

UNIVERSIDADE FEDERAL DO PARANÁ

JULIANA GOMES DA SILVA

ALGORITMOS DE LAGRANGIANO AUMENTADO APLICADOS AO PROBLEMA NÃO
LINEAR CONTÍNUO DE ALOCAÇÃO DE RECURSOS

CURITIBA

2022

JULIANA GOMES DA SILVA

ALGORITMOS DE LAGRANGIANO AUMENTADO APLICADOS AO PROBLEMA NÃO
LINEAR CONTÍNUO DE ALOCAÇÃO DE RECURSOS

Tese apresentada ao Programa de Pós-Graduação em Matemática do Setor de Ciências Exatas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutora em Matemática.

Orientador: Prof. Dr. Luiz Carlos Matioli
Coorientador: Prof. Dr. Paulo Sérgio Marques dos Santos

CURITIBA

2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA CIÊNCIA E TECNOLOGIA

Silva, Juliana Gomes da

Algoritmos de lagrangiano aumentado aplicados ao problema não linear contínuo de alocação de recursos. / Juliana Gomes da Silva. – Curitiba, 2022.

1 recurso on-line : PDF.

Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Matemática.

Orientador: Prof. Dr. Luiz Carlos Matioli.

Coorientador: Prof. Dr. Paulo Sérgio Marques dos Santos.

1. Matemática. 2. Programação não-linear. 3. Alocação de recursos. 4. Método. I. Matioli, Luiz Carlos. II. Santos, Paulo Sérgio Marques dos. III. Universidade Federal do Paraná. Programa de Pós-Graduação em Matemática. IV. Título.

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação MATEMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **JULIANA GOMES DA SILVA** intitulada: **Algoritmos de Lagrangiano aumentado aplicados ao problema não linear contínuo de alocação de recursos**, sob orientação do Prof. Dr. LUIZ CARLOS MATIOLI, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutora está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 17 de Março de 2022.

Assinatura Eletrônica

17/03/2022 16:58:03.0

LUIZ CARLOS MATIOLI

Presidente da Banca Examinadora

Assinatura Eletrônica

17/03/2022 15:53:37.0

SISSY DA SILVA SOUZA

Avaliador Externo (UNIVERSIDADE FEDERAL DO PIAUÍ)

Assinatura Eletrônica

17/03/2022 12:59:50.0

ELIZABETH WEGNER KARAS

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

17/03/2022 12:22:09.0

SANDRA AUGUSTA SANTOS

Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica

18/03/2022 15:58:39.0

LUÍS FELIPE CESAR DA ROCHA BUENO

Avaliador Externo (UNIVERSIDADE FEDERAL DE SÃO PAULO)

ATA DE SESSÃO PÚBLICA DE DEFESA DE DOUTORADO PARA A OBTENÇÃO DO GRAU DE DOUTORA EM MATEMÁTICA

No dia dezessete de março de dois mil e vinte e dois às 08:30 horas, na sala <https://meet.google.com/gym-eoxu-kdn>, remota, foram instaladas as atividades pertinentes ao rito de defesa de tese da doutoranda **JULIANA GOMES DA SILVA**, intitulada: **Algoritmos de Lagrangiano aumentado aplicados ao problema não linear contínuo de alocação de recursos**, sob orientação do Prof. Dr. LUIZ CARLOS MATIOLI. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação MATEMÁTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: LUIZ CARLOS MATIOLI (UNIVERSIDADE FEDERAL DO PARANÁ), SISSY DA SILVA SOUZA (UNIVERSIDADE FEDERAL DO PIAUÍ), ELIZABETH WEGNER KARAS (UNIVERSIDADE FEDERAL DO PARANÁ), SANDRA AUGUSTA SANTOS (UNIVERSIDADE ESTADUAL DE CAMPINAS), LUÍS FELIPE CESAR DA ROCHA BUENO (UNIVERSIDADE FEDERAL DE SÃO PAULO). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de doutora está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, LUIZ CARLOS MATIOLI, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 17 de Março de 2022.

Assinatura Eletrônica

17/03/2022 16:58:03.0

LUIZ CARLOS MATIOLI

Presidente da Banca Examinadora

Assinatura Eletrônica

17/03/2022 15:53:37.0

SISSY DA SILVA SOUZA

Avaliador Externo (UNIVERSIDADE FEDERAL DO PIAUÍ)

Assinatura Eletrônica

17/03/2022 12:59:50.0

ELIZABETH WEGNER KARAS

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

17/03/2022 12:22:09.0

SANDRA AUGUSTA SANTOS

Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica

18/03/2022 15:58:39.0

LUÍS FELIPE CESAR DA ROCHA BUENO

Avaliador Externo (UNIVERSIDADE FEDERAL DE SÃO PAULO)

AGRADECIMENTOS

A Deus, pela saúde concedida.

À minha família, em especial aos meus pais João Alves da Silva e Francisca Gomes da Silva pelo carinho, apoio e incentivo na busca deste sonho.

A Jonathan Cardoso, presente diariamente nessa jornada, me apoiando nos momentos difíceis e comemorando nos momentos de felicidade, obrigada por tudo.

Aos meus orientadores Luiz Carlos Matioli e Paulo Sérgio Marques dos Santos. Obrigada pela orientação, por toda a dedicação, apoio, incentivo e por todas as conversas proveitosas. Vocês são exemplos de excelentes profissionais e de pessoas.

A Elvis Torrealba e Oliver Kolossoski pela parceria na pesquisa e por todas as conversas produtivas, aprendi muito com vocês.

Agradeço especialmente aos professores Elizabeth Wegner Karas, Luís Felipe Cesar da Rocha Bueno, Sandra Augusta Santos e Sissy da Silva Souza, por terem aceito o convite de participarem da banca pela cuidadosa leitura do texto e pelas valiosas contribuições.

Às pessoas maravilhosas que tive o prazer de conhecer em Curitiba. Pela parceria nos estudos e os momentos de descontração, sem dúvidas essa jornada teria sido bem mais difícil e monótona sem vocês.

Ao Programa de Pós-Graduação em Matemática (PPGM) da Universidade Federal do Paraná pela oportunidade de cursar o doutorado. À coordenação do PPGM, e a todos os professores com os quais tive o privilégio aprender.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro.

*“Deep in the human unconscious is a pervasive
need for a logical universe that makes sense.
But the real universe is always one step beyond logic.”*
Frank Herbert, Dune

RESUMO

Propomos uma classe de algoritmos para resolver o problema não linear contínuo de alocação de recursos, o qual é bastante conhecido e pesquisado devido à sua diversificada gama de aplicações. Consideramos problemas em que a função objetivo é convexa e iniciamos analisando o caso de uma única restrição de igualdade e variáveis canalizadas. Este, por sua vez, é conhecido na literatura como problema não linear da mochila. A metodologia proposta consiste no desenvolvimento de algoritmos com uma abordagem híbrida, combinando o método de Lagrangiano aumentado, que penaliza as restrições mais complicadas do problema, com o método de Newton, para resolver os subproblemas gerados a cada passo do algoritmo. Em um segundo momento, estendemos os algoritmos para uma classe mais ampla de problemas incluindo mais de uma restrição, as quais podem ser não lineares. Neste caso, utilizamos uma abordagem semelhante ao método de Gauss-Newton em mínimos quadrados, para possibilitar o uso de informações somente de primeira ordem com respeito às restrições do problema. A maioria dos trabalhos desenvolvidos na literatura da área trata de problemas quadráticos separáveis. A abordagem que propomos é mais geral no sentido que os problemas podem ser não quadráticos e não separáveis. Apresentamos e discutimos propriedades de convergência dos algoritmos propostos e finalizamos a tese com experimentos numéricos, aplicando os algoritmos desenvolvidos através de implementações realizadas no software Matlab.

Palavras-chaves: programação não linear; problema de alocação de recursos; método de Lagrangiano aumentado; método de Newton; problema da mochila.

ABSTRACT

We propose a class of algorithms for solving the continuous nonlinear resource allocation problem, which is well known and researched due to its diverse range of applications. We consider problems where the objective function is convex and we started by analyzing the case of a single equality constraint and simple bounds. This, in turn, is known in the literature as the nonlinear Knapsack problem. The proposed methodology consists of developing algorithms with a hybrid approach combining the augmented Lagrangian method, which penalizes the more complicated constraints of the problem, with Newton's method, to solve the subproblems generated at each step of the algorithm. In a second moment, we extend the algorithms to a broader class of problems including more than one constraint, which can be nonlinear. In this case, we use an approach similar to the Gauss-Newton method for the least squares problem, to enable the use of only first-order information concerning the constraints of the problem. Most works developed in the literature deal with separable quadratic problems. The approach that we propose is more general in the sense that problems can be non-quadratic and non-separable. We present and discuss convergence properties of the proposed algorithms and we finish the thesis with numerical experiments, applying the algorithms developed through implementations carried out with the software Matlab.

Key-words: nonlinear programming; resource allocation problem; augmented Lagrangian method; Newton's method; knapsack problem.

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| FIGURA 1 – PASSOS DO ALGORITMO PARA UM PROBLEMA QUADRÁTICO | 38 |
| FIGURA 2 – PERFIL DE DESEMPENHO PARA TEMPO DE CPU | 70 |
| FIGURA 3 – PERFIL DE DESEMPENHO PARA O NÚMERO DE ITERAÇÕES | 71 |
| FIGURA 4 – CARTEIRA ÓTIMA E CARTEIRAS INDIVIDUAIS | 75 |
| FIGURA 5 – CARTEIRAS ÓTIMAS CONSIDERANDO O PARÂMETRO DE AVERSÃO AO RISCO | 78 |
| FIGURA 6 – PERFIL DE DESEMPENHO PARA TEMPO DE CPU(<i>s</i>) | 80 |

LISTA DE TABELAS

| | |
|--|----|
| TABELA 1 – PROBLEMA DE PROJEÇÃO PONDERADA, $r = 0.9$ | 67 |
| TABELA 2 – PROBLEMA DE AMOSTRAGEM ESTRATIFICADA, $r = 1/n$ | 67 |
| TABELA 3 – PROBLEMA DE OTIMIZAÇÃO DE COMBUSTÍVEL, $r = 0.8$ | 67 |
| TABELA 4 – PROBLEMA DE QUÁRTICA INCLINADA, $r = 0.5$ | 68 |
| TABELA 5 – PROBLEMA DE QUÁRTICA CONVEXA GERAL. $r = 0.8$ | 68 |
| TABELA 6 – ATIVOS NA CARTEIRA E SEUS NOMES | 74 |
| TABELA 7 – CARTEIRA ÓTIMA DETERMINADA PELO ALGORITMO DE LA- GRANGIANO AUMENTADO | 76 |
| TABELA 8 – UMA COMPARAÇÃO COM DIFERENTES VALORES DE β | 80 |
| TABELA 9 – PROBLEMAS QUADRÁTICOS COM RESTRIÇÕES LINEARES | 81 |
| TABELA 10 – CLASSIFICAÇÃO DOS PROBLEMAS | 82 |
| TABELA 11 – TESTES COM O ALGORITMO 9 | 84 |
| TABELA 12 – PROBLEMA 23 COM DIFERENTES DIMENSÕES | 84 |
| TABELA 13 – TESTES LCP | 88 |

LISTA DE ABREVIATURAS E DE SIGLAS

| | |
|-----------|---|
| BS | Método da Bisseção |
| LA | Método de Lagrangiano aumentado |
| LA+Armijo | Método de Lagrangiano aumentado com busca de Armijo |
| LA+MNPA | Método de Lagrangiano aumentado utilizando Newton Projetado com busca de Armijo |
| RF | Método Regula Falsi |
| SC | Método da Secante |

LISTA DE SÍMBOLOS

| | |
|---------------------------------|--|
| \emptyset | O conjunto vazio |
| \mathbb{N} | Conjunto dos números naturais, começando com o zero |
| $K \subset_{\infty} \mathbb{N}$ | Indica que K é um subconjunto infinito de \mathbb{N} |
| \mathbb{R} | Conjunto dos números reais |
| \mathbb{R}_+ | Conjunto dos números reais não negativos |
| \mathbb{R}_{++} | Conjunto dos números reais positivos |
| \mathbb{R}^n | Espaço n -dimensional |
| ∇f | $\equiv \left(\frac{\partial f}{\partial x_i} \right)$ o vetor gradiente da função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ |
| $\nabla^2 f$ | $\equiv \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)$ a matriz Hessiana da função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ |
| ∇h | $\equiv [\nabla h_1 \dots \nabla h_m]_{m \times n}^T$, a matriz Jacobiana da função $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ |
| $\mathcal{N}(A)$ | $\equiv \{y \in \mathbb{R}^n \mid Ay = 0\}$, o núcleo de uma matriz $A \in \mathbb{R}^{m \times n}$ |
| $\ \cdot\ $ | A norma euclidiana, induzida pelo produto interno usual |
| $\ \cdot\ _P$ | A norma induzida pela matriz P |
| $B(x, \delta)$ | A bola aberta de centro x e raio δ |
| $\arg \min_x f(x)$ | O conjunto dos pontos que realizam o mínimo de $f(x)$ |

SUMÁRIO

| | |
|---|-----------|
| INTRODUÇÃO | 14 |
| 1 CONCEITOS PRELIMINARES | 20 |
| 1.1 Problema de minimização | 20 |
| 1.2 Condições de otimalidade | 21 |
| 1.2.1 Problemas irrestritos | 21 |
| 1.2.2 Problemas com restrições | 21 |
| 1.3 Convexidade | 23 |
| 1.4 Operador projeção | 25 |
| 1.5 Método clássico de Lagrangiano aumentado: Restrições de igualdade | 25 |
| 1.5.1 O método de penalização quadrática | 26 |
| 1.5.2 Método clássico de Lagrangiano aumentado | 27 |
| 2 PROBLEMA NÃO LINEAR DA MOCHILA | 31 |
| 2.1 Apresentação do problema e metodologia | 31 |
| 2.1.1 Algoritmo de Lagrangiano aumentado | 32 |
| 2.1.2 Análise de convergência | 34 |
| 2.2 Problema da mochila quadrático | 35 |
| 2.2.1 Convergência no caso quadrático | 38 |
| 2.3 Instâncias do problema da mochila quadrático | 47 |
| 3 PROBLEMA DE ALOCAÇÃO DE RECURSOS | 52 |
| 3.1 Problema quadrático com restrições lineares | 52 |
| 3.1.1 Análise de convergência | 55 |
| 3.2 Problema geral com m restrições | 57 |
| 3.2.1 Metodologia e Algoritmo | 58 |
| 3.2.2 Análise de convergência do Algoritmo 9 | 60 |
| 4 EXPERIMENTOS NUMÉRICOS | 64 |
| 4.1 Problemas não quadráticos | 64 |
| 4.1.1 Descrição dos problemas testes | 65 |
| 4.1.2 Resultados computacionais | 66 |
| 4.1.3 Comparação com algoritmos de localização de raiz | 68 |
| 4.2 Variantes do Algoritmo de Lagrangiano aumentado | 70 |
| 4.3 Otimização de carteira de investimentos | 72 |
| 4.4 Problemas quadráticos separáveis | 78 |
| 4.5 Problemas quadráticos com restrições lineares | 80 |
| 4.6 Problemas gerais | 82 |
| 4.7 Problema de complementaridade linear | 85 |
| 4.7.1 Definição do LCP | 85 |
| 4.7.2 Experimentos numéricos LCP | 87 |
| CONCLUSÃO | 89 |

| | |
|-------------------|----|
| REFERÊNCIAS | 91 |
|-------------------|----|

INTRODUÇÃO

A relevância de estudos direcionados à área de otimização torna-se evidente quando nos deparamos com o crescente desenvolvimento de métodos para resolver problemas em campos diversos como ciência, economia, engenharia e indústria. Em geral, busca-se minimizar ou maximizar algum recurso, com o propósito de reduzir o custo ao menor valor possível, problema notoriamente presente no ramo da computação, ou maximizar os ganhos em determinado processo, como é o caso do setor econômico.

Compreender os métodos estabelecidos na literatura nos fornece conhecimento necessário para avaliar em quais setores podemos aplicá-los, além de apontar direções para o desenvolvimento de pesquisas futuras, investigando melhorias e extensões dos algoritmos. E como novas tecnologias surgem a cada dia, a busca por métodos cada vez mais eficientes é vista como algo crescente nessa área.

Nos problemas de otimização com restrições, o principal objetivo é determinar o menor valor possível para a função objetivo, respeitando um domínio previamente estabelecido. Tal domínio, por vezes, é expresso por funções que definem um conjunto determinado por igualdades e/ou desigualdades.

O problema de alocação de recursos, consiste em determinar uma distribuição ótima de uma quantidade fixa de recursos de modo a minimizar o custo para realizar essa tarefa, ou maximizar o retorno total. O problema tem a seguinte configuração geral

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && g(x) \leq c \\ &&& x \in \Omega, \end{aligned} \tag{1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $c = (c_1, \dots, c_m)^T \in \mathbb{R}^m$ em que c_i é constante para cada $i = 1, \dots, m$, e o conjunto $\Omega \subset \mathbb{R}^n$, em geral, representa uma caixa ou um conjunto de variáveis inteiras.

Esse problema é conhecido por dispor de aplicações em processos de tomada de decisão no mundo real em diversos setores, como em economia (MARKOWITZ, 1952), gestão de produção e estoque (ZIEGLER, 1982) e teoria da computação (GERLA; KLEINROCK, 1977). Uma considerável quantidade de trabalhos tem sido produzida no desenvolvimento de algoritmos que sejam capazes de resolver esse tipo de problema rapidamente, ver por exemplo (PATRIKSSON; STRÖMBERG, 2015).

Na literatura, a classe de problemas (1) em que considera-se a função objetivo f não linear, e a função que descreve o conjunto viável $g : \mathbb{R}^n \rightarrow \mathbb{R}$, ou seja, problemas

com apenas uma restrição de desigualdade (ou igualdade), e limitantes nas variáveis, é comumente chamada de problema não linear da mochila. No artigo (BRETTHAUER; SHETTY, 2002), por exemplo, são relacionadas diversas versões do problema não linear da mochila, e apresentadas as principais abordagens utilizadas para cada tipo de problema. As classes de problemas que os autores abordam englobam problemas contínuos e inteiros, problemas convexos e não convexos, problemas separáveis e não separáveis, além de problemas com restrições adicionais especialmente estruturadas.

É importante salientar que qualquer método de otimização não linear com restrições, desenvolvido para resolver problemas com múltiplas restrições, pode ser aplicado ao problema não linear da mochila. No entanto, dependendo das características de um problema particular, métodos mais eficientes podem ser desenvolvidos tirando proveito da estrutura especial correspondente. A exemplo disso, Bretthauer e Shetty (1997) desenvolveram um algoritmo especializado para uma classe de problemas quadráticos contínuos que chega a ser até 4000 vezes mais rápido do que o *software* de gradiente reduzido de propósito geral LSGRG, proposto por Smith e Lasdon (1992), como destacado em (BRETTHAUER; SHETTY, 2002).

Patriksson (2008) fornece uma revisão da literatura examinando a história e as aplicações do problema de minimizar uma função convexa, separável e diferenciável, sobre um conjunto convexo definido por variáveis limitadas e uma restrição descrita por uma função convexa separável. Além disso, são discutidas abordagens algorítmicas para a solução do problema, constatando que as técnicas mais comuns são baseadas em encontrar o valor ótimo do multiplicador de Lagrange para a restrição explícita, na maioria das vezes por meio do uso de um tipo de procedimento de busca linear. O autor considerou para a análise as referências mais relevantes, especialmente quanto à sua originalidade e resultados numéricos, resumindo com observações sobre possíveis extensões e pesquisas futuras. Com o intuito de ampliar o estudo realizado em 2008, Patriksson e Strömberg (2015) complementam com mais de 20 livros e artigos, fornecendo melhorias para alguns dos algoritmos apresentados. Além disso, avaliam, do ponto de vista numérico, diversos algoritmos de relaxamento (primal) e *breakpoint* (dual), incorporando uma variedade de estratégias de *pegging*, bem como um método quase-Newton.

Como Bretthauer e Shetty (2002) observaram em seu levantamento de artigos, as abordagens básicas para resolver o problema não linear da mochila convexo e separável são os métodos de busca de multiplicadores e métodos de relaxamento ou métodos de *pegging*, que aproveitam a estrutura das condições de Karush-Kuhn-Tucker (KKT) do problema. Enquanto os métodos de busca de multiplicadores de Lagrange mantêm todas as condições de KKT durante suas iterações, exceto a restrição de desigualdade e sua condição de folga complementar correspondente, os métodos de

pegging satisfazem a condições de KKT em todas as iterações, exceto os limites das variáveis, para mais detalhes ver (BRETTHAUER; SHETTY, 2002).

Uma parcela significativa de trabalhos encontrados na literatura se debruça no caso convexo quadrático, sendo uma parte direcionada ao caso quadrático separável, também chamado de quadrático diagonal, ver por exemplo Bayón et al. (2013), Brucker (1984), Cominetti et al. (2014), Frangioni e Gorgone (2013), Kiwiel (2008a), Kiwiel (2008b) e Robinson et al. (1992). Com relação ao caso quadrático não separável, há uma produção considerável, a título de exemplo, Caprara et al. (1999), Dussault et al. (1986), Gallo et al. (1980), Pang (1980), Pardalos e Koor (1990). Também encontramos trabalhos voltados para contextos mais gerais, como o caso convexo separável abordado nos trabalhos de Bretthauer e Shetty (1995), Hoto et al. (2020), Nielsen e Zenios (1992), Padakandla e Sundaresan (2009), Vidal et al. (2016) e Ziegler (1982). Com menos trabalhos relacionados, existem os casos não convexo separável e convexo não separável, em que citamos, para o primeiro Moré e Vavasis (1990), e para o segundo Klastorin (1990). No caso não separável, a estrutura das condições de KKT do problema, crucial no desenvolvimento de métodos para problemas separáveis, se torna especialmente complicada.

Neste trabalho, consideramos o seguinte problema não linear de alocação de recursos

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && h(x) = 0 \\ & && \ell \leq x \leq u, \end{aligned} \tag{2}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa, e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, ambas, pelo menos duas vezes continuamente diferenciáveis. Assumimos $\ell, u \in \mathbb{R}^n$, ou seja, $-\infty < \ell_i$ e $u_i < +\infty$ para todo $i = 1, \dots, n$. Trabalhamos o problema não linear da mochila, caso particular de (2), em que a função que define a restrição de igualdade é linear e a função objetivo é convexa. Consideramos, também, a análise do caso onde não são exigidas hipóteses de convexidade nas funções que definem as restrições de igualdade, para o caso que estas sejam não lineares.

A abordagem proposta é baseada nos métodos de Lagrangiano aumentado para resolver o problema (2), em que penalizamos apenas as restrições de igualdade, h . A cada iteração do algoritmo proposto, resolvemos uma sequência de subproblemas irrestritos projetando, na caixa, a solução encontrada. Para uma classe de problemas mais geral, consideramos uma aproximação da Hessiana da função Lagrangiano aumentado para gerar uma sequência de subproblemas quadráticos com variáveis canalizadas.

Devido à importância do caso quadrático, com diversos trabalhos voltados para este contexto, como apresentado em parágrafos precedentes, consideramos uma aná-

lise do caso onde a função objetivo é quadrática com restrições lineares. Conseguimos estruturar algoritmos embasados na filosofia dos métodos de Lagrangiano aumentado, simplificando o passo mais complicado, que é resolver a sequência de subproblemas gerados.

Neste sentido, os algoritmos propostos nesta tese podem ser aplicados a problemas quadráticos, não quadráticos, separáveis e não separáveis, mas sempre considerando problemas cuja função objetivo é convexa. Os algoritmos são de simples implementação e se mostraram eficientes para os problemas testados.

Resumidamente, as principais contribuições desta tese são:

- Desenvolvimento de uma classe de algoritmos baseados em Lagrangiano aumentado, para resolver diferentes formulações do problema não linear de alocação de recursos;
- Apresentação de uma investigação mais aprofundada do problema de alocação de recursos, com resultados numéricos robustos;
- Análise de convergência dos métodos propostos;
- Análise com mais detalhes para o caso quadrático, com uma proposta de método específico para este caso;
- Experimentos numéricos diversificados. Estes com o objetivo de mostrar que os métodos geram bons resultados e são competitivos na maioria dos casos comparados.

A tese está organizada da seguinte forma:

No Capítulo 1, apresentamos o material básico, mas necessário, para o desenvolvimento do trabalho e para torná-lo autocontido. Destacamos a formulação clássica do método de Lagrangiano aumentado, que envolve apenas restrições de igualdade, onde apontamos a principal vantagem se comparado ao método de penalização quadrática, que embora seja mais simples, apresenta alguns problemas de mau condicionamento. Proporcionamos ao leitor uma base teórica suficiente para o entendimento dos resultados presentes nos demais capítulos.

No Capítulo 2, apresentamos a formulação do problema não linear da mochila e a estrutura do algoritmo proposto para resolver o problema. Analisamos a convergência das sequências geradas pelo algoritmo, com as ideias fundamentadas na teoria de convergência estabelecida por (BIRGIN; MARTÍNEZ, 2014). Realizamos uma análise do caso particular em que consideramos a função objetivo quadrática, chamado de problema da mochila quadrático e conseguimos estabelecer um método mais eficiente,

neste caso a análise de convergência se desvincula das ideias de (BIRGIN; MARTÍNEZ, 2014), não sendo necessário incluir hipóteses extras. Por fim, elencamos algumas instâncias do problema da mochila quadrático, envolvendo inclusive restrições de desigualdade.

No Capítulo 3, estudamos uma generalização do problema apresentado no Capítulo 2, propomos um método também baseado em Lagrangiano aumentado para resolver o problema não linear de alocação de recursos (2). Inicialmente, consideramos a extensão natural do que foi estudado no Capítulo 2, que é o problema quadrático com mais de uma restrição linear e variáveis canalizadas. Finalmente, consideramos a análise do caso geral, em que a função objetivo foi considerada apenas convexa e as restrições de igualdade podem ser não lineares e não convexas, trabalhamos com uma aproximação da Hessiana da função Lagrangiano aumentado, para estabelecermos o algoritmo proposto. A análise de convergência foi baseada nas ideias de (BIRGIN; MARTÍNEZ, 2014).

No Capítulo 4, realizamos experimentos numéricos através da implementação, em Matlab, dos algoritmos propostos, com o objetivo de verificar a robustez e a eficiência dos mesmos. Consideramos problemas não quadráticos e separáveis, descrevemos os problemas considerados para os testes, que foram baseados em (WRIGHT; LIM, 2020). Estudamos variações do método utilizando busca de Armijo e Newton projetado e comparamos com os resultados do método proposto, os resultados estão resumidos nos gráficos de perfil de desempenho. Aplicamos o algoritmo para resolver o problema de otimização de carteiras, consideramos dados do índice Ibovespa para os testes. Este problema é um caso de alocação de recursos, e tem as característica do problema não linear da mochila, que se encaixa na classe não diagonal e não separável. Apresentamos problemas quadráticos separáveis, onde analisamos o comportamento do método ao considerarmos uma aproximação para a inversa da Hessiana da função Lagrangiano aumentado. Em seguida, realizamos testes para o problema quadrático com restrições lineares, geramos problemas com diferentes dimensões, e comparamos os resultados do método proposto com a rotina *quadprog*, do Matlab. Consideramos problemas mais complexos em que a função objetivo é convexa, mas as restrições de igualdade podem não ser convexas, presentes em (HOCK; SCHITTKOWSKI, 1981). Por fim, realizamos testes preliminares envolvendo o Problema de Complementaridade Linear, nesses testes iniciais percebemos que o algoritmo proposto resolve os problemas com uma precisão boa.

Finalmente, apresentamos as conclusões da tese.

Os resultados obtidos com o estudo do problema não linear da mochila originaram o artigo *Augmented Lagrangian algorithms for solving the continuous nonlinear resource allocation problem*, em um trabalho conjunto com Elvis Manuel Rodriguez

Torrealba, Luiz Carlos Matioli, Oliver Kolossoski e Paulo Sérgio Marques dos Santos, ver (TORREALBA et al., 2022).

1 CONCEITOS PRELIMINARES

Neste capítulo, apresentamos uma breve revisão de definições e resultados referentes a minimização de funções, condições de otimalidade, convexidade e projeção ortogonal. Além disso, realizamos um estudo a respeito do método clássico de Lagrangiano aumentado. Fornecemos, dessa forma, uma base de definições e resultados clássicos da literatura, com o intuito de contextualizar os problemas e métodos presentes neste trabalho, facilitando a leitura e o entendimento. Para construção deste capítulo utilizamos como principais referências (BAZARAA et al., 2006), (BERTSEKAS, 1982a), (BIRGIN; MARTÍNEZ, 2014), (IZMAILOV; SOLODOV, 2014, 2012), (NOCEDAL; WRIGHT, 2006) e (RIBEIRO; KARAS, 2013).

1.1 PROBLEMA DE MINIMIZAÇÃO

Considere um conjunto $D \subset \mathbb{R}^n$, $f : D \rightarrow \mathbb{R}$ uma função arbitrária e o problema de otimização geral,

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && x \in D. \end{aligned} \tag{1.1}$$

O conjunto D é chamado de *conjunto viável* do problema, os pontos de D são chamados de *pontos viáveis*, e f é chamada de *função objetivo*.

Definição 1.1.1. Dizemos que $x^* \in D$ é

(i) *minimizador global* de (1.1), se

$$f(x^*) \leq f(x), \text{ para todo } x \in D;$$

(ii) *minimizador local* de (1.1), se existe $\delta > 0$ tal que

$$f(x^*) \leq f(x), \text{ para todo } x \in B(x^*, \delta) \cap D.$$

Quando as desigualdades na Definição 1.1.1 forem estritas para $x \neq x^*$, dizemos que x^* é *minimizador estrito* (global ou local, respectivamente). O valor de f em um minimizador global (ou local), será chamado de *mínimo global* (ou local).

O próximo teorema garante a existência de solução global sob certas condições, para a demonstração consulte o Teorema 1.2.1 em (IZMAILOV; SOLODOV, 2014).

Teorema 1.1.1. (Teorema de Weierstrass) *Considere $D \subset \mathbb{R}^n$ um conjunto compacto não vazio e $f : D \rightarrow \mathbb{R}^n$ uma função contínua. Então o problema de minimizar f em D possui uma solução global em D .*

1.2 CONDIÇÕES DE OTIMALIDADE

1.2.1 Problemas irrestritos

Se no problema (1.1) $D = \mathbb{R}^n$ ele é chamado de *irrestrito*. Neste caso, podemos caracterizar um minimizador de um problema irrestrito por meio de condições necessárias e suficientes, como veremos nos teoremas seguintes. Sendo os resultados a seguir clássicos de otimização contínua, omitiremos as demonstrações, que podem ser encontradas em (RIBEIRO; KARAS, 2013). Teoremas 2.9, 2.12 e 2.14.

Teorema 1.2.1. (Condição necessária de primeira ordem) *Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* é um minimizador local de f , então*

$$\nabla f(x^*) = 0.$$

Dizemos que o ponto x^* é *estacionário* se $\nabla f(x^*) = 0$. Assim, concluímos do Teorema 1.2.1 que todo minimizador local é um ponto estacionário.

