i = H_j y_i, \quad \forall i \in D \Rightarrow \alpha_i s_i = H_j \alpha_i y_i, \quad \forall i \in D,$$

e então

$$\sum_{i \in D} \alpha_i s_i = H_j \sum_{i \in D} \alpha_i y_i \rightarrow s_j = H_j y_j;$$

assim, a matriz H_j satisfará $j + 1$ equações secantes, e não apenas j equações como enuncia o Teorema 3.4.1.

Dado que

$$H_{j+1} = H_j + \frac{(s_j - H_j y_j)(s_j - H_j y_j)^T}{(s_j - H_j y_j)^T y_j},$$

não será possível atualizar a matriz.

Portanto, caso isso ocorra ao longo da implementação, será necessário implementar uma salvaguarda, para que seja oferecida uma outra forma de obtenção da matriz H_{j+1} . É necessário observar que, caso a escolha seja por repetir a matriz da iteração anterior

($H_{j+1} = H_j$) o método estagnará, uma vez que:

$$\begin{aligned}
 H_j y_j &= s_j \\
 H_j [g(x_{j+1}) - g(x_j)] &= s_j \\
 H_j g(x_{j+1}) - H_j g(x_j) &= s_j \\
 H_j g(x_{j+1}) + s_j &= s_j \\
 H_j g(x_{j+1}) &= 0 \\
 H_{j+1} g(x_{j+1}) &= 0 \\
 s_{j+1} &= 0,
 \end{aligned}$$

fazendo com que $x_{j+1} = x_j$.

Outras escolhas para a atualização da matriz H_{j+1} podem ser utilizadas, tais como $H_{j+1} = J(x_{j+1})^{-1}$, $H_{j+1} = I_n$, etc..

3.5 Algoritmo padrão

Denotaremos $\mathbb{N} = \{0, 1, \dots\}$ e $\|\cdot\|$ como a norma euclidiana. Suponhamos que $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ possui derivadas parciais contínuas.

Seja $\{\eta_k\}$ uma seqüência tal que $\eta_k > 0$ para todo $k \in \mathbb{N}$ e

$$\sum_{k=0}^{\infty} \eta_k = \eta < \infty. \tag{3.21}$$

Sejam τ_{min}, τ_{max} tais que $0 < \tau_{min} < \tau_{max} < 1$ e $0 < \gamma < 1$. Dado $x_0 \in \mathbb{R}^n$ um ponto inicial arbitrário, o algoritmo que nos permite obter x_{k+1} a partir de x_k é dado a seguir.

Algoritmo 3.5.1.

Passo 1. Seja $d_k = -H_k F(x_k)$.

Passo 2. Seja $\alpha = 1$.

Passo 3. Se

$$\|F(x_k + \alpha d_k)\| \leq [1 - \alpha\gamma] \|F(x_k)\| + \eta_k \quad (3.22)$$

faça $\alpha_k = \alpha$ e calcule o novo iterando como $x_{k+1} = x_k + \alpha_k d_k$.

Passo 4. Se (3.22) não é satisfeita, escolha $\alpha_{new} \in [\tau_{min}\alpha, \tau_{max}\alpha]$, faça $\alpha = \alpha_{new}$ e volte ao Passo 3.

Durante a implementação computacional, no caso de (3.22) não ser satisfeito, o novo valor para α foi calculado como $\alpha_{new} = \frac{\alpha}{2}$; o valor de γ utilizado foi $\gamma = 10^{-4}$. A escolha por uma busca linear não-monótona foi adotada devido à observação de que, durante os testes numéricos preliminares, o valor da norma da função do sistema (quando aplicada nos pontos gerados pelo método ISR1) em alguns momentos era superior ao valor da norma no ponto anterior (especialmente durante as iterações iniciais); entretanto, caso aceitássemos esse novo ponto, o método conseguia fazer com que os próximos iterandos reduzissem significativamente o valor da norma da função, tornando bastante interessante a possibilidade do uso de uma busca linear que permitisse tal fato.

A seqüência η_k foi tomada como $\eta_k = \frac{ftip}{(k+1)^{1.1}}$, onde $ftip$ é dado por

$$\begin{aligned} ftip &= \|F(x_0)\|, & \text{se } k = 0, \\ ftip &= \min \{\|F(x_k)\|, ftip\}, & \text{se } k \text{ é múltiplo de } 10. \end{aligned}$$

Se k não é múltiplo de 10, $ftip$ não se altera.

Em todas as nossas implementações numéricas, é declarada a convergência do Algoritmo 3.5.1 quando $\|F(x_k)\|_\infty \leq 10^{-5}$ e $k < 200$.

Durante a implementação numérica, caso

$$|y_k^T (s_k - H_k y_k)| \leq c_1 \|s_k - H_k y_k\|_2 \max \{\|s_k - H_k y_k\|_2, \|y_k\|_2\} \quad (3.23)$$

tomamos $H_{k+1} = H_k$ ao invés de utilizarmos a fórmula de atualização (3.10).

A primeira iteração, assim como os reinícios realizados a cada p iterações, são feitos por uma iteração do tipo Newton; para tanto, calcula-se a fatoraçoão LU da matriz Jacobiana da função F . No caso da Jacobiana ser uma matriz esparsa, é utilizada uma fatoraçoão LU esparsa. Para a implementação computacional foram utilizadas rotinas em Fortran, disponíveis no pacote IMSL [17]. Essas rotinas são denominadas DLFTXG (para o cálculo da fatoraçoão) e DLFSXG para a resolução do sistema linear.

A rotina DLFTXG calcula a fatoraçoão LU seguindo o algoritmo proposto em [17] (estratégia de Markowitz simétrica), a qual escolhe pivôs que tentam reduzir o preenchimento da matriz ao mesmo tempo que mantém a estabilidade numérica. As matrizes de permutação das linhas e das colunas são determinadas pela estratégia de Markowitz [23].

A biblioteca IMSL para Fortran (International Mathematical & Statistical Library) contém cerca de 100 subrotinas e funções (em formato Fortran) que permitem acessar implementações de métodos numéricos (matemáticos e estatísticos) de alta qualidade. Dentre as rotinas matemáticas, pode-se destacar: Álgebra Linear, análise de autovalores, interpolação e aproximação, integração e diferenciação, resolução de equações diferenciáveis, transformadas de Laplace e Fourier, resolução de sistemas não-lineares, etc.².

3.5.1 Detalhamento do algoritmo

Considere k_0 o múltiplo de p mais próximo de k (inferiormente). Seja $I_C = \{l < p \mid H_{k_0+l} \neq H_k\}$.

Conforme visto anteriormente, H_{k+1} é uma correção de posto um de H_k e, por esta razão, a implementação é feita tomando em conta que

$$H_k = H_{k_0} + \sum_{l \in I_C} \begin{bmatrix} v_{k_0+l} v_{k_0+l}^T \\ v_{k_0+l}^T y_{k_0+l} \end{bmatrix},$$

²As rotinas IMSL são inseridas da mesma forma que rotinas comuns do Fortran, desde que a biblioteca IMSL esteja encadeada ao arquivo executável. Maiores informações estão disponíveis na página da internet: <http://www.utexas.edu/its/rc/answers/imsl/imslf.html>.

Assim, é necessário armazenar apenas os vetores $v_{k_0+l} = s_{k_0+l} - H_{k_0+l}y_{k_0+l}$ e y_{k_0+l} , para todo $l \in I_C$. Portanto, a expressão para d_k pode ser escrita como

$$d_k = -H_{k_0}F(x_k) - \sum_{l \in I_C} \left[\frac{v_{k_0+l}^T F(x_k)}{v_{k_0+l}^T y_{k_0+l}} v_{k_0+l} \right];$$

conseqüentemente, o *Passo 1* do Algoritmo 3.5.1 pode ser reescrito da seguinte maneira:

- Cálculo de $d_k = H_{k_0}F(x_k)$:

- Resolver o sistema (determinar p_k)

$$L_{k_0}p_k = P_{k_0}F(x_k);$$

- Resolver o sistema (determinar q_k)

$$U_{k_0}q_k = p_k$$

- Cálculo de d_k

$$d_k = Q_{k_0}q_k$$

- Adicionando as demais parcelas

Para $l \in I_C$

$$d_k = d_k + \frac{v_{k_0+l}^T F(x_k)}{v_{k_0+l}^T y_{k_0+l}} v_{k_0+l}$$

Faça $d_k = -d_k$;

onde $P_{k_0}J(x_{k_0})Q_{k_0} = L_{k_0}U_{k_0}$ correspondem às matrizes da fatoração LU com pivoteamento total da matriz Jacobiana em x_{k_0} .

3.6 Resultados de convergência

Muitos trabalhos presentes na literatura apresentam resultados similares aos aqui expostos; entretanto, até onde sabemos, a análise de convergência para métodos quase-Newton implementados com uma busca linear não-monótona ainda não foi publicada.

Iniciamos esta seção com algumas observações:

Observação 3.6.1. *Devido aos fatos de que a função F é contínua e que $\eta_k > 0, \forall k$, temos que o Algoritmo 3.5.1 está bem definido, uma vez que a condição (3.22) é satisfeita para um valor de α suficientemente pequeno. Portanto, a busca linear não-monótona terá êxito em toda iteração k , independentemente da escolha de d_k .*

Observação 3.6.2. *Apartir da desigualdade (3.22), temos que,*

$$\|F(x_{k+1})\| \leq [1 - \gamma\alpha_k] \|F(x_k)\| + \eta_k \leq \|F(x_k)\| + \eta_k$$

e portanto, temos que $\|F(x_{k+1})\| \leq \|F(x_0)\| + \eta_k + \dots + \eta_0 \leq \|F(x_0)\| + \eta$ para todo k . Logo, a seqüência $\{x_k\}$ gerada pelo Algoritmo 3.5.1 está contida em $L = \{x \in \mathbb{R}^n \mid \|F(x)\| \leq \|F(x_0)\| + \eta\}$.

A seguir, apresentamos os resultados de convergência obtidos para o Algoritmo 3.5.1, quando d_k é determinado a partir de um método quase-Newton qualquer, não necessariamente o ISR1, utilizado na nossa implementação do Algoritmo 2.2.1. Tais resultados estão baseados naqueles demonstrados por Birgin, Krejič e Martínez em [10].

Os Lemas 3.6.1 e 3.6.2 são resultados teóricos necessários para a demonstração do Teorema 3.6.1. Tanto os lemas quanto o teorema acima citados encontram-se demonstrados em [10], e estão aqui escritos apenas para facilitar a leitura dos resultados consequentes.

Lema 3.6.1. *Seja $\{x_k\}$ uma seqüência gerada pelo Algoritmo 3.5.1. Se, para alguma seqüência de índices $K_0 \subset \mathbb{N}$, $\lim_{k \in K_0} \|F(x_k)\| = 0$, então*

$$\lim_{k \rightarrow \infty} \|F(x_k)\| = 0. \quad (3.24)$$

Em particular, se x_ é um ponto limite de $\{x_k\}$ tal que $\|F(x_*)\| = 0$, então todo ponto limite da seqüência é uma solução de (3.1).*

Demonstração. Ver [10], Lemma 1. ■

Lema 3.6.2. *Seja $\{x_k\}$ uma seqüência gerada pelo Algoritmo 3.5.1 e suponha que todos os pontos limites da seqüência $\{x_k\}$ são soluções do sistema (3.1). Além disso, suponha que x_* é um ponto limite tal que $J(x_*)$ é inversível e que*

$$\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0. \quad (3.25)$$

Então, toda a seqüência converge para x_ .*

Demonstração. Ver [10], Lemma 2. ■

O parâmetro θ , presente no enunciado do teorema a seguir, controla a inexatidão da obtenção da direção de descida d_k , uma vez que o trabalho [10] versa sobre métodos Newton-inexatos. Como utilizamos um método quase-Newton (exato) para obter d_k , consideramos $\theta = 0$.

Teorema 3.6.1. *Suponha que $\{x_k\}$ seja uma seqüência gerada pelo Algoritmo 3.5.1 e que existe $M > 0$ tal que, para uma seqüência infinita de índices $k_1 \subset \{0, 1, 2, \dots\}$,*

$$\|J(x_k)d_k\| \leq \theta \|F(x_k)\|$$

e

$$d_k \leq M.$$

Então, qualquer ponto limite de $\{x_k\}_{k \in K_1}$ é uma solução do sistema (3.1). Além disso, se um ponto limite de $\{x_k\}_{k \in K_1}$ existe, então $\{F(x_k)\} \rightarrow 0$ e todo ponto limite de $\{x_k\}$ é uma solução de (3.1).

Demonstração. Ver [10], Theorem 2. ■

O teorema a seguir é uma consequência direta do Teorema 3.6.1.

Teorema 3.6.2. *Suponha que $\{x_k\}$ seja gerada pelo Algoritmo 3.5.1. Seja K_1 o conjunto de índices k múltiplos de p . Suponha que, para todo $k \in K_1$, $J(x_k)$ é inversível e que $\|J(x_k)^{-1}\| \leq c$. Então, qualquer ponto limite de $\{x_k\}_{k \in K_1}$ é uma solução do sistema (3.1). Além disso, se um ponto limite de $\{x_k\}_{k \in K_1}$ existe, então $\{\|F(x_k)\|\}$ converge para 0 e todo ponto limite de $\{x_k\}$ é uma solução do sistema (3.1).*

Demonstração. Pela Observação 2, temos que a seqüência $\{F(x_k)\}$ está contida em $L = \{x \in \mathbb{R}^n \mid \|F(x)\| \leq \|F(x_0)\| + \eta\}$; logo, a seqüência $\{F(x_k)\}$ é limitada. Para todo $k \in K_1$, temos que $\|J(x_k)^{-1}\| \leq c$

$$\|d_k\| = \|-H_k F(x_k)\| = \|J(x_k)^{-1} F(x_k)\| \leq \|J(x_k)^{-1}\| \|F(x_k)\| \leq c [\|F(x_0)\| + \eta];$$

logo, $\{d_k\}_{k \in K_1}$ é limitada.

Assim, tomando $\theta = 0$, estamos sob as hipóteses do Teorema 3.6.1. Portanto, se um ponto limite de $\{x_k\}_{k \in K_1}$ existe, então $F(x_k) \rightarrow 0$ e todo ponto limite de $\{x_k\}$ é uma solução de (3.1). ■

Teorema 3.6.3. *Suponha que $\{x_k\}$ seja uma seqüência gerada pelo Algoritmo 3.5.1, que o conjunto de nível $L = \{x \in \mathbb{R}^n \mid \|F(x)\| \leq \|F(x_0)\| + \eta\}$ seja limitado e que $J(x)$ seja inversível para todo $x \in L$. Então, existe $x_* \in \mathbb{R}^n$ uma solução de (3.1) tal que*

$$\lim_{k \rightarrow \infty} x_k = x_*.$$

Demonstração. Dado que $J(x)$ é contínua e inversível para todo $x \in L$ e L é um conjunto compacto, então existe $c > 0$ tal que $\|J(x)^{-1}\| \leq c$ para todo $x \in L$. Pela Observação 2, a seqüência $\{x_k\}$ está contida em L , logo, $\{x_k\}$ possui pelo menos um ponto limite. Pelo Teorema 3.6.2, todos os pontos limites são soluções do sistema (3.1).

Para concluir a demonstração deste teorema, basta provar que toda a seqüência converge para um desses pontos limites.

Inicialmente, vamos provar que a seqüência $\{\|H_k\|\}$ é limitada. Caso $k \equiv 0 \pmod{p}$, a limitação é obtida de forma direta uma vez que

$$\|H_k\| = \|J(x_k)^{-1}\| \leq c, \quad \text{para todo } k \equiv 0 \pmod{p}.$$

De uma forma geral, ou temos que $H_{k+1} = H_k$ ou

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

Portanto, devido à garantia de limitação inferior do denominador (ver (3.23)), temos que

$$\|H_{k+1}\| \leq \|H_k\| + \frac{\|s_k - H_k y_k\| \|s_k - H_k y_k\|}{|(s_k - H_k y_k)^T y_k|} \leq \|H_k\| + \frac{1}{c_1}.$$

Assim, se k é múltiplo de p e $l \in \{1, \dots, p-1\}$,

$$\|H_{k+l}\| \leq \|H_k\| + \frac{l}{c_1} \leq c + \frac{l}{c_1},$$

o que implica que

$$\|H_{k+l}\| < c + \frac{p-1}{c_1}, \quad \forall k \in \mathbb{N}. \quad (3.26)$$

Logo $\{H_k\}$ é limitada, e pelo Teorema 3.6.2, temos que $\{F(x_k)\} \rightarrow 0$. Portanto

$$\|d_k\| = \|-H_k F(x_k)\| \leq \|-H_k\| \|F(x_k)\|$$

e então $\|d_k\| \rightarrow 0$.

Como $\alpha_k \leq 1$, para todo k , temos

$$\|x_{k+1} - x_k\| = \|s_k\| = \|\alpha_k d_k\| \leq \alpha_k \|d_k\|$$

e conseqüentemente, $\{\|x_{k+1} - x_k\|\} \rightarrow 0$.

Logo, as hipóteses do Lema 3.6.2 são satisfeitas, e temos que toda a seqüência converge para o ponto limite. ■

A definição a seguir será utilizada na demonstração do Teorema 3.6.4, que versa sobre a convergência superlinear. O Teorema 3.6.4, assim como o Teorema 3.6.3, são válidos para qualquer método quase-Newton que satisfaça essa definição. No Teorema 3.6.6 demonstraremos que o método ISR1, em particular, é um dos métodos quase-Newton que a satisfaz.

Definição 3.6.1. *A seqüência de matrizes $\{H_k\}$ satisfaz a propriedade WBD (Weak Bounded Deterioration) inversa (ver [10]), se existem constantes $d_1, d_2 > 0$ independentes de k , tais que $\forall k = 0, 1, \dots$,*

$$\|H_{k+1} - J(x_{k+1})^{-1}\| \leq d_1 \|H_k - J(x_k)^{-1}\| + d_2 \|s_k\|. \quad (3.27)$$

O próximo teorema diz que, sob as hipóteses do Teorema 3.6.3, o Algoritmo 3.5.1 converge superlinearmente para uma solução do sistema (3.1).

