Métodos de Descida em Otimização Multiobjetivo

Publicações Matemáticas

Métodos de Descida em Otimização Multiobjetivo

L. M. Graña Drummond
FACC – UFRJ

B. F. Svaiter
IMPA

impa

30º Colóquio Brasileiro de Matemática

Impresso no Brasil / Printed in Brazil
Capa: Noni Geiger / Sérgio R. Vaz

30º Colóquio Brasileiro de Matemática

- Aplicações Matemáticas em Engenharia de Produção - Leonardo J. Lustosa e Fernanda M. P. Raupp
- Boltzmann-type Equations and their Applications - Ricardo Alonso
- Dissipative Forces in Celestial Mechanics - Sylvio Ferraz-Mello, Clodoaldo Grotta-Ragazzo e Lucas Ruiz dos Santos
- Economic Models and Mean-Field Games Theory - Diogo A. Gomes, Levon Nurbekyan and Edgard A. Pimentel
- Generic Linear Recurrent Sequences and Related Topics - Letterio Gatto
- Geração de Malhas por Refinamento de Delaunay - Marcelo Siqueira, Afonso Paiva e Paulo Pagliosa
- Global and Local Aspects of Levi-flat Hypersurfaces - Arturo Fernández Pérez e Jiří Lebl
- Introdução às Curvas Elípticas e Aplicações - Parham Salehyan
- **Métodos de Descida em Otimização Multiobjetivo - L. M. Graña Drummond e B. F. Svaiter**
- Modern Theory of Nonlinear Elliptic PDE - Boyan Slavchev Sirakov
- Novel Regularization Methods for Ill-posed Problems in Hilbert and Banach Spaces - Ismael R. Bleyer e Antonio Leitão
- Probabilistic and Statistical Tools for Modeling Time Series - Paul Doukhan
- Tópicos da Teoria dos Jogos em Computação - R. C. S. Schouery, O. Lee, F. K. Miyazawa e E. C. Xavier
- Topics in Spectral Theory - Carlos Tomei

ISBN: 978-85-244-0402-3

Distribuição: IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
E-mail: ddic@impa.br
http://www.impa.br

Conteúdo

Introdução

Em otimização multiobjetivo dois ou mais objetivos devem ser minimizados simultaneamente. A minimização multiobjetivo modela diversos problemas em engenharia [6], estatísticas [3], análise ambiental [24, 7], etc. Um exemplo típico é o problema de projetar uma peça de alguma máquina: procura-se minimizar seu custo de produção, maximizar sua resistência e, em alguns casos, também pode ser importante minimizar seu peso. Observe que neste exemplo, alguns desses objetivos podem ser conflitantes.

Em otimização multiobjetivo, quase nunca existe um minimizador comum a todas as funções a serem minimizadas. Por este motivo, a noção clássica de otimalidade não é conveniente neste contexto. Uma definição adequada de otimalidade no caso multiobjetivo foi proposta pelo economista e sociólogo italiano Vilfredo Pareto, que viveu nos séculos XIX e XX. Sua noção de otimalidade é, ainda hoje, a prevalente na otimização multiobjetivo [29].

Considere o problema de minimizar várias funções escalares dependentes de um mesmo argumento. Um argumento particular é chamado de ótimo de Pareto quando não há outro argumento que diminua o valor de uma das funções sem aumentar o valor das restantes.

Uma das principais estratégias para a resolução de problemas de otimização multiobjetivo é a técnica de escalarização [14, 15, 30], que consiste em resolver um ou mais problemas de minimização escalar, de modo que os ótimos obtidos sejam pontos de Pareto para o problema original. Uma natural vantagem desta abordagem é o fato de que dispomos de vários métodos eficientes para resolver problemas escalares. Entre as diversas técnicas de escalarização, destaca-se o assim chamado método dos pesos [17, 20, 22, 27], em que se minimiza uma combinação linear (não-negativa) dos objetivos, sendo que os "pesos" adequados nem sempre se conhecem a priori, de forma que os problemas escalares gerados pelo método podem ser ilimitados e, portanto, sequer possuir ótimos. Recentemente, desenvolveram-se técnicas de escalarização "adaptativas", em que, de alguma maneira, os parâmetros são escolhidos obtendo-se uma boa qualidade de aproximação ao longo do procedimento [5, 8]. Também recentemente foram criados outros algoritmos que não escalarizam a função objetivo do problema vetorial original (veja, e.g., [1, 2] para uma visão geral do assunto). Alguns desses procedimentos baseiam-se em estratégias heurísticas [23, 28], sem provas de convergência conhecidas, sendo que resultados empíricos mostram que, como no caso escalar, a aproximação é bastante lenta [31]. Existem também outras técnicas de otimização multiobjetivo que não requerem o uso de parâmetros e utilizam um certo ordenamento pré especificado dos componentes da função objetivo [26].

Os métodos acima mencionados, quando bem sucedidos, computam uma solução ótima. No entanto, devido à natureza do problema, em geral existem outros ótimos com valores funcionais não comparáveis. Por este motivo, é desejável que se obtenham todos os valores ótimos, para que o proponente do problema possa escolher um deles juntamente com um ponto de Pareto com tal valor funcional. Em suma, resolver um problema de otimização multiobjetivo consiste em achar todos seus ótimos de Pareto, um conjunto cuja imagem é chamada *parede ótima*. Achar a parede ótima é um problema difícil. Geralmente conseguem-se aproximações dela. De posse de uma dessa aproximações, um de seus pontos é escolhido pelo proponente do problema com algum critério.

Aproximar a parede ótima com grande precisão é computacionalmente custoso e geralmente resulta no cálculo de muitos pontos que, em sua maioria, não serão utilizados. Uma estratégia para contornar este inconveniente pode ser calcular a parede ótima sem extrema precisão e, feita a escolha de um desses pontos, refinar seu valor funcional, preservando as propriedades que levaram à sua escolha, isto é, procurar uma aproximação com valores funcionais menores. Esta é uma das principais motivações para o desenvolvimento de métodos de descida para otimização multiobjetivo.

O primeiro método de descida multiobjetivo foi proposto em [10] no ano 2000, e consistia em uma extensão do método de Cauchy. Posteriormente, versões multiobjetivo dos métodos de Newton e dos gradientes projetados foram desenvolvidas em [9] e [13], respectivamente. Este métodos foram também estendidos à otimização vetorial em [19, 16, 11, 12, 18].

O crescente interesse de pesquisadores e estudantes brasileiros por métodos de descida em otimização multiobjetivo e a falta de bibliografia sobre o assunto em língua portuguesa nos motivaram a escrever estas notas. Acreditamos que as ideias essenciais destes métodos estão presentes nas duas extensões que aqui apresentamos: os métodos de Cauchy e de Newton para otimização multiobjetivo. Aos leitores entusiastas recomendamos prosseguir os estudos no assunto com os livros [29, 25, 21] e os artigos [10, 19, 16, 9, 11, 12, 13, 18].

Capítulo 1

Otimização multiobjetivo

1.1 Definições e notação

O problema de otimização multiobjetivo consiste em minimizar *simultaneamente* as componentes de uma função

$$f : \mathbb{R}^n \to \mathbb{R}^m, \;\; f(x) = \big(f_1(x), \ldots, f_m(x)\big)$$

em um conjunto viável $C \subset \mathbb{R}^n$. Como discutido na introdução, raramente há um minimizador comum a todas as funções componentes e, por isto, a noção clássica de otimalidade não é adequada a este tipo de problema. A noção apropriada neste contexto é a de Pareto. Um ponto $\bar{x} \in C$ é um *ótimo de Pareto* para o problema acima definido se $f(\bar{x})$ é minimal em $f(C) = \{f(x) \mid x \in C\}$ na ordem parcial:

$$u, v \in \mathbb{R}^m, \;\; u \leq v \iff u_j \leq v_j, \;\; j = 1, \ldots, m. \tag{1.1}$$

Portanto, $\bar{x} \in C$ é ótimo de Pareto se

$$\forall x \in C, \;\; f(x) \leq f(\bar{x}) \implies f(x) = f(\bar{x}),$$

o que equivale à condição

$$\nexists x \in C, \;\; f_j(x) \leq f_j(\bar{x}), \;\; j = 1, \ldots, m \text{ e } f_{j_0}(x) < f_{j_0}(\bar{x}) \text{ para algum } j_0.$$

o problema de encontrar um ótimo de Pareto para f em C será denotado por

$$\min f(x), \;\; x \in C.$$

A desigualdade estrita em $\mathbb{R}^m$ se define como:

$$u, v \in \mathbb{R}^m, \;\; u < v \iff u_j < v_j, \;\; j = 1, \ldots, m. \tag{1.2}$$

Um ponto $\bar{x} \in C$ é um *ótimo fraco de Pareto* se *não* existe x viável tal que

$$f(x) < f(\bar{x}).$$

Segue-se desta definição que todo ótimo de Pareto é um ótimo fraco de Pareto. Observe que para $m = 1$, a otimização multiobjetivo é a otimização escalar e os ótimos de Pareto coincidem com os ótimos fracos de Pareto e são os ótimos da minimização escalar.

Os conjuntos dos números reais não-negativos e dos reais (estritamente) positivos serão designados por $\mathbb{R}_+$ e $\mathbb{R}_{++}$, respectivamente. A ordem parcial (1.1) pode ser caracterizada como uma inclusão no cone (convexo, fechado e com interior não vazio) $\mathbb{R}^m_+$:

$$u \leq v \iff v - u \in \mathbb{R}^m_+.$$

A desigualdade estrita entre vetores definida em (1.2) pode ser caracterizada pela inclusão em $\mathbb{R}^m_{++}$ (o interior de $\mathbb{R}^m_+$):

$$u < v \iff v - u \in \mathbb{R}^m_{++}.$$

Neste trabalho, identificamos $\mathbb{R}^1$ com $\mathbb{R}$ e $\mathbb{R}^n$ com $\mathbb{R}^{n\times 1}$, para $n \in \mathbb{N}$. Portanto, o produto interno canônico entre $x, y \in \mathbb{R}^n$ vem dado por $x^\top y$, onde $x^\top$ é o transposto de x. Denotamos a norma euclideana de $\mathbb{R}^n$ por $\|\cdot\|$:

$$\|x\| = \sqrt{x^\top x}.$$

Para uma matriz $A \in \mathbb{R}^{m\times n}$, $\|A\|$ denota sua norma como operador em relação à norma euclideana:

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

1.2 Direções de descida e pontos críticos

Nestas notas $f : \mathbb{R}^n \to \mathbb{R}^m$ é uma função diferenciável e vamos considerar o problema de otimização multiobjetivo irrestrito

$$\min f(x), \quad x \in \mathbb{R}^n. \tag{1.3}$$

Definição 1.1. *Um vetor $v \in \mathbb{R}^n$ é uma* direção de descida *para f em $x \in \mathbb{R}^n$ se*

$$\nabla f_j(x)^\top v < 0, \quad j = 1, \ldots, m.$$

Um ponto $x \in \mathbb{R}^n$ é estacionário *ou* crítico *para f se* não *há direções de descida em x.*

As direções de descida podem ser caracterizadas pela função $\varphi_x \colon \mathbb{R}^n \to \mathbb{R}$,

$$\varphi_x(v) := \max_{j=1,\ldots,m} \nabla f_j(x)^\top v. \tag{1.4}$$

Trivialmente,

$$v \text{ é uma direção de descida} \iff \varphi_x(v) < 0.$$

Sendo o máximo de funções lineares, φ_x é convexa e positivamente homogênea (de grau 1):

$$\begin{aligned} &\varphi_x(\lambda v) = \lambda \varphi_x(v)\,, \\ &\varphi_x(u + v) \leq \varphi_x(u) + \varphi_x(v), \end{aligned} \qquad \forall \lambda \geq 0 \text{ e } u, v \in \mathbb{R}^n. \tag{1.5}$$

Como $u \mapsto \max_{j=1,\ldots,m} u_j$ é Lipschitz contínua com constante 1, também temos

$$|\varphi_x(u) - \varphi_y(v)| \quad \leq \quad \|Jf(x)u - Jf(y)v\| \tag{1.6}$$

para todo $u, v \in \mathbb{R}^n$.