Teorema 1.2.2. (Condição necessária de segunda ordem) *Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* é um minimizador local de f , então $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é semidefinida positiva, isto é,*

$$d^T \nabla^2 f(x^*) d \geq 0,$$

para todo $d \in \mathbb{R}^n$.

Teorema 1.2.3. (Condição suficiente de segunda ordem) *Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* é um ponto estacionário da função f e $\nabla^2 f(x^*)$ é definida positiva, então x^* é minimizador local estrito de f .*

1.2.2 Problemas com restrições

O problema a ser tratado nessa subseção tem a seguinte formulação

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && h(x) = 0 \\ &&& g(x) \leq 0, \end{aligned} \tag{1.2}$$

em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ são funções de classe \mathcal{C}^2 . Note que neste caso o conjunto viável é dado por $D = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0\}$.

Definição 1.2.1. *A função Lagrangiano associada ao problema (1.2), é dada por*

$$\mathcal{L}(x, \lambda, \mu) := f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x) \tag{1.3}$$

em que $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}^p$ são os multiplicadores de Lagrange associados às restrições de igualdade e desigualdade, respectivamente.

Definição 1.2.2. *Considere $\bar{x} \in D$, dizemos que a condição de desigualdade correspondente ao índice $j \in \{1, \dots, p\}$ é ativa no ponto \bar{x} quando $g_j(\bar{x}) = 0$. Denotaremos por $I(\bar{x})$ o conjunto de índices das restrições de desigualdade ativas no ponto $\bar{x} \in D$, ou seja,*

$$I(\bar{x}) = \{j \in \{1, \dots, p\} \mid g_j(\bar{x}) = 0\}.$$

Note que as restrições de igualdade não são incluídas nesta definição, pois todas são sempre ativas em todo ponto viável.

Definição 1.2.3. *Dizemos que a condição de qualificação de independência linear (LICQ) é satisfeita em \bar{x} se o conjunto formado pelos gradientes das restrições ativas $\{\nabla h_i(\bar{x}) \mid i = 1, \dots, m\} \cup \{\nabla g_j(\bar{x}) \mid j \in I(\bar{x})\}$ é linearmente independente.*

Devido a sua simplicidade, tal condição é comumente utilizada na literatura. Vale lembrar que existem outras condições de qualificação, como por exemplo, a condição de qualificação de Slater ou a condição de qualificação de Mangasarian-Fromovitz, para mais detalhes a respeito consultar (NOCEDAL; WRIGHT, 2006) e (RIBEIRO; KARAS, 2013).

Estabelecemos adiante as condições necessárias de primeira ordem que caracterizam minimizadores locais para o problema com restrições.

O teorema seguinte pode ser enunciado usando qualquer umas das condições de qualificação, por simplicidade, vamos nos restringir a LICQ.

Teorema 1.2.4. (Teorema 12.1, (NOCEDAL; WRIGHT, 2006)) *Suponha que x^* é uma solução local do problema (1.2) e que LICQ é satisfeita em x^* . Então existem vetores $\lambda^* \in \mathbb{R}^m$ e $\mu^* \in \mathbb{R}^p$, tais que*

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) &= 0, \\ h_i(x^*) &= 0, \quad i = 1, \dots, m, \\ g_j(x^*) &\leq 0, \quad j = 1, \dots, p, \\ \mu_j^* &\geq 0, \quad j = 1, \dots, p, \\ \mu_j^* g_j(x^*) &= 0, \quad j = 1, \dots, p. \end{aligned} \tag{1.4}$$

As condições (1.4) são conhecidas como condições de Karush-Kuhn-Tucker, de forma abreviada, condições de KKT.

Consideremos o caso particular em que o problema de otimização possui apenas restrições de igualdade, ou seja,

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && h(x) = 0, \end{aligned} \tag{1.5}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções de classe \mathcal{C}^2 .

As condições de KKT se restringem às duas primeiras linhas de (1.4), adaptando a função Lagrangiano apenas para as restrições de igualdade.

Estabelecemos nos resultados seguintes as condições de otimalidade de segunda ordem para o problema com restrições de igualdade. Neste trabalho não abordamos as condições de otimalidade de segunda ordem para problemas gerais (com restrições de igualdade e desigualdade), pois os problemas com restrições de desigualdade foram tratados tais como os de igualdade por meio da adição de variáveis de folga.

Observação 1.2.1. *Considere a função $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Denotamos a matriz Jacobiana de h no ponto x por $\nabla h(x)$.*

Teorema 1.2.5. (Teorema 7.45, (RIBEIRO; KARAS, 2013)) (Condições necessárias de segunda ordem) *Suponha que x^* é um minimizador local do problema (1.5) e que a condição de qualificação de independência linear é satisfeita em x^* . Então existe $\lambda^* \in \mathbb{R}^m$ tal que*

$$d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) d \geq 0,$$

para todo $d \in \mathcal{N}(\nabla h(x^*))$.

Teorema 1.2.6. (Teorema 7.46, (RIBEIRO; KARAS, 2013)) (Condições suficientes de segunda ordem) *Sejam $x^* \in \mathbb{R}^n$ e $\lambda^* \in \mathbb{R}^m$ tais que $h(x^*) = 0$ e $\nabla f(x^*) + \nabla h(x^*)^T \lambda^* = 0$. Suponha que*

$$d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) d > 0,$$

para todo $d \in \mathcal{N}(\nabla h(x^*)) \setminus \{0\}$, então x^* é um minimizador local estrito do problema (1.5).

1.3 CONVEXIDADE

Nesta seção, definimos conjuntos e funções convexas. Veremos nos teoremas que seguem propriedades de funções convexas que justificam o motivo de serem tão apreciadas em otimização.

Definição 1.3.1. *Um conjunto $C \subset \mathbb{R}^n$ é dito convexo se para quaisquer $x, y \in C$ e $t \in [0, 1]$, tem-se que $tx + (1 - t)y \in C$.*

Geometricamente, isso significa que o segmento $[x, y] = \{tx + (1 - t)y \mid t \in [0, 1]\}$ está inteiramente contido no conjunto C .

Definição 1.3.2. *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : C \rightarrow \mathbb{R}$ é dita convexa em C quando*

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y),$$

para todos $x, y \in C$ e $t \in [0, 1]$.

A função f é dita estritamente convexa em C quando a desigualdade anterior é estrita para todos $x \neq y$ e $t \in (0, 1)$.

Quando a função, além de convexa, é diferenciável, podemos mostrar que f é convexa em C se, e somente se,

$$f(y) \geq f(x) + \nabla f(x)^T(y - x),$$

para cada $x, y \in C$. Conseguimos, assim, uma caracterização mais simples para definir funções convexas. Perceba que segue diretamente dessa equivalência que todo ponto estacionário é minimizador global de f .

Teorema 1.3.1. (Teorema 3.12, (RIBEIRO; KARAS, 2013)) *Sejam $C \subset \mathbb{R}^n$ convexo e $f : C \rightarrow \mathbb{R}$ uma função convexa. Se x^* é minimizador local de f , então x^* é minimizador global de f . Além disso, se f é estritamente convexa, o minimizador é único.*

Concluimos a partir dos resultados anteriores que se a função objetivo é convexa, minimizadores locais e globais para problemas irrestritos são simples de caracterizar.

O teorema seguinte fornece outra maneira de caracterizar funções convexas.

Teorema 1.3.2. (Teorema 3.16, (RIBEIRO; KARAS, 2013)) *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função de classe \mathcal{C}^2 e $C \subset \mathbb{R}^n$ convexo.*

- (i) *Se $\nabla^2 f(x) \geq 0$, para todo $x \in C$, então f é convexa em C .*
- (ii) *Se f é convexa em C e $\text{int}C \neq \emptyset$, então $\nabla^2 f(x) \geq 0$, para todo $x \in C$.*

Definição 1.3.3. *Seja $D \subset \mathbb{R}^n$ um subconjunto convexo. Uma função $f : D \rightarrow \mathbb{R}$ é dita fortemente convexa com módulo $\sigma > 0$, quando para quaisquer $x, y \in D$ e $t \in [0, 1]$, tem-se*

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) - \sigma t(1 - t)\|x - y\|^2.$$

Teorema 1.3.3. (Teorema 3.4.19, (IZMAILOV; SOLODOV, 2014)) *(Compacidade do conjunto de nível de uma função fortemente convexa) Suponha que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja fortemente convexa em \mathbb{R}^n . Então o conjunto de nível*

$$L_{f, \mathbb{R}^n}(c) := \{x \in \mathbb{R}^n \mid f(x) \leq c\}$$

é compacto para todo $c \in \mathbb{R}$.

O próximo resultado, consequência do teorema anterior, assegura que o problema de minimizar uma função fortemente convexa em um conjunto fechado não vazio sempre tem uma solução.

Corolário 1.3.1. (Corolário 3.4.20, (IZMAILOV; SOLODOV, 2014)) *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função fortemente convexa e $\emptyset \neq D \subset \mathbb{R}^n$ um conjunto fechado qualquer. Então f tem um minimizador em D . Se D é convexo, o minimizador é único.*

1.4 OPERADOR PROJEÇÃO

Uma projeção ortogonal do ponto $x \in \mathbb{R}^n$ sobre um conjunto $D \subset \mathbb{R}^n$ é um ponto de D que está mais próximo de x , onde a distância é medida pela norma euclidiana. Ou seja, uma projeção de x sobre D é uma solução global do problema

$$\begin{aligned} &\text{minimizar} && \|y - x\| \\ &\text{sujeito a} && y \in D. \end{aligned}$$

A existência da projeção é garantida quando consideramos o conjunto D fechado.

Teorema 1.4.1. (Teorema 3.2.32, (IZMAILOV; SOLODOV, 2014)) *Considere $D \subset \mathbb{R}^n$ não vazio, convexo e fechado e $x \in \mathbb{R}^n$. A projeção de x sobre D , denotada por proj_D , existe e é única.*

Além disso, $\bar{x} = \text{proj}_D(x)$ se, e somente se,

$$\bar{x} \in D, \quad \langle x - \bar{x}, y - \bar{x} \rangle \leq 0, \quad \forall y \in D.$$

Como consequência do teorema anterior, temos

$$\|\text{proj}_D(x) - \text{proj}_D(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

Assim, proj_D é uma função contínua.

O próximo resultado, é uma combinação dos Teoremas 3.9 e 3.15, de (RIBEIRO; KARAS, 2013), que fornece uma condição necessária e suficiente de otimalidade quando desejamos minimizar uma função convexa em um conjunto convexo e fechado. A hipótese da função f ser convexa é essencial para a recíproca do teorema.

Teorema 1.4.2. *Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa diferenciável e $D \subset \mathbb{R}^n$ convexo e fechado. Então x^* é minimizador local de f em D se, e somente se*

$$\text{proj}_D(x^* - \nabla f(x^*)) = x^*.$$

1.5 MÉTODO CLÁSSICO DE LAGRANGIANO AUMENTADO: RESTRIÇÕES DE IGUALDADE

Para a construção desta seção consideramos como base a referência (NOCEDAL; WRIGHT, 2006). Todos os resultados estão devidamente indicados, as demonstrações foram omitidas.

Considere o problema com restrição de igualdade

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && h_i(x) = 0, \quad i = 1, 2, \dots, m \\ & && x \in \mathbb{R}^n, \end{aligned} \tag{1.6}$$

em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$, para cada $i = 1, 2, \dots, m$ são funções de classe \mathcal{C}^2 .

Em otimização com restrições de igualdade, métodos importantes substituem o problema original por uma sequência de subproblemas irrestritos nos quais as restrições são representadas por termos adicionados à função objetivo. O método de penalização quadrática, por exemplo, adiciona um múltiplo do quadrado da violação de cada restrição na função objetivo. Devido à sua simplicidade e por ser intuitivo, é uma abordagem bastante usada na prática, embora apresente algumas desvantagens de mau condicionamento. O método de Lagrangiano aumentado ou método dos multiplicadores reduz a possibilidade de mau condicionamento ao introduzir estimativas explícitas do multiplicador de Lagrange na função a ser minimizada.

1.5.1 O método de penalização quadrática

A função de penalização quadrática para o problema (1.6) é definida por

$$Q(x, r) := f(x) + \frac{r}{2} \|h(x)\|^2 = f(x) + \frac{r}{2} \sum_{i=1}^m h_i(x)^2, \tag{1.7}$$

onde $r > 0$ é o parâmetro de penalidade. Fazendo $r \rightarrow \infty$, penalizamos a violação da restrição com gravidade severa. É intuitivo considerar uma sequência de valores $\{r_k\}_{k \in \mathbb{N}}$ com $r_k \rightarrow \infty$ quando $k \rightarrow \infty$, e determinar um minimizador aproximado x^k de $Q(x, r_k)$ para cada $k \in \mathbb{N}$.

O algoritmo seguinte, é fundamentado na estrutura geral para algoritmos baseados na função de penalização quadrática (1.7), presente em Nocedal e Wright (2006, p 501).

Algoritmo 1: MÉTODO DE PENALIZAÇÃO QUADRÁTICA

Passo 0: Considere $r_0 > 0$, $\{\epsilon_k\}_{k \in \mathbb{N}}$ uma sequência não negativa tal que $\epsilon_k \rightarrow 0$, e um ponto inicial $x_s^0 \in \mathbb{R}^n$. Faça $k = 0$.

Passo 1: Obtenha um ponto estacionário aproximado x^k de $Q(\cdot, r_k)$, começando em x_s^k e parando quando $\|\nabla_x Q(x^k, r_k)\| \leq \epsilon_k$.

Passo 2: Escolha um novo parâmetro de penalidade $r_{k+1} > r_k$.

Passo 3: Escolha $x_s^{k+1} \in \mathbb{R}^n$.

Passo 4: Faça $k = k + 1$, volte ao **Passo 1**.

Em geral, x_s^{k+1} é escolhido como sendo o ponto estacionário aproximado da função de penalização quadrática, determinado no **Passo 1**, ou seja, $x_s^{k+1} = x^k$, ver (NOCEDAL; WRIGHT, 2006).

Como observado por Nocedal e Wright (2006), a sequência de parâmetros $\{r_k\}_{k \in \mathbb{N}}$ pode ser escolhida de acordo com a dificuldade em minimizar a função de penalidade a cada iteração. Quando a minimização da função de penalidade for custosa, para algum $k \in \mathbb{N}$, escolhamos r_{k+1} com um acréscimo modesto comparado a r_k , por exemplo $r_{k+1} = 1.5 r_k$. Por outro lado, quando a minimização se mostra barata, podemos tentar um crescimento mais ambicioso do parâmetro, por exemplo $r_{k+1} = 10 r_k$.

Segundo Nocedal e Wright (2006, p. 501) não há garantia de que $\|\nabla_x Q(x^k, r_k)\| \leq \epsilon_k$ seja sempre satisfeita no Algoritmo 1, pois os iterados podem se afastar da região viável quando o parâmetro de penalidade não for suficientemente grande. Implementações práticas geralmente incluem salvaguardas que aumentam o parâmetro de penalidade quando a violação da restrição não decresce rápido o suficiente, ou quando os iterados parecem divergir. No teorema seguinte, temos a análise dos resultados de convergência, quando permitimos minimizações inexatas (mas cada vez mais precisas) de $Q(\cdot, r_k)$.

Teorema 1.5.1. (Teorema 17.2, (NOCEDAL; WRIGHT, 2006)) *Suponha que as tolerâncias e os parâmetros de penalidade gerados pelo Algoritmo 1 satisfazem $\epsilon_k \rightarrow 0$ e $r_k \rightarrow \infty$. Se um ponto limite x^* da sequência $\{x^k\}_{k \in \mathbb{N}}$ é inviável, então será um ponto estacionário da função $\|h(x)\|^2$. Por outro lado, se o ponto limite x^* é um ponto viável e os gradientes das restrições $\nabla h_i(x^*)$ são linearmente independentes, então x^* é um ponto KKT para o problema (1.6). Para tais pontos, temos que para qualquer $\mathcal{K} \subset \mathbb{N}$ tal que $\lim_{k \in \mathcal{K}} x^k = x^*$, vale*

$$\lim_{k \in \mathcal{K}} r_k h_i(x^k) = \lambda_i^* \quad (1.8)$$

onde λ^* é o multiplicador que satisfaz as condições de KKT para o problema com restrições de igualdade (1.6).

Como é possível observar no Teorema 1.5.1, os minimizadores aproximados x^k de $Q(x; r_k)$ não satisfazem completamente as condições de viabilidade $h_i(x) = 0$. Em vez disso, são apenas perturbados (equação 1.8). Assim, temos as condições de viabilidade satisfeitas quando $r \rightarrow \infty$. No entanto, quando permitimos que o parâmetro de penalidade cresça ilimitadamente, estamos propensos a problemas de mau condicionamento da Hessiana $\nabla_{xx}^2 Q(x, r_k)$. Para maiores detalhes a respeito do mau condicionamento da matriz Hessiana consultar o Capítulo 17 da referência (NOCEDAL; WRIGHT, 2006).

1.5.2 Método clássico de Lagrangiano aumentado

Uma forma de evitar o crescimento ilimitado do parâmetro de penalidade, é por meio do acréscimo de um termo de penalização diferenciável (quadrático, por exemplo) à função Lagrangiano associada ao problema, e não apenas à função objetivo.

A função Lagrangiano aumentado clássica de Hestenes (1969) e Powell (1969), associada ao problema (1.6), é definida por

$$(x, \lambda, r) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \mapsto L(x, \lambda, r) \in \mathbb{R}$$

sendo

$$L(x, \lambda, r) := f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \frac{r}{2} \sum_{i=1}^m h_i(x)^2, \quad (1.9)$$

onde λ é o vetor de multiplicadores de Lagrange associado às restrições de igualdade e r o parâmetro de penalidade positivo.

Se denotarmos x^{k+1} como o minimizador aproximado de $L(x, \lambda^k, r_k)$, segue da condição de otimalidade para minimização irrestrita (Teorema 1.2.1), que

$$0 \approx \nabla_x L(x^{k+1}, \lambda^k, r_k) = \nabla f(x^{k+1}) + \sum_{i=1}^m [\lambda_i^k + r_k h_i(x^{k+1})] \nabla h_i(x^{k+1}).$$

Comparando o resultado anterior com a condição de otimalidade ($\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$) para o problema com restrição de igualdade (1.6), podemos deduzir que

$$\lambda_i^* \approx \lambda_i^k + r_k h_i(x^{k+1}), \text{ para cada } i = 1, \dots, m. \quad (1.10)$$

Ainda podemos reescrever como sendo

$$h_i(x^{k+1}) \approx \frac{1}{r_k} (\lambda_i^* - \lambda_i^k), \text{ para cada } i = 1, \dots, m.$$

Percebemos que quando λ^k se aproxima do multiplicador ótimo λ^* , a inviabilidade em x^{k+1} se torna muito menor do que $1/r_k$, em vez de proporcional a $1/r_k$, como ocorre em (1.8). A relação (1.10) sugere uma fórmula para atualização do vetor multiplicador de Lagrange, usando o minimizador aproximado x^{k+1} . Podemos considerar

$$\lambda^{k+1} = \lambda^k + r_k h(x^{k+1}). \quad (1.11)$$

Apresentamos a estrutura geral do algoritmo clássico de Lagrangiano aumentado, presente em Nocedal e Wright (2006, p 515), motivado especialmente pela discussão anterior.

Algoritmo 2: MÉTODO CLÁSSICO DE LAGRANGIANO AUMENTADO

Passo 0: Considere $r_0 > 0$, $\lambda^0 \in \mathbb{R}^m$ e $x^0 \in \mathbb{R}^n$. Faça $k = 0$.

Passo 1: Obtenha, possivelmente de forma aproximada,

$$x^{k+1} \text{ solução de } \min\{L(x, \lambda^k, r_k) \mid x \in \mathbb{R}^n\}.$$

considerando x^k como ponto inicial.

Passo 2: Atualize o multiplicador de Lagrange usando (1.11) para obter λ^{k+1} .

Passo 3: Escolha um novo parâmetro de penalidade $r_{k+1} \geq r_k$.

Passo 4: Faça $k = k + 1$, volte ao **Passo 1**.

As principais considerações que devemos ter em mente para a atualização do parâmetro de penalidade, além de regras práticas específicas de atualização podem ser consultadas em Bertsekas (1982a, p. 123).

O resultados seguintes justificam o uso da função Lagrangiano aumentado e dos métodos de multiplicadores para problemas que envolvem restrições de igualdade.

O primeiro teorema, cuja demonstração é detalhada em Nocedal e Wright (2006, p. 157) no Teorema 17.5, considera que tenhamos conhecimento prévio do vetor multiplicador de Lagrange λ^* . Apesar de que em problemas práticos geralmente conseguimos apenas estimativas para λ^* , o resultado e sua prova sugerem ser possível obter uma boa aproximação de um minimizador da função Lagrangiano aumentado, mesmo que o parâmetro de penalidade não seja particularmente grande, desde que tenhamos uma estimativa razoável para o multiplicador de Lagrange.

Teorema 1.5.2. (Teorema 17.5, (NOCEDAL; WRIGHT, 2006)) *Seja x^* uma solução local para (1.6) para o qual LICQ é satisfeita, e as condições de segunda ordem no Teorema 1.2.6 são satisfeitas para $\lambda = \lambda^*$. Então existe um valor limite \bar{r} tal que para todo $r \geq \bar{r}$, x^* é um minimizador local de $L(x, \lambda^*, r)$.*

No teorema seguinte, embasado no Teorema 17.6 de (NOCEDAL; WRIGHT, 2006), considerando $\lambda \neq \lambda^k$, são estabelecidas condições sob as quais existe um minimizador da função Lagrangiano aumentado que fica perto de x^* , além de atribuir *error bounds* tanto para x^{k+1} quanto para λ^{k+1} , obtidos a partir da resolução do subproblema a cada iteração.

Teorema 1.5.3. *Suponha que as hipóteses do Teorema 1.5.2 são satisfeitas em x^* e λ^* e seja \bar{r} escolhido como nesse Teorema. Então, existem escalares positivos δ , ϵ e M tais que as afirmações seguintes são válidas:*

a) Para todo λ^k e r_k satisfazendo

$$\|\lambda^k - \lambda^*\| \leq r_k \delta, \quad r_k \geq \bar{r}, \quad (1.12)$$

o problema

$$\min_x L(x, \lambda^k, r_k) \text{ sujeito a } \|x - x^*\| \leq \epsilon,$$

possui uma única solução x^{k+1} . Além disso, temos

$$\|x^{k+1} - x^*\| \leq \frac{M}{r_k} \|\lambda^k - \lambda^*\|. \quad (1.13)$$

b) Para todo λ^k e r_k satisfazendo (1.12), temos

$$\|\lambda^{k+1} - \lambda^*\| \leq \frac{M}{r_k} \|\lambda^k - \lambda^*\|, \quad (1.14)$$

onde λ^{k+1} é dado por (1.11).

2 PROBLEMA NÃO LINEAR DA MOCHILA

Neste capítulo, apresentamos os resultados obtidos com o estudo do problema não linear da mochila, o qual faz parte de uma classe mais geral do problema não linear de alocação de recursos. Consideramos problemas com uma restrição de igualdade linear e variáveis canalizadas. Estabelecemos uma classe de algoritmos de simples implementação, capaz de resolver uma gama de problemas de forma eficiente e robusta. Estudamos especialmente o caso quadrático, para o qual obtemos bons resultados de convergência para a sequência gerada pelo algoritmo. No caso não quadrático, estruturamos nossos resultados teóricos embasados em Birgin e Martínez (2014).

Na Seção 2.1, apresentamos a formulação do problema trabalhado neste capítulo, o algoritmo proposto para resolvê-lo e estabelecemos resultados de convergência. Na Seção 2.2, analisamos o comportamento do problema quando consideramos a função objetivo quadrática convexa, valendo-se da estrutura do problema, propomos um método mais simples para este caso particular. E por fim, na Seção 2.3, analisamos a estrutura do algoritmo proposto para algumas instâncias do problema da mochila quadrático, envolvendo inclusive uma restrição de desigualdade.

2.1 APRESENTAÇÃO DO PROBLEMA E METODOLOGIA

Foram consideradas algumas versões do problema não linear da mochila, mas todas são casos especiais do seguinte problema,

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && b^T x = c \\ &&& \ell \leq x \leq u, \end{aligned} \tag{2.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa e, pelo menos, de classe \mathcal{C}^2 , $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ e $\ell, u \in \mathbb{R}^n$ com $\ell < u$.

A ideia principal dos algoritmos propostos neste capítulo é penalizar apenas a restrição de igualdade do problema (2.1) utilizando a função Lagrangiano aumentado clássica de Hestenes (1969) e Powell (1969), que é dada por

$$L(x, \lambda, r) = f(x) + \lambda(b^T x - c) + \frac{r}{2}(b^T x - c)^2, \tag{2.2}$$

em que $\lambda \in \mathbb{R}$ é uma estimativa para o multiplicador de Lagrange e r um parâmetro de penalidade positivo. Deste modo, utilizamos a abordagem de Lagrangiano aumentado apenas para penalizar a restrição de igualdade. Considerando λ e r fixados teremos o

seguinte problema

$$\begin{aligned} & \text{minimizar} && L(x, \lambda, r) \\ & \text{sujeito a} && x \in \Omega, \end{aligned} \tag{2.3}$$

onde o conjunto $\Omega := \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$, e L é a função Lagrangiano aumentado (2.2). Como Ω é um conjunto convexo, f é uma função convexa e h é uma função afim, segue que o problema (2.3) é um problema convexo, como detalhado na próxima seção.

Nossa estratégia para resolver o problema (2.3) é minimizar a função Lagrangiano aumentado diretamente, desconsiderando as restrições de caixa em um primeiro momento, ou seja, encontramos um ponto intermediário \bar{x} tal que $\nabla_x L(\bar{x}, \lambda, r) = 0$, utilizando o método de Newton. E em seguida projetamos \bar{x} na caixa Ω , conseguindo a próxima iterada x . Desta forma, os algoritmos propostos aplicados ao problema não linear da mochila (2.1) são um tipo de Lagrangiano aumentado combinado com um método de Newton projetado para resolver o subproblema gerado.

2.1.1 Algoritmo de Lagrangiano aumentado

Apresentamos nesta subseção a estrutura geral do algoritmo proposto para resolver o problema não linear da mochila (2.1), que é embasada nas ideias de (BIRGIN; MARTÍNEZ, 2014) e (DINIZ-EHRHARDT et al., 2004).

Em síntese, a cada iteração do algoritmo proposto, resolvemos (possivelmente de forma aproximada) um subproblema gerado pela ideia do Lagrangiano aumentado, projetamos o ponto obtido na caixa $\Omega = \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$, atualizamos os multiplicadores de Lagrange (com salvaguardas) e o parâmetro de penalidade, e por fim, verificamos se o critério de parada é satisfeito.

No Algoritmo 3, proj_Ω é a projeção ortogonal sobre o conjunto Ω e $\text{proj}_{[\lambda_{\min}, \lambda_{\max}]}$ é a projeção sobre o intervalo $[\lambda_{\min}, \lambda_{\max}]$. Nas primeiras k_0 iterações do algoritmo, o parâmetro de penalidade r_k é mantido fixo, e a partir da iterada $k_0 + 1$ verificamos se houve um ganho suficiente na viabilidade e decidimos se mantemos, caso o ganho seja satisfatório, ou se atualizamos o parâmetro, aumentando seu valor para a próxima iteração, caso não consigamos um progresso satisfatório. As regras de atualização do parâmetro de penalidade e do vetor de multiplicadores de Lagrange (aproximado) foram embasadas em (BIRGIN; MARTÍNEZ, 2014).

Algoritmo 3: LAGRANGIANO AUMENTADO - RESTRIÇÃO LINEAR

Passo 0: Considere $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}$, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$, $\gamma > 1$, $r_0 > 0$, $\theta \in (0, 1)$, $k_0 \in \mathbb{N}$ e $\varepsilon \geq 0$. Faça $k = 0$.

Passo 1: Determine de forma aproximada

$$\bar{x}^k \in \arg \min \{L(x, \lambda_k, r_k) \mid x \in \mathbb{R}^n\}. \quad (2.4)$$

Passo 2: Calcule

$$x^{k+1} = \text{proj}_{\Omega}(\bar{x}^k). \quad (2.5)$$

Passo 3: Calcule uma nova aproximação do multiplicador de Lagrange

$$\lambda_{k+1} = \text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda_k + r_k(b^T x^{k+1} - c)). \quad (2.6)$$

Passo 4: Atualize o parâmetro de penalidade

$$r_{k+1} = \begin{cases} r_k, & \text{se } k \leq k_0 \text{ ou } |h(x^{k+1})| < \theta|h(x^k)| \\ \gamma r_k, & \text{caso contrário,} \end{cases} \quad (2.7)$$

onde $h(x) = b^T x - c$.

Passo 5: Teste a convergência

Se $|b^T x^{k+1} - c| < \varepsilon$ e $\|x^{k+1} - x^k\| < \varepsilon$, então pare. Caso contrário, faça $k = k + 1$ e retorne ao **Passo 1**.

O problema (2.4) estabelecido no **Passo 1** será chamado no decorrer do texto como Subproblema gerado pelo Algoritmo 3. A seguir, apresentamos nossa abordagem para resolver o subproblema. Como discutido anteriormente, a estratégia consiste em minimizar diretamente a função Lagrangiano aumentado, desconsiderando as restrições de caixa, e então projetar a solução encontrada na caixa, prosseguindo dessa forma a cada iteração.

Note que derivando a função Lagrangiano aumentado definida em (2.2), em relação a x , mantendo λ e r fixados, temos

$$\begin{aligned} \nabla_x L(x, \lambda, r) &= \nabla f(x) + rbb^T x + \lambda b - rc b \\ &= \nabla f(x) + rbb^T x + (\lambda - rc)b. \end{aligned} \quad (2.8)$$

Derivando novamente em relação a x , obtemos a Hessiana da função Lagrangiano aumentado, que tem a forma

$$\nabla_{xx}^2 L(x, \lambda, r) = \nabla^2 f(x) + rbb^T. \quad (2.9)$$

Como $\nabla^2 f$ e rbb^T são semidefinidas positivas, a Hessiana em (2.9) é semidefinida positiva, portanto a função Lagrangiano aumentado L estabelecida em (2.2) é

uma função convexa. Logo, o **Passo 1** do Algoritmo 3, é resolvido estabelecendo

$$\bar{x}^k \text{ tal que } \nabla_x L(\bar{x}^k, \lambda_k, r_k) = 0,$$

sendo esse o passo mais complicado do algoritmo proposto.

Observação 2.1.1. *O Passo 4 é embasado no método de Lagrangiano aumentado estabelecido por Birgin e Martínez (2014, p. 33), para resolver um problema similar. A ideia é controlar o quão eficiente o parâmetro de penalidade penaliza a restrição de igualdade.*

2.1.2 Análise de convergência

Nesta subseção discutimos as propriedades de convergência do Algoritmo 3, que são fundamentadas na teoria de convergência estabelecida para o método de Lagrangiano aumentado clássico por Birgin e Martínez (2014).

