



CONVERGÊNCIA DE MÉTODOS DE DESCIDA PARA FUNÇÕES NÃO
CONVEXAS COM APLICAÇÕES À TEORIA DE COMPORTAMENTO

João Carlos de Oliveira Souza

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

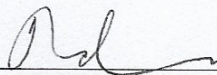
Rio de Janeiro
Novembro de 2016

CONVERGÊNCIA DE MÉTODOS DE DESCIDA PARA FUNÇÕES NÃO
CONVEXAS COM APLICAÇÕES À TEORIA DE COMPORTAMENTO

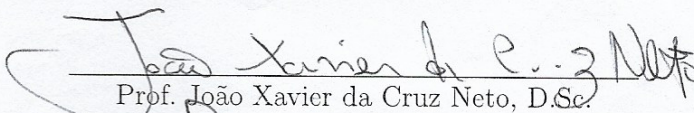
João Carlos de Oliveira Souza

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

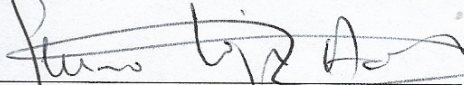
Examinada por:



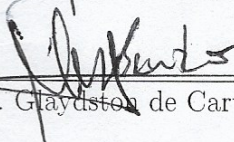
Prof. Paulo Roberto Oliveira, D.Sc.



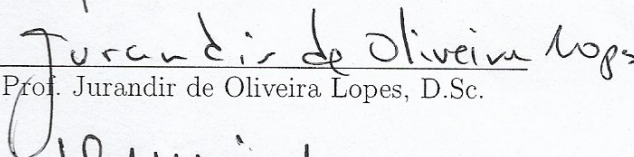
Prof. João Xavier da Cruz Neto, D.Sc.



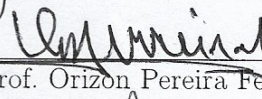
Prof. Genaro López Acedo, Dr.



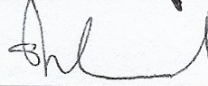
Prof. Gláuston de Carvalho Bento, D.Sc.



Prof. Jurandir de Oliveira Lopes, D.Sc.



Prof. Orizon Pereira Ferreira, D.Sc.



Prof. Susana Scheimberg de Makler, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
NOVEMBRO DE 2016

Souza, João Carlos de Oliveira

Convergência de métodos de descida para funções não convexas com aplicações à teoria de comportamento/João Carlos de Oliveira Souza. – Rio de Janeiro: UFRJ/COPPE, 2016.

IX, 112 p. 29,7cm.

Orientadores: Paulo Roberto Oliveira

João Xavier da Cruz Neto

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2016.

Referências Bibliográficas: p. 101 – 112.

1. método do ponto proximal. 2. método de máxima descida. 3. diferença de funções convexas. 4. propriedade de Kurdyka-Lojasiewicz. 5. variedades de Hadamard. 6. variational rationality. 7. teoria de comportamento. I. Oliveira, Paulo Roberto *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

À minha esposa Lara.
“–But this is touching, Severus,
said Dumbledore seriously.
–Have you grown to care for the
boy, after all?
–For him? shouted Snape.
–Expecto Patronum!
From the tip of his wand burst
the silver doe. She landed on the
office floor, bounded once across
the office, and soared out of the
window. Dumbledore watched her
fly away, and as her silvery glow
faded he turned back to Snape,
and his eyes were full of tears.
–After all this time?
–Always, said Snape.”
J.K. Rowling, *Harry Potter and
the Deathly Hallows*, 2007.

Agradecimentos

Agradeço primeiramente a Deus, pela vida e saúde.

Aos familiares e amigos que sempre estiveram ao meu lado, apoiando-me nos meus sonhos. Agradeço, de modo especial, à minha mãe, Domingas, ao meu pai, Carlos Henrique, e aos meus irmãos, José Rodolfo e Maria Luiza.

À minha esposa por embarcar comigo nessa aventura e pela paciência nesses últimos anos. A ela e aos meus pais dedico uma frase atribuída ao poeta Rabindranath Tagore que sintetiza meu agradecimento: “a noite abre as flores em segredo e deixa que o dia receba os agradecimentos”.

Agradeço aos meus orientadores Dr. Paulo Roberto Oliveira e Dr. João Xavier da Cruz Neto pelos valiosos ensinamentos e pela ótima parceria que não é somente científica.

Agradeço aos membros da banca examinadora Dr. Glaydston de Carvalho Bento, Dr. Jurandir de Oliveira Lopes, Dr. Orizon Pereira Ferreira e Dra. Susana Scheimberg de Makler pelo tempo dedicado à leitura deste trabalho e pelas valiosas sugestões que ajudaram a melhorar o texto. Em especial, agradeço Dr. Genaro López Acedo, representando todos os professores do grupo de Análise Matemática da Universidade de Sevilha e aos funcionários do IMUS, pela acolhida e toda a atenção dada no maravilhoso ano que tive em Sevilha. Agradeço também Dr. Antoine Soubeyran pela parceria e pela recepção na minha visita à Universidade de Aix-Marseille - Escola de Economia de Aix-Marseille.

Agradeço à Universidade Federal do Rio de Janeiro - UFRJ, em particular aos professores e funcionários do Programa de Engenharia de Sistemas e Computação - PESC, mas principalmente a Gutierrez pelo apoio prestado sempre de forma cortês e eficiente. Agradeço também aos colegas de cursos pelo convívio e boas conversas, em especial o Pedro Jorge (amigo de graduação, mestrado e doutorado) pela imensa ajuda no período que não estive no Rio de Janeiro.

Agradeço à Universidade Federal do Piauí - UFPI, em particular ao CEAD e ao Departamento de Matemática, pela liberação e apoio para cursar o doutorado.

Finalmente, agradeço ao CNPq, em especial ao programa Ciências sem Fronteiras, pelo apoio financeiro (processo 203360/2014-1) no Doutorado Sanduíche na Universidade de Sevilha, Espanha.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

CONVERGÊNCIA DE MÉTODOS DE DESCIDA PARA FUNÇÕES NÃO CONVEXAS COM APLICAÇÕES À TEORIA DE COMPORTAMENTO

João Carlos de Oliveira Souza

Novembro/2016

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Programa: Engenharia de Sistemas e Computação

Apresentamos extensões para variedades de Hadamard do método do ponto proximal para diferença de funções convexas e do método de máxima descida para funções continuamente diferenciáveis que satisfazem a propriedade de Kurdyka-Łojasiewicz. Usando o método de máxima descida propomos um algoritmo para calcular o centro de massa Riemanniano de um conjunto de dados em variedades de Hadamard. Também apresentamos um método linearizado proximal generalizado para diferença de funções convexas que usa uma quase distância como regularização. Usando a propriedade de Kurdyka-Łojasiewicz provamos a convergência global da sequência. Como aplicação, usando a abordagem “variational rationality” (VR) apresentamos uma nova versão, dinâmica, do problema de produção ótimo de uma companhia. Finalmente, propomos uma nova abordagem de convergência do método do ponto proximal em otimização multiobjetivo que amplia a aplicação do método para funções vetoriais localmente Lipschitz. Como aplicação estudamos o famoso problema de compromisso usando a abordagem (VR) de comportamento humano.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

CONVERGENCE OF DESCENT METHODS FOR NONCONVEX FUNCTIONS
WITH APPLICATIONS TO BEHAVIORAL SCIENCE

João Carlos de Oliveira Souza

November/2016

Advisors: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Department: Systems Engineering and Computer Science

We present extensions to Hadamard manifolds of the proximal point method for difference of convex functions and the steepest descent method for continuously differentiable functions which satisfy the Kurdyka-Lojasiewicz property. Using the steepest descent method, we propose an algorithm for computing the Riemannian center of mass of a set of data points on Hadamard manifolds. We also present a generalized proximal linearized method for difference of convex functions which uses a quasi distance as regularization. Using the Kurdyka-Lojasiewicz property, we prove the global convergence of the sequence. As an application, by using the variational rationality (VR) approach, we give a new dynamic version of the optimal size of firm problem. Finally, we propose a new approach for convergence of the proximal point method in multiobjective optimization extending its application for vector-valued locally Lipschitz functions. As an application, we study the famous compromise problem using the (VR) approach of human behavioral.

Sumário

1	Introdução	1
2	Preliminares	4
2.1	Conceitos e resultados de otimização	4
2.2	Diferença de funções convexas (DC)	9
2.3	Variedades Riemannianas	12
2.3.1	Variedades diferenciáveis	12
2.3.2	Métrica Riemanniana	12
2.3.3	Conexão Riemanniana	13
2.3.4	Geodésicas e aplicação exponencial	14
2.3.5	Curvatura e fórmulas de variação	15
2.3.6	Variedades de Hadamard	16
2.4	Otimização em variedades Riemannianas	17
2.4.1	Convexidade	17
2.4.2	Subdiferencial	18
3	Métodos de descida em variedades de Riemann	21
3.1	MPP para funções DC em variedades de Hadamard	22
3.1.1	Versão exata	24
3.1.2	Versões inexatas	31
3.1.3	Aplicação em problemas de maximização com restrições	39
3.2	MMD em variedades de Hadamard	39
3.2.1	Análise de convergência	42
3.2.2	Aplicação: centro de massa Riemanniano	47
4	MPP generalizado para funções DC	53
4.1	Quase distância	53
4.2	Algoritmo generalizado	54
4.3	Análise de convergência	56
4.4	Convergência global	59
4.4.1	Experimento numérico	60

4.5	Aplicação 1: convergência linear do MPP - caso convexo	61
4.6	Aplicação 2: problema de produção - caso unidimensional	63
4.6.1	Um modelo de companhia com retornos crescentes a curto prazo	65
4.6.2	A abordagem variational rationality: simples formulação . . .	67
4.6.3	Algoritmo proximal como transições satisfatórias	70
4.6.4	Limites	73
5	MPP para otimização multiobjetivo	75
5.1	MPP em otimização multiobjetivo	76
5.2	MPP para funções vetoriais DC	78
5.2.1	O algoritmo	79
5.2.2	Análise de convergência	83
5.3	MPP para funções vetoriais localmente Lipschitz	84
5.3.1	O algoritmo	85
5.3.2	Análise de convergência	88
5.4	Aplicação em grupos dinâmicos	92
5.4.1	Problema de produção: dinâmica de melhoria cooperativa . .	92
5.4.2	Problema de compromisso	95
6	Conclusão	98
	Referências Bibliográficas	101

Capítulo 1

Introdução

Existe um amplo campo de estudo em otimização que consiste em estender métodos já existentes para contextos mais gerais. Neste trabalho, estendemos dois métodos clássicos de otimização para três cenários diferentes e, além disso, apresentamos algumas aplicações desses métodos em diferentes problemas. Os algoritmos estudados serão o método de máxima descida (MMD) (ou método do gradiente) e o método do ponto proximal (MPP). Tais métodos serão estendidos para os seguintes contextos: otimização em variedade de Riemann, algoritmos com regularizações “*like-distance*” e otimização multiobjetivo.

As justificativas para tais extensões são tanto teóricas como do ponto de vista de aplicações. Por exemplo, problemas de otimização restritos podem ser considerados como irrestritos do ponto de vista da geometria Riemanniana e problemas não convexos no sentido clássico podem tornar-se convexos através da introdução de uma métrica Riemanniana apropriada; veja [45]. Métodos proximais que usam uma quase distância (veja definição no Capítulo 4) como regularização ao invés da distância Euclideana são mais apropriados para aplicações, por exemplo em Teoria de Comportamento usando a abordagem (VR), onde o custo para estar apto a mudar de uma posição atual para outra posição e o custo para estar apto a permanecer na posição atual não necessariamente são simétricos e iguais a zero, respectivamente; veja por exemplo [26, 27, 101]. Otimização multiobjetivo tem como campo de aplicações, por exemplo a engenharia (especialmente otimização “*truss*”, “*design*”, exploração espacial), estatística, ciência de gestão, etc.; veja [59] e suas referências.

Uma das extensões consiste em propor versões do MMD e MPP em variedades de Hadamard para diferentes classes de funções não convexas. Resultados de convergência global para o MMD em variedades de Riemann foram estudados em Cruz Neto et al.[43], em particular, para o caso em que a função objetivo é convexa (continuamente diferenciável) e por Papa Quiroz et al.[107] para o caso quase convexo, ambos no contexto das variedades de Riemann com curvatura não-negativa. Sem

hipóteses na curvatura da variedade, apenas resultados parciais de convergência foram obtidos em [43]. Estudaremos a convergência global do MMD em variedades de Hadamard (que possui curvatura não positiva) para funções, não necessariamente convexas, que satisfazem a propriedade de Kurdyka-Łojasiewicz. Ainda no cenário das variedades de Hadamard iremos propor uma extensão de um MPP para uma classe de funções que contém as funções convexas. Tais funções, conhecidas como funções DC, são as funções que podem ser escritas como a diferença de duas funções convexas. Nesse caso, o MPP proposto em Ferreira e Oliveira [54] se torna um caso particular.

Outro tipo de extensão proposta diz respeito somente ao MPP. O método do ponto proximal para resolver um problema de minimização foi introduzido na literatura de otimização por Martinet [98] e popularizado por Rockafellar [112]. A ideia do método é, a partir de um ponto inicial dado, gerar uma sequência de pontos que são mínimos da função objetivo acrescida de uma regularização (quadrado da função distância). Diversos autores tem proposto variações desse método substituindo a função distância na regularização por aplicações que não satisfazem todos os axiomas da função distância mas preservam algumas de suas boas propriedades tais como continuidade, coercividade, etc.. Essas variações são conhecidas como MPP generalizados e tem diversas aplicações em diferentes áreas; veja por exemplo [26, 27, 35, 41, 51, 101, 106]. Iremos propor uma versão do MPP para funções DC com uma regularização via uma quase distância. O algoritmo proposto generaliza o método analisado por Sun et al. [126].

O terceiro tipo de extensão diz respeito a propor versões de métodos de otimização escalar, ou seja, quando a imagem da função objetivo assume valores reais, para funções que assumem valores vetoriais. Esse campo de estudo é conhecido como otimização vetorial ou, em um caso particular, otimização multiobjetivo. O estudo do MPP para otimização vetorial ou multiobjetivo não tem obtido muitos avanços além de funções quase convexas nos últimos anos. Iremos propor uma nova abordagem para convergência do MPP para funções vetoriais que recupera, no contexto multiobjetivo de dimensão finita, os resultados existentes para os casos convexas (Bonnell et al. [33]) e quase convexas (Apolinário et al. [7] e Bento et al. [21]), e além disso, nos permite estender a aplicação do MPP para funções vetoriais localmente Lipschitz.

Finalmente, estudaremos diferentes aplicações do método do ponto proximal para resolver problemas de teoria de comportamento. A saber, iremos considerar, usando uma abordagem dinâmica, o problema estático de tamanho de produção em duas versões: na primeira consideraremos o caso unidimensional e na segunda abordaremos o mesmo problema sob a perspectiva de grupos dinâmicos cooperativos. Ainda como aplicação do MPP iremos estudar o famoso problema de compromisso,

onde um grupo de agentes tenta minimizar a distância entre suas posições atuais e o ponto ideal para o grupo. O campo de aplicações desses problemas são vários, tais como Teoria de Localização, Teoria de Utilidade, Teoria de Consumidor, Teoria de Tomada de Decisão, Psicologia e Ciências de Comportamento; veja [62, 63, 85, 86, 104, 105]. Uma aplicação do MMD em variedades de Hadamard também é apresentada. Usaremos o MMD para calcular o centro de massa Riemanniano de um conjunto de dados. Esse problema tem diversas aplicação tais como análise estatística de formas, imagens, redes sensoriais, visão computacional e várias outras aplicações que usam análise de dados; veja [4, 5, 12] e suas referências.

A organização deste trabalho é da seguinte forma. No capítulo 2 apresentamos alguns conceitos e resultados preliminares de otimização escalar e vetorial, teoria de funções DC, notação e principais ferramentas de otimização em variedades de Riemann. No capítulo 3 estudamos o MMD e o MPP em variedades de Hadamard para duas classes diferentes de funções não convexas. Na primeira parte, uma primeira extensão do MPP para funções DC em variedades de Hadamard é proposto. Um simples experimento numérico do algoritmo, em uma variedade de curvatura constante é proposto, bem como uma aplicação do método para resolver um problema de maximização com restrições convexas. Na segunda parte, a convergência global do MMD para funções não convexas que satisfazem a propriedade de Kurdyka-Lojasiewicz é proposta para variedades de Hadamard. Como aplicação do método propomos um algoritmo para encontrar o centro de massa Riemanniano de um conjunto de dados. No capítulo 4, um MPP generalizado usando uma quase distância como regularização é proposto para encontrar um ponto crítico de uma função DC. Uma aplicação de MPP para funções DC é dada para produzir uma nova demonstração da convergência linear do MPP clássico proposto em [112] para funções fortemente convexas. Ainda usando o MPP generalizado apresentamos uma nova versão (dinâmica) do problema de produção ótimo de uma companhia usando a recente abordagem “*Variational Rationality*” (VR) de Soubeyran [117–119]. No capítulo 5 propomos uma nova abordagem para convergência do MPP para otimização multiobjetivo que nos permite estudar tal método no contexto das funções DC multiobjetivo e, mais geral, funções vetoriais localmente Lipschitz. Como aplicação discutimos uma versão dinâmica do famoso problema de compromisso usando a abordagem (VR) de comportamento humano. Finalmente, a conclusão e os trabalhos futuros são apresentados no capítulo 6.

Capítulo 2

Preliminares

Nesse capítulo apresentaremos alguns conceitos e as principais propriedades necessárias para compreender os resultados obtidos nos próximos capítulos, bem como referências para maiores informações e demonstrações dos resultados exibidos.

2.1 Conceitos e resultados de otimização

Funções escalares

Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ denotaremos o *domínio* de f como $\text{dom}(f) = \{x \in \mathbb{R}^n : f(x) \neq +\infty\}$. Dizemos que f é *própria* se $\text{dom}(f) \neq \emptyset$. A função f é dita *semicontínua inferiormente* (resp. *superiormente*) em um ponto $x \in \mathbb{R}^n$, se para qualquer sequência $\{x^k\}$ em \mathbb{R}^n , tem-se

$$f(x) \leq \liminf_{k \rightarrow +\infty} f(x^k) \quad (\text{resp.} \quad \limsup_{k \rightarrow +\infty} f(x^k) \leq f(x)).$$

Dizemos que f é uma função *convexa*, se para todo $x, y \in \mathbb{R}^n$ e $\lambda \in [0, 1]$, temos

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

Quando a desigualdade acima é estrita dizemos que f é *estritamente convexa*. Por outro lado, se existe $m > 0$ tal que

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - m\lambda(1 - \lambda)\|x - y\|^2,$$

para todo $x, y \in \mathbb{R}^n$, então dizemos que f é *fortemente convexa* com constante m . Quando

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$$

dizemos que f é *quase convexa*.

A seguir apresentamos algumas definições de subdiferencial de uma função em um ponto:

1. O subdiferencial de Fenchel-Moreau de f em x , denotado por $\partial f(x)$, é definido da seguinte forma:

$$\partial f(x) = \begin{cases} \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x \rangle, \quad \forall y \in \mathbb{R}^n\}, & \text{se } x \in \text{dom}(f); \\ \emptyset, & \text{se } x \notin \text{dom}(f). \end{cases}$$

2. O subdiferencial de Fréchet de f em x , denotado por $\partial_F f(x)$, é definido da seguinte forma:

$$\partial_F f(x) = \begin{cases} \{v \in \mathbb{R}^n : \liminf_{\substack{y \rightarrow x \\ y \neq x}} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|x - y\|} \geq 0\}, & \text{se } x \in \text{dom}(f); \\ \emptyset, & \text{se } x \notin \text{dom}(f). \end{cases}$$

3. O subdiferencial de Mordukhovich (ou limite) de f em x , denotado por $\partial_L f(x)$, é definido da seguinte forma:

$$\partial_L f(x) = \begin{cases} \{v \in \mathbb{R}^n : \exists x^k \rightarrow x, f(x^k) \rightarrow f(x), v^k \in \partial_F f(x^k) \rightarrow v\}, & \text{se } x \in \text{dom}(f); \\ \emptyset, & \text{se } x \notin \text{dom}(f). \end{cases}$$

Se f é uma função própria, semicontínua inferiormente e convexa em $x \in \text{dom}(f)$, então $\partial f(x) = \partial_F f(x) \neq \emptyset$.

Dado um conjunto convexo $C \subset \mathbb{R}^n$ e $x \in C$ denotamos o *cone normal* de C em x por $N_C(x) = \{w \in \mathbb{R}^n : \langle w, y - x \rangle \leq 0, \quad \forall y \in C\}$. Se $x \notin C$ dizemos que $N_C(x) = \emptyset$.

A *função indicadora* de um conjunto $\Omega \subset \mathbb{R}^n$ denotada por $\delta_\Omega : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é definida como

$$\delta_\Omega(x) = \begin{cases} 0, & \text{se } x \in \Omega \\ +\infty, & \text{se } x \notin \Omega. \end{cases}$$

Assim, temos que, quando Ω é um conjunto convexo e fechado, temos que $\delta_\Omega(\cdot)$ é semicontínua inferiormente e convexa. Além disso, $\partial \delta_\Omega(x) = N_\Omega(x)$.

Dado um conjunto $C \subset \mathbb{R}^n$ não vazio consideramos a *função distância ao conjunto* C $d : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $d_C(x) := \inf\{\|x - c\| : c \in C\}$.

Dizemos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função *Lipschitz contínua* com constante L , se existe $L > 0$ tal que para todo $x, y \in \mathbb{R}^n$, vale

$$|f(x) - f(y)| \leq L\|x - y\|.$$

Quando a propriedade acima é válida em uma vizinhança U de um ponto $\hat{x} \in \mathbb{R}^n$

dizemos que f é *localmente Lipschitz* em \hat{x} . Quando f é localmente Lipschitz em todos os pontos dizemos apenas que f é localmente Lipschitz.

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função localmente Lipschitz em $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ uma direção. A *derivada direcional de Clarke* de f em x na direção de d , denotada por $f^\circ(x, d)$, é definida da seguinte forma

$$f^\circ(x, d) := \limsup_{\substack{y \rightarrow x \\ t \downarrow 0}} \frac{f(y + td) - f(y)}{t}.$$

Com isso, definimos o *subdiferencial de Clarke* de f em x , denotado por $\partial_C f(x)$, da seguinte forma

$$\partial_C f(x) := \{w \in \mathbb{R}^n : \langle w, d \rangle \leq f^\circ(x, d), \quad \forall d \in \mathbb{R}^n\}.$$

A seguir, apresentamos alguns resultados clássicos de otimização.

Proposição 2.1.1 *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ e x um ponto tal que f é finito. Então os subdiferenciais $\partial_L f(x)$ e $\partial_F f(x)$ são fechados, com $\partial_F f(x)$ convexo e $\partial_F f(x) \subset \partial_L f(x)$.*

Demonstração: Veja [113, Theorem 8.6]. ■

Proposição 2.1.2 *Se uma função própria $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ tem um ponto de mínimo local em x , então $0 \in \partial_F f(x)$ e $0 \in \partial_L f(x)$. Se f é convexa, essas condições são necessárias e suficientes para um ponto ser mínimo global. Além disso, se $f = f_1 + f_2$ com f_2 continuamente diferenciável, a condição $0 \in \partial_F f(x)$ assume a forma $-\nabla f_2(x) \in \partial_L f_1(x)$.*

Demonstração: Veja [113, Theorem 10.1]. ■

Proposição 2.1.3 *Se f_1 é localmente Lipschitz em x , f_2 é semicontínua inferiormente com $f_2(x)$ finito, então*

$$\partial_L(f_1 + f_2)(x) \subset \partial_L f_1(x) + \partial_L f_2(x).$$

Demonstração: Veja [113, página 431]. ■

Proposição 2.1.4 *Sejam $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R}_+$ Lipschitz contínuas em x , então*

$$\partial_L(f_1 \cdot f_2)(x) \subset f_2(x)\partial_L f_1(x) + f_1(x)\partial_L f_2(x).$$

Demonstração: Veja [100, Theorem 7.1]. ■

Proposição 2.1.5 *Sejam $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ semicontínuas inferiormente, com uma delas Lipschitz contínua em $x \in \text{dom}(f_1) \cap \text{dom}(f_2)$. Então, para todo $\delta > 0$ e $\gamma > 0$, temos*

$$\partial_F(f_1 + f_2)(x) \subset A + \gamma \overline{B}(0, 1),$$

onde $A = \cup \{\partial_F f_1(x_1) + \partial_F f_2(x_2) : x_i \in B(x, \delta), |f_i(x_i) - f_i(x)| \leq \delta, i = 1, 2\}$.

Demonstração: Veja [100, Proposition 2.7]. ■

Dizemos que uma aplicação ponto-conjunto $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ é localmente limitada em um ponto $x \in \mathbb{R}^n$, se para alguma vizinhança $V \in \mathcal{N}(x)$ o conjunto $S(V) \subset \mathbb{R}^m$ é limitada, onde $\mathcal{N}(x)$ é o conjunto de todas as vizinhanças de x . Uma aplicação é dita localmente limitada (em \mathbb{R}^n) se essa condição for verificada para todo $x \in \mathbb{R}^n$.

Proposição 2.1.6 *Uma aplicação $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ é localmente limitada se, e somente se, $S(B)$ é limitado para todo conjunto limitado B . Isso é equivalente a propriedade: quaisquer seqüências $\{x^k\}$ e $\{v^k\}$ tais que $v^k \in S(x^k)$ e $\{x^k\}$ é limitada, então $\{v^k\}$ é limitada.*

Demonstração: Veja [113, Proposition 5.15]. ■

Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é localmente semicontínua inferior em \hat{x} , um ponto onde $f(\hat{x})$ é finito, se existe $\epsilon > 0$ tal que os conjuntos da forma $\{x \in B(\hat{x}, \epsilon) : f(x) \leq \alpha\}$, com $\alpha \leq f(\hat{x}) + \epsilon$, são fechados. A local semicontinuidade de f em \hat{x} pode ser interpretada como a propriedade local do epígrafo de f em $(\hat{x}, f(\hat{x}))$ ser fechado.

Proposição 2.1.7 *Suponha que $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ localmente semicontínua inferior em x com $f(x)$ finito. Então as seguintes afirmações são equivalentes:*

1. f é localmente Lipschitz em x ;
2. a aplicação $\partial_F f : x \mapsto \partial_F f(x)$ é localmente limitada em x ;
3. a aplicação $\partial_L f : x \mapsto \partial_L f(x)$ é localmente limitada em x ;

Além disso, quando uma dessas condições se verificam, temos que $\partial_L f(x)$ é não vazio e compacto.

Demonstração: Veja [113, Theorem 9.13]. ■

Funções multiobjetivos

A seguir exibimos alguns conceitos e resultados de otimização multiobjetivo. Maiores detalhes e demonstrações podem ser encontradas, por exemplo, em [94].

No espaço Euclidiano m -dimensional \mathbb{R}^m , a ordem parcial “ \preceq ” em \mathbb{R}^m induzida pelo cone Pareto \mathbb{R}_+^m é dado por $y \preceq z$ (ou $z \succeq y$) se, e somente se, $z - y \in \mathbb{R}_+^m$ com sua relação associada “ \prec ” dada por $y \prec z$ (ou $z \succ y$) se, e somente se, $z - y \in \mathbb{R}_{++}^m$, onde

$$\mathbb{R}_+^m := \{x \in \mathbb{R}^m : x_j \geq 0, j \in \mathcal{I}\}, \quad \mathbb{R}_{++}^m := \{x \in \mathbb{R}^m : x_j > 0, j \in \mathcal{I}\},$$

e $\mathcal{I} := \{1, \dots, m\}$.

Dada uma função vetorial $F := (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$, um ponto $x^* \in \mathbb{R}^n$ é chamado de *solução Pareto* ou *Pareto eficiente* de F , se não existe $x \in \mathbb{R}^n$ tal que $F(x) \preceq F(x^*)$ e $F(x) \neq F(x^*)$, ou seja, não existe $x \in \mathbb{R}^n$ tal que $f_i(x) \leq f_i(x^*)$, para todo $i = 1, \dots, m$ e $f_{i_0}(x) < f_{i_0}(x^*)$, para pelo menos um $i_0 \in \{1, \dots, m\}$. Dizemos que $x^* \in \mathbb{R}^n$ é *solução Pareto fraca* ou *Pareto eficiente fraca* de F , se não existe $x \in \mathbb{R}^n$ tal que $F(x) \prec F(x^*)$, isto é, não existe $x \in \mathbb{R}^n$ tal que $f_i(x) < f_i(x^*)$, para todo $i = 1, \dots, m$. Denotaremos o problema de encontrar os pontos Pareto fracos de F e conjunto dos pontos Pareto fracos de F , respectivamente por

$$\min_w F(x) \quad \text{e} \quad \arg \min_w F(x).$$

Um ponto $x \in \mathbb{R}^n$ é chamado de *Pareto crítico* de F se existir uma função componente f_i de F em que a derivada direcional de Clarke de f_i em x na direção de $y - x$ é não negativa, para todo $y \in \mathbb{R}^n$, com $i \in \mathcal{I}$. Esse conceito generaliza o de Pareto crítico para funções multiobjetivo diferenciáveis dado por:

“um ponto $x \in \mathbb{R}^n$ é um ponto Pareto crítico de F , se

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

onde

$$JF(x) := (\nabla f_1(x), \dots, \nabla f_m(x))^\top$$

é uma matriz $m \times n$ que denota o Jacobiano de F em $x \in \mathbb{R}^n$ e $\text{Im}(JF(x))$ denota a imagem do Jacobiano de F no ponto $x \in \mathbb{R}^n$ dada por

$$\text{Im}(JF(x)) := \{JF(x)v = (\langle \nabla f_1(x), v \rangle, \dots, \langle \nabla f_m(x), v \rangle) : v \in \mathbb{R}^n\}.$$

Esse último conceito por sua vez generaliza, para otimização multiobjetivo, o conceito de ponto crítico escalar “gradiente igual a zero”.

Dizemos que uma função vetorial $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é \mathbb{R}_+^m -convexa (resp. \mathbb{R}_+^m -quase convexa) se, para todo $x, y \in \mathbb{R}^n$ e $\lambda \in [0, 1]$

$$F(\lambda x + (1-\lambda)y) \preceq \lambda F(x) + (1-\lambda)F(y) \quad (\text{resp. } F(\lambda x + (1-\lambda)y) \preceq \max\{F(x), F(y)\}).$$

Note que o conceito de aplicação \mathbb{R}_+^m -convexa (resp. \mathbb{R}_+^m -quase convexa) é equivalente a convexidade (resp. quase convexidade) componente a componente.

Fejér convergência

Dizemos que uma sequência $\{y^k\}$ é *Fejér convergente* a um conjunto não vazio $U \subset \mathbb{R}^n$ se, para todo $k \in \mathbb{N}$,

$$\|y^{k+1} - y\| \leq \|y^k - y\|, \quad \forall y \in U.$$

O seguinte resultado é clássico e sua prova pode ser encontrada, por exemplo em [115, Teorema 2.7].

Proposição 2.1.8 *Seja $U \subset \mathbb{R}^n$ um conjunto não vazio e $\{y^k\}$ uma sequência Fejér convergente a U . Então, $\{y^k\}$ é limitada. Além disso, se um ponto de acumulação y de $\{y^k\}$ pertence a U , então $\{y^k\}$ converge para y .*

2.2 Diferença de funções convexas (DC)

Nesta seção apresentaremos alguns conceitos e resultados básicos de uma classe especial de funções, não necessariamente convexas, que é a classe das funções que são escritas como a diferença de duas funções convexas, ou funções DC, definidas no \mathbb{R}^n . Mais precisamente, $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma *função DC* quando existem $g, h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ funções convexas tal que $f(x) = g(x) - h(x)$, $\forall x \in \mathbb{R}^n$. Denotaremos por $\Gamma_0(\mathbb{R}^n)$ o cone das funções próprias, semicontínua inferiormente e convexas definidas no \mathbb{R}^n e o espaço vetorial das funções DC será denotado por $DC(\mathbb{R}^n)$. Como g e h podem assumir o valor $+\infty$, naturalmente tem-se convenção que $+\infty - (+\infty) = +\infty$. O espaço das funções DC é o menor espaço vetorial que contém todas as funções convexas e contínuas definidas em um determinado conjunto; veja [11]. Também temos que $DC(\mathbb{R}^n)$ contém o espaço $C^{1,1}$ das funções diferenciáveis cujo gradiente é localmente Lipschitz e os espaços das funções lower- C^2 ; veja [64, Teorema 2.4]. Por simplicidade, optamos por apresentar os resultados no espaço \mathbb{R}^n , mas resultados envolvendo diferença de funções convexas (ou côncavas) já foram estudados em contextos mais gerais, como por exemplo espaços de Hilbert ou Banach, veja [11] e suas referências, e espaços de Alexandrov de dimensão finita com curvatura limitada inferiormente, veja [6, 110].

Denotaremos um *problema de otimização DC* da seguinte forma:

$$\min_{x \in \mathbb{R}^n} f(x), \tag{2.1}$$

onde $f \in DC(\mathbb{R}^n)$, ou seja, $f(x) = g(x) - h(x)$, com $g, h \in \Gamma_0(\mathbb{R}^n)$.

Os exemplos a seguir podem ser encontrados em [72].

Exemplo 2.2.1 Seja $A \in \mathbb{R}^{n \times n}$ uma matriz simétrica e considere $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$Q(x) = \frac{1}{2} \langle Ax, x \rangle.$$

Obviamente, Q é uma função DC em \mathbb{R}^n e há várias formas de encontrar matrizes positivas semidefinidas A^+ e A^- , tal que, $A = A^+ - A^-$ e

$$Q(x) = \frac{1}{2} \langle A^+ x, x \rangle - \frac{1}{2} \langle A^- x, x \rangle.$$

Exemplo 2.2.2 Seja $S \subset \mathbb{R}^n$ um subconjunto não vazio. Denotando por d_S^2 o quadrado da distância ao conjunto S , temos que para todo $x \in \mathbb{R}^n$

$$d_S^2(x) = \|x\|^2 - (\|x\|^2 - d_S^2(x)).$$

Conforme dito em [72], a função $h(x) = \frac{1}{2}\|x\|^2 - \frac{1}{2}d_S^2(x)$ é convexa qualquer que seja S não vazio. Portanto, a função d_S^2 é uma função DC qualquer que seja S não vazio, porém d_S^2 é convexa somente quando S é não vazio e convexo.

Exemplo 2.2.3 Seja $A \in \mathbb{R}^{n \times n}$ uma matriz simétrica positiva definida e λ_M o maior autovalor de A . Assim, temos que

$$\frac{\lambda_M}{2} = \max_{\|x\| \leq 1} \langle Ax, x \rangle,$$

cujas reformulação dual é dada por

$$-\frac{\lambda_M}{2} = \min_{x^* \in \mathbb{R}^n} \{ \|x^*\| - \frac{1}{2} \langle A^{-1} x^*, x^* \rangle \}.$$

Portanto, calcular λ_M pode ser visto como um problema de minimização de uma função DC.

Seja $f \in DC(\mathbb{R}^n)$, então existem funções convexas $g, h : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que $f(x) = g(x) - h(x)$. Com isso, algumas propriedades de f são herdadas das funções g e h . Assim, temos, por exemplo, que f é localmente Lipschitz em \mathbb{R}^n . Com isso, a derivada direcional de f , $d \mapsto f'(x, d)$, existe e $f'(x, d) = g'(x, d) - h'(x, d)$, para todo $x, d \in \mathbb{R}^n$. Além disso, temos que $\partial_C f(x) = \partial g(x) - \partial h(x)$, para todo x ; veja [11]. Porém, diferentemente de funções convexas, o subdiferencial $\partial_C f(x)$ não necessariamente se reduz a $\{\nabla f(x)\}$ quando f é diferenciável em $x \in \mathbb{R}^n$, conforme [11, Exemplo 5.3]. No caso das funções DC, temos que $\partial_C f(x)$ se reduz quase sempre a $\{\nabla f(x)\}$ quando f é diferenciável; veja [11]. Uma outra propriedade que as funções convexas tem e não é válido para as funções DC é que o limite de funções

DC não necessariamente é uma função DC, conforme exemplo abaixo. Porém, sabe-se que toda função contínua definida em um conjunto convexo e compacto C é limite uniforme de uma sequência de funções DC; veja [11, Proposição 2.2].

Exemplo 2.2.4 *Sejam $f_n : \mathbb{R} \rightarrow \mathbb{R}$ dadas por $f_n(x) = \min_{k=1, \dots, n} \{|x - \frac{1}{k}|\}$. Para $n \in \mathbb{N}$, f_n é uma função DC, mas $f = \lim_{n \rightarrow +\infty} f_n$ não é uma função DC.*

O exemplo acima também serve para mostrar que o ínfimo (e o supremo) de uma família infinita de funções DC não é necessariamente uma função DC; veja [72]. As funções DC também gozam de algumas propriedades que não são válidas para funções convexas. Por exemplo, uma função por ser uma função DC ao longo de uma reta (ou seja, $t \mapsto f(a + t(b - a))$ é uma função DC, para todo $a, b \in \mathbb{R}$) sem ser uma função DC em todo \mathbb{R}^n ; veja [72]. Dizemos que uma função é *localmente DC* em \mathbb{R}^n se, para todo $x^0 \in \mathbb{R}^n$, existir uma vizinhança convexa V de x^0 , funções convexas g_V e h_V , tais que, $f(x) = g_V(x) - h_V(x)$, para todo $x \in V$.

Teorema 2.2.1 (Hartman) *Toda função localmente DC em \mathbb{R}^n é globalmente DC em \mathbb{R}^n .*

Demonstração: Veja [71]. ■

Como consequência do teorema acima, temos que toda função de classe C^2 é uma função DC. Algumas propriedades de funções convexas também são válidas para funções DC, conforme proposição abaixo.

Proposição 2.2.1 *Sejam $f_i \in DC(\mathbb{R}^n)$, $i = 1, \dots, m$. Temos que:*

1. $f(x) = \sum_{i=1}^m \lambda_i f_i(x)$, $\lambda_i \in \mathbb{R}$, é uma função DC;
2. $f(x) = \prod_{i=1}^m f_i(x)$ é uma função DC;
3. $f(x) = \max_{1 \leq i \leq m} f_i(x)$ é uma função DC;
4. $f(x) = \min_{1 \leq i \leq m} f_i(x)$ é uma função DC;
5. Se f é uma função DC e g uma função convexa, então $f \circ g$ é uma função DC.

Demonstração: Veja [64]. ■

2.3 Variedades Riemannianas

Nesta seção iremos introduzir algumas notações e propriedades fundamentais em variedades Riemannianas. Esses fatos básicos serão apresentados sem demonstrações e podem ser encontrados em qualquer livro de Geometria Riemanniana, por exemplo [38], [114].

2.3.1 Variedades diferenciáveis

Seja M uma variedade diferenciável e conexa m -dimensional. O *espaço tangente* a M em p , m -dimensional, será denotado por T_pM e o $TM = \cup_{p \in M} T_pM$ denota o *fibrado tangente* de M . Um *campo de vetores* X em M é uma aplicação $X : M \rightarrow TM$ que associa a cada $p \in M$ um $X_p \in T_pM$. Denotaremos por $\mathcal{X}^k(M)$ o espaço de campos de vetores em M de classe C^k , para $k \geq 0$. Se $k = \infty$, denotaremos por $\mathcal{X}(M)$. Neste trabalho, somente consideraremos variedades conexas. Assim, a partir de agora omitiremos o termo conexa sem que haja risco de confusão.