Teorema 3.6.4. *Suponha que sejam válidas as hipóteses do Teorema 3.6.3. Além disso, suponha que $J(x)$ é Lipschitz contínua em um conjunto convexo que contém o conjunto de nível L , e que o método quase-Newton gere uma seqüência de matrizes H_k que satisfaça a propriedade WBD inversa. Então, $\alpha_k = 1$ para k suficientemente grande e $\{x_k\}$ converge superlinearmente para x^* , solução de (3.1).*

Demonstração. Assim como feito no teorema anterior, pode-se demonstrar que as matrizes H_k são limitadas, isto é

$$\|H_k\|_2 \leq M, \quad \forall k \in \mathbb{N}.$$

Da propriedade WBD, temos que existem $d_1, d_2 > 0$ tais que para todo k ,

$$\|H_{k+1} - J(x_{k+1})^{-1}\| \leq d_1 \|H_k - J(x_k)^{-1}\| + d_2 \|s_k\|.$$

Como $\|s_k\| \rightarrow 0$ e $H_k = J(x_k)^{-1}$ sempre que k é múltiplo de p , então

$$\lim_{k \rightarrow \infty} H_k = J(x^*)^{-1}.$$

Além disso, $H_k = B_k^{-1}$, $B_k \rightarrow J(x^*)$ e $x_{k+1} = x_k - \alpha_k B_k^{-1} F(x_k)$ para todo k . Portanto, tomando $\theta_k = 0, \forall k$, a conclusão da demonstração segue de Teorema 3.6.5, o qual foi demonstrado por Birgin, Krejič e Martínez [10]. ■

Teorema 3.6.5. *Suponha que sejam válidas as hipóteses do Teorema 3.6.4. Além disso, suponha que*

$$\|B_k d_k + F(x_k)\| \leq \theta_k \|F(x_k)\|$$

para todo k suficientemente grande, onde $\lim_{k \rightarrow \infty} \theta_k = 0$. Então, $\{x_k\}$ converge superlinearmente para x^* , solução de (3.1).

Demonstração. Ver [10], Theorem 3. ■

A seguir, demonstraremos que o método quase-Newton ISR1 satisfaz a propriedade WBD inversa.

Teorema 3.6.6. *As matrizes H_k geradas pelo método ISR1 (3.10) satisfazem a propriedade WBD inversa.*

Demonstração. Considerando que as atualizações de posto um são realizadas (ver (3.23)) somente se

$$|y_k^T (s_k - H_k y_k)| > c_1 \|s_k - H_k y_k\|_2 \max \{\|s_k - H_k y_k\|_2, \|y_k\|_2\} \quad (3.28)$$

Então, para os valores de k tais que $H_{k+1} \neq H_k$,

$$H_{k+1} - J(x_{k+1})^{-1} =$$

$$\begin{aligned}
&= H_k - J(x_k)^{-1} + J(x_k)^{-1} - J(x_{k+1})^{-1} + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)} \\
&= J(x_k)^{-1} - J(x_{k+1})^{-1} + H_k - J(x_k)^{-1} \\
&\quad + \frac{(J(x_k)^{-1} y_k - H_k y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)} + \frac{(s_k - J(x_k)^{-1} y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)} \\
&= J(x_k)^{-1} - J(x_{k+1})^{-1} \\
&\quad + (H_k - J(x_k)^{-1}) \left[I - \frac{y_k (s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)} \right] + \frac{(s_k - J(x_k)^{-1} y_k)(s_k - H_k y_k)^T}{y_k^T (s_k - H_k y_k)}.
\end{aligned}$$

Portanto,

$$\begin{aligned}
&\|H_{k+1} - J(x_{k+1})^{-1}\| \leq \|J(x_k)^{-1} - J(x_{k+1})^{-1}\| \\
&\quad + \|H_k - J(x_k)^{-1}\| \left[1 + \frac{\|y_k\| \|s_k - H_k y_k\|}{|y_k^T (s_k - H_k y_k)|} \right] + \frac{\|s_k - J(x_k)^{-1} y_k\| \|s_k - H_k y_k\|}{|y_k^T (s_k - H_k y_k)|} \\
&\leq \|J(x_k)^{-1} - J(x_{k+1})^{-1}\| \\
&\quad + \|H_k - J(x_k)^{-1}\| \left[1 + \frac{1}{c_1} \right] + \frac{\|s_k - J(x_k)^{-1} y_k\| \|s_k - H_k y_k\|}{|y_k^T (s_k - H_k y_k)|}
\end{aligned}$$

Como $J(x)$ é Lipschitz contínua e $J(x)^{-1}$ é limitada, então, $J(x)^{-1}$ é também Lipschitz contínua; então, existe $c_2 > 0$ tal que

$$\|J(x)^{-1} - J(y)^{-1}\| \leq c_2 \|x - y\|, \quad (3.29)$$

para todo $x, y \in \mathbb{R}^n$. Isso implica que

$$\|s_k - J(x_k)^{-1} y_k\| \leq \frac{c_2}{2} \|s_k\|^2, \quad (3.30)$$

para todo $k = 0, 1, \dots$. Logo, por (3.23), (3.29) e (3.30),

$$\|H_{k+1} - J(x_{k+1})^{-1}\| \leq c_2 \|s_k\| + \|H_k - J(x_k)^{-1}\| \left[1 + \frac{1}{c_1} \right] + \frac{c_2}{2c_1} \frac{\|s_k\|^2}{\max\{\|s_k - H_k y_k\|_2, \|y_k\|_2\}}.$$

Entretando, como $\|H_k\| \leq M_4, \forall k$, onde $M_4 = c + (p - 1)/c_1$ (ver desigualdade (3.26)), temos que $\|s_k\| \leq 2M_4\|y_k\| \leq 2M_4 \max \{\|s_k - H_k y_k\|_2, \|y_k\|_2\}$. Logo

$$\|H_{k+1} - J(x_{k+1})^{-1}\| \leq c_2\|s_k\| + \|H_k - J(x_k)^{-1}\| \left[1 + \frac{1}{c_1}\right] + \frac{c_2}{c_1} M_4 \|s_k\|. \quad (3.31)$$

Quando $H_{k+1} = H_k$, temos

$$\begin{aligned} H_{k+1} - J(x_{k+1})^{-1} &= H_k - J(x_k)^{-1} + J(x_k)^{-1} - J(x_{k+1})^{-1} \\ \|H_{k+1} - J(x_{k+1})^{-1}\| &\leq \|H_k - J(x_k)^{-1}\| + \|J(x_k)^{-1} - J(x_{k+1})^{-1}\|. \end{aligned}$$

De (3.29), temos que

$$\|J(x_k)^{-1} - J(x_{k+1})^{-1}\| \leq c_2 \|x_k - x_{k+1}\|.$$

Logo

$$\|H_{k+1} - J(x_{k+1})^{-1}\| \leq \|H_k - J(x_k)^{-1}\| + c_2 \|s_k\| \quad (3.32)$$

Portanto, de (3.31) e (3.32), temos que as matrizes H_k satisfazem a propriedade WBD inversa com

$$d_1 = \max \left\{ 1, 1 + \frac{1}{c_1} \right\} = 1 + \frac{1}{c_1} \text{ e } d_2 = \max \left\{ c_2, c_2 + M_4 \frac{c_2}{c_1} \right\} = c_2 + M_4 \frac{c_2}{c_1}.$$

■

3.7 Conclusão

Apresentamos neste capítulo o método quase-Newton ISR1; tal método será utilizado em nossas implementações numéricas como acelerador do método Lagrangiano Aumentado, conforme descrições futuras (Capítulos 5 e 6).

Na Seção 3.6, demonstramos resultados de convergência global para este método, quando implementado com uma busca linear não-monótona. Tais resultados permitirão assegurar a convergência do método proposto neste trabalho, que é o método Lagrangiano Aumentado Acelerado.

Capítulo 4

Método Lagrangiano Aumentado Acelerado

Uma vez feita uma análise cuidadosa dos métodos Lagrangiano Aumentado e ISR1, bem como um estudo das propriedades de convergência e implementação computacional dos mesmos, podemos descrever o tema considerado central da Tese, que é usar um método quase-Newton (o ISR1, no caso) para acelerar a convergência do método Lagrangiano Aumentado. Neste capítulo descrevemos, de uma forma geral, como é introduzido o acelerador quase-Newton no método Lagrangiano Aumentado adotado. Apresentamos o algoritmo proposto bem como um esquema ilustrativo de como ele funciona.

Os experimentos numéricos são apresentados nos próximos dois capítulos, pois a implementação do algoritmo para problemas só com restrições de igualdade é bem mais simples do que a implementação para problemas com restrições gerais. Além disso, o desempenho do algoritmo não é o mesmo em ambos os casos e merece comentários específicos.

Uma outra possibilidade para obter um algoritmo com boas qualidades de convergência para resolver problemas de minimização com restrições, seria utilizar o método Lagrangiano Aumentado como acelerador para o método quase-Newton (invertendo-se a ordem do algoritmo proposto), o que não se justifica teoricamente, como veremos na Seção 5.7.4. No Capítulo 5 são apresentados também testes numéricos, indicativos de que a ordenação da seqüência de passos do algoritmo que foi proposto é realmente melhor. Tais testes se encontram nesse capítulo, por utilizarem problemas específicos que

lá são descritos.

Inicialmente, apresentamos os argumentos utilizados para fundamentar o algoritmo proposto, descritos na Seção 4.2. Na Seção 4.3 apresentamos a justificativa para a convergência do método proposto, e na seção seguinte, apresentamos um exemplo numérico para ilustrar o comportamento do acelerador e do método Lagrangiano Aumentado quando são implementados em separado.

4.1 Introdução

No Capítulo 2, apresentamos o algoritmo Lagrangiano Aumentado para obter a solução do problema de minimização com restrições gerais (2.1). Uma outra opção para encontrar a solução para o problema de minimização em questão é resolver o sistema não-linear formado pelas condições de otimalidade do problema, que são conhecidas como condições KKT [33].

Os métodos mais comuns para resolver sistemas não-lineares são os métodos de Newton e quase-Newton [19], [24], conforme descritos no Capítulo 3. Em geral, resolver o sistema não-linear associado não se mostra uma alternativa muito eficiente, uma vez que a diferença entre minimizadores e maximizadores não é por eles detectada. Entretanto, se o ponto inicial está suficientemente próximo do minimizador, há uma grande chance do método convergir para o minimizador e, melhor ainda, com mais rapidez.

A implementação dos algoritmos do tipo Lagrangiano Aumentado está diretamente relacionada com as estratégias de modificação de um ou mais parâmetros de penalidade: se o valor dos parâmetros são altos, o problema de minimização associado torna-se difícil de ser resolvido; por outro lado, o aumento desse parâmetro de forma lenta ocasiona uma convergência lenta do processo.

Motivados por essas observações o algoritmo que propomos tem o objetivo de combinar os dois métodos, explorando as boas qualidades de cada um deles; combinamos o método Lagrangiano Aumentado descrito no Capítulo 2, que tem a excelente característica de convergência global para um minimizador, com o método Inverso de Correção de Posto Um descrito no Capítulo 3, bastante simples do ponto de vista de implementação computacional. O ISR1 é aplicado na resolução do sistema de otimalidade implementado com uma busca linear não-monótona, com o propósito de acelerar a convergência do

método Lagrangiano Aumentado.

4.2 O algoritmo proposto

Assim, o algoritmo proposto consiste em integrar o método Lagrangiano Aumentado com o método quase-Newton ISR1. A idéia então, é que o método Lagrangiano Aumentado possa fornecer bons pontos iniciais para garantir a convergência do método quase-Newton para um minimizador. Essa integração está descrita no Algoritmo 4.2.1 a seguir, o qual apresenta a idéia geral dessa união. O ponto inicial para execução do Passo 1 do Algoritmo 4.2.1 é o mesmo ponto inicial que seria fornecido para o método Lagrangiano Aumentado, implementado sem o acelerador; para iniciar a execução do Passo 2 desse algoritmo, utilizamos a última aproximação da solução fornecida de pelo Passo 1. O esquema da Figura 4.1 esclarece melhor a idéia do método.

Algoritmo 4.2.1. *Lagrangiano Aumentado Acelerado:*

Passo 1 - *Execute uma única iteração externa do método Lagrangiano Aumentado para a resolução do problema de minimização. Utilize a solução obtida como ponto inicial para o próximo Passo.*

Passo 2 - *Execute um determinado número de iterações quase-Newton na resolução do sistema de otimalidade. Se o valor da norma do sistema, no iterando atual, diminuir de acordo com um critério previamente adotado, repita este Passo.*

Caso contrário, retorne ao Passo 1, após efetuar as atualizações padrões para as estimativas dos vetores de multiplicadores de Lagrange e para o parâmetro de penalidade.

No esquema a seguir (Figura 4.1), ser solução significa que:

- (1) Os vetores x_0, λ_0 e μ_0 satisfazem o teste inicial de convergência do método Lagrangiano Aumentado;
- (2) Os vetores x_k, λ_k e μ_k satisfazem o critério de parada do método Lagrangiano Aumentado ou o teste inicial de convergência do método quase-Newton;
- (3) Os vetores x_k, λ_k e μ_k satisfazem o critério de parada do método quase-Newton.

Em todos os casos, “Parar” significa que o ponto de mínimo foi encontrado. Caso o critério de permanência (4.1) não for satisfeito, o método quase-Newton é interrompido,

e os vetores x_j, λ_j e μ_j que serão repassados para o método Lagrangiano Aumentado serão aqueles que apresentaram o menor valor da norma da Função KKT durante a execução do quase-Newton.

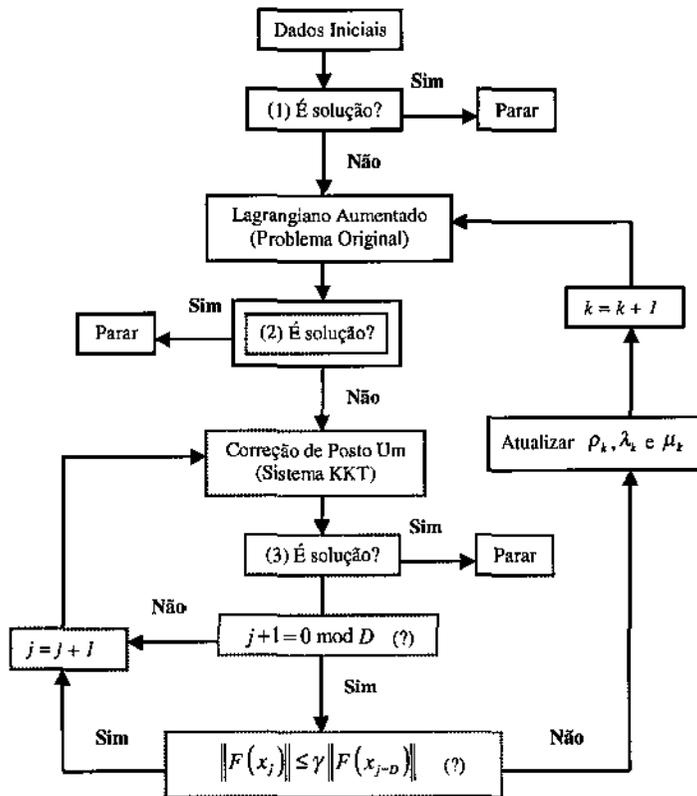


Figura 4.1: Esquema descritivo do Lagrangiano Aumentado Acelerado.

Conforme podemos observar no esquema da Figura 4.1, o número de iterações de ISR1 que serão executadas no Passo 2 do algoritmo é determinado pelo seguinte teste:

$$\|F(x_k)\| \leq \gamma \|F(x_{k-D})\|; \quad \text{se } k \equiv 0 \pmod{D}, \quad (4.1)$$

onde o parâmetro D determina a periodicidade em que o teste será realizado. Se o ponto x_k obtido não satisfaz a condição (4.1), é encerrada a iteração externa e retornamos ao *Passo 1* do algoritmo (ou seja, retornamos ao método Lagrangiano Aumentado).

Em outras palavras, a continuação ou não da execução do método quase-Newton é determinada por meio da realização de um teste periódico, no qual verificamos se a norma

da função KKT está sofrendo um certo percentual γ de decréscimo em relação a um valor da norma, previamente estabelecido. Se esse decréscimo não é verificado, é interrompida a execução do método quase-Newton e retorna-se ao Lagrangiano Aumentado.

É necessário salientar a diferença entre os tipos de iterações:

- iteração externa do método Lagrangiano Aumentado implementado sozinho: consiste na obtenção da solução do subproblema (2.7), com os vetores λ_k e μ_k fixos;
- iteração interna do método Lagrangiano Aumentado implementado sozinho: consiste em cada uma das iterações realizadas para resolver cada subproblema;
- iteração externa do método Lagrangiano Aumentado Acelerado: consiste na execução dos Passos 1 e 2 do Algoritmo 4.2.1, ou seja, na realização de uma iteração externa do método Lagrangiano Aumentado e em seguida, na execução do método quase-Newton até que um critério de convergência seja satisfeito ou até que o critério de permanência seja violado.

4.3 Análise Teórica

Temos assegurada a convergência do algoritmo proposto, uma vez que existem resultados de convergência global para cada um dos dois métodos conjugados: Lagrangiano Aumentado e ISR1. Tal resultado pode ser melhor observado utilizando a seguinte análise:

- Caso estejamos na primeira execução do Passo 1 do Algoritmo 4.2.1, temos a convergência assegurada devido aos resultados teóricos existentes para o método Lagrangiano Aumentado.
- Caso estejamos no Passo 2 do Algoritmo 4.2.1, o ponto fornecido pelo método Lagrangiano Aumentado é adotado como ponto inicial da execução do método ISR1, e portanto, a convergência está assegurada devido aos resultados apresentados no Capítulo 3 desta Tese.
- Por fim, caso estejamos no Passo 1 do Algoritmo 4.2.1 uma segunda (ou mais) vezes, basta considerar o ponto fornecido pelo método ISR1 como um novo ponto inicial do método Lagrangiano Aumentado, e pode-se garantir a convergência pelo mesmo motivo que o primeiro item.

4.4 Exemplo numérico

Analisaremos aqui o comportamento dos métodos expostos neste trabalho, trabalhando em separado e depois em conjunto: o método ISR1, o método Lagrangiano Aumentado e o método Lagrangiano Aumentado Acelerado.

