



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

DANILO RODRIGUES DE SOUZA

Métodos Quase-Newton com Busca Linear de Wolfe para Otimização Multiobjetivo

Goiânia
2023



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Danilo Rodrigues de Souza

3. Título do trabalho

Métodos Quase-Newton com Busca Linear de Wolfe para Otimização Multiobjetivo

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Leandro Da Fonseca Prudente, Professor do Magistério Superior**, em 02/03/2023, às 15:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Danilo Rodrigues De Souza, Discente**, em 02/03/2023, às 17:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3565210** e o código CRC **3F0646A9**.

DANILO RODRIGUES DE SOUZA

Métodos Quase-Newton com Busca Linear de Wolfe para Otimização Multiobjetivo

Tese apresentada ao Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito para obtenção do título de Doutor em Matemática.

Área de concentração: otimização.

Orientador: Professor Doutor Leandro da Fonseca Prudente

Goiânia
2023

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Souza, Danilo Rodrigues de
Métodos Quase-Newton com Busca Linear de Wolfe para
Otimização Multiobjetivo [manuscrito] / Danilo Rodrigues de Souza. -
2023.
138 f.: il.

Orientador: Prof. Dr. Leandro da Fonseca Prudente.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de
Matemática e Estatística (IME), Programa de Pós-Graduação em
Matemática, Goiânia, 2023.
Bibliografia. Apêndice.
Inclui gráfico, tabelas, algoritmos.

1. Otimização multiobjetivo. 2. Métodos quase-Newton. 3. BFGS. 4.
Busca linear de Wolfe. 5. Convergência superlinear. I. Prudente,
Leandro da Fonseca, orient. II. Título.

CDU 51



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE TESE

Ata nº **02** da sessão de Defesa de Tese de **Danilo Rodrigues de Souza**, que confere o título de Doutor em Matemática, **na área de concentração de Otimização**.

Ao décimo sexto dia do mês de fevereiro do ano de dois mil e vinte e três, a partir das nove horas, via Web videoconferência, realizou-se a sessão pública de Defesa de Tese intitulada "**Métodos Quase-Newton com Busca Linear de Wolfe para Otimização Multiobjetivo**." Os trabalhos foram instalados pelo Orientador e Presidente da banca, Professor Doutor **Leandro da Fonseca Prudente - IME/UFG** com a participação dos demais membros da Banca Examinadora: Professor Doutor **Max Leandro Nobre Gonçalves - IME/UFG** - membro titular interno, Professor Doutor **Orizon Pereira Ferreira - IME/UFG** - membro titular interno, Professor Doutor **Ernesto Julian Goldberg Birgin - IME/USP** membro titular externo e a Professora Doutora **Sandra Augusta Santos - IMECC/UNICAMP**, membra titular externa. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese, tendo sido o candidato **APROVADO** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Leandro da Fonseca Prudente - IME/UFG**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao décimo sexto dia do mês de fevereiro do ano de dois mil e vinte e três.

TÍTULO SUGERIDO PELA BANCA

Métodos Quase-Newton com Busca Linear de Wolfe para Otimização Multiobjetivo



Documento assinado eletronicamente por **Sandra Augusta Santos, Usuário Externo**, em 16/02/2023, às 11:25, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leandro Da Fonseca Prudente, Professor do Magistério Superior**, em 16/02/2023, às 11:55, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Max Leandro Nobre Gonçalves, Professor do Magistério Superior**, em 23/02/2023, às 15:42, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ernesto Julian Goldberg Birgin, Usuário Externo**, em 23/02/2023, às 15:48, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Orizon Pereira Ferreira, Professora do Magistério Superior**, em 23/02/2023, às 16:59, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3456480** e o código CRC **F81309C3**.

Referência: Processo nº 23070.001903/2023-84

SEI nº 3456480

Dedico este trabalho a Deus, minha fonte de força e inspiração. À minha família e amigos, que sempre estiveram ao meu lado me apoiando e me ajudando a crescer. Em especial, à minha amada esposa, que tem sido meu pilar de força e minha fonte de amor.

Agradecimentos

Primeiramente, sou grato a Deus por todas as bênçãos em minha vida, por me guiar e me proteger em todos os momentos. Toda honra e glória sejam dadas a Ele.

Agradeço aos meus pais, Waldir R. de Souza e Maria de Lurdes Rodrigues, meus pilares, que me deram condições de sempre estudar. O apoio e presença de vocês significam muito para mim e me ajudaram a superar desafios e momentos difíceis. Agradeço à minha esposa e parceira de vida, Gabriele F. de Souza Mendes, que desde 2013 compartilha comigo as alegrias e os desafios do caminho que escolhemos juntos. Gostaria de expressar gratidão ao meu irmão Leandro R. de Souza, a minha cunhada Marielly Magalhães Bueno e a minha sobrinha Helena Magalhães Rodrigues, por serem uma parte tão importante da minha vida. Agradeço especialmente pelo carinho que vocês têm dado à minha sobrinha, que ilumina nossas vidas com sua doçura e alegria.

Aos meus queridos amigos: Allyssom Marques, Clauciney Pereira, Débora Correia, Filipe Cabrera, Stéfany Barreto e Wandesson Pereira. Agradeço a cada um de vocês por serem uma parte tão importante da minha vida. A amizade, lealdade e presença de vocês sempre me fizeram sentir apoiado e amado, mesmo nos momentos mais difíceis. Vocês me ensinaram a valorizar a amizade e a importância de ter pessoas em quem posso confiar.

Gostaria de expressar minha profunda gratidão ao meu orientador Leandro F. Prudente. Seu apoio, sabedoria e orientação foram fundamentais para minha formação acadêmica e profissional. Você me ensinou muito mais do que apenas matemática, me mostrou como ser um bom pesquisador e profissional. Agradeço especialmente pela paciência que teve comigo, por sempre estar disponível para tirar minhas dúvidas e por me ajudar a desenvolver a pesquisa. Suas orientações me permitiram realizar um trabalho acadêmico de qualidade e contribuir para o avanço da matemática. Desejo a você todo o sucesso e reconhecimento que merece por sua dedicação e excelência como professor e orientador.

Agradeço aos professores do IME-UFG, que durante minha formação acadêmica foram de suma importância. Um agradecimento especial aos professores,

Glaydston de Carvalho Bento, Max Leandro Nobre Gonçalves, Orizon Ferreira Pereira e Romildo Silva Pina. Agradeço também aos amigos matemáticos que fiz ao longo dos meus estudos, Ana Maria, Ana Paula, Ângelo Guimarães, Cícero Rumão, Elianderson Meneses, Fábio Sobré, Luana de Carvalho, Nathanni Pádua, Rafael Marlon, Raquel Pereira, Robson Lousa, Thamara Policarpo; Agradeço por cada momento compartilhado, por cada risada, por cada conversa sincera. Suas amizades me ajudaram a crescer e me tornar uma pessoa melhor.

Por fim, quero esboçar minha gratidão à CAPES, por financiar meus estudos e pesquisas.

“O mundo pertence aos otimistas; os pessimistas são meros espectadores.”

François Guizot,
Historiador francês que dedicou-se a pensamentos políticos.

Resumo

de Souza, Danilo Rodrigues. **Métodos Quase-Newton com Busca Linear de Wolfe para Otimização Multiobjetivo**. Goiânia, 2023. 141p. Tese de Doutorado. Programa de Pós-Graduação em Matemática, Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Propomos três métodos tipo BFGS com busca linear de Wolfe para otimização multiobjetivo irrestrita. Os algoritmos são bem definidos mesmo para problemas gerais não convexos. O primeiro mimetiza o método BFGS clássico para otimização escalar, para o qual a convergência global e R-linear para um ponto Pareto ótimo são estabelecidas para problemas fortemente convexos. Na análise de convergência local, a taxa é Q-superlinear. Os outros dois algoritmos são versões globalmente convergentes do método BFGS para problemas não convexos. Finalmente, caracterizamos explicitamente de maneira não assintótica a convergência local superlinear do método BFGS para otimização multiobjetivo.

Palavras-chave

Otimização multiobjetivo, otimização multicritério, otimalidade Pareto, métodos quase-Newton, BFGS, busca linear de Wolfe, convergência superlinear, convergência local, taxa de convergência.

Abstract

de Souza, Danilo Rodrigues. **Quasi-Newton Methods with Wolfe Line Search for Multiobjective Optimization**. Goiânia, 2023. 141p. PhD. Thesis. Programa de Pós-Graduação em Matemática, Instituto de Matemática e Estatística, Universidade Federal de Goiás.

We propose three BFGS-type methods with Wolfe line search for unconstrained multiobjective optimization. The algorithms are well defined even for general nonconvex problems. The first one mimics the classical BFGS method for scalar optimization, for which global convergence and R-linear convergence to a Pareto optimal point are established for strongly convex problems. In the local convergence analysis, the rate is Q-superlinear. The other two algorithms are globally convergent versions of the BFGS method for nonconvex problems. Finally, we explicitly characterize in a non-asymptotic way the superlinear local convergence of the BFGS method for multiobjective optimization.

Keywords

Multiobjective optimization, multicriteria optimization, Pareto optimality, quasi-Newton methods, BFGS, Wolfe line search, superlinear convergence, local convergence, rate of convergence.

Sumário

Introdução	14
1 Otimização Multiobjetivo	17
1.1 Notação	17
1.2 Conceitos fundamentais de otimização multiobjetivo	17
1.3 Condições de Wolfe padrão para otimização multiobjetivo	21
2 Métodos quase-Newton	26
2.1 Método BFGS clássico	26
2.2 BFGS multiobjetivo	33
3 Novos métodos BFGS para otimização multiobjetivo	46
3.1 Uma extensão que mimetiza o método clássico	50
3.1.1 Análise de convergência	52
3.1.2 Convergência local superlinear	58
3.1.3 Experimentos Numéricos	70
3.2 Um método BFGS modificado para problemas não convexos	80
3.2.1 Análise de convergência	81
3.2.2 Convergência local superlinear	83
3.2.3 Um método mBFGS praticável	88
3.3 Um método BFGS cauteloso para problemas não convexos	91
3.3.1 Análise de convergência	92
3.3.2 Convergência local superlinear	94
3.3.3 Experimentos Numéricos	95
4 Convergência local superlinear não assintótica do método BFGS para otimização multiobjetivo	99
4.1 Resultados preliminares	101
4.2 Convergência superlinear não assintótica	112
4.3 Sobre a limitação de ξ_k	119
4.4 Uma comparação entre taxas	121
5 Conclusão	124
Referências Bibliográficas	125
A Apêndice	132

Introdução

Em otimização multiobjetivo, buscamos minimizar duas ou mais funções objetivo simultaneamente. Tipicamente, em um problema dessa natureza, não existe um único ponto que minimiza todas as funções simultaneamente. Nesse caso, diz-se que os objetivos são conflitantes e usamos o conceito de *otimalidade Pareto* para caracterizar uma solução do problema. Um ponto é chamado *Pareto ótimo* se nenhum valor funcional pode ser melhorado sem piorar outro. Problemas de otimização multiobjetivo aparecem em muitas áreas da ciência. Recomendamos ao leitor [2, 73] e suas referências para algumas aplicações práticas interessantes.

Nas últimas duas décadas, vários métodos iterativos para otimização escalar foram estendidos e analisados para otimização multiobjetivo. Essa linha de pesquisa se iniciou em [29] com a extensão do método do gradiente (*steepest descent method*); veja também [23]. Outros algoritmos incluindo Newton [10, 28, 37, 76], quase-Newton [1, 44, 52, 58, 61, 65, 66], gradientes conjugados [34, 35, 50], gradiente projetado [12, 25, 30, 31, 36] e métodos proximais [4, 7, 8, 9, 11] também foram estendidos. Como característica comum, esses métodos possuem propriedades de convergência e não escalarizam o problema a ser resolvido, sendo alternativas interessantes para métodos de escalarização [24] e abordagens heurísticas [45].

Algoritmos quase-Newton são uma das classes mais populares de métodos para resolver um problema de otimização escalar irrestrita. O primeiro método quase-Newton surgiu em 1959 com o trabalho de Davidon [17] e foi popularizado quatro anos depois por Fletcher e Powell [27]. Desde então, os algoritmos quase-Newton têm atraído a atenção da comunidade científica principalmente porque evitam cálculos de Hessianas e possuem bom desempenho numérico. Existem diversos artigos sobre o assunto (que são muitos para listar), incluindo contribuições dos nomes mais proeminentes da comunidade de otimização. Em um método quase-Newton, a direção de busca é calculada com base em um modelo quadrático da função objetivo, em que a Hessiana verdadeira é substituída por alguma aproximação que é atualizada a cada iteração. O esquema de atualização quase-Newton mais popular é a fórmula BFGS, descoberta independentemente por Broyden [5], Fletcher [26], Goldfarb [32] e Shanno [71] em 1970. Sob hipóteses apropriadas, o algoritmo BFGS

com uma busca linear adequada é globalmente e superlinearmente convergente para problemas fortemente convexos [6, 62]. Por outro lado, o método pode não convergir quando aplicado em funções não convexas gerais [14, 15, 54].

O método BFGS para otimização multiobjetivo foi proposto inicialmente em [65] e posteriormente estudado em [1, 44, 52, 58, 61, 66]. Similarmente ao caso escalar, a direção de busca é definida como a solução de um problema envolvendo modelos quadráticos das funções objetivo, em que as atualizações BFGS são usadas no lugar das Hessianas verdadeiras. Apesar da existência dos trabalhos supracitados, a teoria de convergência do método BFGS para problemas de otimização multiobjetivo ainda pode ser considerada incompleta. Destacamos a seguir algumas deficiências existentes nessas referências: *(i)* os algoritmos são geralmente projetados para problemas fortemente convexos; *(ii)* as atualizações BFGS e suas inversas são frequentemente assumidas como uniformemente limitadas (o que parece irreal, veja [60, Section 6.4]); *(iii)* algumas etapas intermediárias que são cruciais na análise de convergência são frequentemente assumidas sem prova. Um dos objetivos do presente trabalho é superar esses inconvenientes.

Frequentemente, o método BFGS escalar usa busca linear de Wolfe, veja [60, Capítulo 6]. No contexto vetorial, as condições de Wolfe foram recentemente estendidas por Lucambio-Pérez e Prudente em [50]. Na presente tese, associamos busca linear de Wolfe com métodos do tipo BFGS para problemas multiobjetivo irrestritos. Destacamos que as condições de Wolfe são cruciais para preservar as aproximações Hessianas definidas positivas e gerar direções de descida. Como uma consequência, os algoritmos propostos são bem definidos mesmo para problemas não convexas gerais. Até onde sabemos, este é o primeiro trabalho a associar BFGS e Wolfe para otimização multiobjetivo.

O primeiro algoritmo proposto busca mimetizar, para o contexto multiobjetivo, o método BFGS escalar clássico e suas propriedades. Estabelecemos a convergência global e R-linear para um ponto Pareto ótimo, para problemas fortemente convexos. Na análise de convergência local, supondo que as funções objetivo são localmente fortemente convexas com Hessianas Lipschitz contínuas, provamos que a taxa de convergência é Q-superlinear. Como resultado intermediário, mostramos que o passo unitário satisfaz as condições de Wolfe após um número suficientemente grande de iterações. Além disso, uma condição do tipo Dennis-Moré para otimização multiobjetivo surge de forma muito clara. Todas as hipóteses consideradas são extensões naturais daquelas feitas para o caso escalar. Essa proposta foi recentemente publicada em [64].

Por mais de três décadas, a convergência do método BFGS para otimização não convexa esteve em aberto. Durante esse tempo, alguns trabalhos surgiram

com o intuito de jogar luz sobre essa questão propondo versões globalizadas do método BFGS escalar. Inspirados nos trabalhos de Li e Fukushima [46,47] propomos dois algoritmos do tipo BFGS para otimização multiobjetivo não convexa que são *globalmente convergentes* com taxa Q-superlinear. O primeiro desses algoritmos utiliza uma versão modificada da fórmula de atualização BFGS, enquanto o segundo adota um critério cauteloso para promover as atualizações.

É bem conhecido que a taxa de convergência assintótica local do método BFGS escalar é Q-superlinear [62]. Recentemente, surgiram alguns trabalhos na literatura que obtiveram convergência local superlinear explícita (não assintótica) para métodos quase-Newton escalares. Em particular, Rodomanov e Nesterov [68,69] exploraram a convergência local do método BFGS clássico e mostraram, de forma explícita, que a taxa de convergência é Q-superlinear. Inspirados nesses trabalhos, caracterizamos explicitamente a taxa de convergência superlinear do método BFGS multiobjetivo.

Este trabalho está estruturado do seguinte modo. No Capítulo 1, apresentamos as notações, os conceitos e resultados básicos para otimização multiobjetivo, bem como discutimos as condições de Wolfe e de Zoutendijk. No Capítulo 2, apresentamos uma breve revisão sobre métodos quase-Newton para otimização escalar e multiobjetivo, focando em métodos do tipo BFGS. Adicionalmente, provamos alguns resultados técnicos autorais para otimização multiobjetivo que são úteis ao longo do trabalho. No Capítulo 3, encontram-se os três métodos quase-Newton propostos para otimização multiobjetivo. O método BFGS com busca linear de Wolfe que mimetiza o método clássico está na Seção 3.1. Nas Seções 3.2 e 3.3 propomos os algoritmos tipo BFGS globalmente convergentes para problemas não convexos. Experimentos numéricos ilustrando as potenciais vantagens práticas de nossas propostas são apresentados. A taxa de convergência não assintótica do método BFGS multiobjetivo é discutida no Capítulo 4. Finalmente, conclusões são feitas no Capítulo 5.

Otimização Multiobjetivo

1.1 Notação

Denote, respectivamente, por \mathbb{R} , \mathbb{R}_+ e \mathbb{R}_{++} os conjuntos dos números reais, dos reais não negativos e dos reais estritamente positivos. Usualmente, \mathbb{R}^n e $\mathbb{R}^{n \times p}$ denotam o conjunto dos vetores coluna reais n -dimensionais e o conjunto das matrizes reais $n \times p$, respectivamente. A matriz identidade de dimensão $n \times n$ será denotada por I_n . Consideraremos a seguinte ordem parcial induzida por \mathbb{R}_+^m : sejam $u, v \in \mathbb{R}^m$, se $u \leq v$ então $v - u \in \mathbb{R}_+^m := \mathbb{R}_+ \times \mathbb{R}_+ \times \dots \times \mathbb{R}_+$; se $u < v$ então $v - u \in \mathbb{R}_{++}^m := \mathbb{R}_{++} \times \mathbb{R}_{++} \times \dots \times \mathbb{R}_{++}$. Para $A \in \mathbb{R}^{n \times n}$, $A \succ 0$ (resp. $A \prec 0$) significa que A é definida positiva (resp. negativa). Ao longo deste trabalho, vamos usar $\|\cdot\|_1$ para denotar a norma da soma (norma 1) e $\|\cdot\|$ para a norma Euclidiana vetorial ou matricial (norma 2). Em particular, a norma Euclidiana de uma matriz $A \in \mathbb{R}^{n \times p}$, é definida como o maior autovalor da matriz $(A^T A)^{1/2}$. A cardinalidade de um conjunto C é denotado por $|C|$. As funções *ceiling* e *floor* serão denotadas por $\lceil \cdot \rceil$ e $\lfloor \cdot \rfloor$, respectivamente; isto é, se $x \in \mathbb{R}$, então $\lceil x \rceil$ é o menor inteiro maior que ou igual a x e $\lfloor x \rfloor$ é o maior inteiro menor que ou igual a x . Dadas duas seqüências reais $\{a_k\}$ e $\{b_k\}$ (com $b_k > 0$ para todo k), escrevemos $a_k = o(b_k)$ se $\lim_{k \rightarrow \infty} a_k/b_k = 0$. Se $K = \{k_1, k_2, \dots\} \subseteq \mathbb{N}$, com $k_j < k_{j+1}$ para todo $j \in \mathbb{N}$, então denotamos $K \subsetneq \mathbb{N}$.

1.2 Conceitos fundamentais de otimização multiobjetivo

Em otimização multiobjetivo, o intuito é minimizar várias funções simultaneamente. Dada $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, uma função continuamente diferenciável, denotaremos esse problema por

$$\min_{x \in \mathbb{R}^n} F(x). \quad (1-1)$$

Nesse caso, dificilmente existirá um único ponto que minimiza todos os objetivos simultaneamente. Desse modo, no cenário multiobjetivo, a otimalidade é chamada

de *eficiência Pareto* ou *otimalidade Pareto* e estamos interessados em encontrar um ponto *Pareto ótimo* de F . Um ponto $x^* \in \mathbb{R}^n$ é um *Pareto ótimo* (resp. *Pareto ótimo fraco*) de F se não existe outro $x \in \mathbb{R}^n$ com $F(x) \leq F(x^*)$ e $F(x) \neq F(x^*)$ (resp. $F(x) < F(x^*)$). Isto é, não existe $x \in \mathbb{R}^n$ tal que $F_j(x) \leq F_j(x^*)$ (resp. $F_j(x) < F_j(x^*)$), para todo $j = 1, \dots, m$, e $F_{j_0}(x) < F_{j_0}(x^*)$ para pelo menos um $j_0 \in \{1, \dots, m\}$. Esses conceitos também são definidos localmente: dizemos que $x^* \in \mathbb{R}^n$ é *Pareto ótimo local* (resp. *Pareto ótimo fraco local*) de F se existe uma vizinhança $U \subset \mathbb{R}^n$ de x^* tal que x^* é Pareto ótimo (resp. Pareto ótimo fraco) de F restrita a U . Uma condição necessária para $x^* \in \mathbb{R}^n$ ser um ponto Pareto ótimo fraco local é

$$-(\mathbb{R}_{++}^m) \cap \text{Im}(JF(x^*)) = \emptyset, \quad (1-2)$$

em que $\text{Im}(JF(x^*))$ denota o conjunto imagem da Jacobiana de F em x^* . Um ponto x^* que satisfaz (1-2) é chamado de *Pareto crítico*. Observe que, se $x \in \mathbb{R}^n$ não é um Pareto crítico então existe $d \in \mathbb{R}^n$ tal que $JF(x)d < 0$. Nesse caso, $\nabla F_j(x)^T d < 0$ para todo $j = 1, \dots, m$ e d é dita uma *direção de descida* de F em x , isto é, existe $\varepsilon > 0$ tal que $F(x + \alpha d) < F(x)$ para todo $\alpha \in [0, \varepsilon]$. Dados $x, y \in \mathbb{R}^n$, diremos que x domina y quando $F(y) - F(x) \in \mathbb{R}_+^m \setminus \{0\}$. Dizemos que $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é convexa (resp. fortemente convexa) se suas componentes $F_j : \mathbb{R}^n \rightarrow \mathbb{R}$ são convexas (resp. fortemente convexas), para todo $j = 1, \dots, m$. O próximo resultado relaciona os conceitos de criticalidade, otimalidade e convexidade.

Teorema 1.1. [28, Teorema 3.1] *As seguintes afirmações valem:*

- i. se x^* é Pareto ótimo fraco local de F , então x^* é um ponto Pareto crítico;
- ii. se F é convexa e x^* é Pareto crítico de F , então x^* é um ponto Pareto ótimo fraco;
- iii. se F é duas vezes continuamente diferenciável, $\nabla^2 F_j(x) > 0$ para todo $j \in \{1, \dots, m\}$ e todo $x \in \mathbb{R}^n$, e $x^* \in \mathbb{R}^n$ é Pareto crítico de F , então x^* é um ponto Pareto ótimo.

Agora, vamos definir a generalização da derivada direcional para otimização multiobjetivo. Seja $\mathcal{D} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ definida por

$$\mathcal{D}(x, d) := \max_{j=1, \dots, m} \nabla F_j(x)^T d. \quad (1-3)$$

A função \mathcal{D} caracteriza as direções de descida de F em x . De fato, é fácil ver que se $\mathcal{D}(x, d) < 0$, então d é uma direção de descida de F em x . Além disso, x é um ponto Pareto crítico se e somente se $\mathcal{D}(x, d) \geq 0$ para todo $d \in \mathbb{R}^n$. A seguir, apresentaremos outras importantes propriedades da função \mathcal{D} que podem ser trivialmente obtidas a partir da definição.

Lema 1.1. *Os seguintes itens são válidos:*

- (i) para qualquer $x \in \mathbb{R}^n$ e $\alpha \geq 0$, temos $\mathcal{D}(x, \alpha d) = \alpha \mathcal{D}(x, d)$;
- (ii) a função $(x, d) \mapsto \mathcal{D}(x, d)$ é contínua.

Um conceito muito importante em otimização escalar é a direção de máxima descida (direção oposta ao gradiente). Para o caso multiobjetivo, a direção de máxima descida (*steepest descent direction*), introduzida em [29, 59], é dada a partir da solução do seguinte problema:

$$\min_{d \in \mathbb{R}^n} \max_{j=1, \dots, m} \nabla F_j(x)^T d + \frac{1}{2} d^T I_n d. \quad (1-4)$$

Note que, para cada $x \in \mathbb{R}^n$, a função objetivo de (1-4) é fortemente convexa e, portanto, esse problema tem uma única solução ótima a qual denotamos por $d_{SD}(x)$ e chamamos de *direção de máxima descida generalizada*. Vamos denotar o valor ótimo do problema (1-4) por $\theta_{SD}(x)$, isto é,

$$d_{SD}(x) := \operatorname{argmin}_{d \in \mathbb{R}^n} \max_{j=1, \dots, m} \nabla F_j(x)^T d + \frac{1}{2} \|d\|^2.$$

e

$$\theta_{SD}(x) := \max_{j=1, \dots, m} \nabla F_j(x)^T d_{SD}(x) + \frac{1}{2} \|d_{SD}(x)\|^2.$$

Agora, observe que o problema (1-4) é equivalente ao seguinte problema de otimização quadrática convexa:

$$\begin{aligned} \min_{(t, d) \in \mathbb{R} \times \mathbb{R}^n} \quad & t + \frac{1}{2} \|d\|^2 \\ \text{s. a.} \quad & \nabla F_j(x)^T d \leq t, \quad \forall j = 1, \dots, m. \end{aligned} \quad (1-5)$$

A única solução de (1-5) é dada por $(t, d) := (\mathcal{D}(x, d_{SD}(x)), d_{SD}(x))$. Como (1-5) é convexo e possui apenas restrições lineares, então existe multiplicador de Lagrange $\lambda^{SD}(x) \in \mathbb{R}^m$ tal que a tripla $(t, d, \lambda) := (\mathcal{D}(x, d_{SD}(x)), d_{SD}(x), \lambda^{SD}(x)) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$ satisfaz o sistema Karush-Kuhn-Tucker:

$$\sum_{j=1}^m \lambda_j = 1, \quad \sum_{j=1}^m \lambda_j [\nabla F_j(x) + d] = 0,$$

e

$$\lambda_j \geq 0, \quad \nabla F_j(x)^T d \leq t, \quad \lambda_j [\nabla F_j(x)^T d - t] = 0,$$

para todo $j = 1, \dots, m$. Portanto, após algumas manipulações

$$\sum_{j=1}^m \lambda_j^{SD}(x) = 1, \quad \lambda_j^{SD}(x) \geq 0, \quad \forall j = 1, \dots, m, \quad (1-6)$$

$$d_{SD}(x) = - \sum_{j=1}^m \lambda_j^{SD}(x) \nabla F_j(x), \quad (1-7)$$

e

$$\theta_{SD}(x) = -\frac{1}{2} \|d_{SD}(x)\|^2. \quad (1-8)$$

Observação 1.1. A direção de máxima descida generalizada recupera a definição clássica, isto é, se $m = 1$, então $d_{SD}(x) = -\nabla F(x)$.

O seguinte lema mostra que a direção $d_{SD}(x)$ e o valor ótimo $\theta_{SD}(x)$ podem ser usados para caracterizar os pontos Pareto críticos do problema (1-1). Além disso, no item (iv) temos uma propriedade muito útil para os resultados que apresentamos nesse trabalho.

Lema 1.2. *Sejam $d_{SD}(x)$ e $\theta_{SD}(x)$ a solução e o valor ótimo do problema (1-4), respectivamente. Então:*

- i. x é Pareto crítico se e somente se $d_{SD}(x) = 0$;
- ii. se x não é Pareto crítico, então temos $d_{SD}(x) \neq 0$ e $\mathcal{D}(x, d_{SD}(x)) = -\|d_{SD}(x)\|^2 < 0$ (em particular, $d_{SD}(x)$ é direção de descida de F em x);
- iii. as funções $x \mapsto d_{SD}(x)$ e $x \mapsto \theta_{SD}(x)$ são contínuas.
- iv. para qualquer $x \in \mathbb{R}^n$, $-d_{SD}(x)$ é o elemento de menor norma do conjunto

$$\left\{ u \in \mathbb{R}^n \mid u = \sum_{j=1}^m \lambda_j \nabla F_j(x), \sum_{j=1}^m \lambda_j = 1, \lambda_j \geq 0 \forall j = 1, \dots, m \right\},$$

isto é, no envelope convexo de $\{\nabla F_1(x), \dots, \nabla F_m(x)\}$.

Demonstração. Para os itens (i), (ii) e (iii), veja [38, Lema 3.3]. Para o item (iv), veja [74, Corolário 2.3]. ■

Pelo Lema 1.2, se x é um ponto não crítico, existem direções de descida de F em x (em particular, $d_{SD}(x)$). Assim, se $d \in \mathbb{R}^n$ é tal que $JF(x)d < 0$, então podemos calcular $\alpha > 0$ de modo que $F(x + \alpha d) < F(x)$. Uma discussão mais cuidadosa está feita na Seção 1.3.

1.3 Condições de Wolfe padrão para otimização multiobjetivo

Procedimentos de busca linear são ferramentas fundamentais para otimização. Conceitos como condições de Armijo e de Wolfe são frequentemente usados em algoritmos para minimização. Em otimização escalar, isto é, $m = 1$ em (1-1), a condição de Armijo exige que o comprimento de passo $\alpha > 0$ satisfaça

$$F(x + \alpha d) \leq F(x) + c_1 \alpha \nabla F(x)^T d, \quad (1-9)$$

no qual $c_1 \in (0, 1)$. Essa condição, também conhecida como *decrécimo suficiente*, não é suficiente para evitar que passos muito pequenos sejam aceitos, podendo causar uma deterioração do algoritmo. Assim, uma segunda condição é comumente imposta, conhecida como *condição de curvatura*, a qual exige que $\alpha > 0$ satisfaça

$$\nabla F(x + \alpha d)^T d \geq c_2 \nabla F(x)^T d, \quad c_2 \in (c_1, 1), \quad (1-10)$$

e que, em particular, evita a escolha de passos muito pequenos. A união das condições de decréscimo suficiente e de curvatura é conhecida como as *condições de Wolfe padrão*. É possível provar que, se F é suave e limitada inferiormente, então existem comprimentos de passo que satisfazem as condições de Wolfe padrão. Recomendamos ao leitor visitar [60, Seção 3.1], para uma discussão cuidadosa sobre o assunto.

Em otimização multiobjetivo, além da direção $d_{SD}(\cdot)$, os autores em [29] também estenderam o conceito de condição de Armijo, usando-o para estudar a convergência do método do gradiente (*steepest descent method*). Sejam $c_1 \in (0, 1)$ um parâmetro algorítmico e d uma direção de descida de F em x , um passo $\alpha > 0$ satisfaz a condição tipo Armijo se

$$F(x + \alpha d) \leq F(x) + c_1 \alpha JF(x)d. \quad (1-11)$$

Surgiram também outros métodos de otimização multiobjetivo que usam decréscimo suficiente, incluindo Newton [10, 28, 37, 49], quase-Newton [61, 65] e métodos de gradiente projetado [30, 31, 36]. Com respeito às condições de Wolfe para otimização multiobjetivo, uma extensão foi proposta em [50], em que os autores a usaram para estudar a convergência do método dos gradientes conjugados.

Definição 1.1. [50, Definição 3.1] *Sejam $d \in \mathbb{R}^n$ uma direção de descida de F em x e $c_1, c_2 \in \mathbb{R}$ tais que $0 < c_1 < c_2 < 1$. Dizemos que α satisfaz as condições de*

Wolfe padrão, se

$$F_j(x + \alpha d) \leq F_j(x) + c_1 \alpha \mathcal{D}(x, d), \quad \forall j = 1, \dots, m, \quad (1-12)$$

$$\mathcal{D}(x + \alpha d, d) \geq c_2 \mathcal{D}(x, d), \quad (1-13)$$

em que \mathcal{D} está definida em (1-3).

Note que, as m desigualdades em (1-12), diferentemente da condição (1-11), estipulam que α deve fornecer um decréscimo suficiente, proporcional a α e $\mathcal{D}(x, d)$, em cada função objetivo. Essas desigualdades compõem uma condição do tipo Armijo, que no caso escalar, recupera a definição clássica (1-9). Além disso, a condição de curvatura (1-13), que estende (1-10), envolve as informações de todas as funções em uma única desigualdade. Assim, as condições de Wolfe padrão acrescentam apenas mais uma desigualdade na condição de Armijo, independentemente do número m de objetivos.

Observação 1.2. A seguir discutimos sobre a condição de curvatura (1-13). Assuma que no procedimento de busca linear, $\alpha_t > 0$ é um tamanho de passo teste que satisfaz (1-12). Se α_t é tal que $\nabla F_j(x + \alpha_t d)^T d$ é *fortemente negativo* (menor que $c_2 \mathcal{D}(x, d)$) para todo $j = 1, \dots, m$ (consequentemente $\mathcal{D}(x + \alpha_t d, d)$ também será), isso é um sinal que podemos reduzir simultaneamente todos os objetivos ao nos movermos na direção de d e assim, α_t é descartado pela condição (1-13). Por outro lado, se existe $j \in \{1, \dots, m\}$ tal que $\nabla F_j(x + \alpha_t d)^T d$ é apenas fracamente negativo ou até mesmo positivo (mais precisamente, maior que ou igual a $c_2 \mathcal{D}(x, d)$), suspeitamos que não podemos esperar um decréscimo significativo no objetivo F_j ao longo da direção d . Assim, faz sentido finalizar a busca linear aceitando α_t como o tamanho de passo, isto é, $\alpha := \alpha_t$. Note que, nesse último caso, temos $\mathcal{D}(x + \alpha_t d, d) \geq \nabla F_j(x + \alpha_t d)^T d \geq c_2 \mathcal{D}(x, d)$ e portanto, a condição (1-13) é satisfeita para α_t . Em particular, observe que a condição (1-13) descarta tamanhos de passos inaceitavelmente pequenos que satisfazem (1-12). Portanto, as condições de Wolfe padrão impedem que o método tome tamanhos de passos artificialmente pequenos, o que pode estar sob risco ao usar a condição de Armijo. Essa propriedade é ilustrada nas Figuras 1.1 e 1.2 em que, por simplicidade, consideramos duas funções objetivo, de modo que destacamos os tamanhos de passos aceitáveis ao considerar as condições de Armijo e Wolfe padrão. Note que, na situação ilustrada nas figuras, $c_2 \mathcal{D}(x, d) = c_2 \nabla F_1(x)^T d$ define a inclinação desejada em ambos os objetivos.

Em [50], os autores mostraram que, sob hipóteses razoáveis, existem intervalos de passos positivos que satisfazem tais condições, veja o resultado seguinte.

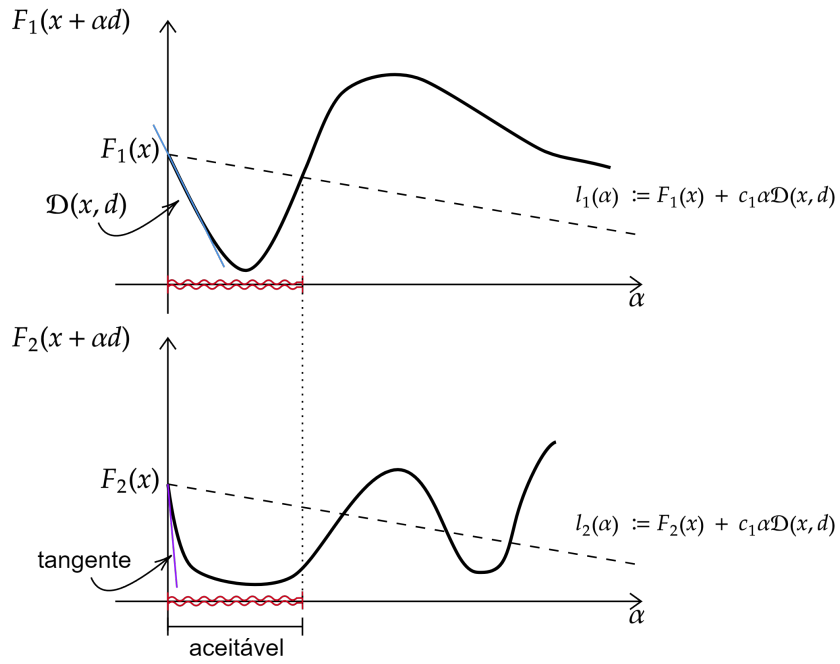


Figura 1.1: Tamanhos de passos aceitáveis quando a condição de Armijo é considerada.

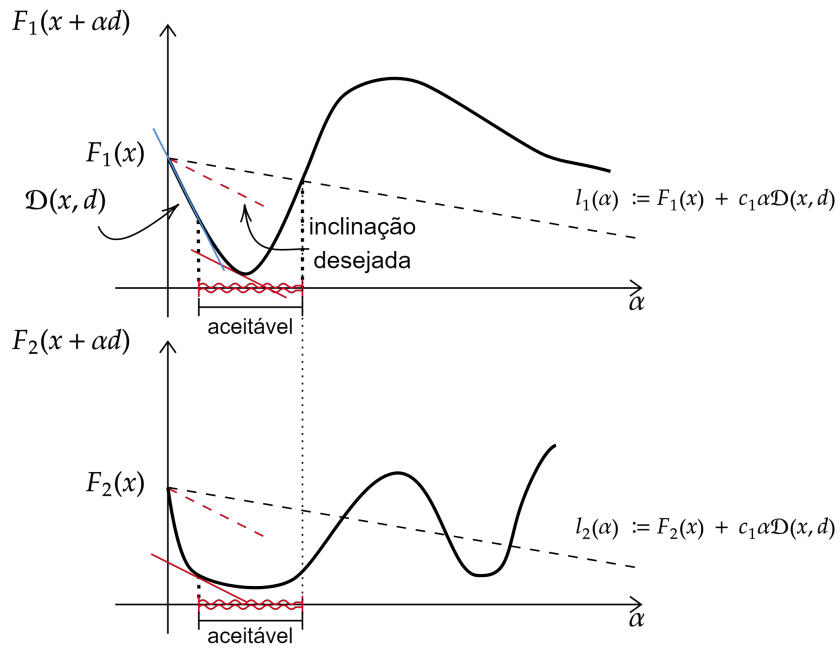


Figura 1.2: Tamanhos de passos aceitáveis quando as condições de Wolfe padrão são consideradas.

Proposição 1.1. [50, Proposição 3.2] *Sejam $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função continuamente diferenciável e d uma direção de descida de F em x . Se existe $\mathcal{A} \in \mathbb{R}^m$ tal que*

$$F(x + \alpha d) \geq \mathcal{A}, \quad \forall \alpha > 0,$$

então existe intervalo com tamanhos de passos positivos satisfazendo as condições de Wolfe padrão (1-12) e (1-13).

Também em [50], os autores estenderam para otimização multiobjetivo a condição de Zoutendijk, uma importante ferramenta para analisar a convergência de métodos de descida, a qual discutiremos a seguir. Considere um método geral de busca linear

$$x^{k+1} = x^k + \alpha_k d^k, \quad (1-14)$$

no qual d^k é uma direção de descida de F em x^k e α_k é um comprimento de passo que satisfaz as condições de Wolfe padrão (1-12)–(1-13). No caso escalar ($m = 1$) e sob hipóteses razoáveis, é possível mostrar que um método do tipo (1-14) satisfaz

$$\sum_{k \geq 0} \frac{(\nabla F(x^k)^T d^k)^2}{\|d^k\|^2} < \infty, \quad (1-15)$$

que é conhecida como condição de Zoutendijk (veja [60, Seção 3.2]). Se β_k é o ângulo entre d^k e a direção de máxima descida $-\nabla F(x^k)$, ou seja, $\cos \beta_k = \frac{-\nabla F(x^k)^T d^k}{\|\nabla F(x^k)\| \|d^k\|}$, então (1-15) implica que

$$\lim_{k \rightarrow \infty} \cos^2 \beta_k \|\nabla F(x^k)\|^2 \rightarrow 0. \quad (1-16)$$

Logo, para mostrar que um método de descida do tipo (1-14) é globalmente convergente, basta garantir que as direções d^k são tais que $\cos \beta_k$ sempre se mantém afastado do zero para todo k (isto é, $\|\nabla F(x^k)\| \xrightarrow{k \rightarrow \infty} 0$). Em outras palavras, um método será globalmente convergente desde que as direções de busca nunca estejam *muito próximas da ortogonalidade* com o gradiente. Em particular, o método do gradiente (no qual a direção de busca é $-\nabla F(x^k)$) com busca linear de Wolfe possui essa propriedade. Na Seção 2.1, discutiremos cuidadosamente o impacto da condição (1-15) no método BFGS, que é o objeto de estudo deste trabalho.

No caso multiobjetivo, a condição de Zoutendijk é dada por

$$\sum_{k \geq 0} \frac{\mathcal{D}(x^k, d^k)^2}{\|d^k\|^2} < \infty. \quad (1-17)$$

Para um método do tipo (1-14), a condição (1-17) é cumprida sob hipóteses que são

extensões naturais das consideradas no caso escalar.

Hipótese 1.1. (i) Existe um conjunto aberto \mathcal{N} tal que $\mathcal{L}(x^0) := \{x \in \mathbb{R}^n \mid F(x) \leq F(x^0)\} \subset \mathcal{N}$ e a Jacobiana JF é Lipschitz contínua em \mathcal{N} com constante $L > 0$, isto é, $\|JF(x) - JF(y)\| \leq L\|x - y\|$ para todo $x, y \in \mathcal{N}$. (ii) F é limitada inferiormente em $\mathcal{L}(x^0)$.

De agora em diante, denotaremos por $\mathcal{L}(x^0)$, o conjunto de nível $\{x \in \mathbb{R}^n \mid F(x) \leq F(x^0)\}$.

Proposição 1.2. [50, Proposição 3.3] Considere um método geral (1-14), em que d^k é uma direção de descida de F em x^k e α_k satisfaz as condições de Wolfe padrão (1-12)–(1-13). Se a Hipótese 1.1 é satisfeita, então vale a condição de Zoutendijk (1-17).

No Capítulo 3, ficará evidenciado o quão útil esse resultado é para a análise de convergência dos métodos propostos neste trabalho.

Métodos quase-Newton

Métodos quase-Newton se tornaram uma interessante alternativa aos métodos Newtonianos, pois além de não necessitarem do cálculo da Hessiana, possuem um bom desempenho na prática. O primeiro método quase-Newton surgiu em 1959 com o trabalho de Davidon [17] e foi posteriormente popularizado por Fletcher e Powell [27]. Curiosamente, o artigo de Davidon [17] não foi aceito para publicação e permaneceu como um relatório técnico por mais de trinta anos até aparecer na primeira edição da *SIAM Journal on Optimization* em 1991 [18]. Em um método quase-Newton, a direção da busca é calculada com base em um modelo quadrático da função objetivo, no qual a Hessiana verdadeira é substituída por alguma aproximação que é atualizada a cada iteração. O esquema de atualização quase-Newton mais eficaz é a fórmula BFGS que foi proposta independentemente em 1970 por Broyden [5], Fletcher [26], Goldfarb [32] e Shanno [71]. Esse esquema de atualização, quando utilizado em um algoritmo quase-Newton, dá origem ao *método BFGS*. Neste capítulo, trataremos dos métodos quase-Newton para otimização irrestrita, com foco no método BFGS. Na primeira seção discutiremos o caso escalar e, na segunda seção, o caso multiobjetivo juntamente com alguns resultados autorais.

2.1 Método BFGS clássico

Suponha que $m = 1$ e considere um ponto $x \in \mathbb{R}^n$. Em um método quase-Newton, a direção de busca a partir de x é dada por

$$d = -B^{-1}\nabla F(x),$$

em que $B \in \mathbb{R}^{n \times n}$ é uma aproximação para a Hessiana. Note que se $B = I_n$ obtemos a direção oposta ao gradiente, e se $B = \nabla^2 F(x)$, a direção de Newton. Diferentes escolhas para B geram diferentes métodos, porém somente quando $B \succ 0$ podemos garantir que d é uma direção de descida de F em x . Do ponto de vista prático, o método BFGS tem sido o mais promissor dos métodos

quase-Newton, sendo amplamente estudado desde a sua concepção. Assim, vamos estudar cuidadosamente as propriedades do método BFGS (Algoritmo 2.1 a seguir).

Algoritmo 2.1. O método BFGS com busca linear de Wolfe

Sejam $c_1 \in (0, 1/2)$, $c_2 \in (c_1, 1)$, $x^0 \in \mathbb{R}^n$ e $B_0 \succ 0$ dados. Faça $k \leftarrow 0$.

Passo 1. *Calcule a direção de busca*

Faça

$$d^k = -B_k^{-1} \nabla F(x^k). \quad (2-1)$$

Passo 2. *Critério de parada*

Se $\nabla F(x^k) = 0$, então PARE.

Passo 3. *Procedimento de busca linear*

Calcule o comprimento de passo $\alpha_k > 0$ (primeiro tente $\alpha_k = 1$) tal que (1-9)–(1-10) sejam satisfeitos e faça

$$x^{k+1} := x^k + \alpha_k d^k.$$

Passo 4. *Prepare a próxima iteração*

Calcule

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \quad (2-2)$$

em que $y_k := \nabla F(x^{k+1}) - \nabla F(x^k)$ e $s_k := x^{k+1} - x^k$. Faça $k \leftarrow k + 1$ e vá ao

Passo 1.

Primeiramente, note que B_0 pode ser escolhida como qualquer matriz simétrica e definida positiva. Em particular, pode-se tomar $B_0 = I_n$ ou $B_0 = \nabla^2 F(x^0)$, no caso em que a Hessiana da função objetivo é definida positiva. Além disso, o método só está bem definido (gerando direção de descida) se a matriz B_{k+1} se manter simétrica e definida positiva. Essa condição é satisfeita se B_k é simétrica e definida positiva, e $s_k^T y_k > 0$. De fato, tomando $z \in \mathbb{R}^n$ arbitrário e não nulo, temos

$$\begin{aligned} z^T B_{k+1} z &= z^T B_k z - \frac{z^T B_k s_k s_k^T B_k z}{s_k^T B_k s_k} + \frac{z^T y_k y_k^T z}{s_k^T y_k} \\ &= \|z\|_{B_k}^2 - \frac{\langle z, s_k \rangle_{B_k}^2}{\|s_k\|_{B_k}^2} + \frac{(z^T y_k)^2}{s_k^T y_k}, \end{aligned}$$

no qual $\langle u, v \rangle_{B_k} := u^T B_k v$ e $\|u\|_{B_k} := \sqrt{\langle u, u \rangle_{B_k}}$. Como B_k é simétrica e definida positiva, o produto interno $\langle \cdot, \cdot \rangle_{B_k}$ e a norma $\|\cdot\|_{B_k}$ estão bem definidos e, pela desigualdade de Cauchy-Schwarz, temos $\langle z, s_k \rangle_{B_k}^2 \leq \|z\|_{B_k}^2 \|s_k\|_{B_k}^2$. Assim, por

$s_k^T y_k > 0$, obtemos

$$z^T B_{k+1} z \geq \|z\|_{B_k}^2 - \|z\|_{B_k}^2 + \frac{(z^T y_k)^2}{s_k^T y_k} = \frac{(z^T y_k)^2}{s_k^T y_k} \geq 0.$$

Por fim, suponha por absurdo que $z^T B_{k+1} z = 0$, logo

$$\|z\|_{B_k}^2 - \frac{\langle z, s_k \rangle_{B_k}^2}{\|s_k\|_{B_k}^2} = 0 \quad \text{e} \quad z^T y_k = 0.$$

Por um lado,

$$\langle z, s_k \rangle_{B_k} = \|z\|_{B_k} \|s_k\|_{B_k} \iff \exists \lambda \in \mathbb{R} \text{ tal que } z = \lambda s_k.$$

Por outro, $z^T y_k = 0$, logo

$$\lambda s_k^T y_k = 0 \iff \lambda = 0.$$

Absurdo, pois $z \neq 0$.