2.3.2 Métrica Riemanniana

Seja M uma variedade diferenciável m -dimensional. Para cada ponto $p \in M$ denotaremos por g uma *métrica Riemanniana* de M . Assim, para cada $p \in M$, a métrica Riemanniana determinará um *produto interno* no espaço tangente T_pM , $\langle u, v \rangle_p := g(u, v)$, com $u, v \in T_pM$, que varia diferencialmente com p . Quando não houver risco de confusão omitiremos o índice p do produto interno. Uma variedade com uma métrica Riemanniana será chamada de *variedade Riemanniana*. Definimos a *norma* de um vetor $u \in T_pM$ como $\|u\| := \langle u, u \rangle^{1/2}$.

Sejam M uma variedade Riemanniana e $f : M \rightarrow \mathbb{R}$ uma função de classe C^1 . Definimos o *gradiente* de f como o único campo $grad f \in \mathcal{X}^0(M)$ dado por

$$\langle grad f_p, v \rangle = df_p.v,$$

para cada $p \in M$ e $v \in T_pM$, onde df representa o diferencial de f .

Seja $c : [a, b] \rightarrow M$ uma curva C^∞ por partes ligando os pontos p e q em M , ou seja, $c(a) = p$ e $c(b) = q$. O *comprimento da curva* c , denotado por $L(c)$, é dado por

$$L(c) = \int_a^b \|c'(t)\| dt$$

e o *comprimento de arco* de c , denotado por $s(t)$, é dado por

$$s(t) = \int_a^t \|c'(t)\| dt,$$

onde $\|c'(t)\| = (\langle \frac{dc}{dt}, \frac{dc}{dt} \rangle)^{1/2}$. Dados $p, q \in M$, denotaremos por C_{pq} o conjunto de todas as curvas C^∞ por partes ligando os pontos p e q .

Assim, se M é uma variedade Riemanniana e $p, q \in M$, a *distância Riemanniana* de p a q , denotada por $d(p, q)$ é dada por

$$d(p, q) = \inf_{c \in C_{pq}} L(c).$$

A função distância $d : M \times M \rightarrow \mathbb{R}$ acima é contínua e induz a topologia original em M . O conjunto $B(p, r) = \{q \in M; d(p, q) < r\}$ é chamado de *bola métrica* de centro $p \in M$ e raio $r > 0$ e seu fecho é dado por $\bar{B}(p, r) = \{q \in M; d(p, q) \leq r\}$.

Dados M e N variedades Riemannianas, um difeomorfismo $\phi : M \rightarrow N$ de classe C^∞ é chamado de *isometria* se

$$\langle u, v \rangle_p = \langle d\phi_p(u), d\phi_p(v) \rangle_{\phi(p)},$$

onde $p \in M$ e $u, v \in T_pM$.

2.3.3 Conexão Riemanniana

Seja M uma variedade Riemanniana. Denotaremos por ∇ a *conexão de Levi-Civita* de M e por $\nabla_Y X$ a *derivada covariante* de X por Y , com $X \in \mathcal{X}^1(M)$ e $Y \in \mathcal{X}^0(M)$. Observe que $(\nabla_Y X)_p$ depende somente de Y_p e do valor de X ao longo de uma curva em M tangente a X_p . Assim, denotaremos este vetor apenas como $\nabla_{Y_p} X$.

Considere uma curva $c : [a, b] \rightarrow M$ de classe C^∞ e $X : [a, b] \rightarrow TM$ um campo de classe C^r ($r \geq 1$) ao longo da curva c , ou seja, $X(t) = X(c(t)) \in T_{c(t)}M$. A *derivada covariante* de X ao longo de c será denotada por $\frac{DX}{dt} = \nabla_{c'} X$. Um campo X ao longo de c é dito *paralelo*, se $\frac{DX}{dt} = \nabla_{c'} X = 0$. O *transporte paralelo* ao longo de uma curva c será denotado $P(c)_a^b$, ou simplesmente P_{pq} , quando $c(a) = p$, $c(b) = q$ e estiver claro que se trata da curva c .

Seja M uma variedade Riemanniana e $X \in \mathcal{X}^1(M)$. O *diferencial* do campo X é o operador linear $A_X : \mathcal{X}^0(M) \rightarrow \mathcal{X}^0(M)$ dado por $A_X(Y) = \nabla_Y X$, e para cada ponto $p \in M$, temos definida uma aplicação linear $A_X(p) : T_pM \rightarrow T_pM$

$$v \mapsto A_X(p).v = \nabla_v X.$$

Em particular, se $X = \text{grad } f$, onde $f : M \rightarrow \mathbb{R}$ é uma função de classe C^2 , então $A_X(p) = \text{Hess } f_p$ é a *hessiana* de f em p .

2.3.4 Geodésicas e aplicação exponencial

Uma curva $\gamma : I \rightarrow M$ é chamada de *geodésica* quando

$$\nabla_{\gamma'(t)}\gamma'(t) = 0, \forall t \in I \subset \mathbb{R}.$$

Denotaremos $\gamma''(t) = \frac{D\gamma'(t)}{dt}$, assim γ é geodésica, se $\gamma''(t) = 0$, para todo $t \in I$. Segue da definição que, se γ é geodésica, então $\|\gamma'(t)\|$ é constante, ou seja, γ tem velocidade constante. Logo, o comprimento de arco de γ a partir de um ponto inicial t_0 é dado por $s(t) = \|\gamma'(t)\|(t - t_0)$. Se $\|\gamma'(t)\| = 1$ dizemos que γ é *parametrizada pelo comprimento de arco* ou *normalizada*. A equação que define uma geodésica é uma equação diferencial ordinária não linear de segunda ordem, então uma geodésica γ é determinada por sua posição p e sua velocidade v em p . Quando for conveniente, denotaremos tal geodésica simplesmente por γ_v . A restrição de uma geodésica a um intervalo fechado e limitado é chamado de *segmento geodésico* e denotaremos o conjunto de todos os segmentos geodésicos ligando dois pontos p e q por Γ_{pq} . Um segmento geodésico ligando p a q em M é dito *minimal* se seu comprimento é igual a $d(p, q)$. Nesse caso, a geodésica é chamada de *geodésica minimizante*.

Variedades Riemannianas cujas geodésicas estão definidas para todo $t \in \mathbb{R}$, ou seja, $I = \mathbb{R}$ são chamadas de *variedades Riemannianas completas*. Assim, prova-se que para todo $a \in \mathbb{R}$, com $a > 0$ a igualdade

$$\gamma_{av}(t) = \gamma_v(at), \quad (2.2)$$

é satisfeita para todo $t \in \mathbb{R}$. Neste trabalho iremos considerar apenas variedades Riemannianas completas.

Em uma variedade Riemanniana completa M , para cada $p \in M$, a *aplicação exponencial* em p , $\exp_p : T_pM \rightarrow M$ é definida por

$$\exp_p v = \gamma_v(1),$$

onde γ_v é uma geodésica em M tal que $\gamma_v(0) = p$. Segue de (2.2) que $\gamma_v(t) = \exp_p tv$.

A aplicação exponencial é uma função de classe C^∞ e um difeomorfismo numa vizinhança Ω da origem em T_pM . O conjunto $\exp_p \Omega = \tilde{\Omega}$ é chamado uma *vizinhança normal* de p . Se $\tilde{\Omega}$ é uma vizinhança normal de cada um de seus pontos, então dizemos que $\tilde{\Omega}$ é uma *vizinhança totalmente normal*. Se $B_\epsilon(0) = \{v \in T_pM; \|v\| < \epsilon\}$ é tal que $\bar{B}_\epsilon(0) \subset \Omega$, chamamos $\exp_p B_\epsilon(0) = B(p, \epsilon)$ a *bola normal* ou *geodésica* de centro p e raio $\epsilon > 0$ que, nesse caso, coincide com a bola métrica.

Teorema 2.3.1 (Teorema de Hopf-Rinow) *Seja M uma variedade Riemanniana. As afirmações a seguir são equivalentes:*

1. Para cada ponto $p \in M$, \exp_p está definida em todo T_pM , ou seja, M é uma variedade Riemanniana completa.
 2. (M, d) é um espaço métrico completo, onde d é uma distância Riemanniana.
 3. Os subconjuntos limitados e fechados de M são compactos.
- Além disso, cada uma das afirmações acima implica em:
4. Para cada dois pontos $p, q \in M$ existe um segmento geodésico γ ligando p a q com $L(\gamma) = d(p, q)$, ou seja, γ é uma geodésica minimizante.

Demonstração: Veja [38, Teorema 2.8] ou [114, Teorema 1.1]. ■

Seja M uma variedade Riemanniana completa. Um *triângulo geodésico* em M formado pelos pontos $p_1, p_2, p_3 \in M$, denotado por $\Delta(p_1, p_2, p_3)$, é o conjunto formado pelos três pontos p_1, p_2 e p_3 chamados de *vértices*, e três segmentos geodésicos minimizantes γ_{i+1} ligando os pontos p_{i+1} a p_{i+2} , $i = 1, 2, 3(\text{mod } 3)$, chamados de *lados*.

2.3.5 Curvatura e fórmulas de variação

O *tensor curvatura* R de uma variedade Riemanniana M é dado por $R(X, Y) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[Y, X]} Z$, onde $X, Y, Z \in \mathcal{X}^r(M)$, $r \geq 2$ e o *colchete* $[X, Y] = YX - XY$. Assim, a *curvatura seccional* $K(X, Y)$ segundo o espaço gerado por X e Y é definida por

$$K(X, Y) = \frac{\langle R(X, Y)Y, X \rangle}{\|X\|^2 \|Y\|^2 - \langle X, Y \rangle^2},$$

onde $\|X\|^2 = \langle X, X \rangle$. Se, para cada par X, Y , $K(X, Y) \leq 0$ (respectivamente, $K(X, Y) \geq 0$) então dizemos que M é uma variedade Riemanniana de curvatura não positiva (respectivamente, não negativa) e denotaremos apenas por $K \leq 0$ (respectivamente, $K \geq 0$).

Seja M uma variedade Riemanniana e γ uma geodésica em M . Um campo J ao longo de γ é chamado *campo de Jacobi* se ele satisfaz a equação

$$\nabla_{\gamma'} \nabla_{\gamma'} J + R(J, \gamma') \gamma' = 0,$$

onde R é o tensor curvatura de M .

Seja M uma variedade Riemanniana e $\gamma : [a, b] \rightarrow M$ uma geodésica em M . Uma *variação* de γ é uma função $\alpha : [a, b] \times (-\epsilon, \epsilon) \rightarrow M$ de classe C^∞ tal que $\alpha(t, 0) \equiv \gamma(t)$. O campo de vetores ao longo de γ definido por $V(t) = \frac{\partial \alpha}{\partial s}(t, 0)$ é o *campo variacional* de α . Se a variação é tal que, para todo s , a curva $\alpha(\cdot, s)$ é uma geodésica, então o campo $J(t) = \frac{\partial \alpha}{\partial s}(t, s)$ é um campo de Jacobi ao longo dessa

geodésica. A fórmula de primeira variação de arco sobre a família de geodésicas $\alpha_s : [a, b] \rightarrow M$, dadas por $c_s(t) = \alpha(t, s)$, onde $s \in (-\epsilon, \epsilon)$, é

$$L'(\gamma) = \frac{d}{ds} L(c_s)|_{s=0} = \langle V, \frac{\gamma'}{\|\gamma'\|} \rangle|_a^b$$

e a fórmula de segunda variação de arco é dada por

$$\begin{aligned} L''(\gamma) &= \frac{d^2}{ds^2} L(c_s)|_{s=0} \\ &= \frac{1}{\|\gamma'\|} \int_a^b \{ \|\nabla_{\gamma'} V^\perp\|^2 - K(V, \gamma') (\|V\|^2 \|\gamma'\|^2 - \langle V, \gamma' \rangle^2) \} dt + \langle \nabla_V V, \frac{\gamma'}{\|\gamma'\|} \rangle|_a^b, \end{aligned}$$

onde $V^\perp = V - \langle V, \frac{\gamma'}{\|\gamma'\|} \rangle \frac{\gamma'}{\|\gamma'\|}$ denota a componente normal de V com relação à γ' .

2.3.6 Variedades de Hadamard

Uma variedade Riemanniana completa (e conexa), simplesmente conexa, com curvatura $K \leq 0$ é chamada de *variedade de Hadamard*. O teorema a seguir garante que uma variedade de Hadamard tem a mesma topologia e estrutura diferenciável dos espaços Euclidianos \mathbb{R}^n .

Teorema 2.3.2 (Teorema de Cartan-Hadamard) *Se M é uma variedade de Hadamard, então M é difeomorfa ao espaço Euclidiano \mathbb{R}^n , $n = \dim M$. Mais precisamente, $\exp_p : T_p M \rightarrow M$ é um difeomorfismo de classe C^∞ para cada $p \in M$.*

Demonstração: Veja [38, Teorema 3.1] ou [114, Teorema 4.1]. ■

Seja M uma variedade de Hadamard e $q \in M$. Pelo Teorema de Cartan-Hadamard podemos definir a inversa da aplicação exponencial $\exp_q^{-1} : M \rightarrow T_q M$ e obtém-se a seguinte relação entre distância Riemanniana e aplicação exponencial

$$d(p, q) = \|\exp_q^{-1} p\|. \quad (2.3)$$

Sendo \exp_q^{-1} uma aplicação de classe C^∞ , segue de (2.3) que a função $d^2(\cdot, q)$ também é de classe C^∞ . Além disso, foi provado em [54] que

$$\text{grad} \frac{1}{2} d^2(p, q) = -\exp_p^{-1} q. \quad (2.4)$$

Além de propriedades topológicas e diferenciáveis, algumas propriedades geométricas similares às dos espaços Euclidianos também são obtidas em variedades de Hadamard, como o teorema a seguir.

Teorema 2.3.3 (Lei dos cossenos) *Seja M uma variedade de Hadamard e $\Delta(x_1, x_2, x_3)$ um triângulo geodésico. Denote por $\gamma_{i+1} : [0, l_{i+1}] \rightarrow M$ o segmento*

geodésico ligando x_{i+1} a x_{i+2} e tome $l_{i+1} := L(\gamma_{i+1})$, $\theta_{i+1} = \angle(\gamma'_{i+1}(0), -\gamma'_i(l_i))$, onde $i = 1, 2, 3(\text{mod } 3)$. Então

$$\theta_1 + \theta_2 + \theta_3 \leq \pi, \quad (2.5)$$

$$l_{i+1}^2 + l_{i+2}^2 - 2l_{i+1}l_{i+2} \cos \theta_{i+2} \leq l_i^2, \quad (2.6)$$

$$l_{i+2} \leq l_{i+1} \cos \theta_{i+2} + l_i \cos \theta_i. \quad (2.7)$$

Se tivermos $K < 0$ as desigualdades acima são estritas.

Demonstração: Os resultados acima são consequências do Teorema de Toponogov. Veja [114, Teorema 4.2]. ■

Em termos da distância e da aplicação exponencial, a inequação (2.6) pode ser reescrita como:

$$d^2(x_{i+1}, x_{i+2}) + d^2(x_{i+2}, x_i) - 2\langle \exp_{x_{i+2}}^{-1} x_{i+1}, \exp_{x_{i+2}}^{-1} x_i \rangle \leq d^2(x_i, x_{i+1}), \quad (2.8)$$

pois $\langle \exp_{x_{i+2}}^{-1} x_{i+1}, \exp_{x_{i+2}}^{-1} x_i \rangle = d(x_{i+1}, x_{i+2})d(x_i, x_{i+2}) \cos \theta_{i+2}$.

2.4 Otimização em variedades Riemannianas

2.4.1 Convexidade

Nesta seção apresentaremos alguns conceitos e resultados clássicos da teoria de otimização em variedades Riemannianas que serão utilizados no próximo capítulo com bastante frequência. Os resultados apresentados e outros resultados de convexidade em variedades Riemanniana podem ser encontrados, por exemplo, em [129]. A partir de agora M será sempre uma variedade de Hadamard. Muitos dos resultados a seguir são válidos para variedades Riemannianas, mas focaremos apenas nas variedades de Hadamard.

Um subconjunto $C \subset M$ é dito *convexo* se, qualquer segmento geodésico com pontos finais em C está contido em C , ou seja, se $\gamma : [a, b] \rightarrow M$ é um segmento geodésico tal que $\gamma(a) = p$ e $\gamma(b) = q$, com $p, q \in C$, então $\gamma((1-t)a + tb) \in C$, para todo $t \in [0, 1]$.

Seja $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$. Dizemos que f é *própria*, se $f \not\equiv +\infty$. O *domínio* de f é definido por $\text{dom}(f) = \{x \in M : f(x) \neq +\infty\}$. Seja $C \subset M$ um conjunto convexo. Dizemos que f é *convexa* em C (respectivamente, *estritamente convexa*) se, para todo $p, q \in C$ e todo segmento geodésico $\gamma : [a, b] \rightarrow M$, ligando p e

q , a composição $f \circ \gamma : [a, b] \rightarrow \mathbb{R}$ é uma função real convexa (respectivamente, estritamente convexa), ou seja,

$$f(\gamma(t)) \leq (1-t)f(p) + tf(q), \quad (2.9)$$

para todo $p, q \in C$, $t \in [0, 1]$ e $\gamma \in \Gamma_{pq}$ (resp. $f(\gamma(t)) < (1-t)f(p) + tf(q)$). Uma relação equivalente a (2.9) é:

$$(f \circ \gamma)(tx + (1-t)y) \leq t(f \circ \gamma)(x) + (1-t)(f \circ \gamma)(y), \quad (2.10)$$

para todo $x, y \in [a, b]$ e $t \in [0, 1]$. Uma prova da equivalência entre (2.9) e (2.10) pode ser encontrada em [129, página 61].

Dizemos que f é *fortemente convexa* se existir uma constante positiva m tal que

$$f(\gamma(t)) \leq (1-t)f(p) + tf(q) - \frac{m}{2} \|\gamma'(t)\|^2 t(1-t),$$

para todo $p, q \in M$ e $t \in [0, 1]$.

Proposição 2.4.1 *Seja A um conjunto aberto convexo e $f : A \rightarrow \mathbb{R}$ uma função de classe C^2 . A função f é convexa (resp. estritamente convexa) em A se $\text{Hess } f$ é semidefinida positiva (resp. definida positiva) em A .*

Demonstração: Veja [129, página 82]. ■

Dizemos que uma função $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$ é *coerciva* se $f(x) \rightarrow +\infty$ sempre que $d(x, p) \rightarrow +\infty$, para algum $p \in M$.

Proposição 2.4.2 *Seja $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função coerciva, convexa e semi-contínua inferiormente. Então f tem um minimizador.*

Demonstração: Veja [12, Lema 2.3]. ■

2.4.2 Subdiferencial

Dado $\epsilon \geq 0$, o ϵ -subdiferencial de uma função convexa f em $x \in \text{dom}(f)$ é definido por

$$\partial_\epsilon f(x) = \{u \in T_x M ; f(y) \geq f(x) + \langle u, \exp_x^{-1} y \rangle - \epsilon, \forall y \in M\}. \quad (2.11)$$

Quando $\epsilon = 0$, denotaremos o subdiferencial exato de f em $x \in \text{dom}(f)$, $\partial_0 f(x)$, por $\partial f(x)$, ou seja,

$$\partial f(x) = \{u \in T_x M ; f(y) \geq f(x) + \langle u, \exp_x^{-1} y \rangle, \forall y \in M\}. \quad (2.12)$$

Segue direto da definição (2.11) que, se $0 \leq \epsilon_1 \leq \epsilon_2$, então

$$\partial f(x) \subseteq \partial_{\epsilon_1} f(x) \subseteq \partial_{\epsilon_2} f(x).$$

Além disso, segue de (2.12) que, $0 \in \partial f(x)$ se, e somente se, x é ponto de mínimo de f .

O teorema a seguir será importante para garantir a boa definição dos algoritmos que serão apresentados no próximo capítulo. Os resultados a seguir são apresentados para a versão exata do subdiferencial, mas também são válidos para ϵ -subdiferencial e a adaptação das provas são simples.

Proposição 2.4.3 *Seja $f : M \rightarrow \mathbb{R}$ uma função convexa. Então, para todo $x \in M$, $\partial f(x)$ é um conjunto não vazio, convexo e compacto.*

Demonstração: Veja [129, Teoremas 4.5 e 4.6]. ■

Uma caracterização do subdiferencial de uma função convexa f em um ponto $x \in M$ também pode ser obtida através das derivadas direcionais de f . A *derivada direcional* de f em $x \in M$ na direção de $v \in T_x M$ é definida por

$$f'(x, v) := \lim_{t \rightarrow 0^+} \frac{f(\gamma(t)) - f(x)}{t} = \inf_{t > 0} \frac{f(\gamma(t)) - f(x)}{t}, \quad (2.13)$$

onde $\gamma : [-\delta, \delta] \rightarrow M$ é um segmento geodésico tal que $\gamma(0) = x$ e $\gamma'(0) = v$.

Proposição 2.4.4 *Sejam $f : M \rightarrow \mathbb{R}$ uma função convexa, $x \in M$ e $v \in T_x M$. Assim, $v \in \partial f(x)$ se, e somente se, $f'(x, w) \geq \langle v, w \rangle$, para todo $w \in T_x M$.*

Demonstração: Veja [129, Teorema 4.8]. ■

Dizemos que uma função f é *semicontínua inferiormente* (respectivamente, *semicontínua superiormente*) em $x \in \text{dom}(f)$, se $x^k \rightarrow x$ implicar que $f(x) \leq \liminf_{k \rightarrow +\infty} f(x^k)$ (respectivamente, $\limsup_{k \rightarrow +\infty} f(x^k) \leq f(x)$).

Proposição 2.4.5 *Seja $f : M \rightarrow \mathbb{R}$ uma função convexa. Então, a aplicação $f' : TM \rightarrow \mathbb{R}$ é convexa e semicontínua superiormente em TM , ou seja, se $\{(x^k, v^k)\}$ converge para $(x, v) \in TM$, então*

$$\limsup_{k \rightarrow +\infty} f'(x^k, v^k) \leq f'(x, v).$$

Demonstração: Veja [25, Proposição 3.1]. ■

Proposição 2.4.6 *Sejam $f : M \rightarrow \mathbb{R}$ uma função convexa, $x^k \rightarrow x$, $v^k \rightarrow v$ e $\epsilon_k \rightarrow 0$, com $x^k, x \in M$, $v^k, v \in TM$, $\epsilon_k \geq 0$, $v^k \in \partial_{\epsilon_k} f(x^k)$, para todo $k \geq 0$. Então $v \in \partial f(x)$.*

Demonstração: O resultado segue diretamente de (2.11) usando as hipóteses do teorema. ■

O resultado a seguir ainda é válido para ϵ -subdiferencial e a prova é análoga à versão exata abaixo.

Proposição 2.4.7 *Sejam $f : M \rightarrow \mathbb{R}$ uma função convexa e $\{x^k\} \subset M$ uma sequência limitada. Se a sequência $\{v^k\} \subset TM$ é tal que $v^k \in \partial f(x^k)$ para cada $k \in \mathbb{N}$, então $\{v^k\}$ também é limitada.*

Demonstração: Veja [25, Proposição 3.2]. ■

O próximo resultado também será usado na boa definição dos algoritmos definidos no capítulo seguinte.

Proposição 2.4.8 *Se $f : M \rightarrow \mathbb{R}$ é uma função convexa, então para todo $x \in M$ e $\lambda > 0$, existe um único ponto, denotado por $p_\lambda(x)$, tal que*

$$f(p_\lambda(x)) + \frac{\lambda}{2}d^2(p_\lambda(x), x) = f_\lambda(x)$$

caracterizado por $\lambda(\exp_{p_\lambda(x)}^{-1} x) \in \partial f(p_\lambda(x))$, onde $f_\lambda(x) = \inf_{y \in M} \{f(y) + \lambda d^2(x, y)\}$.

Demonstração: Veja [54, Lema 4.2]. ■

Capítulo 3

Métodos de descida em variedades de Riemann

Neste capítulo estudaremos dois tipos de métodos de descida para resolver um problema de minimização irrestrito em variedades de Hadamard: o Método de Máxima Descida (MMD), também conhecido como método do gradiente, e o Método do Ponto Proximal (MPP). Durante as últimas décadas problemas de otimização em variedades Riemannianas se tornaram bastante populares e vários métodos tem sido estendidos para esse contexto para resolver diferentes tipos de problemas, por exemplo problemas de minimização usando o método do gradiente [43, 44, 107, 116], método do subgradiente [53, 70], método de Newton [3, 50, 55, 87, 116], método do ponto proximal [20, 22, 23, 47, 54, 107, 134]; problemas de viabilidade convexa [25, 135, 136]; problemas envolvendo campos vetoriais [45, 88, 91, 133]; problemas de desigualdade variacional [90, 92, 103]; problemas de otimização multiobjetivo [19, 24]; problemas de região de confiança [1, 14]; problemas de equilíbrio [46, 48, 77, 78]; ferramentas de análise convexa [89, 129], além de diversas aplicações que podem ser encontradas nas referências dos trabalhos citados acima.

É consenso que a primeira adaptação de problemas de otimização para variedades Riemannianas foi feita por Luenberger [95]. Em 1972, Luenberger estabeleceu resultados de convergência do método do gradiente projetado no espaço formado pela imagem inversa de valores regulares. Outro desenvolvimento importante foi o trabalho de Gabay [60]. Em 1982, Gabay analisou métodos de descida ao longo de geodésicas, incluindo o método do gradiente projetado e métodos de gradiente reduzido, para escolhas especiais de sistemas de coordenadas em uma variedade diferenciável. Nesse mesmo sentido, um esforço ainda maior para estender métodos de otimização para o cenário Riemanniano foi feito por Smith [116]. Em 1993, Smith desenvolveu algumas técnicas de otimização em variedades de Riemann utilizando uma linguagem geométrica sem nenhuma hipótese sobre o sistema local de coordenadas. Essa abordagem tem sido utilizada na literatura desde então. Dentre vários

outros autores destacamos aqui os livros dedicados à otimização em variedades Riemannianas de Udriste [129] e Absil et al. [2], e uma ferramenta de otimização em variedades para o Matlab, veja Boumal et al. [34].

Outro importante desenvolvimento no campo de métodos de otimização em variedades apareceu em 1998. Na mesma época, Cruz Neto et al. [43] e Ferreira e Oliveira [53] sugeriram a influência da curvatura seccional da variedade na convergência de algoritmos para resolver problemas de minimização. Em [43], os autores apresentam uma extensão do método do gradiente proposto em [95] para o contexto mais geral das variedades de Riemann com curvatura não negativa para o caso em que a função objetivo é continuamente diferenciável e convexa. Mais precisamente, provou-se convergência global da sequência gerada pelo método usando a regra de Armijo e um passo fixo para escolha do passo em cada iterada. Em [53], os autores provam a convergência do método subgradiente para funções convexas não diferenciáveis no mesmo ambiente que [43].

No cenário das variedades de Riemann tem sido usual adotar as variedades com curvatura não negativa como o ambiente apropriado para lidar com métodos do tipo gradiente ou subgradiente para resolver problemas de otimização em variedades. Bem como variedades de Hadamard (que possuem curvatura não positiva) o ambiente satisfatório para desenvolver as teorias de método do ponto proximal (proposto primeiramente em [54]), campos de vetores monótonos e desigualdade variacional; veja por exemplo [88, 91] e veja também [78, Theorem 5.1]. Os resultados deste capítulo deram origem aos trabalhos [120] e [125].

3.1 MPP para funções DC em variedades de Hadamard

O estudo de propriedades de diferença de funções convexas não é tão recente, destacando os trabalhos de P. Hartman[71] em 1959, J.F. Toland [128] em 1979, J.-B. Hiriart-Urruty [72] em 1985 e Pham Dinh Tao [109] em 1986. Porém, o estudo de algoritmos para resolver problemas de otimização DC é bem mais recente onde se destacam os vários trabalhos de Pham Dinh Tao com diversos autores. Em 2003, Sun et al. [126] apresenta uma versão do método do ponto proximal para resolver problemas de otimização DC utilizando apenas propriedades das duas funções convexas separadamente. Mais tarde, esse trabalho foi complementado com os resultados de Moudafi e Maingé [102].

Nesta seção estenderemos o algoritmo proposto em Sun et al. [126] para o cenário das variedades de Hadamard para resolver um problema de minimização irrestrito onde a função objetivo é escrita como a diferença entre duas funções convexas. Além

disso, consideraremos duas versões inexatas para o algoritmo. Na primeira delas, substituiremos a definição de subdiferencial pelo conceito de ϵ -subdiferencial. Na segunda, ao invés de resolvermos cada subproblema de forma exata iremos considerar um erro controlado na solução de cada subproblema, o que torna o algoritmo mais eficiente do ponto de vista prático ou computacional.

A teoria de funções DC tem sido estudada a bastante tempo. Porém, o estudo de algoritmos para resolver problemas de otimização DC é relativamente novo. Podemos dividir o estudo de algoritmos para problemas de otimização DC basicamente em duas abordagens: usando ferramentas e métodos de otimização combinatória (veja, por exemplo [73] e suas referências) e usando uma abordagem de análise convexa. Essa última, que será o foco deste capítulo, tem sido muito menos estudada que a primeira e, até onde sabemos, o primeiro algoritmo desse tipo para resolver um problema de otimização DC foi proposto por Pham Dinh Tao [109] em 1986. Tal algoritmo, do tipo subgradiente, é baseado na teoria de dualidade para funções DC e estende o clássico método do subgradiente para funções convexas. Em 2003, Sun et al. [126] propuseram um algoritmo para funções DC que combina uma fase subgradiente com uma fase do tipo proximal. Esse método generaliza o clássico método do ponto proximal para funções convexas. Algoritmos para funções DC são esperados convergir para soluções locais. Em um problema de otimização DC, a função objetivo admite infinitas decomposições que podem influenciar qualitativamente o algoritmo. Uma decomposição “ideal” para se obter convergência para uma solução global é um problema em aberto. Por isso, é importante estudar diferentes algoritmos para essa classe de funções. A seguir, estudaremos métodos que utilizam apenas propriedades decorrentes da convexidade das funções componentes para encontrar um ponto crítico da função objetivo (não necessariamente convexa).

No decorrer dessa seção consideraremos o seguinte problema de otimização DC:

$$\min_{x \in M} f(x) = g(x) - h(x), \quad (3.1)$$

onde M uma variedade de Hadamard de dimensão finita e $f : M \rightarrow \mathbb{R}$ uma função DC própria e limitada inferiormente, tal que, $f(x) = g(x) - h(x)$, com $g, h : M \rightarrow \mathbb{R}$ funções próprias, convexas e semicontínuas inferiormente. Além disso, estamos supondo que o conjunto S dos pontos críticos de f é não vazio, ou seja, $S = \{x \in M; \partial h(x) \cap \partial g(x) \neq \emptyset\}$ é não vazio.

3.1.1 Versão exata

Observe que, se x é um ponto crítico de f , então podemos tomar $w \in T_x M$ tal que $w \in \partial h(x) \cap \partial g(x)$. Assim, defina

$$y = \exp_x cw,$$

para algum $c > 0$. Como $w \in \partial g(x)$, pela definição de y , temos que

$$\frac{1}{c} \exp_x^{-1} y \in \partial g(x).$$

Reciprocamente, se

$$\frac{1}{c} \exp_x^{-1} y \in \partial g(x),$$

com $y = \exp_x cw$, $w \in \partial h(x)$ e $c > 0$, temos que $w \in \partial h(x) \cap \partial g(x)$. Portanto, podemos enunciar o seguinte resultado.

Proposição 3.1.1 *Uma condição necessária e suficiente para que $x \in M$ seja ponto crítico de f é que $\frac{1}{c} \exp_x^{-1} y \in \partial g(x)$, com $y = \exp_x cw$, $w \in \partial h(x)$ e $c > 0$.*

A proposição acima é a motivação para definição do seguinte algoritmo.

Algoritmo 3.1 *Passo 1: Tome um ponto inicial $x^0 \in M$ e $\{c_k\}$ uma sequência limitada de números reais positivos tal que $0 < b \leq c_k \leq c$, para todo $k \geq 0$.*

Passo 2: Tome

$$w^k \in \partial h(x^k) \text{ e defina } y^k := \exp_{x^k}(c_k w^k). \quad (3.2)$$

Passo 3: Calcule

$$x^{k+1} := \arg \min_{x \in M} \{g(x) + \frac{1}{2c_k} d^2(x, y^k)\}. \quad (3.3)$$

Se $x^{k+1} = x^k$, pare. Caso contrário, faça $k = k + 1$ e retorne ao Passo 2.

Observação 3.1.1 *As expressões (3.2) e (3.3) estão bem definidas em virtude das Proposições 2.4.3 e 2.4.8, respectivamente. Note que as expressões em (3.2) e (3.3) são equivalentes a*

$$\frac{1}{c_k} \exp_{x^k}^{-1} y^k \in \partial h(x^k) \quad (3.4)$$

e

$$\frac{1}{c_k} \exp_{x^{k+1}}^{-1} y^k \in \partial g(x^{k+1}), \quad (3.5)$$

respectivamente. Assim, se $x^{k+1} = x^k$, então $\frac{1}{c_k} \exp_{x^k}^{-1} y^k \in \partial h(x^k) \cap \partial g(x^k)$, em virtude de (3.4) e (3.5). Isso implica que x^{k+1} é ponto crítico de f .

Observação 3.1.2 Note ainda que, sendo $f(x) = g(x) - h(x)$, se $h \equiv 0$, então em (3.2) teremos $y^k = x^k$ e (3.3) se torna

$$x^{k+1} = \arg \min_{x \in M} \left\{ f(x) + \frac{1}{2c_k} d^2(x, x^k) \right\},$$

recuperando o método do ponto proximal estudado em [54] para problemas de minimização convexa em variedades de Hadamard, cuja convergência é obtida com a adição da hipótese $\sum_{k=0}^{\infty} c_k = +\infty$. Portanto, de uma certa forma, o Algoritmo 3.1 é mais geral que o proposto em [54], no sentido que o Algoritmo 3.1 resolve uma classe mais ampla de problemas. Além disso, se $M = \mathbb{R}^n$ e $d(x, y) = \|x - y\|$, então (3.2) e (3.3) coincidem com o algoritmo proposto em [126].

Exemplo 3.1.1 Denotando $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$ e $\langle u, v \rangle = p(x)uv$, com $p : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$, verifica-se que o par $\mathcal{H} = (\mathbb{R}_{++}, \langle \cdot, \cdot \rangle)$ é uma variedade de Hadamard cujos símbolos de Christoffel e a equação geodésica são dadas por

$$\Gamma(x) = \frac{1}{2} p^{-1}(x) \frac{d}{dx} p(x) = \frac{d}{dx} \ln \sqrt{p(x)} \quad e \quad \frac{d^2 x}{dt^2} + \Gamma(x) \left(\frac{dx}{dt} \right)^2 = 0,$$

respectivamente. Assim, dada uma função duas vezes diferenciável $f : \mathbb{R}_{++} \rightarrow \mathbb{R}$, o gradiente e a hessiana de f em $(\mathbb{R}_{++}, \langle \cdot, \cdot \rangle)$ são dadas por

$$\text{grad } f(x) = p^{-1}(x) f'(x) \quad e \quad \text{hess } f(x) = f''(x) - \Gamma(x) f'(x),$$

respectivamente, onde f' e f'' denotam as derivadas usuais em \mathbb{R} , veja [129, página 20]. Dessa forma, para o caso particular onde $p(x) = \frac{1}{x^2}$, temos

$$\Gamma(x) = -x^{-1}, \quad \text{grad } f(x) = x^2 h'(x) \quad e \quad \text{hess } f(x) = f''(x) + x^{-1} f'(x).$$

Além disso, a aplicação $\varphi : \mathbb{R} \rightarrow \mathbb{R}_{++}$ dada por $\varphi(x) = e^x$ é uma isometria entre o espaço Euclidiano \mathbb{R} e a variedade \mathcal{H} , cuja a distância Riemanniana $d : \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R}_+$ é dada por

$$d(x, y) = \left| \ln \left(\frac{x}{y} \right) \right|$$

e a única geodésica $\gamma : \mathbb{R} \rightarrow \mathbb{R}_{++}$, com $\gamma(0) = x$ e $\gamma'(0) = v$ é dada por

$$\gamma(t) = x e^{\frac{v}{x} t}.$$

Considere as funções $f, g, h : \mathcal{H} \rightarrow \mathbb{R}$ dadas por $g(x) = x^4$, $h(x) = 2x^2 - 3$ e $f(x) = g(x) - h(x) = x^4 - 2x^2 + 3$ podemos verificar que as funções g e h são limitadas inferiormente, contínuas e convexas em \mathcal{H} e que f não é convexa em \mathcal{H}

tendo ponto de mínimo em $x^* = 1$ com $f^* = f(x^*) = 2$. Logo, o método do ponto proximal proposto em [54] não pode ser aplicado à função f . Por outro lado, o Algoritmo 3.1 toma a seguinte forma: dados $x^0 \in \mathbb{R}_{++}$ e $0 < b \leq c_k \leq c$, calcule $x^{k+1} \in \mathbb{R}_{++}$ tal que

$$x^{k+1} = \arg \min_{x \in \mathbb{R}_{++}} \left\{ x^4 + \frac{1}{2c_k} \left(\ln \frac{x}{x^k} - 4(x^k)^2 \right)^2 \right\}.$$

A seguir apresentamos um gráfico do tipo "distância da solução" versus "número de iterações" com o comportamento do algoritmo sob diferentes escolhas de ponto inicial x^0 e sequência $\{c_k\}$ constante.

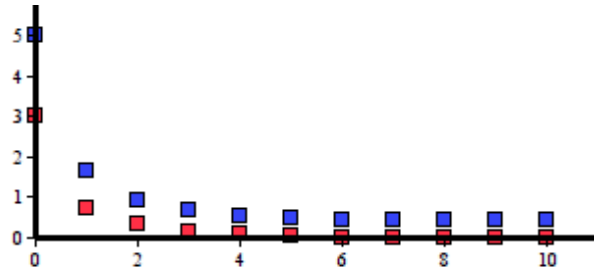


Figura 3.1: $x^0 = 3$ e $c_k = 1$; $x^0 = 5$ e $c_k = 0.5$

Exemplo 3.1.2 Denote por \mathbb{S}^n o conjunto das matrizes simétricas, \mathbb{S}_+^n o cone das matrizes simétricas semi-definidas positivas e \mathbb{S}_{++}^n o cone das matrizes simétricas positivas definidas $n \times n$. Dadas $X, Y \in \mathbb{S}_+^n$ dizemos que $Y \succeq X$ (ou $X \preceq Y$) se $Y - X \in \mathbb{S}_+^n$. Da mesma forma, definimos $Y \succ X$ (ou $X \prec Y$) se $Y - X \in \mathbb{S}_{++}^n$. Dessa forma, o par $M = (\mathbb{S}_{++}^n, \langle \cdot, \cdot \rangle)$, com a métrica induzida pela Hessiana Euclideana de $\varphi(X) = -\ln \det X$, dada por

$$\langle U, V \rangle = \text{tr}(V\varphi''(X)U) = \text{tr}(VX^{-1}UX^{-1}), \quad X \in M, \quad U, V \in T_X M,$$

é uma variedade de Hadamard cuja a única geodésica ligando dois pontos $X, Y \in M$ é dada por

$$\gamma(t) = X^{\frac{1}{2}}(X^{-\frac{1}{2}}YX^{-\frac{1}{2}})^t X^{\frac{1}{2}},$$

veja [82, Teorema 1.2]. Logo,

$$\gamma'(0) = X^{\frac{1}{2}} \ln(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}) X^{\frac{1}{2}},$$

de onde obtemos as expressões da aplicação exponencial, sua inversa e do quadrado da função distância dadas por

$$\exp_X V = X^{\frac{1}{2}} e^{(X^{-\frac{1}{2}} V X^{-\frac{1}{2}})} X^{\frac{1}{2}}, \quad \exp_X^{-1} Y = X^{\frac{1}{2}} \ln(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) X^{\frac{1}{2}}$$

e

$$d^2(X, Y) = \text{tr}(\ln^2(X^{-\frac{1}{2}}YX^{-\frac{1}{2}})) = \sum_{i=1}^n \ln^2 \lambda_i[X^{-\frac{1}{2}}YX^{-\frac{1}{2}}],$$

onde $\lambda_i[X^{-\frac{1}{2}}YX^{-\frac{1}{2}}]$ denota o i -ésimo autovalor da matriz $X^{-\frac{1}{2}}YX^{-\frac{1}{2}}$. Com isso, o gradiente e hessiana de uma função duas vezes diferenciável $F : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ são dadas, respectivamente, por

$$\text{grad } F(X) = XF'(X)X, \quad \text{hess } F(X)(V, V) = \text{tr}(VF''(X)V) + \text{tr}(F'(X)VX^{-1}V),$$

onde F' e F'' denotam o gradiente e hessiana Euclideana, respectivamente.

