Universidade Federal do Piauí
Centro de Ciências da Natureza
Pós-graduação em Matemática
Mestrado em Matemática

Um Estudo do Método Proximal Gradiente com Buscas Lineares

Maiara Veras de Brito

Teresina - 2026

Maiara Veras de Brito

Dissertação de Mestrado:

Um Estudo do Método Proximal Gradiente com Buscas Lineares

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Matemática, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientador:
Prof. Dr. Ray Victor Guimarães Serra.
Co-Orientador:
Prof. Dr. Sandoel de Brito Vieira.

Teresina - 2026

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Um Estudo do Método Proximal Grandiente com Buscas Lineares

Maiara Veras de Brito

Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**, outorgado pela Universidade Federal do Piauí.

A citação de qualquer trecho deste trabalho é permitida, desde que seja feita em conformidade com as normas da ética científica.

Dissertação aprovada em 11 de fevereiro de 2026.

Banca Examinadora:

Documento assinado digitalmente
gov.br RAY VICTOR GUIMARAES SERRA
Data: 09/03/2026 13:41:17-0300
Verifique em https://validar.iti.gov.br

____

Prof. Dr. Ray Victor Guimarães Serra – Orientador

Documento assinado digitalmente
gov.br SANDOEL DE BRITO VIEIRA
Data: 09/03/2026 13:09:07-0300
Verifique em https://validar.iti.gov.br

____

Prof. Dr. Sandoel de Brito Vieira - Coorientador

Documento assinado digitalmente
gov.br PAULO SERGIO MARQUES DOS SANTOS
Data: 09/03/2026 13:30:35-0300
Verifique em https://validar.iti.gov.br

____

Prof. Dr. Paulo Sergio Marques dos Santos – UFDPar

Documento assinado digitalmente
gov.br SISSY DA SILVA SOUZA
Data: 09/03/2026 13:21:50-0300
Verifique em https://validar.iti.gov.br

____

Prof. Drª. Sissy da Silva Souza - UFDPar

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Divisão de Representação da Informação

B862e Brito, Maiara Veras de.
Um Estudo do Método Proximal Gradiente com Buscas Lineares / Maiara Veras de Brito. -- 2026.
122 f.

Dissertação (Mestrado) – Universidade Federal do Piauí, Programa de Pós-Graduação em Matemática, Teresina, 2026.
"Orientador: Prof. Dr. Ray Victor Guimarães Serra".
"Coorientador: Prof. Dr. Sandoel de Brito Vieira".

1. Otimização. 2. Método Gradiente. 3. Método Ponto Proximal. 4. Método Proximal Gradiente. 5. Busca Linear. I. Serra, Ray Victor Guimarães. II. Vieira, Sandoel de Brito. III. Título.

CDD 519.6

Bibliotecária: Francisca das Chagas Dias Leite – CRB3/1004

Dedico este trabalho aos meus pais.

Agradecimentos

Agradeço primeiramente a Deus, pois dEle, por Ele e para Ele são todas as coisas. Sou imensamente grata por me mostrar que com fé e perseverança conseguimos conquistar todos os nossos sonhos, por me ajudar nos momentos em que achei que não conseguiria e principalmente por me fazer entender que as coisas acontecem no Seu tempo. Obrigada por todas as oportunidades Jesus.

A minha mãe Francisca das Chagas por ter sido meu alicerce em todos esses anos. Obrigada por todo o apoio e por acreditar sempre em mim, mesmo quando eu acho que não consigo. Seu abraço e a vontade de te dar uma vida melhor é o meu gás para continuar todos os dias, ainda quero dar muito orgulho a senhora. Ao meu pai, José Coelho (*in memoriam*) por sempre ter se preocupado com a minha educação. Sinto saudades do seu abraço pai. Sua pequena chegou ao mestrado. O cuidado de vocês e a educação que me deram me ensinaram muito a ser a pessoa que sou hoje e sou muito grata por tudo que fizeram por mim, amo vocês. A minha irmã Maisa, por todo o incentivo e companheirismo.

Agradeço ao professor Amaral e professor Raimundinho, que foram meus professores de matemática durante a escola, obrigada por todas as aulas e preparatórios de olimpíadas. A OBMEP me "abriu" portas, e sempre vou ser grata a vocês por terem me ajudado a realizar esse sonho. Espero ajudar muitas pessoas através da OBMEP assim como vocês fizeram comigo. Obrigada pela amizade, por terem acreditado em mim e por me incentivarem a seguir em frente.

Agradeço aos meus professores da UFDPar, em especial a professora Sissy, professor Paulo e ao professor Pedro Jorge. Obrigada por todo o isentivo a seguir a vida acadêmica e por cada ensinamento partilhado em suas aulas. Vocês nos inspiram.

A todo o corpo docente do Departamento de Matemática da UFPI, especialmente àqueles que participaram da minha formação durante esses dois anos, Ray, José Francisco, Barnabé, Sandoel, Vitaliano, Gilson, Joel, e Gleyson. Professor José Francisco, obrigada por cada ensinando matemático proporcionado na disciplina de $\mathbb{R}^n$ e pelas conversas de incentivo. Não poderia deixar de agradecer também, à secretária da Pós-Graduação Larisse e a tia Cruz obrigada por todas as conversas e pela amizade.

Agradeço ao meu Orientador Professor Ray Serra por todo apoio ao longo desses dois anos, por os ensinamentos matemáticos, as conversas de incentivo e todos os conselhos. Admiro muito o senhor pelo empenho com seu trabalho e por ter sido um ótimo orientador. Muito obrigada pela paciência e por todo o tempo dedicado a este trabalho.

Agradeço também aos componentes da banca professora Sissy da Silva Souza. Obrigada por todo apoio ao longo da graduação e por me incentivar a seguir a vida acadêmica, como sempre falo, a senhora é a minha referência de mulher na matemática. Ao professor Paulo Sérgio Marques dos Santos, obrigada por sempre me incentivar a continuar nos estudos, agradeço por todos os conselhos. E ao meu Co-Orientador professor Sandoel de Brito Vieira, obrigada pelos conhecimentos partilhados na orientação e na disciplina de Estruturas Algébricas, principalmente os Teoremas de Sylow, no qual gostei bastante. Obrigada pelas conversas de incentivo, ter professores que nos ajudam a acreditar que podemos conseguir é muito importante nessa caminhada tão difícil.

A minha prima Naiane, obrigada por ter sido minha família em Teresina. As minhas amigas, Jayne, Erika, Duda, Joice e Alane, por a amizade que dura desde o 7 ano do fundamental. Ter o apoio de vocês foi muito importante nessa jornada.

Aos amigos que fiz durante o mestrado e que compartilharam essa jornada comigo. Pedro Henrique, Jerfersson, Ricael, Natália, Andreina, Angélica, Ismael, Vinícius, Júnior, Elliel, Gean, Gustavo, Ana Júlia, Manu, Ana Luísa, João Vinícius, Erisvaldo e Igor. Em especial a Kátia, Lucas, Vitor, Kawanardo, Stefane e Ysadora. Obrigada por sempre ir ao quadro meu amigo Vitor, a Kátia por sempre nos incentivar à continuar e pelos bolos maravilhosos, o Lucas por nos tirar boas risadas, Stefane por todas as palavras positivas e a Ysadora por sempre estar presente com palavras doces. Dividir esses dois anos com vocês fez tudo se tornar mais leve. Muito obrigada pela amizade de vocês.

Por fim, gostaria de agradecer a OBMEP-PICME, pelo apoio financeiro.

"Deus não colocaria em seu coração o desejo de um sonho impossível ou um propósito inalcançável. Ele já sabe onde você vai chegar, Ele só precisa te preparar antes".

- Santa Teresinha.

"É justo que muito custe o que muito vale".

Santa Teresa D'Ávila.

Resumo

Neste trabalho, estudamos o Método Proximal Gradiente, conhecido também na literatura como forward-backward splitting (FBS). Este método visa resolver problemas de otimização convexa não-suaves em que a função objetivo é decomposta na soma de uma função convexa diferenciável e outra convexa não-suave. Em geral, na literatura, a hipótese de Lipschitz continuidade do gradiente da função diferenciável é comumente empregada para garantir a convergência do método para uma solução ótima do problema. No entanto, o foco principal deste trabalho é investigar variações do método que empregam estratégias de pesquisas lineares, garantindo a convergência das sequências geradas sem a suposição de Lipschitz continuidade global. A análise engloba a investigação de três procedimentos de Busca Linear que permitem relaxar significativamente tal suposição; um desses procedimentos é empregado em dois métodos distintos, incluindo uma versão acelerada. Ao longo do trabalho, as propriedades, vantagens e desvantagens desses procedimentos são analisadas, juntamente com a análise de convergência das sequências geradas pelo Método Proximal Gradiente. Adicionalmente, um estudo detalhado sobre a complexidade dos métodos é apresentado.

Palavras-chave: Otimização, Método Gradiente, Método Ponto Proximal, Método Proximal Gradiente, Busca Linear.

Abstract

In this work, we investigate the Proximal Gradient Method, well-known in the literature as forward-backward splitting (FBS). This method aims to solve nonsmooth convex optimization problems where the objective function is decomposed into the sum of a differentiable convex function and another nonsmooth convex function. Generally, in the literature, the hypothesis of Lipschitz continuity of the gradient of the differentiable function is commonly employed to guarantee the convergence of the method to an optimal solution of the problem. However, the main focus of this work is to investigate variations of the method that employ linesearch strategies, ensuring the convergence of the generated sequences without the assumption of global Lipschitz continuity. The analysis encompasses the investigation of three linesearch procedures that significantly relax such an assumption; one of these procedures is employed in two distinct methods, including an accelerated version. Throughout the work, the properties, advantages, and disadvantages of these procedures are analyzed, alongside the convergence analysis of the sequences generated by the Proximal Gradient Method. Additionally, a detailed study on the complexity of the methods is presented.

Keywords: Optimization, Gradient Method, Proximal Point Method, Proximal Gradient Method, Line Search.

Sumário

Introdução

Os problemas de otimização desempenham um papel fundamental em diversas áreas da ciência e engenharia, desde aprendizado de máquina e processamento de sinais até finanças e física. Frequentemente, a formulação desses problemas envolve a minimização de funções objetivo que apresentam estruturas complexas, combinando termos suaves (diferenciáveis) com termos que podem ser não-diferenciáveis ou de difícil manipulação computacional.

Neste trabalho, focamos em uma classe particular de problemas de otimização convexa, onde a função objetivo é expressa como a soma de duas funções, uma das quais é diferenciável e a outra pode não ser diferenciável. Mais especificamente, consideramos o problema na forma

$$\min_{x \in \mathbb{R}^n} F(x) = g(x) + h(x), \tag{1}$$

onde $g : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ é uma função convexa e diferenciável e $h : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ é uma função convexa, possivelmente não-diferenciável, mas cujo operador proximal é eficientemente computável.

Essa estrutura aditiva é ubíqua em diversas aplicações. Por exemplo, em problemas de aprendizado de máquina com regularização, $g(x)$ pode representar o termo de erro (e.g., erro quadrático) e $h(x)$ o termo de regularização (e.g., norma ℓ_1 para promover esparsidade ou a função indicadora de um conjunto convexo para impor restrições). A presença do termo não-diferenciável $h(x)$ pode ser usado em métodos que envolvem o operador proximal, por exemplo no Método do Ponto Proximal, porém este fato também pode impedir à aplicação direta à métodos clássicos baseados no gradiente, como o Método do Gradiente, que exige a diferenciabilidade da função objetivo completa.

Diante deste cenário, métodos que conseguem desacoplar as propriedades das funções g e h são essenciais. O Método Proximal Gradiente surge como uma ferramenta poderosa e eficiente para resolver problemas da forma (1), combinando passos do gradiente para a componente suave g com passos do operador proximal para a componente h.

Esta dissertação tem como inspiração estudar os resultados propostos nos trabalhos feitos por Bello Cruz, J.Y., Nghia, T.T.A em [9] e Bello Cruz, J.Y.; LI, G; Nghia, T.T.A em [10]. No primeiro capítulo, a revisão de conceitos clássicos relacionados aos métodos de descida e método do ponto proximal são baseados em [24, 29]. Enquanto resultados auxiliares podem ser consultados também em [26, 27], onde abordaremos conceitos preliminares que serão de fundamental importância para compreensão deste trabalho, bem como uma explicação dos métodos de otimização que utilizaremos ao decorrer do nosso estudo.

O segundo capítulo será dedicado aos procedimentos de Buscas Lineares vistos principalmente em [9, 10]. No método proximal gradiente, escolher o comprimento de passo é crucial para garantir a convergência e eficiência computacional. Dessa forma, existem duas formas essenciais para determinar esses parâmetros a Busca Linear implícita e a Busca Linear explícita, que diferem principalmente na forma como o operador proximal é utilizado durante o processo iterativo. Na Busca Linear implícita, o tamanho do passo é determinado por meio da resolução repetida de subproblemas proximais até que uma condição de aceitabilidade seja satisfeita, o que, embora robusto, pode acarretar um custo computacional elevado. Por outro lado, a Busca Linear explícita avalia a condição de aceitação sem a necessidade de múltiplas computações do operador proximal, o que torna o método computacionalmente mais eficiente.

O terceiro capítulo introduz quatro métodos do tipo proximal gradiente. Os dois primeiros utilizam a Busca Linear 1, sendo o segundo uma variante acelerada, cujo principal diferencial reside na obtenção de melhores cotas de complexidade iterativa, resultando em convergência mais rápida. O terceiro método tem como novidade a utilização da Busca Linear explicita, sendo o único neste trabalho que utiliza tal busca. Já o quarto método utiliza uma Busca Linear implícita sendo, com os mesmos resultado dos anteriores exceto a prova de complexidade.

O quarto capítulo apresenta o estudo sobre a taxa de complexidade, este fato é muito importante no estudo de otimização, pois descreve a velocidade com que um algoritmo se aproxima de uma solução ótima. Trazemos o resultado de complexidade dos três primeiros métodos abordados. Por último, temos o apêndice, nele está uma coleção de resultados importantes de Análise convexa e Análise no $\mathbb{R}^n$ que podem ser encontradas em [3, 26, 27] ao qual recorremos ao longo deste trabalho.

Capítulo 1

Noções e Conceitos Preliminares

Neste capítulo, apresentaremos os conceitos, definições e resultados principais que serão necessários para o nosso trabalho. Ressaltamos que a revisão realizada neste capítulo, bem como a utilização de resultados clássicos da literatura pertinente ao longo do trabalho, serão complementadas pelo Apêndice A. A revisão de conceitos clássicos relacionados aos Métodos de Descida e Método Ponto Proximal são baseados em [24, 29]. Enquanto resultados auxiliares podem ser consultados também em [3, 26, 27].

1.1 Conceitos Preliminares

Inicialmente, apresentamos algumas definições e resultados necessários para o nosso trabalho. Começaremos com a definição de função semicontínua inferiormente. Tal conceito será essencial, por exemplo, quando falarmos do Método Ponto Proximal na Seção 1.2.2.

Definição 1.1. *(Função Semicontínua inferiormente) Dizemos que a função* $f : D \to \mathbb{R}$ *é semicontínua inferiormente (sci) no ponto* $x \in D \subset \mathbb{R}^n$*, quando para qualquer sequência* $x^k \to x$*, com* $k \to \infty$ *tem-se,*

$$\liminf_{k \to \infty} f(x^k) \geqslant f(x).$$

Em outras palavras, a função f é sci em x se para todo $\xi \in (-\infty, f(x)]$, podemos encontrar uma vizinhança V de x tal que $f(V) \subset (\xi, +\infty]$.

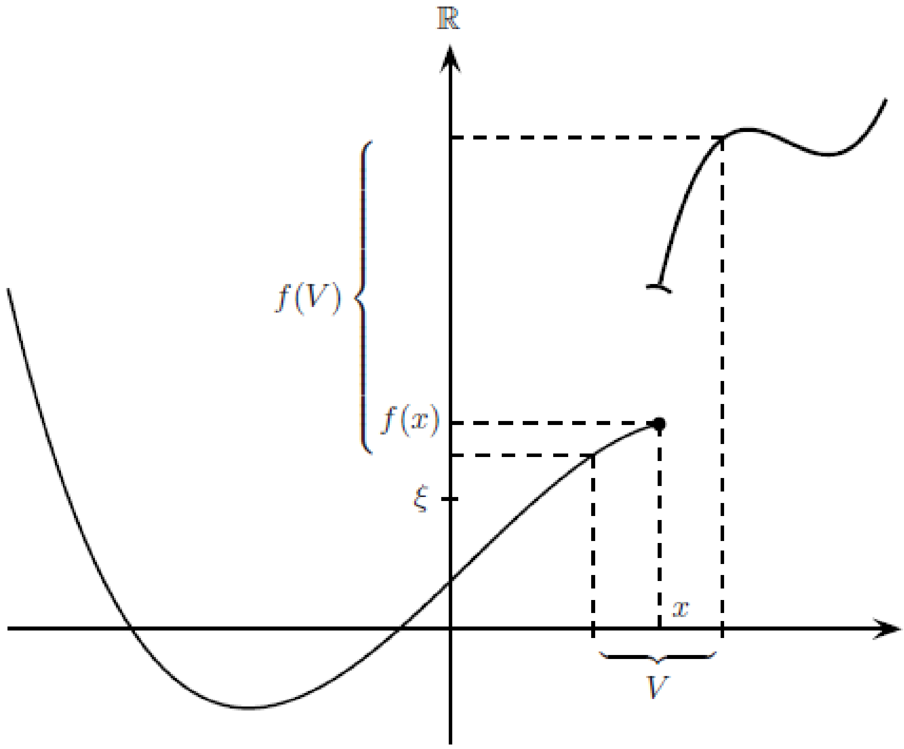

Figura 1.1: Função sci. [3]

Agora, relembramos algumas definições e propriedades comuns em Otimização. Uma função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} := \overline{\mathbb{R}}$ é dita ser própria, se o seu domínio efetivo, denotado por

$$\mathsf{dom}(f) := \{x \in \mathbb{R}^n \mid f(x) < +\infty\},$$

é não-vazio. A seguir, recordamos a definição de derivada direcional, a fim de relembrar certas propriedades e conceitos que serão úteis.

Definição 1.2. *(Derivada direcional) Para qualquer* $x \in$ *dom(f), a derivada direcional de* f *em* x *na direção* d *é,*

$$f'(x; d) := \lim_{t \to 0^+} \frac{f(x + td) - f(x)}{t},$$

sempre que tal limite existe (embora possa ser infinito).

Proposição 1.1 ([5], Proposição 17.2). *Seja* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *uma função própria semicontínua inferior e convexa (ver Definição A.4). Então, para* $x \in$ *dom(f) e* $y \in \mathbb{R}^n$ *as seguintes igualdades são verdadeiras:*

(i) $\phi : \mathbb{R}_{++} \to (-\infty, +\infty]$, *dada por* $\phi(t) \mapsto \frac{f(x+ty)-f(x)}{t}$ *é não-decrescente.*

(ii) $f'(x;y)$ *existe e* $f'(x;y) = \inf_{t \in \mathbb{R}_{++}} \left(\dfrac{f(x+ty) - f(x)}{t} \right)$.

(iii) $f'(x; y - x) + f(x) \leqslant f(y)$.

Demonstração. Para provar o item (i), tome $\alpha < \beta$. Daí, considere $t = \frac{\alpha}{\beta} \in (0,1)$ e $z = x + \beta y$. Se $f(z) = +\infty$, então $\phi(\beta) = +\infty$, o que por sua vez implicaria que $\phi(\alpha) \leqslant \phi(\beta)$. Agora suponha que $f(z) < +\infty$, então segue da Definição A.4 que

$$
\begin{aligned}
f(x + \alpha y) = f\left(x - \frac{\alpha x}{\beta} + \frac{\alpha x}{\beta} + \alpha y \right) &= f\left((1 - \frac{\alpha}{\beta})x + \frac{\alpha}{\beta}(x + \beta y) \right) \\
&\leqslant \frac{\alpha}{\beta} f(x + \beta y) + (1 - \frac{\alpha}{\beta}) f(x) \\
&= \frac{\alpha}{\beta} f(x + \beta y) + f(x) - \frac{\alpha}{\beta} f(x) \\
&= f(x) + \frac{\alpha}{\beta} \left(f(x + \beta y) - f(x) \right).
\end{aligned}
$$

Portanto, reorganizando os termos, obtemos que

$$
f(x + \alpha y) - f(x) \leqslant \frac{\alpha}{\beta} \left(f(x + \beta y) - f(x) \right),
$$

o que por sua vez, implica que

$$
\frac{f(x + \alpha y) - f(x)}{\alpha} \leqslant \frac{f(x + \beta y) - f(x)}{\beta}.
$$

Logo, $\phi(\alpha) \leqslant \phi(\beta)$, concluindo assim que $\phi(t)$ é não-decrescente, para $t > 0$.

O item (ii) segue diretamente de (i), visto que o ínfimo existe e coincide com a derivada direcional

$$
f'(x;y) = \inf_{t \in \mathbb{R}_+} \left(\frac{f(x+ty) - f(x)}{t} \right) = \inf_{t > 0} \phi(t).
$$

Agora, iremos provar o item (iii). Seja $x \in \operatorname{dom}(f)$ e qualquer $y \in \mathbb{R}^n$. Se $y \notin \operatorname{dom}(f)$, então $f(y) = +\infty$. Logo, $f'(x; y - x) + f(x) \leqslant f(y) = +\infty$, ocorre trivialmente. Se $y \in \operatorname{dom}(f)$, como f é convexa, para qualquer $\alpha \in [0,1)$ vale,

$$
\begin{aligned}
f((1-\alpha)x + \alpha y) = f(x + \alpha(y - x)) &\leqslant (1-\alpha) f(x) + \alpha f(y) \\
&= f(x) - \alpha f(x) + \alpha f(y) \\
&= \alpha(f(y) - f(x)) + f(x),
\end{aligned}
$$

o que por sua vez, implica que

$$
\frac{f(x + \alpha(y - x)) - f(x)}{\alpha} \leqslant f(y) - f(x).
$$

Agora, tomando $\alpha \to 0$ o lado esquerdo converge para a derivada direcional, sendo assim temos,

$$f'(x; y - x) + f(x) \leqslant f(y),$$

como queríamos demonstrar. □

A seguir relembramos a definição de globalmente Lipschitz contínua do vetor gradiente $\nabla f(\cdot)$, a qual será essencial no estudo de complexidade pior-caso dos métodos abordados neste trabalho.

Definição 1.3. *(Lipschitz contínuidade do gradiente) Dizemos que ∇f é globalmente Lipschitz contínuo, se existe $L > 0$ tal que,*

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

Em particular, dizemos que ∇f é localmente Lipschitz contínuo em x^*, se existe $\varepsilon > 0$ tal que,

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in B_\varepsilon(x^*),$$

onde $B_\varepsilon(x^*)$ é a bola fechada em $\mathbb{R}^n$ com centro x^* e raio ε.

Definição 1.4. *(Quasi-Fejér) Seja $S \subset \mathbb{R}^n$ um conjunto não-vazio. Uma sequência $(x^k) \subset \mathbb{R}^n$ é dita ser Quasi-Fejér convergente para o conjunto S se para todo $x \in S$, existe uma sequência $(\varepsilon_k) \subset \mathbb{R}_+$ tal que, para todo $k \in \mathbb{N}$ temos que*

$$\|x^{k+1} - x\| \leqslant \|x^k - x\| + \varepsilon_k^2; \quad \sum_{k=0}^{\infty} \varepsilon_k^2 < +\infty.$$

A sequência é dita ser Fejér convergente em relação ao conjunto S, se na relação acima $\varepsilon_k \equiv 0$ para todo $k \in \mathbb{N}$.

Lema 1.1 ([25], Teorema 4.1)**.** *Se (x^k) é Quasi-Fejér convergente para S, então as seguintes afirmações são válidas:*

(i) A sequência (x^k) é limitada.

(ii) Se um ponto de acumulação de (x^k) pertence a S, então (x^k) converge para um ponto em S.

Demonstração. Provaremos inicialmente o item (i). Por hipótese, temos que (x^k) é Quasi-Fejér convergente, sendo assim

$$\begin{aligned} \|x^{k+1} - x\| \leqslant \|x^k - x\| + \varepsilon_k^2 \leqslant \cdots \leqslant \|x^0 - x\| + \sum_{j=0}^{k} \varepsilon_j^2 \\ \leqslant \|x^0 - x\| + \sum_{j=0}^{\infty} \varepsilon_j^2 < \infty. \end{aligned}$$

Daí, temos que existe $P \in \mathbb{R}$ tal que

$$\|x^k - x\| \leqslant \|x^0 - x\| + \sum_{k=0}^{\infty} \varepsilon_k^2 := P,$$

para qualquer $x \in S$. Sendo assim, temos que a sequência (x^k) está contida na bola fechada de centro x e raio P, logo (x^k) é limitada. Em particular, temos que (x^k) possui ponto de acumulação.

Agora, para provar o item (ii), suponha que existe ponto de acumulação x^* de (x^k), o qual pertence ao conjunto S. Ou seja, existe subsequência $(x^{k_j}) \subset (x^k)$ convergindo para $x^* \in S$. Então,

$$0 \leqslant \|x^{k+1} - x^*\| \leqslant \|x^k - x^*\| + \varepsilon_k^2, \quad \forall k \in \mathbb{N}. \tag{1.1}$$

Dessa forma,

$$\|x^{k+1} - x^*\| \leqslant \|x^k - x^*\| + \varepsilon_k^2 \leqslant \cdots \leqslant \|x^0 - x^*\|^2 + \sum_{k=0}^{j+1} \varepsilon_k^2.$$

Em particular, como a série $\sum_{k=0}^{\infty} \varepsilon_k^2$ é convergente, temos $\varepsilon_k^2 \to 0$ quando $k \to \infty$. Logo, dado $\delta > 0$ arbitrário pequeno, podemos escolher k_1 suficientemente grande tal que

$$\sum_{j=k_1}^{k+1} \varepsilon_j^2 < \frac{\delta}{2}.$$

Uma vez que $x^{k_j} \to x^*$, temos que existe k_2 tal que

$$\|x^{k_j} - x^*\|^2 < \frac{\delta}{2}.$$

Daí, tomando $k_j \geqslant max\{k_1, k_2\}$, obtemos que, para todo $k \geqslant k_j$,

$$\|x^k - x^*\|^2 \leqslant \|x^{k_j} - x^*\|^2 + \sum_{j=k_j}^{k-1} \varepsilon_j^2 \leqslant \|x^{k_j} - x^*\|^2 + \sum_{j=k_j}^{\infty} \varepsilon_j^2 < \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Logo, $\|x^k - x^*\|^2 < \delta$, ou seja, (x^k) converge a x^* que é uma solução do problema. □

1.2 Métodos de Otimização

Nesta seção, abordaremos uma breve descrição dos principais métodos empregados neste trabalho. Mais especificamente, focaremos nos métodos de descida, com ênfase no Método de Cauchy (ou Método do Gradiente), Método do Ponto Proximal e no clássico Método Proximal Gradiente.