Agora, usando um argumento indutivo, as condições para manter o Algoritmo 2.1 bem definido são válidas. De fato, o método inicia com uma matriz B_0 simétrica e definida positiva (portanto, d^0 é direção de descida) e, uma vez que $0 < c_2 < 1$, temos

$$\begin{aligned} y_k^T s_k &= \alpha_k \left(\nabla F(x^{k+1})^T d^k - \nabla F(x^k)^T d^k \right) \stackrel{(1-10)}{\geq} \alpha_k \left(c_2 \nabla F(x^k)^T d^k - \nabla F(x^k)^T d^k \right) \\ &= \alpha_k (c_2 - 1) \nabla F(x^k)^T d^k > 0, \end{aligned}$$

sempre que d^k for uma direção de descida. Vale a pena ressaltar que se a função objetivo é estritamente convexa, então $s_k^T y_k > 0$ independentemente da busca linear utilizada. Uma vantagem prática do método BFGS consiste no fato de que a inversa da matriz B_{k+1} no Passo 4 possui fórmula fechada dada por:

$$H_{k+1} := B_{k+1}^{-1} = (I_n - \rho_k s_k y_k^T) H_k (I_n - \rho_k y_k s_k^T) + \rho_k s_k s_k^T, \quad (2-3)$$

em que $\rho_k := 1/s_k^T y_k$. Logo, o método pode trabalhar diretamente com as aproximações da inversa da Hessiana, simplificando o cálculo da direção de busca no Passo 1.

O Algoritmo 2.1, além de bem definido para problemas não convexos, converge e sua taxa de convergência é superlinear, uma leitura mais detalhada pode ser encontrada em [6, 62]. Por outro lado, o método pode não convergir para funções gerais não convexas, como foi mostrado em [14, 15, 54]. Vale ressaltar que em [15]

foi mostrado que o método BFGS, fazendo busca exata para calcular o tamanho de passo, ciclou entre pontos não estacionários ao tentar minimizar uma função polinomial que, ao longo de cada direção de busca, é fortemente convexa. Apesar de não abordarmos com detalhes a convergência do Algoritmo 2.1, faremos uma discussão geral.

Como discutido na Seção 1.3, métodos de descida que utilizam as condições de Wolfe, sob algumas hipóteses razoáveis, geram uma sequência de iterações que satisfaz a condição de Zoutendijk (1-15). Assim, é possível obter a convergência do método BFGS sem usar hipótese de convexidade forte, para isso, basta garantir que a direção de busca (2-1) nunca esteja *muito próxima da ortogonalidade* com o gradiente. Para tal, é suficiente que o número de condição das matrizes B_k seja uniformemente limitado, isto é, existe $M > 0$ tal que

$$\|B_k\| \|B_k^{-1}\| \leq M, \quad \forall k. \quad (2-4)$$

De fato, a limitação acima garante que $\frac{\sigma_k^{\max}}{\sigma_k^{\min}} \leq M$ para todo k , nos quais σ_k^{\max} e σ_k^{\min} são, respectivamente, o maior e menor autovalor da matriz B_k . Logo,

$$\begin{aligned} \cos \beta_k &= \frac{\nabla F(x^k)^T B_k^{-1} \nabla F(x^k)}{\|\nabla F(x^k)\| \|B_k^{-1} \nabla F(x^k)\|} \geq \frac{\frac{1}{\sigma_k^{\max}} \|\nabla F(x^k)\|^2}{\|\nabla F(x^k)\| \|B_k^{-1} \nabla F(x^k)\|} \\ &= \frac{\frac{1}{\sigma_k^{\max}} \|\nabla F(x^k)\|^2}{\|\nabla F(x^k)\| \frac{1}{\sigma_k^{\min}} \|\nabla F(x^k)\|} = \frac{\sigma_k^{\min}}{\sigma_k^{\max}} \geq \frac{1}{M}, \end{aligned}$$

em que β_k é o ângulo entre d^k e $-\nabla F(x^k)$. Portanto, se a condição (2-4) é válida, então, por (1-15), temos

$$\lim_{k \rightarrow \infty} \|\nabla F(x^k)\| \rightarrow 0, \quad (2-5)$$

ou, equivalentemente, todo ponto limite de $\{x^k\}$ é estacionário. Neste trabalho, dizemos que um método é *globalmente convergente* se

$$\liminf_{k \rightarrow \infty} \|\nabla F(x^k)\| \rightarrow 0,$$

que é uma condição mais fraca do que (2-5). Desse modo, o Algoritmo 2.1 é globalmente convergente. Porém, é conhecido que o método BFGS não necessariamente encontra pontos estacionários para problemas não convexos gerais, logo não é possível estabelecer um limitante para o número de condição das aproximações Hessianas B_k . Até mesmo para problemas convexos parece não ser possível obter tal limitação, veja [60, Seção 6.4]. Assim, para estabelecer a convergência do método BFGS, outros caminhos são necessários.

O Teorema 2.1 a seguir, sob hipóteses adequadas, mostra em particular que,

dada uma quantidade de iterações, podemos tomar uma fração $p \in (0, 1)$ e garantir que $\cos \beta_k$ fica uniformemente limitado longe do zero para uma fração p das iterações. Em outras palavras, é possível provar que a direção d^k não fica muito próxima da ortogonalidade com o gradiente para uma fração p das iterações. Note que $s_k = \alpha_k d^k$ e $-\nabla F(x) = B_k d^k$, assim, também podemos considerar a ortogonalidade entre s_k e $B_k s_k$ e obter o mesmo resultado. Até onde sabemos, este resultado foi originalmente mostrado em [6, Teorema 2.1] que generalizou [62, Lema 4], em que o autor mostrou que $\cos \beta_k$ fica limitado longe do zero para pelo menos metade das iterações. Vale ressaltar que em [6, Teorema 2.1], é considerada uma busca linear geral que inclui a busca de Wolfe.

Teorema 2.1. [6, Teorema 2.1] *Sejam $a, b \in \mathbb{R}$ constantes positivas e $\{B_k\}$ a sequência gerada pela fórmula BFGS (2-2), em que B_0 é simétrica e definida positiva, e em que y_k e s_k satisfazem*

$$\frac{s_k^T y_k}{s_k^T s_k} \geq a > 0, \quad (2-6)$$

$$\frac{\|y_k\|^2}{s_k^T y_k} \leq b, \quad \forall k \geq 0. \quad (2-7)$$

Então, para qualquer $p \in (0, 1)$ existem constantes $\delta_1, \delta_2, \delta_3 > 0$ tais que, para qualquer $k > 0$, as relações

$$\cos \beta_i \geq \delta_1, \quad (2-8)$$

$$\delta_2 \leq \frac{s_i^T B_i s_i}{s_i^T s_i} \leq \delta_3, \quad (2-9)$$

$$\delta_2 \leq \frac{\|B_i s_i\|}{\|s_i\|} \leq \frac{\delta_3}{\delta_1} \quad (2-10)$$

valem para pelo menos $\lceil p(k+1) \rceil$ valores de $i \in [0, k]$.

As iterações que satisfazem (2-8)–(2-10) compreendem um conjunto de boas iterações. O Teorema 2.1 estabelece que se as condições de atualização (2-6)–(2-7) são satisfeitas, então uma fração p das iterações são boas. Como p pode ser escolhido próximo de 1, entendemos então que a maioria das iterações são boas. A Hipótese 2.1 a seguir fornece condições suficientes para que as condições (2-6)–(2-7) sejam válidas.

Hipótese 2.1. (i) F é duas vezes continuamente diferenciável. (ii) O conjunto de nível $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n \mid F(x) \leq F(x^0)\}$ é convexo e existem constantes $a, b > 0$ tais que

$$a\|z\|^2 \leq z^T \nabla^2 F(x) z \leq b\|z\|^2, \quad (2-11)$$

para todo $z \in \mathbb{R}^n$ e $x \in \mathcal{L}(x^0)$.

Observação 2.1. A Hipótese 2.1 é usual para provar a convergência do método BFGS clássico. Além disso, a condição (2-11) garante que, no conjunto de nível $\mathcal{L}(x^0)$, a função objetivo é fortemente convexa, seu gradiente é Lipschitz, $\nabla^2 F(x)$ é definida positiva e F possui um minimizador único x^* .

Apesar de não ser possível garantir que o $\cos \beta_k$ não fica limitado longe do zero em todas as iterações, sob a Hipótese 2.1, é possível mostrar que existe uma subsequência de índices $\{\ell_k\}_{k=1,2,\dots}$ tal que $\cos \beta_{\ell_k} \geq \delta > 0$, para alguma constante $\delta \in \mathbb{R}_+$. Assim, pela condição de Zoutendijk, o limite (1-16) garante que $\liminf_{k \rightarrow \infty} \|\nabla F(x^k)\| = 0$. Como o problema é fortemente convexo em $\mathcal{L}(x^0)$, esse último limite garante que $x^k \rightarrow x^*$. Além da convergência, pode-se mostrar que a taxa de convergência das iterações é R-linear. Em particular, a sequência converge com rapidez suficiente para que

$$\sum_{k=0}^{\infty} \|x^k - x^*\| < \infty. \quad (2-12)$$

A finitude da soma acima, até onde sabemos, foi provada pela primeira vez em [62, Lema 5]. É possível mostrar, sob hipóteses adicionais formalmente descritas a seguir, que a taxa de convergência do método BFGS é Q-superlinear.

Hipótese 2.2. (i) F é duas vezes continuamente diferenciável. (ii) A sequência $\{x^k\}$ gerada pelo Algoritmo 2.1 converge para um ponto crítico x^* . (iii) $\nabla^2 F(x^*)$ é definida positiva e L -Lipschitz contínua em x^* . Portanto, existe uma vizinhança U de x^* e existem constantes positivas μ , L , e L_2 tais que

$$\mu \|z\|^2 \leq z^T \nabla^2 F(x) z \leq L \|z\|^2, \quad (2-13)$$

e

$$\|\nabla^2 F(x) - \nabla^2 F(x^*)\| \leq L_2 \|x - x^*\|, \quad (2-14)$$

para todo $z \in \mathbb{R}^n$ e $x \in U$.

Note que a Hipótese 2.2 provê propriedades locais do Algoritmo 2.1. Assim, para analisar a taxa de convergência do método BFGS, é suficiente que a função objetivo F seja fortemente convexa na vizinhança de x^* e a Hessiana seja Lipschitz contínua em x^* . Logo, sem perda de generalidade, é considerado que $\{x^k\} \subset U$, isto é, (2-13)–(2-14) valem em x^k para todo $k \geq 0$. Como a Hipótese 2.2 é mais restritiva que a Hipótese 2.1, então (2-12) é válido. Em particular, a condição (2-12)

é essencial para provar que vale o limite de Dennis-Moré [20], dado por:

$$\lim_{k \rightarrow \infty} \frac{\| (B_k - \nabla^2 F(x^*)) d^k \|}{\| d^k \|} = 0. \quad (2-15)$$

Vale ressaltar que (2-15) é necessário e suficiente para que a taxa de convergência de um método quase-Newton seja Q-superlinear. Um fator surpreendente em (2-15) é que as aproximações B_k não necessariamente convergem para $\nabla^2 F(x^*)$, sendo suficiente que B_k aproxime cada vez mais precisamente $\nabla^2 F(x^*)$ ao longo das direções de busca d^k . Como consequência do limite de Dennis-Moré, o comprimento de passo $\alpha_k = 1$ satisfaz (1-9)–(1-10) para k suficientemente grande, e além disso temos

$$\lim_{k \rightarrow \infty} \frac{\| x^{k+1} - x^* \|}{\| x^k - x^* \|} = 0,$$

isto é, o método BFGS converge de maneira assintótica a uma taxa Q-superlinear.

Por mais de três décadas a convergência do método BFGS para problemas não convexos esteve em aberto. Durante esse tempo, alguns trabalhos surgiram com o intuito de jogar luz sobre essa questão, propondo versões globalizadas do método BFGS (veja [46, 47]). Além do próprio Algoritmo 2.1, essas globalizações também serviram de inspiração para propormos alguns algoritmos do tipo BFGS para otimização multiobjetivo, veja Seções 3.2 e 3.3 do Capítulo 3. A fim de uma melhor compreensão dos Algoritmos 3.2 e 3.3 nas seções supracitadas, vamos destacar, nas observações a seguir, quais as principais características das versões globalizadas.

Observação 2.2. Em [46], os autores propuseram uma modificação do método BFGS escalar para otimização irrestrita, que é globalmente convergente mesmo para problemas não convexos. A principal diferença entre a versão modificada e o método clássico está na estrutura da atualização (2-2), que no trabalho citado, é:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\gamma_k \gamma_k^T}{s_k^T \gamma_k}, \quad (2-16)$$

em que $\gamma_k := y_k + r_k s_k$, $r_k \in [0, C]$ e $C \in \mathbb{R}_{++}$ é um parâmetro algorítmico. Nessa proposta, também se faz necessário que o comprimento de passo satisfaça as condições de Wolfe padrão (1-9)–(1-10) para garantir que (2-16) se mantenha definida positiva. Apesar de não entrarmos em detalhes quanto à prova de convergência, vale destacar que os autores provaram que o algoritmo converge com taxa Q-superlinear, além de proporem um parâmetro r_k praticável.

Observação 2.3. Em [47], a modificação proposta pelos autores pode ser considerada mais simples. Considerando que $s_k^T y_k > 0$ é o elemento principal para manter a atualização (2-2) sempre definida positiva, eles propuseram o seguinte esquema de atualização

$$B_{k+1} := \begin{cases} B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}, & \text{se } \frac{s_k^T y_k}{\|s_k\|^2} \geq \varepsilon \|\nabla F(x^k)\|^\omega, \\ B_k, & \text{caso contrário,} \end{cases}$$

nos quais $\varepsilon, \omega > 0$ são parâmetros algorítmicos. Com essa modificação, o algoritmo somente irá atualizar a matriz que aproxima a Hessiana, se uma cautelosa condição for satisfeita. Desse modo, o algoritmo proposto não depende de calcular passos satisfazendo as condições de Wolfe com o intuito de manter B_{k+1} definida positiva, bastando que o comprimento de passo satisfaça uma condição do tipo Armijo para obter a convergência global do método. Além disso, a convergência Q-superlinear também é obtida sob hipóteses razoáveis.

A discussão feita nesta seção é de suma importância para um bom entendimento do Capítulo 3, no qual propomos uma extensão do Algoritmo 2.1 para otimização multiobjetivo seguida por dois outros algoritmos que foram inspirados em [46] e [47]. O presente trabalho é o primeiro a propor um método BFGS com busca de Wolfe para otimização multiobjetivo (Algoritmo 3.1), mimetizando o Algoritmo 2.1. Na seção seguinte, em particular, faremos uma breve revisão da literatura a respeito de métodos quase-Newton para otimização multiobjetivo e provaremos alguns resultados técnicos que serão importantes ao longo do trabalho.

2.2 BFGS multiobjetivo

Existem diversos métodos quase-Newton para otimização multiobjetivo, veja [1, 44, 52, 58, 61, 65, 66]. Nesta seção nos concentraremos naqueles que se propuseram a estender o método BFGS para otimização multiobjetivo e buscaram mimetizar suas propriedades.

Um método quase-Newton para otimização multiobjetivo, assim como no caso escalar, é inspirado no método de Newton. Como visto na Seção 1.2, diversos conceitos clássicos de otimização escalar foram estendidos para o contexto multiobjetivo. Em particular, os autores em [28] estenderam o método de Newton para otimização multiobjetivo, no qual a direção de busca de F a partir de um ponto x

é dada como sendo a solução do seguinte problema

$$\min_{d \in \mathbb{R}^n} \max_{j=1, \dots, m} \nabla F_j(x)^T d + \frac{1}{2} d^T B_j d, \quad (2-17)$$

em que $B_j = \nabla^2 F_j(x)$ para todo $j = 1, \dots, m$. Observe que no caso escalar, (2-17) recupera a direção de Newton clássica. Além disso, note que se $B_j = I_n$ para todo $j = 1, \dots, m$, então (2-17) corresponde à direção de máxima descida (1-4). Finalmente, para obtermos uma *direção quase-Newton multiobjetivo*, basta tomarmos $B_j \in \mathbb{R}^{n \times n}$ como alguma aproximação de $\nabla^2 F_j(x)$ para todo $j = 1, \dots, m$. Agora, se $B_j \succ 0$ para todo $j = 1, \dots, m$, então a função $d \mapsto \nabla F_j(x)^T d + \frac{1}{2} d^T B_j d$ é fortemente convexa e conseqüentemente (2-17) possui uma única solução. Vamos denotar o valor ótimo do problema (2-17) por $\theta(x)$, isto é,

$$d(x) := \operatorname{argmin}_{d \in \mathbb{R}^n} \max_{j=1, \dots, m} \nabla F_j(x)^T d + \frac{1}{2} d^T B_j d, \quad (2-18)$$

e

$$\theta(x) := \max_{j=1, \dots, m} \nabla F_j(x)^T d(x) + \frac{1}{2} d(x)^T B_j d(x). \quad (2-19)$$

Além disso, note que (2-17) é equivalente ao seguinte problema de otimização quadrática convexa:

$$\begin{aligned} \min_{(t, d) \in \mathbb{R} \times \mathbb{R}^n} \quad & t \\ \text{s. a.} \quad & \nabla F_j(x)^T d + \frac{1}{2} d^T B_j d \leq t, \quad \forall j = 1, \dots, m. \end{aligned} \quad (2-20)$$

A única solução de (2-20) é dada por $(t, d) := (\theta(x), d(x))$. Como (2-20) é convexo e possui um ponto de Slater, por exemplo, $(1, 0) \in \mathbb{R} \times \mathbb{R}^n$, então existe multiplicador de Lagrange $\lambda(x) \in \mathbb{R}^m$ tal que a tripla $(t, d, \lambda) := (\theta(x), d(x), \lambda(x)) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$ satisfaz o sistema Karush-Kuhn-Tucker:

$$\sum_{j=1}^m \lambda_j = 1, \quad \sum_{j=1}^m \lambda_j [\nabla F_j(x) + B_j d] = 0,$$

e

$$\lambda_j \geq 0, \quad \nabla F_j(x)^T d + \frac{1}{2} d^T B_j d \leq t, \quad \lambda_j \left[\nabla F_j(x)^T d + \frac{1}{2} d^T B_j d - t \right] = 0,$$

para todo $j = 1, \dots, m$. Portanto, após algumas manipulações algébricas obtemos

$$d(x) = - \left[\sum_{j=1}^m \lambda_j(x) B_j \right]^{-1} \sum_{j=1}^m \lambda_j(x) \nabla F_j(x), \quad (2-21)$$

$$\sum_{j=1}^m \lambda_j(x) = 1, \quad \lambda_j(x) \geq 0, \quad \forall j = 1, \dots, m, \quad (2-22)$$

e

$$\begin{aligned} \theta(x) &= \sum_{j=1}^m \lambda_j(x) \nabla F_j(x)^T d(x) + \frac{1}{2} d(x)^T \left(\sum_{j=1}^m \lambda_j(x) B_j \right) d(x) \\ &\stackrel{(2-21)}{=} -\frac{1}{2} d(x)^T \left[\sum_{j=1}^m \lambda_j(x) B_j \right] d(x). \end{aligned} \quad (2-23)$$

O próximo resultado, similarmente ao Lema 1.2, mostra que a direção $d(x)$ e o valor ótimo $\theta(x)$ podem ser utilizados para caracterizar pontos Pareto críticos do problema (1-1).

Lema 2.1. *Seja $d : \mathbb{R}^n \rightarrow \mathbb{R}^n$ e $\theta : \mathbb{R}^n \rightarrow \mathbb{R}$ dados por (2-18) e (2-19), respectivamente. Assuma que $B_j \succ 0$ para todo $j = 1, \dots, m$. Então:*

- (i) *x é Pareto crítico se e somente se $d(x) = 0$ e $\theta(x) = 0$;*
- (ii) *se x não é Pareto crítico, então $d(x) \neq 0$ e $\mathcal{D}(x, d(x)) < \theta(x) < 0$ (em particular, $d(x)$ é uma direção de descida de F em x).*

Demonstração. Veja [28, Lemma 3.2] e [61, Lemma 2]. ■

Em otimização escalar, um método de busca linear pode se deteriorar caso as direções de busca sejam *quase* ortogonais com os gradientes, isto é, caso os cossenos dos ângulos entre tais direções não se mantenham uniformemente afastados do zero. Comumente a análise de convergência passa por monitorar os ângulos entre as direções de busca e os gradientes. Tecnicamente, impõe-se que

$$\nabla F(x)^T d(x) \leq \Gamma \|\nabla F(x)\| \|d(x)\|,$$

em que $\Gamma \in (0, 1)$ é um parâmetro algorítmico. Essa imposição é conhecida na literatura como *condição de ângulo*, para uma discussão cuidadosa sobre o assunto recomendamos ao leitor visitar [53, Seção 6.1]. Já em otimização multiobjetivo, a condição de ângulo foi estudada em [33], com o intuito de obter convergência de métodos tipo Newton. O seguinte resultado, inspirado por essa condição, relaciona o cosseno do ângulo entre a direção de busca $d(x)$ em (2-21) e $B_j d(x)$, com a direção de máxima descida generalizada (1-7) e a derivada direcional generalizada (1-3). Ele é uma importante ferramenta para a análise de convergência dos algoritmos propostos no Capítulo 3.

Proposição 2.1. *Dado $x \in \mathbb{R}^n$ e B_j definida positiva para $j = 1, \dots, m$, considere $d(x)$ e $\theta(x)$ como definidos em (2-18) e (2-19), respectivamente. Seja $\beta_j(x)$ o ângulo entre as direções $d(x)$ e $B_j d(x)$, para $j = 1, \dots, m$. Então*

$$\mathcal{D}(x, d(x)) \leq -\frac{\delta(x)}{2} \left\| \sum_{j=1}^m \lambda_j(x) \nabla F_j(x) \right\| \|d(x)\| \leq -\frac{\delta(x)}{2} \|d_{SD}(x)\| \|d(x)\|,$$

em que $\delta(x) := \min_{j=1, \dots, m} \cos \beta_j(x)$.

Demonstração. Inicialmente, note que se x é ponto Pareto crítico, então pelo Lema 2.1, $d(x) = 0$ e o resultado segue trivialmente. Agora, se x não é Pareto crítico, então pelas definições de $\delta(x)$, $\cos \beta_j(x)$ e $d(x)$, temos

$$\delta(x) \leq \cos \beta_j(x) = \frac{d(x)^T B_j d(x)}{\|d(x)\| \|B_j d(x)\|}, \quad \forall j = 1, \dots, m.$$

Logo,

$$d(x)^T B_j d(x) \geq \delta(x) \|d(x)\| \|B_j d(x)\|, \quad \forall j = 1, \dots, m.$$

Assim, pelo Lema 2.1(ii) e (2-23), obtemos

$$-\mathcal{D}(x, d(x)) > -\theta(x) = \frac{1}{2} \sum_{j=1}^m \lambda_j(x) d(x)^T B_j d(x) \geq \sum_{j=1}^m \lambda_j(x) \frac{\delta(x)}{2} \|d(x)\| \|B_j d(x)\|.$$

Portanto, a desigualdade triangular, juntamente com (2-21)–(2-22), e o Lema 1.2(iv), implicam que

$$\begin{aligned} -\mathcal{D}(x, d(x)) &> \frac{\delta(x)}{2} \|d(x)\| \left\| \sum_{j=1}^m \lambda_j(x) B_j d(x) \right\| \\ &= \frac{\delta(x)}{2} \|d(x)\| \left\| \sum_{j=1}^m \lambda_j(x) \nabla F_j(x) \right\| \geq \frac{\delta(x)}{2} \|d(x)\| \|d_{SD}(x)\|, \end{aligned}$$

obtendo o resultado desejado. ■

O próximo resultado é uma importante ferramenta para a análise de convergência de métodos multicritério. O Lema 1.2(iv) garante que, para qualquer $x \in \mathbb{R}^n$, $-d_{SD}(x)$ é o elemento de menor norma do envelope convexo de $\{\nabla F_1(x), \dots, \nabla F_m(x)\}$. Porém, em diversas situações, surge a necessidade de considerarmos uma combinação convexa distinta de (1-6)–(1-7). Assim, o resultado a seguir estabelece uma maneira de relacionar a norma de $d_{SD}(x)$ com a norma de outra combinação convexa dos gradientes.

Lema 2.2. [33, Lema 6] *Sejam $x \in \mathbb{R}^n$ e a, b constantes positivas. Assuma que*

$$aI_n \preceq B_j \preceq bI_n, \quad \forall j = 1, \dots, m.$$

Seja $\lambda(x) := (\lambda_1(x), \dots, \lambda_m(x)) \in \mathbb{R}^m$ o multiplicador de Lagrange do problema (2-20) que satisfaz (2-21)–(2-22), e defina

$$\nabla F_\lambda(x) := \sum_{j=1}^m \lambda_j(x) \nabla F_j(x).$$

Então,

$$\frac{a}{b} \|\nabla F_\lambda(x)\| \leq \|d_{SD}(x)\| \leq \|\nabla F_\lambda(x)\|,$$

em que $d_{SD}(x)$ é dado em (1-7). Além disso, pelas definições de $\theta(x)$ e $\theta_{SD}(x)$ em (2-23) e (1-8), respectivamente, temos

$$\frac{1}{b} |\theta_{SD}(x^k)| \leq |\theta(x^k)| \leq \frac{1}{a} |\theta_{SD}(x^k)|.$$

Agora, munidos da direção quase-Newton para otimização multicritério, podemos estabelecer uma estrutura geral de métodos quase-Newton para otimização multiobjetivo.

Algoritmo 2.2. Um algoritmo quase-Newton

Sejam $x^0 \in \mathbb{R}^n$ e $B_j^0 \succ 0$, para todo $j = 1, \dots, m$, dados. Faça $k \leftarrow 0$.

Passo 1. *Calcule a direção de busca*

Calcule $d^k := d(x^k)$ e $\theta(x^k)$ como em (2-18) e (2-19), respectivamente.

Passo 2. *Critério de parada*

Se $\theta(x^k) = 0$, então PARE.

Passo 3. *Procedimento de busca linear*

Calcule o comprimento de passo $\alpha_k > 0$ e faça $x^{k+1} := x^k + \alpha_k d^k$.

Passo 4. *Prepare a próxima iteração*

Para cada $j = 1, \dots, m$, defina B_j^{k+1} . Faça $k \leftarrow k + 1$ e vá ao **Passo 1**.

Observação 2.4. O Algoritmo 2.2 não especifica como o comprimento de passo α_k é calculado e nem mesmo como as matrizes B_j^{k+1} , para $j = 1, \dots, m$, são definidas. Diferentes escolhas para a matriz B_j^{k+1} geram diferentes métodos. Além disso, a maneira de escolher o comprimento de passo pode gerar diferentes resultados de convergência.

Como discutido na Seção 2.1, parece não ser possível estabelecer limitação no número de condição das atualizações BFGS que aproximam as Hessianas.

Alternativamente, apresentamos um resultado muito importante para o método BFGS clássico, o Teorema 2.1. Nesse resultado, a direção s_k se mantém não *muito próxima da ortogonalidade* com $B_k s_k$, para uma fração p arbitrária das iterações. Agora, supondo que o Algoritmo 2.2 define B_j^{k+1} para todo $j = 1, \dots, m$, como em (2-2), podemos estender o Teorema 2.1 para o caso multiobjetivo. Até onde sabemos, este resultado é a primeira versão estendida do Teorema 2.1 a surgir na literatura.

Teorema 2.2. *Sejam $a, b \in \mathbb{R}$ constantes positivas e considere $\{x^k\}$ gerada pelo Algoritmo 2.2 com*

$$B_j^{k+1} = B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{y_j^k (y_j^k)^T}{s_k^T y_j^k}, \quad (2-24)$$

para todo $j = 1, \dots, m$ e $k \geq 0$. Seja $p \in (0, 1)$ qualquer e β_j^k o ângulo entre as direções s_k e $B_j^k s_k$, para todo $j = 1, \dots, m$. Se

$$\frac{s_k^T y_j^k}{s_k^T s_k} \geq a > 0 \quad e \quad \frac{\|y_j^k\|^2}{s_k^T y_j^k} \leq b, \quad \forall j = 1, \dots, m,$$

então existem constantes $\delta_1, \delta_2, \delta_3 > 0$ tais que, para qualquer $k > 0$, as relações

$$\cos \beta_j^i \geq \delta_1, \quad (2-25)$$

$$\delta_2 \leq \frac{s_i^T B_j^i s_i}{s_i^T s_i} \leq \delta_3, \quad (2-26)$$

$$\delta_2 \leq \frac{\|B_j^i s_i\|}{\|s_i\|} \leq \frac{\delta_3}{\delta_1}, \quad (2-27)$$

valem para pelo menos $\lceil p(k+1) \rceil$ valores de $i \in \{0, 1, \dots, k\}$ e para todo $j = 1, \dots, m$.

Demonstração. Inicialmente, vamos provar que para um $j = 1, \dots, m$ fixado, porém arbitrário, e para qualquer $\bar{p} \in (0, 1)$, existem constantes $\delta_j^1, \delta_j^2, \delta_j^3 > 0$ tais que, para qualquer $k > 0$, as relações

$$\cos \beta_j^i \geq \delta_j^1, \quad (2-28)$$

$$\delta_j^2 \leq \frac{s_i^T B_j^i s_i}{s_i^T s_i} \leq \delta_j^3, \quad (2-29)$$

$$\delta_j^2 \leq \frac{\|B_j^i s_i\|}{\|s_i\|} \leq \frac{\delta_j^3}{\delta_j^1} \quad (2-30)$$

valem para pelo menos $\lceil \bar{p}(k+1) \rceil$ valores de $i \in \{0, 1, \dots, k\}$. A demonstração desses fatos é análoga à prova do [6, Teorema 2.1]. Contudo, por completude, vamos mostrá-la aqui.

Dado um $\bar{p} \in (0, 1)$ qualquer e um $j = 1, \dots, m$ fixado, porém arbitrário, defina $q_k := \frac{s_k^T B_j^k s_k}{s_k^T s_k}$, $a_k := \frac{s_k^T y_j^k}{s_k^T s_k}$ e $b_k := \frac{(y_j^k)^T y_j^k}{s_k^T y_j^k}$. Definindo a função $\psi : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ por:

$$\psi(B) := \text{Tr}(B) - \ln(\text{Det}(B)), \quad (2-31)$$

em que \ln é o logaritmo natural, note que, dada $B \in \mathbb{R}^{n \times n}$ definida positiva então $\psi(B) > 0$, pois

$$\psi(B) = \sum_{l=1}^n (\sigma_l - \ln \sigma_l), \quad (2-32)$$

em que σ_l , para $l = 1, \dots, n$, são os autovalores de B . Assim, calculando o traço e o determinante de B_j^{k+1} temos

$$\text{Tr}(B_j^{k+1}) = \text{Tr}(B_j^k) - \frac{\|B_j^k s_k\|^2}{s_k^T B_j^k s_k} + \frac{\|y_j^k\|^2}{s_k^T y_j^k}$$

e

$$\text{Det}(B_j^{k+1}) = \text{Det}(B_j^k) \frac{s_k^T y_j^k}{s_k^T B_j^k s_k}. \quad (2-33)$$

Por hipótese, $a_k \geq a$, $b_k \leq b$ e $\cos \beta_j^k = \frac{s_k^T B_j^k s_k}{\|s_k\| \|B_j^k s_k\|}$, então

$$\begin{aligned} 0 < \psi(B_j^{k+1}) &= \text{Tr}(B_j^k) - \frac{\|B_j^k s_k\|^2}{s_k^T B_j^k s_k} + \frac{\|y_j^k\|^2}{s_k^T y_j^k} - \ln(\text{Det}(B_j^k)) - \ln\left(\frac{s_k^T y_j^k}{s_k^T B_j^k s_k}\right) \\ &= \psi(B_j^k) - \frac{\|B_j^k s_k\|^2 \|s_k\|^2}{(s_k^T B_j^k s_k)^2} \frac{s_k^T B_j^k s_k}{\|s_k\|^2} + b_k - \ln\left(\frac{s_k^T y_j^k}{s_k^T s_k} \frac{s_k^T s_k}{s_k^T B_j^k s_k}\right) \\ &= \psi(B_j^k) - \frac{1}{\cos^2 \beta_j^k} q_k + b_k - \ln(a_k) + \ln(q_k) \\ &= \psi(B_j^k) + [b_k - \ln(a_k) - 1] + \left[1 - \frac{q_k}{\cos^2 \beta_j^k} + \ln\left(\frac{q_k}{\cos^2 \beta_j^k}\right)\right] \\ &\quad + \ln(\cos^2 \beta_j^k) \\ &\leq \psi(B_j^0) + (b - \ln(a) - 1)(k+1) + \sum_{i=0}^k \ln(\cos^2 \beta_j^i) \\ &\quad + \sum_{i=0}^k \left[1 - \frac{q_i}{\cos^2 \beta_j^i} + \ln\left(\frac{q_i}{\cos^2 \beta_j^i}\right)\right]. \end{aligned} \quad (2-34)$$

Definindo a função $u : \mathbb{R} \rightarrow \mathbb{R}$ por

$$u(t) = 1 - t + \ln t, \quad (2-35)$$

note que $u(t)$ é não positiva para todo $t > 0$, atinge seu valor máximo em $t = 1$ e satisfaz $u(t) \rightarrow -\infty$ tanto para $t \rightarrow 0$ quanto para $t \rightarrow +\infty$. Assim, vamos definir $\eta_j^i \geq 0$ por

$$\eta_j^i := -\ln(\cos^2 \beta_j^i) - \left[1 - \frac{q_i}{\cos^2 \beta_j^i} + \ln\left(\frac{q_i}{\cos^2 \beta_j^i}\right) \right].$$

Logo, por (2-34) temos

$$\frac{1}{k+1} \sum_{i=0}^k \eta_j^i < \frac{\psi(B_j^0)}{k+1} + (b-1-\ln a). \quad (2-36)$$

Seja J_j^{k+1} o conjunto formado dos $[\bar{p}(k+1)]$ índices correspondentes aos $[\bar{p}(k+1)]$ menores valores de η_j^i , para $i \leq k$, e seja $\bar{\eta}_j^{k+1} := \max_{i \in J_j^{k+1}} \{\eta_j^i\}$. Então

$$\begin{aligned} \frac{1}{k+1} \sum_{i=1}^k \eta_j^i &= \frac{1}{k+1} \left[\sum_{\substack{i=0, \\ i \in J_j^{k+1}}}^k \eta_j^i + \sum_{\substack{i=0, \\ i \notin J_j^{k+1}}}^k \eta_j^i \right] \\ &\geq \frac{1}{k+1} \left[\bar{\eta}_j^{k+1} + \sum_{\substack{i=0, \\ i \notin J_j^{k+1}}}^k \eta_j^i \right] \\ &\geq \frac{1}{k+1} \left[\bar{\eta}_j^{k+1} + \bar{\eta}_j^{k+1}(k+1 - [\bar{p}(k+1)]) \right] \\ &\geq \bar{\eta}_j^{k+1}(1 - \bar{p}), \end{aligned}$$

em que a última desigualdade é válida, pois

$$\begin{aligned} 1 + (k+1)\bar{p} &\geq [\bar{p}(k+1)] \Rightarrow \bar{\eta}_j^{k+1} + \bar{\eta}_j^{k+1}(k+1)\bar{p} \geq \bar{\eta}_j^{k+1}[\bar{p}(k+1)] \Rightarrow \\ &\Rightarrow \bar{\eta}_j^{k+1} - \bar{\eta}_j^{k+1}[\bar{p}(k+1)] + \bar{\eta}_j^{k+1}(k+1) \geq -\bar{\eta}_j^{k+1}(k+1)\bar{p} + \bar{\eta}_j^{k+1}(k+1) \Rightarrow \\ &\Rightarrow \bar{\eta}_j^{k+1} + \bar{\eta}_j^{k+1}(k+1 - [\bar{p}(k+1)]) \geq (k+1)\bar{\eta}_j^{k+1}(1 - \bar{p}). \end{aligned}$$

Logo, por (2-36) e para todo $i \in J_j^{k+1}$, temos que

$$\begin{aligned} \eta_j^i &< \bar{\eta}_j^{k+1} \leq \frac{1}{1-\bar{p}} \frac{1}{k+1} \sum_{l=0}^k \eta_j^l \\ &\leq \frac{1}{1-\bar{p}} \left[\frac{\psi(B_j^0)}{k+1} + (b-1-\ln(a)) \right] \\ &\leq \frac{1}{1-\bar{p}} \left[\psi(B_j^0) + (b-1-\ln(a)) \right] := \delta_j^0. \end{aligned}$$

Assim,

$$\begin{aligned} -\ln(\cos^2 \beta_j^i) &\leq -\ln(\cos^2 \beta_j^i) - \left[1 - \frac{q_i}{\cos^2 \beta_j^i} + \ln\left(\frac{q_i}{\cos^2 \beta_j^i}\right) \right] = \eta_j^i \\ &< \delta_j^0, \end{aligned} \quad (2-37)$$

em que a primeira desigualdade segue, pois a função $u(t)$ em (2-35) é não positiva para $t > 0$. Dessa maneira,

$$\cos \beta_j^i \geq e^{-\delta_j^0/2} := \delta_j^1. \quad (2-38)$$

Para garantir que $\delta_j^1 > 0$, basta assumir, sem perda de generalidade, que $b - \ln(a) - 1$ é positivo. Por (2-37) também temos que

$$1 - \frac{q_i}{\cos^2 \beta_j^i} + \ln\left(\frac{q_i}{\cos^2 \beta_j^i}\right) > -\delta_j^0, \quad \forall i \in J_j^{k+1}.$$

Pelas propriedades da função $u(t)$ em (2-35), segue que

$$0 < \tilde{\delta}_j^2 \leq \frac{q_i}{\cos^2 \beta_j^i} \leq \delta_j^3, \quad \forall i \in J_j^{k+1},$$

para constantes $\tilde{\delta}_j^2$ e δ_j^3 . Desse modo, usando (2-38) obtemos

$$\delta_j^2 := (\delta_j^1)^2 \tilde{\delta}_j^2 \leq q_i \leq \delta_j^3, \quad \forall i \in J_j^{k+1},$$

como $\frac{\|B_j^i s_i\|}{\|s_i\|} = \frac{q_i}{\cos \beta_j^i}$, temos que

$$\delta_j^2 \leq \frac{\|B_j^i s_i\|}{\|s_i\|} \leq \frac{\delta_j^3}{\delta_j^1}, \quad \forall i \in J_j^{k+1}.$$

Agora, vamos mostrar que para qualquer $p \in (0, 1)$ e para qualquer $k \geq 1$, as condições (2-25), (2-26) e (2-27) são válidas para pelo menos $\lceil p(k+1) \rceil$ valores

de $i \in \{0, 1, \dots, k\}$, simultaneamente para todos os objetivos. Inicialmente, defina $\bar{p} := (1 - \frac{1-p}{m})$ e note que, para cada objetivo $j = 1, \dots, m$, foi mostrado que existem constantes $\delta_j^1, \delta_j^2, \delta_j^3 > 0$ tais que as condições (2-28), (2-29) e (2-30) valem para pelo menos $\lceil \bar{p}(k+1) \rceil$ valores de $i \in \{0, 1, \dots, k\}$. Defina $\delta_1 := \min_{j=1, \dots, m} \{\delta_j^1\}$, $\delta_2 := \min_{j=1, \dots, m} \{\delta_j^2\}$, $\delta_3 := \max_{j=1, \dots, m} \{\delta_j^3\}$,

$$\mathcal{G}_j^k := \left\{ i \in \{0, 1, \dots, k\} \mid \cos \beta_j^i \geq \delta_1, \delta_2 \leq \frac{s_i^T B_i s_i}{s_i^T s_i} \leq \delta_3 \text{ e } \delta_2 \leq \frac{\|B_j^i s_i\|}{\|s_i\|} \leq \frac{\delta_3}{\delta_1} \right\}$$

e $\mathcal{B}_j^k := \{\ell \in \{0, 1, \dots, k\} \mid \ell \notin \mathcal{G}_j^k\}$, para todo $j = 1, \dots, m$. Note que \mathcal{B}_j^k é o complementar de \mathcal{G}_j^k , assim $k+1 = |\mathcal{G}_j^k| + |\mathcal{B}_j^k|$, para todo $j = 1, \dots, m$. Dessa maneira, pela definição de \bar{p} e usando algumas propriedades das funções *ceiling* e *floor*, temos

$$\begin{aligned} |\mathcal{B}_j^k| &\geq \lceil \bar{p}(k+1) \rceil = \lceil (1 - \frac{1-p}{m})(k+1) \rceil = \lceil k+1 + (-\frac{1-p}{m})(k+1) \rceil \\ &= k+1 + \lceil -\frac{1-p}{m}(k+1) \rceil \\ &= k+1 - \lfloor \frac{1-p}{m}(k+1) \rfloor, \end{aligned} \quad (2-39)$$

logo,

$$|\mathcal{B}_j^k| = k+1 - |\mathcal{G}_j^k| \stackrel{(2-39)}{\leq} k+1 - k - 1 + \lfloor \frac{1-p}{m}(k+1) \rfloor = \lfloor \frac{1-p}{m}(k+1) \rfloor,$$

para todo $j = 1, \dots, m$. Consequentemente,

$$|\bigcup_{j=1}^m \mathcal{B}_j^k| \leq m \lfloor \frac{1-p}{m}(k+1) \rfloor \leq (1-p)(k+1).$$

Além disso, perceba que o conjunto $\bigcup_{j=1}^m \mathcal{B}_j^k$ é o complementar do conjunto $\bigcap_{j=1}^m \mathcal{G}_j^k$, assim, $k+1 = |\bigcap_{j=1}^m \mathcal{G}_j^k| + |\bigcup_{j=1}^m \mathcal{B}_j^k|$. Portanto, segue que

$$|\bigcap_{j=1}^m \mathcal{G}_j^k| \geq k+1 - (1-p)(k+1) = p(k+1),$$

como a cardinalidade de um conjunto sempre é um número inteiro, completamos a prova. ■

Note que o resultado acima mostra que $\cos \beta_j^k$ se mantém afastado do zero para uma fração arbitrária p de iterações, *simultaneamente* para todos os objetivos. Além disso, assim como no caso escalar, a validade do Teorema 2.2 não depende da escolha do tamanho de passo α_k , tampouco da definição dos vetores s_k e y_j^k para

$j = 1, \dots, m$. Essa independência quanto à definição dos vetores é crucial para que esse resultado possa ser utilizado para estudar a convergência do Algoritmos 3.1 e 3.2 no capítulo seguinte.

Agora, vamos analisar alguns métodos quase-Newton que utilizaram a atualização BFGS (2-24) para calcular B_j^{k+1} . Essa análise tem como objetivo mostrar algumas das motivações que nos levaram a também propor uma extensão do método BFGS. Até onde sabemos, [65] foi o primeiro trabalho a propor um método quase-Newton multiobjetivo, como o Algoritmo 2.2, no qual B_j^{k+1} é calculada por (2-24).

Inspirados em uma versão do BFGS escalar mencionada na Observação 2.3, em [65] os autores propuseram um método quase-Newton para otimização multiobjetivo que não necessariamente atualiza as aproximações das Hessianas a cada iteração, a saber:

$$B_j^{k+1} := \begin{cases} B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{y_j^k (y_j^k)^T}{s_k^T y_j^k}, & \text{se } s_k^T y_j^k \geq \varepsilon \min\{1, |\theta(x^k)|\}, \\ B_j^k, & \text{caso contrário,} \end{cases} \quad (2-40)$$

para todo $j = 1, \dots, m$, no qual $\theta(x^k)$ está definido em (2-19) e $\varepsilon > 0$ é um parâmetro algorítmico. Além disso, eles impuseram que o tamanho de passo satisfaça uma condição do tipo Armijo. Pelo que sabemos, até o presente trabalho, atualizações cautelosas como em (2-40), eram as únicas alternativas aos métodos BFGS quando aplicados a problemas multiobjetivo não convexos. Usando hipóteses pouco razoáveis, os autores mostraram que todo ponto de acumulação da sequência gerada pelo método proposto é Pareto crítico. Destacamos as seguintes hipóteses utilizadas: existe uma constante positiva $c \in \mathbb{R}$ tal que $\|B_j^k\| \leq c$ para todo $j = 1, \dots, m$ em todas as k -iterações e, para k suficientemente grande, o comprimento de passo unitário é aceito por uma condição do tipo Armijo.

Em [61] foi proposto um método BFGS multiobjetivo que atualiza as matrizes B_j^k em todas as iterações. Do ponto de vista prático, o algoritmo não está bem definido para problemas não estritamente convexos. Isso ocorre, pois, como foi discutido na Seção 2.1, no caso escalar só é possível garantir que a atualização (2-2) é simétrica e definida positiva caso a matriz B_k seja simétrica e definida positiva e $s_k^T y_k > 0$. Como é sabido, essa última condição pode ser obtida se a função objetivo for estritamente convexa ou então se o tamanho de passo satisfizer as condições de Wolfe. Contudo, no caso multiobjetivo, a única maneira de garantir que $s_k^T y_j^k > 0$ para todo $j = 1, \dots, m$, é supor que a função objetivo seja estritamente convexa, pois mesmo que o comprimento de passo satisfaça as condições de Wolfe multiobjetivo, ainda assim $s_k^T y_j^k$ pode ser negativo para algum $j = 1, \dots, m$. Uma discussão mais

aprofundada sobre esse assunto está feita no Capítulo 3 (veja o Exemplo 3.1). Agora, do ponto de vista teórico, as hipóteses usadas para provar que o método converge e que a taxa de convergência é Q-superlinear, são pouco razoáveis. Dentre elas, vale destacar a suposição de que existem constantes positivas $a, b \in \mathbb{R}$ tais que

$$aI_n \preceq B_j^k \preceq bI_n,$$

para todo $j = 1, \dots, m$ em todas as k -iterações e a suposição de que vale o limite de Dennis–Moré (2-15) para cada objetivo. Note que essa hipótese implica que o número de condição das matrizes B_j^k é uniformemente limitado e, como discutimos na Seção 2.1, não parece ser possível garantir essa propriedade nem mesmo em problemas escalares convexos.

Por fim, em [52] os autores propuseram uma versão multiobjetivo do método BFGS onde a busca linear é não monótona. Apesar do algoritmo proposto só estar bem definido para problemas estritamente convexos, do ponto de vista teórico, eles avançaram quanto à convergência Q-superlinear. Até onde sabemos, foram os primeiros a mostrar que o limite de Dennis–Moré (2-15) vale para todos os objetivos simultaneamente. Porém, das hipóteses utilizadas para obter a convergência Q-superlinear, duas podem ser consideradas pouco razoáveis. A primeira delas é a limitação inferior dos autovalores da matriz B_j^k , isto é, supor que existe uma constante positiva $a \in \mathbb{R}$ tal que $B_j^k \succeq aI_n$, para todo $j = 1, \dots, m$ em todas as k -iterações. A segunda é supor que $\sum_{k=0}^{\infty} \|x^k - x^*\| < \infty$, e como discutimos na Seção 2.1, essa hipótese está diretamente ligada ao limite de Dennis–Moré.

Em síntese, destacamos algumas deficiências existentes nas referências supracitadas: (i) os algoritmos são geralmente projetados para problemas fortemente convexos; (ii) as atualizações BFGS e suas inversas são frequentemente assumidas como uniformemente limitadas; (iii) algumas importantes etapas intermediárias (limite de Dennis–Moré, admissibilidade do passo unitário e validade da condição (2-12)) na análise de convergência são frequentemente assumidos sem prova. Como o leitor poderá observar na Seção 3.1, todos esses inconvenientes são superados pelo método proposto. A análise de convergência desse método utiliza hipóteses que são extensões naturais das consideradas no caso escalar (Hipóteses 2.1 e 2.2). No próximo capítulo, além da extensão já mencionada, também propomos dois algoritmos inspirados em [46] e [47], que globalizam o método BFGS para otimização multiobjetivo, isto é, as sequências geradas por esse métodos são tais que

$$\liminf_{k \rightarrow \infty} \|d_{SD}(x^k)\| = 0. \quad (2-41)$$

Em outras palavras, no presente trabalho, dizemos que um método multiobjetivo é

globalmente convergente se satisfaz (2-41).

Novos métodos BFGS para otimização multiobjetivo

Neste capítulo propomos três algoritmos quase-Newton, do tipo BFGS, para otimização multiobjetivo. O primeiro deles estende o Algoritmo 2.1 para o caso multicritério, enquanto os outros dois, inspirados por [46] e [47], respectivamente, estendem versões globalizadas do método BFGS. Desse modo, este capítulo é composto por três seções, nas quais estudaremos a convergência dos Algoritmos 3.1, 3.2 e 3.3, respectivamente. Os desempenhos numéricos desses algoritmos também foram investigados. Os Algoritmos 3.1 e 3.2 são baseados no Algoritmo 2.2, que atualiza as matrizes B_j^k em todas as iterações. Em contrapartida, o Algoritmo 3.3 usa uma condição cautelosa para atualizar as matrizes B_j^k . Um fator comum a todos esses algoritmos é a utilização do comprimento de passo satisfazendo as condições de Wolfe padrão (1-12)–(1-13), o que nos permite utilizar a condição de Zoutendijk para analisar a convergência desses métodos. O leitor pode notar que para cada algoritmo proposto neste capítulo, as hipóteses utilizadas na análise de convergência são distintas. Assim, para garantir a validade da condição de Zoutendijk (1-17) em cada caso, fizemos a análise separadamente.