Dadas as funções $F, G, H : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$, $G(X) = e^{\text{tr}X}$, $H(X) = \ln \det X$ e $F(X) = G(X) - H(X) = e^{\text{tr}X} - \ln \det X$, verificamos que H, G são convexas e F não é convexa em \mathbb{S}_{++}^n . Assim, o método do ponto proximal proposto em [54] não pode ser aplicado à função f em \mathbb{S}_{++}^n . Por outro lado, o Algoritmo 3.1 toma a seguinte forma: dados $X^0 \in \mathbb{S}_{++}^n$ e $0 < b \leq c_k \leq c$, calcule $X^{k+1} \in \mathbb{S}_{++}^n$ tal que

$$X^{k+1} = \arg \min_{X \in \mathbb{S}_{++}^n} \left\{ e^{\text{tr}X} + \frac{1}{2c_k} \text{tr}(\ln^2(X^{-\frac{1}{2}}X^kX^{-\frac{1}{2}})) \right\}.$$

A seguir, mostraremos que o Algoritmo 3.1 possui propriedades similares aos algoritmos propostos em [54] e [126]. Mais precisamente, mostraremos que o Algoritmo 3.1 é de descida e, para uma quantidade suficientemente grande de passos, iteradas consecutivas estão arbitrariamente próximas.

Teorema 3.1.1 *A sequência $\{x^k\}$ gerada pelo Algoritmo 3.1 satisfaz um dos itens a seguir:*

1. O algoritmo para em um ponto crítico de f ;
2. f decresce estritamente ao longo da sequência, ou seja, $f(x^{k+1}) < f(x^k)$, para todo $k \geq 0$.

Demonstração: O primeiro item está provado na Observação 3.1.1. Agora, suponha $x^{k+1} \neq x^k$. Aplicando as inclusões (3.4) e (3.5), para cada $k \geq 0$, na definição de subdiferencial (2.12), temos que

$$h(x) - h(x^k) \geq \frac{1}{c_k} \langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x \rangle, \quad \forall x \in M, \forall k \geq 0$$

e

$$g(x) - g(x^{k+1}) \geq \frac{1}{c_k} \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x \rangle, \quad \forall x \in M, \forall k \geq 0.$$

Adicionando as duas últimas inequações, substituindo x por x^{k+1} na primeira delas

e x por x^k na segunda, temos

$$g(x^k) - h(x^k) - g(x^{k+1}) + h(x^{k+1}) \geq \frac{1}{c_k} (\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle + \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle),$$

para todo $k \geq 0$. Sendo $f(x) = g(x) - h(x)$, a última desigualdade pode ser vista como

$$f(x^k) - f(x^{k+1}) \geq \frac{1}{c_k} (\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle + \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle). \quad (3.6)$$

Agora iremos obter uma cota inferior para o lado direito da desigualdade (3.6). Assim, para cada $k \geq 0$, considere o triângulo geodésico $\triangle(y^k, x^k, x^{k+1})$, com o ângulo $\theta = \angle(\exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1})$. Pelo Teorema 2.3.3 e (2.8), obtemos que

$$d^2(y^k, x^k) + d^2(x^k, x^{k+1}) - 2\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle \leq d^2(y^k, x^{k+1}), \quad \forall k \geq 0.$$

Logo,

$$\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle \geq \frac{1}{2}d^2(y^k, x^k) + \frac{1}{2}d^2(x^k, x^{k+1}) - \frac{1}{2}d^2(y^k, x^{k+1}), \quad \forall k \geq 0. \quad (3.7)$$

De forma análoga, considerando o triângulo geodésico $\triangle(y^k, x^{k+1}, x^k)$, com $\theta = \angle(\exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k)$, obtemos que

$$\langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle \geq \frac{1}{2}d^2(y^k, x^{k+1}) + \frac{1}{2}d^2(x^k, x^{k+1}) - \frac{1}{2}d^2(y^k, x^k), \quad \forall k \geq 0. \quad (3.8)$$

Somando as inequações (3.7) e (3.8), temos

$$\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle + \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle \geq d^2(x^k, x^{k+1}).$$

Substituindo a inequação acima em (3.6), concluímos que

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{c_k}d^2(x^k, x^{k+1}), \quad \forall k \geq 0. \quad (3.9)$$

Sendo $\{c_k\}$ uma sequência de pontos positivos e $x^{k+1} \neq x^k$, ou seja, $d(x^k, x^{k+1}) > 0$, segue de (3.9) que $f(x^{k+1}) < f(x^k)$, para todo $k \geq 0$. ■

Tendo em vista o primeiro item do teorema anterior, nos resultados desta seção iremos considerar que $\{x^k\}$ não satisfaz o critério de parada, ou seja, que $\{x^k\}$ é uma sequência com infinitos pontos e $x^{k+1} \neq x^k$, para todo $k \geq 0$, pois no caso em que $x^{k+1} = x^k$ os resultados são óbvios.

Corolário 3.1.1 Se $\{x^k\}$ é a sequência gerada pelo Algoritmo 1, então $\{f(x^k)\}$ é convergente.

Demonstração: Esse resultado segue imediatamente do segundo item do teorema anterior juntamente com o fato de f ser limitada inferiormente. ■

O resultado a seguir prova que, se f for contínua então a imagem de todos os pontos de acumulação da sequência gerada pelo Algoritmo 3.1 são iguais.

Corolário 3.1.2 Suponha que $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.1. Se f é contínua e $\{x^k\}$ possui ponto de acumulação, então $\lim_{k \rightarrow \infty} f(x^k) = f(\bar{x})$, para todo ponto de acumulação \bar{x} de $\{x^k\}$.

Demonstração: Sejam $\bar{x} \in M$ um ponto de acumulação de $\{x^k\}$ e $\{x^{k_j}\}$ uma subsequência convergindo para \bar{x} . Como f é contínua, então $f(x^{k_j}) \rightarrow f(\bar{x})$ quando $j \rightarrow +\infty$. Por outro lado, pelo Corolário 3.1.1 temos que $\{f(x^k)\}$ é convergente. Portanto, temos que $f(x^k) \rightarrow f(\bar{x})$ quando $k \rightarrow +\infty$. ■

A resultado a seguir será utilizado no teorema de convergência do Algoritmo 1 e é uma propriedade clássica de métodos proximais.

Proposição 3.1.2 Se $\{x^k\}$ é gerada pelo Algoritmo 1, então $\sum_{k=0}^{+\infty} d^2(x^k, x^{k+1}) < \infty$. Em particular, $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$.

Demonstração: Pela definição do algoritmo, temos que $\frac{1}{c} \leq \frac{1}{c_k}$. Além disso, de (3.9), temos que

$$\frac{1}{c_k} d^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}), \quad \forall k \geq 0.$$

Logo,

$$\frac{1}{c} d^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}), \quad \forall k \geq 0.$$

Somando a desigualdade acima com k variando de 0 até $n - 1$, temos

$$\frac{1}{c} \sum_{k=0}^{n-1} d^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n).$$

Fazendo $n \rightarrow +\infty$ na desigualdade acima, lembrando que f é limitada inferiormente, temos que

$$\sum_{k=0}^{+\infty} d^2(x^k, x^{k+1}) < \infty,$$

e, em particular, $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$. ■

O próximo resultado é simples mas será importante no teorema de convergência desta seção.

Lema 3.1.1 *No Algoritmo 3.1, se $\{x^k\}$ é limitada, então $\{y^k\}$ também é limitada.*

Demonstração: Supondo $\{x^k\}$ limitada, então existe $p \in M$ e $r > 0$ tal que $d(x^k, p) \leq r$, para todo $k \geq 0$. Pela definição do algoritmo, $w^k \in \partial h(x^k)$. Sendo h convexa, pela Proposição 2.4.7, temos que $\{w^k\}$ é limitada, ou seja, existe $\delta > 0$ tal que $\|w^k\| \leq \delta$, para todo $k \geq 0$. Sendo M uma variedade de Hadamard, segue de (3.2) que

$$\|w^k\| = \frac{1}{c_k} d(x^k, y^k) \leq \delta.$$

Assim, $d(x^k, y^k) \leq \delta b$, para todo $k \geq 0$ e

$$d(y^k, p) \leq d(y^k, x^k) + d(x^k, p) \leq \delta b + r, \quad \forall k \geq 0.$$

Portanto, $\{y^k\}$ é limitada. ■

No que diz respeito a funções DC é natural se obter resultados de convergência para soluções locais. Ao contrário do método proximal estudado em [54] para funções convexas, não obtemos que a sequência gerada pelo algoritmo é limitada. Assim, o que obtemos é que os pontos de acumulação da sequência são pontos críticos da função em consideração.

Teorema 3.1.2 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 3.1. Então todos os pontos de acumulação de $\{x^k\}$, caso existam, são pontos críticos de f .*

Demonstração: Suponha que $\{x^k\}$ tenha um ponto de acumulação x^* e tome $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ convergindo para x^* . Pelo Lema 3.1.1 podemos supor sem perda de generalidade que existe uma subsequência $\{y^{k_j}\}$ de $\{y^k\}$ convergindo para y^* . Aplicando (3.4) e (3.5) em (2.12) e usando o fato que, para todo $k \in \mathbb{N}$, $c_k \leq c$, temos que

$$h(z) \geq h(x^{k_j}) + \frac{1}{c} \langle \exp_{x^{k_j}}^{-1} y^{k_j}, \exp_{x^{k_j}}^{-1} z \rangle, \quad \forall z \in M$$

e

$$g(z) \geq g(x^{k_j+1}) + \frac{1}{c} \langle \exp_{x^{k_j+1}}^{-1} y^{k_j}, \exp_{x^{k_j+1}}^{-1} z \rangle, \quad \forall z \in M.$$

Pela Proposição 4.3.1, se $\{x^{k_j}\}$ converge para x^* , então a sequência $\{x^{k_j+1}\}$ terá o mesmo comportamento. Assim, fazendo $j \rightarrow +\infty$ nas duas desigualdades acima, usando o fato de que h e g são semicontínuas inferiormente, temos que

$$h(z) \geq h(x^*) + \frac{1}{c} \langle \exp_{x^*}^{-1} y^*, \exp_{x^*}^{-1} z \rangle, \quad \forall z \in M$$

e

$$g(z) \geq g(x^*) + \frac{1}{c} \langle \exp_{x^*}^{-1} y^*, \exp_{x^*}^{-1} z \rangle, \quad \forall z \in M,$$

ou seja,

$$\frac{1}{c} \exp_{x^*}^{-1} y^* \in \partial h(x^*) \cap \partial g(x^*).$$

Portanto, temos que x^* é ponto crítico de f . ■

3.1.2 Versões inexatas

Nesta seção iremos apresentar duas versões inexatas do Algoritmo 3.1. Na primeira delas iremos considerar uma versão aproximada obtida através da substituição do subdiferencial por uma versão aproximada do subdiferencial, chamada ϵ -subdiferencial. Na segunda versão iremos considerar uma solução aproximada de cada subproblema do método.

ϵ -subdiferencial

A seguir, relembremos a definição de ϵ -subdiferencial de f em um ponto $x \in M$ dada em (2.11)

$$\partial_\epsilon f(x) = \{u \in T_x M ; f(y) \geq f(x) + \langle u, \exp_x^{-1} y \rangle - \epsilon, \forall y \in M\}.$$

Denotaremos $\partial_0 f(x) = \partial f(x)$, ou seja, 0-subdiferencial será o subdiferencial exato. Segue diretamente da definição acima que, se $0 \leq \epsilon_1 \leq \epsilon_2$, então

$$\partial_{\epsilon_1} f(x) \subseteq \partial_{\epsilon_2} f(x).$$

Assim, para $\epsilon > 0$, $\partial_\epsilon f(x)$ é uma expansão de $\partial f(x)$. Portanto, o uso de elementos no subdiferencial aproximados $\partial_\epsilon f(x)$ ao invés do subdiferencial exato $\partial f(x)$ fornece um maior grau de liberdade para o algoritmo que é muito útil em várias aplicações, principalmente computacionais. Por essa razão iremos considerar o seguinte algoritmo inexato que nesse sentido é mais abrangente que o Algoritmo 3.1. Antes disso, diremos que um ponto $x \in M$ é um ϵ -ponto crítico de uma função DC $f(x) = g(x) - h(x)$ se, dado $\epsilon \geq 0$, $\partial_\epsilon h(x) \cap \partial_\epsilon g(x) \neq \emptyset$. Quando $\epsilon = 0$, temos exatamente a definição de ponto crítico analisado na seção anterior.

A seguir mostraremos que o Algoritmo 3.2 tem propriedades similares ao Algoritmo 3.1.

Teorema 3.1.3 *A sequência $\{x^k\}$ gerada pelo Algoritmo 3.2 satisfaz um dos itens a seguir:*

1. O algoritmo para em um ϵ -ponto crítico de f ;
2. f satisfaz $f(x^{k+1}) < f(x^k) + 2\epsilon_k$, para todo $k \geq 0$.

Algoritmo 3.2 *Passo 1:* Dados um ponto inicial $x^0 \in M$, $\{c_k\}$ uma sequência limitada de números reais positivos tal que $0 < b \leq c_k \leq c$, para todo $k \geq 0$ e $\{\epsilon_k\}$ uma sequência de números não negativos, ou seja, $\epsilon_k \geq 0$ para todo $k \geq 0$.

Passo 2: Tome

$$w^k \in \partial_{\epsilon_k} h(x^k) \text{ e defina } y^k := \exp_{x^k}(c_k w^k). \quad (3.10)$$

Passo 3: Calcule $x^{k+1} \in M$ tal que

$$\frac{1}{c_k} \exp_{x^{k+1}}^{-1} y^k \in \partial_{\epsilon_k} g(x^{k+1}) \quad (3.11)$$

Se $x^{k+1} = x^k$, pare. Caso contrário, faça $k := k + 1$ e retorne ao Passo 2.

Demonstração: Se $x^{k+1} = x^k$, segue de (3.10) e (3.11) que

$$\frac{1}{c_k} \exp_{x^k}^{-1} y^k \in \partial_{\epsilon_k} h(x^k) \cap \partial_{\epsilon_k} g(x^k),$$

ou seja, x^{k+1} é um ϵ -ponto crítico de f . Agora, suponha $x^{k+1} \neq x^k$ para todo $k \geq 0$. Usando (3.10) e (3.11) em (2.11), temos que

$$h(x) \geq h(x^k) + \frac{1}{c_k} \langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x \rangle - \epsilon_k, \forall x \in M$$

e

$$g(x) \geq g(x^{k+1}) + \frac{1}{c_k} \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x \rangle - \epsilon_k, \forall x \in M.$$

Adicionando as duas últimas inequações, substituindo x por x^{k+1} na primeira delas e x por x^k na segunda, temos

$$h(x^{k+1}) + g(x^k) \geq h(x^k) + g(x^{k+1}) + \frac{1}{c_k} \langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle + \frac{1}{c_k} \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle - 2\epsilon_k,$$

ou seja,

$$g(x^k) - h(x^k) \geq g(x^{k+1}) - h(x^{k+1}) + \frac{1}{c_k} (\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle + \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle) - 2\epsilon_k.$$

Sendo $f(x) = g(x) - h(x)$, temos que

$$f(x^k) - f(x^{k+1}) + 2\epsilon_k \geq \frac{1}{c_k} (\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle + \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle).$$

Precedendo de forma análoga ao Teorema 4.3.1 para obter uma cota inferior para o lado direito da desigualdade acima, temos que

$$f(x^k) + 2\epsilon_k \geq f(x^{k+1}) + \frac{1}{c_k} d^2(x^k, x^{k+1}). \quad (3.12)$$

Sendo $\{c_k\}$ uma seqüência de pontos positivos e $x^{k+1} \neq x^k$ para todo k , ou seja, $d(x^k, x^{k+1}) > 0$, então $f(x^{k+1}) < f(x^k) + 2\epsilon_k$. ■

A partir de agora, nos resultados desta seção iremos considerar que $\{x^k\}$ não satisfaz o critério de parada, ou seja, que $\{x^k\}$ é uma seqüência com infinitos pontos, pois no caso em que $x^{k+1} = x^k$ os resultados são óbvios.

Corolário 3.1.3 *Se $\{x^k\}$ é uma seqüência gerada pelo Algoritmo 3.2 e $\sum_{k=0}^{+\infty} \epsilon_k < \infty$, então $\{f(x^k)\}$ é convergente.*

Demonstração: Pelo segundo item do teorema anterior, temos que, para todo $k \geq 0$,

$$\begin{aligned} f(x^k) &< f(x^{k-1}) + 2\epsilon_{k-1} \\ &< f(x^{k-2}) + 2\epsilon_{k-1} + 2\epsilon_{k-2} \\ &\vdots \\ &< f(x^0) + 2 \sum_{n=0}^{k-1} \epsilon_n \\ &\leq f(x^0) + 2 \sum_{n=0}^{+\infty} \epsilon_n < \infty. \end{aligned}$$

Além disso, sendo f limitada inferiormente podemos afirmar que $\{f(x^k)\}$ é limitada, logo possui pelo menos um ponto de acumulação. Mostraremos que $\{f(x^k)\}$ possui único ponto de acumulação. Para isso, suponha que $\{f(x^k)\}$ admite dois pontos de acumulação distintos, digamos $f_1 < f_2$. Considere $f(x^{k_j})$ e $f(x^{k_l})$ duas subsequências convergindo para f_1 and f_2 , respectivamente. Tome $\epsilon = \frac{f_2 - f_1}{4}$, então existem $k_{j_0}, k_{l_0}, \bar{k} \in \mathbb{N}$ tais que

$$\sum_{k=\bar{k}}^{+\infty} \epsilon_k < \epsilon, \quad (3.13)$$

$$f(x^{k_j}) < f_1 + \epsilon, \forall k_j \geq k_{j_0} \quad (3.14)$$

e

$$f_2 - \epsilon < f(x^{k_l}), \forall k_l \geq k_{l_0}. \quad (3.15)$$

Denote por $k_0 = \max\{k_{j_0}, k_{l_0}, \bar{k}\}$. Pelo teorema anterior, para todo $k \geq k_0$, temos

$$f(x^k) \leq f(x^{k_0}) + 2 \sum_{n=k_0}^{k-1} \epsilon_n \leq f(x^{k_0}) + 2 \sum_{n=k_0}^{+\infty} \epsilon_n.$$

Sendo $k_0 \geq \bar{k}$, por (3.13), temos que $f(x^k) \leq f(x^{k_0}) + 2\epsilon$. Além disso, $k_0 \geq k_{j_0}$, logo

por (3.14), temos que $f(x^k) \leq f(x^{k_0}) + 2\epsilon < f_1 + 3\epsilon = f_2 - \epsilon$, ou seja,

$$f(x^k) < f_2 - \epsilon, \forall k \geq k_0. \quad (3.16)$$

Por outro lado, $k_0 \geq k_{l_0}$. Logo, por (3.15), para todo $k \geq k_0$ temos que

$$f_2 - \epsilon < f(x^k),$$

que contraria (3.16). Portanto, $\{f(x^k)\}$ possui único ponto de acumulação implicando que $\{f(x^k)\}$ é convergente. ■

Corolário 3.1.4 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.2. Suponha que f seja contínua, $\{x^k\}$ possua ponto de acumulação e $\sum_{k=0}^{+\infty} \epsilon_k < \infty$. Então $\lim_{k \rightarrow \infty} f(x^k) = f(\bar{x})$, para todo ponto de acumulação \bar{x} de $\{x^k\}$.*

Demonstração: Análogo ao Corolário 3.1.2 . ■

O resultado a seguir é clássico para métodos do tipo proximal e será utilizado no teorema de convergência do Algoritmo 3.2.

Proposição 3.1.3 *Se $\{x^k\}$ é gerada pelo Algoritmo 3.2 e $\sum_{k=0}^{+\infty} \epsilon_k < \infty$, então $\sum_{k=0}^{+\infty} d^2(x^k, x^{k+1}) < \infty$. Em particular, $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$.*

Demonstração: Pela definição do algoritmo, temos que $\frac{1}{c} \leq \frac{1}{c_k}$, para todo $k \geq 0$. Além disso, de (3.12), temos que

$$\frac{1}{c_k} d^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}) + 2\epsilon_k.$$

Logo,

$$\frac{1}{c} d^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}) + 2\epsilon_k.$$

Somando a desigualdade acima com k variando de 0 a $n-1$, temos

$$\frac{1}{c} \sum_{k=0}^{n-1} d^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n) + 2 \sum_{k=0}^{n-1} \epsilon_k.$$

Fazendo $n \rightarrow +\infty$ na desigualdade acima, lembrando que f é limitada inferiormente e utilizando a hipótese que $\sum_{k=0}^{+\infty} \epsilon_k < \infty$, temos a convergência da série

$$\sum_{k=0}^{+\infty} d^2(x^k, x^{k+1}) < \infty,$$

e, em particular, segue que $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$. ■

Conforme mencionado anteriormente, a Proposição 2.4.7 também é válida para ϵ -subdiferencial. Com isso, um resultado análogo ao Lema 3.1.1 pode ser obtido para o Algoritmo 3.2. Assim, no Algoritmo 3.2, se a sequência $\{x^k\}$ é limitada, então a sequência $\{y^k\}$ também será limitada.

Teorema 3.1.4 *Se $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.2 e $\sum_{k=0}^{+\infty} \epsilon_k < \infty$, então todo ponto de acumulação de $\{x^k\}$, caso exista, é ponto crítico de f .*

Demonstração: Suponha que $\{x^k\}$ tenha um ponto de acumulação $x^* \in M$. Assim, existem subsequências $\{x^{k_j}\}$ e $\{y^{k_j}\}$ de $\{x^k\}$ e $\{y^k\}$ convergindo para x^* e y^* , respectivamente. Sem perda de generalidade, podemos usar o mesmo índice nas duas subsequências (podemos extrair uma outra subsequência caso necessário). Assim, aplicando (3.10) e (3.11) em em (2.12), obtemos

$$h(z) \geq h(x^{k_j}) + \frac{1}{c} \langle \exp_{x^{k_j}}^{-1} y^{k_j}, \exp_{x^{k_j}}^{-1} z \rangle - \epsilon_{k_j}, \quad \forall z \in M \quad (3.17)$$

$$g(z) \geq g(x^{k_j+1}) + \frac{1}{c} \langle \exp_{x^{k_j+1}}^{-1} y^{k_j}, \exp_{x^{k_j+1}}^{-1} z \rangle - \epsilon_{k_j}, \quad \forall z \in M, \quad (3.18)$$

usando que $\frac{1}{c_k} \geq \frac{1}{c}$, para todo $k \geq 0$. Da hipótese do teorema, segue que $\lim_{k \rightarrow \infty} \epsilon_k = 0$. Além disso, segue da Proposição 3.1.3 que $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$ e como $x^{k_j} \rightarrow x^*$, então $x^{k_j+1} \rightarrow x^*$. Logo, aplicando o limite em (3.17) e (3.18), com $j \rightarrow +\infty$, tendo em mente que h e g são funções semicontínuas inferiormente, obtemos

$$\frac{1}{c} \exp_{x^*}^{-1} y^* \in \partial h(x^*) \cap \partial g(x^*).$$

Portanto, x^* é ponto crítico de f . ■

Observação 3.1.3 *Note que, mesmo considerando no Algoritmo 3.2 elementos pertencente ao ϵ -subdiferencial das funções componentes h e g , o resultado de convergência é obtido para um ponto crítico (e não para um ϵ -ponto crítico) da função f .*

Solução aproximada em cada subproblema

Conforme observado por Rockafellar [112], para que um algoritmo proximal seja prático é importante que se considere soluções aproximadas dos subproblemas do método. Algoritmos do tipo proximal para funções DC onde são consideradas soluções aproximadas de cada subproblemas ainda não foram propostos. Por essa razão, apresentamos a seguir uma versão inexata do Algoritmo 3.2, onde cada subproblema é resolvido aproximadamente levando em consideração um erro controlado.

Algoritmo 3.3 *Passo 1: Dados um ponto inicial $x^0 \in M$ e $\{c_k\}$ uma seqüência limitada de números reais positivos tal que $0 < b \leq c_k \leq c$, para todo $k \geq 0$.*

Passo 2: Determine

$$w^k \in \partial h(x^k) \text{ e tome } y^k = \exp_{x^k}(c_k w^k). \quad (3.19)$$

Passo 3: Calcule $x^{k+1} \in M$ e $e^{k+1} \in T_{x^{k+1}}M$ tais que

$$e^{k+1} \in \partial g(x^{k+1}) - \frac{1}{c_k} \exp_{x^{k+1}}^{-1} y^k, \quad (3.20)$$

onde

$$\|e^{k+1}\| \leq \eta d(x^{k+1}, x^k), \quad \eta c_k \in [0, 1). \quad (3.21)$$

Se $x^{k+1} = x^k$, pare. Caso contrário, faça $k := k + 1$ e retorne ao Passo 2.

Observação 3.1.4 *Note que, quando $x^{k+1} = x^k$, (3.21) obviamente implica que $e^{k+1} = 0$ e o Algoritmo 3.3 se reduz ao Algoritmo 3.1 (versão exata). Além disso, fazendo $\eta = 0$ em (3.21) também obtemos $e^{k+1} = 0$ e estamos no Algoritmo 3.1.*

Teorema 3.1.5 *A seqüência $\{x^k\}$ gerada pelo Algoritmo 3.3 satisfaz um dos itens a seguir:*

1. *O algoritmo para em um ponto crítico de f ;*
2. *f decresce estritamente ao longo da seqüência, ou seja, $f(x^{k+1}) < f(x^k)$, $\forall k \geq 0$.*

Demonstração: Se $x^{k+1} = x^k$, como vimos anteriormente o Algoritmo 3.3 se reduz ao Algoritmo 3.1, onde $x^{k+1} = x^k$ implica em x^{k+1} ponto crítico de f . Agora, suponha $x^{k+1} \neq x^k$ para todo $k \geq 0$. Usando (3.19) e (3.20) em (2.12), temos que, para todo $k \geq 0$, valem as desigualdades

$$h(x) \geq h(x^k) + \frac{1}{c_k} \langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x \rangle, \quad \forall x \in M$$

e

$$g(x) \geq g(x^{k+1}) + \frac{1}{c_k} \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x \rangle + \langle e^{k+1}, \exp_{x^{k+1}}^{-1} x \rangle, \quad \forall x \in M.$$

Adicionando essas duas inequações, substituindo x por x^{k+1} na primeira delas e x por x^k na segunda, tendo em mente que $f(x) = g(x) - h(x)$, obtemos

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq \frac{1}{c_k} \left(\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1} \rangle + \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k \rangle \right) \\ &\quad + \langle e^{k+1}, \exp_{x^{k+1}}^{-1} x^k \rangle, \quad \forall k \geq 0. \end{aligned}$$

De forma análoga ao Teorema 4.3.1 obtemos uma cota inferior para o lado direito da desigualdade acima e concluímos que

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{c_k} d^2(x^k, x^{k+1}) + \langle e^{k+1}, \exp_{x^{k+1}}^{-1} x^k \rangle, \quad \forall k \geq 0.$$

Usando a desigualdade de Cauchy-Schwarz, temos que

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{c_k} d^2(x^k, x^{k+1}) - \|e^{k+1}\| d(x^k, x^{k+1}), \quad \forall k \geq 0.$$

Logo, por (3.21), na definição do algoritmo, temos

$$f(x^k) \geq f(x^{k+1}) + \frac{(1 - \eta c_k)}{c_k} d^2(x^k, x^{k+1}), \quad \forall k \geq 0. \quad (3.22)$$

Usando o fato que $\{c_k\}$ uma sequência de pontos positivos, $\eta c_k < 1$ e $x^{k+1} \neq x^k$, ou seja, $d(x^k, x^{k+1}) > 0$ em (3.22), temos que $f(x^{k+1}) < f(x^k)$, para todo $k \geq 0$. ■

Nos resultados a seguir iremos considerar que $\{x^k\}$ não satisfaz o critério de parada, ou seja, que $\{x^k\}$ é uma sequência com infinitos pontos, pois no caso em que $x^{k+1} = x^k$ os resultados são óbvios.

Corolário 3.1.5 *Se $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.3, então $\{f(x^k)\}$ é convergente.*

Demonstração: Análoga ao Corolário 3.1.1. ■

Corolário 3.1.6 *Suponha $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.3. Se f é contínua e $\{x^k\}$ possui ponto de acumulação, então $\lim_{k \rightarrow \infty} f(x^k) = f(\bar{x})$, para todo ponto de acumulação \bar{x} de $\{x^k\}$.*

Demonstração: Análogo ao Corolário 3.1.2. ■

Nos resultados a seguir iremos considerar $\eta > 0$, pois se $\eta = 0$ o Algoritmo 3.3 se reduz ao Algoritmo 3.1, cujo resultado de convergência foi provado no Teorema 3.1.2.

Proposição 3.1.4 *Se $\{x^k\}$ é gerada pelo Algoritmo 3.3, então $\sum_{k=0}^{+\infty} d^2(x^k, x^{k+1}) < \infty$ e, em particular, $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$.*

Demonstração: Pela definição do algoritmo, temos que $\eta c_k < 1$ e $\frac{1}{c} \leq \frac{1}{c_k}$, para todo $k \geq 0$. Com isso, temos que $\frac{(1 - \eta c)}{c} \leq \frac{(1 - \eta c_k)}{c_k}$, para todo $k \geq 0$. Logo, de (3.22), temos que

$$\frac{(1 - \eta c)}{c} d^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}).$$

Somando a desigualdade acima com k variando de 0 até $n - 1$, obtemos

$$\frac{(1 - \eta c)}{c} \sum_{k=0}^{n-1} d^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n).$$

Fazendo $n \rightarrow +\infty$ na desigualdade acima, lembrando que f é limitada inferiormente, temos a convergência da série

$$\sum_{k=0}^{+\infty} d^2(x^k, x^{k+1}) < \infty,$$

donde segue que $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$. ■

Em virtude do Lema 3.1.1, no Algoritmo 3.3, se a sequência $\{x^k\}$ é limitada, então a sequência $\{y^k\}$ também será limitada.

Teorema 3.1.6 *Se $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.3, então todo ponto de acumulação de $\{x^k\}$, caso exista, é ponto crítico de f .*

Demonstração: Seja $x^* \in M$ um ponto de acumulação da sequência $\{x^k\}$. Assim, considere $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ convergindo para x^* . Sem perda de generalidade, podemos tomar $\{y^{k_j}\}$ uma subsequência de $\{y^k\}$ convergindo para y^* . Aplicando (3.19) e (3.20) em (2.12), segue que para todo $k \geq 0$, valem as desigualdades

$$h(z) \geq h(x^{k_j}) + \frac{1}{c} \langle \exp_{x^{k_j}}^{-1} y^{k_j}, \exp_{x^{k_j}}^{-1} z \rangle, \quad \forall z \in M \quad (3.23)$$

e

$$g(z) \geq g(x^{k_j+1}) + \frac{1}{c} \langle \exp_{x^{k_j+1}}^{-1} y^{k_j}, \exp_{x^{k_j+1}}^{-1} z \rangle - \|e^{k_j+1}\| d(z, x^{k_j+1}), \quad \forall z \in M, \quad (3.24)$$

aplicando em ambas as desigualdades acima que $\frac{1}{c_k} \geq \frac{1}{c}$, para todo $k \geq 0$, e a desigualdade de Cauchy-Schwarz em (3.24). Substituindo (3.21) em (3.24), obtemos

$$g(z) \geq g(x^{k_j+1}) + \frac{1}{c} \langle \exp_{x^{k_j+1}}^{-1} y^{k_j}, \exp_{x^{k_j+1}}^{-1} z \rangle - \eta d(x^{k_j}, x^{k_j+1}) d(z, x^{k_j+1}), \quad \forall z \in M. \quad (3.25)$$

Segue da Proposição 3.1.4 que $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$. Assim, se $\{x^{k_j}\}$ converge para x^* , então $\{x^{k_j+1}\}$ também converge para x^* . Logo, aplicando o limite em (3.23) e (3.25) com $j \rightarrow +\infty$, tendo em mente que as funções h e g são semicontínuas inferiormente, temos que

$$\frac{1}{c} \exp_{x^*}^{-1} y^* \in \partial h(x) \text{ e } \frac{1}{c} \exp_{x^*}^{-1} y^* \in \partial g(x),$$

respectivamente. Portanto, x^* é ponto crítico de f . ■

3.1.3 Aplicação em problemas de maximização com restrições

Considere o problema de maximizar uma função $h : M \rightarrow \mathbb{R}$ convexa e semicontínua inferiormente restrita a um subconjunto convexo e fechado $C \subset M$. Resumidamente

$$\max_{x \in C} h(x). \quad (3.26)$$

Os resultados anteriores deste capítulo ainda continuam válidos se considerarmos que a função objetivo f assume valores $+\infty$. Para isso, basta incluirmos algumas hipóteses tais como as funções envolvidas serem próprias, o ponto inicial do algoritmo pertencer ao interior do domínio da função em consideração e a convenção $+\infty - (+\infty) = +\infty$; veja mais detalhes no próximo capítulo. Dessa forma, o problema (3.26) pode ser reescrito como um problema irrestrito de minimização de diferença de funções convexas, ou seja, (3.26) é equivalente ao seguinte problema:

$$- \min_{x \in M} \{\delta_C(x) - h(x)\}, \quad (3.27)$$

onde $\delta_C(x)$ é a função indicadora, definida por $\delta_C(x) = 0$, se $x \in C$ e $\delta_C(x) = +\infty$, caso contrário. Considere $N_C(x)$ o cone normal do conjunto C no ponto $x \in C$ definido por

$$N_C(x) := \{u \in T_x M; \langle u, \exp_x^{-1} y \rangle \leq 0, \quad \forall y \in C\}.$$

Então

$$\partial \delta_C(x) = N_C(x), \quad \forall x \in C.$$

Nesse contexto o Algoritmo 3.1 tem a seguinte forma: dados $x^0 \in M$ e uma sequência $\{c_k\}$ de números positivos tal que $0 < b \leq c_k \leq c$, para todo $k \geq 0$, tome $w^k \in \partial h(x^k)$ e defina $y^k = \exp_{x^k}(c_k w^k)$. Em seguida, calcule $x^{k+1} \in M$ tal que

$$\langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} y \rangle \leq 0, \quad \forall y \in C. \quad (3.28)$$

Segue de [54, Corolário 3.1] que, para cada k , existe um único ponto $x^{k+1} \in M$ que é solução de (3.28) e tal ponto é a projeção de y^k sobre o conjunto C .

3.2 MMD em variedades de Hadamard

Nesta seção abordaremos o método de máxima descida também presente na literatura como método do gradiente. Os resultados da primeira parte deste capítulo

seguem a mesma ideia de [9, 18]. Como mencionamos anteriormente, foi em 1972 com Luenberger [95] que o método de máxima descida ao longo de geodésicas foi primeiramente estudado. Mas, foi somente em 1993 que Smith [116] obteve os primeiros resultados de convergência desse método no contexto das variedades de Riemann. Em 1998, Cruz Neto et al.[43] observa a influência da curvatura da variedade para obter a convergência global da sequência gerada pelo método de máxima descida (com o passo de Armijo e passo fixo) para funções continuamente diferenciáveis e convexas definidas em uma variedade de Riemann completa de dimensão finita com curvatura não-negativa. Esse resultado foi estendido para funções quase-convexas (continuamente diferenciáveis) por Papa Quiroz et al.[107] em 1998. Dessa forma, tem sido comum considerar o contexto das variedades de Riemann com curvatura não-negativa o ambiente apropriado para se estudar a convergência global de métodos como gradiente e subgradiente, veja por exemplo [43, 44, 53, 107]. No decorrer dessa seção mostraremos que o método de máxima descida pode ser estendido para as variedades de Hadamard usando a propriedade de Kurdyka-Lojasiewicz. Como aplicação usaremos o método de máxima descida para encontrar um centro de massa Riemanniano com relação a um conjunto de dados em uma variedade de Hadamard. Esse problema tem sido amplamente estudado no contexto das variedades de Riemann com curvatura não-negativa.

Considere o problema de minimização irrestrito

$$\min_{x \in M} f(x) \tag{3.29}$$

onde M é uma variedade de Hadamard de dimensão finita e $f : M \rightarrow \mathbb{R}$ é uma função continuamente diferenciável com gradiente Lipschitz de constante $L > 0$. Assuma que o conjunto solução de (3.29) é não-vazio.

O método de máxima descida gera uma sequência da seguinte forma:

Algoritmo 3.4 *Passo 1: Escolha $x^0 \in M$ e $\delta_1, \delta_2 > 0$ tal que $L\delta_1 + \delta_2 < 1$;*
Passo 2: Dado $x^k \in M$, se x^k é um ponto crítico de f , então faça $x^{k+p} = x^k$ para todo $p \in \mathbb{N}$;
Passo 3: Caso contrário, tome como próxima iterada $x^{k+1} \in M$ tal que

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad}f(x^k)), \tag{3.30}$$

onde

$$t_k \in \left(\delta_1, \frac{2}{L}(1 - \delta_2) \right). \tag{3.31}$$

Observação 3.2.1 *A boa definição da igualdade (3.30) segue do Teorema de Cartan-Hadamard, mais precisamente, pelo fato de, em uma variedade de Hada-*

mand, a aplicação exponencial é um difeomorfismo (global). Além disso, segue de (3.30) que

$$t_k \|\text{grad}f(x^k)\| = d(x^{k+1}, x^k), \quad \forall k \geq 0. \quad (3.32)$$

Se $x^{k+1} = x^k$, segue de (3.32) que $\text{grad}f(x^k) = 0$, uma vez que $t_k > 0$, ou seja, x^k é ponto crítico de f . Com isso, temos um critério de parada prático para o Algoritmo 3.4.

