

Universidade Federal do Amazonas
Instituto de Ciências Exatas
Programa de Pós-Graduação em Matemática
Mestrado em Matemática

Um método de gradiente não
monótono para problemas de
otimização multiobjetivo com
restrições

Dainara Silva de Souza

Manaus - AM
Abril de 2024

Universidade Federal do Amazonas
Instituto de Ciências Exatas
Programa de Pós-Graduação em Matemática
Mestrado em Matemática

Dainara Silva de Souza

Um método de gradiente não monótono para problemas de otimização multiobjetivo com restrições

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

Orientador: Dr. Roberto Cristovão Mesquita Silva

Manaus - AM

Abril de 2024

Ficha Catalográfica

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

S729m Souza, Dainara Silva de
Um método de gradiente não monótono para problemas de
otimização multiobjetivo com restrições / Dainara Silva de Souza .
2024
52 f.: il. color; 31 cm.

Orientador: Roberto Cristovão Mesquita Silva
Dissertação (Mestrado em Matemática) - Universidade Federal do
Amazonas.

1. Método do Gradiente. 2. Convergência Linear. 3. Otimização
Multiobjetivo. 4. Busca Linear Não Monótona. 5. Otimalidade de
Pareto. I. Silva, Roberto Cristovão Mesquita. II. Universidade
Federal do Amazonas III. Título

Dedico este trabalho a você, familiar ou amigo, que contribuiu muito na minha caminhada. Sem vocês eu nada seria.

Agradecimentos

Em primeiro lugar agradeço a Deus, pela minha vida, minha saúde, pela minha cura, e por me permitir ultrapassar todos os obstáculos encontrados, que não foram poucos.

Ao meu marido e filho, por me encorajarem em todos os momentos em que pensei em desistir, ao meu marido por me dar todo o suporte para trilhar este mestrado, ao meu filho por ser a minha maior força, mesmo sem saber. Sem vocês na minha vida, nada disso seria possível.

Aos meus pais e irmãos, que me incentivaram nos momentos difíceis, principalmente ao meu irmão que me recebeu em sua casa durante esse processo.

Aos poucos amigos que criei, que sempre estiveram ao meu lado, pela amizade incondicional e pelo apoio demonstrado ao longo de todo o período de tempo em que me dediquei a este trabalho, em especial, ao meu amigo Alex por toda a contribuição, nunca vou me esquecer.

Ao professor Dr. Roberto Cristóvão, por ter sido além de meu orientador, um amigo, que sempre me entendeu diante de todas as adversidades.

E, por último, ao professor Dr. Rui Marques Carvalho por toda a disposição e ajuda que me dedicou durante o processo de construção do trabalho, o meu verdadeiro obrigado.

Resumo

Nesta dissertação, consideramos um método de gradiente não monótono para problemas de Otimização Multiobjetivo com restrições suaves. Sob suposições suaves, demonstramos a estacionariedade de Pareto do ponto de acumulação da sequência gerada por este método, e provamos a convergência da sequência completa para uma solução ótima de Pareto fraco do problema quando a função é convexa. Impondo algumas suposições sobre os gradientes das funções objetivo e as direções de busca linear fornecemos a convergência da sequência de valores da função objetivo para o valor ideal. O ponto inicial nos resultados de convergência estabelecidos aqui podem ser qualquer um no conjunto de restrições. Além disso, mostramos os resultados numéricos ao aplicar este método.

Palavras-chave: Método do Gradiente, Convergência Linear, Otimização Multiobjetivo, Busca Linear Não Monótona, Otimalidade de Pareto.

Abstract

In this dissertation, we consider a nonmonotone gradient method for Multiobjective Optimization problems with smooth constraints. Under mild assumptions, we demonstrate Pareto stationarity of the accumulation point of the sequence generated by this method, and we prove the convergence of the full sequence to a weak Pareto optimal solution of the problem is proven when the function is convex. Further, imposing some assumptions on the gradients of the objective functions and the search directions, we provide the linear convergence of the function value sequence to the optimal value. The initial point, in the our convergence results established can be any one in the constraint set. Furthermore, we show the numerical results when applying this method.

Keywords: Gradient Method; Linear Convergence; Multiobjective Optimization; Nonmonotone Line Search; Pareto Optimality.

Sumário

Introdução	10
1 Preliminares	14
1.1 O espaço \mathbb{R}^n	14
1.2 Continuidade em \mathbb{R}^n	19
1.3 Função Diferenciável	20
1.4 Elementos de análise Convexa	21
1.5 Otimização Vetorial	23
1.6 Ordenamento de soluções	23
1.7 Busca Linear	24
2 Otimização Multiobjetivo	27
2.1 Pareto Otimalidade	28
2.2 O Algoritmo de Gradiente Não Monótono	29
2.2.1 Algoritmo 1	30
3 Análise de Convergência	36
3.1 O caso não convexo	37
3.2 O caso convexo	39
4 Experimentos Numéricos	43
5 Considerações Finais	47
Referências Bibliográficas	48

Introdução

A Otimização Escalar é a área da Matemática que desenvolve métodos para encontrar os mínimos de funções reais $f : \mathbb{R}^n \rightarrow \mathbb{R}$ restritas a um subconjunto do domínio $C \subseteq \mathbb{R}^n$. O conjunto C normalmente é chamado conjunto viável [29].

A Otimização Multiobjetivo estende o conceito de otimização escalar, e consiste em minimizar simultaneamente as funções componentes $f_j : \mathbb{R}^n \subseteq \mathbb{R}, \forall j \in \{1, \dots, m\}$, de uma função vetorial $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ restrita a um subconjunto $C \subseteq \mathbb{R}^n$ [20].

Em problemas de otimização com um único objetivo, a meta é encontrar uma solução ótima global. Se a função objetivo desses problemas for multimodal, poderia existir mais de um ótimo global. Neste caso, todos os ótimos são equivalentes. Por outro lado, em problemas de Otimização Multiobjetivo, determinar o conjunto de soluções da fronteira de Pareto é tão importante quanto preservar a diversidade neste conjunto. Um algoritmo eficiente para otimização multiobjetivo deve considerar ambos os aspectos.

Um problema de Otimização Multiobjetivo trabalha com dois espaços (das variáveis e dos objetivos) ao invés de um. Problemas de objetivo simples trabalham unicamente no espaço de variáveis, pois procuram apenas uma solução no espaço de objetivos. Manter a diversidade em ambos espaços complica mais o problema, dado que a proximidade de duas soluções no espaço de variáveis não implica proximidade no espaço de objetivos.

O problema de otimização multiobjetivo pode ser formulado da seguinte maneira:

$$\min_{x \in C} F(x) \tag{1}$$

onde $C \subset \mathbb{R}^n$ é um conjunto de restrições e $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é uma função vetorial.

Este tipo de problema pode ser encontrado em diversas áreas, como engenharia, economia, finanças, radioterapia e assim por diante. Tendo em vista suas extensas aplicações, a pesquisa sobre os algoritmos numéricos para resolver problemas de otimização multiobjetivo tem recebido muita atenção e muitos métodos iterativos foram propostos, incluindo método do gradiente projetado [14] [40] [41], método de descida mais íngreme [2] [17], método do ponto proximal [8, 31, 30], método do gradiente conjugado [26], método de Newton [16, 38], método de região de confiança [7, 34] e assim por diante.

Entre esses métodos para resolver otimização multiobjetivo, métodos de escalarização [31, 4, 6] e métodos de descida [14, 17, 38, 18] são dois principais tipos diferentes de

abordagens. Baseado na técnica de escalarização, estes métodos calculam o ótimo de Pareto ou soluções de Pareto fraco, escolhendo alguns parâmetros com antecedência e reformulando o valor do vetor do problema original nos valores escalares parametrizados.

O método de ponderação é amplamente utilizado na técnica de escalarização, que minimiza uma combinação linear dos objetivos com o vetor de “pesos”. Porém, ao utilizar este método para alguns problemas, a maioria dos parâmetros pode dar origem a problemas escalares ilimitados (e, portanto, insolúveis).

Normalmente, os métodos de descida não requerem nenhuma informação de parâmetro, a priori, a regra de Armijo é uma estratégia de pesquisa usada com frequência para calcular o tamanho do passo de iteração t_k (ver, por exemplo, [14, 2, 33, 18]).

Dado $\gamma \in (0, 1)$, o tamanho do passo t_k é definido como:

$$t_k = \max\{2^{-j} : j \in \mathbb{N}, F(x_k + 2^{-j}v_k) \preceq F(x_k) + \gamma 2^{-j} JF(x_k)v_k\} \quad (2)$$

onde v_k é a direção de busca e $JF(x_k)$ é a matriz jacobiana de F em x_k . Recentemente Fliege e outros [18] estudaram o método do gradiente descendente para problemas de otimização multiobjetivo irrestrita com cada função objetivo sendo convexa e seu gradiente sendo Lipschitz contínuo, no qual outra regra de busca linear foi considerada:

Dado $\gamma \in (0, 1)$, o tamanho do passo t_k satisfaz:

$$t_k = \max\{2^{-j} : j \in \mathbb{N}, F(x_k + 2^{-j}v_k)\} \preceq F(x_k) + 2^{-j} JF(x_k)v_k + 2^{-j} \frac{\alpha}{2} \|v_k\|^2 \varrho \quad (3)$$

onde ϱ é o vetor de \mathbb{R}^m .

Sob as regras de busca linear acima, a sequência dos valores objetivos da função vetorial é estritamente monótona decrescente, e essa monotonicidade desempenha um papel importante na análise de convergência do algoritmo. Mas como afirmado por Grippo et al [21] para a otimização escalar, a imposição de monotonicidade dos valores da função pode diminuir consideravelmente a taxa de convergência nos estágios intermediários do processo de minimização.

Para melhorar esta situação, algumas técnicas de busca linear não monótona foram propostas em otimização escalar [21, 11, 36, 39], que foram verificados numericamente que podem aumentar a possibilidade de encontrar a solução ótima e melhorar a velocidade de convergência de algoritmos.

Observe que a técnica de busca linear não monótona introduzida por Zhang e Hager [39] que requer que uma média dos valores sucessivos da função diminua, tem se mostrado mais eficiente do que os esquemas monótonos ou não monótonos tradicionais em termos de menos avaliações de função e gradiente, em média.

Recentemente, técnicas de busca linear não monótona também foram aplicadas à otimização multiobjetivo [41, 27, 28, 33]. Particularmente, Fazzio e Schuverdt [15] estenderam a regra de pesquisa linear não monótona em [39] de otimização escalar para otimização multiobjetivo e alteraram a regra de Armijo [2] da seguinte forma:

Dado $\gamma \in (0, 1)$, $C_0 = F(x_0)$, $Q_0 = 1$, $0 \leq \eta_{min} \leq \eta_{max} \leq 1$, $\eta_k \in [\eta_{min}, \eta_{max}]$ o tamanho do passo t_k satisfaz:

$$t_k = \max\{2^{-j} : j \in \mathbb{N}, F(x_k + 2^{-j}v_k) \preceq C_k + \gamma 2^{-j} JF(x_k)v_k\} \quad (4)$$

e atualizar

$$C_{k+1} = \frac{\eta_k Q_k C_k + F(x_k + 1)}{Q_{k+1}} \quad e \quad Q_{k+1} = \eta_k Q_k + 1 \quad (5)$$

Com base na regra de busca linear [4], os autores em [15] consideraram um método de gradiente projetado não monótono para o problema de otimização multiobjetivo restrito (1) quando C é um conjunto fechado e convexo e F é uma função vetorial continuamente diferenciável. Eles mostraram a estacionariedade dos pontos de acumulação das sequências geradas pelo algoritmo proposto, e então estabeleceram a convergência dos pontos fracos de Pareto quando a função F é convexa.

Muito recentemente em [41], adotando a regra de busca linear não monótona [4] considerou-se um método de projeção do gradiente com sequencia somatória quadrada dada exogenamente no cálculo da direção de busca para o mesmo problema de Otimização Multiobjetivo quando F é convexo. Provou-se a convergência da sequência completa gerada pelo algoritmo para um ponto ideal de Pareto fraco. Além disso, sob alguma suposição apropriada de continuidade de Lipschitz do gradiente das funções objetivos, estabeleceu-se o resultado da convergência linear da proposta do método.

Nesta dissertação, considerou-se o problema de otimização multiobjetivo (1) no caso em que C é um conjunto fechado e convexo, e cada função f_i é continuamente diferenciável com o gradiente sendo Lipschitz contínuo. Aplicando a ideia na técnica de busca linear não

monótona (4) e (5) para a condição de diminuição suficiente (3), propõem-se um algoritmo não monótono para resolver o problema de otimização multiobjetivo aqui considerado. Sob hipóteses adequadas, mostramos que qualquer ponto de acumulação da sequência gerada pelo método proposto é Pareto estacionário, e toda a sequência converge para um ponto ótimo de Pareto fraco quando a função objetivo é convexa. Além disso, sob algumas suposições adicionais nos gradientes das funções objetivo, o resultado da convergência linear para este método também é estabelecida.

Este trabalho está organizado como segue. No capítulo 1 são descritas as notações e resultados preliminares que serão utilizados para a melhor compreensão. No capítulo 2 apresenta-se o conceito da Otimização Multiobjetivo, a definição de Pareto Otimalidade. No capítulo 3 prova-se a análise de convergência do algoritmo tanto para o caso convexo quanto para o caso não convexo. No capítulo 4 descrevemos o algoritmo aqui estudado bem como apresentamos alguns resultados referentes aos experimentos numéricos e no capítulo 5 conclui-se o trabalho e aponta-se sugestões para trabalhos futuros.