Os Algoritmos 3.1 e 3.2 têm em comum a estrutura da fórmula de atualização das matrizes B_j^k para $j = 1, \dots, m$, dada por

$$\begin{aligned}
 B_j^{k+1} := & B_j^k - \frac{(\rho_j^k)^{-1} B_j^k s_k s_k^T B_j^k}{\left((\rho_j^k)^{-1} - s_k^T y_j^k \right)^2 + (\rho_j^k)^{-1} s_k^T B_j^k s_k} \\
 & + \frac{(s_k^T B_j^k s_k) y_j^k (y_j^k)^T}{\left((\rho_j^k)^{-1} - s_k^T y_j^k \right)^2 + (\rho_j^k)^{-1} s_k^T B_j^k s_k} \\
 & + \left((\rho_j^k)^{-1} - s_k^T y_j^k \right) \frac{y_j^k s_k^T B_j^k + B_j^k s_k (y_j^k)^T}{\left((\rho_j^k)^{-1} - s_k^T y_j^k \right)^2 + (\rho_j^k)^{-1} s_k^T B_j^k s_k},
 \end{aligned} \tag{3-1}$$

em que $s_k := x^{k+1} - x^k$ e $\rho_j^k > 0$ para todo $j = 1, \dots, m$. Os Algoritmos 3.1 e 3.2 usam

definições distintas para ρ_j^k e y_j^k , para $j = 1, \dots, m$, as quais estão especificadas, respectivamente, nas Seções 3.1 e 3.2. Note que se $y_j^k = \nabla F_j(x^{k+1}) - \nabla F_j(x^k)$ e $(\rho_j^k)^{-1} = s_k^T y_j^k$, então a expressão em (3-1) recupera a fórmula BFGS clássica (2-24) para F_j . Agora, vamos mostrar as motivações que nos levaram à fórmula de atualização (3-1). Como visto na Seção 2.1, o método BFGS clássico, com intuito de reduzir o custo computacional, utiliza a fórmula (2-3) para atualizar a inversa de B_{k+1} . Denote por H_j^k a aproximação para $[\nabla^2 F_j(x^k)]^{-1}$, para cada $j = 1, \dots, m$. Assim, a atualização BFGS para a inversa é naturalmente definida por

$$H_j^{k+1} := \left(I_n - \rho_j^k s_k (y_j^k)^T \right) H_j^k \left(I_n - \rho_j^k y_j^k s_k^T \right) + \rho_j^k s_k s_k^T, \quad (3-2)$$

com $\rho_j^k > 0$. Tomando a inversa de H_j^{k+1} em (3-2), obtemos a fórmula de atualização B_j^{k+1} em (3-1). De fato, fixando $j = 1, \dots, m$ arbitrário, seja $H_j^0 \succ 0$ e defina $B_j^0 = (H_j^0)^{-1}$. Por indução, suponha que para $i = 1, \dots, k$, $H_j^i \succ 0$ e $(H_j^i)^{-1} = B_j^i$. Vamos mostrar que $H_j^{k+1} \succ 0$ e $(H_j^{k+1})^{-1} = B_j^{k+1}$. Tomando $z \in \mathbb{R}^n$ não nulo, uma vez que $\rho_j^k > 0$, é fácil ver, por (3-2), que $z^T H_j^{k+1} z > 0$ e logo $H_j^{k+1} \succ 0$. Para calcular a inversa, vamos usar a fórmula de Sherman-Morrison-Woodbury, veja [60, Apêndice A.1]. A fórmula estabelece que: dada $\hat{A} = A + UV^T$, em que A é uma matriz quadrada $n \times n$ não singular, U e V são matrizes $n \times \ell$ com $1 \leq \ell \leq n$, então \hat{A} é não singular se, e somente se, a matriz $I_n + V^T A^{-1} U$ é não singular. Além disso, sob a equivalência anteriormente mencionada, tem-se

$$\hat{A}^{-1} = A^{-1} - A^{-1} U (I_n + V^T A^{-1} U)^{-1} V^T A^{-1}.$$

Agora, reescrevendo (3-2) como

$$H_j^{k+1} = H_j^k + \begin{bmatrix} H_j^k y_j^k & s_k \end{bmatrix} \begin{bmatrix} 0 & -\rho_j^k \\ -\rho_j^k & (\rho_j^k)^2 (y_j^k)^T H_j^k y_j^k + \rho_j^k \end{bmatrix} \begin{bmatrix} (y_j^k)^T H_j^k \\ s_k^T \end{bmatrix},$$

uma vez que H_j^{k+1} é não singular e definindo

$$A := H_j^k, \quad U := \begin{bmatrix} H_j^k y_j^k & s_k \end{bmatrix}$$

e

$$V^T := \begin{bmatrix} 0 & -\rho_j^k \\ -\rho_j^k & (\rho_j^k)^2 (y_j^k)^T H_j^k y_j^k + \rho_j^k \end{bmatrix} \begin{bmatrix} (y_j^k)^T H_j^k \\ s_k^T \end{bmatrix},$$

temos que a matriz $I_n + V^T B_j^k U$ é não singular. Desse modo,

$$(H_j^{k+1})^{-1} = B_j^k - B_j^k U \left(I_n + V^T B_j^k U \right)^{-1} V^T B_j^k.$$

Agora, para mostrar que a expressão acima pode ser escrita como em (3-1), note que

$$\begin{aligned} & \left(I_n + V^T B_j^k U \right)^{-1} V^T B_j^k \\ &= \left(I_n + \begin{bmatrix} 0 & -\rho_j^k \\ -\rho_j^k & (\rho_j^k)^2 (y_j^k)^T H_j^k y_j^k + \rho_j^k \end{bmatrix} \begin{bmatrix} (y_j^k)^T H_j^k \\ s_k^T \end{bmatrix} B_j^k \begin{bmatrix} H_j^k y_j^k & s_k \end{bmatrix} \right)^{-1} V^T B_j^k \\ &= \left(\begin{bmatrix} 0 & -\rho_j^k \\ -\rho_j^k & (\rho_j^k)^2 (y_j^k)^T H_j^k y_j^k + \rho_j^k \end{bmatrix}^{-1} + \begin{bmatrix} (y_j^k)^T H_j^k \\ s_k^T \end{bmatrix} \begin{bmatrix} y_j^k & B_j^k s_k \end{bmatrix} \right)^{-1} \begin{bmatrix} (y_j^k)^T \\ s_k^T B_j^k \end{bmatrix}. \end{aligned}$$

Por outro lado, definindo

$$Z := \left(\begin{bmatrix} 0 & -\rho_j^k \\ -\rho_j^k & (\rho_j^k)^2 (y_j^k)^T H_j^k y_j^k + \rho_j^k \end{bmatrix}^{-1} + \begin{bmatrix} (y_j^k)^T H_j^k \\ s_k^T \end{bmatrix} \begin{bmatrix} y_j^k & B_j^k s_k \end{bmatrix} \right),$$

obtemos

$$\begin{aligned} Z^{-1} &= \left(\begin{bmatrix} (\rho_j^k)^2 (y_j^k)^T H_j^k y_j^k + \rho_j^k & \rho_j^k \\ \rho_j^k & 0 \end{bmatrix} \frac{-1}{(\rho_j^k)^2} + \begin{bmatrix} (y_j^k)^T H_j^k \\ s_k^T \end{bmatrix} \begin{bmatrix} y_j^k & B_j^k s_k \end{bmatrix} \right)^{-1} \\ &= \left(\begin{bmatrix} (\rho_j^k)^2 (y_j^k)^T H_j^k y_j^k + \rho_j^k & \rho_j^k \\ \rho_j^k & 0 \end{bmatrix} \frac{-1}{(\rho_j^k)^2} + \begin{bmatrix} (y_j^k)^T H_j^k y_j^k & (y_j^k)^T s_k \\ s_k^T y_j^k & s_k^T B_j^k s_k \end{bmatrix} \right)^{-1} \\ &= \begin{bmatrix} -(\rho_j^k)^{-1} & (y_j^k)^T s_k - (\rho_j^k)^{-1} \\ s_k^T y_j^k - (\rho_j^k)^{-1} & s_k^T B_j^k s_k \end{bmatrix}^{-1} \\ &= \begin{bmatrix} s_k^T B_j^k s_k & (\rho_j^k)^{-1} - (y_j^k)^T s_k \\ (\rho_j^k)^{-1} - s_k^T y_j^k & -(\rho_j^k)^{-1} \end{bmatrix} \frac{-1}{\left((\rho_j^k)^{-1} - s_k^T y_j^k \right)^2 + (\rho_j^k)^{-1} s_k^T B_j^k s_k} \\ &= \begin{bmatrix} s_k^T B_j^k s_k & (\rho_j^k)^{-1} - (y_j^k)^T s_k \\ (\rho_j^k)^{-1} - s_k^T y_j^k & -(\rho_j^k)^{-1} \end{bmatrix} \frac{1}{\text{Det } Z}. \end{aligned}$$

Portanto,

$$\begin{aligned}
\left(H_j^{k+1}\right)^{-1} &= B_j^k - B_j^k U \left(I_n + V^T B_j^k U\right)^{-1} V^T B_j^k \\
&= B_j^k - B_j^k U Z^{-1} \begin{bmatrix} (y_j^k)^T \\ s_k^T B_j^k \end{bmatrix} \\
&= B_j^k - B_j^k U \begin{bmatrix} s_k^T B_j^k s_k & (\rho_j^k)^{-1} - (y_j^k)^T s_k \\ (\rho_j^k)^{-1} - s_k^T y_j^k & -(\rho_j^k)^{-1} \end{bmatrix} \begin{bmatrix} (y_j^k)^T \\ s_k^T B_j^k \end{bmatrix} \frac{1}{\text{Det } Z} \\
&= B_j^k - B_j^k U \begin{bmatrix} s_k^T B_j^k s_k (y_j^k)^T + (\rho_j^k)^{-1} s_k^T B_j^k - (y_j^k)^T s_k s_k^T B_j^k \\ (\rho_j^k)^{-1} (y_j^k)^T - s_k^T y_j^k (y_j^k)^T - (\rho_j^k)^{-1} s_k^T B_j^k \end{bmatrix} \frac{1}{\text{Det } Z} \\
&= B_j^k - \begin{bmatrix} y_j^k & B_j^k s_k \end{bmatrix} \begin{bmatrix} s_k^T B_j^k s_k (y_j^k)^T + (\rho_j^k)^{-1} s_k^T B_j^k - (y_j^k)^T s_k s_k^T B_j^k \\ (\rho_j^k)^{-1} (y_j^k)^T - s_k^T y_j^k (y_j^k)^T - (\rho_j^k)^{-1} s_k^T B_j^k \end{bmatrix} \frac{1}{\text{Det } Z} \\
&= B_j^k - \frac{1}{\text{Det } Z} \left(s_k^T B_j^k s_k y_j^k (y_j^k)^T + (\rho_j^k)^{-1} y_j^k s_k^T B_j^k - y_j^k (y_j^k)^T s_k s_k^T B_j^k + \right. \\
&\quad \left. + (\rho_j^k)^{-1} B_j^k s_k (y_j^k)^T - B_j^k s_k s_k^T y_j^k (y_j^k)^T - (\rho_j^k)^{-1} B_j^k s_k s_k^T B_j^k \right),
\end{aligned}$$

obtendo a expressão em (3-1).

3.1 Uma extensão que mimetiza o método clássico

Nesta seção, descrevemos detalhadamente um algoritmo para otimização multiobjetivo que mimetiza o método BFGS escalar. Os resultados desta seção estão publicados em [64].

Algoritmo 3.1. Um algoritmo BFGS com busca linear de Wolfe

Sejam $c_1 \in (0, 1/2)$, $c_2 \in (c_1, 1)$, $x^0 \in \mathbb{R}^n$, e $B_j^0 \succ 0$, para todo $j = 1, \dots, m$.
Inicialize $k \leftarrow 0$.

Passo 1. *Calcule a direção de busca*

Calcule $d^k := d(x^k)$ e $\theta(x^k)$ como em (2-18) e (2-19), respectivamente.

Passo 2. *Critério de parada*

Se $\theta(x^k) = 0$, então PARE.

Passo 3. *Procedimento de busca linear*

Calcule o comprimento de passo $\alpha_k > 0$ (primeiro tente $\alpha_k = 1$) tal que as condições de Wolfe padrão (1-12)–(1-13) sejam satisfeitas e faça $x^{k+1} := x^k + \alpha_k d^k$.

Passo 4. *Prepare a próxima iteração*

Para cada $j = 1, \dots, m$, defina B_j^{k+1} como em (3-1), em que $s_k = x^{k+1} - x^k$, $y_j^k := \nabla F_j(x^{k+1}) - \nabla F_j(x^k)$ e

$$\rho_j^k := \begin{cases} 1/s_k^T y_j^k, & \text{se } s_k^T y_j^k > 0, \\ 1/(\mathcal{D}(x^{k+1}, s_k) - \nabla F_j(x^k)^T s_k), & \text{caso contrário.} \end{cases} \quad (3-3)$$

Faça $k \leftarrow k + 1$ e vá ao **Passo 1**.

Observação 3.1. (i) Na prática, no Passo 1, d^k e $\theta(x^k)$ são calculados resolvendo o problema quadrático convexo de otimização escalar (2-20). (ii) No Passo 2, pelo Lema 2.1(i), o Algoritmo 3.1 para em uma iteração k se, e somente se, x^k é Pareto crítico. (iii) No Passo 3, o comprimento de passo α_k deve satisfazer as condições de Wolfe padrão (1-12)–(1-13), as quais foram cuidadosamente discutidas na Seção 1.3. Na busca linear, sempre testamos primeiro se $\alpha_k = 1$ satisfaz essas condições, sendo crucial na obtenção de uma rápida taxa de convergência. (iv) No Passo 4, ρ_j^k é definido de modo a sempre se manter positivo, veja o Teorema 3.1 a seguir. Além disso, se $\rho_j^k = 1/s_k^T y_j^k$, então (3-1) recupera a atualização BFGS clássica (2-24) para F_j . Isso ocorre, por exemplo, quando F_j é estritamente convexa. Além disso, no caso escalar ($m = 1$), a condição de curvatura (1-13) recupera a condição clássica (1-10) e implica que $s_k^T y_1^k > 0$, logo $\rho_1^k = 1/s_k^T y_1^k$ e o Algoritmo 3.1 se torna o algoritmo BFGS clássico com busca linear de Wolfe padrão (Algoritmo 2.1).

O próximo teorema mostra que o Algoritmo 3.1 está bem definido sem impor qualquer hipótese de convexidade sobre a função objetivo. Essa prova consiste basicamente em mostrar que B_j^{k+1} será definida positiva sempre que B_j^k for definida positiva, para cada $j = 1, \dots, m$.

Teorema 3.1. *Suponha que F é limitada inferiormente em $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n \mid F(x) \leq F(x^0)\}$. Então, o Algoritmo 3.1 está bem definido.*

Demonstração. Por indução, suponha que B_j^k é definida positiva para todo $j = 1, \dots, m$ (que trivialmente é válido para $k = 0$). Portanto, o subproblema no Passo 1 tem solução. Se x^k é Pareto crítico, então pelo Lema 2.1(i), o Algoritmo 3.1 finaliza no Passo 2. Logo, assumiremos que x^k não é Pareto crítico. Nesse caso, o Lema 2.1(ii) implica que d^k é uma direção de descida de F em x^k . Como F é limitado inferiormente em $\mathcal{L}(x^0)$, pela Proposição 1.1, existem intervalos de tamanho de passo positivo satisfazendo as condições (1-12)–(1-13) no Passo 3 e assim x^{k+1} pode ser definido apropriadamente. Agora, mostraremos que B_j^{k+1} em (3-1) ou, equivalentemente, H_j^{k+1} em (3-2), se mantém definida positiva para todo $j = 1, \dots, m$. Como $s_k = \alpha_k d^k$, pela definição de \mathcal{D} , Lema 1.1(i) e (1-13), temos

$$\begin{aligned} \mathcal{D}(x^{k+1}, s_k) - \nabla F_j(x^k)^T s_k &\geq \alpha_k [\mathcal{D}(x^{k+1}, d^k) - \mathcal{D}(x^k, d^k)] \\ &\geq -\alpha_k (1 - c_2) \mathcal{D}(x^k, d^k) > 0, \end{aligned}$$

e assim ρ_j^k definido em (3-3) é positivo para todo $j = 1, \dots, m$. Seja $j \in \{1, \dots, m\}$ e $0 \neq z \in \mathbb{R}^n$. Então, por (3-2),

$$z^T H_j^{k+1} z = \left(z - \rho_j^k z^T s_k y_j^k \right)^T H_j^k \left(z - \rho_j^k z^T s_k y_j^k \right) + \rho_j^k (z^T s_k)^2 > 0,$$

Portanto, H_j^{k+1} é definida positiva para cada $j = 1, \dots, m$. ■

O seguinte exemplo ilustra que, em otimização multicritério, em contrapartida ao caso escalar, $s_k^T y_j^k$ pode ser não positivo para algum $j \in \{1, \dots, m\}$ mesmo quando uma busca linear de Wolfe é usada. Isso justifica nossa escolha para ρ_j^k em (3-3), pois uma condição chave para a boa definição do Algoritmo 3.1 é que $\rho_j^k > 0$ para todo $j = 1, \dots, m$ (veja o Teorema 3.1).

Exemplo 3.1. *Sejam $\beta \geq 1$, $c_1 = 10^{-4}$, $c_2 = 0.9$, $x^0 = 0$, e $B_1^0 = B_2^0 = 1$ dados e considere a aplicação do Algoritmo 3.1 para o problema (1-1) com $F: \mathbb{R} \rightarrow \mathbb{R}^2$*

definida por

$$F_1(x) := \frac{x^2}{3} - x \quad e \quad F_2(x) := \begin{cases} -x, & \text{se } x < 0 \\ (1 - \beta)x^3 + (\beta - 1)x^2 - x, & \text{se } 0 \leq x < 1, \\ -\beta x + \beta - 1, & \text{se } 1 \leq x < 2, \\ \beta x^2 - 5\beta x + 5\beta - 1, & \text{se } x \geq 2. \end{cases}$$

Note que F é continuamente diferenciável e limitada inferiormente em \mathbb{R} . Um cálculo direto mostra que $d^0 = 1$ e $\alpha_0 = 1$ satisfazem as condições Wolfe padrão (1-12)–(1-13), implicando que $x^1 = 1$. Então, temos que $s_0^T y_1^0 = 2/3$ e $s_0^T y_2^0 = 1 - \beta \leq 0$. Se tomarmos $\rho_j^k = 1/s_k^T y_j^k$ em (3-1), o algoritmo falha ao tentar dividir por zero, no caso quando $\beta = 1$. Para $\beta > 1$, teremos também $B_2^1 = 1 - \beta < 0$.

3.1.1 Análise de convergência

Nesta seção, apresentamos os resultados de convergência para o Algoritmo 3.1. Como é usual no caso escalar, veja Hipótese 2.1, assumimos que as funções objetivo são fortemente convexas, conforme formalmente estabelecido a seguir.

Hipótese 3.1. (i) F é duas vezes continuamente diferenciável. (ii) O conjunto de nível $\mathcal{L}(x^0)$ é convexo e existem constantes $\mu, L > 0$ tais que

$$\mu \|z\|^2 \leq z^T \nabla^2 F_j(x) z \leq L \|z\|^2, \quad \forall j = 1, \dots, m, \quad (3-4)$$

para todo $z \in \mathbb{R}^n$ e $x \in \mathcal{L}(x^0)$.

Note que, sob a Hipótese 3.1, $s_k^T y_j^k > 0$ e assim $\rho_j^k = 1/s_k^T y_j^k$ para todo $k \geq 0$ e $j = 1, \dots, m$. Nesse caso, (3-1) sempre recupera a atualização BFGS clássica (2-24). Além disso, as hipóteses do Teorema 3.1 e da Proposição 1.2 são trivialmente satisfeitas. Por fim, como consequência dessa suposição, as hipóteses do Teorema 2.2 também são satisfeitas. De fato, defina

$$\bar{G}_j^k := \int_0^1 \nabla^2 F_j(x^k + \tau s_k) d\tau,$$

e perceba que

$$\bar{G}_j^k s_k = y_j^k, \quad \forall j = 1, \dots, m. \quad (3-5)$$

Assim,

$$\mu \|z\|^2 \leq z^T \bar{G}_j^k z \leq L \|z\|^2 \quad \forall j = 1, \dots, m, \quad (3-6)$$

para todo $z \in \mathbb{R}^n$ e $k \geq 0$. Portanto,

$$\frac{s_k^T y_j^k}{\|s_k\|^2} = \frac{s_k^T \bar{G}_j^k s_k}{\|s_k\|^2} \geq \mu \quad \text{e} \quad \frac{(y_j^k)^T y_j^k}{s_k^T y_j^k} = \frac{s_k^T \bar{G}_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} = \frac{z_k^T \bar{G}_j^k z_k}{z_k^T z_k} \leq L, \quad (3-7)$$

em que $z_k = (\bar{G}_j^k)^{\frac{1}{2}} s_k$. Agora, sob essas hipóteses, estamos prontos para provar o principal resultado dessa seção.

Teorema 3.2. *Suponha que a Hipótese 3.1 vale e que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.1. Então, $\{x^k\}$ converge para um ponto Pareto ótimo x^* de F .*

Demonstração. Pelo Teorema 2.2, existe uma constante $\delta > 0$ e $K \subset \mathbb{N}$ tal que

$$\cos \beta_j^k \geq \delta, \quad \forall k \in K \quad \text{e} \quad \forall j = 1, \dots, m.$$

Logo, a Proposição 2.1 implica que

$$\mathcal{D}(x^k, d^k) \leq -\frac{\delta}{2} \|d^k\| \|d_{SD}(x^k)\|, \quad \forall k \in K.$$

Assim, pela Proposição 1.2, temos

$$\infty > \sum_{k \geq 0} \frac{\mathcal{D}(x^k, d^k)^2}{\|d^k\|^2} \geq \sum_{k \in K} \frac{\mathcal{D}(x^k, d^k)^2}{\|d^k\|^2} \geq \sum_{k \in K} \frac{\delta^4}{4} \|d_{SD}(x^k)\|^2,$$

e conseqüentemente

$$\lim_{k \in K} d_{SD}(x^k) = 0. \quad (3-8)$$

Note que, como $\mathcal{L}(x^0)$ é compacto e $x^k \in \mathcal{L}(x^0)$ para todo $k \in K$, existem $K_1 \subset \mathbb{N}$ e $x^* \in \mathcal{L}(x^0)$ tais que $\lim_{k \in K_1} x^k = x^*$. Desse modo, por (3-8) e pelo Lema 1.2(iii), obtemos $d_{SD}(x^*) = 0$. Assim, pelo Teorema 1.1(iii), concluímos que x^* é um Pareto ótimo.

Agora, vamos mostrar que $\lim_{k \rightarrow \infty} x^k = x^*$. Para isso, suponha por contra-dição que existam $\bar{x} \in \mathcal{L}(x^0)$ com $\bar{x} \neq x^*$ e $K_2 \subset \mathbb{N}$ tais que $\lim_{k \in K_2} x^k = \bar{x}$. Primeiro, afirmamos que $F(\bar{x}) \neq F(x^*)$. De fato, se $F(\bar{x}) = F(x^*)$, pela Hipótese 3.1, para todo $t \in [0, 1]$, temos

$$F_j(t\bar{x} + (1-t)x^*) < tF_j(\bar{x}) + (1-t)F_j(x^*) = F_j(x^*), \quad \forall j = 1, \dots, m,$$

contradizendo o fato de que x^* é ponto Pareto ótimo. Desse modo, $F(\bar{x}) \neq F(x^*)$ vale, como afirmado. Agora, como x^* é Pareto ótimo, existe $j_0 \in \{1, \dots, m\}$ tal que $F_{j_0}(x^*) < F_{j_0}(\bar{x})$. Assim, lembrando que $\lim_{k \in K_1} x^k = x^*$ e $\lim_{k \in K_2} x^k = \bar{x}$, podemos escolher $k_1 \in K_1$ e $k_2 \in K_2$ com $k_1 < k_2$ em que $F_{j_0}(x^{k_1}) < F_{j_0}(x^{k_2})$. Isso contradiz

a condição de decréscimo suficiente (1-12) que, em particular, implica que $\{F_j(x^k)\}$ é uma sequência decrescente para todo $j = 1, \dots, m$. Portanto, $\lim_{k \rightarrow \infty} x^k = x^*$, como queríamos demonstrar. ■

No restante dessa seção, nosso objetivo é mostrar que $\{x^k\}$ converge para x^* suficientemente rápido de modo que

$$\sum_{k \geq 0} \|x^k - x^*\|^\nu < \infty, \quad \forall \nu > 0. \quad (3-9)$$

Até onde sabemos, este é o primeiro trabalho a estabelecer esse resultado para otimização multiobjetivo. No caso escalar, variantes desse resultado podem ser encontradas em, [62, Lema 5], [6, Teorema 3.1] e [46, Lema 3.6]. Como discutido na Seção 2.1, (3-9) desempenha um papel importante na convergência superlinear. Para estabelecer essa condição, apresentamos alguns resultados técnicos.

Lema 3.1. *Suponha que a Hipótese 3.1 vale e que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.1. Seja x^* como no Teorema 3.2. Então, para todo $k \geq 0$, temos*

- (i) $\|x^k - x^*\| \leq \frac{2}{\mu} \|d_{SD}(x^k)\|;$
- (ii) $\|s_k\| \geq \frac{(1 - c_2)}{2L} \delta_k \|d_{SD}(x^k)\|$, em que δ_k é dado como na Proposição 2.1 e c_2 é dado como na Definição 1.1.

Demonstração. Seja $k \geq 0$ dado e defina a função de valor escalar $F_{SD}: \mathbb{R}^n \rightarrow \mathbb{R}$ por

$$F_{SD}(x) := \sum_{j=1}^m \lambda_j^{SD}(x^k) F_j(x),$$

em que $\lambda^{SD}(x^k)$ e $d_{SD}(x^k)$ são definidos em (1-6)–(1-7). Desse modo, por (1-6) e (3-4), é fácil ver que

$$\int_0^1 (1 - \tau) z^T \nabla^2 F_{SD}(x^k + \tau z) z d\tau \geq \frac{\mu}{2} \|z\|^2,$$

em que $z := x^* - x^k$. Considerando que $d_{SD}(x^k) = -\nabla F_{SD}(x^k)$ e avaliando a integral acima por partes ($\int u dv = uv - \int v du$ com $u = 1 - \tau$ e $dv = z^T \nabla^2 F_{SD}(x^k + \tau z) z d\tau$), obtemos

$$\begin{aligned}
\frac{\mu}{2}\|z\|^2 &\leq \int_0^1 (1-\tau)z^T \nabla^2 F_{SD}(x^k + \tau z)z d\tau \\
&= (1-\tau)\langle z, \nabla F_{SD}(x^k + \tau z) \rangle \Big|_0^1 - \int_0^1 \langle z, \nabla F_{SD}(x^k + \tau z) \rangle (-1) d\tau \\
&= -\langle z, \nabla F_{SD}(x^k) \rangle + \int_0^1 \langle z, \nabla F_{SD}(x^k + \tau z) \rangle d\tau \\
&= F_{SD}(x^k + \tau z) \Big|_0^1 - \langle z, \nabla F_{SD}(x^k) \rangle \\
&= F_{SD}(x^*) - F_{SD}(x^k) + d_{SD}(x^k)^T(x^* - x^k).
\end{aligned}$$

Como $F_j(x^*) \leq F_j(x^k)$ para todo $j = 1, \dots, m$, temos que $F_{SD}(x^*) - F_{SD}(x^k) \leq 0$ e consequentemente

$$\frac{\mu}{2}\|x^* - x^k\|^2 \leq d_{SD}(x^k)^T(x^* - x^k) \leq \|d_{SD}(x^k)\| \|x^* - x^k\|,$$

mostrando a parte (i).

Vamos considerar agora a parte (ii). Perceba que, por (1-13) e pela definição de \mathcal{D} , segue que

$$\begin{aligned}
(c_2 - 1)\mathcal{D}(x^k, d^k) &\leq \mathcal{D}(x^{k+1}, d^k) - \mathcal{D}(x^k, d^k) \\
&\leq \max_{j=1, \dots, m} (\nabla F_j(x^{k+1}) - \nabla F_j(x^k))^T d^k = \max_{j=1, \dots, m} (y_j^k)^T d^k \\
&\stackrel{(3-5)}{=} \max_{j=1, \dots, m} s_k^T \bar{G}_j^k d^k,
\end{aligned}$$

em que a segunda desigualdade segue pelo fato de que, para qualquer $u, v \in \mathbb{R}^m$, temos $\max_{j=1, \dots, m} \{u_j - v_j\} \geq \max_{j=1, \dots, m} \{u_j\} - \max_{j=1, \dots, m} \{v_j\}$. Logo,

$$(c_2 - 1)\mathcal{D}(x^k, d^k) \leq \alpha_k \max_{j=1, \dots, m} (d^k)^T \bar{G}_j^k d^k \stackrel{(3-6)}{\leq} L\alpha_k \|d^k\|^2 = L\|s_k\| \|d^k\|,$$

Portanto, usando a Proposição 2.1 e levando em conta que $c_2 < 1$, temos

$$-(c_2 - 1) \frac{\delta_k}{2} \|d^k\| \|d_{SD}(x^k)\| \leq L\|s_k\| \|d^k\|,$$

concluindo a prova. ■

Teorema 3.3. *Suponha que a Hipótese 3.1 vale e que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.1. Seja x^* como no Teorema 3.2. Então, $\{x^k\}$ converge*

R -linearmente para x^* e, conseqüentemente, (3-9) vale.

Demonstração. Seja $\lambda^{SD}(x^*) \in \mathbb{R}^m$ um multiplicador da direção de máxima descida generalizada associado a x^* como em (1-6)–(1-7), isto é, $d_{SD}(x^*) = -\sum_{j=1}^m \lambda_j^{SD}(x^*) \nabla F_j(x^*)$. Vamos definir a função de valor escalar $F_*: \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$F_*(x) := \sum_{j=1}^m \lambda_j^{SD}(x^*) F_j(x).$$

Note que, pelo Lema 1.2(i), temos

$$\nabla F_*(x^*) = \sum_{j=1}^m \lambda_j^{SD}(x^*) \nabla F_j(x^*) = -d_{SD}(x^*) = 0. \quad (3-10)$$

Agora, fazendo uma expansão de segunda ordem da série de Taylor da função F_j em torno de x^* , temos que existe $\tau_j \in (0, 1)$ tal que

$$F_j(x^k) = F_j(x^*) + \nabla F_j(x^*)^T (x^k - x^*) + \frac{1}{2} (x^k - x^*)^T \nabla^2 F_j(x^* + \tau_j(x^k - x^*)) (x^k - x^*).$$

Logo, pela hipótese de convexidade forte (3-4), temos

$$\begin{aligned} \frac{\mu}{2} \|x^k - x^*\|^2 + \nabla F_j(x^*)^T (x^k - x^*) &\leq F_j(x^k) - F_j(x^*) \\ &= \frac{1}{2} (x^k - x^*)^T \nabla^2 F_j(x^* + \tau_j(x^k - x^*)) (x^k - x^*) \\ &\quad + \nabla F_j(x^*)^T (x^k - x^*) \\ &\leq \frac{L}{2} \|x^k - x^*\|^2 + \nabla F_j(x^*)^T (x^k - x^*), \end{aligned}$$

para todo $j = 1, \dots, m$ e $k \geq 0$. Assim, multiplicando essa expressão por $\lambda_j^{SD}(x^*)$, somando todos os índices $j = 1, \dots, m$, e considerando (1-6) e (3-10), obtemos

$$\frac{\mu}{2} \|x^k - x^*\|^2 \leq F_*(x^k) - F_*(x^*) \leq \frac{L}{2} \|x^k - x^*\|^2, \quad \forall k \geq 0. \quad (3-11)$$

Pelo lado direito de (3-11) e pelo Lema 3.1(i), obtemos

$$F_*(x^k) - F_*(x^*) \leq \frac{2L}{\mu^2} \|d_{SD}(x^k)\|^2, \quad \forall k \geq 0. \quad (3-12)$$

Por outro lado, similar a (3-11), a condição de decréscimo suficiente (1-12) implica

$$F_*(x^{k+1}) \leq F_*(x^k) + c_1 \alpha_k \mathcal{D}(x^k, d^k), \quad \forall k \geq 0.$$

Subtraindo o termo $F_*(x^*)$ em ambos os lados dessa desigualdade e usando a

Proposição 2.1, temos

$$F_*(x^{k+1}) - F_*(x^*) \leq F_*(x^k) - F_*(x^*) - \frac{c_1}{2} \delta_k \|s_k\| \|d_{SD}(x^k)\|,$$

para todo $k \geq 0$, em que $\delta_k = \min_{j=1, \dots, m} \cos \beta_j^k$. Desse modo, pelo Lema 3.1(ii), segue que

$$\begin{aligned} F_*(x^{k+1}) - F_*(x^*) &\leq F_*(x^k) - F_*(x^*) - \frac{c_1(1-c_2)}{4L} \delta_k^2 \|d_{SD}(x^k)\|^2 \\ &\stackrel{(3-12)}{\leq} \left(1 - \frac{c_1(1-c_2)\mu^2}{8L^2} \delta_k^2\right) (F_*(x^k) - F_*(x^*)), \quad \forall k \geq 0. \end{aligned} \quad (3-13)$$

Para cada $k \geq 0$, defina $\bar{r}_k := 1 - c_1(1-c_2)\mu^2\delta_k^2/(8L^2)$. É fácil ver que $\bar{r}_k \in (0, 1]$, para todo $k \geq 0$.

Agora, dado $p \in (0, 1)$, o Teorema 2.2, garante que existe uma constante $\delta > 0$ tal que, para qualquer $k \geq 1$, o número de elementos $\ell \in \{0, 1, \dots, k\}$ para os quais $\delta_\ell \geq \delta$ é pelo menos $\lceil p(k+1) \rceil$. Desse modo, definindo $\mathcal{G}_k := \{\ell \in \{0, 1, \dots, k\} \mid \delta_\ell \geq \delta\}$, temos $|\mathcal{G}_k| \geq \lceil p(k+1) \rceil$ e

$$\bar{r}_\ell \leq 1 - \frac{c_1(1-c_2)\mu^2\delta^2}{8L^2} := \bar{r} < 1, \quad \forall \ell \in \mathcal{G}_k.$$

Portanto, por (3-13) e considerando que $F_*(x^0) - F_*(x^*) > 0$, obtemos, para todo $k \geq 1$,

$$\begin{aligned} F_*(x^{k+1}) - F_*(x^*) &\leq \left[\prod_{\ell=0}^k \bar{r}_\ell \right] (F_*(x^0) - F_*(x^*)) \leq \left[\prod_{\ell \in \mathcal{G}_k} \bar{r}_\ell \right] (F_*(x^0) - F_*(x^*)) \\ &\leq \left[\prod_{\ell \in \mathcal{G}_k} \bar{r} \right] (F_*(x^0) - F_*(x^*)) \leq \bar{r}^{\lceil p(k+1) \rceil} (F_*(x^0) - F_*(x^*)), \end{aligned}$$

na qual a segunda desigualdade segue do fato de que $\bar{r}_\ell \leq 1$ para todo $\ell \notin \mathcal{G}_k$. Portanto, definindo $r^{1/p} := \bar{r}$, obtemos

$$F_*(x^{k+1}) - F_*(x^*) \leq r^{k+1} (F_*(x^0) - F_*(x^*)), \quad \forall k \geq 1. \quad (3-14)$$

Combinando o lado esquerdo de (3-11) com a desigualdade acima, encontramos

$$\|x^{k+1} - x^*\| \leq \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*)) \right]^{1/2} (r^{1/2})^{k+1},$$

e assim $\{x^k\}$ converge R-linearmente para x^* . Finalmente, ao somar essa expressão

em k e considerando que $r < 1$ e $\nu > 0$, obtemos

$$\sum_{k \geq 0} \|x^{k+1} - x^*\|^\nu \leq \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*)) \right]^{\frac{\nu}{2}} \sum_{k \geq 0} \left(r^{\frac{\nu}{2}} \right)^{k+1} < \infty.$$

■

Observação 3.2. Por (3-14) temos que $\{F_*(x^k)\}$ converge R-linearmente para $F_*(x^*)$.

3.1.2 Convergência local superlinear

Nesta seção estudaremos as propriedades de convergência local do Algoritmo 3.1. Os resultados aqui apresentados também se aplicam a problemas não convexos, embora não seja possível estabelecer convergência global neste caso geral. Assumiremos que $\{x^k\}$ converge para um ponto ótimo Pareto local x^* e provaremos, sob hipóteses razoáveis, que a taxa de convergência é Q-superlinear.

Hipótese 3.2. (i) F é duas vezes continuamente diferenciável. (ii) A sequência $\{x^k\}$ gerada pelo Algoritmo 3.1 converge para um ponto Pareto ótimo local x^* . (iii) Para cada $j = 1, \dots, m$, $\nabla^2 F_j(x^*)$ é definida positiva e Lipschitz contínua em x^* . Portanto, existe uma vizinhança U de x^* e existem constantes positivas μ , L , e L_2 tais que

$$\mu \|z\|^2 \leq z^T \nabla^2 F_j(x) z \leq L \|z\|^2, \quad (3-15)$$

e

$$\|\nabla^2 F_j(x) - \nabla^2 F_j(x^*)\| \leq L_2 \|x - x^*\|, \quad (3-16)$$

para todo $j = 1, \dots, m$, $z \in \mathbb{R}^n$ e $x \in U$.

Essencialmente, a Hipótese 3.2 estabelece que, em uma vizinhança U de x^* , F é fortemente convexa e as Hessianas $\nabla^2 F_j$ ($j = 1, \dots, m$) são Lipschitz contínuas em x^* . Ao longo desta seção, assumiremos, sem perda de generalidade, que $\{x^k\} \subset U$, isto é, (3-15) e (3-16) valem em x^k para todo $k \geq 0$. Como a Hipótese 3.2 é mais restritiva que a Hipótese 3.1, os resultados da seção anterior permanecem verdadeiros e serão usados aqui sem maiores explicações.

Observação 3.3. Destacamos aqui que a Lipschitz continuidade da Hessiana na Hipótese 3.2, pode ser substituída por uma Hölder continuidade (veja a Hipótese 3.4 na próxima seção). Essa mudança, dependendo do valor do expoente ν , pode enfraquecer a Hipótese 3.2. Contudo, o leitor pode notar que mesmo havendo um

possível enfraquecimento dessa hipótese, a validade dos resultados discutidos ao longo da presente seção não são afetados. Além disso, utilizamos a condição (3-16) na Hipótese 3.2, pois estamos propondo um método que mimetiza o BFGS clássico (veja a Hipótese 2.2).

O próximo teorema estabelece que a condição de Dennis–Moré (2-15) vale individualmente para cada função objetivo F_j (veja (3-18) abaixo). Apesar da prova deste resultado ser semelhante ao caso escalar, para fins de completude do trabalho, iremos incluí-la aqui. Vale ressaltar que, assim como no caso escalar, (3-9) é essencial para que o limite de Dennis–Moré seja válido, veja Seção 2.1. Também incluímos no enunciado do teorema uma etapa intermediário (veja (3-17) abaixo) que será evocado nos próximos resultados.

Teorema 3.4. *Suponha que a Hipótese 3.2 é válida. Então,*

$$\lim_{k \rightarrow \infty} \frac{s_k^T B_j^k s_k}{s_k^T \nabla^2 F_j(x^*) s_k} = 1 \quad \forall j = 1, \dots, m, \quad (3-17)$$

e

$$\lim_{k \rightarrow \infty} \frac{\|(B_j^k - \nabla^2 F_j(x^*))d^k\|}{\|d^k\|} = 0, \quad \forall j = 1, \dots, m, \quad (3-18)$$

ou, de modo equivalente a (3-18),

$$\lim_{k \rightarrow \infty} \frac{\|(B_j^k - \nabla^2 F_j(x^k))d^k\|}{\|d^k\|} = 0, \quad \forall j = 1, \dots, m. \quad (3-19)$$

Além disso, as sequências $\{\|B_j^k\|\}$ e $\{\|H_j^k\|\}$ são limitadas para todo $j = 1, \dots, m$.

Demonstração. Inicialmente, para cada $j \in 1, \dots, m$ e $k \geq 0$, definimos

$$\begin{aligned} \tilde{s}_k &:= \nabla^2 F_j(x^*)^{1/2} s_k, & \tilde{y}_k &:= \nabla^2 F_j(x^*)^{-1/2} y_j^k, & \tilde{a}_k &:= \frac{\tilde{s}_k^T \tilde{y}_k}{\|\tilde{s}_k\|^2}, & \tilde{b}_k &:= \frac{\|\tilde{y}_k\|^2}{\tilde{s}_k^T \tilde{y}_k}, \\ \tilde{B}^k &:= \nabla^2 F_j(x^*)^{-1/2} B_j^k \nabla^2 F_j(x^*)^{-1/2}, & \cos \tilde{\beta}_k &:= \frac{\tilde{s}_k^T \tilde{B}^k \tilde{s}_k}{\|\tilde{s}_k\| \|\tilde{B}^k \tilde{s}_k\|}, & \text{e} & & \tilde{q}_k &:= \frac{\tilde{s}_k^T \tilde{B}^k \tilde{s}_k}{\|\tilde{s}_k\|^2}. \end{aligned}$$

Note que a Hipótese 3.2 garante que $\tilde{s}_k^T \tilde{y}_k = s_k^T y_j^k > 0$ para todo $j = 1, \dots, m$ e $k \geq 0$, logo \tilde{b}_k está bem definido. Agora, dado um j fixado, porém arbitrário, a condição (3-5) garante que

$$y_j^k - \nabla^2 F_j(x^*) s_k = (\bar{G}_j^k - \nabla^2 F_j(x^*)) s_k.$$

Por simplicidade, denote $\bar{G}_k := \bar{G}_j^k$, assim temos

$$\tilde{y}_k - \tilde{s}_k = \nabla^2 F_j(x^*)^{-1/2} (\bar{G}_k - \nabla^2 F_j(x^*)) \nabla^2 F_j(x^*)^{-1/2} \tilde{s}_k.$$

Logo, por Cauchy–Schwarz, obtemos

$$\begin{aligned}
\|\tilde{y}_k - \tilde{s}_k\| &\leq \|\nabla^2 F_j(x^*)^{-1/2}\|^2 \|\bar{G}_k - \nabla^2 F_j(x^*)\| \|\tilde{s}_k\| \\
&\leq \|\nabla^2 F_j(x^*)^{-1/2}\|^2 \|\tilde{s}_k\| \int_0^1 \|\nabla^2 F_j(x^k + \tau\alpha_k d^k) - \nabla^2 F_j(x^*)\| d\tau \\
&\stackrel{(3-16)}{\leq} \|\nabla^2 F_j(x^*)^{-1/2}\|^2 \|\tilde{s}_k\| \int_0^1 L_2 \|x^k + \tau\alpha_k d^k - x^*\| d\tau \\
&\leq \|\nabla^2 F_j(x^*)^{-1/2}\|^2 \|\tilde{s}_k\| L_2 \max\{\|x^{k+1} - x^*\|, \|x^k - x^*\|\}.
\end{aligned}$$

Desse modo, definindo $\varepsilon_k := \max\{\|x^{k+1} - x^*\|, \|x^k - x^*\|\}$ e $\bar{c} := L_2 \|\nabla^2 F_j(x^*)^{-1/2}\|^2$, obtemos

$$\frac{\|\tilde{y}_k - \tilde{s}_k\|}{\|\tilde{s}_k\|} \leq \bar{c}\varepsilon_k \implies \|\tilde{y}_k\|^2 - 2\tilde{s}_k^T \tilde{y}_k + \|\tilde{s}_k\|^2 \leq \bar{c}^2 \varepsilon_k^2 \|\tilde{s}_k\|^2. \quad (3-20)$$

Além disso, como $|\|\tilde{y}_k\| - \|\tilde{s}_k\|| \leq \|\tilde{y}_k - \tilde{s}_k\|$, temos

$$(1 - \bar{c}\varepsilon_k) \|\tilde{s}_k\| \leq \|\tilde{y}_k\| \leq \|\tilde{s}_k\| (1 + \bar{c}\varepsilon_k). \quad (3-21)$$

Assim,

$$\begin{aligned}
(1 - \bar{c}\varepsilon_k)^2 \|\tilde{s}_k\|^2 - 2\tilde{s}_k^T \tilde{y}_k + \|\tilde{s}_k\|^2 &\stackrel{(3-21)}{\leq} \|\tilde{y}_k\|^2 - 2\tilde{s}_k^T \tilde{y}_k + \|\tilde{s}_k\|^2 \\
&\stackrel{(3-20)}{\leq} \bar{c}^2 \varepsilon_k^2 \|\tilde{s}_k\|^2.
\end{aligned}$$

Consequentemente,

$$2\tilde{s}_k^T \tilde{y}_k \geq (1 - 2\bar{c}\varepsilon_k + \bar{c}^2 \varepsilon_k^2 + 1 - \bar{c}^2 \varepsilon_k^2) \|\tilde{s}_k\|^2 = 2(1 - \bar{c}\varepsilon_k) \|\tilde{s}_k\|^2.$$

Segue, pelas definições de \tilde{a}_k e \tilde{b}_k , que

$$\tilde{a}_k = \frac{\tilde{s}_k^T \tilde{y}_k}{\|\tilde{s}_k\|^2} \geq 1 - \bar{c}\varepsilon_k \quad (3-22)$$

e

$$\begin{aligned}
\tilde{b}_k &= \frac{\|\tilde{y}_k\|^2}{\tilde{s}_k^T \tilde{y}_k} = \frac{1}{\tilde{a}_k} \frac{\|\tilde{y}_k\|^2}{\|\tilde{s}_k\|^2} \stackrel{(3-21)}{\leq} \frac{(1 + \bar{c}\varepsilon_k)^2}{\tilde{a}_k} \stackrel{(3-22)}{\leq} \frac{1 + 2\bar{c}\varepsilon_k + \bar{c}^2 \varepsilon_k^2}{1 - \bar{c}\varepsilon_k} \\
&= 1 + \frac{3\bar{c} + \bar{c}^2 \varepsilon_k}{1 - \bar{c}\varepsilon_k} \varepsilon_k.
\end{aligned}$$

Agora, como $\{x^k\}$ converge para x^* , temos $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Logo, para k suficiente-

mente grande, temos que existe uma constante $c > 3\bar{c}$ tal que

$$\tilde{b}_k \leq 1 + c\varepsilon_k. \quad (3-23)$$

Pelas propriedades da função $u(t)$, definida em (2-35), temos

$$\frac{-x}{1-x} - \ln(1-x) = 1 - \frac{1}{1-x} + \ln\left(\frac{1}{1-x}\right) = u\left(\frac{1}{1-x}\right) \leq 0, \quad (3-24)$$

para todo $x \in \mathbb{R} \setminus \{1\}$. Com isso, para k suficientemente grande, podemos assumir que $\bar{c}\varepsilon_k < \frac{1}{2}$, e assim

$$\ln(\tilde{a}_k) \stackrel{(3-22)}{\geq} \ln(1 - \bar{c}\varepsilon_k) \geq \frac{-\bar{c}\varepsilon_k}{1 - \bar{c}\varepsilon_k} \geq -2\bar{c}\varepsilon_k \geq -2c\varepsilon_k. \quad (3-25)$$

Multiplicando ambos os lados (direito e esquerdo) da fórmula de atualização (2-24) por $(\nabla^2 F_j(x^*))^{-1/2}$ e agrupando os termos de modo apropriado, obtemos

$$\tilde{B}^{k+1} = \tilde{B}^k - \frac{\tilde{B}^k \tilde{s}_k \tilde{s}_k^T \tilde{B}^k}{\tilde{s}_k^T \tilde{B}^k \tilde{s}_k} + \frac{\tilde{y}_k \tilde{y}_k^T}{\tilde{s}_k^T \tilde{y}_k}.$$

Dessa maneira, podemos deduzir de (2-34) que

$$\begin{aligned} 0 < \psi(\tilde{B}^{k+1}) &= \psi(\tilde{B}^k) + \left[\tilde{b}_k - \ln(\tilde{a}_k) - 1 \right] + \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k} + \ln\left(\frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k}\right) \right] \\ &\quad + \ln(\cos^2 \tilde{\beta}_k) \\ &\leq \psi(\tilde{B}^k) + 3c\varepsilon_k + \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k} + \ln\left(\frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k}\right) \right] + \ln(\cos^2 \tilde{\beta}_k), \end{aligned} \quad (3-26)$$

na qual a última desigualdade segue de (3-23) e (3-25). Somando expressão acima em k e usando (3-9), temos

$$\sum_{k \geq 0} \left(\ln\left(\frac{1}{\cos^2 \tilde{\beta}_k}\right) - \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k} + \ln\left(\frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k}\right) \right] \right) \leq \psi(\tilde{B}^0) + 3c \sum_{k \geq 0} \varepsilon_k < \infty.$$

Como $u(t)$ é uma função não positiva para $t > 0$, veja (2-35), então o termo entre colchetes é não positivo, e como $\ln(1/\cos^2 \tilde{\beta}_k) \geq 0$ para todo $k \geq 0$, temos

$$\lim_{k \rightarrow \infty} \ln \frac{1}{\cos^2 \tilde{\beta}_k} = 0, \quad \lim_{k \rightarrow \infty} \left(1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k} + \ln\left(\frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k}\right) \right) = 0,$$

implicando que

$$\lim_{k \rightarrow \infty} \cos \tilde{\beta}_k = 1, \quad \text{e} \quad \lim_{k \rightarrow \infty} \tilde{q}_k = 1. \quad (3-27)$$

Pela definição de \tilde{q}_k , o limite (3-27) implica (3-17).

