

UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Algoritmos Quase-Newton para Otimização Multiobjetivo

Osenildo Marques Maciel

Manaus - 2016

UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Algoritmos Quase-Newton para Otimização Multiobjetivo

Osenildo Marques Maciel

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Matemática da UFAM como requisito para a obtenção parcial do grau de MESTRE em Matemática na área de concentração OTIMIZAÇÃO.

Orientador: Prof. Dr. Sandro Dimy
Barbosa Bitar

Manaus - 2016

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

M152a Maciel, Osenildo Marques
Algoritmos Quase-Newton para Otimização Multiobjetivo /
Osenildo Marques Maciel. 2016
55 f.: il.; 31 cm.

Orientador: Sandro Dimy Barbosa Bitar
Dissertação (Mestrado em Matemática - Otimização) -
Universidade Federal do Amazonas.

1. Multiobjetivo. 2. Pareto. 3. Quase-Newton. 4. Convexas. 5. Não convexas. I. Bitar, Sandro Dimy Barbosa II. Universidade Federal do Amazonas III. Título

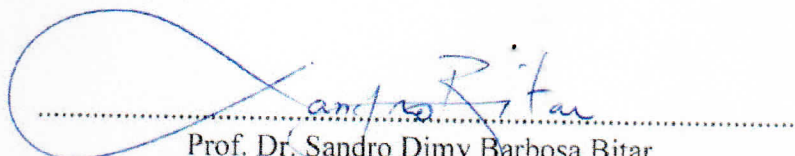
OSENILDO MARQUES MACIEL

ALGORITMOS QUASE-NEWTON PARA OTIMIZAÇÃO MULTIOBJETIVO

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Matemática, área de concentração em Matemática Aplicada.

Aprovado por:

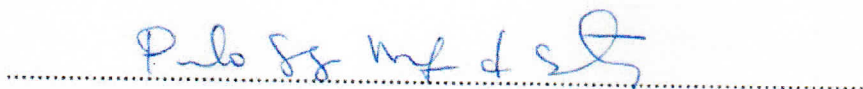
BANCA EXAMINADORA



Prof. Dr. Sandro Dimy Barbosa Bitar
Universidade Federal do Amazonas – UFAM
Orientador



Prof. Dr. Roberto Cristovão Mesquita Silva
Universidade Federal do Amazonas – UFAM



Prof. Dr. Paulo Sérgio Marques dos Santos
Universidade Federal do Piauí – UFPI

*Este trabalho é dedicado às Professoras Terezinha
de Jesus Maciel e Raimunda Nonata Maciel
...minhas tias, minhas mães, meu tudo...*

Agradecimentos

Em primeiro lugar, agradeço a Deus e a Jesus Cristo, pois tinha certeza que atendiam minhas orações;

Aos meus incansáveis professores de mestrado, Professora Morgana, Professor Roberto, Professor Alfredo, Professor Anselmo, Professora Sheila e aquele que aprimorou sua paciência, meu orientador Professor Sandro;

Aos meus amigos de mestrado, em especial Marcinha, Dani, Ezequiel, Suellem, Fábio e Felipe. Aos amigos da secretaria, em especial ao Ari. Ao Raimundo e Abraão, meus sinceros agradecimentos pelas orientações quanto aos exercícios;

À minha família, Ivane, Helen, Vitória e Diego. A cada momento que deixei de participar da vida de vocês, para dedicar-me aos estudos, sempre pensei em vocês. Vocês foram minha energia, minha inspiração;

À Fapeam e Semed, pelo suporte financeiro;

E a todos que contribuíram direta e indiretamente para a realização deste trabalho.

Resumo

Neste trabalho, apresentam-se caracterizações de soluções para Otimização Multiobjetivo Irrestrita para os casos de funções convexas e não convexas. A fundamentação teórica do caso convexo discorre sobre uma solução local, obtida através da resolução de um problema convexo e algumas hipóteses adicionais. Para o caso não convexo, mostramos que o algoritmo tem convergência global, no qual os fundamentos teóricos asseguram que a condição de curvatura é obtida.

Palavras-chaves: Otimização Multiobjetivo, Condições de Otimalidade Pareto-ótimo, Método Quase-Newton BFGS, Funções Convexas e não convexas.

Abstract

In this work, characterization are presented solutions for unconstrained multiobjective optimization for the cases of convex and non-convex function. The theoretical foundation of the convex case discusses a local solution obtained by solving a convex problem and some additional assumptions. For nonconvex case we show that the algorithm have a global convergence, in which the theoretical foundations ensure that curvature condition is obtained.

Keywords: Multiobjective Optimization, optimality conditions Pareto-optimal, Quasi-Newton BFGS method, Convex functions and not convex.

Sumário

1	Fundamentos de Otimização	1
1.1	Problema Geral	1
1.2	Elementos de Análise convexa e algumas definições	2
1.3	Métodos Quase-Newton	8
1.3.1	O método BFGS	8
1.3.2	Convergência do Método Quase-Newton BFGS	14
2	Fundamentos de Otimização Multiobjetivo	18
2.1	Problema Geral	18
2.2	Equivalências entre alguns problemas de Otimização	18
2.3	Ordenamento de soluções no \mathbb{R}^n	20
2.4	Pareto Otimalidade	23
2.5	Um resultado sobre convexidade em Otimização Multiobjetivo	26
3	Alguns Métodos do tipo Quase-Newton para Otimização Multiobjetivo	28
3.1	Algumas definições	28
3.1.1	Direção de descida	28
3.1.2	Algumas definições para o caso convexo	29
3.2	Um algoritmo do tipo Quase-Newton para Otimização Multiobjetivo (caso convexo)	40
3.3	Um algoritmo Quase-Newton para Otimização Multiobjetivo (caso não convexo)	48
4	Considerações Finais	53
	Referências	54

Introdução

Historicamente, a teoria Multiobjetivo (ou Vetorial) teve início a partir de pesquisas do economista britânico Francis Ysidro Edgeworth (1845-1926), que introduziu a ideia de otimalidade em 1881 [1] que depois foi generalizado pelo engenheiro italiano Vilfredo Pareto (1848-1923) na obra "Cours d'économie politique (1896-1897) [2]. Pareto lançou nesta obra a ideia de que uma sociedade está em estado ótimo se o bem estar de um indivíduo acarreta na piora do bem estar de outro indivíduo. Este conceito chama-se Ótimo Pareto, utilizada para o estudo da Otimização Multiobjetivo. A Otimização Multiobjetivo (ou Vetorial), embora tenha suas origens no século XIX, tornou-se uma ferramenta prática de amplo alcance apenas no últimos anos do século XX, em parte devido ao aumento computacional disponível.

Seguindo a mesma linha de pesquisa explorada por trabalhos anteriores, nos últimos 16 anos, encontramos alguns artigos que utilizam basicamente a mesma ideia (caso convexo). Em 2000 [3], os autores apresentaram um algoritmo utilizando o método de Newton, no qual seu embasamento teórico principal eram os cones. Em 2009, [4], os autores também apresentaram um algoritmo com convergência local, utilizando o método de Newton e o cone paretiano. Em 2014, [5] apresentaram um trabalho utilizando o método de Newton e cones pontiagudos. Também em 2014, Ziga Povalej [6], segue o embasamento teórico de [4], e implementa um algoritmo usando um método Quase-Newton. Para o caso não convexo, em 2001, utilizando o método Quase-Newton BFGS, [7] propõe um sistema de atualização da matriz aproximação da Hessiana, no qual garante que a condição de curvatura não será alterada. E finalmente em 2011, Shaojian Qu^a, Mark Goh e Felix Chan [8], propuseram uma tolerância no sistema proposto por Fukushima em [7] e com algumas hipóteses adicionais, implementam um algoritmo com convergência global.

Nesta dissertação, abordam-se as condições de otimalidade para problemas de Otimização Multiobjetivo com funções convexas segundo [6] e não convexas segundo [8].

Objetivos

Objetivo Geral

Apresentar condições de otimalidade para problemas de Otimização Multi-objetivo para funções convexas com convergência local e funções não convexas com convergência global.

Objetivos Específicos

Esta dissertação têm alguns objetivos específicos, tais como, destacar elementos de Otimização Multiobjetivo, apresentar resultados sobre as condições de Otimalidade para problemas de Otimização convexo e não convexo.

Descrição da dissertação

Este trabalho está organizado em três capítulos. O capítulo 1 trata de alguns elementos matemáticos e resultados da Otimização, assim como alguns resultados envolvendo o método Quase-Newton, com o uso da fórmula BFGS. No capítulo 2, mostram-se os fundamentos da Otimização Multiobjetivo. No capítulo 3, apresentam-se alguns resultados sobre Otimalidade de problemas Multiobjetivo para funções convexas, bem com sua ordem de convergência e também para funções não convexas e sua convergência global. Por último, as perspectivas de trabalhos futuros são apresentados na conclusão deste trabalho.

1 Fundamentos de Otimização

Neste capítulo apresentam-se algumas definições, teoremas e proposições que serão utilizados no desenvolvimento deste trabalho. Na seção 1.1, apresentamos conceitos e resultados relacionados à Otimização. Na seção 1.2 trabalhamos com alguns conceitos de convexidade de funções. Na seção 1.3, apresentamos alguns conceitos e explicitamos o método Quase Newton BFGS, bem como sua convergência.

1.1 Problema Geral

A Otimização Clássica estuda problemas do tipo

$$\begin{cases} \min f(x) \\ \text{s.a. } x \in D \end{cases} \quad (1.1)$$

onde $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Para tanto, algumas hipóteses podem ser assumidas para f . Nesse problema, o conjunto D é o conjunto dos pontos de interesse ou pontos viáveis. Em problemas irrestritos, assume-se que $D = \mathbb{R}^n$.

As definições a seguir são encontradas em [9]

Definição 1.1.1. Dizemos que um ponto $\bar{x} \in D$ é

1. *Minimizador global do problema 1.1, se*

$$f(\bar{x}) \leq f(x), \quad \forall x \in D$$

2. *Minimizador local do problema 1.1, se existe uma vizinhança U de \bar{x} tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in D \cap U$$

Se para todo $x \neq \bar{x}$ as desigualdades acima forem estritas, \bar{x} será chamado de minimizador estrito global ou local respectivamente.

Teorema 1.1.1 (Teorema de Weierstrass). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função contínua e $X \subset \mathbb{R}^n$ um conjunto compacto (limitado e fechado no \mathbb{R}^n) e não vazio. Então f admite um minimizador global em X .*

Demonstração. Veja [9]. □

Corolário 1.1.1. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função contínua e suponhamos que existe $c \in \mathbb{R}$ tal que o conjunto de nível $L_c = \{x \in \mathbb{R}^n; f(x) \leq c\}$ seja compacto e não vazio. Então f tem um minimizador global.*

Demonstração. Veja [9]. □

1.2 Elementos de Análise convexa e algumas definições

Nesta seção apresentaremos alguns conceitos de classes de funções envolvendo convexidade, utilizando as referências [9], [10] e [11].

Definição 1.2.1. *Um conjunto $D \subset \mathbb{R}^n$ é chamado de conjunto convexo, se para quaisquer $x, y \in D$ e $\alpha \in [0, 1]$, tem-se que $\alpha x + (1 - \alpha)y \in D$. O ponto $\alpha x + (1 - \alpha)y$, onde $\alpha \in [0, 1]$, chama-se combinação convexa de x e y , com parâmetro α*

Definição 1.2.2. *Seja a função $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, no qual D é um conjunto convexo. Dizemos que a função f é:*

1. **Convexa** em D , quando para quaisquer $x, y \in D$ e $\alpha \in [0, 1]$ tem-se

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

2. **Estritamente Convexa** em D , quando para quaisquer $x, y \in D$, com $x \neq y$ e $\alpha \in (0, 1)$ tem-se

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y).$$

3. **Fortemente Convexa** em D , com módulo $\gamma > 0$, quando para quaisquer $x, y \in D$ e $\alpha \in [0, 1]$, tem-se

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \gamma \alpha(1 - \alpha)\|x - y\|^2.$$

Definição 1.2.3. *Seja a função $g : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, onde D é um conjunto convexo. Dizemos que a função g é convexa, quando $g = (g_1, \dots, g_m)^T : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, é convexa.*

Definição 1.2.4. Dizemos que

$$\min f(x) \text{ sujeito a } x \in D$$

é um problema de minimização convexo quando $D \subset \mathbb{R}^n$ é um conjunto convexo e $f : D \rightarrow \mathbb{R}$ é uma função convexa no conjunto D .

Teorema 1.2.1 (Teorema de minimização convexa). *Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, no qual D é um conjunto convexo. Então, todo minimizador local do problema da definição 1.2.4 é global. Além disso, o conjunto de minimizadores é convexo. Se f é estritamente convexa, o minimizador é único.*

Demonstração. Veja [12]. □

Teorema 1.2.2 (Caracterização de funções convexas diferenciáveis). *Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ e f duas vezes diferenciável em D , no qual D é um conjunto convexo e aberto. As propriedades seguintes são equivalentes.*

1. A função f é convexa no conjunto D

2. Para todo $x, y \in D$,

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

3. Para todo $x, y \in D$

$$(\nabla f(y) - \nabla f(x))^T(y - x) \geq 0$$

4. A matriz Hessiana de f é semidefinida positiva em todo ponto de D , i.e.,

$$d^T \nabla^2 f(x) d \geq 0, \forall x \in \mathbb{R}^n$$

Demonstração. Veja [12] □

Teorema 1.2.3 (Caracterização de funções convexas diferenciáveis estritas). *Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ e f duas vezes diferenciável no conjunto convexo e aberto D . As propriedades seguintes são equivalentes.*

1. A função f é estritamente convexa no conjunto D

2. Para todo $x, y \in D$, tais que $x \neq y$

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

3. Para todo $x, y \in D$, tais que $x \neq y$

$$(\nabla f(y) - \nabla f(x))^T (y - x) > 0$$

4. A matriz Hessiana de f é definida positiva em todo ponto de D , i.e.,

$$d^T \nabla^2(x) d > 0, \forall x \in \mathbb{R}^n \setminus \{0\}.$$

Demonstração. Veja [12]. □

Abaixo apresentamos duas definições que segundo [13], [14] e [15], formalizam os conceitos de funções uniformemente convexas.

Definição 1.2.5. Dizemos que uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uniformemente convexa com módulo $\varsigma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, se ς é crescente, se anula somente em 0, e além disso,

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

para todo $x, y \in \mathbb{R}^n$ e $t \in (0, 1)$.

Definição 1.2.6. Dizemos que uma função f diferenciável é uniformemente convexa se existem constantes positivas α e β tais que

$$\alpha\|x-y\|^2 \leq (x-y)^T (\nabla f(x) - \nabla f(y)) \leq \beta\|x-y\|^2$$

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

Os teoremas do Valor Intermediário e do Valor Médio e as fórmulas de Taylor, são resultados clássicos da Análise Matemática, utilizados em grande parte nas demonstrações de convergência. Por suas importâncias, apresentamo-los a seguir.

Teorema 1.2.4 (Teorema do Valor Intermediário). *Seja $f : X \rightarrow \mathbb{R}$ uma função contínua definida em um conjunto conexo $X \subset \mathbb{R}^n$. Se existirem $m, n \in X$ e $d \in \mathbb{R}$ tais que $f(m) < d < f(n)$, então existe $e \in X$ tal que $f(e) = d$.*

Demonstração. Veja [10]. □

Teorema 1.2.5 (Fórmula de Taylor de Primeira Ordem). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável e $x, h \in \mathbb{R}^n$. Então*

$$f(x+h) = f(x) + \nabla f(x)^T h + o(h)$$

com $\lim_{h \rightarrow 0} \frac{o(h)}{\|h\|} = 0$.

Demonstração. Veja [10]. □

O polinômio $p(x) = f(x) + \nabla f(x)^T h$ é denominado de Taylor de primeira ordem.

Teorema 1.2.6 (Fórmula de Taylor de Segunda Ordem). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função duas vezes diferenciável e $x, h \in \mathbb{R}^n$. Então*

$$f(x+h) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x) h + o^2(h)$$

$$\text{com } \lim_{h \rightarrow 0} \frac{o^2(h)}{\|h\|^2} = 0$$

Demonstração. Veja [10]. □

O polinômio $p(x) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x) h$ é chamado polinômio de Taylor de segunda ordem.

Teorema 1.2.7 (Teorema do Valor Médio com resto integral). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável e $x, h \in \mathbb{R}^n$. Então*

$$f(x+h) - f(x) = \int_0^1 \nabla f(x+th)^T h dt$$

Mais ainda, se f é duas vezes diferenciável, temos que

$$\nabla f(x+h) - \nabla f(x) = \int_0^1 \nabla^2 f(x+th)^T h dt \quad (1.2)$$

Demonstração. Veja [10]. □

Os seguintes teoremas irão nos ajudar no entendimento da convergência local de alguns problemas de Otimização. Apresentamos a seguir.

Teorema 1.2.8. *Sejam $\|\cdot\|_1$ e $\|\cdot\|_2$ duas normas quaisquer no \mathbb{R}^n . Existem constantes positivas α e β tal que*

$$\alpha \|v\|_1 \leq \|v\|_2 \leq \beta \|v\|_1$$

para quaisquer $v \in \mathbb{R}^n$.

Demonstração. Ver [16] □

Teorema 1.2.9. *Sejam $\|\cdot\|_1$ e $\|\cdot\|_2$ duas normas quaisquer em $\mathbb{R}^{n \times n}$. Existem constantes positivas α e β tal que*

$$\alpha \|A\|_1 \leq \|A\|_2 \leq \beta \|A\|_1, \quad \forall A \in \mathbb{R}^{n \times n}$$

Demonstração. Ver [16] □

A seguir, mostramos definições e teoremas segundo [17] sobre Continuidade Uniforme e Conjuntos de funções Equicontínuas, que nos auxiliam no entendimento do trabalho.

Definição 1.2.7. *Uma função $f : X \rightarrow \mathbb{R}$ diz-se Uniformemente Contínua quando, para cada $\epsilon > 0$ tal que $x, y \in X$,*

$$\|x - y\| < \delta \Rightarrow \|f(x) - f(y)\| < \epsilon$$

Teorema 1.2.10. *Seja X um conjunto compacto. Toda função contínua $f : X \rightarrow \mathbb{R}$ é Uniformemente Contínua.*

Demonstração. Ver [17] □

Definição 1.2.8. *Seja E um conjunto de funções $f : X \rightarrow \mathbb{R}$ todas com o mesmo domínio $X \subset \mathbb{R}$. Dado $x_0 \in \mathbb{R}$, diremos que o conjunto E é equicontínuo no ponto x_0 quando, dado arbitrariamente $\epsilon > 0$ existir $\delta > 0$ tal que*

$$x \in X, \|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| < \epsilon$$

qualquer que seja $f \in E$. O número δ escolhido a partir do ϵ é o mesmo para todas as funções f do conjunto E .