Para estabelecer os resultados de convergência referentes ao Algoritmo 3, consideramos a seguinte hipótese, em que cada x^k será a projeção em Ω de uma solução, possivelmente aproximada, do Subproblema gerado pelo Algoritmo 3.

Hipótese 2.1.1. (Hipótese 5.1, (BIRGIN; MARTÍNEZ, 2014)) *Para todo $k \in \mathbb{N}$, obtemos $x^k \in \Omega$ tal que*

$$L(x^k, \lambda_k, r_k) \leq L(x, \lambda_k, r_k) + \varepsilon_k, \text{ para todo } x \in \Omega,$$

sendo a sequência de tolerâncias $\{\varepsilon_k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+$ limitada.

A Hipótese 2.1.1 assegura um decréscimo suficiente, a cada iteração, da função Lagrangiano aumentado restrita à caixa.

No teorema seguinte, adaptação do Teorema 5.1 em (BIRGIN; MARTÍNEZ, 2014), usamos a Hipótese 2.1.1, e é evidente que mesmo considerando a sequência de tolerâncias $\{\varepsilon_k\}_{k \in \mathbb{N}}$ com valores possivelmente grandes obtemos, no limite, um minimizador global da medida de inviabilidade. Como trabalharemos com uma generalização desse problema no Capítulo 3, omitimos a demonstração que segue como caso particular do Teorema 3.2.1.

Teorema 2.1.1. *Considere $\{x^k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 3, satisfazendo a Hipótese 2.1.1. Seja x^* um ponto limite de $\{x^k\}_{k \in \mathbb{N}}$. Então x^* é um minimizador global do problema*

$$\begin{array}{ll} \text{minimizar} & h(x)^2 \\ \text{sujeito a} & x \in \Omega, \end{array}$$

onde $h(x) = b^T x - c$.

No próximo teorema, adaptação do Teorema 5.2 em (BIRGIN; MARTÍNEZ, 2014), constatamos que, assumindo que a sequência $\{\varepsilon_k\}_{k \in \mathbb{N}}$ converge para zero, é possível mostrar que, no caso viável, o algoritmo encontra minimizadores globais do problema (2.1). A demonstração foi omitida devido ao fato do resultado ser um caso particular do Teorema 3.2.2, cuja demonstração é detalhada no Capítulo 3.

Teorema 2.1.2. *Considere $\{x^k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 3, satisfazendo a Hipótese 2.1.1 com $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Seja x^* um ponto limite de $\{x^k\}_{k \in \mathbb{N}}$. Suponha que o problema (2.1) é viável. Então x^* é um minimizador global de (2.1).*

Como é possível notar, os resultados apresentados dependem fortemente da Hipótese 2.1.1. Para o caso quadrático, é possível eliminar a Hipótese 2.1.1 na demonstração de convergência. Aprofundamos essa discussão na próxima seção.

2.2 PROBLEMA DA MOCHILA QUADRÁTICO

Uma parcela considerável de artigos é voltada para o caso particular do problema (2.1), onde considera-se a função objetivo quadrática com variáveis contínuas ou inteiras, chamado de problema da mochila quadrático, ver (BAYÓN et al., 2013), (BRUCKER, 1984), (CAPRARA et al., 1999), (COMINETTI et al., 2014), (DUSSAULT et al., 1986), (FRANGIONI; GORGONE, 2013), (GALLO et al., 1980), (KIWIEL, 2008a), (KIWIEL, 2008b), (PANG, 1980), (PARDALOS; KOVOOR, 1990) e (ROBINSON et al., 1992). Podemos citar como exemplos de técnicas primárias utilizadas para resolver este problema os métodos quase-Newton e Newton, método do gradiente projetado, *branch and bound*, e técnicas de relaxamento.

Trabalhamos nesta seção com o problema da mochila quadrático e a seguinte formulação,

$$\begin{aligned} & \text{minimizar} && \frac{1}{2}x^T P x - a^T x \\ & \text{sujeito a} && b^T x = c \\ & && \ell \leq x \leq u, \end{aligned} \tag{2.10}$$

em que $P \in \mathbb{R}^{n \times n}$ é uma matriz simétrica e definida positiva, a e b são vetores em \mathbb{R}^n e $c \in \mathbb{R}$. Problemas semelhantes ao definido em (2.10) surgem em diversas aplicações, incluindo treinamento de máquinas de vetores de suporte, como é possível conferir em (DAI; FLETCHER, 2006) e (COMINETTI et al., 2014).

Existem trabalhos na literatura relacionando métodos de Lagrangiano aumentado com programação quadrática, por exemplo (BUENO et al., 2019). Nossa proposta de método se diferencia por explorar a álgebra linear de forma eficiente ao considerarmos subproblemas irrestritos, como detalharemos ao longo desta seção.

Consideramos a análise do caso particular do Algoritmo 3 para o problema quadrático estabelecido em (2.10). A função Lagrangiano aumentado para este caso é tal como a definida em (2.2), com a especificação da função objetivo quadrática, ou seja, $f(x) = \frac{1}{2}x^T Px - a^T x$.

Supondo fixados o multiplicador de Lagrange e o parâmetro de penalidade, o gradiente da função Lagrangiano aumentado em relação a x , calculado em (2.8), é dado por

$$\nabla_x L(x, \lambda, r) = (P + rbb^T)x + v(\lambda, r), \quad (2.11)$$

com $v(\lambda, r) = (\lambda - rc)b - a$. Consequentemente, obtemos a Hessiana,

$$\nabla_{xx}^2 L(x, \lambda, r) = P + rbb^T. \quad (2.12)$$

Como P é uma matriz definida positiva, segue que a Hessiana $P + rbb^T$ é definida positiva, e portanto inversível. Esta informação é muito importante, pois em geral, resolve-se o subproblema restrito por meio de algum método clássico como, por exemplo, Gradiente Projetado (DINIZ-EHRHARDT et al., 2004) e o método de Newton, porém, neste caso, olhando para a equação (2.8) vemos que o subproblema, desconsiderando a caixa, tem uma forma fechada. Para obter a inversa da matriz $P + rbb^T$ recorreremos à fórmula de Sherman-Morrison (SHERMAN; MORRISON, 1950), a qual estabelece que, dada uma matriz $Q \in \mathbb{R}^{n \times n}$ inversível e dois vetores arbitrários $u, v \in \mathbb{R}^n$ tais que $1 + v^T Q^{-1} u \neq 0$, então a matriz $Q + uv^T$ é inversível e sua inversa é dada por

$$(Q + uv^T)^{-1} = Q^{-1} - \frac{Q^{-1}uv^T Q^{-1}}{1 + v^T Q^{-1}u}.$$

Aplicando ao nosso problema, identificando na fórmula acima $Q = P$, $u = rb$ e $v = b$, segue que $1 + rb^T P^{-1} b \neq 0$ e a inversa da matriz $P + rbb^T$ é

$$(P + rbb^T)^{-1} = P^{-1} - \alpha P^{-1} bb^T P^{-1}, \quad (2.13)$$

sendo $\alpha = \frac{r}{1 + rb^T P^{-1} b}$.

Nos casos em que a matriz inversa P^{-1} é fácil de determinar, temos uma forma elegante e simples para calcular a inversa da matriz $P + rbb^T$. Usamos essa informação para estabelecer algoritmos que resolvem o problema (2.10), em que minimizamos a função Lagrangiano aumentado diretamente via fórmula fechada, e obtemos o novo iterado, projetando esse ponto na caixa Ω .

Cabe enfatizar que a tarefa mais complicada do Algoritmo 3 é determinar uma solução do subproblema (2.4), para o qual apresentamos uma fórmula fechada. De fato,

utilizando as condições de otimalidade de primeira ordem e a equação (2.11), obtemos

$$(P + rbb^T)x = a + (rc - \lambda)b, \quad (2.14)$$

usando a equação (2.13), podemos escrever

$$x = (P^{-1} - \alpha P^{-1}bb^T P^{-1})[a + (rc - \lambda)b], \quad (2.15)$$

equivalentemente,

$$x = P^{-1}a + (rc - \lambda)P^{-1}b - \alpha P^{-1}bb^T P^{-1}a - \alpha(rc - \lambda)P^{-1}bb^T P^{-1}b. \quad (2.16)$$

Denotando

$$P^{-1}a = y \quad \text{e} \quad P^{-1}b = z, \quad (2.17)$$

a solução do subproblema (2.4) na iteração k , é dada por

$$\bar{x}^k = y + [(r_k c - \lambda_k) - \alpha_k(b^T y + (r_k c - \lambda_k)b^T z)]z, \quad (2.18)$$

onde $\alpha_k = \frac{r_k}{1 + r_k b^T z}$.

Note que o ponto \bar{x}^k em (2.18) é determinado resolvendo $\nabla_x L(x, \lambda, r) = 0$. Como a função Lagrangiano aumentado L é uma função quadrática e estritamente convexa, resolver $\nabla_x L(x, \lambda, r) = 0$ é equivalente a aplicar o método de Newton para o problema $\min\{L(x, \lambda, r) \mid x \in \mathbb{R}^n\}$, que aparece no **Passo 1** do Algoritmo 3. No entanto, para o caso geral, o ponto \bar{x}^k é determinado pelo método de Newton, em todas as k iterações do Algoritmo 3.

Observação 2.2.1. *Há dois aspectos a serem destacados. O primeiro é que os sistemas lineares (2.17) dependem exclusivamente dos dados do problema, então uma única solução de cada um é computada. Além disso, como P é definida positiva, podemos usar a Decomposição de Cholesky $P = G^T G$ e então resolver quatro sistemas lineares triangulares dados por $G^T G y = a$ e $G^T G z = b$. O segundo ponto é que a solução do subproblema (2.4), determinada na equação (2.18), depende apenas do parâmetro de penalidade r_k , e do multiplicador λ_k .*

O exemplo seguinte ilustra de forma simples os passos do Algoritmo 3.

Exemplo 2.2.1. *Considere o problema*

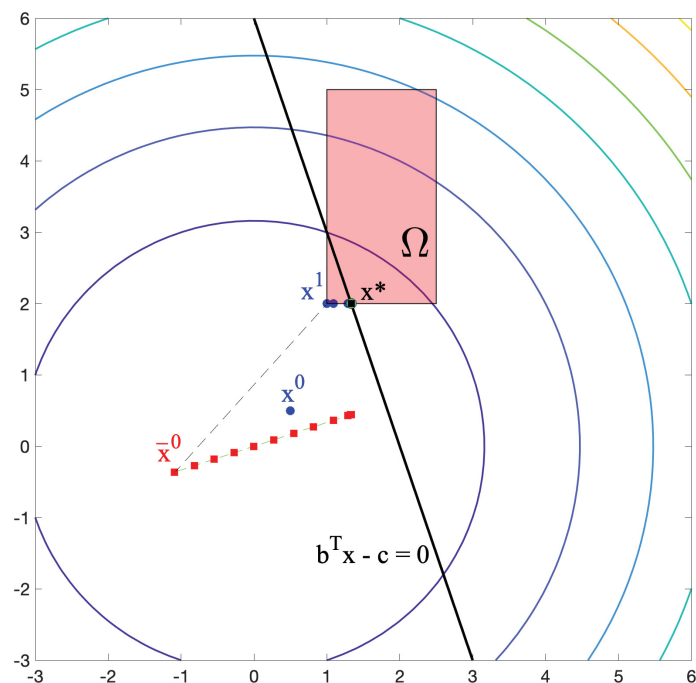
$$\begin{aligned} \text{minimizar} \quad & \frac{1}{2}(x_1^2 + x_2^2) \\ \text{sujeito a} \quad & 3x_1 + x_2 = 6 \\ & 1 \leq x_1 \leq 2.5 \\ & 2 \leq x_2 \leq 5. \end{aligned} \quad (2.19)$$

Trata-se de um problema quadrático, com o vetor $x \in \mathbb{R}^2$, a matriz $P = I \in \mathbb{R}^{2 \times 2}$, $a = (0, 0)^T$, $b = (3, 1)^T$, $c = 6$, $\ell = (1, 2)^T$ e $u = (2.5, 5)^T$.

Para inicializar o algoritmo tomamos $r_0 = 1$, $\lambda_0 = 10$, a acurácia $\varepsilon = 10^{-6}$, o parâmetro $k_0 = 50$ e a caixa $[\lambda_{\min}, \lambda_{\max}] = [-10, 10]$. Partindo do ponto inicial $x^0 = (0.5, 0.5)^T$, o Algoritmo 3 determina a solução em 18 iterações para resolver o problema (2.19).

Na FIGURA 1, estão representadas as curvas de nível da função objetivo do problema (2.19), a restrição de igualdade é destacada pela reta na cor preto e conjunto Ω é representado pela caixa na cor salmão.

FIGURA 1 – PASSOS DO ALGORITMO PARA UM PROBLEMA QUADRÁTICO



FONTE: A autora.

Partindo do ponto inicial x^0 , o algoritmo determina \bar{x}^0 , solução do subproblema gerado no **Passo 1**, esse ponto é então projetado na caixa Ω determinando a próxima iterada x^1 . Assim, na FIGURA 1, os quadrados vermelhos destacam as soluções dos subproblemas e os círculos azuis representam a projeção na caixa Ω . O ponto $x^* = (1.33, 2)^T$ é a solução do problema (2.19) fornecida pelo Algoritmo 3.

2.2.1 Convergência no caso quadrático

Nesta subseção, estabelecemos propriedades de convergência do Algoritmo 3 aplicado ao caso particular onde a função objetivo é quadrática dada em (2.10). Neste

caso particular, aproveitamos a fórmula de Sherman-Morrison para deduzir a inversa da matriz $P + rbb^T$, por meio de algumas manipulações algébricas conseguimos especificar limitantes superiores para a diferença dos iterados consecutivos \bar{x}^{k-1} e \bar{x}^k , e a partir disso verificamos a convergência das sequências $\{x^k\}_{k \in \mathbb{N}}$ e $\{\bar{x}^k\}_{k \in \mathbb{N}}$ geradas pelo algoritmo. Mostramos que quando a sequência dos parâmetros de penalidade é limitada, temos a viabilidade do ponto limite da sequência $\{x^k\}_{k \in \mathbb{N}}$. E por fim, consideramos a análise da otimalidade do ponto limite da sequência $\{x^k\}_{k \in \mathbb{N}}$.

Teorema 2.2.1. *Suponha que a função objetivo f é quadrática, como em (2.10). Então as sequências $\{x^k\}_{k \in \mathbb{N}}$ e $\{\bar{x}^k\}_{k \in \mathbb{N}}$, geradas pelo Algoritmo 3 são convergentes.*

Demonstração. Analisamos primeiro o caso em que a sequência $\{r_k\}$ é limitada. Por simplicidade, vamos considerar $r_k = r$ para todo $k \in \mathbb{N}$. Segue da igualdade (2.14)

$$\begin{aligned}\bar{x}^k &= (P + rbb^T)^{-1}[a - (\lambda_k - rc)b] \\ &= (P + rbb^T)^{-1}a - (P + rbb^T)^{-1}(\lambda_k - rc)b,\end{aligned}\tag{2.20}$$

analogamente,

$$\begin{aligned}\bar{x}^{k-1} &= (P + rbb^T)^{-1}[a - (\lambda_{k-1} - rc)b] \\ &= (P + rbb^T)^{-1}a - (P + rbb^T)^{-1}(\lambda_{k-1} - rc)b,\end{aligned}\tag{2.21}$$

assim, das equações (2.20) e (2.21), temos

$$\bar{x}^k - \bar{x}^{k-1} = (\lambda_{k-1} - \lambda_k)(P + rbb^T)^{-1}b.\tag{2.22}$$

Usando a fórmula de Sherman-Morrison (2.13),

$$\begin{aligned}(P + rbb^T)^{-1}b &= \left(P^{-1} - \frac{rP^{-1}bb^TP^{-1}}{1 + rb^TP^{-1}b} \right) b \\ &= \left(\frac{P^{-1} + rP^{-1}bb^TP^{-1} - rP^{-1}bb^TP^{-1}}{1 + rb^TP^{-1}b} \right) b \\ &= \frac{P^{-1}b}{1 + rb^TP^{-1}b}.\end{aligned}\tag{2.23}$$

Combinando (2.22) com (2.23), obtemos

$$\bar{x}^k - \bar{x}^{k-1} = (\lambda_{k-1} - \lambda_k) \frac{P^{-1}b}{1 + rb^TP^{-1}b}.\tag{2.24}$$

Por outro lado, como a projeção é não expansiva, segue que

$$\|x^{k+1} - x^k\| = \|\text{proj}_\Omega(\bar{x}^k) - \text{proj}_\Omega(\bar{x}^{k-1})\| \leq \|\bar{x}^k - \bar{x}^{k-1}\| = |\lambda_k - \lambda_{k-1}| \frac{\|P^{-1}b\|}{1 + rb^TP^{-1}b},\tag{2.25}$$

sendo que a última igualdade segue de (2.24).

Note que se $r_k = r$ para todo $k \in \mathbb{N}$, pela regra de atualização do parâmetro de penalidade (2.7), a desigualdade $|h(x^{k+1})| < \theta |h(x^k)|$ vale para todo $k \geq k_0 + 1$, sendo $k_0 \in \mathbb{N}$ definido no Algoritmo 3. Lembrando que $h(x) = b^T x - c$, segue que

$$0 \leq |b^T x^{k+1} - c| < \theta |b^T x^k - c| < \theta^2 |b^T x^{k-1} - c| < \dots < \theta^{k-k_0} |b^T x^{k_0+1} - c|. \quad (2.26)$$

Consideremos $\bar{\lambda}_{k+1} := \lambda_k + r(b^T x^{k+1} - c)$, para cada $k \in \mathbb{N}$, de forma equivalente podemos escrever

$$\frac{\bar{\lambda}_{k+1} - \lambda_k}{r} = b^T x^{k+1} - c, \text{ para todo } k, \quad (2.27)$$

uma vez que o parâmetro r é não nulo. Combinando (2.26) e (2.27), obtemos

$$\frac{|\bar{\lambda}_{k+1} - \lambda_k|}{r} < \theta \frac{|\bar{\lambda}_k - \lambda_{k-1}|}{r} < \dots < \theta^{k-k_0} \frac{|\bar{\lambda}_{k_0+1} - \lambda_{k_0}|}{r}. \quad (2.28)$$

Usando a definição de λ_{k+1} e o fato que a projeção é não expansiva

$$\begin{aligned} |\lambda_{k+1} - \lambda_k| &= |\text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda_k + r(b^T x^{k+1} - c)) - \text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda_k)| \\ &\leq |\lambda_k + r(b^T x^{k+1} - c) - \lambda_k| \\ &= r|b^T x^{k+1} - c| \\ &= |\bar{\lambda}_{k+1} - \lambda_k|. \end{aligned} \quad (2.29)$$

Segue das desigualdades (2.28) e (2.29) que

$$|\lambda_{k+1} - \lambda_k| \leq |\bar{\lambda}_{k+1} - \lambda_k| < \theta^{k-k_0} |\bar{\lambda}_{k_0+1} - \lambda_{k_0}|. \quad (2.30)$$

Como $\theta \in (0, 1)$, concluímos que a sequência $\{\lambda_{k+1} - \lambda_k\}_{k \in \mathbb{N}}$ converge para zero. Portanto, tomando-se o limite na igualdade (2.25), temos

$$\|x^k - x^{k-1}\| \rightarrow 0. \quad (2.31)$$

Considere $p \in \mathbb{N}^*$, arbitrário, combinando a desigualdade triangular com as equações (2.25) e (2.30)

$$\begin{aligned} \|x^{k+p+1} - x^{k+1}\| &\leq \|\bar{x}^{k+p} - \bar{x}^k\| \\ &= \|\bar{x}^{k+p} - \bar{x}^{k+p-1} + \bar{x}^{k+p-1} - \dots + \bar{x}^{k+1} - \bar{x}^k\| \\ &\leq \|\bar{x}^{k+p} - \bar{x}^{k+p-1}\| + \dots + \|\bar{x}^{k+1} - \bar{x}^k\| \\ &< (\theta^{k+p-k_0-1} + \dots + \theta^{k-k_0}) |\bar{\lambda}_{k_0+1} - \lambda_{k_0}| \frac{\|P^{-1}b\|}{1 + rb^T P^{-1}b}. \end{aligned} \quad (2.32)$$

Como $\theta \in (0, 1)$, a série geométrica $\sum_{j \in \mathbb{N}} \theta^j$ é convergente, então $\sum_{j=k}^{k+p-k_0-1} \theta^{j-k_0}$ converge para zero, quando $k \rightarrow \infty$. O que implica que $\{x^k\}_{k \in \mathbb{N}}$ e $\{\bar{x}^k\}_{k \in \mathbb{N}}$ são sequências de Cauchy, logo são convergentes.

Agora vamos analisar o caso onde $r_k \rightarrow \infty$. Sabemos pela equação (2.14) que

$$\bar{x}^k = (P + r_k b b^T)^{-1} [a - (\lambda_k - r_k c) b].$$

Utilizando a fórmula de Sherman-Morrison e fazendo algumas manipulações algébricas, obtemos

$$\begin{aligned} \bar{x}^k &= \left(P^{-1} - \frac{r_k P^{-1} b b^T P^{-1}}{1 + r_k b^T P^{-1} b} \right) a - (\lambda_k - r_k c) \left(P^{-1} - \frac{r_k P^{-1} b b^T P^{-1}}{1 + r_k b^T P^{-1} b} \right) b \\ &= P^{-1} a - \left(\frac{r_k P^{-1} b b^T P^{-1}}{1 + r_k b^T P^{-1} b} \right) a - (\lambda_k - r_k c) \left(\frac{P^{-1} b + r_k P^{-1} b b^T P^{-1} b - r_k P^{-1} b b^T P^{-1} b}{1 + r_k b^T P^{-1} b} \right) \\ &= P^{-1} a - \left(\frac{r_k P^{-1} b b^T P^{-1}}{1 + r_k b^T P^{-1} b} \right) a - (\lambda_k - r_k c) \left(\frac{P^{-1} b}{1 + r_k b^T P^{-1} b} \right), \end{aligned} \quad (2.33)$$

de modo similar,

$$\bar{x}^{k-1} = P^{-1} a - \left(\frac{r_{k-1} P^{-1} b b^T P^{-1}}{1 + r_{k-1} b^T P^{-1} b} \right) a - (\lambda_{k-1} - r_{k-1} c) \left(\frac{P^{-1} b}{1 + r_{k-1} b^T P^{-1} b} \right). \quad (2.34)$$

Segue das equações (2.33) e (2.34) que

$$\begin{aligned} \bar{x}^k - \bar{x}^{k-1} &= \left[\frac{r_{k-1}}{1 + r_{k-1} b^T P^{-1} b} - \frac{r_k}{1 + r_k b^T P^{-1} b} \right] P^{-1} b b^T P^{-1} a \\ &\quad + \left[\frac{(\lambda_{k-1} - r_{k-1} c)}{1 + r_{k-1} b^T P^{-1} b} - \frac{(\lambda_k - r_k c)}{1 + r_k b^T P^{-1} b} \right] P^{-1} b \\ &= \left[\frac{1}{\frac{1}{r_{k-1}} + b^T P^{-1} b} - \frac{1}{\frac{1}{r_k} + b^T P^{-1} b} \right] P^{-1} b b^T P^{-1} a \\ &\quad + \left[\frac{\left(\frac{\lambda_{k-1}}{r_{k-1}} - c \right)}{\frac{1}{r_{k-1}} + b^T P^{-1} b} - \frac{\left(\frac{\lambda_k}{r_k} - c \right)}{\frac{1}{r_k} + b^T P^{-1} b} \right] P^{-1} b. \end{aligned} \quad (2.35)$$

Tomando a norma em (2.35) e usando a desigualdade triangular, temos

$$\begin{aligned} \|\bar{x}^k - \bar{x}^{k-1}\| &\leq \left| \frac{1}{\frac{1}{r_k} + b^T P^{-1} b} - \frac{1}{\frac{1}{r_{k-1}} + b^T P^{-1} b} \right| \|b^T P^{-1} a\| \|P^{-1} b\| \\ &\quad + \left| \frac{\left(\frac{\lambda_k}{r_k} - c \right)}{\frac{1}{r_k} + b^T P^{-1} b} - \frac{\left(\frac{\lambda_{k-1}}{r_{k-1}} - c \right)}{\frac{1}{r_{k-1}} + b^T P^{-1} b} \right| \|P^{-1} b\|. \end{aligned} \quad (2.36)$$

Defina para cada $k \in \mathbb{N}$

$$C_k := \frac{1}{\frac{1}{r_k} + b^T P^{-1} b} \quad \text{e} \quad D_k := \frac{\left(\frac{\lambda_k}{r_k} - c\right)}{\frac{1}{r_k} + b^T P^{-1} b}. \quad (2.37)$$

Substituindo na desigualdade (2.36), temos

$$\|\bar{x}^k - \bar{x}^{k-1}\| \leq |C_k - C_{k-1}| \|b^T P^{-1} a\| \|P^{-1} b\| + |D_k - D_{k-1}| \|P^{-1} b\|.$$

Segue que,

$$\begin{aligned} \|\bar{x}^{k+1} - \bar{x}^k + \bar{x}^k - \bar{x}^{k-1}\| &\leq \|\bar{x}^{k+1} - \bar{x}^k\| + \|\bar{x}^k - \bar{x}^{k-1}\| \\ &\leq (|C_{k+1} - C_k| + |C_k - C_{k-1}|) \|b^T P^{-1} a\| \|P^{-1} b\| \\ &\quad + (|D_{k+1} - D_k| + |D_k - D_{k-1}|) \|P^{-1} b\|. \end{aligned}$$

De modo geral, considerando $p \in \mathbb{N} \setminus \{0\}$, arbitrário, combinando a não expansividade da projeção com essas informações, obtemos

$$\begin{aligned} \|x^{k+p+1} - x^{k+1}\| &\leq \|\bar{x}^{k+p} - \bar{x}^k\| \\ &= \|\bar{x}^{k+p} - \bar{x}^{k+p-1} + \bar{x}^{k+p-1} - \dots + \bar{x}^{k+1} - \bar{x}^k\| \\ &\leq \|\bar{x}^{k+p} - \bar{x}^{k+p-1}\| + \dots + \|\bar{x}^{k+1} - \bar{x}^k\| \\ &\leq \|b^T P^{-1} a\| \|P^{-1} b\| \sum_{i=k}^{k+p-1} |C_{i+1} - C_i| + \|P^{-1} b\| \sum_{i=k}^{k+p-1} |D_{i+1} - D_i|. \end{aligned} \quad (2.38)$$

Segue da definição de C_i em (2.37)

$$|C_{i+1} - C_i| = \left| \frac{1}{\frac{1}{r_{i+1}} + b^T P^{-1} b} - \frac{1}{\frac{1}{r_i} + b^T P^{-1} b} \right| = \left| \frac{r_{i+1}}{1 + r_{i+1} b^T P^{-1} b} - \frac{r_i}{1 + r_i b^T P^{-1} b} \right|,$$

como $r_{i+1} = \gamma r_i$, para todo $i > k_0$

$$\begin{aligned} |C_{i+1} - C_i| &= \left| \frac{\gamma r_i}{1 + \gamma r_i b^T P^{-1} b} - \frac{r_i}{1 + r_i b^T P^{-1} b} \right| \\ &= \left| \frac{\gamma r_i (1 + r_i b^T P^{-1} b) - r_i (1 + \gamma r_i b^T P^{-1} b)}{(1 + \gamma r_i b^T P^{-1} b)(1 + r_i b^T P^{-1} b)} \right| \\ &= \left| \frac{\gamma r_i + \gamma r_i^2 b^T P^{-1} b - r_i - \gamma r_i^2 b^T P^{-1} b}{(1 + \gamma r_i b^T P^{-1} b)(1 + r_i b^T P^{-1} b)} \right| \\ &= \left| \frac{(\gamma - 1) r_i}{1 + r_i b^T P^{-1} b + \gamma r_i b^T P^{-1} b + \gamma r_i^2 (b^T P^{-1} b)^2} \right|, \end{aligned}$$

note que $1 + r_i b^T P^{-1} b + \gamma r_i b^T P^{-1} b \geq 1$, pois a matriz P é definida positiva, $\gamma > 1$ e $r_i > 0$, o que implica

$$|C_{i+1} - C_i| \leq \frac{(\gamma - 1)}{\gamma r_i (b^T P^{-1} b)^2},$$

além disso, segundo a regra de atualização do parâmetro de penalidade, $r_i = \gamma^{i-k_0} r_{k_0}$, o que nos leva a seguinte desigualdade

$$|C_{i+1} - C_i| \leq \frac{(\gamma - 1)}{r_{k_0} (b^T P^{-1} b)^2} \left(\frac{1}{\gamma}\right)^{i-k_0+1}.$$

Como $\gamma > 1$ a série $\sum_{i=0}^{\infty} \frac{(\gamma - 1)}{r_{k_0} (b^T P^{-1} b)^2} \left(\frac{1}{\gamma}\right)^{i-k_0+1}$ é convergente, portanto $\sum_{i=0}^{\infty} |C_{i+1} - C_i|$ também é convergente (critério de comparação de séries). Logo, segundo o Critério de Cauchy para séries, segue que

$$\sum_{i=k}^{k+p-1} |C_{i+1} - C_i| \rightarrow 0, \text{ quando } k \rightarrow \infty. \quad (2.39)$$

Consideremos agora a análise para D_i , da definição em (2.37) segue que

$$|D_{i+1} - D_i| = \left| \frac{\frac{\lambda_{i+1}}{r_{i+1}} - c}{\frac{1}{r_{i+1}} + b^T P^{-1} b} - \frac{\frac{\lambda_i}{r_i} - c}{\frac{1}{r_i} + b^T P^{-1} b} \right| = \left| \frac{\lambda_{i+1} - r_{i+1}c}{1 + r_{i+1}b^T P^{-1} b} - \frac{\lambda_i - r_i c}{1 + r_i b^T P^{-1} b} \right|,$$

como $r_{i+1} = \gamma r_i$, para todo $i \geq k_0$

$$\begin{aligned} |D_{i+1} - D_i| &= \left| \frac{\lambda_{i+1} - \gamma r_i c}{1 + \gamma r_i b^T P^{-1} b} - \frac{\lambda_i - r_i c}{1 + r_i b^T P^{-1} b} \right| \\ &= \left| \frac{\lambda_{i+1} - \gamma r_i c + \lambda_{i+1} r_i b^T P^{-1} b - \lambda_i + r_i c - \gamma \lambda_i r_i b^T P^{-1} b}{(1 + \gamma r_i b^T P^{-1} b)(1 + r_i b^T P^{-1} b)} \right| \\ &= \left| \frac{(\lambda_{i+1} - \lambda_i) + (1 - \gamma) r_i c + (\lambda_{i+1} - \gamma \lambda_i) r_i b^T P^{-1} b}{1 + r_i b^T P^{-1} b + \gamma r_i b^T P^{-1} b + \gamma r_i^2 (b^T P^{-1} b)^2} \right| \\ &\leq \frac{|\lambda_{i+1} - \lambda_i| + (\gamma - 1) r_i |c| + |\lambda_{i+1} - \gamma \lambda_i| r_i b^T P^{-1} b}{1 + r_i b^T P^{-1} b + \gamma r_i b^T P^{-1} b + \gamma r_i^2 (b^T P^{-1} b)^2}, \end{aligned}$$

como $\lambda_i \in [\lambda_{\min}, \lambda_{\max}]$ e $\gamma > 1$, existem $L_1, L_2 \geq 0$ tais que $|\lambda_{i+1} - \lambda_i| \leq L_1$ e $|\lambda_{i+1} - \gamma \lambda_i| \leq L_2$. Além disso, temos $1 + r_i b^T P^{-1} b + \gamma r_i b^T P^{-1} b \geq 1$, pois a matriz P é definida positiva, $\gamma > 1$ e $r_i > 0$, assim

$$|D_{i+1} - D_i| \leq \frac{L_1}{\gamma r_i^2 (b^T P^{-1} b)^2} + \frac{(\gamma - 1)|c|}{\gamma r_i (b^T P^{-1} b)^2} + \frac{L_2 b^T P^{-1} b}{\gamma r_i (b^T P^{-1} b)^2},$$

para cada $i > k_0$, temos $r_i = \gamma^{i-k_0} r_{k_0}$, logo

$$|D_{i+1} - D_i| \leq \frac{L_1}{r_{k_0}^2 (b^T P^{-1} b)^2} \left(\frac{1}{\gamma}\right)^{2i-2k_0+1} + \left[\frac{(\gamma-1)|c|}{r_{k_0} (b^T P^{-1} b)^2} + \frac{L_2 b^T P^{-1} b}{r_{k_0} (b^T P^{-1} b)^2} \right] \left(\frac{1}{\gamma}\right)^{i-k_0+1},$$

Como $\gamma > 1$, as séries

$$\sum_{i=0}^{\infty} \frac{L_1}{r_{k_0}^2 (b^T P^{-1} b)^2} \left(\frac{1}{\gamma}\right)^{2i-2k_0+1} \text{ e } \sum_{i=0}^{\infty} \left[\frac{(\gamma-1)|c|}{r_{k_0} (b^T P^{-1} b)^2} + \frac{L_2 b^T P^{-1} b}{r_{k_0} (b^T P^{-1} b)^2} \right] \left(\frac{1}{\gamma}\right)^{i-k_0+1},$$

são convergentes, portanto $\sum_{i=0}^{\infty} |D_{i+1} - D_i|$ também é convergente (critério de comparação de séries). Logo, segundo o Critério de Cauchy para séries, segue que

$$\sum_{i=k}^{k+p-1} |D_{i+1} - D_i| \rightarrow 0, \text{ quando } k \rightarrow \infty. \quad (2.40)$$

Além disso, os termos $|b^T P^{-1} b|$ e $\|P^{-1} b\|$ são limitados, logo por (2.38), (2.39) e (2.40) temos que $\{x^k\}_{k \in \mathbb{N}}$ e $\{\bar{x}^k\}_{k \in \mathbb{N}}$ são seqüências de Cauchy, portanto são convergentes. \square

Para os próximos resultados desse capítulo assumiremos que a seqüência de parâmetros de penalidade $\{r_k\}_{k \in \mathbb{N}}$ é limitada.