Observe que um vetor $v \in \mathbb{R}^n$ é uma *direção de descida* para f em $x \in \mathbb{R}^n$ se ele é uma direção de descida para todas as componentes de f. A i-ésima linha de $Jf(x) \in \mathbb{R}^{m \times n}$, o Jacobiano de f em x, é $\nabla f_i(x)^\top$. Portanto, v é uma direção de descida para f em x se

$$Jf(x)\, v < 0,$$

onde a desigualdade acima é a relação definida em (1.2).

Segue-se da Definição 1.1 que x é crítico se

$$\max_{j=1,\ldots,m} \nabla f_j(x)^\top v \geq 0 \quad \text{para todo } v \in \mathbb{R}^n. \tag{1.7}$$

Esta condição tem a seguinte caracterização geométrica: a imagem do Jacobiano em x não corta o interior do ortante negativo,

$$Jf(x)(\mathbb{R}^n) \cap \Big(-\mathbb{R}^m_{++} \Big) = \emptyset,$$

onde $-\mathbb{R}^m_{++} = \{-w \mid w \in \mathbb{R}^m_{++}\}$. Para $m = 1$, temos $f : \mathbb{R}^n \to \mathbb{R}$ e a criticalidade de x coincide com a condição clássica $\nabla f(x) = 0$.

Numa direção de descida há decréscimo local da função f, como mostra o resultado seguinte.

Lema 1.2. *Sejam $f\colon \mathbb{R}^n \to \mathbb{R}^m$ diferenciável e $v \in \mathbb{R}^n$ uma direção de descida para f em x. Dado $\sigma \in (0,1)$, existe $\bar{t} > 0$ tal que*

$$f(x + tv) < f(x) + \sigma t Jf(x)v < f(x) \quad \textit{para todo } t \in (0, \bar{t}].$$

Demonstração. Por hipótese, $\nabla f_j(x)^\top v < 0$ para $j = 1, \ldots, m$. Como $\sigma \in (0,1)$, temos que

$$\lim_{t \to 0} \frac{f_j(x + tv) - f_j(x)}{t} = \nabla f_j(x)^\top v < \sigma \nabla f_j(x)^\top v, \quad j = 1, \ldots, m.$$

Portanto, para cada $j = 1, \ldots, m$ existe $\bar{t}_j > 0$ tal que

$$f_j(x + tv) - f_j(x) < \sigma t \nabla f_j(x)^\top v \quad \text{para todo } t \in (0, \bar{t}_j].$$

Para terminar a prova, tome $\bar{t} = \min\{\bar{t}_1, \ldots, \bar{t}_m\}$. □

O Lema acima nos permite provar que a estacionariedade é uma condição necessária para a otimalidade de Pareto e, em certas condições, também suficiente.

Corolário 1.3. *Seja $f : \mathbb{R}^n \to \mathbb{R}^m$ diferenciável*

1. *Se $\bar{x}$ é um ótimo de Pareto fraco então $\bar{x}$ é estacionário para f.*

2. *Se f é convexa componente a componente, então $\bar{x}$ é um ótimo fraco de Pareto se e somente se $\bar{x}$ é estacionário.*

3. *Se f é estritamente convexa componente a componente, então $\bar{x}$ é um ótimo de Pareto se e somente se $\bar{x}$ é estacionário.*

Demonstração. Para provar o item 1 observe que se $\bar{x}$ é não estacionário então existe uma direção de descida, e pelo Lemna 1.2, este ponto não é um ótimo fraco de Pareto.

Suponha que f é convexa componente a componente. Se $\bar{x}$ não é um ótimo fraco de Pareto, então existe x tal que $f_j(x) < f_j(\bar{x})$ para $j = 1, \ldots, m$. Neste caso, segue-se da convexidade de f que

$$\nabla f_j(\bar{x})^\top (x - \bar{x}) \leq f_j(x) - f_j(\bar{x}) < 0, \qquad j = 1, \ldots, m$$

e $\bar{x}$ não é estacionário. O item 2 decorre deste resultado e do item 1.

Suponha que f é estritamente convexa componente a componente. Se $\bar{x}$ é um ótimo fraco de Pareto e não é um ótimo de Pareto, então existe x tal que $f_j(x) \leq f_j(\bar{x})$ para $j = 1, \ldots, m$ e $f_{j_0}(x) < f_{j_0}(\bar{x})$ para algum j_0. Em particular, $x \neq \bar{x}$ e da convexidade estrita de f segue-se que

$$f_j\left(\frac{x + \bar{x}}{2}\right) < \frac{f_j(x) + f_j(\bar{x})}{2} \leq f_j(\bar{x}).$$

Provamos que, para funções com componentes estritamente convexas, a otimalidade de Pareto fraca é equivalente à otimalidade de Pareto. O item 3 decorre deste resultado e do item 2. □

Em vista do item 1 do Corolário 1.3, uma condição necessária para que um ponto x seja ótimo de Pareto (fraco ou não) é que tal ponto seja estacionário. Podemos então definir um método geral para otimização multiobjetivo irrestrita (diferenciável) que emprega direções de descida como direções de busca.

Algoritmo 1 (Método de descida)

1. tome $\sigma \in (0, 1)$, $x^0 \in \mathbb{R}^n$; $k \leftarrow 0$;
2. se x^k é crítico, PARE; caso contrário:
3. encontre v^k direção de descida para f em x^k;
4. encontre $t_k > 0$ tal que

$$f(x^k + t_k v^k) \leq f(x^k) + \sigma t_k J f(x^k) v^k;$$

5. $x^{k+1} := x^k + t_k v^k$, $k \leftarrow k+1$ e VÁ PARA 2;

Apenas propriedades muito gerais podem ser provadas para uma sequência $\{x^k\}$ gerada por esse algoritmo. Notemos que a sequência $\{f(x^k)\}$ é estritamente decrescente. Se a sequência $\{x^k\}$ é finita, o último iterado é crítico. O caso de interesse é, portanto, aquele em que essa sequência é infinita. Em minimização escalar, métodos de descida aplicados a funções com conjuntos de nível limitados geram sequências limitadas. Em nosso contexto, define-se um *conjunto de nível* de f como:

$$\{x \in \mathbb{R}^n \ \mid \ f(x) \leq y\},$$

onde $y \in \mathbb{R}^m$.

Proposição 1.4. *Seja $\{x^k\}$ uma sequência infinita gerada pelo Algoritmo 1.*

1. *Se a sequência $\{f(x^k)\}$ é limitada, então ela é convergente e*
$$\begin{aligned}\sum_{k=0}^{\infty} t_k \left|\nabla f_j(x^k)^\top v_k\right| &= \sum_{k=0}^{\infty} t_k \left(-Jf(x^k)v^k\right)_j \\ &\leq \frac{f_j(x^0) - \lim_{k\to\infty} f_j(x^k)}{\sigma} < \infty,\end{aligned}$$
para $j = 1, \ldots, m$.

2. *Se $\bar{x}$ é um ponto de acumulação de $\{x^k\}$, então*
$$\lim_{k\to\infty} f(x^k) = f(\bar{x}), \ \ f(\bar{x}) \leq f(x^k) \ \ \text{para todo } \ k.$$
Em particular, f é constante no conjunto de pontos de acumulação de $\{x^k\}$.

3. *Se x^0 pertence a um conjunto de nível limitado, então $\{x^k\}$ é limitada e $\{f(x^k)\}$ converge.*

Demonstração. Como $\{f(x^k)\}$ é decrescente componente a componente, se $\{f(x^k)\}$ é limitada, então, para $j = 1, \ldots, m$, $\{f_j(x^k)\}$ é uma sequência decrescente e limitada e, portanto, convergente.

Como estamos supondo que a sequência $\{x_k\}$ é infinita, cada x^k é não estacionário. Segue-se do Passo 3 do Algoritm 1 que cada v_k é uma direção de descida em x^k, o que prova a igualdade entre os somatórios. Segue-se do Passo 4 que para cada N

$$\sum_{k=0}^{N} t_k \left(-Jf(x^k)v^k\right)_j \leq \frac{f_j(x^0) - f_j(x^{N+1})}{\sigma}.$$

Para terminar a prova do item 1, observe que cada $\{f_j(x^k)\}_{k\in\mathbb{N}}$ é decrescente e convergente.

Para provar o item 2, suponha que $\{x^{k_i}\}$ é uma subsequência de $\{x^k\}$ que converge a $\bar{x}$. Segue-se da continuidade de f que $\{f_j(x^{k_i})\}$ converge a $f_j(\bar{x})$ para $j = 1, \ldots, m$. Como, $\{f_j(x^k)\}_{k\in\mathbb{N}}$ é decrescente para todo j, cada uma dessas sequências converge ao limite da subsequência correspondente, sendo este limite o seu ínfimo.

O item 3 decorre trivialmente do item 1. $\square$

Se os passos t_k são muito pequenos, a sequência gerada pode convergir a um ponto não estacionário. Este fenômeno ocorre já na otimização escalar (ver [4]). Para escolher um passo t_k adequado, prescrevemos o *backtracking* usual. Neste caso, o Passo 4 do Algoritmo 1 toma a seguinte forma:

4 a. $t := 1$;

4 b. se $f(x^k + tv^k) \leq f(x^k) + \sigma t Jf(x^k)v^k$, então

$$t_k := t \quad \textit{Fim} \text{ (do backtracking);}$$

4 c. caso contrário, $t \leftarrow t/2$ e VÁ PARA 4 b;

O Lema 1.2 garante que o procedimento acima tem terminação finita. Observe que no lugar do fator 1/2 para a redução de t no Passo 4 b pode-se utilizar qualquer outro número no intervalo $(0, 1)$. Agora o problema principal, a ser tratado na próxima seção, é a escolha da direção de descida v^k.

Capítulo 2

O método de Cauchy multiobjetivo

Nesta seção propomos uma extensão ao caso multiobjetivo do clássico método de máxima descida. Primeiro definimos a direção de máxima descida multiobjetivo e estudamos suas propriedades. Em seguida definimos o algoritmo de Cauchy multiobjetivo e estudamos suas propriedades de convergência no caso geral. Finalizamos o capítulo com a análise de convergência do método no caso convexo, isto é, quando f é convexa componente a componente.

2.1 A direção de Cauchy multiobjetivo

Definimos a *direção de Cauchy multiobjetivo* ou *direção de máxima descida multiobjetivo* para f em x como o vetor $\Lambda f(x)$,

$$\Lambda f(x) = \arg\min \max_{j=1,\dots,m} \nabla f_j(x)^\top v + \frac{1}{2}\|v\|^2, \quad v \in \mathbb{R}^n. \tag{2.1}$$

Se $m = 1$ então $\Lambda f(x) = -\nabla f(x)$, e recuperamos a direção de máxima descida "clássica". Como a função objetivo do problema de minimização que define $\Lambda f(x)$ é fortemente convexa e contínua, ela possui um único minimizador e, portanto, $\Lambda f(x)$ está bem definida.