A execução do Passo 1 antes do Passo 2 no Algoritmo 4.2.1 é de fundamental importância para a convergência do mesmo para um minimizador do problema original. Se o método quase-Newton (ISR1) for aplicado diretamente, ele poderá convergir para um ponto de máximo, ou até mesmo não convergir (caso o ponto inicial não esteja suficientemente perto do minimizador). Na seção 5.7.4 voltaremos a esse tema, fazendo a resolução de um conjunto de problemas teste utilizando as duas ordens possíveis para os Passos.

O problema a seguir, embora bastante simples, ilustra bem o comportamento dos três métodos em questão: Lagrangiano Aumentado sozinho, ISR1 sozinho e Lagrangiano Aumentado Acelerado. A título de ilustração, consideramos o problema de otimização:

$$\begin{aligned} \min & [(x_1 - 2)^2 + (x_2 - 1)^2] \sin [(x_1 - 2)^2 + (x_2 - 1)^2] \\ \text{s.a. } & x_1 \geq 0 \\ & x_2 \geq 0 \\ & x_1 - 2x_2 + 2 \geq 0 \\ & x_1 - 3x_2 - 4 \geq 0 \\ & 2x_1 + 5x_2 - 12.73 \leq 0 \end{aligned} \tag{4.2}$$

Acrescentando às restrições de desigualdade variáveis de folga x_3, \dots, x_7 , ao quadrado, obtemos o problema equivalente:

$$\begin{aligned} \min & [(x_1 - 2)^2 + (x_2 - 1)^2] \sin [(x_1 - 2)^2 + (x_2 - 1)^2] \\ \text{s.a. } & x_1 - x_3^2 = 0 \\ & x_2 - x_4^2 = 0 \\ & x_1 - 2x_2 + 2 - x_5^2 = 0 \\ & x_1 - 3x_2 - 4 - x_6^2 = 0 \\ & 2x_1 + 5x_2 - 12.73 + x_7^2 = 0 \end{aligned} \tag{4.3}$$

Resolvendo o problema acima utilizando os três algoritmos, com $(5, 1)^T$ como ponto

inicial e $\varepsilon = 10^{-4}$ como a precisão para os critérios de parada, obtivemos os seguintes resultados:

- Algoritmo 2.2.1 (Lagrangiano Aumentado): foram necessárias 6 iterações externas, e 23 iterações internas, no total (destas, 8 foram feitas na primeira iteração externa).
Solução encontrada: $(5.29; 0.43)^T$ (minimizador global).
- Algoritmo 3.5.1 (ISR1): foram necessárias 9 iterações.
Solução encontrada: $(4.81; 0.63)^T$ (um dos maximizadores globais - ver Figura 4.2).
- Algoritmo 4.2.1 (Lagrangiano Aumentado Acelerado pelo método ISR1): foi executada apenas 1 iteração externa, ou seja, o Lagrangiano Aumentado foi acionado apenas uma vez, tendo realizado 8 iterações internas. O método quase-Newton ISR1 executou 6 iterações, não sendo necessário fazer recomeços (pois $6 < D$).
Solução encontrada: $(5.29; 0.43)^T$ (minimizador global).

Observe que o método ISR1 sozinho realizou apenas 9 iterações no total; porém, convergiu para um maximizador global. O método Lagrangiano Aumentado sozinho convergiu para o minimizador global, mas precisou realizar um número razoável de iterações externas. Essa divergência de comportamento, neste caso particular, ocorre porque a função possui um único minimizador global e infinitos maximizadores globais (trecho de arco - ver Figura 4.2).

Todavia, quando executamos apenas uma iteração externa do Lagrangiano Aumentado, e fornecemos o ponto obtido como ponto inicial para o ISR1, percebemos que este realiza 6 iterações no total (sem recomeços) e converge para o minimizador global, reduzindo significativamente o número de iterações internas realizadas pelo Lagrangiano Aumentado. Em outras palavras, a execução da primeira iteração pelo método Lagrangiano Aumentado foi capaz de fornecer um bom ponto inicial para o método ISR1, uma vez que com esse novo ponto tal método conseguiu convergir rapidamente para um minimizador.

No gráfico da esquerda da Figura 4.2, estão desenhadas as curvas de nível da função objetivo, assim como o polígono referente à região factível; o ponto inicial está marcado em vermelho; em preto está sinalizado o minimizador global e em azul está marcado o maximizador global para o qual o ISR1 converge. A segunda figura, além do desenho das curvas de nível e da região factível no plano $x - y$, apresenta o desenho da superfície gerada pela função objetivo original.

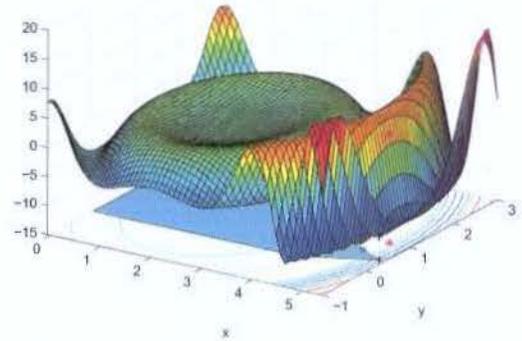
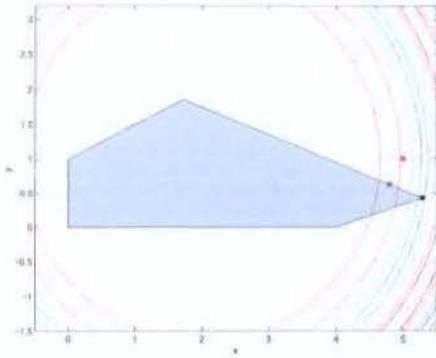


Figura 4.2: Exemplo Numérico - Aceleração do método Lagrangiano Aumentado.

Após essa descrição do algoritmo proposto neste trabalho, passaremos aos Capítulos 5 e 6 que apresentam detalhes da implementação computacional, teste numéricos e comentários sobre o desempenho deste algoritmo. O Capítulo 5 é destinado à resolução de problemas de otimização com restrições de igualdade, e o Capítulo 6 para problemas de otimização com restrições gerais.

Capítulo 5

Método Lagrangiano Aumentado Acelerado - Problemas com Restrições de Igualdade

5.1 Introdução

Consideraremos o problema de minimização com restrições de igualdade definido por:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & h(x) = 0 \end{aligned} \tag{5.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ representa as restrições de igualdade. As funções f e h são consideradas continuamente diferenciáveis; em geral, são funções não convexas.

Considerando a penalização de todas as restrições de igualdade, adicionando-as à função objetivo com o auxílio de uma função de penalização (que depende das restrições, do vetor de multiplicadores de Lagrange e do parâmetro de penalização), o método Lagrangiano Aumentado encontra a solução do problema de minimização original através de uma seqüência de minimizações irrestritas.

Neste capítulo fazemos uma análise do comportamento do algoritmo Lagrangiano Au-

mentado Acelerado proposto no capítulo anterior, quando aplicado à resolução de problemas de minimização com restrições de igualdade. Os mesmos problemas são resolvidos com o método Lagrangiano Aumentado sozinho, e com o método Lagrangiano Aumentado Acelerado pelo método de Newton (programado da mesma forma que a descrita no algoritmo proposto, utilizando o método de Newton no Passo 2 do Algoritmo 4.2.1, ao invés do método ISR1); para comparação dos resultados, é utilizado o traçamento do perfil de desempenho de cada um dos 3 algoritmos, quando aplicado na resolução de um conjunto de problemas de minimização com restrições de igualdade propostos por [34].

Na Seção 2 fazemos uma breve descrição dos métodos Lagrangiano Aumentado específicos para problemas de restrições de igualdade. Apresentamos, na Seção 3, um resumo dos resultados teóricos mais recentes obtidos para o método Lagrangiano Aumentado (sozinho) implementado. Na Seção 4 fazemos uma breve descrição do algoritmo Lagrangiano Aumentado utilizado e das subrotinas destinadas à resolução dos subproblemas. Apresentamos os detalhes da implementação do método Lagrangiano Aumentado Acelerado na Seção 5. Por fim, na Seção 6, apresentamos e analisamos os resultados numéricos obtidos.

5.2 O método Lagrangiano Aumentado

Os métodos Lagrangiano Aumentado foram descritos no Capítulo 2 deste trabalho, quando destinados à resolução de problemas de minimização com restrições gerais. Para problemas com apenas restrições de igualdade, pode-se observar que as condições de Lagrange [24] para (5.1) são dadas pelo sistema de equações não-lineares de dimensão $(n + m)$ a seguir:

$$\begin{cases} \nabla f(x) + J_h(x)^T \lambda = 0 \\ h(x) = 0. \end{cases} \quad (5.2)$$

Supondo que uma das soluções do sistema (5.2) é dada por $(x_*^T; \lambda_*^T)^T$ e definindo a função Lagrangiana da maneira usual,

$$l(x, \lambda) = f(x) + \lambda^T h(x),$$

temos que, na solução de (5.2),

$$\nabla l(x_*; \lambda_*) = 0.$$

Como pode ser observado, x_* é um ponto estacionário da função Lagrangiana (mas pode não ser um minimizador).

De acordo com os comentários anteriores, ocorrem algumas simplificações no Algoritmo 2.2.1, devido à ausência das restrições de desigualdade:

- O subproblema do Passo 1 torna-se irrestrito, e pode ser reescrito como

$$\min f(x) + \lambda_k^T h(x) + \frac{\rho_k}{2} \|h(x)\|_2^2.$$

- Os estimadores dos multiplicadores de Lagrange ficam restritos à seqüência de vetores λ_k .
- O vetor σ_k não existe nessa nova formulação, fazendo com que o Passo 3 desse algoritmo consista apenas em um teste para a verificação (ou não) de uma redução suficiente no valor da norma infinito das restrições de igualdade.
- O critério de parada se reduz a

$$\max \{ \|h(x_k)\|_\infty, \|\nabla f(x_k) + J_h(x_k)^T \lambda_k\|_\infty \} \leq \varepsilon_k.$$

Caso essa desigualdade seja satisfeita, é declarada a convergência do algoritmo, e a execução é interrompida.

Assim, o algoritmo Lagrangiano Aumentado implementado para problemas de minimização com restrições de igualdade é dado a seguir.

Algoritmo 5.2.1. - Lagrangiano Aumentado - restrições de igualdade

Considere o problema de minimização (5.1). Assuma que x_0 é um ponto factível, $\varepsilon > 0$, $\gamma > 1$, $-\infty < \lambda_{\min} < \lambda_{\max} < \infty$, $-\infty < \mu_{\min} < \mu_{\max} < \infty$, $\tau \in (0, 1)$, $0 < \rho_1 \in \mathbb{R}$, $\lambda_1 \in \mathbb{R}^{m_1}$, $\lambda_1 \in [\lambda_{\min}, \lambda_{\max}]^{m_1}$ e $\mu_1 \in \mathbb{R}^{m_2}$, $\mu_1 \in [\mu_{\min}, \mu_{\max}]^{m_2}$. Seja $\{\varepsilon_k\}_{k \in \mathbb{N}}$ uma seqüência de números positivos que converge para zero. Tome inicialmente o contador de iterações externas igual a 1 ($k = 1$).

Passo 1 - Resolver o subproblema

$$\begin{aligned} \min \quad & f(x) + \lambda_k^T h(x) + \frac{\rho_k}{2} \|h(x)\|_2^2 \\ \text{s.a.} \quad & l \leq x \leq u \end{aligned} \tag{5.3}$$

com λ_k e ρ_k fixos, utilizando x_{k-1} como ponto inicial e atribuindo a x_k a solução encontrada.

Passo 2 - *Critério de Parada*

Se

$$\max \{ \|h(x_k)\|_\infty, \|\nabla f(x_k) + J_h(x_k)^T \lambda_k\|_\infty \} \leq \varepsilon_k.$$

tome $x^* = x_k$ e pare a execução do algoritmo.

Passo 3 - *Atualização do vetor de multiplicadores*

Faça $\bar{\lambda}_{k+1} = \lambda_k + \rho_k h(x_k)$;

calcule λ_{k+1} com sendo a projeção de $\bar{\lambda}_{k+1}$ em $[\lambda_{\min}, \lambda_{\max}]^{m1}$.

Passo 4 - *Atualização do parâmetro de penalização*

Se $\|h(x_k)\|_\infty \leq \tau \|h(x_{k-1})\|_\infty$, tome $\rho_{k+1} = \rho_k$.

Caso contrário, $\rho_{k+1} = \gamma \rho_k$.

Passo 5 - *Atualização do contador*

Se $k + 1$ é inferior ao limite máximo de iterações externas, faça $k = k + 1$ e volte para o Passo 1.

Caso contrário, a execução do algoritmo é interrompida e é declarado que o algoritmo não converge.

5.3 Análise teórica

A convergência global dos métodos Lagrangiano Aumentado foi provada por Conn, Gould e Toint [15] e Conn, Gould, Sartenaer e Toint [14], para problemas com restrições de igualdade e restrições lineares usando a hipótese de independência linear de todos os gradientes das restrições ativas.

Um estudo recente sobre algoritmos Lagrangiano Aumentado para problemas de mini-

mização com restrições de igualdade e caixa foi realizado por Andreani, Birgin, Martínez e Schuverdt [2], o qual fornece os seguintes resultados teóricos para o Algoritmo 2.2.1:

- São caracterizados os pontos limites inactíveis (quando existem), utilizando uma hipótese mais fraca que a condição de Independência Linear.
- Pontos factíveis são pontos KKT; a demonstração desse resultado segue através da hipótese de que tais pontos satisfazem a condição de Dependência Linear Positiva Constante (CPLD). Recentemente, foi demonstrado que essa condição é uma Restrição de Qualificação [3] e que é mais geral que a condição de regularidade.
- A hipótese de regularidade é utilizada para demonstrar a limitação dos parâmetros de penalização.

O primeiro resultado apresentado em [2] determina que o Algoritmo 2.2.1 sempre converge para pontos estacionários do problema irrestrito de minimizar o quadrado das inactibilidades.

Teorema 5.3.1. *Suponha que a sequência $\{x_k\}$ é gerada pelo Algoritmo 2.2.1 e que x_* é um ponto limite. Então, x_* é um ponto estacionário do problema*

$$\min \|h(x)\|_2^2 \quad (5.4)$$

Demonstração. Ver [2], Theorem 3.1. ■

O segundo resultado versa sobre que condições um ponto limite factível da sequência gerada pelo Algoritmo 2.2.1 é um ponto KKT. O resultado principal diz que um ponto limite factível é um ponto KKT se satisfaz a condição CPLD.

Definição 5.3.1. CPLD para restrições de igualdade.

Dizemos que um ponto x satisfaz a Condição de Dependência Linear Positiva Constante (CPLD) em relação às restrições de igualdade se a existência de $I_h \subset \{1, \dots, m\}$, $\lambda_i \in \mathbb{R}$ para todo $i \in I_h$ tais que

$$\sum_{i \in I_h} \lambda_i \nabla h_i(x) = 0$$

com $\sum_{i \in I_h} |\lambda_i| > 0$, implica que os gradientes

$$\{\nabla h_i(z)\}_{i \in I_h}$$

são linearmente dependentes para todo z em uma vizinhança de x .

Teorema 5.3.2. *Suponha que $\{x_k\}$ é uma seqüência gerada pelo Algoritmo 2.2.1 e que x_* é um ponto limite factível que satisfaz a restrição de qualificação CPLD. Então, x_* é um ponto KKT.*

Demonstração. Ver [2], Theorem 4.1. ■

O terceiro e último resultado versa sobre a limitação do parâmetro de penalização. Levando em conta que as estimativas dos multiplicadores de Lagrange são limitadas (devido à projeção em uma caixa fixa, em toda iteração), são necessárias as seguintes hipóteses:

- H1. A seqüência $\{x_k\}$, gerada pelo Algoritmo 2.2.1 converge para um ponto KKT $x_* \in \Omega$;
- H2. $\nabla^2 f(x)$ e $\nabla^2 h_i(x)$ existem e são Lipschitz contínuas;
- H3. A matriz

$$\begin{pmatrix} \left[\begin{array}{c} \nabla^2 f(x_*) + \sum_{j=1}^{m_1} [\lambda_*]_j \nabla^2 h_j(x_*) \\ J_h(x_*) \end{array} \right] & J_h(x_*)^T \\ & \mathbf{O} \end{pmatrix}$$

é não-singular;

Assumindo que são válidas as hipóteses acima, é possível demonstrar o seguinte teorema.

Teorema 5.3.3. *Assuma que a seqüência $\{x_k\}$ é gerada pelo Algoritmo 2.2.1 e que ε_k (limites superiores dos testes de parada, para cada iteração k) é tal que*

$$\varepsilon_k = \min \{\varepsilon'_k, \|h(x_k)\|_\infty\}$$

onde $\{\varepsilon'_k\}$ é uma seqüência decrescente que tende a zero. Então, a seqüência de parâmetros de penalização $\{\rho_k\}$ é limitada.

Demonstração. Ver [2], Theorem 5.1. ■

5.4 Algoritmos para resolução dos subproblemas

Inicialmente, o primeiro algoritmo do tipo Lagrangiano Aumentado que foi implementado neste trabalho foi o Milonga [6] (Mixed Integrated Lagrangian Optimization Nonlinear Generalized Algorithm), que foi escrito por J. M. Martínez e E. G. Birgin; tal rotina destina-se à minimização de problemas com restrições de igualdade e caixa. Portanto, a cada iteração externa, é resolvido o subproblema

$$\begin{aligned} \min f(x) + h(x)^T \lambda + \frac{\rho}{2} \|h(x)\|_2^2 \\ \text{s.a } l \leq x \leq u, \end{aligned} \tag{5.5}$$

onde as estimativas para o vetor de multiplicadores de Lagrange (λ) e para o parâmetro de penalidade (ρ) são atualizadas após cada iteração externa.

Milonga foi desenvolvido com propósito didático, o que facilita o primeiro contacto com as rotinas computacionais do tipo Lagrangiano Aumentado; entretando, tal rotina não faz uso de um sistema sofisticado para o armazenamento e a manipulação de matrizes, fazendo com que o mesmo não seja destinado à resolução de problemas de minimização de grande porte.

Uma vez realizado o primeiro contato computacional com o método Lagrangiano Aumentado através da rotina Milonga, adotamos em nossos testes numéricos a rotina Algencan (ver Seção 2.2), a qual utiliza a subrotina Gencan na resolução dos subproblemas do método Lagrangiano Aumentado. Escrita por J. M. Martínez e E. G. Birgin, a rotina Gencan também é baseada no algoritmo descrito em [6]. Entretanto, ela é voltada para problemas de minimização com restrições de caixa para problemas de grande porte, sendo assim mais condizente com a forma de implementação do método quase-Newton.