Capítulo 1

Preliminares

Neste capítulo, iremos introduzir alguns resultados preliminares, definições e provar alguns resultados que serão necessários para o desenvolvimento do restante dessa dissertação. Iniciamos falando um pouco sobre o espaço \mathbb{R}^n , bem como alguns tópicos de sua Topologia.

1.1 O espaço \mathbb{R}^n

Nesse trabalho utilizaremos os vetores do espaço \mathbb{R}^n como vetor(coluna), de modo que $x^T = (x_1, \dots, x_n)$, com cada $x_i \in \mathbb{R}$.

Definição 1.1. *Um produto interno em \mathbb{R}^n é uma aplicação que faz corresponder a cada par $x, y \in \mathbb{R}^n$ um número real, indicado por $\langle x, y \rangle$ de modo que, para quaisquer $x, y, z \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ se tenham :*

1. $\langle x, y \rangle = \langle y, x \rangle$;
2. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$;
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle = \langle x, \alpha y \rangle$;
4. $x \neq 0 \Rightarrow \langle x, x \rangle > 0$

Um exemplo de produto interno em \mathbb{R}^n é o produto interno canônico, que é dado por:

$$x^T y := \langle x, y \rangle = x_1 y_1 + \dots + x_n y_n,$$

onde $x^T = (x_1, \dots, x_n)$ e $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

Esse será o produto interno que faremos uso no decorrer dessa dissertação. Através do produto interno definido anteriormente, podemos definir a norma euclidiana de um vetor $x \in \mathbb{R}^n$, como: $\|x\| = \sqrt{x_1^2, \dots, x_n^2}$. O número $\|x\|$ chama-se a norma euclidiana do vetor x , e essa goza das seguintes propriedades:

1. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in \mathbb{R}^n$ (Desigualdade Triangular)
2. $\|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall x \in \mathbb{R}^n$ e $\forall \alpha \in \mathbb{R}$
3. $x \neq 0 \Rightarrow \|x\| > 0 \quad \forall x \in \mathbb{R}^n$

Definida, a norma euclidiana apresentaremos alguns conceitos a ela relacionados, tais como bola aberta, bola fechada, esfera e distância. Os quais formalizaremos a seguir:

Definição 1.2. Seja $a \in \mathbb{R}^n$ e $r > 0$ um número real, definimos:

1. **Bola aberta:** Conjunto dos pontos $\{x \in \mathbb{R}^n; \text{tais que } \|x - a\| < r\}$.
2. **Bola Fechada:** Conjunto dos pontos $\{x \in \mathbb{R}^n; \text{tais que } \|x - a\| \leq r\}$.
3. **Esfera:** Conjuntos dos pontos $\{x \in \mathbb{R}^n \text{ tais que } \|x - a\| = r\}$

Definição 1.3. Denotamos a função distância $d: \mathbb{R}^n \rightarrow \mathbb{R}$ de um ponto $x \in \mathbb{R}^n$ para um certo conjunto $C \subset \mathbb{R}^n$ como sendo

$$d_C(x) := \inf \{ \|x - c\| : c \in C \}.$$

Agora iremos introduzir algumas definições de seqüências no espaço \mathbb{R}^n , bem como alguns resultados a elas relacionados.

Definição 1.4. Uma seqüência em \mathbb{R}^n é uma aplicação $x: \mathbb{N} \rightarrow \mathbb{R}^n$, onde o valor que essa aplicação assume no número k é indicado como x^k e chama-se o k -ésimo termo da seqüência. Usaremos as notações $\{x^k\}$ ou $x^k_{k \in \mathbb{N}}$ para indicar uma seqüência cujo o k -ésimo termo é $x^k \in \mathbb{R}^n$.

Definição 1.5. Definimos uma subsequência de $\{x^k\}$ como sendo uma restrição da seqüência a um subconjunto infinito $\mathbb{N}^1 = k_1 < k_2 < \dots < k_j < \dots \subset \mathbb{N}$ e usaremos as seguintes notações em sua representação: $\{x^k\}_{k \in \mathbb{N}^1}$ ou ainda $\{x^{k_i}\}_{i \in \mathbb{N}}$.

Definição 1.6. Dizemos que uma sequência $\{x^k\}$ é limitada quando existe uma constante real $M > 0$ tal que $\|x^k\| \leq M \forall k \in \mathbb{N}$.

Definição 1.7. Diz-se que o ponto $a \in \mathbb{R}^n$ é limite da sequência de pontos $\{x^k\} \in \mathbb{R}^n$ quando, para todo $\epsilon > 0$ dado, existe $k_0 \in \mathbb{N}$ tal que $k > k_0 \Rightarrow \|x_k - a\| < \epsilon$. Neste caso, diz-se também que $\{x^k\}$ converge para a ou tende para a e escreve-se $\lim_{k \rightarrow \infty} x^k = a$ ou simplesmente $x^k \rightarrow a$.

Quando o limite acima mencionado existe, dizemos que a sequência $\{x^k\}$ é convergente. Caso contrário, dizemos que a sequência é divergente.

Definição 1.8. Dizemos que o ponto a é um valor de aderência para a sequência $\{x^k\}$ se existe alguma subsequência de $\{x^k\}$ que converge para a .

Um fato a ser observado é que uma sequência $\{x^k\}$ em \mathbb{R}^n equivale a n sequências de números reais. Assim para cada $k \in \mathbb{N}$ podemos escrever $(x^k)^T = (x_{k1}, \dots, x_{kn})$ em que cada sequência x_{ki} é uma sequência de números reais. Usando desse raciocínio podemos enunciar o seguinte teorema.

Teorema 1.9. Uma sequência $\{x^k\} \in \mathbb{R}^n$ converge para um ponto $a^T = (a_1, \dots, a_n)$ se, e somente se, para cada $i = 1, \dots, n$ tem-se $\lim_{k \rightarrow \infty} x_{ki} = a_i$, ou seja, cada coordenada de x^k converge para a coordenada correspondente de a .

Demonstração. Para a demonstração ver [24] pág. 16. □

Teorema 1.10. (Bolzano-Weierstrass) - Toda sequência limitada em \mathbb{R}^n possui uma subsequência convergente.

Demonstração. Ver em [24] pág. 11. □

Definição 1.11. Um ponto $a \in \mathbb{R}^n$ é um ponto de acumulação para o conjunto $X \subset \mathbb{R}^n$ se existe uma sequência de pontos $\{x^k\} \in X$ com $\lim_{k \rightarrow \infty} x^k = a$ e $x^k \neq a \forall k \in \mathbb{N}$.

Definição 1.12. Uma sequência (x_k) chama-se monótona quando se tem $x_k \leq x_{k+1}$ para todo $k \in \mathbb{N}$ ou então $x_{k+1} \leq x_k$ para todo k . No primeiro caso, diz-se que (x_k) é monótona não decrescente e, no segundo, que (x_k) é monótona não-crescente. Se, mais precisamente, tivermos $x_k < x_{k+1}$ (respect. $x_k > x_{k+1}$ para todo $k \in \mathbb{N}$, diremos que a sequência é crescente (respectivamente, decrescente).

Toda sequência monótona não-decrescente (respect. não-crescente) é limitada inferiormente (respect. superiormente) pelo seu primeiro termo. A fim de que ela seja limitada é suficiente que possua uma subsequência limitada. Com efeito, seja $(x_k)_{k \in \mathbb{N}'}$, uma subsequência limitada da sequência monótona (digamos, não-decrescente) (x_k) . Temos $x'_k \leq c$ para todo $k' \in \mathbb{N}'$. Dado qualquer $k \in \mathbb{N}$, existe $k' \in \mathbb{N}'$ tal que $k > k'$. Então $x_k \leq x'_k \leq c$.

O teorema seguinte dá uma condição suficiente para que uma sequência convirja.

Teorema 1.13. - Toda sequência monótona limitada é convergente.

Demonstração. Seja (x_k) monótona, digamos não-crescente, limitada. Escrevamos $X = \{x_1, \dots, x_k, \dots\}$ e $a = \sup X$. Afirmamos que $a = \lim x_k$. Com efeito, dado $\epsilon > 0$, o número $a - \epsilon$ não é cota superior de X . Logo existe $k_0 \in \mathbb{N}$ tal que $a - \epsilon < x_{k_0} \leq a$. Assim, $k > k_0 \Rightarrow a - \epsilon < x_{k_0} \leq x_k < a + \epsilon$ e daí $\lim x_k = a$ \square

Semelhantemente, se (x_k) é não-crescente, limitada então $\lim x_k$ é o ínfimo do conjunto dos valores de x_k .

A seguir, lembramos a definição de sequência quase-Fejér convergente, que tem sido amplamente utilizado para analisar os métodos gradiente e sub gradiente, veja, por exemplo, [14, 4, 15, 10].

Definição 1.14. Diz-se que uma sequência $\{u_k\} \subseteq \mathbb{R}^n$ é quase-Féjer convergente para um conjunto não vazio $U \subseteq \mathbb{R}^n$ se, para cada $u \in U$, existe uma sequência $\{\epsilon_k\} \subseteq \mathbb{R}_+$ com $\sum_{k=0}^{\infty} \epsilon_k < \infty$ tal que

$$\|u_{k+1} - u\|^2 \leq \|u_k - u\|^2 + \epsilon_k \quad (1.1)$$

Exemplo 1.15. Dadas as sequências $\{u_k\} \subseteq \mathbb{R}_+$, $u_k = \sqrt[k]{k}$ e $\{\epsilon_k\} \subset \mathbb{R}$ $\{\epsilon_k\} = \frac{1}{2^k}$, teremos que $\{u_k\}$ é quase-Fejér convergente ao conjunto \mathbb{R}_+ .

O resultado seguinte foi provado em ([22], teorema 4.1) para distâncias mais gerais, apresentamos aqui a prova para a distância Euclidiana.

Proposição 1.16. ([5], Teorema 1) Se $\{u_k\} \subseteq \mathbb{R}^n$ é quase-Féjer convergente para um conjunto não vazio $U \subseteq \mathbb{R}^n$, então $\{u_k\}$ é limitada. Além disso, se um ponto \bar{u} de $\{u_k\}$ pertence a U , então $\lim_{k \rightarrow \infty} u_k = \bar{u}$.

Demonstração. Tome $x \in U$, aplicando a definição [1.14](#) temos

$$\begin{aligned} \|u_k - x\|^2 &\leq \|u_{k-1} - x\|^2 + \varepsilon_{k-1} \\ &\leq \|u_{k-2} - x\|^2 + \varepsilon_{k-1} + \varepsilon_{k-2} \\ &\leq \dots \leq \|u_0 - x\|^2 + \sum_{l=0}^{k-1} \varepsilon_l \\ &\leq \|u_0 - x\|^2 + \beta \end{aligned}$$

onde $\beta = \sum_{k=0}^{\infty} \varepsilon_k < \infty$. Assim, segue que $\{u_k\}$ é limitada, já que

$$\|u^k - x\| \leq \sqrt{\|u^0 - x\|^2 + \beta}$$

e portanto,

$$u^k \in B(x, \sqrt{\|u^0 - x\|^2 + \beta})$$

Seja agora $u \in U$ um ponto de acumulação de $\{u_k\}$ e tome algum $\delta > 0$. Seja $\{u_{k_j}\}$ uma subsequência de $\{u_k\}$ tal que $\lim_{j \rightarrow \infty} u_{k_j} = u$. Visto que $u \in U$, existe $\{\varepsilon_k\}$ satisfazendo as propriedades da definição [1.14](#).

Como $\sum_{k=0}^{\infty} \varepsilon_k < \infty$, tomemos k_0 tal que $\sum_{k=k_0}^{\infty} \varepsilon_k \leq \frac{\delta}{2}$.

Seja \bar{k} tal que $l_{\bar{k}} \geq k_0$ e $\|u_{l_{\bar{k}}} - u\|^2 \leq \frac{\delta}{2}$. Então para algum $k \geq l_{\bar{k}}$,

$$\begin{aligned} \|u^k - u\|^2 &\leq \|u_{l_{\bar{k}}} - u\|^2 + \sum_{i=l_{\bar{k}}}^{k-1} \varepsilon_i \\ &\leq \frac{\delta}{2} + \sum_{i=k_0}^{\infty} \varepsilon_i \\ &\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta \end{aligned}$$

Como δ é arbitrário, isto segue que

$$u = \lim_{k \rightarrow \infty} u_k$$

□

Observação 1.17. *Toda sequência Fejér convergente é quase-Fejér convergente. Para isso, basta tomar $\varepsilon_k = 0$.*

A qualidade de convergência do algoritmo será discutida em termos de convergência linear. Lembre-se de que diz-se que uma sequência $\{u_k\} \subseteq \mathbb{R}^n$ converge linearmente para seu limite u (com taxa θ) se $\theta \in [0, 1)$ e existe algum $\alpha > 0$ que $\|u_k - u\| \leq \alpha\theta^k$ para todo k .

Uma vez abordados os principais conceitos relacionados a sequências em \mathbb{R}^n , apresentamos a definição de continuidade para uma função definida em um subconjunto do espaço \mathbb{R}^n .