Considere agora (3-18). Primeiro, note que

$$\begin{aligned} & \frac{\|\nabla^2 F_j(x^*)^{-1/2}(B_j^k - \nabla^2 F_j(x^*))s_k\|^2}{\|\nabla^2 F_j(x^*)^{1/2}s_k\|^2} \\ &= \frac{\|(\nabla^2 F_j(x^*)^{-1/2}B_j^k \nabla^2 F_j(x^*)^{-1/2} \nabla^2 F_j(x^*)^{1/2} - \nabla^2 F_j(x^*)^{1/2})s_k\|^2}{\|\tilde{s}_k\|^2} \\ &= \frac{\|(\tilde{B}^k \nabla^2 F_j(x^*)^{1/2} - \nabla^2 F_j(x^*)^{1/2})s_k\|^2}{\|\tilde{s}_k\|^2} = \frac{\|(\tilde{B}^k - I_n)\tilde{s}_k\|^2}{\|\tilde{s}_k\|^2} \\ &= \frac{\|\tilde{B}^k \tilde{s}_k\|^2 - 2\tilde{s}_k^T \tilde{B}^k \tilde{s}_k + \|\tilde{s}_k\|^2}{\|\tilde{s}_k\|^2} \\ &= \frac{\tilde{q}_k^2}{\cos^2 \tilde{\beta}_k} - 2\tilde{q}_k + 1. \end{aligned}$$

Por (3-27), o lado direito da expressão anterior converge para 0. Assim,

$$\begin{aligned} & \frac{\|\nabla^2 F_j(x^*)^{-1/2}(B_j^k - \nabla^2 F_j(x^*))s_k\|^2}{\|\nabla^2 F_j(x^*)^{1/2}s_k\|^2} \\ &= \frac{s_k^T (B_j^k - \nabla^2 F_j(x^*)) \nabla^2 F_j(x^*)^{-1/2} \nabla^2 F_j(x^*)^{-1/2} (B_j^k - \nabla^2 F_j(x^*)) s_k}{s_k^T \nabla^2 F_j(x^*)^{1/2} \nabla^2 F_j(x^*)^{1/2} s_k} \\ &= \frac{\left((B_j^k - \nabla^2 F_j(x^*)) s_k \right)^T \nabla^2 F_j(x^*)^{-1} (B_j^k - \nabla^2 F_j(x^*)) s_k}{s_k^T \nabla^2 F_j(x^*) s_k} \\ (3-15) \quad & \geq \frac{1}{L} \frac{\left\| (B_j^k - \nabla^2 F_j(x^*)) s_k \right\|^2}{L \|s_k\|^2}. \end{aligned}$$

Portanto, tomando limites na desigualdade acima, obtemos a condição (3-18). Agora, por (3-18) e pela continuidade de $\nabla^2 F_j(\cdot)$ para cada $j = 1, \dots, m$, temos

$$\begin{aligned} & \lim_{k \rightarrow \infty} \frac{\left\| (B_j^k - \nabla^2 F_j(x^k)) s_k \right\|}{\|s_k\|} \\ & \leq \lim_{k \rightarrow \infty} \frac{\left\| (B_j^k - \nabla^2 F_j(x^*)) s_k \right\|}{\|s_k\|} + \lim_{k \rightarrow \infty} \frac{\left\| (\nabla^2 F_j(x^k) - \nabla^2 F_j(x^*)) s_k \right\|}{\|s_k\|} \\ & = 0, \quad \forall j = 1, \dots, m, \end{aligned}$$

obtendo (3-19). A prova que (3-19) implica (3-18) pode ser obtida de maneira similar.

Finalmente, para todo $k \geq 0$, temos

$$\begin{aligned} 0 < \psi(\tilde{B}^{k+1}) &\stackrel{(3-26)}{\leq} \psi(\tilde{B}^k) + 3c\varepsilon_k \leq \psi(\tilde{B}^0) + 3c \sum_{\ell=0}^k \varepsilon_\ell \\ &\leq \psi(\tilde{B}^0) + 3c \sum_{\ell \geq 0} \varepsilon_\ell. \end{aligned}$$

Logo, por (3-9), existe uma constante positiva $\omega \in \mathbb{R}$ tal que $\psi(\tilde{B}^k) \leq \omega$ para todo $k \geq 0$. Agora, para algum $k \geq 0$ e por (2-32), temos

$$\sigma_n^k - \ln \sigma_n^k \leq \sum_{l=1}^n (\sigma_l - \ln \sigma_l) \leq \omega \quad \text{e} \quad \sigma_1^k - \ln \sigma_1^k \leq \sum_{l=1}^n (\sigma_l - \ln \sigma_l) \leq \omega,$$

em que σ_n^k é o maior e σ_1^k é o menor autovalor de \tilde{B}^k . Como $\ln t < t - \ln t$ para todo $t \in (0, +\infty)$, temos que $\ln \sigma_n^k \leq \omega$ e, conseqüentemente, $\|\tilde{B}^k\| \leq e^\omega$. Logo, pela definição de \tilde{B}^k , temos

$$\begin{aligned} \|B_j^k\| &= \|\nabla^2 F_j(x^*)^{1/2} \nabla^2 F_j(x^*)^{-1/2} B_j^k \nabla^2 F_j(x^*)^{-1/2} \nabla^2 F_j(x^*)^{1/2}\| \\ &= \|\nabla^2 F_j(x^*)^{1/2} \tilde{B}^k \nabla^2 F_j(x^*)^{1/2}\| \stackrel{(3-15)}{\leq} L e^\omega. \end{aligned}$$

Por outro lado, $\sigma_1^k > 0$ e, conseqüentemente, $-\ln \sigma_1^k < \sigma_1^k - \ln \sigma_1^k \leq \omega$. Assim,

$$\frac{1}{\sigma_1^k} \leq e^\omega \quad \implies \quad \|(\tilde{B}^k)^{-1}\| \leq e^\omega,$$

e

$$\begin{aligned} \|H_j^k\| &= \|\nabla^2 F_j(x^*)^{-1/2} \nabla^2 F_j(x^*)^{1/2} H_j^k \nabla^2 F_j(x^*)^{1/2} \nabla^2 F_j(x^*)^{-1/2}\| \\ &= \|\nabla^2 F_j(x^*)^{-1/2} (\tilde{B}^k)^{-1} \nabla^2 F_j(x^*)^{-1/2}\| \stackrel{(3-15)}{\leq} \frac{1}{\mu} e^\omega. \end{aligned}$$

Portanto, $\{\|B_j^k\|\}$ e $\{\|H_j^k\|\}$ são limitadas para todo $j = 1, \dots, m$. ■

Seja $\lambda_k := (\lambda_1^k, \dots, \lambda_m^k) \in \mathbb{R}^m$ o multiplicador de Lagrange do problema (2-20), associado a x^k , que satisfaz (2-21)–(2-23). Com isso, definimos, para todo $k \geq 0$,

$$F_{\lambda_k}(x) := \sum_{j=1}^m \lambda_j^k F_j(x) \quad \text{e} \quad B_{\lambda_k}^k := \sum_{j=1}^m \lambda_j^k B_j^k. \quad (3-28)$$

A seguir, mostramos que a família de funções $\{F_{\lambda_k}(x)\}_{k \geq 0}$ satisfaz uma condição do tipo Dennis–Moré.

Teorema 3.5. *Suponha que a Hipótese 3.2 seja válida. Para cada $k \geq 0$, considere $F_{\lambda_k} : \mathbb{R}^n \rightarrow \mathbb{R}$ e $B_{\lambda_k}^k$ como em (3-28). Então,*

$$\lim_{k \rightarrow \infty} \frac{\|(B_{\lambda_k}^k - \nabla^2 F_{\lambda_k}(x^*))d^k\|}{\|d^k\|} = 0 \quad (3-29)$$

ou, de modo equivalente,

$$\lim_{k \rightarrow \infty} \frac{\|\nabla F_{\lambda_k}(x^k) + \nabla^2 F_{\lambda_k}(x^k)d^k\|}{\|d^k\|} = 0. \quad (3-30)$$

Demonstração. Pelas definições de $B_{\lambda_k}^k$ e F_{λ_k} , usando a desigualdade triangular e considerando (2-22), temos

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\|(B_{\lambda_k}^k - \nabla^2 F_{\lambda_k}(x^*))d^k\|}{\|d^k\|} &= \lim_{k \rightarrow \infty} \frac{\|\sum_{j=1}^m \lambda_j^k (B_j^k - \nabla^2 F_j(x^*))d^k\|}{\|d^k\|} \\ &\leq \lim_{k \rightarrow \infty} \sum_{j=1}^m \lambda_j^k \frac{\|(B_j^k - \nabla^2 F_j(x^*))d^k\|}{\|d^k\|} \\ &\leq \lim_{k \rightarrow \infty} \max_{j=1, \dots, m} \frac{\|(B_j^k - \nabla^2 F_j(x^*))d^k\|}{\|d^k\|}. \end{aligned}$$

Essa desigualdade, junto com (3-18), garante (3-29). Vamos provar agora que (3-29) implica (3-30). Primeiro, note que, por (2-21), temos $B_{\lambda_k}^k d^k = -\nabla F_{\lambda_k}(x^k)$ e portanto (3-30) é equivalente a

$$\lim_{k \rightarrow \infty} \frac{\|(B_{\lambda_k}^k - \nabla^2 F_{\lambda_k}(x^k))d^k\|}{\|d^k\|} = 0. \quad (3-31)$$

Assim, é fácil ver que

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\|(B_{\lambda_k}^k - \nabla^2 F_{\lambda_k}(x^k))d^k\|}{\|d^k\|} &\leq \lim_{k \rightarrow \infty} \frac{\|(B_{\lambda_k}^k - \nabla^2 F_{\lambda_k}(x^*))d^k\|}{\|d^k\|} \\ &\quad + \lim_{k \rightarrow \infty} \|\nabla^2 F_{\lambda_k}(x^*) - \nabla^2 F_{\lambda_k}(x^k)\| \end{aligned} \quad (3-32)$$

e

$$\begin{aligned}
\lim_{k \rightarrow \infty} \|\nabla^2 F_{\lambda_k}(x^*) - \nabla^2 F_{\lambda_k}(x^k)\| &= \lim_{k \rightarrow \infty} \left\| \sum_{j=1}^m \lambda_j^k (\nabla^2 F_j(x^*) - \nabla^2 F_j(x^k)) \right\| \\
&\leq \lim_{k \rightarrow \infty} \sum_{j=1}^m \lambda_j^k \|\nabla^2 F_j(x^*) - \nabla^2 F_j(x^k)\| \\
&\leq \lim_{k \rightarrow \infty} \max_{j=1, \dots, m} \|\nabla^2 F_j(x^*) - \nabla^2 F_j(x^k)\| \\
&= 0,
\end{aligned}$$

em que a última igualdade segue pela continuidade de $\nabla^2 F_j(\cdot)$, para todo $j = 1, \dots, m$. Portanto, (3-31) segue de (3-32) e (3-29). A prova que (3-30) implica (3-29) pode ser obtida de maneira similar. ■

O seguinte resultado auxiliar fornece algumas propriedades relacionadas ao tamanho da direção d^k e ao valor ótimo $\theta(x^k)$.

Lema 3.2. *Suponha que a Hipótese 3.2 seja válida. Então, existem constantes positivas $\bar{\mu}$ e \bar{L} tais que, para todo k grande o suficiente:*

$$\bar{\mu} \|d^k\|^2 \leq |\theta(x^k)| \leq \bar{L} \|d^k\|^2. \quad (3-33)$$

Além disso,

$$\lim_{k \rightarrow \infty} \|d^k\| = 0. \quad (3-34)$$

Demonstração. Por (3-17), existe $\gamma \in (0, 1)$, tal que

$$1 - \gamma \leq \frac{s_k^T B_j^k s_k}{s_k^T \nabla^2 F_j(x^*) s_k} \leq 1 + \gamma, \quad \forall j = 1, \dots, m, \quad (3-35)$$

para todo k grande o suficiente. Por outro lado, para todo $k \geq 0$, e $j = 1, \dots, m$,

$$\frac{1}{L} \frac{s_k^T B_j^k s_k}{\|s_k\|^2} \stackrel{(3-15)}{\leq} \frac{s_k^T B_j^k s_k}{s_k^T \nabla^2 F_j(x^*) s_k} \stackrel{(3-15)}{\leq} \frac{1}{\mu} \frac{s_k^T B_j^k s_k}{\|s_k\|^2}.$$

Assim, usando (3-35) e lembrando que $s_k = \alpha_k d^k$, obtemos

$$\mu(1 - \gamma) \leq \frac{s_k^T B_j^k s_k}{\|s_k\|^2} = \frac{(d^k)^T B_j^k d^k}{\|d^k\|^2} \leq L(1 + \gamma), \quad \forall j = 1, \dots, m,$$

e para todo k grande o suficiente. Com isso, e por (2-22), temos

$$\frac{\mu(1 - \gamma)}{2} \|d^k\|^2 \leq |\theta(x)| \stackrel{(2-23)}{=} \frac{1}{2} \sum_{j=1}^m \lambda_j^k (d^k)^T B_j^k d^k \leq \frac{L(1 + \gamma)}{2} \|d^k\|^2,$$

para todo k grande o suficiente. Definindo $\bar{\mu} := \mu(1 - \gamma)/2$ e $\bar{L} := L(1 + \gamma)/2$, provamos (3-33). Por fim, usando o fato de que $\mathcal{D}(x^k, d^k) < \theta(x) < 0$ (veja o Lema 2.1(ii)), temos

$$0 \leq \lim_{k \rightarrow \infty} \bar{\mu} \|d^k\| \stackrel{(3-33)}{\leq} \lim_{k \rightarrow \infty} \frac{|\theta(x)|}{\|d^k\|} \leq \lim_{k \rightarrow \infty} \frac{|\mathcal{D}(x^k, d^k)|}{\|d^k\|} \stackrel{(1-17)}{=} 0.$$

■

O seguinte resultado mostra que o passo unitário é admissível para k suficientemente grande.

Teorema 3.6. *Suponha que a Hipótese 3.2 seja válida. Então, o comprimento de passo $\alpha_k = 1$ satisfaz as condições de Wolfe padrão (1-12)–(1-13) para k suficientemente grande.*

Demonstração. Seja $j \in \{1, \dots, m\}$ um índice arbitrário. Pela fórmula de Taylor, temos

$$\begin{aligned} F_j(x^k + d^k) &= F_j(x^k) + \nabla F_j(x^k)^T d^k + \frac{1}{2}(d^k)^T \nabla^2 F_j(x^k) d^k + o(\|d^k\|^2) \\ &= F_j(x^k) + \nabla F_j(x^k)^T d^k + \frac{1}{2}(d^k)^T B_j^k d^k \\ &\quad + \frac{1}{2}(d^k)^T \left(\nabla^2 F_j(x^k) - B_j^k \right) d^k + o(\|d^k\|^2) \\ &\stackrel{(3-19)}{=} F_j(x^k) + \nabla F_j(x^k)^T d^k + \frac{1}{2}(d^k)^T B_j^k d^k + o(\|d^k\|^2) \\ &\stackrel{(2-19)}{\leq} F_j(x^k) + t\theta(x^k) + (1 - t)\theta(x^k) + o(\|d^k\|^2), \end{aligned}$$

em que $t := 2c_1 < 1$ e c_1 é um parâmetro algorítmico. Assim, por (3-33) e para k grande o suficiente, temos

$$\begin{aligned} F_j(x^k + d^k) &\leq F_j(x^k) + t\theta(x^k) - \bar{\mu}(1 - t)\|d^k\|^2 + o(\|d^k\|^2), \\ &= F_j(x^k) + t\theta(x^k) + \left[-\bar{\mu}(1 - t) + \frac{o(\|d^k\|^2)}{\|d^k\|^2} \right] \|d^k\|^2. \end{aligned}$$

Tomando k grande o suficiente, obtemos que o termo entre colchetes é negativo, então

$$F_j(x^k + d^k) \leq F_j(x^k) + t\theta(x^k). \quad (3-36)$$

Por outro lado,

$$\begin{aligned}
\theta(x^k) &\stackrel{(2-23)}{=} \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)^T d^k + \frac{1}{2} (d^k)^T \left(\sum_{j=1}^m \lambda_j^k B_j^k \right) d^k \\
&\stackrel{(2-21)}{=} \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)^T d^k - \frac{1}{2} (d^k)^T \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k) \\
&= \frac{1}{2} \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)^T d^k \stackrel{(1-3)}{\leq} \frac{1}{2} \mathcal{D}(x^k, d^k).
\end{aligned} \tag{3-37}$$

Segue por (3-36)–(3-37) e pela definição de t que

$$F_j(x^k + d^k) \leq F_j(x^k) + c_1 \mathcal{D}(x^k, d^k),$$

para todo k suficientemente grande. Como $j \in \{1, \dots, m\}$ é arbitrário, concluímos que o tamanho de passo $\alpha_k = 1$ satisfaz (1-12) para todo k grande o suficiente.

Considere agora a condição de curvatura (1-13). Pela definição de F_{λ_k} em (3-28), temos

$$\begin{aligned}
-\sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)^T d^k &= (d^k)^T \nabla^2 F_{\lambda_k}(x^k) d^k - \left[\nabla^2 F_{\lambda_k}(x^k) d^k + \nabla F_{\lambda_k}(x^k) \right]^T d^k \\
&\stackrel{(3-15)}{\geq} \mu \|d^k\|^2 - \left[\nabla F_{\lambda_k}(x^k) + \nabla^2 F_{\lambda_k}(x^k) d^k \right]^T d^k \\
&\stackrel{(3-30)}{=} \mu \|d^k\|^2 + o(\|d^k\|^2) = \|d^k\|^2 \left[\mu + \frac{o(\|d^k\|^2)}{\|d^k\|^2} \right].
\end{aligned}$$

Assim, por (3-34), para k suficientemente grande, segue que

$$-\sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)^T d^k \geq \frac{\mu}{2} \|d^k\|^2. \tag{3-38}$$

Por outro lado, aplicando o Teorema do Valor Médio na função escalar $\nabla F_{\lambda_k}(\cdot)^T d_k$, existe $v^k := x^k + \tau_k d^k$ para algum $\tau_k \in (0, 1)$ tal que

$$\nabla F_{\lambda_k}(x^k + d^k)^T d_k = \nabla F_{\lambda_k}(x^k)^T d_k + d_k^T \nabla^2 F_{\lambda_k}(v^k) d^k.$$

Logo,

$$\begin{aligned}
\frac{|\nabla F_{\lambda_k}(x^k + d^k)^T d^k|}{\|d^k\|^2} &\leq \frac{\|\nabla F_{\lambda_k}(x^k)^T d^k + (d^k)^T \nabla^2 F_{\lambda_k}(x^k) d^k\|}{\|d^k\|^2} \\
&\quad + \frac{\|(d^k)^T \nabla^2 F_{\lambda_k}(v^k) d^k - (d^k)^T \nabla^2 F_{\lambda_k}(x^k) d^k\|}{\|d^k\|^2} \\
&\leq \frac{\|\nabla F_{\lambda_k}(x^k) + \nabla^2 F_{\lambda_k}(x^k) d^k\|}{\|d^k\|} + \|\nabla^2 F_{\lambda_k}(v^k) - \nabla^2 F_{\lambda_k}(x^k)\|
\end{aligned} \tag{3-39}$$

Agora, pela definição de F_{λ_k} e v^k , e considerando a continuidade de $\nabla^2 F(\cdot)$, obtemos

$$\begin{aligned}
\lim_{k \rightarrow \infty} \|\nabla^2 F_{\lambda_k}(v^k) - \nabla^2 F_{\lambda_k}(x^k)\| &= \lim_{k \rightarrow \infty} \left\| \sum_{j=1}^m \lambda_j^k (\nabla^2 F_j(x^k + \tau_k d^k) - \nabla^2 F_j(x^k)) \right\| \\
&\leq \lim_{k \rightarrow \infty} \sum_{j=1}^m \lambda_j^k \|\nabla^2 F_j(x^k + \tau_k d^k) - \nabla^2 F_j(x^k)\| \\
&\stackrel{(2-22)}{\leq} \lim_{k \rightarrow \infty} \max_{j=1, \dots, m} \|\nabla^2 F_j(x^k + \tau_k d^k) - \nabla^2 F_j(x^k)\| \\
&\stackrel{(3-34)}{=} 0.
\end{aligned}$$

Desse modo, segue por (3-30) e (3-39) que

$$\lim_{k \rightarrow \infty} \frac{|\nabla F_{\lambda_k}(x^k + d^k)^T d^k|}{\|d^k\|^2} = 0.$$

Consequentemente, para k grande o suficiente e c_2 definido no Algoritmo 3.1, temos

$$|\nabla F_{\lambda_k}(x^k + d^k)^T d^k| \leq c_2 \frac{\mu}{4} \|d^k\|^2. \tag{3-40}$$

Portanto,

$$\frac{c_2}{2} \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)^T d^k \stackrel{(3-38)}{\leq} -c_2 \frac{\mu}{4} \|d^k\|^2 \stackrel{(3-40)}{\leq} \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k + d^k)^T d^k.$$

Por fim, da última desigualdade, segue que

$$\begin{aligned}
\mathcal{D}(x^k + d^k, d^k) &\geq \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k + d^k)^T d^k \geq \frac{c_2}{2} \sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)^T d^k \stackrel{(3-37)}{=} c_2 \theta(x^k) \\
&\geq c_2 \mathcal{D}(x^k, d^k),
\end{aligned}$$

para todo k suficientemente grande, na qual a última desigualdade segue do

Lema 2.1(ii), concluindo a prova. ■

O Lema a seguir apresenta uma desigualdade muito útil que será usada em nosso resultado principal. Sua demonstração pode ser encontrada em [21, Lema 4.1.15], contudo, por uma questão de completude, a faremos neste trabalho.

Lema 3.3. *Suponha que a Hipótese 3.2 é válida. Então, para todo $j = 1, \dots, m$, temos:*

$$\begin{aligned} & \|\nabla F_j(x^{k+1}) - \nabla F_j(x^k) - \nabla^2 F_j(x^*)(x^{k+1} - x^k)\| \\ & \leq L_2 \frac{\|x^{k+1} - x^*\| + \|x^k - x^*\|}{2} \|x^{k+1} - x^k\|, \end{aligned}$$

em que L_2 é dado em (3-16).

Demonstração. Lembrando que $s_k = x^{k+1} - x^k$, temos

$$\begin{aligned} & \|\nabla F_j(x^{k+1}) - \nabla F_j(x^k) - \nabla^2 F_j(x^*)(x^{k+1} - x^k)\| \\ & \stackrel{(3-5)}{=} \left\| \int_0^1 \nabla^2 F_j(x^k + \tau s_k) s_k d\tau - \nabla^2 F_j(x^*) s_k \right\| \\ & = \left\| \int_0^1 (\nabla^2 F_j(x^k + \tau s_k) - \nabla^2 F_j(x^*)) s_k d\tau \right\| \\ & \leq \int_0^1 \|(\nabla^2 F_j(x^k + \tau s_k) - \nabla^2 F_j(x^*))\| \cdot \|s_k\| d\tau \\ & \stackrel{(3-16)}{\leq} L_2 \|s_k\| \int_0^1 \|\tau x^{k+1} + (1 - \tau)x^k - x^*\| d\tau = L_2 \|s_k\| \frac{\|x^{k+1} - x^*\| + \|x^k - x^*\|}{2}. \end{aligned}$$

■

Agora estamos preparados para provar a convergência superlinear do Algoritmo 3.1. Esse resultado é baseado em [20, Teorema 2.2].

Teorema 3.7. *Suponha que a Hipótese 3.2 seja válida. Então, $\{x^k\}$ converge para x^* com taxa Q -superlinear.*

Demonstração. Pelo Teorema 3.6, podemos assumir, sem perda de generalidade, que $\alpha_k = 1$ e conseqüentemente $d^k = x^{k+1} - x^k$, para todo k . Assim, por (2-21), temos $B_{\lambda_k}^k(x^{k+1} - x^k) = -\nabla F_{\lambda_k}(x^k)$ e então

$$\begin{aligned} (B_{\lambda_k}^k - \nabla^2 F_{\lambda_k}(x^*))(x^{k+1} - x^k) &= \nabla F_{\lambda_k}(x^{k+1}) - \nabla F_{\lambda_k}(x^k) \\ &\quad - \nabla^2 F_{\lambda_k}(x^*)(x^{k+1} - x^k) - \nabla F_{\lambda_k}(x^{k+1}). \end{aligned}$$

Desse modo,

$$\begin{aligned} \frac{\|\nabla F_{\lambda_k}(x^{k+1})\|}{\|x^{k+1} - x^k\|} &\leq \frac{\|(B_{\lambda_k}^k - \nabla^2 F_{\lambda_k}(x^*)) (x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} \\ &+ \frac{\|\nabla F_{\lambda_k}(x^{k+1}) - \nabla F_{\lambda_k}(x^k) - \nabla^2 F_{\lambda_k}(x^*) (x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|}. \end{aligned} \quad (3-41)$$

Note que o segundo termo do lado direito é menor ou igual a

$$\max_{j=1, \dots, m} \frac{\|\nabla F_j(x^{k+1}) - \nabla F_j(x^k) - \nabla^2 F_j(x^*) (x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|}$$

e, pelo Lema 3.3, essa expressão é menor ou igual a

$$L_2 \frac{\|x^{k+1} - x^*\| + \|x^k - x^*\|}{2}.$$

Assim, aplicando o limite em ambos os lados de (3-41) e usando (3-29), obtemos

$$\lim_{k \rightarrow \infty} \frac{\|\nabla F_{\lambda_k}(x^{k+1})\|}{\|x^{k+1} - x^k\|} = 0. \quad (3-42)$$

Por outro lado, pela definição de F_{λ_k} , pelo Lema 1.2(iv) e Lema 3.1(i), temos que

$$\begin{aligned} \frac{\|\nabla F_{\lambda_k}(x^{k+1})\|}{\|x^{k+1} - x^k\|} &\geq \frac{\|\sum_{j=1}^m \lambda_j^k \nabla F_j(x^{k+1})\|}{\|x^{k+1} - x^*\| + \|x^k - x^*\|} \geq \frac{\|d_{SD}(x^{k+1})\|}{\|x^{k+1} - x^*\| + \|x^k - x^*\|} \\ &\geq \frac{\mu}{2} \frac{\|x^{k+1} - x^*\|}{\|x^{k+1} - x^*\| + \|x^k - x^*\|} = \frac{\mu}{2} \frac{1}{1 + \frac{\|x^k - x^*\|}{\|x^{k+1} - x^*\|}}. \end{aligned}$$

Portanto, usando (3-42), obtemos

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0,$$

e com isso a taxa de convergência é Q-superlinear. ■

3.1.3 Experimentos Numéricos

Esta seção apresenta alguns resultados numéricos para ilustrar as potenciais vantagens práticas do Algoritmo 3.1. Estamos interessados principalmente em verificar a eficácia do uso da busca linear de Wolfe e das aproximações Hessianas que são atualizadas a cada iteração por um esquema BFGS. Para isso, consideramos os seguintes métodos nos testes reportados:

- Algoritmo 3.1 (BFGS-Wolfe): esquema aqui proposto no qual as aproximações Hessianas são atualizadas a cada iteração por (3-1) e os tamanhos dos passos são calculados satisfazendo as condições de Wolfe (1-12)-(1-13).
- Padrão BFGS-Armijo: um algoritmo BFGS em que as aproximações Hessianas são atualizadas, para cada $j = 1, \dots, m$, por (2-40) em que $\varepsilon > 0$ é um parâmetro algorítmico e os comprimentos de passos são calculados satisfazendo uma condição do tipo Armijo (1-12). Em nossos experimentos, usamos $\varepsilon = 10^{-6}$.
- Padrão BFGS-Wolfe: um algoritmo BFGS no qual as aproximações Hessianas são atualizadas por (2-40) e os tamanhos de passos são calculados satisfazendo as condições de Wolfe (1-12)-(1-13).

Todos os algoritmos que testamos foram implementados em Fortran 90. As direções de busca $d(x^k)$ e os valores ótimos $\theta(x^k)$ são calculados ao resolver o subproblema (2-20) usando o Algencan [3], um código de Lagrangiano aumentado para programação não linear geral. Para calcular o tamanho de passo satisfazendo as condições (1-12)-(1-13), usamos o algoritmo proposto em [51]. Este algoritmo envolve várias interpolações polinomiais quadráticas/cúbicas das funções objetivo, combina estratégias de retrocesso e extrapolação e é capaz de calcular o tamanho do passo em um número finito de iterações (internas). Técnicas de interpolação também foram usadas para calcular tamanhos de passos que satisfazem apenas a condição tipo Armijo (1-12). Usamos $c_1 = 10^{-4}$, $c_2 = 0.1$, e tomamos $B_j^0 = I_n$ para todo $j = 1, \dots, m$. Para o caso escalar, existem estratégias mais elaboradas para escolher a matriz inicial, veja [21, Seção 9.4]. Paramos os algoritmos em x^k reportando convergência quando $|\theta(x^k)| \leq 5 \times \text{eps}^{1/2}$, em que $\text{eps} = 2^{-52} \approx 2.22 \times 10^{-16}$ é a precisão da máquina. O número máximo de iterações permitidas foi 2000.

Embora o Algoritmo 3.1 tenha convergência global apenas sob hipóteses de convexidade, experiências numéricas favoráveis também são comumente observadas para problemas não convexos. Assim, o conjunto de problemas teste escolhido inclui problemas convexos e não convexos encontrados na literatura de otimização multiobjetivo. A Tabela 3.1 mostra suas principais características. As duas primeiras colunas contêm o nome do problema e a referência correspondente. As colunas “ n ” e “ m ” informam o número de variáveis e o número de objetivos, respectivamente. “Conv.” indica quando o problema é convexo ou não. Muitos problemas têm restrições de caixa em suas definições originais. Em alguns deles, os objetivos são ilimitados fora da caixa. Nestes casos, acrescentamos um termo que penaliza o não cumprimento das restrições a cada objetivo. Se denotarmos a caixa por $\{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ em que $l, u \in \mathbb{R}^n$, o termo de penalidade é definido por $\frac{\mu}{3} [\|\max\{0, x - u\}\|_3^3 + \|\max\{0, -x + l\}\|_3^3]$, no qual $\mu = 10^{10}$ e o máximo é

calculado componente a componente. Isso força os iterados a permanecerem dentro da caixa. A coluna “Penal.” informa se determinado problema foi penalizado ou não. Os pontos iniciais foram tomados pertencentes às caixas correspondentes. Ressaltamos que as caixas não são consideradas explicitamente pelos algoritmos.

Problema	Ref.	n	m	Conv.	Penal.
AP1	[1]	2	3	S	N
AP2	[1]	1	2	S	N
AP3	[1]	2	2	N	N
AP4	[1]	3	3	S	N
BK1	[40]	2	2	S	N
DD1	[16]	5	2	N	S
DGO1	[40]	1	2	N	N
DGO2	[40]	1	2	S	S
DTLZ1	[19]	7	3	N	S
DTLZ2	[19]	7	3	N	S
DTLZ3	[19]	7	3	N	S
DTLZ4	[19]	7	3	N	S
FA1	[40]	3	3	N	S
Far1	[40]	2	2	N	N
FDS	[28]	5	3	S	N
FF1	[40]	2	2	N	N
Hil1	[39]	2	2	N	N
IKK1	[40]	2	3	S	N
IM1	[40]	2	2	N	S
JOS1	[42]	2	2	S	N
JOS4	[42]	20	2	N	S
KW2	[43]	2	2	N	S
LE1	[40]	2	2	N	N
Lov1	[48]	2	2	S	N
Lov2	[48]	2	2	N	S
Lov3	[48]	2	2	N	N
Lov4	[48]	2	2	N	N
Lov5	[48]	3	2	N	N
Lov6	[48]	6	2	N	S
LTDZ	[40]	3	3	N	S
MGH9 ^a	[57]	3	15	N	S
MGH16 ^a	[57]	4	5	N	N
MGH26 ^a	[57]	4	4	N	N
MGH33 ^a	[57]	10	10	S	N

Problema	Ref.	n	m	Conv.	Penal.
MHBM2	[40]	2	3	S	N
MLF1	[40]	1	2	N	S
MLF2	[40]	2	2	N	N
MMR1	[55]	2	2	N	S
MMR2	[55]	2	2	N	S
MMR3	[55]	2	2	N	S
MMR4	[55]	3	2	N	S
MOP2	[40]	2	2	N	N
MOP3	[40]	2	2	N	N
MOP5	[40]	2	3	N	N
MOP6	[40]	2	2	N	S
MOP7	[40]	2	3	S	N
PNR	[63]	2	2	S	N
QV1	[40]	10	2	N	N
SD	[72]	4	2	S	S
SK1	[40]	1	2	N	N
SK2	[40]	4	2	N	N
SLCDDT1	[70]	2	2	N	N
SLCDDT2	[70]	10	3	S	N
SP1	[40]	2	2	S	N
SSFY2	[40]	1	2	N	N
TKLY1	[40]	4	2	N	S
Toi4 ^a	[75]	4	2	S	N
Toi8 ^a	[75]	3	3	S	N
Toi9 ^a	[75]	4	4	N	N
Toi10 ^a	[75]	4	3	N	N
VU1	[40]	2	2	N	N
VU2	[40]	2	2	S	S
ZDT1	[77]	30	2	S	S
ZDT2	[77]	30	2	N	S
ZDT3	[77]	30	2	N	S
ZDT4	[77]	30	2	N	S
ZDT6	[77]	10	2	N	S
ZLT1	[40]	10	5	S	N

^a Esta é uma adaptação de um problema de otimização escalar para a configuração multiobjetivo que pode ser encontrada em [56].

Tabela 3.1: Lista de problemas teste.

Dado um problema de otimização multiobjetivo, estamos especialmente interessados em estimar sua fronteira Pareto. Para atingir esse objetivo, uma estratégia frequentemente usada é executar o algoritmo disponível a partir de vários pontos iniciais diferentes. Tendo em vista esta aplicação, consideramos 300 pontos iniciais aleatórios para cada problema da Tabela 3.1. Cada instância foi vista como um problema independente e foi resolvida por todos os algoritmos. A Figura 3.1 mostra os resultados usando *performance profiles* [22]. Comparamos os algoritmos com respeito a: (a) número de iterações; (b) tempo de CPU; (c) número de avaliações de funções objetivo; (d) número de avaliações de gradientes.

Iniciamos nossa análise observando que todos os algoritmos se mostraram robustos no conjunto de problemas escolhido. Em particular, isso ilustra a capacidade prática do Algoritmo 3.1 mesmo para problemas não convexos. O Algoritmo 3.1 e os algoritmos Padrão BFGS-Wolfe e Padrão BFGS-Armijo obtiveram sucesso resolvendo 99.8%, 99.7%, 98.6% dos problemas testados, respectivamente. Em relação à eficiência, considerando o número de iterações, o Algoritmo 3.1 (86.2%)

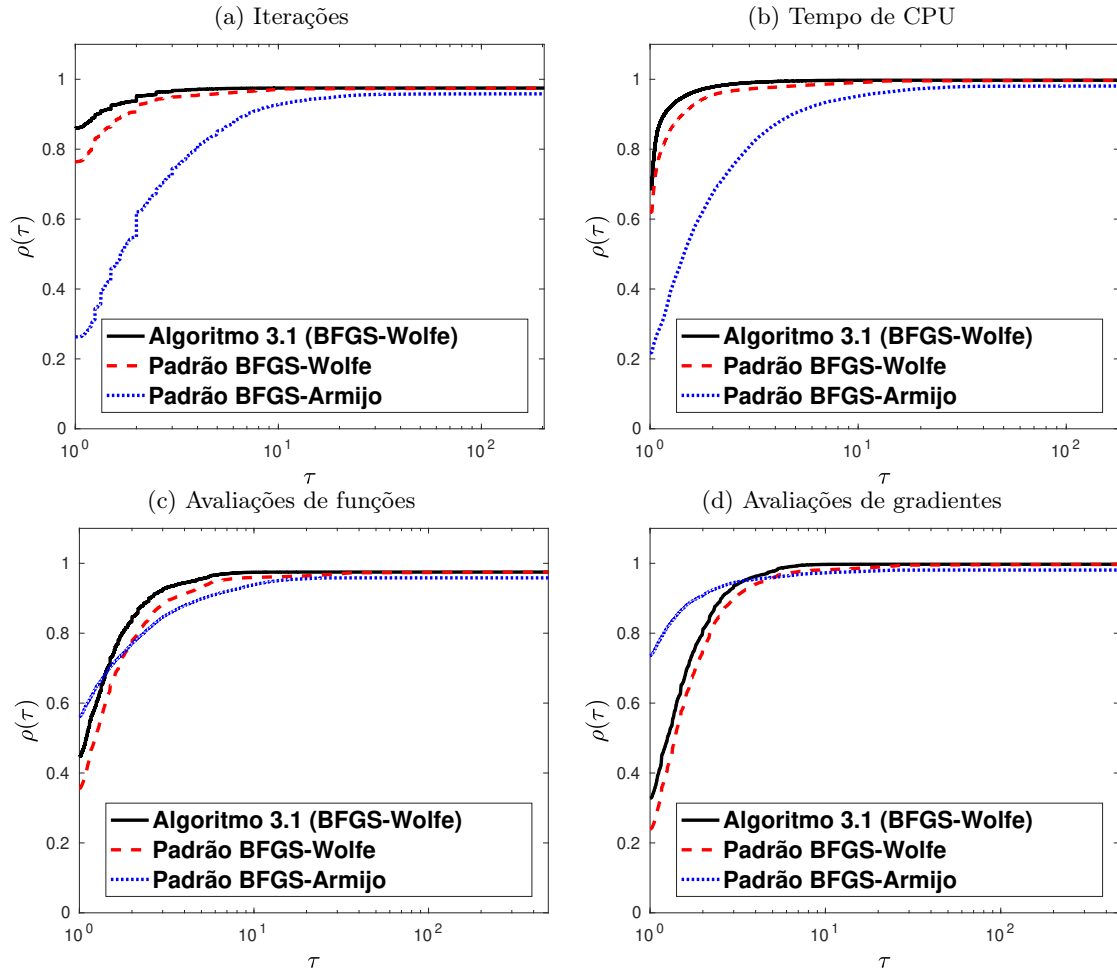


Figura 3.1: Performance profiles considerando 300 pontos de início para cada problema teste usando como medida de desempenho: (a) número de iterações; (b) tempo de CPU; (c) número de avaliações de funções; (d) número de avaliações de gradientes.

teve o melhor desempenho seguido pelos algoritmos Padrão BFGS-Wolfe (76.5%) e Padrão BFGS-Armijo (27.3%), veja a Figura 3.1(a). Esse fato se refletiu diretamente no tempo de CPU (eficiência de 68.7%, 61.6% e 22.6% para o Algoritmo 3.1 e os algoritmos Padrão BFGS-Wolfe e Padrão BFGS-Armijo, respectivamente), como pode ser visto na Figura 3.1(b). Com relação aos números de avaliações de funções e gradientes, o algoritmo Padrão BFGS-Armijo foi o mais eficiente, veja a Figura 3.1(c) e (d). Isso era esperado, pois a busca linear de Wolfe usa mais informações dos objetivos do que a busca linear de Armijo. Entretanto, o algoritmo Padrão BFGS-Armijo foi rapidamente superado pelos demais algoritmos em termos de número de avaliações de funções. A forte correlação entre o número de iterações e o tempo de CPU pode ser explicada pelo fato de que, em nossos experimentos, o custo computacional é amplamente dominado pelas soluções dos subproblemas que fornecem as direções de busca. De fato, nos Algoritmo 3.1 e no algoritmo Padrão

BFGS-Armijo foram gastos, em média, 93.2% e 94.5% do tempo total de CPU na solução de subproblemas, e apenas 2.0% e 0.7% no cálculo de tamanhos de passos, respectivamente. Portanto, pelo menos em nossos testes, o esforço computacional nas buscas lineares pode ser desprezado e o uso de tamanhos de passo de Wolfe se justifica devido ao seu impacto na diminuição do número de iterações e, conseqüentemente, do tempo de CPU. Outra questão de interesse diz respeito ao comprimento dos tamanhos de passos calculados pelos procedimentos de busca linear de Wolfe e Armijo. A Figura 3.2 mostra os histogramas (normalizados pela probabilidade relativa) contendo a distribuição de frequência de todos os tamanhos de passos calculados pelo Algoritmo 3.1 e pelo algoritmo Padrão BFGS-Armijo. Como pode ser visto, em 83.61% das iterações, o tamanho do passo de Wolfe foi maior que ou igual a um. Ressaltamos que, devido ao uso de estratégias de extrapolação, mesmo tamanhos de passos maiores que um podem ser considerados no procedimento de busca linear de Wolfe. Uma frequência similar (78.18%) para o passo unitário foi observado para os tamanhos de passo de Armijo. Contudo, as buscas lineares de Wolfe e Armijo obtiveram tamanhos de passos menores que 0.1 em 8.66% e 18.28% das vezes, respectivamente. Isso corrobora a discussão feita na Observação 1.2, na qual afirmamos que as condições de Wolfe evitam que o método tome tamanhos de passos excessivamente pequenos quando tamanhos de passos *maiores* são possíveis.

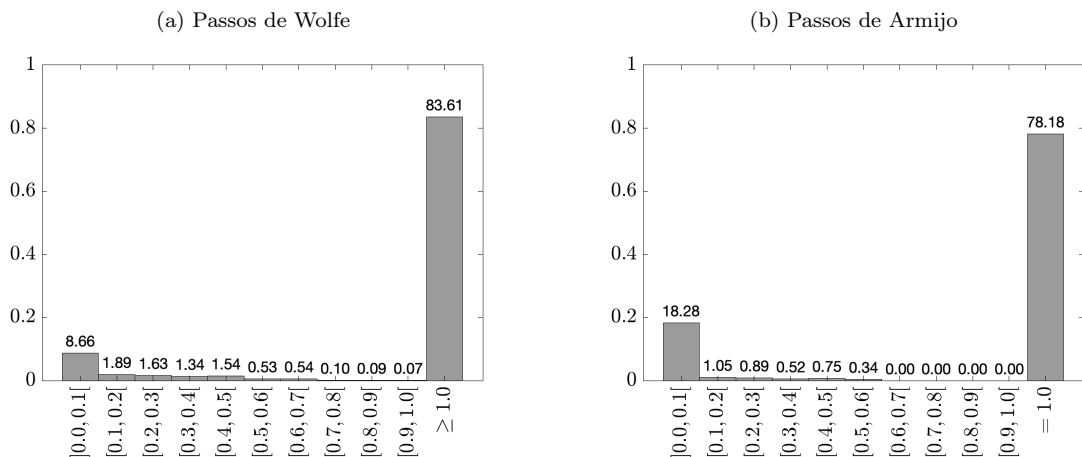


Figura 3.2: Histogramas (normalizados pela probabilidade relativa) contendo a distribuição de frequência de todos os tamanhos de passos calculados por: (a) Algoritmo 3.1 (tamanho de passo de Wolfe); (b) algoritmo Padrão BFGS-Armijo (tamanho de passo de Armijo).

A seguir, comparamos a capacidade dos algoritmos de gerar apropriadamente as fronteiras Pareto. Para tal, usamos as conhecidas métricas *Purity* e *Spread* (Γ e Δ), que explicaremos brevemente aqui. Recomendamos ao leitor [13] para uma explicação detalhada dessas métricas e seus usos em conjunto com o *performance*

profile. Seja $PF_{p,s}$ uma aproximação da fronteira Pareto obtida pelo *solver* s para o problema p . Aqui, $PF_{p,s}$ foi obtido resolvendo o problema p pelo *solver* s a partir de 300 pontos iniciais e então removendo os pontos dominados. Agora, seja PF_p uma aproximação da fronteira Pareto obtida considerando primeiro $\cup_{s \in \mathcal{S}} PF_{p,s}$ e depois removendo os pontos dominados.

- Métrica *Purity*: A métrica *Purity* mede o número de pontos não dominados pertencentes a PF_p encontrados por um determinado *solver*. Dado um *solver* s e um problema p , ele é definido pela razão

$$\bar{t}_{p,s} := \frac{|PF_{p,s} \cap PF_p|}{|PF_p|}.$$

Para analisar a métrica *Purity* usando o perfil de desempenho, definimos $t_{p,s} := 1/\bar{t}_{p,s}$ e, assim, valores menores de $t_{p,s}$ indicam melhores desempenhos. Se $\bar{t}_{p,s} = 0$, então definimos $t_{p,s} := \infty$.

- Métrica *Spread*: Uma métrica *Spread* busca medir a habilidade de um determinado *solver* em obter pontos *bem distribuídos* ao longo da fronteira Pareto. Dado um *solver* s e um problema p , considere que $PF_{p,s} \cap PF_p$ é formado por x_1, \dots, x_N e assumamos que esses pontos são convenientemente ordenados para cada função objetivo j tal que $F_j(x_i) \leq F_j(x_{i+1})$, $i = 1, \dots, N$. Sejam x_0 e x_{N+1} os pontos correspondentes aos menores e maiores valores, respectivamente, de F_j obtidos de PF_p . As métricas Γ e Δ são definidas por

$$\Gamma_{p,s} := \max_{j \in \{1, \dots, m\}} \max_{i \in \{0, \dots, N\}} \delta_{i,j}$$

e

$$\Delta_{p,s} := \max_{j \in \{1, \dots, m\}} \left(\frac{\delta_{0,j} + \delta_{N,j} + \sum_{i=1}^N |\delta_{i,j} - \bar{\delta}_j|}{\delta_{0,j} + \delta_{N,j} + (N-1)\bar{\delta}_j} \right),$$

em que $\delta_{i,j} := |F_j(x_{i+1}) - F_j(x_i)|$ e $\bar{\delta}_j$ ($j = 1, \dots, m$) é a média das distâncias $\delta_{i,j}$, $i = 1, \dots, N$. No *performance profile*, definimos $t_{p,s} := \Gamma_{p,s}$ ou $t_{p,s} := \Delta_{p,s}$, dependendo da métrica escolhida.

Os resultados na Figura 3.3 mostram que o Algoritmo 3.1 e o algoritmo Padrão BFGS-Wolfe superaram o algoritmo Padrão BFGS-Armijo em relação à métrica *Purity*. Nenhuma diferença significativa foi observada para a métrica *Spread*.