5.5 O método Lagrangiano Aumentado Acelerado

As condições de otimalidade de (5.1), dadas por (5.2), são satisfeitas pelas soluções da equação não-linear por blocos $F(x, \lambda)$

$$F(x, \lambda) = \begin{pmatrix} \nabla f(x) + J_h(x)^T \lambda \\ h(x) \end{pmatrix} = 0; \tag{5.6}$$

onde $x \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}^{m_1}$. Ou seja, queremos encontrar (x^*, λ^*) tais que $F(x^*, \lambda^*) = 0$.

A matriz Jacobiana associada à função F é a matriz por blocos

$$J_F(x, \lambda) = \begin{pmatrix} \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) & J_h(x)^T \\ J_h(x) & \mathbf{0} \end{pmatrix}. \quad (5.7)$$

Logo, dadas aproximações iniciais para x e λ , suficientemente próximas do minimizador, os iterandos serão obtidos através do método quase-Newton ISR1 por:

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - H_k \begin{pmatrix} \nabla f(x_k) + J_h(x_k)^T \lambda_k \\ h(x_k) \end{pmatrix}, \quad (5.8)$$

onde a fórmula de atualização da matriz H_k é dada em (3.10).

O primeiro bloco da Jacobiana (5.7) é formado pela combinação das Hessianas da função objetivo e de cada uma das restrições; então, para o conjunto de problemas teste adotado neste capítulo (os quais possuem Hessianas com estrutura esparsa), as estimativas dessas Hessianas foram realizadas de forma discreta, utilizando o método CPR [18], [25]. Buscamos aproximações para as matrizes Hessianas exatas, uma vez que o método Lagrangiano Aumentado não utiliza Hessianas exatas em sua implementação.

O método CPR, desenvolvido por Curtis, Powell e Reid, reduz o número de avaliações dos gradientes na composição das Hessianas, quando as avaliações são feitas utilizando o método de diferenças finitas. A idéia do método CPR consiste em formar grupos de colunas da Hessiana que podem ser estimadas em conjunto, usando apenas uma avaliação do gradiente. Para pertencer a um mesmo grupo, as colunas devem ter todos os seus elementos não-nulos em linhas diferentes.

No exemplo a seguir, ilustramos como foi realizada a estimativa da Hessiana da função objetivo do primeiro problema-teste implementado neste capítulo utilizando o método CPR. As Hessianas das restrições também foram calculadas com o método CPR.

Exemplo 5.5.1. *Dada a função objetivo*

$$F(x) = \sum_{i=1}^{n-1} [100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2];$$

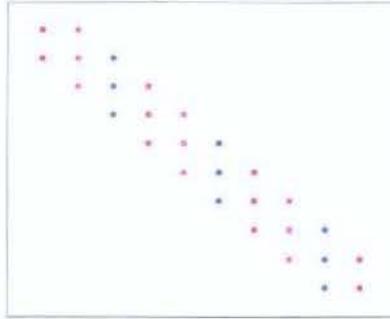


Figura 5.1: Exemplo do método CPR - Estrutura esparsa.

a estrutura de sua matriz Hessiana com $n = 10$ é representada pela Figura 5.1.

São necessários apenas 3 avaliações de gradiente para estimar todas as colunas da matriz Hessiana, uma vez que todas as colunas podem ser reunidas em 3 grupos (diferenciados na figura anterior por cores):

$$\begin{aligned}
 G1 &\approx \frac{g(x + h\mathbf{e}_1 + h\mathbf{e}_4 + h\mathbf{e}_7 + h\mathbf{e}_{10}) - g(x)}{h} \\
 G2 &\approx \frac{g(x + h\mathbf{e}_2 + h\mathbf{e}_5 + h\mathbf{e}_8) - g(x)}{h} \\
 G3 &\approx \frac{g(x + h\mathbf{e}_3 + h\mathbf{e}_6 + h\mathbf{e}_9) - g(x)}{h},
 \end{aligned}$$

o que reduz o número de avaliações da função gradiente de 11 para 4; neste exemplo em particular, independentemente da dimensão adotada, serão sempre necessárias 3 avaliações extras do gradiente, ao invés de n (onde n é o número de variáveis do problema).

As estruturas esparsas das Hessianas da função objetivo e de cada uma das restrições, além de permitir o uso do método CPR em suas estimativas, faz com que o armazenamento dos seus dados possa ser realizado em apenas 3 vetores: um vetor contendo o índice das linhas, outro contendo o índice das colunas, e o terceiro vetor contendo o respectivo elemento não-nulo da matriz em questão. Esse tipo de armazenamento também é utilizado para a matriz Jacobiana das restrições, e conseqüentemente, a Jacobiana do sistema de otimalidade também é armazenada dessa forma.

Todas as rotinas que envolvem essas matrizes foram devidamente adaptadas para trabalhar com esse formato de armazenamento.

Com o intuito de evitar problemas de *overflow* e/ou *underflow* causados pela presença da função exponencial em alguns problemas teste utilizados, a função exponencial foi ligeiramente modificada para

$$\exp(x) = \begin{cases} \exp(10) \left[1 + (x - 10) + \frac{(x - 10)^2}{2} \right], & \text{se } x > 10; \\ \exp(x), & \text{se } -21 \leq x \leq 10; \\ 0, & \text{se } x < -21. \end{cases}$$

5.6 Testes numéricos

Nesta seção apresentamos os resultados obtidos por uma série de testes numéricos realizados com problemas de pequeno [27] e grande porte [34], quando resolvidos pelo método Lagrangiano Aumentado sozinho (Algencan), o método Lagrangiano Aumentado acelerado por Newton, e o método Lagrangiano Aumentado Acelerado (algoritmo proposto, cujo acelerador é o método ISR1).

O teste de permanência no método ISR1 foi realizado a cada 10 iterações ($D = 10$), uma vez que este número mostrou-se o mais adequado durante a realização dos testes numéricos preliminares. Utilizamos os gráficos de perfis de desempenho como ferramenta para avaliação dos três métodos utilizados.

5.6.1 Perfil de desempenho

O perfil de desempenho [22] é uma técnica que permite comparar o desempenho de n_s métodos na resolução de um conjunto de n_p problemas; para tanto, é preciso escolher uma medida $m_{s,p}$ de comparação (que poder ser o número de iterações, o tempo de execução, etc.) que deve ser computada por cada método s em cada problema p .

Após escolhida $m_{s,p}$, determina-se a razão do perfil de desempenho

$$r_{s,p} = \begin{cases} \frac{m_{s,p}}{\min_s m_{s,p}}, & \text{se o método } s \text{ resolve } p \\ r_M, & \text{caso contrário,} \end{cases}$$

onde r_M é um valor arbitrário bem superior aos demais valores de $r_{s,p}$.

Por fim, o perfil de desempenho é observado por meio do gráfico da função de distribuição acumulativa

$$\rho_s(\tau) = \frac{1}{n_p} \#\{p \mid r_{s,p} \leq \tau\},$$

que é contínua, constante por partes e não-decrescente; o valor de $\rho_s(1)$ é o percentual de problemas resolvidos pelo método s com menor valor da medida $m_{s,p}$ escolhida, o que será denotado como a *eficiência* do método. O menor valor de τ para o qual $\rho_s(\tau) = 1$ (ou fica o mais próximo possível de 1) será considerado como *robustez* do método.

O valor de τ corresponde à proporção da menor medida que todos os métodos obtiveram (levando em conta cada um dos problemas). Por exemplo, para um dado método, $\rho(\tau = 1)$ corresponde ao percentual de problemas que o método em questão conseguiu resolver com o menor valor possível para a medida (percentual de problemas em que o método analisado foi mais eficiente). O valor de $\rho(\tau = 2)$ corresponde ao percentual de problemas que o método conseguiu resolver utilizando o dobro da medida mínima (obtido pelo método mais eficiente para cada problema), $\rho(\tau = 3)$ corresponde ao percentual de problemas que o método conseguiu resolver utilizando o triplo da medida mínima (obtido pelo método mais eficiente para cada problema), e assim por diante. O valor da função ρ avaliada em τ suficientemente alto corresponderá ao percentual de problemas que o método conseguiu resolver utilizando um múltiplo alto da medida mínima, o que por conseguinte, permitirá verificar o percentual de problemas (geral) que o método resolveu (em outras palavras, sua robustez).

Assim, dizemos que método \bar{s} é o mais *eficiente* se possuir o maior valor para $\rho_{\bar{s}}(1)$, dentre os métodos testados. De forma análoga, o método \hat{s} é o mais *robusto* se $\rho_{\hat{s}}(\hat{\tau}) = 1$, para $\hat{\tau}$, o menor valor de τ .

Para efeito comparativo, foram desenhados os perfis de desempenho de 3 métodos: o método proposto, o qual foi denominado Lagrangiano Aumentado Acelerado por ISR1,

o método Lagrangiano Aumentado implementado conforme o Algoritmo 2.2.1, e o Lagrangiano Aumentado Acelerado pelo método de Newton.

Nas seções a seguir, são expostos os resultados numéricos para os problemas de pequeno e grande porte, aonde são apresentados os perfis de desempenho dos métodos testados, quando utilizadas 3 medidas de comparação: número de avaliações da função Lagrangiana, tempo de CPU e/ou número total de iterações internas.

5.6.2 Problemas de pequeno porte

Embora nosso objetivo maior seja construir um algoritmo que seja eficiente para problemas de grande porte, realizamos testes com problemas de minimização de pequeno porte com restrições de igualdade e desigualdade, retirados de [27]. Para esses problemas, as restrições de desigualdade foram transformadas em restrições de igualdade por meio da introdução de variáveis de folga ao quadrado. Os problemas resolvidos foram:

1. Problema 10:

$$\begin{array}{ll} \min & x_1 - x_2 \\ \text{s.a} & -3x_1^2 + 2x_1x_2 - x_2^2 + 1 \geq 0 \end{array}$$

2. Problema 11:

$$\begin{array}{ll} \min & (x_1 - 5)^2 + x_2^2 - 25 \\ \text{s.a} & -x_1^2 + x_2 \geq 0 \end{array}$$

3. Problema 12:

$$\begin{array}{ll} \min & 0.5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2 \\ \text{s.a} & 25 - 4x_1^2 - x_2^2 \geq 0 \end{array}$$

4. Problema 22:

$$\begin{array}{ll} \min & (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{s.a} & -x_1^2 + x_2 \geq 0 \\ & -x_1 - x_2 + 2 \geq 0 \end{array}$$

5. Problema 29:

$$\begin{aligned} \min \quad & -x_1x_2x_3 \\ \text{s.a.} \quad & -x_1^2 - 2x_2^2 - 4x_3^2 + 48 \geq 0 \end{aligned}$$

6. Problema 43:

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

7. Problemas 88 a 92 (variando n de 2 a 6):

$$\begin{aligned} \min \quad & \sum_{i=1}^n x_i^2 \\ \text{s.a.} \quad & h(x) \leq 10^{-4} \end{aligned}$$

onde

$$h(x) = \int_0^1 \left[\sum_{j=1}^{30} \alpha_j(s) \rho_j(x) - k_0(s) \right]^2 ds$$

$$\alpha_j(s) = \mu_j^2 \frac{2 \sin(\mu_j)}{\mu_j + \sin(\mu_j) \cos(\mu_j)} \cos(\mu_j s)$$

$$\rho_j(s) = -\mu_j^{-2} \left[\exp \left(-\mu_j^2 \sum_{i=1}^n x_i^2 \right) - 2 \exp \left(-\mu_j^2 \sum_{i=2}^n x_i^2 \right) + \dots + (-1)^{n-1} 2 \exp \left(-\mu_j^2 x_n^2 \right) + (-1)^n \right]$$

$$k_0(s) = 0.5(1 - s^2)$$

$$\mu_j : \text{ Solução de } \mu \tan(\mu) = 1$$

8. Problema 100:

$$\begin{aligned} \min \quad & (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7 \\ \text{s.a.} \quad & 127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0 \\ & 282 - 7x_1 - 3x_2 - 10x_3^2 - x_4 + x_5 \geq 0 \\ & 196 - 23x_1 - x_2^2 + -6x_6^2 + 8x_7 \geq 0 \\ & -4x_1^2 - x_2^2 + 3x_1x_2 - 2x_3^2 - 5x_6 + 11x_7 \geq 0 \end{aligned}$$

9. Problema 113:

$$\begin{aligned}
 \min \quad & x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 \\
 & + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45 \\
 \text{s.a} \quad & 127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0 \\
 & 105 - 4x_1 - 5x_2 + 3x_7 - 9x_8 \geq 0 \\
 & -10x_1 + 8x_2 + 17x_7 - 2x_8 \geq 0 \\
 & 8x_1 - 2x_2 - 5x_9 - 2x_{10} \geq 0 \\
 & -3(x_1 - 2)^2 - 4(x_2 - 3)^2 - 2x_3^2 + 7x_4 + 120 \geq 0 \\
 & -5x_1^2 - 8x_2 - (x_3 - 6)^2 + 2x_4 + 40 \geq 0 \\
 & -0.5(x_1 - 8)^2 - 2(x_2 - 4)^2 - 3x_5^2 + x_6 + 30 \geq 0 \\
 & -x_1^2 - 2(x_2 - 2)^2 + 2x_1x_2 - 14x_5 + 6x_6 \geq 0 \\
 & 3x_1 - 6x_2 - 12(x_9 - 8)^2 + 7x_{10} \geq 0
 \end{aligned}$$

10. Problema 116:

$$\begin{aligned}
 \min \quad & x_{11} + x_{12} + x_{13} \\
 \text{s.a} \quad & x_3 - x_2 \geq 0 \\
 & x_2 - x_1 \geq 0 \\
 & 1 - 0.002x_7 + 0.002x_8 \geq 0 \\
 & 50 \leq x_{11} + x_{12} + x_{13} \leq 250 \\
 & x_{13} - 1.262626x_{10} + 1.231059x_3x_{10} \geq 0 \\
 & x_5 - 0.3475x_2 - 0.975x_2x_5 + 0.0975x_2^2 \geq 0 \\
 & x_6 - 0.3475x_3 - 0.975x_3x_6 + 0.0975x_3^2 \geq 0 \\
 & x_5x_7 - x_1x_8 - x_4x_7 + x_4x_8 \geq 0 \\
 & 1 - 0.002(x_2x_9 + x_5x_8 - x_1x_8 - x_6x_9) - x_5 - x_6 \geq 0 \\
 & x_2x_9 - x_3x_{10} - x_6x_9 - 500x_2 + 500x_6 + x_2x_{10} \geq 0 \\
 & x_2 - 0.9 - 0.002(x_2x_{10} - x_3x_{10}) \geq 0 \\
 & x_4 - 0.03475x_1 - 0.975x_1x_4 + 0.00975x_1^2 \geq 0 \\
 & x_{11} + 1.262626x_8 + 1.231059x_1x_8 \geq 0 \\
 & x_{12} + 1.262626x_9 + 1.231059x_2x_9 \geq 0
 \end{aligned}$$

Os perfis de desempenho a seguir referem-se ao método Langrangiano Aumentado Algen-can, e sua versão acelerada pelo método ISR1 e pelo método de Newton. Nos resultados

aqui apresentados foram utilizados as medidas tempo de execução e número de avaliações da função Lagrangiana para efeito de comparação entre os métodos, conforme ilustra a Figura 5.2 a seguir.

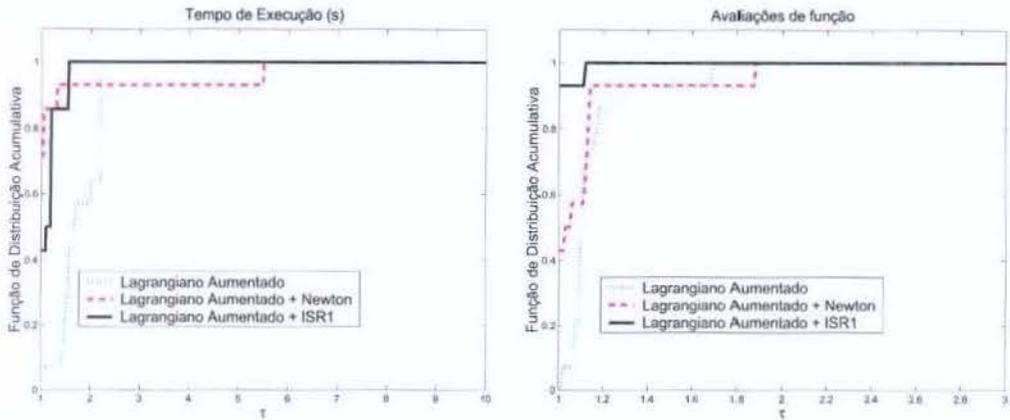


Figura 5.2: *Perfil de Desempenho com Milonga - Problemas de Pequeno Porte.*

Embora os resultados obtidos tenham sido muito bons, o fato de serem problemas de pequeno porte faz com que o comportamento dos métodos testados tenham sido bem similares, dificultando uma análise comparativa de desempenho dos métodos.

5.6.3 Problemas de grande porte

Durante a implementação numérica do Algoritmo 4.2.1, é declarada convergência quando $k < 200$ e $\|F(x_k, \lambda_k)\|_\infty \leq 10^{-5}$ (no caso do método quase-Newton) ou $\max\{\|h(x_k)\|, \|\nabla f(x_k) + J_h(x_k)^T \lambda_k\|\} \leq 10^{-5}$ (no caso do método Lagrangiano Aumentado).

Os problemas-teste escolhidos para os experimentos numéricos são os 18 problemas-teste listados a seguir, propostos por Lukšan e Vlček em [34], Seção 5:

1. Chained Rosenbrock function with trigonometric-exponential constraints;
2. Chained Wood function with Broyden banded constraints;
3. Chained Powell singular function with simplified trigonometric-exponential constraints;

4. Chained Cragg-Levy function with tridiagonal constraints;
5. Generalized Broyden tridiagonal function with five diagonal constraints;
6. Generalized Broyden banded function with exponential constraints;
7. Trigonometric tridiagonal function with simplified five-diagonal constraints;
8. Augmented Lagrangian function with discrete boundary value constraints;
9. Modified Brown function with simplified seven-diagonal constraints;
10. Generalized Brown function with Broyden tridiagonal constraints;
11. Chained HS46 problem;
12. Chained HS47 problem;
13. Chained modified HS48 problem;
14. Chained modified HS49 problem;
15. Chained modified HS50 problem;
16. Chained modified HS51 problem;
17. Chained modified HS52 problem;
18. Chained modified HS53 problem.