1.2 Continuidade em \mathbb{R}^n

Uma aplicação $f : X \rightarrow \mathbb{R}^n$, definida no conjunto $X \subset \mathbb{R}^m$, associa a cada ponto $x \in X$ sua imagem $f(x) = (f_1(x), \dots, f_n(x))$. As funções $f_1, \dots, f_n : X \rightarrow \mathbb{R}$, chamam-se funções-coordenadas de f .

Definição 1.18. Dizemos que $f : X \rightarrow \mathbb{R}^n$ é contínua no ponto $a \in X$, quando:

$$\forall \epsilon > 0; \exists \delta > 0, x \in X, 0 < \|x - a\| < \delta \Rightarrow \|f(x) - f(a)\| < \epsilon$$

Ou seja, $\forall \epsilon > 0; \exists \delta > 0; f(B(a; \delta) \cap X) \subset B(f(a); \epsilon)$.

Note que, pela equivalência das normas, continuidade não depende de qual norma se utiliza. Dizemos que $f : X \rightarrow \mathbb{R}^n$ é contínua em X , quando f for contínua em todos os pontos de X .

Note que a implicação acima é equivalente a cada uma das seguintes afirmações:

1. $x \in B(a, \delta) \cap X \Rightarrow f(x) \in B(f(a), \epsilon) \cap Y$;
2. $f(B(a, \delta) \cap X) \subset B(f(a), \epsilon) \cap Y$.

Dado $X \subset \mathbb{R}^n$, verifica-se facilmente que são contínuas:

1. A aplicação identidade de X ;
2. qualquer aplicação constante $f : X \rightarrow Y \subset \mathbb{R}^m$;
3. a restrição, a um subconjunto de X , de qualquer aplicação contínua $f : X \rightarrow Y \subset \mathbb{R}^m$.

Teorema 1.19. *Uma aplicação $f : X \subset \mathbb{R}^n \rightarrow Y \subset \mathbb{R}^m$ é contínua em $a \in X$ se, e somente se, para toda vizinhança V de $f(a)$ em Y , $f^{-1}(V)$ é uma vizinhança de a em X . Em particular, f é contínua se, e somente se, para todo aberto U de Y , $f^{-1}(U)$ é aberto em X .*

Demonstração. Ver demonstração em [13] pág. 87 □

Observação 1.20. *No enunciado do teorema 1.19, pode-se substituir “aberto” por “fechado”, isto é, pode-se afirmar que: $f : X \subset \mathbb{R}^n \rightarrow Y \subset \mathbb{R}^m$ é contínua se, e somente se, para todo fechado F de Y , $f^{-1}(F)$ é fechado em X . Para provarmos isto, basta notarmos que, pelo Teorema 1.19, f é contínua se, e somente se, para todo subconjunto fechado F de Y , $f^{-1}(Y - F) \subset X$ é aberto em X . Porém, $f^{-1}(Y - F) = f^{-1}(Y) - f^{-1}(F) = X - f^{-1}(F)$, isto é, $f^{-1}(Y - F)$ é aberto em X se, e somente se, $f^{-1}(F)$ é fechado em X .*

Proposição 1.21. *Dados conjuntos $X \subset \mathbb{R}^n$ e $Y \subset \mathbb{R}^m$, uma aplicação $f : X \rightarrow Y$ é contínua em $a \in X$ se, e somente se, para toda sequência $(x_k)_{k \in \mathbb{N}}$ em X satisfazendo $x_k \rightarrow a$, tem-se $f(x_k) \rightarrow f(a)$.*

Demonstração. Ver demonstração em [13] pág. 88 □

1.3 Função Diferenciável

Definição 1.22. *Seja $U \subset \mathbb{R}^n$ um conjunto aberto. Dizemos que uma aplicação $f : U \rightarrow \mathbb{R}^m$ é diferenciável em $x \in U$ se existe uma transformação linear $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$, tal que a aplicação resto, $r = r(h)$, definida pela igualdade*

$$f(x + h) - f(x) = Th + r(h),$$

satisfaz

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

A aplicação f é dita diferenciável em $X \subset U$ quando é diferenciável em cada ponto de X . Diz-se, simplesmente, que f é diferenciável quando é diferenciável em U . É essencial, nesta definição, que o conjunto U seja aberto. De fato, neste caso, para todo $x \in U$, existe $\delta > 0$, tal que $B(x, \delta) \subset U$. Assim, para todo $h \in \mathbb{R}^n$ satisfazendo $\|h\| < \delta$, tem-se $x + h \in U$ e, portanto, $f(x + h)$ está bem definido.

A aplicação T da Definição [1.22](#), quando existe, é única. Isto decorre do seguinte fato. Dado $t \in \mathbb{R}$, $t \neq 0$, tem-se,

$$Th = \frac{T(th)}{t} = \frac{f(x+th) - f(x)}{t} \pm \frac{r(th) \|th\|}{\|h\|}$$

Logo,

$$Th = \lim_{t \rightarrow 0} \frac{f(x+th) - f(x)}{t},$$

o que prova a unicidade de T , uma vez que, nesta igualdade, o segundo membro depende apenas de f , x e h . A aplicação linear T é, então, chamada de derivada de f em x e é denotada por $f'(x)$. Segue-se imediatamente da definição de diferenciabilidade que qualquer aplicação constante $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ é diferenciável, sendo a derivada, em cada $x \in U$, a transformação linear nula $f'(x) = 0$.

Observação 1.23. Consideremos uma aplicação $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ definida num aberto $U \subset \mathbb{R}^n$ é diferenciável em $x \in U$. Neste caso, temos que a aplicação resto, $r(h) = f(x+h) - f(x) - f'(x)h$, satisfaz $\lim_{h \rightarrow 0} \frac{r(h)}{\|h\|} = 0$. Logo,

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

Sendo assim,

$$\lim_{h \rightarrow 0} f(x+h) = \lim_{h \rightarrow 0} (f(x) + f'(x)h + r(h)) = f(x),$$

donde f é contínua em x .

1.4 Elementos de análise Convexa

Nesta seção exibiremos resultados e definições relacionados à análise Convexa. Para isso dividiremos esta seção em duas subseções: na primeira trataremos de conjuntos convexos e de funções convexas.

Definição 1.24. Um conjunto $C \subset \mathbb{R}^n$ é dito convexo, se para todo λ no intervalo $[0,1]$ e para todo $x, y \in C$ tivermos

$$\lambda x + (1 - \lambda)y \in C.$$

Geometricamente, esta definição nos diz que o segmento de reta $[x, y]$ definido por,

$$[x, y] := \lambda x + (1 - \lambda)y : 0 \leq \lambda \leq 1,$$

está inteiramente contido em C , sempre que os pontos extremos x e y estão em C .

Exemplo 1.25. O espaço euclidiano \mathbb{R}^n e a bola aberta de centro $a \in \mathbb{R}^n$ e raio δ , denotada por $B(a, \delta)$, são conjuntos convexos.

Definição 1.26. Um conjunto $K \subset \mathbb{R}^n$ chama-se cone quando para todo $t \in \mathbb{R}_+$ obtemos:

$$d \in K \quad \Rightarrow \quad td \in K.$$

Definição 1.27. Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $\bar{x} \in C$. O cone normal (cone de direções normais) no ponto x em relação ao conjunto C é dado por

$$\aleph_C(\bar{x}) = \{d \in \mathbb{R}^n : \langle d, x - \bar{x} \rangle \leq 0 \quad x \in C\}.$$

Outro aspecto relevante a ser destacado é a definição de função convexa, que é um conceito bastante importante em otimização com o qual obtemos resultados também muito importantes.

Definição 1.28. (Funções Convexas) - Seja C um conjunto convexo. Uma função f é dita convexa em C se para todo $x, y \in C$ e $\lambda \in [0, 1]$ obtemos

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

Definição 1.29. O epígrafo da função $f : C \subset \mathbb{R}^n \rightarrow \mathbb{R}$ é definido como o conjunto

$$E_f = \{(x, b) \in C \times \mathbb{R} : f(x) \leq b\}$$

Definição 1.30. Seja $D \subset \mathbb{R}^n$ um conjunto não-vazio e $f : D \rightarrow \mathbb{R}$. Dizemos que um ponto $\bar{x} \in D$ é:

1. **Minimizador Global** de f em D se

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

2. **Minimizador Local** de f em D se existe uma vizinhança V de \bar{x} tal que

$$f(\bar{x}) \leq f(x) \quad \forall x \in V \cap D;$$

Observação 1.31. Se \bar{x} é minimizador global (respectivamente local) para a função f em D , então $f(\bar{x})$ é chamado valor ótimo global (respectivamente local). Perceba que o problema de minimizar uma função pode ter várias soluções (globais), isto é, vários minimizadores (globais). Porém cada problema apresenta apenas um único valor ótimo.

Dizemos que

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

é um problema de minimização convexo quando $D \subset \mathbb{R}^n$ é um conjunto convexo e a função $f : D \rightarrow \mathbb{R}$ é convexa em D . A seguir apresentamos um teorema de grande importância em análise convexa, uma vez que relaciona mínimos globais com a convexidade de uma função.

Teorema 1.32. Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \rightarrow \mathbb{R}$ uma função convexa em D . Então todo minimizador local de f em D é um minimizador global. Além disso, o conjunto dos minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.

Demonstração. Para uma demonstração ver [23] pág. 79. □

1.5 Otimização Vetorial

Trazemos aqui algumas definições e resultados relacionados à funções vetoriais, as quais denotaremos por $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Tais resultados darão suporte para a teoria a ser desenvolvida nos capítulos posteriores.

1.6 Ordenamento de soluções

Relembremos que o conjunto \mathbb{R} é totalmente ordenado, o que o diferencia dos demais espaços vetoriais. Uma vez que \mathbb{R} é bem ordenado, temos que dados quaisquer $x, y \in \mathbb{R}$ podemos determinar uma relação de ordem entre os mesmos, utilizando para tal os sinais \leq ou \geq bem como suas representações $<$ ou $>$ que são utilizadas quando x e y são necessariamente distintos. Em espaços onde não existe uma boa ordenação como a apresentada pelo conjunto dos números reais utilizamos a ideia de ordenação parcial, a qual definiremos a seguir.

Definição 1.33. Conjuntos Ordenados - Um conjunto A é dito totalmente ordenado com relação de ordem (\preceq), se dados quaisquer elementos $x, y, z \in A$ sempre vale $x \preceq y$ ou $y \preceq x$ e além disso as seguintes propriedades são satisfeitas:

1. $x \preceq x$ (reflexividade)
2. $x \preceq y$ e $y \preceq z \Rightarrow x \preceq z$ (transitividade)
3. $x \preceq y$ e $y \preceq x \Rightarrow x = y$ (anti-simetria)

Definição 1.34. Conjuntos parcialmente ordenados - Dizemos que A é parcialmente ordenado quando valem as propriedades acima, mas A não é totalmente ordenado. Em outras palavras, não é verdade que dados quaisquer $x, y \in A$ é possível estabelecer entre eles uma relação da forma $x \preceq y$ ou $x \succeq y$. No espaço \mathbb{R}^n a ordenação é dada por:

$$x \preceq y \Rightarrow \{x_i \leq y_i, i = 1, 2, \dots, n\}$$

$$x \prec y \Rightarrow \{x_i < y_i, i = 1, 2, \dots, n\}$$

$$x = y \Rightarrow \{x_i = y_i, i = 1, 2, \dots, n\}.$$

O operador \neq é dado por:

$$x \neq y \Rightarrow \{x_i \neq y_i, \text{ para algum } i = 1, 2, \dots, n\}.$$

Exemplo 1.35. Considerem-se os vetores $x, y, z \in \mathbb{R}^3$:

$$x = [3 \quad 5 \quad 7]$$

$$y = [2 \quad 4 \quad 6]$$

$$z = [3 \quad 4 \quad 5]$$

Verifique que não há uma ordenação total dos vetores x, y, z . Dessa forma, temos que o espaço \mathbb{R}^n com a ordem acima é parcialmente ordenado, porém não é totalmente ordenado, visto que no Exemplo acima os pontos z e x não puderam ser comparados.