1.2.1 Métodos de Descida (Gradiente)

Inicialmente, considere o seguinte problema de minimizar uma função $f : \mathbb{R}^n \to \mathbb{R}$, o qual iremos denotar da seguinte forma,

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1.2}$$

Nesta subseção, assumiremos que o conjunto solução do Problema 1.2, denotado por S^*, é não-vazio.

Agora, estamos interessados em explorar os chamados métodos de descidas para resolver o Problema 1.2, em especial o método do gradiente. Este método consiste em utilizar a direção de descida $d^k = -\nabla f(x^k)$ e gerar uma sequência (x^k) tal que,

$$x^0 \in \mathbb{R}^n, \qquad x^{k+1} = x^k - \alpha_k \nabla f(x^k), \quad k = 0, 1, 2, \ldots$$

onde $\alpha_k > 0$ denota o tamanho do passo. Tal comprimento de passo é tomado de forma que $f(x^{k+1}) \leqslant f(x^k)$. Em seguida, ilustramos a ideia de um algorítimo de descida para uma direção de descida d^k arbitrária.

Método de Descida

Passo 0: Defina $x^0 \in \mathbb{R}^n$, $k := 0$;

Passo 1: Calcule d^k tal que $\langle \nabla f(x^k), d^k \rangle < 0$;

Passo 2: Verifique se $\|\nabla f(x^k)\| < \varepsilon$. Caso contrário, escolha o comprimento de passo $t_k > 0$ tal que $f(x^k + t_k d^k) < f(x^k)$;

Passo 3: Faça $x^{k+1} = x^k + t_k d^k$, $k := k + 1$ e retorne ao Passo 1.

A estrutura de um método de descida, conforme discutida, baseia-se em dois conceitos centrais: a direção de descida d^k e o comprimento de passo t_k. Na sequência, apresentaremos uma das abordagens para a determinação de t_k, realizada através de um dos mais conhecidos procedimentos de Busca Linear: a **Regra de Armijo**. Em resumo, dado $\overline{x} \in \mathbb{R}^n$, uma direção de descida $d \in \mathbb{R}^n$ e $\eta \in (0,1)$, a regra de Armijo consiste em encontra $\overline{t} > 0$ tal que,

$$f(\overline{x} + \overline{t}d) \leqslant f(\overline{x}) + \eta\overline{t}\langle \nabla f(\overline{x}), d\rangle. \tag{1.3}$$

A Regra de Armijo é um procedimento de Busca Linear frequentemente empregado em algoritmos de otimização para determinar um comprimento de passo adequado. Em essência, o procedimento consiste em verificar a chamada "condição de Armijo" dada pela equação (1.3) (ou outra condição de descida) para um valor inicial de comprimento de passo $t > 0$. Caso esta condição não seja satisfeita, adota-se um processo iterativo conhecido como *backtracking*. Nesse processo, o comprimento de passo atual é multiplicado por um fator de redução $\eta \in (0,1)$, com o objetivo de gerar um novo e menor comprimento de passo. A condição é então reavaliada para este novo valor de t, e o processo de redução é repetido até que um comprimento de passo que a satisfaça seja encontrado.

A seguir, trazemos o resultado que apresenta a boa definição da Regra de Armijo.

Teorema 1.1. *Considere uma função diferenciável* $f : \mathbb{R}^n \to \mathbb{R}$*, um ponto* $\overline{x} \in \mathbb{R}^n$*, uma direção de decida* $d \in \mathbb{R}^n$ *e* $\eta \in (0,1)$*. Então, existe* $\delta > 0$ *tal que*

$$f(\overline{x} + td) \leqslant f(\overline{x}) + \eta t\langle \nabla f(\overline{x}), d\rangle,$$

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

Demonstração. Inicialmente, como $d \in \mathbb{R}^n$ é uma direção de descida, segue do Lema A.2 que $\langle \nabla f(\overline{x}), d\rangle \leqslant 0$. Caso, $\langle \nabla f(\overline{x}), d\rangle = 0$ então segue pela definição de direção descida (Definição A.7) que existe $\delta > 0$ tal que para $t \in [0,\delta)$ vale $f(\overline{x} + td) \leqslant f(\overline{x})$. Agora, suponha que $\langle \nabla f(\overline{x}), d\rangle < 0$. Segue da definição de derivada direcional, do fato que f é diferenciável (veja Teorema A.3) e $0 < \eta < 1$, que

$$\lim_{t \to 0} \frac{f(\overline{x} + td) - f(\overline{x})}{t} = f'(\overline{x}, d) = \langle \nabla f(\overline{x}), d\rangle < \eta\langle \nabla f(\overline{x}), d\rangle.$$

Dessa forma, segue da conservação do limite, que existe $\delta > 0$, tal que

$$\frac{f(\overline{x} + td) - f(\overline{x})}{t} < \eta\langle \nabla f(\overline{x}), d\rangle,$$

para todo $t \in (0,\delta)$. Ou seja,

$$f(\overline{x} + td) \leqslant f(\overline{x}) + \eta t\langle \nabla f(\overline{x}), d\rangle,$$

para todo $t \in [0,\delta)$, como queríamos demonstrar. $\square$

Para encontrar o comprimento de passo t_k, utilizamos a Busca linear de Armijo, descrita do seguinte modo:

Procedimento de Busca Linear de Armijo

Dados $\overline{x} \in \mathbb{R}^n, d \in \mathbb{R}^n$ com d sendo direção de decida, $\eta \in (0,1)$. Tome $t = 1$. Se

$$f(\overline{x} + td) > f(\overline{x}) + \eta t \langle \nabla f(\overline{x}), d \rangle.$$

Tome, $t = \eta t$.

A seguir, traremos uma forma geral de gerar uma direção de descida para ser utilizada no método de descida descrito acima. Seja $H : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ uma função continua que associa a cada $x \in \mathbb{R}^n$ uma matriz definida positiva $H(x) \in \mathbb{R}^{n \times n}$. Assim, se $\nabla f(\overline{x}) \neq 0$ temos que $d = -H(x)\nabla f(x)$ é uma direção de decida, visto que a condição suficiente do Lema A.2 é satisfeita, ou seja, $\langle \nabla f(x), d \rangle < 0$. Dessa forma, o método de descida descrito acima pode ser reescrito da seguinte forma utilizando o procedimento de Busca Linear de Armijo.

Método de Descida

Passo 0: Defina $x^0 \in \mathbb{R}^n, k := 0$;

Passo 1: Defina $d^k = -H(x^k)\nabla f(x^k)$;

Passo 2: Verifique se $\left\| \nabla f(x^k) \right\| < \varepsilon$. Caso contrário, compute $t_k > 0$ satisfazendo o procedimento de Busca Linear da Regra de Armijo (1.3).

Passo 3: Faça

$$x^{k+1} = x^k + t_k d^k, \tag{1.4}$$

com $k := k + 1$ e retorne ao passo 1.

Por fim, o Método do Gradiente (ou de Cauchy) é caracterizado por empregar a matriz $H(x) = I_n$ (matriz identidade de ordem n) na determinação da direção de descida. Consequentemente, a direção de descida d^k torna-se a clássica direção de máxima descida, definida por $d^k = -\nabla f(x^k)$.

Para realizarmos a análise de convergência (total e parcial) do Método do Gradiente, ou seja $d^k = -\nabla f(x^k)$ com t_k determinado pela Regra de Armijo, assumimos que a sequência gerada (x^k) é infinita. Em outras palavras, essa suposição implica que o critério de parada baseado no gradiente nulo não é satisfeito em um número finito de iterações, ou seja, $\|\nabla f(x^k)\| \neq 0$ para todo $k \in \mathbb{N}$.

Teorema 1.2. *(Convergência parcial) Considere uma função continuamente diferenciável $f : \mathbb{R}^n \to \mathbb{R}$. Seja (x^k) a sequência gerada pelo Método de descida e (t_k) gerada pela regra de Armijo. Então todo ponto de acumulação de (x^k) é ponto estacionário.*

Demonstração. Sejam (x^k) uma sequência gerada pelo algorítimo e x^* um ponto de acumulação de (x^k), ou seja, existe uma subsequência $(x^{k_j}) \to x^*$. Vamos supor por absurdo que x^* não é estacionário, isto é, $\nabla f(x^*) \neq 0$. Como f é contínua, temos $(f(x^{k_j})) \to f(x^*)$. Porém, como $(f(x^k))$ é monótona não-crescente, $(f(x^k)) \to f(x^*)$. Por outro lado, obtemos da Busca de Armijo (1.3),

$$f(x^{k+1}) = f(x^k + t_k d^k) \leqslant f(x^k) + \eta t_k \langle \nabla f(x^k), d^k \rangle,$$

e pelo método de descida (1.4), temos $d^k = -H(x^k)\nabla f(x^k)$. Agora, como $H(x)$ é definida positiva, obtemos

$$f(x^k) - f(x^{k+1}) \geqslant \eta t_k \langle \nabla f(x^k), H(x^k)\nabla f(x^k) \rangle \geqslant 0.$$

Agora, utilizando o fato de $(f(x^k)) \to f(x^*)$ e o Teorema do Confronto, temos

$$\lim_{k \to +\infty} t_k \langle \nabla f(x^k), H(x^k)\nabla f(x^k) \rangle = 0.$$

Daí, separamos a análise em dois casos. Primeiro, se $0 < \bar{t} \leqslant t_k$, para todo $k \in \mathbb{N}$, então o resultado segue da última igualdade. Caso contrário, suponha sem perda de generalidade que $t_{k_j} \to 0$. Então para $k_j \in \mathbb{N}$ suficientemente grande temos $t_{k_j} < 1$. Nesse caso, note que

$$\lim_{k_j \to +\infty} \langle \nabla f(x_j^k), H(x^{k_j})\nabla f(x^{k_j}) \rangle = \langle \nabla f(x^*), H(x^*)\nabla f(x^*) \rangle \neq 0.$$

Agora, o passo $\frac{t_{k_j}}{\theta}$, com $\theta \in (0, 1)$, existiu e foi recusado pelo procedimento de Busca Linear de Armijo. Daí, por um lado, temos

$$f(x^{k_j} + t_{k_j} d^{k_j}) \leqslant f(x^{k_j}) + \eta t_{k_j} \langle \nabla f(x^{k_j}), d^{k_j} \rangle,$$

por outro lado,

$$f\left(x^{k_j} + \frac{t_{k_j}}{\theta} d^{k_j}\right) > f(x^{k_j}) + \eta \frac{t_{k_j}}{\theta} \langle \nabla f(x^{k_j}), d^{k_j} \rangle.$$

Agora, seja $\sigma(t) = f(x^{k_j} + td^{k_j}) - f(x^{k_j}) - \eta t\langle \nabla f(x^{k_j}), d^{k_j}\rangle$, onde $\sigma(t)$ é contínua. Pelo Teorema do Valor Intermediário (TVI) (A.1) e das últimas duas desigualdades acima, existe $s_{k_j} \in [t_{k_j}, \frac{t_{k_j}}{\theta}]$ tal que $\sigma(s_{k_j}) = 0$, isto é,

$$f(x^{k_j} + s_{k_j} d^{k_j}) - f(x^{k_j}) = \eta s_{k_j} \langle \nabla f(x^{k_j}), d^{k_j}\rangle.$$

Agora, pelo Teorema do Valor Médio (TVM) (A.2), existe $\beta_{k_j} \in (0, 1)$ tal que

$$\begin{aligned} s_{k_j} \langle \nabla f(x^{k_j} + \beta_{k_j} s_{k_j} d^{k_j}), d^{k_j}\rangle &= f(x^{k_j} + s_{k_j} d^{k_j}) - f(x^{k_j}) \\ &= \eta s_{k_j} \langle \nabla f(x^{k_j}), d^{k_j}\rangle. \end{aligned}$$

Sendo assim, como no método de descida, $d^{k_j} = -H(x^{k_j})\nabla f(x^{k_j})$, obtemos

$$-s_{k_j} \langle \nabla f(x^{k_j} + \beta_{k_j} s_{k_j} d^{k_j}), H(x^{k_j})\nabla f(x^{k_j})\rangle = -\eta s_{k_j} \langle \nabla f(x^{k_j}), H(x^{k_j})\nabla f(x^{k_j})\rangle,$$

o que por sua vez, organizando, temos

$$\langle \nabla f(x^{k_j} + \beta_{k_j} s_{k_j} d^{k_j}), H(x^{k_j})\nabla f(x^{k_j})\rangle = \eta \langle \nabla f(x^{k_j}), H(x^{k_j})\nabla f(x^{k_j})\rangle, \tag{1.5}$$

podemos perceber que $t_{k_j} \leqslant s_{k_j} \leqslant \frac{t_{k_j}}{\theta}$, dessa forma, como $t_{k_j} \to 0$, então $s_{k_j} \to 0$. Por consequência $\beta_{k_j} s_{k_j} d^{k_j} \to 0$ e como vimos no início da prova que $x^{k_j} \to x^*$, podemos concluir que $x^{k_j} + \beta_{k_j} s_{k_j} d^{k_j} \to x^*$. Dessa forma, fazendo $k_j \to +\infty$ em (1.5), temos

$$\langle \nabla f(x^*), H(x^*)\nabla f(x^*)\rangle = \eta \langle \nabla f(x^*), H(x^*)\nabla f(x^*)\rangle, \quad \eta \in (0, 1).$$

Entretanto, a última igualdade só é verdade se $\langle \nabla f(x^*), H(x^*)\nabla f(x^*)\rangle = 0$, ou seja, $\nabla f(x^*) = 0$, o que por sua vez gera um absurdo, pois supomos que $\nabla f(x^*) \neq 0$. Dessa forma, todo ponto de acumulação de (x^k) é ponto estacionário. □

Após a demonstração da convergência parcial do Método do Gradiente, a conclusão da convergência total da sequência gerada para uma solução do Problema (1.2) requer, como veremos a seguir, a suposição adicional da convexidade da função objetivo, veja Definição A.4. Tal resultado se baseia fortemente na ideia de convergência Quasi-Fejér (veja Definição 1.4) e no Lema 1.1).

Teorema 1.3. *(Convergência total) Seja* (x^k) *a sequência gerada pelo Método de descida que utiliza a regra de Armijo com* $\eta \in (0, 1)$, *então* (x^k) *converge para uma solução do problema* (1.2).

Demonstração. Uma vez que o conjunto solução S^* é não-vazio, considere $x^* \in S^*$. Então,

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &= \|x^{k+1} - x^k + x^k - x^*\|^2 \\ &= \|x^{k+1} - x^k\|^2 + 2\langle x^{k+1} - x^k, x^k - x^*\rangle + \|x^k - x^*\|^2.\end{aligned}$$

Segue do Passo 3 do Método de Descida (1.4) $x^{k+1} - x^k = -t_k \nabla f(x^k)$. Então,

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &= \| - t_k \nabla f(x^k)\|^2 + 2\langle -t_k \nabla f(x^k), x^k - x^*\rangle + \|x^k - x^*\|^2 \\ &= t_k^2 \|\nabla f(x^k)\|^2 + 2t_k \langle \nabla f(x^k), x^* - x^k\rangle + \|x^k - x^*\|^2. \end{aligned} \tag{1.6}$$

Por outro lado, segue da Desigualdade do Gradiente, veja (A.1) que

$$f(x^*) \geqslant f(x^k) + \langle \nabla f(x^k), x^* - x^k\rangle,$$

o que por sua vez, implica que

$$\langle \nabla f(x^k), x^* - x^k\rangle \leqslant f(x^*) - f(x^k).$$

Sendo assim, substituindo a última desigualdade em (1.6), obtemos que

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leqslant \|x^k - x^*\|^2 + t_k^2 \|\nabla f(x^k)\|^2 + 2t_k (f(x^*) - f(x^k)) \\ &\leqslant \|x^k - x^*\|^2 + t_k^2 \|\nabla f(x^k)\|^2, \end{aligned} \tag{1.7}$$

onde a última desigualdade segue do fato que $f(x^*) - f(x^k) \leqslant 0$ e $t_k > 0$.

Assim, observa-se que a desigualdade (1.7) está bem próxima da desigualdade da Definição 1.4, referente a Quasi-Fejér convergência, restando demonstrar que o termo $t_k^2 \|\nabla f(x^k)\|^2$ é somável. Uma vez que, $t_k \leqslant 1$ (uma condição usualmente garantida pela regra de Busca Linear, como Armijo), segue que $t_k^2 \leqslant t_k$. Multiplicando ambos os lados por $\|\nabla f(x^k)\|^2$, concluímos que

$$t_k^2 \|\nabla f(x^k)\|^2 \leqslant t_k \|\nabla f(x^k)\|^2.$$

Sendo assim, segue do Teste da Comparação de séries que é suficiente provar que

$$\sum_{k=0}^{+\infty} t_k \|\nabla f(x^k)\|^2 < +\infty.$$

De fato, pela regra de Armijo (1.3), temos

$$\sigma t_k \|\nabla f(x^k)\|^2 \leqslant f(x^k) - f(x^{k+1}), \quad \forall\, k \in \mathbb{N}.$$

Somando essa desigualdade para $k = 0, \dots, N$, obtemos

$$\sigma \sum_{k=0}^{N} t_k \|\nabla f(x^k)\|^2 \leqslant f(x^0) - f(x^{N+1}).$$

Agora, como $f(x^*) \leqslant f(x^{N+1})$, visto que $x^* \in S^*$, então $-f(x^{N+1}) \leqslant -f(x^*)$. Substituindo, obtemos que

$$\sigma \sum_{k=0}^{N} t_k \|\nabla f(x^k)\|^2 \leqslant f(x^0) - f(x^*).$$

Passando limite com $N \to \infty$,

$$\sigma \sum_{k=0}^{\infty} t_k \|\nabla f(x^k)\|^2 \leqslant f(x^0) - f(x^*) < +\infty.$$

Em particular, seque que $\sum_{k=0}^{\infty} t_k \|\nabla f(x^k)\|^2 < +\infty$. Logo (x^k) é Quasi-Fejer convergente em relação ao conjunto S^* com $\varepsilon_k := t_k \|\nabla f(x^k)\|$. Dessa forma, segue do Lema 1.1 (i) que (x^k) é limitada. Então, existe $(x^{k_j}) \subset (x^k)$ tal que $(x^{k_j}) \to \bar{x}$. Daí, segue do Teorema 1.2 que $\nabla f(\bar{x}) = 0$. Como f é uma função convexa, segue do Teorema A.1 (Desigualdade do Gradiente) que

$$f(y) \geqslant f(\bar{x}) + \langle \nabla f(\bar{x}), y - \bar{x} \rangle = f(\bar{x}),$$

ou seja,

$$f(y) \geqslant f(\bar{x}), \ y \in \mathbb{R}^n.$$

Logo, $\bar{x} \in S^*$. Pelo Lema 1.1 (ii), temos $(x^k) \to \bar{x}$. Portanto, (x^k) converge para um ponto que é solução do Problema (1.2). □

1.2.2 Método do Ponto Proximal

Nesta subseção, revisitamos o problema de minimização da seção anterior, com a particularidade de que a função objetivo possui contradomínio $\overline{\mathbb{R}} := \mathbb{R} \cup +\infty$. Em resumo, considere o seguinte problema

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1.8}$$

onde $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ é possivelmente não-diferenciável, convexa e contínua. De forma análoga à subseção anterior, assumimos que o conjunto solução do Problema 1.8, denotado por S^*, é não-vazio. Utilizaremos o chamado Método Ponto Proximal para resolver o problema (1.8). Inicialmente, considere o chamado operador proximal para $f(\cdot)$, o qual é dado por

$$\operatorname{prox}_{\alpha f}(x) := \arg\min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\alpha} \|y - x\|^2 \right\}, \tag{1.9}$$

onde a norma, $\|\cdot\|$ é induzida pelo produto interno de $\mathbb{R}^n$ $\langle\cdot,\cdot\rangle$, com $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$.

O Método do Ponto Proximal pode ser interpretado como um método para resolver inclusões monótonas do tipo $0 \in \partial f(x)$, sendo o operador ∂f maximal e monótono. Tendo em vista a condição de otimalidade de (1.9), concluímos a seguinte inclusão monótona:

$$0 \in \partial f(\mathtt{prox}_{\alpha f}(x)) + \frac{1}{\alpha}\left(\mathtt{prox}_{\alpha f}(x) - x\right). \tag{1.10}$$

Sendo assim, podemos dizer que existe um subgradiente z no subdiferencial $\partial f(\text{prox}_{\alpha f}(x))$, tal que

$$z + \frac{1}{\alpha}\left(\text{prox}_{\alpha f}(x) - x\right) = 0. \tag{1.11}$$

A seguir, trazemos o clássico Método do Ponto Proximal para resolver o Problema (1.8)

Método Pronto Proximal (MPP)

Seja a sequência $(\alpha_k) \subset \mathbb{R}_{++}$, tal que $0 < \hat{\alpha} \leqslant \alpha_k \leqslant \overline{\alpha}$ para todo $k \in \mathbb{N}$.

Passo 1: $x^0 \in \mathbb{R}^n$, $k := 0$;

Passo 2: Dado $x^k \in \mathbb{R}^n$;

$$x^{k+1} = \text{prox}_{\alpha_k f}(x^k) := \arg\min_{y\in\mathbb{R}^n}\left\{f(y) + \frac{1}{2\alpha_k}\|y - x^k\|^2\right\}. \tag{1.12}$$

Passo 3: Se $x^{k+1} = x^k$ pare. Caso contrário, faça $k := k+1$ e retorne ao passo 2.

Proposição 1.2. *A sequência* (x^k) *gerada pelo Método do Ponto Proximal dada em* (1.12) *está bem definida.*

Demonstração. A prova é feita por indução. Dado $k \in \mathbb{N}$, considere a função $f_k : \mathbb{R}^n \to \mathbb{R}$ definida por

$$f_k(x) = f(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2.$$

Como $S^* \neq \emptyset$, temos que f possui mínimo, em particular, f é limitada inferiormente. Sendo assim, $f_k(\cdot)$ é também limitada inferiormente. Por outro lado, temos que $f(\cdot)$ é uma função convexa e continua, (α_k) é limitada e $\|\cdot - x^k\|^2$ é continua e estritamente convexa. Logo, podemos concluir que $f_k(\cdot)$ é também continua (pois a soma de duas funções continuas é continua), estritamente convexa (pois é a soma de uma função convexa com uma estritamente convexa). Por fim, seja (y^j) tal que $(\|y^j\|) \to \infty$, quando $j \to \infty$. Mas, $\lim\limits_{j\to\infty} f(y^j) > -\infty$, pois $f(\cdot)$ é limitada inferiormente. Agora,

$$\|y^j\| = \|y^j - x^k + x^k\| \leqslant \|y^j - x^k\| + \|x^k\|,$$

isto é,

$$\|y^j\| \leqslant \|y^j - x^k\| + \|x^k\|.$$

Uma vez que $\|y^j\| \to \infty$, segue da última desigualdade acima que $\|y^j - x^k\| \to \infty$, quando $j \to +\infty$. Então, uma vez que $\alpha_k > \hat{\alpha} > 0$, temos

$$\lim_{j\to\infty} f_k(y^j) = \lim_{j\to\infty} \left(f(y^j) + \frac{1}{2\alpha_k}\|y^j - x^k\|^2 \right) = +\infty.$$

Logo, segue da Definição (A.6) que $f_k(\cdot)$ é coerciva. Sendo assim, como $f_k(x)$ é contínua e coerciva, então possui mínimo, veja Teorema (A.6). Finalmente, segue do fato de $f_k(x)$ ser estritamente convexa que tal minimizador é único. Logo, x^{k+1} dado em (1.12) existe e é único, ou seja, o método está bem definido. □

A seguinte observação é um fato muito útil e bem conhecido em relação ao Método do Ponto Proximal. Tal fato, é bem útil no que diz respeito ao critério de parada do método.

Proposição 1.3. *Seja* x^{k+1} *definido como em* (1.12)*, temos que* $x^{k+1} = x^k$ *se, e somente se,* x^k *é solução do Problema 1.8.*

Demonstração. Inicialmente, lembre que $x^{k+1} \in \mathbb{R}^n$ é uma solução global de (1.8), se e somente se,

$$0 \in \partial f(x^{k+1}).$$

Vamos supor primeiramente que $x^{k+1} = x^k$. Tendo em vista a condição de otimalidade do subproblema proximal dada em (1.10) com $x = x^k, \alpha = \alpha_k$, e $\text{prox}_{\alpha f}(x) = x^{k+1}$, temos

$$0 \in \partial f(x^{k+1}) + \frac{1}{\alpha_k}(x^{k+1} - x^k). \tag{1.13}$$

Sendo assim, podemos dizer que existe um subgradiente z_k no subdiferencial $\partial f(x^{k+1})$, tal que

$$z_k + \frac{1}{\alpha_k}(x^{k+1} - x^k) = 0. \tag{1.14}$$

Agora, desde que $x^{k+1} = x^k$, concluímos que $z_k = 0$. Utilizando a Definição de Subgradiente (veja Definição (A.8)), temos

$$f(y) \geqslant f(x^{k+1}) + \langle z_k, y - x^{k+1}\rangle, \quad \forall y \in \mathbb{R}^n.$$

o que por sua vez, como $z_k = 0$, implica em $f(y) \geqslant f(x^{k+1})$. Logo $x^k = x^{k+1}$ é solução global de (1.8).