O valor ótimo deste problema de minimização será chamado de $\theta_f(x)$, isto é

$$\theta_f(x) = \min \max_{j=1,\ldots,m} \nabla f_j(x)^\top v + \frac{1}{2}\|v\|^2, \quad v \in \mathbb{R}^n. \tag{2.2}$$

Segue-se de (1.4), (2.1) e (2.2) que

$$\Lambda f(x) = \arg\min_{v\in\mathbb{R}^n} \varphi_x(v) + \frac{1}{2}\|v\|^2 \tag{2.3}$$

$$\begin{aligned} \theta_f(x) &= \min_{v\in\mathbb{R}^n} \varphi_x(v) + \frac{1}{2}\|v\|^2 \\ &= \varphi_x(\Lambda f(x)) + \frac{1}{2}\|\Lambda f(x)\|^2. \end{aligned} \tag{2.4}$$

Note que o par $(v^*, \tau^*) = (\Lambda f(x), \varphi_x(\Lambda f(x)))$ e o escalar $\theta_f(x)$ são, respectivamente, a solução e o valor ótimo do seguinte problema de programação quadrática (convexa) nas variáveis $(v, \tau) \in \mathbb{R}^n \times \mathbb{R}$:

$$\begin{array}{ll} \text{minimize} & \tau + \frac{1}{2}\|v\|^2 \\ \text{sujeito a} & \nabla f_j(x)^\top v - \tau \le 0, \quad j = 1, \ldots, m. \end{array} \tag{2.5}$$

Usaremos esta reformulação dos problemas em (2.3) e (2.4) para mostrar que a direção de Cauchy é uma combinação convexa das direções de máxima descida de cada componente da função objetivo.

Lema 2.1. *Para todo $x \in \mathbb{R}^n$ existe $w \in \mathbb{R}^m$ tal que*

$$w \ge 0, \quad \sum_{j=1}^m w_j = 1, \quad \Lambda f(x) = -\sum_{j=1}^m w_j \nabla f_j(x).$$

Demonstração. As restrições do problema (2.5) são lineares. Logo, existem multiplicadores de Lagrange $w \in \mathbb{R}^m_+$ satisfazendo, juntamente com (v^*, τ^*), as condições de Karush-Kuhn-Tucker. Em particular

$$v^* + \sum_{j=1}^m w_j \nabla f_j(x) = 0, \;\; 1 - \sum_{j=1}^m w_j = 0.$$

Para terminar a prova, observe que $\Lambda f(x) = v^*$. □

Note que no lema anterior, $\Lambda f(x)$ é a direção de Cauchy clássica para a função $\sum_{j=1}^{m} w_j f_j$ no ponto x. Estes pesos w_j dependem do ponto x e estão definidos implicitamente como multiplicadores de Lagrange. A seguir relacionaremos os valores de $\|\Lambda f(x)\|$, $\varphi_x(\Lambda f(x))$ e $\theta_f(x)$.

Lema 2.2. *Para todo $x \in \mathbb{R}^n$*

$$\varphi_x(\Lambda f(x)) = 2\theta_f(x) = -\|\Lambda f(x)\|^2 \geq -\left(\max_{j=1,\ldots,m} \|\nabla f_j(x)\|\right)^2.$$

Portanto, $\theta_f(x) \leq 0$ e $\|\Lambda_f(x)\| \leq \|Jf(x)\|$ para todo $x \in \mathbb{R}^n$.

Demonstração. Segue-se de (2.3) que a função

$$(0, \infty) \ni t \mapsto \varphi_x(t\Lambda f(x)) + \frac{1}{2}\|t\Lambda f(x))\|^2$$

atinge seu mínimo em $t = 1$. Como φ_x é positivamente homogênea,

$$0 = \frac{d}{dt}\Big(t\varphi_x(\Lambda f(x)) + \frac{1}{2}\|t\Lambda f(x))\|^2\Big)_{t=1} = \varphi_x(\Lambda f(x)) + \|\Lambda f(x)\|^2$$

o que, em vista de (2.4), prova as duas igualdades.

Usando (2.2) e a desigualdade de Cauchy-Schwarz

$$\begin{aligned}\theta_f(x) &\geq \min_v \max_{j=1,\ldots,m} -\|v\|\|\nabla f_j(x)\| + \frac{1}{2}\|v\|^2 \\ &= -\frac{1}{2}\left(\max_{j=1,\ldots,m} \|\nabla f_j(x)\|\right)^2. \end{aligned}$$ □

Vamos agora mostrar que

$$\Lambda f(x) = 0 \iff x \text{ é estacionário}$$

e que em pontos não críticos a direção de Cauchy $\Lambda f(x)$ é de descida. Este resultado nos interessa, pois mostra que no Algoritmo 1 podemos empregar a direção de Cauchy como a direção de descida: $v^k = \Lambda f(x^k)$.

Proposição 2.3. *Para todo $x \in \mathbb{R}^n$ as seguintes condições são equivalentes:*

(a) *o ponto x é não estacionário;*

(b) $\theta_f(x) < 0$*;*

(c) $\Lambda f(x) \neq 0$*;*

(d) $\Lambda f(x)$ *é uma direção de descida.*

Demonstração. (a)⇒(b): Se x é não estacionário, existe $v \in \mathbb{R}^n$ tal que $\nabla f_j(x)^\top v < 0$ para $j = 1, \ldots, m$. Logo $\varphi_x(v) < 0$ e segue-se de (2.4) e (1.5) que para todo $\tau > 0$

$$\theta_f(x) \leq \varphi_x(\tau v) + \frac{1}{2}\|\tau v\|^2 = \tau\Big(\varphi_x(v) + \frac{1}{2}\tau\|v\|^2\Big).$$

Portanto, $\lim_{\tau \to 0^+} \theta_f(x)/\tau \leq \varphi_x(v) < 0$ e vale (b).
(b)⇒(c) e (c)⇒(d) seguem-se do Lema 2.2.
(d) ⇒ (a) segue-se trivialmente da definição de estacionariedade. □

2.2 O método de Cauchy multiobjetivo

Estamos agora em condições de definir a extensão do clássico método de máxima descida para o problema irrestrito de otimização multiobjetivo (1.3).

Algoritmo 2 (Método de Cauchy ou de Máxima Descida Multiobjetivo)

1. escolha $\sigma \in (0, 1/2)$ e $x^0 \in \mathbb{R}^n$; $k \leftarrow 0$;

2. calcule $v^k := \Lambda\, f(x^k)$, se $v_k = 0$, PARE;

3 a. $t \leftarrow 1$;

3 b. se $f(x^k + tv^k) \leq f(x^k) + \sigma t J f(x^k) v^k$, então

$$t_k := t \quad \textit{Fim} \text{ (do Passo 3);}$$

3 c. caso contrário, $t \leftarrow t/2$ e VÁ PARA 3 b;

4. $x^{k+1} := x^k + t_k v^k$, $k \leftarrow k + 1$ e VÁ PARA 2.

Vamos analisar uma iteração $k+1$ genérica deste algoritmo. Se $v^k = 0$, então, em virtude da Proposição 2.3, x^k é estacionário e o algoritmo para no Passo 2; caso contrário, v^k é uma direção de descida e, pelo Lema 1.2, a busca de Armijo executada no Passo 3 tem terminação finita, o algoritmo gera x^{k+1} e prossegue para a iteração seguinte. A escolha de t_k garante um decréscimo de f como prescrito no Passo 4 do Algoritmo 1. Portanto, o Algoritmo 2 é um caso particular do Algoritmo 1.

2.3 Análise de convergência no caso geral

Para estabelecer a estacionariedade dos pontos de acumulação das sequências geradas pelo Algoritmo 2, no caso $\mathcal{C}^1$, usaremos dois resultados técnicos que provamos a seguir.

Uma vez que Λf é computada usando-se os gradientes das f_j, $j = 1, \ldots, m$, é natural supor que f seja continuamente diferenciável para obter a continuidade da aplicação $x \mapsto \Lambda f(x)$.

Proposição 2.4. *Se f é continuamente diferenciável, então Λf e θ_f são contínuas.*

Demonstração. Dados $x, y \in \mathbb{R}^n$, para qualquer $v \in \mathbb{R}^n$

$$\begin{aligned}
\varphi_y(v) + \frac{1}{2}\|v\|^2 &\geq \varphi_x(v) + \frac{1}{2}\|v\|^2 - \|Jf(x) - Jf(y)\|\|v\| \\
&\geq \theta_f(x) + \frac{1}{2}\|v - \Lambda f(x)\|^2 - \|Jf(x) - Jf(y)\|\|v\|,
\end{aligned}$$

onde a primeira desigualdade segue de (1.6) e a segunda segue-se de que a função minimizada em (2.4) é fortemente convexa com módulo de convexidade 1. Tomando acima $v = \Lambda f(y)$, concluímos que

$$\theta_f(y) \geq \theta_f(x) + \frac{1}{2}\|\Lambda f(y) - \Lambda f(x)\|^2 - \|Jf(x) - Jf(y)\|\|\Lambda f(y)\|.$$

Trocando os papéis de x e y obtemos

$$\theta_f(x) \geq \theta_f(y) + \frac{1}{2}\|\Lambda f(y) - \Lambda f(x)\|^2 - \|Jf(x) - Jf(y)\|\|\Lambda f(x)\|.$$

Somando as duas últimas desigualdades, temos que

$$\|\Lambda f(y) - \Lambda f(x)\|^2 \leq \|Jf(x) - Jf(y)\| \Big(\|\Lambda f(x)\| + \|\Lambda f(y)\| \Big).$$

A continuidade de Λ_f segue desta desigualdade, da última desigualdade do Lema 2.2 e da continuidade de Jf.

Para provar a continuidade de θ_f observe que, segundo o Lema 2.2,

$$\theta_f(x) = -\frac{1}{2}\|\Lambda_f(x)\|^2$$

e use a continuidade de Λ_f. □

Lema 2.5. *Se f é continuamente diferenciável e $\bar{x}$ não é crítico, então para qualquer $\sigma \in (0,1)$ existem $\bar{t} > 0$ e $\delta > 0$ tais que*

$$f(x + t\Lambda f(x)) \leq f(x) + \sigma t Jf(x)\, \Lambda f(x) \quad \forall x \in B(\bar{x}, \delta),\ t \in (0, \bar{t}).$$

Demonstração. Em vista da Proposição 2.3, $\theta_f(\bar{x}) < 0$. Logo, pela Proposição 2.4, existe $r_0 > 0$ tal que

$$\theta_f(x) \leq \theta_f(\bar{x})/2 \qquad \forall x \in B(\bar{x}, r_0). \tag{2.6}$$

Sejam

$$M_0 = \sup\{\|\Lambda f(x)\| \mid \|x - \bar{x}\| \leq r_0\}, \qquad \varepsilon = \frac{1-\sigma}{M_0}|\theta_f(\bar{x})|.$$

Observe que em virtude da Proposição 2.4, $M_0 < \infty$.

Existe $r_1 \in (0, r_0)$ tal que

$$x, x' \in B(\bar{x}, r_1) \Longrightarrow \|\nabla f_j(x) - \nabla f_j(x')\| < \varepsilon,\ j = 1, \ldots, m.$$

Para $x \in B(\bar{x}, r_1/2)$ e $t \in (0, r_1/(2M_0))$, temos $x + t\Lambda_f(x) \in B(\bar{x}, r_1)$ e, pelo item 1 do Lema 4.3 e pela definição de ε,

$$\begin{aligned} f_j(x + t\Lambda f(x)) &\leq f_j(x) + t\nabla f_j(x)^\top \Lambda_f(x) + t\varepsilon M_0 \\ &\leq f_j(x) + t\big[\nabla f_j(x)^\top \Lambda_f(x) + (1-\sigma)|\theta_f(\bar{x})|\big] \end{aligned}$$

para $j = 1, \ldots, m$. Portanto, pelo Lema 2.2, (1.4) e (2.6),

$$\nabla f_j(x)^\top \Lambda f(x) \leq \varphi_x(\Lambda_f(x)) = 2\theta_f(x) \leq \theta_f(\bar{x}).$$

Combinando as duas últimas desigualdades, obtemos

$$\begin{aligned} f_j(x + t\Lambda f(x)) &\leq f_j(x) + \sigma t \nabla f_j(x)^\top \Lambda_f(x) \\ &\qquad + t(1-\sigma)\left[\nabla f_j(x)^\top \Lambda f(x) + |\theta_f(\bar{x})|\right] \\ &\leq f_j(x) + \sigma t \nabla f_j(x)^\top \Lambda_f(x), \end{aligned}$$

onde a última desigualdade decorre do fato de que $\theta_f(\bar{x}) < 0$.