Teorema 2.2.2. *Considere a função objetivo quadrática, como em (2.10), e suponha que a seqüência de parâmetros de penalidade $\{r_k\}_{k \in \mathbb{N}}$ é limitada. Então o limite x^* da seqüência $\{x^k\}_{k \in \mathbb{N}}$ definida pelo Algoritmo 3 é um ponto viável.*

Demonstração. Como a seqüência de parâmetros de penalidade é limitada, podemos considerar $r_k = r$ para todo $k \in \mathbb{N}$. Pela regra de atualização do parâmetro de penalidade (2.7), a desigualdade $|h(x^{k+1})| < \theta |h(x^k)|$ vale para todo $k \geq k_0 + 1$, sendo $k_0 \in \mathbb{N}$ definido no Algoritmo 3. Lembrando que $h(x) = b^T x - c$, então

$$0 \leq |b^T x^{k+1} - c| < \theta |b^T x^k - c| < \theta^2 |b^T x^{k-1} - c| < \dots < \theta^{k-k_0} |b^T x^{k_0+1} - c|. \quad (2.41)$$

Consideremos $\bar{\lambda}_{k+1} := \lambda_k + r(b^T x^{k+1} - c)$, para cada $k \in \mathbb{N}$, de modo equivalente podemos escrever

$$\frac{\bar{\lambda}_{k+1} - \lambda_k}{r} = b^T x^{k+1} - c, \text{ para todo } k \in \mathbb{N}, \quad (2.42)$$

uma vez que o parâmetro r é não nulo. Combinando (3.14) e (3.15), obtemos

$$\frac{|\bar{\lambda}_{k+1} - \lambda_k|}{r} < \theta \frac{|\bar{\lambda}_k - \lambda_{k-1}|}{r} < \dots < \theta^{k-k_0} \frac{|\bar{\lambda}_{k_0+1} - \lambda_{k_0}|}{r}, \quad (2.43)$$

sendo $\theta \in (0, 1)$

$$\lim_{k \rightarrow \infty} \left(\frac{\bar{\lambda}_{k+1} - \lambda_k}{r} \right) = 0.$$

Como a igualdade (3.15) é válida para cada $k \in \mathbb{N}$, segue que

$$\lim_{k \rightarrow \infty} h(x^{k+1}) = \lim_{k \rightarrow \infty} (b^T x^{k+1} - c) = 0,$$

pela continuidade de h , segue que $h(x^*) = 0$. Além disso, $x^* \in \Omega$, pois Ω é fechado e $\{x^k\} \subset \Omega$. Concluimos que x^* é um ponto viável. \square

Seguem alguns resultados auxiliares que utilizamos para demonstrar a otimalidade do ponto limite da sequência $\{x^k\}_{k \in \mathbb{N}}$ gerada pelo Algoritmo 3, aplicado ao caso quadrático.

Considere a função $f(x) = \frac{1}{2}x^T P x - a^T x$ quadrática, tal como em (2.10) e defina o conjunto $\bar{\Omega} := \{x \in \mathbb{R}^n \mid \ell \leq x \leq u, h(x) = 0\}$. O problema principal pode então ser escrito como sendo

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && x \in \bar{\Omega}. \end{aligned} \tag{2.44}$$

Assumimos no Algoritmo 3 que a norma utilizada é a induzida pela matriz simétrica definida positiva P .

Fornecemos a seguir a definição que será útil para estabelecer os resultados de convergência.

Definição 2.2.1. *O operador projeção denotado por Π_Z^P de modo que $\text{proj}_Z = \Pi_Z^P$ sobre o conjunto Z usa a norma $\|\cdot\|_P$ sendo P uma matriz simétrica definida positiva.*

Lema 2.2.1. *Seja $\hat{x} = \Pi_{[\ell, u]}^P(x)$ e $h(\hat{x}) = 0$. Então*

$$\hat{x} = \Pi_{\bar{\Omega}}^P(x).$$

Demonstração. Segue da definição que

$$\|\hat{x} - x\|_P \leq \|y - x\|_P, \quad \forall y \in [\ell, u],$$

em particular, como $\hat{x} \in \bar{\Omega}$, temos

$$\|\hat{x} - x\|_P \leq \|y - x\|_P \quad \forall y \in \bar{\Omega}.$$

\square

Lema 2.2.2. *Seja $z \in \bar{\Omega}$. Então z é uma solução para o problema (2.44) se, e somente se, $z = \Pi_{\bar{\Omega}}^P(P^{-1}a)$.*

Demonstração. Considere $z, y \in \bar{\Omega}$, temos que

$$\begin{aligned}
 f(z) - f(y) &= \left[\frac{1}{2} \langle z, Pz \rangle - \langle a, z \rangle \right] - \left[\frac{1}{2} \langle y, Py \rangle - \langle a, y \rangle \right] \\
 &= \left[\frac{1}{2} \langle z, z \rangle_P - \langle P^{-1}a, z \rangle_P \right] - \left[\frac{1}{2} \langle y, y \rangle_P - \langle P^{-1}a, y \rangle_P \right] \\
 &= \left[\frac{1}{2} \|z\|_P^2 + \frac{1}{2} \|P^{-1}a\|_P^2 - \langle P^{-1}a, z \rangle_P \right] - \left[\frac{1}{2} \|y\|_P^2 + \frac{1}{2} \|P^{-1}a\|_P^2 - \langle P^{-1}a, y \rangle_P \right] \\
 &= \frac{1}{2} \|z - P^{-1}a\|_P^2 - \frac{1}{2} \|y - P^{-1}a\|_P^2.
 \end{aligned}$$

Concluimos que $f(z) \leq f(y)$ se, e somente se, $\|z - P^{-1}a\|_P \leq \|y - P^{-1}a\|_P$. \square

Ainda no contexto do problema da mochila quadrático, considere as sequências $\{\bar{x}^k\}_{k \in \mathbb{N}}$ e $\{x^k\}_{k \in \mathbb{N}}$ geradas pelo Algoritmo 3, então pelo Teorema 2.2.1 existem \bar{x} e x^* tais que $\bar{x} = \lim_{k \rightarrow \infty} \bar{x}^k$ e $x^* = \lim_{k \rightarrow \infty} x^k$. Ao considerar o Algoritmo 3 com a norma induzida pela matriz P segue que $x^{k+1} = \Pi_{\bar{\Omega}}^P(\bar{x}^k)$, **Passo 2.** Como o operador projeção é contínuo, temos $x^* = \Pi_{\bar{\Omega}}^P(\bar{x})$. Logo, considerando a sequência $\{r_k\}_{k \in \mathbb{N}}$ limitada, usando o Teorema 2.2.2 mostramos que $h(x^*) = 0$, pelo Lema 2.2.1 segue que $x^* = \Pi_{\bar{\Omega}}^P(\bar{x})$.

Teorema 2.2.3. $x^* = \Pi_{\bar{\Omega}}^P(\bar{x})$ é uma solução do problema (2.44).

Demonstração. Vamos mostrar primeiramente que $x^* = \Pi_{\bar{\Omega}}^P(P^{-1}a)$, e concluir a prova por meio do Lema 2.2.2. Para isso, recordemos que $\nabla_x L(\bar{x}^k, \lambda_k, r_k) = 0$ para todo $k \in \mathbb{N}$, ou seja,

$$P\bar{x}^k - a + \lambda_k b + r_k(b^T \bar{x}^k - c)b = 0 \Leftrightarrow P\bar{x}^k - a = -[\lambda_k + r_k(b^T \bar{x}^k - c)]b.$$

Como $\{\|P\bar{x}^k - a\|\}$ é limitada, temos que $\{[\lambda_k + r_k(b^T \bar{x}^k - c)]\|b\|\}$ também é limitada. Consideremos $\mu_k := \lambda_k + r_k(b^T \bar{x}^k - c)$, para cada k . Portanto, temos que (sem perda de generalidade) existe $\mu \in \mathbb{R}$ tal que

$$0 = \lim_{k \rightarrow \infty} (P\bar{x}^k - a + \mu_k b) = P\bar{x} - a + \mu b,$$

o que implica

$$\bar{x} = P^{-1}(a - \mu b).$$

Pela caracterização do operador projeção, temos que

$$\langle \bar{x} - x^*, P(y - x^*) \rangle = \langle \bar{x} - x^*, y - x^* \rangle_P \leq 0, \quad \forall y \in \bar{\Omega}. \quad (2.45)$$

Assim,

$$\begin{aligned}
\langle \bar{x} - x^*, P(y - x^*) \rangle &= \langle P^{-1}(a - \mu b) - x^*, P(y - x^*) \rangle \\
&= \langle P^{-1}a - P^{-1}\mu b - x^*, P(y - x^*) \rangle \\
&= \langle P^{-1}a - x^*, P(y - x^*) \rangle - \langle P^{-1}\mu b, P(y - x^*) \rangle \\
&= \langle P^{-1}a - x^*, P(y - x^*) \rangle - \langle \mu b, y - x^* \rangle \\
&= \langle P^{-1}a - x^*, P(y - x^*) \rangle - \mu[\langle b, y \rangle - \langle b, x^* \rangle] \\
&= \langle P^{-1}a - x^*, P(y - x^*) \rangle - \mu(c - c) \\
&= \langle P^{-1}a - x^*, P(y - x^*) \rangle \\
&= \langle P^{-1}a - x^*, y - x^* \rangle_P.
\end{aligned}$$

Segue de (2.45) e da igualdade anterior que

$$\langle P^{-1}a - x^*, y - x^* \rangle_P \leq 0, \quad \forall y \in \bar{\Omega}.$$

Portanto $x^* = \Pi_{\bar{\Omega}}^P(P^{-1}a)$. □

Observação 2.2.2. *Na literatura, a eficácia da abordagem de Lagrangiano aumentado está associada à limitação da sequência dos parâmetros de penalidade. Hipóteses para assegurar essa limitação são usuais na literatura veja, por exemplo, em Birgin et al. (2012) e em Birgin e Martínez (2014). De fato, a introdução do parâmetro auxiliar k_0 no Algoritmo 3 (**Passo 4**) foi uma medida adotada para manter o parâmetro de penalidade fixo nas primeiras iterações. Tal abordagem garantiu bons resultados numéricos, conforme será discutido no Capítulo 4.*

Note que no caso quadrático estabelecemos a convergência da sequência gerada pelo Algoritmo de Lagrangiano aumentado (Algoritmo 3) independente da Hipótese 2.1.1 ser satisfeita. Diferentemente do caso geral, em que para desenvolver a teoria de convergência tal hipótese foi crucial.

2.3 INSTÂNCIAS DO PROBLEMA DA MOCHILA QUADRÁTICO

Nesta seção, vamos aplicar quatro instâncias do Algoritmo 3 para diferentes problemas particulares da mochila quadrático. Lembrando que estabelecemos uma fórmula fechada em (2.18) para a solução do subproblema (2.4), dada por

$$\bar{x}^k = y + [(r_k c - \lambda_k) - \alpha_k (b^T y + (r_k c - \lambda_k) b^T z)] z,$$

onde $\alpha_k = \frac{r_k}{1 + r_k b^T z}$, $Py = a$ e $Pz = b$. Usamos tal informação para formular algoritmos especializados, simplificando o **Passo 1** do Algoritmo 3. Além disso, consideramos, também, a análise de casos com restrições de desigualdade.

Primeira Instância

Um caso particular do problema (2.1), muito usual na literatura do problema da mochila, é o caso quadrático diagonal, no qual a matriz P definida tal como em (2.10) é diagonal, ou seja, $P = \text{diag}(p_1, p_2, \dots, p_n)$. Denotando,

$$\hat{z} = \left(\frac{b_1}{p_1}, \dots, \frac{b_n}{p_n} \right)^T, \quad (2.46)$$

obtemos que, cada componente do vetor \bar{x}^k é dada por

$$\bar{x}_i^k = \frac{a_i}{p_i} + \left[(r_k c - \lambda_k) - \frac{r_k}{1 + r_k b^T \hat{z}} \sum_{j=1}^n \left(\frac{b_j a_j}{p_j} + (r_k c - \lambda_k) \frac{b_j^2}{p_j} \right) \right] \frac{b_i}{p_i}. \quad (2.47)$$

Assim, o algoritmo para este caso especial pode ser escrito da seguinte forma:

Algoritmo 4: LAGRANGIANO AUMENTADO - PRIMEIRA INSTÂNCIA

Passo 0: Considere $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}$, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$, $\gamma > 1$, $r_0 > 0$, $\theta \in (0, 1)$, $k_0 \in \mathbb{N}$ e $\varepsilon \geq 0$. Obtenha \hat{z} por (2.46) e $\frac{a_i}{p_i}$, para cada $i = 1, \dots, n$. Faça $k = 0$.

Passo 1: Determine

$$\bar{x}^k \text{ por (2.47).}$$

Os **Passos 2, 3, 4 e 5** seguem como no Algoritmo 3.

Note que o Algoritmo 4 executa apenas operações simples tais como adição, multiplicação e divisão. Portanto é bem simples de implementar, rápido e pode resolver problemas com altas dimensões.

Segunda Instância

Nesta instância, consideramos o problema (2.1) quadrático como na primeira instância, mas agora com restrição de desigualdade. Ou seja,

$$\begin{aligned} \text{minimizar} \quad & f(x) := \frac{1}{2} x^T P x - a^T x \\ \text{sujeito a} \quad & b^T x \leq c \\ & \ell \leq x \leq u. \end{aligned} \quad (2.48)$$

O problema (2.48) pode ser tratado como o problema (2.1) apenas adicionando uma variável de folga na restrição de desigualdade, que passa a ser $b^T x + s = c$ com $s \in \mathbb{R}_+$. Dessa forma, o problema pode ser reescrito como

$$\begin{aligned} \text{minimizar} \quad & f(x) := \frac{1}{2} x^T P x - a^T x \\ \text{sujeito a} \quad & b^T x + s = c \\ & \ell \leq x \leq u \\ & 0 \leq s. \end{aligned} \quad (2.49)$$

Assim, o problema (2.3) é reformulado como sendo

$$\begin{aligned} & \text{minimizar} && L(x, s, \lambda, r) \\ & \text{sujeito a} && (x, s) \in \Omega_1, \end{aligned}$$

em que $L(x, s, \lambda, r) = f(x) + \lambda(b^T x + s - c) + \frac{r}{2}(b^T x + s - c)^2$ é a função Lagrangiano aumentado que penaliza a restrição de igualdade em (2.49), e definimos o conjunto $\Omega_1 := \{(x, s) \in \mathbb{R}^n \times \mathbb{R} \mid \ell \leq x \leq u, 0 \leq s\}$.

O processo é similar ao realizado para o problema (2.3). Fixados λ_k e r_k , determinamos (\bar{x}^k, \bar{s}_k) como uma solução de $\nabla_{(x,s)} L(x, s, \lambda_k, r_k) = 0$ e então projetamos na caixa Ω_1 . Assim, derivando a função Lagrangiano aumentado com relação a x e s e igualando a zero, obtemos

$$\begin{aligned} \nabla f(x) + [\lambda_k + r_k(b^T x - c + s)]b &= 0, \\ \lambda_k + r_k(b^T x - c + s) &= 0. \end{aligned} \tag{2.50}$$

Logo para o caso quadrático onde a matriz $P = \text{diag}(p_1, p_2, \dots, p_n)$, a solução para \bar{x}^k coincide com (2.47) se $\bar{s}_k = 0$ e, por conta de (2.50), \bar{x}^k será apenas o vetor $P^{-1}a$ no caso de $\bar{s}_k \neq 0$, assim, cada componente do vetor \bar{x}^k é dada por

$$\bar{x}_i^k = \begin{cases} \frac{a_i}{p_i} + \left[(r_k c - \lambda_k) - \alpha_k \sum_{j=1}^n \left(\frac{b_j a_j}{p_j} + (r_k c - \lambda_k) \frac{b_j^2}{p_j} \right) \right] \frac{b_i}{p_i}, & \text{se } \bar{s}_k = 0 \\ \frac{a_i}{p_i}, & \text{caso contrário,} \end{cases} \tag{2.51}$$

em que $\alpha_k = \frac{r_k}{1 + r_k b^T \hat{z}}$, com o vetor \hat{z} por (2.46). Aqui consideramos o termo $\bar{s}_k = -\frac{\lambda_k}{r_k} - b^T \bar{x}^k + c$.

Então, o algoritmo de Lagrangiano aumentado se torna:

Algoritmo 5: LAGRANGIANO AUMENTADO - SEGUNDA INSTÂNCIA

Passo 0: Considere $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}$, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$, $\gamma > 1$, $r_0 > 0$, $\theta \in (0, 1)$, $k_0 \in \mathbb{N}$ e $\varepsilon \geq 0$. Obtenha \hat{z} por (2.46) e $\frac{a_i}{p_i}$, para cada $i = 1, \dots, n$. Faça $k = 0$.

Passo 1: Determine

$$(\bar{x}^k, \bar{s}_k) \text{ por (2.51) .}$$

Passo 2: Calcule

$$(x^{k+1}, s_{k+1}) = \text{proj}_{\Omega_1}(\bar{x}^k, \bar{s}_k), \text{ usando a mesma filosofia de (2.5).}$$

Passo 3: Atualize

$$\lambda_{k+1} = \text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda_k + r_k(b^T x^{k+1} - c + s_{k+1})).$$

Passo 4: Atualize o parâmetro de penalidade $r_{k+1} \geq r_k$, usando a mesma filosofia do **Passo 4** do Algoritmo 3, com $\bar{h}(x, s) = b^T x - c + s$.

O **Passo 5** segue como no Algoritmo 3.

Terceira Instância

Nessa instância, consideramos o problema (2.1) quadrático com a matriz P simétrica, definida positiva mas não necessariamente diagonal. Como discutido na Observação 2.2.1, podemos usar a Decomposição de Cholesky para resolver os sistemas $Py = a$, $Pz = b$, dados na equação (2.17). Assim, o algoritmo tem a forma:

Algoritmo 6: LAGRANGIANO AUMENTADO - TERCEIRA INSTÂNCIA

Passo 0: Considere $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}$, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$, $\gamma > 1$, $r_0 > 0$, $\theta \in (0, 1)$, $k_0 \in \mathbb{N}$ e $\varepsilon \geq 0$. Resolva $Py = a$, $Pz = b$. Faça $k = 0$.

Passo 1: Determine $\alpha_k = \frac{r_k}{1 + r_k b^T z}$ e

$$\bar{x}^k = y + [(r_k c - \lambda_k) - \alpha_k(b^T y + (r_k c - \lambda_k)b^T z)]z.$$

Os **Passos 2, 3, 4 e 5** seguem como no Algoritmo 3.

Quarta Instância

Para esta instância consideramos o problema com restrição de desigualdade, semelhante a (2.49), sendo a matriz P simétrica, definida positiva mas não necessariamente diagonal. Neste caso, resolvemos (2.50) de modo similar à segunda instância, exceto que \bar{x}^k é determinado como segue. Se $s_k = 0$, então \bar{x}^k é determinado por (2.18), caso contrário, \bar{x}^k é a solução de $\nabla f(x) = 0$, mais especificamente, $\bar{x}^k = P^{-1}a$.

Formalmente,

$$\bar{x}^k = \begin{cases} y + [(r_k c - \lambda_k) - \alpha_k (b^T y + (r_k c - \lambda_k) b^T z)] z, & \text{se } s_k = 0 \\ P^{-1} a, & \text{caso contrário,} \end{cases} \quad (2.52)$$

em que $\alpha_k = \frac{r_k}{1 + r_k b^T z}$, $y = P^{-1} a$, $z = P^{-1} b$ e \bar{s}_k é tal como na segunda instância.

Algoritmo 7: LAGRANGIANO AUMENTADO - QUARTA INSTÂNCIA

Passo 0: Considere $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}$, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$, $\gamma > 1$, $r_0 > 0$,

$\theta \in (0, 1)$, $k_0 \in \mathbb{N}$ e $\varepsilon \geq 0$. Resolva $Py = a$, $Pz = b$. Faça $k = 0$.

Passo 1: Determine

$$(\bar{x}^k, \bar{s}_k) \text{ por (2.52).}$$

Os **Passos 2, 3, 4 e 5** seguem como no Algoritmo 5.

Apresentamos dessa forma extensões e especializações do Algoritmo de Lagrangiano aumentado para resolver o problema da mochila quadrático. Finalizamos aqui a discussão teórica do problema da mochila.

3 PROBLEMA DE ALOCAÇÃO DE RECURSOS

Até o momento consideramos problemas com uma restrição de igualdade linear, atribuindo uma atenção especial ao caso onde a função objetivo é quadrática. Assim, uma extensão natural do que foi trabalhado é realizar uma análise do caso onde a função objetivo é quadrática e possui mais de uma restrição de igualdade linear. Neste capítulo, apresentamos uma proposta para resolver problemas desse tipo, utilizando uma metodologia de Lagrangiano aumentado. Além disso, examinamos uma classe de problemas mais geral, onde consideramos a função objetivo convexa e as restrições de igualdade sendo não lineares e não necessariamente convexas.

Na Seção 3.1, apresentamos o problema quadrático com m restrições lineares, estruturamos o algoritmo proposto para resolvê-lo e estabelecemos resultados de convergência. Na Seção 3.2, apresentamos uma abordagem para resolver um problema mais geral em que a função objetivo é convexa e as restrições não são necessariamente lineares nem convexas.

3.1 PROBLEMA QUADRÁTICO COM RESTRIÇÕES LINEARES

Nesta seção, consideramos o problema de alocação de recursos em que a função objetivo é quadrática com restrições de igualdade lineares e caixa, como formalizado a seguir

$$\begin{aligned} \text{minimizar} \quad & f(x) = \frac{1}{2}x^T Px - a^T x \\ \text{sujeito a} \quad & h(x) = Ax - b = 0 \\ & \ell \leq x \leq u, \end{aligned} \tag{3.1}$$

em que $P \in \mathbb{R}^{n \times n}$ é simétrica e definida positiva, $a \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ e $\ell, u \in \mathbb{R}^n$ com $\ell < u$. Logo, a função Lagrangiano aumentado que penaliza as restrições de igualdade é definida como

$$L(x, \lambda, r) = \frac{1}{2}x^T Px - a^T x + \lambda^T (Ax - b) + \frac{r}{2} \|Ax - b\|^2, \tag{3.2}$$

em que $\lambda \in \mathbb{R}^m$ é o vetor de multiplicadores de Lagrange e $r > 0$ é o parâmetro de penalidade.

Começamos por analisar o gradiente e a Hessiana da função Lagrangiano aumentado. Considerando λ e r fixados e derivando a função Lagrangiano aumentado, L , em relação a x , obtemos o gradiente e a Hessiana, dados respectivamente por

$$\nabla_x L(x, \lambda, r) = (P + rA^T A)x - a + A^T(\lambda - rb), \tag{3.3}$$

e

$$\nabla_{xx}^2 L(x, \lambda, r) = P + rA^T A. \quad (3.4)$$

Como P é uma matriz definida positiva, e $A^T A$ é semidefinida positiva, segue que a Hessiana da função Lagrangiano aumentado é definida positiva e portanto inversível. Determinamos a inversa da matriz em questão por meio da Identidade de Woodbury (WOODBURY, 1950). Assumindo que existam as inversas de $W + UCV$, W e C , obtemos

$$(W + UCV)^{-1} = W^{-1} - W^{-1}U(C^{-1} + VW^{-1}U)^{-1}VW^{-1},$$

em que $W \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{m \times m}$, $U \in \mathbb{R}^{n \times m}$ e $V \in \mathbb{R}^{m \times n}$.

Desse modo, aplicando ao nosso problema, identificando $W = P$, $C = I$ (matriz identidade $m \times m$), $U = rA^T$ e $V = A$, a inversa da Hessiana da função Lagrangiano aumentado estabelecida em (3.4) tem a forma

$$(P + rA^T A)^{-1} = P^{-1} - rP^{-1}A^T(I + rAP^{-1}A^T)^{-1}AP^{-1}. \quad (3.5)$$

De modo geral, a dimensão das variáveis nos problemas considerados supera a quantidade de restrições. Portanto, ao aplicar a Identidade de Woodbury contornamos o problema de calcular a inversa da matriz $P + rA^T A$ de dimensão $n \times n$, para calcular a inversa da matriz $I + rAP^{-1}A$ de dimensão $m \times m$ a qual é inferior a $n \times n$ para o caso em que $m \leq n$.

Como observado anteriormente, a Hessiana da função Lagrangiano aumentado é definida positiva, portanto a função Lagrangiano aumentado definida em (3.2) é convexa. Podemos, então, aplicar uma abordagem semelhante à aplicada no Capítulo 2. Lembrando que a filosofia do método consistia em primeiramente resolver o problema de minimização irrestrito $\min_{x \in \mathbb{R}^n} L(x, \lambda, r)$, utilizando fortemente o fato da função Lagrangiano aumentado ser convexa, e em seguida projetar a solução encontrada na caixa $\Omega = \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$.

Utilizando as condições de otimalidade de primeira ordem e a igualdade estabelecida em (3.3), segue que

$$(P + rA^T A)x = a - A^T(\lambda - rb). \quad (3.6)$$

Usando a fórmula de Woodbury (3.5) para calcular a inversa da matriz $P + rA^T A$, obtemos

$$x = [P^{-1} - rP^{-1}A^T(I + rAP^{-1}A^T)^{-1}AP^{-1}](a - A^T(\lambda - rb)),$$

ou seja,

$$x = P^{-1}a - P^{-1}A^T(\lambda - rb) - rP^{-1}A^T(I + rAP^{-1}A^T)^{-1}AP^{-1}a \\ + rP^{-1}A^T(I + rAP^{-1}A^T)^{-1}AP^{-1}A^T(\lambda - rb),$$

onde podemos denotar

$$P^{-1}a = y \quad \text{e} \quad P^{-1}A^T = N \quad \text{e} \quad AP^{-1}A^T = M.$$

Dessa maneira, a solução do problema de determinar \bar{x}^k de modo que seja satisfeita $\nabla_x L(\bar{x}^k, \lambda^k, r_k) = 0$, na iteração k , pode ser escrita por meio da seguinte fórmula

$$\bar{x}^k = y + (r_k N(I + r_k M)^{-1} M - N)\lambda^k + r_k(N - r_k N(I + r_k M)^{-1} M)b \\ - r_k N(I + r_k M)^{-1} Ay \\ = y + (r_k N(I + r_k M)^{-1} M - N)(\lambda^k - r_k b) - r_k N(I + r_k M)^{-1} Ay.$$

Assim, estabelecemos o algoritmo de Lagrangiano aumentado aplicado ao caso quadrático com m restrições lineares.

Algoritmo 8: LAGRANGIANO AUMENTADO - QUADRÁTICO RESTRIÇÕES LINEARES

Passo 0: Considere $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}^m$, $\lambda^0 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\gamma > 1$, $r_0 > 0$, $\theta \in (0, 1)$, $k_0 \in \mathbb{N}$ e $\varepsilon \geq 0$. Resolva $Py = a$, $P^{-1}A^T = N$, $AP^{-1}A^T = M$. Faça $k = 0$.

Passo 1: Determine

$$\bar{x}^k = y + (r_k N(I + r_k M)^{-1} M - N)(\lambda^k - r_k b) - r_k N(I + r_k M)^{-1} Ay.$$

Passo 2: Calcule

$$x^{k+1} = \text{proj}_{\Omega}(\bar{x}^k).$$

Passo 3: Estime uma nova aproximação do multiplicador de Lagrange

$$\lambda^{k+1} = \text{proj}_{[\lambda_{\min}, \lambda_{\max}]^m}(\lambda^k + r_k(Ax^{k+1} - b)).$$

Passo 4: Atualize o parâmetro de penalidade

$$r_{k+1} = \begin{cases} r_k, & \text{se } k \leq k_0 \text{ ou } \|Ax^{k+1} - b\| < \theta \|Ax^k - b\| \\ \gamma r_k, & \text{caso contrário.} \end{cases}$$

Passo 5: Teste a convergência

Se $\|Ax^{k+1} - b\| < \varepsilon$ e $\|x^{k+1} - x^k\| < \varepsilon$, então pare. Caso contrário, faça $k = k + 1$ e retorne ao **Passo 1**.