Reciprocamente, vamos supor que x^k é solução global de (1.8), ou seja, para todo $y \in \mathbb{R}^n$ temos $f(y) \geqslant f(x^k)$. Em particular, tomando $y = x^{k+1}$, obtemos que

$$f(x^{k+1}) \geqslant f(x^k). \tag{1.15}$$

Por outro lado, usando a desigualdade do subgradiente em x^{k+1} para qualquer $z \in \partial f(x^{k+1})$, vale

$$f(x^k) \geqslant f(x^{k+1}) + \langle z, x^k - x^{k+1} \rangle. \tag{1.16}$$

Daí, segue da última desigualdade e de (1.15), que

$$0 \geqslant f(x^k) - f(x^{k+1}) \geqslant \langle z, x^k - x^{k+1} \rangle.$$

Agora, segue novamente da condição de otimalidade do subproblema proximal dada em (1.11), temos que $z_k = -\frac{1}{\alpha_k}(x^{k+1} - x^k)$. Fazendo a substituição do valor de z_k em z na desigualdade anterior,

$$\begin{aligned} 0 &\geqslant \langle -\frac{1}{\alpha_k}(x^{k+1} - x^k), x^k - x^{k+1} \rangle \\ &= \frac{1}{\alpha_k}\langle -(x^{k+1} - x^k), x^k - x^{k+1} \rangle \\ &= \frac{1}{\alpha_k}\|x^{k+1} - x^k\|^2 \geqslant 0. \end{aligned}$$

Por fim, desde que $\alpha_k > 0$ e $\|x^{k+1} - x^k\|^2 \geqslant 0$, obtemos que $x^{k+1} = x^k$, como queríamos demonstrar. □

Teorema 1.4. *Seja* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *uma função convexa. Suponha que o conjunto* S^* *de minimizadores de* f *em* $\mathbb{R}^n$*, é não-vazio. Então, a sequência* (x^k) *gerada pelo MPP converge para um ponto* $x^* \in S^*$.

Demonstração. Provaremos que para todo $k \in \mathbb{N}$ e todo $\overline{x} \in S^*$ vale a desigualdade,

$$\|x^{k+1} - \overline{x}\|^2 \leqslant \|x^k - \overline{x}\|^2 - \|x^{k+1} - x^k\|^2.$$

Temos por hipótese que $S^* \neq 0$, logo podemos pegar $\overline{x} \in S^*$. Portanto,

$$\begin{aligned} \|x^k - \overline{x}\|^2 &= \|x^k - x^{k+1} + x^{k+1} - \overline{x}\|^2 \\ &= \|x^k - x^{k+1}\|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - \overline{x} \rangle + \|x^{k+1} - \overline{x}\|^2, \end{aligned}$$

sendo assim,

$$\|x^k - \overline{x}\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - \overline{x}\|^2 = 2\langle x^k - x^{k+1}, x^{k+1} - \overline{x} \rangle. \tag{1.17}$$

Por (1.11), podemos concluir que existe $u^{k+1} \in \partial f(x^{k+1})$ tal que $u^{k+1} + \dfrac{1}{\alpha_k}(x^{k+1} - x^k) = 0$. Então,

$$x^k - x^{k+1} = \alpha_k u^{k+1},$$

substituindo a igualdade anterior do lado direito em (1.17), obtemos

$$\|x^k - \bar{x}\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - \bar{x}\|^2 = 2\alpha_k \langle u^{k+1}, x^{k+1} - \bar{x}\rangle. \tag{1.18}$$

Por outro lado, como $u^{k+1} \in \partial f(x^{k+1})$, então obtemos pela desigualdade do subgradiente que

$$f(y) \geqslant f(x^{k+1}) + \langle u^{k+1}, y - x^{k+1}\rangle, \quad \forall y \in \mathbb{R}^n.$$

O que por sua vez, implica que

$$f(x^{k+1}) - f(y) \leqslant \langle u^{k+1}, x^{k+1} - y\rangle, \quad \forall y \in \mathbb{R}^n.$$

Tomando $y = \bar{x}$, temos

$$f(x^{k+1}) - f(\bar{x}) \leqslant \langle u^{k+1}, x^{k+1} - \bar{x}\rangle.$$

Uma vez que $\bar{x} \in S^*$, então $f(\bar{x}) \leqslant f(x^{k+1})$, logo $f(x^{k+1}) - f(\bar{x}) \geqslant 0$. Daí, concluímos que $\langle u^{k+1}, x^{k+1} - \bar{x}\rangle \geqslant 0$. Sendo assim, por (1.18) e pelo fato de $\alpha_k > 0$, podemos concluir que

$$\|x^k - \bar{x}\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - \bar{x}\|^2 \geqslant 0,$$

o que por sua vez, implica que

$$\|x^{k+1} - \bar{x}\|^2 \leqslant \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2.$$

Dessa forma, podemos observar que a desigualdade que acabamos de provar nos diz que a sequência (x^k) é Fejér convergente em relação ao conjunto S^*, pois obtemos que

$$0 \leqslant \|x^{k+1} - \bar{x}\|^2 \leqslant \|x^k - \bar{x}\|^2.$$

Agora provaremos que $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$. Temos que (x^k) é Fejér convergente em relação à S^*. Dessa forma, segue do Lema 1.1 (i) que (x^k) é limitada, ou seja existe $(x^{k_j}) \subset (x^k)$ convergente, digamos que $x^{k_j} \to x^*$. Vimos anteriormente que $(\|x^k - \bar{x}\|)$ é monótona, não-crescente (pela definição de Fejér convergente), limitada (pois, é limitada por cima por $\|x^0 - \bar{x}\|^2$ e por baixo por zero). Logo, $(\|x^k - \bar{x}\|)$ é convergente. Como

$$0 \leqslant \|x^{k+1} - x^k\|^2 \leqslant \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2,$$

quando $k \to \infty$, temos que $(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2)$ converge para zero, pois $(\|x^k - \bar{x}\|)$ é convergente. Sendo assim, pelo Teorema do Confronto, concluímos que $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$.

Por fim, provaremos que (x^k) possui ponto de acumulação em S^*. Temos que $x^{k_j} \to x^*$ e que $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$, dessa forma, $x^{k_j+1} \to x^*$.
Agora, fazendo uma manipulação algébrica e utilizando a Desigualdade Triangular, temos

$$0 \leqslant \|x^{k_j+1} - x^*\| = \|x^{k_j+1} - x^{k_j} + x^{k_j} - x^*\| \leqslant \|x^{k_j+1} - x^{k_j}\| + \|x^{k_j} - x^*\|,$$

o que por sua vez, implica que

$$0 \leqslant \|x^{k_j+1} - x^*\| \leqslant \|x^{k_j+1} - x^{k_j}\| + \|x^{k_j} - x^*\|.$$

Quando $k_j \to \infty$, temos que $\lim\limits_{k_j\to\infty} \|x^{k_j+1} - x^{k_j}\| \to 0$ (pois, $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$) e $\|x^{k_j} - x^*\| \to 0$, pois $x^{k_j} \to x^*$. Portanto, pelo Teorema do Confronto, obtemos que $\lim\limits_{k_j\to\infty} \|x^{k_j+1} - x^*\| = 0$.

Daí, vimos anteriormente que a condição de otimalidade do subproblema proximal dada em (1.10) com $x = x^k, \alpha = \alpha_k$ e $\operatorname{prox}_{\alpha f}(x) = x^{k+1}$ nos fornece a seguinte inclusão

$$-\frac{1}{\alpha_k}(x^{k+1} - x^k) \in \partial f(x^{k+1}).$$

Substituindo x^k por x^{k_j}, temos

$$-\frac{1}{\alpha_{k_j}}(x^{k_j+1} - x^{k_j}) \in \partial f(x^{k_j+1}),$$

o que por sua vez nos diz que existe $u^{k_j+1} \in \partial f(x^{k_j+1})$ tal que

$$u^{k_j+1} = -\frac{1}{\alpha_{k_j}}(x^{k_j+1} - x^{k_j}).$$

Temos que (α_k) é limitada e $\|x^{k_j+1} - x^{k_j}\| \to 0$, quando $k_j \to \infty$. Logo, tomando o limite com $k_j \to \infty$ na igualdade anterior, temos que $u^{k_j+1} = -\dfrac{1}{\alpha_k}(x^{k_j+1} - x^{k_j}) \to 0$. Temos também que $\|x^{k_j+1} - x^*\| \to 0$ e $u^{k_j+1} \in \partial f(x^{k+1})$. Dessa forma, pelo Fato A.1 concluímos que $0 \in \partial f(x^*)$. Então, segue do Teorema A.10 e da convexidade de f que $x^* \in S^*$. Agora, novamente pelo Lema 1.1 (ii) temos que a sequência (x^k) converge para x^*, o qual é um ponto solução do Problema (1.8).

□

1.2.3 Introdução ao Método Proximal Gradiente

Nesta subseção, focamos em uma classe particular de problemas de otimização convexa, onde a função objetivo é expressa como a soma de duas funções, uma das quais suposta ser continuamente diferenciável e a outra possivelmente não-diferenciável. Mais especificamente, consideramos o problema na forma:

$$\min_{x \in \mathbb{R}^n} F(x) = g(x) + h(x), \tag{1.19}$$

onde $g : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa e diferenciável e $h : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ é uma função convexa, própria e possivelmente não-diferenciável.

Tal problema, na literatura relacionada é chamado de "composite problems". Por exemplo, em problemas de aprendizado de máquina com regularização, $g(x)$ pode representar o termo de erro (i.e., erro quadrático) e $h(x)$ o termo de regularização (i.e., norma ℓ_1 para promover esparsidade ou a função indicadora de um conjunto convexo para impor restrições). Para mais detalhes, veja [9, 10, 11, 12, 13]. A presença do termo não-diferenciável $h(x)$ impede a aplicação direta de métodos clássicos, como o Método do Gradiente, que exigem a diferenciabilidade da função objetivo completa.

Diante deste cenário, métodos que conseguem desacoplar as propriedades das funções g e h são essenciais. O Método Proximal Gradiente surge como uma ferramenta poderosa e eficiente para resolver problemas da forma (1.19), combinando passos do gradiente para a componente suave g com passos do operador proximal para a componente h. No decorrer deste trabalho, detalharemos este método e exploraremos suas propriedades de convergência.

A etapa principal no método de proximal gradiente envolve avaliar o Operador Proximal definido como segue:

$$\mathrm{prox}_{\alpha h}(x) := \arg\min_{y \in \mathbb{R}^n} \left\{ h(y) + \frac{1}{2\alpha}\|y - x\|^2 \right\}.$$

Método Proximal Gradiente Clássico (MPG-C)

Seja $x_0 \in \text{dom}(h)$. Calcule $\lambda_k > 0$ e

$$\tilde{x}_k := \text{prox}_{\alpha_k h}\left(x^k - \lambda_k \nabla g(x^k)\right). \tag{1.20}$$

Escolha $\beta_k \in (0, 1]$ e calcule

$$x_{k+1} := x^k + \beta_k(\tilde{x}_k - x^k).$$

Os coeficientes λ_k e β_k, referidos como tamanhos de passo, podem ser determinados com base em procedimentos de Busca Linear retrocedentes. Tais estratégias são essenciais sempre que a continuidade global L-Lipschitz do gradiente (Ver Definição 1.3) de g falha ou mesmo quando calcular um limite superior aceitável para L é desafiador. Essa situação é frequentemente encontrada em numerosas aplicações; por exemplo, em problemas inversos baseados em normas não euclidianas [16, 30] ou distâncias de Bregman, como a divergência de Kullback–Leibler [15, 19, 31, 33]. Além disso, mesmo quando L é conhecido, Buscas Lineares podem permitir passos mais longos em direção à solução usando informações locais em cada iteração.

Existem várias escolhas possíveis para esses tamanhos de passo, cada uma impactando o desempenho do algoritmo de maneiras diferentes; veja, por exemplo, [9, 8, 31, 33]. É importante notar que, para calcular o tamanho de passo λ_k usando uma Busca Linear retrocedente em cada iteração k, o operador proximal pode precisar ser avaliado várias vezes dentro do procedimento, nesse caso temos a chamada Busca Linear implícita. Por outro lado, o tamanho de passo β_k pode ser selecionado avaliando o operador proximal apenas uma vez por iteração. Neste contexto, nos referiremos a Busca Linear explícita para descrever um procedimento de retrocesso que determina β_k após definir λ_k como uma constante para todo k. Tal dilema abordado acima, em relação a possibilidade ou não do operador proximal ser resolvido mais de uma vez por iteração é um dos principais tópicos a serem discutidos nos capítulos seguintes.

O tipo de estratégia explícita, apresentada pela primeira vez em [9], é particularmente vantajosa, especialmente nos casos em que avaliar o operador proximal é desafiador. A função h pode ser frequentemente complexa o suficiente para que o operador proximal correspondente não tenha uma solução analítica.

Capítulo 2

Procedimentos de Buscas Lineares

Antes de expormos os procedimentos de Buscas Lineares, tema central deste capítulo, enunciaremos algumas hipóteses relativas às funções g e h, que desempenharão papel essencial no decorrer deste trabalho.

Estamos interessados em resolver o seguinte problema de otimização:

$$\min_{x \in \mathbb{R}^n} F(x) = g(x) + h(x). \tag{2.1}$$

onde $g, h \in \Gamma_0(\mathbb{R}^n)$, de forma que $\Gamma_0(\mathbb{R}^n)$ é o conjunto de funções próprias, semi-contínuas inferior e convexas em $\mathbb{R}^n$. A função g é suposta ser diferenciável no $\mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$.

Agora, mostraremos as hipóteses principais para este trabalho, que envolvem características para as funções g e h.

A1. $g, h \in \Gamma_0(\mathbb{R}^n)$ com $\mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h) \neq \emptyset$ e $\mathrm{dom}(h) \subseteq \mathrm{dom}(g)$.

A2. A função g é diferenciável ao longo de um conjunto aberto que contém dom(h). Além disso, o gradiente ∇g é uniformemente contínuo em qualquer subconjunto limitado de dom(h) e mapeia qualquer subconjunto limitado de dom(h) para um conjunto limitado de $\mathbb{R}^n$.

A3 - g é diferenciável em todo ponto de $\mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h), \nabla g$ é uniformemente contínua em qualquer subconjunto compacto de $\mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$.

A4 - Para qualquer $x \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$, existe $L_c F(x) = \{x \in \mathbb{R}^n | \ c \leqslant F(x)\} \subset \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$ e para qualquer subconjunto compacto de x de $L_c F(x)$, temos

$$d(x;\mathbb{R}^n \setminus \mathrm{int}(\mathrm{dom}(g))) > 0. \tag{2.2}$$

Sob essas hipóteses, definimos o operador proximal forward-backward

$$J : [\mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)] \times \mathbb{R}_{++} \to \mathrm{dom}(h),$$

por

$$J(x,\alpha) := \mathrm{prox}_{\alpha h}\big(x - \alpha\nabla g(x)\big), \tag{2.3}$$

O lema apresentado a seguir, é consequência direta de [[5], Teorema 23.47] e será empregado na prova da boa-definição Busca Linear 3.

Lema 2.1. *Seja* $h \in \Gamma_0(\mathbb{R}^n)$ *e seja* $\alpha > 0$. *Então, para todo* $x \in \mathrm{dom}(h)$, *temos*

$$\mathrm{prox}_{\alpha h}(x) \to x, \qquad \textit{quando } \alpha \to 0. \tag{2.4}$$

Observação 2.1. *Note que segue da condição de otimalidade vista em*(1.10) *com* $f = h$ *que*

$$0 \in \partial h(\mathrm{prox}_{\alpha h}(z)) + \frac{1}{\alpha}(\mathrm{prox}_{\alpha h}(z) - z),$$

o que por sua vez, reorganizando e lembrando que $\alpha > 0$, *temos*

$$\frac{z - \mathit{prox}_{\alpha h}(z)}{\alpha} \in \partial h(\mathrm{prox}_{\alpha h}(z)), \quad \forall\, z \in \mathbb{R}^n. \tag{2.5}$$

Fazendo $z = x - \alpha\nabla g(x)$, *obtemos*

$$\frac{x - \alpha\nabla g(x) - \mathit{prox}_{\alpha h}(x - \alpha\nabla g(x))}{\alpha} \in \partial h(\mathrm{prox}_{\alpha h}(x - \alpha\nabla g(x)), \tag{2.6}$$

o que por sua vez, implica que

$$\frac{x - \mathit{prox}_{\alpha h}(x - \alpha\nabla g(x))}{\alpha} \in \nabla g(x) + \partial h(\mathrm{prox}_{\alpha h}(x - \alpha\nabla g(x)).$$

Agora, como $J(x,\alpha) = \mathrm{prox}_{\alpha h}\big(x - \alpha\nabla g(x)\big)$, *obtemos*

$$\frac{x - J(x,\alpha)}{\alpha} \in \nabla g(x) + \partial h(J(x,\alpha)). \tag{2.7}$$

Seja S^* *o conjunto de soluções ótimas do Problema* (2.1) *e* x^* *um ponto no* $\mathtt{int}(\mathtt{dom}(g)) \cap \mathtt{dom}(h)$. *Pelo Teorema A.10, podemos concluir que* $x^* \in S^*$ *se, e somente se,*

$$0 \in \partial(g+h)(x^*) = \nabla g(x^*) + \partial h(x^*). \tag{2.8}$$

É importante salientar também a seguinte observação, na qual utilizaremos com recorrência na prova da boa definição das Buscas Lineares, bem como no critério de parada dos métodos a serem analisados.

Observação 2.2. *Considere* $x \in \mathbb{R}^n$ *e* $\alpha \in \mathbb{R}_{++}$. *Logo,* $x \in S^*$, *isto é,* x *é uma solução global para o Problema* (2.1) *se, e somente se,* $x = J(x, \alpha)$.

Primeiramente vamos admitir que $x \in S^*$, *sendo assim, temos que* $0 \in \nabla g(x) + \partial h(x)$, *ou seja,* $-\nabla g(x) \in \partial h(x)$. *Multiplicando por* $\alpha > 0$, *temos*

$$-\alpha \nabla g(x) \in \alpha \partial h(x),$$

o que por sua vez, implica que

$$x - \alpha \nabla g(x) - x \in \alpha \partial h(x).$$

Utilizando a caracterização fundamental do Operador Proximal dada em (1.10), *segue que*

$$x = \mathit{prox}_{\alpha h}(z) \Leftrightarrow z - x \in \alpha \partial h(x).$$

Agora, fazendo $z = x - \alpha \nabla g(x)$, *obtemos que* $x = \mathit{prox}_{\alpha h}(x - \alpha \nabla g(x))$. *Dessa forma, por* (2.3), *que diz que* $J(x, \alpha) = \mathit{prox}_{\alpha h}(x - \alpha \nabla g(x))$ *concluímos que* $x = J(x, \alpha)$.

Agora, se $x = J(x, \alpha)$ *então por* (2.7) *concluímos que*

$$\frac{x - x}{\alpha} \in \nabla g(x) + \partial h(x),$$

sendo assim, $0 \in \nabla g(x) + \partial h(x)$, *ou seja,* x *é ponto crítico para o Problema* (2.1), *o que por sua vez pela convexidade da função objetivo implica que* $x \in S^*$.

O seguinte Lema será de grande importância para nosso estudo, pois será essencial para mostrar que a primeira regra de Busca Linear está bem definida.

Lema 2.2. *Para qualquer* $x \in$ *int(dom(g))* $\cap$ *dom(h) , temos:*

$$\frac{\alpha_2}{\alpha_1}\|x - J(x, \alpha_1)\| \geqslant \|x - J(x, \alpha_2)\| \geqslant \|x - J(x, \alpha_1)\|. \tag{2.9}$$

Para todo $\alpha_2 \geqslant \alpha_1 > 0$.

Demonstração. Utilizando (2.5) com $z = x - \alpha \nabla g(x)$, obtemos

$$\frac{x - \alpha \nabla g(x) - \mathrm{prox}_{\alpha h}(x - \alpha \nabla g(x))}{\alpha} \in \partial h(\mathrm{prox}_{\alpha h}(x - \alpha \nabla g(x))).$$

Agora utilizando (2.3), temos

$$\frac{x - \alpha\nabla g(x) - J(x,\alpha)}{\alpha} \in \partial h(J(x,\alpha)).$$

Para qualquer $\alpha > 0$, $\alpha_2 \geqslant \alpha_1 > 0$ arbitrário, da inclusão anterior e utilizando a monotonicidade de $\partial h(x)$, (veja (A.2)) temos:

$$\begin{aligned}
0 &\leqslant \left\langle \frac{x - \alpha_2\nabla g(x) - J(x,\alpha_2)}{\alpha_2} - \frac{x - \alpha_1\nabla g(x) - J(x,\alpha_1)}{\alpha_1}, J(x,\alpha_2) - J(x,\alpha_1) \right\rangle \\
&= \left\langle \frac{x - J(x,\alpha_2)}{\alpha_2} - (\nabla g(x)) - \frac{x - J(x,\alpha_1)}{\alpha_1} + (\nabla g(x)), J(x,\alpha_2) - J(x,\alpha_1) \right\rangle \\
&= \left\langle \frac{x - J(x,\alpha_2)}{\alpha_2} - \frac{x - J(x,\alpha_1)}{\alpha_1}, (x - J(x,\alpha_1)) - (x - J(x,\alpha_2)) \right\rangle \\
&= -\frac{1}{\alpha_2}\langle x - J(x,\alpha_2), x - J(x,\alpha_2)\rangle + \frac{1}{\alpha_2}\langle x - J(x,\alpha_2), x - J(x,\alpha_1)\rangle \\
&\quad - \frac{1}{\alpha_1}\langle x - J(x,\alpha_1), x - J(x,\alpha_1)\rangle + \frac{1}{\alpha_1}\langle x - J(x,\alpha_1), x - J(x,\alpha_2)\rangle \\
&= -\frac{1}{\alpha_2}\|x - J(x,\alpha_2)\|^2 - \frac{1}{\alpha_1}\|x - J(x,\alpha_1)\|^2 + \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2}\right)\langle x - J(x,\alpha_2), x - J(x,\alpha_1)\rangle.
\end{aligned}$$

Multiplicando por α_2, temos:

$$0 \leqslant -\|x - J(x,\alpha_2)\|^2 - \frac{\alpha_2}{\alpha_1}\|x - J(x,\alpha_1)\|^2 + (\frac{\alpha_2}{\alpha_1} + 1)\langle x - J(x,\alpha_2), x - J(x,\alpha_1)\rangle.$$

Agora, utilizando a desigualdade de Cauchy-Schwarz, temos

$$\begin{aligned}
0 &\leqslant -\|x - J(x,\alpha_2)\|^2 - \frac{\alpha_2}{\alpha_1}\|x - J(x,\alpha_1)\|^2 + (\frac{\alpha_2}{\alpha_1} + 1)\|x - J(x,\alpha_2)\|\|x - J(x,\alpha_1)\| \\
&= \frac{\alpha_2}{\alpha_1}\|x - J(x,\alpha_1)\| \left(\|x - J(x,\alpha_2)\| - \|x - J(x,\alpha_1)\|\right) \\
&\quad - \|x - J(x,\alpha_2)\| \cdot \left(\|x - J(x,\alpha_2)\| - \|x - J(x,\alpha_1)\|\right) \\
&= \left(\|x - J(x,\alpha_2)\| - \|x - J(x,\alpha_1)\|\right) \cdot \left(\frac{\alpha_2}{\alpha_1}\|x - J(x,\alpha_1)\| - \|x - J(x,\alpha_2)\|\right),
\end{aligned}$$

Note que para a desigualdade anterior ser verdadeira, a multiplicação entre ambos os fatores devem ser ou os dois negativos ou os dois positivos. Para o caso positivo, note que as duas desigualdades em (2.9) são imediatas.

Agora, vamos supor por contradição que ambos os termos são negativos. Para facilitar o argumento, vamos denotar $a = \|x - J(x,\alpha_1)\|$ e $b = \|x - J(x,\alpha_2)\|$. Então, obtemos da desigualdade anterior que

$$0 \leqslant (b - a)(ka - b), \tag{2.10}$$

com $k = \frac{\alpha_2}{\alpha_1} \geqslant 1$, pois $\alpha_2 \geqslant \alpha_1 > 0$.

Desde que, $k \geqslant 1$ e $a > 0$ temos que $ka \geqslant a$, ou seja, $0 > ka - b \geqslant a - b$. Sendo assim, temos $a - b < 0$, então $b - a > 0$, o que por sua vez implica em uma contradição, pois a multiplicação de um número negativo $(ka - b)$ com um positivo $(b - a)$ resultaria em um número negativo, o que contrária a nossa suposição de ambos os termos em (2.10) serem negativos. $\square$

A proposição apresentado a seguir, é consequência direta de [[9], Proposição 2.3].

Proposição 2.1. *Sejam* $g, h : \mathbb{R}^n \to \overline{\mathbb{R}}$ *duas funções que satisfazem A1. Suponha que* $\overline{dom(h)} \subset dom(g)$*,* g *é diferenciável em um conjunto aberto contendo* $\overline{dom(h)}$*, e* ∇g *é contínuo em* $dom(h)$*. Então, a hipótese A2 é satisfeita. Consequentemente, se dom(h) for fechado, a validade da hipótese A2 é equivalente à afirmação de que* g *é diferenciável em um conjunto aberto contendo dom(h) e que seu gradiente é contínuo em* $dom(h)$*.*

Nas próximas seções apresentaremos os procedimentos de Busca Linear estudados neste trabalho.

2.1 Procedimento de Busca Linear 1

Nesta seção, iremos dar início à explicação dos procedimentos de Busca Linear, abordando o primeiro dos três procedimentos cruciais para a determinação do comprimento de passo. O primeiro procedimento de Busca Linear foi abordado em [9] para Problema (2.1). Sua definição formal é estabelecida a seguir.

Busca Linear 1 [9]

Dado $x \in \mathrm{dom}(h)$*,* $\sigma > 0$*,* $\theta \in (0, 1)$ *and* $\delta \in (0, 1/2)$*.*

Entrada: *Seja* $\alpha = \sigma$ *e* $J(x, \alpha) := \mathrm{prox}_{\alpha h}(x - \alpha \nabla f(x))$.

Enquanto:

$$\alpha \|\nabla g(J(x, \alpha)) - \nabla g(x)\| > \delta \|J(x, \alpha) - x\|. \tag{2.11}$$

Faça: $\alpha = \theta\alpha$.

Fim enquanto.

Saída α.

A saída α dessa Busca Linear 1 será denotada por $\mathrm{LS1}(x, \sigma, \theta, \delta)$. Primeiramente, é fundamental observar que a Busca Linear 1, conforme abordada, apresenta a particularidade de que, a cada tentativa de um comprimento de passo α para satisfazer a desigualdade da busca, o subproblema proximal (2.3) deve ser computado novamente. Tal procedimento, é chamado na literatura relacionada de **Regra de Busca Linear Implícita.**

O seguinte Lema mostra a boa definição do procedimento de Busca Linear 1, ou seja, mostramos que em um número finito de passos é possível achar α satisfazendo (2.11).