Definição 1.2.9. *Um conjunto E de funções $f : X \rightarrow \mathbb{R}$ chama-se Equicontínuo quando E é Equicontínuo em todos os pontos $x_0 \in X$.*

Definição 1.2.10. *Um conjunto E de funções $f : X \rightarrow \mathbb{R}$ chama-se Uniformemente Equicontínuo quando, para cada $\epsilon > 0$ dado, existe $\delta > 0$ tal que $x, y \in X$, $\|x - y\| < \delta \Rightarrow \|f(x) - f(y)\| < \epsilon$, seja qual for $f \in E$.*

Teorema 1.2.11. *Seja $K \subset \mathbb{R}$ um conjunto compacto. Todo conjunto Equicontínuo de funções $f : K \rightarrow \mathbb{R}$ é Uniformemente Equicontínuo.*

Demonstração. Ver [17] □

As condições de Otimalidade de KKT são amplamente utilizados em resoluções de problemas de Otimização. Abaixo mostramos suas condições de regularidade e no

final o Teorema de KKT, encontrado em [12].

Seja o problema de Otimização com restrições de igualdade e desigualdade

$$\begin{cases} \min f(x) \\ \text{s.a. } x \in D \end{cases} \quad (1.3)$$

$D = \{x \in \mathbb{R}^n; h(x) = 0, g(x) \leq 0\}$ onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções dadas.

Definição 1.2.11 (Condição de Linearidade). *Dizemos que o problema (1.3) satisfaz a Condição de Linearidade quando as funções h e g são funções afins, i.e.,*

$$\begin{aligned} h(x) &= Ax - a, \quad A \in \mathbb{R}(l, n), \quad a \in \mathbb{R}^l \\ g(x) &= Bx - b, \quad B \in \mathbb{R}(m, n), \quad b \in \mathbb{R}^m \end{aligned} \quad (1.4)$$

Definição 1.2.12 (Condição de Regularidade de Mangasarian-Fromovitz). *Dizemos que o ponto $\bar{x} \in D$ satisfaz a Condição de Regularidade de Mangasarian-Fromovitz, quando*

$$\{\nabla h_i(\bar{x}), i = 1, \dots, l\} \quad (1.5)$$

é um conjunto linearmente independente e existe $\bar{d} \in \text{Ker } \nabla f(\bar{x})$, tal que $\nabla f_i(\bar{x})^T \bar{d} < 0$, $\forall i \in I(\bar{x})$, no qual $I(\bar{x}) = \{i = 1, \dots, m; g_i(\bar{x}) = 0\}$ é o conjunto dos índices das restrições ativas no ponto $\bar{x} \in D$

Definição 1.2.13 (Condição de Regularidade de Slater). *Dizemos que o ponto $\hat{x} \in D$ satisfaz a condição de Slater se*

$$h(\hat{x}) = 0 \text{ e } g_i(\hat{x}) < 0, \quad i = 1, \dots, m \quad (1.6)$$

no qual h é uma função afim e g é uma função convexa.

Teorema 1.2.12 (Condições de Karush-Kuhn-Tucker). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ funções diferenciáveis no ponto $\bar{x} \in \mathbb{R}^n$, e seja $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$ uma função diferenciável numa vizinhança do ponto \bar{x} , com derivada contínua nesse ponto. Seja \bar{x} um minimizador local do problema (1.3). Então, sob qualquer uma das três condições de regularidade das restrições (de linearidade (1.4), de Mangasarian-Fromovitz (1.5) ou de Slater (1.6)), tem-se que existem $\bar{\lambda} \in \mathbb{R}^l$ e $\bar{\mu} \in \mathbb{R}_+^m$, tais que*

$$L'_x(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$$

$$\bar{\mu}_i g_i(\bar{x}) = 0, \quad i = 1, \dots, m$$

Demonstração. Ver [12]

□

1.3 Métodos Quase-Newton

Nesta seção, apresenta-se o método que servirá de alicerce para o entendimento deste trabalho: o método Quase-Newton BFGS. Historicamente, o início das pesquisas sobre método Quase-Newton é bastante controverso, porém a base mais simples, são as aproximações secantes. Alguns métodos Quase-Newton, que foram estudados e divulgados por Powell e Davidon em meados de 1960, visam melhorar o desempenho do método Gradiente e baratear o custo do método de Newton. Segundo [11] o método de Newton comporta-se melhor que o método Gradiente, com respeito à ordem de convergência e os métodos Quase-Newton procuram não ficar distantes da ordem de convergência do método de Newton. Como o cálculo da matriz Hessiana no algoritmo de Newton é caro computacionalmente, Quase-Newton procura a aproximação para a matriz Hessiana. Chamamos estes cálculos de atualizações de uma matriz B_k para aproximar-se da matriz Hessiana $\nabla^2 f(x)$. Um dos métodos clássicos que atualizam estas B'_k s é o método DFP (Davidon, Fletcher e Powell). Para maiores esclarecimentos deste método, consulte [18], [19] e [20].

O método BFGS (Broyden, Fletcher, Goldfarb e Shanno) possui todas as características teóricas do método DFP, porém os resultados implementares são superiores. A diferença entre os métodos consiste nas suas atualizações. Segundo [11] o método DFP realiza a atualização buscando a inversa da Hessiana, enquanto que o BFGS atualiza a própria Hessiana. Sobre o método BFGS, neste capítulo, serão provadas, teoricamente suas funcionalidades e por fim, o estudo da ordem de convergência.

1.3.1 O método BFGS

O método BFGS, segundo [18], realiza as atualizações usando as informações do gradiente da função f que por sua vez, indicam as informações da matriz Hessiana no decorrer da direção de busca. Defina o seguinte modelo quadrático convexo

$$m_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d \quad (1.7)$$

onde B_k é uma matriz $n \times n$ simétrica, definida positiva e qua será atualizada a cada iteração. O modelo $m_k(d)$ e $\nabla m_k(d)$, com $d = 0$, coincidem com $f(x^k)$ e $\nabla f(x^k)$ respectivamente. De fato, com $d = 0$ na equação 1.7, temos

$$m_k(0) = f(x^k)$$

e calculando o gradiente, com $d = 0$, em (1.7), obtemos

$$\nabla m_k(0) = \nabla f(x^k)$$

O minimizador do modelo quadrático (1.7) é

$$d_k = -B_k^{-1} \nabla f(x^k)$$

que irá ser a nova direção de busca do iterado $x^{k+1} = x^k + \alpha_k d_k$. Supondo que um novo iterado x^{k+1} foi gerado, construímos um novo modelo quadrático, da forma

$$m_{k+1}(d) = f(x^{k+1}) + \nabla f(x^{k+1})^T d + \frac{1}{2} d^T B_{k+1} d$$

Um dos requisitos para uma boa aproximação para a função f é que o gradiente de m_{k+1} coincida com o gradiente da função objetivo f nas duas últimas iterações x^k e x^{k+1} . Verificamos que $\nabla m_{k+1}(0) = \nabla f(x^{k+1})$ e que também

$$\nabla m_{k+1}(-\alpha_k d_k) = \nabla f(x^{k+1}) - \alpha_k B_{k+1} d_k = \nabla f(x^k)$$

ou seja,

$$B_{k+1} \alpha_k d_k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

$$B_{k+1} s^k = y^k \tag{1.8}$$

onde

$$s^k = x^{k+1} - x^k \text{ e } y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

A equação (1.8) é chamada de Secante.

Algoritmo 1.3.1 (Algoritmo Geral dos Métodos Quase-Newton).

- **Passo 1.** Dados $x_0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n \times n}$ (simétrica e definida positiva) e $\epsilon \in [0, 1)$
- **Passo 2.** Se $\|\nabla f^k\| \leq \epsilon$, PARE.
- **Passo 3.** Calcule $d^k = -B_k \nabla f^k$
- **Passo 4.** Encontre o tamanho do passo $\alpha_k > 0$ com uma busca linear e defina $x^{k+1} := x^k + \alpha_k d^k$
- **Passo 5.** Atualize B_k para B_{k+1} tal que satisfaça a equação secante (1.8).
- **Passo 6.** Faça $k := k + 1$ é vá para o Passo 2.

Segundo [15] e [11], diferentes soluções de (1.8) originam diferentes métodos. Estamos interessados em soluções facilmente computáveis, ou seja, que realize pouco esforço computacional. Considere $B_k \in \mathbb{R}^{n \times n}$ definida positiva e $s_k^T y_k > 0$. Vamos

procurar uma matriz atualização B_{k+1} para (1.8) com posto dois a partir da matriz B_k . Deste modo, devem existir escalares $a_k, b_k \in \mathbb{R}$ e os vetores $u_k, v_k \in \mathbb{R}^n$ tais que

$$B_{k+1} = B_k + a_k u_k u_k^T + b_k v_k v_k^T \quad (1.9)$$

Pré-multiplicando (1.9) por s^k pela direita, temos

$$B_{k+1} s_k = B_k s_k + a_k (u_k^T s_k) u_k + b_k (v_k^T s_k) v_k$$

ou seja, podemos escrever

$$a_k (u_k^T s_k) u_k = y^k \quad e \quad b_k (v_k^T s_k) v_k = -B_k s_k \quad (1.10)$$

i.e.,

$$u_k = \frac{q_k}{a_k (u_k^T s_k)} \quad e \quad v_k = \frac{-B_k s_k}{b_k (v_k^T s_k)} \quad (1.11)$$

Multiplicando (1.10) por s_k^T pela esquerda, obtemos

$$a_k (u_k^T s_k)^2 = s_k^T y_k \quad e \quad b_k (v_k^T s_k)^2 = -s_k^T B_k s_k$$

ou seja,

$$u_k^T s_k = \sqrt{\frac{s_k^T y_k}{a_k}} \quad e \quad v_k^T s_k = \sqrt{\frac{-s_k^T B_k s_k}{b_k}} \quad (1.12)$$

Substituindo (1.12) em (1.11), com $a_k = 1$ e $b_k = -1$, obtemos

$$u_k = \frac{y_k}{\sqrt{s_k^T y_k}} \quad e \quad v_k = \frac{B_k s_k}{\sqrt{s_k^T B_k s_k}} \quad (1.13)$$

Substituindo os valores de u_k e v_k obtidos em (1.12) na equação (1.9), com $a_k = 1$ e $b_k = -1$ temos

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \quad (1.14)$$

onde

$$s^k = x^{k+1} - x^k \quad e \quad y^k = \nabla f(x^{k+1}) - \nabla f(x^k) \quad (1.15)$$

A expressão (1.14) é chamada de fórmula BFGS (Broyden, Fletcher, Goldfarb, Shanno).

Proposição 1.3.1. *Seja a fórmula B_{k+1} em (1.14). Suponha que no algoritmo 1.3.1 o tamanho do passo α_k é obtido por uma minimização local de $f(x^k + \alpha d^k)$ e que B_k é uma matriz simétrica e definida positiva. Então,*

- i) B_{k+1} satisfaz a equação secante;
- ii) B_{k+1} mantém a simetria;
- iii) $s_k^T y_k > 0$ e $s_k^T B_k s_k > 0$;
- iv) B_{k+1} é definida positiva

Demonstração. i) Pré multiplicando a equação (1.14) por s_k temos

$$\begin{aligned} B_{k+1} s_k &= B_k s_k + \frac{y_k y_k^T}{s_k^T y_k} s_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} s_k \\ &= B_k s_k + y_k - B_k s_k \\ &= y_k \end{aligned}$$

Portanto, a matriz B_{k+1} satisfaz a equação secante (1.8).

ii) Basta provar que cada uma das matrizes parcelas de B_{k+1} é uma matriz simétrica.

A matriz B_k é simétrica por hipótese. Na segunda matriz de (1.14) temos

$$\left(\frac{y_k y_k^T}{s_k^T y_k} \right)^T = \frac{y_k y_k^T}{s_k^T y_k}$$

Na terceira matriz de (1.14), temos

$$\begin{aligned} \left(\frac{B_k s_k (s_k^T B_k)}{s_k^T B_k s_k} \right)^T &= \frac{(s_k^T B_k)^T (B_k s_k)^T}{s_k^T B_k s_k} \\ &= \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \end{aligned}$$

Portanto, a matriz B_{k+1} é simétrica.

iii) Pelo algoritmo (1.3.1) o passo α_k , é obtido pela minimização local de $f(x^k + \alpha d^k)$. Logo

$$\nabla f(x^k + \alpha d^k)^T d^k = \nabla f(x^{k+1}) d(x^k) = 0 \quad (1.16)$$

pelo algoritmo (1.3.1)

$$\begin{aligned} s_k &= x^{k+1} - x^k \\ &= \alpha_k d(x^k) \\ &= -B_k \nabla f(x^k) \end{aligned} \quad (1.17)$$

Utilizando as relações (1.15), (1.17), (1.16) e a hipótese que B_k é uma matriz definida positiva temos

$$\begin{aligned}
 y_k^T s_k &= (\nabla f(x^{k+1}) - \nabla f(x^k))^T s_k \\
 &= (\nabla f(x^{k+1}))^T s_k - (\nabla f(x^k))^T s_k \\
 &= -(\nabla f(x^k))^T s_k \\
 &= (\nabla f(x^k))^T B_k \nabla f(x^k) > 0
 \end{aligned} \tag{1.18}$$

Desta forma, concluímos que os vetores $y_k, s_k \neq 0$. Novamente, usando a hipótese de que B_k é definida positiva, temos

$$s_k^T B_k s_k > 0$$

iv) Tome $w \in \mathbb{R}^n \setminus \{0\}$. Pré multiplicando a equação (1.14) por w pela esquerda e direita temos

$$\begin{aligned}
 w^T B_{k+1} w &= w^T B_k w + \frac{w^T y_k y_k^T w}{s_k^T y_k} - \frac{w^T B_k s_k s_k^T B_k w}{s_k^T B_k s_k} \\
 &= w^T B_k w - \frac{(w^T B_k s_k)^2}{s_k^T B_k s_k} + \frac{(y_k y_k^T)^2}{s_k^T y_k}
 \end{aligned} \tag{1.19}$$

Por (1.18), $\frac{(y_k^T w)^2}{s_k^T y_k} = M$ é positivo. Existe uma matriz $Q \in \mathbb{R}^{n \times n}$ tal que $B_k = Q Q^T$. Defina $u = Q^T w$ e $v = Q^T s_k$. Logo

$$\begin{aligned}
 u^T u &= w^T Q Q^T w = w^T B_k w \\
 v^T v &= s_k^T Q Q^T s_k = s_k^T B_k s_k \\
 u^T v &= w^T Q Q^T s_k = w^T B_k s_k
 \end{aligned} \tag{1.20}$$

Substituindo as relações de (1.20) em (1.19) e pela desigualdade de Cauchy-Schwarz temos

$$\begin{aligned}
 w^T B_{k+1} w &= u^T u - \frac{(u^T v)^2}{v^T v} + M \\
 &= \frac{(u^T u)(v^T v) - (u^T v)^2}{v^T v} + M > 0
 \end{aligned}$$

Portanto, a matriz B_{k+1} é definida positiva. \square

Observação 1.3.1

A convergência de um método de Otimização está diretamente ligada ao comprimento do passo em cada iteração. Uma condição bastante conhecida, que define o valor de α_k que baseia-se num valor mínimo de redução da função objetivo f é a **Condição de Armijo**, que possui a seguinte fórmula recursiva

$$f(x^k + \alpha_k d^k) \leq f(x^k) + c_1 \alpha_k \nabla f(x^k)^T d^k \quad (1.21)$$

definida para uma constante $c_1 \in (0, 1)$. Para valores muito pequenos de α_k , (1.21) é satisfeita. Desta forma, apenas a condição (1.21) não garante a convergência, pois α_k pode se tornar muito pequeno. Para corrigir este problema, introduzimos o que chamamos de **Condição de Curvatura**, que é escrita da forma

$$\nabla f(x^k + \alpha_k d^k)^T d^k \geq c_2 \nabla f(x^k)^T d^k \quad (1.22)$$

definida para uma constante $c_2 \in (0, 1)$. Juntos a Condição de Armijo e a Condição de Curvatura são conhecidas como **Condições de Wolfe**, ou seja

$$\begin{aligned} f(x^k + \alpha_k d^k) &\leq f(x^k) + c_1 \alpha_k \nabla f(x^k)^T d^k \\ \nabla f(x^k + \alpha_k d^k)^T d^k &\geq c_2 \nabla f(x^k)^T d^k \end{aligned} \quad (1.23)$$

com $0 < c_1 < c_2 < 1$.

Veremos mais à frente, que quando a função f é fortemente convexa, $s_k^T y_k > 0$ será satisfeita para quaisquer dois pontos x^k e x^{k+1} . Para funções não convexas, a escolha do comprimento do passo satisfazendo a condição de Wolfe (em particular o passo de Newton $\alpha_k = 1$), garante que $\nabla f(x^{k+1})^T s_k \geq c_2 \nabla f(x^k)^T s_k$, e por isso, utilizando as hipóteses da Proposição 1.3.1, temos

$$\begin{aligned} y_k^T s_k &= \nabla f(x^{k+1})^T s_k - \nabla f(x^k)^T s_k \geq c_2 \nabla f(x^k)^T s_k - \nabla f(x^k)^T s_k \\ &= (c_2 - 1) \nabla f(x^k)^T s_k \\ &= (c_2 - 1) \alpha_k \nabla f(x^k)^T d^k > 0 \end{aligned} \quad (1.24)$$

que coincide com o item (iii), da Proposição 1.3.1.

1.3.2 Convergência do Método Quase-Newton BFGS

Nesta seção mostra-se a convergência dos métodos Quase-Newton, no qual foi abordado na referência [8].