Para concluir a prova, tome $0 < \delta < r_1/2$ e $0 < \bar{t} < r_1/(2M_0)$. □

O Algoritmo 2 termina com um ponto estacionário ou produz uma sequência infinita $\{x^k\}$ de pontos não estacionários. De agora em diante, supomos que o algoritmo não tem terminação finita e gera as sequências infinitas $\{x^k\}$, $\{v^k\}$ e $\{t_k\}$.

Apresentamos a seguir uma simples aplicação do argumento padrão de convergência para o método de máxima descida em otimização a valores escalares.

Teorema 2.6. *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 2. Se f é continuamente diferenciável, então todo ponto de acumulação de $\{x^k\}$ é estacionário.*

Demonstração. Seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$. Segue-se da Proposição 1.4, itens 2 e 1, que

$$\lim_{k\to\infty} f(x^k) = f(\bar{x}), \quad \lim_{k\to\infty} t_k \nabla f_j(x^k)^\top v^k = 0 \quad \text{para } j = 1, \ldots, m.$$

Logo, pelo Lema 2.2,

$$\lim_{k\to\infty} t_k \varphi_{x^k}(v^k) = \lim_{k\to\infty} t_k \theta_f(x_k) = 0. \tag{2.7}$$

Vamos provar por absurdo que $\bar{x}$ é estacionário. Se ele não for, então, em virtude da Proposição 2.3, $\bar{v} = \Lambda_f(\bar{x}) \neq 0$ e $\theta_f(\bar{x}) < 0$. Existe uma subsequência $\{x^{k_j}\}$ que converge a $\bar{x}$. Segue-se do Lema 2.5 e da definição do Passo 3 que existe $M \in \mathbb{N}$ tal que $t_{k_j} \geq 2^{-M}$ para j suficientemente grande. Pela Proposição 2.4 temos $\theta_f(x^{k_j}) \to \theta_f(\bar{x}) < 0$ quando $j \to \infty$. Portanto $\liminf_{j\to\infty} t_{k_j}\theta_f(x^{k_j}) < 0$ em contradição com (2.7). □

2.4 Análise de convergência no caso convexo

Em minimização escalar irrestrita, uma condição necessária e suficiente para a existência de ótimos é que toda sequência decrescente na imagem da função objetivo seja limitada inferiormente por um elemento da imagem. Vamos precisar de uma propriedade análoga, que suporemos válida:

Hipótese 2.7. *Toda sequência $\{y^k\} \subset f(\mathbb{R}^n)$ decrescente (componente a componente) é limitada inferiormente (componente a componente) por algum $y \in f(\mathbb{R}^n)$.*

Para provar a convergência do método de Cauchy multiobjetivo no caso convexo, além da condição acima, precisaremos utilizar as propriedades da convergência quasi-Fejér.

Definição 2.8. *Uma sequência $\{z^k\} \subset \mathbb{R}^n$ converge* quasi-Fejér *a $C \subset \mathbb{R}^n$ se, para cada $z \in C$ existe uma sequência de números não-negativos $\{\varepsilon_k\}$ tal que*

$$\|z^{k+1} - z\|^2 \leq \|z^k - z\|^2 + \varepsilon_k \quad \text{e} \quad \sum_{k=1}^{\infty} \varepsilon_k < \infty.$$

Lema 2.9. *Se $\{z^k\}$ converge quasi-Fejér a $C \neq \emptyset$ então ela é limitada. Se, adicionalmente, a sequência tem um ponto de acumulação em C, então ela converge a tal ponto.*

Demonstração. Tome $z \in C$ e seja $\{\varepsilon_k\}$ como especificado na Definição 2.8. Para todo $i < j$, $\|z - z^j\|^2 \leq \|z - z^i\|^2 + \sum_{k=i}^{\infty} \varepsilon_k$. Logo,

$$\limsup_{j\to\infty} \|z - z^j\|^2 \leq \|z - z^i\|^2 + \sum_{k=i}^{\infty} \varepsilon_k < \infty.$$

Tomando o $\liminf_{i\to\infty}$ nas desigualdades acima, concluímos que $\limsup_{k\to\infty} \|z - z^k\|^2 \leq \liminf_{k\to\infty} \|z - z^k\|^2$. Portanto existe

$$\lim_{k\to\infty} \|z - z^k\| \in \mathbb{R}.$$

As conclusões seguem trivialmente deste resultado. $\square$

Teorema 2.10. *Se f é convexa componente a componente, diferenciável e satisfaz a Hipótese 2.7, então a sequência $\{x^k\}$ converge a um ótimo fraco de Pareto. Se, adicionalmente, f é fortemente convexa componente a componente, então o limite da sequência é um ótimo de Pareto.*

Demonstração. A sequência $\{f(x^k)\}$ é decrescente. Como f satisfaz a Hipótese 2.7, o conjunto

$$X = \{x \in \mathbb{R}^n \mid f(x) \le f(x^k)\ \forall k \in \mathbb{N}\} \tag{2.8}$$

é não vazio. Vamos provar que a sequência $\{x^k\}$ converge quasi-Fejér a X.

Seja $\hat{x} \in X$. Em virtude do Lema 2.1, para cada $k \in \mathbb{N}$ existe $w^k \in \mathbb{R}^m_+$ tal que

$$\sum_{j=1}^m w_j^k = 1, \qquad v^k = -\sum_{j=1}^m w_j^k \nabla f_j(x^k). \tag{2.9}$$

Logo, da convexidade de cada f_j segue-se que

$$\begin{aligned}(w^k)^\top f(\hat{x}) &\ge \sum_{j=1}^m w_j^k [f_j(x^k) + \nabla f_j(x^k)^\top (\hat{x} - x^k)] \\ &= (w^k)^\top f(x^k) - (v^k)^\top (\hat{x} - x^k),\end{aligned}$$

e, pelo Passo 4,

$$\begin{aligned}(x^{k+1} - x^k)^\top (\hat{x} - x^k) &= t_k (v^k)^\top (\hat{x} - x^k) \\ &\ge t_k (w^k)^\top (f(x^k) - f(\hat{x})) \ge 0,\end{aligned}$$

onde a última desigualdade decorre da hipótese $\hat{x} \in X$. Portanto

$$\begin{aligned}\|\hat{x} - x^{k+1}\|^2 &= \|\hat{x} - x^k\|^2 + \|x^k - x^{k+1}\|^2 + 2(\hat{x} - x^k)^\top (x^k - x^{k+1}) \\ &\le \|\hat{x} - x^k\|^2 + \|x^k - x^{k+1}\|^2.\end{aligned} \tag{2.10}$$

Das desigualdades $0 < t_k \leq 1$ e de (2.9) segue-se que, para todo k,

$$\begin{aligned}
\|x^k - x^{k+1}\|^2 \leq t_k \|v^k\|^2 = t_k (v^k)^\top v^k &= \sum_{j=1}^m -t_k w_j^k \nabla f_j(x^k)^\top v^k \\
&= \sum_{j=1}^m w_j^k t_k \left|\nabla f_j(x^k)^\top v^k\right| \\
&\leq \sum_{j=1}^m t_k \left|\nabla f_j(x^k)^\top v^k\right|.
\end{aligned}$$

Logo, para todo N,

$$\begin{aligned}
\sum_{k=0}^N \|x^k - x^{k+1}\|^2 &\leq \sum_{k=0}^N \sum_{j=1}^m t_k \left|\nabla f_j(x^k)^\top v^k\right| \\
&= \sum_{j=1}^m \sum_{k=0}^N t_k \left|\nabla f_j(x^k)^\top v^k\right| \\
&\leq \sum_{j=1}^m \sum_{k=0}^\infty t_k \left|\nabla f_j(x^k)^\top v^k\right|.
\end{aligned}$$

Como f satisfaz a Hipótese 2.7, segue-se da Proposição 1.4 que o termo da última desigualdade, que não depende de N, é finito. Logo $\sum_{k=0}^\infty \|x^k - x^{k+1}\| < \infty$ e em vista de (2.10), $\{x^k\}$ converge quasi-Fejér ao conjunto X definido em (2.8). Portanto, a sequência $\{x^k\}$ é limitada e possui um ponto de acumulação $\bar{x}$. Pelo item 2 da Proposição 1.4, tal ponto pertence a X. Usando novamente a convergência quasi-Fejér de $\{x^k\}$ a X e o Lema 2.9, concluímos que tal sequência converge a $\bar{x}$.

Cada f_j é convexa e diferenciável; portanto, f é continuamente diferenciável. Logo, pelo Teorema 2.6, tal ponto é estacionário. Para concluir a prova, combine estes resultados com o Corolário 1.3. $\square$

Capítulo 3

O método de Newton multiobjetivo

Nesta seção estudamos um método de Newton para o problema (1.3) supondo, adicionalmente, que f é duas vezes diferenciável e que os Hessianos $\nabla^2 f_j(x)$ são positivos definidos para todo $x \in \mathbb{R}^n$ e $j = 1, \ldots, m$. Observe que esta hipótese é utilizada no caso $m = 1$ (minimização escalar) para que o método de Newton esteja bem definido e para que a direção de Newton seja uma direção de descida. Assim como no caso escalar, a hipótese dos Hessianos serem positivos definidos poderia ser relaxada usando-se técnicas de regiões de confiança.

3.1 A direção de Newton multiobjetivo

A aproximação de segunda ordem no ponto x de $f_j(x+s) - f_j(x)$ é

$$\nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s, \qquad j = 1, \ldots, m.$$

Definimos a *direção de Newton multiobjetivo* para f no ponto x como

$$s_f(x) = \arg\min_{s \in \mathbb{R}^n} \max_{j=1,\ldots,m} \nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s. \tag{3.1}$$

A direção $s_f(x)$ está bem definida em virtude da hipótese dos Hessianos das f_j serem positivos definidos. Se $m = 1$, recuperamos o passo de Newton para minimização escalar. O valor ótimo do problema de minimização em (3.1) será chamado de $\theta_f^N(x)$, ou seja

$$\begin{aligned}
\theta_f^N(x) &= \min_{s \in \mathbb{R}^n} \max_{j=1,\dots,m} \nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s \\
&= \max_{j=1,\dots,m} \nabla f_j(x)^\top s_f(x) + \frac{1}{2} s_f(x)^\top \nabla^2 f_j(x) s_f(x).
\end{aligned} \tag{3.2}$$

Note que o par $(s^*, \tau^*) = (s_f(x), \theta_f^N(x))$ é a solução do seguinte problema de programação quadrática (convexa) nas variáveis $(s, \tau) \in \mathbb{R}^n \times \mathbb{R}$:

$$\begin{aligned}
&\text{minimize} \quad \tau \\
&\text{sujeito a} \quad \nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s - \tau \leq 0, \quad j = 1, \dots, m.
\end{aligned} \tag{3.3}$$

Usaremos esta reformulação dos problemas (3.1) e (3.2) para mostrar que a direção de Newton multiobjetivo é a direção de Newton (clássica) de uma combinação convexa das componentes da função objetivo.