Proposição 3.2.1 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.4. Se $x^{k+1} = x^k$, então o algoritmo para em x^k que é um ponto crítico de f .*

Observação 3.2.2 *Note que em (3.31) o passo pode variar desde que permaneça no intervalo determinado. Mesmo assim, esse passo é denominado na literatura como passo fixo. Nos resultados desta seção iremos considerar o método de máxima descida com o passo fixo como em (3.31), mas ressaltamos que resultados similares podem ser obtidos com o passo clássico conhecido como “busca de Armijo” e definido da seguinte forma:*

$$t_k := \arg \max_t \{ f(\exp_{x^k}(-t \text{grad}f(x^k))) \leq f(x^k) - \alpha t \|\text{grad}f(x^k)\|^2, t = 2^{-j} : j \in \mathbb{N} \},$$

com $\alpha \in (0, 1)$. Dessa forma, quando a função objetivo f tem gradiente Lipschitz, temos que zero não é um ponto de acumulação da sequência $\{t_k\}$. Além disso, pela definição da busca de Armijo, a sequência gerada pelo método de máxima descida satisfaz

$$f(x^{k+1}) + \alpha t_k^2 \|\text{grad}f(x^k)\|^2 \leq f(x^k), \quad \forall k \geq 0,$$

que é uma desigualdade importante para a convergência do método, conforme veremos na análise de convergência.

A seguir, apresentamos um resultado que mostra que o Algoritmo 3.4 satisfaz a desigualdade acima sem a necessidade de nenhum procedimento adicional, como por exemplo a busca de Armijo. A proposição seguinte é a motivação da escolha do passo como em (3.31) ao invés da busca de Armijo.

Proposição 3.2.2 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.4. Então, existe uma constante $\beta > 0$ tal que*

$$f(x^{k+1}) + \beta t_k^2 \|\text{grad}f(x^k)\|^2 \leq f(x^k), \quad \forall k \geq 0. \quad (3.33)$$

Em particular, vale um dos seguintes itens:

1. $\{x^k\}$ para em um ponto crítico;
2. a sequência $\{f(x^k)\}$ é estritamente decrescente.

Demonstração: Veja [43, Teorema 5.1]. ■

Observação 3.2.3 *De uma forma mais geral, a proposição anterior foi provada em [43] apenas supondo f continuamente diferenciável com gradiente Lipschitz e sem nenhuma hipótese na curvatura da variedade. Porém, para obter os resultados de convergência parcial (no sentido de que pontos de acumulação de $\{x^k\}$ são pontos críticos de f) os autores assumem que os conjuntos de níveis de f são compactos. Substituindo essa hipótese pela convexidade de f juntamente com a não negatividade da curvatura da variedade, eles provam a convergência da sequência para um minimizador de f .*

A seguir, apresentamos um resultado técnico que será usado na análise de convergência do método.

Lema 3.2.1 *Seja $\{a_k\}$ uma sequência de números reais positivos tal que*

$$\sum_{k=1}^{+\infty} a_k^2/a_{k-1} < +\infty.$$

Então, $\sum_{k=1}^{+\infty} a_k < +\infty$.

Demonstração: Veja [22, Lema 4.1]. ■

3.2.1 Análise de convergência

Recentemente, Bento et al. [22] provaram a convergência do método do ponto proximal para variedades de Riemann sem nenhuma hipótese sobre a curvatura desde que a função objetivo satisfaça a propriedade de Kurdyka-Lojasiewicz. Nesta seção substituiremos a hipótese de convexidade da função objetivo analisada em [43] para convergência do método de máxima descida em uma variedade de Riemann com curvatura não negativa pela propriedade de Kurdyka-Lojasiewicz para analisar a convergência do Algoritmo 3.4 em variedades de Hadamard.

Dizemos que uma função semicontínua inferiormente $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$ satisfaz a *propriedade de Kurdyka-Lojasiewicz* no ponto $\bar{x} \in \text{dom } \partial f$ se existem $\eta \in (0, +\infty]$, uma vizinhança U de \bar{x} e uma função contínua e côncava $\varphi : [0, \eta) \rightarrow \mathbb{R}_+$ tal que

$$\varphi(0) = 0, \quad \varphi \in C^1(0, \eta), \quad \varphi'(s) > 0, \quad s \in (0, \eta); \quad (3.34)$$

$$\varphi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1, \quad x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta], \quad (3.35)$$

onde $\text{dist}(0, \partial f(x)) := \inf\{\|v\| : v \in \partial f(x)\}$ e $[\eta_1 < f < \eta_2] := \{x \in M : \eta_1 < f(x) < \eta_2\}$.

Esse conceito foi introduzido por Łojasiewicz [93] para funções reais analíticas e depois estendido por Kurdyka [79] para funções diferenciáveis definidas em uma estrutura o-minimal (veja mais detalhes sobre estrutura o-minimal, por exemplo em Dries e Miller [130]). Para extensões desse conceito para funções não diferenciáveis no contexto Euclidiano veja Bolte et al. [28], Bolte et al. [29] e Attouch et al. [8]. Extensão desse conceito em espaços não lineares podem ser encontradas em Kurdyka et al.[80], Lageman [81], Bolte et al. [31], Cruz Neto et al. [46].

No decorrer desta seção os termos φ , η e U são os mesmo que aparecem na definição acima. Além disso, nos resultados que seguiremos iremos supor que a sequência $\{x^k\}$ gerada pelo Algoritmo 3.4 não satisfaz o critério de parada, ou seja, $\{x^k\}$ é uma sequência de infinitos termos tais que $x^{k+1} \neq x^k$ e $\text{grad } f(x^k) \neq 0$ para todo $k \geq 0$, pois caso contrário os resultados são triviais.

Proposição 3.2.3 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.4. Suponha que $\{x^k\}$ tenha um ponto de acumulação $\tilde{x} \in M$ e f satisfaz a propriedade de Kurdyka-Łojasiewicz no ponto \tilde{x} . Então, dada uma constante $c > 0$, para todo $\rho > 0$ existe $k_0 \in \mathbb{N}$ tal que*

$$f(\tilde{x}) < f(x^k) < f(\tilde{x}) + \eta, \quad \forall k \geq k_0 \quad (3.36)$$

e

$$d(x^{k_0}, \tilde{x}) + 2d(x^{k_0+1}, x^{k_0}) + c\varphi(f(x^{k_0}) - f(\tilde{x})) < \rho. \quad (3.37)$$

Demonstração: Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ convergindo para \tilde{x} . Segue da continuidade de f que $\{f(x^{k_j})\}$ converge para $f(\tilde{x})$. Além disso, segue de (3.33) que

$$f(x^{k+1}) + \beta\delta_1^2 \|\text{grad } f(x^k)\|^2 \leq f(x^k), \quad \forall k \geq 0, \quad (3.38)$$

onde usamos que $t_k > \delta_1$. Daí, segue que $\{f(x^k)\}$ é uma sequência estritamente decrescente e sendo o conjunto solução de (3.29) não vazio, temos que $\{f(x^k)\}$ é convergente e, como $f(x^{k_j}) \rightarrow f(\tilde{x})$ quando $j \rightarrow +\infty$, temos que $f(x^k) \rightarrow f(\tilde{x})$ quando $k \rightarrow +\infty$. Logo, para todo $\eta > 0$, existe $k_0 \in \mathbb{N}$ tal que

$$f(\tilde{x}) < f(x^k) < f(\tilde{x}) + \eta, \quad \forall k \geq k_0.$$

Agora, combinando (3.30) com (3.38), obtemos

$$2d(x^{k+1}, x^k) < \frac{2}{\delta_1\sqrt{\beta}} \sqrt{f(x^k) - f(x^{k+1})}, \quad \forall k \geq 0.$$

Sendo $\{f(x^k)\}$ convergente, segue da desigualdade acima que $d(x^{k+1}, x^k) \rightarrow 0$ quando $k \rightarrow +\infty$. Com isso, levando em consideração que φ é contínua, temos

que

$$d(x^{k_j}, \tilde{x}) + 2d(x^{k_j+1}, x^{k_j}) + c\varphi(f(x^{k_j}) - f(\tilde{x})) \rightarrow 0,$$

quando $j \rightarrow +\infty$. Portanto, podemos assumir (tomando o máximo dos índices se necessário) que

$$d(x^{k_0}, \tilde{x}) + 2d(x^{k_0+1}, x^{k_0}) + c\varphi(f(x^{k_0}) - f(\tilde{x})) < \rho,$$

e isso conclui a demonstração. ■

Proposição 3.2.4 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.4. Se as hipóteses da Proposição 3.2.3 são satisfeitas e existem $k_0 \in \mathbb{N}$ e $\epsilon > 0$ tal que $x^{k_0} \in B(\tilde{x}, \epsilon)$, então existem $\alpha, \beta > 0$ tais que*

$$\frac{d^2(x^{k_0+1}, x^{k_0})}{d(x^{k_0}, x^{k_0-1})} \leq \frac{\alpha}{\beta} [\varphi(f(x^{k_0}) - f(\tilde{x})) - \varphi(f(x^{k_0+1}) - f(\tilde{x}))]. \quad (3.39)$$

Demonstração: Assuma, sem perda de generalidade, que $k_0 \in \mathbb{N}$ é tal que (3.36) se verifica e $\epsilon > 0$ é suficientemente pequeno tal que $B(\tilde{x}, \epsilon) \subset U$. Assim,

$$x^{k_0} \in U \cap [f(\tilde{x}) < f < f(\tilde{x}) + \eta].$$

Como f satisfaz a propriedade de Kurdyka-Lojasiewicz em \tilde{x} , temos que

$$\varphi'(f(x^{k_0}) - f(\tilde{x})) \geq \frac{1}{\|\text{grad}f(x^{k_0})\|}. \quad (3.40)$$

Como φ é côncava e $\varphi' > 0$, lembrando que $f(x^{k_0+1}) \leq f(x^{k_0})$, obtemos

$$\begin{aligned} \varphi(f(x^{k_0}) - f(\tilde{x})) - \varphi(f(x^{k_0+1}) - f(\tilde{x})) &\geq \varphi'(f(x^{k_0}) - f(\tilde{x}))(f(x^{k_0}) - f(x^{k_0+1})) \\ &\geq \beta t_{k_0}^2 \frac{\|\text{grad}f(x^{k_0})\|^2}{\|\text{grad}f(x^{k_0})\|} \\ &\geq \beta \frac{d^2(x^{k_0+1}, x^{k_0})}{\|\text{grad}f(x^{k_0})\|}, \end{aligned} \quad (3.41)$$

onde a segunda desigualdade segue de (3.40) e (3.33), e a terceira vem de (3.30). Além disso, sendo a aplicação $x \mapsto \text{grad}f(x)$ Lipschitz contínua e o transporte paralelo uma isometria, segue da desigualdade triangular que

$$\begin{aligned} \|\text{grad}f(x^{k_0})\| &\leq \|\text{grad}f(x^{k_0}) - P_{x^{k_0-1}, x^{k_0}} \text{grad}f(x^{k_0-1})\| \\ &\quad + \|P_{x^{k_0-1}, x^{k_0}} \text{grad}f(x^{k_0-1})\| \\ &\leq \left(L + \frac{1}{\delta_1}\right) d(x^{k_0}, x^{k_0-1}). \end{aligned} \quad (3.42)$$

Portanto, o resultado desejado segue combinando as inequações (3.42) e (3.41), com $\alpha = (L + \frac{1}{\delta_1})$. \blacksquare

Proposição 3.2.5 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.4. Se as hipóteses da Proposição 3.2.3 são satisfeitas. Então, existem $k_0 \in \mathbb{N}$ e $\varepsilon > 0$ tais que*

$$x^k \in B(\tilde{x}, \varepsilon), \quad \forall k \geq k_0. \quad (3.43)$$

Demonstração: Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ convergindo para \tilde{x} e tome $\varepsilon > 0$ suficientemente pequeno tal que $B(\tilde{x}, \varepsilon) \subset U$. Segue da Proposição 3.2.3, para $\rho = \varepsilon$, que existe $k_0 \in \mathbb{N}$ tal que (3.37) se verifica. Provaremos (3.43) por indução sobre k . Para $k = k_0$, a afirmação segue diretamente de (3.37). Agora, suponha que $x^k \in B(\tilde{x}, \varepsilon)$, para todo $k = k_0 + 1, \dots, k_0 + j - 1$. Então, pela Proposição 3.2.4, obtemos

$$\sqrt{d(x^k, x^{k-1})(\alpha/\beta)[\varphi(f(x^k) - f(\tilde{x})) - \varphi(f(x^{k+1}) - f(\tilde{x}))]} \geq d(x^{k+1}, x^k), \quad (3.44)$$

para $k = k_0 + 1, \dots, k_0 + j - 1$. Note que, para quaisquer números reais $r, s \geq 0$, temos que $r + s \geq 2\sqrt{rs}$. Assim, para $k = k_0 + 1, \dots, k_0 + j - 1$, tomando

$$r = d(x^k, x^{k-1}) \text{ e } s = (\alpha/\beta)[\varphi(f(x^k) - f(\tilde{x})) - \varphi(f(x^{k+1}) - f(\tilde{x}))],$$

pela desigualdade (3.44), segue que

$$2d(x^{k+1}, x^k) \leq d(x^k, x^{k-1}) + \frac{\alpha}{\beta}[\varphi(f(x^k) - f(\tilde{x})) - \varphi(f(x^{k+1}) - f(\tilde{x}))],$$

para $k = k_0 + 1, \dots, k_0 + j - 1$. Somando a última desigualdade de $k = k_0 + 1$ até $k = k_0 + j - 1$, obtemos

$$\begin{aligned} \sum_{i=k_0+1}^{k_0+j-1} d(x^{i+1}, x^i) + d(x^{k_0+j}, x^{k_0+j-1}) &\leq d(x^{k_0+1}, x^{k_0}) \\ &+ \frac{\alpha}{\beta}[\varphi(f(x^{k_0+1}) - f(\tilde{x}))] \\ &- \frac{\alpha}{\beta}[\varphi(f(x^{k_0+j}) - f(\tilde{x}))], \end{aligned}$$

que nos leva a concluir que

$$\sum_{i=k_0+1}^{k_0+j-1} d(x^{i+1}, x^i) \leq d(x^{k_0+1}, x^{k_0}) + \frac{\alpha}{\beta}\varphi(f(x^{k_0}) - f(\tilde{x})) \quad (3.45)$$

porque φ é crescente, $f(x^{k+1}) \leq f(x^k)$, para todo $k \geq 0$, e $d(x, y) \geq 0$ para todo

$x, y \in M$. Agora, usando a desigualdade triangular, temos que

$$\begin{aligned} d(x^{k_0+j}, \tilde{x}) &\leq d(x^{k_0+j}, x^{k_0}) + d(x^{k_0}, \tilde{x}) \\ &\leq d(x^{k_0}, \tilde{x}) + d(x^{k_0+1}, x^{k_0}) + \sum_{i=k_0+1}^{k_0+j-1} d(x^{i+1}, x^i), \end{aligned}$$

que, combinado com (3.45), obtemos

$$d(x^{k_0+j}, \tilde{x}) \leq d(x^{k_0}, \tilde{x}) + 2d(x^{k_0+1}, x^{k_0}) + \frac{\alpha}{\beta} \varphi(f(x^{k_0}) - f(\tilde{x})) < \epsilon,$$

onde a desigualdade do lado direito vem da Proposição 3.2.3 para $\rho = \epsilon$. Portanto, concluímos que $x^{k_0+j} \in B(\tilde{x}, \epsilon)$, e a indução está completa. ■

Proposição 3.2.6 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.4. Se as hipóteses da Proposição 3.2.3 são satisfeitas, então $\sum_{k=0}^{+\infty} d(x^{k+1}, x^k) < +\infty$. Em particular, $\lim_{k \rightarrow +\infty} d(x^{k+1}, x^k) = 0$.*

Demonstração: Tome $k_0, N \in \mathbb{N}$ números naturais tais que $N > k_0$ e (3.43) se verifica. Então, combinando a Proposição 3.2.4 com a Proposição 3.2.5, temos que

$$\frac{d^2(x^{k+1}, x^k)}{d(x^k, x^{k-1})} \leq \frac{\alpha}{\beta} [\varphi(f(x^k) - f(\tilde{x})) - \varphi(f(x^{k+1}) - f(\tilde{x}))], \quad \forall k \geq k_0, \quad (3.46)$$

que implica em

$$\sum_{i=k_0+1}^N \frac{d^2(x^{i+1}, x^i)}{d(x^i, x^{i-1})} \leq \frac{\alpha}{\beta} [\varphi(f(x^{k_0+1}) - f(\tilde{x}))], \quad (3.47)$$

porque $\varphi' > 0$ e $f(x^k) \leq f(\tilde{x})$, para todo $k \geq 0$. Portanto, o resultado desejado segue fazendo $k \rightarrow +\infty$ em (3.47) e aplicando o Lema 3.2.1. ■

A maioria dos resultados de convergência de métodos numéricos em minimização convexa são obtidos supondo a existência de um minimizador local (e com isso global) da função objetivo. No presente cenário não-convexo, a existência de um minimizador local não implica a existência de um minimizador global mesmo se a função em consideração for limitada inferiormente. Dessa forma, não é esperado se obter limitação da sequência das iteradas.

Teorema 3.2.1 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.4. Assuma que $\{x^k\}$ tem um ponto de acumulação $\tilde{x} \in M$ e f satisfaz a propriedade de Kurdyka-Lojasiewicz em \tilde{x} . Então, $\lim_{k \rightarrow +\infty} f(x^k) = f(\tilde{x})$ e $\{x^k\}$ converge para \tilde{x} que é um ponto crítico de f .*

Demonstração: Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ convergindo para \tilde{x} . Da Proposição 3.2.6 segue que $\{x^k\}$ é uma sequência de Cauchy, e com isso, $\{x^k\}$ converge para \tilde{x} quando $k \rightarrow +\infty$. Sendo f contínua, temos que $f(x^k)$ converge para $f(\tilde{x})$ quando $k \rightarrow +\infty$. Além disso, de (3.30), temos que

$$t_k \|\text{grad}f(x^k)\| = d(x^{k+1}, x^k), \quad \forall k \geq 0.$$

Aplicando o limite com $k \rightarrow +\infty$ na igualdade acima obtemos

$$\|\text{grad}f(\tilde{x})\| = 0,$$

uma vez que $\liminf_{k \rightarrow +\infty} t_k > 0$ e $\lim_{k \rightarrow +\infty} d(x^{k+1}, x^k) = 0$. Portanto, \tilde{x} é um ponto crítico de f e a demonstração está concluída. ■

3.2.2 Aplicação: centro de massa Riemanniano

O *centro de massa Riemanniano* (global) ou *média Riemanniana* ou ainda *média de Fréchet* de um conjunto de pontos $\{a_i\}_{i=1}^n$ dados em uma variedade de Riemann M é definida como o conjunto de pontos que minimiza a soma dos quadrados das distâncias aos pontos dados. Esse conceito e suas variações tem uma longa história de aplicações em matemática pura. Recentemente, análise estatística em variedades de Riemann e, em particular, o centro de massa Riemanniano tem sido aplicado em diversas áreas como visão computacional, análise estatística de formas, imagens, redes de sensores e várias outras aplicações de análise de dados, veja Afsari et al. [4] e suas referências.

Esse problema também aparece na literatura em uma forma mais geral chamado *L^p -centro de massa Riemanniano* que considera o problema de calcular o L^p -centro de massa de um conjunto de dados $\{a_i\}_{i=1}^n \subset M$ com relação aos pesos $0 \leq w_i \leq 1$ tal que $\sum_{i=1}^n w_i = 1$. Nesta seção estudaremos o problema de encontrar o L^p -centro de massa Riemanniano em uma variedade de Hadamard de dimensão finita M definido como a solução do seguinte problema:

$$\min_{x \in M} \frac{1}{p} \sum_{i=1}^n w_i d^p(x, a_i), \quad (3.48)$$

para $1 \leq p < \infty$. Se $p = \infty$, o centro de massa é definido como o minimizador de $\max_i d(x, a_i)$ em M . O foco do nosso estudo será limitado aos casos $p = 1$ e $p = 2$ que são os mais comuns na literatura. Normalmente quando nos referimos ao centro de massa de um conjunto de dados não fazemos referência explícita aos pesos a menos que seja necessário. Dizemos que os pontos $a_1, \dots, a_n \in M$ são *colineares* se eles pertencem à mesma geodésica, ou seja, existe $y \in M$, $d \in T_y M$ e $t_i \in \mathbb{R}$,

$i = 1, \dots, n$, tal que $a_i = \exp_y t_i d$, para cada $i = 1, \dots, n$.

Dados a_i e w_i , $i = 1, \dots, n$ como em (3.48) definimos a função $f_p : M \rightarrow \mathbb{R}$ como

$$f_p(x) = \frac{1}{p} \sum_{i=1}^n w_i d^p(x, a_i). \quad (3.49)$$

Verifica-se que

$$\text{grad } f_p(x) = - \sum_{i=1}^n w_i d^{p-2}(x, a_i) \exp_x^{-1} a_i, \quad (3.50)$$

para todo $x \in M$ desde que x não pertença ao “cut locus” de qualquer um dos pontos do conjunto de dados; veja por exemplo [4]. A definição de “cut locus” pode ser encontrada por exemplo em [114, página 102] ou [38, página 295].

Considere $x, y \in M$ (y distinto de x) e $t \mapsto \gamma(t)$ com $\gamma(0) = x$ a geodésica normalizada que faz em x um ângulo β com a geodésica minimal ligando y a x . Segue de [114, Lema 2.9] que

$$\frac{\sin^2 \beta}{d(x, y)} \leq \frac{d^2}{dt^2} d(\gamma(t), y)|_{t=0}. \quad (3.51)$$

Antes de apresentarmos o algoritmo para calcular o centro de massa Riemanniano de um conjunto de dados veremos algumas propriedades importantes das funções f_1 e f_2 e seus respectivos problemas de minimização.

Proposição 3.2.7 *Hess $f_1(x)$ é semi-positiva definida para todo $x \in M$ tal que $x \neq a_i$, $i = 1, \dots, n$. Além disso, se os pontos a_1, \dots, a_n não são colineares, então Hess $f_1(x)$ é positiva definida para todo $x \in M$ tal que $x \neq a_i$, $i = 1, \dots, n$.*

Demonstração: Dado $x \in M$ tal que $x \neq a_i$, $i = 1, \dots, n$. Seja $t \mapsto \gamma(t)$ a geodésica tal que $\gamma(0) = x$ e $\gamma'(0) = v$, e β_i o ângulo que essa geodésica forma com a geodésica que liga os pontos x e a_i . De (3.51), temos que

$$\sum_{i=1}^n \frac{\sin^2 \beta_i}{d(x, a_i)} \leq \sum_{i=1}^n \frac{d^2}{dt^2} d(\gamma(t), a_i)|_{t=0}. \quad (3.52)$$

Assim, denotando por

$$\alpha = \frac{1}{\max_{1 \leq i \leq n} d(x, a_i)} \sum_{i=1}^n \sin^2 \beta_i, \quad (3.53)$$

temos de (3.52) que

$$\frac{d^2}{dt^2} (f_1 \circ \gamma)(t)|_{t=0} \geq \alpha \geq 0. \quad (3.54)$$

Agora, se os pontos a_1, \dots, a_n são não colineares, então existe pelo menos um ponto, digamos a_q , $q \in \{1, \dots, n\}$, tal que $\sin^2 \beta_q > 0$ que implica por (3.53) que $\alpha > 0$. ■

Dizemos que uma função de classe C^2 é de *Morse* se todos os seus pontos críticos são não degenerados, isto é, se hessiana de f tem todos os seus autovalores diferentes de zero. Sabemos que toda função de Morse satisfaz a propriedade de Kurdyka-Lojasiewicz em todos os pontos de seu domínio; veja [46, Teorema 3.8].

Corolário 3.2.1 *Se os pontos a_1, \dots, a_n não são colineares e existe $C \subset M$ tal que $a_i \notin C$, para todo $i = 1, \dots, n$, então a função $f_1 : M \rightarrow \mathbb{R}$ dada por $f_1(x) = \sum_{i=1}^n w_i d(x, a_i)$ satisfaz a propriedade de Kurdyka-Lojasiewicz em todos os pontos de C .*

Demonstração: Como $a_i \notin C$, para todo $i = 1, \dots, n$, então f_1 é de classe C^∞ em C ; veja [129, página 106]. Segue da segunda parte da Proposição 3.2.7 que os autovalores de $Hess f_1(x)$ são todos positivos para todo $x \in C$. Com isso, f_1 é uma função de Morse, donde por [46, Teorema 3.8], temos que f_1 satisfaz a propriedade de Kurdyka-Lojasiewicz em todos os pontos de C . ■

Proposição 3.2.8 *As seguintes afirmações se verificam:*

- (a) *A função $f_1(x) = \sum_{i=1}^n w_i d(x, a_i)$ é convexa;*
- (b) *O problema (3.48), para $p = 1$, sempre tem uma solução;*
- (c) *A solução do problema (3.48), para $p = 1$, é única se os pontos a_1, \dots, a_n são não colineares;*
- (d) *Determine um índice $p \in \{1, \dots, n\}$ tal que $f_1(a_p) = \min_{i=1, \dots, n} f_1(a_i)$. Então, a_p é um minimizador de f_1 em M se, e somente se,*

$$\left\| \sum_{i=1, i \neq p}^n w_i \frac{\exp^{-1} a_p}{d(a_i, a_p)} \right\| \leq w_p.$$

Demonstração: (a) A convexidade de $x \mapsto w_i d(x, a_i)$ está provado em [129, página 106]. Dessa forma, a convexidade de f_1 segue da propriedade que a soma de funções convexas é convexa; veja [129, página 67].

(b) Sendo f_1 uma função contínua e coerciva, então temos que f_1 tem um minimizador global;

(c) Para provar esse item basta notar que se os pontos a_1, \dots, a_n são não colineares, então pela segunda parte da Proposição 3.2.7 existe pelo menos um ponto, digamos a_q , tal que a função $x \mapsto d(x, a_q)$ é estritamente convexa. Logo f_1 é estritamente convexa;

(d) Como f_1 é convexa, temos que um ponto a_p é minimizador de f_1 se, e somente

se, $0 \in \partial f_1(a_p)$. Por outro lado, combinando (3.50) com [89, Lemma 5.2], obtemos que

$$\partial f_1(a_p) = \sum_{i=1, i \neq p}^n w_i \frac{\exp_{a_i}^{-1} a_p}{d(a_i, a_p)} + w_p \mathbb{B}_{a_p},$$

onde \mathbb{B}_{a_p} denota a bola unitária com centro em a_p . Com isso, a_p é um minimizador de f_1 se, e somente se, existe $\xi_p \in \mathbb{B}_{a_p}$ tal que

$$\sum_{i=1, i \neq p}^n w_i \frac{\exp_{a_i}^{-1} a_p}{d(a_i, a_p)} = -w_p \xi_p,$$

e a afirmação segue tomando a norma nos dois lados da igualdade acima levando em consideração que $\|\xi_p\| \leq 1$. ■

Proposição 3.2.9 *Seja C um conjunto compacto tal que $a_i \notin C$ para cada $i = 1, \dots, n$. Então o campo $\text{grad } f_1 : M \rightarrow TM$ é Lipschitz contínuo em C .*

Demonstração: Assuma que $X_i(z) = \frac{\partial d}{\partial x}(z, a_i)$ é o vetor (unitário) tangente do único segmento geodésico minimizante ligando os pontos z e a_i , para cada $i = 1, \dots, n$. Seguindo os passos da demonstração de [68, Proposition 4.1] podemos mostrar que, para todo $x \in C$, o campo X_i é Lipschitz contínuo em $B(x, r_x)$ com constante L_x . Sendo C compacto, pelo Teorema de Hopf-Rinow, temos que o campo X_i é Lipschitz contínuo em C . ■

A seguir, apresentamos algumas propriedades da função f_2 .

Proposição 3.2.10 *As seguintes afirmações se verificam:*

(a) *A função $f_2(x) = \frac{1}{2} \sum_{i=1}^n w_i d^2(x, a_i)$ é estritamente convexa e continuamente diferenciável com seu gradiente Lipschitz em conjuntos compactos;*

(b) *O problema (3.48) sempre tem única solução para o caso $p = 2$.*

Demonstração: (a) A convexidade estrita bem como o fato de f_2 ser continuamente diferenciável está provado em [129, página 111]. A Lipschitz continuidade local de $\text{grad } f_2(\cdot)$ segue de [16, Theorem 3.4] combinado com [16, Example 3.1]. Dessa forma, pelo Teorema de Hopf-Rinow, temos que $\text{grad } f_2(\cdot)$ é Lipschitz contínuo em conjuntos compactos.

(b) Sendo f_2 contínua e coerciva, então f_2 possui minimizador. A unicidade segue da convexidade estrita provada no item anterior. ■

É conhecido que uma variedade de Hadamard é um exemplo de um espaço CAT(0) ou espaço de Hadamard; veja [13, página 14]. Um espaço CAT(0) é caracterizado como um par (M, d) tal que para todo $x, y \in M$, existe uma geodésica

$\gamma : [0, 1] \rightarrow M$ tal que $\gamma(0) = x$ e $\gamma(1) = y$ satisfazendo

$$d^2(\gamma(t), z) \leq (1-t)d^2(x, z) + td^2(y, z) - t(1-t)d^2(x, y), \quad (3.55)$$

para todo $z = \gamma(t)$ com $t \in [0, 1]$; veja [13, página 9].

Como em uma variedade de Hadamard $\|\gamma'(t)\|$ é constante com respeito a t e $d(x, y) = \|\exp_x^{-1} y\|$, então (3.55) pode ser escrito como

$$d^2(\gamma(t), z) \leq (1-t)d^2(x, z) + td^2(y, z) - t(1-t)\|\gamma'(t)\|^2.$$

Essa última desigualdade significa que a função $d^2(\cdot, z)$ é fortemente convexa. Como a soma finita de funções fortemente convexas ainda é uma função fortemente convexa podemos enunciar o seguinte resultado:

Proposição 3.2.11 *A função $f_2(x) = \sum_{i=1}^n w_i d^2(x, a_i)$ satisfaz a propriedade de Kurdyka-Lojasiewicz em todo ponto de M .*

Demonstração: Como toda função fortemente convexa é estritamente convexa (veja [129, página 188]), temos que toda função fortemente convexa é uma função de Morse. Logo, segue de [46, Teorema 3.8], que f_2 satisfaz a propriedade de Kurdyka-Lojasiewicz em M . ■

A seguir, apresentamos uma versão do método de máxima descida para calcular o centro de massa Riemanniano de um conjunto de dados $\{a_i\}_{i=1}^n \subset M$. Dessa forma, denotaremos o índice $p \in \{1, \dots, n\}$ tal que $f(a_p) = \min_{i=1, \dots, n} f(a_i)$, onde $f = f_1$ ou $f = f_2$. Iremos assumir que a_p não é o centro de massa Riemanniano.

Algoritmo 3.5 *Passo 1: Determine $d_p \in T_{a_p}M$ e $t_p > 0$ suficientemente pequeno tal que $f(\exp_{a_p} t_p d_p) < f(a_p)$, e tome $x^0 := \exp_{a_p} t_p d_p$;*

Passo 2: Dado $x^k \in M$, se x^k é um ponto crítico de f , então faça $x^{k+p} = x^k$ para todo $p \in \mathbb{N}$;

Passo 3: Caso contrário, tome como próxima iterada $x^{k+1} \in M$ tal que

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad} f(x^k)), \quad (3.56)$$

onde $\{t_k\}$ é como em (3.31) para $f = f_1$ ou $f = f_2$.

Observação 3.2.4 *Denotando por $L^f(x^0) := \{x \in M : f(x) \leq f(x^0)\}$ notamos que, se $\{x^k\}$ é uma sequência gerada pelo Algoritmo 3.5, então $x^k \in L^f(x^0)$, para todo $k \in \mathbb{N}$ e $a_i \notin L^f(x^0)$, para todo $i = 1, \dots, n$. Além disso, podemos verificar que $L^f(x^0)$ é não vazio e compacto para $f = f_1$ e $f = f_2$. Dessa forma, a Proposição 3.2.9 se verifica para $C = L^f(x^0)$ bem como o item (a) da Proposição 3.2.10 no mesmo conjunto. Isso garante a boa definição do Algoritmo 3.5.*

A seguir apresentamos os resultados de convergência para calcular L^1 e L^2 -centro de massa Riemanniano via Algoritmo 3.5.

Teorema 3.2.2 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.5 para $f = f_2$. Então $\{x^k\}$ converge para o Riemannian L^2 -centro de massa Riemanniano do conjunto de dados $\{a_i\}_{i=1}^n$.*

Demonstração: Como $x^k \in L^{f_2}(x^0)$, para todo $k \in \mathbb{N}$ e $L^{f_2}(x^0)$ é um conjunto compacto, temos que $\{x^k\}$ é uma sequência limitada. Assim, considere \hat{x} um ponto de acumulação de $\{x^k\}$. Pela Proposição 3.2.11, temos que f_2 satisfaz a propriedade de Kurdyka-Lojasiewicz em \hat{x} . Logo, pelo Theorem 3.2.1, obtemos que $\{x^k\}$ converge para \hat{x} que é um ponto crítico de f_2 . Pelo item (a) da Proposição 3.2.10, temos que f_2 é estritamente convexa que significa \hat{x} ser (único) minimizador de f_2 e a prova está completa. ■

Teorema 3.2.3 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.5 para $f = f_1$. Então $\{x^k\}$ converge para o Riemannian L^1 -centro de massa Riemanniano do conjunto de dados $\{a_i\}_{i=1}^n$ desde que os pontos a_i , para $i = 1, \dots, n$, sejam não colineares.*

Demonstração: Similar ao Teorema 3.2.2 tendo em vista o Corolário 3.2.1 e a Proposição 3.2.8. ■

Capítulo 4

MPP generalizado para funções DC

Há uma vasta literatura sobre a classe de funções que podem ser escritas como diferença de funções convexas, ou abreviadamente, funções DC. Essa importante subclasse de funções localmente Lipschitz desperta o interesse tanto da matemática pura como da matemática aplicada. Essa classe surge naturalmente como o menor subespaço vetorial contendo todas as funções contínuas e convexas definidas em um dado conjunto. Além disso, a classe das funções DC é densa no conjunto das funções contínuas definidas num conjunto convexo e compacto X (com a topologia da convergência uniforme em X). Neste capítulo estudaremos o método do ponto proximal em que a norma Euclideana é substituída por uma aplicação “*like-distance*” ou distância generalizada, ou seja, uma função que não necessariamente satisfaz todos os axiomas da função distância, mas preserva boas propriedades como continuidade, coercividade e, em alguns casos, convexidade. Esse tipo de método é normalmente denominado na literatura como *algoritmos generalizados* ou *não lineares*.

4.1 Quase distância

Há um próspero campo de pesquisa em algoritmos generalizados. Métodos proximais usando regularizações não lineares foram estudados por exemplo em Auslender et al. [10], Bento e Soubeyran [26, 27], Burachik e Svaiter [35], Chen e Teboulle [41], Censor e Zenios [40], Eckstein [51], Kiwiel [76], Moreno et al. [101], Pan e Chen [106], Teboulle [127]. Neste trabalho o termo “*like-distance*” ou “generalizado” diz respeito a uma aplicação definida da seguinte forma: dado um conjunto X não vazio, uma aplicação $q : X \times X \rightarrow \mathbb{R}_+$ é chamada de *quase distância* se satisfaz

1. Para todo $x, y \in X$, $q(x, y) = q(y, x) = 0 \Leftrightarrow x = y$;
2. Para todo $x, y, z \in X$, $q(x, z) \leq q(x, y) + q(y, z)$.

O par (X, d) é chamado de *espaço quase métrico*. Um espaço quase métrico cuja quase distância satisfaz a propriedade simétrica $q(x, y) = q(y, x)$ é um espaço métrico. Quase distâncias são aplicações continuamente diferenciáveis e coercivas, mas não necessariamente convexas. Alguns exemplos de quase distâncias podem ser encontrados em [101]. Esse tipo de “*like-distance*” é mais apropriado para aplicações em Teoria Comportamental, onde por exemplo o custo $C(x, y)$ de mudar de uma estratégia x para a estratégia y não necessariamente é o mesmo custo que voltar da estratégia y para x , ou seja, é possível que $C(x, y) \neq C(y, x)$; veja por exemplo [26, 27, 101].

Neste trabalho estamos considerando quase distâncias que satisfazem a seguinte condição:

Condition 1: Existem números reais positivos $\alpha, \beta > 0$ tais que

$$\alpha\|x - y\| \leq q(x, y) \leq \beta\|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \quad (4.1)$$

Observação 4.1.1 *Com a Condição 1 temos que as aplicações $q(\cdot, z)$ e $q(z, \cdot)$ são Lipschitz contínuas em \mathbb{R}^n e $q^2(\cdot, z)$ e $q^2(z, \cdot)$ são localmente Lipschitz em \mathbb{R}^n ; veja [101, Proposições 3.6 e 3.7]. Condição 1 tem sido usado em [26, 27, 101]. Em [101] alguns exemplos de quase distâncias que satisfazem a Condição 1 são apresentados.*

4.2 Algoritmo generalizado

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função DC própria e limitada inferiormente com componentes $g, h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ semicontínuas inferiormente, ou seja, $f(x) = g(x) - h(x)$, com g e h convexas. Nesta seção iremos considerar um algoritmo generalizado que, em cada iteração, lineariza a função f sem minimizá-la diretamente, enquanto minimiza a função g juntamente com a linearização de h e a regularização com uma quase distância. Esse tipo de método linearizado tem sido bastante usado em problemas envolvendo soma de funções convexas; veja [32, 65].

Algoritmo 4.1 *Passo 1: Tome um ponto inicial $x^0 \in \text{int dom}(h)$ e $\{\lambda_k\}$ uma sequência limitada de números reais positivos tal que $\liminf_{k \rightarrow \infty} \lambda_k > 0$.*

Passo 2: Tome

$$w^k \in \partial h(x^k). \quad (4.2)$$

Passo 3: Calcule

$$x^{k+1} \in \arg \min_{x \in \mathbb{R}^n} \left\{ g(x) - \langle w^k, x - x^k \rangle + \frac{1}{2\lambda_k} q^2(x^k, x) \right\}. \quad (4.3)$$

Se $x^{k+1} = x^k$, pare. Caso contrário, faça $k = k + 1$ e retorne ao Passo 2.

A boa definição de (4.2) e (4.3) é garantido pela convexidade das funções h , g e pela Condição 1. Note que quando $q(x, y) = \|x - y\|$ e $h \equiv 0$, então o Algoritmo 4.1 se reduz ao método proposto por Rockafellar [112] para resolver um problema de minimização convexa.

Observação 4.2.1 *Enfatizamos que o Algoritmo 4.1 é diferente do algoritmo DCA proposto em Pham e Souad [109]. Algoritmo 4.1 compartilha a mesma ideia que o algoritmo DCA, mais precisamente, uma aproximação linear de uma das componentes (ou das duas) da função DC $f(x) = g(x) - h(x)$. Porém o Algoritmo 4.1 é mais simples pois a linearização é feita diretamente e não na dual de uma das componentes com em [109]. Além disso, o método do ponto proximal é mais eficiente que o método subgradiente; veja Moudafi e Maingé [102]. Algoritmo 4.1 é bem similar ao método proposto em Sun et al. [126], mas algoritmos com uma regularização usando uma quase distância são mais apropriados para aplicações em Teoria de comportamento, onde o custo de mover de uma posição atual x^k para outra posição x^{k+1} e o custo para permanecer na posição x^k não necessariamente são simétrico e igual a zero, respectivamente.*

No Algoritmo 4.1, o processo iterativo (4.3) poderia ser escrito como

$$x^{k+1} \in \arg \min_{x \in \mathbb{R}^n} \varphi_k(x), \quad (4.4)$$

onde $\varphi_k(x) = g(x) - h(x^k) - \langle w^k, x - x^k \rangle + \frac{1}{2\lambda_k} q^2(x^k, x)$. Nesse caso, para cada $k \geq 0$, segue da convexidade de h que $f(x) \leq \varphi_k(x)$, para todo $x \in \text{dom}(f)$. Assim, em cada iteração (4.4) minimiza uma cota superior da função objetivo f decrescendo o valor de f ao longo da sequência. Esse tipo de método é denominado “*majorization-minimization*” (veja [83]) ou “*successive upper-bound minimization*” (veja [111]). Esse tipo de método tem sido usado com sucesso em diversos tipos de aplicações; veja [5, 49, 52, 61, 84, 96, 97].