Definição 1.3.1. *Seja a sequência $\{x^k\}$ gerada pelo Algoritmo 1.3.1, no qual converge para x^* , i.e., $\lim_{k \rightarrow \infty} \|x^k - x^*\| = 0$. Se existe um número real $\mu \geq 1$ e uma constante positiva β (independente de k), tal que:*

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^\mu} = \beta$$

Dizemos que $\{x_k\}$ é μ -ordem. Em particular:

1. *Quando $\mu = 1$ e $\beta \in (0, 1)$, a sequência $\{x_k\}$ possui taxa de convergência Q-linear.*
2. *Quando $\mu = 1$ e $\beta = 0$ ou $1 < \mu < 2$ e $\beta > 0$, a sequência $\{x_k\}$ tem taxa de convergência Q-superlinear.*
3. *Quando $\mu = 2$, dizemos que x_k tem taxa de convergência Q-quadrática.*

Lema 1.3.1. *Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável no conjunto aberto convexo D . Então para qualquer $u, v, x \in D$, temos*

$$i) \quad \|\nabla f(u) - \nabla f(v) - \nabla^2 f(x)(u - v)\| \leq \left[\sup_{0 \leq t \leq 1} \|\nabla^2 f(v + t(u - v)) - \nabla^2 f(x)\| \right] \|u - v\| \quad (1.25)$$

E além disso, se a matriz Hessiana de f é Lipschitz contínua, com constante γ , em D , então

$$ii) \quad \|\nabla f(u) - \nabla f(v) - \nabla^2 f(x)(u - v)\| \leq \gamma \max\{\|u - x\|, \|v - x\|\} \|u - v\| \quad (1.26)$$

Demonstração. Ver [21]. □

Lema 1.3.2. *Sejam o gradiente e a Hessiana da função f satisfazendo as condições do lema 1.3.1 e que $[\nabla^2 f(x)]^{-1}$ exista para todo $x \in D$. Então existe $\epsilon > 0$ e $\beta > \alpha > 0$ tal que para todo $u, v \in D$, quando $\max\{\|u - x\|, \|v - x\|\} \leq \epsilon$, vale*

$$\alpha \|u - v\| \leq \|\nabla f(u) - \nabla f(v)\| \leq \beta \|u - v\| \quad (1.27)$$

Demonstração. Ver [21]. □

Agora veremos dois resultados de convergência, segundo [22] e [23].

Lema 1.3.3. *Se a convergência da sequência gerada pelo Algoritmo 1.3.1 é Q -superlinear para x^* , então:*

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} = 1$$

Demonstração. Para um inteiro $k \geq 0$, temos

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{\|(x_{k+1} - x_k + (x_k - x^*))\|}{\|x_k - x^*\|} \geq \left| \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} - \frac{\|x_k - x^*\|}{\|x_k - x^*\|} \right|$$

Usando o fato de que a ordem de convergência de $\{x_k\}$ é Q -superlinear e a definição de módulo temos:

$$-\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} - \frac{\|x_k - x^*\|}{\|x_k - x^*\|} \leq \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}$$

Passando o limite quando $k \rightarrow \infty$ temos:

$$0 \leq \lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} - \frac{\|x_k - x^*\|}{\|x_k - x^*\|} \leq 0$$

Portanto,

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} = 1$$

□

Teorema 1.3.1. *Sejam a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável em um conjunto aberto convexo $D \subset \mathbb{R}^n$, um minimizador local $x^* \in D$ com $\nabla^2 f(x^*)$ simétrica e definida positiva e uma vizinhança $N(x^*, \epsilon)$ de x^* tal que $\|\nabla^2 f(\bar{x}) - \nabla^2 f(x)\| \leq \gamma \|\bar{x} - x\|$, $\forall x, \bar{x} \in N(x^*, \epsilon)$. Suponha que B_k é uma sequência de matrizes simétricas definidas positivas. Assuma que a sequência $\{x^k\}$ converge para x^* . Então $\nabla f(x^*) = 0$ e $\{x^k\}$ converge superlinearmente para x^* se, e somente se,*

$$\lim_{k \rightarrow \infty} \frac{\|[B_k - \nabla^2 f(x^*)]s^k\|}{\|s^k\|} = 0$$

Demonstração. A ideia é provar a equivalência

$$\lim_{k \rightarrow \infty} \frac{\|[B_k - \nabla^2 f(x^*)]s_k\|}{\|s_k\|} = 0 \Leftrightarrow \lim_{k \rightarrow \infty} \frac{\|\nabla f(x^{k+1})\|}{\|s_k\|} = 0 \Leftrightarrow \lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0$$

no qual $s_k = x^{k+1} - x^k$.

(\Leftarrow)

Observamos que:

$$\begin{aligned} [B_k - \nabla^2 f(x^*)](x_{k+1} - x_k) &= -\nabla f(x^k) - \nabla^2 f(x^k)(x^{k+1} - x^k) \\ &= [\nabla f(x^{k+1}) - \nabla f(x^k) - \nabla^2 f(x^k)(x^{k+1} - x^k)] - \nabla f(x^{k+1}) \end{aligned}$$

que podemos reescrever,

$$\nabla f(x^{k+1}) = [\nabla f(x^{k+1}) - \nabla f(x^k) - \nabla^2 f(x^k)(x^{k+1} - x^k)] - [B_k - \nabla^2 f(x^*)](x_{k+1} - x_k)$$

Usando a desigualdade triangular e dividindo por $\|s^k\|$, temos:

$$\frac{\|\nabla f(x^k)\|}{\|s^k\|} \leq \frac{\|\nabla f(x^{k+1}) - \nabla f(x^k) - \nabla^2 f(x^k)(x^{k+1} - x^k)\|}{\|s^k\|} + \frac{\|(B_k - \nabla^2 f(x^*))s^k\|}{\|s^k\|} \quad (1.28)$$

Pelo lema 1.3.1, a inequação (1.28) fica:

$$\frac{\|\nabla f(x^k)\|}{\|s^k\|} \leq \frac{\gamma}{2} \|x^{k+1} - x^*\| + \|x^k - x^*\| \frac{\|s^k\|}{\|s^k\|} + \frac{\|(B_k - \nabla^2 f(x^*))s^k\|}{\|s^k\|}$$

Passando ao limite quando $k \rightarrow \infty$, temos:

$$\lim_{k \rightarrow \infty} \frac{\|\nabla f(x^{k+1})\|}{\|s^k\|} = 0 \quad (1.29)$$

Pelo lema 1.3.2, existe $\beta > 0$ e $k_0 \geq 0$ tal que $\forall k \geq k_0$, temos:

$$\|\nabla f(x^{k+1})\| = \|\nabla f(x^{k+1} - \nabla f(x^*))\| \geq \beta \|x^{k+1} - x^*\|$$

Dividindo por $\|x^{k+1} - x^k\|$, temos:

$$\frac{\|\nabla f(x^{k+1})\|}{\|s^k\|} = \frac{\beta \|x^{k+1} - x^*\|}{\|x^{k+1} - x^*\| + \|x^k - x^*\|} = \beta \frac{r^k}{1 + r^k}, \quad \text{onde } r^k = \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|}$$

Passando ao limite quando $k \rightarrow \infty$,

$$\lim_{k \rightarrow \infty} \frac{r^k}{1 + r^k} = 0$$

ou seja,

$$\lim_{k \rightarrow \infty} r^k = 0$$

Portanto, a sequência $\{x^k\}$ é convergente para x^* superlinearmente.

(\Rightarrow)

Pelo lema 1.3.2 temos:

$$\|\nabla f(x^{k+1})\| \leq \bar{\beta} \|x^{k+1} - x^*\|$$

Dividindo por $\|x^k - x^*\|$

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \geq \frac{1}{\bar{\beta}} \frac{\|\nabla f(x^{k+1})\|}{\|x^k - x^*\|}$$

Passando o limite e usando a hipótese que a convergência é superlinear temos:

$$\lim_{k \rightarrow \infty} \frac{1}{\bar{\beta}} \frac{\|\nabla f(x^{k+1})\|}{\|x^k - x^*\|} = 0$$

Reescrevendo, temos:

$$\lim_{k \rightarrow \infty} \frac{1}{\bar{\beta}} \frac{\|\nabla f(x^{k+1})\| \|x^{k+1} - x^k\|}{\|x^{k+1} - x^k\| \|x^k - x^*\|} = 0$$

Pelo Lema 1.3.3, fica:

$$\lim_{k \rightarrow \infty} \frac{\|\nabla f(x^{k+1})\|}{\|x^{k+1} - x^k\|} = 0 \quad (1.30)$$

Note que a sequência $\{x^k\}$ converge para x^* , ou seja, $\lim_{k \rightarrow \infty} \|x^k - x^*\| = 0$. Como a função f é duas vezes diferenciável, temos

$$\lim_{k \rightarrow \infty} \nabla f(x^k) = \nabla f(x^*) \quad (1.31)$$

Por (1.30), e também usando o fato que $\lim_{k \rightarrow \infty} \|x^k - x^*\| = 0$, temos

$$\lim_{k \rightarrow \infty} \|\nabla f(x^{k+1})\| = 0$$

que substituindo em (1.31) temos

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

Observe que:

$$[B_k - \nabla^2 f(x^*)](x^{k+1} - x^k) = [\nabla f(x^{k+1}) - \nabla f(x^k) - \nabla^2 f(x^*)s^k] - \nabla f(x^{k+1})$$

Usando a desigualdade triangular, dividindo por $\|s^k\|$ e usando o lema 1.3.1,

$$\frac{\|[B_k - \nabla^2 f(x^*)]s^k\|}{\|s^k\|} \leq \frac{\gamma(\|x^{k+1} - x^*\| + \|x^k - x^*\|)\|s^k\|}{2\|s^k\|} + \frac{\|\nabla f(x^{k+1})\|}{\|s^k\|}$$

Passando ao limite quando $k \rightarrow \infty$, temos:

$$\lim_{k \rightarrow \infty} \frac{\|[B_k - \nabla^2 f(x^*)]s^k\|}{\|s^k\|} = 0 \quad \square$$

2 Fundamentos de Otimização Multiobje- tivo

Neste capítulo, vamos relacionar alguns aspectos dos modelos matemáticos usados na Otimização Multiobjetivo. Definimos ordem parcial no espaço \mathbb{R}^n e conceitos relacionados às soluções Pareto-ótimo, que serão utilizados neste trabalho.

2.1 Problema Geral

A Otimização Multiobjetivo consiste em minimizar duas ou mais funções objetivos, em geral conflitantes. A solução que minimiza umas das funções objetivo não será, em geral, a solução que minimiza as demais funções objetivos em questão.

Segundo Miettinen [24], um problema de otimização multiobjetivo, é da forma

$$(P) \begin{cases} \min F(x) = (f_1(x), f_2(x), \dots, f_j(x))^T \\ \text{s.a. } x \in S \end{cases}$$

no qual existem $j (\geq 2)$ funções objetivos definidos como $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$. Denotamos o vetor objetivo por $F(x) = (f_1(x), f_2(x), \dots, f_j(x))^T$, o vetor decisão por $x = (x_1, x_2, \dots, x_n)^T$ pertencente ao conjunto $S \subset \mathbb{R}^n$ e $S \neq \emptyset$. De uma forma geral, este problema, não existe uma única solução que é ótima para todas as funções objetivo.

2.2 Equivalências entre alguns problemas de Otimização

No desenvolvimento dos artigos principais, vamos utilizar equivalências entre problemas de otimização, visando tornar os resultados mais fáceis de verificação. Sejam os seguintes problemas

$$(P1) = \left\{ \min_{x \in \mathbb{R}^n} [\max_{i=1, \dots, n} f_i(x)] \right.$$

Defina uma função $K : \mathbb{R}^n \rightarrow \mathbb{R}$ onde $K(x) = \max_{i=1, \dots, n} f_i(x)$. Logo o problema (P1) é reescrito como:

$$(P1) = \left\{ \min_{x \in \mathbb{R}^n} K(x) \right.$$

Dizemos que $x^* \in \mathbb{R}^n$ é solução de (P1), se $K(x^*) \leq K(x)$, $\forall x \in \mathbb{R}^n$.

$$(P2) = \begin{cases} \min t \\ \text{s.a. } f_i(x) \leq t, \quad \forall i = 1, \dots, n \\ (x, t) \in \mathbb{R}^{n+1}. \end{cases}$$

Sejam os conjuntos $A = \{(x, t) \in \mathbb{R}^{n+1}; f_i(x) \leq t\}$ e $U = \{t \in \mathbb{R}; (x, t) \in A, \text{ para algum } x \in \mathbb{R}^n\}$. Dizemos que t^* é solução de (P2), se $t^* \leq t$, para todo $t \in D$ (conjunto viável de (P2)).

Proposição 2.2.1. *Seja $K(x) = \max f_i(x)$ e $t^* = K(x^*)$, onde $x^* = \operatorname{argmin} K(x)$. Então o par (x^*, t^*) é solução de (P2). Reciprocamente, se (x^*, t^*) é solução de (P2), então $t^* = K(x^*)$, i.e., é solução de (P1).*

Demonstração. Vamos demonstrar primeiramente que o par (x^*, t^*) solução de (P1) resolve o problema (P2).

Seja $t^* = K(x^*)$ o valor ótimo de (P1), no qual $x^* \in \mathbb{R}^n$ é o ponto ótimo. Como $K(x) = \max f_i(x)$, $t^* = K(x^*) = \max f_i(x^*)$, ou seja, $f_i(x^*) \leq t^*$. Isto significa que o par (x^*, t^*) pertence ao conjunto A de (P2). Seja $t = K(x)$, no qual $t \in \operatorname{Im}(K) \subset \mathbb{R}$. Como por hipótese, $K(x) \geq f_i(x)$, temos que $f_i(x) \leq t$. Logo, o par (x, t) , pertence ao conjunto A de (P2). Como $K(x^*) \leq K(x)$, temos que $t^* \leq t$, com $t \in \mathbb{R}$ e $(x, t) \in A$. Portanto, t^* é solução de (P2).

Agora vamos demonstrar que t^* solução de (P2) resolve o problema (P1).

Pelo problema (P2), observamos que um ponto $(x, t) \in \mathbb{R}^{n+1}$ é factível, se $f_i(x) \leq t$, $i = 1, 2, \dots, n$, ou seja, $\max_{i=1, \dots, n} f_i(x) = t$, no qual $t \in \operatorname{Im}(K)$. Tomando a solução de (P2) $(x^*, t^*) \in \mathbb{R}^{n+1}$, temos

$$\max_{i=1, \dots, n} f_i(x^*) \leq t^* \quad e \quad t^* \leq t, \quad \forall t \in D$$

Para todo $x \in \mathbb{R}^n$, é verdade que $(x, K(x)) \in A$ e isso implica que $K(x) \in D$. Agora, como $t^* \leq t$, para todo $t \in D$, segue que $K(x^*) \leq K(x)$, para todo $x \in \mathbb{R}^n$. Portanto, a solução t^* de (P2) resolve o problema (P1). \square

2.3 Ordenamento de soluções no \mathbb{R}^n

Nos problemas em Otimização (com somente uma função objetivo), tínhamos uma certa "facilidade" entre comparações de soluções, pois os valores ótimos pertenciam ao conjunto dos números reais \mathbb{R} . Porém, isto não acontece na Otimização Multiobjetivo, pois os valores ótimos estão contidos no \mathbb{R}^n . Desta forma ficamos impossibilitados de fazermos comparações entre esses vetores. Vamos definir o ordenamento desses vetores iniciando um estudo sobre Relações Binárias e Ordens Parciais.

Segundo Sampaio [25], uma relação binária A em \mathbb{R}^n é um subconjunto do produto cartesiano $\mathbb{R}^n \times \mathbb{R}^n$. Assim, diz-se que um elemento $x \in \mathbb{R}^n$ está relacionado com $y \in \mathbb{R}^n$ por A se $(x, y) \in A$.

Definição 2.3.1. *Seja A uma relação binária em \mathbb{R}^n . Dizemos que ela é:*

1. *Reflexiva, se $(x, x) \in A$ para todo $x \in \mathbb{R}^n$*
2. *Anti-Simétrica, se $(x, y) \in A$ se $(y, x) \in A$ implica que $x = y$ para $x, y \in \mathbb{R}^n$*
3. *Transitiva, se $(x, y) \in A$ e se $(y, z) \in A$ implica que $(x, z) \in A$ para $x, y, z \in A$*
4. *Completo ou conectado, se $(x, y) \in A$ ou $(y, x) \in A$ para todo $x, y \in \mathbb{R}^n$ com $x \neq y$*
5. *Linear, se $(x_1, y_1), (x_2, y_2) \in A$ e $t \geq 0$ então (tx_1, tx_2) e $(x_1 + x_2, y_1 + y_2) \in A$*

Definição 2.3.2. *Dizemos que um conjunto $X \subset \mathbb{R}^n$ é parcialmente ordenado, se existe uma relação binária para todo $x, y \in X$ e valem as propriedades de Reflexividade, Anti-Simetria e Transitividade.*

Definição 2.3.3. *Sejam $x, y \in \mathbb{R}^n$. Definimos as relações " \preceq ", " \prec ", " $=$ " e " \neq " da seguinte forma:*

$$\begin{aligned} x \preceq y &\Leftrightarrow x_i \leq y_i, \quad i = 1, \dots, n \\ x \prec y &\Leftrightarrow x_i < y_i, \quad i = 1, \dots, n \\ x = y &\Leftrightarrow x_i = y_i, \quad i = 1, \dots, n \\ x \neq y &\Leftrightarrow \exists i; \quad x_i \neq y_i \end{aligned}$$

De forma análoga, definimos as relações " \succeq " e " \succ ". Se não houver comparação $x \not\preceq y$.

Pela definição 2.3.2, uma relação binária é chamada de **ordem parcial** se é reflexiva, transitiva e anti-simétrica. Vamos definir, segundo [26], a classe de ordens parciais determinadas por cones em \mathbb{R}^m . Os cones no \mathbb{R}^m são definidos como o conjunto $K \subset \mathbb{R}^m$ quando $d \in K$ temos $td \in K$, $\forall t \in \mathbb{R}_+$. Também definimos que o cone $K \subset \mathbb{R}^m$ é **pontiagudo** se $K \cap (-K) = \{0\}$.

Definição 2.3.4. Dado $K \subset \mathbb{R}^m$ um cone não vazio, convexo, fechado e pontiagudo. A relação binária associada a K , denotada por \preceq_K , é definida como:

$$y_1 \preceq_K y_2 \text{ para } y_1, y_2 \in \mathbb{R}^m \Leftrightarrow y_2 - y_1 \in K$$

Neste caso, diz-se que y_1 domina y_2 .