1.7 Busca Linear

Regra de Armijo - no método de Armijo procuramos uma boa redução da função ao longo da direção d sem tentar minimizá-la. Considerando uma função diferenciável $f : \mathbb{R}^n \rightarrow \mathbb{R}$, um ponto $\bar{x} \in \mathbb{R}^n$, uma direção de descida $d \in \mathbb{R}^n$ e $\eta \in (0, 1)$, basicamente a regra de Armijo nos garante que existe $\delta > 0$ tal que:

$$f(\bar{x} + td) \leq f(\bar{x}) + \eta t \nabla f(\bar{x})^T d$$

para todo $t \in [0, \delta)$

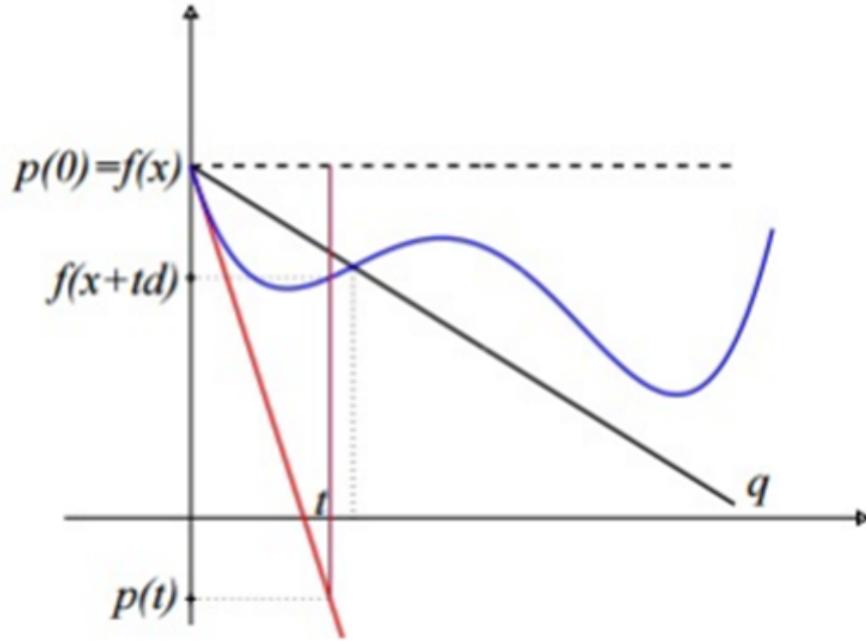


Figura 1.1: Interpretação da Condição de Armijo (Imagem: Captura de tela/ Erasmo Fonseca)

A figura [1.1](#) nos dá uma interpretação geométrica para a busca de Armijo. Considerando que a função $\varphi : [0, \infty) \rightarrow \mathbb{R}$ dada por $\varphi(t) = f(\bar{x} + td)$, a aproximação de primeira ordem de φ em torno de $t = 0$ é $p(t) = \varphi(0) + t\varphi'(0) = f(\bar{x}) + t\nabla f(\bar{x})^T d$. Nesta figura, a reta $q(t)$ é dada por $q(t) = f(\bar{x}) + \eta t \nabla f(\bar{x})^T d$. A condição de Armijo é satisfeita para os pontos tais que φ está abaixo de q .

Tendo do ponto de vista computacional quanto do ponto de vista teórico, é importante que o tamanho do passo t , satisfazendo [1.7](#), não seja muito pequeno. Uma maneira de conseguir isso, é iniciar com $t = 1$ e, se necessário, reduzir até que [1.7](#) seja satisfeita. Podemos ver isso de maneira sintetizada no algoritmo abaixo:

Algoritmo Busca de Armijo - Seja $\bar{x} \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ (direção de descida), $\gamma, \eta \in (0, 1)$.

Passo 1: Defina $t = 1$.

Passo 2: Enquanto $f(\bar{x} + td) \leq f(\bar{x}) + \eta t \nabla f(\bar{x})^T d$, faça $t = \gamma t$.

Embora não encontre um ponto próximo a um minimizador unidirecional, o método de Armijo é muito eficiente para algoritmos bem projetados, pois faz um número muito pequeno de cálculo de função, sendo portanto muito rápido.

O sucesso dos algoritmos que resolvem um problema de otimização dependem da escolha adequada da direção d_k e do tamanho do passo t_k . A diferença essencial entre os métodos está na escolha da direção d_k , pois este exerce uma grande influência no processo iterativo $x_{k+1} = x_k + t_k d_k$.

Capítulo 2

Otimização Multiobjetivo

De acordo com [35], a maior parte dos problemas reais encontrados na área de otimização envolve a obtenção de diversas metas que devem ser atingidas simultaneamente. Elas geralmente são conflitantes, ou seja, uma função objetivo f_1 é conflitante com uma outra função f_2 quando não é possível melhorar o valor de f_1 sem piorar o valor da função f_2 logo não existe uma solução única que otimize todas ao mesmo tempo. Para tal classe de problemas devemos buscar um conjunto de soluções eficientes.

Na vida real é comum a existência de problemas de otimização que consideram mais de um objetivo. Um exemplo de um problema com objetivos conflitantes é a tarefa de comprar um computador, a aquisição ótima é aquela que fornece o custo mínimo enquanto maximiza o desempenho do equipamento, estes objetivos são conflitantes entre si, uma vez que existirão desde computadores com elevado custo e desempenho até aqueles com baixo custo e desempenho. Assim, nenhuma solução que tenha menor custo e desempenho pode ser considerada como superior a outra com maior custo e desempenho.

Contudo, dentre todas as configurações de equipamentos existem algumas que são superiores a outras, isto é, apresentam desempenho maior ou equivalente por um custo menor ou igual. Estas configurações (soluções) que superam outras são conhecidas como soluções não dominadas, enquanto que as configurações que são superadas por pelo menos uma são conhecidas como soluções dominadas [12].

Portanto, uma solução razoável para um problema multiobjetivo é uma solução que não seja dominada por qualquer outra solução. Melhor ainda é ter como resultado o conjunto das soluções não-dominadas.

Problemas dessa natureza são chamados de problemas de otimização multiobjetivo por

envolverem minimização (ou maximização) simultânea de um conjunto de objetivos satisfazendo a um conjunto de restrições. Neste caso, a tomada de decisão será de responsabilidade do analista, que deverá ponderar os objetivos globais do problema e escolher uma entre as soluções do conjunto de soluções eficientes. [1]

Para [9], a otimização multiobjetivo pode ser definida como o problema de achar um vetor de variáveis de decisão cujos elementos representam as funções objetivos. Essas funções formam uma descrição matemática do critério de otimalidade que estão em conflito umas com as outras. Neste caso, o termo “otimizar” significa encontrar um conjunto de soluções que não podem ser melhoradas simultaneamente para o analista.

Enquanto que na otimização mono-objetivo uma solução ótima é claramente identificada, pois o espaço de solução é ordenado, na otimização multiobjetivo, por outro lado, há um conjunto de alternativas, geralmente conhecidas como soluções Pareto-ótimas, que, para [19], também podem ser denominadas como soluções eficientes, ou conjunto admissível do problema.

2.1 Pareto Otimalidade

Segundo [32], o conceito de Pareto-ótimo constitui a origem da busca na otimização multiobjetivo, uma vez que surge da necessidade de comparação entre vetores do \mathbb{R}^n . Desta problemática, tomamos os vetores que “melhoram os resultados” das funções objetivos ao mesmo tempo. Estes elementos irão pertencer ao que chamamos de conjunto **Pareto Ótimo**. Sendo \mathbb{R} o conjunto dos números reais. Definimos por \mathbb{R}_+ o conjunto dos números reais não-negativos e por \mathbb{R}_{++} o conjunto dos números reais estritamente positivos.

Aqui consideramos o problema de otimização multiobjetivo [1] com $C \subseteq \mathbb{R}^n$ sendo um conjunto não vazio fechado e convexo e $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sendo uma função vetorial continuamente diferenciável em um subconjunto aberto de C denotado por:

$$F := (f_1, f_2, \dots, f_m)^T, \quad (2.1)$$

onde cada f_i , $i = 1, \dots, m$, é uma função de valor real e seu gradiente é Lipschitz contínuo com constante $L_i > 0$ em C , ou seja,

$$\| \nabla f_i(x) - \nabla f_i(y) \| \leq L_i \| x - y \| \quad \forall x, y \in C$$

Nós denotamos por $L_{max} := \max\{L_1, \dots, L_m\}$. Diz-se que a função multiobjetivo F é continuamente diferenciável ou convexa se cada função componente f_i com $i = 1, \dots, m$ é continuamente diferenciável ou convexa. Usamos a $JF(x)$ para denotar a matriz jacobiana de F em x , ou seja,

$$JF(x) = (\nabla f_1(x), \dots, \nabla f_m(x))^T.$$

Definição 2.1. Um ponto $x^* \in C$ é dito

1. um ponto ótimo de Pareto (ou solução eficiente de Pareto) de F em C se não existir $x \in C$ tal que $F(x) \preceq F(x^*)$ e $F(x) \neq F(x^*)$,
2. um ponto ótimo de Pareto fraco (ou solução eficiente de Pareto fraca) de F em C se não existir $x \in C$ tal que $F(x) \prec F(x^*)$,
3. um ponto estacionário de Pareto (ou um ponto crítico de Pareto) de F em C se

$$JF(x^*)(C - x^*) \cap (-\mathbb{R}_{++}^m) = \emptyset.$$

Observação 2.2. É conhecido que todo ponto ótimo de Pareto também é um ponto ótimo de Pareto fraco. O ponto ótimo de Pareto fraco também é um ponto estacionário de Pareto mas o inverso nem sempre é verdadeiro. No entanto, se F for convexa, então a estacionariedade de Pareto implica uma otimalidade de Pareto fraco.

2.2 O Algoritmo de Gradiente Não Monótono

Antes de apresentarmos o algoritmo de gradiente não monótono para o problema de otimização multiobjetivo propriamente dito, descrevemos abaixo o algoritmo estudado por Zhang e Hager [39] de onde deriva a técnica aqui estudada.

Então seja:

Inicialização - Escolha a estimativa inicial x_0 e os parâmetros $0 \leq \eta_{min} \leq \eta_{max} \leq 1$, $0 < \delta < \sigma < 1 < \rho$ e $\mu > 0$. Defina $C_0 = f(x_0)$, $Q_0 = 1$, e $k = 0$

Teste de Convergência - Se $\nabla f(x_k)$ for suficientemente pequeno, então pare.

Atualização da busca linear - Defina $x_{k+1} = x_k + \alpha_k d_k$ onde α_k satisfaz a (não monótona) Condições de Wolfe:

$$f(x_k + \alpha_k d_k) \leq C_k + \delta \alpha_k \nabla f(x_k) d_k, \quad (2.2)$$

$$\nabla f(x_k + \alpha_k d_k) d_k \geq \sigma \nabla f(x_k) d_k, \quad (2.3)$$

ou as condições Armijo (não monótonas):

$$\alpha_k = \bar{\alpha}_k \rho^{hk}, \quad (2.4)$$

onde $\bar{\alpha}_k > 0$ é a etapa experimental, e hk é o maior número inteiro tal que [2.2](#) é válido e $\alpha_k \leq \mu$.

Atualização de custos - Escolha $\eta_k \in [\eta_{min}, \eta_{max}]$ e defina

$$Q_{k+1} = \eta_k Q_k + 1, \quad C_{k+1} = \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}}. \quad (2.5)$$

Substitua k por $k + 1$ e retorne ao teste de convergência.

Combinando a técnica de busca linear não monótona de Zhang e Hager [\[39\]](#) acima descrita com a condição de descida [\[3\]](#) que foi recentemente considerada em [\[18\]](#), propôs-se o algoritmo de gradiente não monótono para problema de otimização multiobjetivo [2.2.1](#).

Este algoritmo é formalmente declarado como segue.

2.2.1 Algoritmo 1

PASSO 1: Escolha os parâmetros $\gamma \in (0, 1)$ e $0 \leq \eta_{min} \leq \eta_{max} \leq 1$. Seja $x_0 \in C$ um ponto inicial arbitrário. Defina $C_0 = F(x_0)$, $Q_0 = 1$ e $k = 0$.

PASSO 2: Se $\nabla f_i(x_k) = 0$ para algum $i \in \{1, \dots, m\}$, então pare. Caso contrário, calcule a direção de busca v_k :

$$v_k := \operatorname{argmin}_{v \in C - x_k} \varphi_k(v) \quad (2.6)$$

onde $\varphi_k(v) = \max_{1 \leq i \leq m} \langle \nabla f_i(x_k), v \rangle + \frac{\|v\|^2}{2}$.

PASSO 3: Se $v_k = 0$, então pare. Caso contrário, prossiga para o PASSO 4.

PASSO 4: Calcule o tamanho do passo $t_k \in (0, 1]$ como o máximo de:

$$T_k := \{t = \frac{1}{2^j} | j \in \mathbb{N}, F(x_k + tv_k) \preceq C_k + tJF(x_k)v_k + \frac{\gamma t}{2} \|v_k\|^2 \varrho\} \quad (2.7)$$

onde $\varrho = (1, \dots, 1)^T \in \mathbb{R}^m$. Em seguida, defina:

$$x_{k+1} = x_k + t_k v_k \quad (2.8)$$

PASSO 5: Escolha $\eta_k \in [\eta_{min}, \eta_{max}]$ e defina:

$$Q_{k+1} = \eta_k Q_k + 1, \quad C_{k+1} = \frac{\eta_k Q_k C_k + F(x_{k+1})}{Q_{k+1}} \quad (2.9)$$

Defina $k = k + 1$ e volte para o PASSO 2.

Observe que, (pode-se também ver [15]), para cada k , C_{k+1} pode ser reescrito equivalentemente como:

$$C_{k+1} = \frac{(\eta_k Q_k + 1)C_k + F(x_{k+1}) - C_k}{Q_{k+1}} = C_k + \frac{F(x_{k+1}) - C_k}{Q_{k+1}} \quad (2.10)$$

e assim,

$$C_k - C_{k+1} = \frac{C_k - F(x_{k+1})}{Q_{k+1}} \quad (2.11)$$

Observe que se $\eta_k = 0$ para cada k , então $C_k = F(x_k)$ e a busca linear [2.7] se reduz a monótono que foi recentemente considerada em [18] para o método de descida do gradiente.

Enquanto que se $\eta_k = 1$, para cada k , então $C_k = \frac{1}{k+1} \sum_{i=0}^k F(x_i)$ é a média de todos os valores da função anterior.