Todos os problemas que têm sua designação iniciada por HS foram originalmente propostos por Hock & Schittkowski [27]. O número que aparece logo após HS indica a numeração do problema no trabalho desses autores. Observe que os problemas recebem a indicação de “modified HSxx problem” visto que os autores (Lukšan e Vlček em [34]) realizaram adaptações para que os problemas passem a ter dimensão variável.

Os problemas de 1 a 18 possuem dimensão variável (n); em geral, o número de restrições de igualdade (m) é da mesma magnitude que o número de variáveis. Assim, os sistemas de otimalidade associados a esses problemas tem dimensão $N = n + m$.

Os perfis de desempenho a seguir referem-se aos métodos Langrangeano Aumentado quando implementados utilizando Milonga. Nos resultados aqui apresentados foram utilizados $n = 250$ e $n = 500$, e assim $N \approx 500$ e $N \approx 1000$. A Figura 5.3 refere-se ao perfil de desempenho dos métodos utilizando o tempo de execução (em segundos) como medida de comparação.

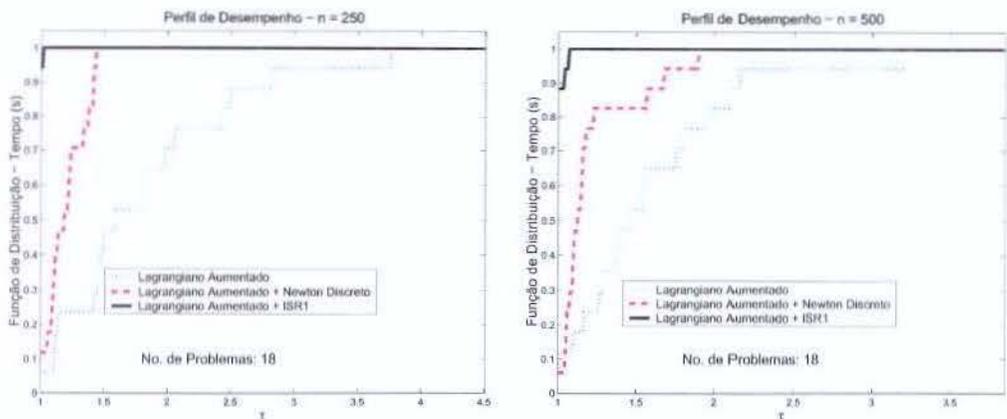


Figura 5.3: Perfil de Desempenho com Milonga - Tempo de execução.

Como pode ser observado, o método Lagrangiano Aumentado Acelerado pelo ISR1 apresentou o melhor desempenho em ambos os casos ($n = 250$ e $n = 500$), uma vez que resolveu em torno de 90% dos problemas com o menor tempo; além disso, a função de distribuição acumulativa do método proposto foi a que atingiu o valor 1 com um menor valor de τ ($\rho(1.1) = 1$ para $n = 250$ e $n = 500$), indicando maior robustez que os demais.

Esse comportamento pode ser explicado pela redução no número de avaliações de função, ocasionado pelo acelerador; isso pode ser observado por meio dos gráficos de perfis de desempenho, quando utilizado o número de avaliações de função como medida de comparação (Figura 5.4).

Como era de se esperar, em relação ao número de iterações internas, o método Lagrangiano Acelerado necessita de menos iterações internas para convergir do que o método Lagrangiano sozinho. Dentre as acelerações testadas, devido ao fato do método de Newton convergir com um número menor de iterações do que os métodos quase-Newton, o método Lagrangiano Aumentado Acelerado com o método de Newton foi aquele que apresentou um menor número de iterações internas (Figura 5.5).

O tempo de execução de todos os métodos é muito alto, e esse valor aumenta conforme a dimensão dos problemas é aumentada. Isso ocorre devido ao fato da rotina que resolve o subproblema em Milonga não ser voltada para a resolução de problemas de grande porte. Dessa forma há uma interferência nos resultados obtidos, uma vez que o acelerador (o método ISR1) está programado de forma a aproveitar a estrutura esparsa do problema,

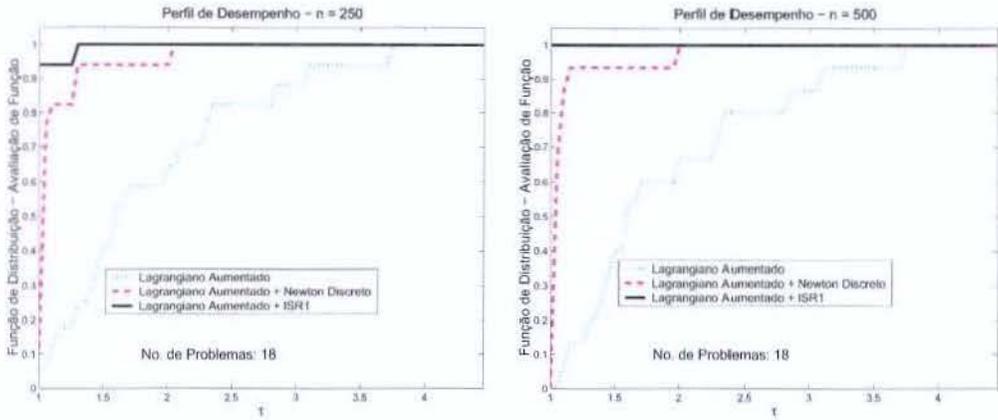


Figura 5.4: Perfil de Desempenho com Milonga - Avaliações de Função.

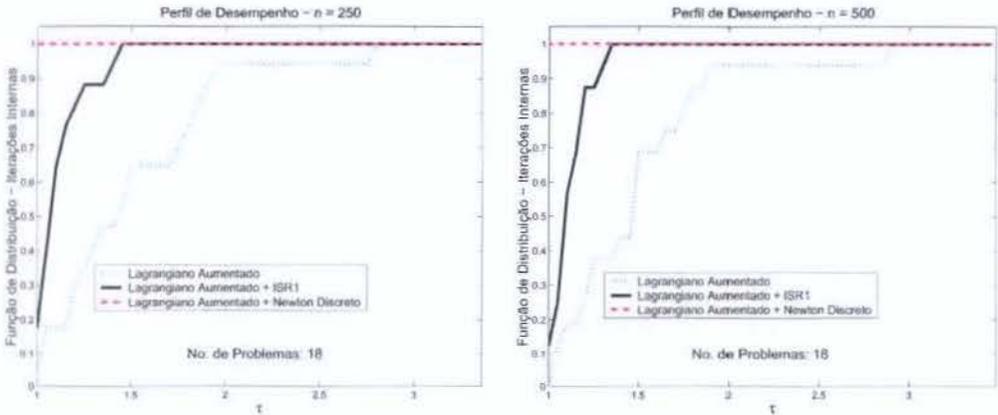


Figura 5.5: Perfil de Desempenho com Milonga - Iterações Internas.

enquanto a rotina utilizada não está.

Sendo assim, fica claro a necessidade do uso de outra rotina computacional (a qual aproveite a estrutura dos problemas em questão) para o cálculo de (5.5). Portanto, de agora em diante, todos os testes com problemas de grande porte serão executados com a rotina Gencan.

Para efeito da análise dos perfis de desempenho, foram utilizados os mesmos métodos:

Lagrangiano Aumentado, Lagrangiano Aumentado Acelerado com ISRI e Lagrangiano Aumentado Acelerado com Newton. Porém, foram utilizadas 3 dimensões: $n = 250$, $n = 500$ e $n = 1000$. A Figura 5.6 a seguir representa os perfis de desempenho obtidos pelos métodos, utilizando como medida o tempo de execução.

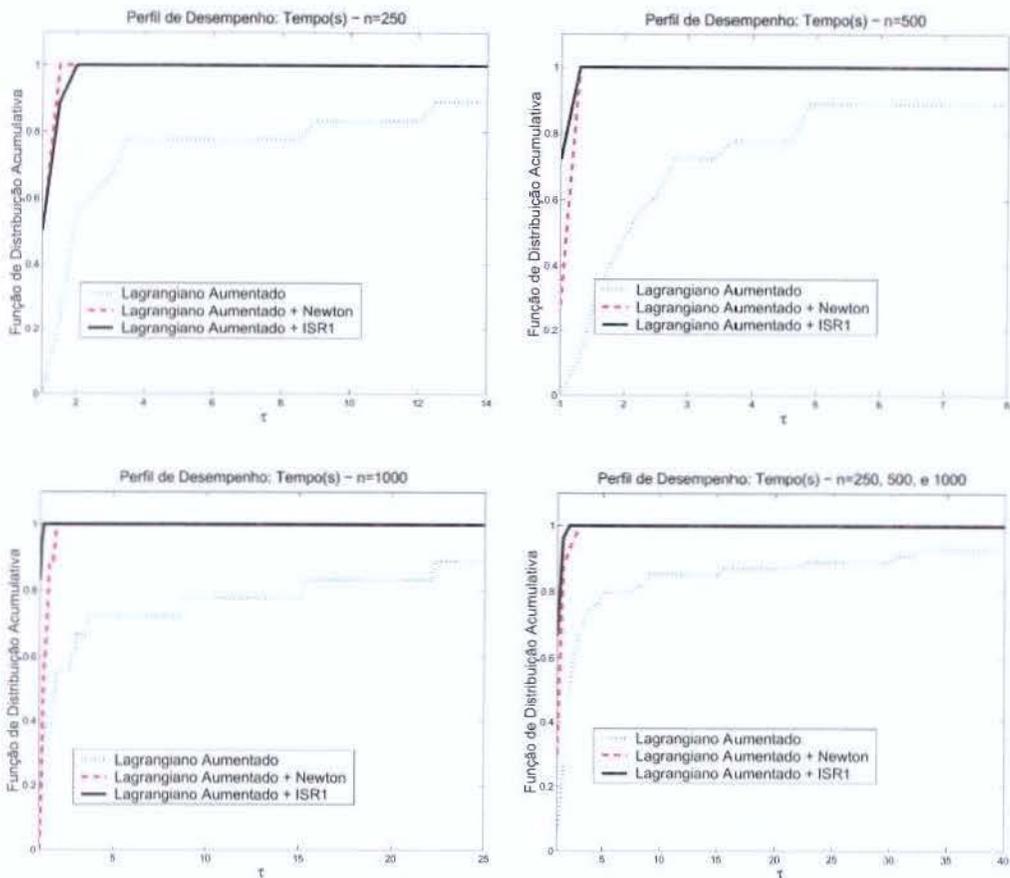


Figura 5.6: Perfil de Desempenho com Gencan - Tempo de execução.

Como pode ser observado, o método Lagrangiano Aumentado Acelerado apresentou um desempenho superior em 2 das dimensões testadas: $n = 250$, $n = 500$ e $n = 1000$; (e no caso em que todos são avaliados conjuntamente, quando traçamos o perfil dos 54 problemas¹ ao mesmo tempo - Figura 5.6, canto direito inferior), tanto em termos de

¹São 18 problemas resolvidos com 3 dimensões diferentes.

robustez quanto em termos de eficiência.

No primeiro gráfico (Figura 5.6, canto superior esquerdo), por ainda tratar-se de problemas com dimensão “pequena”, o método Lagrangiano Acelerado com Newton mostrou-se mais robusto, embora ainda tenha perdido em eficiência para o método proposto.

Em todos os perfis, a função de distribuição acumulativa do método acelerado atingiu o seu valor mais alto com um valor de τ próximo de 2 (observe que há uma variação na escala do eixo τ em cada um dos quatro gráficos).

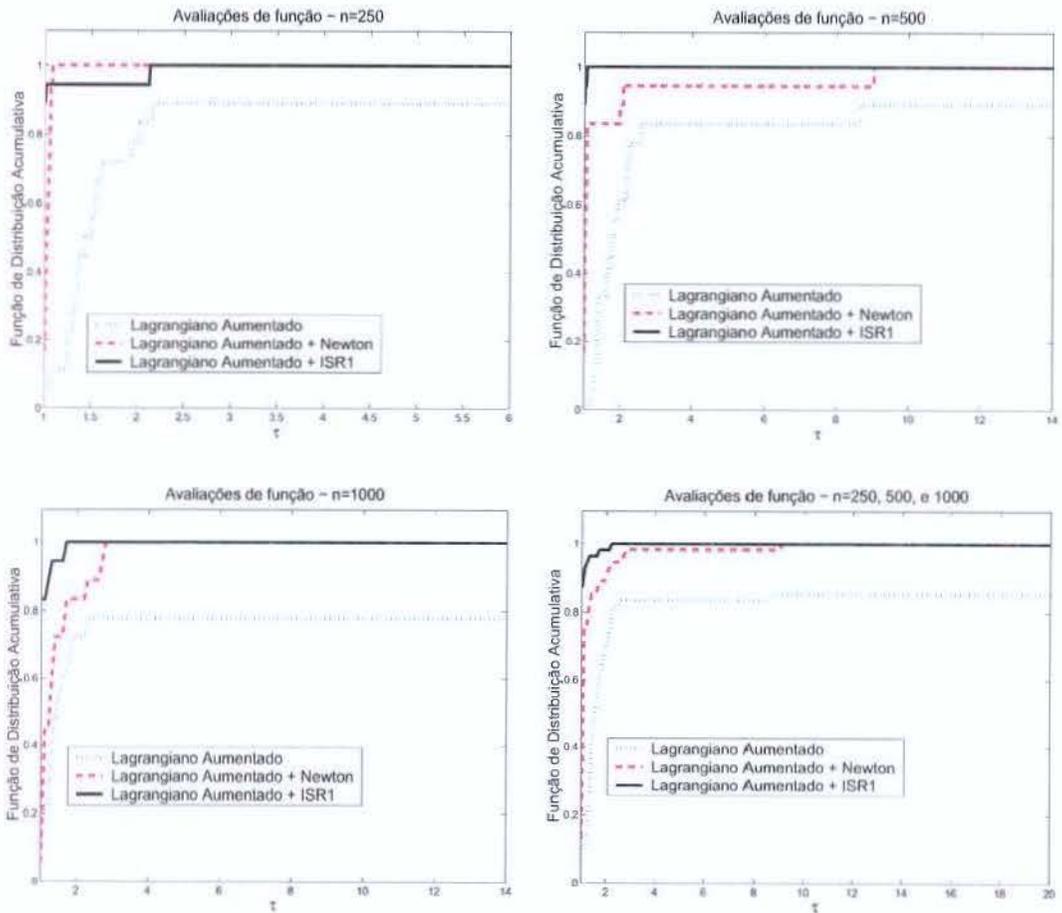


Figura 5.7: Perfil de Desempenho com Gencan - Avaliações de função.

Outro fato que pode ser observado é que, conforme a dimensão aumenta, o método Lagrangiano Aumentado Acelerado com ISR1 torna-se mais eficiente que o método Lagrangiano Aumentado Acelerado com Newton.

Como uma forma de justificar tal comportamento, são apresentados na Figura 5.7 a seguir, os gráficos dos perfis de desempenho dos mesmos métodos quando utilizado o número de avaliações de função como medida de comparação.

A partir desses gráficos, percebe-se que a presença do acelerador reduz, em todos os casos testados, o número total de avaliações de função. Dentre os aceleradores testados, como era de se esperar, o ISR1 mostrou-se mais eficiente que Newton conforme a dimensão dos problemas analisados aumenta (o mesmo comportamento obtido na Figura 5.6).

Fazendo uma comparação entre os gráficos apresentados nas Figuras 5.3, 5.4 e 5.5, referentes à implementação com Milonga, e as Figuras 5.6 e 5.7, referentes à implementação com Algencan, podemos perceber que a superioridade do Algoritmo proposto é maior com Milonga do que com Algencan quando usamos o tempo de execução como medida; esse comportamento é decorrente do fato de Algencan apresentar um desempenho superior para esse conjunto de problemas com estrutura esparsa, fazendo com que o impacto causado pela aceleração quasi-Newton seja menor. Esse fato também pode ser observado quando usamos como medida o número de avaliações de função.

5.6.4 Uma observação sobre a estruturação do algoritmo Lagrangiano Aumentado

Um questionamento que pode ser feito refere-se à ordem em que os *Passos 1* e *2* do Algoritmo 4.2.1 devem ser executados, ou seja, aceleramos o Lagrangiano Aumentado com ISR1 ou aceleramos o ISR1 com o Lagrangiano Aumentado?

Foram realizados testes numéricos com as duas possibilidades a fim de determinar qual deles possui melhor desempenho. O comportamento geral dos resultados obtidos pode ser visto nos gráficos de perfis de desempenho a seguir, onde temos: o *Passo 1* realizado pelo Lagrangiano Aumentado com o *Passo 2* realizado pelo ISR1, o qual será denotado por “Algoritmo Proposto”; e também temos o *Passo 1* realizado pelo ISR1 com o *Passo 2* realizado pelo Lagrangiano Aumentado, o qual será denotado por “Algoritmo Invertido”. É apresentado apenas um perfil para cada medida de comparação (tempo de execução

e número de avaliações de função), sendo levado em conta todos os 54 problemas teste ao mesmo tempo.

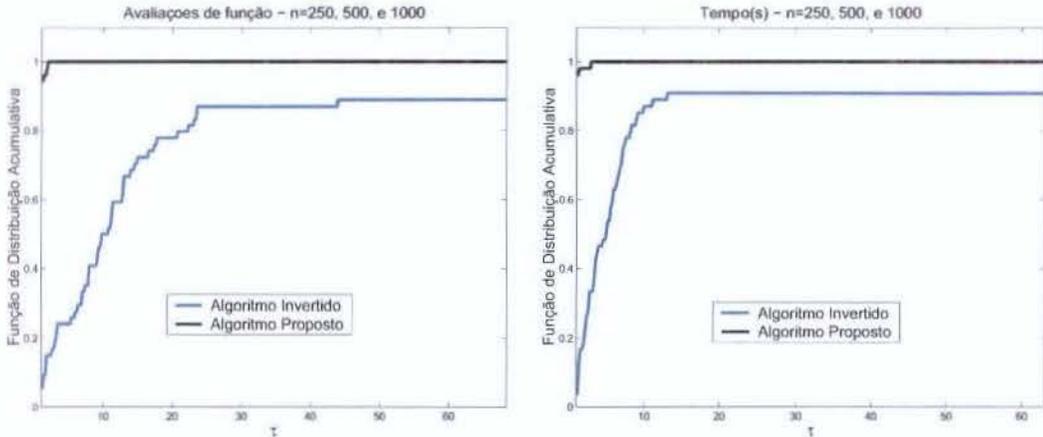


Figura 5.8: *Perfil de Desempenho* - Algoritmo proposto direto e invertido.