Lema 3.1. *Para todo $x \in \mathbb{R}^n$ existe $w \in \mathbb{R}^m$ tal que*

$$\begin{aligned}
s_f(x) &= -\left[\sum_{j=1}^m w_j \nabla^2 f_j(x)\right]^{-1} \sum_{j=1}^m w_j \nabla f_j(x), \\
\theta_f^N(x) &= \sum_{j=1}^m w_j \left[\nabla f_j(x)^\top s_f(x) + \frac{1}{2} s_f(x)^\top \nabla^2 f_j(x) s_f(x)\right], \\
w &\geq 0, \ \sum_{j=1}^m w_j = 1.
\end{aligned}$$

Demonstração. O lagrangeano do problema (3.3) é

$$\mathcal{L}_x((s,\tau), w) = \tau + \sum_{j=1}^m w_j \left[\nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s - \tau\right].$$

Como as restrições deste problema são quadráticas convexas e no ponto $(s,\tau)=(0,1)$ se verifica a condição de Slater, existem multiplicadores de Lagrange $w \in \mathbb{R}^m_+$ que satisfazem, juntamente com (s^*,τ^*), as condições de Karush-Kuhn-Tucker. Em particular,

$$\sum_{j=1}^{m} w_j \left[\nabla f_j(x) + \nabla^2 f_j(x) s^*\right] = 0, \quad 1 - \sum_{j=1}^{m} w_j = 0,$$
$$\sum_{j=1}^{m} w_j \left[\nabla f_j(x)^\top s^* + \frac{1}{2}(s^*)^\top \nabla^2 f_j(x) s^* - \tau^*\right] = 0.$$

Para terminar a prova, observe que $(s^*,\tau^*) = (s_f(x), \theta_f^N(x))$ e use a primeira igualdade acima para caracterizar $s_f(x)$. $\square$

Note que no lema anterior, $s_f(x)$ é direção clássica de Newton para a função $\sum_{j=1}^{m} w_j f_j$ no ponto x. Damos agora cotas superiores para a norma da direção de Newton.

Lema 3.2. *Se $x \in \mathbb{R}^n$ e $\lambda_{\min}$ e $\lambda_{\max}$ são, respectivamente, o menor e o maior dos autovalores dos Hessianos $\nabla^2 f_j(x)$ para $j = 1,\ldots,m$, então*

1. $\nabla f_j(x)^\top s_f(x) \le \theta_f^N(x)$ *para* $j = 1,\ldots,m$;
2. $\dfrac{1}{\lambda_{\max}} |\theta_f(x)| \le -\theta_f^N(x) \le \dfrac{1}{\lambda_{\min}} |\theta_f(x)|$, *com* $\theta_f(x)$ *dado por* (2.4); *em particular,* $\theta_f^N(x) \le 0$;
3. $\dfrac{\lambda_{\min}}{2}\|s_f(x)\|^2 \le -\theta_f^N(x) \le \dfrac{\lambda_{\max}}{2}\|s_f(x)\|^2$;
4. $\|s_f(x)\| \le \dfrac{1}{\lambda_{\min}} \max_{j=1,\ldots,m} \|\nabla f_j(x)\|$.

Demonstração. O item 1 segue da hipótese de que os Hessianos de todas as f_j são positivos definidos e da definição (3.2).

Para todo $s \in \mathbb{R}^n$,

$$\lambda_{\min}\|s\|^2 \le s^\top \nabla^2 f_j(x) s \le \lambda_{\max}\|s\|^2 \tag{3.4}$$

Logo

$$\begin{aligned}\min_{s\in\mathbb{R}^n} \max_j \nabla f_j(x)^\top s + \frac{\lambda_{\min}}{2}\|s\|^2 &\leq \theta_f^N(x) \\ &\leq \min_{s\in\mathbb{R}^n} \max_j \nabla f_j(x)^\top s + \frac{\lambda_{\max}}{2}\|s\|^2\end{aligned}$$

Para provar as duas primeiras desigualdades do item 2, use as desigualdades acima, recorde que $\theta_f(x) \leq 0$ (Lema 2.2) e observe que para todo $\lambda > 0$,

$$\begin{aligned}\min_{s\in\mathbb{R}^n} \max_j \nabla f_j(x)^\top s + \frac{\lambda}{2}\|s\|^2 &= \frac{1}{\lambda} \min_{s\in\mathbb{R}^n} \max_j \nabla f_j(x)^\top \lambda s + \frac{1}{2}\|\lambda s\|^2 \\ &= \frac{\theta_f(x)}{\lambda}.\end{aligned}$$

A última desigualdade do item 2 segue trivialmente da primeira.

Tome w como no Lema 3.1. Usando a primeira e a segunda igualdade desse lema, temos que

$$\sum_{j=1}^m w_j \nabla f_j(x) = -\sum_{j=1}^m w_j \nabla^2 f_j(x) s_f(x)$$

e

$$\theta_f^N(x) = \left[\sum_{j=1}^m w_j \nabla f_j(x)\right]^\top s_f(x) + \frac{1}{2} s_f(x)^\top \left[\sum_{j=1}^m w_j \nabla^2 f_j(x)\right] s_f(x).$$

Logo,

$$\begin{aligned}\theta_f^N(x) &= - s_f(x)^\top \left[\sum_{j=1}^m w_j \nabla^2 f_j(x)\right] s_f(x) \\ &\quad + \frac{1}{2} s_f(x)^\top \left[\sum_{j=1}^m w_j \nabla^2 f_j(x)\right] s_f(x), \\ &= - \frac{1}{2} s_f(x)^\top \left[\sum_{j=1}^m w_j \nabla^2 f_j(x)\right] s_f(x).\end{aligned}$$

Para concluir a prova do item 3, observe que $w \geq 0$, $\sum_{j=1}^m w_j = 1$ e use (3.4).

Como $\lambda_{\min} > 0$, segue-se do Lema 3.1 que

$$\|s_f(x)\| \leq \frac{1}{\lambda_{\min}} \sum_{j=1}^m w_j \|\nabla f_j(x)\| \leq \frac{1}{\lambda_{\min}} \max_{j=1,\ldots,m} \|\nabla f_j(x)\|. \quad \square$$

Vamos agora mostrar que

$$s_f(x) = 0 \iff x \text{ é estacionário}$$

e que em pontos não críticos a direção de Newton $s_f(x)$ é de descida.

Proposição 3.3. *Para todo $x \in \mathbb{R}^n$ as seguintes condições são equivalentes:*

(a) *o ponto x é não estacionário;*

(b) $\theta_f^N(x) < 0$*;*

(c) $s_f(x) \neq 0$*;*

(d) $s_f(x)$ *é uma direção de descida.*

Demonstração. Os itens **(a)** e **(b)** são equivalentes em virtude do Lema 3.2 item 2 e da Proposição 2.3. Por sua vez, os itens **(b)** e **(c)** são equivalentes em vista do item 3 do Lema 3.2.

O item **(d)** implica trivialmente o item **(a)**. Finalmente, observe que pelo item 1 do Lema 3.2, o item **(b)** implica o item **(d)**. $\square$

O próximo lema será utilizado para mostrar a boa definição de uma busca tipo Armijo na direção de Newton, usando-se um critério de aceitação (do passo) adaptado ao contexto.

Lema 3.4. *Se $\bar{x}$ não é crítico, então, para qualquer $\sigma \in (0,1)$ existe $\bar{t} > 0$ tal que*

$$f_j(\bar{x} + ts_f(\bar{x})) \leq f_j(\bar{x}) + \sigma t \theta_f^N(\bar{x}) \quad \forall t \in (0, \bar{t}],$$

para $j = 1, \ldots, m$.

Demonstração. Em virtude da Proposição 3.3, $s_f(\bar{x})$ é uma direção de descida. Logo, pelo Lema 1.2, existe $\bar{t} > 0$ tal que

$$f_j(\bar{x} + ts_f(\bar{x})) < f_j(\bar{x}) + \sigma t \nabla f_j(\bar{x})^\top s_f(\bar{x}) \quad \forall t \in (0, \bar{t}],\ j = 1, \ldots, m.$$

Para concluir a prova, note que

$$\nabla f_j(\bar{x})^\top s_f(\bar{x}) \le \theta_f^N(\bar{x})$$

para $j = 1, \ldots, m$. □

3.2 O método de Newton multiobjetivo

Apresentamos a seguir uma versão multiobjetivo do método de Newton clássico.

Algoritmo 3 (Método de Newton Multiobjetivo)

1. escolha $\sigma \in (0, 1)$ e $x^0 \in \mathbb{R}^n$; $k \leftarrow 0$;

2. calcule $s^k := s_f(x^k)$, se $s^k = 0$ PARE;

3 a. $t \leftarrow 1$;

3 b. se $f_j(x^k + ts^k) \le f_j(x^k) + \sigma t \theta_f^N(x^k)$ para $j = 1, \ldots, m$, então

$$t_k := t \quad \textit{Fim} \text{ (do Passo 3);}$$

3 c. caso contrário, $t \leftarrow t/2$, VÁ PARA 3 b;

4. $x^{k+1} := x^k + t_k s^k$, $k \leftarrow k + 1$ e VÁ PARA 2.

Vamos analisar uma iteração $k + 1$ genérica deste algoritmo: se $s^k = 0$, então ele para no Passo 2 e, em virtude da Proposição 3.3, x^k é estacionário; caso contrário, s^k é uma direção de descida e, pelo Lema 3.4, a busca de Armijo executada no Passo 3 tem terminação finita, o algoritmo gera x^{k+1} e prossegue para a iteração seguinte. Acabamos de verificar que o Algoritmo 3 está bem definido para uma função objetivo duas vezes diferenciável com Hessianos dos seus componentes positivos definidos.

O Algoritmo 3 termina com um ponto estacionário ou produz uma sequência infinita $\{x^k\}$ de pontos não estacionários. Nesta seção e nas subsequentes vamos supor que o algoritmo não tem terminação finita e gera as sequências infinitas $\{x^k\}$, $\{s^k\}$ e $\{t_k\}$. Veremos a seguir que para este algoritmo valem resultados análogos aos da Proposição 1.4.

Proposição 3.5. *As sequências $\{f_j(x^k)\}$, $j = 1, \ldots, m$, são decrescentes e*

$$f_j(x^{k+1}) \leq f_j(x^0) + \sigma \sum_{i=0}^{k} t_i \theta_f^N(x^i), \qquad j = 1, \ldots, m$$

para todo k.

Demonstração. Como estamos supondo que a sequência $\{x^k\}$ é infinita, cada x^k é não estacionário e cada s^k é uma direção de descida em x^k. A proposição segue então da definição do Passo 3. □

Corolário 3.6. *1. se $\{f(x^k)\}$ é limitada, então ela é convergente e para todo j*

$$\sum_{k=0}^{\infty} t_k |\theta_f^N(x^k)| < \frac{f_j(x^0) - \lim_{k\to\infty} f_j(x^k)}{\sigma} < \infty.$$

Em particular, f é constante no conjunto de pontos de acumulação de $\{x^k\}$.

2. Se x^0 pertence a um conjunto de nível limitado, então $\{x^k\}$ é limitada e $\{f(x^k)\}$ converge.

Demonstração. Os itens 1 e 2 seguem da Proposição 3.5 e do fato de que $\theta_f^N(x) < 0$ para todo x não estacionário. □

3.3 Análise de convergência

Para estabelecer a convergência de $\{x^k\}$, no caso $\mathcal{C}^2$, vamos provar primeiro que seus pontos de acumulação são estacionários. Em seguida provaremos que se a sequência "passa" suficientemente perto de um ponto crítico, então ela converge a um ponto crítico próximo.

Os dois próximos resultados serão utilizados para mostrar, no caso $\mathcal{C}^2$, a estacionariedade dos pontos de acumulação de $\{x^k\}$.