Também definimos a relação binária estrita a K denotada por \prec_K , quando K tem interior não vazio, como

$$y_1 \prec_K y_2 \text{ para } y_1, y_2 \in \mathbb{R}^m \Leftrightarrow y_2 - y_1 \in \text{int}(K)$$

Neste caso, diz-se que y_1 domina fracamente (ou estritamente) y_2

Segundo [27], \preceq_K é uma relação binária reflexiva, transitiva e anti-simétrica. Portanto, \preceq_K é uma ordem parcial. Em [26], além de ser uma ordem parcial, é compatível com a soma de vetores e o produto com um escalar positivo, ou seja, dados $y, y_1, y_2 \in \mathbb{R}^m$ e $\lambda \in \mathbb{R}$ tem-se que:

$$y_1 \preceq_K y_2 \Rightarrow \lambda y_1 \preceq_K \lambda y_2$$

e

$$y_1 \preceq_K y_2 \Rightarrow y_1 + y \preceq_K y_2 + y$$

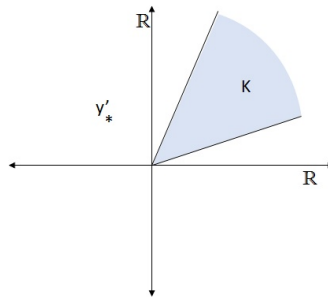


Figura 1 – Cone $K \subset \mathbb{R}^2$.

A Figura 1 ilustra um cone $K \subset \mathbb{R}^2$ fechado, convexo, pontiagudo e com interior não vazio.

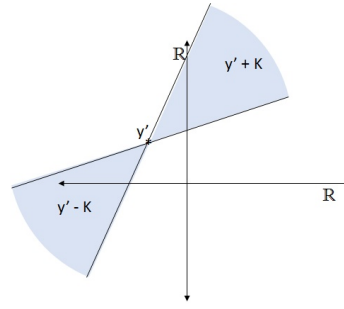


Figura 2 – Cone transladado.

Na Figura 2, temos a ilustração de dois cones transladados: $y' + K$ e $y' - K$. Todos os vetores do cone $y' + K$ são dominados por y' , enquanto que no cone $y' - K$ todos os vetores do cone dominam y' .

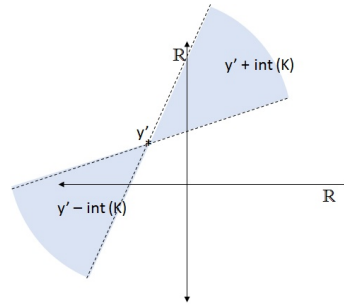


Figura 3 – Interior de um cone transladado.

Na Figura 3, todos os vetores no cone transladado $y' + int(K)$ são dominados estritamente por y' , enquanto que todos os vetores do cone $y' - int(K)$ dominam estritamente y' . Figuras 1, 2 e 3 inspirados no trabalho de [26].

Definição 2.3.5. Dizemos que um ponto $x^* \in \mathbb{R}^n$ é K -minimizador irrestrito (ou K -ótimo) se não existe um $y \in \mathbb{R}^n$ com $F(y) \preceq_K F(x^*)$ e $F(y) \neq F(x^*)$.

Exemplo 2.3.1. Sejam $K = \mathbb{R}_+^2$, $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $F(x, y) = (x^2 + 1, x^2 + y^2)$, temos que $(0, 0)$ é um ponto K -ótimo de F , pois $x^2 + 1 < 1$ e $x^2 + y^2 < 0$ não possuem solução para nenhum par $(x, y) \in \mathbb{R}^2$. Assim, não existe um $(x, y) \in \mathbb{R}^2$ com $F(x, y) \preceq_{\mathbb{R}_+^2} F(0, 0)$ e $F(x, y) \neq F(0, 0)$.

Definição 2.3.6. Dizemos que um ponto $x \in \mathbb{R}^n$ é K -crítico de F quando:

$$Im(JF(x)) \cap -int(K) = \emptyset$$

Definição 2.3.7. *Seja $x \in \mathbb{R}^n$ um ponto não K -crítico. Dizemos que $v \in \mathbb{R}^n$ é uma K -direção de descida de F em x , quando,*

$$JF(x)v \in -\text{int}(K), \text{ ou seja, } JF(x)v \prec_K 0$$

Além disso, existe um $\bar{t} > 0$ tal que $F(x + tv) \prec_K F(x)$ para todo $t \in (0, \bar{t})$.

2.4 Pareto Otimalidade

Nesta seção, vamos definir o conceito de Ótimo Pareto [24]. Este conceito surge da necessidade de comparação entre vetores do \mathbb{R}^n . Desta problemática, tomamos os vetores que "melhoram os resultados" das funções objetivos ao mesmo tempo. Estes elementos irão pertencer ao que chamamos de conjunto Pareto Ótimo.

Seja \mathbb{R} o conjunto dos números reais. Definimos por \mathbb{R}_+ o conjunto dos números reais não-negativos e por \mathbb{R}_{++} o conjunto dos números reais estritamente positivos. Seja $U \subset \mathbb{R}^n$ um conjunto aberto e uma função

$$F : U \rightarrow \mathbb{R}^m$$

Pela Definição 2.3.4, usaremos a ordem parcial induzida dos conjuntos $\mathbb{R}_+^m = \mathbb{R}_+ \times \dots \times \mathbb{R}_+$ (chamado de cone Paretiano do \mathbb{R}^m) e $\mathbb{R}_{++}^m = \mathbb{R}_{++} \times \dots \times \mathbb{R}_{++}$, definindo

$$F(y) \preceq F(z) \Leftrightarrow F(z) - F(y) \in \mathbb{R}_+^m$$

e

$$F(y) \prec F(z) \Leftrightarrow F(z) - F(y) \in \mathbb{R}_{++}^m$$

E o conjunto

$$-\mathbb{R}_{++}^m = \{-u; u \in \mathbb{R}_{++}^m\}$$

Teorema 2.4.1. *Seja F diferenciável em $x \in \mathbb{R}^n$. Uma condição necessária para que x seja Pareto-ótimo é:*

$$\text{Im}(JF(x)) \cap (-\mathbb{R}_{++}^m) = \emptyset$$

Demonstração. Suponha por absurdo que $\text{Im}(JF(x)) \cap (-\mathbb{R}_{++}^m) \neq \emptyset$, isto é, existe uma direção de descida $v \in \mathbb{R}^n$ tal que $JF(x)v \in (-\mathbb{R}_{++}^m)$, isto implica que $\langle \nabla f_i(x), v \rangle < 0, \forall i = 1, \dots, m$. Assim, v é uma direção de descida para cada f_i a partir de x , ou seja, existe $\bar{\lambda} > 0$ tal que $f_i(x + \lambda v) < f_i(x), \forall \lambda \in (0, \bar{\lambda})$. Logo, $F(x + \lambda v) \prec_{\mathbb{R}_+^m} F(x)$ o que é uma contradição, pois supomos que x é Pareto ótimo. \square

Exemplo 2.4.1. *Caso particular do Teorema 2.4.1*

O Teorema 2.4.1 generaliza o conceito de ponto crítico para a otimização clássica. Basta fazer $m = 1$ e $K = \mathbb{R}_+$. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável e $x^* \in \mathbb{R}^n$ uma solução Pareto ótimo. Logo, $Jf(x^*) = \nabla f(x^*)$ e $-\text{int}K = -\mathbb{R}_{++}$, i.e.,

$$\text{Im}(\nabla f(x^*)) \cap (-\mathbb{R}_{++}) = \emptyset$$

Então,

$$\forall d \in \mathbb{R}^n, \nabla f(x^*)^T d \notin (-\mathbb{R}_{++}), \text{ ou seja, } \nabla f(x^*)^T d \geq 0 \quad (2.1)$$

Tome $d = -\nabla f(x^*)$. Em (2.1), temos $-\nabla f(x^*)^T \nabla f(x^*) \geq 0$, isto é, $\|\nabla f(x^*)\|^2 \leq 0$, tem-se, $\nabla f(x^*) = 0$

Exemplo 2.4.2. *Exemplo da não suficiência do Teorema 2.4.1*

Seja a função $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, diferenciável definida como $f(x, y) = x^2 - y^2$ e o ponto $\bar{x} = (0, 0)$

Logo, $Jf(x, y) = \nabla f(x, y) = \begin{pmatrix} 2x \\ -2y \end{pmatrix}$. Desta forma, $Jf(\bar{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, satisfazendo assim a condição do Teorema 2.4.1. Porém para $\hat{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ temos que $f(\hat{x}) = -3 < 0 = f(\bar{x})$.

Agora definiremos, segundo [24], Soluções Pareto.

Definição 2.4.1 (Solução Pareto Ótimo)(ou Ponto eficiente global). *Um vetor $x^* \in S$ é Pareto Ótimo se não existe outro vetor $x \in S$ tal que*

$$f_i(x) \leq f_i(x^*), \quad i = 1, \dots, m \quad \text{e} \quad f_j(x) < f_j(x^*) \quad \text{para algum } j$$

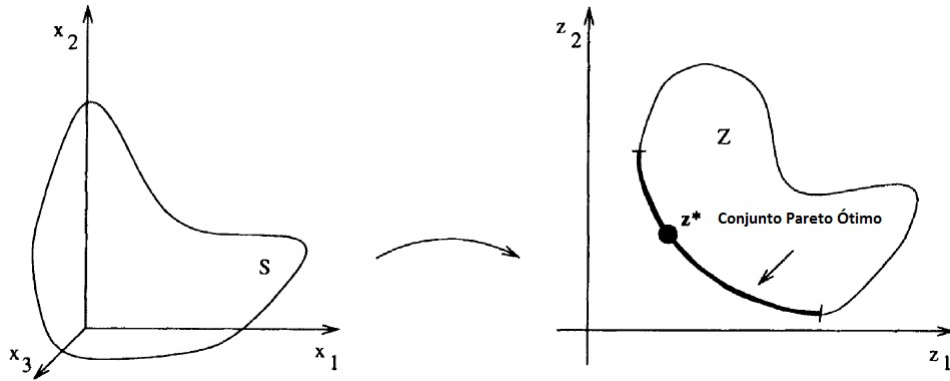


Figura 4 – Ilustração do conjunto Ótimo Pareto.

Na Figura 4, retirada do trabalho de [24], mostra-se uma região factível $S \subset \mathbb{R}^3$ e seu respectivo conjunto imagem $Z \subset \mathbb{R}^2$. Observe que a linha em negrito corresponde ao conjunto Pareto Ótimo. Outras definições sobre Ótimo Pareto também são importantes para nosso estudo.

Definição 2.4.2 (Solução localmente Pareto Ótimo)(ou Ponto eficiente global local). *Uma solução $x^* \in S$ é localmente Pareto-ótimo se existe $\delta > 0$ tal que x^* é Pareto-ótimo em $S \cap N(x^*, \delta)$*

Definição 2.4.3 (Solução Pareto Fraco)(ou Ponto eficiente fraco). *Uma solução $x^* \in S$ é Pareto fraco se não existe um outro ponto $x \in S$ tal que $f_i(x) < f_i(x^*)$, $i = 1, \dots, m$*

Definição 2.4.4 (Solução Pareto Fraco Localmente)(ou Ponto eficiente fraco local). *Uma solução $x^* \in S$ é Pareto fraco localmente, se existe $\delta > 0$ tal que x^* é uma solução Pareto fraco em $S \cap N(x^*, \delta)$.*

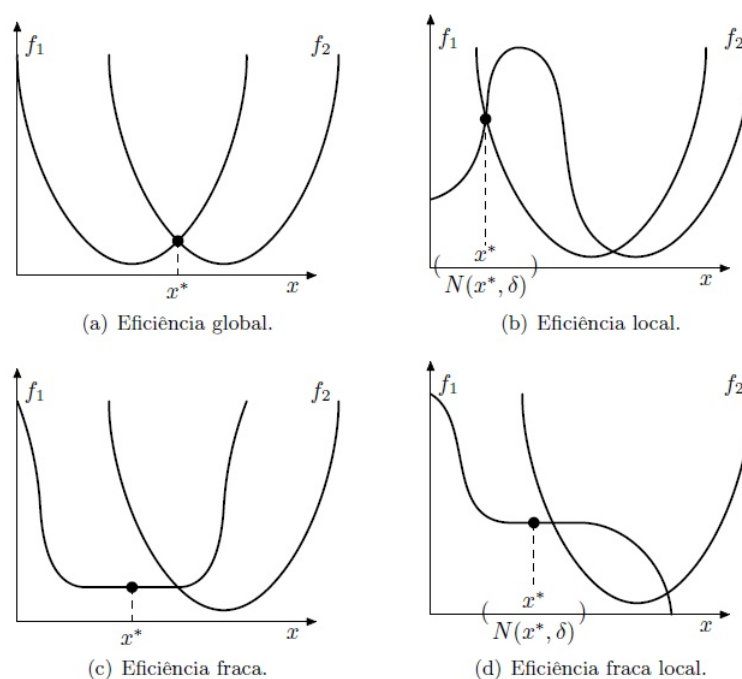


Figura 5 – Tipos de Eficiências.

A Figura 5, encontrada em [25], ilustra as definições anteriores. Se as funções objetivo são convexas, qualquer solução localmente Ótimo Pareto é também Ótimo Pareto, fato que provaremos mais adiante através do Teorema 2.5.1. Observe também que, de uma maneira geral, as soluções Pareto Ótimo local, Pareto ótimo fraco e Pareto ótimo fraco local não são soluções Pareto Ótimo.

2.5 Um resultado sobre convexidade em Otimização Multiobjetivo

A seguir, mostraremos um resultado sobre convexidade, segundo [24]. Seja o Problema de Otimização Multiobjetivo definido como

$$\begin{cases} \min F(x) \\ \text{s.a. } x \in D \end{cases} \quad (2.2)$$

Definição 2.5.1. O problema (2.2) é convexo se todas as funções coordenadas da função vetorial F e o conjunto $D \subseteq \mathbb{R}^n$ forem convexos.

Teorema 2.5.1. Se o problema de otimização multiobjetivo definido em (2.2) for convexo, então toda solução localmente Pareto-ótima é também globalmente Pareto-ótima.

Demonstração. Seja x^* uma solução Pareto ótimo local do problema (P) , i.e., existe uma vizinhança $B(x^*, \delta)$ em x^* , com $\delta > 0$ tal que não exista $x \in B(x^*, \delta) \cap S$ e

$$\begin{cases} f_i(x) \leq f_i(x^*), \forall i = 1, \dots, m \text{ e} \\ f_j(x) < f_j(x^*), \text{ para algum } j \end{cases}$$

Suponha que x^* não seja solução pareto ótimo global, ou seja, existe $x^0 \in S$ tal que:

$$\begin{cases} f_i(x^0) \leq f_i(x^*), \forall i = 1, \dots, m \text{ e} \\ f_j(x^0) < f_j(x^*), \text{ para algum } j \end{cases} \quad (2.3)$$

Defina $\hat{x} = \alpha(x^0) + (1 - \alpha)x^*$, onde $\alpha < \frac{\delta}{\|x^* - x^0\|}$, garantindo assim que $\hat{x} \in B(x^*, \delta)$. Pela convexidade de S , $\hat{x} \in S$. Pela convexidade da função multiobjetivo e por 2.3 temos:

$$f_i(\hat{x}) = f_i(\alpha x^0 + (1 - \alpha)x^*) \leq \alpha f_i(x^0) + (1 - \alpha)f_i(x^*) \leq \alpha f_i(x^*) + (1 - \alpha)f_i(x^*) = f_i(x^*), \forall i = 1, \dots, m.$$

$$f_i(\hat{x}) = f_i(x^*) \Rightarrow f_i(x^*) \leq \alpha f_i(x^0) + (1 - \alpha)f_i(x^*) \Rightarrow f_i(x^*) \leq f_i(x^*)$$

Contradição, pois $f_j(x^*) > f_j(x^0)$ para algum j . Portanto, x^* é solução Pareto ótimo global do problema (P) . \square

3 Alguns Métodos do tipo Quase-Newton para Otimização Multiobjetivo

Nesta seção apresentam-se os principais resultados objetivados neste trabalho que são as resoluções de problemas multiobjetivos de funções convexas e não convexas. Na Seção 3.1 abordaremos algumas definições, lemas e teoremas que irão servir como base para a compreensão dos algoritmos propostos nas Seções 3.2 e 3.3. Na Seção 3.2, trabalharemos com o artigo de Povalej [6] que estende os resultados obtidos por [4] para os métodos Quase-Newton. Com o mesmo suporte teórico de [4], Povalej incrementa um algoritmo com atualizações do método BGFS, aplicado para Otimização Multiobjetivo. Na Seção 3.3, estudaremos um algoritmo Quase-Newton para Otimização Multiobjetivo não convexo. Segundo [8], considerando uma função não convexa $F = (f_1, f_2, \dots, f_m)^T : X \rightarrow \mathbb{R}^m$ continuamente diferenciável, no qual $X \subseteq \mathbb{R}^n$, é um conjunto aberto, mostraremos que o algoritmo tem propriedade convergência global. A partir de agora, em todas as relações de elementos do \mathbb{R}^m , utilizadas para o desenvolvimento teórico, será usado o cone Paretiano.

3.1 Algumas definições

3.1.1 Direção de descida

Definição 3.1.1. *Seja a função $F = (f_1, f_2, \dots, f_m)^T : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, no qual U é um conjunto aberto. As funções f_j , estão definidas na forma*

$$f_j : U \subset \mathbb{R}^n \rightarrow \mathbb{R}, \quad j = 1, 2, \dots, m$$

Assumimos que F é duas vezes diferenciável no conjunto U .