Como pode ser observado, o “Algoritmo Proposto”, que foi utilizado em nossas implementações, possui um desempenho significativamente melhor do que o “Algoritmo Invertido”. A justificativa para esse fato é bastante óbvia: o Algoritmo Invertido inicia a execução do método ISR1 com o ponto inicial x_0 , que na maioria dos casos encontra-se bem mais distante do minimizador do que o ponto inicial do Algoritmo Proposto, pois este já sofreu uma iteração externa do Lagrangiano Aumentado. A convergência dos métodos somente é assegurada para pontos iniciais suficientemente próximos da solução.

Assim, o Algoritmo Inverso tende mais facilmente a buscar pontos que possam não ser minimizadores enquanto executa a passagem pelo método ISR1; isso faz com que ele necessite da intervenção do Lagrangiano Aumentado, no sentido de colocar o método ISR1 na direção correta de um minimizador, executando assim um número maior de iterações externas para conseguir convergir, e conseqüentemente, um número maior de avaliações da função Lagrangiana (como comprovam os resultados numéricos).

Em alguns casos, mesmo com a intervenção do método Lagrangiano Aumentado, o Algoritmo Inverso tende a convergir para um maximizador.

5.7 Conclusões

Neste capítulo apresentamos o algoritmo proposto (método Lagrangiano Aumentado Acelerado pelo método ISR1) destinado à resolução de problemas de minimização com apenas restrições de igualdade.

De acordo com os experimentos numéricos, o novo algoritmo é mais robusto que o método Lagrangiano Aumentado, e também que o Lagrangiano Aumentado Acelerado pelo método de Newton. Isso ocorre com quase todas as medidas de comparação utilizadas (com exceção do número de iterações externas), o que comprova a obtenção do objetivo de acelerar o método Lagrangiano Aumentado.

O sistema a ser solucionado pelo método ISR1 possui um número de variáveis igual a $n + m_1$, o que embora seja um número de variáveis maior do que possui o subproblema do método Lagrangiano Aumentado, o método ISR1 consegue realizar suas iterações em um tempo consideravelmente reduzido. Além disso, por se tratarem de problemas com apenas restrições de igualdade, questões como uma possível reordenação das restrições ou reformulação diferente para o sistema de otimalidade não cabem neste contexto.

O desempenho do método proposto está diretamente relacionado à variação dos parâmetros necessários para a sua implementação, em particular, o parâmetro γ em (4.1). Quando γ é pequeno, o novo algoritmo pode ter um comportamento similar ao do método Lagrangiano Aumentado: em geral, a primeira iteração ISR1 será rejeitada. Por outro lado, se γ é muito grande, o algoritmo Lagrangiano Aumentado pode desempenhar um papel secundário durante o processo, fazendo com que o método ISR1 trabalhe sozinho. Essa situação não é preferível em relação à primeira, uma vez que o método ISR1 pode encontrar um maximizador ao invés de um minimizador conforme já foi descrito anteriormente. Em nossas implementações, o valor de $\gamma = 0.5$ foi aquele que apresentou melhor desempenho em testes numéricos realizados com valores que variavam de 0.1 a 0.9.

Capítulo 6

Método Lagrangiano Aumentado Acelerado - Problemas com Restrições de Desigualdade

Neste capítulo estudaremos o problema de minimização abaixo:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & h(x) = 0 \\ & g(x) \leq 0 \\ & l \leq x \leq u, \end{aligned} \tag{6.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$ representa m_1 as restrições de igualdade, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$ representa as m_2 restrições de desigualdade e $l, u \in \mathbb{R}^n$ são os limites inferior e superior da variável x , respectivamente. As funções f, g e h são continuamente diferenciáveis; em geral, são funções não convexas.

Neste capítulo apresentaremos um método Lagrangiano Aumentado para a resolução de problemas de minimização com restrições de desigualdade, igualdade e caixa. Seguindo a mesma estrutura do capítulo anterior, introduzimos o acelerador quase-Newton para esse tipo de problema, e testamos sua eficiência por meio da resolução de problemas teste.

Uma descrição detalhada da implementação do algoritmo proposto encontra-se na Seção

1. Na seção seguinte, são expostos os perfis de desempenho dos três métodos avaliados (Lagrangiano Aumentado sozinho, acelerado com ISR1 e acelerado com o método de Newton) quando destinados à resolução de problemas de pequeno porte, propostos em [34], problema de esferas rígidas [31] e o problema do passo da montanha [28], [37]. Por fim, na Seção 3, apresentamos as conclusões.

6.1 Acelerador

As mesmas idéias utilizadas para acelerar a convergência de métodos Lagrangiano Aumentado na resolução de problemas de minimização com restrições apenas de igualdade também podem ser aplicadas, aproveitando o Algoritmo 2.2.1; ou seja, podemos analisar qual o efeito da introdução de um método quase-Newton na resolução do sistema KKT associado ao problema (6.1).

A introdução do método quase-Newton será realizada da mesma forma que a descrita no esquema da Figura 4.1: fornecidos os parâmetros iniciais, é realizada apenas uma iteração externa do Algoritmo 2.2.1; caso já tenhamos encontrado a solução de (6.1), paramos a execução. Caso contrário, iniciamos a execução do método quase-Newton para resolver o sistema formado pelas condições de otimalidade do problema de otimização.

Entretanto, as condições de otimalidade do problema (6.1) são representadas por meio de um sistema de equações e inequações (2.5), o que impossibilita a sua resolução por meio de um método quase-Newton. Portanto, faz-se necessário uma reformulação de (2.5), a fim de que sejam eliminadas as desigualdades presentes.

Considere o problema (6.1); para deduzir o sistema não-linear correspondente às suas condições de otimalidade, além das variáveis originais representadas pelo vetor x , são adicionadas variáveis de folga representadas pelos vetores z , a e b , correspondendo respectivamente às restrições de desigualdade, limitação inferior e limitação superior da variável x . Assim, podemos definir um novo problema de minimização equivalente ao

problema (6.1), com apenas restrições de igualdade:

$$\begin{aligned}
 \min \quad & f(x) \\
 \text{s.a} \quad & h(x) = 0 \\
 & g(x) + z^2 = 0 \\
 & x - a^2 - l = 0 \\
 & x + b^2 - u = 0
 \end{aligned} \tag{6.2}$$

com $x \in \mathbb{R}^n$, $z \in \mathbb{R}^{m_2}$, $a \in \mathbb{R}^n$ e $b \in \mathbb{R}^n$.

Temos então que x^* é um minimizador local (global) de (2.1) se e somente se (x^*, z^*, a^*, b^*) é um minimizador local (global) do problema (6.2).

Para escrever as condições de otimalidade para o problema (6.2), é necessário primeiro determinar a matriz Jacobiana das restrições:

$$\begin{bmatrix} J_h(x) & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ J_g(x) & 2 \operatorname{diag}(z) & \mathbf{O} & \mathbf{O} \\ I_n & \mathbf{O} & -2 \operatorname{diag}(a) & \mathbf{O} \\ I_n & \mathbf{O} & \mathbf{O} & 2 \operatorname{diag}(b) \end{bmatrix}; \tag{6.3}$$

tal matriz é formada por blocos, onde I_n é a matriz identidade de ordem n , \mathbf{O} é a matriz nula, $\operatorname{diag}(w)$ é a matriz cuja diagonal é formada pelo vetor w e os demais elementos são nulos; J_h corresponde à matriz Jacobiana da função h e J_g é a matriz Jacobiana da função g .

Então, as condições de otimalidade para o problema (6.2) são representadas por um sistema de equações não-lineares que envolve o cálculo do gradiente da função objetivo,

das restrições e de suas Jacobianas:

$$\left\{ \begin{array}{l} \begin{bmatrix} \nabla f(x) \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} J_h(x)^T & J_g(x)^T & I_n & I_n \\ \mathbf{0} & 2 \operatorname{diag}(z) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -2 \operatorname{diag}(a) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 2 \operatorname{diag}(b) \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \\ r \\ t \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \\ h(x) = 0 \\ g(x) + z^2 = 0 \\ x - a^2 - l = 0 \\ x + b^2 - u = 0. \end{array} \right. \quad (6.4)$$

Os vetores $\lambda \in \mathbb{R}^{m_1}$, $\mu \in \mathbb{R}^{m_2}$, $r \in \mathbb{R}^n$ e $t \in \mathbb{R}^n$ contêm os multiplicadores de Lagrange associados, respectivamente, às restrições $h(x)$, $g(x)$, e aos limites inferiores e superiores da variável x .

Logo, a função KKT associada a (6.2) é uma função por blocos, e pode ser representada por

$$F(x, z, a, b, \lambda, \mu, r, t) = \begin{bmatrix} \nabla f(x) + J_h(x)^T \lambda + J_g(x)^T \mu + r + t \\ 2 \operatorname{diag}(z) \mu \\ -2 \operatorname{diag}(a) r \\ 2 \operatorname{diag}(b) t \\ h(x) \\ g(x) + z^2 \\ x - a^2 - l \\ x + b^2 - u \end{bmatrix}, \quad (6.5)$$

onde $F : \mathbb{R}^{5n+m_1+2m_2} \rightarrow \mathbb{R}^{5n+m_1+2m_2}$, cuja matriz Jacobiana $J_F(x, z, a, b, \lambda, \mu, r, t)$ é dada por:

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^{m_1} \nabla^2 h_i(x) \lambda_i + \sum_{i=1}^{m_2} \nabla^2 g_i(x) \mu_i & \mathbf{0} & \mathbf{0} & \mathbf{0} & J_h(x)^T & J_g(x)^T & I_n & I_n \\ \mathbf{0} & 2 \operatorname{diag}(\mu) & \mathbf{0} & \mathbf{0} & \mathbf{0} & 2 \operatorname{diag}(z) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -2 \operatorname{diag}(r) & \mathbf{0} & \mathbf{0} & \mathbf{0} & -2 \operatorname{diag}(a) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 2 \operatorname{diag}(t) & \mathbf{0} & \mathbf{0} & \mathbf{0} & 2 \operatorname{diag}(b) \\ J_h(x) & \mathbf{0} \\ J_g(x) & 2 \operatorname{diag}(z) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ I_n & \mathbf{0} & -2 \operatorname{diag}(a) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ I_n & \mathbf{0} & \mathbf{0} & 2 \operatorname{diag}(b) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

Pode-se observar que os pontos que satisfazem as condições apresentadas em (2.5) também irão satisfazer

$$\|F(x, z, a, b, \lambda, \mu, r, t)\| = 0. \quad (6.6)$$

Logo, o método quase-Newton será aplicado na resolução de (6.6) na tentativa de acelerar o método Lagrangiano Aumentado (Algoritmo 2.2.1).

6.1.1 Detalhes da implementação

Conforme o descrito no Esquema 4.1 e no Algoritmo 4.2.1, o ponto fornecido após a primeira iteração externa do método Lagrangiano Aumentado será o ponto inicial para a execução do método ISR1. Assim, entre o Passo 1 e 2 do Algoritmo 4.2.1, são fornecidos os valores de:

- x_0 : o último valor para x_k obtido pelo método Lagrangiano Aumentado;
- λ_0 : o último valor para λ_k obtido pelo método Lagrangiano Aumentado;
- μ_0 : o último valor para μ_k obtido pelo método Lagrangiano Aumentado.

Além deles, são necessários outros dados iniciais não fornecidos pelo algoritmo Lagrangiano Aumentado:

- O valor da variável de folga z relativa à restrição $g(x) \leq 0$;

- O valor da variável de folga a relativa à restrição $l \leq x$;
- O valor da variável de folga b relativa à restrição $x \leq u$;
- O vetor de multiplicadores relativo à limitação inferior da variável x : r ;
- O vetor de multiplicadores relativo à limitação superior da variável x : t ;

Assim, ao iniciar a subrotina ISR1 (passagem entre o Passo 1 e o Passo 2 do Algoritmo 4.2.1), faz-se necessário o cálculo de:

- $z_0 = \sqrt{\max\{0, -g(x_0)\}}$;
- $a_0 = \sqrt{\max\{0, x_0 - l\}}$;
- $b_0 = \sqrt{\max\{0, u - x_0\}}$;
- $r_0 = -\max\{0, \rho_k(l - x_0)\}$;
- $t_0 = \max\{0, \rho_k(x_0 - u)\}$.

Observe que z_0 , a_0 e b_0 são variáveis de folga; r_0 e t_0 são estimativas para os multiplicadores de Lagrange das restrições de caixa. A cada passagem entre o Passo 2 e o Passo 1, são necessárias as atualizações das estimativas de ρ , μ e λ , seguindo as mesmas fórmulas sugeridas nos Passos 2 e 3 do Algoritmo 2.2.1:

- Calcule $\bar{\lambda}_{k+1} = \lambda_k + \rho_k h(x_k)$, e tome λ_{k+1} com sendo a projeção de $\bar{\lambda}_{k+1}$ em $[\lambda_{min}, \lambda_{max}]^{m1}$
- Calcule $\bar{\mu}_{k+1} = \max\{0, \mu_k + \rho_k g(x)\}$, e tome μ_{k+1} com sendo a projeção de $\bar{\mu}_{k+1}$ em $[\mu_{min}, \mu_{max}]^{m2}$
- Se $\max\{\|h(x_k)\|_\infty, \|\sigma_k\|_\infty\} \leq \tau \max\{\|h(x_{k-1})\|_\infty, \|\sigma_{k-1}\|_\infty\}$, tome $\rho_{k+1} = \rho_k$. Caso contrário, $\rho_{k+1} = \gamma\rho_k$.

6.2 Resultados numéricos

Foram utilizados vários problemas para a realização dos testes numéricos com o algoritmo proposto, envolvendo pequenas e grandes dimensões. Nas seções a seguir, descrevemos os resultados obtidos para a conjunto de problemas de pequeno porte (retirados da coleção proposta por Hock e Schittkowski [27]), assim como outros dois tipos de conjuntos de problemas de grande porte: os problemas de Esfera Rígida [31] e Passo da Montanha [28], [37].

6.2.1 Problemas de pequeno porte

Os problemas de pequeno-porte extraídos de [27] são alguns dos problemas de minimização que possuem apenas restrições de desigualdade (Veja a Seção 5.7.3 para uma melhor descrição de tais problemas), a saber:

Problema	10	11	12	22	29	43	88	89	90	91	92	100	113	116
n	2	2	2	2	3	4	2	3	4	5	6	7	10	13
m_2	1	1	1	2	1	3	1	1	1	1	1	4	8	15

São apresentados a seguir os perfis de desempenho [22] dos métodos Lagrangiano Aumentado, Lagrangiano Aumentado Acelerado por ISR1, e o Lagrangiano Aumentado Acelerado por Newton. Foram utilizados como medidas de desempenho o número de avaliações de função e o tempo de execução (embora os tempos de execução tenham sido bem pequenos, devido ao tamanho dos problemas).

Como pode-se observar na Figura 6.1, o método Lagrangiano Aumentado Acelerado pelo ISR1 é aquele que apresenta o melhor desempenho (utilizando tanto o tempo de execução quanto o número de avaliações de função como medida). Para esse conjunto especial de problemas, devido à pequena dimensão que suas variáveis possuem, o método Lagrangiano Aumentado Acelerado com Newton foi aquele que conseguiu resolver o maior percentual de problemas com o menor tempo de execução, dentre os três métodos testados.

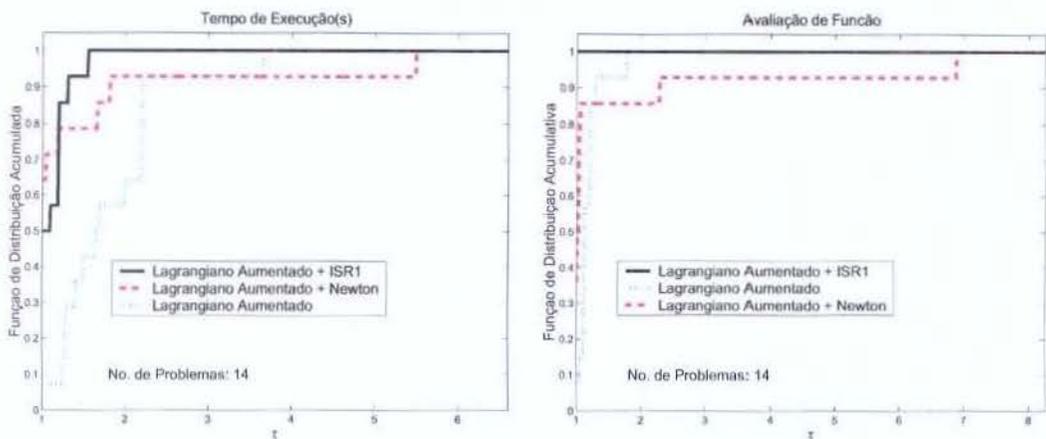


Figura 6.1: *Perfis de Desempenho*: Avaliações de função e tempo de execução.

6.2.2 Problemas de esferas-rígidas

Os problemas de Esfera Rígida [31] pertencem a uma família de problemas de empacotamento de Esferas, os quais são derivados de problemas práticos de Química, Biologia e Física, por exemplo.