A seguir apresentamos um resultado técnico que será usado na análise de convergência do algoritmo.

Proposição 4.2.1 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 4.1. Então existem $\xi^{k+1} \in \partial g(x^{k+1})$ e $\eta^{k+1} \in \partial_L(q(x^k, \cdot))(x^{k+1})$ tais que*

$$w^k = \xi^{k+1} + \frac{q(x^k, x^{k+1})}{\lambda_k} \eta^{k+1}. \quad (4.5)$$

Demonstração: Como em (4.3) x^{k+1} é solução de um problema de minimização,

segue da Proposição 2.1.2 que

$$\begin{aligned}
0 &\in \partial_L \left(g(\cdot) - \langle w^k, \cdot - x^k \rangle + \frac{1}{2\lambda_k} q^2(x^k, \cdot) \right) (x^{k+1}) \\
&\subset \partial_L (g(\cdot) - \langle w^k, \cdot - x^k \rangle) (x^{k+1}) + \frac{1}{2\lambda_k} \partial_L (q^2(x^k, \cdot)) (x^{k+1}) \\
&= \partial_L g(x^{k+1}) - w^k + \frac{1}{2\lambda_k} \partial_L (q^2(x^k, \cdot)) (x^{k+1}),
\end{aligned}$$

onde a segunda inclusão segue da Proposição 2.1.3. Logo, obtemos que

$$w^k \in \partial g(x^{k+1}) + \frac{1}{2\lambda_k} \partial_L (q^2(x^k, \cdot)) (x^{k+1}). \quad (4.6)$$

Aplicando a Proposição 2.1.4 com $f_1(x) = f_2(x) = q(x^k, x)$, temos que

$$\partial_L (q^2(x^k, \cdot)) (x^{k+1}) \subset 2q(x^k, x^{k+1}) \partial_L (q(x^k, \cdot)) (x^{k+1}).$$

Portanto, combinando a inclusão anterior com (4.6) obtemos

$$w^k \in \partial g(x^{k+1}) + \frac{q(x^k, x^{k+1})}{\lambda_k} \partial_L q(x^k, \cdot) (x^{k+1}),$$

e a demonstração está concluída. ■

4.3 Análise de convergência

Agora estabeleceremos os resultados de convergência do algoritmo. Iniciaremos mostrando que o método é de descida. Lembramos que, se f é uma função DC tal que $f(x) = g(x) - h(x)$, um ponto $x^* \in \text{dom}(f)$ é ponto crítico de f se, $\partial h(x^*) \cap \partial g(x^*) \neq \emptyset$.

Teorema 4.3.1 *A sequência $\{x^k\}$ gerada pelo Algoritmo 4.1 satisfaz um dos itens a seguir:*

1. o algoritmo para em um ponto crítico;
2. f decresce estritamente, ou seja, $f(x^{k+1}) < f(x^k)$, $\forall k \geq 0$.

Demonstração: Se $x^{k+1} = x^k$, então $q(x^k, x^{k+1}) = 0$ e segue de (4.5) que $w^k = \xi^{k+1}$, com $\xi^{k+1} \in \partial g(x^k)$ e $w^k \in \partial h(x^k)$. Logo, $\partial h(x^k) \cap \partial g(x^k) \neq \emptyset$, o que mostra que x^k é um ponto crítico de f . Agora, suponha que $x^{k+1} \neq x^k$. Segue de (4.2) e h ser convexa que

$$h(x^{k+1}) \geq h(x^k) + \langle w^k, x^{k+1} - x^k \rangle. \quad (4.7)$$

Por outro lado, segue de (4.3) que

$$g(x^k) \geq g(x^{k+1}) - \langle w^k, x^{k+1} - x^k \rangle + \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}). \quad (4.8)$$

Adicionando (4.7) e (4.8) obtemos

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) > f(x^{k+1}), \quad (4.9)$$

onde na desigualdade estrita usamos o fato que $\lambda_k > 0$ e $q^2(x^k, x^{k+1}) > 0$. \blacksquare

Observação 4.3.1 *Se f limitada inferiormente, segue diretamente do segundo item do teorema anterior que a sequência $\{f(x^k)\}$ é convergente.*

O resultado a seguir mostra que, no que diz respeito à quase distância q , iterações consecutivas estão próximas para uma quantidade de iterações suficientemente grandes.

Proposição 4.3.1 *Se $\{x^k\}$ é a sequência gerada pelo Algoritmo 4.1, então*

$$\sum_{k=0}^{\infty} q^2(x^k, x^{k+1}) < \infty.$$

Em particular, $\lim_{k \rightarrow +\infty} q(x^k, x^{k+1}) = 0$.

Demonstração: De (4.9), temos que

$$\frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}).$$

Somando a desigualdade acima com k variando de 0 até $n - 1$, temos

$$\sum_{k=0}^n \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n).$$

Como $\{\lambda_k\}$ é limitada, digamos $\lambda_k \leq \lambda_+$, e f é limitada inferiormente, temos que

$$\frac{1}{2\lambda_+} \sum_{k=0}^{n-1} q^2(x^k, x^{k+1}) \leq f(x^0).$$

Fazendo $n \rightarrow +\infty$ na desigualdade acima obtemos que $\sum_{k=0}^{\infty} q^2(x^k, x^{k+1})$ é convergente e, em particular, temos que $\lim_{k \rightarrow +\infty} q(x^k, x^{k+1}) = 0$. \blacksquare

Seja h uma função convexa. Combinando as Proposições 2.1.6 e 2.1.7, observamos que se $\{x^k\}$ é uma sequência limitada e $w^k \in \partial h(x^k)$, para todo $k \in \mathbb{N}$, então $\{w^k\}$ também é limitada. Usaremos esse fato no seguinte resultado de convergência do Algoritmo 4.1.

Teorema 4.3.2 *Se $\{x^k\}$ é a seqüência gerada pelo Algoritmo 4.1, então todo ponto de acumulação de $\{x^k\}$, caso exista, é ponto crítico de f .*

Demonstração: Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e tome $\{x^{k_j}\}$ uma subseqüência de $\{x^k\}$ convergindo para \hat{x} . Como $w^k \in \partial h(x^k)$, temos que $\{w^{k_j}\}$ também é limitada e podemos assumir sem perda de generalidade que $\{w^{k_j}\}$ converge para um ponto \hat{w} (podemos extrair uma outra subseqüência se necessário). Segue da definição do algoritmo que

$$0 \in \partial_F \left(g(\cdot) - \langle w^{k_j}, \cdot - x^{k_j} \rangle + \frac{1}{2\lambda_{k_j}} q^2(x^{k_j}, \cdot) \right) (x^{k_{j+1}}).$$

Usando a Proposição 2.1.5, com $f_1(x) = g(x) - \langle w^{k_j}, x - x^{k_j} \rangle$, $f_2(x) = \frac{1}{2\lambda_{k_j}} q^2(x^{k_j}, x)$ e $\gamma = \delta = \frac{1}{k_j}$, temos que

$$0 \in \partial_F(f_1 + f_2)(x^{k_{j+1}}) \subset A + \frac{1}{k_{j+1}} \overline{B}(0, 1), \quad (4.10)$$

onde $A = \{\partial_F f_1(a_1^{k_{j+1}}) + \partial_F f_2(a_2^{k_{j+1}}) : a_i^{k_{j+1}} \in \overline{B}(x^{k_{j+1}}, \frac{1}{k_{j+1}}), |f_i(a_i^{k_{j+1}}) - f_i(x^{k_{j+1}})| < \frac{1}{k_{j+1}}, i = 1, 2\}$. Note que $\partial_F f_1(a_1^{k_{j+1}}) = \partial g(a_1^{k_{j+1}}) - w^{k_{j+1}}$ e

$$\partial_F f_2(a_2^{k_{j+1}}) \subset \partial_L f_2(a_2^{k_{j+1}}) \subset \frac{1}{\lambda_{k_j}} q(x^{k_j}, a_2^{k_{j+1}}) \partial_L(q(x^{k_j}, \cdot))(a_2^{k_{j+1}}),$$

onde as inclusões na última desigualdade vem das Proposições 2.1.1 e 2.1.4, respectivamente. Usando esses fatos em (4.10), temos que

$$w^{k_j} \in \partial g(a_1^{k_{j+1}}) + \frac{q(x^{k_j}, a_2^{k_{j+1}})}{\lambda_{k_j}} \partial_L(q(x^{k_j}, \cdot))(a_2^{k_{j+1}}) + \frac{1}{k_{j+1}} \overline{B}(0, 1).$$

Com isso, podemos afirmar que existem subseqüências $\xi^{k_{j+1}} \in \partial g(a_1^{k_{j+1}})$, $\eta^{k_{j+1}} \in \partial_L(q(x^{k_j}, \cdot))(a_2^{k_{j+1}})$ e $u^{k_{j+1}} \in \overline{B}(0, 1)$ tais que

$$w^{k_j} = \xi^{k_{j+1}} + \frac{q(x^{k_j}, a_2^{k_{j+1}})}{\lambda_{k_j}} \eta^{k_{j+1}} + \frac{u^{k_{j+1}}}{k_{j+1}}, \quad (4.11)$$

onde $\{\eta^{k_j}\}$ e $\{u^{k_j}\}$ são limitadas, mais precisamente, $\|\eta^{k_{j+1}}\| \leq M$ e $\|u^{k_{j+1}}\| \leq 1$, para algum $M > 0$ e para todo $j \in \mathbb{N}$.

Provaremos agora que $\{\xi^{k_j}\}$ converge para \hat{w} . Combinando (4.11) com a desigualdade triangular da norma Euclidian, obtemos

$$0 \leq \|w^{k_j} - \xi^{k_{j+1}}\| \leq \frac{M}{\lambda_{k_j}} q(x^{k_j}, a_2^{k_{j+1}}) + \frac{1}{k_{j+1}}. \quad (4.12)$$

Agora, aplicando a desigualdade triangular na quase distância, temos que

$$\begin{aligned} q(x^{k_j}, a_2^{k_j+1}) &\leq q(x^{k_j}, x^{k_j+1}) + q(x^{k_j+1}, a_2^{k_j+1}) \\ &\leq q(x^{k_j}, x^{k_j+1}) + \frac{\beta}{k_j + 1}, \end{aligned} \quad (4.13)$$

usando que $a_2^{k_j+1} \in \overline{B}(x^{k_j+1}, \frac{1}{k_j+1})$ na segunda desigualdade. Combinando (4.12) e (4.13), levando em consideração que $\liminf_{k \rightarrow +\infty} \lambda_k > 0$ e $q(x^k, x^{k+1}) \rightarrow 0$ quando $k \rightarrow +\infty$, concluímos que

$$\lim_{j \rightarrow +\infty} w^{k_j} = \lim_{j \rightarrow +\infty} \xi^{k_j+1} = \hat{w}.$$

Sendo $a_1^{k_j+1} \in \overline{B}(x^{k_j+1}, \frac{1}{k_j+1})$, temos que $\lim_{j \rightarrow +\infty} a_1^{k_j+1} = \lim_{j \rightarrow +\infty} x^{k_j+1} = \hat{x}$. Com isso, usando o fato que $\xi^{k_j+1} \in \partial g(a_1^{k_j+1})$ e $w^{k_j} \in \partial h(x^{k_j})$, juntamente com o fato que ∂g e ∂h são conjuntos fechados, obtemos que $\hat{w} \in \partial h(\hat{x}) \cap \partial g(\hat{x})$, ou seja, \hat{x} é um ponto crítico de f . ■

4.4 Convergência global

Lidando com métodos de descida para funções convexas é esperado que os algoritmos obtenham convergência global da sequência. Quando as funções objetivo não são convexas (ou quase convexas) esses métodos podem ter comportamento oscilatório e com isso, resultados de convergência parcial são obtidos. Nesta seção iremos adicionar hipóteses razoáveis para obter um resultado de convergência global para o Algoritmo 4.1.

Teorema 4.4.1 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 4.1. Assuma que $\{x^k\}$ seja limitada, f seja contínua e satisfaça a propriedade de Kurdyka-Lojasiewicz em um ponto de acumulação \hat{x} de $\{x^k\}$ e que h seja diferenciável com $\nabla h(\cdot)$ Lipschitz com constante $L > 0$. Então, $\{x^k\}$ converge para \hat{x} que é ponto crítico de f .*

Demonstração: A prova segue diretamente de [9, Theorem 2.9] levando em consideração os seguintes fatos:

1. Como $\{\lambda_k\}$ é limitada, temos que $\lambda_k \leq \lambda_+$, for all k . Com isso, segue de (4.9) que

$$f(x^{k+1}) + \frac{1}{2\lambda_+} q^2(x^k, x^{k+1}) \leq f(x^k), \quad \forall k \geq 0.$$

Logo, segue da Condição 1 que

$$f(x^{k+1}) + a \|x^{k+1} - x^k\|^2 \leq f(x^k), \quad \forall k \geq 0,$$

onde $a = \frac{\alpha^2}{2\lambda_+} > 0$;

2. Segue de (4.5), que para todo $k \geq 0$

$$\|\xi^{k+1} - \nabla h(x^k)\| \leq q(x^k, x^{k+1}) \frac{\|\eta^{k+1}\|}{\lambda_k}, \quad (4.14)$$

onde $\xi^{k+1} \in \partial g(x^{k+1})$ e $\eta^{k+1} \in \partial_L(q(x^k, \cdot))(x^{k+1})$. Sendo $\{x^k\}$ limitada, temos que existe $M > 0$ talque $\|\eta^k\| \leq M$, para todo k . Também temos que $\{\lambda_k\}$ é limitada, então $0 < \lambda_- \leq \lambda_k$, para todo k . Definindo $z^{k+1} = \xi^{k+1} - \nabla h(x^{k+1})$, temos que $z^{k+1} \in \partial_L f(x^{k+1})$ e

$$\begin{aligned} \|z^{k+1}\| &= \|\xi^{k+1} - \nabla h(x^{k+1}) + \nabla h(x^k) - \nabla h(x^k)\| \\ &\leq \|\xi^{k+1} - \nabla h(x^k)\| + \|\nabla h(x^{k+1}) - \nabla h(x^k)\| \\ &\leq \frac{M}{\lambda_-} q(x^k, x^{k+1}) + L\|x^{k+1} - x^k\| \\ &\leq \left(\frac{\beta M}{\lambda_-} + L \right) \|x^{k+1} - x^k\|. \end{aligned}$$

Logo, para cada $k \in \mathbb{N}$ existe $z^{k+1} \in \partial_L f(x^{k+1})$ tal que $\|z^{k+1}\| \leq b\|x^{k+1} - x^k\|$, onde $b = \left(\frac{\beta M}{\lambda_-} + L \right)$.

3. Como \hat{x} é um ponto de acumulação de $\{x^k\}$, então existe $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ convergindo para \hat{x} . Segue da continuidade de f que $\{f(x^{k_j})\}$ converge para $f(\hat{x})$. Logo, pela observação 4.3.1, temos que $\{f(x^k)\}$ converge para $f(\hat{x})$.

E o teorema está provado. ■

4.4.1 Experimento numérico

Considere a função $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $f(x_1, x_2) = x_1^4 + x_2^4 - 2x_1^2 - 2x_2^2$, cujos pontos críticos são $(x_1^*, x_2^*) = (0, 0)$ (máximo local), $(x_1^*, x_2^*) = (1, 1)$, $(x_1^*, x_2^*) = (-1, -1)$, $(x_1^*, x_2^*) = (1, -1)$ e $(x_1^*, x_2^*) = (-1, 1)$ (mínimos globais), com $f^* = \inf_{(x_1, x_2) \in \mathbb{R}^2} f(x_1, x_2) = -2$. O exemplo serve para mostrar que o método pode ser implementado.

Sabemos que f satisfaz a propriedade de Kurdyka-Łojasiewicz pois f é uma função semi-algébrica; veja esse conceito em [8]. A seguir, apresentamos como o Algoritmo 4.1 se comporta para o caso particular em que $q(x, y) = \|x - y\|$.

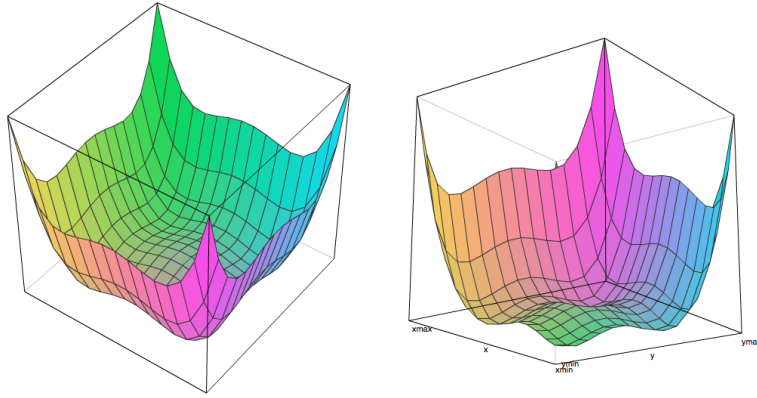


Figura 4.1: Gráfico de $f(x, y)$

Tabela 4.1: $(x_1^0, x_2^0) = (2, -3)$ e $c_k = 1$

k	x_1^k	x_2^k	$f(x_1^k, x_2^k)$
1	1.29585206	-1.50000000	0.0238569
2	1.10354986	-1.16559737	- 1.8239471
3	1.03844506	-1.06027079	- 1.9784392
4	1.01458913	-1.02270175	- 1.9970276
5	1.00558239	-1.00866198	- 1.9995719
6	1.00214283	-1.00332134	- 1.9999373
7	1.00082354	-1.00127593	- 1.9999908
8	1.00031665	-1.00049052	- 1.9999986
9	1.00012178	-1.00018863	- 1.9999998
10	1.00004683	-1.00007254	-1.9999999
15	1.00000039	-1.00000061	-2.0000000
21	1.00000000	-1.00000000	-2.0000000

Nesse exemplo usamos o *software Scilab* com o *toolbox "fmincon"*, para resolver cada minimização nos subproblemas, com critério de parada $\|x^{k+1} - x^k\| < 10^{-7}$.

4.5 Aplicação 1: convergência linear do MPP - caso convexo

Nesta seção consideraremos o caso particular em que $f(x) = g(x) - h(x)$, com g fortemente convexa com constante $\rho > 0$ e h diferenciável com $\nabla h(\cdot)$ Lipschitz com constante $L > 0$. Nesse caso, um ponto x^* é ponto crítico de f se $\nabla h(x^*) \in \partial g(x^*)$. Denotamos S o conjunto dos pontos críticos de f e assumimos $S \neq \emptyset$. Note que a

hipótese de g ser fortemente convexa é razoável pois dada uma função fortemente convexa φ e $f(x) = g(x) - h(x)$, temos que f sempre admite uma decomposição com componentes fortemente convexas, isto é, $f(x) = (g(x) + \varphi(x)) - (h(x) + \varphi(x))$, sabendo que a soma de uma função convexa com uma fortemente convexa é uma função fortemente convexa.

Consideraremos também o caso particular em que $q(x, y) = \|x - y\|$ no Algoritmo 4.1, isto é

$$x^{k+1} \in \arg \min_{x \in \mathbb{R}^n} \left\{ g(x) - \langle \nabla h(x^k), x - x^k \rangle + \frac{1}{2\lambda_k} \|x - x^k\|^2 \right\}. \quad (4.15)$$

Normalmente, taxa de convergência linear para o método do ponto proximal é obtido para o caso em que a função objetivo é fortemente convexa. A seguir, mostramos a convergência linear de (4.15).

Teorema 4.5.1 *Considere $\{x^k\}$ a sequência gerada pelo Algoritmo 4.1 com (4.15) ao invés de (4.3). Se $\rho > 2L$, então existe uma constante $0 < r < 1$ tal que*

$$\|x^{k+1} - x^*\| \leq r \|x^k - x^*\|, \quad \forall k \geq 0, \forall x^* \in S. \quad (4.16)$$

Com isso, a sequência $\{x^k\}$ converge linearmente para um ponto crítico de f .

Demonstração: Segue de (4.15) que existe $z^{k+1} \in \partial g(x^{k+1})$ tal que

$$\nabla h(x^k) = z^{k+1} + \frac{1}{\lambda_k} (x^{k+1} - x^k). \quad (4.17)$$

Tome $x^* \in S$ um ponto crítico de f , ou seja, $\nabla h(x^*) \in \partial g(x^*)$. Sendo g fortemente convexa, temos que $\partial g(\cdot)$ é fortemente monótono e, com isso, segue de (4.17) que

$$\begin{aligned} 0 &\leq \langle x^{k+1} - x^*, z^{k+1} - \nabla h(x^*) \rangle - \rho \|x^{k+1} - x^*\|^2 \\ &= \langle x^{k+1} - x^*, \frac{x^k - x^{k+1}}{\lambda_k} + \nabla h(x^k) - \nabla h(x^*) \rangle - \rho \|x^{k+1} - x^*\|^2. \end{aligned}$$

Assim, sendo $\lambda_k > 0$, usando a desigualdade de Cauchy-Schwarz e o fato que $\nabla h(\cdot)$

é Lipschitz contínuo, obtemos

$$\begin{aligned}
0 &\leq 2\lambda_k \langle x^{k+1} - x^*, \nabla h(x^k) - \nabla h(x^*) \rangle - 2\langle x^{k+1} - x^*, x^{k+1} - x^k \rangle \\
&\quad - 2\lambda_k \rho \|x^{k+1} - x^*\|^2 \\
&\leq 2\lambda_k L \|x^{k+1} - x^*\| \|x^k - x^*\| - \|x^{k+1} - x^*\|^2 - \|x^k - x^{k+1}\|^2 \\
&\quad + \|x^k - x^*\|^2 - 2\lambda_k \rho \|x^{k+1} - x^*\|^2 \\
&\leq 2\lambda_k L (\|x^{k+1} - x^*\|^2 + \|x^k - x^*\|^2) - \|x^{k+1} - x^*\|^2 - \|x^k - x^{k+1}\|^2 \\
&\quad + \|x^k - x^*\|^2 - 2\lambda_k \rho \|x^{k+1} - x^*\|^2 \\
&= (1 + 2\lambda_k L) \|x^k - x^*\|^2 - [1 + 2\lambda_k(\rho - L)] \|x^{k+1} - x^*\|^2 - \|x^k - x^{k+1}\|^2.
\end{aligned}$$

Logo,

$$[1 + 2\lambda_k(\rho - L)] \|x^{k+1} - x^*\|^2 \leq (1 + 2\lambda_k L) \|x^k - x^*\|^2.$$

Assim, tomando $0 < r := \sqrt{\frac{(1 + 2\lambda_k L)}{1 + 2\lambda_k(\rho - L)}} < 1$ obtemos

$$\|x^{k+1} - x^*\| \leq r \|x^k - x^*\|, \quad \forall k \geq 0, \forall x^* \in S.$$

A última desigualdade implica que $\{x^k\}$ é Féjer convergente ao conjunto S . Logo, $\{x^k\}$ é limitada. Combinando o Teorema 4.3.2 com a Proposição 2.1.8, temos que $\{x^k\}$ converge para um ponto crítico de f . ■

A condição $\rho > 2L$ implica que f é fortemente convexa. Logo, o teorema acima pode ser visto como uma nova prova de que o método do ponto proximal converge linearmente para o caso fortemente convexo (veja Rockafellar [112]), pois dada uma função f fortemente convexa com constante $\rho > 0$ temos que f pode ser escrita como uma função DC que satisfaz as hipóteses do teorema acima apenas tomando h uma função convexa e $C^{1,1}$ tal que $2L < \rho$. Assim, $f(x) = g(x) - h(x)$, com $g(x) = f(x) + h(x)$ fortemente convexa.

4.6 Aplicação 2: problema de produção - caso unidimensional

Um dos principais tópicos em Economia e Ciências de Gestão, em Pesquisa Operacional, é determinar o tamanho ótimo de produção ou o tamanho ótimo de uma organização. Esse é um problema difícil tanto no aspecto conceitual como no aspecto técnico. O termo “tamanho ótimo” pode se referir à quantidade produzida de um bem (ou produto final), a variedade de bens (ou produtos finais) produzidos, a quantidade e a qualidade de trabalhadores de diferentes tipos de empregadores, a quantidade de meios usados no processo de produção, assim como o número de

estágios intermediários no processo de produção e as diferentes possíveis localizações em vários países em um processo de produção globalizado. Existe uma ampla literatura sobre esse assunto e vários aspectos devem ser levados em consideração.

Nesta seção iremos considerar, usando uma abordagem dinâmica, o problema estático de tamanho de produção. Esse tamanho se refere aos níveis de produção, isto é, o número de unidades de um produto final que uma companhia produz. Iremos considerar o caso mais difícil, porém mais realista, em que uma companhia possui retorno crescente em curto prazo, quando a função do custo de execução de produção é côncava. O caso mais tradicional de retornos decrescentes de escala de produção, quando o custo de produção é convexo, é bem mais simples.

Denominamos retornos crescentes de escala de produção quando há insumos (“*inputs*”) necessários para se produzir mas isso não influencia a quantidade de bens finais (“*outputs*”) produzidos, ou seja, a função produção $f(K, L)$ satisfaz

$$f(\theta K, \theta L) > \theta f(K, L),$$

para $\theta > 1$. Por exemplo, para escrever um “*software*” necessitamos de “*inputs*” para programadores que não influenciará na quantidade de cópias produzidas.

Quando $f(\theta K, \theta L) < \theta f(K, L)$, para $\theta > 1$, dizemos que o retorno na escala de produção é decrescente. Isso ocorre, por exemplo quando, para aumentar o tamanho de uma companhia requer que se gaste mais recursos para controlar o aumento da burocracia ou fluxo de informações. Se $f(\theta K, \theta L) = \theta f(K, L)$, então dizemos que o retorno é constante.

Exemplo 4.6.1 *A função de produção Cobb-Douglas é dada por*

$$f(K, L) = pL^\alpha K^\beta,$$

onde L é a quantidade de trabalho empregado, K é o capital investido, p é o fator de produtividade total e α, β são constantes associadas às variáveis L e K determinadas pelas tecnologias disponíveis, respectivamente. Assim,

$$f(\theta K, \theta L) = p(\theta L)^\alpha (\theta K)^\beta = \theta^{\alpha+\beta} pL^\alpha K^\beta = \theta^{\alpha+\beta} f(K, L).$$

Portanto,

se $\alpha + \beta < 1$, temos retorno decrescente;

se $\alpha + \beta > 1$, temos retorno crescente;

se $\alpha + \beta = 1$, temos retorno constante.

Usaremos a recente abordagem “*Variational Rationality*” (VR) de A. Soubeyran [117–119] para determinar o tamanho ótimo de produção de uma companhia,

a longo prazo, onde se pode em cada período contratar, demitir e manter trabalhadores na companhia. Essa abordagem oferece uma teoria original e dinâmica do problema de produção.

4.6.1 Um modelo de companhia com retornos crescentes a curto prazo

Para entender melhor como otimização DC funciona em aplicações em Ciência de Gestão vamos examinar um caso simples que pode ser estendido para o caso multiobjetivo que é mais realista. Diferentes variações desse exemplo podem ser encontradas em Soubeyran [117, 118], Bento e Soubeyran [26, 27] e Bao et al. [15]. Mas nenhum desses trabalhos examina o importante caso de retorno crescente, que é o caso mais realista em custo de produção, como fazemos aqui, ou seja, como uma aplicação do método do ponto proximal para otimização DC.

Considere uma companhia hierárquica consistindo de empreendedor, um perfil de trabalhadores e uma sucessão de períodos onde o empreendedor pode contratar, despedir ou manter empregados trabalhando em um ambiente de mudanças. Em cada período o empreendedor escolhe o quanto produzir do mesmo bem final (com uma determinada qualidade) e vender cada unidade desse bem final ao mesmo preço fixo $p > 0$. No período atual a companhia produz $x \in \mathbb{R}_+$ unidades de um bem final e emprega $l(x) \in \mathbb{R}_+$ trabalhadores. Nesse modelo, o tamanho ótimo de produção se refere a x . Para simplificar, a cada trabalhador é solicitado que produza uma unidade de bem final. Então, $l(x) = x$. O lucro atual do empreendedor, $\pi(x) = r(x) - c(x)$, é a diferença entre a receita da companhia $r(x) = px \geq 0$ e o custo de produção $c(x) \geq 0$.

Um exemplo de lucro “a ser aumentado” e “a ser diminuído”

Para produzir uma unidade do bem final cada trabalhador usa um dado conjunto de meios individuais (ferramentas e ingredientes) e um meio coletivo fixo (digamos uma determinada região ou uma infraestrutura). O empreendedor aluga as ferramentas duráveis e compra os ingredientes não duráveis. Seja $\bar{\pi} = \sup \{ \pi(y), y \in X \} < +\infty$ o maior lucro possível que o empreendedor pode esperar obter. Então, $f(x) = \bar{\pi} - \pi(x) \geq 0$ é o atual lucro não obtido que ele pode almejar executar no atual período ou mais tarde. A função lucro $\pi(\cdot)$ é a função “a ser aumentada” enquanto a função lucro não obtida $f(\cdot)$ é a função “a ser diminuída”.

Nas seções anteriores deste capítulo, a função objetivo $f(x) = g(x) - h(x)$ é a diferença de duas funções convexas $h(x)$ e $g(x)$. Nesse exemplo, $f(x)$ representa a

função lucro não obtido que o empreendedor espera obter, isto é,

$$f(x) = \bar{\pi} - \pi(x) = \bar{\pi} + c(x) - r(x),$$

onde $g(x) = \bar{\pi} + c(x)$ e $h(x) = r(x)$. Então, as funções custo e receita, $c(\cdot)$ e $r(\cdot)$, devem ser côncavas para que sejam aplicáveis os resultados obtidos neste capítulo.

Com efeito, em um mercado competitivo perfeito em que o preço p de um bem final é dado, a função receita $r : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ é dada por $r(x) = px$ que é linear, logo é côncava com respeito ao nível de produção x . O que nos resta mostrar é que a função custo $c(\cdot)$ é normalmente côncava em curto prazo. Em geral, para escapar de dificuldades matemáticas, os livros de Economia focam no caso menos usado em que a função custo de produção é convexa em curto prazo.

Custo de produção é côncavo quando a tecnologia da companhia exhibe retorno crescente vindo de economia de escala, economia de especialização, aprender fazendo várias vezes a mesma coisa, capacidades limitadas, falta de tempo para ser capaz de mudar custos fixos a curto prazo que se tornam custos variáveis a longo prazo. No nosso modelo de companhia, custo de produção $c(x) = wx + hx + K$ é a soma de três diferentes custos, a saber:

- i) $w > 0$ é um determinado salário pago a cada trabalhador empregado;
- ii) $h > 0$ é o preço pago ao fornecedor para adquirir cada conjunto de meios usados por cada trabalhador para produzir uma unidade do bem final;
- iii) $K > 0$ é o custo para alugar um meio durável, fixo, coletivo e indivisível.

Esse custo de produção exhibe retorno crescente para escala porque no período atual, antes da produção iniciar, o custo fixo $K > 0$ deve ser pago mesmo se depois nenhum trabalhador seja solicitado para trabalhar, ou seja, $c(0) = K > 0$. Isso implica que o custo unitário de produção $c(x)/x = w + h + K/x$ decresce quando o nível de produção x aumenta. O custo de produção será estritamente côncavo se, por exemplo, o preço $h = h(x)$ de cada conjunto de meios usado pelos trabalhadores decrescer com o número x de conjunto de meios que o empreendedor deve comprar para produzir x unidades do bem final (por exemplo quando fornecedores dão descontos).

Exemplo 4.6.2 *A função de produção de Leontief é uma função de produção que implica que os fatores de produção serão usados em proporções fixas (tecnologicamente pré-determinados). A função é dada por*

$$f(x) = \min_i \left\{ \frac{x_i}{a_i} \right\}, \quad x_i > 0, \quad a_i > 0, \quad i = 1, \dots, n,$$

onde x_i representa quantidade utilizada e a_i uma constante tecnologicamente determinada. Temos que a função de produção de Leontief é côncava.

4.6.2 A abordagem variational rationality: simples formulação

Dinâmicas de “permanecer” e “mudar”

A abordagem “*variational rationality*” (VR) (Soubeyran [117, 118]) modela e unifica vários modelos de dinâmicas de permanecer e mudar que são usadas em Ciência de Comportamento (Economia, Ciência de Gestão, Psicologia, Sociologia, Ciência Política, Teoria de Decisão, Teoria de Jogos, Inteligência Artificial, ...). O termo “permanecer” (“*stay*”) se refere a fases de exploração (obtendo benefícios), repetições temporárias da mesma ação, hábitos temporários, rotinas, regras, normas, ..., enquanto o termo “mudar” (“*change*”) se refere a fases de exploração, processos de aprendizagem e inovação, formação e quebra de hábitos e rotinas, mudanças de acontecimentos (ações), Essa abordagem dinâmica considera entidades (um agente, uma organização ou vários agentes interagindo) que estão a princípio em uma posição indesejável e são incapazes de atingir imediatamente a posição final desejada. O objetivo dessa abordagem é examinar o problema de transição: como tal entidade pode encontrar, construir e usar uma transição viável e aceitável que seja capaz de superar vários obstáculos, dificuldades e resistência intermediárias para mudar com pouco sacrifício intermediário e suficiente satisfação intermediária para sustentar a motivação para mudar e perseverar até alcançar a posição final desejada. Essa abordagem (VR) admite várias variações baseadas na mesma pequena lista de princípios e conceitos gerais.

Os quatro principais conceitos são:

1. permanecer (“*stay*”) e mudar (“*change*”);
2. vale a pena (“*worthwhile*”) permanecer (“*stay*”) e mudar (“*change*”);
3. transição satisfatória (“*worthwhile*”) e “*trap*” variacional;
4. vale a pena se aproximar e alcançar mas não vale a pena sair.

Uma dinâmica “*stay*” e “*change*” se refere a uma sucessão de períodos, onde $k + 1$ é o período atual e k é o período passado com $x = x^k \in X$ sendo uma ação passada (feita) e $y = x^{k+1} \in X$ sendo a ação atual (sendo feita). Uma única mudança (“*change*”) de $x = x^k \in X$ para $y = x^{k+1} \in X$ é denotada por $x \rightsquigarrow y$, $y \neq x$. Enquanto uma única permanência (“*stay*”) em x é denotado por $x \rightsquigarrow y$, $y = x$.

Com base no exemplo anterior, vejamos um simple protótipo da abordagem “*variational rationality*” (VR) para finalmente mostrar como, no limite de uma transição satisfatória, uma companhia pode obter o tamanho ótimo de produção.

Mudanças satisfatórias

A abordagem (VR) começa com a seguinte (ampla) definição de uma mudança satisfatória: uma mudança é satisfatória se a motivação para mudar ao invés de ficar é “suficientemente grande” com respeito a resistência para mudar ao invés de ficar. Essa definição assume várias variações, tantas quantas as definições de motivação (existe mais de cem teorias de motivações em Psicologia), resistência (que inclui diferentes aspectos) e “suficientemente grande” (veja Soubeyran [117, 118]). Vejamos uma simples formulação do conceito de mudança satisfatória.

No nosso exemplo de produção, uma mudança (“*change*”) se refere sair de estando produzido $x \in X = \mathbb{R}_+$ unidades de um bem final no período anterior para produzir $y \in \mathbb{R}_+$ unidades desse bem final no período atual. Permanecer (“*stay*”) é um movimento particular de estando produzido uma dada quantidade $x = x^k$ de um bem final no período anterior para produzir novamente a mesma quantidade $y = x^{k+1} = x^k$ desse mesmo bem final no período atual. Os lucros “a ser aumentado” anterior e atual do empreendedor são os lucros $\pi(x)$ e $\pi(y)$. Seus lucros “a ser diminuído” anterior e atual são os lucros não obtidos $f(x) = \bar{\pi} - \pi(x) \geq 0$ e $f(y) = \bar{\pi} - \pi(y) \geq 0$.

Vantagens para mudar de x para y , se existirem, representam a diferença entre o lucro e o lucro não obtido $A(x, y) = \pi(y) - \pi(x) = f(x) - f(y) \geq 0$.

Inconveniências para mudar de x para y se referem a diferença $I(x, y) = C(x, y) - C(x, x) \geq 0$.

$C(x, y) \geq 0$ modela o custo de estar apto a mudar de x para y . No nosso modelo de produção, $C(x, y)$ modela o custo para contratar, demitir e manter trabalhadores trabalhando para estar apto para mover de produzir x unidades de um bem final para produzir y unidades do bem final, onde y pode ser maior, menor ou igual a x . O custo de contratar $y - x > 0$ trabalhadores é $C(x, y) = \rho^+(y - x)$, onde $\rho^+ > 0$ é o custo de contratar um trabalhador. O custo de demitir $x - y > 0$ trabalhadores é $C(x, y) = \rho^-(x - y)$, onde $\rho^- > 0$ é o custo de demitir um trabalhador. O custo de manter trabalhando $y = x$ trabalhadores é $C(x, x) = \rho^-x$, onde $\rho^- \geq 0$ é o custo de manter trabalhando um período mais um trabalhador. Por simplificação iremos assumir que $\rho^- = 0$. Então, $C(x, x) = 0$ e inconveniência para mudar é dada por

$$I(x, y) = C(x, y) = \left\{ \begin{array}{l} \rho^+(y - x) \text{ if } y \geq x \\ \rho^-(x - y) \text{ if } y \leq x \end{array} \right\} \geq 0.$$

Motivação para mudar $M(x, y) = U[A(x, y)]$ é a utilidade $U[A]$ de vantagens para mudar $A = A(x, y) \geq 0$.

Resistência para mudar $R(x, y) = D[I(x, y)]$ é a inutilidade $D[I]$ de inconveniência para mudar $I = I(x, y) \geq 0$, onde a função utilidade $U[\cdot] : A \in \mathbb{R}_+ \mapsto U[A] \in \mathbb{R}_+$ e a função inutilidade $D[\cdot] : I \in \mathbb{R}_+ \mapsto D[I] \in \mathbb{R}_+$ são estritamente crescentes e zero em zero.

Uma mudança satisfatória de x para y é tal que a motivação para mudar $M(x, y) \in \mathbb{R}_+$ de x para y é maior que a resistência para mudar $R(x, y)$ de x para y ponderado por um determinado raio de satisfação de mudança satisfatória $\xi > 0$, ou seja,

$$M(x, y) \geq \xi R(x, y).$$

No exemplo, a utilidade $U[A]$ de vantagens para mudar e inutilidade $D[I]$ de inconveniências para mudar são linear-quadráticas, mais precisamente, $M = U[A] = A$ e $R = D[I] = I^2$; veja Soubeyran [117, 118] para casos mais gerais. Nesse contexto, uma mudança $x \curvearrowright y$ de produzir novamente a quantidade x de um bem final para produzir uma quantidade diferente y desse mesmo bem final é satisfatória se vantagens para mudar são “suficientemente grandes” com respeito a resistência para mudar, ou seja,

$$A(x, y) = \pi(y) - \pi(x) = f(x) - f(y) \geq \xi R(x, y) = \xi C(x, y)^2,$$

onde $C(x, x) = 0$. Aqui o termo “suficientemente grande” é definido pelo tamanho de $\xi > 0$.