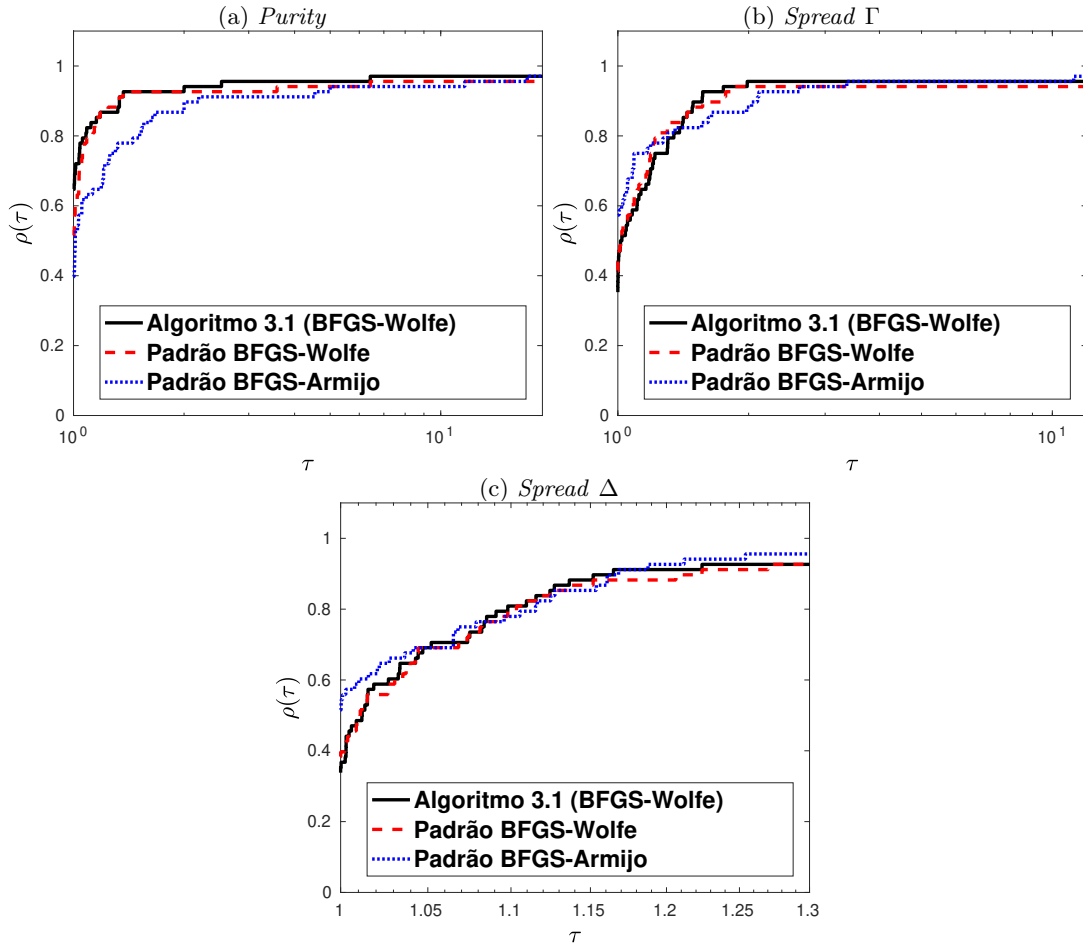


Figura 3.3: Performance profiles com respeito às métricas:
 (a) Purity; (b) Spread Γ ; (c) Spread Δ .

A Figura 3.4 mostra o esboço dos conjuntos imagem e fronteira Pareto obtidos pelo Algoritmo 3.1 para os problemas Hil1, KW2, MMR3 e MOP6. Nos gráficos, um ponto representa uma iteração final, enquanto o início de um segmento de reta representa o ponto inicial correspondente. Os pontos Pareto ótimos estão destacados por um quadrado. Como esperado, uma vez que esses problemas são não convexos, o Algoritmo 3.1 converge para alguns pontos Pareto críticos que não são ótimos.

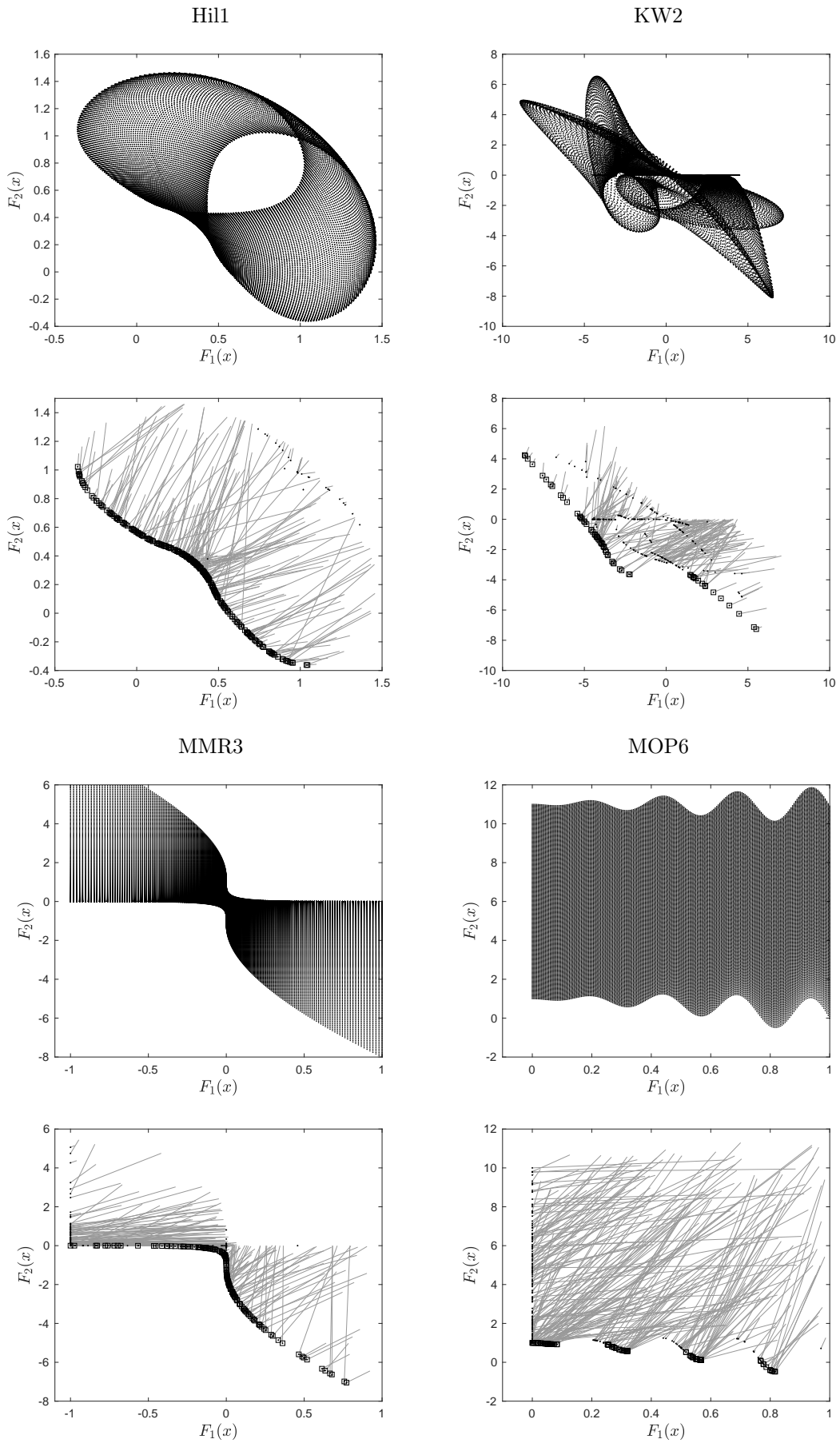


Figura 3.4: Conjunto imagem e a fronteira Pareto obtida pelo Algoritmo 3.1 para os problemas não convexos Hil1, KW2, MMR3 e MOP6.

Por fim, verificamos o comportamento dos métodos à medida que o número de funções objetivo aumenta. Para isso, comparamos os desempenhos do Algoritmo 3.1 e o do algoritmo Padrão BFGS-Armijo nos problemas DTLZ1, DTLZ2, DTLZ3, DTLZ4, MGH26, Toi9 e Toi10 com dimensões maiores. Esses são os problemas da Tabela 3.1 em que a dimensão m é customizável. MGH26, Toi9 e Toi10 são extensões de problemas de otimização escalar também conhecidos como *Trigonometric*, *Shifted TRIDIA* e *Rosenbrock*, respectivamente. As três primeiras colunas da Tabela 3.2 identificam o problema e as dimensões consideradas. Para cada instância, executamos os dois algoritmos a partir de 10 pontos iniciais aleatórios diferentes. Ressaltamos que os algoritmos reportaram convergência em todos os casos. A tabela fornece as médias de: número de iterações (it), tempo de CPU em segundos (tempo), número de avaliações de funções (nfev) e gradientes (ngev). Os melhores valores reportados para cada instância estão destacados em negrito. Ressaltamos que consideramos cada avaliação de um objetivo (resp. gradiente de um objetivo) no cálculo de nfev (resp. ngev). Como pode ser visto, para os problemas MGH26 e DTLZ, o Algoritmo 3.1 superou fortemente o algoritmo Padrão BFGS-Armijo com respeito ao número de iterações e ao tempo de CPU. Tipicamente, levando em consideração essas medidas de performance, o Algoritmo 3.1 usa menos da metade dos recursos computacionais requeridos pelo algoritmo Padrão BFGS-Armijo nesse grupo de problemas. Mesmo com respeito ao número de avaliações de funções e gradientes, uma vantagem para o Algoritmo 3.1 foi, em geral, observada. Em relação aos problemas Toi9 e Toi10, os algoritmos apresentaram um desempenho mais homogêneo. Enquanto o algoritmo Padrão BFGS-Armijo foi o mais eficiente em termos de avaliações de funções e gradientes, o Algoritmo 3.1 sempre exigiu menos iterações, resultando em economia de tempo de CPU em quatro das seis instâncias.

Problema	n	m	Algoritmo 3.1 (BFGS-Wolfe)				Padrão BFGS-Armijo			
			it	tempo	nfev	ngev	it	tempo	nfev	ngev
DTLZ1	500	5	6.5	31.1	84.1	71.7	30.0	118.9	244.4	155.0
	500	8	5.2	37.5	93.0	86.1	10.7	57.3	134.4	93.6
	500	10	2.9	14.4	73.0	68.2	7.8	32.9	131.2	88.0
DTLZ2	500	5	2.7	6.8	49.6	44.8	10.9	36.2	123.8	59.5
	500	8	3.1	11.0	84.0	73.6	8.9	55.5	129.5	79.2
	500	10	3.1	13.5	94.8	87.4	9.1	68.2	167.9	101.0
DTLZ3	500	5	4.7	6.7	57.4	50.1	33.7	53.0	208.0	173.5
	500	8	7.6	21.1	116.9	108.5	81.0	287.1	719.4	656.0
	500	10	5.8	17.6	113.6	105.9	23.5	160.3	326.0	245.0
DTLZ4	500	5	2.4	8.4	54.1	48.7	4.2	14.2	30.4	26.0
	500	8	2.9	13.4	89.8	81.5	5.0	23.5	58.2	48.0
	500	10	2.4	13.5	84.8	76.0	6.0	33.5	98.5	70.0
MGH26	100	100	14.8	4.4	3714.3	3339.6	36.5	9.5	3750.0	3750.0
	200	200	16.3	42.3	9739.5	8056.8	49.1	122.9	10020.0	10020.0
	400	400	21.6	480.2	26987.7	22211.8	59.3	1298.9	24120.0	24120.0
Toi9	100	100	4.5	2.7	1014.5	868.8	5.4	2.4	640.2	640.0
	300	300	4.2	63.4	2387.2	2165.4	4.6	71.7	1765.1	1680.0
	400	400	3.8	119.9	2628.4	2442.9	4.7	134.2	2370.4	2280.0
Toi10	100	99	11.3	9.1	1895.7	1733.9	12.3	11.2	1333.3	1316.7
	200	199	8.7	85.8	3269.6	2937.9	10.8	70.3	2421.6	2348.2
	300	299	10.2	282.9	5357.8	4796.8	15.6	336.4	4992.0	4963.4

Tabela 3.2: Performance do Algoritmo 3.1 (BFGS-Wolfe) e do algoritmo Padrão BFGS-Armijo em instâncias maiores dos problemas DTLZ1, DTLZ2, DTLZ3, DTLZ4, MGH26, Toi9 e Toi10.

3.2 Um método BFGS modificado para problemas não convexos

Nesta seção, descrevemos detalhadamente o algoritmo que estende, para otimização multiobjetivo, o método BFGS escalar modificado (mBFGS) que foi proposto em [46]. Como discutido na Observação 2.2, essa modificação tem como objetivo globalizar o método BFGS para problemas não convexos. Desse modo, no caso multiobjetivo, buscamos mimetizar o algoritmo BFGS modificado, juntamente com as hipóteses usadas para obter os resultados de convergência. Assim como no caso escalar, o Algoritmo 3.2 (veja a seguir) usa um processo de busca linear satisfazendo as condições de Wolfe padrão (1-12)–(1-13). Isso será essencial para garantir que as atualizações das matrizes B_j^k sejam definidas positivas para problemas não convexos. Vale ressaltar que alguns resultados apresentados nesta seção são provados por caminhos alternativos aos utilizados em [46]. Destacamos que os testes numéricos referentes a essa proposta serão apresentados na Seção 3.3.3.

Algoritmo 3.2. Um algoritmo mBFGS com busca linear de Wolfe

Sejam $c_1 \in (0, 1/2)$, $c_2 \in (c_1, 1)$, $C \in (0, +\infty)$, $x^0 \in \mathbb{R}^n$, e $B_j^0 \succ 0$, para todo $j = 1, \dots, m$. Inicialize $k \leftarrow 0$.

Passo 1. Calcule a direção de busca

Calcule $d^k := d(x^k)$ e $\theta(x^k)$ como em (2-18) e (2-19), respectivamente.

Passo 2. Critério de parada

Se $\theta(x^k) = 0$, então PARE.

Passo 3. Procedimento de busca linear

Calcule o comprimento de passo $\alpha_k > 0$ (primeiro tente $\alpha_k = 1$) tal que as condições de Wolfe padrão (1-12)–(1-13) sejam satisfeitas e faça $x^{k+1} := x^k + \alpha_k d^k$.

Passo 4. Prepare a próxima iteração

Para cada $j = 1, \dots, m$, defina

$$\begin{aligned}
 B_j^{k+1} := & B_j^k - \frac{(\rho_j^k)^{-1} B_j^k s_k s_k^T B_j^k}{\left((\rho_j^k)^{-1} - s_k^T \gamma_j^k \right)^2 + (\rho_j^k)^{-1} s_k^T B_j^k s_k} \\
 & + \frac{(s_k^T B_j^k s_k) \gamma_j^k (\gamma_j^k)^T}{\left((\rho_j^k)^{-1} - s_k^T \gamma_j^k \right)^2 + (\rho_j^k)^{-1} s_k^T B_j^k s_k} \\
 & + \left((\rho_j^k)^{-1} - s_k^T \gamma_j^k \right) \frac{\gamma_j^k s_k^T B_j^k + B_j^k s_k (\gamma_j^k)^T}{\left((\rho_j^k)^{-1} - s_k^T \gamma_j^k \right)^2 + (\rho_j^k)^{-1} s_k^T B_j^k s_k},
 \end{aligned} \tag{3-43}$$

em que $s_k = x^{k+1} - x^k$, $\gamma_j^k := y_j^k + r_j^k s_k$, $y_j^k = \nabla F_j(x^{k+1}) - \nabla F_j(x^k)$, $r_j^k \in [0, C]$ e

$$\rho_j^k := \begin{cases} 1/s_k^T \gamma_j^k, & \text{se } s_k^T \gamma_j^k > 0, \\ 1/\left(\mathcal{D}(x^{k+1}, s_k) - \nabla F_j(x^k)^T s_k + r_j^k \|s_k\|^2\right), & \text{caso contrário.} \end{cases} \quad (3-44)$$

Faça $k \leftarrow k + 1$ e vá ao **Passo 1**.

Observação 3.4. (i) A diferença fundamental entre os Algoritmos 3.1 e 3.2 está na maneira de obter a atualização B_j^{k+1} . A fórmula (3-43) corresponde a (3-1) com o vetor y_j^k substituído por γ_j^k e ρ_j^k definido em (3-44). (ii) Analogamente ao Teorema 3.1, é possível mostrar que essa mudança não afeta a boa definição do método. (iii) Uma escolha apropriada para o parâmetro r_j^k , será proposta na Seção 3.2.3. (iv) Se F_j é estritamente convexa, então $\rho_j^k = 1/s_k^T \gamma_j^k$ e, conseqüentemente, (3-43) recupera a atualização BFGS modificada escalar (2-16) para a função F_j . Isso também ocorre no caso escalar ($m = 1$), pois a condição de curvatura (1-13) se reduz à condição clássica (1-10), implicando que $s_k^T \gamma_1^k > 0$ e, em consequência, $\rho_1^k = 1/s_k^T \gamma_1^k$. Nesse caso, o Algoritmo 3.2 se torna o método BFGS modificado com busca linear de Wolfe padrão proposto em [46].

3.2.1 Análise de convergência

Assim como no caso escalar, vamos mostrar que o método converge para problemas não convexos. As hipóteses a seguir são extensões das hipóteses utilizadas em [46].

Hipótese 3.3. (i) F é continuamente diferenciável e $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n \mid F(x) \leq F(x^0)\}$ é limitado. (ii) Existe $\varepsilon > 0$ tal que, para todo $j = 1, \dots, m$ e $k \geq 0$, tem-se $\eta_j^k + r_j^k > \varepsilon$, em que $\eta_j^k := s_k^T y_j^k / \|s_k\|^2$. (iii) Para todo $j = 1, \dots, m$, o gradiente ∇F_j é L -Lipschitz contínuo, isto é,

$$\|\nabla F_j(x) - \nabla F_j(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \quad (3-45)$$

Observe que

$$|\eta_j^k| = \frac{|s_k^T y_j^k|}{\|s_k\|^2} \leq \frac{\|s_k\| \|y_j^k\|}{\|s_k\|^2} = \frac{\|\nabla F_j(x^{k+1}) - \nabla F_j(x^k)\|}{\|s_k\|} \stackrel{(3-45)}{\leq} L.$$

Portanto, uma vez que $r_j^k \in [0, C]$, temos

$$\varepsilon < \eta_j^k + r_j^k \leq L + C, \quad (3-46)$$

para todo $j = 1, \dots, m$, e $k \geq 0$.

O lema a seguir, inspirado por [46, Lema 3.1], fornece um importante resultado para a análise de convergência do Algoritmo 3.2.

Lema 3.4. *Suponha que a Hipótese 3.3 vale e que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.2. Então, para todo $j = 1, \dots, m$ e $k \geq 0$, temos*

- (i) $\varepsilon \|s_k\|^2 \leq s_k^T \gamma_j^k \leq (L + C) \|s_k\|^2$;
- (ii) $\|\gamma_j^k\| \leq (L + C) \|s_k\|$.

Demonstração. Primeiro, pela definição de γ_j^k e de η_j^k , temos

$$s_k^T \gamma_j^k = \left(\frac{s_k^T y_j^k}{\|s_k\|^2} + r_j^k \right) \|s_k\|^2 = (\eta_j^k + r_j^k) \|s_k\|^2.$$

Logo,

$$\varepsilon \|s_k\|^2 \stackrel{(3-46)}{\leq} (\eta_j^k + r_j^k) \|s_k\|^2 \stackrel{(3-46)}{\leq} (L + C) \|s_k\|^2,$$

para todo $j = 1, \dots, m$ e $k \geq 0$, mostrando o item (i). Para provar o item (ii), note que novamente pela definição de γ_j^k , temos

$$\|\gamma_j^k\| = \|y_j^k + r_j^k s_k\| \leq \|y_j^k\| + r_j^k \|s_k\| = \left(\frac{\|y_j^k\|}{\|s_k\|} + r_j^k \right) \|s_k\| \leq (L + C) \|s_k\|,$$

para todo $j = 1, \dots, m$ e $k \geq 0$, em que a última desigualdade segue da Hipótese 3.3(iii). ■

Note que a Hipótese 3.3 garante que $\mathcal{L}(x^0)$ é compacto. Assim, F é limitada inferiormente em $\mathcal{L}(x^0)$, satisfazendo trivialmente as hipóteses da Proposição 1.2 e do Teorema 3.1. Além disso, como consequência do resultado anterior, temos que $s_k^T \gamma_j^k > 0$ e assim $\rho_j^k = 1 / (s_k^T y_j^k + r_j^k \|s_k\|^2)$, para todo $j = 1, \dots, m$ e $k \geq 0$. Assim, sob a Hipótese 3.3, a fórmula de atualização (3-43) recupera a estrutura de atualização escalar dada por

$$B_j^{k+1} = B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{\gamma_j^k (\gamma_j^k)^T}{s_k^T \gamma_j^k}. \quad (3-47)$$

Apesar da fórmula de atualização (3-47), por definição, ser distinta de (2-24), suas estruturas são iguais. Além disso, é fácil ver que a validade do Teorema 2.2 não depende da estrutura particular de y_j^k . Logo, se as hipóteses do Teorema 2.2 são satisfeitas, então o resultado é válido mesmo substituindo (2-24) por (3-47). Por fim, para mostrar que as hipóteses do Teorema 2.2 são válidas nesse contexto, basta

notar que, pelo Lema 3.4, temos

$$\frac{s_k^T \gamma_j^k}{\|s_k\|^2} \geq \frac{\varepsilon \|s_k^T\|^2}{\|s_k\|^2} = \varepsilon =: a,$$

e

$$\frac{\|\gamma_j^k\|^2}{s_k^T \gamma_j^k} \leq \frac{\|\gamma_j^k\|^2}{\varepsilon \|s_k\|^2} \leq \frac{(L+C)^2 \|s_k\|^2}{\varepsilon \|s_k\|^2} = \frac{(L+C)^2}{\varepsilon} =: b,$$

para todo $j = 1, \dots, m$ e $k \geq 0$.

Teorema 3.8. *Suponha que a Hipótese 3.3 vale e que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.2. Então,*

$$\liminf_{k \rightarrow \infty} \|d_{SD}(x^k)\| = 0.$$

Demonstração. Pelo Teorema 2.2, existem uma constante $\delta > 0$ e $K \subseteq \mathbb{N}$ tais que

$$\cos \beta_j^k \geq \delta, \quad \forall k \in K \quad \text{e} \quad \forall j = 1, \dots, m.$$

Logo, a Proposição 2.1 implica que

$$\mathcal{D}(x^k, d^k) \leq -\frac{\delta}{2} \|d^k\| \|d_{SD}(x^k)\|, \quad \forall k \in K.$$

Portanto, pela Proposição 1.2, temos

$$\infty > \sum_{k \geq 0} \frac{\mathcal{D}(x^k, d^k)^2}{\|d^k\|^2} \geq \sum_{k \in K} \frac{\mathcal{D}(x^k, d^k)^2}{\|d^k\|^2} \geq \sum_{k \in K} \frac{\delta^4}{4} \|d_{SD}(x^k)\|^2,$$

e, conseqüentemente,

$$\lim_{k \in K} d_{SD}(x^k) = 0.$$

■

3.2.2 Convergência local superlinear

Nesta seção, estudamos as propriedades de convergência local do Algoritmo 3.2. Observe que o Teorema 3.8 não garante necessariamente a convergência de $\{x^k\}$, o que será assumido aqui. As hipóteses consideradas estão formalmente descritas a seguir.

Hipótese 3.4. (i) F é duas vezes continuamente diferenciável. (ii) A sequência $\{x^k\}$ gerada pelo Algoritmo 3.1 converge para um ponto Pareto ótimo local x^* . (iii)

Para cada $j = 1, \dots, m$, r_j^k é tal que

$$\sum_{k \geq 0} r_j^k < \infty.$$

(iv) Para cada $j = 1, \dots, m$, $\nabla^2 F_j(x^*)$ é definida positiva e Hölder contínua em x^* . Portanto, existe uma vizinhança U de x^* e existem constantes positivas μ , L , M_2 e ν tais que

$$\mu \|z\|^2 \leq z^T \nabla^2 F_j(x) z \leq L \|z\|^2, \quad (3-48)$$

e

$$\|\nabla^2 F_j(x) - \nabla^2 F_j(x^*)\| \leq M_2 \|x - x^*\|^\nu, \quad (3-49)$$

para todo $j = 1, \dots, m$, $z \in \mathbb{R}^n$ e $x \in U$.

Essencialmente, a Hipótese 3.4 coincide com a Hipótese 3.2 acrescida da condição sobre r_j^k . O item (iv) estabelece que F é fortemente convexa em uma vizinhança U de x^* , e as Hessianas $\nabla^2 F_j$ ($j = 1, \dots, m$) são Hölder contínuas em x^* . Ao longo desta seção, assumiremos, sem perda de generalidade, que $\{x^k\} \subset U$, isto é, (3-48) e (3-49) valem em x^k para todo $k \geq 0$. Assim, a Hipótese 3.4 é suficiente para garantir a validade da Hipótese 3.3. De fato, para o item (i) da Hipótese 3.3, basta notar que a convexidade forte de F em U garante que o conjunto de nível $\mathcal{L}(x^0)$ é limitado. Além disso, pela definição de γ_j^k , temos

$$\begin{aligned} \frac{s_k^T \gamma_j^k}{\|s_k\|^2} &= \frac{s_k^T (y_j^k + r_j^k s_k)}{\|s_k\|^2} = \frac{s_k^T y_j^k}{\|s_k\|^2} + \frac{r_j^k \|s_k\|^2}{\|s_k\|^2} \stackrel{(3-5)}{=} \frac{s_k^T \bar{G}_j^k s_k}{\|s_k\|^2} + r_j^k \\ &\stackrel{(3-48)}{\geq} \mu + r_j^k \geq \mu, \quad \forall k \geq 0 \text{ e } j = 1, \dots, m. \end{aligned} \quad (3-50)$$

Logo, tomando $\varepsilon := \mu/2$ e considerando que $r_j^k \geq 0$, obtemos $\eta_j^k + r_j^k \stackrel{(3-50)}{\geq} \mu + r_j^k > \varepsilon$, mostrando o item (ii). O último item segue pela segunda desigualdade em (3-48). Portanto, sob a Hipótese 3.4, os resultados estabelecidos na Seção 3.2.1 permanecem válidos.

Agora, como o objetivo desta seção é provar que o Algoritmo 3.2 converge superlinearmente, vamos proceder como na Seção 3.1.2, mostrando o limite de Dennis–Moré. Iniciaremos estabelecendo que a velocidade com que a sequência $\{x^k\}$ converge para x^* é rápida o suficiente para garantir (3-9). Esse resultado está formalmente descrito no teorema a seguir. A prova é análoga à demonstração do Teorema 3.3 e será omitida.

Teorema 3.9. *Suponha que a Hipótese 3.4 vale e que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.2. Então, $\{x^k\}$ converge R -linearmente para x^* e, consequente-*

mente, vale

$$\sum_{k \geq 0} \|x^k - x^*\|^\nu < \infty$$

e

$$\sum_{k \geq 0} \varepsilon_k < \infty,$$

em que $\varepsilon_k = \max\{\|x^k - x^*\|^\nu, \|x^{k+1} - x^*\|^\nu\}$.

O resultado a seguir estabelece o limite de Dennis-Moré para o Algoritmo 3.2.

Teorema 3.10. *Suponha que a Hipótese 3.4 é válida. Então,*

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

Além disso, as sequências $\{\|B_j^k\|\}$ e $\{\|H_j^k\|\}$ são limitadas, para todo $j = 1, \dots, m$.

Demonstração. Inicialmente, para cada $j \in 1, \dots, m$ e $k \geq 0$, definimos

$$\begin{aligned} \tilde{s}_k &:= \nabla^2 F_j(x^*)^{1/2} s_k, & \tilde{y}_k &:= \nabla^2 F_j(x^*)^{-1/2} y_j^k, & \tilde{a}_k &:= \frac{\tilde{s}_k^T \tilde{\gamma}_k}{\|\tilde{s}_k\|^2} & \text{e} & \tilde{b}_k &:= \frac{\|\tilde{\gamma}_k\|^2}{\tilde{s}_k^T \tilde{\gamma}_k} \\ \tilde{B}^k &:= \nabla^2 F_j(x^*)^{-1/2} B_j^k \nabla^2 F_j(x^*)^{-1/2}, & \cos \tilde{\beta}_k &:= \frac{\tilde{s}_k^T \tilde{B}^k \tilde{s}_k}{\|\tilde{s}_k\| \|\tilde{B}^k \tilde{s}_k\|}, & \tilde{q}_k &:= \frac{\tilde{s}_k^T \tilde{B}^k \tilde{s}_k}{\|\tilde{s}_k\|^2} & \text{e} \\ \tilde{\gamma}_k &:= \nabla^2 F_j(x^*)^{-1/2} \gamma_j^k = \tilde{y}_k + r_j^k \nabla^2 F_j(x^*)^{-1} \tilde{s}_k. \end{aligned}$$

Note que $\tilde{s}_k^T \tilde{\gamma}_k = s_k^T \gamma_j^k \stackrel{(3-50)}{>} 0$ para todo $j = 1, \dots, m$ e $k \geq 0$, logo, \tilde{b}_k está bem definido. Agora, dado um j fixado, porém arbitrário, a condição (3-5) garante que

$$y_j^k - \nabla^2 F_j(x^*) s_k = (\bar{G}_j^k - \nabla^2 F_j(x^*)) s_k.$$

Tomando $\bar{G}_k := \bar{G}_j^k$, temos

$$\tilde{y}_k - \tilde{s}_k = \nabla^2 F(x^*)^{-1/2} (\bar{G}_k - \nabla^2 F(x^*)) \nabla^2 F(x^*)^{-1/2} \tilde{s}_k.$$

Logo, por Cauchy-Schwarz

$$\begin{aligned} \|\tilde{y}_k - \tilde{s}_k\| &\leq \|\nabla^2 F(x^*)^{-1/2}\|^2 \|\bar{G}_k - \nabla^2 F(x^*)\| \|\tilde{s}_k\| \\ &\leq \|\nabla^2 F(x^*)^{-1/2}\|^2 \|\tilde{s}_k\| \int_0^1 \|\nabla^2 F_j(x^k + \tau \alpha_k d^k) - \nabla^2 F(x^*)\| d\tau \\ &\stackrel{(3-49)}{\leq} \|\nabla^2 F(x^*)^{-1/2}\|^2 \|\tilde{s}_k\| \int_0^1 M_2 \|x^k + \tau \alpha_k d^k - x^*\|^\nu d\tau \\ &\leq \|\nabla^2 F(x^*)^{-1/2}\|^2 \|\tilde{s}_k\| M_2 \max\{\|x^{k+1} - x^*\|^\nu, \|x^k - x^*\|^\nu\}. \end{aligned}$$

Desse modo, definindo $\bar{c} := M_2 \|\nabla^2 F(x^*)^{-1/2}\|^2$, obtemos

$$\frac{\|\tilde{y}_k - \tilde{s}_k\|}{\|\tilde{s}_k\|} \leq \bar{c}\varepsilon_k \implies \|\tilde{y}_k\|^2 - 2\tilde{s}_k^T \tilde{y}_k + \|\tilde{s}_k\|^2 \leq \bar{c}^2 \varepsilon_k^2 \|\tilde{s}_k\|^2, \quad (3-51)$$

em que ε_k foi definido no Teorema 3.9. Além disso, como $|\|\tilde{y}_k\| - \|\tilde{s}_k\|| \leq \|\tilde{y}_k - \tilde{s}_k\|$, temos

$$(1 - \bar{c}\varepsilon_k)\|\tilde{s}_k\| \leq \|\tilde{y}_k\| \leq \|\tilde{s}_k\|(1 + \bar{c}\varepsilon_k). \quad (3-52)$$

Assim,

$$\begin{aligned} (1 - \bar{c}\varepsilon_k)^2 \|\tilde{s}_k\|^2 - 2\tilde{s}_k^T \tilde{y}_k + \|\tilde{s}_k\|^2 &\stackrel{(3-52)}{\leq} \|\tilde{y}_k\|^2 - 2\tilde{s}_k^T \tilde{y}_k + \|\tilde{s}_k\|^2 \\ &\stackrel{(3-51)}{\leq} \bar{c}^2 \varepsilon_k^2 \|\tilde{s}_k\|^2. \end{aligned}$$

Consequentemente,

$$2\tilde{s}_k^T \tilde{y}_k \geq (1 - 2\bar{c}\varepsilon_k + \bar{c}^2 \varepsilon_k^2 + 1 - \bar{c}^2 \varepsilon_k^2) \|\tilde{s}_k\|^2 = 2(1 - \bar{c}\varepsilon_k) \|\tilde{s}_k\|^2.$$

Segue, pela expressão acima e pelas definições de \tilde{a}_k e \tilde{b}_k , que

$$\begin{aligned} \tilde{a}_k &= \frac{\tilde{s}_k^T \tilde{\gamma}_k}{\|\tilde{s}_k\|^2} = \frac{\tilde{s}_k^T (\tilde{y}_k + r_j^k \nabla^2 F_j(x^*)^{-1} \tilde{s}_k)}{\|\tilde{s}_k\|^2} \geq \frac{\tilde{s}_k^T \tilde{y}_k}{\|\tilde{s}_k\|^2} + r_j^k \frac{\tilde{s}_k^T \nabla^2 F_j(x^*)^{-1} \tilde{s}_k}{\|\tilde{s}_k\|^2} \\ &\stackrel{(3-48)}{\geq} 1 - \bar{c}\varepsilon_k + \frac{r_j^k}{L} \geq 1 - \bar{c}\varepsilon_k. \end{aligned} \quad (3-53)$$

e

$$\begin{aligned} \tilde{b}_k &= \frac{\|\tilde{y}_k + r_j^k \nabla^2 F_j(x^*)^{-1} \tilde{s}_k\|^2}{\tilde{s}_k^T \tilde{\gamma}_k} \stackrel{(3-53)}{\leq} \frac{1}{1 - \bar{c}\varepsilon_k} \frac{\|\tilde{y}_k\|^2}{\|\tilde{s}_k\|^2} + \frac{2r_j^k \tilde{y}_k^T \nabla^2 F_j(x^*)^{-1} \tilde{s}_k}{\tilde{s}_k^T \tilde{\gamma}_k} \\ &\quad + \frac{1}{1 - \bar{c}\varepsilon_k} \frac{(r_j^k)^2 \|\nabla^2 F_j(x^*)^{-1}\|^2 \|\tilde{s}_k\|^2}{\|\tilde{s}_k\|^2} \\ &\stackrel{(3-52)}{\leq} \frac{(1 + \bar{c}\varepsilon_k)^2}{1 - \bar{c}\varepsilon_k} + \frac{2r_j^k \|y_j^k\| \|\nabla^2 F_j(x^*)^{-1/2}\| \|\nabla^2 F_j(x^*)^{-1}\| \|\nabla^2 F_j(x^*)^{1/2}\| \|s_k\|}{s_k^T \gamma_j^k} \\ &\quad + \frac{(r_j^k)^2 \|\nabla^2 F_j(x^*)^{-1}\|^2}{1 - \bar{c}\varepsilon_k} \\ &\stackrel{(3-50)}{\leq} \frac{1 + 2\bar{c}\varepsilon_k + \bar{c}^2 \varepsilon_k^2}{1 - \bar{c}\varepsilon_k} + 2M_3 r_j^k \frac{\|y_j^k\| \|s_k\|}{\mu \|s_k\|^2} + \frac{(M_3 r_j^k)^2}{1 - \bar{c}\varepsilon_k} \\ &\leq 1 + \frac{3\bar{c} + \bar{c}^2 \varepsilon_k}{1 - \bar{c}\varepsilon_k} \varepsilon_k + \frac{2M_3 L r_j^k}{\mu} + \frac{(M_3 r_j^k)^2}{1 - \bar{c}\varepsilon_k}, \end{aligned}$$

em que $M_3 := \|\nabla^2 F_j(x^*)^{-1}\|$. Agora, como $\{x^k\}$ converge para x^* , temos $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Logo, para k suficientemente grande, podemos assumir que $\bar{c}\varepsilon_k < \frac{1}{2}$,

além disso, temos que existe uma constante $c > \max\{3\bar{c}, \frac{2M_3L}{\mu}\}$ tal que

$$\tilde{b}_k \leq 1 + c\varepsilon_k + cr_j^k + 2(M_3r_j^k)^2. \quad (3-54)$$

Assim,

$$\ln(\tilde{a}_k) \stackrel{(3-53)}{\geq} \ln(1 - \bar{c}\varepsilon_k) \stackrel{(3-24)}{\geq} \frac{-\bar{c}\varepsilon_k}{1 - \bar{c}\varepsilon_k} \geq -2\bar{c}\varepsilon_k \geq -2c\varepsilon_k. \quad (3-55)$$

Agora, multiplicando ambos os lados (direito e esquerdo) da fórmula de atualização (3-47) por $(\nabla^2 F_j(x^*))^{-1/2}$ e agrupando os termos de modo apropriado, obtemos

$$\tilde{B}^{k+1} = \tilde{B}^k - \frac{\tilde{B}^k \tilde{s}_k \tilde{s}_k^T \tilde{B}^k}{\tilde{s}_k^T \tilde{B}^k \tilde{s}_k} + \frac{\tilde{\gamma}_k \tilde{\gamma}_k^T}{\tilde{s}_k^T \tilde{\gamma}_k}.$$

Dessa maneira, podemos deduzir de (2-34) que

$$\begin{aligned} 0 < \psi(\tilde{B}^{k+1}) &= \psi(\tilde{B}^k) + \left[\tilde{b}_k - \ln(\tilde{a}_k) - 1 \right] + \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k} + \ln\left(\frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k}\right) \right] \\ &\quad + \ln(\cos^2 \tilde{\beta}_k) \\ &\leq \psi(\tilde{B}^k) + 3c\varepsilon_k + cr_j^k + 2(M_3)^2(r_j^k)^2 + \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k} + \ln\left(\frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k}\right) \right] \\ &\quad + \ln(\cos^2 \tilde{\beta}_k), \end{aligned}$$

em que a última desigualdade segue por (3-54) e (3-55). Somando essa expressão para $k \geq 0$ e utilizando a Hipótese 3.4 e o Teorema 3.9, obtemos que

$$\begin{aligned} \sum_{k \geq 0} \left(\ln\left(\frac{1}{\cos^2 \tilde{\beta}_k}\right) - \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k} + \ln\left(\frac{\tilde{q}_k}{\cos^2 \tilde{\beta}_k}\right) \right] \right) &\leq \\ &\leq \psi(\tilde{B}^0) + 3c \sum_{k \geq 0} \varepsilon_k + c \sum_{k \geq 0} r_j^k + 2(M_3)^2 \sum_{k \geq 0} (r_j^k)^2 < \infty. \end{aligned}$$

A parte essencial para provar o resultado já está estabelecida. Vamos omitir o restante da prova, pois é análogo à parte final da demonstração do Teorema 3.4. ■

O limite de Dennis-Moré, provado no Teorema 3.10, é a propriedade chave para se obter a convergência superlinear. O leitor pode observar que os demais resultados apresentados na Seção 3.1.2 podem ser facilmente reproduzidos para o Algoritmo 3.2. O teorema a seguir sintetiza os resultados de convergência do Algoritmo 3.2.

Teorema 3.11. *Suponha que a Hipótese 3.4 seja válida. Então, o comprimento de passo $\alpha_k = 1$ satisfaz as condições de Wolfe padrão (1-12)–(1-13) para k suficientemente grande e $\{x^k\}$ converge para x^* com taxa Q -superlinear.*

3.2.3 Um método mBFGS praticável

Nesta seção propomos uma escolha praticável do parâmetro r_j^k e mostramos que essa escolha é suficiente para garantir que o Algoritmo 3.2, sob hipóteses razoáveis, converge com taxa superlinear. Seja $\chi^k := (\chi_1^k, \dots, \chi_m^k) \in \mathbb{R}_+^m$ tal que $\|\chi^k\|_1 = \sum_{j=1}^m \chi_j^k = 1$ e considere a função escalar $F_{\chi^k} : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por:

$$F_{\chi^k}(x) := \sum_{j=1}^m \chi_j^k F_j(x).$$

Para cada $j = 1, \dots, m$ e $k \geq 0$, definimos o parâmetro r_j^k como:

$$r_j^k = \max\{-\eta_j^k, 0\} + \|\nabla F_{\chi^k}(x^k)\|, \quad (3-56)$$

em que $\eta_j^k = s_k^T y_j^k / \|s_k\|^2$. Diferentes escolhas de χ^k geram diferentes definições do parâmetro. Sob a Hipótese 3.3(i), existe $C > 0$ tal que r_j^k , dado em (3-56), satisfaz $r_j^k \in [0, C]$ para todo $j = 1, \dots, m$ e $k \geq 0$. De fato, o conjunto $\mathcal{L}(x^0)$ é compacto e, conseqüentemente, existe uma constante $\bar{M} \in \mathbb{R}$ tal que $\|\nabla F_{\chi^k}(x^k)\| \leq \bar{M}$, para todo $k \geq 0$. Assim,

$$r_j^k \leq |\eta_j^k| + \|\nabla F_{\chi^k}(x^k)\| \stackrel{(3-45)}{\leq} L + \|\nabla F_{\chi^k}(x^k)\| \leq L + \bar{M} =: C.$$

O seguinte resultado prova que o Algoritmo 3.2 converge globalmente, se o parâmetro r_j^k é escolhido como em (3-56).

Teorema 3.12. *Suponha que os parâmetros r_j^k , para todo $j = 1, \dots, m$ e $k \geq 0$, são escolhidos como em (3-56), para algum $(\chi_1^k, \dots, \chi_m^k) \in \mathbb{R}_+^m$ tal que $\sum_{j=1}^m \chi_j^k = 1$. Se os itens (i) e (iii) da Hipótese 3.3 são válidos, então a sequência $\{x^k\}$, gerada pelo Algoritmo 3.2, satisfaz*

$$\liminf_{k \rightarrow \infty} \|d_{SD}(x^k)\| = 0.$$

Demonstração. Suponha, por contradição, que existe $\varepsilon > 0$ tal que $\|d_{SD}(x^k)\| \geq \varepsilon$ para todo $k \geq 0$. Logo, pelo Lema 1.2(iv), obtemos

$$\eta_j^k + r_j^k \geq \|\nabla F_{\chi^k}(x^k)\| \geq \|d_{SD}(x^k)\| \geq \varepsilon, \quad \forall j = 1, \dots, m \quad \text{e} \quad k \geq 0.$$

Assim, o item (ii) da Hipótese 3.3 é válido. Portanto, pelo Teorema 3.8, temos uma contradição. ■

A seguir, discutimos possíveis escolhas de χ^k de modo a obter taxa de convergência superlinear. Uma escolha natural é definir como $\chi^k = (\lambda_1^{SD}(x^k), \dots, \lambda_m^{SD}(x^k))$, veja (1-6). Contudo, do ponto de vista computacional, essa escolha pode *encarecer* o método, pois o subproblema (1-5) deve ser resolvido a cada iteração. Alternativamente, podemos usar os multiplicadores do subproblema (2-20), que são calculados no Passo 1 do Algoritmo 3.2, a cada iteração.

Teorema 3.13. *Suponha que os parâmetros r_j^k , para todo $j = 1, \dots, m$ e $k \geq 0$, são escolhidos como em (3-56), com $\chi^k = (\lambda_1^{SD}(x^k), \dots, \lambda_m^{SD}(x^k))$, veja (1-6). Se os itens (i), (ii) e (iv) da Hipótese 3.4 são válidos, então a sequência $\{x^k\}$ converge com taxa Q-superlinear.*

Demonstração. Primeiro, perceba que, para todo $j = 1, \dots, m$ e k grande o suficiente, a condição (3-48) garante que

$$\eta_j^k = \frac{s_k^T y_j^k}{\|s_k\|^2} \stackrel{(3-5)}{=} \frac{1}{\|s_k\|^2} s_k^T \int_0^1 \nabla^2 F_j(x^k + \tau s_k) d\tau s_k \stackrel{(3-6)}{\geq} \mu > 0. \quad (3-57)$$

Assim, pelo Lema 1.2(i), temos

$$r_j^k \stackrel{(3-57)}{=} \|\nabla F_{SD}(x^k)\| \stackrel{(1-7)}{=} \|d_{SD}(x^k)\| = \|d_{SD}(x^k)\| - \|d_{SD}(x^*)\| \leq L\|x^k - x^*\|,$$

em que a última desigualdade segue do fato de $\|d_{SD}(\cdot)\|$ ser Lipschitz contínua devido à condição (3-48) (veja [74, Teorema 3.1]). Portanto, pelo Teorema 3.9, segue que $\{r_j^k\}$, para $j = 1, \dots, m$, satisfaz o item (iii) da Hipótese 3.4 e, conseqüentemente, a taxa de convergência é Q-superlinear, veja a Seção 3.2.2. ■

Para a escolha de χ^k como os multiplicadores de Lagrange de (2-20), necessitamos de uma hipótese adicional.

Teorema 3.14. *Suponha que os parâmetros r_j^k , para todo $j = 1, \dots, m$ e $k \geq 0$, são escolhidos como em (3-56) com $\chi^k = (\lambda_1(x^k), \dots, \lambda_m(x^k))$, veja (2-22). Se os itens (i), (ii) e (iv) da Hipótese 3.4 são válidos e as sequências $\{\|B_j^k\|\}$, $\{\|H_j^k\|\}$ são limitadas para todo $j = 1, \dots, m$, então a sequência $\{x^k\}$ converge com taxa Q-superlinear.*

Demonstração. Uma vez que $\{\|B_j^k\|\}$ e $\{\|H_j^k\|\}$ são limitadas, então existem constantes positivas $a, b \in \mathbb{R}$ tais que $aI_n \prec B_j^k \prec bI_n$ para todo $j = 1, \dots, m$ e

$k \geq 0$. Assim, pelo Lema 2.2, temos

$$\frac{a}{b} \|\nabla F_{\chi^k}(x^k)\| \leq \|d_{SD}(x^k)\|.$$

Logo,

$$r_j^k \stackrel{(3-57)}{=} \|\nabla F_{\chi^k}(x^k)\| \leq \frac{b}{a} \|d_{SD}(x^k)\| = \frac{b}{a} \left(\|d_{SD}(x^k)\| - \|d_{SD}(x^*)\| \right) \leq \frac{Lb}{a} \|x^k - x^*\|,$$

em que a última desigualdade segue de [74, Teorema 3.1]. De modo análogo ao Teorema 3.13, a sequência $\{x^k\}$ converge a uma taxa Q-superlinear. ■

3.3 Um método BFGS cauteloso para problemas não convexos

Nesta seção, estendemos para otimização multiobjetivo o método proposto em [47], veja a Observação 2.3. Assim como no caso escalar, estamos interessados em estabelecer convergência global do método BFGS em problemas não convexos. O algoritmo proposto não necessariamente atualiza as aproximações B_j^k em todas as iterações. A atualização ocorre quando uma condição cautelosa é satisfeita. Diferentemente do caso escalar, nesta proposta calculamos comprimentos de passos que satisfazem as condições de Wolfe padrão (1-12)–(1-13), o que nos permitirá utilizar a condição de Zoutendijk (1-17) para a análise de convergência do método.

Algoritmo 3.3. Um método BFGS cauteloso (cBFGS)

Sejam $c_1 \in (0, 1/2)$, $c_2 \in (c_1, 1)$, $x^0 \in \mathbb{R}^n$, $\varepsilon, \omega \in \mathbb{R}_{++}$ e $B_j^0 \succ 0$, para todo $j = 1, \dots, m$. Inicialize $k \leftarrow 0$.

Passo 1. Calcule a direção de busca

Calcule $d^k := d(x^k)$ e $\theta(x^k)$ como em (2-18) e (2-19), respectivamente.

Passo 2. Critério de parada

Se $\theta(x^k) = 0$, então PARE.

Passo 3. Procedimento de busca linear

Calcule o comprimento de passo $\alpha_k > 0$ (primeiro tente $\alpha_k = 1$) tal que as condições de Wolfe padrão (1-12)–(1-13) sejam satisfeitas e faça $x^{k+1} := x^k + \alpha_k d^k$.

Passo 4. Prepare a próxima iteração

Seja $d_{SD}(x^k)$ a direção de máxima descida dada pela solução de (1-4) para x^k .

Para cada $j = 1, \dots, m$, defina

$$B_j^{k+1} = \begin{cases} B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{y_j^k (y_j^k)^T}{s_k^T y_j^k}, & \text{se } \min_{j=1, \dots, m} \frac{s_k^T y_j^k}{\|s_k\|^2} \geq \varepsilon \|d_{SD}(x^k)\|^\omega; \\ B_j^k, & \text{caso contrário,} \end{cases} \quad (3-58)$$

em que $y_j^k = \nabla F_j(x^{k+1}) - \nabla F_j(x^k)$ e $s_k = x^{k+1} - x^k$. Faça $k \leftarrow k + 1$ e vá ao

Passo 1.