Tais problemas consistem em maximizar a distância mínima entre dois pontos em uma esfera (de raio unitário) no \mathbb{R}^n , para um conjunto de p pontos.

$$\begin{aligned} \max \min_{i \neq j} \|y^i - y^j\| \\ \text{s.a } \|y^k\| = 1, \quad k = 1, \dots, p \end{aligned} \quad (6.7)$$

Usando a definição de produto interno padrão, (6.7) se torna

$$\begin{aligned} \min \max_{i \neq j} \langle y^i, y^j \rangle \\ \text{s.a } \|y^k\| = 1, \quad k = 1, \dots, p \end{aligned}$$

e introduzindo uma nova variável z , obtemos

$$\begin{aligned} \min z \\ \text{s.a } z \geq \langle y^i, y^j \rangle, \quad \forall i \neq j \\ \|y^k\| = 1, \quad k = 1, \dots, p \end{aligned}$$

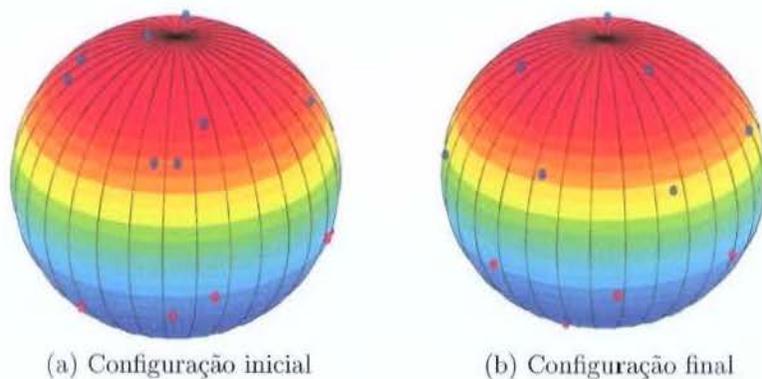


Figura 6.2: Ilustração da configuração inicial e da configuração ótima - Esfera Rígida.

que é um problema de otimização não-linear não convexo e com um grande número de pontos não-ótimos que satisfazem as condições KKT [31].

Nesta formulação, temos $n \times p + 1$ variáveis, p restrições de igualdade e $p(p - 1)/2$ restrições de desigualdade. A dimensão do sistema esparso resolvido pelo método quase-Newton é dada por

$$N = 5np + 5 + p^2.$$

A configuração inicial (ponto inicial para a execução do método Lagrangiano aumentado) é obtida por meio de uma geração aleatória de pontos distintos de norma unitária. Na Figura 6.2 temos a representação de uma configuração inicial e da solução final.

Foram traçados os perfis de desempenho dos três métodos (Lagrangiano Aumentado, Lagrangiano Aumentado Acelerado com ISR1 e acelerado com Newton), utilizando as mesmas duas medidas do conjunto de problemas teste anterior. Na Figura 6.3, temos o perfil de desempenho dos três métodos, quando os vetores x_i possuem dimensão 3 ou 4 e o número de pontos varia de 2 a 25, o que fornece uma dimensão do sistema variando de 40 a 1000. Neste caso, em relação ao número de avaliações de função, o método Lagrangiano Aumentado sem acelerador é aquele que faz uso do menor número de avaliações, apresentando um desempenho superior aos demais; entretanto, o tempo gasto para a execução das iterações quase-Newton é suficientemente pequeno para que o método que apresente melhor desempenho (no sentido de tempo) seja o Lagrangiano Aumentado Acelerado pelo ISR1.

Quando aumentamos o número de pontos (fazendo com que a dimensão do sistema de

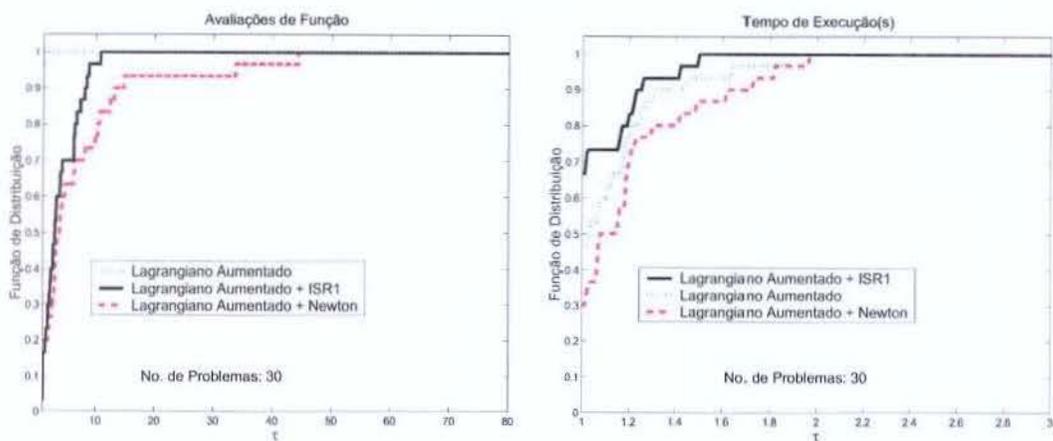


Figura 6.3: *Perfis de Desempenho*: Avaliações de função e tempo de execução - $N \leq 1000$.

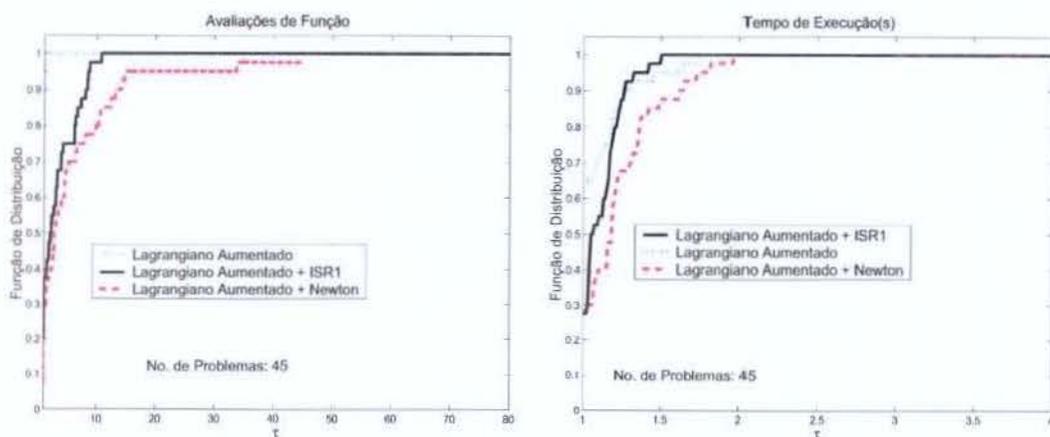


Figura 6.4: *Perfis de Desempenho*: Avaliações de função e tempo de execução - $N \geq 1000$.

otimalidade reformulado também aumente), o desempenho do método Lagrangiano Aumentado Acelerado se altera, conforme mostra a Figura 6.4. Conforme a dimensão do sistema KKT aumenta, menos competitivo torna-se o método acelerado quando comparado com o método Lagrangiano Aumentado sozinho.

6.2.3 Problema do passo da montanha

A idéia do teorema do passo da montanha [28], [37], sem muitos detalhes técnicos, é a seguinte: Sejam I um funcional definido em um espaço de Banach B , e_1 e e_2 dois pontos distintos em B . Considere a família Γ de todos os caminhos em B que conectam e_1 e e_2 e defina

$$c = \min_{\gamma \in \Gamma} \max_{u \in |\gamma|} I(u)$$

que corresponde ao mínimo do valor máximo obtido pelo funcional I ao longo dos caminhos pertencentes a Γ .

A superfície da montanha é representada pelo funcional I , e os vetores $e_1, e_2 \in \mathbb{R}^{ndim}$ são os pontos iniciais e finais do caminho a ser percorrido, respectivamente. Logo, se $c > \max \{I(e_1), I(e_2)\}$, então todos os caminhos precisaram, em algum momento, passar em algum ponto do caminho entre e_1 e e_2 cuja imagem obtida pelo funcional I (que representa o formato da superfície) foi superior a $I(e_1)$ e $I(e_2)$. Portanto, existe um ponto crítico de I no nível c , o qual é chamado de *ponto do passo da montanha*.

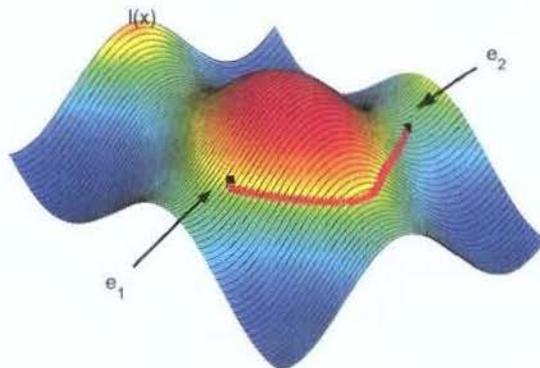


Figura 6.5: Ilustração do caminho ótimo - Passo da Montanha.

A definição de caminho (γ) ligando e_1 a e_2 será aqui adotada como sendo uma seqüência de np pontos, x_1, \dots, x_{np} , de tal sorte que a distância entre cada dois pontos consecutivos

não ultrapasse um valor previamente estipulado ($dmax$):

$$\begin{aligned} \|x_{i+1} - x_i\|_2 &\leq dmax, \quad \text{para } i = 1, \dots, np - 1 \\ \|x_1 - e_1\|_2 &\leq dmax \\ \|e_2 - x_{np}\|_2 &\leq dmax \end{aligned}$$

Portanto, o problema do Passo da Montanha (determinar o nível c) pode ser escrito como:

$$\begin{aligned} \min_{x \in \mathbb{R}^{ndim}} \quad & \max_{1 \leq i \leq np} I(x_i) \\ \text{s.a.} \quad & \|x_{i+1} - x_i\|_2 \leq dmax, \quad \text{para } i = 1, \dots, np - 1 \\ & \|x_1 - e_1\|_2 \leq dmax \\ & \|e_2 - x_{np}\|_2 \leq dmax \end{aligned} \tag{6.8}$$

Introduzindo uma nova variável z , obtemos

$$\begin{aligned} \min \quad & z \\ \text{s.a.} \quad & I(x_i) \leq z, \quad \text{para } i = 1, \dots, np \\ & \|x_{i+1} - x_i\|_2 \leq dmax, \quad \text{para } i = 1, \dots, np - 1 \\ & \|x_1 - e_1\|_2 \leq dmax \\ & \|e_2 - x_{np}\|_2 \leq dmax \end{aligned} \tag{6.9}$$

Para encontrar o caminho ótimo, devemos determinar todas as componentes dos np vetores que determinam o trajeto, além da variável z , o que gera um total de $ndim \times np + 1$ variáveis. O problema reformulado apresenta apenas restrições de desigualdade, sendo np restrições provenientes da reformulação do problema (6.8) no problema (6.9), e as $np + 1$ restrições asseguram que os vetores x_i formaram um caminho:

$$n = ndim \times np + 1 \quad \text{e} \quad m_2 = 2 * np + 1.$$

A dimensão do sistema esparso resolvido pelo método quase-Newton é dada por

$$N = 5 * ndim * np + 4 * np + 7.$$

A superfície utilizada em nossas implementações (funcional I), sugerida originalmente pelos autores do Algencan para esse tipo de problema-teste, foi a função

$$I(x) = \sin(x_1 * x_2) + \sin(x_1 + x_2);$$

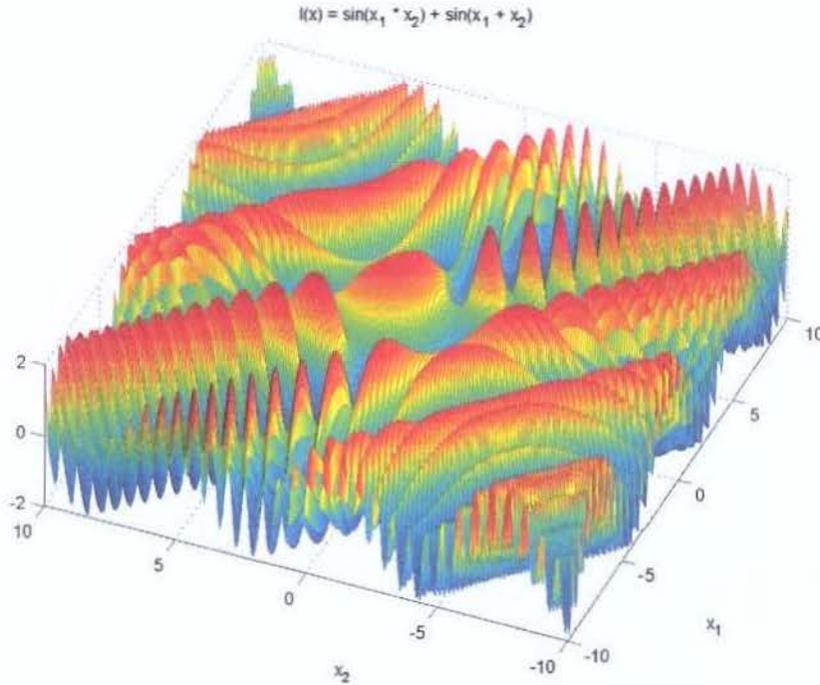


Figura 6.6: Superfície utilizada nos teste numéricos - Passo da Montanha.

cujo gráfico está apresentado na figura a seguir.

Os vetores que indicam o início e o fim do caminho foram adotados como

$$e_1 = (-10, -10, \dots, -10)^T \quad e_2 = (10, 10, \dots, 10)^T;$$

o caminho inicial adotado como chute inicial em nossas implementações consiste na partição em $np + 1$ partes iguais da linha reta que une os pontos e_1 e e_2 . ou seja,

$$x_i^0 = e_1 + i \times \frac{e_2 - e_1}{np + 1} \quad \text{para } i = 1, \dots, np.$$

Assim como para os problemas teste anteriores (esferas rígidas), foram traçados os perfis de desempenho dos três métodos (Lagrangiano Aumentado, Lagrangiano Aumentado Acelerado com ISR1 e acelerado com Newton), utilizando as mesmas duas medidas: número de avaliações da função Lagrangiana e tempo de execução do método. Na

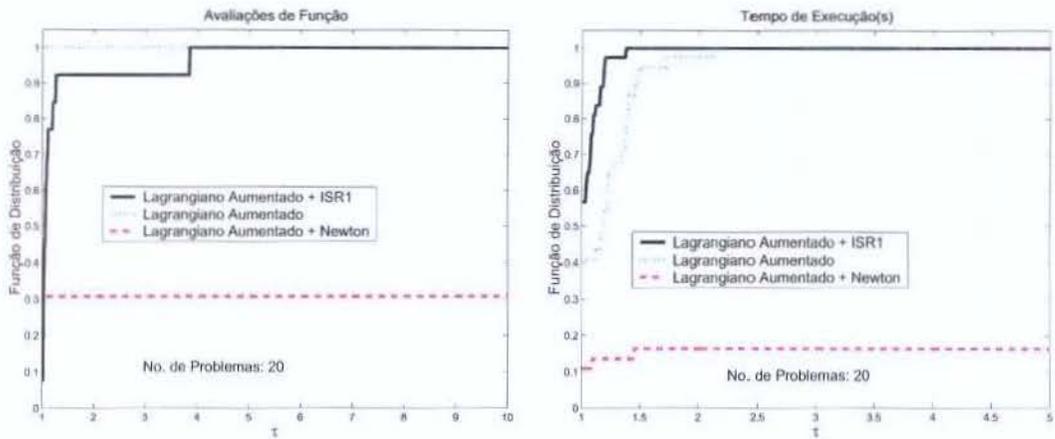


Figura 6.7: *Perfis de Desempenho*: Avaliações de função e tempo de execução - $N \leq 2000$.

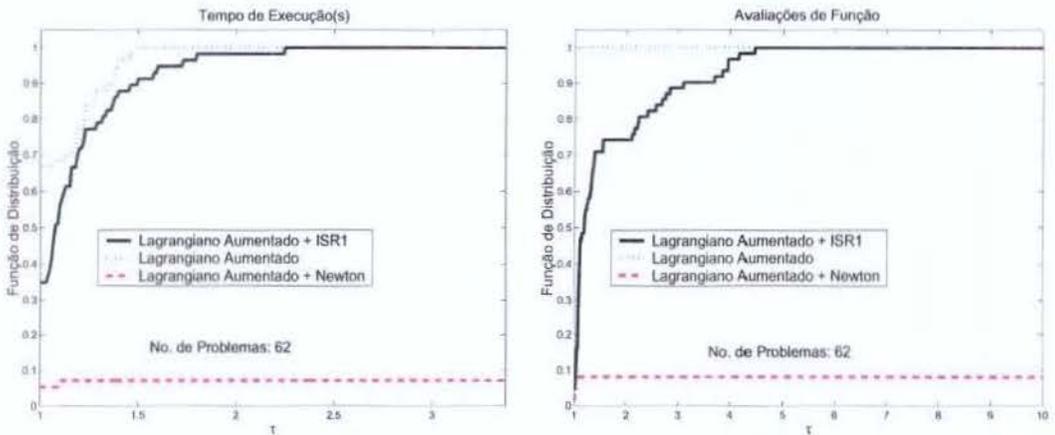


Figura 6.8: *Perfis de Desempenho*: Avaliações de função e tempo de execução - $N \geq 2000$.

Figura 6.7, são apresentados os perfis quando obtidos quando o número de pontos varia entre 10 e 50, e a dimensão dos pontos varia entre 5 e 10; a dimensão do sistema KKT oscila entre 300 e 2700. Como pode ser observado, o método Lagrangiano Aumentado Acelerado apresenta desempenho superior ao método Lagrangiano Aumentado sozinho quando utilizamos como medida o tempo de execução, embora necessite realizar um

número maior de avaliações da função Lagrangiana. Entretanto, quando realizamos testes numéricos com um número maior de pontos (entre 50 e 100, a dimensão do sistema KKT varia entre 2000 e 5000), podemos observar que o método Lagrangiano Aumentado sozinho apresenta melhor comportamento com as duas medidas analisadas (Figura 6.8).

Mesmo utilizando funções diferentes para representar a superfície da montanha, o desempenho do algoritmo proposto não apresenta grandes alterações.

Por exemplo, tomando $I(x) = \sin(x_1) + \cos(x_2)$ como superfície e adotando como pontos inicial e final $e_1 = (5, 5)^T$ e $e_2 = (-0.5, -7.5)^T$, a solução encontrada pode ser visualizada graficamente na Figura 6.9.

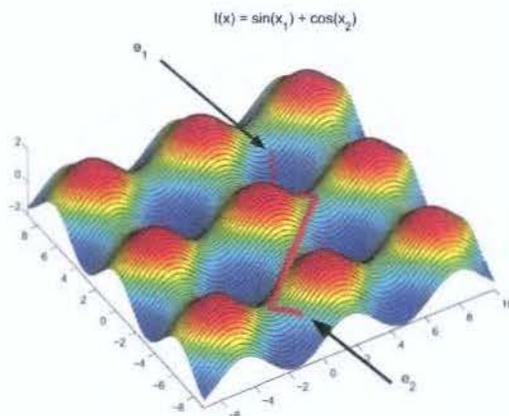


Figura 6.9: Solução do problema do Passo da Montanha - Superfície suave.