Se $x \in U$ não é estacionário (não é crítico), então existe $d \in \mathbb{R}^n$ tal que $\nabla f_j(x)^T d < 0$ para todo $j = 1, 2, \dots, m$. Pelo fato que a função F é continuamente diferenciável, então

$$\lim_{\alpha \rightarrow 0} \frac{f_j(x + \alpha d) - f_j(x)}{\alpha} = \nabla f_j(x)^T d < 0, \quad j = 1, 2, \dots, m$$

Logo, $d \in \mathbb{R}^n$ é uma direção de descida de F em x , i.e., existe $\alpha_0 > 0$ tal que

$$F(x + \alpha d) \prec F(x), \quad \forall \alpha \in (0, \alpha_0]$$

3.1.2 Algumas definições para o caso convexo

Teorema 3.1.1. *Seja $F : U \subset \mathbb{R}^n$ diferenciável.*

1. *Se x^* é um ponto Pareto-ótimo fraco local, então x^* é um ponto estacionário de F .*
2. *Se U é um conjunto convexo, F convexo e $x^* \in U$ é um ponto estacionário de F , então x^* é um ponto Pareto ótimo fraco.*
3. *Se U é um conjunto convexo, $F \in C^2$, $\nabla^2 f_j(x) > 0$ para todo $x \in U$ para todo $j = 1, 2, \dots, m$, e se x^* é um ponto estacionário de F , então x^* é Pareto ótimo.*

Demonstração. 1) Suponha que x^* não é um ponto crítico de F . Então, o Teorema 2.4.1 não acontece, ou seja, existe $t_0 > 0$ tal que

$$F(x^* + ts) \prec F(x^*), \quad \forall t \in (0, t_0]$$

Contradição, pois x^* é Pareto ótimo fraco local. Portanto, x^* é um ponto crítico de F .

2) Como x^* é ponto crítico, isto é,

$$\nabla f_{j_0}(x^*)^T s \geq 0, \quad \forall s \in \mathbb{R}^n \text{ e algum } j_0 \in I \quad (3.1)$$

Tome $s = x - x^*$. Pelo Teorema 1.2.2, temos:

$$f_{j_0}(x) \geq f_{j_0}(x^*) + \nabla f_{j_0}(x)^T (x - x^*), \quad \forall x \in U \text{ e algum } j_0 \in j = 1, 2, \dots, m$$

por (3.1), temos

$$\begin{aligned} f_{j_0}(x) &\geq f_{j_0}(x^*) + \nabla f_{j_0}(x)^T (x - x^*) \\ &\geq f_{j_0}(x^*) \end{aligned} \quad (3.2)$$

$$\forall x \in U \text{ e algum } j_0 \in j = 1, 2, \dots, m$$

Logo, $\nexists y \in U$ tal que $f_{j_0}(y) < f_{j_0}(x^*)$, $j_0 \in j = 1, 2, \dots, m$, ou seja, $\nexists y \in U$ tal que $F(y) \prec F(x^*)$, $j_0 \in I$. Portanto, x^* é Pareto ótimo fraco.

3) Pelo teorema 1.2.3, como a matriz $\nabla^2 f_{j_0}(x)$ é definida positiva, f é estritamente convexa. Logo pelo item (2) do Teorema 3.1.1,

$$f_{j_0}(x) > f_{j_0}(x^*) + \nabla f_{j_0}(x)^T (x - x^*) > f_{j_0}(x^*)$$

Ou seja

$$f_{j_0}(x) > f_{j_0} x^* \text{ para algum } j_0 \in j = 1, 2, \dots, m$$

Logo, não existe $x \in U$ tal que $F(x) \preceq F(x^*)$ e $F(x) \neq f(x^*)$

Portanto, $x^* \in U$ é Pareto ótimo. □

Utilizaremos no caso convexo, a atualização $B_j(x^{k+1})$, tendo como inspiração a fórmula recursiva Quase-Newton BFGS em (1.14), adaptada para o problema Multiobjetivo, isto é,

$$B_j(x^{k+1}) = B_j(x^k) - \frac{B_j(x^k)s_k s_k^T B_j(x^k)}{s_k^T B_j(x^k)s_k} + \frac{y_{jk} y_{jk}^T}{y_{jk}^T s_k} \quad (3.3)$$

onde $j = 1, 2, \dots, m$

Definição 3.1.2. [6] Definimos $d(x)$ como direção Quase-Newton quando $d(x)$ é solução ótima do problema

$$\min_{d \in \mathbb{R}^n} \max_{j=1,2,\dots,m} \nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d \quad (3.4)$$

onde $B_j(x)$ é uma aproximação BFGS para a Hessiana $\nabla^2 f_j(x)$, $j = 1, 2, \dots, m$

Definição 3.1.3. Definimos como valor ótimo e solução as seguintes funções respectivamente

$$\tau(x) = \min_{d \in \mathbb{R}^n} \max_{j=1,\dots,m} \nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d \quad (3.5)$$

$$d(x) = \arg \min_{d \in \mathbb{R}^n} \max_{j=1,\dots,m} \nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d \quad (3.6)$$

Pela Proposição 2.2.1, o problema (3.4) pode ser escrito como um problema equivalente de otimização quadrática convexa

$$\begin{cases} \min & g(t, d) = t, \\ \text{s.a.} & \nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d - t \leq 0, \quad j = 1, \dots, m \\ & (t, d) \in \mathbb{R} \times \mathbb{R}^n. \end{cases} \quad (3.7)$$

O Lagrangiano do problema (3.7) é

$$L((t, d), \lambda) = t + \sum_{j=1}^m \lambda_j \left(\nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d - t \right)$$

Pela condição de otimalidade de KKT, mostraremos os seus multiplicadores. Logo,

$$\sum_{j=1}^m \lambda_j (\nabla f_j(x) + B_j(x) d) = 0 \quad (3.8)$$

$$\lambda_j \geq 0, \quad j = 1, 2, \dots, m \quad (3.9)$$

$$\sum_{j=1}^m \lambda_j = 1 \quad (3.10)$$

$$\nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d \leq t, \quad j = 1, 2, \dots, m \quad (3.11)$$

$$\lambda_j \left(\nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d - t \right) = 0, \quad j = 1, 2, \dots, m \quad (3.12)$$

O problema tem um ponto de Slater $(1, \mathbf{0}) \in \mathbb{R} \times \mathbb{R}^n$. Logo, pelo Teorema 1.2.12, existem os multiplicadores KKT, $\lambda_j = \lambda_j(x)$, $j = 1, 2, \dots, m$. Desta forma, temos em (3.8),

$$d(x) = - \left(\sum_{j=1}^m \lambda_j(x) B_j(x) \right)^{-1} \sum_{j=1}^m \lambda_j(x) \nabla f_j(x) \quad (3.13)$$

A existência dos multiplicadores KKT do problema convexo (3.7), implica que não há brecha de dualidade. O dual, utilizando a função τ , será

$$\begin{aligned} \tau(x) &= \sup_{\lambda \geq 0} \inf_{d \in \mathbb{R}^n} L((t, d), \lambda) \\ &= \sup_{\substack{\lambda \geq 0 \\ \sum \lambda_j = 1}} \inf_{d \in \mathbb{R}^n} \sum_{j=1}^m \lambda_j \left(\nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d \right) \end{aligned} \quad (3.14)$$

No seguinte lema, vamos provar condições para um ponto $x \in U$ ser estacionário (crítico) e algumas propriedades da função τ .

Lema 3.1.1. *Seja a matriz $B_j(x)$ definida positiva para todo $x \in U \subseteq \mathbb{R}^n$, no qual U é um conjunto aberto e a função τ definida em (3.5). Então:*

- a) Para todo $x \in U$, temos $\tau(x) \leq 0$.
 b) As condições seguintes são equivalentes

1. O ponto x não é estacionário (não é crítico).
2. $\tau(x) < 0$.
3. $d(x) \neq 0$.

- c) A função $d : U \rightarrow \mathbb{R}^n$, definida em (3.6), é limitada em conjuntos compactos.
 d) A função $\tau : U \rightarrow \mathbb{R}$, definida em (3.5) é contínua no conjunto $U \subseteq \mathbb{R}^n$.

Demonstração. Provando o item a)

Pela equação (3.5) e como $d \in \mathbb{R}^n$ é qualquer, em particular vale para $d = 0$. Logo,

$$\tau(x) \leq \max_{j=1,\dots,m} \nabla f_j(x)^T 0 + \frac{1}{2} 0^T B_j(x) 0 = 0 \quad (3.15)$$

Equivalências do item b)

Provaremos de (1) para (2)

Como por hipótese, x é um ponto não estacionário,

$$Im(JF(x)) \cap (-\mathbb{R}_{++}^m) \neq \emptyset$$

o que significa que existe $\hat{d} \in \mathbb{R}^n$ tal que $\nabla f_j(x)^T \hat{d} < 0$, $j = 1, 2, \dots, m$

Pela função (3.5), para todo $t > 0$, temos

$$\begin{aligned} \tau(x) &\leq \max_{j=1,2,\dots,m} \left[t \nabla f_j(x)^T \hat{d} + \frac{t^2}{2} \hat{d}^T B_j(x) \hat{d} \right] \\ &= t \max_{j=1,2,\dots,m} \left[\nabla f_j(x)^T \hat{d} + \frac{t}{2} \hat{d}^T B_j(x) \hat{d} \right] \end{aligned}$$

Para t suficientemente pequeno, o termo à direita ficará negativo. Portanto,

$$\tau(x) < 0$$

Provaremos de (2) para (3)

Suponha que $d(x) = 0$. Pela função em (3.5) temos,

$$\tau(x) = \max_{j=1,2,\dots,m} [\nabla f_j(x)^T 0 + \frac{1}{2} 0^T B_j(x) 0] = 0$$

Contradição, pois $\tau(x) < 0$. Portanto, $d(x) \neq 0$

Provaremos de (3) para (1)

Como por hipótese, a matriz aproximação $B_j(x)$ é definida positiva, para $j = 1, 2, \dots, m$, então

$$\begin{aligned} \nabla f_j(x)^T d(x) &< \nabla f_j(x)^T d(x) + \frac{1}{2} d(x)^T B_j(x) d(x) \\ &\leq \max_{j=1,2,\dots,m} \left[\nabla f_j(x)^T d(x) + \frac{1}{2} d(x)^T B_j(x) d(x) \right] = \tau(x) \leq 0 \end{aligned}$$

Logo, $\nabla f_j(x)^T d(x) < 0$, i.e.,

$$Im(JF(x)) \cap (-\mathbb{R}_{++}^m) \neq \emptyset$$

Portanto, x não é um ponto crítico.

Provando o item (c)

Seja $W \subset U$, um conjunto compacto arbitrário. Pelo item (a), tem-se $\tau(x) \leq 0$, para todo $x \in U$. Tomando $d(x)$ e substituindo na função τ , tem-se

$$f_j(x)^T d(x) + \frac{1}{2} d(x)^T B_j(x) d(x) \leq \max_{j=1,2,\dots,m} \left[\nabla f_j(x)^T d(x) + \frac{1}{2} d(x)^T B_j(x) d(x) \right] = \tau(x) \leq 0$$

no qual $j = 1, 2, \dots, m$, ou seja,

$$\frac{1}{2} d(x)^T B_j(x) d(x) \leq -\nabla f_j(x)^T d(x), \quad j = 1, 2, \dots, m \quad (3.16)$$

A função f é duas vezes diferenciável e a matriz aproximação $B_j(x)$ é definida positiva, e para todo $x \in W$ e $j = 1, 2, \dots, m$. Logo, os autovalores de $B_j(x)$ são uniformemente limitados (e distantes) por zero em W , para todo $j = 1, 2, \dots, m$ e existem $K, L \in \mathbb{R}_+$, tal que

$$K = \max_{\substack{x \in W \\ j=1,2,\dots,m}} \|\nabla f_j(x)\| \quad e \quad L = \min_{\substack{x \in W \\ \|u\|=1 \\ j=1,2,\dots,m}} u^T B_j(x) u \quad (3.17)$$

Passando a norma e usando a desigualdade de Cauchy-Schwarz na inequação (3.16), temos

$$\frac{1}{2} \|B_j(x)\| \|d(x)\|^2 \leq \|\nabla f_j(x)\| \|d(x)\|$$

que substituindo os valores de (3.17)

$$\frac{1}{2} L \|d(x)\|^2 \leq \frac{1}{2} \|B_j(x)\| \|d(x)\|^2 \leq \|\nabla f_j(x)\| \|d(x)\| \leq K \|d(x)\|$$

ou seja

$$\|d(x)\| \leq \frac{2K}{L}, \quad \forall x \in W$$

Portanto, a direção Quase-Newton é uniformemente limitada no conjunto compacto W .

Provando o item (d)

Para a continuidade da função τ é suficiente mostrá-la em um conjunto compacto arbitrário $W \subset U$.

Defina uma família de funções $\{\varphi_{x,j}\}_{x \in W, j=1,2,\dots,m}$ onde

$$\varphi_{x,j} : W \rightarrow \mathbb{R}$$

$$\varphi_{x,j}(z) = \nabla f_j(z)^T d(x) + \frac{1}{2} d(x)^T B_j(z) d(x)$$

Vamos mostrar que esta família de funções é uniformemente equicontínua. Primeiro, defina uma cobertura $\{B(z, \delta_z)\}_{z \in W}$. Para qualquer $z \in W$ a bola $B(z, \delta_z)$ é formada de tal forma que, para todos ϵ_z pequenos, existe algum $\delta_z \in \mathbb{R}_+$ tal que para todo $w \in B(z, \delta_z)$ temos

$$\|B_j(w) - \nabla^2 f_j(z)\| < \frac{\epsilon_z}{2} \quad e \quad \|\nabla^2 f_j(w) - \nabla^2 f_j(z)\| < \frac{\epsilon_z}{2} \quad \text{para todo } j = 1, 2, \dots, m$$

Pela compacidade de W , existe uma subcobertura finita de $\{B(z, \delta_z)\}_{z \in W}$, digamos $\{B(z_i, \delta_{z_i})\}$ e $i = 1, 2, \dots, m$.

Podemos reescrever $\varphi_{x,j}$ como

$$\begin{aligned} \varphi_{x,j}(z) &= \nabla f_j(z)^T d(x) + \frac{1}{2} d(x)^T B_j(z) d(x) + \frac{1}{2} d(x)^T \nabla^2 f_j(z) d(x) - \frac{1}{2} d(x)^T \nabla^2 f_j(z) d(x) \\ &= \nabla f_j(z)^T d(x) + \frac{1}{2} d(x)^T \nabla^2 f_j(z) d(x) + \left[\frac{1}{2} d(x)^T (B_j(z) - \nabla^2 f_j(z)) d(x) \right] \end{aligned} \quad (3.18)$$

Na equação (3.18), os termos $\nabla f_j(z)^T d(x) + \frac{1}{2} d(x)^T \nabla^2 f_j(z) d(x)$ são contínuos, pois a função f é C^2 . Estudaremos o termo entre colchetes de (3.18).

Tome $w_1, w_2 \in W$ tal que $\|w_1 - w_2\| < \delta$ no qual $\delta \in \mathbb{R}_{++}$. Então existem índices $i_1, i_2 \in \{1, 2, \dots, l\}$ tal que $w_i \in B(z_{i_k}, \delta_{z_{i_k}}), k = 1, 2$. Logo

$$\begin{aligned} & \left| \frac{1}{2} d(x)^T (B_j(w_1) - \nabla^2 f_j(w_1)) d(x) - \frac{1}{2} d(x)^T (B_j(w_2) - \nabla^2 f_j(w_2)) d(x) \right| \leq \\ & \leq \frac{1}{2} \|d(x)\|^2 (\|B_j(w_1) - \nabla^2 f_j(z_{i_1})\| + \|\nabla^2 f_j(z_{i_1}) - \nabla^2 f_j(w_1)\| + \|B_j(w_2) - \nabla^2 f_j(z_{i_2})\| + \\ & + \|\nabla^2 f_j(z_{i_2}) - \nabla^2 f_j(w_2)\|) \\ & \leq \frac{1}{2} \|d(x)\|^2 (\epsilon_{z_{i_1}} + \epsilon_{z_{i_2}}) \end{aligned}$$

Ou seja,

(I) Pelo Teorema 1.2.10, $\varphi_{x,j}$ é uniformemente contínua para todo $x \in W$ e $j = 1, 2, \dots, m$

(II) A família $\{\varphi_{x,j}\}_{x \in W, j = 1, 2, \dots, m}$ é uniformemente equicontínua.

(III) Pelo Teorema 1.2.11, a família $\{\phi_x = \max_{j=1,2,\dots,m} \varphi_{x,j}\}_{x \in W}$ é uniformemente equicontínua.

Aplicando a definição de função uniformemente equicontínua em (III), no qual, para todo $\epsilon \in \mathbb{R}_+$, existe $\delta \in \mathbb{R}_+$ tal que para todo $y, z \in W$, $\|y - z\| < \delta$ implica $|\phi_x(y) - \phi_x(z)| < \epsilon$ para todo $x \in W$. Logo para $\|y - z\| < \delta$, temos:

$$\begin{aligned} \tau(z) &\leq \max_{j=1,2,\dots,m} \nabla f_j^T(z) d(y) + \frac{1}{2} d(y)^T B_j(z) d(y) = \phi_y(z) \\ &\leq \phi_y(y) + |\phi_y(z) - \phi_y(y)| < \tau(y) + \epsilon \end{aligned}$$

pois $\tau(y) = \max_{j=1,2,\dots,m} \nabla f_j^T(y) d(y) + \frac{1}{2} d(y)^T B_j(y) d(y) = \phi_y(y)$, pela definição de $d(x)$ em 3.1.3. Logo, $\tau(z) - \tau(y) < \epsilon$. De maneira análoga, provamos também que $\tau(y) - \tau(z) < \epsilon$, i.e., $|\tau(z) - \tau(y)| < \epsilon$. Portanto, τ é contínua em W . \square

No próximo teorema, vamos definir como vai ser o tamanho do passo α_k , utilizando uma modificação na Condição de Armijo utilizado em [23] e [18] onde envolve a função τ . A motivação seria o uso da Condição de Armijo clássica

$$f(x + \alpha d(x)) \leq f(x) + b\alpha \nabla f(x)^T d(x), \text{ no qual } b \in (0, 1) \quad (3.19)$$

Utilizamos o passo de Newton ($\alpha = 1$) próximo ao ponto ótimo, onde $B(x)$ é uma boa aproximação para a Hessiana da F , i.e., $B(x)$ será definida positiva e $b \in (0, \frac{1}{2})$, de acordo com [16]. Para o caso escalar $m = 1$, em (3.13), temos

$$-\nabla f(x) = B(x)d(x) \quad (3.20)$$

e

$$\tau(x) = \nabla f(x)^T d(x) + \frac{1}{2}d(x)^T B(x)d(x) \quad (3.21)$$

Substituindo (3.20) em (3.21), temos

$$\begin{aligned} \tau(x) &= \nabla f(x)^T d(x) - \frac{1}{2}\nabla f(x)^T d(x) \\ &= \frac{1}{2}\nabla f(x)^T d(x) \end{aligned} \quad (3.22)$$

Reescrevendo (3.19), temos

$$f(x + \alpha d(x)) \leq f(x) + b\alpha 2 \cdot \frac{1}{2} \nabla f(x)^T d(x)$$

que substituindo (3.22), fica

$$f(x + \alpha d(x)) \leq f(x) + \alpha c \tau(x)$$

no qual $c = 2b \in (0, 1)$.