Note que devido à estrutura do problema, conseguimos simplificar a resolução do subproblema irrestrito de $\min_{x \in \mathbb{R}^n} L(x, \lambda, r)$, presente no **Passo 1**. Detalhamos a escolha dos parâmetros no Capítulo 4, onde abordamos os experimentos numéricos da Tese.

3.1.1 Análise de convergência

Nesta subseção discutimos a convergência das sequências geradas pelo Algoritmo 8, para o caso particular onde consideramos a sequência dos parâmetros de penalidade limitada. De forma semelhante ao que fizemos no Capítulo 2, para o caso quadrático com uma restrição de igualdade linear e variáveis canalizadas, apresentamos as demonstrações para este caso mais complexo.

Proposição 3.1.1. *Considere o problema quadrático (3.1). Suponha que a sequência de parâmetros de penalidade no Algoritmo 8 é limitada. Então as sequências $\{\bar{x}^k\}_{k \in \mathbb{N}}$ e $\{x^k\}_{k \in \mathbb{N}}$, geradas pelo Algoritmo 8, são convergentes.*

Demonstração. Como por hipótese a sequência de parâmetros de penalidade $\{r_k\}_{k \in \mathbb{N}}$ é limitada, vamos considerar, por simplicidade, que $r_k = r$ para todo $k \in \mathbb{N}$.

Segue da igualdade em (3.6),

$$\begin{aligned}\bar{x}^k &= (P + rA^T A)^{-1}[a - A^T(\lambda^k - rb)] \\ &= (P + rA^T A)^{-1}a - (P + rA^T A)^{-1}A^T(\lambda^k - rb),\end{aligned}$$

analogamente,

$$\begin{aligned}\bar{x}^{k-1} &= (P + rA^T A)^{-1}[a - A^T(\lambda^{k-1} - rb)] \\ &= (P + rA^T A)^{-1}a - (P + rA^T A)^{-1}A^T(\lambda^{k-1} - rb).\end{aligned}$$

Logo a diferença entre as iteradas consecutivas tem a forma

$$\bar{x}^k - \bar{x}^{k-1} = (P + rA^T A)^{-1}A^T(\lambda^{k-1} - \lambda^k).$$

Assim,

$$\begin{aligned}\|\bar{x}^k - \bar{x}^{k-1}\| &= \|(P + rA^T A)^{-1}A^T(\lambda^{k-1} - \lambda^k)\| \\ &\leq \|(P + rA^T A)^{-1}A^T\| \|\lambda^k - \lambda^{k-1}\|.\end{aligned}$$

Utilizando o fato que o operador projeção é não expansivo, obtemos

$$\begin{aligned}\|x^{k+1} - x^k\| &= \|\text{proj}_\Omega(\bar{x}^k) - \text{proj}_\Omega(\bar{x}^{k-1})\| \\ &\leq \|\bar{x}^k - \bar{x}^{k-1}\| \\ &= \|(P + rA^T A)^{-1}A^T(\lambda^{k-1} - \lambda^k)\| \\ &\leq \|(P + rA^T A)^{-1}A^T\| \|\lambda^k - \lambda^{k-1}\|.\end{aligned}\tag{3.7}$$

Como $r_k = r$ para todo $k \in \mathbb{N}$, devido à regra de atualização do parâmetro de penalidade (**Passo 4**), vale a desigualdade $\|Ax^{k+1} - b\| < \theta \|Ax^k - b\|$ para todo $k \in \mathbb{N}$ com $k \geq k_0 + 1$, sendo k_0 o mesmo estabelecido no algoritmo. Logo, temos as seguintes desigualdades

$$0 \leq \|Ax^{k+1} - b\| < \theta \|Ax^k - b\| < \theta^2 \|Ax^{k-1} - b\| < \dots < \theta^{k-k_0} \|Ax^{k_0+1} - b\|. \quad (3.8)$$

Consideremos $\bar{\lambda}^{k+1} := \lambda^k + r(Ax^{k+1} - b)$, para cada $k \in \mathbb{N}$, de forma equivalente podemos escrever

$$\frac{\bar{\lambda}^{k+1} - \lambda^k}{r} = Ax^{k+1} - b, \quad (3.9)$$

uma vez que o parâmetro r é não nulo.

Combinando (3.8) e (3.9), obtemos

$$\frac{\|\bar{\lambda}^{k+1} - \lambda^k\|}{r} < \theta \frac{\|\bar{\lambda}^k - \lambda^{k-1}\|}{r} < \dots < \theta^{k-k_0} \frac{\|\bar{\lambda}^{k_0+1} - \lambda^{k_0}\|}{r}. \quad (3.10)$$

Lembrando que $\lambda^{k+1} = \text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda^k + r(Ax^{k+1} - b))$, e utilizando a não expansividade da projeção, segue que

$$\begin{aligned} \|\lambda^{k+1} - \lambda^k\| &= \|\text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda^k + r(Ax^{k+1} - b)) - \text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\lambda^k)\| \\ &\leq \|\lambda^k + r(Ax^{k+1} - b) - \lambda^k\| \\ &= r \|Ax^{k+1} - b\| \\ &= \|\bar{\lambda}^{k+1} - \lambda^k\|. \end{aligned} \quad (3.11)$$

Segue de (3.10) e (3.11) que

$$\|\lambda^{k+1} - \lambda^k\| \leq \|\bar{\lambda}^{k+1} - \lambda^k\| < \theta^{k-k_0} \|\bar{\lambda}^{k_0+1} - \lambda^{k_0}\|. \quad (3.12)$$

Como $\theta \in (0, 1)$, concluímos que a sequência $\{\lambda^{k+1} - \lambda^k\}_{k \in \mathbb{N}}$ converge para zero, quando $k \rightarrow \infty$. Portanto, tomando-se o limite em (3.7),

$$\|x^{k+1} - x^k\| \rightarrow 0.$$

Considere $p \in \mathbb{N} \setminus \{0\}$, arbitrário, combinando a desigualdade triangular com as desigualdades estabelecidas em (3.7) e (3.12)

$$\begin{aligned} \|x^{k+p+1} - x^{k+1}\| &\leq \|\bar{x}^{k+p} - \bar{x}^k\| \\ &= \|\bar{x}^{k+p} - \bar{x}^{k+p-1} + \bar{x}^{k+p-1} - \dots + \bar{x}^{k+1} - \bar{x}^k\| \\ &\leq \|\bar{x}^{k+p} - \bar{x}^{k+p-1}\| + \dots + \|\bar{x}^{k+1} - \bar{x}^k\| \\ &< (\theta^{k+p-k_0-1} + \dots + \theta^{k-k_0}) \|(P + rA^T A)^{-1} A^T\| \|\bar{\lambda}^{k_0+1} - \lambda^{k_0}\|. \end{aligned} \quad (3.13)$$

Como $\theta \in (0, 1)$, a série geométrica $\sum_{j \in \mathbb{N}} \theta^j$ é convergente, então a soma $\sum_{j=k}^{k+p-k_0-1} \theta^{j-k_0}$ converge para zero, quando $k \rightarrow \infty$. Isso implica que $\{x^k\}_{k \in \mathbb{N}}$ e $\{\bar{x}^k\}_{k \in \mathbb{N}}$ são sequências de Cauchy, portanto são convergentes. \square

Corolário 3.1.1. *Considere o problema quadrático (3.1). No caso em que a sequência dos parâmetros de penalidade é limitada, o limite x^* da sequência $\{x^k\}_{k \in \mathbb{N}}$, definida pelo Algoritmo 8, é um ponto viável.*

Demonstração. Como $\{r_k\}_{k \in \mathbb{N}}$ é limitada, podemos considerar $r_k = r$ para todo $k \in \mathbb{N}$. Pela regra de atualização do parâmetro de penalidade estabelecida no **Passo 4** do Algoritmo 8, vale a desigualdade $\|Ax^{k+1} - b\| < \theta \|Ax^k - b\|$ para todo $k \geq k_0 + 1$, sendo $k_0 \in \mathbb{N}$ definido no Algoritmo 8. Assim,

$$0 \leq \|Ax^{k+1} - b\| < \theta \|Ax^k - b\| < \theta^2 \|Ax^{k-1} - b\| < \dots < \theta^{k-k_0} \|Ax^{k_0+1} - b\|. \quad (3.14)$$

Consideremos $\bar{\lambda}^{k+1} := \lambda^k + r(Ax^{k+1} - b)$, para cada $k \in \mathbb{N}$. Como r é não nulo, podemos escrever

$$\frac{\bar{\lambda}^{k+1} - \lambda^k}{r} = Ax^{k+1} - b, \text{ para todo } k \in \mathbb{N}, \quad (3.15)$$

Combinando (3.14) e (3.15), obtemos

$$\frac{|\bar{\lambda}^{k+1} - \lambda^k|}{r} < \theta \frac{|\bar{\lambda}^k - \lambda^{k-1}|}{r} < \dots < \theta^{k-k_0} \frac{|\bar{\lambda}^{k_0+1} - \lambda^{k_0}|}{r}. \quad (3.16)$$

Note que $\theta \in (0, 1)$, logo

$$\lim_{k \rightarrow \infty} \left(\frac{\bar{\lambda}^{k+1} - \lambda^k}{r} \right) = 0.$$

Como a igualdade (3.15) é válida para cada $k \in \mathbb{N}$, temos

$$\lim_{k \rightarrow \infty} (Ax^{k+1} - b) = 0,$$

segue que $Ax^* - b = 0$. Como Ω é compacto e a sequência $\{x^k\}_{k \in \mathbb{N}} \subset \Omega$, logo $x^* \in \Omega$. Portanto, o ponto limite é viável para o problema. \square

Concluimos portanto que a sequência gerada pelo Algoritmo 8 converge para um ponto de acumulação, sendo este ponto viável para o problema.

3.2 PROBLEMA GERAL COM m RESTRIÇÕES

Nesta seção, consideramos uma generalização do problema abordado no Capítulo 2, agora trabalhando com mais de uma restrição, e não exigindo linearidade das mesmas. Segue a formulação geral do problema

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && h_i(x) = 0, \quad \text{com } i = 1, \dots, m \\ &&& \ell \leq x \leq u, \end{aligned} \quad (3.17)$$

em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $h_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, m$, são funções de classe \mathcal{C}^2 . Assumiremos que a função objetivo f é convexa, mas as funções que representam as restrições h_i , com $i = 1, \dots, m$, não são necessariamente convexas.

Um caso particular e relevante do problema (3.17) é do Despacho Hidrotérmico, ver (MARCILIO, s.d.). Por ser de interesse mundial, tem atraído muitas pesquisas no desenvolvimento de novos algoritmos para resolvê-lo. Neste caso, a função objetivo é quadrática e as restrições de igualdade são não lineares e não convexas. Portanto, se trata de um problema difícil de resolver, por ser um problema de grande porte, principalmente, quando aplicado ao caso brasileiro que é predominante em geração hídrica.

3.2.1 Metodologia e Algoritmo

No problema (3.17), se as restrições de igualdade são funções não lineares, o cálculo de segunda derivada torna o problema ainda mais complexo. Devido a isso, usamos uma abordagem semelhante ao método de Gauss-Newton em mínimos quadrados, ou seja, desprezamos a Hessiana com relação a essas funções (h_i). Dessa forma, apesar de exigirmos continuidade das derivadas de segunda ordem, a ideia é considerar um algoritmo que utilize somente informações de primeira ordem das restrições de igualdade, conforme descrito a seguir.

Associado ao problema (3.17), determinamos a função Lagrangiano aumentado que penaliza apenas as restrições de igualdade, que tem a seguinte forma

$$L(x, \lambda, r) = f(x) + \lambda^T h(x) + \frac{r}{2} \|h(x)\|^2, \quad (3.18)$$

em que $\lambda \in \mathbb{R}^m$ é o vetor de multiplicadores de Lagrange associado às restrições de igualdade, e $r > 0$ representa o parâmetro de penalidade.

Usando a metodologia de Lagrangiano aumentado para resolver o problema (3.17), a cada passo do algoritmo, precisamos resolver um subproblema que possui somente restrições simples, $\ell \leq x \leq u$. Ou seja,

$$\begin{aligned} &\text{minimizar} && L(x, \lambda^k, r_k) \\ &\text{sujeito a} && \ell \leq x \leq u. \end{aligned} \quad (3.19)$$

Resolver o problema (3.19) é a tarefa mais complicada no algoritmo. Com o intuito de contornar isso, neste trabalho consideramos a análise de uma abordagem aproximada, por meio de um modelo quadrático.

Considere fixados λ e r , então o gradiente e a Hessiana da função Lagrangiano aumentado, definida em (3.18), em relação à variável x , são dados, respectivamente, por

$$\nabla_x L(x, \lambda, r) = \nabla f(x) + \nabla h(x)^T \lambda + r \nabla h(x)^T h(x), \quad (3.20)$$

e

$$\nabla_{xx}^2 L(x, \lambda, r) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + r \nabla h(x)^T \nabla h(x) + r \sum_{i=1}^m \nabla^2 h_i(x) h_i(x). \quad (3.21)$$

Observamos que os termos que envolvem informação de segunda ordem, nas funções que definem as restrições de igualdade do problema (3.17), em (3.21) são

$$\sum_{i=1}^m \lambda_i \nabla^2 h_i(x) \quad \text{e} \quad r \sum_{i=1}^m \nabla^2 h_i(x) h_i(x).$$

Desprezando estes termos e trabalhando com a Hessiana aproximada, a qual denotamos por $\nabla_{xx}^2 L^a(x, \lambda, r)$, temos,

$$\nabla_{xx}^2 L^a(x, \lambda, r) = \nabla^2 f(x) + r \nabla h(x)^T \nabla h(x). \quad (3.22)$$

Utilizando a aproximação no termo de segunda ordem definido acima e fixados λ^k e r_k , um modelo quadrático da função Lagrangiano aumentado (3.18), no iterando x^k , é estabelecido da seguinte forma

$$m_k(x) = \nabla_x L(x^k, \lambda^k, r_k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T \nabla_{xx}^2 L^a(x^k, \lambda^k, r_k) (x - x^k). \quad (3.23)$$

Considerando $d = x - x^k$ em (3.23), estabelecemos uma aproximação quadrática para o problema (3.19). Dessa forma, temos que resolver um subproblema quadrático a cada iteração do método. Note que na k -ésima iteração do algoritmo temos computados λ^k e r_k , segundo a regra de atualização estabelecida para estes parâmetros. Assim, para calcular x^{k+1} , resolvemos o subproblema quadrático, que tem a seguinte formulação

$$\begin{aligned} &\text{minimizar} && \nabla_x L(x^k, \lambda^k, r_k)^T d + \frac{1}{2} d^T \nabla_{xx}^2 L^a(x^k, \lambda^k, r_k) d \\ &\text{sujeito a} && \ell - x^k \leq d \leq u - x^k, \end{aligned} \quad (3.24)$$

sendo tal subproblema estabelecido a partir da aproximação quadrática em x^k , determinada em (3.23). O processo é repetido até que um candidato para uma solução do problema (3.17) seja determinado. Resumimos as etapas descritas anteriormente no seguinte algoritmo.

Algoritmo 9: LAGRANGIANO AUMENTADO - m RESTRIÇÕES

Passo 0: Considere $\lambda_{\min}, \lambda_{\max} \in \mathbb{R}^m$, $\lambda^0 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\gamma > 1$, $r_0 > 0$, $\theta \in (0, 1)$, $k_0 \in \mathbb{N}$ e $\varepsilon \geq 0$. Faça $k = 0$.

Passo 1: Determine d^k minimizador, possivelmente aproximado, do problema quadrático (3.24).

Passo 2: Faça $x^{k+1} = x^k + d^k$.

Passo 3: Calcule uma nova aproximação do multiplicador de Lagrange

$$\lambda^{k+1} = \text{proj}_{[\lambda_{\min}, \lambda_{\max}]^m}(\lambda^k + r_k h(x^{k+1})).$$

Passo 4: Atualize o parâmetro de penalidade

$$r_{k+1} = \begin{cases} r_k, & \text{se } k \leq k_0 \text{ ou } \|h(x^{k+1})\| < \theta \|h(x^k)\| \\ \gamma r_k, & \text{caso contrário.} \end{cases}$$

Passo 5: Teste a convergência

Se $\|h(x^{k+1})\| < \varepsilon$ e $\|x^{k+1} - x^k\| < \varepsilon$, então pare. Caso contrário, considere $k = k + 1$ e retorne ao **Passo 1**.

Consideramos o vetor de multiplicadores de Lagrange (aproximado) $\lambda^k \in \mathbb{R}^m$ limitado para cada $k \in \mathbb{N}$, pois necessitamos de salvaguardas que garantam sua limitação para estabelecer a teoria de convergência do Algoritmo 9, por conta disso, realizamos a projeção no **Passo 3**.

Nas primeiras k_0 iterações do Algoritmo 9, o parâmetro de penalidade r_k é considerado fixo, e a partir da iterada $k_0 + 1$ verificamos se houve um ganho suficiente na viabilidade e decidimos se mantemos, caso o ganho seja satisfatório, ou atualizamos o parâmetro, aumentando seu valor para a próxima iteração, caso não consigamos um progresso satisfatório. As regras de atualização do parâmetro de penalidade e do vetor de multiplicadores de Lagrange (aproximado) foram embasadas em (BIRGIN; MARTÍNEZ, 2014).

3.2.2 Análise de convergência do Algoritmo 9

Nesta subseção, abordamos a análise de convergência do Algoritmo 9, a qual tem por base a teoria de convergência do método geral de Lagrangiano aumentado desenvolvida por Birgin e Martínez (2014).

Com o intuito de simplificar a notação, nos referimos às restrições de caixa, no problema (3.17), por meio do conjunto $\Omega := \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$.

Observação 3.2.1. Segundo a regra de atualização de x^{k+1} , presente no **Passo 2** do Algoritmo 9, temos $x^{k+1} = x^k + d^k$, onde $\ell - x^k \leq d^k \leq u - x^k$, uma vez que d^k é solução

de (3.24), segue que a sequência $\{x^k\}_{k \in \mathbb{N}}$ gerada pelo Algoritmo 9 está inteiramente contida no conjunto Ω .

Para estabelecer resultados de convergência do Algoritmo 9, consideramos a hipótese a seguir (mesma Hipótese 2.1.1, presente no Capítulo 2). Lembrando que tal hipótese é embasada na Hipótese 5.1 estabelecida em (BIRGIN; MARTÍNEZ, 2014).

Hipótese 3.2.1. Para todo $k \in \mathbb{N}$, obtemos $x^k \in \Omega$ tal que

$$L(x^k, \lambda^k, r_k) \leq L(x, \lambda^k, r_k) + \varepsilon_k, \text{ para todo } x \in \Omega,$$

sendo a sequência de tolerâncias $\{\varepsilon_k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+$ limitada.

O teorema que segue é uma adaptação do Teorema 5.1 presente em (BIRGIN; MARTÍNEZ, 2014).

Teorema 3.2.1. Considere $\{x^k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 9, satisfazendo a Hipótese 3.2.1. Seja x^* um ponto limite de $\{x^k\}_{k \in \mathbb{N}}$. Então, para todo $x \in \Omega$, temos que

$$\|h(x^*)\|^2 \leq \|h(x)\|^2.$$

Demonstração. Inicialmente, note que o conjunto Ω é fechado e que a sequência $\{x^k\}_{k \in \mathbb{N}} \subset \Omega$, conseqüentemente, o ponto limite x^* pertence ao conjunto Ω . Vamos considerar a análise dos dois casos possíveis para a sequência dos parâmetros de penalidade: $\{r_k\}_{k \in \mathbb{N}}$ limitada e $r_k \rightarrow \infty$.

Suponha que a sequência $\{r_k\}_{k \in \mathbb{N}}$ é limitada. Por simplicidade, podemos considerar $r_k = r$, para todo $k \in \mathbb{N}$. Devido à regra de atualização do parâmetro r_k , no **Passo 4** do Algoritmo 9, para cada $k > k_0 + 1$ vale a seguinte desigualdade $\|h(x^{k+1})\| < \theta \|h(x^k)\|$, portanto segue que

$$0 \leq \|h(x^{k+1})\| < \theta \|h(x^k)\| < \dots < \theta^{k-k_0} \|h(x^{k_0+1})\|,$$

para todo $k \in \mathbb{N}$. Conseqüentemente, $\|h(x^k)\| \rightarrow 0$, quando $k \rightarrow \infty$ pois $\theta \in (0, 1)$. Logo, o ponto limite é viável.

Agora, suponha que $r_k \rightarrow \infty$. Seja $K \subseteq \mathbb{N}$ tal que

$$\lim_{k \in K} x^k = x^*.$$

Suponha, por contradição, que existe $\tilde{x} \in \Omega$ tal que

$$\|h(x^*)\|^2 > \|h(\tilde{x})\|^2.$$

Devido à continuidade de h , da limitação da sequência $\{\lambda^k\}_{k \in \mathbb{N}}$ e da sequência $\{r_k\}_{k \in \mathbb{N}}$ ser não decrescente com $r_k \rightarrow \infty$, existem $q > 0$ e $k_2 \in \mathbb{N}$ tais que para todo $k \in K$, $k \geq k_2$, temos satisfeita

$$\left\| h(x^k) + \frac{\lambda^k}{r_k} \right\|^2 > \left\| h(\tilde{x}) + \frac{\lambda^k}{r_k} \right\|^2 + q.$$

Então, para todo $k \in K$, $k \geq k_2$,

$$f(x^k) + \frac{r_k}{2} \left\| h(x^k) + \frac{\lambda^k}{r_k} \right\|^2 > f(\tilde{x}) + \frac{r_k}{2} \left\| h(\tilde{x}) + \frac{\lambda^k}{r_k} \right\|^2 + \frac{r_k q}{2} + f(x^k) - f(\tilde{x}). \quad (3.25)$$

Como $\lim_{k \in K} x^k = x^*$, a função f é contínua, a sequência $\{\varepsilon_k\}_{k \in \mathbb{N}}$ é limitada e $r_k \rightarrow \infty$, existe $k_3 \geq k_2$ tal que, para todo $k \in K$, $k \geq k_3$, vale a seguinte desigualdade

$$\frac{r_k q}{2} + f(x^k) - f(\tilde{x}) > \varepsilon_k. \quad (3.26)$$

Substituindo (3.26) em (3.25), obtemos

$$f(x^k) + \frac{r_k}{2} \left\| h(x^k) + \frac{\lambda^k}{r_k} \right\|^2 > f(\tilde{x}) + \frac{r_k}{2} \left\| h(\tilde{x}) + \frac{\lambda^k}{r_k} \right\|^2 + \varepsilon_k,$$

o que implica em,

$$f(x^k) + (\lambda^k)^T h(x^k) + \frac{r_k}{2} \|h(x^k)\|^2 > f(\tilde{x}) + (\lambda^k)^T h(\tilde{x}) + \frac{r_k}{2} \|h(\tilde{x})\|^2 + \varepsilon_k.$$

Ou seja,

$$L(x^k, \lambda^k, r_k) > L(\tilde{x}, \lambda^k, r_k) + \varepsilon_k,$$

para $k \in K$, $k \geq k_3$, o que contradiz a Hipótese 3.2.1, completando a demonstração. \square

Concluimos do teorema anterior que o Algoritmo 9, com as iterações definidas pela Hipótese 3.2.1, encontra minimizadores da inviabilidade. Portanto, se o problema (3.17) é viável, todo ponto limite de uma sequência gerada pelo algoritmo, sob a Hipótese 3.2.1, é viável. Note que usamos apenas a limitação da sequência $\{\varepsilon_k\}_{k \in \mathbb{N}}$ na prova do Teorema 3.2.1. No próximo teorema, adaptação do Teorema 5.2 em (BIRGIN; MARTÍNEZ, 2014), constatamos que, assumindo que a sequência $\{\varepsilon_k\}_{k \in \mathbb{N}}$ converge para zero, é possível mostrar que, no caso viável, o algoritmo encontra minimizadores globais do problema (3.17).

Teorema 3.2.2. *Considere $\{x^k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 9, satisfazendo a Hipótese 3.2.1 com $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Seja x^* um ponto limite de $\{x^k\}_{k \in \mathbb{N}}$. Suponha que o problema (3.1) é viável. Então x^* é um minimizador global de (3.1).*

Demonstração. Considere o subconjunto $K \subseteq \mathbb{N}$, de modo que

$$\lim_{k \in K} x^k = x^*.$$

Resulta do Teorema 3.2.1 que o ponto limite x^* é viável, uma vez que o problema (3.1) é viável. Considere \tilde{x} escolhido arbitrariamente em Ω de modo que $h(\tilde{x}) = 0$, ou seja, \tilde{x} um ponto viável para o problema (3.1). Uma vez que o algoritmo está sob a Hipótese 3.2.1, temos satisfeita a seguinte desigualdade

$$f(x^k) + (\lambda^k)^T h(x^k) + \frac{r_k}{2} \|h(x^k)\|^2 \leq f(\tilde{x}) + (\lambda^k)^T h(\tilde{x}) + \frac{r_k}{2} \|h(\tilde{x})\|^2 + \varepsilon_k,$$

para todo $k \in \mathbb{N}$. Como \tilde{x} foi tomado de forma que $h(\tilde{x}) = 0$, segue que

$$f(x^k) + (\lambda^k)^T h(x^k) + \frac{r_k}{2} \|h(x^k)\|^2 \leq f(\tilde{x}) + \varepsilon_k,$$

Usando o fato que $\left\{ \frac{r_k}{2} \|h(x^k)\|^2 \right\} \subset \mathbb{R}_+$, obtemos

$$f(x^k) + (\lambda^k)^T h(x^k) \leq f(\tilde{x}) + \varepsilon_k, \quad (3.27)$$

para todo $k \in \mathbb{N}$.

Valendo-se do fato que a sequência $\{\lambda^k\}_{k \in \mathbb{N}}$ é limitada, podemos considerar um subconjunto $K_1 \subseteq K$ e $\lambda^* \in [\lambda_{\min}, \lambda_{\max}]^m$ tais que

$$\lim_{k \in K_1} \lambda^k = \lambda^*.$$

Devido à viabilidade de x^* e pelo fato que $\lim_{k \rightarrow \infty} \varepsilon_k = 0$, tomando os limites na desigualdade (3.27) para $k \in K_1$, obtemos

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

Como \tilde{x} é um ponto arbitrário e viável, concluímos a demonstração. \square

Vale destacar que as demonstrações aqui apresentadas diferem das demonstrações presentes em (BIRGIN; MARTÍNEZ, 2014). Embora tenhamos nos embasado nas ideias apresentadas pelos autores para estabelecer os resultados, aproveitamos a estrutura do problema (3.17), que vem a ser um caso particular do problema abordado em (BIRGIN; MARTÍNEZ, 2014), para construir demonstrações mais simples.

4 EXPERIMENTOS NUMÉRICOS

Neste capítulo, apresentamos experimentos numéricos a partir dos algoritmos desenvolvidos usando a metodologia de Lagrangiano aumentado. Os algoritmos propostos são robustos, no sentido que podem ser usados para resolver uma gama de problemas de alocação de recursos. Ilustramos isso resolvendo problemas quadráticos, não quadráticos e não separáveis. Todos os problemas levam em consideração a convexidade na função objetivo.

Na Seção 4.1, apresentamos resultados obtidos com o algoritmo proposto aplicado a uma coleção de problemas separáveis e não quadráticos. Na Seção 4.2, consideramos a análise de duas variações do algoritmo proposto, uma delas incluindo busca de Armijo e a outra utilizando Newton Projetado com busca. Na Seção 4.3 resolvemos o problema de determinar a carteira de menor risco. Este problema, embora quadrático, é não separável. Na Seção 4.4 ilustramos a eficiência do algoritmo quando aplicado a exemplos quadráticos separáveis. Na Seção 4.5 apresentamos os resultados obtidos aplicando o algoritmo proposto a problemas quadráticos com mais de uma restrição linear e restrições de caixa. Na Seção 4.6, resolvemos problemas mais gerais, com a função objetivo convexa e com restrições não lineares. Na Seção 4.7 apresentamos os experimentos numéricos preliminares referentes à análise do problema de complementaridade linear.

Os experimentos numéricos referentes ao problema com uma restrição linear foram executados no software Matlab R2019 em um notebook CPU Intel Core i7-8550U (1.8 GHz) com 8GB de RAM. Enquanto nos testes relacionados aos problemas que envolvem mais de uma restrição de igualdade, sendo essas não necessariamente convexas, utilizamos o software Matlab R2021 em um Macbook Pro com processador Quad-Core Intel Core i7 (2.8 GHz) e 16GB de RAM.

4.1 PROBLEMAS NÃO QUADRÁTICOS

Nesta seção, apresentamos resultados da execução do algoritmo de Lagrangiano aumentado (Algoritmo 3) para problemas não quadráticos e separáveis de alocação de recursos, embasados no artigo de Wright e Lim (2020). Em todos os problemas abordados, a função objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a ser minimizada é separável, ou seja, pode ser escrita como

$$f(x) = \sum_{j=1}^n f_j(x_j),$$

desta forma, a matriz Hessiana de f , embora varie com x , é diagonal a cada iteração.

No **Passo 1** do Algoritmo 3 precisamos determinar, a cada iteração, \bar{x}^k tal que $\nabla_x L(\bar{x}^k, \lambda_k, r_k) = 0$. Isso gera um subproblema interno, o qual resolvemos por meio do método de Newton.

4.1.1 Descrição dos problemas testes

Para cada problema, definimos a função; as hipóteses do problema; o parâmetro de penalidade r ; a geração de dados, ou seja, a forma como os dados são gerados, com base em (WRIGHT; LIM, 2020).

Problema 1: Projeção l^p ponderada

Função objetivo: $f_j(x_j) = \Gamma_j |x_j - \eta_j|^p$.

Hipóteses: $p \geq 2$; $\Gamma_j > 0$; $b_j = 1$.

Geração de dados: $p \in \{2, 3\}$, Γ_j , η_j foram uniformemente gerados aleatoriamente ao longo do intervalo $(10, 25)$, ℓ_j foram gerados sobre $\left(0, \frac{3\Gamma_j}{2\eta_j}\right)$, e u_j sobre $\left(\ell_j, \frac{3\eta_j}{\Gamma_j}\right)$. Consideramos o parâmetro de penalidade inicial $r_0 = 0.9$.

Problema 2: Amostragem estratificada

Função objetivo: $f_j(x_j) = \Gamma_j/x_j$.

Hipóteses: $\Gamma_j, \ell_j > 0$.

Geração de dados: Γ_j foram uniformemente gerados aleatoriamente ao longo do intervalo $(10000, 20000)$. Similarmente, b_j foi gerado sobre $(10, 50)$, ℓ_j sobre $(100, 200)$, e u_j sobre $(\ell_j, 200)$. Para cada n , consideramos o parâmetro de penalidade inicial $r_0 = 1/n$, onde n representa a dimensão dos problemas.

Problema 3: Otimização de combustível

Função objetivo: $f_j(x_j) = \Gamma_j/x_j^3$.

Hipóteses: $\Gamma_j, \ell_j > 0$.