Lema 2.3. *Se $x \in$ dom(h) então a Busca Linear 1 tem terminação finita.*

Demonstração. Se $x \in S^*$, pela Observação 2.2 temos que $x = J(x, \alpha)$, dessa forma, a desigualdade da busca vista em (2.11) é satisfeita trivialmente na primeira tentativa de comprimento de passo $\alpha = \sigma$.

Agora, se $x \notin S^*$ vamos supor por absurdo que a desigualdade da Busca Linear 1 nunca é satisfeita para nenhum $\alpha := \theta^j \sigma$. Ou seja, para $\alpha \in \mathcal{P} := \{\sigma, \sigma\theta, \sigma\theta^2, \cdots\}$ temos que

$$\alpha \|\nabla g(J(x, \alpha)) - \nabla g(x)\| > \delta \|J(x, \alpha) - x\|. \tag{2.12}$$

Podemos observar que para $j \in \mathbb{N}$ suficientemente grande, quando $\alpha := \theta^j \sigma \in \mathcal{P}$ está suficientemente próximo de zero. Por outro lado, temos pelo Lema 2.2 que $J(x, \alpha)$ é uniformemente limitado. Dessa forma, podemos concluir de (2.12) e da hipótese A2, que

$$\lim_{\alpha \to 0^+} \|x - J(x, \alpha)\| = 0.$$

Sendo assim, como $x \to J(x, \alpha)$ pela igualdade anterior e utilizando a hipótese A2 novamente, junto com a continuidade de $\nabla g(\cdot)$, obtemos que

$$\lim_{\alpha \to 0^+} \|\nabla g(J(x, \alpha)) - \nabla g(x)\| = 0.$$

Dessa forma, utilizando o fato de $\alpha > 0$ e (2.12) concluímos que

$$\|\nabla g(J(x, \alpha)) - \nabla g(x)\| > \frac{\delta \|J(x, \alpha) - x\|}{\alpha} \geqslant 0.$$

Logo, pelo teorema do confronto, podemos concluir que

$$\lim_{\alpha \to 0^+} \frac{\|x - J(x, \alpha)\|}{\alpha} = 0. \tag{2.13}$$

Por outro lado, pela condição de otimalidade e utilizando $z = x - \alpha \nabla g(x)$, temos

$$\frac{x - J(x, \alpha)}{\alpha} \in \nabla g(x) + \partial h(J(x, \alpha)),$$

o que por sua vez tomando $\alpha \to 0$, utilizando (2.13) e levando em conta o resultado de gráfico fechado $\mathsf{Gph}(\partial h)$ presente no Fato A.1 temos

$$0 \in \nabla g(x) + \partial h(x) \subseteq \partial((g+h)(x)).$$

Porém, isso gera uma contradição, pois concluímos que $0 \in \partial(g+h)(x)$, ou seja, que x é ponto crítico do Problema 2.1, o que por sua vez pela convexidade da função objetivo implicaria que tal ponto é solução global. O que leva a uma contradição com a hipótese inicial de $x \notin S^*$. □

2.2 Procedimento de Busca Linear 2

Agora, estudaremos o segundo procedimento de Busca Linear, o qual foi analisado também em [9]. Essa busca torna-se mais eficiente, uma vez que a desigualdade associada ao procedimento de busca requer a resolução de $J(x, \alpha)$ apenas uma única vez.

Busca Linear 2 [9]

Dado $x \in$ *dom(h) e* $\theta \in (0, 1)$.

Entrada: Defina $\beta = 1$, $J_x := J(x, 1) = \text{prox}_h(x - \nabla g(x))$.

Enquanto:

$$(g+h)(x - \beta(x - J_x)) > (g+h)(x) - \beta[h(x) - h(J_x)] \tag{2.14}$$
$$-\beta\langle \nabla g(x), x - J_x\rangle + \frac{\beta}{2}\|x - J_x\|^2.$$

Atualiza: $\beta = \theta\beta$.

Fim enquanto.

Saída: β.

A saída β dessa Busca Linear 2 será denotada por LS2(x, θ). Alguns comentários importantes são necessários. Primeiramente, note que ao resolver o subproblema proximal (2.3) toma-se o valor de $\alpha = 1$. Em seguida, note que o procedimento de Busca Linear é feito efetivamente pelo parâmetro $\beta > 0$, o qual é computado dentro do argumento de $(g+h)(\cdot)$ do lado esquerdo em (2.14). Por fim, note que em comparação com a Busca Linear 1, a Busca Linear 2 tem como principal característica e vantagem o fato que exige a resolução do subproblema (2.3) apenas uma vez por iteração, tal procedimento é chamado

Regra de Busca Linear Explícita. Em geral, esse procedimento de busca se torna mais barato.

Lema 2.4. *Se $x \in$ dom(h), então o procedimento de Busca Linear 2 tem terminação finita.*

Demonstração. Inicialmente, segue da Observação 2.2 que, se $x \in S^*$, então $x = J(x, 1)$. Ou seja, a desigualdade em (2.14) é satisfeita trivialmente e retorna o comprimento de passo $\beta = 1$.

Se $x \notin S^*$, vamos supor por contradição que a Busca Linear 2 não termina após um número finito de passos. Sendo assim, para todo $\beta \in \mathcal{Q} := \{1, \theta, \theta^2, \ldots, \theta^j, \cdots\}$, temos

$$(g+h)\,(x-\beta(x-J_x)) > (g+h)\,(x) - \beta\,[h(x)-h(J_x)] - \beta\langle\nabla g(x), x-J_x\rangle + \frac{\beta}{2}\|x-J_x\|^2.$$

Reorganizando, obtemos

$$\frac{(g+h)\,(x-\beta(x-J_x)) - (g+h)\,(x)}{\beta} + [h(x)-h(J_x)] + \langle\nabla g(x), x-J_x\rangle > \frac{1}{2}\|x-J_x\|^2, \tag{2.15}$$

tomando $\beta \to 0$, temos

$$\lim_{\beta\to 0}\frac{(g+h)(x-\beta(x-J_x)) - (g+h)(x)}{\beta}.$$

Note que o limite acima é a derivada direcional de $g + h$ na direção $J_x - x$, isto é, $(g+h)'(x; J_x - x)$. Como g é diferenciável,

$$\lim_{\beta\to 0}\frac{g(x-\beta(x-J_x)) - g(x)}{\beta} = \langle\nabla g(x), J_x - x\rangle,$$

e h é convexa, então

$$\lim_{\beta\to 0}\frac{h(x-\beta(x-J_x)) - h(x)}{\beta} = h'(x; J_x - x),$$

onde $h'(x; J_x - x)$ é a derivada direcional de h em x na direção $J_x - x$. Sendo assim, a partir de (2.15) e das duas últimas igualdades acima, obtemos que

$$\langle\nabla g(x), J_x - x\rangle + h'(x; J_x - x) + [h(x)-h(J_x)] + \langle\nabla g(x), x-J_x\rangle \geqslant \frac{1}{2}\|x-J_x\|^2,$$

o que por sua vez, implica que

$$h'(x; J_x - x) + [h(x)-h(J_x)] \geqslant \frac{1}{2}\|x-J_x\|^2. \tag{2.16}$$

Pela convexidade de h e pela Proposição 1.1 (iii), concluímos que a derivada direcional $h'(x; J_x - x)$ satisfaz,

$$h(J_x) \geqslant h(x) + h'(x; J_x - x).$$

Assim sendo,

$$h(x) + h'(x; J_x - x) - h(J_x) \leqslant 0, \tag{2.17}$$

daí e de (2.16), temos

$$0 \geqslant h'(x; J_x - x) + h(x) - h(J_x) \geqslant \frac{1}{2}\|x - J_x\|^2.$$

Então,

$$0 \geqslant \frac{1}{2}\|x - J_x\|^2,$$

ou seja, $x = J_x$. Porém, se $x = J_x$ e $J_x = \mathtt{prox}_h(x - \nabla f(x))$, obtemos que $x = \mathtt{prox}_h(x - \nabla f(x))$. Logo,

$$x = \arg\min_{y \in \mathbb{R}^n} \left\{ h(y) + \frac{1}{2}\|y - (x - \nabla g(x))\|^2 \right\}.$$

Segue da condição de otimalidade para este problema que para $y = x$, vale

$$0 \in \partial h(x) + (x - (x - \nabla g(x))) = \partial h(x) + \nabla g(x),$$

dessa forma,

$$0 \in \partial h(x) + \nabla g(x) \subseteq \partial(g + h)(x),$$

sendo assim, $x \in S^*$. Absurdo, pois pela nossa hipótese inicial $x \notin S_*$. $\square$

2.3 Procedimento de Busca Linear 3

Agora, apresentaremos a Busca Linear de Beck-Teboulle [10], a qual denotaremos nessa seção e no restante do trabalho como Busca Linear 3. Tal busca, em geral, na literatura relacionada é útil para métodos forward-backward quando a constante de Lipschitz do vetor gradiente $\nabla g(\cdot)$ não é conhecida ou é difícil de estimar.

Busca Linear 3 [10]

Dada $x \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$, com $\sigma > 0$ e $0 < \theta < 1$.

Entrada: Defina $\alpha = \sigma$ e

$$J(x,\alpha) = \text{prox}_{\alpha h}(x - \alpha\nabla g(x)). \tag{2.18}$$

Enquanto:

$$g(J(x,\alpha)) > g(x) + \langle \nabla g(x), J(x,\alpha) - x\rangle + \frac{1}{2\alpha}\|x - J(x,\alpha)\|^2, \tag{2.19}$$

faça $\alpha = \theta\alpha$.

Fim enquanto:

Saída: α.

A saída α dessa Busca Linear 3 será denotada por $\text{LS3}(x,\sigma,\theta)$. Vamos agora mostrar a boa definição e terminação finita dessa Busca Linear sob as hipóteses A1-A4.

Proposição 2.2. *Suponhamos que as hipóteses* $A1 - A4$ *sejam válidas. Então, para qualquer* $x \in$ *int(dom(g))* $\cap$ *dom(h), temos:*

(i) A Busca Linear 3 acima termina após um número finito de iterações com saída positiva $\bar{\alpha} =$ *LS3*(x,σ,θ).

(ii) $\|x - u\|^2 - \|J(x,\bar{\alpha}) - u\|^2 \geqslant 2\bar{\alpha}[F(J(x,\bar{\alpha})) - F(u)]$ *para qualquer* $u \in \mathbb{R}^n$.

(iii) $F(J(x,\bar{\alpha})) - F(x) \leqslant -\frac{1}{2\bar{\alpha}}\|J(x,\bar{\alpha}) - x\|^2 \leqslant 0$. *Consequentemente, pela hipótese* A4, $J(x,\bar{\alpha}) \in$ *int(dom(g))* $\cap$ *dom(h).*

Demonstração. Para provar o item (i), suponha $x \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$. Podemos perceber que $J(x,\alpha)$ é bem definido para qualquer $\alpha > 0$, pois $\nabla g(x)$ existe devido à hipótese A4. Se $x \in S^*$, então segue da Observação 2.2 que $x = J(x,\sigma)$. Assim, de forma análoga às Buscas Lineares anteriores, temos que o procedimento de Busca Linear 3 é satisfeita trivialmente para $\alpha = \sigma$.

Para $x \notin S^*$, vamos supor por contradição que a Busca Linear não termina após um número finito de passos. Sendo assim, para todo $\alpha \in \mathcal{P} := \{\sigma, \sigma\theta, \sigma\theta^2, \ldots\}$ temos que

$$g\big(J(x,\alpha)\big) > g(x) + \langle \nabla g(x),\, J(x,\alpha) - x\rangle + \frac{1}{2\alpha}\|x - J(x,\alpha)\|^2. \tag{2.20}$$

Utilizando a igualdade (2.18), temos

$$J(x, \alpha) - x = \text{prox}_{\alpha h}(x - \alpha \nabla g(x)) - x,$$

Agora, aplicando a norma de ambos os lados e utilizando a definição de $J(x, \alpha)$ dada em (2.3), obtemos que

$$\begin{aligned}\|J(x, \alpha) - x\| &= \|\text{prox}_{\alpha h}(x - \alpha \nabla g(x)) - x\| \\ &= \|\text{prox}_{\alpha h}(x - \alpha \nabla g(x)) - \text{prox}_{\alpha h}(x) + \text{prox}_{\alpha h}(x) - x\|.\end{aligned}$$

Então, segue da desigualdade triangular que

$$\|J(x, \alpha) - x\| \leqslant \| \text{prox}_{\alpha h}(x - \alpha \nabla f(x)) - \text{prox}_{\alpha h}(x)\| + \| \text{prox}_{\alpha h}(x) - x\|.$$

Como $\text{prox}_{\alpha h}$ é não expansivo (veja (A.1)), temos

$$\|J(x, \alpha) - x\| \leqslant \|x - \alpha \nabla g(x) - x\| + \| \text{prox}_{\alpha h}(x) - x\|,$$

o que por sua vez, implica que

$$\|J(x, \alpha) - x\| \leqslant \alpha \|\nabla g(x)\| + \| \text{prox}_{\alpha h}(x) - x\|. \tag{2.21}$$

Pelo Lema 2.1, temos

$$\|J(x, \alpha) - x\| \longrightarrow 0, \quad \text{quando} \alpha \to 0. \tag{2.22}$$

Por outro lado, como g é convexa temos que $g(y) \geqslant g(x) + \langle \nabla g(x), y - x \rangle$. Sendo assim, fazendo $y = x$ e $x = J(x, \alpha)$, temos

$$g(x) - g(J(x, \alpha)) \geqslant \langle \nabla g(J(x, \alpha)), x - J(x, \alpha) \rangle, \quad \forall\ \alpha \in \mathcal{P}.$$

Segue da última desigualdade e de (2.20) que

$$\begin{aligned}\frac{1}{2\alpha}\|x - J(x, \alpha)\|^2 &< g(J(x, \alpha)) - g(x) - \langle \nabla g(x),\ J(x, \alpha) - x \rangle \\ &< \langle \nabla g(J(x, \alpha)),\ J(x, \alpha) - x \rangle - \langle \nabla g(x),\ J(x, \alpha) - x \rangle \\ &= \langle \nabla g(J(x, \alpha)) - \nabla g(x),\ J(x, \alpha) - x \rangle.\end{aligned}$$

Agora, utilizando a desigualdade de Cauchy-Schwarz,

$$\frac{1}{2\alpha}\|x - J(x, \alpha)\|^2 \leqslant \|\nabla g(J(x, \alpha)) - \nabla g(x)\| \cdot \| J(x, \alpha) - x\|.$$

Desde que $J(x, \alpha) \neq x$, pois estamos no caso em que $x \notin S^*$ (veja Observação 2.2), concluímos que

$$0 < \frac{\|J(x, \alpha) - x\|}{\alpha} < 2\,\|\nabla g(J(x, \alpha)) - \nabla g(x)\|\,, \quad \forall\ \alpha \in \mathcal{P}. \tag{2.23}$$

Como $\|x - J(x, \alpha)\| \to 0$, quando $\alpha \to 0$ pela equação (2.22) e $x \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$ podemos concluir que $J(x, \alpha) \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$ para $\alpha > 0$ suficientemente pequeno.

Devido à continuidade de ∇g em $x \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$ pela afirmação A4, temos que

$$\|\nabla g(J(x, \alpha) - \nabla g(x)\| \to 0, \quad \text{quando} \ \ \alpha \to 0.$$

Sendo assim, utilizando o resultado anterior e (2.23). Pelo Teorema do Confronto, obtemos

$$\lim_{\alpha \to 0} \frac{\|x - J(x, \alpha)\|}{\alpha} = 0. \tag{2.24}$$

Utilizando a condição de otimalidade (2.7) e tomando limite quando $\alpha \to 0$, temos do Lema (2.1) que $\|J(x, \alpha) - x\| \longrightarrow 0$ e por (2.24) obtemos que $\frac{\|x - J(x, \alpha)\|}{\alpha} \to 0$. Sendo assim, pela Definição de gráfico fechado do subdiferencial visto em (A.1), temos $-\nabla g(x) \in \partial h(x)$. Dessa forma, $0 \in \nabla g(x) + \partial h(x)$, o que contradiz a nossa hipótese inicial de $x \notin S^*$. Logo, a Busca Linear termina após um número finito de passos, com resultado $\overline{\alpha}$.

Agora, vamos provar o item (ii). Temos que vale a desigualdade da busca, sendo assim

$$g(J(x, \overline{\alpha})) \leqslant g(x) + \langle \nabla g(x), J(x, \overline{\alpha}) - x \rangle + \frac{1}{2\overline{\alpha}}\|x - J(x, \overline{\alpha})\|^2, \tag{2.25}$$

além disso, pela condição de otimalidade (2.7), tem-se

$$\frac{x - J(x, \overline{\alpha})}{\overline{\alpha}} - \nabla g(x) \in \partial h(J(x, \overline{\alpha})).$$

Escolhendo qualquer $u \in \mathbb{R}^n$, obtemos da definição de subgradiente, com $v \in \partial h(x)$ que $h(u) \geqslant h(y) + \langle v, u - y \rangle$. Dado $v = \frac{x - J(x, \overline{\alpha})}{\overline{\alpha}} - \nabla g(x)$ e $y = J(x, \overline{\alpha})$, tem-se:

$$h(u) \geqslant h(J(x, \overline{\alpha})) + \left\langle \frac{x - J(x, \overline{\alpha})}{\overline{\alpha}} - \nabla g(x), u - J(x, \overline{\alpha}) \right\rangle. \tag{2.26}$$

Podemos perceber também pela desigualdade do gradiente (A.1) em g, que $g(u) \geqslant g(x) + \langle \nabla g(x), u - x \rangle$. Daí, somando com (2.26), temos

$$g(u) + h(u) \geqslant h(J(x, \overline{\alpha})) + g(x) + \left\langle \frac{x - J(x, \overline{\alpha})}{\overline{\alpha}} - \nabla g(x), u - J(x, \overline{\alpha}) \right\rangle + \langle \nabla g(x), u - x \rangle,$$

dessa forma,

$$
\begin{aligned}
(g+h)(u) &\geqslant g(x) + h(J(x,\overline{\alpha})) + \frac{1}{\overline{\alpha}}\langle x - J(x,\overline{\alpha}) - \nabla g(x), u - J(x,\overline{\alpha})\rangle + \langle \nabla g(x), u - x\rangle \\
&= g(x) + h(J(x,\overline{\alpha})) + \frac{1}{\overline{\alpha}}\langle x - J(x,\overline{\alpha}), u - J(x,\overline{\alpha})\rangle + \langle -\nabla g(x), u - J(x,\overline{\alpha})\rangle \\
&+ \langle \nabla g(x), u - x\rangle \\
&= g(x) + h(J(x,\overline{\alpha})) + \frac{1}{\overline{\alpha}}\langle x - J(x,\overline{\alpha}), u - J(x,\overline{\alpha})\rangle + \langle \nabla g(x), J(x,\overline{\alpha}) - u\rangle \\
&+ \langle \nabla g(x), u - x\rangle \\
&= g(x) + h(J(x,\overline{\alpha})) + \frac{1}{\overline{\alpha}}\langle x - J(x,\overline{\alpha}), u - J(x,\overline{\alpha})\rangle + \langle \nabla g(x), J(x,\overline{\alpha}) - x\rangle.
\end{aligned}
$$

Utilizando (2.25), temos

$$
F(u) = (g+h)(u) \geqslant g(J(x,\overline{\alpha})) + h(J(x,\overline{\alpha})) + \frac{1}{\overline{\alpha}}\langle x - J(x,\overline{\alpha}), u - J(x,\overline{\alpha})\rangle - \frac{1}{2\overline{\alpha}}\|J(x,\overline{\alpha}) - x\|^2.
$$

Daí, segue que

$$
\overline{\alpha}(g+h)(u) \geqslant \overline{\alpha}(g(J(x,\overline{\alpha}) + h(J(x,\overline{\alpha})) + \langle x - J(x,\overline{\alpha}), u - J(x,\overline{\alpha})\rangle - \frac{1}{2}\|J(x,\overline{\alpha}) - x\|^2,
$$

dessa forma,

$$
\langle x - J(x,\overline{\alpha}), J(x,\overline{\alpha}) - u\rangle \geqslant \overline{\alpha}\left[F(J(x,\overline{\alpha})) - F(u)\right] - \frac{1}{\overline{\alpha}}\|J(x,\overline{\alpha}) - x\|^2.
$$

Agora, uma vez que $\|x-u\|^2 = \langle x-u, x-u\rangle$, obtemos

$$
\begin{aligned}
\|x-u\|^2 &= \langle (x - J(x,\overline{\alpha})) + (J(x,\overline{\alpha}) - u), (x - J(x,\overline{\alpha})) + (J(x,\overline{\alpha}) - u)\rangle \\
&= \|x - J(x,\overline{\alpha})\|^2 + 2\langle x - J(x,\overline{\alpha}), J(x,\overline{\alpha}) - u\rangle + \|J(x,\overline{\alpha}) - u\|^2.
\end{aligned}
$$

Em particular, temos que

$$
2\langle x - J(x,\overline{\alpha}), J(x,\overline{\alpha}) - u\rangle = \|x-u\|^2 - \|x - J(x,\overline{\alpha})\|^2 - \|J(x,\overline{\alpha}) - u\|^2.
$$

Segue da última desigualdade e por (2.20) que

$$
\|x-u\|^2 - \|x - J(x,\overline{\alpha})\|^2 - \|J(x,\overline{\alpha}) - u\|^2 + \|J(x,\overline{\alpha}) - x\|^2 \geqslant 2\overline{\alpha}\left[F(J(x,\overline{\alpha})) - F(u)\right],
$$

o que por sua vez, implica que

$$
\|x-u\|^2 - \|J(x,\bar{z}) - u\|^2 \geqslant 2\alpha\left[F(J(x,\bar{z})) - F(u)\right].
$$

Como queríamos demonstrar em (ii).

Agora para provar o item (iii), seja $u = x$, na desigualdade do item anterior, temos

$$\|x - x\|^2 - \|J(x, \overline{\alpha}) - x\|^2 \geqslant 2\overline{\alpha}\left[F(J(x, \overline{\alpha})) - F(x)\right],$$

o que por sua vez, implica que

$$F(J(x, \overline{\alpha})) - F(x) \leqslant -\frac{1}{2\overline{\alpha}}\|J(x, \overline{\alpha}) - x\|^2 \leqslant 0.$$

Segue que $J(x, \overline{\alpha})$ pertence ao conjunto de nível inferior $L_c F(x)$ e pela Hipótese A4, $J(x, \overline{\alpha}) \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$, como queríamos demonstrar. $\square$

Capítulo 3

Método Proximal Gradiente com Buscas Lineares Implícita e Explícita

Em geral, na literatura relacionada, é comum que as propostas do Método Proximal Gradiente exijam a hipótese de Lipschitz continuidade global do gradiente de g (Ver Definição 1.3). Tal hipótese é essencial por diversos motivos, entre eles, a possibilidade de escolha do comprimento de passo constante e/ou a limitação inferior do comprimento de passo em processos de Buscas Lineares, como os procedimentos estudados no Capítulo anterior.

Neste capítulo será abordado o Método Proximal Gradiente, utilizando os procedimentos de Buscas Lineares vistos no Capítulo 2. Mais especificamente, na Seção 3.1 apresentaremos o método com a Busca Linear 1, que denominaremos **MPG1**. Na Seção 3.2 será tratada a versão acelerada, denotada por **MPG2**. A Seção 3.3 abordará o método com a Busca Linear 2, denotado por **MPG3** e na Seção 3.4 o correspondente com a Busca Linear 3, será denotado, respectivamente por **MPG4**.

3.1 Método Proximal Gradiente com Procedimento de Busca Linear 1 (MPG1)

Consideremos inicialmente o Método Proximal Gradiente com a Busca Linear 1, no qual denotaremos por (**MPG1**). Como dito anteriormente tal método foi proposto e analisado em [9].

Método Proximal Gradiente 1 (**MPG1**) [9]

Passo de inicialização: Tome $x^0 \in \text{dom}(h)$, $\sigma > 0$, $\theta \in (0,1)$ e $\delta \in \left(0, \frac{1}{2}\right)$.

Passo iterativo: Dado x^k, defina

$$x^{k+1} = J(x^k, \alpha_k) := \text{prox}_{\alpha_k h}(x^k - \alpha_k \nabla g(x^k)), \tag{3.1}$$

com $\alpha_k := \text{LS1}(x^k, \sigma, \theta, \delta)$ e $J(\cdot,\cdot)$ definido como em (2.3).

Critério de parada: Se $x^{k+1} = x^k$, pare. Caso contrário, tome $k = k+1$ e retorne ao passo iterativo.

Lembremos que devido ao Lema (2.3), sabemos que a Busca Linear 1, termina com um número finito de passos, sendo assim, visto que o operador proximal está bem definido, temos que a sequência (x^k) está bem definida. A seguinte desigualdade da Busca Linear 1 também será importante para nosso trabalho,

$$\alpha_k \|\nabla g(x^{k+1}) - \nabla g(x^k)\| \leqslant \delta \|x^{k+1} - x^k\|, \tag{3.2}$$

podemos perceber que se o **MPG1** parar na iteração k então, $x^k = \text{prox}_{\alpha_k h}(x^k - \alpha_k \nabla g(x^k))$, ou seja, $x^k \in S^*$. Caso contrário, mostraremos que a sequência (x^k) gerada pelo **MPG1** converge para algum $\bar{x} \in S^*$. Para verificar está afirmação precisamos da seguinte proposição.

Proposição 3.1. *Seja* $\alpha_k =$ *Busca Linear 1* $(x^k, \sigma, \theta, \delta)$. *Para todo* $k \in \mathbb{N}$ *e* $x \in$ *dom(h), temos:*

(i) $\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x)] + (1-2\delta)\|x^{k+1} - x^k\|^2$.

(ii) $(g+h)(x^{k+1}) - (g+h)(x^k) \leqslant -\left(\frac{(1-\delta)}{\alpha_k}\right)\|x^{k+1} - x^k\|^2$.