Teorema 3.1.2. *Seja $\bar{x} \in U$ um ponto não estacionário de F . Então para qualquer $0 < c < 1$ existe $r > 0$ e $\alpha_0 \in (0, 1]$ tal que $B[\bar{x}, r] \subset U$ e para todo $x \in B[\bar{x}, r]$,*

$$x + \alpha d(x) \in U \text{ e } f_j(x + \alpha d(x)) \leq f_j(x) + c\alpha \tau(x)$$

para todo $\alpha \in [0, \alpha_0]$ e $j = 1, \dots, m$

Demonstração. Como \bar{x} é não crítico, pelo Lema 3.1.1, item (b2), $\tau(\bar{x}) < 0$. Usando a continuidade de τ , temos que dado $\bar{\epsilon} \leq -\frac{\tau(\bar{x})}{2}$, existe $r_{\bar{\epsilon}} > 0$ tal que $\|x - \bar{x}\| \leq r_{\bar{\epsilon}}$, temos

$$|\tau(x) - \tau(\bar{x})| \leq \bar{\epsilon} \leq -\frac{\tau(\bar{x})}{2}$$

Pela definição de módulo e somando todos os membros por $\tau(\bar{x})$ temos

$$\tau(x) \leq \frac{\tau(\bar{x})}{2}, \forall x \in B[\bar{x}, r] \quad (3.23)$$

No qual $B[\bar{x}, r] \subset U$. Em particular, qualquer $x \in B[\bar{x}, r]$ é não crítico.

Pelo Lema 3.1.1 item (c), d é limitado em conjuntos compactos, i.e., d é limitado no conjunto $B[\bar{x}, r]$. Logo, existe $0 < \alpha_1 \leq 1$ tal que $x + \alpha d(x) \in U$ para $x \in B[\bar{x}, r]$ e $\alpha \in [0, \alpha_1]$. Para $x \in B[\bar{x}, r]$ e $\alpha \in [0, \alpha_1]$

$$f_j(x + \alpha d(x)) = f_j(x) + \alpha \nabla f_j(x)^T d(x) + o_j(\alpha d(x), x), \quad j = 1, 2, \dots, m$$

onde

$$\lim_{\alpha \rightarrow 0^+} \frac{o_j(\alpha d(x), x)}{\alpha \|d(x)\|} = 0$$

Usando novamente o fato que d é limitado na bola $B[\bar{x}, r]$,

$$\lim_{\alpha \rightarrow 0^+} \frac{o_j(\alpha d(x), x)}{\alpha} = 0$$

Como $\nabla f_j(x)^T d(x) \leq \tau(x)$, com $\alpha \in [0, \alpha_1]$, $j = 1, 2, \dots, m$, temos:

$$\begin{aligned} f_j(x + \alpha d(x)) &\leq f_j(x) + \alpha \tau(x) + o_j(\alpha d(x), x) \\ &= f_j(x) + \alpha c \tau(x) + \alpha \left[(1 - c) \tau(x) + \frac{o_j(\alpha d(x), x)}{\alpha} \right] \end{aligned}$$

Por (3.23) e tomando α suficientemente pequeno

$$f_j(x + \alpha d(x)) \leq f_j(x) + \alpha c \tau(x)$$

□

Lema 3.1.2. *Seja $V \subset U$ um conjunto convexo e que existam $\delta, \epsilon \in \mathbb{R}_+$ tal que para qualquer $x, y \in V$, com $\|y - x\| < \delta$, temos $\|\nabla^2 f_j(y) - \nabla^2 f_j(x)\| < \frac{\epsilon}{2}$, $j = 1, 2, \dots, m$. Então, para qualquer $x, y \in V$ tal que $\|y - x\| < \delta$, temos:*

$$\|\nabla f_j(y) - \nabla f_j(x) - \nabla^2 f_j(x)(y - x)\| < \frac{\epsilon}{2} \|y - x\| \quad (3.24)$$

e

$$|f_j(y) - (f_j(x) + \nabla f_j(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2 f_j(x) (y - x))| < \frac{\epsilon}{4} \|y - x\|^2 \quad (3.25)$$

Demonstração. **Vamos provar a inequação (3.24)**

Faça $p = y - x$. Em (3.24), devemos mostrar que $\|\nabla f_j(x+p) - \nabla f_j(x) - \nabla^2 f_j(x)(p)\| < \frac{\epsilon}{2}\|p\|$. Por (1.2.7), temos:

$$\begin{aligned} \left\| \int_0^1 \nabla^2 f_j(x+tp)p \, dt - \nabla^2 f_j(x)p \right\| &\leq \int_0^1 \|\nabla^2 f_j(x+tp) - \nabla^2 f_j(x)\| \|p\| \, dt \\ &\leq \|\nabla^2 f_j(y) - \nabla^2 f_j(x)\| \|p\| < \frac{\epsilon}{2}\|p\| \end{aligned} \quad (3.26)$$

Provando a inequação (3.25)

Faça $p = y - x$. Devemos provar que:

$$|f_j(x+p) - (f_j(x) + \nabla f_j(x)^T p + \frac{1}{2} p^T \nabla^2 f_j(x) p)| < \frac{\epsilon}{4} \|p\|^2 \quad \forall j = 1, \dots, m$$

Observe que, pela Fórmula de Taylor de segunda ordem:

$$|f_j(x+p) - (f_j(x) + \nabla f_j(x)^T p + \frac{1}{2} p^T \nabla^2 f_j(x) p)| = o_j^2(p), \quad \forall j = 1, \dots, m$$

Logo, dividindo os dois membros por $\|p\|^2$ e para p suficientemente pequeno temos:

$$\frac{o_j^2(p)}{\|p\|^2} < \frac{\epsilon}{4}, \quad i.e., \quad o_j^2(p) < \frac{\epsilon}{4} \|p\|^2$$

Portanto, $|f_j(x+p) - (f_j(x) + \nabla f_j(x)^T p + \frac{1}{2} p^T \nabla^2 f_j(x) p)| < \frac{\epsilon}{4} \|p\|^2, \quad \forall j = 1, \dots, m \quad \square$

No próximo lema, usaremos um resultado da convergência do método Quase-Newton em otimização clássica. Verifique que a ordem de convergência do método Quase-Newton é superlinear com as hipóteses do Teorema 1.3.1 e que

$$\lim_{k \rightarrow \infty} \frac{\|[B(x^k) - \nabla^2 f(x^*)]s^k\|}{\|s^k\|} = 0 \quad (3.27)$$

Na otimização Multiobjetivo, não podemos conseguir uma das hipóteses da Otimização Clássica que é uma solução x^* para o qual $\nabla f_j(x^*) = 0$, para todo $j = 1, 2, \dots, m$. O limite (3.27), citado em [22], será utilizado como hipótese no decorrer do nosso trabalho. Tomando $\alpha_k = 1$, temos

$$\lim_{k \rightarrow \infty} \frac{\|[B_j(x^k) - \nabla^2 f_j(x^*)]d^k\|}{\|d^k\|} = 0, \quad j = 1, 2, \dots, m$$

Logo, para um $\epsilon \in R_+$, existe $k_0 \in \mathbb{N}$ tal que para todo $k \geq k_0$, podemos escrever

$$\frac{\|[B_j(x^k) - \nabla^2 f_j(x^*)]d^k\|}{\|d^k\|} < \frac{\epsilon}{2}, \quad j = 1, 2, \dots, m$$

Lema 3.1.3. *Seja $V \subset U$ um conjunto convexo e $\delta, \epsilon \in \mathbb{R}_+$ tal que qualquer $x, y \in V$, com $\|y - x\| < \delta$ temos $\|\nabla^2 f_j(y) - \nabla^2 f_j(x)\| < \frac{\epsilon}{2} \quad \forall j = 1, \dots, m$. Seja $\{x^k\}$ e $B_j(x^k)$, $j = 1, \dots, m$ sequências, nas quais, existe $k_0 \in \mathbb{N}$ tal que para todo $k \geq k_0$,*

$$\frac{\|[\nabla^2 f_j(x^k) - B_j(x^k)](y - x^k)\|}{\|y - x^k\|} < \frac{\epsilon}{2} \quad \forall j = 1, \dots, m.$$

Para qualquer x^k , $k \geq k_0$, e qualquer $y \in V$ tal que $\|y - x^k\| < \delta$, temos:

$$\|\nabla f_j(y) - [\nabla f_j(x^k) + B_j(x^k)(y - x^k)]\| < \epsilon \|y - x^k\| \quad (3.28)$$

e

$$\left| f_j(y) - [f_j(x^k) + \nabla f_j(x^k)^T (y - x^k) + \frac{1}{2}(y - x^k)^T B_j(x^k)(y - x^k)] \right| < \frac{\epsilon}{2} \|y - x^k\|^2 \quad (3.29)$$

Demonstração. Provando a inequação (3.28)

Usando a inequação (3.24)

$$\begin{aligned} \|\nabla f_j(y) - [\nabla f_j(x^k) + B_j(x^k)(y - x^k)]\| &\leq \|\nabla f_j(y) - \nabla f_j(x^k) - \nabla^2 f_j(x^k)(y - x^k)\| + \\ &\quad + \|(\nabla^2 f_j(x^k) - B_j(x^k))(y - x^k)\| \\ &< \frac{\epsilon}{2} \|y - x^k\| + \frac{\epsilon}{2} \|y - x^k\| = \epsilon \|y - x^k\| \end{aligned}$$

Provando a inequação (3.29)

Utilizando a inequação (3.25), temos

$$\begin{aligned} &\left| f_j(y) - [f_j(x^k) + \nabla f_j(x^k)^T (y - x^k) + \frac{1}{2}(y - x^k)^T B_j(x^k)(y - x^k)] \right| \leq \\ &\leq \left| f_j(y) - f_j(x^k) - \nabla f_j(x^k)^T (y - x^k) - \frac{1}{2}(y - x^k)^T \nabla^2 f_j(x^k)(y - x^k) \right| + \\ &+ \left| \frac{1}{2}(y - x^k)^T [\nabla^2 f_j(x^k) - B_j(x^k)](y - x^k) \right| < \frac{\epsilon}{4} \|y - x^k\|^2 + \frac{\epsilon}{4} \|y - x^k\|^2 = \frac{\epsilon}{2} \|y - x^k\|^2. \square \end{aligned}$$

Lema 3.1.4. *Seja $x \in U$ e $a, b \in \mathbb{R}_+$ tal que $a \leq b$. Se $aI \leq B_j(x) \leq bI$, $j = 1, \dots, m$, então*

$$\frac{a}{2} \|d(x)\|^2 \leq |\tau(x)| \leq \frac{b}{2} \|d(x)\|^2$$

Demonstração. Defina

$$H := \sum_{j=1}^m \lambda_j(x) B_j(x) \text{ e } v := \sum_{j=1}^m \lambda_j(x) \nabla f_j(x) \quad (3.30)$$

Observe que por hipótese, $aI \leq B_j(x) \leq bI$. Pré-multiplicando por $\sum_{j=1}^m \lambda_j(x)$ temos:

$$\sum_{j=1}^m \lambda_j(x) aI \leq \sum_{j=1}^m \lambda_j(x) B_j(x) \leq \sum_{j=1}^m \lambda_j(x) bI$$

Ou seja,

$$aI \leq H \leq bI \quad (3.31)$$

Substituindo H e v na equação (3.13) temos:

$$d(x) = -H^{-1}v$$

Substituindo na função (3.5) temos:

$$\begin{aligned} \tau(x) &= v^T d(x) + \frac{1}{2} d(x)^T H d(x) \\ &= (-H d(x))^T d(x) + \frac{1}{2} d(x)^T H d(x) \\ &= -d(x)^T H d(x) + \frac{1}{2} d(x)^T H d(x) \\ &= -\frac{1}{2} d(x)^T H d(x) \end{aligned}$$

Tomando a desigualdade (3.31) e pré-multiplicando por $\frac{1}{2} d(x)^T d(x)$ temos:

$$\frac{a}{2} \|d(x)\|^2 \leq -\tau(x) \leq \frac{b}{2} \|d(x)\|^2$$

utilizando o fato que $\tau(x) \leq 0$, concluímos que

$$\frac{a}{2} \|d(x)\|^2 \leq |\tau(x)| \leq \frac{b}{2} \|d(x)\|^2$$

□

Lema 3.1.5. *Sejam $x \in U$ e $0 < a$. Se $aI \leq B_j(x)$ para todo $j = 1, \dots, m$, então*

$$|\tau(x)| \leq \frac{1}{2a} \left\| \sum_{j=1}^m \lambda_j \nabla f_j(x) \right\|^2$$

para todo $\lambda_j \geq 0$, $j = 1, \dots, m$, com $\sum_{j=1}^m \lambda_j = 1$

Demonstração. Seja $\lambda_j \geq 0$, com $j = 1, \dots, m$ e $\sum_{j=1}^m \lambda_j = 1$. Definimos $w :=$

$$\sum_{j=1}^m \lambda_j \nabla f_j(x)$$

Substituindo w na inequação dual (3.14), temos:

$$\begin{aligned}\tau(x) &\geq \inf_{s \in \mathbb{R}^n} \left(w^T s + \sum_{j=1}^m \lambda_j \frac{1}{2} s^T B_j(x) s \right) \\ &\geq \inf_{s \in \mathbb{R}^n} \left(w^T s + \frac{1}{2} a \|s\|^2 \right) \\ &= -\frac{\|w\|^2}{2a}\end{aligned}$$

Como $\tau(x) \leq 0$, $\forall x \in U$, temos:

$$-\frac{\|w\|^2}{2a} \leq \tau(x) \leq 0 \leq -\tau(x) \leq \frac{\|w\|^2}{2a}$$

Portanto,

$$|\tau(x)| \leq \frac{1}{2a} \left\| \sum_{j=1}^m \lambda_j \nabla f_j(x) \right\|^2$$

□

3.2 Um algoritmo do tipo Quase-Newton para Otimização Multiobjetivo (caso convexo)

A seguir, apresentamos o algoritmo para o caso convexo. Inicialmente no passo 1, entramos com um elemento qualquer $x^0 \in U$, uma matriz simétrica definida positiva $B_0 \in \mathbb{R}^{n \times n}$, os valores de $c \in (0, 1)$ e $\epsilon > 0$, suficientemente pequeno. No passo 2 é realizada a resolução do problema (3.4) que é equivalente ao sistema (3.7), no qual também é calculada o valor de $\tau(x^k)$ e $d(x^k)$. Pelo Lema 3.1.1, item (b), se obtermos um valor $x^k \in U$, onde $\tau(x^k) = 0$, o elemento $x^k \in U$ é ponto crítico. Se isso acontecer, o algoritmo para. O parâmetro ϵ serve em nosso algoritmo como critério de parada para implementação computacional. Se $\tau(x^k) < 0$, avançamos para o passo 4, para o cálculo do tamanho do passo α^k . Neste momento, tomamos α^k como o maior valor do conjunto $M = \{\frac{1}{2^n}; n = 0, 1, 2, \dots\}$, que satisfaça as condições do Teorema (3.1.2). De posse do tamanho do passo α^k , nos passos 5 e 6, calculamos a atualização x^{k+1} , onde $x^{k+1} := x^k + \alpha^k d(x^k)$ e também é atualizada a matriz aproximação $B_j(x^{k+1})$, utilizando a fórmula (3.3). Com estas informações volta-se para o passo 2. Por comodidade, em alguns casos, valem as seguintes igualdades: $f(x^k) = f^k$, $\tau(x^k) = \tau^k$, $B_j(x^k) = B_j^k$, $B_j(x^{k+1}) = B_j^{k+1}$, $d(x^k) = d^k$ e $d(x^{k+1}) = d^{k+1}$.

Algoritmo 3.2.1 (Algoritmo Quase-Newton para Otimização Multiobjetivo).

- **Passo 1. Início** Dados $x_0 \in U$, uma matriz simétrica, definida positiva $B_0 \in \mathbb{R}^{n \times n}$ $c \in (0, 1)$ e $\epsilon > 0$. Defina $M = \{1/2^n; n = 0, 1, 2, \dots\}$
- **Passo 2. Direção de descida** Resolver o problema Quase-Newton (3.4) e obter (d^k) e τ^k resolvendo (3.5) e (3.6).
- **Passo 3. Parada** Se $\tau^k > -\epsilon$, então PARE!. Senão, continue no passo 4.
- **Passo 4. Busca linear** Escolha α^k o maior valor de $\alpha \in M$ tal que $x^k + \alpha d^k \in U$ e $f_j(x^k + \alpha d^k) \leq f_j^k + c\alpha\tau^k$, $j = 1, 2, \dots, m$
- **Passo 5. Atualização** Defina $x^{k+1} := x^k + \alpha^k d^k$. Atualize B_j^k para B_j^{k+1} com o método BFGS (3.3), $j = 1, 2, \dots, m$.
- **Passo 6.** Faça $k := k + 1$ e vá para o Passo 2.

Agora estudaremos o teorema principal. Seus principais objetivos são provar a ordem de convergência superlinear e a convergência da sequência $\{x^k\}$ gerada pelo Algoritmo 3.2.1 para um ponto $x^* \in U \subseteq \mathbb{R}^n$, Pareto ótimo local.

Considerações sobre as hipóteses do Teorema 3.2.1

Nas hipóteses de (i) é assegurada sua existência pelo Teorema 1.2.9. A hipótese (ii) é garantida pelo fato que a função é fortemente convexa e C^2 , em um conjunto convexo V , logo, a Hessiana de f é localmente Lipschitz. Em (iii), é válido, pois é assegurada pelo Lema 3.1.3 (Condição de Dennis Moré). Em (iv), a bola fechada $B[x_0, r]$ nos auxiliará por exemplo, a usar a hipótese de que a direção Quase-Newton $d(x)$ é limitada uniformemente em conjuntos compactos, provado no Lema 3.1.1. Na relação da hipótese (v), será calculado o valor de a utilizada, por exemplo, na hipótese (i), no qual os valores de ϵ é dado e $c \in (0, 1)$ e em (vi), $d(x_0)$ vai ser controlado pelos valores de δ, r e $\frac{\epsilon}{a}$.