Observação 3.5. (i) O algoritmo impõe uma cautelosa condição para atualizar as matrizes B_j^k . O esquema (3-58) atualiza as matrizes de modo simultâneo, isto é, se para um dos objetivos a condição $s_k^T y_j^k / \|s_k\|^2 \geq \varepsilon \|d_{SD}(x^k)\|^\omega$ falha, então nenhuma das m matrizes é atualizada. Essa é uma diferença significativa em relação

a [65], em que uma condição cautelosa é tratada em cada objetivo de maneira independente, veja (2-40). (ii) Diferentemente dos Algoritmos 3.1 e 3.2, a boa definição do Algoritmo 3.3 não depende explicitamente da escolha do comprimento de passo, pois as matrizes B_j^k , para $j = 1, \dots, m$, sempre se mantêm definidas positivas. (iii) No caso escalar ($m = 1$), o Algoritmo 3.3 se reduz ao algoritmo proposto em [47].

3.3.1 Análise de convergência

Nesta seção, apresentamos os resultados de convergência global para o Algoritmo 3.3. Assim como no caso escalar [47], não assumimos que as funções objetivo são fortemente convexas. Em contrapartida, supomos que o conjunto de nível $\mathcal{L}(x^0)$ é limitado.

Hipótese 3.5. (i) O conjunto de nível $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n \mid F(x) \leq F(x^0)\}$ é limitado e F é continuamente diferenciável em um aberto $\mathcal{N} \supset \mathcal{L}(x^0)$. (ii) Para $j = 1, \dots, m$, o gradiente ∇F_j é L -Lipschitz contínuos em \mathcal{N} , isto é,

$$\|\nabla F_j(x) - \nabla F_j(y)\| \leq L\|x - y\|, \quad (3-59)$$

para todo $j = 1, \dots, m$ e $x, y \in \mathcal{N}$.

Note que o conjunto de nível $\mathcal{L}(x^0)$ é compacto, assim a função objetivo F é limitada inferiormente em $\mathcal{L}(x^0)$. Logo, como os comprimentos de passo do Algoritmo 3.3 satisfazem as condições de Wolfe padrão (1-12)–(1-13), a condição de Zoutendijk (1-17) é válida, veja a Proposição 1.2. O seguinte teorema é uma extensão direta de [47, Teorema 3.1].

Teorema 3.15. *Suponha que a Hipótese 3.5 vale e que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.3. Se existem constantes positivas $\delta_1, \delta_2, \delta_3 \in \mathbb{R}$ e um conjunto de índices $K \subset \mathbb{N}$, com cardinalidade infinita, de modo que as relações*

$$\|B_j^k s_k\| \leq \delta_1 \|s_k\|, \quad \delta_2 \|s_k\|^2 \leq s_k^T B_j^k s_k \leq \delta_3 \|s_k\|^2, \quad (3-60)$$

valem para todo $k \in K$ e para todo $j = 1, \dots, m$, simultaneamente, então

$$\liminf_{k \rightarrow \infty} \|d_{SD}(x^k)\| = 0.$$

Demonstração. Seja $\delta(x^k)$ e β_j^k como definido na Proposição 2.1. Assim, lembrando

que $s_k = \alpha_k d^k$, temos

$$\cos \beta_j^k = \frac{(d^k)^T B_j^k d^k}{\|d^k\| \|B_j^k d^k\|} = \frac{s_k^T B_j^k s_k}{\|s_k\| \|B_j^k s_k\|} \stackrel{(3-60)}{\geq} \frac{\delta_2 \|s_k\|^2}{\|s_k\| \delta_1 \|s_k\|} = \frac{\delta_2}{\delta_1} =: \delta > 0,$$

para todo $j = 1, \dots, m$ e $k \in K$. Assim, a Proposição 2.1 implica que

$$\mathcal{D}(x^k, d^k) \leq -\frac{\delta}{2} \|d^k\| \|d_{SD}(x^k)\|, \quad \forall k \in K.$$

Portanto, pela Proposição 1.2, temos

$$\infty > \sum_{k \geq 0} \frac{\mathcal{D}(x^k, d^k)^2}{\|d^k\|^2} \geq \sum_{k \in K} \frac{\mathcal{D}(x^k, d^k)^2}{\|d^k\|^2} \geq \sum_{k \in K} \frac{\delta^4}{4} \|d_{SD}(x^k)\|^2,$$

e, conseqüentemente,

$$\lim_{k \in K} d_{SD}(x^k) = 0.$$

■

Teorema 3.16. *Suponha que a Hipótese 3.5 vale e que $\{x^k\}$ é uma seqüência gerada pelo Algoritmo 3.3. Então,*

$$\liminf_{k \rightarrow \infty} \|d_{SD}(x^k)\| = 0.$$

Demonstração. Pelo Teorema 3.15, é suficiente mostrar que existem constantes positivas $\delta_1, \delta_2, \delta_3 \in \mathbb{R}$ e um conjunto de índices $K \subseteq \mathbb{N}$, com cardinalidade infinita, tal que (3-60) seja válido para todo $j = 1, \dots, m$ e $k \in K$. Para isso, defina o seguinte conjunto

$$\bar{K} := \{i \in \mathbb{N} \mid \min_{j=1, \dots, m} \frac{s_i^T y_j^i}{\|s_i\|^2} \geq \varepsilon \|d_{SD}(x^i)\|^\omega\}.$$

Se \bar{K} é finito, então existe $k_0 \in \mathbb{N}$ tal que $B_j^k = B_j^{k_0}$ para todo $j = 1, \dots, m$ e $k \geq k_0$. Como $B_j^{k_0}$ é simétrica e definida positiva, é fácil ver que existem constantes positivas $\delta_1, \delta_2, \delta_3 \in \mathbb{R}$ tais que vale (3-60), para todo $j = 1, \dots, m$ e $k \geq k_0$.

Agora, assumamos que \bar{K} é um conjunto com infinitos elementos. Suponha, por contradição, que existe uma constante positiva $\delta \in \mathbb{R}$ tal que $\|d_{SD}(x^k)\| \geq \delta$, para todo $k \geq 0$. Logo, pelo Lema 1.2(iv) e para cada $j = 1, \dots, m$, temos

$$\frac{s_k^T y_j^k}{\|s_k\|^2} \geq \min_{j=1, \dots, m} \frac{s_k^T y_j^k}{\|s_k\|^2} \geq \varepsilon \|d_{SD}(x^k)\|^\omega \geq \varepsilon \delta^\omega, \quad \forall k \in \bar{K}.$$

Assim, por (3-59), temos

$$\frac{\|y_j^k\|^2}{s_k^T y_j^k} = \frac{\|y_j^k\|^2}{\|s_k\|^2} \leq \frac{L^2}{\varepsilon \delta^\omega}, \quad \forall k \in \bar{K}.$$

Aplicando o Teorema 2.2 na subsequência \bar{K} , existe um conjunto de índices $K \subset \bar{K}$ tal que as condições (2-26) e (2-27) valem (em consequência de (3-60)), para todo $j = 1, \dots, m$ e $k \in K$. Portanto, pelo Teorema 3.15, temos uma contradição. ■

3.3.2 Convergência local superlinear

Nesta seção, mostramos que, sob hipóteses razoáveis, o Algoritmo 3.3 se reduz ao Algoritmo 3.1 para k suficientemente grande, herdando a propriedade de convergência superlinear.

Hipótese 3.6. (i) F é duas vezes continuamente diferenciável. (ii) A sequência $\{x^k\}$ gerada pelo Algoritmo 3.3 converge para um ponto Pareto ótimo local x^* . (iii) Para cada $j = 1, \dots, m$, $\nabla^2 F_j(x^*)$ é definida positiva e Lipschitz contínua em x^* . Portanto, existe uma vizinhança U de x^* e existem constantes positivas μ , L , e L_2 tais que

$$\mu \|z\|^2 \leq z^T \nabla^2 F_j(x) z \leq L \|z\|^2, \quad (3-61)$$

e

$$\|\nabla^2 F_j(x) - \nabla^2 F_j(x^*)\| \leq L_2 \|x - x^*\|, \quad (3-62)$$

para todo $j = 1, \dots, m$, $z \in \mathbb{R}^n$ e $x \in U$.

Note que as Hipóteses 3.2 e 3.6 são análogas. Assim como na Seção 3.1.2 assumimos, sem perda de generalidade, que $\{x^k\} \subset U$, isto é, (3-61)–(3-62) valem em x^k para todo $k \geq 0$.

Teorema 3.17. *Suponha que a Hipótese 3.6 seja válida. Então, $\{x^k\}$ converge para x^* com taxa Q -superlinear.*

Demonstração. Inicialmente, similarmente a (3-7), a condição (3-61) implica que

$$\min_{j=1, \dots, m} \frac{s_k^T y_j^k}{\|s_k\|^2} \geq \mu \quad \forall k \geq 0.$$

Por outro lado,

$$\lim_{k \rightarrow \infty} \|d_{SD}(x^k)\| = 0,$$

pois $d_{SD}(\cdot)$ é contínua e $d_{SD}(x^*) = 0$. Portanto, para k grande o suficiente, temos

$$\min_{j=1,\dots,m} \frac{s_k^T y_j^k}{\|s_k\|^2} \geq \mu \geq \varepsilon \|d_{SD}(x^k)\|^\omega.$$

Logo, para k grande o suficiente, as matrizes B_j^k sempre são atualizadas e, conseqüentemente, o Algoritmo 3.3 se reduz ao Algoritmo 3.1. Portanto, a convergência superlinear segue do Teorema 3.7. ■

3.3.3 Experimentos Numéricos

Esta seção apresenta alguns resultados numéricos para ilustrar as potenciais vantagens práticas dos Algoritmos 3.2 e 3.3. Estamos interessados principalmente em verificar a eficácia desses métodos que possuem convergência global para problemas não convexos. Para tal, consideramos os seguintes métodos nos testes reportados.

- Algoritmo 3.1 (BFGS-Wolfe) e o algoritmo Padrão BFGS-Armijo, veja Seção 3.1.3.
- Algoritmo 3.2 (mBFGS-Wolfe): atualiza as aproximações Hessianas a cada iteração utilizando (3-43), em que

$$r_j^k = \max\left\{-\frac{s_k^T y_j^k}{\|s_k\|^2}, 0\right\} + \|\nabla F_{\lambda_k}(x^k)\|, \quad j = 1, \dots, m, \quad (3-63)$$

e calcula os tamanhos dos passos satisfazendo as condições de Wolfe (1-12)–(1-13).

- Algoritmo 3.3 (cBFGS-Wolfe): atualiza as aproximações Hessianas por (3-58), em que $\varepsilon = 10^{-6}$ e

$$\omega = \begin{cases} 6, & \text{se } \|d_{SD}(x^k)\| < 1, \\ 10^{-4}, & \text{caso contrário,} \end{cases}$$

e os tamanhos de passos são calculados satisfazendo as condições de Wolfe (1-12)–(1-13). A escolha desses parâmetros foi baseada no caso escalar [47].

Como discutido na Seção 3.2.3, a escolha do parâmetro r_j^k é crucial para a aplicabilidade do o Algoritmo 3.2. Em particular, testes preliminares mostraram que o Algoritmo 3.2 com o parâmetro r_j^k escolhido como em (3-63) apresentou desempenho numérico superior do que r_j^k definido em termos de $\|d_{SD}(x^k)\|$.

Nos testes numéricos reportados aqui, os parâmetros algorítmicos, o critério de parada, o cálculo das direções de busca e dos comprimentos de passo foram

considerados como na Seção 3.1.3. O conjunto de problemas teste utilizado se encontra na Tabela 3.1. Assim como na Seção 3.1.3, foram tomados 300 pontos iniciais aleatórios para cada problema.

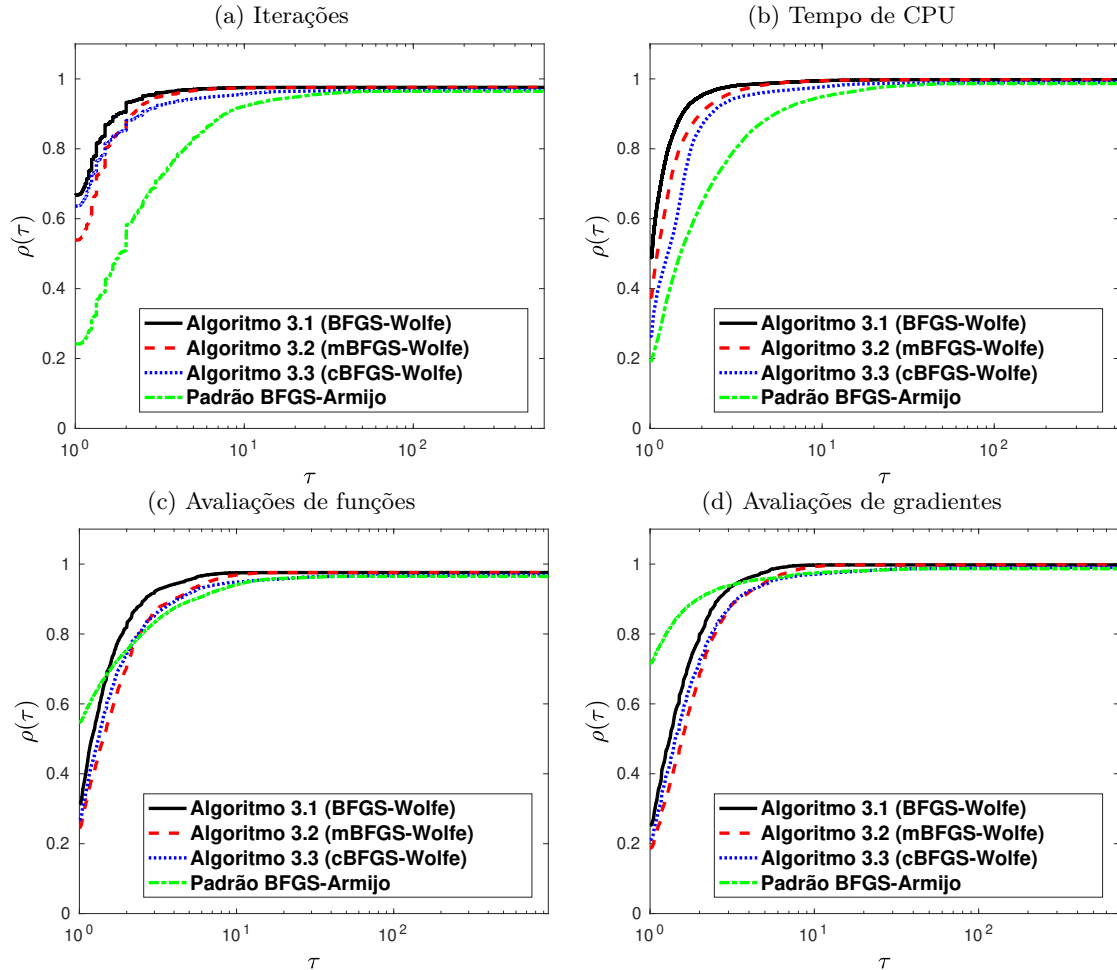


Figura 3.5: Performance profiles considerando 300 pontos iniciais para cada problema teste usando como medida de desempenho: (a) número de iterações; (b) tempo de CPU; (c) número de avaliações de funções; (d) número de avaliações de gradientes.

A Figura 3.5 mostra os resultados usando *performance profiles* [22], comparando os algoritmos com respeito a: (a) número de iterações; (b) tempo de CPU; (c) número de avaliações de funções; (d) número de avaliações de gradientes. Como pode ser observado, todos os algoritmos se mostraram robustos no conjunto de problemas escolhido. Os Algoritmos 3.1 e 3.2 resolveram 99.8% (empate) dos problemas testados, enquanto o Algoritmo 3.3 e o algoritmo Padrão BFGS Armijo resolveram 99.1% e 98.06%, respectivamente. Em relação à eficiência, considerando o número de iterações, o Algoritmo 3.1 (66.8%) teve o melhor desempenho seguido pelo Algoritmo 3.3 e 3.2 (63.6% e 53.9%, resp.) e pelo algoritmo Padrão BFGS-Armijo (24.4%), veja a

Figura 3.5(a). Isso se refletiu diretamente no tempo de CPU (eficiência de 48.9%, 37.5%, 26.4% e 19.3% para os Algoritmos 3.1, 3.2, 3.3 e o algoritmo Padrão BFGS-Armijo, respectivamente), como pode ser visto na Figura 3.5(b). Com respeito aos números de avaliações de funções e gradientes, o algoritmo Padrão BFGS-Armijo foi o mais eficiente, veja as Figuras 3.5(c) e (d). Isso era esperado, pois a busca linear de Wolfe usa mais informações dos objetivos do que a busca da linear de Armijo. Como discutido na Seção 3.1.3, a forte correlação entre o número de iterações e o tempo de CPU pode ser explicada pelo fato de que, em nossos experimentos, o custo computacional é amplamente dominado pelas soluções dos subproblemas que fornecem as direções de busca. Vale ressaltar que esse fato justifica a menor eficiência do Algoritmo 3.3 com respeito ao tempo de CPU, pois esse método requer a solução de dois subproblemas a cada iteração: (2-20) e (1-5).

A seguir, comparamos a capacidade dos algoritmos de gerar apropriadamente as fronteiras Pareto utilizando as métricas *Purity* e *Spread* (Γ e Δ). Os resultados na Figura 3.6 mostram que os Algoritmos 3.1 e 3.2 superaram o algoritmo Padrão BFGS-Armijo em relação à métrica *Purity*, enquanto nenhuma diferença significativa foi observada para a métrica *Spread* Γ e Δ .

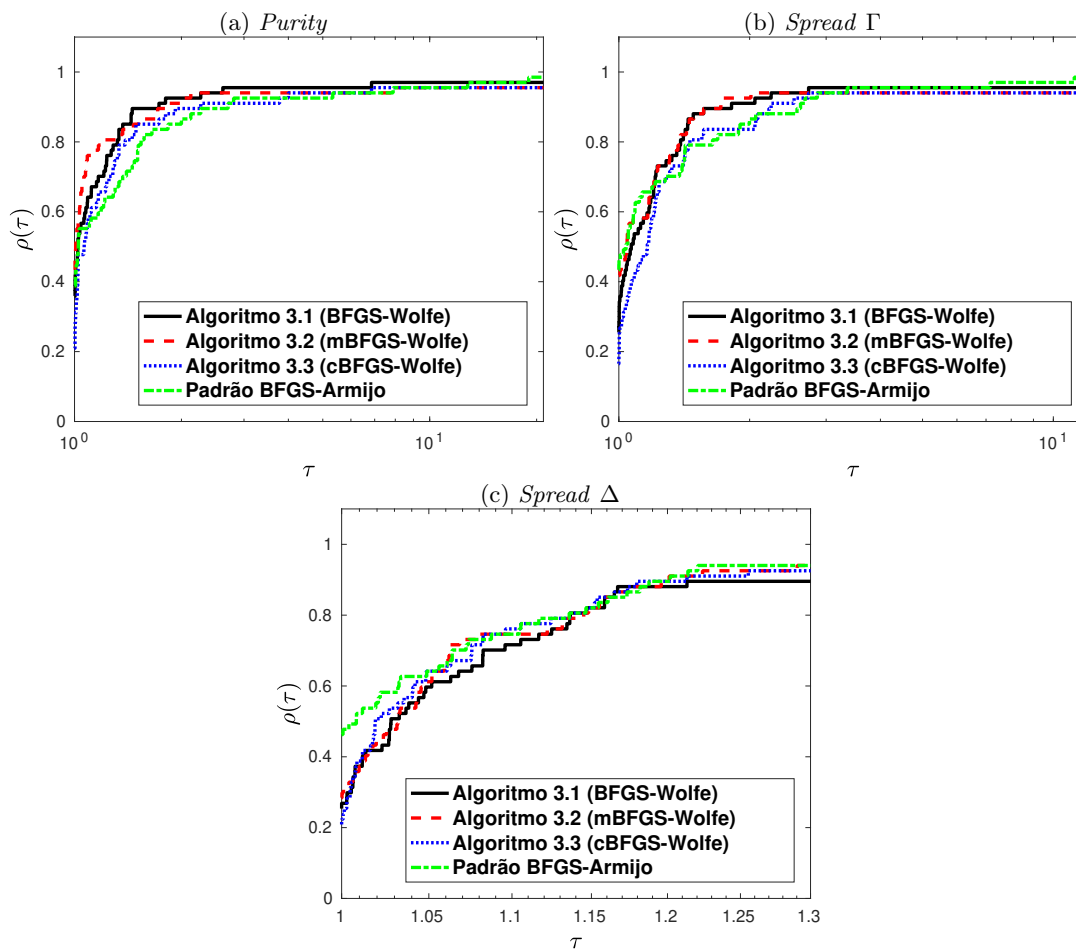


Figura 3.6: Performance profiles com respeito às métricas: (a) *Purity*; (b) *Spread Γ* ; (c) *Spread Δ* .

Convergência local superlinear não assintótica do método BFGS para otimização multiobjetivo

Neste capítulo, discutimos sobre a convergência não assintótica (explícita) do método BFGS para otimização multiobjetivo. Nas Seções 2.1 e 3.1.2, a taxa de convergência local do método BFGS, escalar ou multiobjetivo, é obtida de maneira assintótica, isto é,

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0.$$

Recentemente, surgiram alguns trabalhos na literatura que obtiveram convergência não assintótica de métodos quase-Newton escalares. Em particular, foi proposto por Rodomanov e Nesterov [67] uma versão *gananciosa* do método BFGS escalar e caracterizada de forma explícita sua convergência. Posteriormente, em [69], os mesmos autores exploraram a convergência local do método BFGS clássico e mostraram que a taxa de convergência é $\left(\frac{nL}{k\mu}\right)^{k/2}$, em que μ é o parâmetro de convexidade forte e L é a constante de Lipschitz do gradiente da função objetivo. Em [68], os autores melhoraram esse resultado, mostrando que a taxa de convergência depende do logaritmo do número de condição $\frac{L}{\mu}$. Inspirados nesses trabalhos, buscamos obter a convergência superlinear para o método BFGS multiobjetivo de maneira não assintótica. Consideraremos as hipóteses formalmente descritas a seguir.

Hipótese 4.1. (i) F é duas vezes continuamente diferenciável. (ii) Existem constantes positivas $\mu, L \in \mathbb{R}$ tais que

$$\mu I_n \preceq \nabla^2 F_j(x) \preceq L I_n, \quad (4-1)$$

para todo $j = 1, \dots, m$ e $x \in \mathbb{R}^n$. Isto é, para cada $j = 1, \dots, m$, F_j é μ -fortemente convexa e ∇F_j é L -Lipschitz contínua. (iii) Existe uma constante positiva $L_2 \in \mathbb{R}$

tal que, para cada $j = 1, \dots, m$, $\nabla^2 F_j$ é L_2 -Lipschitz contínua, isto é,

$$\|\nabla^2 F_j(x) - \nabla^2 F_j(y)\| \leq L_2 \|x - y\|, \quad (4-2)$$

para todo $j = 1, \dots, m$, $x, y \in \mathbb{R}^n$.

Observação 4.1. Seja $\chi = (\varrho_1, \dots, \varrho_m) \in \mathbb{R}_+^m$ tal que $\|\chi\|_1 = 1$. Note que, como consequência da condição (4-1), temos

$$\mu I_n \preceq \nabla^2 F_\chi(x) \preceq L I_n, \quad e \quad \frac{1}{L} I_n \preceq \nabla^2 F_\chi(x)^{-1} \preceq \frac{1}{\mu} I_n, \quad \forall x \in \mathbb{R}^n, \quad (4-3)$$

em que $\nabla^2 F_\chi(x)^{-1} = \left(\sum_{j=1}^m \varrho_j \nabla^2 F_j(x) \right)^{-1}$.

A Hipótese 4.1 é ligeiramente mais forte do que as hipóteses consideradas no Capítulo 3. Ressaltamos que essa suposição está alinhada com o caso escalar [67, 68, 69]. Nessas referências, os autores usaram o conceito de função *fortemente auto-concordante* que, até onde sabemos, não está definido para funções vetoriais. Dizemos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é fortemente auto-concordante se existe $M > 0$ tal que

$$\nabla^2 f(v) - \nabla^2 f(u) \leq M \|v - u\|_{\nabla^2 f(z)} \nabla^2 f(w),$$

para quaisquer $u, v, z, w \in \mathbb{R}^n$ em que $\|v - u\|_{\nabla^2 f(z)} = \sqrt{(v - u)^T \nabla^2 f(z) (v - u)}$. Para que o objetivo F_j , para algum j dado, seja fortemente auto-concordante, é suficiente que F_j seja duas vezes continuamente diferenciável e as condições (4-1)–(4-2) sejam válidas, veja [67, Exemplo 4.1]. Assim, assumimos condições suficientes para que os objetivos sejam fortemente auto-concordantes.

Algoritmo 4.1. Método BFGS com convergência superlinear não assintótica

Seja $x^0 \in \mathbb{R}^n$. Faça $B_j^0 = L I_n$, para todo $j = 1, \dots, m$. Inicialize $k \leftarrow 0$.

Passo 1. Calcule a direção de busca

Calcule $d^k := d(x^k)$ e $\theta(x^k)$ como em (2-18) e (2-19), respectivamente.

Passo 2. Critério de parada

Se $\theta(x^k) = 0$, então PARE.

Passo 3. Atualização

Faça $x^{k+1} := x^k + d^k$.

Passo 4. Prepare a próxima iteração

Para cada $j = 1, \dots, m$, defina

$$B_j^{k+1} = B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{y_j^k (y_j^k)^T}{s_k^T y_j^k},$$

em que $y_j^k = \nabla F_j(x^{k+1}) - \nabla F_j(x^k)$ e $s_k = x^{k+1} - x^k$. Faça $k \leftarrow k + 1$ e vá ao **Passo 1**.

Ao longo desse capítulo, faremos o uso de diversos resultados que podem ser encontrados em algum dos três trabalhos supramencionados. A seguir, definimos algumas notações que simplificam em demasiado os cálculos apresentados.

i. Seja $W \in \mathbb{R}^{n \times n}$ uma matriz simétrica e definida positiva. Dado $v \in \mathbb{R}^n$, denotamos:

$$\|v\|_W := \sqrt{v^T W v} \quad \text{e} \quad \|v\|_W^* := \sqrt{v^T W^{-1} v}. \quad (4-4)$$

ii Sejam duas matrizes simétricas $B, G \in \mathbb{R}^{n \times n}$, definimos o produto interno por

$$\langle G, B \rangle := \text{Tr}(GB). \quad (4-5)$$

O determinante de B com respeito a G é dado por: $\text{Det}(G, B) := \text{Det}(GB)$. Além disso, dizemos que ξ é o menor (resp. maior) autovalor de B em relação a G , se $\xi G \preceq B$ (resp. \succeq).

iii Sejam $B, G \in \mathbb{R}^{n \times n}$ matrizes simétricas definidas positivas, definimos a função potencial, *barreira log-det aumentada*, por:

$$\begin{aligned} \Psi(B, G) &:= \langle B^{-1}, G - B \rangle - \ln \text{Det}(B^{-1}, G) \\ &= \ln \text{Det}(G^{-1}, B) - \langle B^{-1}, B - G \rangle \geq 0, \end{aligned} \quad (4-6)$$

em que a positividade está provada em [69].

A ideia de combinar o traço e o determinante para formar uma função potencial para analisar métodos quase-Newton foi originalmente proposta em [6] (veja (2-31)). A função (4-6) foi utilizada em [68, 69] para estudar a convergência superlinear do método BFGS.

4.1 Resultados preliminares

Sob a Hipótese 4.1, o seguinte resultado mostra, em particular, que a função escalar dada por uma combinação convexa dos objetivos F_j é fortemente autoconcordante. Seja $\chi = (\varrho_1, \dots, \varrho_m) \in \mathbb{R}_+^m$ tal que $\|\chi\|_1 = 1$, denotamos a função

escalar $F_\chi : \mathbb{R}^n \rightarrow \mathbb{R}$ por:

$$F_\chi(x) := \sum_{j=1}^m \varrho_j F_j(x). \quad (4-7)$$

Proposição 4.1. *Suponha que vale a Hipótese 4.1. Sejam $\chi_1, \chi_2, \chi_3 \in \mathbb{R}_+^m$ tais que $\|\chi_1\|_1 = \|\chi_2\|_1 = \|\chi_3\|_1 = 1$. Então, existe uma constante positiva $M \in \mathbb{R}$ tal que, para todo $u, v, z, w \in \text{dom}(F)$, vale*

$$\nabla^2 F_{\chi_1}(v) - \nabla^2 F_{\chi_1}(u) \preceq M \|v - u\|_{\nabla^2 F_{\chi_2}(z)} \nabla^2 F_{\chi_3}(w), \quad (4-8)$$

em que $\|v - u\|_{\nabla^2 F_{\chi_2}(z)} \stackrel{(4-4)}{=} \sqrt{(v - u)^T \nabla^2 F_{\chi_2}(z) (v - u)}$.

Demonstração. Note que, para todo u, v, z e $w \in \mathbb{R}^n$, temos

$$\begin{aligned} \nabla^2 F_{\chi_1}(v) - \nabla^2 F_{\chi_1}(u) &\stackrel{(4-2)}{\preceq} L_2 \|v - u\| I_n = L_2 \langle v - u, v - u \rangle^{\frac{1}{2}} I_n \\ &= L_2 \langle I_n (v - u), v - u \rangle^{\frac{1}{2}} I_n \\ &\stackrel{(4-3)}{\preceq} \frac{L_2}{\mu^{1/2}} \langle \nabla^2 F_{\chi_2}(z) (v - u), v - u \rangle^{\frac{1}{2}} I_n \\ &= \frac{L_2}{\mu^{1/2}} \|v - u\|_{\nabla^2 F_{\chi_2}(z)} I_n \stackrel{(4-3)}{\preceq} \frac{L_2}{\mu^{3/2}} \|v - u\|_{\nabla^2 F_{\chi_2}(z)} \nabla^2 F_{\chi_3}(w). \end{aligned}$$

Tomando $M := \frac{L_2}{\mu^{3/2}}$, obtemos o resultado desejado. ■

Note que se tomarmos χ_1, χ_2, χ_3 iguais ao j -ésimo vetor canônico, então a Proposição 4.1 implica que F_j é fortemente auto-concordante. O próximo resultado, inspirado pelo caso escalar [67, Lema 4.2], relaciona a Hessiana da função F_χ (veja (4-7)) à Hessiana média.

Lema 4.1. *Sejam $\chi_1, \chi_2 \in \mathbb{R}_+^m$ tais que $\|\chi_1\|_1 = \|\chi_2\|_1 = 1$. Dados, $x, y \in \mathbb{R}^n$, defina*

$$r := \|y - x\|_{\nabla^2 F_{\chi_2}(x)} \stackrel{(4-4)}{=} \sqrt{\langle y - x, \nabla^2 F_{\chi_2}(x) (y - x) \rangle}. \quad (4-9)$$

Sob as condições da Hipótese 4.1, temos

$$\frac{\nabla^2 F_{\chi_1}(x)}{1 + Mr} \preceq \nabla^2 F_{\chi_1}(y) \preceq (1 + Mr) \nabla^2 F_{\chi_1}(x), \quad (4-10)$$

$$\frac{\nabla^2 F_{\chi_1}(x)}{1 + \frac{Mr}{2}} \preceq \bar{G}_{\chi_1}(x) \preceq (1 + \frac{Mr}{2}) \nabla^2 F_{\chi_1}(x), \quad (4-11)$$

$$\frac{\nabla^2 F_{\chi_1}(y)}{1 + \frac{Mr}{2}} \preceq \bar{G}_{\chi_1}(x) \preceq (1 + \frac{Mr}{2}) \nabla^2 F_{\chi_1}(y), \quad (4-12)$$

em que $\bar{G}_{\chi_1}(x) = \int_0^1 \nabla^2 F_{\chi_1}(x + \tau(y - x)) d\tau$.

Demonstração. Pela Hipótese 4.1, temos que a Proposição 4.1 é válida. Assim, tomando $u = z = w = x$ e $v = y$ na condição (4-8), obtemos

$$\nabla^2 F_{\chi_1}(y) - \nabla^2 F_{\chi_1}(x) \preceq Mr \nabla^2 F_{\chi_1}(x) \implies \nabla^2 F_{\chi_1}(y) \preceq (1 + Mr) \nabla^2 F_{\chi_1}(x).$$

Por outro lado, tomando $v = z = x$ e $u = w = y$ em (4-8), obtemos

$$\nabla^2 F_{\chi_1}(x) - \nabla^2 F_{\chi_1}(y) \preceq Mr \nabla^2 F_{\chi_1}(y) \implies \nabla^2 F_{\chi_1}(x) \preceq (1 + Mr) \nabla^2 F_{\chi_1}(y),$$

provando assim (4-10).

Considere (4-12). Em (4-8) tome $z = x$, $w = u = y$, $v = x + \tau h$, para $\tau \geq 0$, em que $h = y - x$. Assim,

$$\begin{aligned} \nabla^2 F_{\chi_1}(x + \tau h) - \nabla^2 F_{\chi_1}(y) &\preceq M \|x + \tau h - y\|_{\nabla^2 F_{\chi_2}(x)} \nabla^2 F_{\chi_1}(y) \\ &= M \|\tau h - h\|_{\nabla^2 F_{\chi_2}(x)} \nabla^2 F_{\chi_1}(y) = M |1 - \tau| r \nabla^2 F_{\chi_1}(y). \end{aligned}$$

Integrando em ambos os lados, obtemos

$$\int_0^1 \nabla^2 F_{\chi_1}(x + \tau h) d\tau - \nabla^2 F_{\chi_1}(y) \preceq \left. \frac{M(2\tau - \tau^2)}{2} r \nabla^2 F_{\chi_1}(y) \right|_0^1,$$

isto é, $\bar{G}_{\chi_1}(x) \preceq (1 + \frac{Mr}{2}) \nabla^2 F_{\chi_1}(y)$. Por fim, fazendo $v = y$, $z = x$, w arbitrário e $u = x + \tau h$, para $\tau \geq 0$, em (4-8), obtemos:

$$\nabla^2 F_{\chi_1}(y) - \nabla^2 F_{\chi_1}(x + \tau h) \preceq M \|h - \tau h\|_x \nabla^2 F_{\chi_1}(w) = M |1 - \tau| r \nabla^2 F_{\chi_1}(w).$$

Integrando em ambos os lados temos

$$\nabla^2 F_{\chi_1}(y) - \int_0^1 \nabla^2 F_{\chi_1}(x + \tau h) d\tau \preceq \left. \frac{M(2\tau - \tau^2)}{2} r \nabla^2 F_{\chi_1}(w) \right|_0^1,$$

isto é, $\nabla^2 F_{\chi_1}(y) - \bar{G}_{\chi_1}(x) \preceq \frac{Mr}{2} \nabla^2 F_{\chi_1}(w)$, para todo $w \in \mathbb{R}^n$. Tomando $w = x + \tau h$ e integrando novamente obtemos

$$\nabla^2 F_{\chi_1}(y) - \bar{G}_{\chi_1}(x) \preceq \frac{Mr}{2} \int_0^1 \nabla^2 F_{\chi_1}(x + \tau h) d\tau = \frac{Mr}{2} \bar{G}_{\chi_1}(x),$$

obtendo (4-10). A prova de (4-11) é análoga. ■

As demonstrações dos lemas a seguir podem ser encontradas em [69, Lema 6.1] e [69, Lema 2.1], respectivamente. Nesse último, os autores provaram que a atualização B_{k+1} preserva uma determinada condição de limitação. A demonstração para o caso multiobjetivo segue imediatamente do caso escalar, porém, por uma questão de completude, a colocamos neste trabalho.

Lema 4.2. [69, Lema 6.1] *Sejam $A, B \in \mathbb{R}^{n \times n}$ duas matrizes simétricas tais que $0 \prec A \preceq B$. Então, para qualquer $v \in \mathbb{R}^n \setminus \{0\}$, temos*

$$A - \frac{Avv^T A}{v^T Av} \preceq B - \frac{Bvv^T B}{v^T Bv}.$$

Lema 4.3. [69, Lema 2.1] *Suponha que vale a Hipótese 4.1. Se*

$$\frac{1}{a_1} \bar{G}_j^k \preceq B_j^k \preceq a_2 \bar{G}_j^k, \quad j = 1, \dots, m,$$

em que $a_1, a_2 \geq 1$, então

$$\frac{1}{a_1} \bar{G}_j^k \preceq B_j^{k+1} \preceq a_2 \bar{G}_j^k, \quad j = 1, \dots, m.$$

Demonstração. Dado $j = 1, \dots, m$, pelo Lema 4.2, temos

$$\frac{1}{a_1} \bar{G}_j^k - \frac{\frac{1}{a_1} \bar{G}_j^k s_k s_k^T \frac{1}{a_1} \bar{G}_j^k}{s_k^T \frac{1}{a_1} \bar{G}_j^k s_k} \preceq B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} \preceq a_2 \bar{G}_j^k - \frac{a_2 \bar{G}_j^k s_k s_k^T a_2 \bar{G}_j^k}{s_k^T a_2 \bar{G}_j^k s_k}. \quad (4-13)$$

Assim, por um lado

$$\begin{aligned} B_j^{k+1} &= B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{y_j^k (y_j^k)^T}{s_k^T y_j^k} \stackrel{(3-5)}{=} B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{\bar{G}_j^k s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} \\ &\stackrel{(4-13)}{\preceq} a_2 \left(\bar{G}_j^k - \frac{\bar{G}_j^k s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} \right) + \frac{\bar{G}_j^k s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} = a_2 \bar{G}_j^k - (a_2 - 1) \frac{\bar{G}_j^k s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} \\ &\preceq a_2 \bar{G}_j^k, \end{aligned}$$

em que a última desigualdade segue por $a_2 \geq 1$ e $\bar{G}_j^k s_k s_k^T \bar{G}_j^k \succ 0$. Por outro lado,

$$\begin{aligned} B_j^{k+1} &= B_j^k - \frac{B_j^k s_k s_k^T B_j^k}{s_k^T B_j^k s_k} + \frac{y_j^k (y_j^k)^T}{s_k^T y_j^k} \stackrel{(4-13)}{\succeq} \frac{1}{a_1} \left(\bar{G}_j^k - \frac{\bar{G}_j^k s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} \right) + \frac{\bar{G}_j^k s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} \\ &= \frac{1}{a_1} \bar{G}_j^k + \left(1 - \frac{1}{a_1}\right) \frac{\bar{G}_j^k s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} \succeq \frac{1}{a_1} \bar{G}_j^k, \end{aligned}$$

provando o resultado desejado. ■

O resultado a seguir, baseado em [68, Lema 5.2] e [69, Teorema 4.1], relaciona os autovalores das aproximações das Hessianas com os autovalores das Hessianas verdadeiras. O resultado envolve a norma definida em (4-9), a qual será considerada com uma combinação convexa conveniente. Seja $\chi = (\varrho_1, \dots, \varrho_m) \in \mathbb{R}_+^m$ com $\|\chi\|_1 = 1$. Defina:

$$B_\chi^k := \sum_{j=1}^m \varrho_j B_j^k. \quad (4-14)$$

Observação 4.2. Para todo $x \in \mathbb{R}$, temos:

$$1 + x \leq e^x \quad \text{e} \quad 1 - x \leq e^{-x}. \quad (4-15)$$

Teorema 4.1. *Assuma a Hipótese 4.1. Para todo $k \geq 0$ seja*

$$\xi_k := e^{M \sum_{i=0}^{k-1} r_i}, \quad (4-16)$$

em que

$$r_i := \|x^{i+1} - x^i\|_{\nabla^2 F_{\lambda_i}(x^i)} \stackrel{(4-4)}{=} \sqrt{\langle x^{i+1} - x^i, \nabla^2 F_{\lambda_i}(x^i)(x^{i+1} - x^i) \rangle}, \quad (4-17)$$

e $\lambda_i \equiv \lambda(x^i)$ é o multiplicador de Lagrange do subproblema (2-20). Seja $\chi \in \mathbb{R}_+^m$ tal que $\|\chi\|_1 = 1$. Então, para todo $k \geq 0$, temos:

$$\xi_k \leq \left(1 + \frac{Mr_k}{2}\right) e^{M \sum_{i=0}^{k-1} r_i} =: \xi'_k \leq \xi_{k+1}, \quad (4-18)$$

e

$$\frac{1}{\xi_k} \nabla^2 F_\chi(x^k) \preceq B_\chi^k \preceq \xi_k \frac{L}{\mu} \nabla^2 F_\chi(x^k), \quad (4-19)$$

$$\frac{1}{\xi_{k+1}} \bar{G}_\chi^k \preceq B_\chi^k \preceq \xi_{k+1} \frac{L}{\mu} \bar{G}_\chi^k. \quad (4-20)$$

Demonstração. Note que

$$\xi_k = e^{M \sum_{i=0}^{k-1} r_i} \leq \left(1 + \frac{Mr_k}{2}\right) e^{M \sum_{i=0}^{k-1} r_i} \stackrel{(4-15)}{\leq} e^{\frac{Mr_k}{2}} e^{M \sum_{i=0}^{k-1} r_i} \leq e^{M \sum_{i=0}^k r_i} = \xi_{k+1}, \quad (4-21)$$

em que $\xi_0 = 1^\dagger$ e $B_j^0 = LI_n$, para todo $j = 1, \dots, m$. Portanto, para $k = 0$, temos

[†]Estamos considerando que a soma sobre o conjunto vazio é definida como zero.

$B_\chi^0 = LI_n$ e, pela Hipótese 4.1,

$$\nabla^2 F_\chi(x^0) \preceq LI_n \preceq \frac{L}{\mu} \nabla^2 F_\chi(x^0),$$

mostrando que vale (4-19) para $k = 0$. Agora, por indução, suponha que (4-19) é válido para algum $k \geq 0$, logo

$$\begin{aligned} \frac{1}{\xi_{k+1}} \bar{G}_\chi^k &\stackrel{(4-18)}{\preceq} \frac{1}{\xi_k'} \bar{G}_\chi^k = \frac{1}{\xi_k \left(1 + \frac{Mr_k}{2}\right)} \bar{G}_\chi^k \stackrel{(4-11)}{\preceq} \frac{1}{\xi_k} \nabla^2 F_\chi(x^k) \\ &\stackrel{(4-19)}{\preceq} B_\chi^k \\ &\stackrel{(4-19)}{\preceq} \xi_k \frac{L}{\mu} \nabla^2 F_\chi(x^k) \stackrel{(4-11)}{\preceq} \xi_k \frac{L}{\mu} \left(1 + \frac{Mr_k}{2}\right) \bar{G}_\chi^k = \xi_k' \frac{L}{\mu} \bar{G}_\chi^k \stackrel{(4-18)}{\preceq} \xi_{k+1} \frac{L}{\mu} \bar{G}_\chi^k, \end{aligned} \quad (4-22)$$

obtendo assim (4-20). Agora, nos resta mostrar que (4-19) é válido para $k + 1$. Pelo Lema 4.3 e por (4-22), temos

$$\frac{1}{\xi_k'} \bar{G}_\chi^k \preceq B_\chi^{k+1} \preceq \xi_k' \frac{L}{\mu} \bar{G}_\chi^k. \quad (4-23)$$

Assim, por um lado, a condição (4-12) garante que

$$\begin{aligned} B_\chi^{k+1} &\stackrel{(4-23)}{\preceq} \left(1 + \frac{Mr_k}{2}\right) \xi_k' \frac{L}{\mu} \nabla^2 F_\chi(x^{k+1}) = \left(1 + \frac{Mr_k}{2}\right)^2 \xi_k \frac{L}{\mu} \nabla^2 F_\chi(x^{k+1}) \\ &\stackrel{(4-15)}{\preceq} e^{Mr_k} \xi_k \frac{L}{\mu} \nabla^2 F_\chi(x^{k+1}) = e^{Mr_k} e^{\sum_{i=0}^{k-1} Mr_i} \frac{L}{\mu} \nabla^2 F_\chi(x^{k+1}) \\ &= \xi_{k+1} \frac{L}{\mu} \nabla^2 F_\chi(x^{k+1}). \end{aligned}$$

Por outro lado, a mesma condição (4-12) também garante que

$$\begin{aligned} B_\chi^{k+1} &\stackrel{(4-23)}{\succeq} \frac{\nabla^2 F_\chi(x^{k+1})}{\left(1 + \frac{Mr_k}{2}\right) \xi_k'} = \frac{\nabla^2 F_\chi(x^{k+1})}{\left(1 + \frac{Mr_k}{2}\right)^2 \xi_k} \\ &\stackrel{(4-15)}{\succeq} \frac{\nabla^2 F_\chi(x^{k+1})}{e^{Mr_k} \xi_k} = \frac{\nabla^2 F_\chi(x^{k+1})}{\xi_{k+1}}, \end{aligned}$$

concluindo assim a prova. ■

A seguir enunciamos um resultado técnico cuja demonstração pode ser encontrada em [68, Lema 3.3].

Lema 4.4. [68, Lema 3.3] *Para quaisquer números reais $\alpha \geq \beta > 0$, temos*

$\alpha + \frac{1}{\beta} - 1 \geq 1$, e

$$\alpha - \ln \beta - 1 \geq \frac{\sqrt{3}}{2 + \sqrt{3}} \ln \left(\alpha + \frac{1}{\beta} - 1 \right) \geq \frac{6}{13} \ln \left(\alpha + \frac{1}{\beta} - 1 \right). \quad (4-24)$$

O resultado a seguir é usado em [68, Lema 3.4] sem uma prova formal, a qual apresentaremos aqui.

Proposição 4.2. *Para qualquer $k \geq 0$ e $j = 1, \dots, m$ temos*

$$\frac{s_k^T \bar{G}_j^k s_k}{s_k^T B_j^k s_k} \leq \frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k}. \quad (4-25)$$

Demonstração. Considere $u = \left(B_j^k \right)^{\frac{1}{2}} s_k$ e $v = \left(H_j^k \right)^{\frac{1}{2}} \bar{G}_j^k s_k$. Então, por Cauchy-Schwarz, obtemos

$$\begin{aligned} \left(s_k^T \bar{G}_j^k s_k \right)^2 &= \left(s_k^T \left(B_j^k \right)^{\frac{1}{2}} \left(H_j^k \right)^{\frac{1}{2}} \bar{G}_j^k s_k \right)^2 = \langle u, v \rangle^2 \leq \|u\|^2 \|v\|^2 = u^T u v^T v \\ &= s_k^T \left(B_j^k \right)^{\frac{1}{2}} \left(B_j^k \right)^{\frac{1}{2}} s_k s_k^T \bar{G}_j^k \left(H_j^k \right)^{\frac{1}{2}} \left(H_j^k \right)^{\frac{1}{2}} \bar{G}_j^k s_k \\ &= s_k^T B_j^k s_k s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k, \end{aligned}$$

provando o resultado desejado. ■

Uma maneira de mensurar a proximidade entre as matrizes B_j^k e \bar{G}_j^k , ao longo da direção s_k , é:

$$\nu(\bar{G}_j^k, B_j^k) := \frac{\langle (B_j^k - \bar{G}_j^k) H_j^k (B_j^k - \bar{G}_j^k) s_k, s_k \rangle^{1/2}}{\langle \bar{G}_j^k s_k, s_k \rangle^{1/2}} = \frac{\| (B_j^k - \bar{G}_j^k) s_k \|_{B_j^k}^*}{\| s_k \|_{\bar{G}_j^k}}. \quad (4-26)$$

Essa medida, para o caso escalar, foi definida inicialmente em [68]. Agora, vamos provar que a função potencial (4-6) pode ser limitada inferiormente por uma função logarítmica de ν . O roteiro de prova desse resultado não difere do caso escalar [68, Lema 3.4], porém, por uma questão de completude, está feito aqui.