Embora esta seja uma superfície mais suave que a adotada no teste anterior, o método proposto não apresentou um desempenho melhor. Veja Figura 6.10, onde estão traçados os perfis de desempenho dos métodos, para uma dimensão de sistema que varia entre $500 \leq N \leq 3000$, com as mesmas medidas de comparação (tempo de execução e número de avaliações da função Lagrangiana), observa-se que o método Lagrangiano Aumentado apresentou o melhor desempenho. Parte do mau comportamento apresentado pelo método ISR1 pode ser justificado pelo desempenho ruim apresentado pelo método de Newton, uma vez que a matriz Jacobiana do sistema KKT torna-se quase-singular.

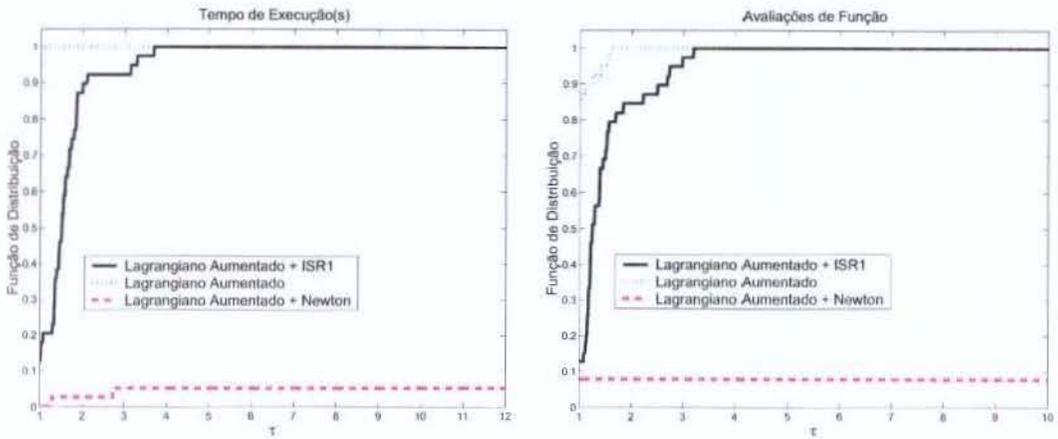


Figura 6.10: *Perfis de Desempenho:* Avaliações de função e tempo de execução - Superfície suave.

6.3 Conclusões

Neste Capítulo apresentamos o método Lagrangiano Aumentado Acelerado destinado à resolução de problemas de minimização com restrições gerais (igualdade, desigualdade e caixa).

Pode-se observar, através dos testes numéricos, que o algoritmo proposto apresenta um bom desempenho quando o sistema KKT associado possui dimensão pequena. Conforme a dimensão do sistema aumenta, o método quase-Newton não consegue convergir rapidamente para um minimizador, fazendo com que os iterandos obtidos não satisfaçam o critério de permanência dentro do método ISR1; dessa forma, são necessárias várias iterações externas do algoritmo proposto (várias intervenções do método Lagrangiano Aumentado) para que o ponto fornecido ao ISR1 permita-o convergir a um minimizador, realizando um elevado número de avaliações de função e tempo de execução.

Uma análise geral do algoritmo proposto e sugestões para melhorar seu desempenho em problemas de minimização com restrições gerais são apresentados no capítulo a seguir.

Capítulo 7

Conclusão

O problema estudado neste trabalho foi

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & h(x) = 0 \\ & g(x) \leq 0 \\ & l \leq x \leq u, \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$ representa m_1 as restrições de igualdade, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$ representa as m_2 restrições de desigualdade e $l, u \in \mathbb{R}^n$ são os limites inferior e superior da variável x , respectivamente.

A nossa proposta para resolver tal tipo de problema foi usar o método quase-Newton como um acelerador para o método Lagrangiano Aumentado. Para tanto, fez-se necessário uma escolha criteriosa sobre como deveria ser realizada a junção desses algoritmos, tanto do ponto de vista teórico quanto do ponto de vista computacional. A idéia básica foi transformar o sistema de otimalidade (sistema KKT) associado ao problema em um sistema de equações; esse sistema reformulado foi resolvido pelo método quase-Newton, tomando como ponto inicial o iterando fornecido pelo método Lagrangiano Aumentado (após sua primeira iteração externa).

No Capítulo 3 apresentamos o método quase-Newton ISR1, utilizado como acelerador para o método Lagrangiano Aumentado. Tal método foi adotado em nossas implementações por se tratar de um método quase-Newton de correção de posto um, que

mantém a simetria das matrizes de aproximação. Embora a teoria de convergência para métodos quase-Newton sem busca linear tenha sido desenvolvida em [3] e [4], a parte original deste capítulo consiste na demonstração da convergência do método ISR1 implementado com a busca linear não-monótona escolhida, que na realidade é válida para qualquer quase-Newton.

O algoritmo proposto foi apresentado no Capítulo 4. Tal capítulo foi escrito com o intuito de expor de forma clara como foi realizada a junção entre o método Lagrangiano Aumentado e o método quase-Newton: quais são os dados iniciais necessários em cada método, e conseqüentemente, como esses dados são transmitidos de um algoritmo para o outro; qual o critério de permanência em cada método, e principalmente, qual deve ser a ordenação entre eles. Testes numéricos mostraram que o método Lagrangiano Aumentado deveria ser executado em primeiro lugar, uma vez que este seria capaz de gerar melhores pontos iniciais para o método ISR1, no sentido de fazer o método quase-Newton convergir para um minimizador.

Fazendo uma análise do algoritmo proposto quanto à resolução de problemas de minimização com apenas restrições de igualdade (Capítulo 5), pode-se concluir que o desempenho apresentado pelo algoritmo proposto (LAA - Lagrangiano Aumentado Acelerado) nos problemas teste utilizados foi muito bom, uma vez que na sua grande maioria conseguiu reduzir tanto o número de avaliações da função Lagrangiana quanto o tempo de execução. Portanto, neste caso, atingimos totalmente o objetivo proposto (aceleração do método Lagrangiano Aumentado) com o algoritmo LAA.

Em relação aos problemas de minimização com restrições de igualdade, desigualdade e caixa, pode-se observar que para problemas com dimensão relativamente pequena (dimensão do sistema KKT inferior a 2000), o algoritmo proposto consegue encontrar a solução com menor tempo de execução do que o método Lagrangiano Aumentado sozinho, embora faça um número maior de avaliações da função Lagrangiana (a diferença entre o número de avaliações da função entre o método Lagrangiano Aumentado sozinho e o acelerado não é tão discrepante). O único problema-teste que não apresentou esse comportamento foi o problema do passo da montanha, quando a superfície foi tomada como sendo $I(x) = \sin(x_1) + \cos(x_2)$. Neste caso, o método Lagrangiano Aumentado não foi acelerado para a maioria das dimensões testadas. Concluímos assim que para esse tamanho de problema, o nosso método é ainda bastante competitivo. No entanto, para problemas com dimensão superior a 2000, o desempenho do método proposto LAA deixou a desejar.

Uma possível justificativa para essa observação consiste no fato de que os problemas usados como problemas teste são considerados computacionalmente difíceis de resolver. Além disso, conforme a dimensão aumenta, a diferença entre o número de variáveis que o Lagrangiano Aumentado trabalha (n) e o número de variáveis que o ISR1 trabalha ($5n + 2m_2 + m_1$) também aumenta, tornando as iterações ISR1 não competitivas com as iterações para a resolução do subproblema dentro do método Lagrangiano Aumentado.

Apresentamos a seguir algumas possibilidades para tentar acelerar de forma mais adequada o método Lagrangiano Aumentado, para que ele se torne também competitivo para problemas de minimização com restrições gerais de maior porte ($N \geq 2000$). Essas possibilidades são algumas de nossas sugestões para dar continuidade a esse trabalho:

- Uma forma de tentar contornar a discrepância no número de variáveis entre os dois métodos seria utilizar outra reformulação para o sistema KKT associado ao problema de minimização: ao invés de utilizarmos variáveis de folga para transformá-lo em um problema de minimização com restrições de igualdade, uma alternativa seria utilizar a função de Fischer-Burmeister [13]

$$\Phi_{FB}(a, b) = \sqrt{a^2 + b^2} - a - b$$

neste caso, exigir a satisfação da restrição de desigualdade, a garantia do sinal do multiplicador e a complementaridade seria assegurada apenas exigindo que a função de Fischer-Burmeister seja igual a zero quando aplicada em cada uma das restrições de desigualdade (e o seu respectivo multiplicador de Lagrange), pois

$$\Phi_{FB}(a, b) = 0 \iff a \geq 0, b \geq 0 \text{ e } ab = 0.$$

Entretando, a derivada da função de Fischer-Burmeister não é contínua, acarretando dificuldades teóricas e possivelmente, computacionais.

- Nos problemas de grande-porte utilizados nos testes-numéricos, o método Lagrangiano Aumentado convergiu com poucas iterações externas, e uma vez que a primeira iteração externa é comum a ambos os métodos (Lagrangiano Aumentado sozinho e o acelerado), o processo de aceleração não é tão eficaz.

Uma forma de contornar esse comportamento seria limitar o número de iterações internas da primeira iteração externa do método Lagrangiano Aumentado, uma vez que a maior parte do esforço computacional (em geral), concentra-se na primeira iteração externa.

Essa limitação deverá ser feita de forma cuidadosa: caso coloquemos um número permitido de iterações internas extremamente pequeno, o ponto fornecido ao ISR1 poderá estar muito longe do minimizador, fazendo com que o ISR1 tenda a convergir a um maximizador (em outras palavras, todas as iterações que o método ISR1 fará inicialmente serão desprezadas). Por outro lado, um número permitido extremamente alto fará com que o método acelerado tenha o mesmo comportamento apresentado nos testes-numéricos deste trabalho, pois a primeira iteração externa do método Lagrangiano Aumentado poderá ser realizada por completo. A proposta aqui é encontrar um número ótimo de iterações a serem permitidas na primeira iteração externa do método Lagrangiano Aumentado.

- Outro possível caminho seria interromper precocemente a resolução dos sistemas lineares através da utilização de métodos tipo Newton Inexato [8].
- Quanto à estrutura da matriz Jacobiana do sistema KKT proveniente da reformulação do problema original, um caminho a seguir seria pensar sobre uma reordenação das posições das restrições e das variáveis, na tentativa de gerar uma matriz Jacobiana que sofra um preenchimento menor que a matriz atual durante a fatoração LU (realizada a cada reinício do método quase-Newton), poupando assim tempo de execução e o espaço destinado ao armazenamento. Embora esparsa, a matriz sofre um grande preenchimento com a fatoração LU.
- Outro caminho possível seria estudar a viabilidade da utilização de um método quase-Newton estruturado, no sentido de gerar uma seqüência de matrizes de aproximação apenas do trecho “mais difícil de calcular” computacionalmente, sendo todo o resto da matriz estimado de forma exata. Este item provavelmente será o primeiro passo a ser desenvolvido, dentre as propostas mencionadas.

Bibliografia

- [1] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, “On augmented Lagrangian with general lower-level constraints”, *submetido*, disponível em www.ime.usp.br/~egbirgin.
- [2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, “Augmented Lagrangian methods under the Constant Positive Linear Dependence Constraint Qualification”, a aparecer em *Mathematical Programming*.
- [3] R. Andreani, J. M. Martínez, and M. L. Schuverdt. “On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification”. *Journal of Optimization Theory and Applications*, vol. 125, pp 743–785, 2005.
- [4] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Computer Science and Applied Mathematics, Academic Press, Califórnia, 1982.
- [5] E. G. Birgin, R. Castillo and J. M. Martínez, “Numerical comparison of augmented Lagrangian algorithms for nonconvex problems”, *Computational Optimization and Applications*, vol. 31, pp. 31–56, 2005.
- [6] E. G. Birgin, J. M. Martínez. “Large-scale active set box-constrained optimization method with spectral projected gradients”, *Computational Optimization and Applications*, vol. 23, pp 101–125, 2002.
- [7] E. G. Birgin, J. M. Martínez e M. Raydan, “Nonmonotone spectral projected gradient methods on convex sets”, *SIAM Journal on Optimization*, vol. 10, pp. 1196–1211, 2000.
- [8] S. C. Eisenstat e H. F. Walker, “Globally convergent inexact-Newton methods”, *SIAM J. Optimization*, vol. 4-2, pp. 393–422, 1994.

- [9] E. G. Birgin, J. M. Martínez e M. Raydan, “Algorithm 813: SPG - Software for convex-constrained optimization”, *ACM Transations on Mathematical Software*, vol. 27, pp. 340-349, 2001.
- [10] E. G. Birgin, N. Krejić, and J. M. Martínez, “Globally convergent inexact quasi-Newton methods for solving nonlinear systems”, *Numerical Algorithms*, vol. 32, pp. 249–260, 2003.
- [11] C. G. Broyden, J. E. Dennis Jr., and J. J. Moré, “On the local and superlinear convergence of quasi-Newton methods”, *J. Inst. Math. Appl.* , vol. 12, pp. 223–245, 1973.
- [12] X. Chen, “Superlinear convergence of smoothing quasi-Newton methods for nonsmooth equations”, *Journal of Computational and Applied Mathematics*, vol. 80, pp. 105–126, 1997.
- [13] B. Chen, X. Chen and C. Kanzow, “A penalized Fischer-Burmeister NCP-function, theoretical investigation and numerical results”, *Mathematical Programming*, vol. 88, pp. 211–216, 2000.
- [14] A. R. Conn, N. I. M. Gold, A. Sartenaer and Ph. L. Toint. “Convergence properties of an Augmented Lagrangian algorithm for optimization with combination of general equality and linear constraints”, *SIAM Journal of Optimization*, vol. 6, pp. 674–703, 1996.
- [15] A. R. Conn, N. I. M. Gold and Ph. L. Toint, “A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simples bounds”, *SIAM Journal on Numerical Analysis*, vol. 28, pp. 545-572, 1991.
- [16] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, “Convergence of quasi-Newton matrices generated by the symmetric rank one update”, *Mathematical Programming*, vol. 50, pp. 177–195, 1991.
- [17] K. Crowe, J. Li Yuan-An Fan, D. Neaderhouser, and P. Smith, “A direct sparse linear equation solver using linked list storage”, IMSL Technical Report 9006, IMSL, Houston, 1990.
- [18] A. R. Curtis, J. D. Powell, and J. K. Reid, “On the estimation of sparse Jacobian matrices”, *J. Inst. Maths Applics*, vol. 13, pp. 117–119, 1974.

- [19] J. E. Dennis Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics in Applied Mathematics 16, SIAM, 1996.
- [20] J. E. Dennis Jr. and J. J. Moré, "Quasi-Newton Methods, Motivation and Theory", *SIAM Review*, vol. 19, pp. 46–89, 1997.
- [21] Zhu Detong, "Nonmonotone backtracking inexact quasi-Newton algorithms for solving smooth nonlinear equations", *Appl. Math. Comput.*, vol. 161-3, pp 875-895, 2003.
- [22] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles", *Mathematical Programming*, vol. 91, pp. 201–213, 2002.
- [23] I. S. Duff, A.M. Erisman, and J.K. Reid, *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford, 1986.
- [24] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, 2nd. Edition, 1987.
- [25] D. Goldfarb, and Ph. L. Toint, "Optimal Estimation of Jacobian and Hessian Matrices That Arise in Finite Difference Calculations", *Mathematics of Computation*, vol. 43-167, pp. 69–88, 1986.
- [26] M. R. Hestenes, "Multiplier and gradient methods", *Journal of Optimization Theory and Applications*, vol. 4, pp. 303-320, 1969.
- [27] W. Hock and K. Schittkowski, "Test examples for nonlinear programming codes", *Lecture Notes in Economic and Mathematical Systems* vol. 187, Springer-Verlag, Berlin-New York, 1981.
- [28] J. Horák, "Constrained mountain pass algorithm for the numerical solution of semi-linear elliptic problems", *Numer. Math.*, vol. 98, pp. 251–276, 2004.
- [29] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer Verlag, New York, 1999.
- [30] Chi-Ming Ip and J. Kyparisis, "Local convergence of quasi-Newton methods for P-differentiable equations", *Mathematical Programming*, vol. 56, pp. 71-89, 1992.
- [31] N. Krejić, J. M. Martínez, M. Mello and E. A. Pilotta, "Validation of an augmented Lagrangian algorithm with a Gauss-Newton Hessian approximation using a set of

- Hard-Spheres problems”, *Computational Optimization and Application*, vol. 16, pp. 247–263, 2000.
- [32] Dong-Hui Li and Masao Fukushima, “A derivative-free line search and global convergence of Broyden-like method for nonlinear equations”, *Optimization Methods and Software*, vol. 13, pp 181-201, 2000.
- [33] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Second Edition, California, 1984.
- [34] L. Lukšan and J. Vlček, Sparse and “Partially Separable Test Problems for Unconstrained and Equality Constrained Optimization”, *Technical Report*, vol. 767, Institute of Computer Science, Academy of Sciences of the Czech Republic, 1999.
- [35] J. M. Martínez e S. A. Santos. *Métodos Computacionais de Otimização*, XX Colóquio Brasileiro de Matemática - IMPA, SBM, 1a. Edição, Rio de Janeiro, 1995.
- [36] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.
- [37] J. J. Moré and T. S. Munson, “Computing mountain passes and transition states”, *Math. Program.*, vol. 100, pp. 151–182, 2004.
- [38] M. J. D. Powell, “A method for nonlinear constraints in minimization problem”, *Optimization*, R.Fletcher (ed.), Academic Press, New York, NY, pp 283-298, 1969.
- [39] R. T. Rockafellar, “Lagrange multipliers and optimality”, *SIAM Review*, vol. 35, pp. 183–238, 1993.
- [40] R. T. Rockafellar, “The multiplier method of Hestenes and Powell applied to convex programming”, *Journal of Optimization, Theory and Applications*, vol. 12, pp. 555-562, 1973.
- [41] L. Qi and Z. Wei. “On the constant positive linear dependence condition and its application to SQP methods”, *SIAM Journal on Optimization*, vol. 10, pp. 963-981, 2000.