Geração de dados: ℓ_j foram uniformemente gerados aleatoriamente ao longo do intervalo $(0.7, 1)$, $u_j = 1.5\ell_j$, $\Gamma_j = \tilde{\Gamma}_j \ell_j^4$, em que $\tilde{\Gamma}_j$ foi gerado sobre $(0.8, 1.2)$. Consideramos o parâmetro de penalidade inicial $r_0 = 0.8$.

Problema 4: Quártica inclinada

Função objetivo: $f_j(x_j) = \frac{1}{4}(1 - x_j)^4 + \Gamma_j(1 - x_j)$.

Hipóteses: $\ell_j = 0$; $u_j = 1$; $0 < \Gamma_j < 1$.

Geração de dados: Γ_j foram uniformemente gerados aleatoriamente ao longo de $(0, 1)$ e classificados em ordem crescente. O parâmetro de penalidade inicial r_0 foi fixado em 0.5.

Problema 5: Quártica convexa geral

Função objetivo: $f_j(x_j) = \alpha_j x_j^4 + \beta_j x_j^3 + \Gamma_j x_j^2 + \eta_j x_j$.

Hipóteses: $\alpha_j > 0$; $8\alpha_j \Gamma_j \geq 3\beta_j^2$.

Geração de dados: $b_j = 1$, $\alpha_j = \frac{\pi_j^2 + \sigma_j^2}{\sqrt{8}}$, $\beta_j = \frac{\pi_j \zeta_j + \sigma_j \chi_j}{\sqrt{3}}$ e $\Gamma_j = \frac{\zeta_j^2 + \chi_j^2}{\sqrt{8}}$, onde π_j , σ_j , ζ_j e χ_j foram uniformemente gerados aleatoriamente ao longo do intervalo $(0, 1)$. $\eta_j = -f'(\tilde{\eta}_j)$, em que $\tilde{\eta}_j$ foi gerado sobre $(0, 1)$, $u_j = \min\{\tilde{\eta}_j, \tilde{u}_j\}$, com \tilde{u}_j uniformemente gerados aleatoriamente ao longo de $(0, \tilde{\eta}_j)$, e l_j foi gerado sobre $(0, u_j)$. O parâmetro de penalidade inicial r_0 foi fixado em 0.8.

Observação 4.1.1. *A escolha do parâmetro de penalidade inicial r_0 , nos problemas anteriormente descritos, foi realizada de forma heurística, assim como a escolha do multiplicador de Lagrange inicial $\lambda_0 = 10$. Os valores para os limitantes λ_{\min} e λ_{\max} foram fixados em -10^{10} e 10^{10} , respectivamente. Consideramos $\gamma = 2$, $\theta = 0.9$, $\varepsilon = 10^{-4}$ e $k_0 = 300$ os parâmetros auxiliares.*

A seguir, apresentamos os resultados da execução do Algoritmo 3 para problemas de alocação de recursos não quadráticos, anteriormente elencados.

4.1.2 Resultados computacionais

Para cada dimensão executamos uma série de instâncias do problema geradas aleatoriamente (mais especificamente, o número de problemas é sempre $10^7/n$). Registramos o número mínimo, máximo, e a média de iterações, bem como os tempos de CPU e o parâmetro de penalidade r utilizados pelo Algoritmo 3 para resolver cada problema, e exibimos os resultados em tabelas.

A ordem e a notação em que os valores são listados nas tabelas são: a dimensão n ; o valor de p , se existir; o mínimo, máximo e a média do número de iterações externas para a coleção de problemas (MinE, MaxE e MedE, respectivamente); os mesmos dados para as iterações internas (MinI, MaxI e MedI, respectivamente); e os mesmos dados para o tempo de CPU (MinT, MaxT e MedT, respectivamente) em segundos.

TABELA 1 – PROBLEMA DE PROJEÇÃO PONDERADA, $r = 0.9$

| n | p | MinE | MaxE | MedE | MinI | MaxI | MedI | MinT | MaxT | MedT |
|---------|-----|------|------|-------|------|------|---------|--------|--------|--------|
| 50000 | 2 | 57 | 62 | 59.6 | 114 | 124 | 119.2 | 0.06 | 0.123 | 0.066 |
| | 3 | 48 | 300 | 58.82 | 831 | 4985 | 1049.13 | 0.24 | 1.752 | 0.329 |
| 100000 | 2 | 60 | 63 | 61.75 | 120 | 126 | 123.5 | 0.168 | 0.247 | 0.178 |
| | 3 | 48 | 180 | 59.16 | 929 | 3385 | 1120.89 | 0.69 | 2.497 | 0.852 |
| 200000 | 2 | 63 | 65 | 63.96 | 126 | 130 | 127.92 | 0.913 | 1.119 | 0.958 |
| | 3 | 54 | 97 | 60.72 | 1041 | 1907 | 1211.84 | 4.934 | 9.293 | 5.881 |
| 500000 | 2 | 66 | 68 | 66.95 | 132 | 136 | 133.9 | 2.618 | 2.763 | 2.672 |
| | 3 | 55 | 77 | 60.5 | 1140 | 1621 | 1299.7 | 15.846 | 22.637 | 18.019 |
| 1000000 | 2 | 65 | 70 | 67.5 | 130 | 140 | 135 | 5.256 | 5.633 | 5.422 |
| | 3 | 58 | 88 | 66.3 | 1270 | 1928 | 1469.1 | 36.984 | 53.036 | 42.362 |

FONTE: Torrealba et al. (2022)

TABELA 2 – PROBLEMA DE AMOSTRAGEM ESTRATIFICADA, $r = 1/n$

| n | MinE | MaxE | MedE | MinI | MaxI | MedI | MinT | MaxT | MedT |
|---------|------|------|-------|------|------|---------|---------|---------|---------|
| 50000 | 78 | 90 | 83.87 | 981 | 2219 | 1485.24 | 2.793 | 7.171 | 4.598 |
| 100000 | 83 | 90 | 86.07 | 1177 | 2479 | 1662.3 | 5.502 | 14.853 | 9.459 |
| 200000 | 85 | 92 | 88.44 | 1457 | 2500 | 1833.88 | 19.036 | 33.272 | 25.141 |
| 500000 | 87 | 108 | 92.3 | 1761 | 3036 | 2152.2 | 64.868 | 107.896 | 77.576 |
| 1000000 | 85 | 100 | 91.1 | 2144 | 2752 | 2308.6 | 135.838 | 193.732 | 161.753 |

FONTE: Torrealba et al. (2022)

TABELA 3 – PROBLEMA DE OTIMIZAÇÃO DE COMBUSTÍVEL, $r = 0.8$

| n | MinE | MaxE | MedE | MinI | MaxI | MedI | MinT | MaxT | MedT |
|---------|------|------|------|------|------|------|-------|-------|-------|
| 50000 | 3 | 3 | 3 | 8 | 8 | 8 | 0.036 | 0.073 | 0.038 |
| 100000 | 3 | 3 | 3 | 8 | 8 | 8 | 0.067 | 0.078 | 0.07 |
| 200000 | 3 | 3 | 3 | 8 | 8 | 8 | 0.154 | 0.172 | 0.158 |
| 500000 | 3 | 5 | 3.25 | 8 | 10 | 8.25 | 0.398 | 0.539 | 0.421 |
| 1000000 | 3 | 12 | 6.6 | 8 | 17 | 11.6 | 0.782 | 2.026 | 1.275 |

FONTE: Torrealba et al. (2022)

TABELA 4 – PROBLEMA DE QUÁRTICA INCLINADA, $r = 0.5$

| n | MinE | MaxE | MedE | MinI | MaxI | MedI | MinT | MaxT | MedT |
|---------|------|------|-------|------|------|-------|--------|--------|--------|
| 50000 | 17 | 19 | 17.63 | 510 | 570 | 528.9 | 1.195 | 1.595 | 1.345 |
| 100000 | 17 | 19 | 18.12 | 510 | 570 | 543.6 | 2.387 | 2.956 | 2.583 |
| 200000 | 19 | 21 | 19.04 | 570 | 630 | 571.2 | 6.685 | 7.668 | 6.942 |
| 500000 | 20 | 20 | 20 | 600 | 600 | 600 | 19.073 | 19.958 | 19.535 |
| 1000000 | 20 | 21 | 20.5 | 600 | 630 | 615 | 37.623 | 40.671 | 39.401 |

FONTE: Torrealba et al. (2022)

TABELA 5 – PROBLEMA DE QUÁRTICA CONVEXA GERAL. $r = 0.8$

| n | MinE | MaxE | MedE | MinI | MaxI | MedI | MinT | MaxT | MedT |
|---------|------|------|--------|-------|-------|----------|---------|---------|---------|
| 50000 | 159 | 216 | 198.30 | 8041 | 11736 | 9845.54 | 23.97 | 40.78 | 31.53 |
| 100000 | 168 | 220 | 204.59 | 9076 | 15271 | 10849.7 | 50.96 | 94.55 | 66.32 |
| 200000 | 203 | 222 | 212.34 | 10361 | 13774 | 11829.06 | 167.21 | 224.38 | 195.26 |
| 500000 | 221 | 227 | 223.4 | 12815 | 13953 | 13349.75 | 511.53 | 556.6 | 536.79 |
| 1000000 | 224 | 233 | 228.2 | 13080 | 15941 | 14179.8 | 1051.04 | 1282.81 | 1148.09 |

FONTE: Torrealba et al. (2022)

Os resultados apresentados nas tabelas anteriores evidenciam que todos os problemas foram resolvidos com êxito utilizando a metodologia do Algoritmo 3. Observe que o número de iterações é praticamente independente da dimensão para cada problema testado.

4.1.3 Comparação com algoritmos de localização de raiz

Segundo Bretthauer e Shetty (2002), especificamente a seção 3.1 do artigo, podemos resolver o Problema (2.1) utilizando um método de multiplicadores para encontrar uma solução de $\phi(\lambda) = b^T x(\lambda) - c = 0$, sendo

$$x_i(\lambda) = \begin{cases} l_i, & \text{se } \tilde{x}_i \leq l_i \\ \tilde{x}_i, & \text{se } l_i < \tilde{x}_i < u_i \\ u_i, & \text{se } \tilde{x}_i \geq u_i. \end{cases}$$

A seguir, comparamos o Método de Lagrangiano aumentado (LA), estabelecido no Algoritmo 3, com três algoritmos clássicos usados para localizar raízes, a saber, Método da Bisseção (BS), Método Regula Falsi (RF) e Método da Secante (SC). As implementações dos métodos BS, RF e SC foram embasadas em (MÜNNICH et al., 2012). Para estes testes, resolvemos os problemas de otimização de combustível, amostragem estratificada, projeção ℓ^2 ponderada e quártica inclinada para obter uma comparação em termos de tempo de CPU.

Com o intuito de facilitar tal comparação vamos utilizar a análise de perfil de desempenho apresentada por Dolan e Moré (2002), que fornece uma maneira de avaliar e comparar o desempenho de um conjunto \mathcal{S} de n_s algoritmos aplicados a um conjunto \mathcal{P} de n_p problemas testes.

Vamos considerar $t_{p,s}$ o tempo computacional necessário para satisfazer o critério de parada do algoritmo na resolução do problema $p \in \mathcal{P}$ pelo esquema $s \in \mathcal{S}$. Se o algoritmo s não satisfazer os critérios de parada, consideramos $t_{p,s} = \infty$. O índice de desempenho será definido por

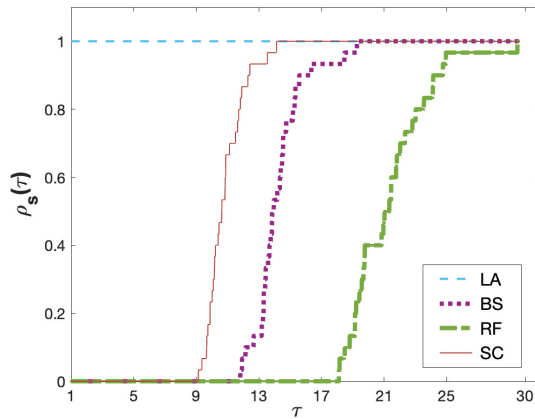
$$r_{p,s} = \frac{t_{p,s}}{\min \{t_{p,j} \mid j \in \mathcal{S}\}}.$$

Para o algoritmo mais eficiente temos $r_{p,s} = 1$, e conforme esse número aumenta, o desempenho do algoritmo piora. Consideramos a avaliação geral do desempenho de um determinado algoritmo s determinada pela função de desempenho $\rho_s : [1, \infty) \rightarrow [0, 1]$, definida por

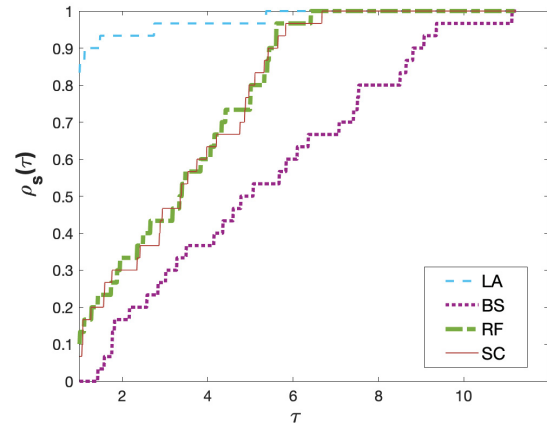
$$\rho_s(\tau) = \frac{1}{n_p} \text{card}\{p \in \mathcal{P} \mid r_{s,p} \leq \tau\}.$$

Assim, $\rho_s(\tau)$ é a probabilidade que uma taxa de performance $r_{p,s}$ esteja dentro de um fator $\tau \in \mathbb{R}$ da melhor taxa de performance. Note que quando $\tau = 1$ a função de desempenho representa a proporção de problemas que o algoritmo s resolve no menor tempo.

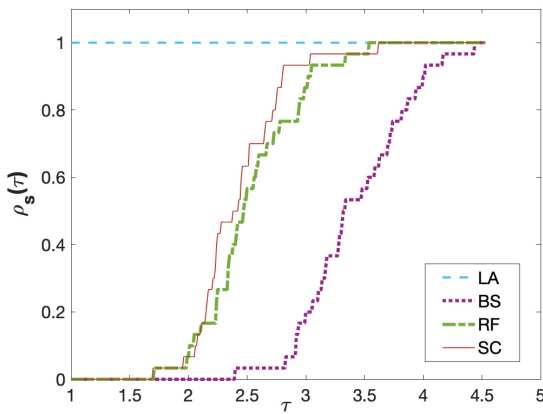
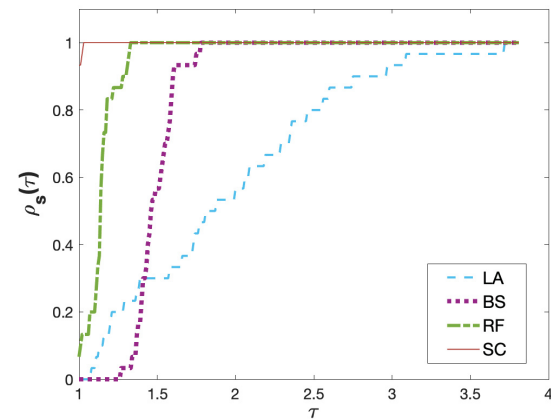
FIGURA 2 – PERFIL DE DESEMPENHO PARA TEMPO DE CPU



(a) Otimização de combustível



(b) Amostragem estratificada

(c) Projeção ℓ^2 ponderada

(d) Quártica inclinada

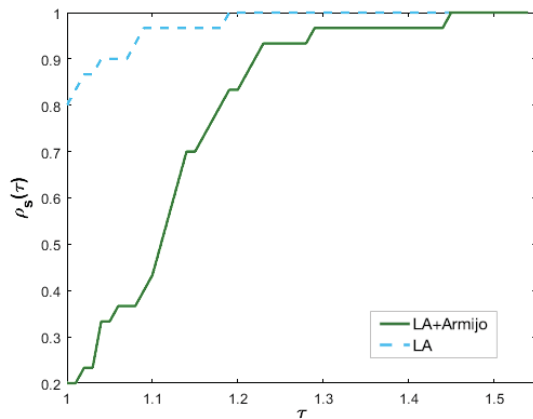
FONTE: Torrealba et al. (2022)

Os gráficos de perfil de desempenho apresentados na FIGURA 2 resumem a comparação de tempo de CPU médio, em segundos, dos quatro esquemas. Para tal comparação foram gerados aleatoriamente 30 problemas de cada classe. Para estes problemas consideramos a dimensão $n = 10000$. Observamos que nestes exemplos o Algoritmo 3 (LA) apresenta um bom desempenho quando comparado aos demais métodos, sendo que apenas no caso quártica inclinada o método é menos eficiente frente aos demais.

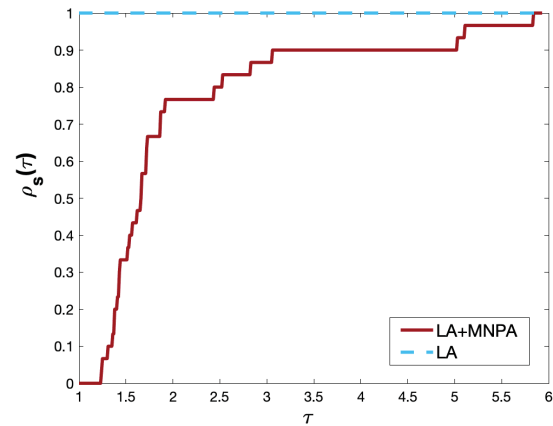
4.2 VARIANTES DO ALGORITMO DE LAGRANGIANO AUMENTADO

Nesta seção, consideramos duas variações do Algoritmo 3. Na primeira alteramos o passo unitário no método de Newton para resolver o subproblema (2.3), incluindo uma busca de Armijo, a qual denotamos por LA+Armijo. Enquanto que na segunda variante substituímos os **Passos 1 e 2** do Algoritmo 3, resolvendo o subproblema (2.3) utilizando o método de Newton Projetado, baseado em (BERTSEKAS, 1982b), mas fazendo a busca clássica de Armijo, identificamos essa variação como LA+MNPA.

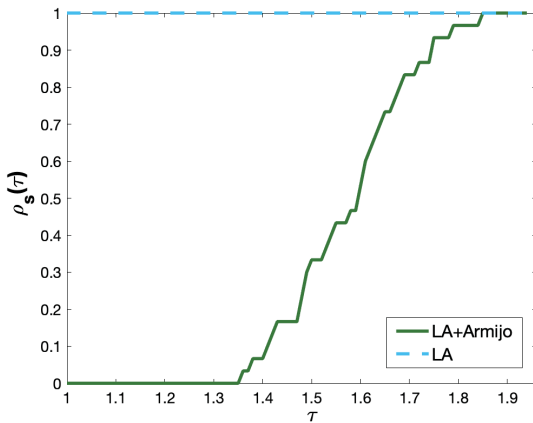
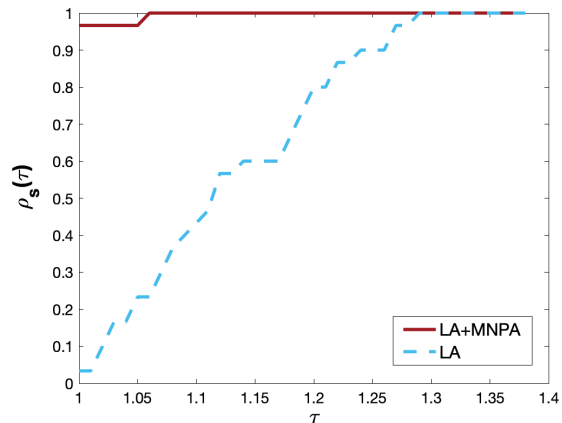
FIGURA 3 – PERFIL DE DESEMPENHO PARA O NÚMERO DE ITERAÇÕES



(a) Amostragem estratificada - LA e LA+Armijo



(b) Amostragem estratificada - LA e LA+PNMA

(c) Projeção ℓ^2 ponderada - LA e LA+Armijo(d) Projeção ℓ^2 ponderada - LA e LA+PNMA

FONTE: Torrealba et al. (2022).

Para os testes computacionais consideramos 30 problemas gerados aleatoriamente, com os mesmos parâmetros estabelecidos para os Problemas 1 e 2 apresentados na Seção 4.1, ou seja, consideramos os problemas de projeção ℓ^2 ponderada e de amostragem estratificada.

Na FIGURA 3 ilustramos a comparação entre o Algoritmo 3 e as variações consideradas. Para uma melhor visualização dos resultados comparamos separadamente, pois a escala é muito diferente.

Podemos inferir que, em geral, ignorar o cálculo do tamanho do passo é benéfico, pois raramente obtemos melhores resultados considerando essa conta, e mesmo quando temos uma redução no número de iterações não justifica o gasto computacional extra em cada iteração.

Também concluímos que alterar o cálculo do subproblema (2.4) e a projeção na caixa presentes no Algoritmo 3 (**Passos 1 e 2**), para uma abordagem que lida com ambos ao mesmo tempo não é necessariamente vantajoso. Isso ocorre porque as fórmulas de projeção nos casos considerados são simples.

4.3 OTIMIZAÇÃO DE CARTEIRA DE INVESTIMENTOS

Nesta seção, aplicamos o algoritmo de Lagrangiano aumentado ao problema de seleção de carteiras, também conhecido na literatura como otimização de portfólios, que embora seja um problema quadrático é não separável. É de conhecimento de todo investidor que existe uma negociação (*trade-off*) entre recompensa e risco em carteiras de investimento. Em outras palavras, para aumentar o retorno esperado do investimento, um investidor deve estar disposto a tolerar riscos mais altos. Na década de 1950, Harry Markowitz introduziu o Modelo de Média-Variância pelo qual foi laureado com o Prêmio Nobel de Economia, ver (MARKOWITZ, 1952).

Uma carteira de investimento é definida pelo vetor $x = (x_1, x_2, \dots, x_n)^T$, onde x_i denota a proporção do investimento a ser feito no ativo i . Assumindo que todos os ativos são investidos e a venda a descoberto (*short-selling*) não é permitida, então o problema deve satisfazer as restrições $\sum_{i=1}^n x_i = 1$ e $x_i \geq 0$ para $i = 1, \dots, n$.

Como os retornos de uma carteira não são conhecidos antecipadamente, e são frequentemente consideradas variáveis aleatórias que seguem uma distribuição normal, Markowitz associou o retorno de uma carteira p ao retorno esperado $\bar{R}_p = \sum_{i=1}^n \bar{R}_i x_i$, onde \bar{R}_i é o retorno esperado do ativo i e x_i a proporção do ativo i na carteira. Além disso, Markowitz associou a variação em uma carteira p pela sua variância esperada $\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \delta_{ij}$ e seu risco por $\sigma_p = \sqrt{\sigma_p^2}$, onde δ_{ij} é a covariância entre os ativos i e j . Se denotamos por $Q_{ij} = \delta_{ij}$ sendo $\delta_{ii} = \delta_i^2$ a variância do ativo i , segue que

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \delta_{ij} = x^T Q x,$$

onde $Q \in \mathbb{R}^{n \times n}$ é uma matriz simétrica, semidefinida positiva, conhecida como matriz de covariâncias. Portanto, o modelo básico de média-variância proposto por Markowitz é dado por

$$\begin{aligned} &\text{minimizar} && x^T Q x \\ &\text{sujeito a} && e^T x = 1 \\ &&& x \geq 0, \end{aligned} \tag{4.1}$$

onde $Q \in \mathbb{R}^{n \times n}$ é a matriz de covariâncias e $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$.

Como é possível notar, o modelo (4.1) é um caso especial do problema (2.1) com a função objetivo sendo $f(x) = x^T Q x$, o vetor $b = e = (1, 1, \dots, 1)^T$, a constante $c = 1$ e o limitante inferior das variáveis dado por $\ell = (0, 0, \dots, 0)^T$. Como $e^T x = 1$ e $x \geq 0$ essas restrições são equivalentes a $e^T x = 1$ e $0 \leq x_i \leq 1$, para cada $i = 1, \dots, n$, conseqüentemente $u = (1, 1, \dots, 1)^T$, embora esse limitante seja omitido na formulação do problema (4.1) estabelecido por Markowitz.

Podemos executar o algoritmo de Lagrangiano aumentado para resolver o modelo (4.1), e mais adiante incluiremos um parâmetro de aversão ao risco neste modelo para obter um novo.

Como o modelo (4.1) não tem termo linear (uma vez que $a = 0$ na função objetivo) obtemos $y = 0$ e $z = Q^{-1}e$ em (2.17). Assim, precisamos resolver um único sistema linear e, usando esta informação na equação (2.18), concluímos que \bar{x}^k , solução do Subproblema gerado pelo Algoritmo 3 na iteração k , é determinado por

$$\begin{aligned}\bar{x}^k &= [(r_k - \lambda_k) - \alpha_k(r_k - \lambda_k)e^T z]z \\ &= (r_k - \lambda_k)[1 - \alpha_k e^T z]z,\end{aligned}$$

em que $e^T = (1, 1, \dots, 1)$ e $\alpha_k = \frac{r_k}{1 + r_k e^T z}$.

Para aplicar o Algoritmo 3 ao modelo (4.1), utilizamos os dados de ativos da bolsa de valores brasileira. Mais especificamente, buscamos os retornos das ações que faziam parte do índice Ibovespa (o índice mais importante do mercado brasileiro) durante os pregões de cinco anos, a saber, de 01/08/2015 a 01/08/2020. Para isso utilizamos o *yahoo finance*, e foram constituídos 75 ativos sendo que 8 deles não tiveram todos os retornos no período e foram excluídos da carteira. Portanto, um total de 67 ativos compõem a carteira que estão listados na TABELA 6 a seguir. Destacamos que ocorreram 1238 pregões em bolsa no período considerado e a dimensão da matriz de covariâncias é de 67 linhas por 67 colunas. Na TABELA 6, as colunas *Ordem*, *Tickers* e *Nome*, são respectivamente, a ordem que os artigos foram baixados do site da Bovespa, um código para o nome do ativo (esse código é padrão na Bovespa) e o nome do ativo.

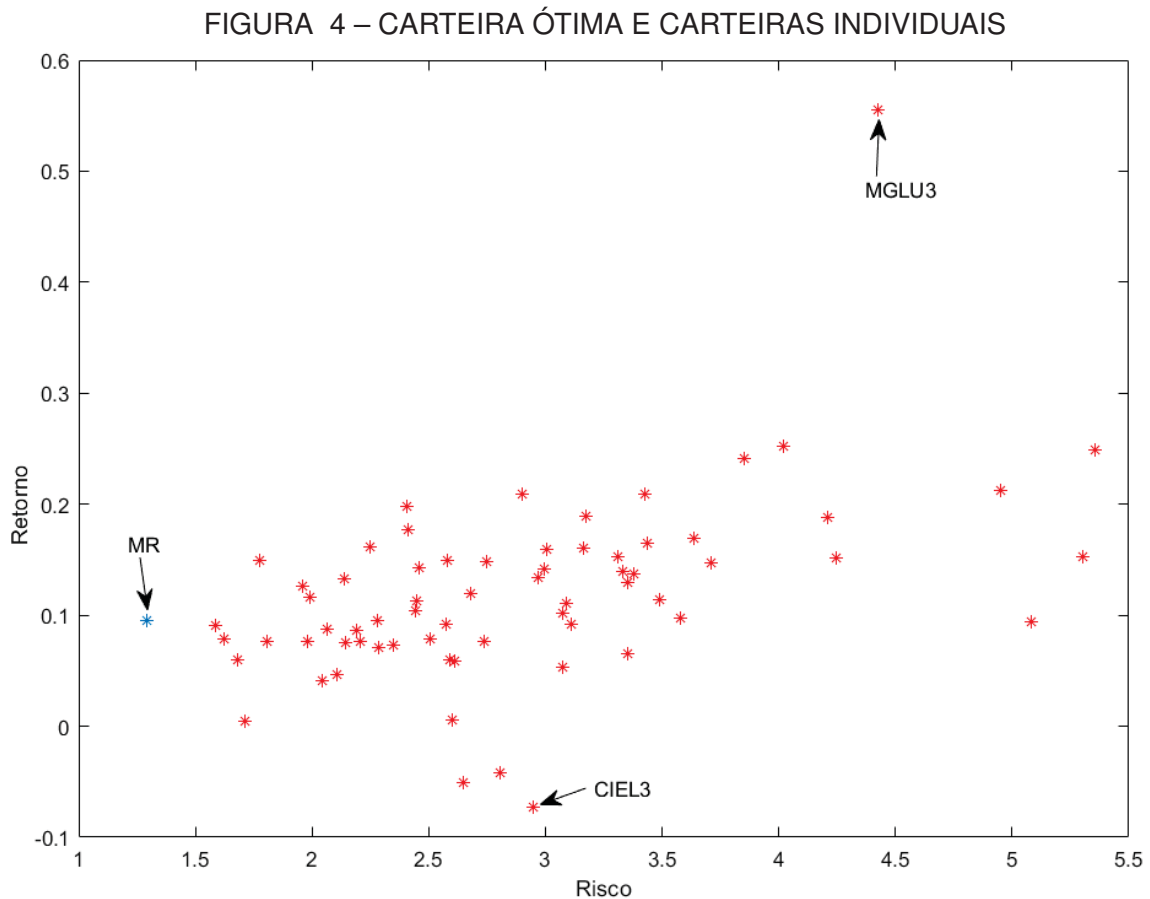
TABELA 6 – ATIVOS NA CARTEIRA E SEUS NOMES

| <i>Ordem</i> | <i>Tickers</i> | <i>Nome</i> | <i>Ordem</i> | <i>Tickers</i> | <i>Nome</i> |
|--------------|----------------|--------------|--------------|----------------|----------------|
| 1 | ABEV3 | AMBEV S/A | 35 | HYPE3 | HYPERA |
| 2 | B3SA3 | B3 | 36 | IGTA3 | IGUATEMI |
| 3 | BBAS3 | BRASIL | 37 | ITSA4 | ITAUSA |
| 4 | BBDC3 | BRADESCO | 38 | ITUB4 | ITAUUNIBANCO |
| 5 | BBDC4 | BRADESCO | 39 | JBSS3 | JBS |
| 6 | BBSE3 | BBSEGURIDADE | 40 | KLBN11 | KLABIN S/A |
| 7 | BEEF3 | MINERVA | 41 | LAME4 | LOJAS AMERIC |
| 8 | BRAP4 | BRADESPAR | 42 | LREN3 | LOJAS RENNER |
| 9 | BRFS3 | BRF SA | 43 | MGLU3 | MAGAZINE LUIZA |
| 10 | BRKM5 | BRASKEM | 44 | MRFG3 | MARFRIG |
| 11 | BRML3 | BR MALLS PAR | 45 | MRVE3 | MRV |
| 12 | BTOW3 | B2W DIGITAL | 46 | MULT3 | MULTIPLAN |
| 13 | CCRO3 | CCR SA | 47 | PCAR3 | P.ACUCAR-CBD |
| 14 | CIEL3 | CIELO | 48 | PETR3 | PETROBRAS |
| 15 | CMIG4 | CEMIG | 49 | PETR4 | PETROBRAS |
| 16 | COGN3 | COGNA ON | 50 | QUAL3 | QUALICORP |
| 17 | CPFE3 | CPFL ENERGIA | 51 | RADL3 | RAIADROGASIL |
| 18 | CSAN3 | COSAN | 52 | RAIL3 | RUMO S.A. |
| 19 | CSNA3 | SID NACIONAL | 53 | RENT3 | LOCALIZA |
| 20 | CVCB3 | CVC BRASIL | 54 | SANB11 | SANTANDER BR |
| 21 | CYRE3 | CYRELA REALT | 55 | SBSP3 | SABESP |
| 22 | ECOR3 | ECORODOVIAS | 56 | SULA11 | SUL AMERICA |
| 23 | EGIE3 | ENGIE BRASIL | 57 | SUZB3 | SUZANO S.A. |
| 24 | ELET3 | ELETROBRAS | 58 | TAE11 | TAESA |
| 25 | ELET6 | ELETROBRAS | 59 | TIMP3 | TIM PART S/A |
| 26 | EMBR3 | EMBRAER | 60 | TOTS3 | TOTVS |
| 27 | ENBR3 | ENERGIAS BR | 61 | UGPA3 | ULTRAPAR |
| 28 | ENGI11 | ENERGISA | 62 | USIM5 | USIMINAS |
| 29 | EQTL3 | EQUATORIAL | 63 | VALE3 | VALE |
| 30 | FLRY3 | FLEURY | 64 | VIVT4 | TELEF BRASIL |
| 31 | GGBR4 | GERDAU | 65 | VVAR3 | VIAVAREJO |
| 32 | GOAU4 | GERDAU MET | 66 | WEGE3 | WEG |
| 33 | GOLL4 | GOL | 67 | YDUQ3 | YDUQS PART |
| 34 | HGTX3 | CIA HERING | — | — | — |

FONTE: Torrealba et al. (2022).