Proposição 3.7. *Se f é duas vezes continuamente diferenciável, então s_f e θ_f^N são contínuas.*

Demonstração. Tome $\bar{x} \in \mathbb{R}^n$ e $r > 0$. Como a função menor autovalor é contínua (Lema 4.1), existe $\bar{\lambda}$ o mínimo do menor dos autovalores de $\nabla^2 f_j$, $j = 1, \ldots, m$, na bola fechada em $B[\bar{x}, r]$ e $\bar{\lambda} > 0$. Portanto

$$v^\top \nabla^2 f_j(z) v \geq \bar{\lambda}\|v\|^2 \qquad \forall z \in B(\bar{x}, r),\ j = 1, \ldots, m,\ v \in \mathbb{R}^n. \quad (3.5)$$

Dados $x, x' \in B(\bar{x}, r)$, sejam $s = s_f(x)$ e $s' = s_f(x')$. Como cada uma das m funções $v \mapsto \nabla f_j(z)^\top v + \frac{1}{2} v^\top \nabla^2 f_j(z) v$ é $\bar{\lambda}$-fortemente convexa para $z \in B(\bar{x}, r)$,

$$\theta_f^N(x') + \frac{1}{2}\bar{\lambda}\|s - s'\|^2 \leq \max_{j=1,\ldots,m} \nabla f_j(x')^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x') s. \quad (3.6)$$

Para simplificar a exposição, definimos

$$\rho(x, x') = \max_j \left\{ \|\nabla f_j(x') - \nabla f_j(x)\| + \frac{1}{2}\|\nabla^2 f_j(x') - \nabla^2 f_j(x)\| \right\}.$$

Usando esta definição e a desigualdade triangular, temos que para $j = 1, \ldots, m$,

$$\begin{aligned} \nabla f_j(x')^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x') s \leq \nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s \\ + \rho(x, x') \max\{\|s\|, \|s\|^2\}. \end{aligned}$$

Decorre do item 4 do Lema 3.2, de f ser $\mathcal{C}^2$ e de (3.5) que $\|s_f(x)\|$ está limitada na bola $B(\bar{x}, r)$. Usando-se este resultado, (3.6) e tomando o máximo em $j = 1, \ldots, m$ na desigualdade acima, concluímos que existe M tal que

$$\theta_f^N(x') + \frac{1}{2}\bar{\lambda}\|s - s'\|^2 \leq \theta_f^N(x) + \rho(x, x') M$$

Como esta desigualdade vale intercambiando x e x' e como $s = s_f(x)$ e $s' = s_f(x')$, temos que

$$|\theta_f^N(x') - \theta_f^N(x)| + \frac{1}{2}\bar{\lambda}\|s_f(x) - s_f(x')\|^2 \leq \rho(x, x') M$$

para todo $x, x' \in B(\bar{x}, r)$. Para terminar a prova, observe que como f é $\mathcal{C}^2$, $\lim_{x' \to x} \rho(x, x') = 0$. □

Lema 3.8. *Se f é $\mathcal{C}^2$ e $\bar{x}$ não é crítico, então, para qualquer $\sigma \in (0,1)$ existem $\bar{t} > 0$ e $\delta > 0$ tais que*

$$f_j(x + ts_f(x)) \leq f_j(x) + \sigma t \theta_f^N(x) \quad \forall x \in B(\bar{x}, \delta),\ t \in (0, \bar{t}],$$

para $j = 1, \ldots, m$.

Demonstração. Em vista da Proposições 3.3 e 3.7, $\theta_f^N(\bar{x}) < 0$ e existe $r_0 > 0$ tal que

$$\theta_f^N(x) \leq \theta_f^N(\bar{x})/2 \qquad \forall x \in B(\bar{x}, r_0). \tag{3.7}$$

Sejam

$$M_0 = \sup\{\|s_f(x)\| \mid \|x - \bar{x}\| \leq r_0\}, \qquad \varepsilon = -\frac{1-\sigma}{2M_0}\theta_f^N(\bar{x}).$$

Observe que como s_f é contínua, $M_0 < \infty$. Dado $\varepsilon > 0$, existe $r_1 \in (0, r_0)$ tal que

$$x, x' \in B(\bar{x}, r_1) \Longrightarrow \|\nabla f_j(x) - \nabla f_j(x')\| < \varepsilon,\ \ j = 1, \ldots, m.$$

Segue-se da condição acima e do item 1 do Lema 4.3 que, para $x, x' \in B(\bar{x}, r_1)$ e $j = 1, \ldots, m$,

$$f_j(x') \leq f_j(x) + \nabla f_j(x)^\top (x' - x) + \varepsilon \|x' - x\| \qquad \forall x, x' \in B(\bar{x}, r_1),\ j = 1, \ldots, m.$$

Tome $x \in B(\bar{x}, r_1/2)$ e $t \in (0, r_1/(2M_0))$. Como $B(\bar{x}, r_1/2) \subset B(\bar{x}, r_0)$, segue-se da definição de M_0 que $\|s_f(x)\| \leq M_0$ e $x + ts_f(x) \in B(\bar{x}, r_1)$. Logo, da desigualdade acima e da definição de $\theta_f^N(x)$ se depreende que

$$\begin{aligned} f_j(x + ts_f(x)) &\leq f_j(x) + t\nabla f_j(x)^\top s_f(x) + t\varepsilon \|s_f(x)\| \\ &\leq f_j(x) + t\theta_f^N(x) + t\varepsilon M_0 \\ &= f_j(x) + \sigma t \theta_f^N(x) + t\left[(1-\sigma)\theta_f^N(x) + \varepsilon M_0\right]. \end{aligned}$$

De definição de ε e de (3.7) segue-se que

$$(1-\sigma)\theta_f^N(x) + \varepsilon M_0 \leq (1-\sigma)\theta_f^N(x) - \frac{1-\sigma}{2}\theta_f^N(\bar{x}) \leq 0.$$

Para concluir a prova, combine as duas inequações acima e tome $\delta = r_1/2$ e $\bar{t} \in (0, r_1/(2M_0))$. □

Teorema 3.9. *Se f é $\mathcal{C}^2$, então todo ponto de acumulação de $\{x^k\}$ é estacionário.*

Demonstração. Seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$. Em virtude do item 1 do Corolário 3.6,

$$\lim_{k\to\infty} t_k|\theta_f^N(x^k)| = 0. \tag{3.8}$$

Vamos supor que $\bar{x}$ é não estacionário, afim de provar por absurdo o teorema. Segue-se dessa suposição e do Lema 3.8 que existem $\bar{t} \in (0,1]$ e $\delta > 0$ tais que, para $j = 1,\ldots,m$,

$$f_j(x + ts_f(x)) \leq f_j(x) + \sigma t\theta_f^N(x) \quad \forall x \in B(\bar{x},\delta),\ t \in (0,\bar{t}).$$

Sendo $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$, existe uma subsequencia $\{x^{k_i}\}$ tal que

$$x^{k_i} \in B(\bar{x},\delta), \qquad i = 1,2,\ldots$$

Combinando os dois resultados acima com a definição do Passo 3, concluímos que $t_{k_i} \geq \bar{t}/2$ para todo i. Segue-se disto e de (3.8) que $\lim_{i\to\infty} \theta_f^N(x^{k_i}) = 0$. Como θ_f^N é contínua (porque f é $\mathcal{C}^2$), $\theta_f^N(\bar{x}) = 0$ e, pela Proposição 3.3, $\bar{x}$ é crítico. Isto contradiz nossa suposição de $\bar{x}$ ser não estacionário. $\square$

O seguinte lema técnico nos fornece uma relação entre as normas do passo de Newton e do gradiente de escalarizações (com pesos no simplex unitário) da função objetivo.

Lema 3.10. *Se $w \in \mathbb{R}^m_+$, $\sum_{j=1}^m w_j = 1$ e $u = \sum_{j=1}^m w_j \nabla f_j(x)$ então*

$$\|s_f(x)\| \leq \frac{\|u\|}{\lambda_{\min}}$$

onde $\lambda_{\min}$ é o menor dos autovalores dos Hessianos $\nabla^2 f_j(x)$, $j = 1,\ldots,m$.

Demonstração. Usando-se as definições de θ_f^N, $\lambda_{\min}$ e u temos

$$\begin{aligned}
\theta_f^N(x) &\geq \min_s \sum_{j=1}^m w_j [\nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s] \\
&\geq \min_s \left[\sum_{j=1}^m w_j \nabla f_j(x) \right]^\top s + \frac{\lambda_{\min}}{2} \|s\|^2 \\
&= \min_s u^\top s + \frac{\lambda_{\min}}{2} \|s\|^2 = -\frac{\|u\|^2}{2\lambda_{\min}}.
\end{aligned}$$

Logo, usando o item 3 do Lema 3.2, a definições de $\lambda_{\min}$ e a desigualdade acima, concluímos que

$$\|s_f(x)\|^2 \leq \frac{2}{\lambda_{\min}}(-\theta_f^N(x)) \leq \left(\frac{\|u\|}{\lambda_{\min}} \right)^2. \qquad \square$$

A convergência superlinear dos métodos do tipo Newton depende de que o passo não seja relaxado e de que haja redução superlinear das normas das direções de busca. Os resultados seguintes garantem este comportamento no método de Newton multiobjetivo em vizinhanças de pontos críticos.

Proposição 3.11. *Se f é $\mathcal{C}^2$, $\bar{x}$ é estacionário e $\sigma, \kappa \in (0, 1)$, então existe $\delta > 0$ tal que, para todo $x \in B(\bar{x}, \delta)$*

1. $f_j(x + s_f(x)) \leq f_j(x) + \sigma \theta_f^N(x)$ *para* $j = 1, \ldots, m$;
2. $\|s_f(x + s_f(x))\| \leq \kappa \|s_f(x)\|$.

Demonstração. Tome $r_0 > 0$. Como a função menor autovalor é contínua (Lema 4.1), existe $\bar{\lambda}$ o mínimo do menor dos autovalores de $\nabla^2 f_j$, $j = 1, \ldots, m$ na bola fechada $B[\bar{x}, r_0]$ e $\bar{\lambda} > 0$. Defina

$$\varepsilon = \bar{\lambda} \min\{1 - \sigma, \kappa\}.$$

Existe $r_1 \in (0, r_0)$ tal que

$$\|\nabla^2 f_j(x) - \nabla^2 f_j(x')\| < \varepsilon \qquad \forall x, x' \in B(\bar{x}, r_1) \tag{3.9}$$

Com $\bar{x}$ é estacionário, usando a Proposição 3.3 e a continuidade de s_f vemos que existe $\delta \in (0, r_1/2)$ tal que

$$\|s_f(x)\| \leq r_1/2 \qquad \forall x \in B(\bar{x}, \delta).$$

Sejam

$$x \in B(\bar{x}, \delta), \qquad s = s_f(x). \tag{3.10}$$

Em virtude da escolha de δ,

$$x + s \in B(\bar{x}, r_1).$$

Tome $\lambda_{\min}$ como o menor dos autovalores dos Hessianos $\nabla^2 f_j(x)$ para $j = 1, \ldots, m$. Como $x, x + s \in B(\bar{x}, r_1)$, pelo item 2 do Lema 4.3

$$\begin{aligned} f_j(x+s) &\leq f_j(x) + \nabla f_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 f_j(x) s + \frac{\varepsilon}{2}\|s\|^2 \\ &\leq f_j(x) + \theta_f^N(x) + \frac{(1-\sigma)\lambda_{\min}}{2}\|s\|^2 \\ &\leq f_j(x) + \sigma\theta_f^N(x), \end{aligned}$$

onde a segunda desigualdade decorre das definições de θ_f^N, ε e $\bar{\lambda}$, e a última decorre do item 3 do Lema 3.2. Ficou demonstrado o item 1.

Para provar o item 2, tome w como no Lema 3.1, para x (e s) como em (3.10) e defina a função auxiliar

$$g : \mathbb{R}^n \to \mathbb{R}, \qquad g(z) = \sum_{j=1}^m w_j f_j(z).$$

Observe que

$$\begin{aligned} \nabla g(z) &= \sum_{j=1}^m w_j \nabla f_j(z), \\ \nabla^2 g(z) &= \sum_{j=1}^m w_j \nabla^2 f_j(z), \\ s &= -\nabla^2 g(x)^{-1} \nabla g(x). \end{aligned}$$

Logo

$$\begin{aligned}\|\nabla^2 g(z) - \nabla^2 g(z')\| &= \left\| \sum_{j=1}^{m} w_j (\nabla^2 f_j(z) - \nabla^2 f_j(z')) \right\| \\ &\leq \sum_{j=1}^{m} w_j \left\| \nabla^2 f_j(z) - \nabla^2 f_j(z') \right\|\end{aligned}$$

e, em virtude de (3.9),

$$\|\nabla^2 g(z) - \nabla^2 g(z')\| < \varepsilon \qquad \forall z, z' \in B(\bar{x}, r_1).$$

Usando as inclusões $x, x+s \in B(\bar{x}, r_1)$, e aplicando o item 1 do Corolário 4.4, concluímos que

$$\|\nabla g(x+s)\| \leq \varepsilon \|s\|.$$

Como $\nabla g(x+s) = \sum_{j=1}^{m} w_j \nabla f_j(x+s)$ e $x+s \in B(\bar{x}, r_1)$, usando a desigualdade acima, a definição de $\bar{\lambda}$ e o Lema 3.10 temos

$$\|s_f(x+s)\| \leq \frac{\varepsilon \|s\|}{\bar{\lambda}}.$$

O item 2 segue desta desigualdade e das definições de ε e s. □

Provaremos agora que, para f de classe $\mathcal{C}^2$, na proximidade de um ponto crítico, a sequência $\{x^k\}$ converge a um ponto estacionário (não necessariamente o mesmo). De posse deste resultado, estabeleceremos facilmente a convergência superlinear.