Por isso, semelhante à seleção dos parâmetros envolvidos em outras regras de busca linear não monótona ([41, 15]), a escolha de η_k controla o grau de não monotonicidade da busca linear [2.7] (a busca linear se aproxima muito da monótono conforme η_k se aproxima de 0, e o esquema torna-se mais não monótono conforme η_k se aproxima de 1).

Agora, ilustramos a validade dos critérios de parada no Algoritmo [2.2.1]. Para isso, precisamos primeiro formular algumas desigualdades úteis. Observe que a função φ_k em [2.7] é fortemente convexa. Então, segue da condição de otimalidade de primeira ordem para $\min_{v \in C - x_k} \varphi_k(v)$ que existe $u_k \in \partial \varphi_k(v_k)$ tal que:

$$\langle u_k, v - v_k \rangle \geq 0, \forall v \in C - x_k \quad (2.12)$$

Então, pela expressão de φ_k e a fórmula para o sub diferencial do máximo de funções convexas (ver, por exemplo [37]), existem, $\emptyset \neq J_k \subseteq \{1, \dots, m\}$ e $\lambda_j^k > 0$ com $j \in J_k$ tal

que:

$$\sum_{j \in J_k} \lambda_j^k = 1, \langle \nabla f_j(x_k), v_k \rangle = \max_{1 \leq i \leq m} \langle \nabla f_i(x_k), v_k \rangle, \forall j \in J_k \quad (2.13)$$

e

$$u_k = v_k + \sum_{j \in J_k} \lambda_j^k \nabla f_j(x_k) \quad (2.14)$$

Combinando [2.12](#) e [2.14](#) resulta:

$$\left\langle v_k + \sum_{j \in J_k} \lambda_j^k \nabla f_j(x_k), v - v_k \right\rangle \geq 0, \forall v \in C - x_k \quad (2.15)$$

Apresentaremos algumas propriedades básicas do algoritmo [2.2.1](#) que serão usadas para a análise de convergência no próximo capítulo.

Proposição 2.3. *Se o algoritmo [2.2.1](#) parar na iteração k , então x_k é um ponto estacionário de Pareto, e ainda um ponto ótimo de Pareto fraco quando F é convexa.*

Demonstração. Tome $k \in \mathbb{N}$. Se $\nabla f_{i_0}(x_k) = 0$ para algum $i_0 \in \{1, \dots, m\}$, então para qualquer $x \in C$, $\langle \nabla f_{i_0}(x_k), x - x_k \rangle = 0$ e assim, $J_F(x_k)(C - x_k) \cap (-\mathbb{R}_{++}^m) = \emptyset$, i.e, x_k é um ponto estacionário de Pareto.

No caso quando $v_k = 0$, segue-se de [2.15](#) que:

$$\left\langle \sum_{j \in J_k} \lambda_j^k \nabla f_j(x_k), x - x_k \right\rangle \geq 0, \forall x \in C \quad (2.16)$$

Isto implica que, para qualquer $x \in C$, existe $j \in J_k \subseteq \{1, \dots, m\}$ tal que $\langle \nabla f_j(x_k), x - x_k \rangle \geq 0$. Consequentemente, $J_F(x_k)(C - x_k) \cap (-\mathbb{R}_{++}^m) = \emptyset$, e x_k é um ponto estacionário de Pareto.

Uma conclusão adicional da proposição [2.3](#) segue imediatamente da equivalência entre Pareto estacionariedade e a otimalidade de Pareto fraca de x_k quando a função objetivo F é convexa. \square

Daqui em diante, supomos que o Algoritmo [2.2.1](#) não possui terminação finita. O resultado a seguir mostra a boa definição do Algoritmo [2.2.1](#) e uma relação útil entre a sequencia gerada $\{x_k\}$ e $\{C_k\}$, onde a prova da conclusão $F(x_k) \preceq C_k$ baseia-se em Zhang e a técnica de Hager para otimização escalar [\[39\]](#).

Lema 2.4. - A regra de busca linear não monótona [2.7](#) é válida e, portanto, o algoritmo [2.2.1](#) está bem definido. A sequência $\{x_k\}$ gerada por ele pertence a C e $F(x_k) \preceq C_k$ para cada $k \in \mathbb{N}$.

Demonstração. Vamos provar por indução. Primeiro considere o caso em que $k = 0$. Usando a continuidade de Lipschitz de ∇f_i com $i \in \{1, \dots, m\}$, pode-se obter que para todo $t \in (0, \frac{\gamma}{L_i}]$,

$$\begin{aligned} f_i(x_0 + tv_0) &\leq f_i(x_0) + t\nabla f_i(x_0)^T v_0 + \frac{L_i}{2} \|tv_0\|^2 \\ &\leq f_i(x_0) + t\nabla f_i(x_0)^T v_0 + \frac{\gamma^t}{2} \|v_0\|^2 \end{aligned}$$

Assim temos que, $\forall t \in (0, \frac{\gamma}{L_{max}}]$,

$$F(x_0 + tv_0) \preceq F(x_0) + tJF(x_0)v_0 + \frac{\gamma^t}{2} \|v_0\|^2 \varrho \quad (2.17)$$

Observe que $C_0 = F(x_0)$. Então [2.17](#) implica que a regra de busca linear [2.7](#) no algoritmo [2.2.1](#) vale para $k = 0$, e assim o tamanho do passo t_0 pode ser calculado pelo fato de que no esquema de retrocesso o tamanho do passo começa em 1 e é reduzido pela metade a cada vez. E pode-se verificar que $x_1 = x_0 + t_0 v_0 \in C$. Observando que a iteração inicial x_0 pertence a C , $v_0 \in C - x_0$ e a convexidade de C .

Além disso, pela definição da direção de busca e pelo fato de que $0 \in C - x_0$, tem-se que

$$\phi_0(v_0) = \max_{1 \leq i \leq m} \langle \nabla f_i(x_0), v_0 \rangle + \frac{\|v_0\|^2}{2} \leq \phi_0(0) = 0$$

o que significa que,

$$t_0 JF(x_0)v_0 + \frac{\gamma^{t_0}}{2} \|v_0\|^2 \varrho \preceq t_0 \left(JF(x_0)v_0 + \frac{\|v_0\|^2}{2} \varrho \right) \preceq 0 \quad (2.18)$$

onde a primeira desigualdade vale porque $\gamma \in (0, 1)$. Isso, junto com a regra de busca linear [2.7](#) (para $k = 0$) e a iteração de atualização [2.8](#), implica que $F(x_1) \preceq C_0 = F(x_0)$. Agora vamos considerar o caso $k = 1$. Sem perda de generalidade, assumimos que $F(x_k) \preceq C_{k-1}$ e $x_k \in C$. Da mesma forma que obtivemos [2.17](#), pode-se obter usando a continuidade de Lipschitz de ∇f_i , que

$$F(x_k + tv_k) \preceq F(x_k) + tJF(x_k)v_k + \frac{\gamma^t}{2} \|v_k\|^2 \varrho \quad (2.19)$$

$\forall t \in (0, \frac{\gamma}{L_{max}}]$ Para cada $i = \{1, \dots, m\}$ defina a função $D_{k,i} : \mathbb{R} \rightarrow \mathbb{R}$ por:

$$D_{k,i}(t) = \frac{tC_k^i + f_i(x_k)}{t + 1}$$

onde C_{k-1}^i é o i -ésimo componente de C_{k-1} . É fácil calcular que

$$D'_{k,i}(t) = \frac{C_{k-1}^i - f_i(x_k)}{(t + 1)^2} \geq 0$$

para todo t (devido a suposição $F(x_k) \preceq C_{k-1}$) e assim,

$$f_i(x_k) = D_{k,i}(0) \preceq D_{k,i}(\eta_{k-1}Q_{k-1}) = \frac{\eta_{k-1}Q_{k-1}C_{k-1}^i + f_i(x_k)}{\eta_{k-1}Q_{k-1} + 1 = C_k^i}$$

o que é equivalente a $f(x_k) \preceq C_k$. Combinando isso com [2.19](#) resulta que a regra de busca linear [2.7](#) vale para k . Então, o tamanho do passo t_k pode ser obtido e $x_{k+1} := x_k + t_k v_k$, e $x_{k+1} \in C$ também pode ser verificado facilmente observando que $v_k \in C - x_k$. \square

Observação 2.5. *Similarmente a [2.18](#) na prova do Lema [2.4](#), para todo $k \in \mathbb{N}$, pode-se obter $t_k JF(x_k)v_k + \frac{\gamma t_k}{2} \|v_k\|^2 \varrho \leq 0$ por $\phi_k(v_k) \preceq 0$, e assim $F(x_{k+1}) \preceq C_k$ devido a [2.7](#).*

Aplicando isso a [2.10](#), segue que para cada k , $C_{k+1} \preceq C_k$, ou seja, $\{C_k\}$ é uma sequência não crescente em \mathbb{R}^m .

Em seguida mostraremos que o tamanho do passo $\{t_k\}$ gerado no algoritmo [2.2.1](#) possui um limite inferior.

Lema 2.6. *O tamanho do passo no algoritmo [2.2.1](#) satisfaz sempre que $t_k \geq t_{min} \equiv \min\{\frac{\gamma}{2L_{min}}, 1\}$.*

Demonstração. Fixe qualquer $k \in \mathbb{N}$. Pela definição de t_k , sabemos que $2t_k$ não satisfaz a desigualdade na regra de busca linear não monótona [2.7](#) do algoritmo [2.2.1](#). Portanto, existe um índice $i \in \{1, \dots, m\}$ tal que

$$f_i(x_k + 2t_k v_k) > C_k^i + 2t_k \nabla f_i(x_k)^T v_k + \frac{\gamma}{2} 2t_k \|v_k\|^2 \geq f_i(x_k) + 2t_k \nabla f_i(x_k)^T v_k + \gamma t_k \|v_k\|^2 \quad (2.20)$$

onde a segunda desigualdade vale graças ao Lema [2.4](#). Por outro lado, pela Lipschitz

continuidade de ∇f_i , temos

$$f_i(x_k + 2t_k v_k) \leq f_i(x_k) + 2t_k \nabla f_i(x_k)^T v_k + \frac{L_i}{2} \|t_k v_k\|^2 \quad (2.21)$$

Combinando [2.20](#) e [2.21](#), obtém-se

$$t_k \geq \frac{\gamma}{2L_i} \geq \frac{\gamma}{2L_{max}}$$

A conclusão segue então observando que t_k é menor que 1. □

Capítulo 3

Análise de Convergência

Neste capítulo nos concentramos na propriedade de convergência do Algoritmo [2.2.1](#) e consideramos quando a função F é não convexa, respectivamente, convexa. O seguinte resultado será utilizado na análise de convergência.

Proposição 3.1. *Suponha que $\eta_{max} < 1$. Então para todo $x \in C$ e $k \in \mathbb{N}$, existe $\{\lambda_j^k\}_{j=1}^m \subseteq [0, 1]$ satisfazendo $\sum_{j=1}^m \lambda_j^k = 1$, e*

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 + 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), x - x_k \rangle + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j)$$

Demonstração. Seja $x \in C$ e $k \in \mathbb{N}$. Por [2.8](#) temos que

$$\begin{aligned} \|x_{k+1} - x\|^2 &= \|x_k - x\|^2 - 2t_k \langle v_k, x - x_k \rangle + t_k^2 \|v_k\|^2 \\ &= \|x_k - x\|^2 - 2t_k \langle v_k, x - x_k - v_k \rangle + t_k(t_k - 2) \|v_k\|^2 \end{aligned} \quad (3.1)$$

Tomando $v = x - x_k$ em [2.15](#) e fazendo $\lambda_j^k = 0$ para $j \notin J_k$, pode-se obter:

$$2t_k \langle v_k, x - x_k - v_k \rangle \geq 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), v_k \rangle - 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), x - x_k \rangle$$

Usando $F(x_{k+1}) \leq C_k + t_k JF(x_k)v_k + \frac{\gamma t_k}{2} \|v_k\|^2$ na última desigualdade, e observando

que $F(x_{k+1}) \preceq C_k$ e $\sum_{j=1}^m \lambda_j^k = 1$ com cada $0 \leq \lambda_j^k \leq 1$ (ver [\[5\]](#)), segue que:

$$\begin{aligned}
& 2t_k \langle v_k, x - x_k - v_k \rangle \\
& \geq 2 \sum_{j=1}^m \lambda_j^k (f_j(x_{k+1}) - C_k^j - \frac{\gamma t_k}{2} \|v_k\|^2) - 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), x - x_k \rangle \\
& \geq 2 \sum_{j=1}^m (f_j(x_{k+1}) - C_k^j) - 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), x - x_k \rangle - \gamma t_k \|v_k\|^2
\end{aligned} \tag{3.2}$$

Aplicando [\[3.1\]](#) a [\[3.1\]](#), obtemos

$$\begin{aligned}
& \|x_{k+1} - x\|^2 \leq \\
& \leq \|x_k - x\|^2 + 2 \sum_{j=1}^m (C_k^j - f_j(x_{k+1})) + 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), x - x_k \rangle + t_k(\gamma + t_k - 2) \|v_k\|^2 \\
& \leq \|x_k - x\|^2 + 2 \sum_{j=1}^m (C_k^j - f_j(x_{k+1})) + 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), x - x_k \rangle
\end{aligned} \tag{3.3}$$

onde a última desigualdade é válida, pois γ e t_k são menores ou iguais a 1. Além disso, pela definição de Q_{k+1} em [\[2.9\]](#), pode-se calcular que:

$$Q_{k+1} = 1 + \sum_{j=0}^k \prod_{l=0}^j \eta_{k-l} \leq 1 + \sum_{j=0}^k \eta_{max}^{j+1} \leq \sum_{j=0}^{\infty} \eta_{max}^j \leq \frac{1}{1 - \eta_{max}} \tag{3.4}$$