Teorema 3.2.1 (Teorema Principal do Caso Convexo). *Sejam V um conjunto convexo, U um conjunto aberto, no qual $V \subset U \subset \mathbb{R}^n$ e $\{x^k\}$ a sequência gerada pelo Algoritmo (3.2.1). Dados $c \in (0, 1)$, $k_0 \in \mathbb{N}$ e $a, b, r, \delta, \epsilon \in \mathbb{R}_+^*$, considere as seguintes hipóteses:*

- (i) $aI \leq B_j(x) \leq bI$ e $aI \leq \nabla^2 f_j(x) \leq bI$, $\forall x \in V$, $j = 1, \dots, m$,
- (ii) $\|\nabla^2 f_j(y) - \nabla^2 f_j(x)\| < \epsilon/2$, $\forall x, y \in V$ com $\|y - x\| < \delta$, $j = 1, \dots, m$,
- (iii) $\|(B_j(x^k) - \nabla^2 f_j(x^k))(y - x^k)\| < (\epsilon/2)\|y - x^k\|$, $\forall k \geq k_0$, $y \in V$, $j = 1, \dots, m$,
- (iv) $B[x_0, r] \subset V$
- (v) $\epsilon/a \leq 1 - c$
- (vi) $\|d(x_0)\| < \min\{\delta, r(1 - (\epsilon/a))\}$

Então, para todo $k \geq k_0$, temos:

- (a) $\|x^k - x_0\| \leq \|d(x_0)\| \frac{1 - (\epsilon/a)^k}{1 - (\epsilon/a)}$
- (b) $\|d(x^k)\| \leq \|d(x_0)\| (\epsilon/a)^k$
- (c) $x^k, x^k + d(x^k) \in B[x_0, r]$, $\|(x^k + d(x^k)) - x^k\| < \delta$ e $\alpha_k = 1$
- (d) $\|d(x^{k+1})\| \leq \|d(x^k)\| (\epsilon/a)$
- (e) A sequência $\{x^k\}$ converge para algum Pareto ótimo local $x^* \in \mathbb{R}^m$
- (f) A ordem de convergência da sequência é superlinear.

Demonstração. Considerando como verdadeiros os itens (a) e (b) provaremos os itens (c) e (d):

Prova do item (c)

Pela desigualdade triangular e utilizando os itens (a) e (b) do Teorema (3.2.1), temos:

$$\begin{aligned} \|x^k + d(x^k) - x_0\| &\leq \|x^k - x_0\| + \|d(x^k)\| \\ &\leq \|d(x_0)\| \frac{1 - (\epsilon/a)^k}{1 - (\epsilon/a)} + \|d(x_0)\| (\epsilon/a)^k \end{aligned}$$

ou seja,

$$\|x^k + d(x^k) - x_0\| \leq \|d(x_0)\| \left(\frac{1 - (\epsilon/a)^{k+1}}{1 - (\epsilon/a)} \right) \quad (3.32)$$

Na desigualdade (3.32), utilizando as hipóteses (v) e (vi) tem-se:

$$\|(x^k + d(x^k)) - x_0\| \leq \frac{1 - (\epsilon/a)^{k+1}}{1 - (\epsilon/a)} r(1 - (\epsilon/a)) = r[1 - (\epsilon/a)^{k+1}] \leq r$$

i.e., $x^k + d(x^k) \in B[x_0, r]$

Do item (a), fica:

$$\|x^k - x_0\| \leq \|d(x_0)\| \frac{1 - (\epsilon/a)^k}{1 - (\epsilon/a)}$$

Utilizando (vi) temos:

$$\|x^k - x_0\| \leq r(1 - \epsilon/a) \frac{1 - (\epsilon/a)^k}{1 - (\epsilon/a)} \leq r[1 - (\epsilon/a)^k] \leq r$$

i.e., $x^k \in B[x_0, r]$

Por (b), utilizando as hipóteses (v) e (vi)

$$\|(x^k + d(x^k)) - x^k\| = \|d(x^k)\| \leq \|d(x_0)\| (\epsilon/a)^k \leq \delta (\epsilon/a)^k \leq \delta$$

ou seja, $\|(x^k + d(x^k)) - x^k\| \leq \delta$

Pelo Lema (3.1.3) na desigualdade (3.29)

$$f_j(x^k + d(x^k)) \leq f_j(x^k) + \nabla f_j(x^k)^T d(x^k) + \frac{1}{2} d(x^k)^T B_j^k d(x^k) + \frac{\epsilon}{2} \|d(x^k)\|^2$$

que substituindo pela função (3.5) vale

$$f_j(x^k + d(x^k)) \leq f_j(x^k) + \tau(x^k) + \frac{\epsilon}{2} \|d(x^k)\|^2 = f_j(x^k) + c\tau(x^k) + (1-c)\tau(x^k) + \frac{\epsilon}{2} \|d(x^k)\|^2$$

Pelo Lema (3.1.4), temos a desigualdade $\frac{a}{2} \|d(x^k)\|^2 \leq |\tau(x^k)|$. Pelo Lema (3.1.1), $\tau(x^k) \leq 0$, $\forall x^k \in U$, então

$$\tau(x^k) \leq -\frac{a}{2} \|d(x^k)\|^2$$

Logo, usando a hipótese (v), fica

$$(1-c)\tau(x^k) + \frac{\epsilon}{2} \|d(x^k)\|^2 \leq -\frac{a}{2}(1-c)\|d(x^k)\|^2 + \frac{a}{2}(1-c)\|d(x^k)\|^2 = 0$$

Ou seja,

$$f_j(x^k + d(x^k)) \leq f_j(x^k) + c\tau(x^k), \quad c \in (0, 1), \quad j = 1, \dots, m$$

Pelo Teorema (3.1.2), concluímos que $\alpha_k = 1$.

Prova do item (d)

Como $\alpha_k = 1$, temos que

$$x^{k+1} = x^k + d(x^k), \quad x^k, x^{k+1} \in B[x_0, r] \text{ e } \|x^{k+1} - x^k\| \leq \delta$$

Definimos

$$v^{k+1} := \sum_{j=1}^m \lambda_j(x^k) \nabla f_j(x^{k+1}) \quad (3.33)$$

Substituindo (3.33) no resultado do Lema (3.1.5), temos

$$|\tau^{k+1}| \leq \frac{1}{2a} \|v^{k+1}\|^2 \quad (3.34)$$

Para $x \in U$ definimos:

$$G^k(x) := \sum_{j=1}^m \lambda_j^k f_j(x) \quad \text{e} \quad H^k := \sum_{j=1}^m \lambda_j^k B_j^k \quad (3.35)$$

No qual podemos escrever

$$\nabla G^k(x) = \sum_{j=1}^m \lambda_j^k \nabla f_j(x) \quad \text{e} \quad \nabla^2 G^k(x) = \sum_{j=1}^m \lambda_j^k \nabla^2 f_j(x)$$

Verificamos que

$$v^{k+1} = \nabla G^k(x^{k+1}) = \sum_{j=1}^m \lambda_j^k \nabla f_j(x^{k+1}) \quad (3.36)$$

Que substituindo na desigualdade (3.13)

$$d^k = -(H^k)^{-1} \nabla G^k(x^k)$$

i.e.,

$$H^k d^k + \nabla G^k(x^k) = 0 \quad (3.37)$$

As hipóteses (ii) e (iii) ficam respectivamente

$$\begin{aligned} \|\nabla^2 G^k(y) - \nabla^2 G^k(x)\| &< \frac{\epsilon}{2} \\ \|(H^k - \nabla^2 G^k(x^k))(y - x^k)\| &< \frac{\epsilon}{2} \|y - x^k\| \end{aligned}$$

Desta forma, podemos reescrever a inequação (3.28) do Lema (3.1.3) como

$$\|\nabla G^k(x^k + d^k) - (\nabla G^k(x^k) + H^k d^k)\| < \epsilon \|d^k\|$$

que substituindo (3.37) e usando a relação (3.36) temos

$$\|v^{k+1}\| = \|\nabla G^k(x^{k+1})\| < \epsilon \|d^k\|$$

Substituindo em (3.34) temos

$$|\tau^{k+1}| \leq \frac{1}{2a} \|v^{k+1}\|^2 < \frac{\epsilon^2}{2a} \|d^k\|$$

Pelo Lema (3.1.4) e o item (i) obtemos

$$\frac{a}{2} \|d(x^{k+1})\|^2 \leq |\tau(x^{k+1})| < \frac{\epsilon^2}{2a} \|d^k\|^2$$

ou seja

$$\|d^{k+1}\| < \frac{\epsilon}{a} \|d^k\|$$

Prova do item (a)

Vamos provar por indução. Devemos verificar se a inequação vale para $k = 0$ e $k := k + 1$.

Verificando quando $k = 0$, temos

$$\|x_0 - x_0\| \leq \|d(x_0)\| \frac{1 - (\epsilon/a)^0}{1 - (\epsilon/a)}$$

o que torna a sentença verdadeira.

Verificando para $k := k + 1$. Como foi demonstrado na prova do item (c), inequação (3.32) e $\alpha_k = 1$, temos

$$\|(x^k + d^k) - x_0\| \leq \frac{1 - (\epsilon/a)^{k+1}}{1 - (\epsilon/a)} \|d(x_0)\|$$

Ou seja,

$$\|x^{k+1} - x_0\| \leq \frac{1 - (\epsilon/a)^{k+1}}{1 - (\epsilon/a)} \|d(x_0)\|$$

Prova do item (b)

Provaremos por indução

Para $k = 0$ temos

$$\|d(x_0)\| \leq (\epsilon/a)^0 \|d(x_0)\|$$

O que torna a sentença verdadeira.

Para $k := k + 1$ temos que provar $\|d(x^{k+1})\| \leq (\epsilon/a)^{k+1} \|d(x_0)\|$.

Utilizando a inequação do item (d) temos:

$$\|d(x^{k+1})\| \leq (\epsilon/a) \|d(x^k)\| \leq (\epsilon/a)(\epsilon/a)^k \|d(x_0)\| = (\epsilon/a)^{k+1} \|d(x_0)\|$$

Prova do item (e)

Pelos itens provados anteriormente, é verdade que

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\| = \sum_{k=0}^{\infty} \|d(x^k)\| \leq \|d(x_0)\| \sum_{k=0}^{\infty} (\epsilon/a)^k < \infty$$

Logo, $\{x^k\}$ é uma sequência de Cauchy, i.e., existe $x^* \in \mathbb{R}^n$ tal que

$$\lim_{k \rightarrow \infty} x^k = x^*$$

Veja que, como τ é uma função contínua temos

$$\lim_{k \rightarrow \infty} \tau(x^k) = \tau(x^*) \quad (3.38)$$

Mais ainda, pelo item (b) e (v)

$$\lim_{k \rightarrow \infty} \|d(x^k)\| = 0 \quad (3.39)$$

Pelo Lema (3.1.4) temos

$$\frac{a}{2} \|d(x^k)\|^2 \leq |\tau(x^k)| \leq \frac{b}{2} \|d(x^k)\|^2$$

Que passando o limite, quando $k \rightarrow \infty$

$$\frac{a}{2} \lim_{k \rightarrow \infty} \|d(x^k)\|^2 \leq \lim_{k \rightarrow \infty} |\tau(x^k)| \leq \frac{b}{2} \lim_{k \rightarrow \infty} \|d(x^k)\|^2$$

substituindo (3.38) e (3.39)

$$0 \leq |\tau(x^*)| \leq 0$$

Logo

$$\tau(x^*) = 0$$

As equivalências de (b) no Lema 3.1.1 também valem para as negações. Logo, x^* é ponto estacionário de F e por (c) do Teorema 3.1.1, x^* é Pareto ótimo de F restrito

em V , i.e., x^* é Pareto ótimo local.

Prova do item (f)

Neste item provamos a ordem de convergência superlinear do método Quase-Newton do algoritmo (3.2.1)

Definimos

$$r^k = \|d(x_0)\| \frac{(\epsilon/a)^k}{1 - (\epsilon/a)} \quad e \quad \delta^k = \|d(x_0)\| (\epsilon/a)^k$$

Pela desigualdade triangular, (iv) e (vi), temos

$$B_2[x^k, r^k] \subset B_1[x_0, r] \subset V \quad (3.40)$$

Dado $\varphi \in \mathbb{R}_+^*$ arbitrário, definimos $\hat{\epsilon} := \min\{\frac{a\varphi}{1 + 2\varphi}, \epsilon\}$

Para k suficientemente grande temos:

$$\|\nabla^2 f_j(y) - \nabla^2 f_j(x)\| < \frac{\hat{\epsilon}}{2}, \quad \forall x, y \in B[x^k, r^k] \quad \text{com} \quad \|y - x\| < \delta^k \quad (3.41)$$

$$\|(B_j^l - \nabla^2 f_j^l)(y - x^l)\| < \frac{\hat{\epsilon}}{2}, \quad \forall y \in B[x^k, r^k] \quad e \quad l \geq k, \quad j = 1, \dots, m \quad (3.42)$$

As hipóteses (i) e (vi) são satisfeitas trocando os valores de ϵ, r, δ e x_0 por $\hat{\epsilon}, r^k, \delta^k$ e x^k respectivamente. O item (iv) será satisfeito por (3.40) e os itens (ii) e (iii) são as inequações (3.41) e (3.42) respectivamente. O item (vi) é verdadeiro pelas definições de r^k, δ^k e $\hat{\epsilon}$ e (v) pelas definições de ϵ, a e c .

Tomando o item (a), aplicando o limite quando $k \rightarrow \infty$ e trocando x_0 e ϵ por \hat{x}_0 e $\hat{\epsilon}$ temos

$$\|x^* - x^k\| \leq \|d(x^k)\| \frac{1}{1 - (\hat{\epsilon}/a)}$$

A desigualdade acima também vale para $k := k + 1$ e aplicando (d) temos

$$\|x^* - x^{k+1}\| \leq \|d(x^{k+1})\| \frac{1}{1 - (\hat{\epsilon}/a)} \leq \|d(x^k)\| \frac{(\hat{\epsilon}/a)}{1 - (\hat{\epsilon}/a)} \quad (3.43)$$

Usando a definição de x^{k+1} e a inequação (3.43), temos

$$\begin{aligned} \|x^* - x^k\| &\geq \|x^{k+1} - x^k\| - \|x^* - x^{k+1}\| \\ &\geq \|d(x^k)\| - \|d(x^k)\| \frac{(\hat{\epsilon}/a)}{1 - (\hat{\epsilon}/a)} \\ &= \|d(x^k)\| \left[1 - \frac{(\hat{\epsilon}/a)}{1 - (\hat{\epsilon}/a)} \right] \\ &= \|d(x^k)\| \left[\frac{1 - (\hat{\epsilon}/a) - (\hat{\epsilon}/a)}{1 - (\hat{\epsilon}/a)} \right] \end{aligned}$$

Logo

$$\|x^* - x^k\| \geq \|d(x^k)\| \frac{1 - (2\hat{\epsilon}/a)}{1 - (\hat{\epsilon}/a)} \quad (3.44)$$

Pela definição de $\hat{\epsilon}$, temos que $(1 - (2\hat{\epsilon}/a)) > 0$. De fato

$$\hat{\epsilon} = \frac{a\varphi}{1 + 2\varphi} \Rightarrow \hat{\epsilon} + 2\hat{\epsilon}\varphi = a\varphi$$

Ou seja

$$\varphi = \frac{\hat{\epsilon}}{a - 2\hat{\epsilon}}$$

Dividindo por a e pela substituição de ϵ por $\hat{\epsilon}$

$$\varphi = \frac{\epsilon/a}{1 - (2\epsilon/a)} \quad (3.45)$$

Como por hipótese φ , ϵ e $a \in \mathbb{R}_+^*$, o denominador $1 - (2\epsilon/a) > 0$. Mais ainda, $(\epsilon/a) < 1/2$, e quanto menor o valor de (ϵ/a) , temos $0 < \varphi < 1$

De (3.43) e (3.44) temos

$$\|x^* - x^{k+1}\| \leq \|x^* - x^k\| \frac{\hat{\epsilon}/a}{1 - (2\hat{\epsilon}/a)}$$

De fato, pois de (3.44) temos

$$\|d(x^k)\| \leq \|x^* - x^k\| \frac{1 - (\hat{\epsilon}/a)}{1 - (2\hat{\epsilon}/a)}$$

que substituindo em (3.43) fica

$$\|x^* - x^{k+1}\| \leq \|x^* - x^k\| \frac{(\hat{\epsilon}/a)}{1 - (\hat{\epsilon}/a)} \frac{1 - (\hat{\epsilon}/a)}{1 - (2\hat{\epsilon}/a)} = \|x^* - x^k\| \frac{(\hat{\epsilon}/a)}{1 - (2\hat{\epsilon}/a)}$$

Pela substituição de ϵ por $\hat{\epsilon}$ e pela equação (3.45), o quociente $\frac{(\hat{\epsilon}/a)}{1 - (2\hat{\epsilon}/a)} = \varphi$

Logo

$$\|x^* - x^{k+1}\| \leq \varphi \|x^* - x^k\|$$

Como $\varphi \in \mathbb{R}_+^*$ é arbitrário e está tendendo a zero para (ϵ/a) pequeno, a sequência $\{x^k\}$ converge superlinearmente para x^* \square

3.3 Um algoritmo Quase-Newton para Otimização Multiobjetivo (caso não convexo)

Nesta seção, iniciamos o estudo sobre Otimização de problemas Multiobjetivo com funções não convexas e com convergência global. Uma das principais dificuldades para incrementação deste algoritmo, é a garantia da positividade da matriz atualização BFGS Quase-Newton $B_j(x^{k+1})$ com a condição de curvatura $y_j^{k(T)} s^k > 0$, para $j = 1, 2, \dots, m$. Fukushima em 2001 em [28] propôs uma modificação na atualização BFGS $B(x^k)$, através de um sistema de atualização. Também em 2001, em outro trabalho de Fukushima [7], propôs uma outra modificação da atualização BFGS. Seu trabalho era voltado para Otimização Clássica, no qual a modificação, era um sistema da forma

$$B(x^{k+1}) = \begin{cases} B(x^k) - \frac{B(x^k)s_k s_k^T B(x^k)}{s_k^T B(x^k)s_k} + \frac{y_k y_k^T}{y_k^T s_k}, & \text{se } y_k^T s_k > 0 \\ B(x^k), & \text{se } y_k^T s_k \leq 0 \end{cases} \quad (3.46)$$

Com este sistema, garante-se a positividade da matriz atualização e a condição de curvatura $y_k^T s_k > 0$. Computacionalmente, $y_k^T s_k > 0$ é substituída pela condição $y_k^T s_k > \eta$, onde $\eta > 0$ é suficientemente pequeno.