Na FIGURA 4, MR é carteira ótima determinada pelo algoritmo de Lagrangiano aumentado (Algoritmo 3) aplicado ao modelo (4.1), a qual tem um retorno de 0.0953 e risco de 1.2920 por dia. Os outros pontos são carteiras investindo 100% em cada ativo individual. A carteira MGLU3 é determinada pelo investimento de 100% na companhia

MAGAZINE LUIZA e esta é a carteira com maior retorno, que é 0.5548. A carteira CIEL3 é determinada pelo investimento de 100% na companhia CIELO e o menor retorno dentre todas carteiras, além disso, é negativo, sendo seu valor igual a -0.0727 . Vale destacar que, para o período considerado, o risco de investir 100% em MGLU3 é 242,52%, maior do que investir na carteira MR ótima.



FONTE: Torrealba et al. (2022).

Outro fato importante a se notar é a respeito da diversificação. Todas as carteiras determinadas pelo investimento de 100% em qualquer um dos ativos individuais têm um risco maior do que a carteira ótima que é formada pela combinação de vários ativos como na carteira MR. Destacamos isso adiante na TABELA 7 que mostra a carteira MR ótima.

Na TABELA 7, denotamos por *ord* a ordem dos ativos, *prop* é a proporção de cada ativo na carteira MR, \bar{R}_i é o retorno esperado do ativo i , σ_i é o risco do ativo i . Por exemplo, a ordem 1 é o ativo 1, que de acordo com a TABELA 6 é AMBEV S/A, e a proporção desse ativo para a carteira MR ótima é 0.0576 cujo retorno e risco esperado são 0.0047 e 1.7124, respectivamente. Similarmente, para os demais ativos em ordem 2 a 67. Adicionalmente, destacamos os ativos CIELO, na ordem 14, e MAGAZINE LUIZA

na ordem 43, ambos com proporção igual a 0 na carteira MR ótima. Finalmente, na ordem 68, destacamos o retorno esperado e o risco da carteira MR ótima.

TABELA 7 – CARTEIRA ÓTIMA DETERMINADA PELO ALGORITMO DE LAGRANGIANO AUMENTADO

| <i>ord</i> | <i>prop</i> | \bar{R}_p | σ_p | <i>ord</i> | <i>prop</i> | \bar{R}_p | σ_p |
|------------|-------------|----------------|---------------|------------|-------------|---------------|---------------|
| 1 | 0.0576 | 0.0047 | 1.7124 | 35 | 0.0189 | 0.0863 | 2.1909 |
| 2 | 0 | 0.1980 | 2.4079 | 36 | 0.0122 | 0.0705 | 2.2879 |
| 3 | 0 | 0.1019 | 3.0746 | 37 | 0 | 0.0873 | 2.0634 |
| 4 | 0 | 0.0781 | 2.5037 | 38 | 0.0629 | 0.0756 | 2.1456 |
| 5 | 0.0135 | 0.0916 | 2.5727 | 39 | 0 | 0.1142 | 3.4868 |
| 6 | 0.0406 | 0.0466 | 2.1076 | 40 | 0.0063 | 0.0408 | 2.0433 |
| 7 | 0.0251 | 0.0586 | 2.6114 | 41 | 0 | 0.1198 | 2.6801 |
| 8 | 0 | 0.1895 | 3.1730 | 42 | 0 | 0.1124 | 2.4509 |
| 9 | 0 | -0.0508 | 2.6502 | 43 | 0 | 0.5548 | 4.4254 |
| 10 | 0 | 0.1399 | 3.3334 | 44 | 0 | 0.1418 | 2.9966 |
| 11 | 0 | 0.0594 | 2.5923 | 45 | 0.0164 | 0.1486 | 2.7454 |
| 12 | 0.0038 | 0.2412 | 3.8495 | 46 | 0.0013 | 0.0734 | 2.3459 |
| 13 | 0.0032 | 0.0761 | 2.7399 | 47 | 0.0098 | 0.0944 | 5.0836 |
| 14 | 0 | -0.0727 | 2.9457 | 48 | 0 | 0.1297 | 3.3507 |
| 15 | 0 | 0.0918 | 3.1109 | 49 | 0.0335 | 0.1368 | 3.3773 |
| 16 | 0 | 0.0650 | 3.3541 | 50 | 0.0097 | 0.1109 | 3.0877 |
| 17 | 0.0505 | 0.0759 | 1.8092 | 51 | 0.0382 | 0.1164 | 1.9933 |
| 18 | 0 | 0.1765 | 2.4126 | 52 | 0 | 0.1696 | 3.6387 |
| 19 | 0 | 0.1885 | 4.2068 | 53 | 0 | 0.2090 | 2.8992 |
| 20 | 0 | 0.0979 | 3.5769 | 54 | 0 | 0.1042 | 2.4418 |
| 21 | 0 | 0.1593 | 3.0071 | 55 | 0 | 0.1493 | 2.5815 |
| 22 | 0 | 0.1344 | 2.9696 | 56 | 0.0129 | 0.1431 | 2.4609 |
| 23 | 0.0950 | 0.0908 | 1.5856 | 57 | 0.0621 | 0.0952 | 2.2818 |
| 24 | 0 | 0.2519 | 4.0214 | 58 | 0.1023 | 0.0787 | 1.6206 |
| 25 | 0.0336 | 0.2092 | 3.4264 | 59 | 0 | 0.0769 | 2.2067 |
| 26 | 0.0122 | -0.0421 | 2.8073 | 60 | 0.0095 | 0.0536 | 3.0732 |
| 27 | 0 | 0.0767 | 1.9780 | 61 | 0 | 0.0060 | 2.5992 |
| 28 | 0.0411 | 0.1267 | 1.9587 | 62 | 0.0124 | 0.1521 | 4.2458 |
| 29 | 0.0373 | 0.1498 | 1.7752 | 63 | 0.0303 | 0.1601 | 3.1633 |
| 30 | 0.0492 | 0.1332 | 2.1380 | 64 | 0.0679 | 0.0597 | 1.6814 |
| 31 | 0.0154 | 0.1531 | 3.3097 | 65 | 0 | 0.2122 | 4.9515 |
| 32 | 0 | 0.1468 | 3.7112 | 66 | 0.0055 | 0.1621 | 2.2510 |
| 33 | 0 | 0.2491 | 5.3558 | 67 | 0.0025 | 0.1647 | 3.4376 |
| 34 | 0.0074 | 0.1521 | 5.3011 | 68 | - | 0.0953 | 1.2920 |

FONTE: Torrealba et al. (2022).

Uma das premissas no modelo de Markowitz (4.1) é que todos os investidores preferem um retorno maior e um risco menor. Dessa forma, podemos combinar isso em uma única função objetivo, incorporando um parâmetro de tolerância ao risco denotado por ρ , então o modelo (4.1) se torna

$$\begin{aligned} & \text{minimizar} && x^T Q x - \frac{1}{\rho} \bar{R}^T x \\ & \text{sujeito a} && e^T x = 1 \\ & && x \geq 0, \end{aligned} \tag{4.2}$$

onde $\bar{R} = (\bar{R}_1, \bar{R}_2, \dots, \bar{R}_n)^T$ é o vetor de retorno esperado dos n ativos, $\rho \in (0, \infty)$ representa a preferência do investidor individual, também conhecido como parâmetro de aversão ao risco.

Destacamos aqui, que os dois sistemas em (2.17), que são $y = Q^{-1} \bar{R}$ e $z = Q^{-1} e$, têm a mesma matriz de coeficientes, assim precisamos realizar uma única decomposição de matriz. Usando as equações (2.17) e (2.18) podemos determinar \bar{x}^k no **Passo 1** do Algoritmo 3, que é dado por

$$\begin{aligned} \bar{x}^k &= \frac{y}{\rho} + \left[(r_k - \lambda_k) - \alpha_k \left(e^T \frac{y}{\rho} + (r_k - \lambda_k) e^T z \right) \right] z \\ &= \frac{y}{\rho} + \left[-\frac{\alpha_k}{\rho} e^T y + (r_k - \lambda_k) (1 - \alpha_k e^T z) \right] z, \end{aligned}$$

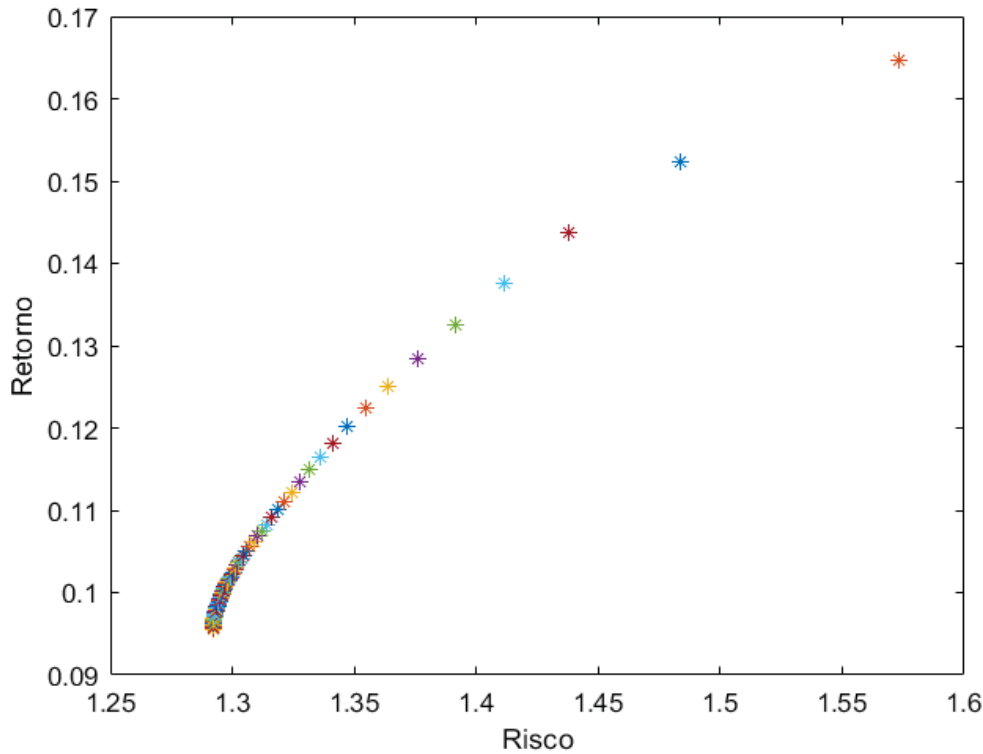
onde $e = (1, 1, \dots, 1)^T$ e $\alpha_k = \frac{r_k}{(1 + r_k e^T z)}$.

Executamos o algoritmo de Lagrangiano aumentado para resolver o problema (4.2), iniciando com $\lambda_0 = 1$ e o parâmetro de penalidade é mantido constante, ou seja, $r_k = r_0 = 1$ para todo $k \in \mathbb{N}$, escolhidos de forma heurística. Os dados usados são os mesmos do primeiro exemplo e os ativos são os fornecidos na TABELA 6. Para cada valor fixado de ρ no problema (4.2) teremos uma carteira ótima determinada pelo algoritmo. Se o investidor for avesso ao risco ele escolherá um valor grande para ρ , o que significa que deseja minimizar o risco. Por outro lado, se o investidor for mais tolerante ao risco ele escolherá um valor pequeno para ρ dando mais peso ao retorno.

Com o intuito de ilustrar a relação entre retorno e risco ao variar o parâmetro de tolerância ao risco, exibimos na FIGURA 5, 996 carteiras determinadas pelo algoritmo de Lagrangiano aumentado variando ρ de 100 a 0.5, por etapas de tamanho 0.1, ou seja, a primeira carteira foi determinada considerando $\rho = 100$, que é a carteira no canto inferior esquerdo da FIGURA 5 e tem retorno esperado de 0.0954 e o risco de 1.2920 e é praticamente igual à carteira MR determinada pelo problema (4.1). A próxima é determinada por $\rho = 99.9$ e assim por diante até $\rho = 0.5$ que é a carteira no canto superior direito da FIGURA 5 e tem retorno esperado de 0.1647 e o risco

1.5729. Observe que a última carteira, determinada fixando $\rho = 0.5$, é quase 22% mais arriscada do que a primeira determinada por $\rho = 100$.

FIGURA 5 – CARTEIRAS ÓTIMAS CONSIDERANDO O PARÂMETRO DE AVERSÃO AO RISCO



FONTE: Torrealba et al. (2022).

4.4 PROBLEMAS QUADRÁTICOS SEPARÁVEIS

Nesta seção, resolvemos o problema quadrático separável utilizando a metodologia do Algoritmo 3. Ao longo desta seção consideramos o problema da mochila determinado da seguinte forma

$$\begin{aligned} &\text{minimizar} && \frac{1}{2}x^T P x - a^T x \\ &\text{sujeito a} && b^T x = c \\ &&& \ell \leq x \leq u, \end{aligned} \tag{4.3}$$

onde $P \in \mathbb{R}^{n \times n}$ é uma matriz definida positiva e diagonal, ou seja, $P = \text{diag}(p_1, \dots, p_n)$, com $p_i > 0$ para cada $i = 1, \dots, n$.

Começamos esta seção com um resultado que estabelece uma tolerância na resolução aproximada do subproblema gerado pelo Algoritmo 3 aplicado ao problema (4.3). Este resultado é especialmente uma contribuição do ponto de vista numérico, pois ao permitir essa tolerância ganhamos a possibilidade de um grau de liberdade na implementação.

Proposição 4.4.1. *Considere a matriz $P^{-1} - \beta\alpha P^{-1}bb^T P^{-1}$, onde $\beta > 0$ é um parâmetro e α é dado por (2.18). Então essa matriz é uma inversa aproximada da matriz $P + rbb^T$, que é a Hessiana do problema quadrático estabelecida em (2.12). E a precisão da aproximação depende de $r - \beta r$, onde r é o parâmetro de penalidade estabelecido no Algoritmo 3.*

Demonstração. De fato, calculando o produto das duas matrizes, obtemos:

$$\begin{aligned}
[P + rbb^T][P^{-1} - \beta\alpha P^{-1}bb^T P^{-1}] &= [P + rbb^T] \left[P^{-1} - \frac{\beta r}{1 + rb^T P^{-1}b} P^{-1}bb^T P^{-1} \right] \\
&= I - \frac{\beta r(bb^T P^{-1})}{1 + rb^T P^{-1}b} + r(bb^T P^{-1}) \\
&\quad - \frac{\beta r^2(b(b^T P^{-1}b)b^T P^{-1})}{1 + rb^T P^{-1}b} \\
&= I - \frac{\beta r(1 + rb^T P^{-1}b)}{1 + rb^T P^{-1}b} (bb^T P^{-1}) + rbb^T P^{-1} \\
&= I + (r - \beta r)bb^T P^{-1}.
\end{aligned}$$

Tal resultado implica que o erro cometido pela aproximação da inversa de $P + rbb^T$ pela matriz $P^{-1} - \beta\alpha P^{-1}bb^T P^{-1}$ depende unicamente de quão pequeno $|r - \beta r|$ é. \square

Observação 4.4.1. *Considere o sistema em (2.14). Suponha que adicionamos um parâmetro multiplicador β na inversa da matriz $(P + rbb^T)^{-1}$, ou seja, resolvemos com a matriz $P^{-1} - \beta\alpha P^{-1}bb^T P^{-1}$. Então x calculado satisfaz o sistema $\nabla_x L(\cdot, \lambda, r) = 0$ aproximadamente com o erro dependendo de $r - \beta r$. Isso significa que tomando β de forma que essa diferença seja suficientemente próxima de zero na implementação, ainda podemos resolver o subproblema gerado pelo algoritmo de Lagrangiano aumentado usando tal x , e então projetá-lo na caixa. Para essa classe de problemas, utilizamos essa informação em nossos experimentos numéricos, uma vez que foi constatada a possibilidade de obter uma convergência mais rápida.*

Realizamos dois tipos de testes para o problema (4.3), para os quais conhecemos a solução. Os dados utilizados foram a matriz $P = \text{diag}(1, 1, \dots, 1) \in \mathbb{R}^{n \times n}$, os vetores $a = (2, 3, \dots, n + 1)^T \in \mathbb{R}^n$, $b = (1, 1, \dots, 1)^T \in \mathbb{R}^n$, $\ell = (0, 0, \dots, 0)^T \in \mathbb{R}^n$, $u = (10, 10, \dots, 10)^T \in \mathbb{R}^n$ e a constante $c = 1$.

O primeiro teste destaca a importância da escolha de β , baseada na Proposição 4.4.1. Nesse teste variamos β de 0.1 a 1, para o problema com dimensão $n = 10^7$, fixado o parâmetro de penalidade em $r = 0.1$. Os resultados são apresentados na TABELA 8. Perceba que quando $\beta = 1$ o algoritmo de Lagrangiano aumentado não

alcança solução em 1000 iterações (o número máximo de iterações considerado nos testes), reforçando que para este problema quadrático diagonal métodos quase-Newton têm uma performance melhor do que o método de Newton.

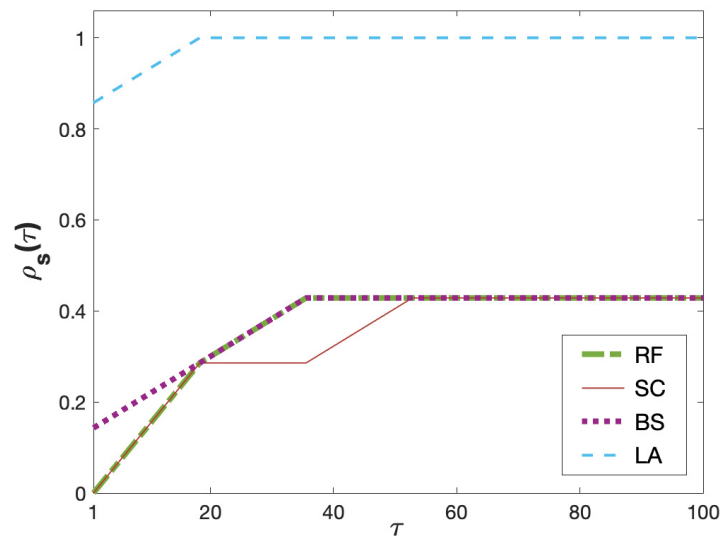
TABELA 8 – UMA COMPARAÇÃO COM DIFERENTES VALORES DE β

| β | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|--------------|------|------|------|-------|-------|-------|-------|-------|-------|------|
| Iterações | 64 | 74 | 87 | 105 | 129 | 164 | 223 | 341 | 693 | 1000 |
| Tempo de CPU | 6.87 | 7.85 | 9.18 | 11.09 | 13.61 | 17.11 | 23.35 | 35.58 | 72.13 | 103 |

FONTE: Torrealba et al. (2022).

No segundo teste fixamos $\beta = 0.1$ e resolvemos este problema usando o algoritmo de Lagrangiano aumentado para diferentes dimensões, a saber, $n = 10, 50, 100, 500, 1000, 5000, 10000$, com precisão $\varepsilon = 10^{-4}$, o parâmetro de penalidade $r = 0.1$ e multiplicador inicial $\lambda_0 = 10$. Ilustramos os resultados na FIGURA 6.

FIGURA 6 – PERFIL DE DESEMPENHO PARA TEMPO DE CPU(s)



FONTE: Torrealba et al. (2022).

O perfil de desempenho exibido na FIGURA 6 mostra que o Algoritmo de Lagrangiano aumentado (LA) apresenta um comportamento superior quando comparado aos métodos de Bisseção (BS), Regula Falsi (RF) e Secante (SC), para os problemas testados.

4.5 PROBLEMAS QUADRÁTICOS COM RESTRIÇÕES LINEARES

Nesta seção, apresentamos os resultados obtidos com o Algoritmo de Lagrangiano aumentado (Algoritmo 8) aplicado a problemas quadráticos com restrições lineares.

Utilizamos a metodologia proposta na Seção 3.1, do Capítulo 3.

Foram gerados de forma randômica problemas quadráticos com restrições lineares de igualdade e caixa, para diferentes dimensões. Aumentamos o número de restrições conforme ampliamos a dimensão dos problemas. Comparamos os resultados obtidos executando o Algoritmo 8 com os resultados obtidos executando a rotina *quadprog*, própria do MATLAB, para resolver problemas quadráticos.

Estabelecemos a matriz que define a função objetivo utilizando o comando *rand* do MATLAB, consideramos a parte simétrica da matriz randômica gerada e em seguida somamos uma matriz identidade multiplicada à dimensão do problema, para garantir que a matriz em questão fosse simétrica definida positiva.

Para cada problema foram estabelecidos o parâmetro de penalidade e o vetor de multiplicadores de Lagrange inicial de forma heurística, ou seja, testamos os problemas para diferentes combinações de r e λ^0 e selecionamos aqueles que apresentaram os melhores resultados na resolução de cada problema.

Na TABELA 9 estão apresentados os resultados dos testes numéricos. A dimensão dos problemas considerados é representada por n , e a quantidade de restrições por m . O número de iterações (externas) do Algoritmo 8 é dada por k , e o parâmetro de penalidade para cada problema por r . Resolvemos cada problema 100 vezes para extrair os tempos mínimo, máximo e a média, que estão representados por MinT, MaxT e MedT, respectivamente. De modo análogo, representamos os tempos obtidos com *quadprog* por MinTq, MaxTq e MedTq, para os respectivos problemas. Por fim, comparamos as soluções x^* obtidas pelo Algoritmo 8 com as soluções x^q fornecidas pela rotina *quadprog*.

TABELA 9 – PROBLEMAS QUADRÁTICOS COM RESTRIÇÕES LINEARES

| n | m | k | r | MinT | MaxT | MedT | MinTq | MaxTq | MedTq | $\ x^* - x^q\ $ |
|-------|------|-----|--------|--------|--------|--------|--------|---------|--------|-----------------|
| 50 | 20 | 3 | 10^5 | 0.0001 | 0.0025 | 0.0002 | 0.0017 | 0.09060 | 0.0046 | 8.96e-07 |
| 100 | 50 | 3 | 10^5 | 0.0003 | 0.0035 | 0.0006 | 0.0024 | 0.08740 | 0.0059 | 1.17e-07 |
| 500 | 200 | 3 | 10^5 | 0.0054 | 0.0117 | 0.0060 | 0.0306 | 0.12480 | 0.0383 | 5.24e-08 |
| 1000 | 500 | 4 | 10^4 | 0.0416 | 0.0559 | 0.0432 | 0.1661 | 0.27280 | 0.1773 | 2.56e-08 |
| 5000 | 1000 | 3 | 10^4 | 3.3229 | 4.4558 | 3.8245 | 7.5173 | 11.8472 | 9.4884 | 5.70e-07 |
| 10000 | 5000 | 12 | 10^2 | 33.269 | 50.647 | 35.909 | 62.763 | 122.927 | 90.446 | 1.06e-07 |

FONTE: A autora.

No Algoritmo 8 utilizamos como ponto inicial $x^0 \in \mathbb{R}^n$ um vetor de uns, fixamos $\lambda^0 = (10, \dots, 10)^T \in \mathbb{R}^m$ e os valores do parâmetro de penalidade inicial estão listados na quarta coluna da TABELA 9. A acurácia foi definida em $\varepsilon = 10^{-6}$. Consideramos $\gamma = 2$, $\theta = 0.9$ e a constante $k_0 = 50$, perceba que o número de iterações sempre

ficou abaixo dessa estimativa, por conta disso não houve atualização do parâmetro de penalidade r . Para garantirmos a limitação do vetor de multiplicadores de Lagrange a caixa $[\lambda_{\min}, \lambda_{\max}]^m$ em que projetamos λ^k a cada iteração do método, foi determinada pelos vetores $\lambda_{\min} = -(10^{10}, \dots, 10^{10})^T$ e $\lambda_{\max} = (10^{10}, \dots, 10^{10})^T$ em \mathbb{R}^m .

Como é possível observar nos resultados da TABELA 9, o algoritmo proposto apresentou bons resultados quando comparado com a rotina quadprog do Matlab, para a classe de problemas testados.

4.6 PROBLEMAS GERAIS

Consideramos problemas da coleção (HOCK; SCHITTKOWSKI, 1981) cuja função objetivo é convexa, para a análise do comportamento do Algoritmo 9. Vale ressaltar que para problemas com restrições de desigualdade adicionamos variáveis de folga e trabalhamos tal como nos problemas com restrições de igualdade.

Semelhante ao realizado por Hock e Schittkowski (1981), consideramos o seguinte esquema de classificação dos problemas: OCD-K. Na tabela a seguir, descrevemos as abreviaturas que podem substituir as letras O, C, D e K que foram consideradas.

TABELA 10 – CLASSIFICAÇÃO DOS PROBLEMAS

| | |
|---|--|
| O | Informação sobre a função objetivo |
| Q | Função objetivo quadrática |
| P | Função objetivo polinômio generalizado |
| G | Função objetivo geral |
| C | Informação sobre as funções restrições |
| L | Funções lineares |
| Q | Funções quadráticas |
| P | Funções polinômios generalizados |
| G | Funções gerais |
| D | Regularidade dos problemas |
| R | Problema regular |
| I | Problema não regular |
| K | Informação a respeito da solução |
| T | Solução exata conhecida |
| P | Solução exata não conhecida |

FONTE: Adaptado de Hock e Schittkowski (1981).

Um problema é dito regular se a primeira e a segunda derivada de todas as funções envolvidas existem na região viável, caso contrário é classificado como não regular. Consideramos apenas problemas regulares nos nossos testes, é possível observar que a maioria dos problemas relacionados em (HOCK; SCHITTKOWSKI, 1981) são regulares. Nos problemas cuja informação a respeito da solução é classificada como *solução exata não conhecida*, significa basicamente que a solução do problema só pode ser obtida numericamente. A solução, seja ela exata ou numérica, está listada nas especificações de cada problema do livro (HOCK; SCHITTKOWSKI, 1981). Nos problemas em que realizamos alterações, como acrescentar restrições de caixa, comparamos a solução obtida pelo Algoritmo 9 com a solução obtida por meio da rotina *fmincon* do Matlab.

O ponto inicial foi considerado conforme listado nas especificações dos problemas em (HOCK; SCHITTKOWSKI, 1981), a acurácia foi fixada em $\varepsilon = 10^{-6}$, consideramos $\gamma = 2$, $\theta = 0.9$, o parâmetro auxiliar $k_0 = 50$ e a caixa $[\lambda_{\min}, \lambda_{\max}]^m$ é determinada pelos vetores $\lambda_{\min} = -(10^{10}, \dots, 10^{10})^T$ e $\lambda_{\max} = (10^{10}, \dots, 10^{10})^T$ em \mathbb{R}^m .

Na TABELA 11 estão listados os problemas testados com o Algoritmo 9. Na primeira coluna estão relacionados os problemas conforme a numeração especificada em (HOCK; SCHITTKOWSKI, 1981). A classificação de cada problema é representada por *Classe*, conforme a TABELA 10. A dimensão é dada por n , de acordo com o problema original, lembrando que em alguns problemas adicionamos variáveis de folga. O número de iterações (externas) é dado por k , e r representa o parâmetro de penalidade escolhido para cada problema. O vetor de multiplicadores de Lagrange inicial é definido a partir do parâmetro λ , multiplicado a um vetor de uns. Resolvemos cada problema 100 vezes para extrair os tempos mínimo, máximo e a média do Algoritmo 9, que estão representados por MinT, MaxT e MedT, respectivamente. De modo análogo, listamos os tempos da *fmincon*, representado por MinTf, MaxTf e MedTf.

TABELA 11 – TESTES COM O ALGORITMO 9

| | <i>Classe</i> | <i>n</i> | <i>k</i> | <i>r</i> | λ | MinT | MaxT | MedT | MinTf | MaxTf | MedTf |
|-------|---------------|----------|----------|----------|-----------|---------|---------|---------|---------|--------|---------|
| P. 22 | QQR-T | 2 | 2 | 10^6 | 1 | 0.00235 | 0.07785 | 0.00579 | 0.00981 | 0.3973 | 0.02266 |
| P. 23 | QQR-T | 2 | 8 | 10^2 | 1 | 0.00827 | 0.14474 | 0.01549 | 0.00983 | 0.2791 | 0.02002 |
| P. 32 | QPR-T | 3 | 4 | 10 | 2 | 0.00471 | 0.10225 | 0.00901 | 0.00965 | 0.2799 | 0.02055 |
| P. 48 | QQR-T | 5 | 8 | 10 | 8 | 0.00839 | 0.16134 | 0.01640 | 0.00923 | 0.2609 | 0.02009 |
| P. 53 | QQR-T | 5 | 12 | 10^2 | 63 | 0.01190 | 0.15343 | 0.02298 | 0.00821 | 0.2598 | 0.01751 |
| P. 55 | GLR-T | 6 | 3 | 10 | 4 | 0.00378 | 0.09493 | 0.00730 | 0.02149 | 0.3044 | 0.03698 |
| P. 60 | PPR-P | 3 | 10 | 1 | 1 | 0.01144 | 0.16692 | 0.02160 | 0.01055 | 0.2744 | 0.02052 |
| P. 61 | QQR-P | 3 | 15 | 1 | 91 | 0.01949 | 0.18331 | 0.03639 | 0.01861 | 0.2920 | 0.03236 |
| P. 76 | QQR-P | 4 | 5 | 10^3 | 1 | 0.00596 | 0.14976 | 0.01120 | 0.01310 | 0.2990 | 0.02590 |
| P. 77 | PGR-P | 5 | 10 | 10 | 1 | 0.01090 | 0.14700 | 0.01931 | 0.04119 | 0.3310 | 0.05888 |

FONTE: A autora.