Transições satisfatórias

Uma transição é uma sucessão de únicos “permanecer” e “mudar” $x^0 \curvearrowright x^1 \curvearrowright \dots x^k \curvearrowright x^{k+1} \curvearrowright \dots$, onde $x^{k+1} \neq x^k$ ou $x^{k+1} = x^k$, para cada $k \in \mathbb{N}$.

Uma transição satisfatória é uma transição tal que “permanecer” ou “mudar” é satisfatório, isto é, $x^{k+1} \in W_{\xi_{k+1}}(x^k)$, $k \in \mathbb{N}$, que significa

$$\begin{aligned} A(x^k, x^{k+1}) &= \pi(x^{k+1}) - \pi(x^k) \\ &= f(x^k) - f(x^{k+1}) \\ &\geq \xi_{k+1} R(x^k, x^{k+1}) \\ &= \xi_{k+1} C(x^k, x^{k+1})^2, \quad k \in \mathbb{N}. \end{aligned}$$

Limites como trap variacionais

Dizemos que um ponto $x^* \in X$ é um “trap” variacional (forte) se é:

- i) um ponto aspirante $x^* \in W_{\xi_{k+1}}(x^k)$, $k \in \mathbb{N}$, vale a pena (“*worthwhile*”) atingir qualquer posição de uma transição;
- ii) um “*trap*” estacionário $W_{\xi_*}(x^*) = \{x^*\}$, onde não vale a pena mover para qualquer outra posição $y \neq x^*$, dado que o raio de satisfação de mudar converge, ou seja, $\lim_{k \rightarrow +\infty} \xi_{k+1} = \xi_* > 0$;
- iii) um ponto que vale a pena se aproximar, ou seja, que converge para um ponto aspirante.

Em outras palavras, x^* é um “*trap*” variacional se:

- i) $A(x^k, x^*) = \pi(x^*) - \pi(x^k) = f(x^k) - f(x^*) \geq \xi_{k+1}R(x^k, x^*) = \xi_{k+1}C(x^k, x^*)^2$, para todo $k \in \mathbb{N}$;
- ii) $A(x^*, y) = \pi(y) - \pi(x^*) = f(x^*) - f(y) < \xi_*R(x^*, y) = \xi_*C(x^*, y)^2$, para todo $y \neq x^*$;
- iii) É um ponto limite da transição satisfatória, ou seja, $\lim_{k \rightarrow +\infty} x^k = x^*$.

Um *trap* variacional fraco não necessita ser um ponto aspirante.

4.6.3 Algoritmo proximal como transições satisfatórias

Para mostrar como o método do ponto proximal pode ser visto como um exemplo de transição satisfatória, apresentamos uma formulação específica da abordagem (VR) onde a a função utilidade de vantagem para mudar e inutilidade de inconveniência para mudar são linear-quadrática, mais especificamente

$$M = U[A] = A \quad \text{e} \quad R = D[I] = I^2 = C^2,$$

onde $C(x, y) = q(x, y) \geq 0$ é uma quase distância; esse caso foi usado em Moreno et al. [101] e casos mais gerais podem ser encontrados em Bento e Soubeyran [26, 27].

A formulação proximal de uma mudança satisfatória

No contexto linear-quadrática, motivação e resistência para mudar são $M(x, y) = A(x, y) = \pi(y) - \pi(x) = f(x) - f(y)$ e $R(x, y) = q(x, y)^2$. Isso define:

1. Lucro proximal “a ser aumentado” $P_\xi(y/x) = \pi(y) - \xi R(x, y)$, que é a diferença entre o atual lucro “a ser aumentado” $\pi(y)$ e a atual resistência (com um peso) para mudar $R(x, y)$, onde o peso $\xi > 0$ determina a importância do atual lucro “a ser aumentado” e a atual resistência para mudar;

2. Lucro proximal “a ser diminuído” $Q_\xi(y/x) = f(y) + \xi R(x, y)$, que é a soma do atual lucro “a ser diminuído” $f(y)$ e a atual resistência (com um peso) para mudar $R(x, y)$.

Então, uma mudança $x \rightsquigarrow y \in W_\xi(x)$ é satisfatória se, mudando de x para y , o lucro proximal “a ser aumentado” aumenta, ou seja,

$$P_\xi(y/x) \geq P_\xi(x/x)$$

e lucro proximal “a ser diminuído” decresce, isto é,

$$Q_\xi(y/x) \leq Q_\xi(x/x).$$

Isso vem das seguintes equivalências:

$$\begin{aligned} y \in W_\xi(x) &\iff M(x, y) \geq \xi R(x, y) \\ &\iff \pi(y) - \pi(x) = f(x) - f(y) \geq \xi R(x, y) \\ &\iff P_\xi(y/x) \geq P_\xi(x/x) \\ &\iff Q_\xi(y/x) \leq Q_\xi(x/x). \end{aligned}$$

Método proximal como exemplo de transição satisfatória

Uma transição é uma sucessão de únicos “*stay*” e “*change*”, $x^0 \rightsquigarrow x^1 \rightsquigarrow \dots x^k \rightsquigarrow x^{k+1} \rightsquigarrow \dots$, onde $x^{k+1} \neq x^k$ ou $x^{k+1} = x^k$, para cada $k \in \mathbb{N}$.

Uma transição satisfatória é uma transição tal que cada “*stay*” ou “*change*” é satisfatório que em termos dos lucros proximais para mudar, se tornam:

$$x^{k+1} \in W_{\xi_{k+1}}(x^k) = \left\{ \begin{array}{l} y \in X, \text{ tal que} \\ P_{\xi_{k+1}}(y/x^k) \geq P_{\xi_{k+1}}(x^k/x^k), \text{ ou seja,} \\ \pi(y) - \xi_{k+1}R(x_n, y) \geq \pi(x^k), \text{ ou,} \\ Q_{\xi_{k+1}}(y/x^k) \leq Q_{\xi_{k+1}}(x^k/x^k), \text{ ou seja,} \\ f(y) + \xi_{k+1}R(x^k, y) \leq f(x^k) \end{array} \right\},$$

onde cada $\xi_{k+1} > 0$, $k \in \mathbb{N}$ pode ser escolhido e $R(x^k, y) = q(x^k, y)^2$.

Nesse contexto,

$$x^{k+1} \in W_{\xi_{k+1}}(x^k) \iff f(x^{k+1}) + \xi_{k+1}q(x^k, x^{k+1})^2 \leq f(x^k),$$

onde $\xi_{k+1} = 1/2\lambda_k > 0$.

Uma mudança satisfatória é exata se $x^{k+1} \in \arg \max \{P_{\xi_{k+1}}(y/x_n), y \in X\}$. Neste capítulo consideramos apenas mudança satisfatória exatas, mas como o

método proximal admite versões inexatas também podemos considerar mudanças satisfatórias inexatas.

Uma mudança satisfatória inexata é qualquer mudança satisfatória “suficientemente próxima” a uma mudança satisfatória exata, onde o termo “suficientemente próxima” pode ter diferentes interpretações; veja uma justificativa para isso em Bento e Soubeyran [26].

Quando a função lucro não é conhecida

Normalmente, o empreendedor não conhece toda a função lucro $\pi(\cdot)$. Nesse caso, em cada período atual $k + 1$, uma avaliação aproximada $\tilde{\pi}(\cdot/x)$, onde $x = x^k$, da função lucro é obtida. Isso requer considerar uma formulação mais complexa da abordagem (VR), onde experiências passadas e atuais avaliações são incluídas no processo de mudança satisfatória; veja Soubeyran [119]. No nosso exemplo iremos descartar o papel da experiência passada para focar nossa atenção no processo de atual avaliação, quando o empreendedor conhece desde o início toda a função receita $r(\cdot) = -g(\cdot)$, mas não conhece completamente a função custo de execução $c(\cdot)$.

Então, em cada período é necessário fazer uma avaliação aproximada da função custo de execução $c(\cdot)$, em termos de uma função $\tilde{c}(\cdot/x^k)$ que globalmente estima superiormente essa função custo $c(\cdot) = -h(\cdot)$, ou seja,

$$\tilde{c}(y/x^k) \geq c(y), \quad \forall y \in X \quad \text{com} \quad \tilde{c}(x^k/x^k) = c(x^k).$$

Logo, a função avaliação $\tilde{\pi}(\cdot/x) : y \in X \mapsto \tilde{\pi}(y/x) = r(x) - \tilde{c}(y/x^k)$ estima inferiormente a função custo “a ser aumentada” $\pi(\cdot) = r(\cdot) - c(\cdot)$, pois $\tilde{\pi}(y/x) \leq \pi(y)$, para todo $y \in X$ e $\tilde{\pi}(x/x) = \pi(x)$.

De forma similar, a função avaliação $\tilde{f}(\cdot/x) : y \in X \mapsto \tilde{f}(y/x) = g(y) - \tilde{h}(y/x)$ estima superiormente a função lucro “a ser diminuída” $f(\cdot) = g(\cdot) - h(\cdot)$, pois $\tilde{f}(y/x) \geq f(y)$, para todo $y \in X$, com $\tilde{f}(x/x) = f(x)$, onde $\tilde{h}(y/x) \leq h(y)$, para todo $y \in X$ com $\tilde{h}(x/x) = h(x)$.

Iremos supor que, no período atual $k + 1$, o empreendedor conhece a função resistência para mudar $R(x, y)$. Assim, dado essa estrutura de conhecimento, onde em cada período o empreendedor faz uma estimação inferior de sua função lucro $\tilde{\pi}(\cdot/x)$, uma mudança $x = x^k \curvearrowright y$ é satisfatória se $\tilde{A}(x, y) = \tilde{\pi}(y/x) - \pi(x) = f(x) - \tilde{f}(y/x) \geq \xi R(x, y)$.

Assim, a versão proximal dessa condição de mudança satisfatória é

$$\tilde{P}_\xi(y/x) = \tilde{\pi}(y/x) - \xi R(x, y) \geq \pi(x) = \tilde{P}_\xi(x/x),$$

ou

$$\tilde{Q}_\xi(y/x) = \tilde{f}(y/x) + \xi R(x, y) \leq f(x) = \tilde{Q}_\xi(x/x).$$

Esse processo avaliação comportamental condiz a parte matemática desse capítulo.

4.6.4 Limites

Quando pontos críticos são “trap” variacionais

Um ponto é variacional “trap” fraco quando é ao mesmo tempo um ponto limite de uma transição satisfatória e um “trap” estacionário de onde não vale a pena mudar. Isso modela a abordagem e o limite de um processo de “stay” e “change” satisfatório. Normalmente, um ponto crítico não é um “trap” estacionário. Assim, naturalmente surge: Quando um ponto crítico limite do método proximal é um “trap” variacional?

Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ dizemos que f é fracamente convexa se existe $\rho > 0$ tal que, para todo $x, y \in \mathbb{R}^n$ e $\lambda \in [0, 1]$, temos

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) + \rho\lambda(1 - \lambda)\|x - y\|^2. \quad (4.18)$$

A função f é dita localmente fracamente convexa em x se existe $\epsilon > 0$ tal que f é fracamente convexa em $B(x, \epsilon)$. Dizemos simplesmente que f é localmente fracamente convexa que o é em todos os pontos de seu domínio.

A resposta do questionamento anterior segue das seguintes proposições.

Proposição 4.6.1 *Seja f uma função fracamente convexa. Se x^* é um ponto crítico de f , então*

$$f(x^*) \leq f(y) + \frac{\rho}{\alpha^2} q^2(x^*, y) \quad \forall y \in \mathbb{R}^n, \quad (4.19)$$

onde $\alpha > 0$ satisfaz a Condição 1.

Demonstração: Segue de [131, Proposição 4.8] e (4.1). ■

Proposição 4.6.2 *Seja f uma função fracamente convexa. Se x^* é um ponto crítico de f e $\lambda > \frac{\rho}{\alpha^2}$, então $W_\lambda(x^*) = \{x^*\}$.*

Demonstração: Combinando (4.19) com $\lambda > \frac{\rho}{\alpha^2}$, temos que

$$f(x^*) \leq f(y) + \frac{\rho}{\alpha^2} q^2(x^*, y) < f(y) + \lambda q^2(x^*, y) \quad \forall y \neq x^*.$$

O resultado segue da desigualdade anterior juntamente com a definição de $W_\lambda(x)$. ■

Observação 4.6.1 *Segue da definição que toda função convexa é fracamente convexa. Além disso, funções de classe $C^{1,1}$ e funções lower- C^2 são funções localmente fracamente convexas e fracamente convexas (localmente Lipschitz), respectivamente.*

Sabe-se que toda função de classe $C^{1,1}$ e lower- C^2 são funções DC; veja [72, página 48].

Tamanho ótimo de produção

No nosso exemplo de produção um “*trap*” variacional $x^* \in X$ define um tamanho ótimo de produção de uma companhia onde o empreendedor contrata e demite menos e menos trabalhadores até finalmente parar de contratar e demitir trabalhadores quando a resistência para mudar vence a motivação para mudar. Isso oferece uma teoria original do problema de tamanho ótimo de produção em termos da abordagem (VR), onde o empreendedor obtém o ótimo no limite sem muitos sacrifícios durante a transição.

Capítulo 5

MPP para otimização multiobjetivo

Neste capítulo apresentaremos uma nova abordagem para convergência do método do ponto proximal em problemas de otimização multiobjetivo de dimensão finita. O método proximal para otimização vetorial foi primeiramente estudado em Bonnel et al. [33]. A partir desse trabalho, vários outros autores estudaram o método do ponto proximal no cenário vetorial ou multiobjetivo, mas todos seguindo a abordagem de convergência do método proposta em [33]; veja por exemplo Apolinário et al. [7], Bento et al. [21], Ceng e Yao [39], Choung et al. [42], Gregório e Oliveira [69], Villacorta e Oliveira [132]. Tal abordagem é baseada na condição de otimalidade de primeira ordem do problema escalarizado. Essa abordagem permitiu o estudo de algoritmos para funções convexas e quase convexas como veremos na seção a seguir.

A nova abordagem proposta aqui, para analisar a convergência do método do ponto proximal em problemas multiobjetivos, nos permite ir além de funções quase convexas (e conseqüentemente funções convexas). Mais precisamente analisaremos dois casos não convexas: no primeiro iremos propor um algoritmo proximal para funções DC multiobjetivo e no segundo caso para funções vetoriais localmente Lipschitz. Em particular, também analisamos o caso quase convexo recuperando os principais resultados de [7], [21] e [33] (esse último para o caso multiobjetivo de dimensão finita). Essa nova abordagem combina o fato de cada iteração do método é uma solução eficiente fraca de um problema multiobjetivo com uma condição necessária para um ponto ser solução eficiente fraca de um problema de otimização multiobjetivo. Este capítulo deu origem aos trabalhos [123] e [124] submetidos para possível publicação.

5.1 MPP em otimização multiobjetivo

Otimização multiobjetivo é o processo de otimizar simultaneamente duas ou mais funções reais objetivo. Normalmente, um único ponto não minimiza toda as funções objetivo ao mesmo tempo, isto é, não existe um minimizador ideal, e com isso, o conceito de otimalidade é substituído pelo conceito de Pareto otimalidade. Há um amplo campo de pesquisa que consiste em estender para o contexto vetorial, método iterativos para funções escalares, por exemplo método do gradiente [56, 67], método do gradiente projetado [58, 66], método subgradiente [17], método de Newton [57], método do ponto proximal [33].

A seguir, brevemente descrevemos como o método proximal (exato) analisado em [33] fornece uma solução para um problema de otimização vetorial. O método é usado para encontrar uma solução Pareto fraca de uma aplicação $F : X \rightarrow Y$ de um espaço de Hilbert X com valores em um espaço de Banach Y contendo um cone \mathcal{C} fechado, convexo e pontudo com interior não vazio, onde “pontudo” significa que $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$, com respeito à ordem parcial “ $\preceq_{\mathcal{C}}$ ” induzida pelo cone \mathcal{C} . O algoritmo gera uma sequência que calcula na k -ésima iteração uma solução Pareto fraca de $F_k : X \rightarrow Y$ definida como

$$F_k(x) = F(x) + \lambda_k \|x - x^k\|^2 \varepsilon^k$$

sujeito ao conjunto $\Omega_k = \{x \in X : F(x) \preceq_{\mathcal{C}} F(x^k)\}$, onde $\{\lambda_k\}$ é uma sequência limitada de números reais positivos e ε^k é tomado no interior de \mathcal{C} tal que $\|\varepsilon^k\| = 1$, para todo $k \geq 0$. A ideia por trás do resultados é baseada na condição de otimalidade de primeira ordem do problema escalar

$$\min_{x \in \Omega_k} \eta_k(x), \tag{5.1}$$

onde $\eta_k(x) = \langle F(x), z^k \rangle + \frac{\lambda_k}{2} \langle \varepsilon^k, z^k \rangle \|x - x^k\|^2$ e $\{z^k\}$ é uma sequência que pertence ao cone polar positivo $\mathcal{C}^+ \subset Y^*$ dado por $\mathcal{C}^+ = \{z \in Y^* : \langle y, z \rangle \geq 0, \forall y \in \mathcal{C}\}$ tal que $\|z^k\| = 1$, para todo $k \geq 0$, e Y^* é o dual de Y com $\langle \cdot, \cdot \rangle : Y \times Y^* \rightarrow \mathbb{R}$. Então, (5.1) implica que

$$0 \in \partial \psi_k(x^{k+1}) + \lambda_k \langle \varepsilon^k, z^k \rangle (x^{k+1} - x^k), \tag{5.2}$$

onde $\psi_k(x) = \langle F(x), z^k \rangle + \delta_{\Omega_k}(x)$, com $\partial \psi_k$ denotando o subdiferencial de ψ_k em x no sentido clássico de análise convexa e $\delta_{\Omega_k}(\cdot)$ é a função indicadora, isto é, $\delta_{\Omega_k}(x) = 0$, se $x \in \Omega_k$, e $\delta_{\Omega_k}(x) = +\infty$, caso contrário. Bonnel et al. [33] estabelece que toda sequência gerada por esse algoritmo converge (na topologia fraca de X) para uma solução Pareto fraca de F sob as seguintes hipóteses:

(A1) (Convexidade e semicontinuidade inferior) F é \mathcal{C} -convexa, ou seja, $F((1-t)x +$

$ty) \preceq_{\mathcal{C}} (1-t)F(x) + tF(y)$, para todo $x, y \in X$ e $t \in [0, 1]$; e F é positivamente semicontínua inferior que significa que, para todo $z \in \mathcal{C}^+$, a função escalar $x \mapsto \langle F(x), z \rangle$ é semicontínua inferiormente;

(A2) (Completeza) O conjunto $(F(x^0) - \mathcal{C}) \cap F(X)$ é \mathcal{C} -completo, isto é, para toda sequência $\{a^k\} \subset X$, com $a^0 = x^0$, tal que $F(a^{k+1}) \preceq_{\mathcal{C}} F(a^k)$, para todo $k \in \mathbb{N}$, existe $a \in X$ tal que $F(a) \preceq_{\mathcal{C}} F(a^k)$, para todo $k \in \mathbb{N}$.

A hipótese (A1) garante que o conjunto Ω_k é convexo e fechado, para todo $k \in \mathbb{N}$. Assim, (5.2) pode ser visto como

$$\alpha_k(x^k - x^{k+1}) \in \partial(\langle F(\cdot), z^k \rangle)(x^{k+1}) + N_{\Omega_k}(x^{k+1}), \quad (5.3)$$

onde $\alpha_k = \lambda_k \langle \varepsilon^k, z^k \rangle$ e $N_{\Omega_k}(x^{k+1})$ denota o cone normal de Ω_k em $x^{k+1} \in \Omega_k$ no sentido clássico de análise convexa. Nessa abordagem, a convexidade dos conjuntos Ω_k desempenha um papel importante, pois sem essa propriedade o cone normal $N_{\Omega_k}(x^{k+1})$ é, em geral, difícil de ser encontrado.

Outros autores consideraram variantes do método proposto em [33] para problemas convexos de otimização vetorial e multiobjetivo, por exemplo Ceng e Yao [39], Choung et al. [42], Gregório e Oliveira [69], Villacorta e Oliveira [132]. Recentemente, o caso \mathbb{R}_+^m -quase convexo foi estudado em Apolinário et al. [7] e Bento et al. [21]. Nesses trabalhos, os métodos propostos calculam na k -ésima iteração um ponto x^{k+1} satisfazendo

$$0 \in \partial g(F(x^{k+1})) + \alpha_k(x^{k+1} - x^k) + N_{\Omega_k}(x^{k+1}), \quad (5.4)$$

onde $g : \mathbb{R}^m \rightarrow \mathbb{R}$ é uma função escalarização, ∂g denota um subdiferencial de g e $\{\alpha_k\}$ é uma sequência de números reais positivos. Em ambos [7] e [21], a convexidade de Ω_k é uma consequência da \mathbb{R}_+^m -quase convexidade de F . A nossa abordagem apresentada ao longo deste capítulo permite a possibilidade de não convexidade do conjunto Ω_k que, como veremos na última seção, desempenha um papel importante em aplicações.

Sabe-se que métodos do tipo proximal para otimização vetorial encontra separadamente uma solução do problema por vez, e não o conjunto solução inteiro. Como mencionado em Fukuda e Graña Drummond [59], e Fliege et al. [57], podemos obter uma tipo de aproximação do conjunto solução aplicando o método para diferentes pontos iniciais. Esse tipo de ideia também foi proposta em Burachik et al. [36].

5.2 MPP para funções vetoriais DC

Nesta seção iremos propor um algoritmo proximal para encontrar um ponto Pareto crítico de uma função vetorial DC $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ restrito a um conjunto (não vazio) convexo e fechado $\mathcal{D} \subset \mathbb{R}^n$. Denotaremos esse problema da seguinte forma:

$$\min_{x \in \mathcal{D}} F(x). \quad (5.5)$$

O algoritmo proposto em [33], restrito ao caso multiobjetivo de dimensão finita, gera uma sequência que satisfaz

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \varepsilon^k : x \in \Omega_k \right\}, \quad (5.6)$$

onde $\{\lambda_k\} \subset \mathbb{R}_{++}$, $\{\varepsilon^k\} \subset \mathbb{R}_+^m$, $\Omega_k = \{x \in \mathbb{R}^n : F(x) \preceq F(x^k)\}$ e “ argmin_w ” denota o conjunto de soluções Pareto fraca e “ \preceq ” a ordem parcial induzida pelo cone Pareto \mathbb{R}_+^m .

Note que o conjunto restrição Ω_k em (5.6) impõe o método ser de descida no sentido da ordem parcial “ \preceq ”, ou seja, $F(x^{k+1}) \preceq F(x^k)$, para todo $k \geq 0$. Recentemente, Ji et al. [75] estudou a convergência de um método proximal para funções vetoriais $F = (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ tal que, para cada $i \in \mathcal{I}$, $f_i(\cdot)$ é uma função DC, ou seja, $f_i(\cdot)$ pode ser escrita como diferença de duas funções convexas, digamos $f_i(x) = g_i(x) - h_i(x)$. Na k -ésima iterada o algoritmo proposto em [75] calcula um ponto $x^{k+1} \in S$ solução do seguinte subproblema (escalar):

$$\min_{x \in S} \left[\max_{i \in \mathcal{I}} [g_i(x) - \langle v_i^k, x - x^k \rangle] + \frac{\theta_k}{2} \|x - x^k\|^2 \right], \quad (5.7)$$

onde $v_i^k \in \partial h_i(x^k)$, com ∂h_i denotando um subdiferencial de h_i em x^k , $\theta_k > 0$ para todo k e, S um conjunto convexo e compacto. Note que (5.7) pode ser reescrito como

$$\min_{x \in S} \xi \left(F_k(x) + \frac{\theta_k}{2} \|x - x^k\|^2 e \right), \quad (5.8)$$

onde $F_k(x) = G(x) - V_k(x - x^k)$, $V_k \in \partial H(x^k)$, $e = (1, \dots, 1) \in \mathbb{R}^m$ e $\xi : \mathbb{R}^m \rightarrow \mathbb{R}$ é uma função escalarizadora dada por $\xi(x) = \max_{1 \leq j \leq m} \langle x, \beta_j \rangle$, onde $\{\beta_j\}$ é a base canônica do espaço \mathbb{R}^m . A terminação finita de um algoritmo similar a (5.8) foi anteriormente analisada em Bento et al. [21]. Claramente, a sequência gerada por (5.7) (e conseqüentemente (5.8)) não satisfaz $F(x^{k+1}) \preceq F(x^k)$, para todo $k \in \mathbb{N}$ como em (5.6). Essa propriedade de descida do método desempenha um papel importante para uma ampla classe de aplicações; veja por exemplo Bento and Soubeyran [26].

Antes de apresentar o algoritmo vejamos um resultado técnico que fornece uma condição necessária para um ponto ser um Pareto crítico de uma função vetorial.

No decorrer deste capítulo, denotaremos ∂_C o subdiferencial de Clarke definido no capítulo 2.

Lema 5.2.1 *Sejam $a, b \in \mathbb{R}_+^m$ tais que pelo menos um deles é não nulo. Assuma que Ω é um conjunto não vazio, convexo e fechado. Se $-(U^\top a + V^\top b) \in N_\Omega(x)$, para algum $U, V \in \partial_C F(x)$, então x é um ponto Pareto crítico de F .*

Demonstração: Provaremos apenas o caso em que $a, b \in \mathbb{R}_+^m \setminus \{0\}$. A prova do caso em que um dos vetores $a, b \in \mathbb{R}_+^m$ é zero segue o mesmo argumento. Tome $x \in \Omega$ e $a, b \in \mathbb{R}_+^m \setminus \{0\}$ tais que $-(U^\top a + V^\top b) \in N_\Omega(x)$. Suponhamos, por contradição, que x não seja um ponto Pareto crítico de F . Então, existe $y \in \Omega$ tal que

$$U(y - x) \prec 0 \quad \text{and} \quad V(y - x) \prec 0.$$

Logo, temos que $\langle a, U(y - x) \rangle < 0$ e $\langle b, V(y - x) \rangle < 0$, pois $a, b \in \mathbb{R}_+^m \setminus \{0\}$. Com isso, temos que $\langle U^\top a, y - x \rangle < 0$ e $\langle V^\top b, y - x \rangle < 0$. Somando as duas últimas desigualdades obtemos

$$\langle U^\top a + V^\top b, y - x \rangle < 0,$$

que contradiz o fato que $-(U^\top a + V^\top b) \in N_\Omega(x)$, e o resultado está provado. \blacksquare

5.2.1 O algoritmo

Nesta seção, assumimos que $G, H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são \mathbb{R}_+^m -convexas e H continuamente diferenciável. Assim, em (5.5), iremos considerar a função vetorial $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função DC dada por $F(x) := G(x) - H(x)$. No algoritmo a seguir iremos tomar $z \in \mathbb{R}_+^m \setminus \{0\}$ fixo e as sequências auxiliares $\{\lambda_k\} \subset \mathbb{R}_{++}$ e $\{\varepsilon^k\} \subset \mathbb{R}_+^m$ tais que $\|\varepsilon^k\| = 1$, para todo $k \geq 0$ e $\{\lambda_k\}$ é uma sequência limitada satisfazendo

$$\liminf_{k \in \mathbb{N}} \lambda_k \langle \varepsilon^k, z \rangle > 0. \quad (5.9)$$

A hipótese (5.9) pode ser facilmente verificada se $z \in \mathbb{R}_{++}^m$ e $\inf_{k \in \mathbb{N}} \lambda_k > 0$.

Algoritmo 5.1 *Passo 1: Escolha $x^0 \in \mathcal{D}$.*

Passo 2: Dado x^k , se x^k é um ponto Pareto crítico, então faça $x^{k+p} = x^k$, para todo $p \in \mathbb{N}$.

Passo 3: Caso contrário, tome como próxima iterada $x^{k+1} \in \mathcal{D}$ tal que

$$x^{k+1} \in \arg \min \left\{ \langle G(x) - JH(x^k)(x - x^k) + \frac{\lambda_k}{2} \|x - x^k\|^2 \varepsilon^k, z \rangle : x \in \Omega_k \right\}, \quad (5.10)$$

onde $\Omega_k = \{x \in \mathcal{D} : F(x) \preceq F(x^k)\}$.

Note que o Algoritmo 5.1 resolve em cada iteração um problema escalar. Essa abordagem é bastante conhecida na literatura e é chamada de *escalarização*. Como mencionado em [33, Remark 5], não existe uma diferença substancial entre a apresentação (no formato vetorial) de (5.6) e (5.10) (na forma escalarizada), pois toda solução do subproblema escalar (5.10) é uma solução Pareto fraca do subproblema (vetorial) (5.6) com $F(x) = G(x) - JH(x^k)(x - x^k)$; veja [94, Proposição 2.2]. Além disso, se $H \equiv 0$, levando em consideração [33, Teorema 2.1], o Algoritmo 5.1 coincide (no cenário multiobjetivo de dimensão finita) com algoritmo proposto em [33] para otimização vetorial convexa.

Vale a pena mencionar também que no Algoritmo 5.1, temos que $x^{k+1} \in \Omega_k$, para todo $k \geq 0$, isto é,

$$F(x^{k+1}) \preceq F(x^k), \quad (5.11)$$

que é uma propriedade importante para algumas aplicações, como veremos no final deste capítulo. Por outro lado, o algoritmo proposto em [75] para funções DC vetoriais satisfaz

$$\sum_{i \in \mathcal{I}} \alpha_i (f_i(x^{k+1}) - f_i(x^k)) \leq 0, \quad (5.12)$$

para algum $\alpha_i \geq 0$ com $\sum_{i \in \mathcal{I}} \alpha_i = 1$. Claramente, a propriedade de descida (5.11) implica em (5.12).

Observação 5.2.1 *Como mencionado em Huang e Yang [74], as funções vetoriais*

$$F(\cdot) \quad e \quad e^{F(\cdot)} := (e^{f_1(\cdot)}, \dots, e^{f_m(\cdot)}),$$

tem o mesmo conjunto de pontos Pareto fraco, onde e^α denota a aplicação exponencial calculada em $\alpha \in \mathbb{R}$. Esse resultado também se verifica para pontos Pareto críticos. Com isso, no que diz respeito a pontos Pareto críticos, podemos assumir sem perda de generalidade que $F \succ 0$. Por outro lado, toda função DC não negativa admite uma decomposição não negativa; veja [72]. Portanto, também podemos assumir sem perda de generalidade que $G \succeq 0$ e $H \succeq 0$.

Proposição 5.2.1 *O Algoritmo 5.1 está bem definido.*

Demonstração: Seja $\{x^k\}$ dada por (5.10) e $\phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ definida por

$$\phi(x) = \langle G(x), z \rangle - [JH(x^k)(x - x^k)]^\top z + \frac{\lambda_k}{2} \langle \varepsilon^k, z \rangle \|x - x^k\|^2 + I_{\Omega_k}(x), \quad (5.13)$$

onde I é a função indicadora. Como $G \succeq 0$ e tendo em vista que $(\lambda_k/2) \langle \varepsilon^k, z \rangle > 0$, segue que ϕ é coerciva. Como Ω_k é fechado, temos que ϕ é semicontínua inferior-

mente. Assim, existe $\tilde{x} \in \Omega_k$ tal que

$$\tilde{x} \in \operatorname{argmin}_{x \in \Omega_k} \phi(x).$$

Portanto, podemos tomar $x^{k+1} := \tilde{x}$ e a prova está concluída. \blacksquare

A seguir, iremos apresentar um resultado proposto por Minami [99] que fornece uma condição necessária para um ponto ser Pareto fraco de um problema de otimização vetorial. Aqui ficaremos restrito ao caso multiobjetivo de dimensão finita. Seja $\mathcal{D} \subset \mathbb{R}^n$ um conjunto não vazio e considere o problema de encontrar os pontos Pareto fraco de uma função vetorial $F = (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ em Ω , denotado por

$$\min_w \{F(x) : x \in \Omega\}, \quad (5.14)$$

onde

$$\Omega = \{x \in \mathcal{D} : s_j(x) \leq 0, \quad j \in \mathcal{I}\} \quad \text{com} \quad s_j : \mathbb{R}^n \rightarrow \mathbb{R}. \quad (5.15)$$

Teorema 5.2.1 *Assuma que Ω em (5.15) é convexo e fechado, e em (5.14) e (5.15) as funções $f_j, s_j : \mathbb{R}^n \rightarrow \mathbb{R}$, $j \in \mathcal{I}$, são localmente Lipschitz. Se $x^* \in \Omega$ é um ponto Pareto fraco de (5.14), então existem números reais $u_j \geq 0$, $v_j \geq 0$, com $j \in \mathcal{I}$, e $\tau > 0$ tais que*

$$0 \in \sum_{j \in \mathcal{I}} u_j \partial_C f_j(x^*) + \sum_{j \in \mathcal{I}} v_j \partial_C s_j(x^*) + \tau \partial_C d_{\mathcal{D}}(x^*), \quad (5.16)$$

$$\text{com} \quad \sum_{j \in \mathcal{I}} (u_j + v_j) = 1 \quad \text{e} \quad v_j s_j(x^*) = 0, \quad j \in \mathcal{I}. \quad (5.17)$$

Demonstração: Veja Minami [99, Theorem 3.1]. \blacksquare

A seguir utilizamos a estrutura vetorial do problema para produzir uma relação que utilizaremos no teorema de convergência. Esse resultado “faz o papel” da condição de otimalidade de primeira ordem do problema escalarizado usado na abordagem proposta em [33].

Proposição 5.2.2 *Para todo $k \in \mathbb{N}$, existem $A_k, B_k \in \mathbb{R}^{m \times n}$, $u^k, v^k \in \mathbb{R}_+^m$, $w^k \in \mathbb{R}^m$ e $\tau_k \in \mathbb{R}_{++}$ tais que*

$$[A_k - JH(x^{k-1})]^\top u^k + \lambda_{k-1} \langle \varepsilon^{k-1}, u^k \rangle (x^k - x^{k-1}) + B_k^\top v^k + \tau_k w^k = 0, \quad (5.18)$$

onde

$$A_k \in \partial_C G(x^k), \quad B_k \in \partial_C F(x^k), \quad w^k \in B[0, 1] \cap N_{\mathcal{D}}(x^k) \quad \text{e} \quad \|u^k + v^k\|_1 = 1. \quad (5.19)$$

Demonstração: Segue da definição do Algoritmo 5.1 que, para cada $k \in \mathbb{N}$, x^k é uma solução do problema escalar

$$\min \left\{ \left\langle G(x) - JH(x^{k-1})(x - x^{k-1}) + \frac{\lambda_{k-1}}{2} \|x - x^{k-1}\|^2 \varepsilon^{k-1}, z \right\rangle : x \in \Omega_{k-1} \right\}.$$

Logo, por [94, Proposição 2.2], temos que x^k é uma solução Pareto fraca do problema vetorial correspondente

$$\min_w \{ G(x) - JH(x^{k-1})(x - x^{k-1}) + \frac{\lambda_{k-1}}{2} \|x - x^{k-1}\|^2 \varepsilon^{k-1} : x \in \Omega_{k-1} \}.$$

Note que as funções

$$(f_k)_j(x) = g_j(x) - \langle \nabla h(x^k), x - x^k \rangle + \frac{\lambda_k}{2} \|x - x^k\|^2 \varepsilon_j^k, \text{ com } j \in \mathcal{I}, \quad (5.20)$$

e

$$(s_k)_j(x) = f_j(x) - f_j(x^k), \text{ with } j \in \mathcal{I}, \quad (5.21)$$

são localmente Lipschitz, para cada $j \in \mathcal{I}$, pois as funções (5.20) e (5.21) são funções DC, para cada $j \in \mathcal{I}$; veja [72, página 40]. Assim, as hipóteses do Teorema 5.2.1 estão satisfeita. Logo, vale (5.16) para as funções (5.20) e (5.21), ou seja,

$$0 \in [A_k - JH(x^{k-1})]^\top u^k + \lambda_{k-1} \langle \varepsilon^{k-1}, u^k \rangle (x^k - x^{k-1}) + B_k^\top v^k + \tau \partial_C d_{\mathcal{D}}(x^k), \quad \forall k \in \mathbb{N},$$

onde $A_k \in \partial_C G(x^k)$ e $B_k \in \partial_C F(x^k)$. Como $x^k \in \mathcal{D}$, o resultado segue do fato que $\partial_C d_{\mathcal{D}}(x^k) = B[0, 1] \cap N_{\mathcal{D}}(x^k)$; veja Burke et al. [37, Teorema 1]. \blacksquare

Observação 5.2.2 *Note que de (5.17) as sequências $\{u^k\}$, $\{v^k\}$ e $\{w^k\}$ são limitadas. De acordo com Bolte et al. [30, Remark 1], $\partial_C f$ é limitado em conjuntos compactos. Assim, temos que $\{A_k\}$ e $\{B_k\}$ são sequências limitadas desde que $\{x^k\}$ seja limitada. Com isso, se $\{\lambda_k\}$ e $\{x^k\}$ forem limitadas segue de (5.18) que $\{\tau_k\}$ também é limitada.*

Como uma consequência da Proposição 5.2.2, temos os seguintes critérios de paradas para o Algoritmo 5.1.

Corolário 5.2.1 *Seja $k_0 \in \mathbb{N}$ tal que $x^{k_0+1} = x^{k_0}$ ou $u^{k_0} = 0$. Então, x^{k_0} é um ponto Pareto crítico de F .*

Demonstração: Se existe $k_0 \in \mathbb{N}$ tal que $x^{k_0+1} = x^{k_0}$ (resp. $u^{k_0} = 0$), então segue de (5.18) que

$$[A_{k_0} - JH(x^{k_0-1})]^\top u^{k_0} + B_{k_0}^\top v^{k_0} + \tau_{k_0} w^{k_0} = 0. \quad (\text{resp. } B_{k_0}^\top v^{k_0} + \tau_{k_0} w^{k_0} = 0).$$

Como $\tau_{k_0} > 0$ e $w^{k_0} \in N_{\mathcal{D}}(x^{k_0})$, a última igualdade implica que

$$-[A_{k_0} - JH(x^{k_0-1})]^\top u^{k_0} - B_{k_0}^\top v^{k_0} \in N_{\mathcal{D}}(x^{k_0}) \quad (\text{resp. } -B_{k_0}^\top v^{k_0} \in N_{\mathcal{D}}(x^{k_0})).$$

Portanto, o resultado segue do Lema 5.2.1 levando em consideração que $u^{k_0}, v^{k_0} \in \mathbb{R}_+^m \setminus \{0\}$ (resp. $v^{k_0} \in \mathbb{R}_+^m \setminus \{0\}$). ■

5.2.2 Análise de convergência

Nesta seção faremos a análise de convergência do Algoritmo 5.1. Note que se o algoritmo termina em uma quantidade finita de passos, o mesmo termina em um ponto Pareto crítico. Com isso, iremos supor que a sequência $\{x^k\}$ gerada pelo Algoritmo 5.1 é infinita. Logo, iremos assumir que $x^{k+1} \neq x^k$ e $u^k \neq 0$, para todo $k \in \mathbb{N}$, em virtude do Corolário 5.2.1.