Demonstração. Vamos provar primeiramente o item (i). Temos da condição de otimalidade do nosso subproblema (2.7), que

$$\frac{x^k - J(x^k, \alpha_k)}{\alpha_k} \in \nabla g(x^k) + \partial h(J(x^k, \alpha_k)),$$

no qual por meio do **MPG1**, temos que $x^{k+1} = J(x^k, \alpha_k)$, onde α_k é o comprimento de passo que satisfaz a Busca Linear 1. Dessa forma, utilizando a inclusão anterior obtemos

$$\frac{x^k - x^{k+1}}{\alpha_k} \in \nabla g(x^k) + \partial h(x^{k+1}). \tag{3.3}$$

Por outro lado, segue da convexidade de h e da desigualdade do subgradiente (A.8) que, dado $w \in \partial h(y)$, temos

$$h(x) \geqslant h(y) + \langle w, x-y\rangle, \quad \forall\, x \in \mathbb{R}^n,\ y \in \operatorname{dom}(h).$$

Daí, tomando $y = x^{k+1}$ juntamente com (3.3), obtemos

$$h(x) \geqslant h(x^{k+1}) + \left\langle \frac{x^k - x^{k+1}}{\alpha_k} - \nabla g(x^k), x - x^{k+1} \right\rangle. \tag{3.4}$$

Agora por outro lado, usando a convexidade de g (veja (A.1)), temos

$$g(x) \geqslant g(y) + \langle \nabla g(y), x-y\rangle, \quad \forall x \in \mathbb{R}^n, y \in \operatorname{dom}(h). \tag{3.5}$$

Somando (3.4) e (3.5) com $y = x^k$ e $x \in \operatorname{dom}(h) \subseteq \operatorname{dom}(g)$, temos:

$$g(x) + h(x) \geqslant h(x^{k+1}) + g(x^k) + \left\langle \frac{x^k - x^{k+1}}{\alpha_k} - \nabla g(x^k), x - x^{k+1} \right\rangle + \langle \nabla g(x^k), x - x^k\rangle,$$

o que por sua vez, implica que

$$\begin{aligned}
(g+h)(x) \geqslant\ & h(x^{k+1}) + g(x^k) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, x - x^{k+1}\rangle - \langle \nabla g(x^k), x - x^{k+1}\rangle \\
& + \langle \nabla g(x^k), x + x^{k+1} - x^{k+1} - x^k\rangle \\
=\ & h(x^{k+1}) + g(x^k) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, x - x^{k+1}\rangle - \langle \nabla g(x^k), x - x^{k+1}\rangle \\
& + \langle \nabla g(x^k), x - x^{k+1}\rangle + \langle \nabla g(x^k), x^{k+1} - x^k\rangle \\
=\ & h(x^{k+1}) + g(x^k) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, x - x^{k+1}\rangle + \langle \nabla g(x^k), x^{k+1} - x^k\rangle \\
=\ & h(x^{k+1}) + g(x^k) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, x - x^{k+1}\rangle \\
& + \langle \nabla g(x^k) + \nabla g(x^{k+1}) - \nabla g(x^{k+1}), x^{k+1} - x^k\rangle.
\end{aligned}$$

Dessa forma,

$$\begin{aligned}
(g+h)(x) \geqslant h(x^{k+1}) + g(x^k) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, x - x^{k+1}\rangle + \langle \nabla g(x^{k+1}), x^{k+1} - x^k\rangle \quad (3.6)\\
+ \langle \nabla g(x^k) - \nabla g(x^{k+1}), x^{k+1} - x^k\rangle.
\end{aligned}$$

Agora, note que segue da desigualdade de Cauchy-Shcwarz, que

$$\|\langle \nabla g(x^k) - \nabla g(x^{k+1}), x^{k+1} - x^k\rangle\| \leqslant \|\nabla g(x^k) - \nabla g(x^{k+1})\| \|x^{k+1} - x^k\|,$$

ou ainda,

$$-\|\langle \nabla g(x^k) - \nabla g(x^{k+1}), x^{k+1} - x^k\rangle\| \geqslant -\|\nabla g(x^k) - \nabla g(x^{k+1})\| \|x^{k+1} - x^k\|.$$

Agora, segue que

$$(g+h)(x) \geqslant h(x^{k+1}) + g(x^k) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, x - x^{k+1}\rangle + \langle \nabla g(x^{k+1}), x^{k+1} - x^k\rangle \tag{3.7}$$
$$-\|\nabla g(x^k) - \nabla g(x^{k+1})\|\|x^{k+1} - x^k\|.$$

Pelo procedimento de Busca Linear 1, temos que

$$\|\nabla g(x^{k+1}) - \nabla g(x^k)\| \leqslant \frac{\delta}{\alpha_k}\|x^{k+1} - x^k\|^2. \tag{3.8}$$

Multiplicando (3.8) por (-1) e utilizando (3.7), temos

$$(g+h)(x) \geqslant g(x^k) + h(x^{k+1}) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, x - x^{k+1}\rangle + \langle \nabla g(x^{k+1}), x^{k+1} - x^k\rangle$$
$$- \frac{\delta}{\alpha_k}\|x^{k+1} - x^k\|^2,$$

reorganizando e usando $\alpha > 0$, temos

$$\langle x^k - x^{k+1}, x^{k+1} - x\rangle \geqslant \alpha_k[g(x^k) + h(x^{k+1}) - (g+h)(x) + \langle \nabla g(x^{k+1}), x^{k+1} - x^k\rangle] \tag{3.9}$$
$$-\delta\|x^{k+1} - x^k\|^2.$$

Por outro lado, sabemos que

$$\begin{aligned} 2\langle x^k - x^{k+1}, x^{k+1} - x\rangle &= \|x^k - x^{k+1} + x^{k+1} - x\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - x\|^2 \\ &= \|x^k - x\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - x\|^2, \end{aligned}$$

usando (3.9) e a igualdade acima, temos

$$\begin{aligned} \|x^k - x\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - x\|^2 \geqslant{} & 2\alpha_k[g(x^k) + h(x^{k+1}) - (g+h)(x) \\ & + \langle \nabla g(x^{k+1}), x^{k+1} - x^k\rangle] - 2\delta\|x^{k+1} - x^k\|^2, \end{aligned}$$

o que por sua vez, implica que

$$\begin{aligned} \|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant{} & 2\alpha_k[g(x^k) + h(x^{k+1}) - (g+h)(x) + \langle \nabla g(x^{k+1}), x^{k+1} - x^k\rangle] \\ & + (1 - 2\delta)\|x^{k+1} - x^k\|^2. \end{aligned}$$

Desde que g é convexa, segue da desigualdade do gradiente (veja (A.5)), com $y = x^k$ e $x = x^{k+1}$ que

$$g(x^k) - g(x^{k+1}) \geqslant \langle \nabla g(x^{k+1}), x^k - x^{k+1}\rangle,$$

o que por sua vez, implica que

$$-2\alpha_k(g(x^k) - g(x^{k+1})) \leqslant 2\alpha_k\langle \nabla g(x^{k+1}), x^{k+1} - x^k\rangle.$$

Portanto,

$$\begin{aligned}\|x^k - x\|^2 - \|x^{k+1} - x\|^2 &\geqslant 2\alpha_k[g(x^k) + h(x^{k+1}) - (g+h)(x)] + 2\alpha_k(g(x^{k+1}) - g(x^k)) \\ &\quad + (1-2\delta)\|x^{k+1} - x^k\|^2 \\ &= 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x)] + (1-2\delta)\|x^{k+1} - x^k\|^2,\end{aligned}$$

sendo assim,

$$\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x)] + (1-2\delta)\|x^{k+1} - x^k\|^2,$$

como queríamos demonstrar.

Para verificar o item (ii), vamos substituir no item anterior $x = x^k$, sendo assim

$$\|x^k - x^k\|^2 - \|x^k - x^{k+1}\|^2 \geqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x^k)] + (1-2\delta)\|x^{k+1} - x^k\|^2,$$

o que por sua vez, implica que

$$\begin{aligned}2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x^k)] &\leqslant -\|x^k - x^{k+1}\|^2 - (1-2\delta)\|x^{k+1} - x^k\|^2 \\ &= -(1+1-2\delta)\|x^{k+1} - x^k\|^2 \\ &= -(2-2\delta)\|x^{k+1} - x^k\|^2.\end{aligned}$$

Reorganizando os termos, obtemos que

$$(g+h)(x^{k+1}) - (g+h)(x^k) \leqslant -\frac{(1-\delta)}{\alpha_k}\|x^{k+1} - x^k\|^2. \tag{3.10}$$

□

Note que segue da Proposição 3.1 item (i) com $x = \overline{x} \in S^*$ que a sequência (x^k) é Fejér convergente em relação ao conjunto solução S^*. De fato, segue da Proposição 3.1 item (i) que

$$\|x^k - \overline{x}\|^2 - \|x^{k+1} - \overline{x}\|^2 \geqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(\overline{x})] + (1-2\delta)\|x^{k+1} - x^k\|^2. \tag{3.11}$$

Sabemos que $[(g+h)(x^{k+1}) - (g+h)(\overline{x})] \geqslant 0$, já que $\overline{x} \in S^*$ e $(1-2\delta)\|x^{k+1} - x^k\|^2 \geqslant 0$, pois $\delta \in (0, \frac{1}{2})$. Daí,

$$\|x^{k+1} - \overline{x}\|^2 \leqslant \|x^k - \overline{x}\|^2.$$

Logo, a sequência gerada pelo **MPG1** é Fejér convergente no conjunto S^*.

A Proposição 3.1 (ii) mostra que o **MPG1** é um método de descida, visto que o valor da função de custo $g+h$ em cada iteração está diminuindo. Além disso pela Proposição 3.1 (i) a sequência gerada pelo **MPG1** é Fejér convergente para o conjunto de solução ótima S^* sempre que $S^* \neq \emptyset$. Esta observação é de fato o centro do seguinte resultado principal desta seção, onde provamos a convergência da sequência (x^k) gerada pelo **MPG1** e também que $((g+h)(x^k))$ é uma sequência minimizante de $g+h$ sem a suposição de Lipschitz em $\nabla g(\cdot)$. Até onde sabemos, este resultado melhora [[7], Teorema 1.2] e até os resultados clássicos para o método do gradiente com Buscas Lineares; veja, por exemplo, [14], Proposição 1.3.3, [1]]. Além disso, mostraremos que a sequência $((g+h)(x^k))$ converge para o valor ínfimo quando o conjunto de soluções está vazio.

Teorema 3.1. *Sejam (x^k) e (α_k) as sequências geradas pelo **MPG1**. As seguintes afirmações são verdadeiras:*

(i) Se $S^* \neq \emptyset$, então (x^k) converge para um ponto em $S^* \neq \emptyset$. Além disso,

$$\lim_{k\to+\infty}(g+h)(x^k) = \min_{x\in\mathbb{R}}(g+h)(x).$$

(ii) Se $S^* = \emptyset$ então temos

$$\lim_{k\to+\infty}\|x^k\| = +\infty \quad e \quad \lim_{k\to+\infty}(g+h)(x^k) = \inf_{x\in\mathbb{R}^n}(g+h)(x).$$

Demonstração. Vamos provar primeiro o item (i). Dado $S^* \neq \emptyset$. Pela Proposição 3.1 (i), para qualquer $\bar{x} \in S^*$, temos

$$\|x^k-\bar{x}\|^2-\|x^{k+1}-\bar{x}\|^2 \geqslant 2\alpha_k[(g+h)(x^{k+1})-(g+h)(\bar{x})]+(1-2\delta)\|x^{k+1}-x^k\|^2, \quad (3.12)$$

o que por sua vez, implica que

$$\|x^k-\bar{x}\|^2-\|x^{k+1}-\bar{x}\|^2 \geqslant (1-2\delta)\|x^{k+1}-x^k\|^2 \geqslant 0. \quad (3.13)$$

Dessa forma, podemos concluir pela Definição (1.4) com $\varepsilon_k \equiv 0$ que a sequência (x^k) é Fejér convergente. Pelo Lema 1.1 (i) garantimos também que (x^k) é limitada. Pela desigualdade (3.12), temos

$$\begin{aligned}0 \leqslant 2\alpha_k[(g+h)(x^{k+1})-(g+h)(\bar{x})] &\leqslant \|x^k-\bar{x}\|^2-\|x^{k+1}-\bar{x}\|^2\\ &= (\|x^k-\bar{x}\|+\|x^{k+1}-\bar{x}\|)(\|x^k-\bar{x}\|-\|x^{k+1}-\bar{x}\|),\end{aligned}$$

como (x^k) é limitada, existe $M > 0$ tal que $\|x^k - \bar{x}\| < M,\ \forall k \in \mathbb{N}$. Sendo assim

$$0 \leqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(\bar{x})] \leqslant 2M(\|x^k - \bar{x}\| - \|x^{k+1} - \bar{x}\|), \tag{3.14}$$

por outro lado, utilizando a desigualdade triangular, temos:

$$\|x^k - \bar{x}\| = \|x^k - x^{k+1} + x^{k+1} - \bar{x}\| \leqslant \|x^k - x^{k+1}\| + \|x^{k+1} - \bar{x}\|,$$

o que por sua vez, implica que

$$\|x^k - x^{k+1}\| \geqslant \|x^k - \bar{x}\| - \|x^{k+1} - \bar{x}\|,$$

daí, substituindo em (3.14), temos

$$0 \leqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(\bar{x})] \leqslant 2M\|x^k - x^{k+1}\|,$$

onde $M = \left\{ \sup \|x^k - \bar{x}\| | k \in \mathbb{N} < +\infty \right\}$. Sendo assim,

$$(g+h)(x^{k+1}) - (g+h)(\bar{x}) \leqslant \frac{M}{\alpha_k}\|x^k - x^{k+1}\|. \tag{3.15}$$

Como a sequência (x^k) é Fejér Convergente em relação a S^*, então $(\|x^k - \bar{x}\|)$ é convergente. Sendo assim, pelo Teorema do Confronto e a desigualdade (3.13), temos que $\|x^{k+1} - x^k\| \to 0$. Pois, tomando $k \to +\infty$, obtemos que $\|x^k - \bar{x}\|^2 \to a$ e $\|x^{k+1} - \bar{x}\|^2 \to a$, logo

$$a - a \geqslant \lim_{k \to +\infty} (1 - 2\delta)\|x^{k+1} - x^k\| \geqslant 0.$$

Portanto, $\lim\limits_{k \to +\infty} \|x^{k+1} - x^k\| = 0$.

Como (x^k) é limitada, existe uma subsequência (x^{n_k}) convergente. Vejamos os dois casos possíveis.

Caso 1: Vamos supor que a sequência de comprimento de passos (α_k) definida pelo **MPG1** não converge para zero. Dessa forma, existe uma sequência (α_{n_k}) e $\beta > 0$ tal que,

$$\alpha_{n_k} \geqslant \beta. \tag{3.16}$$

Como (x^k) é limitada e $\|x^{k+1} - x^k\| \to 0$, utilizando a nossa afirmação A2 e (3.2), temos

$$\alpha_{n_k}\|\nabla g(x^{n_k+1}) - \nabla g(x^{n_k})\| \leqslant \delta\|x^{n_k+1} - x^{n_k}\|,$$

por (3.16), temos

$$\beta\|\nabla g(x^{n_k+1}) - \nabla g(x^{n_k})\| \leqslant \alpha_{n_k}\|\nabla g(x^{n_k+1}) - \nabla g(x^{n_k})\| \leqslant \delta\|x^{n_k+1} - x^{n_k}\|,$$

o que por sua vez nos permite concluir que,

$$\beta\|\nabla g(x^{n_k+1}) - \nabla g(x^{n_k})\| \leqslant \delta\|x^{n_k+1} - x^{n_k}\|,$$

passando limite com $k \to +\infty$, temos pelo teorema do confronto que

$$\lim_{k\to+\infty} \|\nabla g(x^{n_k+1}) - \nabla g(x^{n_k})\| = 0.$$

Segue da condição de otimalidade vista em (2.5), substituindo $z = x^{n_k} - \alpha_{n_k}\nabla g(x^{n_k})$ e levando em conta o **MPG1** onde $x^{n_k+1} = J(x^{n_k}, \alpha^{n_k})$, temos

$$\frac{x^{n_k} - \alpha_{n_k}\nabla g(x^{n_k}) - x^{n_k+1}}{\alpha_{n_k}} \in \partial h(x^{n_k+1}),$$

reorganizando,

$$\frac{x^{n_k} - x^{n_k+1}}{\alpha_{n_k}} - \nabla g(x^{n_k}) \in \partial h(x^{n_k+1}),$$

o que por sua vez, implica que

$$\frac{x^{n_k} - x^{n_k+1}}{\alpha_{n_k}} - \nabla g(x^{n_k}) + \nabla g(x^{n_k+1}) \in \nabla g(x^{n_k+1}) + \partial h(x^{n_k+1}) \subseteq \partial(g+h)(x^{n_k+1}). \quad (3.17)$$

A sequência (x^{n_k+1}) converge para $x^* \in S^*$ e como $\|x^{n_k} - x^{n_k+1}\| \to 0$ quando $k \to +\infty$ na desigualdade anterior, utilizando a nossa afirmação A2 de continuidade para o ∇g temos,

$$\frac{x^{n_k} - x^{n_k+1}}{\alpha_{n_k}} - \nabla g(x^{n_k}) + \nabla g(x^{n_k+1}) \in \partial(g+h)(x^{n_k+1}).$$

Agora utilizando o Teorema do Gráfico fechado (veja Fato (A.1)), podemos concluir que $0 \in \partial(g+h)(x^*)$, logo $x^* \in S^*$. Para completar a prova do item (i) temos que $((g+h)(x^k))$ é monótona não-crescente pela Proposição 3.1 (ii), sendo assim, de (3.15) e (3.16) temos

$$(g+h)(x^{k+1}) - (g+h)(\bar{x}) \leqslant \frac{M}{\alpha_k}\|x^k - x^{k+1}\| \leqslant \frac{M}{\beta}\|x^k - x^{k+1}\|,$$

passando limite quando $k \to +\infty$ e lembrando que $\|x^k - x^{k+1}\| \to 0$, temos

$$\lim_{k\to+\infty} (g+h)(x^{k+1}) - (g+h)(\bar{x}) \leqslant 0,$$

como $\bar{x} \in S^*$, obtemos

$$\lim_{k\to+\infty} (g+h)(x^{k+1}) \leqslant \min_{x\in\mathbb{R}^n} (g+h)(x).$$

Por outro lado, sabemos que

$$\min_{x\in\mathbb{R}^n} (g+h)(x) \leqslant \lim_{k\to+\infty} (g+h)(x^{k+1}),$$

logo,

$$\lim_{k\to+\infty}(g+h)(x^k) = \min_{x\in\mathbb{R}^n}(g+h)(x),$$

como queríamos demonstrar.

Caso 2: Agora vamos supor que $\lim\limits_{k\to+\infty}\alpha_{n_k} = 0$. De forma que, $\widehat{\alpha}_{n_k} := \frac{\alpha_{n_k}}{\theta} > \alpha_{n_k} > 0$ e $\widehat{x}^{n_k} := J(x^{n_k}, \widehat{\alpha}_{n_k})$. Ou seja, $\widehat{\alpha}_{n_k}$ não passa a desigualdade da busca, pois é maior que α_{n_k}. Deste fato e do Lema (2.2), temos

$$\|x^{n_k} - J(x^{n_k}, \widehat{\alpha}_{n_k})\| \leqslant \frac{\widehat{\alpha}_{n_k}}{\alpha_{n_k}}\|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\|,$$

sendo assim, temos

$$\|x^{n_k} - \widehat{x}^{n_k}\| \leqslant \frac{\alpha_{n_k}}{\alpha_{n_k}\theta}\|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| = \frac{1}{\theta}\|x^{n_k} - x^{n_k+1}\|.$$

Como (x^{n_k}) é limitada então a subsequência $(\widehat{x}^{n_k})$ também é limitada. Temos da Busca Linear 1 que,

$$\widehat{\alpha}_{n_k}\|\nabla g(\widehat{x}^{n_k}) - \nabla g(x^{n_k})\| > \delta\|\widehat{x}^{n_k} - x^{n_k}\|, \tag{3.18}$$

como $\widehat{\alpha}_{n_k} \to 0$ e as sequências (x^{n_k}) e $(\widehat{x}^{n_k})$ são limitadas, por a desigualdade anterior e assumindo a nossa Hipótese A2, temos que $\lim\limits_{k\to+\infty}\|\widehat{x}^{n_k} - x^{n_k}\| = 0$. E que $(\widehat{x}^{n_k})$ converge para $x^* \in S^*$. Sendo assim, pela nossa hipótese A2 novamente e (3.18), temos

$$\lim_{k\to+\infty}\|\nabla g(\widehat{x}^{n_k}) - \nabla g(x^{n_k})\| = 0. \tag{3.19}$$

De (3.18), temos

$$\|\nabla g(\widehat{x}^{n_k}) - \nabla g(x^{n_k})\| > \frac{\delta}{\widehat{\alpha}^{n_k}}\|\widehat{x}^{n_k} - x^{n_k}\|,$$

passando limite quando $k \to +\infty$, uma vez que δ é uma constante, pelo Teorema do Confronto, temos

$$\lim_{k\to+\infty}\frac{\|\widehat{x}^{n_k} - x^{n_k}\|}{\widehat{\alpha}^{n_k}} = 0. \tag{3.20}$$

Usando a condição de otimalidade vista em (2.5) com $z = x^{n_k} - \widehat{\alpha}_{n_k}\nabla f(x^{n_k})$, temos

$$\frac{x^{n_k} - \widehat{\alpha}_{n_k}\nabla g(x^{n_k}) - \widehat{x}^{n_k}}{\widehat{\alpha}_{n_k}} = \frac{x^{n_k} - \widehat{x}^{n_k}}{\widehat{\alpha}_{n_k}} - \nabla f(x^{n_k}) \in \partial h(\widehat{x}^{n_k}).$$

Sendo assim, acrescentando $\nabla g(\widehat{x}^{n_k})$, temos

$$\frac{x^{n_k} - \widehat{x}^{n_k}}{\widehat{\alpha}_{n_k}} - \nabla g(x^{n_k}) + \nabla g(\widehat{x}^{n_k}) \in \partial h(\widehat{x}^{n_k}) + \nabla g(\widehat{x}^{n_k}) \subseteq \partial(g+h)(\widehat{x}^{n_k}),$$

quando $k \to +\infty$, obtemos de (3.20) que $\frac{x^{n_k} - \widehat{x}^{n_k}}{\widehat{\alpha}_{n_k}} \to 0$, temos também de (3.19) que $\nabla g(x^{n_k}) \to \nabla g(\widehat{x}^{n_k})$. Dessa forma, utilizando Fato A.1 e lembrando que $\widehat{x}^{n_k} \to x^*$, obtemos que $0 \in \partial(g+h)(x^*)$, logo $x^* \in S^*$. Agora precisamos verificar a seguinte igualdade $\lim\limits_{k\to+\infty}(g+h)(x^k) = \min\limits_{x\in\mathbb{R}^n}(g+h)(x)$. Utilizando o Lema 2.2, sabendo que $\widehat{\alpha}_{n_k} := \frac{\alpha_{n_k}}{\theta} > \alpha_{n_k} > 0$ e $\widehat{x}^{n_k} := J(x^{n_k}, \widehat{\alpha}_{n_k})$, temos

$$\begin{aligned}\|x^{n_k} - \widehat{x}^{n_k}\| &= \|x^{n_k} - J(x^{n_k}, \widehat{\alpha}_{n_k})\| = \|x^{n_k} - J(x^{n_k}, \frac{\alpha_{n_k}}{\theta})\| \\ &\geqslant \|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| = \|x^{n_k} - x^{n_k+1}\|.\end{aligned}$$

Sendo assim,

$$\|x^{n_k} - \widehat{x}^{n_k}\| \geqslant \|x^{n_k} - x^{n_k+1}\|. \tag{3.21}$$

Dividindo a desigualdade anterior por α_{n_k}, passando limite quando $k \to +\infty$ e utilizando (3.20), temos que $\lim\limits_{k\to+\infty} \dfrac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} = 0$. Pela Proposição 3.1 (ii) temos que a sequência $((g+h)(x^k))$ é decrescente, desde fato e de (3.15), obtemos

$$\begin{aligned}0 = \lim_{k\to+\infty} M\frac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} &\geqslant \lim_{k\to+\infty}(g+h)(x^{n_k}) - \lim_{k\to+\infty}(g+h)(x^*) \\ &= \lim_{k\to+\infty}(g+h)(x^k) - \lim_{k\to+\infty}(g+h)(x^*) \geqslant 0,\end{aligned}$$

pois $x^* \in S^*$, logo

$$0 \geqslant \lim_{k\to+\infty}(g+h)(x^k) - \lim_{k\to+\infty}(g+h)(x^*) \geqslant 0,$$

dessa forma,

$$\lim_{k\to+\infty}(g+h)(x^k) - \lim_{k\to+\infty}(g+h)(x^*) = 0.$$

Portanto,

$$\lim_{k\to+\infty}(g+h)(x^k) = \min_{x\in\mathbb{R}}(g+h)(x).$$

Agora iremos demonstrar o item (ii). Como vimos anteriormente na prova do item (i), concluímos que se $S^* \neq \emptyset$, então a sequência (x^k) é limitada. Além disso, todos os seus pontos de acumulação são soluções. Agora, suponha que $S^* = \emptyset$, dessa forma qualquer subsequência de (x^k) é ilimitada, pois caso contrário, teria ponto de acumulação e tal ponto seria solução. Logo, $\|x^k\| \to +\infty$ quando $k \to +\infty$. Além disso, segue da definição de ínfimo que,

$$s := \lim_{k\to+\infty}(g+h)(x^k) \geqslant \inf_{x\in\mathbb{R}^n}(g+h)(x),$$

de forma que s existe pois $((g+h)(x^k))$ é monótona não-crescente pela Proposição 3.1 (ii). Vamos supor que $s > \inf\limits_{x\in\mathbb{R}^n}(g+h)(x)$. Dessa forma, podemos concluir que o seguinte conjunto de nível

$$S_{lev}(x^0) := \{x \in \mathrm{dom}(g) : (g+h)(x) \leqslant (g+h)(x^k), \quad \forall k \in \mathbb{N}\},$$

não é vazio. Agora, aplicando a Proposição 3.1 (i) com $x \in S_{lev}(x^0)$,

$$\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x)] + (1-2\delta)\|x^{k+1} - x^k\|^2 \geqslant 0.$$

onde, na última desigualdade usamos que temos $(g+h)(x^{k+1})-(g+h)(x) \geqslant 0$ e $\delta \in (0, \frac{1}{2})$. Portanto, concluímos que a sequência (x^k) é Fejér convergente para $S_{lev}(x^0)$. Sendo assim, pelo Lema 1.1 (i) temos que a sequência (x^k) é limitada. O que por sua vez é uma contradição, pois supomos que a sequência era ilimitada. Logo, $s = \inf_{x \in \mathbb{R}^n}(g+h)(x)$, ou seja

$$\lim_{k \to +\infty}(g+h)(x^k) = \inf_{x \in \mathbb{R}^n}(g+h)(x).$$

□

3.2 Método Proximal Gradiente Acelerado com Procedimento de Busca Linear 1 (MPG2)

Nesta seção, será abordada uma versão alternativa do **MPG1** visto na seção anterior. Mais especificamente, será analisada uma versão acelerada do método estudado, novamente independentemente da hipótese de continuidade Lipschitz do gradiente de g. Por fim, destaca-se que os resultados de complexidade de pior-caso relativos aos métodos serão apresentados no Capítulo 4. O Método Proximal Gradiente 2 com Busca Linear 1 será denotado por **MPG2** e é descrito como segue.