Teorema 3.12. *Suponha que f é $\mathcal{C}^2$ e que $\bar{x}$ é um ponto crítico. Dado $r > 0$, existe $\delta \in (0, r)$ tal que se $x^{k_0} \in B(\bar{x}, \delta)$ para algum k_0, então*

1. *$t_k = 1$ para $k \geq k_0$;*

2. *a sequência $\{x^k\}$ converge a um ponto estacionário em $B(\bar{x}, r)$;*

3. *a convergência de $\{x^k\}$ é superlinear.*

Demonstração. Tome σ como especificado no Passo 1 do Algoritmo 3. Em virtude da Proposição 3.11 existe $\delta_1 \in (0, r)$ tal que

$$\begin{aligned} &f_j(x + s_f(x)) \leq f_j(x) + \sigma\theta_f^N(x) \\ &\|s_f(x + s_f(x))\| \leq \frac{\|s_f(x)\|}{2} \end{aligned} \qquad \forall x \in B(\bar{x}, \delta_1), \qquad (3.11)$$

onde a primeira desigualdade vale para $j = 1, \ldots, m$. Defina

$$\begin{aligned} \Omega(\tau) &= \left\{ x \in B(\bar{x}, \tau) \mid \|s_f(x)\| < \frac{\delta_1 - \tau}{2} \right\}, \ \tau \in (0, \delta_1), \\ \Omega &= \bigcup_{0<\tau<\delta_1} \Omega(\tau). \end{aligned}$$

Afirmamos que

$$x^k \in \Omega \implies t_k = 1, \ x^{k+1} \in \Omega. \qquad (3.12)$$

Com efeito, suponha que $x^k \in \Omega$. Como $\Omega \subset B(\bar{x}, \delta_1)$, segue-se da primeira desigualdade de (3.11) e da definição do Passo 3 que $t_k = 1$. Logo, $x^{k+1} = x^k + s^k$. Por hipótese, $x^k \in \Omega(\tau)$ para algum $\tau \in (0, \delta_1)$. Então,

$$\|\bar{x} - x^{k+1}\| \leq \|\bar{x} - x^k\| + \|s^k\| < \tau + \frac{\delta_1 - \tau}{2}.$$

Logo, para

$$\tau' = \tau + \frac{\delta_1 - \tau}{2} = \frac{\tau + \delta_1}{2} < \delta_1$$

temos $x^{k+1} \in B(\bar{x}, \tau')$ e $\tau' \in (0, \delta_1)$. Por sua vez, a segunda desigualdade de (3.11) nos dá a cota

$$\|s^{k+1}\| \leq \frac{\|s^k\|}{2} < \frac{\delta_1 - \tau}{4} = \frac{\delta_1 - \tau'}{2},$$

o que conclui a prova de (3.12).

Observe que Ω é aberto e $\bar{x} \in \Omega$; logo, existe $\delta > 0$ tal que $B(\bar{x}, \delta) \subset \Omega$. Se $x^{k_0} \in B(\bar{x}, \delta)$, então, por (3.12), $x^k \in \Omega \subset B(\bar{x}, \delta_1)$ e

$t_k = 1$ para $k \geq k_0$. Além disso, em virtude da segunda desigualdade em (3.11), temos

$$\sum_{i=k_0}^{\infty} \|x^{i+1} - x^i\| = \sum_{i=0}^{\infty} \|s^{k_0+i}\| \leq \|s^{k_0}\|(1 + 1/2 + 1/4 + \cdots)$$
$$= 2\|s^0\|.$$

Isto prova que $\{x^k\}$ é uma sequência de Cauchy. Logo,

$$\exists \lim_{k\to\infty} x^k = x^* \in \overline{\Omega} \subset B[\bar{x}, \delta_1] \subset B(\bar{x}, r).$$

O ponto x^* é, em particular, um ponto de acumulação da sequência $\{x^k\}$. Logo, pelo Teorema 3.9, x^* é um ponto estacionário.

O item 3 segue-se dos itens 1 e 2 deste Teorema, do item 2 da Proposição 3.11 e do item 2 do Lema 4.5. $\square$

Teorema 3.13. *Suponha que f é $\mathcal{C}^2$ e que os Hessianos $\nabla^2 f_j$, $j = 1, \ldots, m$, são localmente Lipschitz contínuos. Se $\{x^k\}$ converge, então a convergência é quadrática.*

Demonstração. Suponha que $\{x^k\}$ converge a x^*. Pelo Teorema 3.9, x^* é crítico, e aplicando o Teorema 3.12 com $\bar{x} = x^*$, concluímos que $t_k = 1$ para k suficientemente grande. Dado $r > 0$, existe L tal que

$$\|\nabla^2 f_j(x) - \nabla^2 f_j(x')\| \leq L\|x - x'\| \quad \forall x, x' \in B(x^*, r),\ j = 1, \ldots, m.$$

Para cada $k \in \mathbb{N}$, sejam $w^k \in \mathbb{R}^m$ como no Lema 3.1 para $x = x^k$ e g^k como definida na prova da Proposição 3.11 para $w = w^k$, isto é

$$g^k : \mathbb{R}^n \to \mathbb{R}, \quad g^k(z) = \sum_{j=1}^{m} w_j^k f_j(z).$$

Existe K tal que $x^k \in B(x^*, r)$ e $t_k = 1$ para $k \geq K$. Logo, como $x^{k+1} - x^k = s^k$ e $s^k = -\nabla^2 g^k(x^k)^{-1} \nabla g^k(x^k)$ é a direção de Newton para g^k em x^k, pelo item 2 do Corolário 4.4,

$$\|\nabla g^k(x^{k+1})\| \leq \frac{L}{2}\|x^{k+1} - x^k\|^2.$$

Segue-se desta desigualdade e do Lema 3.10 que

$$\|x^{k+2} - x^{k+1}\| = \|s^{k+1}\| \leq \frac{\|\nabla g^k(x^{k+1})\|}{\bar{\lambda}} \leq \frac{L}{2\bar{\lambda}}\|x^{k+1} - x^k\|^2,$$

onde $\bar{\lambda} > 0$ é o mínimo do menor dos autovalores de $\nabla^2 f_j$, $j = 1, \ldots, m$ na bola fechada $B[x^*, r]$. Por último, aplicamos o item 3 do Lema 4.5 e vemos que $\{x^k\}$ converge quadraticamente. □

Corolário 3.14. *Se f é $\mathcal{C}^2$ e x^0 pertence a um conjunto de nível limitado de f, então a sequência $\{x^k\}$ converge superlinearmente a um ponto estacionário. Se, adicionalmente, os Hessianos $\nabla^2 f_j$, $j = 1, \ldots, m$, são localmente Lipschitz contínuos, então a convergência é quadrática.*

Demonstração. Primeiro use o item 2 do Corolário 3.6 para concluir que a sequência $\{x^k\}$ tem um ponto de acumulação $\bar{x}$. Segue-se do Teorema 3.9 que tal ponto é crítico. Como existe uma subsequência $\{x^{k_i}\}$ que converge a $\bar{x}$, em virtude do Teorema 3.12, a sequência $\{x^k\}$ converge superlinearmente a algum ponto, que tem de ser $\bar{x}$.

A convergência quadrática decorre da primeira parte deste corolário e do Teorema 3.13

□

Capítulo 4

Resultados Auxiliares

Lema 4.1. *A função que leva cada matriz simétrica $A \in \mathbb{R}^{n\times n}$ em seu menor autovalor é continua.*

Demonstração. Lembremos que o *espectro* de $A \in \mathbb{R}^{n\times n}$ é o conjunto de seus autovalores, isto é

$$\sigma(A) = \{\lambda \mid \exists v \neq 0,\ Av = \lambda v\}.$$

Se $A \in \mathbb{R}^{n\times n}$ é simétrica, então $\sigma(A) \subset \mathbb{R}$ e podemos ordenar os seus autovalores como $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, contados com as respectivas multiplicidades, e existe uma base ortonormal $(v_i)_{i=1,\ldots,m}$ tal que $Av_i = \lambda_i v_i$. Se $v = \sum_{i=1}^{n} \alpha_i v_i$ tem norma 1, então $\sum_{i=1}^{m} \alpha_i^2 = 1$ e portanto,

$$v^\top Av = \sum_{i=1}^{m} \lambda_i \alpha_i^2 \geq \lambda_1 \sum_{i=1}^{m} \alpha_i^2 = \lambda_1.$$

Logo,

$$\min \sigma(A) = \min_{\|v\|=1} v^\top Av.$$

Se $B \in \mathbb{R}^{n\times n}$ é simétrica, então

$$\begin{aligned}\min \sigma(B) \leq v_1^\top Bv_1 &= v_1^\top Av_1 + v_1^\top (B-A)v_1 \\ &\leq \lambda_1 + \|B-A\|.\end{aligned}$$

Acabamos de provar que

$$\min \sigma(B) - \min \sigma(A) \leq \|B - A\|$$

Intercambiando os papéis de A e B obtemos

$$|\min \sigma(B) - \min \sigma(A)| \leq \|B - A\|,$$

o que conclui a prova. □

O próximo lema segue trivialmente do Teorema Fundamental do Cálculo.

Lema 4.2. *Se $\psi : \mathbb{R} \to \mathbb{R}$ é continuamente diferenciável, então*

$$\psi(1) = \psi(0) + \psi'(0) + \int_0^1 (\psi'(t) - \psi'(0))dt.$$

Se, adicionalmente, ψ é duas vezes continuamente diferenciável, então

$$\psi(1) = \psi(0) + \psi'(0) + \frac{1}{2}\psi''(0) + \int_0^1 \int_0^t (\psi''(\xi) - \psi''(0))\, d\xi\, dt.$$

A seguir estabelecemos cotas para os erros das aproximações lineares e quadráticas de uma função g e das aproximações lineares de ∇g.

Lema 4.3 (Erros de aproximacões locais)**.** *Sejam $U \subset \mathbb{R}^n$ um aberto convexo, $g : U \to \mathbb{R}$ de classe $\mathcal{C}^1$ e $x, x+v \in U$.*

1. *Se $\|\nabla g(z) - \nabla g(z')\| < \varepsilon$ para todo $z, z' \in U$, então*

$$|g(x+v) - \big(g(x) + \nabla g(x)^\top v\big)| \leq \varepsilon \|v\|.$$

2. *Se g é $\mathcal{C}^2$ e $\|\nabla^2 g(z) - \nabla^2 g(z')\| < \varepsilon$ para todo $z, z' \in U$, então*

$$\left| g(x+v) - \left(g(x) + \nabla g(x)^\top v + \frac{1}{2} v^\top \nabla^2 g(x) v \right) \right| \leq \frac{\varepsilon}{2} \|v\|^2,$$
$$\|\nabla g(x+v) - \big(\nabla g(x) + \nabla^2 g(x) v\big)\| \leq \varepsilon \|v\|.$$

3. Se g é $\mathcal{C}^2$ e $\nabla^2 g$ é L-Lipschitz continuo em U, então

$$\|\nabla g(x+v) - \big(\nabla g(x) + \nabla^2 g(x)v\big)\| \leq \frac{L}{2}\|v\|^2.$$

Demonstração. O item 1 segue da primeira parte do Lema 4.2 com

$$\psi(t) = g(x+tv).$$

A primeira desigualdade do item 2 segue da segunda parte do Lema 4.2 com ψ como acima definida.