Proposição 5.2.3 *As seguintes propriedades se verificam:*

- i) a sequência $\{\langle F(x^k), z \rangle\}$ é estritamente decrescente;
- ii) $\lim_{k \rightarrow +\infty} \|x^{k+1} - x^k\| = 0$.

Demonstração: Como H é diferenciável e \mathbb{R}_+^m -convexa, temos que

$$\langle JH(x^k)(x^{k+1} - x^k), z \rangle \leq \langle H(x^{k+1}) - H(x^k), z \rangle, \quad \forall k \in \mathbb{N}. \quad (5.22)$$

Por outro lado, segue de (5.10) na definição do algoritmo que, para cada $k \in \mathbb{N}$,

$$\langle G(x^{k+1}) - JH(x^k)(x^{k+1} - x^k) + \frac{\lambda_k}{2} \|x^{k+1} - x^k\|^2 \varepsilon^k, z \rangle \leq \langle G(x^k), z \rangle, \quad (5.23)$$

que combinado com (5.22) e $F(x) = G(x) - H(x)$, implica

$$\langle F(x^{k+1}), z \rangle + \frac{\lambda_k}{2} \langle \varepsilon^k, z \rangle \|x^{k+1} - x^k\|^2 \leq \langle F(x^k), z \rangle, \quad \forall k \in \mathbb{N}. \quad (5.24)$$

Com isso, usando o fato que $(\lambda_k/2) \langle \varepsilon^k, z \rangle > 0$ e $x^{k+1} \neq x^k$, para todo $k \in \mathbb{N}$ temos que o primeiro item está provado. Para provar o segundo item observe que $F \succeq 0$ e, pelo item anterior, temos que $\{\langle F(x^k), z \rangle\}$ é convergente. Além disso, segue de (5.24) que

$$0 \leq \frac{\lambda_k}{2} \langle \varepsilon^k, z \rangle \|x^{k+1} - x^k\|^2 \leq \langle F(x^k), z \rangle - \langle F(x^{k+1}), z \rangle.$$

Portanto, o resultado segue aplicando o limite com $k \rightarrow +\infty$ na desigualdade acima, assumindo que (5.9) se verifica. ■

Para o caso particular $m = 1$, o resultado a seguir coincide com o resultado de convergência apresentado em [126].

Teorema 5.2.2 *Todo ponto de acumulação de $\{x^k\}$, caso exista, é um ponto Pareto crítico de F .*

Demonstração: Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e considere $\{x^{k_l}\}$ uma subsequência de $\{x^k\}$ convergindo para \hat{x} . Da Proposição 5.2.2, temos que existem seqüências $\{A_k\}, \{B_k\} \subset \mathbb{R}^{m \times n}$, $\{u^k\}, \{v^k\} \subset \mathbb{R}_+^m$, $\{w^k\} \subset \mathbb{R}^m$ e $\{\tau_k\} \subset \mathbb{R}_{++}$ satisfazendo (5.18). Como $\{x^{k_l}\}$ e $\{\lambda_k\}$ são limitadas, segue da Observação 5.2.2 que $\{A_k\}, \{B_k\}, \{u^k\}, \{v^k\}, \{w^k\}$ e $\{\tau_k\}$ também são limitadas. Assim, podemos assumir sem perda de generalidade que $A_{k_l} \rightarrow \hat{A}$, $B_{k_l} \rightarrow \hat{B}$, $u^{k_l} \rightarrow \hat{u}$, $v^{k_l} \rightarrow \hat{v}$, $w_{k_l} \rightarrow \hat{w}$ e $\tau_{k_l} \rightarrow \hat{\tau}$ (usamos o mesmo índice pois podemos extrair outras subsequências se necessário). De (5.18), temos

$$[A_{k_l} - JH(x^{k_l-1})]^\top u^{k_l} + \gamma_{k_l}(x^{k_l} - x^{k_l-1}) + B_{k_l}^\top v^{k_l} + \tau_{k_l} w^{k_l} = 0, \quad (5.25)$$

onde $\gamma_{k_l} = \lambda_{k_l-1} \langle \varepsilon^{k_l-1}, u^{k_l} \rangle$ e $\|u^{k_l} + v^{k_l}\|_1 = 1$. Note que $\{\gamma_k\}$ é limitada. Esse fato, combinado com a Proposição 5.2.3 implica que $\gamma_{k_l}(x^{k_l} - x^{k_l-1})$ converge para 0 quando $l \rightarrow +\infty$. Sabendo que $\partial_C f_i(x) = \partial_C g_i(x) - \nabla h_i(x)$, para cada $i \in \mathcal{I}$, e que $\partial_C f(\cdot)$ e $N_{\mathcal{D}}(\cdot)$ são fechados, aplicando o limite em (5.25) com $l \rightarrow +\infty$, temos que

$$\hat{C}^\top \hat{u} + \hat{B}^\top \hat{v} + \hat{\tau} \hat{w} = 0, \quad (5.26)$$

onde $\hat{C} := (\hat{A} - JH(\hat{x})) \in \partial_C F(\hat{x})$, $\hat{B} \in \partial_C F(\hat{x})$ e $\hat{\tau} \hat{w} \in N_{\mathcal{D}}(\hat{x})$. Com isso, segue de (5.26) que

$$-(\hat{C}^\top \hat{u} + \hat{B}^\top \hat{v}) \in N_{\mathcal{D}}(\hat{x}). \quad (5.27)$$

Como $\hat{u}, \hat{v} \in \mathbb{R}_+^m$ e $\|u^{k_l} + v^{k_l}\|_1 = 1$, para todo $l \in \mathbb{N}$, temos que $\hat{u} \neq 0$ ou $\hat{v} \neq 0$. Portanto, o resultado segue combinando (5.27) com o Lema 5.2.1. ■

5.3 MPP para funções vetoriais localmente Lipschitz

Nesta seção usaremos a abordagem estudada na seção anterior para analisar a convergência do método do ponto proximal em um ambiente mais amplo que o considerado na seção anterior. Mais precisamente, estudaremos a convergência do método para funções vetoriais localmente Lipschitz. Como a abordagem apresentada é nova faremos uma análise do caso quase convexo (e conseqüentemente o caso convexo) para mostrar que nossa abordagem recupera os resultados já existentes na literatura.

5.3.1 O algoritmo

Nesta seção iremos considerar $\mathcal{D} \subset \mathbb{R}^n$ um conjunto não vazio, convexo e fechado, e $F = (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função vetorial tal que cada função componente (escalar) $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i \in \mathcal{I}$, é localmente Lipschitz. Além disso, pela Observação 5.2.1, iremos assumir, sem perda de generalidade, que $F \succ 0$.

A seguir, definimos o método do ponto proximal para encontrar um ponto Pareto crítico de F em \mathcal{D} . Para isso, considere as sequências $\{\lambda_k\} \subset \mathbb{R}_{++}$ limitada e $\{\varepsilon^k\} \subset \mathbb{R}_{++}^m$ tais que

$$\liminf_{k \in \mathbb{N}} \lambda_k > 0, \quad \liminf_{k \in \mathbb{N}} \varepsilon_j^k > 0 \text{ e } \|\varepsilon^k\| = 1, \quad \forall k \geq 0, \quad j = 1, \dots, m. \quad (5.28)$$

Algoritmo 5.2 *Passo 1: Escolha $x^0 \in \mathcal{D}$.*

Passo 2: Dado x^k , se x^k é um ponto Pareto crítico, então faça $x^{k+p} = x^k$, para todo $p \in \mathbb{N}$.

Passo 3: Caso contrário, tome como próxima iterada $x^{k+1} \in \mathcal{D}$ tal que

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \varepsilon^k : x \in \Omega_k \right\}, \quad (5.29)$$

onde $\Omega_k = \{x \in \mathcal{D} : F(x) \preceq F(x^k)\}$.

Dado um problema de otimização multiobjetivo (PM) de uma função vetorial $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ restrito a um conjunto Ω e uma função escalar $g : \mathbb{R}^m \rightarrow \mathbb{R}$ definimos o problema de otimização escalar (PE) correspondente a (PM) como

$$\min\{g(F(x)) : x \in \Omega\}. \quad (5.30)$$

Dizemos que (PE) é uma *representação escalar fraca* de (PM), se $F(y) \prec F(x)$ implica em $g(F(y)) < g(F(x))$.

Exemplo 5.3.1 *Considere a função escalar $g : \mathbb{R}^m \rightarrow \mathbb{R}$ dada por*

$$g(y) = \max_{1 \leq i \leq m} \langle y, e_i \rangle, \quad (5.31)$$

onde $\{e_i\}$ é a base canônica do \mathbb{R}^m . Nesse caso, claramente (PE) é uma *representação escalar fraca* de (PM).

Proposição 5.3.1 *Para os problemas (PM) e (PE), se verifica:*

$$\operatorname{argmin}\{g(F(x)) : x \in \Omega\} \subseteq \operatorname{argmin}_w\{F(x) : x \in \Omega\},$$

se (PE) é uma *representação escalar fraca* de (PM).

Demonstração: [94, page 87]. ■

Seja \mathcal{G} uma família de funções do \mathbb{R}^m em \mathbb{R} . Dizemos que essa família é uma *escalarização completa fraca* de (PM) se, para toda solução Pareto fraca x^* de (PM), existe $g \in \mathcal{G}$ tal que x^* é solução escalar (PE) correspondente a g e (PM), e além disso

$$\arg \min\{g(F(x)) : x \in \Omega\} \subseteq \arg \min_w\{F(x) : x \in \Omega\}.$$

Exemplo 5.3.2 *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto não vazio e $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função vetorial tal que $F(x) \succ 0$, para todo $x \in \Omega$. A família de funções escalares $\{g_z\}_{z \in \mathbb{R}_{++}^m}$ dadas por*

$$g_z(y) = \max_{1 \leq i \leq m} \frac{\langle y, e_i \rangle}{\langle z, e_i \rangle} \quad (5.32)$$

é uma escalarização completa fraca do (PM) associado, onde $\{e_i\}$ é a base canônica do \mathbb{R}^m . Com efeito, dados $z \in \mathbb{R}_{++}^m$, se $F(y) \succ F(x)$, temos $F(y) - F(x) \in \mathbb{R}_{++}^m$. Logo,

$$\frac{\langle F(y) - F(x), e_i \rangle}{\langle z, e_i \rangle} > 0, \quad \forall i = 1, \dots, m.$$

Então,

$$\frac{\langle F(y), e_i \rangle}{\langle z, e_i \rangle} > \frac{\langle F(x), e_i \rangle}{\langle z, e_i \rangle}, \quad \forall i = 1, \dots, m,$$

e, em particular, $g_z(F(y)) > g_z(F(x))$. Assim, pela Proposição 5.3.1, temos que

$$\arg \min\{g_z(F(x)) : x \in \Omega\} \subseteq \arg \min_w\{F(x) : x \in \Omega\},$$

para todo $z \in \mathbb{R}_{++}^m$. Agora, para cada $x^ \in \arg \min_w\{F(x) : x \in \Omega\}$ tome $z = F(x^*) \succ 0$ e defina $g_z(x) = \max_{1 \leq i \leq m} \frac{\langle x, e_i \rangle}{\langle z, e_i \rangle}$. Pela definição de ponto Pareto fraco, temos que*

$$g_z(F(y)) = \max_{1 \leq i \leq m} \frac{\langle F(y), e_i \rangle}{\langle z, e_i \rangle} \geq 1,$$

e, além disso, $g_z(F(x^)) = 1$. Logo, $x^* \in \arg \min\{g_z(F(x)) : x \in \Omega\}$. Portanto, provamos que (5.32), com $z \in \mathbb{R}_{++}^m$, é uma escalarização completa fraca.*

Proposição 5.3.2 *O Algoritmo 5.2 está bem definido.*

Demonstração: O ponto inicial $x^0 \in \mathcal{D}$ é escolhido previamente. Assumindo que o algoritmo dispõe da k -ésima iterada, mostraremos a seguir que a $(k+1)$ -ésima iterada existe. Tome $z \in \mathbb{R}_{++}^m$, e defina a função escalar

$$g_z(y) = \max_{1 \leq i \leq m} \frac{\langle y, e_i \rangle}{\langle z, e_i \rangle},$$

onde $\{e_i\}$ é a base canônica do \mathbb{R}^m . Observe que $\min\{g_z(F_k(x)) : x \in \Omega_k\}$ é uma representação escalar fraca de $\min_w\{F_k(x) : x \in \Omega_k\}$, para todo $z \in \mathbb{R}_{++}^m$, onde

$F_k(x) = F(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \varepsilon^k$. Logo, pela Proposição 5.3.1, temos que

$$\arg \min \{g_z(F_k(x)) : x \in \Omega_k\} \subseteq \arg \min_w \{F_k(x) : x \in \Omega_k\}. \quad (5.33)$$

Iremos provar que $\arg \min \{g_z(F_k(x)) : x \in \Omega_k\}$ é não vazio. Com efeito,

$$\begin{aligned} g_z(F_k(x)) &= \max_{1 \leq i \leq m} \frac{\langle F(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \varepsilon^k, e_i \rangle}{\langle z, e_i \rangle} \\ &\geq \frac{\langle F(x), e_i \rangle}{\langle z, e_i \rangle} + \frac{\lambda_k \langle \varepsilon^k, e_i \rangle}{2 \langle z, e_i \rangle} \|x - x^k\|^2, \end{aligned} \quad (5.34)$$

para $i = 1, \dots, m$. Note que $\langle z, e_i \rangle > 0$ e $\lambda_k \langle \varepsilon^k, e_i \rangle > 0$, para cada $i = 1, \dots, m$ e $k \in \mathbb{N}$. Sendo $F \succ 0$, também temos que $\langle F(x), e_i \rangle > 0$, para $i = 1, \dots, m$. Combinando esses fatos com (5.34), temos que $g_z(F_k(\cdot))$ é coerciva, ou seja, $\lim_{\|x\| \rightarrow +\infty} g_z(F_k(x)) = +\infty$. Logo, sendo $g_z(F_k(\cdot))$ contínua e Ω_k fechado, temos que existe um ponto $\tilde{x} \in \Omega_k$ tal que

$$\tilde{x} \in \arg \min \{g_z(F_k(x)) : x \in \Omega_k\}.$$

Portanto, por (5.33) podemos tomar $x^{k+1} = \tilde{x}$ como a $(k+1)$ -ésima iteração, e a prova está concluída. \blacksquare

Proposição 5.3.3 *Para todo $k \in \mathbb{N}$, existem $A_k \in \mathbb{R}^{m \times n}$, $u^k, v^k \in \mathbb{R}_+^m$, $w^k \in \mathbb{R}^m$ e $\tau_k \in \mathbb{R}_{++}$ tais que*

$$A_k^\top (u^k + v^k) + \lambda_{k-1} \langle \varepsilon^{k-1}, u^k \rangle (x^k - x^{k-1}) + \tau_k w^k = 0, \quad (5.35)$$

onde

$$w^k \in B[0, 1] \cap N_{\mathcal{D}}(x^k) \quad e \quad \|u^k + v^k\|_1 = 1, \quad \forall k \in \mathbb{N}. \quad (5.36)$$

Demonstração: Da definição do algoritmo, temos que x^k é uma solução Pareto fraca do problema

$$\min_w \{F_{k-1}(x) : x \in \Omega_{k-1}\},$$

onde $F_{k-1}(x) = F(x) + \frac{\lambda_{k-1}}{2} \|x - x^{k-1}\|^2 \varepsilon^{k-1}$. Denotando $G_{k-1}(x) = F(x) - F(x^{k-1})$, segue da local Lipschitz continuidade de F , que todas as funções componentes

$$(g_{k-1})_j(\cdot) = f_j(\cdot) - f_j(x^{k-1}), \quad \text{com } j \in \mathcal{I}, \quad (5.37)$$

e

$$(f_{k-1})_j(\cdot) = f_j(\cdot) + \frac{\lambda_{k-1}}{2} \|\cdot - x^{k-1}\|^2 \varepsilon_j^{k-1}, \quad \text{com } j \in \mathcal{I}, \quad (5.38)$$

são localmente Lipschitz. Com isso, o resultado segue, similar à Proposição 5.2.2, aplicando o Teorema 5.2.1, para cada $k \in \mathbb{N}$ fixo, com $s_j = (g_{k-1})_j$ e $f_j = (f_{k-1})_j$ dadas em (5.37) e (5.38), respectivamente, levando em consideração que

$$\partial_C d_{\mathcal{D}}(x^k) = B[0, 1] \cap N_{\mathcal{D}}(x^k), \quad \forall k \in \mathbb{N}.$$

Dessa forma, $A_k^\top = [a_1^k \dots a_m^k]^\top$, onde $a_j^k \in \partial_C f_j(x^k)$, com $j \in \mathcal{I}$. \blacksquare

Como consequência do resultado anterior temos o seguinte critério de parada para o Algoritmo 5.2.

Corolário 5.3.1 *Seja $k_0 \in \mathbb{N}$ tal que $x^{k_0+1} = x_0^k$ ou $u^{k_0} = 0$. Então, x^{k_0} é um ponto Pareto crítico de F .*

Demonstração: Se existe $k_0 \in \mathbb{N}$ tal que $x^{k_0+1} = x_0^k$ ou $u^{k_0} = 0$, então segue de (5.35) que

$$A_{k_0}^\top v^{k_0} + \tau_{k_0} w^{k_0} = 0.$$

Como $\tau_{k_0} > 0$ e $w^{k_0} \in N_{\mathcal{D}}(x^{k_0})$, temos que

$$-A_{k_0}^\top v^{k_0} \in N_{\mathcal{D}}(x^{k_0}).$$

De (5.36) temos que $v^{k_0} \in \mathbb{R}_+^m \setminus \{0\}$. Sendo $A_{k_0} \in \partial_C F(x^{k_0})$, o resultado segue aplicando o Lema 5.2.1, com $a = 0$, $V = A_{k_0}$, $b = v^{k_0}$ e $x = x^{k_0}$. \blacksquare

5.3.2 Análise de convergência

Nesta seção iremos supor a seguinte hipótese na função F e no ponto inicial x^0 conhecida como \mathbb{R}_+^m -completeza:

(Hipótese de \mathbb{R}_+^m -Completeza) Para toda sequência $\{a^k\} \subset \mathbb{R}^n$, com $a^0 = x^0$ tal que $F(a^{k+1}) \preceq F(a^k)$, para todo $k \geq 0$, existe $a \in \mathbb{R}^n$ satisfazendo

$$F(a) \preceq F(a^k) \quad \forall k \geq 0.$$

Pela definição, o Algoritmo 5.2 para se estivermos em um ponto Pareto crítico. Para evitar a análise do caso mais óbvio, em virtude do Corolário 5.3.1 iremos supor que a sequência $\{x^k\}$ gerada pelo Algoritmo 5.2 satisfaz $x^{k+1} \neq x^k$ e $u^k \neq 0$, para todo $k \in \mathbb{N}$.

Caso localmente Lipschitz

Teorema 5.3.1 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 5.2. Então, todo ponto de acumulação de $\{x^k\}$, caso exista, é um ponto Pareto crítico de F .*

Demonstração: Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e considere $\{x^{k_l}\}$ uma subsequência de $\{x^k\}$ convergindo para \hat{x} . Segue do Exemplo 5.3.2 que a escalarização dada em (5.32) é uma escalarização completa fraca do problema vetorial associado. Sendo x^{k_l} , para cada $l \in \mathbb{N}$, uma solução Pareto fraca de

$$\min_w \left\{ F(x) + \frac{\lambda_{k_l-1}}{2} \|x - x^{k_l-1}\|^2 \varepsilon^{k_l-1} : x \in \Omega_{k_l-1} \right\},$$

temos que existe $\{z^{k_l}\} \subset \mathbb{R}_{++}^m$ tal que

$$\max_{1 \leq j \leq m} \left\{ \frac{\langle F(x^{k_l}) + \frac{\lambda_{k_l-1}}{2} \|x^{k_l} - x^{k_l-1}\|^2 \varepsilon^{k_l-1}, e_j \rangle}{\langle z^{k_l}, e_j \rangle} \right\} \leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle}, \quad (5.39)$$

para todo $l \in \mathbb{N}$. Como a desigualdade acima se mantém inalterada com a multiplicação de um escalar positivo, podemos assumir sem perda de generalidade que $\|z^{k_l}\| = 1$, para todo $l \in \mathbb{N}$, e com isso podemos supor que $z^{k_l} \rightarrow \hat{z}$ quando $l \rightarrow +\infty$ (podemos extrair uma outra subsequência se necessário). Segue de (5.39), aplicando a desigualdade de Cauchy-Schwarz, que

$$\frac{\langle F(x^{k_l}), e_i \rangle}{\langle z^{k_l}, e_j \rangle} + \frac{\lambda_{k_l-1}}{2} \|x^{k_l} - x^{k_l-1}\|^2 \langle \varepsilon^{k_l-1}, e_j \rangle \leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle},$$

para cada $j = 1, \dots, m$. Então, em particular, temos que

$$\frac{\langle F(x^{k_l}), e_j \rangle}{\langle z^{k_l}, e_j \rangle} + \frac{\lambda_{k_l-1}}{2} \|x^{k_l} - x^{k_l-1}\|^2 \min_{1 \leq j \leq m} \langle \varepsilon^{k_l-1}, e_j \rangle \leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle},$$

para cada $j = 1, \dots, m$. Mais uma vez, como a desigualdade acima se verifica para todo $j = 1, \dots, m$, em particular também se verifica para o índice onde o máximo do primeiro termo do lado esquerdo da desigualdade é atingido. Logo,

$$\begin{aligned} \frac{\lambda_{k_l-1}}{2} \|x^{k_l} - x^{k_l-1}\|^2 \min_{1 \leq j \leq m} \langle \varepsilon^{k_l-1}, e_j \rangle &\leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle} \\ &- \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l}), e_j \rangle}{\langle z^{k_l}, e_j \rangle}. \end{aligned} \quad (5.40)$$

Como $z^{k_l} \rightarrow \hat{z}$ quando $l \rightarrow +\infty$ e $\{F(x^k)\}$ é não crescente com $F \succ 0$, temos que o lado direito de (5.40) converge para 0 quando $l \rightarrow +\infty$. Com isso, sendo $0 < a \leq \lambda_k$ e $0 < c \leq \varepsilon_j^k$, para todo $k \in \mathbb{N}$ e $j = 1, \dots, m$, aplicando o limite com $l \rightarrow +\infty$ em (5.40) obtemos

$$(x^{k_l} - x^{k_l-1}) \rightarrow 0 \text{ quando } l \rightarrow +\infty. \quad (5.41)$$

Agora, aplicando a Proposição 5.3.3 para a subsequência $\{x^{k_l}\}$, temos que existem

$\{A_{k_l}\} \subset \mathbb{R}^{m \times n}$, $\{u^{k_l}\}$, $\{v^{k_l}\} \subset \mathbb{R}_+^m$, $\{w^{k_l}\} \subset \mathbb{R}^m$ e $\{\tau_{k_l}\} \subset \mathbb{R}_{++}$ satisfazendo

$$A_{k_l}^\top (u^{k_l} + v^{k_l}) + \lambda_{k_l-1} \langle \varepsilon^{k_l-1}, u^{k_l} \rangle (x^{k_l} - x^{k_l-1}) + \tau_{k_l} w^{k_l} = 0. \quad (5.42)$$

Note que sendo $\{\lambda_{k_l}\}$ e $\{x^{k_l}\}$ limitadas, similar à Observação 5.2.2, podemos assumir sem perda de generalidade que $A_{k_l} \rightarrow \hat{A}$, $u^{k_l} \rightarrow \hat{u}$, $v^{k_l} \rightarrow \hat{v}$ e $\tau_{k_l} \rightarrow \hat{\tau}$ quando $l \rightarrow +\infty$. Como $\{\lambda_{k_l-1} \langle \varepsilon^{k_l-1}, u^{k_l} \rangle\}$ é limitada, segue de (5.41) que $\lambda_{k_l-1} \langle \varepsilon^{k_l-1}, u^{k_l} \rangle (x^{k_l} - x^{k_l-1})$ converge para 0 quando $l \rightarrow +\infty$. Assim, fazendo $l \rightarrow +\infty$ em (5.42), temos que

$$\hat{A}^\top \hat{y} + \hat{\tau} \hat{w} = 0, \quad (5.43)$$

onde $\mathbb{R}_+^m \setminus \{0\} \ni \hat{y} := \hat{u} + \hat{v}$, $\hat{A} \in \partial_C F(\hat{x})$ e $\hat{w} \in N_{\mathcal{D}}(\hat{x})$, pois $\partial_C F(\cdot)$ e $N_{\mathcal{D}}(\cdot)$ são fechados. Logo, segue de (5.43) que

$$-\hat{A}^\top \hat{y} \in N_{\mathcal{D}}(\hat{x}).$$

Portanto, pelo Lema 3.2.1, temos que \hat{x} é um ponto Pareto crítico de F . ■

Caso quase convexo

Nesta seção iremos estudar a convergência do Algoritmo 5.2 supondo adicionalmente que $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é \mathbb{R}_+^m -quase convexa. Recentemente, esse caso foi analisado por Apolinário et al. [7] para o contexto multiobjetivo de dimensão finita. No algoritmo proposto em [7] na k -ésima iterada tem-se que

$$0 \in \partial_C \left(\langle F(\cdot), z^k \rangle + \frac{\lambda_k}{2} \langle \varepsilon^k, z^k \rangle \|\cdot - x^k\|^2 \right) (x^{k+1}) + N_{\Omega_k}(x^{k+1}). \quad (5.44)$$

O caso \mathbb{R}_+^m -quase convexo para otimização multiobjetivo também foi estudado por Bento et al. [21] em que os autores propõe o seguinte processo iterativo

$$x^{k+1} \in \arg \min_{x \in \mathbb{R}^n} f \left(F(x) + \delta_{\Omega_k} + \frac{\lambda_k}{2} \|x - x^k\|^2 e \right), \quad (5.45)$$

onde $e = (1, \dots, 1) \in \mathbb{R}^m$, a função escalarização $f : \mathbb{R}^m \rightarrow \mathbb{R}$ é dada por $f(y) = \max_{1 \leq i \leq m} \langle y, e_i \rangle$ e $\{e_i\}$ é a base canônica do \mathbb{R}^m .

Observação 5.3.1 *A análise de convergência de ambos os métodos (5.44) e (5.45) é baseada no conceito de Fejér convergência, usando a mesma abordagem proposta em [33]. Nesses trabalhos, a função escalarização desempenha um papel importante nas demonstrações pois a análise de convergência é baseada na condição de otimalidade de primeira ordem do problema escalarizado; veja [7, Proposição 3.4.1], [33, Teorema 3.1] e [21, Teorema 4.1].*

A seguir, veremos que usando nossa abordagem obtemos o mesmo resultado de convergência como em [7] e [21] para a sequência gerada pelo Algoritmo 5.2 para o caso \mathbb{R}_+^m -quase convexa utilizando o conceito de Fejér convergência, mas sem depender da função escalarização.

Teorema 5.3.2 *A sequência $\{x^k\}$ gerada pelo Algoritmo 5.2 converge para um ponto Pareto crítico de F .*

Demonstração: Dividimos a prova em cinco passos.

Passo1 (Fejér convergência): Defina $E \subset \mathcal{D}$ como

$$E = \{x \in \mathcal{D} : F(x) \preceq F(x^k), \quad \forall k \in \mathbb{N}\}.$$

Segue da hipótese de \mathbb{R}_+^m -completeza de F em x^0 que E é não vazio. Considere um ponto arbitrário $x^* \in E$. Logo, $x^* \in \Omega_k$, para todo $k \in \mathbb{N}$. Denotando por $\gamma_{k+1} = \lambda_k \langle \varepsilon^k, u^{k+1} \rangle$, temos que $\gamma_{k+1} > 0$, para todo $k \in \mathbb{N}$, pois $\lambda_k > 0$, $\varepsilon^k \in \mathbb{R}_{++}^m$ e $u^k \in \mathbb{R}_+^m \setminus \{0\}$, para todo $k \in \mathbb{N}$. Como

$$\|x^k - x^*\|^2 = \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x^*\|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - x^* \rangle,$$

segue de (5.35) que

$$\begin{aligned} \|x^k - x^*\|^2 &= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x^*\|^2 \\ &\quad + \frac{2}{\gamma_{k+1}} \langle A_{k+1}^\top (u^{k+1} + v^{k+1}) + \tau_{k+1} w^{k+1}, x^{k+1} - x^* \rangle \\ &= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x^*\|^2 \\ &\quad + \frac{2}{\gamma_{k+1}} \sum_{i=1}^m (u_i^{k+1} + v_i^{k+1}) \langle a_i^{k+1}, x^{k+1} - x^* \rangle + \tau_{k+1} \langle w^{k+1}, x^{k+1} - x^* \rangle, \end{aligned} \tag{5.46}$$

onde $a_i^{k+1} \in \partial_C f_i(x^{k+1})$, para todo k e $i = 1, \dots, m$. Por outro lado, sendo F uma função \mathbb{R}_+^m -quase convexa, $x^* \in \Omega_{k+1}$ e $\gamma_k > 0$, para todo k , temos que

$$\frac{2}{\gamma_{k+1}} \sum_{i=1}^m (u_i^{k+1} + v_i^{k+1}) \langle a_i^{k+1}, x^{k+1} - x^* \rangle \geq 0. \tag{5.47}$$

Além disso, sendo $w^{k+1} \in N_{\mathcal{D}}(x^{k+1})$ e $\tau_k > 0$, obtemos

$$\tau_{k+1} \langle w^{k+1}, x^{k+1} - x^* \rangle \geq 0. \tag{5.48}$$

Logo, usando (5.47) e (5.48) em (5.46), temos

$$\|x^{k+1} - x^k\|^2 \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2, \quad \forall k \in \mathbb{N} \quad (5.49)$$

que implica $\|x^{k+1} - x^*\| \leq \|x^k - x^*\|$, para todo $k \in \mathbb{N}$ e $x^* \in E$. Isso mostra que $\{x^k\}$ é Fejér convergente ao conjunto E .

Passo 2 (Os pontos de acumulação de $\{x^k\}$ pertencem a E): Como $\{x^k\}$ é Fejér convergente ao conjunto E , segue da Proposição 2.1.8 que $\{x^k\}$ é limitada. Seja x^* um ponto de acumulação de $\{x^k\}$. Pela definição do algoritmo temos que $F(x^{k+1}) \preceq F(x^k)$, para todo k . Logo, da continuidade de F , temos que $F(x^*) \preceq F(x^k)$, para todo k , que significa que $x^* \in E$.

Passo 3 (Convergência da sequência): Esse passo segue diretamente da Proposição 2.1.8 tendo em vista os passos anteriores.

Passo 4 (Proximidade das iteradas consecutivas): Considere $\{x^k\}$ convergindo para \hat{x} . Da desigualdade triangular, temos que

$$\|x^{k+1} - x^k\| \leq \|x^{k+1} - \hat{x}\| + \|x^k - \hat{x}\|, \quad \forall k \in \mathbb{N}. \quad (5.50)$$

Como o lado direito de (5.50) converge para zero quando $k \rightarrow +\infty$, temos que

$$\lim_{k \rightarrow +\infty} \|x^{k+1} - x^k\| = 0. \quad (5.51)$$

Passo 5 (Ponto limite Pareto crítico): A prova desse passo segue o mesmo argumento que o Teorema 5.3.1 a partir de (5.42). Isso termina a prova do teorema. ■

Observação 5.3.2 *Sabemos que sob a hipótese de \mathbb{R}_+^m -convexidade, os conceitos Pareto fraco e Pareto crítico são equivalentes. Assim, no contexto \mathbb{R}_+^m -convexo, pelo teorema anterior, temos que o Algoritmo 5.2 converge para um ponto Pareto fraco de F recuperando do resultado obtido em [33] para o caso multiobjetivo de dimensão finita.*

5.4 Aplicação em grupos dinâmicos

5.4.1 Problema de produção: dinâmica de melhoria cooperativa

Nesta seção retomamos o problema produção discutido no capítulo anterior, agora sob a perspectiva do Algoritmo 5.1 para funções DC multiobjetivo. Esse caso é mais realista. Nesse sentido, consideramos $\mathcal{I} = \{1, 2, \dots, m\}$ um grupo de produtores. O objetivo de cada um deles é uma função retorno “a ser aumentada” (função lucro,

utilidade) $\tilde{f}_i(x) = \tilde{g}_i(x) - \tilde{h}_i(x)$, onde $\tilde{g}_i(x) \in \mathbb{R}_+$ e $\tilde{h}_i(x) \in \mathbb{R}_+$ se referem à suas receitas e custos escalares.

A variável de decisão do grupo é o vetor $x \in \mathbb{R}^n$ que deve satisfazer algumas restrições. Cada agente quer um retorno tão alto quanto possível. O objetivo do grupo \mathcal{I} é se aproximar e atingir um ponto Pareto crítico. A função objetivo do grupo é $\tilde{F}(x) = \tilde{G}(x) - \tilde{H}(x) \in \mathbb{R}^m$, onde $\tilde{F}(x) = (\tilde{f}_1(x), \dots, \tilde{f}_m(x))$, $\tilde{G}(x) = (\tilde{g}_1(x), \dots, \tilde{g}_m(x))$ e $\tilde{H}(x) = (\tilde{h}_1(x), \dots, \tilde{h}_m(x))$. Em um contexto dinâmico cooperativo, todos os agentes de um grupo aceitam mudar de uma posição anterior $x = x^k$ para a próxima $y = x^{k+1}$ somente se seus retornos não decrescerem, ou seja, se

$$\tilde{f}_i(x^{k+1}) \geq \tilde{f}_i(x^k), \quad \forall i \in \mathcal{I}.$$

Caso contrário, alguns agentes saem do grupo ou resistem em mudar. Isso define uma dinâmica de melhoria cooperativa

$$x^{k+1} \in \Omega(x^k),$$

onde

$$\Omega(x^k) = \left\{ y \in \mathcal{D} : \tilde{F}(y) \succeq \tilde{F}(x^k) \right\}.$$

O problema de grupos dinâmicos cooperativos é encontrar uma dinâmica de melhoria cooperativa $x^{k+1} \in \Omega(x^k)$ que converge (se aproxima e atinge) para um ponto Pareto crítico.

Na prática, em um contexto dinâmico, na maioria dos problemas retornos crescentes prevalecem, vindo de custos fixos, aprendendo com a repetição, etc.. Isso significa que cada receita marginal e função custo é decrescente. Então, todas as funções receitas \tilde{g}_i e funções custos \tilde{h}_i são funções côncavas. Nesse caso, o retorno de cada agente é uma diferença de duas funções côncavas.

Considere as funções retornos “a ser diminuída” $f_i(y) = -\tilde{f}_i(y)$, com $g_i(y) = -\tilde{g}_i(y)$ e $h_i(y) = -\tilde{h}_i(y)$, que é escrita como diferença de funções convexas $f_i(y) = g_i(y) - h_i(y)$. Dessa forma, fornecemos um algoritmo multiobjetivo DC que se aproxima e atinge um ponto Pareto crítico, com a importante condição: o algoritmo segue uma dinâmica de melhoria cooperativa $x^{k+1} \in \Omega(x^k)$. Esse contexto comportamental foi estudado nos trabalhos Bento et al. [21], Bonnel et al. [33] e Choung et al. [42]. Porém, a convexidade dos conjuntos $\Omega(x^k)$, para todo $k \in \mathbb{N}$, é necessária para obter resultados de convergência. Na nossa abordagem não utilizamos a hipótese de convexidade de $\Omega(x^k)$ o que nos fornece uma abordagem não convexa tanto do ponto de vista da função objetivo quanto do ponto de vista do conjunto $\Omega(x^k)$. Uma motivação em um contexto dinâmico para considerar o conjunto de restrições $\Omega(x^k)$ é dado em Bento et al. [21]. O algoritmo proposto por Ji et al. [75]

para otimização multiobjetivo DC não pode ser visto como uma dinâmica de melhoria cooperativa, pois não segue $x^{k+1} \in \Omega(x^k)$, e isso justifica por que o Algoritmo 5.1 é melhor adaptado para aplicações, por exemplo em Ciência de Comportamento, que o método (5.7).

Nesta seção, como no capítulo anterior, estamos supondo que os agentes tem conhecimento limitado (parcial). Também assumimos que eles não conhecem a função custo enquanto ele estão cientes de suas funções receitas. Por isso, consideramos uma aproximação linear da função objetivo no algoritmo.

Processo de mudança satisfatória

Consideramos o contexto de dinâmica de comportamento humano a nível de um agente com vários objetivos ou a nível de uma organização onde cada agente tem seu próprio grupo dinâmico de objetivos. Usaremos a recente abordagem (VR) proposta por Soubeyran [117–119], onde em cada período, agentes aceitam mudar se suas mudanças são satisfatórias. Esse é o caso em que suas motivações (vetoriais) para mudar $M = U[A] \in \mathbb{R}^m$ são suficientemente grande com respeito às suas resistências (vetoriais) para mudar $R = D[I] \in \mathbb{R}_+^m$, isto é, $M \geq \xi R$. Suficientemente grande significa que $\xi > 0$ suficientemente grande (raio de satisfação).

A seguir, mostramos rapidamente como, no contexto da abordagem (VR), o Algoritmo 5.1 representa um processo “*stay*” e “*change*” satisfatório.

Consideramos o caso linear-quadrático onde a função utilidade $U[A] = A$ de suas vantagens para mudar $A \in \mathbb{R}^m$ são suficientemente grandes com respeito à função inutilidade $D[I] = I^2$ de suas inconveniências para mudar $I \in \mathbb{R}_+^m$, ou seja, $A \geq \xi I^2$, onde I^2 é o vetor de quadrados de cada componente do vetor I . Em cada período, vantagens para mudar partindo de “repetindo a última ação x ” para “fazendo uma nova ação y ” são definidas como a diferença $A = A(x, y) = P(y) - P(x)$ entre a função vetorial retorno atual (a ser aumentada) $P(y) \in \mathbb{R}^m$ para fazer a ação y e a função vetorial retorno atual $P(x) \in \mathbb{R}^m$ para repetir a última ação x . Aqui, $F(\cdot) = -P(\cdot)$ representa uma função vetorial de necessidades não satisfeitas. Então, $A(x, y) = F(x) - F(y)$. Inconveniência para mudar $I(x, y) = C(x, y) - C(x, x)$ se refere a diferença entre o custo $C(x, y) \in \mathbb{R}_+^m$ de estar apto a mudar de x para y e o custo $C(x, x) \in \mathbb{R}_+^m$ de estar apto a repetir a ação x . Neste capítulo, $C(x, y) = \|y - x\|^2 \varepsilon$, onde $\varepsilon \in \mathbb{R}_{++}^m$ é um vetor de ações. Então, uma mudança de x para y é satisfatória se $A(x, y) \geq \xi C(x, y)^2$, ou seja,

$$F(x) - F(y) \succeq \xi \|y - x\|^2 \varepsilon.$$

Seja $z \in \mathbb{R}_{++}^n$ um vetor de pesos que ajuda a adicionar diferentes vantagens para mudar e diferentes inconveniências para mudar e que permite comparar suas for-

mulações escalarizadas. Nesse caso, uma mudança é satisfatória se

$$\langle F(x) - F(y), z \rangle \geq \xi \|y - x\|^2 \langle \varepsilon, z \rangle.$$

Seja k e $k + 1$ o período anterior e atual, respectivamente., onde $x = x^k$ e $y = x^{k+1}$. Com isso, no período atual, uma mudança de repetir a ação anterior x^k para fazer uma nova ação x^{k+1} é satisfatória se

$$\langle F(x^k) - F(x^{k+1}), z \rangle \geq \xi_k \|x^{k+1} - x^k\|^2 \langle \varepsilon^k, z \rangle,$$

onde $\{\varepsilon^k\}$ e $\xi_k = \lambda_k/2$ podem ser tomadas como no Algoritmo 5.1.