Método Proximal Gradiente 2 (**MPG2**) [9]

Passo de inicialização: Considere $\Omega = \mathtt{dom}(h)$, $x^{-1} = x^0 \in \text{dom}(h)$, $t_0 = 1$, $\theta \in (0,1)$, $\alpha_1 = \sigma$ e $\delta \in (0, \frac{1}{2})$.

Passo iterativo: Dado t_k e x^k, defina

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}, \tag{3.22}$$

$$y^k = x^k + \left(\frac{t_k - 1}{t_{k+1}}\right)\left(x^k - x^{k-1}\right), \quad \widetilde{y}^k = P_\Omega(y^k), \tag{3.23}$$

$$x^{k+1} = J\left(\widetilde{y}^k, \alpha_k\right) := \text{prox}_{\alpha_k g}(\widetilde{y}^k - \alpha_k \nabla f(\widetilde{y}^k)), \tag{3.24}$$

com $\alpha_k := \text{LS1}(\widetilde{y}^k, \alpha_{k-1}, \theta, \delta)$.

Critério de parada: Se $x^{k+1} = \widetilde{y}^k$, pare. Caso contrário, faça $k = k+1$ e retorne ao passo iterativo.

Podemos observar de (3.23) e (3.24) que $\widetilde{y}^k$ e $x^k \in \mathtt{dom}(h)$. Como consequência direta do Lema (2.3), α_k que satisfaz a condição de Busca Linear 1 é sempre positivo e não-crescente, visto que $\alpha_k : = \text{LS1}(\widetilde{y}^k, \alpha_{k-1}, \theta, \delta)$. Além disso, temos que se $x^{k+1} = \widetilde{y}^k$, então x^{k+1} é solução ótima. Um resultado da Busca Linear 1 que será importante para nosso trabalho será:

$$\alpha_k \|\nabla g(x_{k+1}) - \nabla g(\widetilde{y}^k)\| \leqslant \delta \|x^{k+1} - \widetilde{y}^k\|, \tag{3.25}$$

com $\delta \in \left(0, \frac{1}{2}\right)$.

Proposição 3.2. *Seja* α_k *definido no* ***MPG2*** *e* $x \in \text{dom}(h)$. *Então,*

$$(g+h)(x) - (g+h)(x^{k+1}) \geqslant \frac{1}{2\alpha_k}\left(\|x^{k+1} - x\|^2 - \|y^k - x\|^2\right), \quad \forall k \in \mathbb{N}. \tag{3.26}$$

Demonstração. Dada a condição de otimalidade vista em (2.5), vamos substituir $z = \widetilde{y}^k - \alpha_k \nabla g(\widetilde{y}^k)$. Sendo assim, temos

$$\frac{\widetilde{y}^k - \alpha_k \nabla g(\widetilde{y}^k) - \text{prox}_{\alpha h}\left(\widetilde{y}^k - \alpha_k \nabla g(\widetilde{y}^k)\right)}{\alpha_k} \in \partial h(\text{prox}_{\alpha h}(\widetilde{y}^k - \alpha_k \nabla g(\widetilde{y}^k))).$$

Então, utilizando informações do **MPG2**, obtemos

$$\frac{\widetilde{y}^k - x^{k+1}}{\alpha_k} - \nabla g(\widetilde{y}^k) \in \partial h(x^{k+1}). \tag{3.27}$$

Pela definição de subgradiente dada em (A.8), temos

$$h(y) \geqslant h(x^{k+1}) + \langle v, y - x^{k+1}\rangle, \quad \forall y \in \mathbb{R}^n,\ v \in \partial h(x^{k+1}).$$

Tomando $y = x$ e utilizando (3.27), obtemos

$$h(x) - h(x^{k+1}) \geqslant \left\langle \frac{\tilde{y}^k - x^{k+1}}{\alpha_k} - \nabla g(\tilde{y}^k),\ x - x^{k+1} \right\rangle, \quad \forall x \in \mathrm{dom}(h)\,. \tag{3.28}$$

Agora, usando a convexidade de g, pela Desigualdade do gradiente (A.5), temos

$$g(x) - g(y) \geqslant \langle \nabla g(y), x - y\rangle, \quad \forall x \in \mathbb{R}^n,\ y \in \mathrm{dom}(h)\,.$$

Tomando $y = \tilde{y}^k$, temos

$$g(x) - g(\tilde{y}^k) \geqslant \langle \nabla g(\tilde{y}^k), x - \tilde{y}^k\rangle. \tag{3.29}$$

Somando (3.28) e (3.29), com algumas manipulações algébricas, temos

$$\begin{aligned}(g+h)(x) &\geqslant g(\tilde{y}^k) + h(x^{k+1}) + \left\langle \frac{\tilde{y}^k - x^{k+1}}{\alpha_k} - \nabla g(\tilde{y}^k), x - x^{k+1} \right\rangle + \langle \nabla g(\tilde{y}^k), x - \tilde{y}^k\rangle \\ &= g(\tilde{y}^k) + h(x^{k+1}) + \frac{1}{\alpha_k}\langle \tilde{y}^k - x^{k+1}, x - x^{k+1}\rangle + \langle \nabla g(\tilde{y}^k), x^{k+1} - \tilde{y}^k\rangle \\ &= g(\tilde{y}^k) + h(x^{k+1}) + \frac{1}{\alpha_k}\langle \tilde{y}^k - x^{k+1}, x - x^{k+1}\rangle \\ &\quad + \langle \nabla g(\tilde{y}^k) - \nabla g(x^{k+1}), x^{k+1} - \tilde{y}^k\rangle + \langle \nabla g(x^{k+1}), x^{k+1} - \tilde{y}^k\rangle.\end{aligned}$$

Em particular, utilizando a desigualdade de Cauchy–Schwarz e a Busca Linear 1 (2.11), temos

$$\begin{aligned}-\|\langle \nabla g(\tilde{y}^k) - \nabla g(x^{k+1}), x^{k+1} - \tilde{y}^k\rangle\| &\geqslant - \|\nabla g(\tilde{y}^k) - \nabla g(x^{k+1})\| \|x^{k+1} - \tilde{y}^k\| \\ &\geqslant - \frac{\delta}{\alpha_k}\|x^{k+1} - \tilde{y}^k\|^2.\end{aligned}$$

Sendo assim, temos

$$\begin{aligned}(g+h)(x) &\geqslant g(\tilde{y}^k) + h(x^{k+1}) + \frac{1}{\alpha_k}\langle \tilde{y}^k - x^{k+1}, x - x^{k+1}\rangle - \frac{\delta}{\alpha_k}\|x^{k+1} - \tilde{y}^k\|^2 \\ &\quad + \langle \nabla g(x^{k+1}), x^{k+1} - \tilde{y}^k\rangle,\end{aligned}$$

reorganizando e multiplicando por α_k, temos

$$\begin{aligned}\langle \tilde{y}^k - x^{k+1}, x^{k+1} - x\rangle &\geqslant \alpha_k[g(\tilde{y}^k) + h(x^{k+1}) - (g+h)(x)] - \delta\|x^{k+1} - \tilde{y}^k\|^2 \\ &\quad + \alpha_k\langle \nabla g(x^{k+1}), x^{k+1} - \tilde{y}^k\rangle.\end{aligned} \tag{3.30}$$

Podemos observar que

$$2\langle \tilde{y}^k - x^{k+1}, x^{k+1} - x\rangle = \|\tilde{y}^k - x\|^2 - \|x^{k+1} - x\|^2 - \|\tilde{y}^k - x^{k+1}\|^2,$$

combinando a igualdade anterior e (3.30), temos

$$\|\widetilde{y}^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[g(\widetilde{y}^k) + h(x^{k+1}) - (g+h)(x)] + (1-2\delta)\|x^{k+1} - \widetilde{y}^k\|^2$$
$$+ \alpha_k\langle \nabla g(x^{k+1}), x^{k+1} - \widetilde{y}^k\rangle.$$

Agora, desde que $\delta \in \left(0, \dfrac{1}{2}\right)$, temos que $1 - 2\delta > 0$, logo obtemos

$$\|\widetilde{y}^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[g(\widetilde{y}^k) + h(x^{k+1}) - (g+h)(x)] + \alpha_k\langle \nabla g(x^{k+1}), x^{k+1} - \widetilde{y}^k\rangle. \tag{3.31}$$

Daí, segue da convexidade de g e do Teorema A.5 com $y = \widetilde{y}^k$, e $x = x^{k+1}$, que

$$g(\widetilde{y}^k) - g(x^{k+1}) \geqslant \langle \nabla g(x^{k+1}), \widetilde{y}^k - x^{k+1}\rangle,$$

reorganizando,

$$g(x^{k+1}) - g(\widetilde{y}^k) \leqslant \langle \nabla g(x^{k+1}), x^{k+1} - \widetilde{y}^k\rangle.$$

Tendo em vista a desigualdade anterior e (3.31), obtemos

$$\|\widetilde{y}^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[g(\widetilde{y}^k) + h(x^{k+1}) - (g+h)(x) + g(x^{k+1}) - g(\widetilde{y}^k)]$$
$$= 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x)].$$

Como por (3.23) $\widetilde{y}^k = P_\Omega(y^k)$ então, pela definição de projeção (A.3), podemos concluír que $\|\widetilde{y}^k - x\| \leqslant \|y^k - x\|$ para todo $x \in \mathrm{dom}(h)$. Assim, da desigualdade anterior, temos

$$\|y^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[(g+h)(x^{k+1}) - (g+h)(x)],$$

multiplicando por (-1), temos

$$(g+h)(x) - (g+h)(x^{k+1}) \geqslant \frac{1}{2\alpha_k}(\|x^{k+1} - x\|^2 - \|y^k - x\|^2),$$

como queríamos demonstrar. □

3.3 Método Proximal Gradiente com Busca Linear 2 (MPG3).

Vimos anteriormente que o **MPG1**, o qual utiliza o procedimento de Busca Linear 1, durante o processo de atualização da iteração $x^{k+1} = J(x^k, \alpha_k)$ precisava, em cada iteração, computar o operador proximal dado em (2.3) a cada tentativa de comprimento de passo α_k para satisfazer a Busca Linear 1. Por este motivo, o **MPG1** com tal procedimento de Busca Linear pode muitas vezes ser ineficiente, caso o proximal não seja fácil de calcular. Para tentar melhorar essa limitação, vamos estudar uma modificação do método por meio da Busca Linear 2. Como observado anteriormente o procedimento de Busca Linear 2 analisado na Seção 2.2 tem a vantagem de resolver apenas uma vez o operador proximal, a fim de computar $J(x^k, 1)$ dado em (2.3) para todas tentativas de satisfazer a Busca Linear. Assim, como o **MPG1**, iremos abordar alguns resultados auxiliares, bem como a análise de convergência do método.

Método Proximal Gradiente 3 (MPG3) [9]

Passo de inicialização: Tome $x^0 \in$ dom(h) e $\theta \in (0, 1)$.

Passo iterativo: Dado x^k, defina

$$J_k = \mathsf{Prox}_h(x^k - \nabla g(x^k)), \tag{3.32}$$

$$x^{k+1} = x^k - \beta_k(x^k - J_k), \tag{3.33}$$

com β_k:= Busca Linear 2 (x^k, θ) e $J_k := J(x^k, 1)$ como em (2.3).

Critério de parada: Se $x^{k+1} = x^k$, pare. Caso contrário, faça $k = k + 1$ e retorne ao passo iterativo.

Como consequência direta do Lema (2.4) e da convexidade da função $h(\cdot)$ temos que $x^k \in$ dom(h) para todo $k \in \mathbb{N}$. Além disso, segue-se da Busca Linear 2 (2.14) que para $x = x^k$, $J_x = J_k$ e $\beta = \beta_k$, temos

$$(g+h)(x^k-\beta_k(x^k-J_k)) \leqslant (g+h)(x^k)-\beta_k[h(x^k)-h(J_k)]-\beta_k\langle\nabla g(x^k), x^k-J_k\rangle+\frac{\beta_k}{2}\|x^k-J_k\|^2,$$

Daí, utilizando (3.33),

$$(g + h)(x^{k+1}) \leqslant (g + h)(x^k) - \beta_k[h(x^k) - h(J_k)] - \beta_k\langle\nabla g(x^k), x^k - J_k\rangle + \frac{\beta_k}{2}\|x^k - J_k\|^2. \tag{3.34}$$

Em seguida, iremos obter alguns resultados auxiliares para o **MPG3** que serão essenciais na análise de convergência.

Proposição 3.3. *Seja $x \in$ dom(h). Então temos que*

$$\|x^{k+1}-x\|^2 \leqslant \|x^k-x\|^2+2[(g+h)(x^k)-(g+h)(x^{k+1})]+2\beta_k[(g+h)(x)-(g+h)(x^k)], \ \forall k \in \mathbb{N}.$$

Demonstração. Dado $x \in \text{dom}(h)$, para simplicidade nas contas, vamos definir

$$A_k := 2\langle x^k - x^{k+1}, x^k - x\rangle = \|x^{k+1} - x^k\|^2 + \|x^k - x\|^2 - \|x^{k+1} - x\|^2.$$

Além disso, substituindo x^{k+1} dado em (3.33) na igualdade anterior, temos

$$\begin{aligned}\frac{A_k}{2} &= \langle x^k - (x^k - \beta_k(x^k - J_k)), x^k - x\rangle\\ &= \langle x^k - (x^k - \beta_k x^k - \beta_k J_k), x^k - x\rangle\\ &= \langle \beta_k(x^k - J_k), x^k - x\rangle\\ &= \beta_k\langle x^k - J_k, x^k - x\rangle,\end{aligned}$$

reorganizando, temos

$$\frac{A_k}{2\beta_k} = \langle x^k - J_k, x^k - x\rangle.$$

Somando e subtraindo $\nabla g(x^k)$, temos

$$\frac{A_k}{2\beta_k} = \langle \nabla g(x^k), x^k - x\rangle + \langle x^k - J_k - \nabla g(x^k), x^k - x\rangle.$$

Agora, somando e subtraindo J_k,

$$\begin{aligned}\frac{A_k}{2\beta_k} &= \langle \nabla g(x^k), x^k - x\rangle + \langle x^k - J_k - \nabla g(x^k), J_k - x\rangle + \langle x^k - J_k - \nabla g(x^k), x^k - J_k\rangle\\ &= \langle \nabla g(x^k), x^k - x\rangle + \langle x^k - J_k - \nabla g(x^k), J_k - x\rangle\\ &\quad - \langle \nabla g(x^k), x^k - J_k\rangle + \langle x^k - J_k, x^k - J_k\rangle,\end{aligned}$$

ou seja,

$$\frac{A_k}{2\beta_k} = \langle \nabla g(x^k), x^k - x\rangle + \langle x^k - J_k - \nabla g(x^k), J_k - x\rangle - \langle \nabla g(x^k), x^k - J_k\rangle + \|x^k - J_k\|^2. \quad (3.35)$$

De (3.32) e pela condição de otimalidade (2.6) (vale lembrar que neste caso, $\alpha = 1$) temos que $x^k - \nabla g(x^k) - J_k \in \partial h(J_k)$. Dessa forma, aplicando a desigualdade do gradiente (A.1) para g e do subgradiente (A.8) em (3.35) para h nos dois primeiros termos do lado direito da igualdade acima, temos

$$\frac{A_k}{2\beta_k} \geqslant g(x^k) - g(x) + h(J_k) - h(x) - \langle \nabla g(x^k), x^k - J_k\rangle + \|x^k - J_k\|^2. \quad (3.36)$$

Sendo assim, organizando (3.34) temos,

$$(g + h)(x^{k+1}) - (g + h)(x^k) + \beta_k[h(x^k) - h(J_k)] - \frac{\beta_k}{2}\|x^k - J_k\|^2 \leqslant -\beta_k\langle \nabla g(x^k), x^k - J_k\rangle,$$

o que por sua vez, dividindo por $\beta_k > 0$ nos fornece que

$$\frac{1}{\beta_k}\left[(g+h)(x^{k+1}) - (g+h)(x^k)\right] + h(x^k) - h(J_k) - \frac{1}{2}\|x^k - J_k\|^2 \leqslant -\langle \nabla g(x^k), x^k - J_k\rangle.$$

Somando a desigualdade anterior a (3.36), obtemos

$$\begin{aligned}
\frac{A_k}{2\beta_k} &\geqslant g(x^k) + h(J_k) - (g+h)(x) + \frac{1}{\beta_k}[(g+h)(x^{k+1}) - (g+h)(x^k)] \\
&\quad + h(x^k) - h(J_k) - \frac{1}{2}\|x^k - J_k\|^2 + \|x^k - J_k\|^2 \\
&= g(x^k) + h(J_k) - (g+h)(x) + \frac{1}{\beta_k}[(g+h)(x^{k+1}) - (g+h)(x^k)] + h(x^k) - h(J_k) \\
&\quad + \frac{1}{2}\|x^k - J_k\|^2 \\
&\geqslant [(g+h)(x^k) - (g+h)(x)] + \frac{1}{\beta_k}[(g+h)(x^{k+1}) - (g+h)(x^k)] + \frac{1}{2}\|x^k - J_k\|^2.
\end{aligned}$$

Portanto, concluímos que

$$-\frac{A_k}{2\beta_k} \leqslant [(g+h)(x) - (g+h)(x^k)] + \frac{1}{\beta_k}[(g+h)(x^k) - (g+h)(x^{k+1})] - \frac{1}{2}\|x^k - J_k\|^2.$$

Sendo assim, como $A_k = \|x^{k+1} - x^k\|^2 + \|x^k - x\|^2 - \|x^{k+1} - x\|^2$, podemos substituir A_k na desigualdade acima, e encontramos

$$\begin{aligned}
\|x^{k+1} - x\|^2 \leqslant \|x^{k+1} - x^k\|^2 + \|x^k - x\|^2 + 2\beta_k[(g+h)(x) - (g+h)(x^k)] \\
+ 2[(g+h)(x^k) - (g+h)(x^{k+1})] - \beta_k\|x^k - J_k\|^2.
\end{aligned}$$

Como por (3.33) temos que $x^{k+1} - x^k = \beta_k(x^k - J_k)$ e $\beta_k^2 \leqslant \beta_k$. Podemos concluir que,

$$\begin{aligned}
\|x^{k+1} - x\|^2 &\leqslant \|x^k - x\|^2 + \|\beta_k(x^k - J_k)\|^2 - \beta_k\|x^k - J_k\|^2 \\
&\quad + 2[(g+h)(x^k) - (g+h)(x^{k+1})] + 2\beta_k[(g+h)(x) - (g+h)(x^k)] \\
&= \|x^k - x\|^2 + (\beta_k^2 - \beta_k)\|x^k - J_k\|^2 + 2[(g+h)(x^k) - (g+h)(x^{k+1})] \\
&\quad + 2\beta_k[(g+h)(x) - (g+h)(x^k)] \\
&\leqslant \|x^k - x\|^2 + 2[(g+h)(x^k) - (g+h)(x^{k+1})] \\
&\quad + 2\beta_k[(g+h)(x) - (g+h)(x^k)],
\end{aligned}$$

ou seja,

$$\|x^{k+1} - x\|^2 \leqslant \|x^k - x\|^2 + 2[(g+h)(x^k) - (g+h)(x^{k+1})] + 2\beta_k[(g+h)(x) - (g+h)(x^k)], \forall k \in \mathbb{N}.$$

$\square$

Vale ressaltar que se usarmos $x = x^k \in \text{dom(h)}$ na proposição anterior, temos

$$\begin{aligned}\|x^{k+1} - x^k\|^2 &\leqslant \|x^k - x^k\|^2 + 2[(g+h)(x^k) - (g+h)(x^{k+1})] \\ &\quad + 2\beta_k[(g+h)(x^k) - (g+h)(x^k)] \\ &= 2[(g+h)(x^k) - (g+h)(x^{k+1})].\end{aligned}$$

o que por sua vez, implica que

$$(g+h)(x^k) - (g+h)(x^{k+1}) \geqslant \frac{1}{2}\|x^{k+1} - x^k\|^2 \geqslant 0, \tag{3.37}$$

sendo assim,

$$(g+h)(x^k) - (g+h)(x^{k+1}) \geqslant 0.$$

Logo,

$$(g+h)(x^k) \geqslant (g+h)(x^{k+1}).$$

Podemos concluir então, que o **MPG3** também é um método de descida.

A seguir se encontra o resultado principal desta seção, cujo a afirmação é semelhante ao Teorema 3.1.

Teorema 3.2. *Seja (x^k) a sequência gerada pelo **MPG3**, as seguintes afirmações valem:*

(i) Se $S^* \neq \emptyset$ então (x^k) é Quasi-Fejér convergente para S^* e converge para um ponto em S^*.

(ii) Se $S^* = \emptyset$ então temos

$$\lim_{k\to\infty} \|x^k\| = +\infty \quad e \quad \lim_{k\to\infty} (g+h)(x^k) = \inf_{x\in\mathbb{R}^n} (g+h)(x). \tag{3.38}$$

Demonstração. Para justificar (i), vamos supor que $S^* \neq \emptyset$. Empregando a Proposição 3.3 com $x = x^* \in S^* \subseteq \text{dom(h)}$, temos

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leqslant \|x^k - x^*\|^2 + 2[(g+h)(x^k) - (g+h)(x^{k+1})] \\ &\quad + 2\beta_k[(g+h)(x^*) - (g+h)(x^k)],\end{aligned}$$

como $x^* \in S^*$, então $(g+h)(x^*) \leqslant (g+h)(x^k)$. Sendo assim, vale a seguinte desigualdade

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - x^*\|^2 + 2[(g+h)(x^k) - (g+h)(x^{k+1})], \quad \forall k \in \mathbb{N}. \tag{3.39}$$

Por (3.37), seja $\varepsilon_k = 2[(g+h)(x^k) - (g+h)(x^{k+1})] \geqslant 0$. Além disso, aplicando somatório de $k = 0, 1, 2, \ldots, N$, temos

$$\sum_{k=0}^{N} \varepsilon_k = 2 \sum_{k=0}^{N} [(g+h)(x^k) - (g+h)(x^{k+1})],$$

utilizando soma telescópica no lado direito da igualdade, obtemos

$$\sum_{k=0}^{N} \varepsilon_k \leqslant 2[(g+h)(x^0) - (g+h)(x^{N+1})], \tag{3.40}$$

como a função $(g+h)(x)$ é monótona não-crescente e limitada inferiormente, temos pelo Teorema da convergência monótona que o $\lim\limits_{k\to\infty}(g+h)(x^k)$ existe. Dessa forma, passando limite com $N \to \infty$ em (3.40) temos

$$\sum_{k=0}^{\infty} \varepsilon_k \leqslant 2[(g+h)(x^0) - \lim_{k\to\infty}(g+h)(x^{k+1})].$$

Como $\lim\limits_{k\to\infty}(g+h)(x^{k+1}) \geqslant \inf\limits_{x\in\mathbb{R}^n}(g+h)(x)$, podemos concluir da desigualdade anterior que

$$\begin{aligned}\sum_{k=0}^{\infty} \varepsilon_k &\leqslant 2[(g+h)(x^0) - \inf_{x\in\mathbb{R}^n}(g+h)(x)] \\ &= 2[(g+h)(x^0) - (g+h)(x^*)],\end{aligned}$$

pois o ínfimo é atingido em $x^* \in S^*$. Sendo assim,

$$\sum_{k=0}^{\infty} \varepsilon_k \leqslant 2[(g+h)(x^0) - (g+h)(x^*)] < +\infty.$$

Dessa forma, podemos concluir que ε_k é somável e finito. Por (3.39), temos

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - x^*\|^2 + \varepsilon_k.$$

Logo, pela Definição (1.4), a sequência (x^k) é Quasi-Féjer convergente em S^*. Pelo Lema 1.1 (i), temos que essa sequência (x^k) é limitada, sendo assim, possui ponto de acumulação. Seja $\bar{x}$ um ponto de acumulação de (x^k). Então, existe uma subsequência (x^{n_k}) de (x^k) que converge para $\bar{x}$.

Agora, dividiremos nossa análise em dois casos. O primeiro trata do caso que o comprimento de passo β_k é limitado inferiormente por uma constante positiva, enquanto o segundo caso admite que o comprimento de passo pode convergir a zero, ao menos em alguma subsequência.