Lema 4.5. [68, Lema 3.4] *Suponha que vale a Hipótese 4.1. Para algum $j = 1, \dots, m$ e $k \geq 0$, temos*

$$\Psi(B_j^k, \bar{G}_j^k) - \Psi(B_j^{k+1}, \bar{G}_j^k) \geq \frac{6}{13} \ln \left(1 + \nu(\bar{G}_j^k, B_j^k)^2 \right).$$

Demonstração. Note que

$$\begin{aligned}
(H_j^k - H_j^{k+1}) \bar{G}_j &\stackrel{(3-2)}{=} \left[H_j^k - \left((I - \rho_j^k s_k (y_j^k)^T) H_j^k (I - \rho_j^k y_j^k s_k^T) + \rho_j^k s_k s_k^T \right) \right] \bar{G}_j \\
&= \left[\frac{H_j^k y_j^k s_k^T}{s_k^T y_j^k} + \frac{s_k (y_j^k)^T H_j^k}{s_k^T y_j^k} - \frac{s_k (y_j^k)^T H_j^k y_j^k s_k^T}{(s_k^T y_j^k)^2} - \frac{s_k s_k^T}{s_k^T y_j^k} \right] \bar{G}_j \\
&= \left[\frac{H_j^k y_j^k s_k^T + s_k (y_j^k)^T H_j^k}{s_k^T y_j^k} - \left(\frac{(y_j^k)^T H_j^k y_j^k}{s_k^T y_j^k} + 1 \right) \frac{s_k s_k^T}{s_k^T y_j^k} \right] \bar{G}_j \\
&\stackrel{(3-5)}{=} \left[\frac{H_j^k \bar{G}_j^k s_k s_k^T + s_k s_k^T \bar{G}_j^k H_j^k}{s_k^T \bar{G}_j^k s_k} \right. \\
&\quad \left. - \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{s_k s_k^T}{s_k^T \bar{G}_j^k s_k} \right] \bar{G}_j \\
&= \frac{H_j^k \bar{G}_j^k s_k s_k^T \bar{G}_j^k + s_k s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k} \\
&\quad - \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{s_k s_k^T \bar{G}_j^k}{s_k^T \bar{G}_j^k s_k}.
\end{aligned}$$

Consequentemente,

$$\begin{aligned}
\text{Tr} \left((H_j^k - H_j^{k+1}) \bar{G}_j \right) &= \left[\text{Tr} \left(H_j^k \bar{G}_j^k s_k s_k^T \bar{G}_j^k + s_k s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k \right) \right] \frac{1}{s_k^T \bar{G}_j^k s_k} \\
&\quad - \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{\text{Tr} \left(s_k s_k^T \bar{G}_j^k \right)}{s_k^T \bar{G}_j^k s_k} \\
&= \frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} - 1.
\end{aligned}$$

Além disso,

$$\begin{aligned}
\text{Det} \left(H_j^{k+1} B_j^k \right) &= \left[\text{Det} \left(B_j^{k+1} H_j^k \right) \right]^{-1} \stackrel{(2-33)}{=} \left[\text{Det}(B_j) \frac{s_k^T y_j^k}{s_k^T B_j^k s_k} \text{Det} \left(H_j^k \right) \right]^{-1} \\
&\stackrel{(3-5)}{=} \frac{s_k^T B_j^k s_k}{s_k^T \bar{G}_j^k s_k}.
\end{aligned}$$

Assim, tomando $\alpha := \frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} \stackrel{(4-25)}{\geq} \frac{s_k^T \bar{G}_j^k s_k}{s_k^T B_j^k s_k} =: \beta$, temos

$$\begin{aligned}
\Psi(B_j^k, \bar{G}_j^k) - \Psi(B_j^{k+1}, \bar{G}_j^k) &\stackrel{(4-6)}{=} \ln \text{Det} \left((\bar{G}_j^k)^{-1} B_j^k \right) - \langle H_j^k, B_j^k - \bar{G}_j^k \rangle \\
&\quad - \ln \text{Det} \left((\bar{G}_j^k)^{-1} B_j^{k+1} \right) + \langle H_j^{k+1}, B_j^{k+1} - \bar{G}_j^k \rangle \\
&\stackrel{(4-5)}{=} \langle H_j^k - H_j^{k+1}, \bar{G}_j^k \rangle + \ln \frac{\text{Det } B_j^k}{\text{Det } B_j^{k+1}} \\
&= \text{Tr} \left(\left(H_j^k - H_j^{k+1} \right) \bar{G}_j^k \right) + \ln \text{Det} \left(H_j^{k+1} B_j^k \right) \\
&= \frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} - 1 + \ln \left(\frac{s_k^T B_j^k s_k}{s_k^T \bar{G}_j^k s_k} \right) \\
&= \alpha - \ln \beta - 1 \stackrel{(4-24)}{\geq} \frac{6}{13} \ln \left(\alpha + \frac{1}{\beta} - 1 \right).
\end{aligned}$$

Por fim, basta notar que

$$\begin{aligned}
\alpha + \frac{1}{\beta} - 1 &= \frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + \frac{s_k^T B_j^k s_k}{s_k^T \bar{G}_j^k s_k} - 1 = \frac{s_k^T \left(B_j^k + \bar{G}_j^k H_j^k \bar{G}_j^k \right) s_k}{s_k^T \bar{G}_j^k s_k} - 1 \\
&= \frac{s_k^T \left(B_j^k + \bar{G}_j^k H_j^k \bar{G}_j^k - \bar{G}_j^k \right) s_k}{s_k^T \bar{G}_j^k s_k} \\
&= \frac{s_k^T (B_j^k - \bar{G}_j^k) H_j^k (B_j^k - \bar{G}_j^k) s_k + s_k^T \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} \stackrel{(4-26)}{=} 1 + \nu(\bar{G}_j^k, B_j^k)^2,
\end{aligned}$$

obtendo o resultado desejado. ■

A medida definida em (4-26) é a razão entre a norma de $(B_j^k - \bar{G}_j^k)s_k$, em relação a H_j^k , e a norma de s_k , em relação a \bar{G}_j^k . Para mudar as métricas correspondentes para H_j^{k+1} e B_j^k , respectivamente, a expressão é penalizada com o menor autovalor de B_j^k em relação a \bar{G}_j^k . O resultado a seguir foi estabelecido em [68, Lema 3.5], contudo, por uma questão de completude vamos apresentá-lo aqui.

Lema 4.6. [68, Lema 3.5] *Dado algum $k \geq 0$, sejam \bar{G}_j^k e B_j^k tais que $\frac{1}{\xi} \bar{G}_j^k \preceq B_j^k$ para alguma constante $\xi > 0$ e para todo $j = 1, \dots, m$. Então,*

$$\nu(\bar{G}_j^k, B_j^k)^2 \geq \frac{1}{1 + \xi} \frac{\|(B_j^k - \bar{G}_j^k)s_k\|_{B_j^{k+1}}^*}{\|s_k\|_{B_j^k}} = \frac{1}{1 + \xi} \frac{\langle (B_j^k - \bar{G}_j^k) H_j^{k+1} (B_j^k - \bar{G}_j^k) s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle}.$$

Demonstração. Inicialmente, note que $H_j^{k+1}\bar{G}_j^k s_k = s_k$. De fato,

$$\begin{aligned} H_j^{k+1} &\stackrel{(3-2)}{=} \left(I_n - \frac{s_k(y_j^k)^T}{s_k^T y_j^k} \right) H_j^k \left(I_n - \frac{y_j^k s_k^T}{s_k^T y_j^k} \right) + \frac{s_k s_k^T}{s_k^T y_j^k} \\ &\stackrel{(3-5)}{=} H_j^k - \frac{H_j^k \bar{G}_j^k s_k s_k^T + s_k s_k^T \bar{G}_j^k H_j^k}{s_k \bar{G}_j^k s_k} + \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{s_k s_k^T}{s_k^T \bar{G}_j^k s_k}. \end{aligned}$$

Consequentemente,

$$\begin{aligned} H_j^{k+1} \bar{G}_j^k s_k &= H_j^k \bar{G}_j^k s_k - \frac{H_j^k \bar{G}_j^k s_k s_k^T \bar{G}_j^k s_k + s_k s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k \bar{G}_j^k s_k} \\ &\quad + \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{s_k s_k^T \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} = s_k. \end{aligned} \quad (4-27)$$

Agora, note que

$$\begin{aligned} \frac{\langle (B_j^k - \bar{G}_j^k) H_j^{k+1} (B_j^k - \bar{G}_j^k) s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} &= \frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} + \frac{\langle H_j^{k+1} \bar{G}_j^k s_k, \bar{G}_j^k s_k \rangle}{\langle B_j^k s_k, s_k \rangle} \\ &\quad - \frac{\langle H_j^{k+1} \bar{G}_j^k s_k, B_j^k s_k \rangle + \langle B_j^k s_k, H_j^{k+1} \bar{G}_j^k s_k \rangle}{\langle B_j^k s_k, s_k \rangle} \\ &\stackrel{(4-27)}{=} \frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} + \frac{\langle s_k, \bar{G}_j^k s_k \rangle}{\langle B_j^k s_k, s_k \rangle} \\ &\quad - \frac{\langle s_k, B_j^k s_k \rangle + \langle B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} \\ &= \frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} + \frac{\langle s_k, \bar{G}_j^k s_k \rangle}{\langle B_j^k s_k, s_k \rangle} - 2. \end{aligned} \quad (4-28)$$

Além disso, também temos

$$\begin{aligned} \langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle &= s_k^T B_j^k H_j^k B_j^k s_k - \frac{s_k^T B_j^k H_j^k \bar{G}_j^k s_k s_k^T B_j^k s_k + s_k^T B_j^k s_k s_k^T \bar{G}_j^k H_j^k B_j^k s_k}{s_k \bar{G}_j^k s_k} \\ &\quad + \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{s_k^T B_j^k s_k s_k^T B_j^k s_k}{s_k^T \bar{G}_j^k s_k} \\ &= -s_k^T B_j^k s_k + \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{(s_k^T B_j^k s_k)^2}{s_k^T \bar{G}_j^k s_k}. \end{aligned}$$

Logo,

$$\frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} = -1 + \left(\frac{s_k^T \bar{G}_j^k H_j^k \bar{G}_j^k s_k}{s_k^T \bar{G}_j^k s_k} + 1 \right) \frac{s_k^T B_j^k s_k}{s_k^T \bar{G}_j^k s_k}. \quad (4-29)$$

Denotando $\nu := \nu(\bar{G}_j^k, B_j^k)$, temos

$$\nu^2 \stackrel{(4-26)}{=} \frac{\langle (B_j^k - \bar{G}_j^k) H_j^k (B_j^k - \bar{G}_j^k) s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} = \frac{\langle B_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} + \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} - 2. \quad (4-30)$$

Consequentemente, por $\frac{1}{\xi} \bar{G}_j^k \preceq B_j^k$, temos

$$\begin{aligned} (1 + \xi)\nu^2 &= \left(\frac{\langle \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} \xi + 1 \right) \nu^2 = \left(\frac{\langle B_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} \xi + 1 \right) \nu^2 \\ &\geq \left(\frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} + 1 \right) \nu^2 \\ &\stackrel{(4-30)}{=} \left(\frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} + 1 \right) \left(\frac{\langle B_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} + \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} - 2 \right) \\ &= \left(\frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} + 1 \right) \frac{\langle B_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} + \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle^2}{\langle \bar{G}_j^k s_k, s_k \rangle^2} \\ &\quad - \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} - 2 \\ &\stackrel{(4-29)}{\geq} \frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} + \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle^2}{\langle \bar{G}_j^k s_k, s_k \rangle^2} - \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} - 1. \end{aligned} \quad (4-31)$$

Portanto,

$$\begin{aligned} (1 + \xi)\nu^2 &- \frac{\langle (B_j^k - \bar{G}_j^k) H_j^{k+1} (B_j^k - \bar{G}_j^k) s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} \\ &\stackrel{(4-28)}{=} (1 + \xi)\nu^2 - \frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} - \frac{\langle s_k, \bar{G}_j^k s_k \rangle}{\langle B_j^k s_k, s_k \rangle} + 2 \\ &\stackrel{(4-31)}{\geq} \frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} + \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle^2}{\langle \bar{G}_j^k s_k, s_k \rangle^2} - \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} \\ &\quad - 1 - \frac{\langle B_j^k H_j^{k+1} B_j^k s_k, s_k \rangle}{\langle B_j^k s_k, s_k \rangle} - \frac{\langle s_k, \bar{G}_j^k s_k \rangle}{\langle B_j^k s_k, s_k \rangle} + 2 \\ &\stackrel{(4-25)}{\geq} \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle^2}{\langle \bar{G}_j^k s_k, s_k \rangle^2} - 2 \frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} + 1 \\ &= \left(\frac{\langle \bar{G}_j^k H_j^k \bar{G}_j^k s_k, s_k \rangle}{\langle \bar{G}_j^k s_k, s_k \rangle} - 1 \right)^2 \geq 0. \end{aligned}$$

■

4.2 Convergência superlinear não assintótica

Nesta seção, vamos provar de maneira não assintótica que a taxa de convergência do Algoritmo 4.1 é Q-superlinear. Para tal, ao longo desta seção vamos supor que

$$\xi_k \leq \sqrt{\frac{3}{2}}, \quad \forall k \geq 0, \quad (4-32)$$

em que ξ_k está definido em (4-16). Na Seção 4.3, fazemos uma análise de como garantir que vale (4-32). Por essa condição, temos

$$1 + \frac{Mr_k}{2} \stackrel{(4-15)}{\leq} e^{\frac{Mr_k}{2}} \leq \left(e^{M \sum_{i=0}^k r_i} \right)^{1/2} \leq \left(\frac{3}{2} \right)^{1/4}. \quad (4-33)$$

Além disso, (4-32) também implica que as sequências $\{\|B_j^k\|\}$ e $\{\|H_j^k\|\}$, para todo $j = 1, \dots, m$, são limitadas. De fato, dado $\chi \in \mathbb{R}_+^m$ tal que $\|\chi\|_1 = 1$, por (4-19) e (4-1), temos

$$\frac{1}{\xi_k} \mu I_n \preceq \frac{1}{\xi_k} \nabla^2 F_\chi(x^k) \preceq B_\chi^k \preceq \xi_k \frac{L}{\mu} \nabla^2 F_\chi(x^k) \preceq \xi_k \frac{L^2}{\mu} I_n,$$

para todo $k \geq 0$. Assim, por (4-32), obtemos

$$\frac{1}{\sqrt{\frac{3}{2}}} \mu I_n \preceq \frac{1}{\xi_k} \mu I_n \preceq B_\chi^k \preceq \xi_k \frac{L^2}{\mu} I_n \preceq \sqrt{\frac{3}{2}} \frac{L^2}{\mu} I_n.$$

Definindo

$$a := \sqrt{\frac{2}{3}} \mu \quad \text{e} \quad b := \sqrt{\frac{3}{2}} \frac{L^2}{\mu}, \quad (4-34)$$

temos

$$a I_n \preceq B_\chi^k \preceq b I_n \quad \text{e} \quad \frac{1}{b} I_n \preceq (B_\chi^k)^{-1} \preceq \frac{1}{a} I_n, \quad (4-35)$$

para todo $k \geq 0$. Esse resultado é fundamental para o desenvolvimento da análise feita neste capítulo, pois em diversos pontos das demonstrações precisamos substituir a matriz B_χ^k (ou $(B_\chi^k)^{-1}$) pela identidade I_n .

O lema a seguir, inspirado em [28, Lema 4.3], é crucial para mostrar o principal resultado desta seção (Teorema 4.2).

Lema 4.7. *Suponha que vale a Hipótese 4.1. Dado $\chi = (\varrho_1, \dots, \varrho_m) \in \mathbb{R}_+^m$ tal que*

$\|\chi\|_1 = 1$, então

$$2|\theta(x^k)| \leq \left(\left\| \nabla F_\chi(x^k) \right\|_{B_\chi^k}^* \right)^2, \quad k \geq 0. \quad (4-36)$$

Demonstração. Sabemos que a Lagrangiana do problema (2-20), para algum x^k , é dada por:

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

em que $\lambda := (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$. O problema dual associado é:

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m} \inf_{(t, d) \in \mathbb{R} \times \mathbb{R}^n} L((t, d), \lambda) \\ \text{s. a. } \lambda \geq 0. \end{aligned} \quad (4-37)$$

Como o problema (2-20) tem solução e existem multiplicadores de Lagrange $\lambda(x^k) = (\lambda_1(x^k), \dots, \lambda_m(x^k)) \in \mathbb{R}^m$, então o conjunto de soluções do problema dual (4-37) coincide com o conjunto de multiplicadores de Lagrange do problema primal (2-22), além disso, não há *gap* de dualidade, isto é, $\theta(x^k)$ (veja (2-23)) é o valor ótimo de (4-37) (veja [41, Teorema 5.2.18]). Logo,

$$\begin{aligned} \theta(x^k) &= \max_{\lambda \in \mathbb{R}_+^m} \inf_{(t, d) \in \mathbb{R} \times \mathbb{R}^n} L((t, d), \lambda) \\ &= \inf_{d \in \mathbb{R}^n} \sum_{j=1}^m \lambda_j(x^k) \left(\nabla F_j(x^k)^T d + \frac{1}{2} d^T B_j^k d \right) \\ &\geq \inf_{d \in \mathbb{R}^n} \sum_{j=1}^m \varrho_j \left(\nabla F_j(x^k)^T d + \frac{1}{2} d^T B_j^k d \right) \\ &= \inf_{d \in \mathbb{R}^n} \left(\nabla F_\chi(x^k)^T d + \frac{1}{2} d^T B_\chi^k d \right) \\ &= - \frac{\left(\left\| \nabla F_\chi(x^k) \right\|_{B_\chi^k}^* \right)^2}{2}, \end{aligned}$$

em que a segunda igualdade é válida, pois os multiplicadores do problema (2-20) coincidem com as soluções de (4-37) e $\sum_{j=1}^m \lambda_j(x^k) = 1$. Além disso, a última igualdade vale, pois $d \mapsto \nabla F_\chi(x^k)^T d + \frac{1}{2} d^T B_\chi^k d$ é uma função fortemente convexa. ■

O próximo resultado estabelece uma versão preliminar do teorema sobre a convergência superlinear do Algoritmo 4.1. Este resultado foi inspirado em [68, Lema 5.4, e Teorema 5.1], contudo, como estamos tratando do contexto

multiobjetivo, há algumas diferenças. A primeira delas é o surgimento de m (quantidade de objetivos) na exponencial. Além disso, estabelecemos o resultado com a norma euclidiana da direção de máxima descida generalizada, diferentemente do caso escalar, em que foi utilizada uma norma do gradiente definida a partir da Hessiana em um ponto.

Teorema 4.2. *Suponha que vale a Hipótese 4.1. Para todo $k \geq 1$, temos*

$$\|d_{SD}(x^k)\| \leq \left[\frac{81L^4}{16\mu^4} \left(e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 \right) \right]^{k/2} \sqrt{\frac{3}{2}} \frac{L}{\mu} \|d_{SD}(x^0)\|.$$

Demonstração. Sem perda de generalidade, assuma que $x^{i+1} - x^i \neq 0$ para todo $0 \leq i \leq k$. Denote $\Psi_j^i := \Psi(B_j^i, \bar{G}_j^i)$, $\tilde{\Psi}_j^{i+1} := \Psi(B_j^{i+1}, \bar{G}_j^i)$ e $\nu_j^i := \nu(\bar{G}_j^i, B_j^i)$. Sejam $j = 1, \dots, m$ e $0 \leq i \leq k-1$ fixados, porém arbitrários. Pelo Lema 4.5, temos

$$\frac{6}{13} \ln \left(1 + (\nu_j^i)^2 \right) \leq \Psi_j^i - \tilde{\Psi}_j^{i+1} = \Psi_j^i - \Psi_j^{i+1} + \Delta_j^i, \quad (4-38)$$

em que $\Delta_j^i := \Psi_j^{i+1} - \tilde{\Psi}_j^{i+1}$. Além disso,

$$\begin{aligned} \Delta_j^i &\stackrel{(4-6)}{=} \ln \text{Det}((\bar{G}_j^{i+1})^{-1} B_j^{i+1}) - \langle (B_j^{i+1})^{-1}, B_j^{i+1} - \bar{G}_j^{i+1} \rangle \\ &\quad - \ln \text{Det}((\bar{G}_j^i)^{-1} B_j^{i+1}) + \langle (B_j^{i+1})^{-1}, B_j^{i+1} - \bar{G}_j^i \rangle \\ &= \langle (B_j^{i+1})^{-1}, \bar{G}_j^{i+1} - \bar{G}_j^i \rangle + \ln \text{Det}((\bar{G}_j^{i+1})^{-1} B_j^{i+1}) \\ &\quad - \ln \text{Det}((\bar{G}_j^i)^{-1} B_j^{i+1}) \\ &= \langle (B_j^{i+1})^{-1}, \bar{G}_j^{i+1} - \bar{G}_j^i \rangle + \ln \left(\frac{\text{Det}((\bar{G}_j^{i+1})^{-1} B_j^{i+1})}{\text{Det}((\bar{G}_j^i)^{-1} B_j^{i+1})} \right) \\ &= \langle (B_j^{i+1})^{-1}, \bar{G}_j^{i+1} - \bar{G}_j^i \rangle + \ln \left(\frac{\text{Det}(\bar{G}_j^{i+1})^{-1}}{\text{Det}(\bar{G}_j^i)^{-1}} \right) \\ &= \langle (B_j^{i+1})^{-1}, \bar{G}_j^{i+1} - \bar{G}_j^i \rangle + \ln \text{Det}((\bar{G}_j^{i+1})^{-1} \bar{G}_j^i). \end{aligned}$$

Note que

$$\begin{aligned} \bar{G}_j^i &\stackrel{(4-12)}{\succeq} \frac{1}{1 + \frac{Mr_i}{2}} \nabla^2 F_j(x^{i+1}) \stackrel{(4-11)}{\succeq} \frac{1}{1 + \frac{Mr_i}{2}} \frac{1}{1 + \frac{Mr_{i+1}}{2}} \bar{G}_j^{i+1} \stackrel{(4-15)}{\succeq} e^{-\frac{M}{2}(r_i+r_{i+1})} \bar{G}_j^{i+1} \\ &\stackrel{(4-15)}{\succeq} \left(1 - \frac{M}{2}(r_i + r_{i+1}) \right) \bar{G}_j^{i+1}. \end{aligned}$$

Portanto, $\bar{G}_j^{i+1} - \bar{G}_j^i \preceq \frac{M}{2}(r_i + r_{i+1})\bar{G}_j^{i+1}$. Assim,

$$\begin{aligned}
\sum_{i=0}^{k-1} \langle (B_j^{i+1})^{-1}, \bar{G}_j^{i+1} - \bar{G}_j^i \rangle &\leq \frac{M}{2} \sum_{i=0}^{k-1} (r_i + r_{i+1}) \langle (B_j^{i+1})^{-1}, \bar{G}_j^{i+1} \rangle \\
&\stackrel{(4-20)}{\leq} n \frac{M}{2} \sum_{i=0}^{k-1} \xi_{i+2} (r_i + r_{i+1}) \\
&\stackrel{(4-18)}{\leq} n \xi_{k+1} \frac{M}{2} \sum_{i=0}^{k-1} (r_i + r_{i+1}) \\
&\leq n \xi_{k+1} \frac{M}{2} \sum_{i=0}^k 2r_i \stackrel{(4-18)}{=} n \xi_{k+1} \ln \xi_{k+1}.
\end{aligned}$$

Consequentemente,

$$\begin{aligned}
\sum_{i=0}^{k-1} \Delta_j^i &= \sum_{i=0}^{k-1} \langle (B_j^{i+1})^{-1}, \bar{G}_j^{i+1} - \bar{G}_j^i \rangle + \sum_{i=0}^{k-1} \ln \text{Det} \left((\bar{G}_j^{i+1})^{-1} \bar{G}_j^i \right) \\
&\leq n \xi_{k+1} \ln \xi_{k+1} + \ln \prod_{i=0}^{k-1} \text{Det} \left((\bar{G}_j^{i+1})^{-1} \bar{G}_j^i \right) \\
&= n \xi_{k+1} \ln \xi_{k+1} + \ln \text{Det} \left(\bar{G}_j^k \right)^{-1} \bar{G}_j^0.
\end{aligned} \tag{4-39}$$

Agora, somando (4-38) de 0 até $k-1$, obtemos

$$\begin{aligned}
\frac{6}{13} \sum_{i=0}^{k-1} \ln \left(1 + (\nu_j^i)^2 \right) &\leq \Psi_j^0 - \Psi_j^k + \sum_{i=0}^{k-1} \Delta_j^i \stackrel{(4-6)}{\leq} \Psi_j^0 + \sum_{i=0}^{k-1} \Delta_j^i \\
&\stackrel{(4-6)}{=} \ln \text{Det} \left((\bar{G}_j^0)^{-1} L I_n \right) - \left\langle \frac{1}{L} I_n, L I_n - \bar{G}_j^0 \right\rangle + \sum_{i=0}^{k-1} \Delta_j^i \\
&\stackrel{(4-39)}{\leq} \ln \left(\text{Det} \left((\bar{G}_j^k)^{-1} \bar{G}_j^0 \right) \text{Det} \left((\bar{G}_j^0)^{-1} L I_n \right) \right) \\
&\quad - \left\langle \frac{1}{L} I_n, L I_n - \bar{G}_j^0 \right\rangle + n \ln \xi_{k+1}^{\xi_{k+1}} \\
&= \ln \text{Det} \left((\bar{G}_j^k)^{-1} L I_n \right) - \left\langle \frac{1}{L} I_n, L I_n - \bar{G}_j^0 \right\rangle + n \ln \xi_{k+1}^{\xi_{k+1}} \\
&\stackrel{(4-1)}{\leq} n \ln \frac{L}{\mu} + n \ln \xi_{k+1}^{\xi_{k+1}} = n \ln \left(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu} \right).
\end{aligned} \tag{4-40}$$

Pela condição (4-20), vale o Lema 4.6, assim:

$$(\nu_j^i)^2 \geq \frac{1}{1 + \xi_{i+1}} \frac{s_i^T (B_j^i - \bar{G}_j^i) H_j^{i+1} (B_j^i - \bar{G}_j^i) s_i}{s_i^T B_j^i s_i} \stackrel{(4-35)}{\geq} \frac{1}{1 + \xi_{i+1}} \frac{1}{b^2} \frac{\| (B_j^i - \bar{G}_j^i) s_i \|^2}{\| s_i \|^2}. \tag{4-41}$$

Pela fórmula de Taylor, temos $\nabla F_j(x^{i+1}) = \nabla F_j(x^i) + \bar{G}_j^i(x^{i+1} - x^i)$. Logo, pela condição (3-28) e lembrando que $d^i = x^{i+1} - x^i = s_i$, temos

$$\begin{aligned} \nabla F_{\lambda_i}(x^{i+1}) &= \nabla F_{\lambda_i}(x^i) + \bar{G}_{\lambda_i}^i(x^{i+1} - x^i) \stackrel{(2-21)}{=} -B_{\lambda_i}^i(x^{i+1} - x^i) + \bar{G}_{\lambda_i}^i(x^{i+1} - x^i) \\ &= -\left(B_{\lambda_i}^i - \bar{G}_{\lambda_i}^i\right) s_i. \end{aligned} \quad (4-42)$$

Além disso, para todo $j = 1, \dots, m$, temos

$$\begin{aligned} \frac{|\theta(x^{i+1})|}{|\theta(x^i)|} &= \frac{2|\theta(x^{i+1})|}{2|\theta(x^i)|} \stackrel{(2-23)}{=} \frac{2|\theta(x^{i+1})|}{s_i^T B_{\lambda_i}^i s_i} \stackrel{(4-36)}{\leq} \frac{\left(\|\nabla F_{\lambda_i}(x^{i+1})\|_{B_{\lambda_i}^{i+1}}^*\right)^2}{s_i^T B_{\lambda_i}^i s_i} \\ &\stackrel{(4-42)}{=} \frac{\left(\|(B_{\lambda_i}^i - \bar{G}_{\lambda_i}^i)s_i\|_{B_{\lambda_i}^{i+1}}^*\right)^2}{s_i^T B_{\lambda_i}^i s_i} \stackrel{(4-35)}{\leq} \frac{1}{a^2} \max_{j=1, \dots, m} \frac{\|(B_j^i - \bar{G}_j^i)s_i\|^2}{\|s_i\|^2}. \end{aligned} \quad (4-43)$$

Logo,

$$\max_{j=1, \dots, m} (\nu_j^i)^2 \stackrel{(4-41)}{\geq} \frac{1}{1 + \xi_{i+1}} \frac{1}{b^2} \max_{j=1, \dots, m} \frac{\|(B_j^i - \bar{G}_j^i)s_i\|^2}{\|s_i\|^2} \stackrel{(4-43)}{\geq} \frac{1}{1 + \xi_{i+1}} \frac{a^2 |\theta(x^{i+1})|}{|\theta(x^i)|}.$$

Assim,

$$\begin{aligned} \prod_{i=0}^{k-1} \max_{j=1, \dots, m} (\nu_j^i)^2 &\geq \prod_{i=0}^{k-1} \frac{1}{1 + \xi_{i+1}} \frac{a^2 |\theta(x^{i+1})|}{|\theta(x^i)|} = \frac{|\theta(x^k)|}{|\theta(x^0)|} \left(\frac{a}{b}\right)^{2k} \prod_{i=0}^{k-1} \frac{1}{1 + \xi_{i+1}} \\ &\stackrel{(4-18)}{\geq} \frac{|\theta(x^k)|}{|\theta(x^0)|} \left(\frac{a}{b}\right)^{2k} \frac{1}{(1 + \xi_k)^k}. \end{aligned} \quad (4-44)$$

Note que, como $\ln(\cdot)$ é crescente e $(\nu_j^i)^2$ é positivo, temos: $\ln(1 + \max_{j=1, \dots, m} (\nu_j^i)^2) = \ln(\max_{j=1, \dots, m} \{1 + (\nu_j^i)^2\}) = \max_{j=1, \dots, m} \ln(1 + (\nu_j^i)^2)$. Desse modo,

$$\begin{aligned} \frac{6}{13} \sum_{i=0}^{k-1} \ln(1 + \max_{j=1, \dots, m} (\nu_j^i)^2) &= \frac{6}{13} \sum_{i=0}^{k-1} \max_{j=1, \dots, m} \ln(1 + (\nu_j^i)^2) \leq \frac{6}{13} \sum_{i=0}^{k-1} \sum_{j=1}^m \ln(1 + (\nu_j^i)^2) \\ &= \frac{6}{13} \sum_{j=1}^m \sum_{i=0}^{k-1} \ln(1 + (\nu_j^i)^2) \stackrel{(4-40)}{\leq} mn \ln\left(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu}\right). \end{aligned}$$

Consequentemente, pela convexidade de $t \mapsto \ln(1 + e^t)$, segue que

$$\begin{aligned}
\frac{13}{6} \frac{mn}{k} \ln \left(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu} \right) &\geq \frac{1}{k} \sum_{i=0}^{k-1} \ln(1 + \max_{j=1, \dots, m} (\nu_j^i)^2) = \frac{1}{k} \sum_{i=0}^{k-1} \ln \left(1 + e^{\ln \max_j (\nu_j^i)^2} \right) \\
&\geq \ln \left(1 + e^{\frac{1}{k} \sum_{i=0}^{k-1} \ln \max_j (\nu_j^i)^2} \right) = \ln \left(1 + e^{\ln \left(\prod_{i=0}^{k-1} \max_j (\nu_j^i)^2 \right)^{\frac{1}{k}}} \right) \\
&= \ln \left(1 + \left(\prod_{i=0}^{k-1} \max_j (\nu_j^i)^2 \right)^{\frac{1}{k}} \right) \\
&\stackrel{(4-44)}{\geq} \ln \left(1 + \left(\frac{|\theta(x^k)|}{|\theta(x^0)|} \left(\frac{a}{b} \right)^{2k} \frac{1}{(1 + \xi_k)^k} \right)^{\frac{1}{k}} \right) \\
&= \ln \left(1 + \frac{|\theta(x^k)|^{\frac{1}{k}}}{|\theta(x^0)|^{\frac{1}{k}}} \left(\frac{a}{b} \right)^2 \frac{1}{(1 + \xi_k)} \right).
\end{aligned}$$

Rearranjando, obtemos:

$$|\theta(x^k)| \leq \left[\frac{b^2(1 + \xi_k)}{a^2} \left(e^{\frac{13}{6} \frac{mn}{k} \ln \left(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu} \right)} - 1 \right) \right]^k |\theta(x^0)|.$$

Além disso, por (1-8) e pelo Lema 2.2, temos que $|\theta_{SD}(x^k)| = \frac{1}{2} \|d_{SD}(x^k)\|^2$ e $\frac{1}{b} |\theta_{SD}(x^k)| \leq |\theta(x^k)| \leq \frac{1}{a} |\theta_{SD}(x^k)|$, respectivamente. Assim, segue que

$$\|d_{SD}(x^k)\| \leq \left[\frac{b^2(1 + \xi_k)}{a^2} \left(e^{\frac{13}{6} \frac{mn}{k} \ln \left(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu} \right)} - 1 \right) \right]^{k/2} \sqrt{\frac{b}{a}} \|d_{SD}(x^0)\|.$$

Por fim, por (4-32) e (4-34), temos

$$\begin{aligned}
\|d_{SD}(x^k)\| &\leq \left[\frac{9 \left(\sqrt{\frac{3}{2}} \frac{L^2}{\mu} \right)^2}{4 \left(\sqrt{\frac{2}{3}} \mu \right)^2} \left(e^{\frac{13}{6} \frac{mn}{k} \ln \left(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu} \right)} - 1 \right) \right]^{k/2} \sqrt{\frac{\sqrt{\frac{3}{2}} \frac{L^2}{\mu}}{\sqrt{\frac{2}{3}} \mu}} \|d_{SD}(x^0)\| \\
&= \left[\frac{81L^4}{16\mu^4} \left(e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 \right) \right]^{k/2} \sqrt{\frac{3}{2}} \frac{L}{\mu} \|d_{SD}(x^0)\|.
\end{aligned}$$

■

O corolário a seguir expressa a convergência superlinear usando uma norma alternativa. Essa mudança em relação ao Teorema 4.2 nos permite comparar com o resultado que foi obtido no caso escalar [68, Teorema 5.1], que utiliza uma norma definida a partir da Hessiana. Além disso, essa maneira de expressar a convergência superlinear será útil para compararmos a taxa que obtivemos nesta seção com a obtida no Apêndice A (Teorema A.1).

Corolário 4.1. *Sob as hipóteses do Teorema 4.2, temos*

$$\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \leq \left[\left(\frac{3L}{2\mu} \right)^4 \left(e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 \right) \right]^{\frac{k}{2}} \sqrt{\frac{3}{2}} \left(\frac{L}{\mu} \right)^{\frac{3}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^*,$$

em que $\lambda_k^{SD} := \lambda^{SD}(x^k)$ é um multiplicador de Lagrange associado a x^k do problema (1-5).

Demonstração. Para todo $k \geq 0$, temos

$$\begin{aligned} \left(\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \right)^2 &\stackrel{(4-4)}{=} \langle d_{SD}(x^k), \nabla^2 F_{\lambda_k^{SD}}(x^k)^{-1} d_{SD}(x^k) \rangle \\ &\stackrel{(4-3)}{\leq} \frac{1}{\mu} \langle d_{SD}(x^k), I_n d_{SD}(x^k) \rangle = \frac{1}{\mu} \|d_{SD}(x^k)\|^2. \end{aligned}$$

Por outro lado,

$$\begin{aligned} \left(\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \right)^2 &\stackrel{(4-4)}{=} \langle d_{SD}(x^k), \nabla^2 F_{\lambda_k^{SD}}(x^k)^{-1} d_{SD}(x^k) \rangle \\ &\stackrel{(4-3)}{\geq} \frac{1}{L} \langle d_{SD}(x^k), I_n d_{SD}(x^k) \rangle = \frac{1}{L} \|d_{SD}(x^k)\|^2. \end{aligned}$$

Portanto, pelo Teorema 4.2, segue que

$$\begin{aligned} \|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* &\leq \sqrt{\frac{1}{\mu}} \|d_{SD}(x^k)\| \\ &\leq \sqrt{\frac{1}{\mu}} \left[\frac{81L^4}{16\mu^4} \left(e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 \right) \right]^{\frac{k}{2}} \sqrt{\frac{3}{2}} \frac{L}{\mu} \|d_{SD}(x^0)\| \\ &\leq \left[\left(\frac{3L}{2\mu} \right)^4 \left(e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 \right) \right]^{\frac{k}{2}} \sqrt{\frac{3}{2}} \left(\frac{L}{\mu} \right)^{\frac{3}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^*. \end{aligned}$$

■

4.3 Sobre a limitação de ξ_k

A limitação de ξ_k (veja (4-32)) é crucial para garantir a validade dos resultados que obtivemos neste capítulo e também no Apêndice A. Contudo, até onde sabemos, a Hipótese 4.1 não é suficiente para obter tal limitação. Assim, faremos a suposição adicional de que o passo unitário satisfaz as condições de Wolfe padrão, conforme formalmente descrita a seguir.

Hipótese 4.2. *O passo unitário satisfaz as condições de Wolfe padrão, isto é, dados parâmetros algorítmicos $0 < c_1 < c_2 < 1$, para todo $k \geq 0$, temos que*

$$\begin{aligned} F_j(x^k + d^k) &\leq F_j(x^k) + c_1 \mathcal{D}(x^k, d^k), \quad \forall j = 1, \dots, m, \\ \mathcal{D}(x^k + d^k, d^k) &\geq c_2 \mathcal{D}(x^k, d^k). \end{aligned}$$

Observação 4.3. (i) Vale ressaltar que na análise local de convergência assintótica, veja Seção 3.1.2, provamos que o passo unitário satisfaz as condições de Wolfe padrão (1-12)–(1-13) (Teorema 3.6). (ii) Sob as Hipóteses 4.1 e 4.2, o Algoritmo 4.1 coincide com o Algoritmo 3.1 tomando $B_j^0 = LI_n$ para todo $j = 1, \dots, m$.

Como consequência da Hipótese 4.2, a sequência $\{x^k\}$ gerada pelo Algoritmo 4.1 converge para um ponto Pareto ótimo x^* . Assim, todos os resultados obtidos na Seção 3.1 são válidos e serão usados aqui sem mais explicações. Em particular as sequências $\{\|B_j^k\|\}$ e $\{\|H_j^k\|\}$ são limitadas, veja Teorema 3.4. Logo, dado $\chi \in \mathbb{R}_+^m$ com $\|\chi\|_1 = 1$, existem constantes positivas $a, b \in \mathbb{R}$ tais que, para todo $k \geq 0$, vale

$$aI_n \preceq B_\chi^k \preceq bI_n \quad \text{e} \quad \frac{1}{b}I_n \preceq (B_\chi^k)^{-1} \preceq \frac{1}{a}I_n, \quad (4-45)$$

em que B_χ^k está definido em (4-14). Essa limitação será crucial para mostrarmos o resultado a seguir, que estabelece uma limitação para ξ_k . Apesar dessa limitação também ser necessária no caso escalar, veja [69, Teorema 4.1] e [68, Teorema 5.1], o roteiro de prova utilizado aqui segue uma linha distinta.

Proposição 4.3. *Suponha que valem as Hipóteses 4.1 e 4.2 e $x^0 \in \mathbb{R}^n$ é tal que*

$$\frac{4ML}{\mu^{3/2}} \|d_{SD}(x^0)\| \frac{1}{1 - q^{1/2}} \leq \frac{\ln \frac{3}{2}}{2}, \quad (4-46)$$

em que $q := 1 - \frac{(1 - c_2)c_1 a^2 \mu^2}{8L^2 b^2}$. Para todo $k \geq 0$, temos:

$$\xi_{k+1} \leq \sqrt{\frac{3}{2}}. \quad (4-47)$$

Demonstração. Por (4-45), segue que

$$\cos \beta_j^k = \frac{s_k^T B_j^k s_k}{\|s_k\| \|B_j^k s_k\|} \geq \frac{a \|s_k\|^2}{\|s_k\| \|B_j^k\| \|s_k\|} \geq \frac{a}{b},$$

para todo $j = 1, \dots, m$ e $k \geq 0$. Com isso, $\delta_k := \min_{j=1, \dots, m} \cos \beta_j^k \geq a/b$. Logo,

$$\begin{aligned} F_*(x^{k+1}) - F_*(x^*) &\stackrel{(3-13)}{\leq} \left(1 - \frac{(1-c_2)\delta_k^2 c_1 \mu^2}{8L^2}\right) (F_*(x^k) - F_*(x^*)) \\ &\leq \left(1 - \frac{(1-c_2)\frac{a^2}{b^2} c_1 \mu^2}{8L^2}\right) (F_*(x^k) - F_*(x^*)) \\ &= \left(1 - \frac{(1-c_2)c_1 a^2 \mu^2}{8L^2 b^2}\right) (F_*(x^k) - F_*(x^*)) \\ &\leq q^k (F_*(x^0) - F_*(x^*)). \end{aligned}$$

Além disso, por (3-11), temos

$$\|x^{k+1} - x^*\| \leq \left[\frac{2}{\mu} (F_*(x^{k+1}) - F_*(x^*))\right]^{\frac{1}{2}} \leq \left(q^{\frac{1}{2}}\right)^{k+1} \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*))\right]^{\frac{1}{2}}.$$

Logo,

$$\begin{aligned} \sum_{k \geq 0} \|x^{k+1} - x^*\| &\leq \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*))\right]^{\frac{1}{2}} \sum_{k=0}^{\infty} \left(q^{\frac{1}{2}}\right)^{k+1} \\ &= \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*))\right]^{\frac{1}{2}} \frac{q^{\frac{1}{2}}}{1 - q^{\frac{1}{2}}}. \end{aligned} \tag{4-48}$$

Por (4-46), temos

$$\begin{aligned} 2ML^{1/2} \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*))\right]^{\frac{1}{2}} \frac{1}{1 - q^{\frac{1}{2}}} &\stackrel{(3-12)}{\leq} 2ML^{1/2} \left[\frac{2}{\mu} \frac{2L}{\mu^2} \|d_{SD}(x^0)\|^2\right]^{\frac{1}{2}} \frac{1}{1 - q^{\frac{1}{2}}} \\ &\leq \frac{\ln \frac{3}{2}}{2}. \end{aligned}$$

Agora, para mostrar (4-47), resta provar que $M \sum_{i=0}^k r_i \leq \frac{\ln \frac{3}{2}}{2}$. De fato,

$$\begin{aligned}
M \sum_{i=0}^k r_i &= M \sum_{i=0}^k \langle x^{i+1} - x^i, \nabla^2 F_{\lambda_i}(x^i)(x^{i+1} - x^i) \rangle^{1/2} \\
&\stackrel{(4-3)}{\leq} ML^{1/2} \sum_{i=0}^k \langle x^{i+1} - x^i, I_n(x^{i+1} - x^i) \rangle^{1/2} = ML^{1/2} \sum_{i=0}^k \|x^{i+1} - x^i\| \\
&\leq 2ML^{1/2} \sum_{i=0}^k \max\{\|x^{i+1} - x^*\|, \|x^i - x^*\|\} \\
&\stackrel{(4-48)}{\leq} 2ML^{1/2} \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*)) \right]^{\frac{1}{2}} \max \left\{ \frac{q^{\frac{1}{2}}}{1 - q^{\frac{1}{2}}}, \frac{1}{1 - q^{\frac{1}{2}}} \right\} \\
&\leq 2ML^{1/2} \left[\frac{2}{\mu} (F_*(x^0) - F_*(x^*)) \right]^{\frac{1}{2}} \frac{1}{1 - q^{\frac{1}{2}}},
\end{aligned}$$

em que a última desigualdade segue, pois $q < 1$. Portanto, por (4-21), temos

$$\xi_{k+1} = e^{M \sum_{i=0}^k r_i} \leq e^{\frac{\ln \frac{3}{2}}{2}} = \sqrt{\frac{3}{2}} \quad \forall k \geq 0.$$

■

4.4 Uma comparação entre taxas

Nesta seção, comparamos a taxa de convergência obtida no Teorema 4.2, com a obtida no Teorema A.1, que é dada por:

$$\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \leq \left(\frac{44L^{11}mn}{\mu^{11}k} \right)^{\frac{k}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^*. \quad (4-49)$$

Essa comparação é interessante para que possamos ver como a taxa de convergência apresentada neste capítulo é *melhor*, quando comparada com a que obtivemos no Apêndice A.

Por um lado, o Corolário 4.1 garante que:

$$\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \leq \left[\left(\frac{3L}{2\mu} \right)^4 \left(e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 \right) \right]^{\frac{k}{2}} \sqrt{\frac{3}{2}} \left(\frac{L}{\mu} \right)^{\frac{3}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^*.$$

Agora, vamos mostrar que a convergência superlinear *se inicia* quando $k \geq K_0$, em

que

$$K_0 := \frac{31L^4mn}{\mu^4} \ln \frac{2L}{\mu} \geq \left(\frac{3L}{2\mu}\right)^4 \frac{13}{6} mn \ln \frac{2L}{\mu} \geq 1. \quad (4-50)$$

Para isso, note que $e^t \leq \frac{1}{1-t} = 1 + \frac{t}{1-t}$, para qualquer $t < 1$. Assim,

$$\begin{aligned} e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 &\leq \frac{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}}{1 - \frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} \stackrel{(4-50)}{\leq} \frac{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}}{1 - \left(\frac{2\mu}{3L}\right)^4} \leq \frac{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}}{1 - \left(\frac{2}{3}\right)^4} \\ &= \frac{81}{65} \frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu} = \frac{27mn}{10k} \ln \frac{2L}{\mu}. \end{aligned} \quad (4-51)$$

Além disso,

$$\left(\frac{L}{\mu}\right)^{\frac{3}{2}} = e^{\frac{3}{2} \ln \frac{L}{\mu}} \leq e^{\frac{3}{2} \ln \frac{2L}{\mu}} \stackrel{(4-50)}{\leq} e^{\frac{k}{7}} = (e^{\frac{2}{7}})^{\frac{k}{2}} \leq \left(\frac{3}{2}\right)^{\frac{k}{2}} \quad \text{e} \quad \left(\frac{3}{2}\right)^{\frac{1}{2}} \stackrel{(4-50)}{\leq} \left(\frac{3}{2}\right)^{\frac{k}{2}}. \quad (4-52)$$

Portanto,

$$\begin{aligned} \|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* &\leq \left[\left(\frac{3L}{2\mu}\right)^4 \left(e^{\frac{13}{6} \frac{mn}{k} \ln \frac{2L}{\mu}} - 1 \right) \right]^{\frac{k}{2}} \sqrt{\frac{3}{2}} \left(\frac{L}{\mu}\right)^{\frac{3}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \\ &\stackrel{(4-51)}{\leq} \left[\left(\frac{3L}{2\mu}\right)^4 \frac{27mn}{10k} \ln \frac{2L}{\mu} \right]^{\frac{k}{2}} \sqrt{\frac{3}{2}} \left(\frac{L}{\mu}\right)^{\frac{3}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \\ &\stackrel{(4-52)}{\leq} \left[\left(\frac{3L}{2\mu}\right)^4 \frac{27mn}{10k} \ln \frac{2L}{\mu} \right]^{\frac{k}{2}} \left(\frac{3}{2}\right)^k \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \\ &= \left[\left(\frac{3}{2}\right)^6 \left(\frac{L}{\mu}\right)^4 \frac{27mn}{10k} \ln \frac{2L}{\mu} \right]^{\frac{k}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \\ &\leq \left(\frac{31L^4mn}{\mu^4k} \ln \frac{2L}{\mu} \right)^{\frac{k}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^*. \end{aligned}$$

Assim, o *momento de início* da convergência superlinear ocorre quando

$$\frac{31L^4mn}{\mu^4k} \ln \frac{2L}{\mu} \leq 1,$$

isto é, quando $k \geq K_0$.

Por outro lado, a segunda taxa de convergência superlinear (4-49), se *inicia* quando

$$\frac{44L^{11}mn}{\mu^{11}k} \leq 1,$$

isto é, quando $k \geq \frac{44L^{11}mn}{\mu^{11}} =: \bar{K}_0$. Portanto, ao compararmos K_0 e \bar{K}_0 , vemos uma melhora significativa da taxa obtida na Seção 4.2 em relação à apresentada no Apêndice A.

Conclusão

Neste trabalho, propusemos três novos esquemas de atualizações BFGS para problemas de otimização multiobjetivo, a saber (3-1), (3-43) e (3-58). Quando os tamanhos dos passos satisfazem as condições de Wolfe e as aproximações iniciais são definidas positivas, os dois primeiros esquemas geram aproximações Hessianas definidas positivas. Até onde sabemos, os Algoritmos 3.1 e 3.2 são os primeiros algoritmos do tipo BFGS projetados para problemas multiobjetivo não convexos que atualizam as aproximações Hessianas a cada iteração. Por outro lado, a atualização cautelosa adotada na terceira proposta se mantém definida positiva, independentemente de busca linear. Como consequência, os Algoritmos 3.1, 3.2 e 3.3 estão bem definidos mesmo para problemas gerais não convexos. Fornecemos um estudo abrangente das principais propriedades de convergência global e local dos métodos, usando hipóteses que são extensões naturais daquelas feitas para o caso escalar. Em particular, os Algoritmos 3.2 e 3.3 são globalmente convergentes mesmo para problemas multiobjetivo não convexos. Os experimentos numéricos sugerem que as técnicas propostas aqui promovem um aperfeiçoamento prático do método BFGS multiobjetivo. Esperamos que essas técnicas também possam ser úteis para outras variantes de métodos quase-Newton para otimização multiobjetivo.