Para a solução de problemas de Otimização Multiobjetivo não convexas, seguiremos a mesma linha de raciocínio de [7] com o sistema 3.46, porém usando a condição η , utilizada em [8]. Logo

$$B_j(x^{k+1}) = \begin{cases} B_j(x^k) - \frac{B_j(x^k)s_k s_k^T B_j(x^k)}{s_k^T B_j(x^k)s_k} + \frac{y_{jk} y_{jk}^T}{y_{jk}^T s_k}, & \text{se } y_{jk}^T s_k > \epsilon \min\{-\tau(x^k), 1\} \\ B_j(x^k), & \text{se } y_{jk}^T s_k \leq \epsilon \min\{-\tau(x^k), 1\} \end{cases} \quad (3.47)$$

no qual $j = 1, 2, \dots, m$ e $\tau(x)$ é definida em (3.5)

Desta forma, para qualquer $j = 1, 2, \dots, m$, $B_j(x^{k+1})$ herda a limitação positiva de $B_j(x^k)$.

Definição 3.3.1. *Definimos como direção Quase-Newton, para ϵ suficientemente pequeno, a solução ótima $d(x)$, do problema*

$$SP_\epsilon(x) = \begin{cases} \min t \\ \text{s.a. } \nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d \leq t, \quad j = 1, 2, \dots, m \\ \|d\| \leq 1, \quad t \leq -\epsilon, \quad d \in \mathbb{R}^n \end{cases} \quad (3.48)$$

Onde $B_j(x)$ é uma matriz aproximação para a Hessiana de cada f_j obtida pelo sistema BFGS Quase-Newton em (3.47). A limitação $\|d\| \leq 1$ é usada para eliminar o caso de $\|d\| \rightarrow \infty$

Lema 3.3.1. *Seja $x \in X \subset \mathbb{R}^n$, no qual X é um conjunto aberto. Então,*

- (i) *Se $\epsilon = 0$, o conjunto viável do problema (3.48) é não vazio.*
- (ii) *Se $B_j(x)$, $j \in j = 1, 2, \dots, m$ uma matriz definida positiva, $\epsilon > 0$ suficientemente pequeno e a região viável do problema (3.48) é não vazio.*
- (iii) *Se a região viável do problema é vazio, então x é uma boa estimativa de ponto crítico de F .*

Demonstração. Provando o item (i)

Como por hipótese $\epsilon = 0$, temos pelo problema

$$\nabla f_j(x)^T d + \frac{1}{2} d^T B_j(x) d \leq 0$$

A desigualdade é satisfeita quando $d = 0$. Portanto, a região factível do problema (3.48) não pode ser vazia.

Provando o item (ii)

Tome $d(x)$ um ponto viável do problema (3.48). A seguinte inequação é verdadeira

$$\nabla f_j(x)^T d(x) + \frac{1}{2} d(x)^T B_j(x) d(x) \leq t$$

ou seja,

$$\nabla f_j(x)^T d(x) \leq t - \frac{1}{2} d(x)^T B_j(x) d(x) \leq t \leq -\epsilon < 0, \quad j = 1, 2, \dots, m$$

Pela Teorema 2.4.1, x é não crítico e $d(x)$ é direção de descida de F .

Provando o item (iii)

Se a região viável do problema (3.48) é vazia com $\epsilon > 0$ suficientemente pequeno, então chegamos a conclusão que não existe direção de descida. Logo, x é ponto crítico de F . Suponha que exista direção de descida $\bar{d} \in \mathbb{R}^n$, i.e., $\nabla f_j(x)^T \bar{d} < 0$. Logo, existe um escalar positivo $\bar{\alpha}$ tal que $\forall \alpha \in (0, \bar{\alpha}]$

$$\begin{aligned} \nabla f_j(x)^T \alpha \bar{d} + \frac{1}{2} \alpha \bar{d}^T B_j(x) \alpha \bar{d} &= \alpha \nabla f_j(x)^T \bar{d} + \frac{\alpha^2}{2} \bar{d}^T B_j(x) \bar{d} \\ &= \alpha \left(\nabla f_j(x)^T \bar{d} + \frac{\alpha}{2} \bar{d}^T B_j(x) \bar{d} \right) < 0 \end{aligned}$$

Isto significa que a direção $\alpha\bar{d} \in \mathbb{R}^n$ é viável para o problema (3.48). Contradição, pois o conjunto viável é vazio. Pelo Teorema 2.4.1, existe $j_0 = j_0(d)$, tal que

$$\nabla f_{j_0}(x)^T d \geq 0, \text{ ou seja, } F'_{j_0}(x, d) \succeq 0$$

Portanto, x é um ponto crítico. □

Considerações sobre o algoritmo

Neste algoritmo, tomamos inicialmente um elemento $x^0 \in X \subseteq \mathbb{R}^n$, onde X é um conjunto aberto, dado $\epsilon > 0$ suficientemente pequeno e uma matriz $B(x^0)$ definida positiva. No passo 2, resolve-se o problema (3.48). Pelo Lema 3.3.1, se $t > -\epsilon$, $d(x)$ não é viável e portanto, a região factível do problema é vazia. Logo, $\{x^k\}$ é uma boa aproximação ao ponto crítico. Se $t \leq -\epsilon$, prosseguimos no passo 3, onde calcula-se o tamanho do passo α^k , no qual, existe $\bar{\alpha}_j > 0$, tal que $\alpha_j \in (0, \bar{\alpha}_j]$, $x^k + \alpha_j d(x^k) \in X$ e $f_j(x^k + \alpha_j d(x^k)) \leq f_j(x^k) + \beta\tau(x^k)$, $j = 1, 2, \dots, m$, onde $\beta \in (0, 1/2)$. Seja $\bar{\alpha} = \min_{j=1,2,\dots,m} \bar{\alpha}_j$, então temos $x^k + \alpha_j d(x^k) \in X$ e $\forall \alpha \in (0, \bar{\alpha}]$, $f_j(x^k + \alpha d(x^k)) \leq f_j(x^k) + \beta\tau(x^k)$, $j = 1, 2, \dots, m$. Escolhemos $\alpha^k \in (0, \bar{\alpha}]$. No passo 4, faça $x^{k+1} := x^k + \alpha^k d(x^k)$, no qual $d(x^k)$ é solução ótima do problema (3.48). Atualize $B_j(x^k)$, de acordo com o sistema (3.47). Faça $k := k + 1$ e vá para o passo 2.

Algoritmo 3.3.1 (Algoritmo Quase-Newton para otimização multiobjetivo (caso não convexo)).

- **Passo 1. Início** Sejam $x^0 \in X \subseteq \mathbb{R}^n$, no qual X é um conjunto aberto, ϵ suficientemente pequeno e positivo, $c > 0$ e uma matriz $B(x^0)$ definida positiva.
- **Passo 2. Direção de descida** Resolver o problema (3.48), i.e., até $t \leq -\epsilon$. Caso contrário, PARE.
- **Passo 3. Busca linear** Escolha α^k pela condição de Armijo, tal que $x^k + \alpha^k d \in X$ e faça $x^{k+1} := x^k + \alpha^k d(x^k)$, no qual $d(x^k)$ é a solução ótima do problema.
- **Passo 4. Atualização** Gerar a matriz definida $B_j(x^{k+1})$, de acordo com o sistema (3.47). Faça $k := k + 1$. Vá para o Passo 1.

Lema 3.3.2. Defina $\psi(x) = \sup_{\|d\| \leq 1} \max_{j \in I} \{-\nabla f_j(x)^T d\}$. Um ponto $x^* \in X \subseteq \mathbb{R}^n$ é crítico de F se, e somente se, $\psi(x^*) = 0$.

Demonstração. Como $x^* \in X \subseteq \mathbb{R}^n$ é ponto crítico de F , existe $j_0 = j_0(d)$, para todo $d \in \mathbb{R}^n$ tal que $\nabla f_{j_0}(x^*)^T d \geq 0$. Logo, $\psi(x^*) = \sup_{\substack{\|d\| \leq 1 \\ d \in \mathbb{R}^n}} \{-\nabla f_{j_0}(x^*)^T d \leq 0\} = 0$

Reciprocamente

$\psi(x^*) = \sup_{\|d\| \leq 1} \max_{j \in I} \{-\nabla f_j(x^*)^T d\} = 0$, i.e., existe $j_0 = j_0(d)$ tal que $-\nabla f_{j_0}(x^*)^T d \leq 0$, ou seja, $\nabla f_{j_0}(x^*)^T d \geq 0$, para todo $d \in \mathbb{R}^n$, com $\|d\| \leq 1$. Portanto, $x^* \in X \subseteq \mathbb{R}^n$ é ponto crítico de F . \square

Teorema 3.3.1 (Teorema Principal do Caso não Convexo). *Seja o conjunto de nível $L_0 = \{x \in \mathbb{R}^n; F(x) \preceq F(x^0)\}$ limitado e para k suficientemente grande tal que o passo $\alpha_k = 1$ é aceito e uma constante real $c > 0$ tal que $\|B_j(x)\| \leq c$, para todo $j = 1, 2, \dots, m$. Então todo ponto de acumulação, gerado pelo Algoritmo 3.3.1 é um ponto crítico da função F .*

Demonstração. Seja \bar{d}_k o máximo do problema (3.48) para $x := x^k$, temos por (3.5):

$$\tau(x^k) \leq \max_{j=1,2,\dots,m} \left[\nabla f_j(x^k)^T \bar{d}_k + \frac{1}{2} \bar{d}_k^T B_j(x^k) \bar{d}_k \right]$$

ou seja,

$$\begin{aligned} -\tau(x^k) &\geq \max_{j=1,2,\dots,m} \left[-\nabla f_j(x^k)^T \bar{d}_k - \frac{1}{2} \bar{d}_k^T B_j(x^k) \bar{d}_k \right] \\ &\geq \max_{0 \leq \alpha \leq 1} \max_{j=1,2,\dots,m} \left[-\alpha \nabla f_j(x^k)^T \bar{d}_k - \frac{\alpha^2}{2} \bar{d}_k^T B_j(x^k) \bar{d}_k \right] \\ &\geq \max_{0 \leq \alpha \leq 1} \left[-\alpha \psi(x^k) - \frac{1}{2} \alpha^2 c \right] \\ &\geq \frac{1}{2} \psi(x^k) \min \left\{ 1, \frac{\psi(x^k)}{c} \right\} \end{aligned} \quad (3.49)$$

Sem perda de generalidade, assuma que a subsequência de $\{x^k\}_{k \in \mathbb{N}}$, convirja para x^* . Por hipótese, $\alpha_k = 1$ é aceito para algum k suficientemente grande. Logo

$$\tau(x^k) = f_j(x^{k+1}) - f_j(x^k) + o(\|d^k\|^2)$$

Pelo Lema (3.3.2) e definição de $\tau(x^k)$

$$\frac{1}{2} \sum_{\substack{k \in K \\ k \geq \bar{k}}} \psi(x^k) \min \left\{ 1, \frac{\psi(x^k)}{c} \right\} \leq \sum_{k \in K} \max_{j \in \{1,2,\dots,m\}} [f_j(x_j^k) - f_j(x^{k+1})] \leq \max_{j \in \{1,2,\dots,m\}} [f_j(x^0) - f_j(x^*)] < +\infty$$

Portanto,

$$\sum_{k \in K} \psi(x^k) \min \left\{ 1, \frac{\psi(x^k)}{c} \right\} < +\infty$$

Pelo Lema (3.3.2), basta provarmos que $\psi(x^*) = 0$. Suponha que $\psi(x^*) > 0$ (pois $\psi(x^*) < 0$ não convém, pelo Lema 3.3.1), com $\beta > 0$ e $\epsilon_0 > 0$ tal que, para todo

$0 < \epsilon \leq \epsilon_0$ e todo $\|x^k - x^*\| \leq \epsilon$, temos

$$\psi(x^k) \geq \beta > 0$$

Isto significa que

$$\sum_{k \in K} \psi(x^k) \min\left\{1, \frac{\psi(x^k)}{c}\right\} \geq \sum \beta \min\left\{1, \frac{\beta}{c}\right\} = +\infty$$

Absurdo. Portanto, $\psi(x^*) = 0$ e x^* é ponto crítico de F .

□

4 Considerações Finais

O uso de conceitos de Otimalidade Pareto é fundamental para a resolução de problemas de Otimização Multiobjetivo. Neste trabalho, apresentamos estratégias do tipo Quase-Newton para a solução de problemas convexos e não convexos. Para o caso convexo, estudamos condições para que o algoritmo (caso convexo), esteja bem definido, com convergência local, fazendo uso da atualização BFGS e garantindo a condição de curvatura. Mostramos também para este caso, que a ordem de convergência é superlinear. Para o caso não convexo, mostramos a convergência global do algoritmo, suprimindo a necessidade da condição de curvatura e da positividade da matriz atualização BFGS, através de um sistema de atualização adequado. Desta forma, mostramos que os pontos de acumulação do algoritmo são críticos.

Sugestões para trabalhos futuros

Para dar continuidade na pesquisa, abordando aspectos não estudados na presente dissertação, seguem algumas sugestões para trabalhos futuros.

- Investigar ordem de convergência para problemas de otimização multiobjetivo não convexo.
- Implementar computacionalmente os algoritmos dos casos convexos e não convexos e comparar suas funcionalidades com diferentes funções.

Referências

- 1 EDGEWORTH, F. Y. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. [S.l.]: Kegan Paul, 1881.
- 2 PARETO, V. *Cours d'économie politique*. [S.l.]: Librairie Droz, 1964.
- 3 FLIEGE, J.; SVAITER, B. F. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, Springer, v. 51, n. 3, p. 479–494, 2000.
- 4 FLIEGE, J.; DRUMMOND, L. G.; SVAITER, B. F. Newton's method for multiobjective optimization. *SIAM Journal on Optimization*, SIAM, v. 20, n. 2, p. 602–626, 2009.
- 5 DRUMMOND, L. G.; RAUPP, F.; SVAITER, B. A quadratically convergent newton method for vector optimization. *Optimization*, Taylor & Francis, v. 63, n. 5, p. 661–677, 2014.
- 6 POVALEJ, Ž. Quasi-newton's method for multiobjective optimization. *Journal of Computational and Applied Mathematics*, Elsevier, v. 255, p. 765–777, 2014.
- 7 LI, D.-H.; FUKUSHIMA, M. On the global convergence of the bfgs method for nonconvex unconstrained optimization problems. *SIAM Journal on Optimization*, SIAM, v. 11, n. 4, p. 1054–1064, 2001.
- 8 QU, S.; GOH, M.; CHAN, F. T. Quasi-newton methods for solving multiobjective optimization. *Operations Research Letters*, Elsevier, v. 39, n. 5, p. 397–399, 2011.
- 9 IZMAILOV, A.; SOLODOV, M. *Otimização, volume 2: métodos computacionais*. [S.l.]: IMPA, 2007.
- 10 LIMA, E. L. *Curso de Análise, vol. 2.(6a edição)*. [S.l.: s.n.], 2000.
- 11 RIBEIRO, A. A.; KARAS, E. W. Um curso de otimização. *Universidade de Curitiba, Curitiba, Brasil*, 2001.
- 12 IZMAILOV, A.; SOLODOV, M. *Otimização, volume 1: condições de otimalidade, elementos de análise convexa e de dualidade*. [S.l.]: IMPA, 2005.
- 13 BAUSCHKE, H. H.; COMBETTES, P. L. *Convex analysis and monotone operator theory in Hilbert spaces*. [S.l.]: Springer Science & Business Media, 2011.
- 14 DAI, Y. On the nonmonotone line search. *Journal of Optimization Theory and Applications*, Springer, v. 112, n. 2, p. 315–330, 2002.
- 15 PANONCELI, D. M. Um estudo de buscas unidirecionais aplicadas ao método bfgs. 2015.

- 16 JR, J. E. D.; SCHNABEL, R. B. *Numerical methods for unconstrained optimization and nonlinear equations*. [S.l.]: Siam, 1996.
- 17 LIMA, E. L. *Análise real*. [S.l.]: Impa, 2004.
- 18 NOCEDAL, J.; WRIGHT, S. *Numerical optimization*. [S.l.]: Springer Science & Business Media, 2006.
- 19 LUENBERGER, D. G.; YE, Y. *Linear and nonlinear programming*. [S.l.]: Springer, 1984.
- 20 MARTINEZ, J. M.; SANTOS, S. A. Métodos computacionais de otimização. *Colóquio Brasileiro de Matemática, Apostilas*, v. 20, 1995.
- 21 SUN, W.; YUAN, Y.-X. *Optimization theory and methods: nonlinear programming*. [S.l.]: Springer Science & Business Media, 2006.
- 22 DENNIS JR, J. E.; MORÉ, J. J. Quasi-newton methods, motivation and theory. *SIAM review*, SIAM, v. 19, n. 1, p. 46–89, 1977.
- 23 DENNIS, J. E.; MORÉ, J. J. A characterization of superlinear convergence and its application to quasi-newton methods. *Mathematics of computation*, v. 28, n. 126, p. 549–560, 1974.
- 24 MIETTINEN, K. *Nonlinear Multiobjective Optimization*. [S.l.]: Springer Science & Business Media, 1999.
- 25 SAMPAIO, P. R. *Teoria, métodos e aplicações de otimização multiobjetivo*. Tese (Doutorado) — Universidade de São Paulo, 2011.
- 26 VILLACORTA, K. D. V. *MÉTODO DE PONTO PROXIMAL PARA OTIMIZAC AO VETORIAL E MÉTODO DE REGIAO DE CONFIANCA PARA OTIMIZAC AO MULTIOBJETIVO*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2011.
- 27 GÖPFERT, A. et al. *Variational methods in partially ordered spaces*. [S.l.]: Springer Science & Business Media, 2006.
- 28 LI, D.-H.; FUKUSHIMA, M. A modified bfgs method and its global convergence in nonconvex minimization. *Journal of Computational and Applied Mathematics*, Elsevier, v. 129, n. 1, p. 15–35, 2001.