Caso 1: A sequência (β_{n_k}) não converge para zero, isto é, existe algum $\beta > 0$ e uma subsequência de (β_{n_k}) (sem reclassificação) tal que

$$\beta_{n_k} \geqslant \beta, \quad \forall k \in \mathbb{N}. \tag{3.41}$$

Reorganizando a Proposição 3.3 com $x = x^* \in S^*$ e fazendo a divisão por 2, obtemos

$$\beta_k[(g+h)(x^k)-(g+h)(x^*)] \leqslant \frac{1}{2}[\|x^k-x^*\|^2-\|x^{k+1}-x^*\|^2]+[(g+h)(x^k)-(g+h)(x^{k+1})],$$

aplicando o somatório em $k = 0, 1, 2, 3, \dots, m$ na desigualdade acima, temos

$$\begin{aligned}\sum_{k=0}^{m} \beta_k[(g+h)(x^k) - (g+h)(x^*)] \leqslant{} & \frac{1}{2}\sum_{k=0}^{m}[\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2] \\ & + \sum_{k=0}^{m}[(g+h)(x^k) - (g+h)(x^{k+1})].\end{aligned}$$

Calculando a soma telescópica no lado direito, obtemos que

$$\begin{aligned}\sum_{k=0}^{m} \beta_k[(g+h)(x^k) - (g+h)(x^*)] \leqslant{} & \frac{1}{2}[\|x^0 - x^*\|^2 - \|x^{m+1} - x^*\|^2] \\ & + [(g+h)(x^0) - (g+h)(x^{m+1})] \\ \leqslant{} & \frac{1}{2}\|x^0 - x^*\|^2 + (g+h)(x^0) - (g+h)(x^{m+1}).\end{aligned}$$

Como $(g+h)(x^{m+1}) \geqslant (g+h)(x^*)$, pois $x^* \in S^*$, temos

$$\sum_{k=0}^{m} \beta_k[(g+h)(x^k) - (g+h)(x^*)] \leqslant \frac{1}{2}\|x^0 - x^*\|^2 + (g+h)(x^0) - (g+h)(x^*).$$

Tomando limite com $m \to +\infty$, temos

$$\sum_{k=0}^{\infty} \beta_k \left[(g+h)(x^k) - (g+h)(x^*)\right] < +\infty.$$

Como a série é convergente, então seu termo geral tende a zero. Em particular, temos que

$$\lim_{k\to\infty} \beta_{n_k} [(g+h)(x^{n_k}) - (g+h)(x^*)] = 0,$$

e como $\beta_{n_k} \geqslant \beta > 0$, podemos concluir que

$$\lim_{k\to\infty} [(g+h)(x^{n_k}) - (g+h)(x^*)] = 0,$$

o que por sua vez, implica que

$$\lim_{k\to\infty} (g+h)(x^{n_k}) = (g+h)(x^*). \tag{3.42}$$

Agora, temos que $g + h$ é semicontínua inferiormente em dom(h) com $x^{n_k} \to \overline{x}$. Sendo assim, podemos concluir que $\lim_{k\to\infty} \inf(g + h)(x^{n_k}) \geqslant (g + h)(\overline{x})$. Como $x^* \in S^*$, segue que $(g + h)(x^*) \leqslant (g + h)(\overline{x})$. Sendo assim, combinando essas informações com (3.42), obtemos

$$(g + h)(x^*) \leqslant (g + h)(\overline{x}) \leqslant \lim_{k\to\infty} \inf(g + h)(x^{n_k}) = \lim_{k\to\infty} (g + h)(x^{n_k}) = (g + h)(x^*).$$

Então,

$$(g + h)(\overline{x}) = (g + h)(x^*).$$

Logo, $\overline{x} \in S^*$.

Caso 2: $\lim_{k\to\infty} \beta_k = 0$. Vamos definir $\hat{\beta}_k = \dfrac{\beta_k}{\theta} > 0$ e

$$\hat{y}^k := x^k - \hat{\beta}_k(x^k - J_k) = (1 - \hat{\beta}_k)x^k + \hat{\beta}_k J_k = x^k - \hat{\beta}_k x^k + \hat{\beta}_k J_k. \tag{3.43}$$

Note que uma vez que o comprimento de passo β_k é o primeito escalar da forma $\beta_k = \theta^j$ que satisfaz a desigualdade da Busca Linear 2 em(2.14), temos que $\hat{\beta}_k$ não satisfaz a desigualdade da busca. Sendo assim, utilizando a Busca Linear 2 (2.14) com $\hat{y}^k$ dado em (3.43) e $x = x^k$, temos

$$(g+h)(\hat{y}^k) > (g+h)(x^k) - \hat{\beta}_k[h(x^k) - h(J_k)] - \hat{\beta}_k\langle \nabla g(x^k), x^k - J_k\rangle + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2. \tag{3.44}$$

Reorganizando os termos,

$$0 > -\hat{\beta}_k\langle \nabla g(x^k), x^k - J_k\rangle + (g+h)(x^k) - (g+h)(\hat{y}^k) - \hat{\beta}_k[h(x^k) - h(J_k)] + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2,$$

o que por sua vez, implica que

$$0 > -\hat{\beta}_k\langle \nabla g(x^k), x^k - J_k\rangle + g(x^k) - g(\hat{y}^k) + h(x^k) - h(\hat{y}^k) - \hat{\beta}_k[h(x^k) - h(J_k)] + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2.$$

Desde que g é uma função convexa, utilizando a desigualdade do gradiente (A.5), com $f = g$, temos

$$g(x^k) - g(\hat{y}^k) \geqslant \langle \nabla g(\hat{y}^k), x^k - \hat{y}^k\rangle.$$

Daí e da desigualdade anterior, temos

$$0 > -\hat{\beta}_k\langle \nabla g(x^k), x^k - J_k\rangle + \langle \nabla g(\hat{y}^k), x^k - \hat{y}^k\rangle + h(x^k) - h(\hat{y}^k) - \hat{\beta}_k[h(x^k) - h(J_k)] + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2. \tag{3.45}$$

Pela convexidade de h, obtemos que $h(\hat{y}^k) \leqslant (1-\hat{\beta}_k)h(x^k) + \hat{\beta}_k h(J_k)$. Além disso, temos por (3.43) que $x^k - \hat{y}^k = -\hat{\beta}_k(J_k - x^k) = \hat{\beta}_k(x^k - J_k)$. Sendo assim, fazendo essas substituições em (3.45), resulta que

$$
\begin{aligned}
0 > &-\hat{\beta}_k\langle \nabla g(x^k), x^k - J_k\rangle + \langle \nabla g(\hat{y}^k), \hat{\beta}_k(x^k - J_k)\rangle + h(x^k) - ((1-\hat{\beta}_k)h(x^k) + \hat{\beta}_k h(J_k)) \\
&- \hat{\beta}_k[h(x^k) - h(J_k)] + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \\
&= \hat{\beta_k}\langle \nabla g(\hat{y}^k) - \nabla g(x^k), x^k - J_k\rangle + h(x^k) - h(x^k) + \hat{\beta}_k h(x^k) - \hat{\beta}_k h(J_k) \\
&- \hat{\beta}_k[h(x^k) - h(J_k)] + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \\
&= \hat{\beta_k}\langle \nabla g(\hat{y}^k) - \nabla g(x^k), x^k - J_k\rangle + \hat{\beta}_k[h(x^k) - h(J_k)] - \hat{\beta}_k[h(x^k) - h(J_k)] \\
&+ \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \\
&= \hat{\beta_k}\langle \nabla g(\hat{y}^k) - \nabla g(x^k), x^k - J_k\rangle + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \\
&= \langle \nabla g(\hat{y}^k) - \nabla g(x^k), x^k - J_k\rangle + \frac{1}{2}\|x^k - J_k\|^2,
\end{aligned}
$$

o que por sua vez, implica que

$$\frac{1}{2}\|x^k - J_k\|^2 < -\langle \nabla g(\hat{y}^k) - \nabla g(x^k), x^k - J_k\rangle.$$

Utilizando a desigualdade de Cauchy-Schwarz, temos

$$\frac{1}{2}\|x^k - J_k\|^2 \leqslant \|\nabla g(\hat{y}^k) - \nabla g(x^k)\|\|x^k - J_k\|,$$

podemos dividir por $\|x^k - J_k\| \neq 0$, pois caso contrário, teríamos $0 = 0$ (absurdo), pois a desigualdade da busca como vimos anteriormente não ocorre. Dessa forma,

$$\frac{1}{2}\|x^k - J_k\| \leqslant \|\nabla g(\hat{y}^k) - \nabla g(x^k)\|. \tag{3.46}$$

Como $\mathsf{Prox}_h(\cdot)$ é não expansivo (veja o Corolário (A.1)), obtemos da definição dada em (3.32) que

$$
\begin{aligned}
\|J_k - J_0\| &\leqslant \|x^k - \nabla g(x^k) - (x^0 - \nabla g(x^0))\| \\
&\leqslant \|x^k - x^0\| + \|\nabla g(x^k) - \nabla g(x^0)\|.
\end{aligned}
$$

Agora, segue da limitação de (x^k), uma vez que (x^k) é limitada, veja Lema 1.1, que a sequência (J_k) também é limitada pela hipótese A2. Temos juntamente com (3.43) e o fato de $\beta_k \to 0$ (por hipótese no caso 2), que $\|\hat{y}^k - x^k\| \to 0$ quando $k \to \infty$. Como

∇g é uniformemente contínuo em conjuntos limitados, podemos concluir que $\|\nabla g(\hat{y}^k) - \nabla g(x^k)\| \to 0$ quando $k \to \infty$. Sendo assim, pelo Teorema do Confronto em (3.46), temos

$$\lim_{k\to\infty} \|x^k - J_k\| = 0. \tag{3.47}$$

Novamente, como ∇g é uniformemente contínuo em conjuntos limitados, a desigualdade anterior implica que

$$\lim_{k\to\infty} \|\nabla g(x^k) - \nabla g(J_k)\| = 0. \tag{3.48}$$

Utilizando (2.5) com $z = x^k - \nabla g(x^k)$ e lembrando que no procedimento de Busca Linear 2 temos $\alpha = 1$, podemos concluir que

$$x^k - \nabla g(x^k) - \text{prox}_{\alpha h}(x^k - \nabla g(x^k)) \in \partial h(\text{prox}_{\alpha h}(x^k - \nabla g(x^k))),$$

como $\text{prox}_{\alpha h}(x^k - \nabla g(x^k)) = J_k$, temos

$$x^k - \nabla g(x^k) - J_k \in \partial h(J_k).$$

Daí, somando $\nabla g(J_k)$ em ambos os lados da inclusão, temos

$$x^k - \nabla g(x^k) - J_k + \nabla g(J_k) \in \nabla g(J_k) + \partial h(J_k) \subseteq \partial(g + h)(J_k).$$

Passando limite sobre a subsequência (x^{n_k}) na inclusão acima e lembrando que $(x^{n_k}) \to \bar{x}$, obtemos do Fato (A.1), (3.47) e (3.48) que $0 \in \partial(g + h)(\bar{x})$. Logo, $\bar{x} \in S^*$. Podemos perceber que nos dois casos listados, qualquer ponto de acumulação de (x^k) pertence ao conjunto solução S^*. Dessa forma, pelo Lema 1.1 (ii) temos que (x^k) converge para uma solução ótima em S^*, o que completa a prova de (i).

A prova do item (ii) segue de forma análoga à prova do Teorema 3.1 (i), a qual será omitida por simplicidade. □

De (3.38), e também do Teorema 3.1, pode-se questionar se no caso $S^* \neq \emptyset$, vale

$$\lim_{k\to\infty} (g + h)(x^k) = \min_{x\in\mathbb{R}^n} (g + h)(x). \tag{3.49}$$

Não conseguimos dizer que vale a resposta em geral, mas quando $g + h$ é contínua no domínio h em condições finitas ou a sequência (β_k) é limitada inferiormente por uma constante positiva, a igualdade (3.49) é verdadeira, com uma complexidade adicional discutida no próximo capítulo.

3.4 Método Proximal Gradiente 4 (MPG4)

Agora, iremos definir o chamado método forward-backward com o procedimento de Busca Linear 3, estudado na Seção 2.3. Tal método foi proposto em [10].

Método Proximal Gradiente 4 (MPG4) [10]

Passo 0: Escolha $x^0 \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$, $\sigma > 0$ e $0 < \theta < 1$.

Passo iterativo: Dado x^k, defina

$$x^{k+1} := J(x^k, \alpha_k) = \mathrm{prox}_{\alpha h}\big(x^k - \alpha_k \nabla g(x^k)\big)\,, \tag{3.50}$$

com

$$\alpha_{-1} := \sigma \quad \text{e} \quad \alpha_k := \mathrm{LS3}(x^k, \alpha_{k-1}, \theta). \tag{3.51}$$

Critério de parada: Se $x^{k+1} = x^k$, pare. Caso contrário, faça $k = k + 1$ e retorne ao passo iterativo.

O seguinte resultado, que é uma consequência direta da Proposição 2.2 desempenha um papel central em nosso estudo. Vamos mostrar agora, a boa definição do método **MPG4**.

Corolário 3.1. *Suponha* $x^0 \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$*. A sequência* (x^k) *gerada pelo método **MPG4** é bem definida, com* $x^k \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$*, e* g *é diferenciável em torno de* x^k *para todo* $k \in \mathbb{N}$*. Além disso, para qualquer* $x \in \mathbb{R}^n$*, temos:*

(i) $\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k \big[F(x^{k+1}) - F(x)\big]$.

(ii) $F(x^{k+1}) - F(x^k) \leqslant -\frac{1}{2\alpha_k} \|x^{k+1} - x^k\|^2$.

Demonstração. Dado $x^k \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$, temos pela Proposição 2.2 (iii) com $x = x^k, \bar{\alpha} = \alpha_k$ e $J(x, \bar{\alpha}) = x^{k+1}$ que $x^{k+1} \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$ e g é diferenciável em qualquer x^k pelas suposições de A3. Logo, a sequência (x^k) gerada pelo método **MPG4** é bem definida. Para provar o item (i) vamos substituir $x = x^k$ e $\overline{\alpha} = \alpha_k$ na Proposição 2.2 (ii),

$$\|x^k - u\|^2 - \|J(x^k, \alpha_k) - u\|^2 \geqslant 2\alpha_k\,[F(J(x^k, \alpha_k)) - F(u)], \quad \forall u \in \mathbb{R}^n.$$

Agora, vamos substituir $u = x$ e $J(x^k, \alpha_k) = x^{k+1}$ na desigualdade acima,

$$\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geqslant 2\alpha_k[F(x^{k+1}) - F(x)].$$

Como queríamos demonstrar no item (i).

Para provar o item (ii) vamos fazer as seguintes substituições $J(x, \overline{\alpha}) = x^{k+1}$ e $x = x^k$ na Proposição 2.2 (iii). Sendo assim, temos

$$F(x^{k+1}) - F(x^k) \leqslant -\frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2.$$

□

O seguinte resultado comprova a convergência global do método **MPG4**, sem supor a continuidade Lipschitz do gradiente ∇g (Ver Definição 1.3), conforme em ([9], Teorema 4.2) e ([32], Teorema 3.18) com novas hipóteses simplificadas (A1-A4). A prova é semelhante à de ([9], Teorema 4.2), com algumas modificações e ideias de [32].

Teorema 3.3. *Seja (x^k) a sequência gerada pelo método **MPG4**. As seguintes afirmações são válidas:*

(i) Se $S^* \neq \emptyset$, então (x^k) converge para um ponto em S^*. Além disso,

$$\lim_{k\to\infty} F(x^k) = \min_{x\in\mathbb{R}^n} F(x). \tag{3.52}$$

(ii) Se $S^* = \emptyset$, então temos:

$$\lim_{k\to\infty}\|x^k\| = +\infty \quad e \quad \lim_{k\to\infty} F(x^k) = \inf_{x\in\mathbb{R}^n} F(x). \tag{3.53}$$

Demonstração. Vamos provar o item (i). Suponha que $S^* \neq \emptyset$, então pelo Corolário 3.1 (i) fazendo a substituição de $x = x^*$, temos que para qualquer $x^* \in S^*$, vale

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \geqslant 2\alpha_k[F(x^{k+1}) - F(x^*)] \geqslant 0. \tag{3.54}$$

Podemos observar pela desigualdade (3.54) que a sequência (x^k) é Féjer convergente em relação ao conjunto S^*, pois $\|x^k - x^*\| \geqslant \|x^{k+1} - x^*\|^2$. Logo, pelo Lema 1.1 (ii), (x^k) é limitada. Vamos definir $M := \sup\{\|x^k - x^*\| : k \in \mathbb{N}\} < \infty$ para algum $x^* \in S^*$. Reorganizando (3.54), temos que

$$0 \leqslant 2\alpha_k[F(x^{k+1}) - F(x^*)] \leqslant \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2,$$

o que por sua vez, implica que

$$0 \leqslant 2\alpha_k[F(x^{k+1}) - F(x^*)] \leqslant (\|x^k - x^*\| + \|x^{k+1} - x^*\|)(\|x^k - x^*\| - \|x^{k+1} - x^*\|).$$

Pela desigualdade triangular inversa $\|x^k - x^{k+1}\| \geqslant |\|x^k - x^*\| - \|x^{k+1} - x^*\||$, temos:

$$0 \leqslant 2\alpha_k[F(x^{k+1}) - F(x^*)] \leqslant (\|x^k - x^*\| + \|x^{k+1} - x^*\|)\|x^k - x^{k+1}\|,$$

sendo assim, como $M = \sup\{\|x^k - x^*\| : k \in \mathbb{N}\}$, temos

$$0 \leqslant 2\alpha_k[F(x^{k+1}) - F(x^*)] \leqslant 2M\|x^k - x^{k+1}\|.$$

Agora, dividindo por 2 e usando o fato que $\alpha_k > 0$, temos

$$0 \leqslant [F(x^{k+1}) - F(x^*)] \leqslant \frac{M\|x^k - x^{k+1}\|}{\alpha_k}. \tag{3.55}$$

Reorganizando a desigualdade do Corolário 3.1 (ii), temos que

$$F(x^k) - F(x^{k+1}) \geqslant \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2.$$

Como $\sigma \geqslant \alpha_k$, temos da desigualdade anterior que

$$2\sigma[F(x^k) - F(x^{k+1})] \geqslant 2\alpha_k[F(x^k) - F(x^{k+1})] \geqslant \|x^k - x^{k+1}\|^2.$$

Agora, desde que a sequência $(F(x^k))$ seja monótona não-crescente e limitada inferiormente por $F(x^*)$ com $(x^* \in S^*)$. Para $k = 0, 1, 2, \ldots, N$, temos

$$2\sigma\sum_{k=0}^{N}[F(x^k) - F(x^{k+1})] \geqslant 2\alpha_k\sum_{k=0}^{N}[F(x^k) - F(x^{k+1})] \geqslant \sum_{k=0}^{N}\|x^k - x^{k+1}\|^2,$$

dessa forma,

$$2\sigma\sum_{k=0}^{N}[F(x^k) - F(x^{k+1})] \geqslant \sum_{k=0}^{N}\|x^k - x^{k+1}\|^2.$$

Utilizando a série telescópica do lado esquerdo da equação anterior, temos

$$2\sigma[F(x^0) - F(x^{N+1})] \geqslant \sum_{k=0}^{N}\|x^k - x^{k+1}\|^2. \tag{3.56}$$

Temos que $F(x^*) \leqslant F(x^k)$, para todo $k \in \mathbb{N}$. Concluímos assim, que $-F(x^*) \geqslant -F(x^k)$. Em particular $-F(x^*) \geqslant -F(x^{N+1})$. Como $\sigma > 0$, vale

$$2\sigma[F(x^0) - F(x^*)] \geqslant 2\sigma[F(x^0) - F(x^{N+1})],$$

o que por sua vez, combinando com (3.56) e a desigualdade anterior, temos

$$2\sigma[F(x^0) - F(x^*)] \geqslant \sum_{k=0}^{N}\|x^k - x^{k+1}\|^2.$$

Tomando o limite quando $N \to +\infty$,

$$\sum_{k=0}^{\infty} \|x^k - x^{k+1}\|^2 \leqslant 2\sigma[F(x^0) - F(x^*)] < +\infty. \tag{3.57}$$

Logo, a série é convergente, ou seja, $\|x^k - x^{k+1}\|^2 \to 0$ quando $k \to \infty$. Vimos anteriormente que (x^k) é uma sequência Féjer convergente em relação ao conjunto solução S^*. Portanto, pelo Lema 1.1 (i), é limitada. Sendo assim, vamos supor que $\overline{x}$ é um ponto de acumulação dessa sequência (x^k) e que (x^{n_k}) é uma subsequência de (x^k) de forma que $x^{n_k} \to \overline{x}$.

Temos que F é semicontínua inferiormente, e que $(F(x^k))$ é monótona não-crescente, sendo assim, como $x^{n_k} \to \bar{x}$, temos que

$$F(x^k) \geqslant \liminf_{kj \to \infty} F(x^{kj}) \geqslant F(\bar{x}), \quad \forall k \in \mathbb{N}.$$

Em particular, vale para $k = 0$, logo $F(x^0) \geqslant F(\bar{x})$. Sendo assim, como sabemos que $x^{n_k} \to \bar{x}$ e (x^{n_k}) é uma subsequência de (x^k) onde $(x^k) \subset \text{int}(\text{dom}(g)) \cap \text{dom}(h)$. Podemos concluir por A3 e A4 que g é diferenciável em $\bar{x} \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$. De fato, se $\bar{x} \in \text{dom}(h)$, sabemos que $h(x^k) < +\infty$, para todo $k \in \mathbb{N}$, pois h é semicontínua inferiormente, logo

$$+\infty > \liminf h(x^{k_j}) \geqslant h(\bar{x}) \;\therefore\; h(\bar{x}) < +\infty,$$

isto é, $\bar{x} \in \text{dom}(h)$. Agora, se $\bar{x} \in \text{int}(\text{dom}(g))$: como $x^{k_j} \in \text{int}(\text{dom}(g))$, isto é, existe $\varepsilon_k > 0$ com $B(x^{k_j}, \varepsilon_k) \subset \text{dom}(g)$. Por outro lado, $x^{k_j} \to \bar{x}$, isto é, para k_j suficientemente grande temos $\bar{x} \in B(x^{k_j}, \varepsilon_k) \subset \text{dom}(h)$.

$$\therefore\; \bar{x} \in \text{int}(\text{dom}(g)) \cap \text{dom}(h).$$

Como (α_k) é monótona não-crescente por construção vista na Busca Linear 3 (veja, (3.51)), no qual $\alpha_k \leqslant \alpha_{k-1}$ e além disso (α_k) é limitada inferiormente por zero. Portanto, podemos concluir que a sequência (α_k) é monótona e limitada, logo convergente. Suponha que $\lim\limits_{k \to \infty} \alpha_k = \lim\limits_{k \to \infty} \alpha_{n_k} = \alpha \geqslant 0$.

Agora, precisamos verificar os casos $\alpha > 0$ e $\alpha = 0$.

Caso 1: $\alpha > 0$. Utilizando $z = x^{n_k} - \alpha_{n_k} \nabla g(x^{n_k})$ em (2.5) e a definição de x^{n_k+1} dada em (3.50), obtemos,

$$\frac{x^{n_k} - \alpha_{n_k} \nabla g(x^{n_k}) - x^{n_k+1}}{\alpha_{n_k}} \in \partial h(x^{n_k+1}),$$

o que por sua vez, implica que

$$\frac{x^{n_k} - x^{n_k+1}}{\alpha_{n_k}} - \nabla g(x^{n_k}) + \nabla g(x^{n_k+1}) \in \nabla g(x^{n_k+1}) + \partial h(x^{n_k+1}). \tag{3.58}$$

Agora, segue do Corolário 3.1 (ii), que

$$\frac{1}{2}\frac{\|x^{n_k+1} - x^{n_k}\|^2}{\alpha_{n_k}} \leqslant F(x^{n_k}) - F(x^{n_k+1}).$$

Aplicando somatório de $k = 0, 1, \dots, N$, na desigualdade anterior, temos

$$\frac{1}{2}\sum_{k=0}^{N}\left(\frac{\|x^{n_k+1} - x^{n_k}\|^2}{\alpha_{n_k}}\right) \leqslant \sum_{k=0}^{N}(F(x^{n_k}) - F(x^{n_k+1})).$$

Agora, utilizando soma telescópica no lado direito da desigualdade anterior, obtemos

$$\frac{1}{2}\sum_{k=0}^{N}\left(\frac{\|x^{n_k+1} - x^{n_k}\|^2}{\alpha_{n_k}}\right) \leqslant F(x^0) - F(x^{N+1}). \tag{3.59}$$

Como $(F(x^k))$ é monótona não-crescente, temos que $F(x^{N+1}) \geqslant F(x^*)$, que por sua vez, implica que $-F(x^{N+1}) \leqslant -F(x^*)$. Utilizando essa informação com (3.59), obtemos

$$\frac{1}{2}\sum_{k=0}^{N}\left(\frac{\|x^{n_k+1} - x^{n_k}\|^2}{\alpha_{n_k}}\right) \leqslant F(x^0) - F(x^*).$$

Aplicando limite quando $N \to \infty$, temos

$$\frac{1}{2}\sum_{k=0}^{\infty}\left(\frac{\|x^{n_k+1} - x^{n_k}\|^2}{\alpha_{n_k}}\right) \leqslant F(x^0) - F(x^*) < +\infty. \tag{3.60}$$

Logo, a série é convergente, ou seja, seu termo geral vai a zero, o que implica que $\frac{\|x^{n_k+1} - x^{n_k}\|^2}{\alpha_{n_k}} \to 0$. Porém, como $\alpha_{n_k} > \alpha > 0$, temos que $\frac{\|x^{n_k+1} - x^{n_k}\|}{\alpha_{n_k}} \to 0$. Pela mesma razão, utilizando (3.57), concluímos que $\|x^{n_k+1} - x^{n_k}\| \to 0$, quando $k \to \infty$ e pela continuidade do gradiente em g vista em A3, podemos concluir que $\|\nabla g(x^{n_k+1}) - \nabla g(x^{n_k})\| \to 0$, quando $k \to \infty$. Como $x^{n_k} \to \bar{x}$ e $\|x^{n_k+1} - x^{n_k}\| \to 0$, quando $k \to \infty$, então $x^{n_k+1} \to \bar{x}$.

Portanto, utilizando essas informações e o Teorema do gráfico fechado do subdiferencial (veja (A.1), temos que tomando o limite em (3.58) com $n_k \to +\infty$, que $0 \in \nabla g(\bar{x}) + \partial h(\bar{x})$. Portanto, $\bar{x} \in S^*$. Além disso, como $(F(x^k))$ é monótona não-crescente, passando limite quando $k \to \infty$ em (3.55), obtemos

$$0 \leqslant \lim_{k\to\infty}\left(F(x^k) - F(x^*)\right) \leqslant \lim_{k\to\infty}\left(M\frac{\|x^k - x^{k+1}\|}{\alpha_k}\right).$$

Vimos anteriormente que uma consequência direta de (3.57) é o fato que $\|x^k - x^{k+1}\| \to 0$, quando $k \to \infty$. Sendo assim, como (α_k) é uma sequência limitada, obtemos

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

Sendo assim, pelo Teorema do Confronto, obtemos que $\big(F(x^k) - F(x^*)\big) \to 0$, quando $k \to \infty$. Dessa forma, como $x^* \in S^*$, então

$$\lim_{k\to\infty} F(x^k) - \min_{x\in\mathbb{R}^n} F(x) = 0,$$

ou seja, $\lim\limits_{k\to\infty} F(x^k) = \min\limits_{x\in\mathbb{R}^n} F(x)$.