Por isso e por [\[2.11\]](#) obtém-se então que

$$\sum_{j=1}^m (C_k^j - f_j(x_{k+1})) = \sum_{j=1}^m Q_{k+1} (C_k^j - C_{k+1}^j) \leq \frac{1}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j) \tag{3.5}$$

Assim, a desigualdade desejada pode ser obtida combinando [\[3.3\]](#) e [\[3.5\]](#) □

3.1 O caso não convexo

Teorema 3.2. *Assuma que $\eta_{max} < 1$ e que F é limitada inferiormente. Então todo ponto de acumulação, se houver, da sequência $\{x_k\}$ gerada pelo algoritmo [\[2.2.1\]](#) é um ponto Pareto estacionário.*

Demonstração. Seja x^* um ponto de acumulação de $\{x_k\}$ como $\{x_k\} \subseteq C$ (pelo Lema [\[2.4\]](#)

e C é fechado, segue que $x^* \in C$. Suponha que x^* não seja um ponto Pareto estacionário. Então, por definição, existe $\hat{x} \in C$ tal que $JF(x^*)(\hat{x} - x^*) \in -\mathbb{R}_{++}^m$, isto é,

$$\max_{1 \leq i \leq m} \langle \nabla f_i(x^*), \hat{x} - x^* \rangle < 0 \quad (3.6)$$

Aplicando a proposição [3.1](#) com $x = \hat{x}$, temos que para todo k ,

$$\|x_{k+1} - \hat{x}\|^2 \leq \|x_k - \hat{x}\|^2 + 2t_k \max_{1 \leq i \leq m} \langle \nabla f_i(x_k), \hat{x} - x_k \rangle + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j) \quad (3.7)$$

Tome $N \in \mathbb{N}$. Some [3.7](#) de $k = 0, \dots, N$, deriva-se

$$\begin{aligned} & -2 \sum_{k=0}^N t_k \max_{1 \leq i \leq m} \langle \nabla f_i(x_k), \hat{x} - x_k \rangle \\ & \leq \sum_{k=0}^N (\|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2) + \frac{2}{1 - \eta_{max}} \sum_{k=0}^N \left(\sum_{j=1}^m (C_k^j - C_{k+1}^j) \right) \\ & = \|x_0 - \hat{x}\|^2 - \|x_{N+1} - \hat{x}\|^2 + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m \left(\sum_{k=0}^N (C_k^j - C_{k+1}^j) \right) \\ & \leq \|x_0 - \hat{x}\|^2 + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_0^j - C_{N+1}^j) \end{aligned} \quad (3.8)$$

Como F é limitado inferiormente, podemos deduzir de $F(x_k) \leq C_k$ para todo k no Lema [2.4](#) que C_k é também limitado inferiormente. Então segue de [3.8](#) que:

$$-2 \sum_{k=0}^N t_k \max_{1 \leq i \leq m} \langle \nabla f_i(x_k), \hat{x} - x_k \rangle < +\infty \quad (3.9)$$

Além disso, sabemos do Lema [2.6](#) que $t_k \geq t_{min} > 0$ para todo k . Isso, junto com [3.9](#) implica que

$$\lim_{k \rightarrow \infty} \left(- \max_{1 \leq i \leq m} \langle \nabla f_i(x_k), \hat{x} - x_k \rangle \right) = 0 \quad (3.10)$$

Entretanto, para qualquer subsequência $\{x_{j_k}\}$ de $\{x_k\}$ tal que $\lim_{k \rightarrow \infty} x_{j_k} = x^*$, pode-se obter por [3.6](#) que:

$$\lim_{k \rightarrow \infty} \left(- \max_{1 \leq i \leq m} \langle \nabla f_i(x_{j_k}), \hat{x} - x_{j_k} \rangle \right) = - \max_{1 \leq i \leq m} \langle \nabla f_i(x^*), \hat{x} - x^* \rangle > 0$$

o que é uma contradição com [3.10](#). Logo x^* é um ponto de Pareto Estacionário. \square

3.2 O caso convexo

Nesta subseção, consideramos o algoritmo [2.2.1](#) supondo a convexidade de F para obter resultados de convergência mais fortes.

Sabe-se que a condição $\{x \in C : F(x) \preceq F(x_k), \forall k \in \mathbb{N}\} \neq \emptyset$ tem sido muito usado na análise de convergência de muitos algoritmos para otimização de vetores, veja, por exemplo o método do gradiente projetado [\[14, 40, 10\]](#), método de descida mais íngreme [\[2\]](#), método do ponto proximal [\[8, 30, 3\]](#).

Esta suposição está relacionada com a completude da imagem de F , ou seja, todas as seqüências não crescentes com relação a imagem de F tem um limite inferior. E a completude é uma suposição padrão para garantir a existência de pontos eficientes para problemas de otimização vetorial [\[25\]](#).

Enquanto para algoritmo não monótono, considerando C_k ao invés de $F(x_k)$. Fazio e Schuverdt [\[15\]](#) utilizaram a seguinte suposição para provar a convergência do algoritmo ali proposto:

(A₁) O conjunto $T := \{x \in C : F(x) \preceq C_k, \forall k \in \mathbb{N}\}$ é não vazio.

Recentemente, também usou-se essa suposição para mostrar a convergência de outro método de gradiente projetado não monótono [\[41\]](#). Em seguida, aplicando a hipótese (A₁) novamente, estudamos a convergência do Algoritmo [2.2.1](#) considerado aqui.

Teorema 3.3. *Assuma que (A₁) vale e $\eta_{max} < 1$. Então, a seqüência $\{x_k\}$ gerada pelo algoritmo [2.2.1](#) converge para um ponto ótimo de Pareto fraco.*

Demonstração. Primeiro mostramos que $\{x_k\}$ é quase-Fejér convergente para T . Tome qualquer $x \in T$. Pela desigualdade do gradiente de cada f_j temos que, para cada k ,

$$\langle \nabla f_j(x_k), x - x_k \rangle \leq f_j(x) - f_j(x_k) \leq C_k^j - f_j(x_k).$$

A partir disso e da proposição [3.1](#), pode-se obter que, para todo k ,

$$\begin{aligned}
& \|x_{k+1} - x\|^2 \\
& \leq \|x_k - x\|^2 + 2t_k \sum_{j=1}^m \lambda_j^k \langle \nabla f_j(x_k), x - x_k \rangle + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j) \\
& \leq \|x_k - x\|^2 + 2t_k \sum_{j=1}^m \lambda_j^k (C_k^j - f_j(x_k)) + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j) \\
& \leq \|x_k - x\|^2 + 2 \sum_{j=1}^m (C_k^j - f_j(x_k)) + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j)
\end{aligned} \tag{3.11}$$

onde a última desigualdade é estabelecida, pela conclusão $F(x_k) \preceq C_k$ no Lema [2.4](#), o tamanho do passo $t_k \leq 1$ e $\lambda_j^k \leq 1$. Para cada k , seja

$$\varepsilon_k := 2 \sum_{j=1}^m (C_k^j - f_j(x_k)) + \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j)$$

Note que $C_{k+1} \preceq C_k$ (ver a observação [2.5](#)), então temos $\varepsilon_k \geq 0$, e $\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 + \varepsilon_k$ por [3.11](#) para todo k .

Vamos provar que $\sum_{k=0}^{\infty} \varepsilon_k < +\infty$. Por [3.8](#) na prova do teorema [3.1](#) e observando que $x \in T$, temos que para qualquer $N \in \mathbb{N}$,

$$\sum_{k=0}^N \left(\frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j) \right) = \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_0^j - C_{N+1}^j) \leq \frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_0^j - f_j(x))$$

e também

$$\sum_{k=0}^{\infty} \left(\frac{2}{1 - \eta_{max}} \sum_{j=1}^m (C_k^j - C_{k+1}^j) \right) < +\infty \tag{3.12}$$

deixando N ir para $+\infty$.

Além disso, por [2.10](#) e [3.4](#), tem-se que para cada $j \in \{1, \dots, m\}$ e $k \in \mathbb{N}$,

$$\begin{aligned}
C_k^j - f_j(x_k) &= C_{k-1}^j - f_j(x_k) + \frac{f_j(x_k) - C_{k-1}^j}{Q_k} = C_{k-1}^j - f_j(x_k) \frac{Q_{k-1}}{Q_k} \\
&= (C_{k-1}^j - C_k^j) (Q_{k-1}) \leq \frac{\eta_{max}}{1 - \eta_{max}} (C_{k-1}^j - C_k^j)
\end{aligned} \tag{3.13}$$

Então, definindo $C_{-1} = C_0$, segue de [3.13](#) que para algum $N \in \mathbb{N}$,

$$\sum_{k=0}^N \left(2 \sum_{j=1}^m (C_k^j - f_j(x_k)) \right) \leq \sum_{k=0}^N \left(\frac{2\eta_{max}}{1 - \eta_{max}} \sum_{j=1}^m (C_{k-1}^j - C_k^j) \right)$$

$$= \frac{2\eta_{max}}{1 - \eta_{max}} \sum_{j=1}^m (C_0^j - C_N^j) \leq \frac{2\eta_{max}}{1 - \eta_{max}} \sum_{j=1}^m (C_0^j - f_j(x))$$

Por isso,

$$\sum_{k=0}^{\infty} \left(2 \sum_{j=1}^m (C_k^j - f_j(x_k)) \right) < +\infty$$

Combinando isso, com [3.12](#) resulta que $\sum_{k=0}^{\infty} \varepsilon_k < +\infty$ e conseqüentemente $\{x_k\}$ é quase-Fejér convergente para T .

Então pela proposição [1.16](#), temos que $\{x_k\}$ é limitada e portanto tem um ponto de acumulação denotado por x^* , que é um ponto ótimo de Pareto fraco graças ao teorema [3.2](#) (observe que a prova do teorema [3.2](#) ainda vale quando a suposição “ F é limitada inferiormente” é substituída por (A_1) e a convexidade de F).

Na prova restante, mostramos que $x^* \in T$. Por isso e pela proposição [1.16](#) novamente, a convergência de $\{x_k\}$ para x^* é então obtida. Seja $\{x_{j_k}\}$ uma subsequência de $\{x_k\}$ tal que $\lim_{k \rightarrow \infty} x_{j_k} = x^*$. Tome qualquer $k_0 \in \mathbb{N}$. Pelo Lema [2.4](#) e a propriedade não crescente de $\{C_k\}$ tem-se que para todo $k \geq k_0$, $F(x_{j_k}) \leq C_{j_k} \leq C_k \leq C_{k_0}$. Como F é contínua, obtém-se então que $F(x^*) = \lim_{k \rightarrow \infty} F(x_{j_k}) \leq C_{k_0}$. Assim, $x^* \in T$ como k_0 é arbitrário. \square

Ao final desta secção, damos um resultado de convergência linear do algoritmo [2.2.1](#). Recentemente, provou-se a convergência linear do algoritmo não monótono estudado em [\[41\]](#) usando a seguinte condição sobre a direção de busca e os gradientes de f_i :

(A₂) Existe uma constante positiva c tal que para cada $i \in \{1, \dots, m\}$ e $k \in \mathbb{N}$,

$$\langle \nabla f_i(x_k), v_k \rangle \leq -c \|\nabla f_i(x_k)\|$$

Neste trabalho, empregamos técnicas semelhantes às utilizadas em [\[41\]](#), estabelecemos a convergência linear do algoritmo [2.2.1](#).

Em primeiro lugar, podemos obter o seguinte lema útil sob a hipótese (A₂).

Lema 3.4. *Suponha que (A₂) seja válido. Então existe $\alpha > 0$ tal que:*

$$f_i(x_{k+1}) \leq C_k^i - \alpha \|\nabla f_i(x_k)\|$$

para cada $i \in \{1, \dots, m\}$ e $k \in \mathbb{N}$.

Demonstração. Tome $i \in \{1, \dots, m\}$ e $k \in \mathbb{N}$. Pela definição da direção de busca v_k e observando que $0 \in C - x_k$, podemos obter $\varphi_k(v_k) \leq \varphi_k(0) = 0$ que $\frac{\|v_k\|^2}{2} \leq -\langle \nabla f_i(x_k), v_k \rangle$.

Combinando isso, com a regra de busca linear não monótona [2.7](#) produz

$$f_i(x_{k+1}) \leq C_k^i + t_k(1 - \gamma)\langle \nabla f_i(x_k), v_k \rangle. \quad (3.14)$$

Então, aplicando o limite inferior de t_k no Lema [2.6](#) e a hipótese (A_2) a [3.14](#) pode-se obter

$$f_i(x_{k+1}) \leq C_k^i - ct_{min}(1 - \gamma) \| f_i(x_k) \|,$$

onde t_{min} é definido como no Lema [2.6](#). Consequentemente, a conclusão é válida com $\alpha := ct_{min}(1 - \gamma)$. \square

Abaixo, enunciamos o resultado da convergência linear do algoritmo [2.2.1](#). Aplicando o lema [2.4](#), lema [3.4](#) e teorema [3.3](#), o seguinte teorema pode ser obtido usando técnicas semelhantes como o usado em [[41](#)], Teorema 2] e, portanto omitimos a prova aqui para simplificar.