Isso prova que o Algoritmo 5.1 é um exemplo específico de um processo “*stay*” e “*change*” satisfatório. Nesse contexto de grupos dinâmicos cooperativos, a seção 5.2.2 garante que quando todos os agentes requerem, em cada período, que seus retornos não decresçam, uma transição “*stay*” e “*change*” satisfatória se aproxima e atinge (converge para) um ponto Pareto crítico.

5.4.2 Problema de compromisso

Nesta seção apresentamos uma aplicação do Algoritmo 5.2 para função multiobjetivo localmente Lipschitz. Considere o famoso problema de compromisso, onde um grupo de agente tenta minimizar a distância de suas atuais posições para o ponto ideal para o grupo; veja Gearhart [62]. Nesse caso, as distâncias são localmente Lipschitz; veja por exemplo [104, 105]. Em uma ampla classe de aplicações tais distâncias são usadas como funções objetivos, por exemplo Teoria de Localização, Teoria de Utilidade, Teoria de Consumidor, etc.. Daremos uma formulação dinâmica do problema (estático) de compromisso de grupo para modelar como, partindo de um ponto inicial, um grupo de agentes com retornos inter-relacionados podem se aproximar e atingir uma transição aceitável, um limite desejado, definido como uma solução de compromisso. Esse é um problema muito importante relacionado a Jogos Dinâmicos Cooperativos.

Considere um grupo de produtores $i \in \mathcal{I} = \{1, \dots, m\}$. A variável de decisão do grupo é o vetor $x \in \mathbb{R}^n$ que deve satisfazer algumas restrições $x = (x_1, \dots, x_n) \in \mathcal{D} \subset \mathbb{R}^n$. O objetivo de cada um deles é um retorno “a ser aumentado” (lucro, utilidade) $h_i(x) \in \mathbb{R}_+$. O objetivo (vetorial) do grupo é $H(x) \in \mathbb{R}^m$, onde $H(x) = (h_1(x), \dots, h_m(x))$. Assim, o subconjunto de retornos vetoriais viáveis do grupo, ou seja, o subespaço retorno do grupo é $H(\mathcal{D}) = \{H(x) : x \in \mathcal{D}\} \subset \mathbb{R}^m$. Cada agente quer um retorno tão grande quanto possível.

Suponha que o retorno máximo de cada agente do grupo é limitado superiormente, ou seja, $\bar{h}_i = \sup \{h_i(x) : x \in \mathcal{D}\} < +\infty$, $i \in \mathcal{I}$. Assim, o retorno vetorial

$\bar{H} = (\bar{h}_1, \dots, \bar{h}_m)$ é o retorno ideal (ou utópico) do grupo. Normalmente, o retorno vetorial ideal não é viável, que significa que $\bar{H} \notin H(\mathcal{D})$. Considere o retorno “a ser diminuído”

$$f_i(x) = \bar{h}_i - h_i(x) \geq 0, \quad i \in \mathcal{I},$$

que se refere, em Psicologia, à função “*gap*” de insatisfação f_i . Essas funções medem quanto cada retorno individual $h_i(x)$, com $x \in \mathcal{D}$, falha em atingir seu valor máximo (ideal ou utópico) \bar{h}_i . Esses “*gap*” (vetoriais) de insatisfações

$$F(x) = (f_1(x), \dots, f_m(x)) = \bar{H} - H(x) \succeq 0$$

geram arrependimentos ou insatisfações com respeito ao retorno ideal. Uma solução de compromisso (com relação a uma norma) é algum ponto viável $x^* \in \mathcal{D}$ que minimiza toda a insatisfação do grupo, em outras palavras, minimiza a distância entre o retorno vetorial ideal \bar{H} e o subespaço retorno $H(\mathcal{D})$. Para problemas de compromisso em tomada decisão multiobjetivo veja os conhecidos trabalhos Gearhart [62] e Goetzmann et al. [63].

Usando esse modelo estático de compromisso, vamos considerar um modelo dinâmico simples. Isso inclui um ponto inicial, uma transição aceitável e alguns limites desejados. Esse modelo dinâmico considera que transições são aceitáveis se, em cada período, todos os membros de um grupo melhora seus retornos. Caso contrário, alguns agentes saem do grupo ou resistem em mudar. O limite desejado do grupo \mathcal{I} é se aproximar e atingir um ponto limite que por sua vez se aproxima tanto quanto possível do ponto ideal. Em um contexto dinâmico cooperativo, todos os agentes de um grupo aceitam mudar de uma posição anterior $x = x^k$ para a próxima $y = x^{k+1}$ somente se seus retornos não decrescerem, ou seja, se

$$h_i(x^k) \leq h_i(x^{k+1}), \quad \forall i \in \mathcal{I} \quad \iff \quad H(x^k) \preceq H(x^{k+1}).$$

Isso define uma dinâmica de melhoramento cooperativo $x^{k+1} \in \Omega(x^k)$, onde

$$\Omega(x^k) = \{x \in \mathcal{D} : H(x^k) \preceq H(x)\}.$$

O problema de grupo dinâmico cooperativo é encontrar uma dinâmica de melhoramento $x^{k+1} \in \Omega(x^k)$ que se aproxima e atinge (converge para) um posição limite desejada próxima o suficiente do ponto ideal; veja Lewin [85, 86] para detalhes sobre problemas de gestão com “grupos dinâmicos” e “mudança organizacional” em Psicologia e Ciências de Gestão.

A natureza do problema de grupos dinâmicos depende fortemente da natureza das funções objetivos que determinam as propriedades do conjunto de melhoramento

e outras restrições relevantes. Funções objetivos podem ser convexas ou côncavas, quase convexas ou quase côncavas, diferença de convexas ou diferença de côncavas, Lipschitz ou localmente Lipschitz. Na recente abordagem (VR) de comportamento humano (veja Soubeyran [117–119]), funções retornos Lipschitz e localmente Lipschitz são interessantes por duas razões:

1. Significam que quando inconveniências para mudar são baixas não se pode esperar grandes vantagens para mudar (Não existe almoço grátis!);
2. Essas funções são facilmente estimadas localmente.

O primeiro item é uma hipótese bastante razoável. O segundo item ajuda agentes com poucas informações, que conhecem suas funções retornos apenas em um dado ponto, a estarem aptos a encontrar, em cada período, mudanças de melhoramento. Além disso, funções Lipschitz f tem funções estimativas superior côncavas

$$y \in \mathcal{D} \mapsto u_0(y) = f(x_0) - L\|y - x_0\|,$$

em cada $x_0 \in \mathcal{D}$. Tudo isso nos estimula a analisar propriedades de convergência de um algoritmo para funções multiobjetivos localmente Lipschitz.

Capítulo 6

Conclusão

Neste trabalho estudamos dois tipos de métodos de descida para diferentes tipos de problemas e em diferentes ambientes. Para alguns deles foram apresentados simples experimentos numéricos para demonstrar suas viabilidades computacionais e para todos os métodos foram apresentadas diferentes aplicações, tanto teóricas como aplicações práticas em teoria comportamental.

Os métodos estudados foram o método de máxima descida (MMD) e o método do ponto proximal (MPP). Para o MMD apresentamos uma extensão do método para o cenário das variedades de Hadamard bem como uma aplicação para calcular L^p -centro de massa Riemanniano de um conjunto de dados. Nesse mesmo cenário de variedades de Hadamard, um MPP foi proposto para encontrar um ponto crítico de uma função não convexa (não diferenciável) escrita com a diferença de duas funções convexas. Os resultados ampliam a aplicação do MPP para uma classe mais ampla que funções convexas. Como aplicação apresentamos como o método pode ser usado para resolver um problema de maximização com restrições em uma variedade de Hadamard. Ainda sobre o MPP foi proposto um algoritmo generalizado usando uma quase distância como regularização. Como aplicação desse algoritmo propomos como resolver, de uma forma dinâmica, o problema de produção de uma companhia usando a recente abordagem “*variational rationality*” (VR) para comportamentos humanos. Ainda sobre o MPP, agora no contexto de otimização multiobjetivo, apresentamos uma nova análise de convergência do método, diferente das abordagens existentes, que possibilitou ampliar a aplicação do MPP para funções localmente Lipschitz com valores vetoriais e que quando restrita ao caso convexo ou quase convexo recuperou os resultados existentes na literatura. Como aplicações desse método, consideramos os problemas de grupos dinâmicos e problemas de compromisso, relativos à Teoria de Comportamento.

Como consequência dos resultados obtidos escrevemos alguns trabalhos científicos que foram publicados e estão submetidos ou em preparação para possível publicação. Além disso, listamos a seguir algumas propostas de trabalhos futuros

consequências dos resultados obtidos.

Trabalhos futuros

- O Algoritmo 3.1 para diferença de funções convexas é um algoritmo do tipo proximal-subgradiente, pois dada uma função $f(x) = g(x) - h(x)$, o método faz um passo subgradiente para a função convexa h e usa essa informação para calcular a próxima iteração via método proximal na função g . Análise de convergência para algoritmos do tipo proximal-subgradiente ou proximal-gradiente para soma de funções convexas tem sido amplamente estudado por vários autores. As técnicas usadas na convergência do Algoritmo 3.1 serão usadas para estudar a convergência de um método do tipo proximal-gradiente em variedades de Riemann para soma de funções convexas. Esse método pode ser usado para resolver problemas de localização em variedade de Riemann tais como o problema de localização de Fermat-Weber, em que dados m pontos distintos no \mathbb{R}^n , um ponto é solução do problema de Fermat-Weber se minimiza em \mathbb{R}^n o problema

$$\min f(x) = \sum_{i=1}^m w_i \|x - a_i\|,$$

onde $w_i \geq 0$ são pesos. Quando os pontos a_i são substituídos por conjuntos convexos e fechados Ω_i temos uma forma mais geral do problema anterior conhecido como problema de Fermat-Torricelli ou problema de Heron.

- Ainda sobre o Algoritmo 3.1 podemos estudar taxa de convergência ou complexidade do algoritmo para a sequência $\{f(x^k)\}$ que é convergente. Além disso, propor uma busca para acelerar a convergência dessa sequência tal como: dada $f(x) = g(x) - h(x)$ tomando $\{c_k\}$ uma sequência auxiliar, $\alpha > 0$, $0 < \beta < 1$, $w^k \in \partial h(x^k)$, $z^k := \exp_{x^k}(c_k w^k)$,

$$y^k := \arg \min_{x \in M} \left\{ g(x) + \frac{1}{2c_k} d^2(x, z^k) \right\}$$

e $d^k := \exp_{x^k}^{-1} y^k$. Enquanto $f(\exp_{y^k} c_k d^k) > f(y^k) - \alpha c_k \|d^k\|^2$, faça $c_k = \beta c_k$. Tome $x^{k+1} := \exp_{y^k} c_k d^k$. Portanto, devemos ter

$$f(x^{k+1}) \leq f(y^k) - \alpha c \|d^k\|^2.$$

Isso acelera a convergência da sequência $\{f(x^k)\}$ comparada com a gerada pelo Algoritmo 3.1 que coincide com $\{f(y^k)\}$ acima.

- Na aplicação tanto do Algoritmo 5.1 quanto do Algoritmo 5.2 em grupos dinâmicos, o custo $C(x, y) \in \mathbb{R}_+^m$ de estar apto a mudar de x para y é dado por

$C(x, y) = \|y - x\|^2 \varepsilon$. Nesse caso, o custo de estar apto a mudar de x para y e o custo de estar apto a mudar de y para x são iguais, ou seja, $C(x, y) = C(y, x)$. No caso mais realista, temos que $C(x, y)$ pode ser diferente de $C(y, x)$ e o custo $C(x, x) \in \mathbb{R}_+^m$ de estar apto a repetir a ação x pode ser diferente do vetor zero. Portanto, uma versão dos Algoritmo 5.1 e 5.2 em que $C(x, y) = q^2(x, y)\varepsilon$, onde $q(\cdot, \cdot)$ é uma quase distância é mais realista para aplicações em Teoria Comportamental.

Referências Bibliográficas

- [1] ABSIL, P. A., BAKER, C. G., GALLIVAN, K. A., “Trust-region methods on Riemannian manifolds”, *Foundations of Computational Mathematics*, v. 7(3), pp. 303–330, 2007.
- [2] ABSIL, P. A., MAHONY, R., SEPULCHRE, R., “Optimization algorithms on matrix manifold”, *Princeton University Press*, 2009.
- [3] ADLER, R.L., DEDIEU, J.P., MARGULIES, J.Y., MARTENS, M., SHUB, M., “Newton’s method on Riemannian manifolds and a geometric model for the human spine”, *IMA J. Numer. Anal.*, v. 22, pp. 359–390, 2002.
- [4] AFSARI, B., TRON, R., VIDAL, R., “On the convergence of gradient descent for finding the Riemannian center of mass”, *SIAM J. Control Optim.*, v. 51(3), pp. 2230–2260, 2013.
- [5] AHN, S., FESSLER, J.A., BLATT, D., HERO, A.O., “Convergent incremental optimization transfer algorithms: Application to tomography”, *IEEE T. Med. Imaging.*, v. 25, pp. 283–296, 2006.
- [6] AMBROSIO, L., BERTRAND, J., “DC Calculus”, *Preprint*, 2015.
- [7] APOLINÁRIO, H.C.F., PAPA QUIROZ, E.A., OLIVEIRA, P.R., “A Scalarization Proximal Point Method for Quasiconvex Multiobjective Minimization”, *J. Glob. Optim.*, v. 64, pp. 79–96, 2016.
- [8] ATTOUCH, H., REDONT, P., BOLTE, J., SOUBEYRAN, A., “Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Lojasiewicz inequality”, *Math. Oper. Res.*, v. 35(2), pp. 438–457, 2010.
- [9] ATTOUCH, H., BOLTE, J., SVAITER, B.F., “Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods”, *Math. Program.*, v. 137(1-2), pp. 91–129, 2013.

- [10] AUSLENDER, A., TEBoulLE, M., BEN-TIBA, S., “Interior Proximal and Multiplier Methods Based on Second Order Homogeneous Functionals”, *Math. Oper. Research*, v. 24, pp. 645–668, 1999.
- [11] BAČÁK, M., BORWEIN, J.M., “On difference convexity of locally Lipschitz functions”, *Optimization*, v. 60(8-9), pp. 961–978, 2011.
- [12] BAČÁK, M., “Computing medians and means in Hadamard spaces”, *SIAM J. Optim.*, v. 24(3), pp. 1542–1566, 2014.
- [13] BAČÁK, M., “Convex analysis and optimization in Hadamard spaces”, *Walter de Gruyter GmbH, Berlin*, v. 22, 2014.
- [14] BAKER, C. G., ABSIL, P. A., GALLIVAN, K. A., “An implicit trust-region method on Riemannian manifolds”, *IMA J. Numer. Anal.*, v. 28(4), pp. 665–689, 2008.
- [15] BAO, T., MORDUKHOVICH, B.S., SOUBEYRAN, A., “Variational Analysis in Psychological Modelling”, *J. Optim. Theory Appl.*, v. 164, pp. 290–315, 2015.
- [16] BARANI, A., HOSSEINI, S., POURYAYEVALI, M.R., “On the metric projection onto ϕ -convex subsets of Hadamard manifolds”, *Rev. Mat. Complut.*, v. 26(2), pp. 815–826, 2013.
- [17] BELLO CRUZ, J.Y., “A subgradient method for vector optimization problems”, *SIAM J. Optim.*, v. 23, pp. 2169–2182, 2013.
- [18] BENTO, G. C., CRUZ NETO, J.X., OLIVEIRA, P. R., “Convergence of inexact descent methods for nonconvex optimization on Riemannian manifolds”, *Preprint*, 2011.
- [19] BENTO, G. C., CRUZ NETO, J.X., “A subgradient method for multiobjective optimization on Riemannian manifolds”, *J. Optim. Theory Appl.*, v. 159(1), pp. 125–137, 2013.
- [20] BENTO, G. C., CRUZ NETO, J.X., “Finite termination of the proximal point method for convex functions on Hadamard manifolds”, *Optimization*, v. 63(9), pp. 1281–1288, 2014.
- [21] BENTO, G. C., CRUZ NETO, J.X., SOUBEYRAN, A., “A proximal point-type method for multicriteria optimization”, *Set-Valued Var. Anal.*, v. 22, pp. 557–573, 2014.

- [22] BENTO, G. C., CRUZ NETO, J.X., OLIVEIRA, P.R., “A New Approach to the Proximal Point Method: Convergence on General Riemannian Manifolds”, *J. Optim. Theory Appl.*, v. 168, pp. 743–755, 2016.
- [23] BENTO, G. C., FERREIRA, O. P., OLIVEIRA, P. R., “Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds”, *Nonlinear Anal.*, v. 73, pp. 564–572, 2010.
- [24] BENTO, G. C., FERREIRA, O. P., OLIVEIRA, P. R., “Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds”, *J. Optim. Theory Appl.*, v. 154(1), pp. 88–107, 2012.
- [25] BENTO, G. C., MELO, J. G., “Subgradient Method for Convex Feasibility on Riemannian Manifolds”, *J. Optim. Theory Appl.*, v. 152, pp. 773–785, 2012.
- [26] BENTO, G.C., SOUBEYRAN, A., “Generalized inexact proximal algorithms: Routine’s formation with resistance to change, following worthwhile changes”, *J. Optim. Theory Appl.*, v. 166, pp. 172–187, 2016.
- [27] BENTO, G. C., SOUBEYRAN, A., “A generalized inexact proximal point method for nonsmooth functions that satisfies Kurdyka-Lojasiewicz inequality”, *Set-Valued Var. Anal.*, v. 23, pp. 501–517, 2015.
- [28] BOLTE, J., DANIILIDIS, J.A., LEWIS, A., “The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems”, *SIAM J. Optim.*, v. 17(4), pp. 1205–1223, 2006.
- [29] BOLTE, J., DANIILIDIS, J.A., LEWIS, A., SHIOTA, M., “Clarke subgradients of stratifiable functions”, *SIAM J. Optim.*, v. 18(2), pp. 556–572, 2007.
- [30] BOLTE, J., DANIILIDIS, J.A., LEWIS, A., SHIOTA, M., “Clarke critical values of subanalytic Lipschitz continuous functions”, *Ann. Polon. Math.*, v. 87, pp. 13–25, 2005.
- [31] BOLTE, J., DANIILIDIS, A., LEY, O., MAZET, L., “Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity”, *Trans. Am. Math. Soc.*, v. 362, pp. 3319–3363, 2010.
- [32] BOLTE, J., SABACH, S., TEBoulLE, M., “Proximal alternating linearized minimization for nonconvex and nonsmooth problems”, *Math. Program.*, v. 146, pp. 459–494, 2014.
- [33] BONNEL, H., IUSEM, A. N., SVAITER, B. F., “Proximal methods in vector optimization”, *SIAM J. Optim.*, v. 15, pp. 953–970, 2005.

- [34] BOUMAL, N., MISHRA, B., ABSIL, P. A., SEPULCHRE, R., “Manopt, a matlab toolbox for optimization on manifolds”, *Journal of Machine Learning Research*, v. 15(1), pp. 1455–1459, 2014.
- [35] BURACHIK, R.S., SVAITER, B.F., “A relative error tolerance for a family of generalized proximal point methods”, *Math. Oper. Res.*, v. 26, pp. 816–831, 2001.
- [36] BURACHIK, R.S., KAYA, C.Y., RIZVI, M.M., “A new scalarization technique to approximate Pareto fronts of problems with disconnected feasible sets”, *J. Optim. Theory Appl.*, v. 162, pp. 428–446, 2014.
- [37] BURKE, J.V., FERRIS, M.C., QIAN, M., “On the Clarke subdifferential of the distance function of a closed set”, *J. Math. Anal. Appl.*, v. 166, pp. 199–213, 1992.
- [38] DO CARMO, M. P., “Riemannian Geometry”, *Birkhauser, Boston*, 1992.
- [39] CENG, L. C., YAO, J.C., “Approximate proximal methods in vector optimization”, *Eur. J. Oper. Res.*, v. 183, pp. 1–19, 2007.
- [40] CENSOR, Y., ZENIOS, S. A., “Proximal minimization algorithm with D-functions”, *J. Optim. Theory Appl.*, v. 73, pp. 451–464, 1992.
- [41] CHEN, G., TEBoulLE, M., “Convergence analysis of proximal-like optimization algorithm using Bregman functions”, *SIAM J. Optim.*, v. 3, pp. 538–543, 1993.
- [42] CHOUNG, T.D., MORDUKHOVICH, B.S., YAO, J.C., “Hybrid approximate proximal algorithms for efficient solutions in vector optimization”, *J. Non-linear Convex Anal.*, v. 12, pp. 257–286, 2011.
- [43] CRUZ NETO, J. X., LIMA, L.L., OLIVEIRA, P.R., “Geodesic algorithms in Riemannian geometry”, *Balkan J. Geom. Appl.*, v. 3(2), pp. 89–100, 1998.
- [44] CRUZ NETO, J. X., FERREIRA, O.P., LUCAMBIO PEREZ, L.R., “A proximal regularization of the steepest descent method in Riemannian manifold”, *Balkan J. Geom. Appl.*, v. 4(2), pp. 1–8, 1999.
- [45] CRUZ NETO, J. X., FERREIRA, O.P., LUCAMBIO PÉREZ, L.R., NÉMETH, S. Z., “Convex-and monotone-transformable mathematical programming problems and a proximal-like point method”, *J. Glob. Optim.*, v. 35(1), pp. 53–69, 2006.

- [46] CRUZ NETO, J. X., OLIVEIRA, P. R., SOARES JR, P. A., SOUBEYRAN, A., “Learning how to play nash, potential games and alternating minimization method for structured nonconvex problems on Riemannian manifolds”, *Journal of Convex Analysis*, v. 20(2), pp. 395–438, 2013.
- [47] CRUZ NETO, J. X., OLIVEIRA, P. R., SOARES JR, P. A., SOUBEYRAN, A., “Proximal Point Method on Finslerian Manifolds and the “Effort-Accuracy” Trade-off”, *J. Optim. Theory Appl.*, v. 162(3), pp. 873–891, 2014.
- [48] CRUZ NETO, J. X., SANTOS, P. S. M., SOARES JR, P. A., “An extragradient method for equilibrium problems on Hadamard manifolds”, *Optim. Lett.*, v. 10(6), pp. 1327–1336, 2016.
- [49] DAUBECHIES, I., DEFRISE, M., DE MOL, C., “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”, *Commun. Pur. Appl. Math.*, v. 57, pp. 1413–1457, 2004.
- [50] DEDIEU, J.P., PRIOURET, P., MALAJOVICH, G., “Newton’s method on Riemannian manifolds: covariant α -theory”, *IMA J. Numer. Anal.*, v. 23, pp. 395–419, 2003.
- [51] ECKSTEIN, J., “Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming”, *Math. Oper. Res.*, v. 18, pp. 202–226, 1993.
- [52] ERDOGAN, H., FESSLER, J.A., “Ordered subsets algorithms for transmission tomography”, *Phys. Med. Biol.*, v. 44, pp. 2835–2851, 1999.
- [53] FERREIRA, O.P., OLIVEIRA, P. R., “Subgradient algorithm on Riemannian manifolds”, *J. Optim. Theory Appl.*, v. 97, pp. 93–104, 1998.
- [54] FERREIRA, O.P., OLIVEIRA, P.R., “Proximal Point Algorithms on Riemannian Manifolds”, *Optimization*, v. 51, n.2, pp. 257–270, 2002.
- [55] FERREIRA, O.P., SVAITER, B.F., “Kantorovich’s theorem on Newton’s method in Riemannian manifolds”, *J. Complexity*, v. 18(1), pp. 304–329, 2002.
- [56] FLIEGE, J., SVAITER, B.F., “Steepest descent methods for multicriteria optimization”, *Math. Methods Oper. Res.*, v. 51, pp. 479–494, 2000.
- [57] FLIEGE, J., GRAÑA DRUMMOND, L.M., SVAITER, B.F., “Newton’s method for multiobjective optimization”, *SIAM J. Optim.*, v. 20, pp. 602–626, 2009.

- [58] FUKUDA, E.H., GRAÑA DRUMMOND, L.M., “On the convergence of the projected gradient method for vector optimization”, *Optimization*, v. 60, pp. 1009–1021, 2011.
- [59] FUKUDA, E.H., GRAÑA DRUMMOND, L.M., “A survey on multiobjective decent methods”, *Optimization*, v. 34, pp. 585–620, 2014.
- [60] GABAY, D., “Minimizing a differentiable function over a differential manifold”, *J. Optim. Theory Appl.*, v. 37(2), pp. 177–219, 1982.
- [61] GASSO, G., RAKOTOMAMONJY, A., CANU, S., “Recovering sparse signals with non-convex penalties and DC programming”, *IEEE T. Signal Process.*, v. 57, pp. 4686–4698, 2009.
- [62] GEARHART, W.B., “Compromise solutions and estimation of the non inferior set”, *J. Optim. Theory Appl.*, v. 47, pp. 29–47, 1979.
- [63] GOETZMANN, K.S., BUSING, B., MATUSCHKE, J., “Multicriteria optimization and compromise solutions”, *MDS Colloquium*, 2011.
- [64] GINCHEV, I., GINTCHEVA, D., “Characterization and recognition of d.c. functions”, *J. Glob. Optim.*, v. 57, pp. 633–647, 2013.
- [65] GOLDFARB, D., MA, S., SCHEINBERG, K., “Fast alternating linearization methods for minimizing the sum of two convex functions”, *Math. Program.*, v. 141, pp. 349–382, 2013.
- [66] GRAÑA DRUMMOND, L.M., IUSEM, A.N., “A projected gradient method for vector optimization problems”, *Comput. Optim. Appl.*, v. 28, pp. 5–29, 2004.
- [67] GRAÑA DRUMMOND, L.M., SVAITER, B.F., “A steepest descent method for vector optimization”, *J. Comput. Appl. Math.*, v. 175, pp. 395–414, 2005.
- [68] GREENE, R.E., SHIOHAMA, K., “Convex Functions on Complete Noncompact Manifolds: Topological Structure”, *Invent. Math.*, v. 63, pp. 129–157, 1981.
- [69] GREGÓRIO, R., OLIVEIRA, P.R., “A logarithmic-quadratic proximal point scalarization method for multiobjective programming”, *J. Glob. Optim.*, v. 49, pp. 281–291, 2011.

- [70] GROHS, P., HOSSEINI, S., “ ε -subgradient algorithms for locally Lipschitz functions on Riemannian manifolds”, *Advances in Computational Mathematics*, v. 42(2), pp. 333–360, 2016.
- [71] HARTMAN, P., “On functions representable as a difference of convex functions”, *Pac. J. Math.*, v. 9, pp. 707–713, 1959.
- [72] HIRIART-URRUTY, J.B., “Generalized differentiability, duality and optimization for problems dealing with difference of convex functions, Convexity and Duality in Optimization”, *Lecture Notes in Economics and Mathematical Systems*, v. 256, pp. 37–70, 1986.
- [73] HORST, R., TUY, H., “Global Optimization (Deterministic Approaches)”, *Springer-Verlag, Berlin New York*, second edition, 1993.
- [74] HUANG, X.X., YANG, X.Q., “Duality for multiobjective optimization via nonlinear Lagrangian functions”, *J. Optim. Theory Appl.*, v. 120(1), pp. 111–127, 2004.
- [75] JI, Y., GOH, M., DE SOUZA, R., “Proximal point algorithms for multi-criteria optimization with the difference of convex objective functions”, *J. Optim. Theory Appl.*, v. 169, pp. 280–289, 2016.
- [76] KIWIEL, K.C., “Proximal minimization methods with generalized Bregman functions”, *SIAM J. Control Optim.*, v. 35, pp. 1142–1168, 1997.
- [77] KRISTÁLY, A., “Location of Nash equilibria: a Riemannian geometrical approach”, *Proceedings of the American Mathematical Society*, v. 138(5), pp. 1803–1810, 2010.
- [78] KRISTÁLY, A., “Nash-type equilibria on Riemannian manifolds: a variational approach”, *Journal de Mathématiques Pures et Appliquées*, v. 101(5), pp. 660–688, 2014.
- [79] KURDYKA, K., “On gradients of functions definable in o-minimal structures”, *Ann. Inst. Fourier*, v. 48, pp. 769–783, 1998.
- [80] KURDYKA, K., MOSTOWSKI, T., PARUSINSKI, A., “Proof of the gradient conjecture of R.Thom”, *Ann. Math.*, v. 152, pp. 763–792, 2000.
- [81] LAGEMAN, C., “Convergence of gradient-like dynamical systems and optimization algorithms”, *Ph.D. Thesis*, Universitt Wrzburg, 2007.
- [82] LANG, S., “Fundamentals of differential geometry”, *Graduate Texts in Mathematics*, *Springer-Verlag: New York*, v. 191, 1999.

- [83] LANGE, K., HUNTER, D.R., YANG, I., “Optimization transfer using surrogate objective functions”, *J. Comput. Graph. Stat.*, v. 9, pp. 1–20, 2000.
- [84] LEE, D.D., SEUNG, H.S., “Algorithms for non-negative matrix factorizations”, *Adv. NIPS.*, 2001.
- [85] LEWIN, K., “Frontiers in group dynamics: Concept, method and reality in social science, social equilibria and social change”, *Human Relations*, v. 1, pp. 5–41, 1947.
- [86] LEWIN, K., “Field Theory in Social Science”, *Harper and Row, New York*, 1951.
- [87] LI, C., WANG, J.H., “Newton’s method for sections on Riemannian Manifolds: Generalized Coveriant α -Theory”, *J. Complexity*, v. 24, pp. 423–451, 2008.
- [88] LI, C., LÓPEZ, G., MARTÍN-MÁRQUEZ, V., “Monotone vector fields and the proximal point algorithm on Hadamard manifolds”, *J. Lond. Math. Soc.*, v. 79, pp. 663–683, 2009.
- [89] LI, C., MORDUKHOVICH, B.S., WANG, J., YAO, J.C., “Weak sharp minima on Riemannian manifolds”, *SIAM J. Optim.*, v. 21(4), pp. 1523–1560, 2011.
- [90] LI, S. L., LI, C., LIOU, Y. C., YAO, J. C., “Existence of solutions for variational inequalities on Riemannian manifolds”, *Nonlinear Analysis: Theory, Methods and Applications*, v. 71(11), pp. 5695–5706, 2009.
- [91] LI, C., LÓPEZ, G., MARTÍN-MÁRQUEZ, V., WANG, J. H., “Resolvents of set-valued monotone vector fields in Hadamard manifolds”, *Set-Valued Var. Anal.*, v. 19(3), pp. 361–383, 2011.
- [92] LI, C., YAO, J. C., “Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal point algorithm”, *SIAM J. Control Optim.*, v. 50(4), pp. 2486–2514, 2012.
- [93] LOJASIEWICZ, S., “Une propri t  topologique des sous-ensembles analytiques reles”, *Les quations aux Drives Partielles, ditions du centre National de la Recherche Scientifique*, pp. 87–89, 1963.
- [94] LUC, D.T., “Theory of Vector Optimization”, *Lecture Notes in Econom. and Math. Syst., Springer-Verlag, New York*, 1989.
- [95] LUENBERGER, D. G., “The gradient projection method along geodesics”, *Management Science*, v. 18(11), pp. 620–631, 1972.

- [96] MAIRAL, J., BACH, F., PONCE, J., SAPIR, G., “Online learning for matrix factorization and sparse coding”, *J. Mach. Learn. Res.*, v. 11, pp. 19–60, 2010.
- [97] MAIRAL, J., “Incremental majorization-minimization optimization with application to large-scale machine learning”, *SIAM J. on Optim.*, v. 25(2), pp. 829–855, 2015.
- [98] MARTINET, B., “Regularisation d’inéquations variationnelles par approximations succesives”, *Rev. Française d’Inform. Recherche Oper.*, v. 4, pp. 154–159, 1970.
- [99] MINAMI, M., “Weak Pareto-Optimal Necessary Conditions in a Nondifferentiable Multiobjective Program on a Banach Space”, *J. Optim. Theory Appl.*, v. 41, pp. 451–461, 1983.
- [100] MORDUKHOVICH, B.S., SHAO, Y., “Nonsmooth Sequential Analysis in Asplund Spaces”, *Transactions of the American Mathematical Society*, v. 348, pp. 1235–1280, 1996.
- [101] MORENO, F.G., OLIVEIRA, P. R., SOUBEYRAN, A., “A proximal point algorithm with quasi distance. Application to habit’s formation”, *Optimization*, v. 61, pp. 1383–1403, 2012.
- [102] MOUDAFI, A., MAINGÉ, P-E., “On the convergence of an approximate proximal method for DC functions”, *Journal of Computational Mathematics*, v. 24, pp. 475–480, 2006.
- [103] NÉMETH, S.Z., “Variational inequalities on Hadamard manifolds”, *Nonlinear Anal.*, v. 52, pp. 1491–1498, 2003.
- [104] OPRICOVIC, S., TZENG, G-H., “Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS”, *Eur. J. Oper. Res.*, v. 156, pp. 445–455, 2004.
- [105] OPRICOVIC, S., TZENG, G-H., “Extended VIKOR method in comparison with outranking methods”, *Eur. J. Oper. Res.*, v. 178, pp. 514–529, 2007.
- [106] PAN, S., CHEN, J.S., “Entropy-like proximal algorithms based on a second-order homogeneous distance function for quasi-convex programming”, *J. Glob. Optim.*, v. 39, pp. 555–575, 2007.
- [107] PAPA QUIROZ, E. A., QUISPE, E.M., OLIVEIRA, P. R., “Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds”, *J. Math. Anal. Appl.*, v. 341, pp. 467–477, 2008.

- [108] PAPA QUIROZ, E. A., OLIVEIRA, P. R., “Proximal point method for minimization quasiconvex locally Lipschitz functions on Hadamard manifolds”, *Nonlinear Anal.*, v. 75, pp. 5924–5932, 2012.
- [109] PHAM, D. T., “Algorithms for solving a class of nonconvex optimization problems, methods of subgradients”, in *Fermat Days 85, Mathematics for Optimization*, pp. 249–270, 1986.
- [110] PERELMAN, G., “DC Structure on Alexandrov Space”, *Preprint*, 2003.
- [111] RAZAVIYAYN, M., SANJABI, M., LUO, Z.Q., “A Stochastic Successive Minimization Method for Nonsmooth Nonconvex Optimization with Applications to Transceiver Design in Wireless Communication Networks”, *Math. Program.*, v. 157(2), pp. 515–545, 2016.
- [112] ROCKAFELLAR, R. T., “Monotone operators and the proximal point algorithm”, *SIAM J. Control. Optim.*, v. 14, pp. 877–898, 1976.
- [113] ROCKAFELLAR, R.T., WETS, R.J-B., “Variational Analysis”, *Springer, Berlin*, 1998.
- [114] SAKAI, T., “Riemannian Geometry”, *Translations of Mathematical Monographs, Am. Math. Soc., Providence*, v. 149, 1996.
- [115] SCHOTT, D., “Basic properties of Fejer monotone sequences”, *Rostocker Mathematische Kolloquium*, v. 49, pp. 57–74, 1995.
- [116] SMITH, S. T., “Optimization techniques on Riemannian manifolds”, *Fields institute communications*, v. 3(3), pp. 113–135, 1994.
- [117] SOUBEYRAN, A., “Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors”, *GREQAM, Aix Marseille University*, 2009.
- [118] SOUBEYRAN, A., “Variational rationality and the ”unsatisfied man”: routines and the course pursuit between aspirations, capabilities and beliefs”, *GREQAM, Aix Marseille University*, 2010.
- [119] SOUBEYRAN, A., “Variational rationality, Part 1. A theory of worthwhile stay and change approach-avoidance transitions ending in traps”, *GREQAM, Aix Marseille University*, 2016.
- [120] SOUZA, J. C., OLIVEIRA, P. R., “A proximal point algorithm for DC functions on Hadamard manifolds”, *J. Glob. Optim.*, v. 63, pp. 797–810, 2015.

- [121] SOUZA, J. C., OLIVEIRA, P. R., SOUBEYRAN, A., “Global convergence of a proximal linearized algorithm for difference of convex functions”, *Optim. Lett.*, v. 10(7), pp. 1529–1539, 2016.
- [122] SOUZA, J. C., OLIVEIRA, P. R., SOUBEYRAN, A., “A generalized proximal linearized algorithm for DC functions with application to the optimal size of the firm problem”, *Submetido*, 2016.
- [123] SOUZA, J. C., BENTO, G.C., CRUZ NETO, J.X., SOUBEYRAN, A., “Vectorial improving in DC programming”, *Submetido*, 2016.
- [124] SOUZA, J. C., BENTO, G.C., CRUZ NETO, J.X., LÓPEZ, G., SOUBEYRAN, A., “The proximal point method for locally Lipschitz functions in multiobjective optimization”, *Submetido*, 2016.
- [125] SOUZA, J. C., BENTO, G.C., CRUZ NETO, J.X., OLIVEIRA, P. R., “Convergence of the steepest descent method for computing Riemannian center of mass on Hadamard manifolds”, *Preprint*, 2016.
- [126] SUN, W., SAMPAIO, R. J. B., CANDIDO, M. A. B., “A proximal point algorithm for minimization of DC function”, *J. of Comput. Math.*, v. 21, pp. 451–462, 2003.
- [127] TEBOULLE, M., “Entropic proximal mappings with applications to nonlinear programming”, *Math. Oper. Res.*, v. 17, pp. 670–690, 1992.
- [128] TOLAND, J. F., “On subdifferential calculus and duality in nonconvex optimization”, *Bull. Soc. Math. France*, v. 60, pp. 177–183, 1979.
- [129] UDRISTE, C., “Convex Functions and Optimization Algorithms on Riemannian Manifolds”, *Mathematics and Its Applications, Kluwer Academic, Dordrecht*, v. 297, 1994.
- [130] VAN DEN DRIES, L., MILLER, C., “Geometric categories and o-minimal structures”, *Duke Math. J.*, v. 84, pp. 497–540, 1996.
- [131] VIAL, J-P., “Strong and Weak Convexity of Sets and Functions”, *Math. Oper. Res.*, v. 8, pp. 231–259, 1983.
- [132] VILLACORTA, K.D.V., OLIVEIRA, P. R., “An interior proximal method in vector optimization”, *Eur. J. Oper. Res.*, v. 214, pp. 485–492, 2011.
- [133] WANG, J., LÓPEZ, G., MARTÍN-MÁRQUEZ, V., LI, C., “Monotone and accretive vector fields on Riemannian manifolds”, *J. Optim. Theory Appl.*, v. 146(3), pp. 691–708, 2010.

- [134] WANG, J., LI, C., LÓPEZ, G., YAO, J. C., “Convergence analysis of inexact proximal point algorithms on Hadamard manifolds”, *J. Glob. Optim.*, v. 61(3), pp. 553–573, 2015.
- [135] WANG, X., LI, C., YAO, J. C., “Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures”, *J. Optim. Theory Appl.*, v. 164(1), pp. 202–217, 2015.
- [136] WANG, X., LI, C., WANG, J., YAO, J. C., “Linear Convergence of Subgradient Algorithm for Convex Feasibility on Riemannian Manifolds”, *SIAM J. Optim.*, v. 25(4), pp. 2334–2358, 2015.