Nos Problemas 22 e 23 foram adicionadas restrições de caixa e variáveis de folga, uma vez que o problema original apresentava somente restrições de desigualdade. No Problema 32 foram adicionados limitantes superiores às variáveis, neste problema a solução fornecida pela *fmincon* não se aproximou tão bem da solução encontrada pelo Algoritmo 9. Constatamos que a solução fornecida pelo algoritmo proposto apresentou qualidade superior. Modificamos as restrições dos Problemas 48, 53 e 76 para se tornarem quadráticas, pois nos problemas originais havia apenas restrições lineares. No Problema 55, a rotina *fmincon* não alcançou a solução fornecida em (HOCK; SCHITTKOWSKI, 1981).

A configuração da função objetivo no Problema 23 nos permite fazer alterações de modo a trabalhar com diferentes dimensões do problema. Dessa forma, resolvemos variações do problema utilizando o Algoritmo 9. Na TABELA 12 apresentamos os resultados obtidos, na primeira coluna estão as dimensões consideradas, na segunda coluna o número de iterações (externas) que o método executou, e nas três últimas colunas estão relacionados os tempos mínimo, máximo e média; analogamente ao que fizemos para o problema original, resolvemos o problema 100 vezes para extrair tais valores.

TABELA 12 – PROBLEMA 23 COM DIFERENTES DIMENSÕES

| <i>n</i> | <i>k</i> | MinT | MaxT | MedT |
|----------|----------|---------|--------|--------|
| 10 | 11 | 0.02214 | 0.1346 | 0.0384 |
| 100 | 14 | 0.05556 | 0.1755 | 0.0799 |
| 500 | 28 | 1.08750 | 1.7902 | 1.3729 |
| 1000 | 32 | 5.30980 | 7.1397 | 6.0748 |
| 5000 | 15 | 0.11165 | 0.2390 | 0.1589 |

FONTE: A autora.

Utilizamos os mesmos parâmetros do problema original, ou seja, o parâmetro de penalidade foi fixado em $r = 10^2$, e o vetor de multiplicadores de Lagrange inicial λ^0 foi escolhido como sendo um vetor de uns, cuja dimensão é igual ao número de restrições de igualdade.

4.7 PROBLEMA DE COMPLEMENTARIDADE LINEAR

O Problema de Complementaridade Linear, de forma abreviada LCP¹, se refere a um sistema de desigualdades com uma rica teoria matemática, uma variedade de algoritmos desenvolvidos, além de possuir diversas aplicações, ver por exemplo Cottle et al. (2009).

É possível encontrar publicações abordando exemplos e casos especiais do problema de complementaridade linear já na década de 40, como ilustrado com o trabalho de Du Val (1940), possivelmente a referência mais antiga contendo explicitamente uma declaração ao LCP. Mas somente a partir da década de 60, o LCP alcançou notoriedade em estudos direcionados, como observado por Billups e Murty (2000).

Para uma leitura completa e detalhada a respeito de LCP são ótimas referências o livro (COTTLE et al., 2009), e os artigos (COTTLE; DANTZIG, 1970), (BILLUPS; MURTY, 2000), que são as principais fontes de consulta para a elaboração deste capítulo.

4.7.1 Definição do LCP

O LCP resume-se em determinar um vetor em um espaço de dimensão finita (real), satisfazendo um certo sistema de desigualdades. Nomeadamente, dado um vetor $q \in \mathbb{R}^n$ e uma matriz $M \in \mathbb{R}^{n \times n}$, o problema de complementaridade linear consiste em encontrar um vetor $z \in \mathbb{R}^n$ tal que

$$z \geq 0 \tag{4.4}$$

$$q + Mz \geq 0 \tag{4.5}$$

$$z^T(q + Mz) = 0, \tag{4.6}$$

ou mostrar que não existe tal vetor z . Na literatura, o problema definido acima é normalmente denotado apenas por $\text{LCP}(q, M)$.

Um vetor z que satisfaz as desigualdades em (4.4) e (4.5) é dito viável. O conjunto de vetores viáveis do $\text{LCP}(q, M)$ é chamado de sua região viável, denotada por $\text{FEA}(q, M)$. Considere

$$w = q + Mz. \tag{4.7}$$

¹ Do inglês: Linear Complementarity Problem

Perceba que um vetor viável z do $\text{LCP}(q, M)$ satisfaz a condição (4.6) se, e somente se

$$z_i w_i = 0, \quad \text{para todo } i = 1, \dots, n. \quad (4.8)$$

A condição (4.8) é geralmente usada no lugar de (4.6). As variáveis z_i e w_i são conhecidas como par complementar, além disso, são chamadas complementares uma da outra. Seguindo essa nomenclatura, o vetor satisfazendo (4.8) é chamado complementar. Portanto, o LCP corresponde a encontrar um vetor que é tanto viável como complementar, esse vetor é chamado uma solução do LCP. Note ainda que se $q \geq 0$, então o $\text{LCP}(q, M)$ sempre tem solução, com o vetor nulo sendo uma solução trivial.

O vetor w tal como estabelecido anteriormente normalmente é usado em outra forma de expressar o $\text{LCP}(q, M)$, a saber, como o problema de encontrar vetores não negativos w e z em \mathbb{R}^n satisfazendo (4.7) e (4.8). Desse modo, uma formulação equivalente pode ser expressa como

$$w \geq 0, \quad z \geq 0 \quad (4.9)$$

$$w = q + Mz \quad (4.10)$$

$$z^T w = 0, \quad (4.11)$$

sendo essa maneira de representar o problema particularmente útil na discussão dos algoritmos para resolver LCP.

O contexto de programação quadrática configura um ambiente significativamente importante para problemas de complementaridade linear, pois em seus primórdios o LCP encontrava-se estreitamente relacionado ao estudo de programas lineares quadráticos, ver (COTTLE et al., 2009). Neste sentido, note que podemos ainda escrever o problema de complementaridade linear como um problema de minimização quadrático por meio da seguinte formulação

$$\begin{aligned} &\text{minimizar} && z^T(Mz + q) \\ &\text{sujeito a} && Mz + q \geq 0 \\ &&& z \geq 0. \end{aligned} \quad (4.12)$$

Observe que a função objetivo em (4.12) é limitada inferiormente por zero, no conjunto viável. Além disso, o vetor z é uma solução do $\text{LCP}(q, M)$ se, e somente se, é um minimizador global de (4.12) com o valor da função objetivo igual a zero.

Note ainda que nesses problemas as restrições de desigualdade se apresentam como parte da função a ser minimizada. Exigiremos algumas características da função objetivo a ser minimizada, como a convexidade, para os testes numéricos apresentados na próxima seção.

4.7.2 Experimentos numéricos LCP

Nesta seção apresentamos os experimentos numéricos preliminares referentes à resolução do problema de complementaridade linear. Utilizamos a metodologia presente no Capítulo 3, mais precisamente, o Algoritmo 9. Ao trabalharmos apenas com restrições lineares, a Hessiana aproximada, estabelecida em (3.22), coincide com a Hessiana da função Lagrangiano aumentado, pois, como é possível observar, não há termos de segunda ordem envolvendo as restrições de igualdade para serem desprezados.

O código disponível em (ALMQVIST, 2015), foi desenvolvido para resolver o problema de complementaridade linear associado ao problema de mecânica de contato. Além do algoritmo para resolver o problema, o autor também disponibilizou a metodologia para a geração dos dados de entrada que definem o $LCP(q, M)$, a qual adaptamos para gerar problemas cuja função objetivo seja convexa. Conforme alteramos a dimensão conseguimos gerar diferentes problemas, para os quais a matriz $M \in \mathbb{R}^{n \times n}$ é simétrica semidefinida positiva.

Relacionando o problema de mecânica de contato com a formulação matemática apresentada em (4.12), o termo $Mz + q$ está associado à folga existente entre dois corpos e z é a pressão de contato correspondente. Idealmente, buscamos determinar z de modo que $z^T(Mz + q) = 0$. Neste problema, a força entre dois objetos é complementar à distância entre eles, ver (BILLUPS; MURTY, 2000).

Os dados de entrada foram gerados para diferentes dimensões do problema de complementaridade linear por meio de uma rotina adaptada a partir de (ALMQVIST, 2015), os resultados obtidos estão resumidos na TABELA 13, na qual n representa a dimensão considerada; k o número de iterações que o método proposto utilizou; resolvemos cada problema 100 vezes para extrair MinT, MaxT e MedT, que são o tempo mínimo, máximo e médio, respectivamente; e por fim, comparamos a qualidade da solução por meio do valor da função objetivo no ponto z^k , encontrado pelo Algoritmo 9, com o ponto \bar{y} , encontrado utilizando a rotina *quadprog*.

TABELA 13 – TESTES LCP

| n | k | MinT | MaxT | MedT | $\ z^k - z^{k-1}\ $ | $\ Mz^k + q\ $ | $(z^k)^T(Mz^k + q)$ | $\bar{y}^T(M\bar{y} + q)$ |
|-----|-----|--------|--------|--------|---------------------|----------------|---------------------|---------------------------|
| 10 | 4 | 0.0049 | 0.1109 | 0.0107 | 2.73e-07 | 4.41e-15 | 6.28e-07 | 3.96e-06 |
| 50 | 5 | 0.0163 | 0.2050 | 0.0277 | 1.51e-07 | 3.50e-16 | 5.80e-15 | 8.26e-08 |
| 100 | 7 | 0.0622 | 0.2585 | 0.0765 | 3.33e-07 | 7.89e-16 | 1.08e-09 | 1.05e-08 |
| 150 | 9 | 0.1997 | 0.3959 | 0.2437 | 2.67e-08 | 9.49e-16 | 1.11e-12 | 1.43e-08 |
| 200 | 13 | 0.4914 | 0.9355 | 0.7013 | 4.73e-07 | 1.64e-15 | 3.02e-11 | 1.40e-08 |
| 250 | 10 | 0.5650 | 0.7159 | 0.6137 | 2.65e-07 | 1.96e-15 | 1.61e-09 | 1.62e-08 |
| 300 | 7 | 0.7202 | 1.0104 | 0.8706 | 1.84e-07 | 2.89e-15 | 3.52e-13 | 7.07e-09 |
| 350 | 12 | 1.2764 | 2.2038 | 1.6800 | 7.12e-07 | 3.64e-15 | -2.52e-13 | 2.63e-08 |
| 400 | 37 | 6.2142 | 7.1829 | 6.6539 | 2.63e-07 | 4.46e-15 | 2.09e-13 | 7.67e-09 |
| 450 | 8 | 1.3539 | 1.9524 | 1.5095 | 3.39e-07 | 6.95e-15 | 2.98e-12 | 8.11e-09 |
| 500 | 10 | 2.1015 | 3.0821 | 2.6366 | 5.36e-07 | 7.15e-15 | 1.08e-10 | 1.69e-08 |

FONTE: A autora.

A tolerância foi fixada em $\varepsilon = 10^{-6}$, consideramos os parâmetros auxiliares $\gamma = 2$, $\theta = 0.9$ e $k_0 = 50$. O vetor de multiplicadores de Lagrange inicial $\lambda^0 = (10, \dots, 10)^T \in \mathbb{R}^n$. A caixa $[\lambda_{\min}, \lambda_{\max}]^n$ foi determinada pelos vetores $\lambda_{\min} = -(10^{10}, \dots, 10^{10})^T$ e $\lambda_{\max} = (10^{10}, \dots, 10^{10})^T$ em \mathbb{R}^n . Estabelecemos o parâmetro de penalidade inicial $r = 10^9$, como é possível notar na TABELA 13, o número de iterações ficou abaixo de 50, portanto não houve atualização do parâmetro de penalidade nos problemas testados. As escolhas do parâmetro de penalidade e do vetor de multiplicadores de Lagrange iniciais foram realizadas de forma heurística.

Reforçamos que os resultados na TABELA 13 são referentes a estudos preliminares do problema de complementaridade linear, por conta disso não realizamos testes comparativos com algoritmos presentes na literatura específicos para resolver essa classe de problemas. A partir dos testes realizados percebemos que o algoritmo proposto é robusto, no sentido que resolve os problemas gerados. O próximo passo é realizar melhorias no método analisando formas mais eficientes de resolver o subproblema gerado no algoritmo, com o intuito de desenvolver um método competitivo frente a algoritmos consolidados.

CONCLUSÃO

Neste trabalho, propusemos uma classe de algoritmos baseados em Lagrangiano aumentado para resolver o problema não linear de alocação de recursos. Consideramos o caso particular conhecido como problema não linear da mochila, onde reduzimos o problema a uma sequência de subproblemas, porém ao invés de resolvermos cada subproblema restrito, como é usual quando se trata da filosofia de Lagrangiano aumentado, resolvemos o problema penalizando apenas a restrição de igualdade e projetando a solução na caixa. Destacamos que o custo computacional de projetar um ponto na caixa é baixo, relativamente aos demais cálculos executados pelos algoritmos propostos.

Como resultado, observamos que o algoritmo apresenta um bom desempenho em diversas aplicações envolvendo o problema não linear da mochila. Realizamos testes para os casos em que a função objetivo é quadrática e não quadrática, separável e não separável, e em todos os casos os subproblemas são resolvidos em tempos relativamente rápidos e o ponto limite alcançado é satisfatório. O algoritmo também se mostra bastante robusto, mantendo seu desempenho tanto quando a Hessiana da função Lagrangiano aumentado é constante ou variável.

No caso em que a função objetivo é quadrática, também conseguimos provar a convergência do método proposto sem a necessidade da Hipótese 2.1.1, gerando uma nova prova, independente da clássica teoria de Lagrangiano aumentado desenvolvida em (BIRGIN; MARTÍNEZ, 2014).

Consideramos extensões do problema em que trabalhamos o caso não linear de alocação de recursos com a função objetivo quadrática e m restrições lineares de igualdade. Propusemos um método específico para resolver o problema e conseguimos bons resultados teóricos e numéricos. Além disso, apresentamos uma generalização com a função objetivo convexa, e as funções que definem as restrições não necessariamente convexas. Neste caso consideramos uma aproximação da Hessiana da função Lagrangiano aumentado, desprezando os termos de segunda ordem das restrições, e resolvemos os subproblemas quadráticos gerados utilizando a rotina *quadprog*. Para a análise de convergência nos embasamos em (BIRGIN; MARTÍNEZ, 2014). Consideramos problemas do livro (HOCK; SCHITTKOWSKI, 1981) para realizar os testes com o algoritmo proposto, e concluímos a partir dos resultados que o método é de fato robusto.

Como trabalhos futuros, uma das propostas de continuidade de pesquisa consiste em estudar melhorias para o Algoritmo 3, especialmente visando estabelecer

estratégias que melhorem a resolução do subproblema gerado, que é o passo de maior custo computacional. No caso geral, em que trabalhamos com uma aproximação da Hessiana da função Lagrangiano aumentando desprezando o termo que envolve informações de segunda ordem das funções que definem as restrições de igualdade, vamos analisar situações em que a Hessiana da função h possa ser aproximada por alguma matriz com boas propriedades. Além disso, buscaremos aprimorar a resolução do subproblema gerado pelo método aplicado ao problema de complementaridade linear. Apresentamos os testes preliminares e embora o método resolva bem ainda há o que melhorar em termos de sua eficiência.

REFERÊNCIAS

ALMQVIST, A. **An LCP solution of the linear elastic contact mechanics problem.** [S.l.: s.n.], mai. 2015. DOI: [10.13140/RG.2.1.3960.7200](https://doi.org/10.13140/RG.2.1.3960.7200).

BAYÓN, L.; GRAU, J. M.; RUIZ, M. M.; SUÁREZ, P. M. An exact algorithm for the continuous quadratic knapsack problem via infimal convolution. In: I. ZELINKA, V. S.; ABRAHAM, A. (Ed.). **Handbook of Optimization: From Classical to Modern Approach.** Berlim: Springer-Verlag, 2013. v. 38. P. 97–127.

BAZARAA, M. S.; SHERALI, H. D.; SHETTY, C. M. **Nonlinear programming: theory and algorithms.** 3. ed. New Jersey: John Wiley & Sons, 2006. ISBN 0471486000.

BERTSEKAS, D. P. **Constrained optimization and Lagrange multiplier methods.** New York: Academic Press, 1982. ISBN 9780120934805.

_____. Projected Newton methods for optimization problems with simple constraints. **Journal on Control and Optimization**, Society for Industrial & Applied Mathematics (SIAM), v. 20, n. 2, p. 221–246, mar. 1982. DOI: [10.1137/0320018](https://doi.org/10.1137/0320018).

BILLUPS, S. C.; MURTY, K. G. Complementarity problems. **Journal of Computational and Applied Mathematics**, Elsevier BV, v. 124, n. 1-2, p. 303–318, dez. 2000. DOI: [10.1016/s0377-0427\(00\)00432-5](https://doi.org/10.1016/s0377-0427(00)00432-5).

BIRGIN, E. G.; MARTÍNEZ, J. M. **Practical augmented Lagrangian methods for constrained optimization.** Philadelphia, PA: Society for Industrial e Applied Mathematics, 2014. DOI: [10.1137/1.9781611973365](https://doi.org/10.1137/1.9781611973365).

BIRGIN, E. G.; FERNÁNDEZ, D.; MARTÍNEZ, J. M. The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. **Optimization Methods and Software**, Informa UK Limited, v. 27, n. 6, p. 1001–1024, dez. 2012. DOI: [10.1080/10556788.2011.556634](https://doi.org/10.1080/10556788.2011.556634).

BRETTAUER, K. M.; SHETTY, B. The nonlinear knapsack problem – algorithms and applications. **European Journal of Operational Research**, Elsevier BV, v. 138, n. 3, p. 459–472, mai. 2002. DOI: [10.1016/s0377-2217\(01\)00179-5](https://doi.org/10.1016/s0377-2217(01)00179-5).

BRETTTHAUER, K. M.; SHETTY, B. Quadratic resource allocation with generalized upper bounds. **Operations Research Letters**, Elsevier BV, v. 20, n. 2, p. 51–57, fev. 1997. DOI: [10.1016/s0167-6377\(96\)00039-9](https://doi.org/10.1016/s0167-6377(96)00039-9).

_____. The nonlinear resource allocation problem. **Operations Research**, Institute for Operations Research e the Management Sciences (INFORMS), v. 43, n. 4, p. 670–683, ago. 1995. DOI: [10.1287/opre.43.4.670](https://doi.org/10.1287/opre.43.4.670).

BRUCKER, P. An $O(n)$ algorithm for quadratic knapsack problems. **Operations Research Letters**, Elsevier BV, v. 3, n. 3, p. 163–166, ago. 1984. DOI: [10.1016/0167-6377\(84\)90010-5](https://doi.org/10.1016/0167-6377(84)90010-5).

BUENO, L. F.; HAESER, G.; SANTOS, L.-R. Towards an efficient augmented Lagrangian method for convex quadratic programming. **Computational Optimization and Applications**, Springer Science e Business Media LLC, v. 76, n. 3, p. 767–800, dez. 2019. DOI: [10.1007/s10589-019-00161-2](https://doi.org/10.1007/s10589-019-00161-2).

CAPRARA, A.; PISINGER, D.; TOTH, P. Exact solution of the quadratic knapsack problem. **Journal on Computing**, Institute for Operations Research e the Management Sciences (INFORMS), v. 11, n. 2, p. 125–137, mai. 1999. DOI: [10.1287/ijoc.11.2.125](https://doi.org/10.1287/ijoc.11.2.125).

COMINETTI, R.; MASCARENHAS, W. F.; SILVA, P. J. S. A Newton's method for the continuous quadratic knapsack problem. **Mathematical Programming Computation**, Springer Science e Business Media LLC, v. 6, n. 2, p. 151–169, fev. 2014. DOI: [10.1007/s12532-014-0066-y](https://doi.org/10.1007/s12532-014-0066-y).

COTTLE, R. W.; DANTZIG, G. B. A generalization of the linear complementarity problem. **Journal of Combinatorial Theory**, Elsevier BV, v. 8, n. 1, p. 79–90, jan. 1970. DOI: [10.1016/s0021-9800\(70\)80010-2](https://doi.org/10.1016/s0021-9800(70)80010-2).

COTTLE, R. W.; PANG, J.-S.; STONE, R. E. **The Linear Complementarity Problem**. Philadelphia, PA: Society for Industrial & Applied Mathematics (SIAM), 2009. ISBN 9780898716863.

DAI, Y. H.; FLETCHER, R. New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. **Mathematical Programming**, Springer Science e Business Media LLC, v. 106, n. 3, p. 403–421, mai. 2006. DOI: [10.1007/s10107-005-0595-2](https://doi.org/10.1007/s10107-005-0595-2).

DINIZ-EHRHARDT, M. A.; GOMES-RUGGIERO, M. A.; MARTÍNEZ, J. M.; SANTOS, S. A. Augmented Lagrangian algorithms based on the spectral projected gradient method for solving nonlinear programming problems. **Journal of Optimization Theory and Applications**, Springer Science e Business Media LLC, v. 123, n. 3, p. 497–517, dez. 2004. DOI: [10.1007/s10957-004-5720-5](https://doi.org/10.1007/s10957-004-5720-5).

DOLAN, E. D.; MOREÉ, J. J. Benchmarking optimization software with performance profiles. **Mathematical Programming**, Springer Science e Business Media LLC, v. 91, n. 2, p. 201–213, jan. 2002. DOI: [10.1007/s101070100263](https://doi.org/10.1007/s101070100263).

DU VAL, P. The unloading problem for plane curves. **American Journal of Mathematics**, JSTOR, v. 62, n. 1, p. 307–311, 1940. DOI: [10.2307/2371454](https://doi.org/10.2307/2371454).

DUSSAULT, J. P.; FERLAND, J. A.; LEMAIRE, B. Convex quadratic programming with one constraint and bounded variables. **Mathematical Programming**, Springer Science e Business Media LLC, v. 36, n. 1, p. 90–104, out. 1986. DOI: [10.1007/bf02591992](https://doi.org/10.1007/bf02591992).

FRANGIONI, A.; GORGONE, E. A library for continuous convex separable quadratic knapsack problems. **European Journal of Operational Research**, Elsevier BV, v. 229, n. 1, p. 37–40, ago. 2013. DOI: [10.1016/j.ejor.2013.02.038](https://doi.org/10.1016/j.ejor.2013.02.038).

GALLO, G.; HAMMER, P. L.; SIMEONE, B. Quadratic knapsack problems. In: PADBERG, M. W. (Ed.). **Mathematical Programming Studies**. Berlin, Heidelberg: Springer, 1980. v. 12. P. 132–149. DOI: [10.1007/bfb0120892](https://doi.org/10.1007/bfb0120892).

GERLA, M.; KLEINROCK, L. On the topological design of distributed computer networks. **IEEE Transactions on Communications**, Institute of Electrical e Electronics Engineers (IEEE), v. 25, n. 1, p. 48–60, jan. 1977. DOI: [10.1109/tcom.1977.1093709](https://doi.org/10.1109/tcom.1977.1093709).

HESTENES, M. R. Multiplier and gradient methods. **Journal of Optimization Theory and Applications**, Springer Science e Business Media LLC, v. 4, n. 5, p. 303–320, nov. 1969. DOI: [10.1007/bf00927673](https://doi.org/10.1007/bf00927673).

HOCK, W.; SCHITTKOWSKI, K. **Test Examples for Nonlinear Programming Codes**. New York: Springer-Verlag Berlin Heidelberg, 1981. ISBN 3540105611.

HOTO, R.; MATIOLI, L.; SANTOS, P. A penalty algorithm for solving convex separable knapsack problems. **Applied Mathematics and Computation**, Elsevier BV, v. 387, p. 124855, dez. 2020. DOI: [10.1016/j.amc.2019.124855](https://doi.org/10.1016/j.amc.2019.124855).

IZMAILOV, A.; SOLODOV, M. **Otimização: condições de otimalidade, elementos de análise convexa e de dualidade**. 3. ed. Rio de Janeiro: IMPA, 2014. v. 1. ISBN 9788524403897.

_____. **Otimização: métodos computacionais**. 2. ed. Rio de Janeiro: IMPA, 2012. v. 2. ISBN 9788524402685.

KIWIEL, K. C. Breakpoint searching algorithms for the continuous quadratic knapsack problem. **Mathematical Programming**, Springer Science e Business Media LLC, v. 112, n. 2, p. 473–491, abr. 2008. DOI: [10.1007/s10107-006-0050-z](https://doi.org/10.1007/s10107-006-0050-z).

_____. Variable fixing algorithms for the continuous quadratic knapsack problem. **Journal of Optimization Theory and Applications**, Springer Science e Business Media LLC, v. 136, n. 3, p. 445–458, mar. 2008. DOI: [10.1007/s10957-007-9317-7](https://doi.org/10.1007/s10957-007-9317-7).

KLASTORIN, T. On a discrete nonlinear and nonseparable knapsack problem. **Operations Research Letters**, Elsevier BV, v. 9, n. 4, p. 233–237, jul. 1990. DOI: [10.1016/0167-6377\(90\)90067-f](https://doi.org/10.1016/0167-6377(90)90067-f).

MARCILIO, D. C. **Otimização do despacho hidrotérmico utilizando o método de Lagrangeano aumentado com gradiente espectral projetado**. Tese (Doutorado) – Setores de Tecnologia e de Ciências Exatas, Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná. Curitiba-PR, p. 122. 2014.

MARKOWITZ, H. Portfolio selection. **The Journal of Finance**, JSTOR, v. 7, n. 1, p. 77–91, mar. 1952. DOI: [10.2307/2975974](https://doi.org/10.2307/2975974).

MORÉ, J. J.; VAVASIS, S. A. On the solution of concave knapsack problems. **Mathematical Programming**, Springer Science e Business Media LLC, v. 49, p. 397–411, nov. 1990. DOI: [10.1007/bf01588800](https://doi.org/10.1007/bf01588800).

MÜNNICH, R. T.; SACHS, E. W.; WAGNER, M. Numerical solution of optimal allocation problems in stratified sampling under box constraints. **AStA Advances in Statistical Analysis**, Springer Science e Business Media LLC, v. 96, n. 3, p. 435–450, jul. 2012. DOI: [10.1007/s10182-011-0176-z](https://doi.org/10.1007/s10182-011-0176-z).

NIELSEN, S. S.; ZENIOS, S. A. Massively parallel algorithms for singly constrained convex programs. **ORSA Journal on Computing**, Institute for Operations Research e

the Management Sciences (INFORMS), v. 4, n. 2, p. 166–181, mai. 1992. DOI: [10.1287/ijoc.4.2.166](https://doi.org/10.1287/ijoc.4.2.166).

NOCEDAL, J.; WRIGHT, S. **Numerical optimization**. 2. ed. New York: Springer Series in Operations Research and Financial Engineering, set. 2006. ISBN 0387303030.

PADAKANDLA, A.; SUNDARESAN, R. Separable convex optimization problems with linear ascending constraints. **Journal on Optimization**, Society for Industrial & Applied Mathematics (SIAM), v. 20, n. 3, p. 1185–1204, ago. 2009. DOI: [10.1137/07069729x](https://doi.org/10.1137/07069729x).

PANG, J. S. A new and efficient algorithm for a class of portfolio selection problems. **Operations Research**, Institute for Operations Research e the Management Sciences (INFORMS), v. 28, n. 3, p. 754–767, jun. 1980. DOI: [10.1287/opre.28.3.754](https://doi.org/10.1287/opre.28.3.754).

PARDALOS, P. M.; KOVOOR, N. An algorithm for a singly constrained class of quadratic programs subject to upper and lower bounds. **Mathematical Programming**, Springer Science e Business Media LLC, v. 46, n. 1-3, p. 321–328, jan. 1990. DOI: [10.1007/bf01585748](https://doi.org/10.1007/bf01585748).

PATRIKSSON, M. A survey on the continuous nonlinear resource allocation problem. **European Journal of Operational Research**, Elsevier BV, v. 185, n. 1, p. 1–46, fev. 2008. DOI: [10.1016/j.ejor.2006.12.006](https://doi.org/10.1016/j.ejor.2006.12.006).

PATRIKSSON, M.; STRÖMBERG, C. Algorithms for the continuous nonlinear resource allocation problem—New implementations and numerical studies. **European Journal of Operational Research**, Elsevier BV, v. 243, n. 3, p. 703–722, jun. 2015. DOI: [10.1016/j.ejor.2015.01.029](https://doi.org/10.1016/j.ejor.2015.01.029).

POWELL, M. J. D. A method for nonlinear constraints in minimization problems, in Optimization. In: FLETCHER, R. (Ed.). New York: Academic Press, 1969. P. 283–298.

RIBEIRO, A. A.; KARAS, E. W. **Otimização contínua : aspectos teóricos e computacionais**. São Paulo: Cengage Learning, 2013. ISBN 9788522115013.

ROBINSON, A. G.; JIANG, N.; LERME, C. S. On the continuous quadratic knapsack problem. **Mathematical Programming**, Springer Science e Business Media LLC, v. 55, n. 1-3, p. 99–108, abr. 1992. DOI: [10.1007/bf01581193](https://doi.org/10.1007/bf01581193).

SHERMAN, J.; MORRISON, W. J. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. **The Annals of Mathematical Statistics**, v. 21, n. 1, p. 124–127, mar. 1950.

SMITH, S.; LASDON, L. Solving large sparse nonlinear programs using GRG. **ORSA Journal on Computing**, Institute for Operations Research and the Management Sciences (INFORMS), v. 4, n. 1, p. 2–15, fev. 1992. DOI: [10.1287/ijoc.4.1.2](https://doi.org/10.1287/ijoc.4.1.2).

TORREALBA, E.; SILVA, J.; MATIOLI, L.; KOLOSSOSKI, O.; SANTOS, P. Augmented Lagrangian algorithms for solving the continuous nonlinear resource allocation problem. **European Journal of Operational Research**, Elsevier BV, v. 299, n. 1, p. 46–59, mai. 2022. DOI: [10.1016/j.ejor.2021.11.027](https://doi.org/10.1016/j.ejor.2021.11.027).

VIDAL, T.; JAILLET, P.; MACULAN, N. A decomposition algorithm for nested resource allocation problems. **Journal on Optimization**, Society for Industrial & Applied Mathematics (SIAM), v. 26, n. 2, p. 1322–1340, jan. 2016. DOI: [10.1137/140965119](https://doi.org/10.1137/140965119).

WOODBURY, M. A. **Inverting modified matrices**. Technical Report 42, Statistical Research Group, Princeton University, Princeton, NJ, 1950.

WRIGHT, S. E.; LIM, S. Solving nested-constraint resource allocation problems with an interior point method. **Operations Research Letters**, Elsevier BV, v. 48, n. 3, p. 297–303, mai. 2020. DOI: [10.1016/j.orl.2020.04.001](https://doi.org/10.1016/j.orl.2020.04.001).

ZIEGLER, H. Solving certain singly constrained convex optimization problems in production planning. **Operations Research Letters**, Elsevier BV, v. 1, n. 6, p. 246–252, dez. 1982. DOI: [10.1016/0167-6377\(82\)90030-x](https://doi.org/10.1016/0167-6377(82)90030-x).