Teorema 3.5. *Suponha que (A_1) , (A_2) sejam válidos e $\eta_{max} < 1$. Seja $\{x_k\}$ a sequência gerada pelo algoritmo [2.2.1](#). Então, existe $\theta \in (0, 1)$ tal que*

$$F(x_k) - F(x^*) \preceq \theta^k (F(x_0) - F(x^*)) \text{ para todo } k,$$

onde $x^* := \lim_{k \rightarrow \infty} x_k$ é um ponto ótimo de Pareto fraco. Especificamente, a taxa de convergência θ é dada por

$$\theta = 1 - \frac{\alpha(1 - \eta_{max})}{\alpha + (\rho + \|x^*\|)(1 + 2L_{max})}$$

com α como no lema [3.4](#), a constante ρ satisfazendo $\{x_k\} \subseteq \rho B$ e L_{max} é a máxima constante de Lipschitz L_i de ∇f_i , i.e, $L_{max} := \max\{L_1, \dots, L_m\}$.

Capítulo 4

Experimentos Numéricos

Abaixo mostraremos alguns experimentos numéricos em \mathbb{R} e \mathbb{R}^2 devido a sua rápida resposta e visualização em relação a espaços maiores, no entanto, deixamos claro que esse algoritmo funciona em \mathbb{R}^n . Usamos para a implementação a linguagem de programação Python.

Primeiramente, testamos o algoritmo [2.2.1](#) em funções descontínuas e conseguimos demonstrar a imagem da função, bem como pontos ótimo de Pareto e ótimo de Pareto fraco. O exemplo [4.1](#) e [4.2](#) apresentam tal resultado.

Exemplo 4.1. A figura abaixo corresponde a imagem de f , sendo que $f_1(x) = (x_1 - 2)^2 - (x_2 + 3)^2$ e $f_2(x) = \frac{\text{sen}x_1 + \text{sen}x_2}{100(1 + x_1^2 + x_2^2)}$.

Exemplo 4.2. A figura abaixo corresponde a imagem de f , sendo que $f_1(x) = \frac{1}{4}(x_1 - 1)^4 + 2(x_2 - 2)^4$ e $f_2(x) = (x_2 - x_1)^2 + (1 - x_1)^2$.

Exemplo 4.3. A figura abaixo corresponde a imagem de f , sendo que $f_1(x) = \frac{1}{x_1^2 + x_2^2 + 1}$ e $f_2(x) = x_1^2 + 3x_2^2 + 1$.

Exemplo 4.4. O gráfico abaixo corresponde a imagem de f com destaque para a imagem do ponto crítico, sendo que $f(x) = x^3 - 3x^2$, calculada para valores de x em $[-1, 3]$.

Exemplo 4.5. O gráfico abaixo corresponde a imagem de f com destaque para a imagem do ponto crítico, sendo que $f(x) = \text{sen}(x) + \cos(x)$, calculada para valores de x em $[-6, 6]$.

Exemplo 4.6. O gráfico abaixo corresponde a imagem de f com destaque para a imagem dos pontos críticos, sendo que $f_1(x) = -x + 2\text{sen}x$ e $f_2(x) = x + \text{sen}x$.

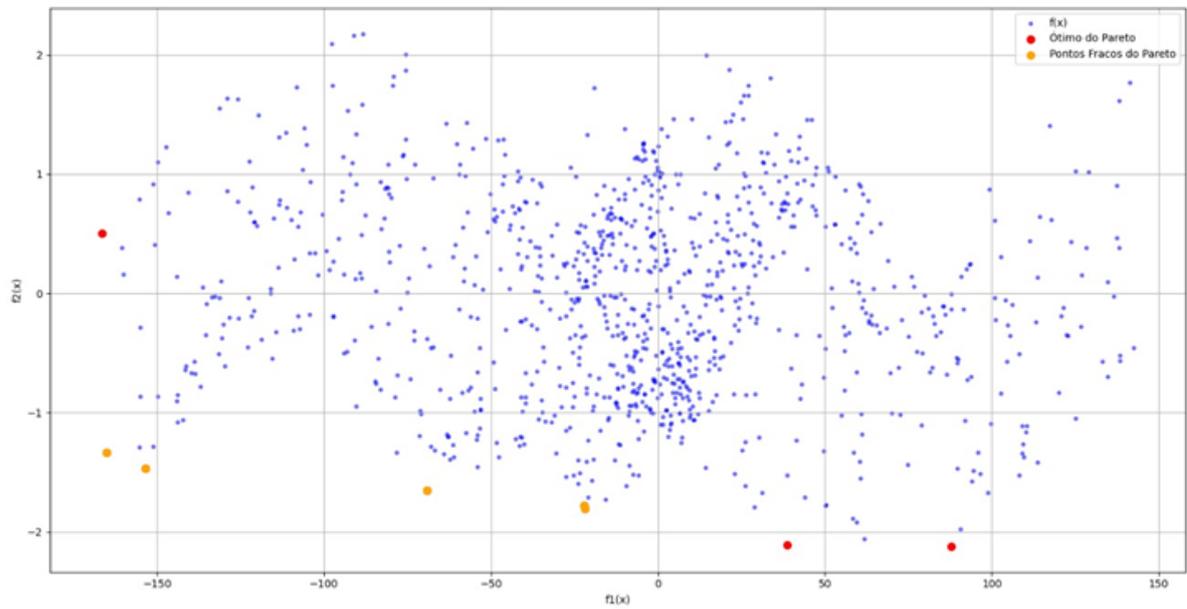


Figura 4.1: Imagem de $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ do Exemplo 2.11

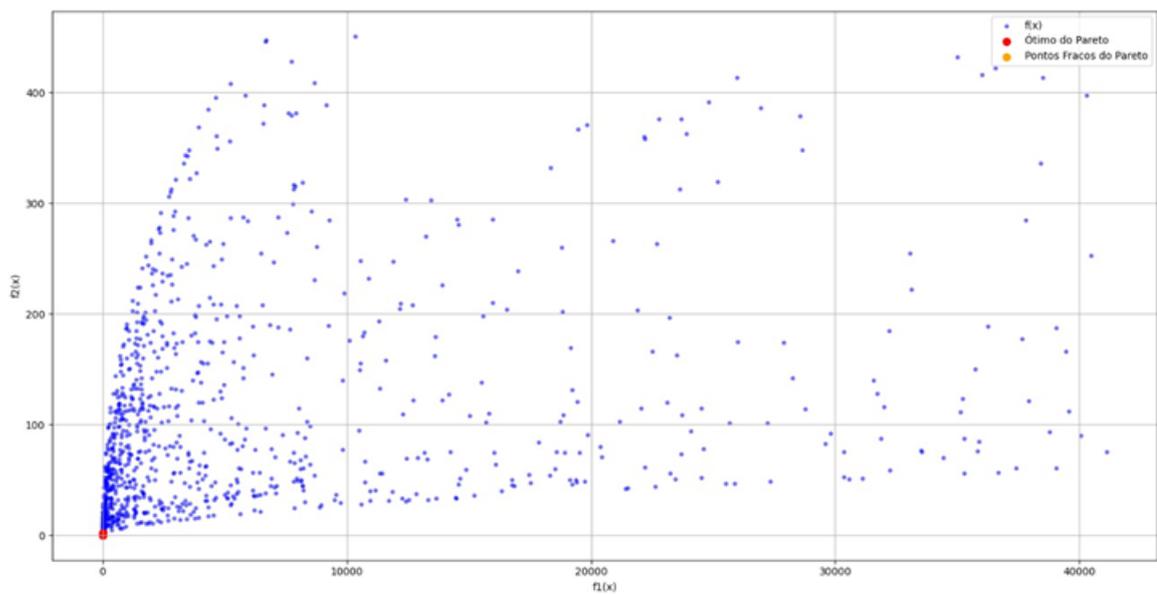


Figura 4.2: Imagem de $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ do Exemplo 2.12

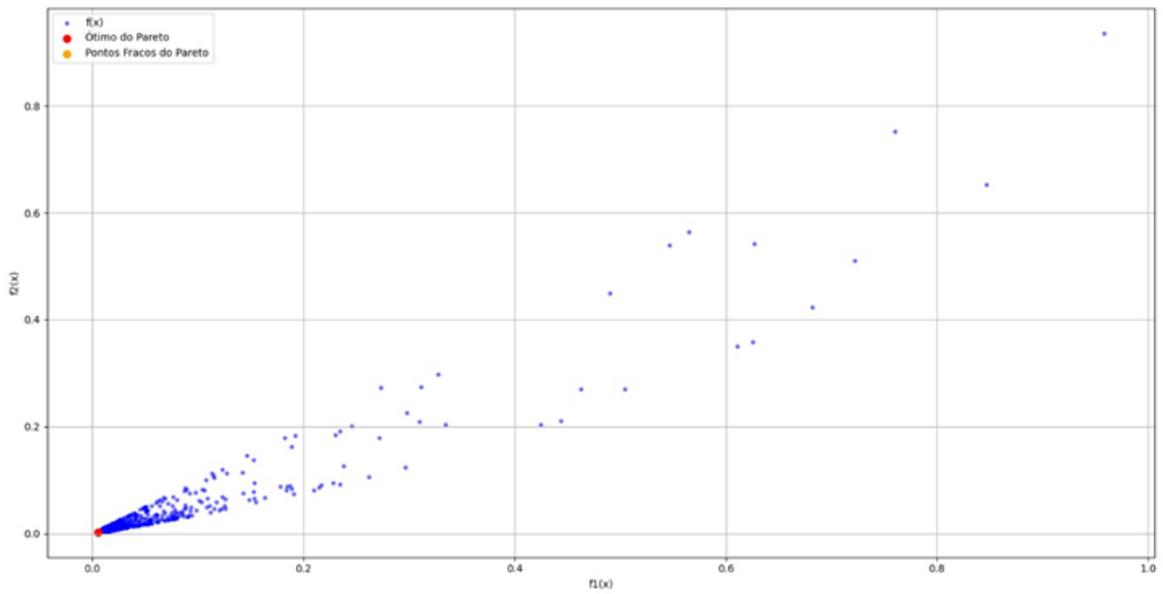


Figura 4.3: Imagem de $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ do Exemplo 2.13