Por fim, caracterizamos explicitamente a convergência local superlinear do método BFGS para otimização multiobjetivo. Para isso, assumimos que o passo unitário satisfaz as condições de Wolfe. Acreditamos que seja possível caracterizar a região de aceite do passo unitário em termos da norma da direção de máxima descida generalizada. Além do mais, a taxa de convergência que obtivemos não se reduz (tomando $m = 1$) à taxa obtida no caso escalar [68]. Esses são indicativos que os resultados obtidos no Capítulo 4 podem ser aprimorados, o que pretendemos explorar em trabalhos futuros.

Referências Bibliográficas

- [1] ANSARY, M. A.; PANDA, G. **A modified Quasi-Newton method for vector optimization problem.** *Optimization*, 64(11):2289–2306, 2015.
- [2] BHASKAR, V.; GUPTA, S. K.; RAY, A. K. **Applications of multiobjective optimization in chemical engineering.** *Rev. Chem. Eng.*, 16(1):1–54, 2000.
- [3] BIRGIN, E.; MARTÍNEZ, J. **Practical Augmented Lagrangian Methods for Constrained Optimization.** SIAM, Philadelphia, 2014.
- [4] BONNEL, H.; IUSEM, A. N.; SVAITER, B. F. **Proximal methods in vector optimization.** *SIAM J. Optim.*, 15(4):953–970, 2005.
- [5] BROYDEN, C. G. **The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations.** *IMA J. Appl. Math.*, 6(1):76–90, 03 1970.
- [6] BYRD, R. H.; NOCEDAL, J. **A Tool for the Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization.** *SIAM J. Numer. Anal.*, 26(3):727–739, June 1989.
- [7] CENG, L. C.; MORDUKHOVICH, B. S.; YAO, J. C. **Hybrid approximate proximal method with auxiliary variational inequality for vector optimization.** *J. Optimiz. Theory App.*, 146(2):267–303, 2010.
- [8] CENG, L. C.; YAO, J. C. **Approximate proximal methods in vector optimization.** *Eur. J. Oper. Res.*, 183(1):1–19, 2007.
- [9] CHUONG, T. D. **Generalized proximal method for efficient solutions in vector optimization.** *Numer. Funct. Anal. Optim.*, 32(8):843–857, 2011.
- [10] CHUONG, T. D. **Newton-like methods for efficient solutions in vector optimization.** *Comput. Optim. Appl.*, 54(3):495–516, 2013.
- [11] CHUONG, T. D.; MORDUKHOVICH, B. S.; YAO, J. C. **Hybrid approximate proximal algorithms for efficient solutions in vector optimization.** *J. Nonlinear Convex Anal.*, 12(2):257–285, 8 2011.

- [12] CRUZ, J. B.; PÉREZ, L. L.; MELO, J. **Convergence of the projected gradient method for quasiconvex multiobjective optimization.** *Nonlinear Anal.*, 74(16):5268 – 5273, 2011.
- [13] CUSTÓDIO, A. L.; MADEIRA, J. F. A.; VAZ, A. I. F.; VICENTE, L. N. **Direct Multisearch for Multiobjective Optimization.** *SIAM J. Optim.*, 21(3):1109–1140, 2011.
- [14] DAI, Y.-H. **Convergence properties of the BFGS algorithm.** *SIAM J. Optim.*, 13(3):693–701, 2002.
- [15] DAI, Y.-H. **A perfect example for the BFGS method.** *Math. Program.*, 138(1-2):501–530, 2013.
- [16] DAS, I.; DENNIS, J. **Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems.** *SIAM J. Optim.*, 8(3):631–657, 1998.
- [17] DAVIDON, W. C. **Variable Metric Method for Minimization.** *Research and Development Report, No. ANL-5990, Argonne Natational Lab., Argonne, Illinois, 1959.*
- [18] DAVIDON, W. C. **Variable Metric Method for Minimization.** *SIAM J. Optim.*, 1(1):1–17, 1991.
- [19] DEB, K.; THIELE, L.; LAUMANN, M.; ZITZLER, E. **Scalable Test Problems for Evolutionary Multiobjective Optimization.** In: Abraham, A.; Jain, L.; Goldberg, R., editors, *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, p. 105–145. Springer, London, 2005.
- [20] DENNIS, J. E.; MORÉ, J. J. **A Characterization of Superlinear Convergence and Its Application to Quasi-Newton Methods.** *Math. Comp.*, 28(126):549–560, 1974.
- [21] DENNIS, J. E.; SCHNABEL, R. B. **Numerical Methods for Unconstrained Optimization and Nonlinear Equations.** SIAM, Philadelphia, 1996.
- [22] DOLAN, E. D.; MORÉ, J. J. **Benchmarking optimization software with performance profiles.** *Math. Program.*, 91(2):201–213, 2002.
- [23] DRUMMOND, L.; SVAITER, B. **A steepest descent method for vector optimization.** *J. Comput. Appl. Math.*, 175:395–414, 03 2005.

- [24] EICHFELDER, G. **Adaptive Scalarization Methods in Multiobjective Optimization**. Springer Berlin Heidelberg, 2008.
- [25] FAZZIO, N. S.; SCHUVERDT, M. L. **Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems**. *Optim. Lett.*, 13(6):1365–1379, 2019.
- [26] FLETCHER, R. **A new approach to variable metric algorithms**. *Comput. J.*, 13(3):317–322, 01 1970.
- [27] FLETCHER, R.; POWELL, M. J. D. **A Rapidly Convergent Descent Method for Minimization**. *Comput. J.*, 6(2):163–168, 08 1963.
- [28] FLIEGE, J.; GRAÑA DRUMMOND, L. M.; SVAITER, B. F. **Newton’s Method for Multiobjective Optimization**. *SIAM J. Optim.*, 20(2):602–626, 2009.
- [29] FLIEGE, J.; SVAITER, B. F. **Steepest descent methods for multicriteria optimization**. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.
- [30] FUKUDA, E. H.; GRAÑA DRUMMOND, L. M. **On the convergence of the projected gradient method for vector optimization**. *Optimization*, 60(8-9):1009–1021, 2011.
- [31] FUKUDA, E. H.; GRAÑA DRUMMOND, L. M. **Inexact projected gradient method for vector optimization**. *Comput. Optim. Appl.*, 54(3):473–493, 2013.
- [32] GOLDFARB, D. **A family of variable-metric methods derived by variational means**. *Math. Comput.*, 24:23–26, 1970.
- [33] GONÇALVES, M. L. N.; LIMA, F. S.; PRUDENTE, L. F. **Globally convergent newton-type methods for multiobjective optimization**. *Comput. Optim. Appl.*, 83(2):403–434, 2022.
- [34] GONÇALVES, M. L. N.; PRUDENTE, L. F. **On the extension of the Hager–Zhang conjugate gradient method for vector optimization**. *Comput. Optim. Appl.*, 76(3):889–916, 2020.
- [35] GONÇALVES, M.; LIMA, F.; PRUDENTE, L. **A study of Liu–Storey conjugate gradient methods for vector optimization**. *Appl. Math. Comput.*, 425:127099, 2022.
- [36] GRAÑA DRUMMOND, L. M.; IUSEM, A. N. **A Projected Gradient Method for Vector Optimization Problems**. *Comput. Optim. Appl.*, 28(1):5–29, 2004.

- [37] GRAÑA DRUMMOND, L. M.; RAUPP, F. M. P.; SVAITER, B. F. **A quadratically convergent Newton method for vector optimization.** *Optimization*, 63(5):661–677, 2014.
- [38] GRAÑA DRUMMOND, L. M.; SVAITER, B. F. **A steepest descent method for vector optimization.** *J. Comput. Appl. Math.*, 175(2):395 – 414, 2005.
- [39] HILLERMEIER, C. **Generalized homotopy approach to multiobjective optimization.** *J. Optimiz. Theory App.*, 110(3):557–583, 2001.
- [40] HUBAND, S.; HINGSTON, P.; BARONE, L.; WHILE, L. **A review of multiobjective test problems and a scalable test problem toolkit.** *IEEE T. Evolut. Comput.*, 10(5):477–506, 2006.
- [41] IZMAILOV, A.; SOLODOV, M. **Otimização, volume 1: condições de otimalidade, elementos de análise convexa e de dualidade.** Impa, 2005.
- [42] JIN, Y.; OLHOFFER, M.; SENDHOFF, B. **Dynamic Weighted Aggregation for Evolutionary Multi-Objective Optimization: Why Does It Work and How?** In: *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, GECCO'01, p. 1042–1049, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [43] KIM, I.; DE WECK, O. **Adaptive weighted-sum method for bi-objective optimization: Pareto front generation.** *Struct. Multidiscip. O.*, 29(2):149–158, Feb 2005.
- [44] LAI, K. K.; MISHRA, S. K.; RAM, B. **On q-Quasi-Newton's Method for Unconstrained Multiobjective Optimization Problems.** *Mathematics*, 8(4), 2020.
- [45] LAUMANN, M.; THIELE, L.; DEB, K.; ZITZLER, E. **Combining convergence and diversity in evolutionary multiobjective optimization.** *Evol. Comput.*, 10(3):263–282, 2002.
- [46] LI, D.-H.; FUKUSHIMA, M. **A modified BFGS method and its global convergence in nonconvex minimization.** *J. Comput. Appl. Math.*, 129(1):15–35, 2001. *Nonlinear Programming and Variational Inequalities*.
- [47] LI, D.-H.; FUKUSHIMA, M. **On the Global Convergence of the BFGS Method for Nonconvex Unconstrained Optimization Problems.** *SIAM J. Optim.*, 11(4):1054–1064, 2001.

- [48] LOVISON, A. **Singular Continuation: Generating Piecewise Linear Approximations to Pareto Sets via Global Analysis.** *SIAM J. Optim.*, 21(2):463–490, 2011.
- [49] LU, F.; CHEN, C.-R. **Newton-like methods for solving vector optimization problems.** *Appl. Anal.*, 93(8):1567–1586, 2014.
- [50] LUCAMBIO PÉREZ, L. R.; PRUDENTE, L. F. **Nonlinear Conjugate Gradient Methods for Vector Optimization.** *SIAM J. Optim.*, 28(3):2690–2720, 2018.
- [51] LUCAMBIO PÉREZ, L. R.; PRUDENTE, L. F. **A Wolfe Line Search Algorithm for Vector Optimization.** *ACM Trans. Math. Softw.*, 45(4):23, 2019.
- [52] MAHDAVI-AMIRI, N.; SADAGHIANI, F. S. **A superlinearly convergent nonmonotone quasi-Newton method for unconstrained multiobjective optimization.** *Optim. Methods Softw.*, 0(0):1–25, 2020.
- [53] MARTÍNEZ, J. M.; SANTOS, S. A. **Métodos computacionais de otimização.** Sociedade Brasileira de Matemática, Rio de Janeiro, Brasil, 1 edition, 1995.
- [54] MASCARENHAS, W. F. **The BFGS method with exact line searches fails for non-convex objective functions.** *Math. Program.*, 99(1):49–61, 2004.
- [55] MIGLIERINA, E.; MOLHO, E.; RECCHIONI, M. **Box-constrained multi-objective optimization: A gradient-like method without a priori scalarization.** *Eur. J. Oper. Res.*, 188(3):662–682, 2008.
- [56] MITA, K.; FUKUDA, E. H.; YAMASHITA, N. **Nonmonotone line searches for unconstrained multiobjective optimization problems.** *J. Global Optim.*, 75:63–90, 2019.
- [57] MORÉ, J. J.; GARBOW, B. S.; HILLSTROM, K. E. **Testing Unconstrained Optimization Software.** *ACM Trans. Math. Softw.*, 7(1):17–41, Mar. 1981.
- [58] MOROVATI, V.; BASIRZADEH, H.; POURKARIMI, L. **Quasi-Newton methods for multiobjective optimization problems.** *4OR-Q J Oper Res*, 16(3):261–294, Nov. 2017.
- [59] MUKAI, H. **Algorithms for multicriterion optimization.** *IEEE Trans. Automat. Contr.*, 25(2):177–186, 1980.
- [60] NOCEDAL, J.; WRIGHT, S. J. **Numerical Optimization.** Springer, New York, NY, USA, second edition, 2006.

- [61] POVALEJ, Z. **Quasi-Newton's method for multiobjective optimization.** *J. Comput. Appl. Math.*, 255:765–777, 2014.
- [62] POWELL, M. J. D. **Some global convergence properties of a variable metric algorithm for minimization without exact line searches.** *Nonlinear Programming, SIAM-AMS Proceedings*, 4:53–72, 1976.
- [63] PREUSS, M.; NAUJOKS, B.; RUDOLPH, G. **Pareto Set and EMOA Behavior for Simple Multimodal Multiobjective Functions.** In: Runarsson, T. P.; Beyer, H.-G.; Burke, E.; Merelo-Guervós, J. J.; Whitley, L. D.; Yao, X., editors, *Parallel Problem Solving from Nature - PPSN IX*, p. 513–522, Berlin, Heidelberg, 2006. Springer.
- [64] PRUDENTE, L. F.; SOUZA, D. R. **A Quasi-Newton Method with Wolfe Line Searches for Multiobjective Optimization.** *J. Optim. Theory Appl.*, 194:1107–1140, 2022.
- [65] QU, S.; GOH, M.; CHAN, F. T. **Quasi-Newton methods for solving multiobjective optimization.** *Oper. Res. Lett.*, 39(5):397–399, 2011.
- [66] QU, S.; LIU, C.; GOH, M.; LI, Y.; JI, Y. **Nonsmooth multiobjective programming with quasi-Newton methods.** *Eur. J. Oper. Res.*, 235(3):503 – 510, 2014.
- [67] RODOMANOV, A.; NESTEROV, Y. **Greedy Quasi-Newton Methods with Explicit Superlinear Convergence.** *SIAM J. Optim.*, 31(1):785–811, 2021.
- [68] RODOMANOV, A.; NESTEROV, Y. **New Results on Superlinear Convergence of Classical Quasi-Newton Methods.** *J. Optim. Theory Appl.*, 188, 03 2021.
- [69] RODOMANOV, A.; NESTEROV, Y. **Rates of superlinear convergence for classical quasi-Newton methods.** *Math. Program.*, 02 2021.
- [70] SCHÜTZE, O.; LAUMANN, M.; COELLO COELLO, C. A.; DELLNITZ, M.; TALBI, E.-G. **Convergence of stochastic search algorithms to finite size Pareto set approximations.** *J. Global Optim.*, 41(4):559–577, 2008.
- [71] SHANNO, D. F. **Conditioning of quasi-Newton methods for function minimization.** *Math. Comput.*, 24:647–656, 1970.
- [72] STADLER, W.; DAUER, J. **Multicriteria Optimization In Engineering: A Tutorial And Survey.** In: *Structural Optimization: Status And Promise*, chapter 10, p. 209–249. American Institute of Aeronautics and Astronautics, 1992.

- [73] STEWART, T.; BANDTE, O.; BRAUN, H.; CHAKRABORTI, N.; EHRGOTT, M.; GÖBELT, M.; JIN, Y.; NAKAYAMA, H.; POLES, S.; DI STEFANO, D. **Real-World Applications of Multiobjective Optimization**. In: Branke, J.; Deb, K.; Miettinen, K.; Słowiński, R., editors, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, p. 285–327. Springer, Berlin, Heidelberg, 2008.
- [74] SVAITER, B. F. **The multiobjective steepest descent direction is not Lipschitz continuous, but is Hölder continuous**. *Oper. Res. Lett.*, 46(4):430–433, 2018.
- [75] TOINT, P. L. **Test problems for partially separable optimization and results for the routine PSPMIN**. *The University of Namur, Department of Mathematics, Belgium, Tech. Rep.*, 1983.
- [76] WANG, J.; HU, Y.; WAI YU, C. K.; LI, C.; YANG, X. **Extended Newton methods for multiobjective optimization: majorizing function technique and convergence analysis**. *SIAM J. Optim.*, 29(3):2388–2421, 2019.
- [77] ZITZLER, E.; DEB, K.; THIELE, L. **Comparison of Multiobjective Evolutionary Algorithms: Empirical Results**. *Evol. Comput.*, 8(2):173–195, 2000.

Uma versão preliminar da convergência explícita do Algoritmo 4.1

Neste apêndice, assim como no Capítulo 4, vamos desenvolver os resultados necessários para mostrar, de maneira não assintótica, que o Algoritmo 4.1 possui taxa de convergência Q-superlinear. Vale destacar, que a análise apresentada aqui nos leva a uma taxa *inferior* quando comparada àquela que calculamos na Seção 4.2, veja a Seção 4.4. Contudo, o roteiro de prova desenvolvido aqui usa técnicas alternativas, as quais acreditamos que podem ser úteis para futuros estudos de métodos quase-Newton.

Ao longo deste apêndice, vamos considerar as hipóteses feitas no Capítulo 4, isto é, vamos supor que vale a Hipótese 4.1 e a condição (4-32). Iniciaremos, com o incremento de alguns resultados iniciais, calculando a taxa de convergência linear do Algoritmo 4.1.

Considere a função de uma variável $\omega : (-1, +\infty) \rightarrow \mathbb{R}$, definida por:

$$\omega(t) := t - \ln(1+t) \geq 0. \quad (\text{A-1})$$

Note que ω é uma função convexa, decrescente em $(-1, 0]$ e crescente em $[0, +\infty)$. Além disso, temos:

$$\frac{t^2}{2(1+t)} \leq \omega(t) \leq \frac{t^2}{2+t}, \quad \forall t \geq 0. \quad (\text{A-2})$$

Os próximos dois lemas podem ser encontrados em [69].

Lema A.1. [69, Lema 2.3] *Para quaisquer números reais $\alpha \geq \beta > 0$, temos*

$$\alpha - \ln \beta - 1 \geq \omega(\sqrt{\alpha\beta - 2\beta + 1}).$$

Agora, dado $\chi \in \mathbb{R}_+^m$ tal que $\|\chi\|_1 = 1$, definimos uma maneira de medir a

proximidade entre as matrizes B_χ^k e \bar{G}_χ^k , ao longo da direção s_k , por:

$$\varphi(\bar{G}_\chi^k, B_\chi^k) := \frac{\langle (B_\chi^k - \bar{G}_\chi^k)(\bar{G}_\chi^k)^{-1}(B_\chi^k - \bar{G}_\chi^k)s_k, s_k \rangle^{1/2}}{\langle B_\chi^k(\bar{G}_\chi^k)^{-1}B_\chi^k s_k, s_k \rangle^{1/2}} = \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|_{\bar{G}_\chi^k}^*}{\|s_k\|_{\bar{G}_\chi^k}^*}. \quad (\text{A-3})$$

Quando $\chi = e_j$ em que e_j é o j -ésimo vetor canônico denotaremos, por simplicidade, $\varphi(\bar{G}_{e_j}^k, B_{e_j}^k) = \varphi(\bar{G}_j^k, B_j^k)$. A medida (A-3) se difere da que definimos em (4-26), pois aqui utilizamos a matriz \bar{G}_χ^k nas normas do numerador e denominador. Assim como no Lema 4.5, é possível provar que a função potencial (A-3) é limitada inferiormente.

Lema A.2. [69, Lema 2.4] *Suponha que vale a Hipótese 4.1 e que existam constantes $\xi, \eta \geq 1$, tais que*

$$\frac{1}{\xi}\bar{G}_j^k \preceq B_j^k \preceq \eta\bar{G}_j^k, \quad \forall j = 1, \dots, m,$$

para algum $k \geq 0$. Então,

$$\Psi(\bar{G}_j^k, B_j^k) - \Psi(\bar{G}_j^k, B_j^{k+1}) \geq \omega \left(\frac{\varphi(\bar{G}_j^k, B_j^k)}{\xi} \right).$$

O resultado a seguir foi inspirado em [69, Lema 4.2].

Lema A.3. *Suponha que vale a Hipótese 4.1. Seja $\lambda_k := \lambda(x^k) \in \mathbb{R}^m$ o multiplicador de Lagrange do problema (2-20). Para todo $k \geq 0$, temos:*

$$\|\nabla F_{\lambda_k}(x^{k+1})\|_{\nabla^2 F_{\lambda_k}(x^{k+1})}^* \leq \left(1 + \frac{Mr_k}{2}\right) \varphi_k \|\nabla F_{\lambda_k}(x^k)\|_{\nabla^2 F_{\lambda_k}(x^k)}^*, \quad (\text{A-4})$$

em que $\varphi_k := \varphi(\bar{G}_{\lambda_k}^k, B_{\lambda_k}^k)$ é dado por (A-3) e r_k está definido em (4-17).

Demonstração. Lembrando que $s_k = d^k$, temos

$$\begin{aligned}
\|\nabla F_{\lambda_k}(x^{k+1})\|_{\nabla^2 F_{\lambda_k}(x^{k+1})}^* &= \langle \nabla F_{\lambda_k}(x^{k+1}), \nabla^2 F_{\lambda_k}(x^{k+1})^{-1} \nabla F_{\lambda_k}(x^{k+1}) \rangle^{\frac{1}{2}} \\
&\stackrel{(4-12)}{\preceq} \sqrt{1 + \frac{Mr_k}{2}} \langle \nabla F_{\lambda_k}(x^{k+1}), (\bar{G}_{\lambda_k}^k)^{-1} \nabla F_{\lambda_k}(x^{k+1}) \rangle^{\frac{1}{2}} \\
&\stackrel{(4-42)}{=} \sqrt{1 + \frac{Mr_k}{2}} \langle (B_{\lambda_k}^k - \bar{G}_{\lambda_k}^k) s_k, (\bar{G}_{\lambda_k}^k)^{-1} (B_{\lambda_k}^k - \bar{G}_{\lambda_k}^k) s_k \rangle^{\frac{1}{2}} \\
&\stackrel{(A-3)}{=} \sqrt{1 + \frac{Mr_k}{2}} \varphi_k \langle B_{\lambda_k}^k (\bar{G}_{\lambda_k}^k)^{-1} B_{\lambda_k}^k s_k, s_k \rangle^{\frac{1}{2}} \\
&= \sqrt{1 + \frac{Mr_k}{2}} \varphi_k \langle B_{\lambda_k}^k s_k, (\bar{G}_{\lambda_k}^k)^{-1} B_{\lambda_k}^k s_k \rangle^{\frac{1}{2}} \\
&\stackrel{(2-21)}{=} \sqrt{1 + \frac{Mr_k}{2}} \varphi_k \langle \nabla F_{\lambda_k}(x^k), (\bar{G}_{\lambda_k}^k)^{-1} \nabla F_{\lambda_k}(x^k) \rangle^{\frac{1}{2}} \\
&\stackrel{(4-11)}{\preceq} \left(1 + \frac{Mr_k}{2}\right) \varphi_k \langle \nabla F_{\lambda_k}(x^k), \nabla^2 F_{\lambda_k}(x^k)^{-1} \nabla F_{\lambda_k}(x^k) \rangle^{\frac{1}{2}} \\
&= \left(1 + \frac{Mr_k}{2}\right) \varphi_k \|\nabla F_{\lambda_k}(x^k)\|_{\nabla^2 F_{\lambda_k}(x^k)}^*,
\end{aligned}$$

obtendo assim o resultado desejado. \blacksquare

O leitor pode notar que, no lado esquerdo da desigualdade (A-4), a combinação convexa dos gradientes em x^{k+1} e a norma utilizada dependem dos multiplicadores de Lagrange do problema (2-20) no ponto x^k . Isso dificulta obter uma limitação para $\|\nabla F_{\lambda_k}(x^{k+1})\|_{\nabla^2 F_{\lambda_k}(x^{k+1})}^*$ usando argumentos de recursividade, pois há um *cruzamento de índices* na desigualdade. Nos resultados seguintes superamos esse inconveniente.

Lema A.4. *Suponha que vale a Hipótese 4.1 e também a condição (4-32). Considere $\chi \in \mathbb{R}_+^m$ com $\|\chi\|_1 = 1$, então, para todo $k \geq 0$, temos*

$$\|\nabla F_{\lambda_k}(x^k)\|_{\nabla^2 F_{\lambda_k}(x^k)}^* \leq \frac{b L^{1/2}}{a \mu^{1/2}} \|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^*, \quad (\text{A-5})$$

$$\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \leq \frac{L^{1/2}}{\mu^{1/2}} \|\nabla F_{\chi}(x^k)\|_{\nabla^2 F_{\chi}(x^k)}^*, \quad (\text{A-6})$$

em que a, b estão definidos em (4-34).

Demonstração. Por (4-35) e pelo Lema 2.2, temos

$$\frac{a}{b} \|\nabla F_{\lambda_k}(x^k)\| \leq \|d_{SD}(x^k)\| \leq \|\nabla F_{\lambda_k}(x^k)\|, \quad \forall k \geq 0.$$

Assim,

$$\begin{aligned}
 \|\nabla F_{\lambda_k}(x^k)\|_{\nabla^2 F_{\lambda_k}(x^k)}^* &= \langle \nabla F_{\lambda_k}(x^k), \nabla^2 F_{\lambda_k}(x^k)^{-1} \nabla F_{\lambda_k}(x^k) \rangle^{\frac{1}{2}} \\
 &\stackrel{(4-3)}{\leq} \frac{1}{\mu^{1/2}} \langle \nabla F_{\lambda_k}(x^k), I_n \nabla F_{\lambda_k}(x^k) \rangle^{\frac{1}{2}} = \frac{1}{\mu^{1/2}} \|\nabla F_{\lambda_k}(x^k)\| \\
 &\leq \frac{1}{\mu^{1/2}} \frac{b}{a} \|d_{SD}(x^k)\| = \frac{1}{\mu^{1/2}} \frac{b}{a} \langle d_{SD}(x^k), I_n d_{SD}(x^k) \rangle^{\frac{1}{2}} \\
 &\stackrel{(4-3)}{\leq} \frac{1}{\mu^{1/2}} \frac{b}{a} L^{1/2} \langle d_{SD}(x^k), \nabla^2 F_{\lambda_k^{SD}}(x^k)^{-1} d_{SD}(x^k) \rangle^{\frac{1}{2}} \\
 &= \frac{b}{a} \frac{L^{1/2}}{\mu^{1/2}} \|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^*, \quad \forall k \geq 0.
 \end{aligned}$$

Além disso, o Lema 1.2 (iv) garante que $\|d_{SD}(x^k)\| \leq \|\nabla F_{\chi}(x^k)\|$. Logo,

$$\begin{aligned}
 \|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* &= \langle d_{SD}(x^k), \nabla^2 F_{\lambda_k^{SD}}(x^k)^{-1} d_{SD}(x^k) \rangle^{\frac{1}{2}} \\
 &\stackrel{(4-3)}{\leq} \frac{1}{\mu^{1/2}} \langle d_{SD}(x^k), I_n d_{SD}(x^k) \rangle^{\frac{1}{2}} = \frac{1}{\mu^{1/2}} \|d_{SD}(x^k)\| \\
 &\leq \frac{1}{\mu^{1/2}} \|\nabla F_{\chi}(x^k)\| = \frac{1}{\mu^{1/2}} \langle \nabla F_{\chi}(x^k), I_n \nabla F_{\chi}(x^k) \rangle^{\frac{1}{2}} \\
 &\stackrel{(4-3)}{\leq} \frac{L^{1/2}}{\mu^{1/2}} \langle \nabla F_{\chi}(x^k), \nabla^2 F_{\chi}(x^k)^{-1} \nabla F_{\chi}(x^k) \rangle^{\frac{1}{2}} \\
 &= \frac{L^{1/2}}{\mu^{1/2}} \|\nabla F_{\chi}(x^k)\|_{\nabla^2 F_{\chi}(x^k)}^*, \quad \forall k \geq 0.
 \end{aligned}$$

■

Lema A.5. *Suponha que valem a Hipótese 4.1 e a condição (4-32). Assim, para todo $k \geq 0$, temos:*

$$\|d_{SD}(x^{k+1})\|_{\nabla^2 F_{\lambda_{k+1}^{SD}}(x^{k+1})}^* \leq \frac{b}{a} \frac{L}{\mu} \left(1 + \frac{Mr_k}{2}\right) \varphi_k \|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^*. \quad (\text{A-7})$$

Demonstração. Para todo $k \geq 0$, temos

$$\begin{aligned}
 \left(\|d_{SD}(x^{k+1})\|_{\nabla^2 F_{\lambda_{k+1}^{SD}}(x^{k+1})}^* \right)^2 &\stackrel{(A-6)}{\leq} \frac{L}{\mu} \left(\|\nabla F_{\lambda_k}(x^{k+1})\|_{\nabla^2 F_{\lambda_k}(x^{k+1})}^* \right)^2 \\
 &\stackrel{(A-4)}{\leq} \frac{L}{\mu} \left[\left(1 + \frac{Mr_k}{2}\right) \varphi_k \|\nabla F_{\lambda_k}(x^k)\|_{\nabla^2 F_{\lambda_k}(x^k)}^* \right]^2 \\
 &\stackrel{(A-5)}{\leq} \frac{b^2 L^2}{a^2 \mu^2} \left[\left(1 + \frac{Mr_k}{2}\right) \varphi_k \|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \right]^2,
 \end{aligned}$$

provando o resultado desejado. ■

A seguir, apresentamos o último resultado antes de mostrar a convergência superlinear. Este resultado relaciona a medida de proximidade (A-3) usando diferentes escolhas de combinações convexas.

Lema A.6. *Suponha que vale a Hipótese 4.1 e a condição (4-32). Assim, para todo $k \geq 0$, temos:*

$$\varphi_k^2 \leq \frac{3bL^2}{2a\mu^2} \max_{j=1,\dots,m} (\varphi_j^k)^2, \quad (\text{A-8})$$

em que $\varphi_k := \varphi(\bar{G}_{\lambda_k}^k, B_{\lambda_k}^k)$ e $\varphi_j^k := \varphi(\bar{G}_j^k, B_j^k)$.

Demonstração. Por um lado, note que

$$\begin{aligned} \varphi(\bar{G}_\chi^k, B_\chi^k)^2 &= \frac{\langle (B_\chi^k - \bar{G}_\chi^k)(\bar{G}_\chi^k)^{-1}(B_\chi^k - \bar{G}_\chi^k)s_k, s_k \rangle}{\langle B_\chi^k(\bar{G}_\chi^k)^{-1}B_\chi^k s_k, s_k \rangle} \stackrel{(4-3)}{\leq} \frac{1}{\mu} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\langle B_\chi^k(\bar{G}_\chi^k)^{-1}B_\chi^k s_k, s_k \rangle} \\ &\stackrel{(4-20)}{\leq} \frac{1}{\mu} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\frac{1}{\xi_{k+1}} \langle \bar{G}_\chi^k(\bar{G}_\chi^k)^{-1}B_\chi^k s_k, s_k \rangle} = \frac{\xi_{k+1}}{\mu} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\langle B_\chi^k s_k, s_k \rangle} \\ &\stackrel{(4-35)}{\leq} \frac{\xi_{k+1}}{a\mu} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\|s_k\|^2} \stackrel{(4-32)}{\leq} \frac{\sqrt{\frac{3}{2}}}{a\mu} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\|s_k\|^2}. \end{aligned} \quad (\text{A-9})$$

Por outro lado, temos

$$\begin{aligned} \varphi(\bar{G}_\chi^k, B_\chi^k)^2 &= \frac{\langle (B_\chi^k - \bar{G}_\chi^k)(\bar{G}_\chi^k)^{-1}(B_\chi^k - \bar{G}_\chi^k)s_k, s_k \rangle}{\langle B_\chi^k(\bar{G}_\chi^k)^{-1}B_\chi^k s_k, s_k \rangle} \stackrel{(4-3)}{\geq} \frac{1}{L} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\langle B_\chi^k(\bar{G}_\chi^k)^{-1}B_\chi^k s_k, s_k \rangle} \\ &\stackrel{(4-20)}{\geq} \frac{1}{L} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\frac{L}{\mu} \xi_{k+1} \langle \bar{G}_\chi^k(\bar{G}_\chi^k)^{-1}B_\chi^k s_k, s_k \rangle} = \frac{\mu}{L^2 \xi_{k+1}} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\langle B_\chi^k s_k, s_k \rangle} \\ &\stackrel{(4-35)}{\geq} \frac{\mu}{bL^2 \xi_{k+1}} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\|s_k\|^2} \stackrel{(4-32)}{\geq} \frac{\mu}{bL^2 \sqrt{\frac{3}{2}}} \frac{\|(B_\chi^k - \bar{G}_\chi^k)s_k\|^2}{\|s_k\|^2}. \end{aligned} \quad (\text{A-10})$$

Assim,

$$\begin{aligned}
\varphi_k^2 &\stackrel{\text{(A-9)}}{\leq} \frac{\sqrt{\frac{3}{2}} \|(B_{\lambda_k}^k - \bar{G}_{\lambda_k}^k) s_k\|^2}{a\mu \|s_k\|^2} = \frac{\sqrt{\frac{3}{2}}}{a\mu} \left(\frac{\|\sum_{j=1}^m \lambda_j(x^k)(B_{\lambda_k}^k - \bar{G}_{\lambda_k}^k) s_k\|}{\|s_k\|} \right)^2 \\
&\stackrel{\text{(2-22)}}{\leq} \frac{\sqrt{\frac{3}{2}}}{a\mu} \left(\max_{j=1, \dots, m} \frac{\|(B_j^k - \bar{G}_j^k) s_k\|}{\|s_k\|} \right)^2 = \frac{\sqrt{\frac{3}{2}}}{a\mu} \max_{j=1, \dots, m} \frac{\|(B_j^k - \bar{G}_j^k) s_k\|^2}{\|s_k\|^2} \\
&\stackrel{\text{(A-10)}}{\leq} \frac{\sqrt{\frac{3}{2}} bL^2 \sqrt{\frac{3}{2}}}{a\mu \mu} \max_{j=1, \dots, m} \varphi(B_j^k, \bar{G}_j^k)^2 = \frac{3 bL^2}{2 a\mu^2} \max_{j=1, \dots, m} (\varphi_j^k)^2.
\end{aligned}$$

■

Finalmente, o próximo resultado estabelece de maneira explícita que o Algoritmo 4.1 possui taxa de convergência Q-superlinear.

Teorema A.1. *Suponha que valem a Hipótese 4.1 e a condição (4-32). Assim, para todo $k \geq 1$, temos*

$$\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* \leq \left(\frac{44L^{11}mn}{\mu^{11}k} \right)^{\frac{k}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^*.$$

Demonstração. Dados $k \geq 1$ e $j = 1, \dots, m$, defina $\varphi_j^k = \varphi(\bar{G}_j^k, B_j^k)$, $\Psi_j^k := \Psi(\bar{G}_j^k, B_j^k)$ e $\tilde{\Psi}_j^{k+1} := \Psi(\bar{G}_j^k, B_j^{k+1})$, em que Ψ é dada por (4-6). Seja $0 \leq i \leq k-1$ arbitrário. Assim, por (4-20) e pelo Lema A.2, temos

$$\begin{aligned}
\Psi_j^i - \tilde{\Psi}_j^{i+1} &= \Psi(\bar{G}_j^i, B_j^i) - \Psi(\bar{G}_j^i, B_j^{i+1}) \geq \omega\left(\frac{\varphi(\bar{G}_j^i, B_j^i)}{\xi_{i+1}}\right) = \omega\left(\frac{\varphi_j^i}{\xi_{i+1}}\right) \\
&\stackrel{\text{(A-1)}}{=} \frac{\varphi_j^i}{\xi_{i+1}} - \ln\left(1 + \frac{\varphi_j^i}{\xi_{i+1}}\right) \geq 0.
\end{aligned} \tag{A-11}$$

Por outro lado, $\xi_{i+1} \stackrel{\text{(4-32)}}{\leq} 2$ implica que $2B_j^i - \bar{G}_j^i \stackrel{\text{(4-20)}}{\succeq} 0$, assim

$$\begin{aligned}
(\varphi_j^i)^2 &\stackrel{\text{(A-3)}}{=} \frac{\langle (B_j^i - \bar{G}_j^i)(\bar{G}_j^i)^{-1}(B_j^i - \bar{G}_j^i) s_i, s_i \rangle}{\langle B_j^i (\bar{G}_j^i)^{-1} B_j^i s_i, s_i \rangle} = 1 + \frac{-2\langle B_j^i s_i, s_i \rangle + \langle \bar{G}_j^i s_i, s_i \rangle}{\langle B_j^i (\bar{G}_j^i)^{-1} B_j^i s_i, s_i \rangle} \\
&\leq 1.
\end{aligned}$$

Com isso,

$$\omega \left(\frac{\varphi_j^i}{\xi_{i+1}} \right) \geq \omega \left(\sqrt{\frac{2}{3}} \varphi_j^i \right) \stackrel{(A-2)}{\geq} \frac{\left(\sqrt{\frac{2}{3}} \varphi_j^i \right)^2}{2 \left(1 + \sqrt{\frac{2}{3}} \varphi_j^i \right)} \geq \frac{\left(\varphi_j^i \right)^2}{3 \left(1 + \sqrt{\frac{2}{3}} \right)}.$$

Logo,

$$\frac{1}{(3 + \sqrt{6})} \left(\varphi_j^i \right)^2 \leq \omega \left(\frac{\varphi_j^i}{\xi_{i+1}} \right) \stackrel{(A-11)}{\leq} \Psi_j^i - \tilde{\Psi}_j^{i+1} = \Psi_j^i - \Psi_j^{i+1} + \Delta_j^i, \quad (A-12)$$

em que $\Delta_j^i := \Psi_j^{i+1} - \tilde{\Psi}_j^{i+1}$. Além disso,

$$\begin{aligned} \Delta_j^i &= \Psi_j^{i+1} - \tilde{\Psi}_j^{i+1} \stackrel{(4-6)}{=} \langle (\bar{G}_j^{i+1})^{-1}, B_j^{i+1} - \bar{G}_j^{i+1} \rangle - \ln \text{Det}((\bar{G}_j^{i+1})^{-1} B_j^{i+1}) \\ &\quad - \langle (\bar{G}_j^i)^{-1}, B_j^{i+1} - \bar{G}_j^i \rangle + \ln \text{Det}((\bar{G}_j^i)^{-1} B_j^{i+1}) \\ &= \langle (\bar{G}_j^{i+1})^{-1} - (\bar{G}_j^i)^{-1}, B_j^{i+1} \rangle - \ln \text{Det}((\bar{G}_j^{i+1})^{-1} B_j^{i+1}) \\ &\quad + \ln \text{Det}((\bar{G}_j^i)^{-1} B_j^{i+1}) \\ &= \langle (\bar{G}_j^{i+1})^{-1} - (\bar{G}_j^i)^{-1}, B_j^{i+1} \rangle + \ln \left(\frac{\text{Det}((\bar{G}_j^i)^{-1} B_j^{i+1})}{\text{Det}((\bar{G}_j^{i+1})^{-1} B_j^{i+1})} \right) \\ &= \langle (\bar{G}_j^{i+1})^{-1} - (\bar{G}_j^i)^{-1}, B_j^{i+1} \rangle + \ln \left(\frac{\text{Det}(\bar{G}_j^i)^{-1}}{\text{Det}(\bar{G}_j^{i+1})^{-1}} \right) \\ &= \langle (\bar{G}_j^{i+1})^{-1} - (\bar{G}_j^i)^{-1}, B_j^{i+1} \rangle + \ln \text{Det}((\bar{G}_j^i)^{-1} \bar{G}_j^{i+1}). \end{aligned}$$

Agora, vamos estimar $\sum_{i=0}^{k-1} \Delta_j^i$. Para isso, note que

$$\bar{G}_j^{i+1} \stackrel{(4-11)}{\succeq} \frac{\nabla^2 F_j(x^{k+1})}{1 + \frac{Mr_{i+1}}{2}} \stackrel{(4-12)}{\succeq} \frac{1}{u_i} \bar{G}_j^i, \quad (A-13)$$

em que $u_i := \left(1 + \frac{Mr_{i+1}}{2}\right) \left(1 + \frac{Mr_i}{2}\right) \geq 1$. Logo,

$$\begin{aligned}
 \langle (\bar{G}_j^{i+1})^{-1} - (\bar{G}_j^i)^{-1}, B_j^{i+1} \rangle &= \langle (\bar{G}_j^{i+1})^{-1}, B_j^{i+1} \rangle - \langle (\bar{G}_j^i)^{-1}, B_j^{i+1} \rangle \\
 &\stackrel{(A-13)}{\leq} \langle (\bar{G}_j^{i+1})^{-1}, B_j^{i+1} \rangle - u_i^{-1} \langle (\bar{G}_j^{i+1})^{-1}, B_j^{i+1} \rangle \\
 &= (1 - u_i^{-1}) \langle (\bar{G}_j^{i+1})^{-1}, B_j^{i+1} \rangle \\
 &\stackrel{(4-20)}{\leq} (1 - u_i^{-1}) \xi_{i+2} \frac{L}{\mu} \langle (\bar{G}_j^{i+1})^{-1}, \bar{G}_j^{i+1} \rangle \\
 &\stackrel{(4-32)}{\leq} (1 - u_i^{-1}) \sqrt{\frac{3}{2}} \frac{L}{\mu} n,
 \end{aligned}$$

e, conseqüentemente, temos

$$\begin{aligned}
 \sum_{i=0}^{k-1} \Delta_j^i &\leq \sqrt{\frac{3}{2}} \frac{nL}{\mu} \sum_{i=0}^{k-1} (1 - u_i^{-1}) + \sum_{i=0}^{k-1} \ln(\text{Det}(\bar{G}_j^i)^{-1} \bar{G}_j^{i+1}) \\
 &= \sqrt{\frac{3}{2}} \frac{nL}{\mu} \sum_{i=0}^{k-1} (1 - u_i^{-1}) + \ln \prod_{i=0}^{k-1} \text{Det}(\bar{G}_j^i)^{-1} \bar{G}_j^{i+1} \\
 &= \sqrt{\frac{3}{2}} \frac{nL}{\mu} \sum_{i=0}^{k-1} (1 - u_i^{-1}) + \ln \text{Det}(\bar{G}_j^0)^{-1} \bar{G}_j^k.
 \end{aligned}$$

Agora, obtemos

$$\begin{aligned}
 \sum_{i=0}^{k-1} (1 - u_i^{-1}) &= \sum_{i=0}^{k-1} 1 - \left[\left(1 + \frac{Mr_{i+1}}{2}\right) \left(1 + \frac{Mr_i}{2}\right) \right]^{-1} \stackrel{(4-15)}{\leq} \sum_{i=0}^{k-1} 1 - e^{-\frac{M(r_i+r_{i+1})}{2}} \\
 &\stackrel{(4-15)}{\leq} \frac{M}{2} \sum_{i=0}^{k-1} r_i + r_{i+1} = M \left(\frac{r_0 + r_k}{2} + \sum_{i=1}^{k-1} r_i \right) \leq M \sum_{i=0}^k r_i \\
 &\stackrel{(4-32)}{\leq} \frac{\ln \frac{3}{2}}{2}.
 \end{aligned}$$

Portanto,

$$\sum_{i=0}^{k-1} \Delta_j^i \leq \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} + \ln \text{Det}(\bar{G}_j^0)^{-1} \bar{G}_j^k. \tag{A-14}$$

Somando (A-12) de 0 até $k - 1$, e usando que $\Psi_j^k \geq 0$, obtemos

$$\begin{aligned}
 \frac{1}{(3 + \sqrt{6})} \sum_{i=0}^{k-1} (\varphi_j^i)^2 &\leq \Psi_j^0 - \Psi_j^k + \sum_{i=0}^{k-1} \Delta_j^i \leq \Psi_j^0 + \sum_{i=0}^{k-1} \Delta_j^i = \Psi(\bar{G}_j^0, LI_n) + \sum_{i=0}^{k-1} \Delta_j^i \\
 &\stackrel{(4-6)}{=} \langle (\bar{G}_j^0)^{-1}, LI_n - \bar{G}_j^0 \rangle - \ln \text{Det} \left((\bar{G}_j^0)^{-1} LI_n \right) + \sum_{i=0}^{k-1} \Delta_j^i \\
 &\stackrel{(A-14)}{\leq} \langle (\bar{G}_j^0)^{-1}, LI_n - \bar{G}_j^0 \rangle + \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} - \ln \frac{\text{Det}(\bar{G}_j^0)^{-1} LI_n}{\text{Det}(\bar{G}_j^0)^{-1} \bar{G}_j^k} \\
 &= \langle (\bar{G}_j^0)^{-1}, LI_n - \bar{G}_j^0 \rangle + \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} - \ln \text{Det} \left((\bar{G}_j^k)^{-1} LI_n \right) \\
 &\stackrel{(4-1)}{\leq} \langle (\bar{G}_j^0)^{-1}, LI_n - \bar{G}_j^0 \rangle + \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} - \ln \text{Det} \left(\frac{1}{L} I_n LI_n \right) \\
 &= \langle (\bar{G}_j^0)^{-1}, LI_n - \bar{G}_j^0 \rangle + \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} \\
 &\stackrel{(4-1)}{\leq} \langle (\bar{G}_j^0)^{-1}, \frac{L}{\mu} \bar{G}_j^0 - \bar{G}_j^0 \rangle + \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} \\
 &= \frac{L}{\mu} \text{Tr}(I_n) - \text{Tr}(I_n) + \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} \leq n \frac{L}{\mu} + \sqrt{\frac{3}{2}} \frac{nL}{\mu} \frac{\ln \frac{3}{2}}{2} \\
 &= \left(1 + \frac{\ln \frac{3}{2}}{2} \sqrt{\frac{3}{2}} \right) \frac{nL}{\mu}.
 \end{aligned}$$

Consequentemente,

$$\begin{aligned}
 \sum_{i=0}^{k-1} \max_{j=1, \dots, m} (\varphi_j^i)^2 &\leq \sum_{j=1}^m \sum_{i=0}^{k-1} (\varphi_j^i)^2 \leq m (3 + \sqrt{6}) \left(1 + \frac{\ln \frac{3}{2}}{2} \sqrt{\frac{3}{2}} \right) \frac{nL}{\mu} \\
 &\leq 7 \frac{mnL}{\mu}.
 \end{aligned} \tag{A-15}$$

Finalmente, por (A-7) e pela desigualdade das médias[‡], obtemos

$$\begin{aligned}
\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* &\leq \frac{bL}{a\mu} \left(1 + \frac{Mr_{k-1}}{2}\right) \varphi_{k-1} \|d_{SD}(x^{k-1})\|_{\nabla^2 F_{\lambda_{k-1}^{SD}}(x^{k-1})}^* \\
&\leq \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \prod_{i=0}^{k-1} \frac{bL}{a\mu} \left(1 + \frac{Mr_i}{2}\right) \varphi_i \\
&\stackrel{(4-33)}{\leq} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \left[\left(\frac{b}{a}\right)^{2k} \left(\frac{L}{\mu}\right)^{2k} \left(\frac{3}{2}\right)^{\frac{k}{2}} \prod_{i=0}^{k-1} \varphi_i^2 \right]^{\frac{1}{2}} \\
&\leq \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \left[\left(\frac{b}{a}\right)^2 \left(\frac{L}{\mu}\right)^2 \left(\frac{3}{2}\right)^{\frac{1}{2}} \frac{1}{k} \sum_{i=0}^{k-1} \varphi_i^2 \right]^{\frac{k}{2}} \\
&\stackrel{(A-8)}{\leq} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \left[\left(\frac{b}{a}\right)^3 \left(\frac{L}{\mu}\right)^4 \left(\frac{3}{2}\right)^{\frac{3}{2}} \frac{1}{k} \sum_{i=0}^{k-1} \max_{j=1, \dots, m} (\varphi_j^i)^2 \right]^{\frac{k}{2}} \\
&\stackrel{(A-15)}{\leq} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \left[\left(\frac{b}{a}\right)^3 \left(\frac{L}{\mu}\right)^4 \left(\frac{3}{2}\right)^{\frac{3}{2}} \frac{1}{k} \frac{7mnL}{\mu} \right]^{\frac{k}{2}} \\
&= \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \left[\left(\frac{b}{a}\right)^3 \left(\frac{L}{\mu}\right)^5 7 \left(\frac{3}{2}\right)^{\frac{3}{2}} mn \frac{1}{k} \right]^{\frac{k}{2}} \\
&\leq \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \left[13 \left(\frac{b}{a}\right)^3 \left(\frac{L}{\mu}\right)^5 mn \frac{1}{k} \right]^{\frac{k}{2}}.
\end{aligned}$$

Portanto, por (4-34), temos

$$\begin{aligned}
\|d_{SD}(x^k)\|_{\nabla^2 F_{\lambda_k^{SD}}(x^k)}^* &\leq \left[13 \left(\frac{3L^2}{2\mu^2}\right)^3 \left(\frac{L}{\mu}\right)^5 mn \frac{1}{k} \right]^{\frac{k}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^* \\
&\leq \left(\frac{44L^{11}mn}{\mu^{11}k}\right)^{\frac{k}{2}} \|d_{SD}(x^0)\|_{\nabla^2 F_{\lambda_0^{SD}}(x^0)}^*.
\end{aligned}$$

■

[‡]A desigualdade das médias afirma que a média aritmética é maior que ou igual à média geométrica.