Para provar a segunda desigualdade do item 2 e o item 3, tome $u \in \mathbb{R}^r$ um vetor unitário qualquer e defina

$$\psi(t) = u^\top(\nabla g(x+tv) - \nabla g(x) - t\nabla^2 g(x)v).$$

Como $\psi'(t) = u^\top(\nabla^2 g(x+tv) - \nabla^2 g(x))v$,

$$\begin{aligned}
\psi(1) &= u^\top(\nabla g(x+v) - \nabla g(x) - \nabla^2 g(x)v) \\
&= \psi(0) + \psi'(0) + \int_0^1 (\psi'(t) - \psi'(0))\,dt = \int_0^1 \psi'(t)\,dt.
\end{aligned}$$

Logo, no item 2 e no item 3 temos, respectivamente,

$$\begin{aligned}
|u^\top(\nabla g(x+v) - \nabla g(x) - \nabla^2 g(x)v)| &\leq \varepsilon\|v\|, \\
|u^\top(\nabla g(x+v) - \nabla g(x) - \nabla^2 g(x)v)| &\leq \frac{L}{2}\|v\|^2.
\end{aligned}$$

A conclusões seguem-se do fato de que estas desigualdades valem para qualquer vetor u unitário. □

As cotas de erro acima enunciadas nos permitem limitar o gradiente de uma função $\mathcal{C}^2$ após um passo de Newton para sua minimização.

Corolário 4.4 (Resíduo após o passo de Newton)**.** *Seja $U \subset \mathbb{R}^n$ um aberto convexo, $g : U \to \mathbb{R}$ de classe $\mathcal{C}^2$. Se $\nabla^2 g(x)$ é não singular para algum $x \in U$ e que*

$$s = -\nabla^2 g(x)^{-1}\nabla g(x), \qquad x + s \in U,$$

então

1. *se* $\|\nabla^2 g(z) - \nabla^2 g(z')\| \leq \varepsilon$ *para todo* $z, z' \in U$ *então*
$$\|\nabla g(x+s)\| \leq \varepsilon \|s\|;$$

2. *se* $\nabla^2 g$ *é L-Lipschitz continuo em* U, *então*
$$\|\nabla g(x+s)\| \leq \frac{L}{2}\|s\|^2.$$

Demonstração. Em virtude da definição de s,

$$\nabla g(x+s) = \nabla g(x+s) - (\nabla g(x) + \nabla^2 g(x) s).$$

As conclusões seguem-se da segunda desigualdade do item 2 e do item 3 do Lema 4.3. □

Lembremos que uma sequência $\{z_k\}$ converge a z^* *superlinearmente* se ela converge a tal ponto e se para todo $\kappa > 0$,

$$\|z^{k+1} - z^*\| \leq \kappa \|z^k - z^*\|$$

para k suficientemente grande. A sequência $\{z_k\}$ converge a z^* *quadraticamente* se ela converge a tal ponto e se existe $C > 0$ tal que

$$\|z^{k+1} - z^*\| \leq C\|z^k - z^*\|^2$$

para k suficientemente grande.

Lema 4.5. *Suponha que* $\{z^k\}$ *converge a* z^* *e seja*

$$\rho_k = \|z^{k+1} - z^k\|, \qquad k = 0, 1, \ldots$$

1. *Se existem* $\kappa \in (0,1)$ *e* K *tais que*
$$\rho_{k+1} \leq \kappa \rho_k, \qquad k \geq K$$
então $\|z^* - z^k\| \leq \dfrac{1}{1-\kappa}\rho_k$ *para* $k \geq K$.

2. *Se para todo* $\kappa \in (0,1)$ *existe* K *tal que*
$$\rho_{k+1} \leq \kappa \rho_k \qquad \forall k \geq K$$
então $\{z^k\}$ *converge superlinearmente a* z^*.

3. Se existem C e K tais que

$$\rho_{k+1} \leq C\rho_k^2 \qquad \forall k \geq K$$

então $\{z^k\}$ converge quadraticamente a z^*.

Demonstração. Todos os itens valem trivialmente se existir uma subsequência $\{\rho_{k_i}\}$ tal que $\rho_{k_i} = 0$ para todo i, pois isto implica (nos três itens) que $\rho_k = 0$ para k suficientemente grande. Suporemos então que $\rho_k \neq 0$ para k suficientemente grande.

Para provar o item 1, tome $\ell > k \geq K$, e use a desigualdade triangular para concluir que

$$\|z^\ell - z^k\| \leq \sum_{i=k}^{\ell-1} \rho_i \leq \rho_k \sum_{i=0}^{\ell-k-1} \kappa^i \leq \rho_k \frac{1}{1-\kappa}.$$

Para terminar a prova, tome o limite para $\ell \to \infty$ na desigualdade acima.

Para provar o item 2, tome $\kappa \in (0, 1/2)$ e seja K como na premissa deste item. Segue-se de nossas hipóteses que $\rho_k \neq 0$ para $k \geq K$. Se $k \geq K$ então $\rho_{k+1} \leq \kappa\rho_k$ e pelo item 1

$$\|z^* - z^{k+1}\| \leq \rho_{k+1} \frac{1}{1-\kappa} \leq \rho_k \frac{\kappa}{1-\kappa}$$

Usando este resultado e a desigualdade triangular, temos

$$\|z^* - z^k\| \geq \|z^{k+1} - z^k\| - \|z^* - z^{k+1}\| \geq \left(1 - \frac{\kappa}{1-\kappa}\right)\rho_k$$

Combinando as duas últimas desigualdade concluímos que

$$\frac{\|z^* - z^{k+1}\|}{\|z^* - z^k\|} \leq \frac{\kappa}{1-2\kappa}.$$

Como κ é um número arbitrário em $(0, 1/2)$, a sequência $\{z^k\}$ converge superlinearmente a z^*.

Para provar o item 3, primeiro observe que, em vista da convergência de $\{z^k\}$, vale a premissa do item 2. Logo $\{z^k\}$ converge superlinearmente a z^* e existe $K' \geq K$ tal que

$$\|z^* - z^{k+1}\| \leq \frac{1}{2}\|z^* - z^k\|, \quad \rho_{k+1} \leq C\rho_k^2 \qquad \forall k \geq K'.$$

Usando a primeira desigualdade acima e a desigualdade triangular concluímos que

$$\begin{aligned}
\rho_k + \frac{1}{2}\|z^* - z^k\| &\geq \|z^{k+1} - z^k\| + \|z^* - z^{k+1}\| \\
&\geq \|z^* - z^k\| \\
&\geq \|z^{k+1} - z^k\| - \|z^* - z^{k+1}\| \\
&\geq \rho_k - \frac{1}{2}\|z^* - z^k\|
\end{aligned}$$

Logo

$$2\rho_k \geq \|z^* - z^k\| \geq \frac{2}{3}\rho_k, \qquad k \geq K'.$$

Portanto, para $k \geq K'$

$$\|z^* - z^{k+1}\| \leq 2\rho_{k+1} \leq 2C\rho_k^2 \leq 2C\left(\frac{3}{2}\|z^* - z^k\|\right)^2$$

o que prova a convergência quadrática. $\square$

Bibliografia

[1] J. Branke, K. Deb, K. Miettinen, and R. Slowinski, editors. *Practical Approaches to Multi-Objective Optimization*. Dagstuhl Seminar, 2006. Seminar 06501. Forthcoming.

[2] J. Branke, K. Deb, K. Miettinen, and R. E. Steuer, editors. *Practical Approaches to Multi-Objective Optimization*. Dagstuhl Seminar, 2004. Seminar 04461. Electronically available at `http://drops.dagstuhl.de/portals/index.php?semnr=04461`.

[3] E. Carrizosa and J. B. G. Frenk. Dominating sets for convex functions with some applications. *Journal of Optimization Theory and Applications*, 96(2):281–295, 1998.

[4] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996.

[5] G. Eichfelder. An adaptive scalarization method in multiobjective optimization. *SIAM Journal on Optimization*, 19(4):1694–1718, 2009.

[6] H. Eschenauer, J. Koski, and A. Osyczka. *Multicriteria Design Optimization: Procedures and Applications*. Springer-Verlag, Berlin, 1990.

[7] J. Fliege. OLAF – A general modeling system to evaluate and optimize the location of an air polluting facility. *OR Spektrum*, 23:117–136, 2001.

[8] J. Fliege. An efficient interior-point method for convex multicriteria optimization problems. *Mathematics of Operations Research*, 31(4):825–845, 2006.

[9] J. Fliege, L. M. Graña Drummond, and B. F. Svaiter. Newton's method for multiobjective optimization. *SIAM Journal on Optimization*, 20(2):602–626, 2009.

[10] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.

[11] Ellen H. Fukuda and L. M. Graña Drummond. On the convergence of the projected gradient method for vector optimization. *Optimization*, 60(8-9):1009–1021, 2011.

[12] Ellen H. Fukuda and L. M. Graña Drummond. Inexact projected gradient method for vector optimization. *Computational Optimization and Applications*, 54(3):473–493, 2013.

[13] Ellen H. Fukuda and Luis Mauricio Graña Drummond. A survey on multiobjective descent methods. *Pesquisa Operacional*, 34(3):585–620, 2014.

[14] S. Gass and T. Saaty. The computational algorithm for the parametric objective function. *Naval Research Logistics Quarterly*, 2(1-2):39–45, 1955.

[15] A. M. Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications*, 22:618–630, 1968.

[16] L. M. Graña Drummond and A. N. Iusem. A projected gradient method for vector optimization problems. *Computational Optimization and Applications*, 28(1):5–29, 2004.

[17] L. M. Graña Drummond, N. Maculan, and B. F. Svaiter. On the choice of parameters for the weighting method in vector optimization. *Mathematical Programming*, 111(1-2), 2008.

[18] L. M. Graña Drummond, F. M. P. Raupp, and B. F. Svaiter. A quadratically convergent Newton method for vector optimization. *Optimization*, 2011. Accepted for publication.

[19] L. M. Graña Drummond and B. F. Svaiter. A steepest descent method for vector optimization. *Journal of Computational and Applied Mathematics*, 175(2):395–414, 2005.

[20] J. Jahn. Scalarization in vector optimization. *Mathematical Programming*, 29:203–218, 1984.

[21] J. Jahn. *Mathematical Vector Optimization in Partially Ordered Linear Spaces*. Verlag Peter D. Lang, Frankfurt, 1986.

[22] J. Jahn. *Vector Optimization – Theory, Applications and Extensions*. Springer, Erlangen, 2003.

[23] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10:263–282, 2002.

[24] T. M. Leschine, H. Wallenius, and W.A. Verdini. Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions. *European Journal of Operational Research*, 56:278–289, 1992.

[25] D. T. Luc. Theory of vector optimization. In *Lecture Notes in Economics and Mathematical Systems*, 319, Berlin, 1989. Springer.

[26] K. Miettinen and M. M. Mäkelä. Interactive bundle-based method for nondifferentiable multiobjective optimization: Nimbus. *Optimization*, 34:231–246, 1995.

[27] K. M. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Norwell, 1999.

[28] S. Mostaghim, J. Branke, and H. Schmeck. Multi-objective particle swarm optimization on computer grids. Technical report. 502, Institut AIFB, University of Karlsruhe (TH), December 2006.

[29] Y. Sawaragi, H. Nakayama, and T. Tanino. *Theory of Multiobjective Optimization.* Academic Press, Orlando, 1985.

[30] L. Zadeh. Optimality and non-scalar-valued performance criteria. *IEEE Transactions on Automatic Control*, 8(1):59–60, 1963.

[31] E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation*, 8:173–195, 2000.