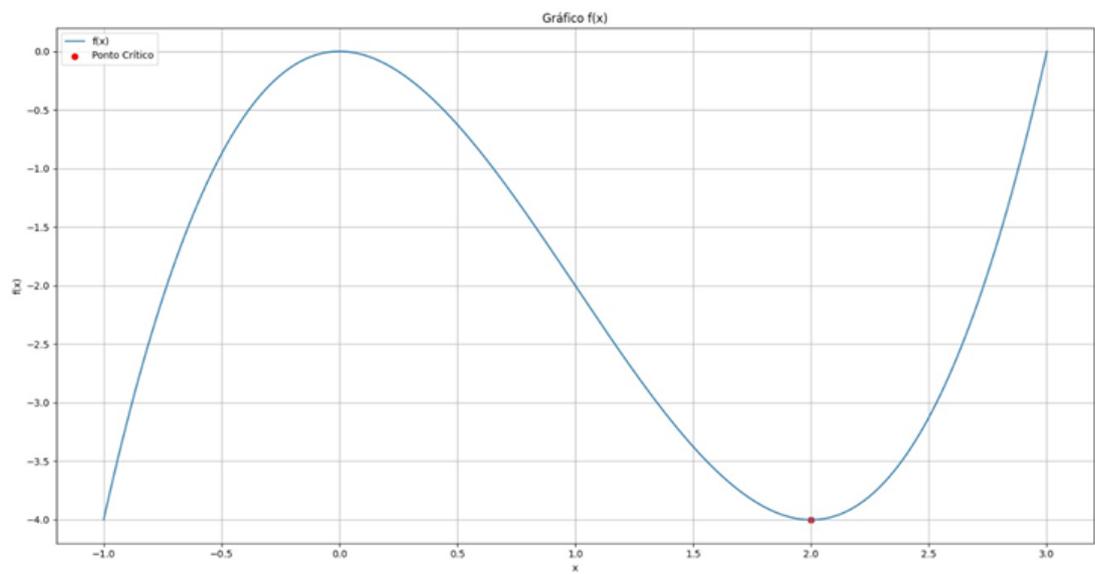


Figura 4.4: Imagem de $f : \mathbb{R} \rightarrow \mathbb{R}$ do Exemplo 2.14

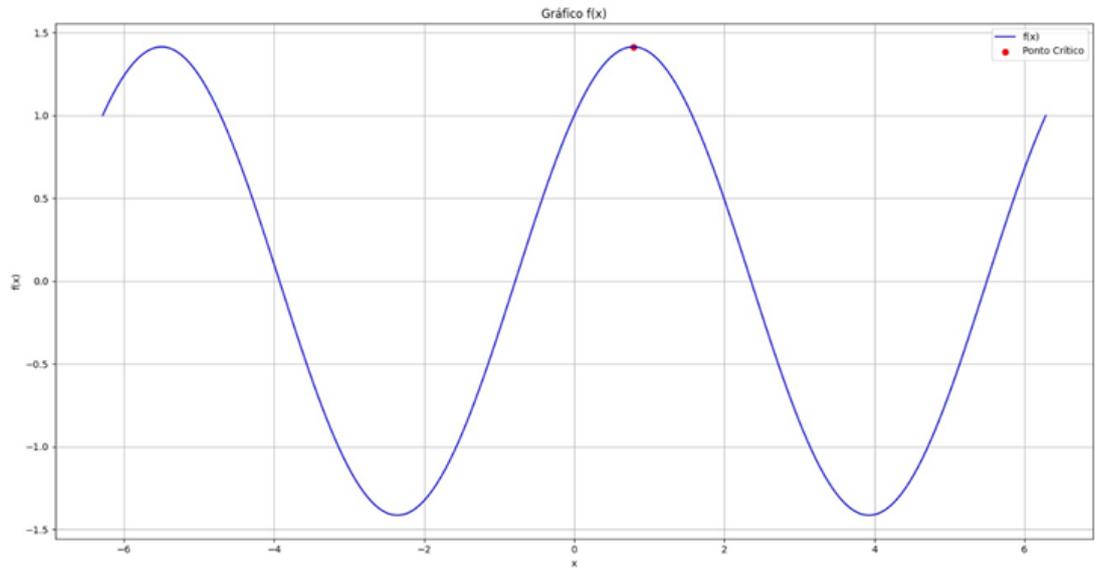


Figura 4.5: Imagem de $f : \mathbb{R} \rightarrow \mathbb{R}$ do Exemplo 2.15

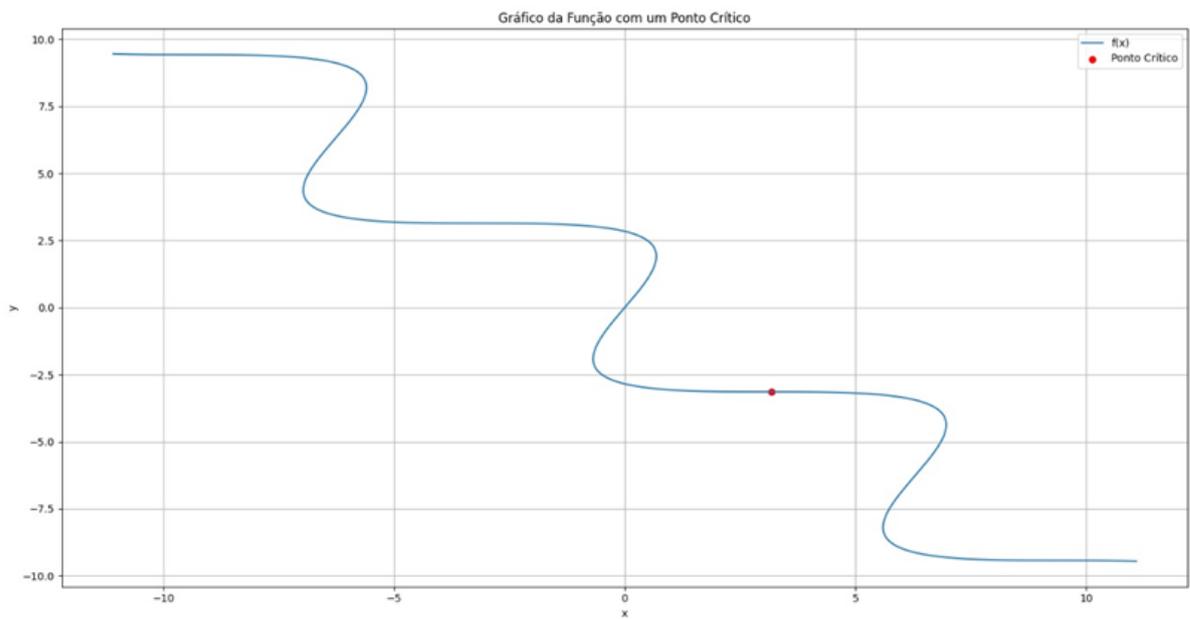


Figura 4.6: Imagem de $f : \mathbb{R} \rightarrow \mathbb{R}^2$ do Exemplo 2.16

Capítulo 5

Considerações Finais

A pesquisa em Otimização Multiobjetivo tem ganhado bastante enfoque nos últimos anos, conseqüentemente vários métodos vem surgindo, após ampla pesquisa bibliográfica, foi verificado numericamente que as estratégias não monótonas, muitas vezes, funcionam melhor do que as monótonas.

Apresentamos então um método de gradiente equipado com a técnica de busca linear não monótona de Zhang e Hager (no caso multiobjetivo) para otimização multiobjetivo suave.

Estabelecemos a estacionariedade de Pareto do ponto de acumulação da seqüência gerada pelo método proposto para problemas de otimização multiobjetivo não convexo.

Quando a função multiobjetivo era convexa, a convergência da seqüência gerada para um ponto ótimo de Pareto fraco e ainda mais a convergência linear dos valores da função para o valor ótimo foram comprovados.

Para verificar a boa funcionalidade do algoritmo aqui proposto fizemos a implementação na linguagem de programação Python e obtivemos os resultados numéricos aqui apresentados.

Visando trabalhos futuros podemos incluir vários aspectos, por exemplo:

- 1) É interessante investigar a linearidade da taxa de convergência da seqüência gerada pelo algoritmo não monótono (em vez da seqüência de valor de função);*
- 2) Investigar os métodos de gradiente acelerado que propõem novas técnicas não monótonas para encontrar boas frentes de Pareto para problemas de otimização multiobjetivo.*

Referências Bibliográficas

- [1] ARROYO, J. E. C., ET AL. *Heurísticas e metaheurísticas para otimização combinatória multiobjetivo*. Doutorado em Engenharia Elétrica (Tese de doutorado)-Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, Campinas (2002).
- [2] BENTO, G. C., NETO, J. C., OLIVEIRA, P. R., AND SOUBEYRAN, A. *The self regulation problem as an inexact steepest descent method for multicriteria optimization*. European Journal of Operational Research 235, 3 (2014), 494–502.
- [3] BENTO, G. D. C., CRUZ NETO, J. X., LOPEZ, G., SOUBEYRAN, A., AND SOUZA, J. *The proximal point method for locally lipschitz functions in multiobjective optimization with application to the compromise problem*. SIAM Journal on Optimization 28, 2 (2018), 1104–1120.
- [4] BRITO, A. S., NETO, J. C., SANTOS, P., AND SOUZA, S. *A relaxed projection method for solving multiobjective optimization problems*. European Journal of Operational Research 256, 1 (2017), 17–23.
- [5] BURACHIK, R., GRAÑA DRUMMOND, L., IUSEM, A. N., AND SVAITER, B. F. *Full convergence of the steepest descent method with inexact line searches*. Optimization 32, 2 (1995), 137–146.
- [6] BURACHIK, R. S., KAYA, C. Y., AND RIZVI, M. *A new scalarization technique and new algorithms to generate pareto fronts*. SIAM Journal on Optimization 27, 2 (2017), 1010–1034.
- [7] CARRIZO, G. A., LOTITO, P. A., AND MACIEL, M. C. *Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem*. Mathematical Programming 159 (2016), 339–369.

- [8] CENG, L.-C., AND YAO, J.-C. *Approximate proximal methods in vector optimization*. European Journal of Operational Research 183, 1 (2007), 1–19.
- [9] COELLO, C. A. *An updated survey of ga-based multiobjective optimization techniques*. ACM Computing Surveys (CSUR) 32, 2 (2000), 109–143.
- [10] CRUZ, J. B., PÉREZ, L. L., AND MELO, J. *Convergence of the projected gradient method for quasiconvex multiobjective optimization*. Nonlinear Analysis: Theory, Methods & Applications 74, 16 (2011), 5268–5273.
- [11] DAI, Y.-H. *On the nonmonotone line search*. Journal of Optimization Theory and Applications 112 (2002), 315–330.
- [12] DE CASTRO, R. E. *Otimização de estruturas com multi-objetivos via algoritmos genéticos*. Rio de Janeiro-RJ (2001).
- [13] DE LIMA, R. F. *Topologia e análise no espaço r n*. Coleção Textos Universitário (2015).
- [14] DRUMMOND, L. G., AND IUSEM, A. N. *A projected gradient method for vector optimization problems*. Computational Optimization and applications 28 (2004), 5–29.
- [15] FAZZIO, N. S., AND SCHUVERDT, M. L. *Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems*. Optimization Letters 13 (2019), 1365–1379.
- [16] FLIEGE, J., DRUMMOND, L. G., AND SVAITER, B. F. *Newton’s method for multiobjective optimization*. SIAM Journal on Optimization 20, 2 (2009), 602–626.
- [17] FLIEGE, J., AND SVAITER, B. F. *Steepest descent methods for multicriteria optimization*. Mathematical methods of operations research 51 (2000), 479–494.
- [18] FLIEGE, J., VAZ, A. I. F., AND VICENTE, L. N. *Complexity of gradient descent for multiobjective optimization*. Optimization Methods and Software 34, 5 (2019), 949–959.
- [19] FONSECA, C. M., AND FLEMING, P. J. *An overview of evolutionary algorithms in multiobjective optimization*. Evolutionary computation 3, 1 (1995), 1–16.

- [20] GRANA DRUMMOND, L., AND SVAITER, B. *Métodos de descida em otimização multiobjetivo*. Publicações Matemáticas, IMPA 30 (2015).
- [21] GRIPPO, L., LAMPARIELLO, F., AND LUCIDI, S. *A nonmonotone line search technique for newton's method*. SIAM journal on Numerical Analysis 23, 4 (1986), 707–716.
- [22] IUSEM, A. N., SVAITER, B. F., AND TEBoulLE, M. *Entropy-like proximal methods in convex programming*. Mathematics of Operations Research 19, 4 (1994), 790–814.
- [23] IZMAILOV, A., S. M. *Otimização*, volume 1. IMPA, 2009.
- [24] LIMA, E. *Curso de Análise volume 2(décima primeira edição)*. Springer, 2015.
- [25] LUC, D. T. *Theory of vector optimization*. Springer, 1989.
- [26] LUCAMBIO PÉREZ, L., AND PRUDENTE, L. *Nonlinear conjugate gradient methods for vector optimization*. SIAM Journal on Optimization 28, 3 (2018), 2690–2720.
- [27] MAHDAVI-AMIRI, N., AND SALEHI SADAGHIANI, F. *A superlinearly convergent nonmonotone quasi-newton method for unconstrained multiobjective optimization*. Optimization Methods and Software 35, 6 (2020), 1223–1247.
- [28] MITA, K., FUKUDA, E. H., AND YAMASHITA, N. *Nonmonotone line searches for unconstrained multiobjective optimization problems*. Journal of Global Optimization 75 (2019), 63–90.
- [29] NOCEDAL, J., AND WRIGHT, S. J. *Numerical optimization*. Springer, 1999.
- [30] PAPA QUIROZ, E., AND CRUZADO ACUÑA, S. *An inexact scalarization proximal point method for multiobjective quasiconvex minimization in euclidean spaces*.
- [31] PAPA QUIROZ, E. A., APOLINÁRIO, H. C. F., VILLACORTA, K. D., AND OLIVEIRA, P. R. *A linear scalarization proximal point method for quasiconvex multiobjective minimization*. Journal of Optimization Theory and Applications 183 (2019), 1028–1052.
- [32] PARETO, V. *Cours d'économie politique, vol. 1*. Librairie Droz, 1964.

- [33] QU, S., JI, Y., JIANG, J., AND ZHANG, Q. *Nonmonotone gradient methods for vector optimization with a portfolio optimization application*. European Journal of Operational Research 263, 2 (2017), 356–366.
- [34] THOMANN, J., AND EICHFELDER, G. *A trust-region algorithm for heterogeneous multiobjective optimization*. SIAM Journal on Optimization 29, 2 (2019), 1017–1047.
- [35] TICONA, W. G. C. *Algoritmos evolutivos multi-objetivo para a reconstrução de árvores filogenéticas*. *PhD thesis, Universidade de São Paulo, 2003*.
- [36] TOINT, P. L. *An assessment of nonmonotone linesearch techniques for unconstrained optimization*. SIAM Journal on Scientific Computing 17, 3 (1996), 725–739.
- [37] URRUTY, J.-B. H., AND LEMARÉCHAL, C. *Convex analysis and minimization algorithms*. Springer-Verlag, 1993.
- [38] WANG, J., HU, Y., WAI YU, C. K., LI, C., AND YANG, X. *Extended newton methods for multiobjective optimization: majorizing function technique and convergence analysis*. SIAM Journal on Optimization 29, 3 (2019), 2388–2421.
- [39] ZHANG, H., AND HAGER, W. W. *A nonmonotone line search technique and its application to unconstrained optimization*. SIAM journal on Optimization 14, 4 (2004), 1043–1056.
- [40] ZHAO, X., KÖBIS, M. A., YAO, Y., AND YAO, J.-C. *A projected subgradient method for nondifferentiable quasiconvex multiobjective optimization problems*. Journal of Optimization Theory and Applications 190, 1 (2021), 82–107.
- [41] ZHAO, X., AND YAO, J.-C. *Linear convergence of a nonmonotone projected gradient method for multiobjective optimization*. Journal of Global Optimization (2022), 1–18.