Caso 2: $\alpha = 0$. Defina $x = x^{n_k}$, $\hat{\alpha}_{n_k} = \frac{\alpha_{n_k}}{\theta} > \alpha_{n_k} > 0$ e $\hat{x}^{n_k} = J(x^{n_k}, \hat{\alpha}_{n_k}) \in \text{dom}(h)$. Fazendo essas substituições no procedimento de Busca Linear 3 (veja (2.19)), obtemos

$$g(\hat{x}^{n_k}) > g(x^{n_k}) + \langle \nabla g(x^{n_k}), \hat{x}^{n_k} - x^{n_k}\rangle + \frac{1}{2\hat{\alpha}_{n_k}}\|x^{n_k} - \hat{x}^{n_k}\|^2. \tag{3.61}$$

De acordo com o Lema 2.2 e lembrando que para o **MPG4** temos $x^{n_k+1} = J(x^{n_k}, \alpha_{n_k})$ (veja (3.50)), concluímos que

$$\begin{aligned}
\|x^{n_k} - \hat{x}^{n_k}\| = \|x^{n_k} - J(x^{n_k}, \hat{\alpha}_{n_k})\| &\leqslant \frac{\hat{\alpha}_{n_k}}{\alpha_{n_k}}\|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| \\
&= \frac{\frac{\alpha_{n_k}}{\theta}}{\alpha_{n_k}}\|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| = \frac{1}{\theta}\|x^{n_k} - x^{n_k+1}\|,
\end{aligned}$$

o que por sua vez, implica que

$$0 \leqslant \|x^{n_k} - \hat{x}^{n_k}\| \leqslant \frac{1}{\theta}\|x^{n_k} - x^{n_k+1}\|.$$

Como $\theta > 0$ e $\|x^{n_k} - x^{n_k+1}\| \to 0$ quando $k \to \infty$, concluímos pelo Teorema do Confronto que $\|x^{n_k} - \hat{x}^{n_k}\| \to 0$. Como $x^{n_k} \to \overline{x}$ então $\hat{x}^{n_k} \to \overline{x}$ quando $k \to +\infty$. Ou seja, a sequência $(\hat{x}^{n_k})$ converge para algum ponto $\overline{x} \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$.

Agora, vamos definir o conjunto $\mathcal{A} := \{\bar{x}, (x^{n_k})\}$. No qual, denotamos anteriormente $\overline{x}$ como um ponto de acumulação de (x^k) e (x^{n_k}) uma subsequência de (x^k) que converge para $\overline{x}$. Sendo assim, segue que

$$\mathcal{A} \subset S_{lev}(x^0) := \{x \in \mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h) : F(x) \leqslant F(x^0)\},$$

com tal conjunto de nível da função $F(x) = (g+h)(x)$, sendo um conjunto compacto. De fato, uma vez que $(F(x^k))$ é monótona não-crescente, então vale

$$(g+h)(x^{n_k}) \leqslant (g+h)(x^0), \quad \forall\, k \in \mathbb{N}.$$

Logo, $x^{n_k} \in S_{lev}(x^0)$ para todo $k \in \mathbb{N}$. Como F é semicontínua inferior, então

$$(g+h)(\overline{x}) \leqslant \liminf_{k\to\infty}(g+h)(x^{n_k}) \leqslant (g+h)(x^0).$$

Dessa forma, $(g+h)(\overline{x}) \leqslant (g+h)(x^0)$. Portanto, $\overline{x} \in S_{lev}(x^0)$. Ou seja, $A \subseteq S_{lev}(x^0)$ é um conjunto compacto. Por A4, temos que

$$d(A, \mathbb{R}^n/\text{int}(\text{dom}(g))) > 0.$$

Sendo assim, os pontos de $\mathcal{A}$ estão dentro do interior do domínio de g, de forma que não estão na "borda".

Dessa forma, temos as seguintes informações $x^{n_k} \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$, $\|x^{n_k} - \hat{x}^{n_k}\| \to 0$ e $d(A, \mathbb{R}^n/\text{int}(\text{dom}(g))) > 0$, sendo assim, para k suficientemente grande, temos que $\hat{x}^{n_k} \in \text{int}(\text{dom}(g))$, portanto $\hat{x}^{n_k} \in \text{int}(\text{dom}(g)) \cap \text{dom}(h)$ para k suficientemente grande.

Temos pela Hipótese A3 que g é diferenciável em $\text{int}(\text{dom}(g)) \cap \text{dom}(h)$, sendo assim $\nabla g(x)$ existe para todo $x^{n_k} \in \text{int}(\text{dom}(g))$. Como $\hat{x}^{n_k} \in \text{int}(\text{dom}(g))$ para k suficientemente grande, então $\nabla g(\hat{x}^{n_k})$ existe e está bem definido. Agora utilizando a desigualdade do gradiente (A.5) em g, obtemos

$$g(x^{n_k}) \geqslant g(\hat{x}^{n_k}) + \langle \nabla g(\hat{x}^{n_k}), x^{n_k} - \hat{x}^{n_k} \rangle.$$

Reorganizando (3.61), e utilizando a desigualdade anterior, vale

$$\begin{aligned}
\frac{1}{2\hat{\alpha}_{n_k}}\|x^{n_k} - \hat{x}^{n_k}\|^2 &< \langle \nabla g(x^{n_k}), x^{n_k} - \hat{x}^{n_k} \rangle + g(\hat{x}^{n_k}) - g(x^{n_k}) \\
&= \langle \nabla g(x^{n_k}), x^{n_k} - \hat{x}^{n_k} \rangle - \langle \nabla g(\hat{x}^{n_k}), x^{n_k} - \hat{x}^{n_k} \rangle \\
&= \langle \nabla g(x^{n_k}) - \nabla g(\hat{x}^{n_k}), x^{n_k} - \hat{x}^{n_k} \rangle.
\end{aligned}$$

Utilizando a desigualdade de Cauchy–Schwarz, temos

$$\frac{1}{2\hat{\alpha}_{n_k}}\|x^{n_k} - \hat{x}^{n_k}\|^2 \leqslant \|\nabla g(x^{n_k}) - \nabla g(\hat{x}^{n_k})\| \cdot \|x^{n_k} - \hat{x}^{n_k}\|.$$

Agora, note que $\|x^{n_k} - \hat{x}^{n_k}\| \neq 0$, pois caso contrário teríamos uma contradição na desigualdade anterior, ou ainda segue do fato de que x^{n_k} é a iterada obtida com α_k e $\hat{x}^{n_k}$ é obtida com $\hat{\alpha}_k$. Então, segue que

$$\|\nabla g(x^{n_k}) - \nabla g(\hat{x}^{n_k})\| \geqslant \frac{1}{2\hat{\alpha}_{n_k}}\|x^{n_k} - \hat{x}^{n_k}\| \geqslant 0. \tag{3.62}$$

Como o conjunto $\{\bar{x}, (x^{n_k}), (\hat{x}^{n_k})\}$ é um subconjunto compacto de int(dom(g)) $\cap$ dom(h) e $\|x^{n_k} - \hat{x}^{n_k}\| \to 0$ quando $k \to \infty$, podemos concluir pela Hipótese A3 que $\|\nabla g(x^{n_k}) - \nabla g(\hat{x}^{n_k})\| \to 0$ quando $k \to \infty$. Sendo assim, como $\hat{\alpha}_{n_k} > \alpha_{n_k} > 0$, passando limite quando $k \to \infty$ em (3.62) e utilizando o Teorema do Confronto, concluímos que

$$\lim_{k\to\infty} \frac{\|x^{n_k} - \hat{x}^{n_k}\|}{\hat{\alpha}_{n_k}} = 0. \tag{3.63}$$

Substituindo $z = x^{n_k} - \hat{\alpha}_{n_k}\nabla g(x^{n_k})$ em (2.5), e lembrando das seguintes substituições $\text{Prox}_{\alpha h}(x^{n_k} - \hat{\alpha}_{n_k}\nabla g(x^{n_k})) = J(x^{n_k}, \hat{\alpha}_{n_k})$ e $J(x^{n_k}, \hat{\alpha}_{n_k}) = \hat{x}^{n_k}$, obtemos que

$$\frac{x^{n_k} - \hat{x}^{n_k}}{\hat{\alpha}_{n_k}} - \nabla g(x^{n_k}) \in \partial h(\hat{x}^{n_k}),$$

acrescentando $\nabla g(\hat{x}^{n_k})$ em ambos os lados da inclusão anterior, vale

$$\frac{x^{n_k} - \hat{x}^{n_k}}{\hat{\alpha}_{n_k}} - \nabla g(x^{n_k}) + \nabla g(\hat{x}^{n_k}) \in \nabla g(\hat{x}^{n_k}) + \partial h(\hat{x}^{n_k}).$$

Temos as seguintes informações $\hat{x}^{n_k} \to \bar{x}$, $\|x^{n_k} - \hat{x}^{n_k}\| \to 0$ e que pela Hipótese A2 o gradiente é uniformemente contínuo. Sendo assim, podemos concluir que $\|\nabla g(x^{n_k}) - \nabla g(\hat{x}^{n_k})\| \to 0$ quando $k \to \infty$.

Utilizando as informações anteriores, (3.63) e o Fato (A.1), quando $k \to \infty$, obtemos

$$\frac{x^{n_k} - \hat{x}^{n_k}}{\hat{\alpha}_{n_k}} - \nabla g(x^{n_k}) + \nabla g(x^{n_k}) \in \nabla g(\bar{x}) + \partial h(\bar{x}).$$

Dessa forma, $0 \in \nabla g(\bar{x}) + \partial h(\bar{x})$. Logo, $\bar{x} \in S^*$.

Além disso, fazendo algumas modificações no Lema (2.9) com $x = x^{n_k}$, $\alpha_1 = \alpha_{n_k}$ e $\alpha_2 = \hat{\alpha}_{n_k}$, e recordando que $J(x^{n_k}, \hat{\alpha}_{n_k}) = \hat{x}^{n_k}$ e $x^{n_k+1} = J(x^{n_k}, \alpha_{n_k})$, obtemos

$$\|x^{n_k} - \hat{x}^{n_k}\| = \|x^{n_k} - J(x^{n_k}, \hat{\alpha}_{n_k})\| \geqslant \|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| = \|x^{n_k} - x^{n_k+1}\|,$$

sendo assim,

$$\|x^{n_k} - \hat{x}^{n_k}\| \geqslant \|x^{n_k} - x^{n_k+1}\|.$$

Como $\hat{\alpha}_{n_k} > \alpha_{n_k} > 0$, então

$$\frac{\|x^{n_k} - \hat{x}^{n_k}\|}{\hat{\alpha}_{n_k}} \geqslant \frac{\|x^{n_k} - x^{n_k+1}\|}{\hat{\alpha}_{n_k}} > 0.$$

Utilizando (3.63) e o Teorema do confronto, temos que $\lim\limits_{k\to\infty} \dfrac{\|x^{n_k} - x^{n_k+1}\|}{\hat{\alpha}_{n_k}} = 0$. Dessa forma, como $\hat{\alpha}_{n_k} = \frac{\alpha_{n_k}}{\theta}$, onde $\theta > 0$. Obtemos que, $\lim\limits_{k\to\infty} \dfrac{\theta\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} = 0$. Logo, $\lim\limits_{k\to\infty} \dfrac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} = 0$.

Sabemos que $(F(x^k))$ é monótona não-crescente e $\lim\limits_{k\to\infty} \dfrac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} = 0$. Sendo assim, aplicando limite quando $k \to \infty$, em (3.55), temos

$$0 \leqslant \lim_{k\to\infty} \left(F(x^{n_k+1}) - F(x^*)\right) \leqslant \lim_{k\to\infty} M \frac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} = 0.$$

o que por sua vez, pelo Teorema do confronto implica que, $\lim\limits_{k\to\infty} \left(F(x^{n_k+1}) - F(x^*)\right) = 0$. Portanto,

$$\lim_{k\to\infty} F(x^k) = \min_{x\in\mathbb{R}^n} F(x).$$

Portanto, em ambos os casos analisados acima, obtemos (3.52). Além disso, concluímos que dado qualquer ponto de acumulação da sequência (x^k), a qual provamos que é Fejér convergente em relação ao conjunto solução S^*, pertence ao conjunto S^*.

A prova do item (ii) segue de maneira análoga à prova do Teorema 3.1 (ii). Como vimos anteriormente na prova do item (i), concluímos que se $S^* \neq \emptyset$, então a sequência (x^k) é limitada. Além disso, todos os seus pontos de acumulação são soluções. Agora, suponha que $S^* = \emptyset$, dessa forma qualquer subsequência de (x^k) é ilimitada, pois caso contrário, teria ponto de acumulação e tal ponto seria solução. Logo, $\|x^k\| \to +\infty$ quando $k \to +\infty$. Além disso, segue da definição de ínfimo que,

$$s := \lim_{k\to+\infty} F(x^k) \geqslant \inf_{x\in\mathbb{R}^n} F(x),$$

de forma que s existe pois $(F(x^k))$ é monótona não-crescente pelo Corolário 3.1 (ii) e também é limitada inferiormente. Vamos supor que $s > \inf\limits_{x\in\mathbb{R}^n}(g+h)(x)$. Dessa forma, podemos concluir que o seguinte conjunto

$$S_{lev}(x^0) := \{x \in \mathrm{dom}(g) : F(x) \leqslant F(x^k), \quad \forall k \in \mathbb{N}\},$$

não é vazio. Agora, aplicando o Corolário 3.1 (i) com $x \in S_{lev}(x^0)$, obtemos que

$$\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \ \geqslant\ 2\alpha_k \left[F(x^{k+1}) - F(x)\right] > 0,$$

o que por sua vez, implica que $\|x^k - x\|^2 > \|x^{k+1} - x\|^2$, ou seja, a sequência (x^k) é Fejér convergente para $S_{lev}(x^0)$. Chegamos em um absurdo, pois pelo Lema 1.1 (i) a sequência (x^k) é limitada. Entretanto, supomos anteriormente que era ilimitada. Logo, $s = \inf\limits_{x\in\mathbb{R}^n}(g+h)(x)$. Sendo assim,

$$\lim_{k\to+\infty} F(x^k) = \inf_{x\in\mathbb{R}^n} F(x).$$

□

A proposição a seguir mostra que, quando g é localmente Lipschitz contínua (Ver Definição 1.3), o tamanho do passo α_k é limitado inferiormente por um número positivo. A segunda parte desse resultado coincide em [[7], Observação 1.2].

Proposição 3.4. *Seja* (x^k) *e* (α_k) *as duas sequências geradas pelo método* ***MPG4***. *Suponha que* $S^* \neq \emptyset$ *e que a sequência* (x^k) *está convergindo para algum* $x^* \in S^*$. *Se* g *é Localmente lipschitz contínua em torno de* x^* *com módulo* L, *então existe algum* $K \in \mathbb{N}$ *tal que*

$$\alpha_k \geqslant \min\left\{\alpha_K, \frac{\theta}{L}\right\}, \quad \forall\ k > K. \tag{3.64}$$

Consequentemente, para $k > K$, *a Busca Linear LS3*$(x^k, \alpha_{k-1}, \theta)$ *precisa de no máximo* $\log_\theta(\min\{1, \frac{\theta}{\alpha_k L}\})$ *passos.*

Além disso, se ∇g *for globalmente Lipschitz contínua em* $\mathrm{int}(\mathrm{dom}(g)) \cap \mathrm{dom}(h)$ *com módulo uniforme* L, *então*

$$\alpha_k \geqslant \min\left\{\alpha_1, \frac{\theta}{L}\right\}, \quad \forall\ k > K.$$

Nesse caso, a Busca Linear LS3$(x^k, \alpha_{k-1}, \theta)$ *precisa de no máximo* $\log_\theta(\min\{1, \frac{\theta}{\sigma L}\})$ *passos para qualquer* k.

Demonstração. Para justificar, vamos supor que $S^* \neq \emptyset$, que a sequência (x^k) converge para algum $x^* \in S^*$ e que ∇g é Localmente Lipschitz contínua em torno de x^* (Ver Definição 1.3) com constante $L > 0$.

Seja algum $\varepsilon > 0$, tal que

$$\|\nabla g(x) - \nabla g(y)\| \leqslant L\|x - y\|,\ \forall\ x, y \in B_\varepsilon(x^*). \tag{3.65}$$

Onde $B_\varepsilon(x^*)$ é a bola fechada em $\mathbb{R}^n$ com centro em x^* e raio ε.

Como (x^k) converge para x^*, então existe $k \in \mathbb{N}$ tal que

$$\|x^k - x^*\| \leqslant \frac{\theta\varepsilon}{2+\theta} < \varepsilon, \quad \forall\ k > K. \tag{3.66}$$

Com $0 < \theta < 1$ definido na Busca Linear 3. Sendo assim, vamos assumir que

$$\alpha_k \geqslant \min\left\{\alpha_{k-1}, \frac{\theta}{L}\right\}, \quad \forall\ k > K. \tag{3.67}$$

Supondo por contradição, que

$$\alpha_k < \min\left\{\alpha_{k-1}, \frac{\theta}{L}\right\}.$$

Concluímos que, $\alpha_k < \alpha_{k-1}$. Dessa forma, o loop da Busca Linear 3 em (x^k, α_{k-1}) precisa de mais uma iteração. Sendo assim, seja $\hat{\alpha}_k := \frac{\alpha_k}{\theta}$ e $\hat{x}^k := J(x^k, \hat{\alpha}_k)$, temos de (3.61) que

$$g(\hat{x}^k) > g(x^k) + \left\langle \nabla g(x^k), \hat{x}^k - x^k \right\rangle + \frac{1}{2\hat{\alpha}_k}\|x^k - \hat{x}^k\|^2. \tag{3.68}$$

Além disso, pelo Lema (2.9), obtemos

$$\begin{aligned} \|x^k - \hat{x}^k\| = \|x^k - J(x^k, \hat{\alpha}_k)\| &\leqslant \frac{\hat{\alpha}_k}{\alpha_k}\|x^k - J(x^k, \alpha_k)\| \\ &= \frac{\frac{\alpha_k}{\theta}}{\alpha_k}\|x^k - x^{k+1}\| \\ &= \frac{1}{\theta}\|x^k - x^{k+1}\|. \end{aligned}$$

Sendo assim,

$$\|x^k - \hat{x}^k\| \leqslant \frac{1}{\theta}\|x^k - x^{k+1}\|.$$

Sabemos que pela desigualdade triangular, vale $\|\hat{x}^k - x^*\| \leqslant \|\hat{x}^k - x^k\| + \|x^k - x^*\|$. Daí e da desigualdade anterior, temos

$$\|\hat{x}^k - x^*\| \leqslant \frac{1}{\theta}\|x^k - x^{k+1}\| + \|x^k - x^*\|.$$

Utilizando a desigualdade triangular novamente, $\|x^k - x^{k+1}\| \leqslant \|x^k - x^*\| + \|x^* - x^{k+1}\|$, obtemos

$$\|\hat{x}^k - x^*\| \leqslant \frac{1}{\theta}(\|x^k - x^*\| + \|x^* - x^{k+1}\|) + \|x^k - x^*\|.$$

Utilizando (3.66) na desigualdade anterior, temos

$$\begin{aligned} \|\hat{x}^k - x^*\| &\leqslant \frac{1}{\theta} \cdot \left(\frac{\theta\varepsilon}{2+\theta} + \frac{\theta\varepsilon}{2+\theta}\right) + \frac{\theta\varepsilon}{2+\theta} \\ &= \frac{1}{\theta} \cdot \frac{2\theta\varepsilon}{2+\theta} + \frac{\theta\varepsilon}{2+\theta} = \frac{2\theta\varepsilon}{2\theta+\theta^2} + \frac{\theta^2\varepsilon}{2\theta+\theta^2} \\ &= \frac{\theta^2\varepsilon + 2\theta\varepsilon}{2\theta+\theta^2} = \frac{(\theta^2+2\theta)\varepsilon}{\theta^2+2\theta} = \varepsilon. \end{aligned}$$

Logo, $\|\hat{x}^k - x^*\| \leqslant \varepsilon$.

Dessa forma da desigualdade anterior e de (3.66) podemos concluir que $x^k, \hat{x}^k \in B_\varepsilon(x^*)$. Logo, qualquer ponto intermediário está dentro da bola, ou seja, a combinação convexa $x^k + t(\hat{x}^k - x^k) \in B_\varepsilon(x^*)$, com $t \in [0, 1]$. Aplicando o Teorema Fundamental do Cálculo em g, temos

$$g(\hat{x}^k) - g(x^k) = \int_0^1 \langle \nabla g(x^k + t(\hat{x}^k - x^k)), \hat{x}^k - x^k \rangle dt,$$

acrescentando $-\langle \nabla g(x^k), \hat{x}^k - x^k \rangle$ em ambos os lados da igualdade,

$$\begin{aligned} g(\hat{x}^k) - g(x^k) - \langle \nabla g(x^k), \hat{x}^k - x^k \rangle &= \int_0^1 \langle \nabla g(x^k + t(\hat{x}^k - x^k)), \hat{x}^k - x^k \rangle dt \\ &\quad - \langle \nabla g(x^k), \hat{x}^k - x^k \rangle \\ &= \int_0^1 \langle \nabla g\ (x^k + t(\hat{x}^k - x^k)) - \nabla g(x^k), \hat{x}^k - x^k \rangle dt. \end{aligned}$$

Aplicando Cauchy-Schwarz na integral,

$$\begin{aligned} g(\hat{x}^k) - g(x^k) - \langle \nabla g(x^k), \hat{x}^k - x^k \rangle &\leqslant \int_0^1 \|\nabla g(x^k + t(\hat{x}^k - x^k)) - \nabla g(x^k)\| \cdot \|\hat{x}^k - x^k\| dt \\ &= \int_0^1 \|\nabla g(t(\hat{x}^k - x^k))\| \cdot \|\hat{x}^k - x^k\| dt \\ &= \int_0^1 t\|\nabla g(\hat{x}^k) - \nabla g(x^k)\| \cdot \|\hat{x}^k - x^k\| dt. \end{aligned}$$

Utilizando a condição (3.65), temos

$$\begin{aligned} g(\hat{x}^k) - g(x^k) - \langle \nabla g(x^k), \hat{x}^k - x^k \rangle &\leqslant \int_0^1 tL \cdot \|\hat{x}^k - x^k\| \cdot \|\hat{x}^k - x^k\| dt \\ &= \int_0^1 tL\|\hat{x}^k - x^k\|^2 dt \\ &= \frac{L}{2}\|\hat{x}^k - x^k\|^2. \end{aligned}$$

O que por sua vez, implica que

$$g(\hat{x}^k) - g(x^k) - \left\langle \nabla g(x^k), \hat{x}^k - x^k \right\rangle \leqslant \frac{L}{2}\|\hat{x}^k - x^k\|^2.$$

Daí e de (3.68), concluímos

$$\frac{1}{2\hat{\alpha}_k}\|\hat{x}^k - x^k\|^2 \leqslant \frac{L}{2}\|\hat{x}^k - x^k\|^2,$$

como $\|x^k - \hat{x}^k\| \neq 0$, pois se fosse igual bateria a desigualdade da busca. Sendo assim, temos que $\frac{1}{2\hat{\alpha}_k} \leqslant \frac{L}{2}$, isto é $\hat{\alpha_k} \geqslant \frac{1}{L}$. Além disso, como $\hat{\alpha}_k := \frac{\alpha_k}{\theta}$, obtemos $\alpha_k \geqslant \frac{\theta}{L}$.

Chegamos em um absurdo, pois supomos que $\alpha_k < \frac{\theta}{L}$. Se existe algum $H > K$ com $H \in \mathbb{N}$ tal que $\alpha_H > \frac{\theta}{L}$, obtemos de (3.67) que $\alpha_k \geqslant \frac{\theta}{L}$ para todo $k \geqslant H$. Caso contrário, $\alpha_k < \frac{\theta}{L}$ para todo $k > K$, o que implica que $\alpha_k = \alpha_{k-1} = \alpha_K$ para todo $k > K$, devido a (3.67) e também à propriedade decrescente da sequência (α_k). Em ambos os casos, temos (3.64).

Agora, suponha que a Busca Linear LS3$(x^k, \alpha_{k-1}, \theta)$ precisa de d repetições. Ou seja, $\alpha_k = \alpha_{k-1}\theta^d$. Sendo assim, utilizando (3.64), temos que

$$\alpha_K \theta^d \geqslant \alpha_k = \alpha_{k-1}\theta^d \geqslant \min\left\{\alpha_K, \frac{\theta}{L}\right\},$$

o que por sua vez, implica que

$$\alpha_K \theta^d \geqslant \min\left\{\alpha_K, \frac{\theta}{L}\right\},$$

dividindo por $\alpha_K > 0$, temos $\theta^d \geqslant \min\left\{1, \frac{\theta}{\alpha_k L}\right\}$. Como ambos os lados da desigualdade são positivos, podemos aplicar logaritmo na base θ. Sendo assim, temos

$$\log_\theta(\theta^d) \leqslant \log_\theta\left(\min\left\{1, \frac{\theta}{\alpha_K L}\right\}\right),$$

lembrando que o sinal mudou, pois $0 < \theta < 1$. Utilizando identidade trigonométrica, temos que a desigualdade anterior, se torna

$$d \leqslant \log_\theta\left(\min\left\{1, \frac{\theta}{\alpha_K L}\right\}\right).$$

Vamos supor que ∇g seja globalmente Lipschitz com constante L, no int(dom(g)) $\cap$ dom(h) $\subset$ int(dom(g)). Usando a Proposição 2.2 (iii), podemos repetir o argumento acima sem nos preocuparmos com ε, k e substituir (3.67) por

$$\alpha_k \geqslant \min\left\{\sigma, \frac{\theta}{L}\right\}.$$

Sendo assim, sob a hipótese de que o gradiente de g é globalmente Lipschitz contínuo, a Busca Linear é bem definida e termina após um número finito de reduções em cada iteração. Além disso, os tamanhos de passo gerados pelo método admitem um limite inferior positivo uniforme, não se aproximando de zero ao longo do processo iterativo. □

Em seguida, apresentamos a complexidade assintótica $o(k^{-1})$ do método **MPG4** [[20], Teorema 3], mas sob as suposições consideradas nesse trabalho e considerando que ∇g é Localmente Lipschitz contínuo. A complexidade permanece válida ao substituir a condição local de Lipschitz por a de que o tamanho do passo (α_k) é limitado por um número positivo. Veja a Proposição 3.4. Esta ideia foi iniciada em [[9], Teorema 4.3] em dimensões finitas e estendida a diferentes tipos de Busca Linear em espaço de Hilbert em [[32], Corolário 3.20]. Agora, estudaremos a convergência sublinear do **MPG4**.

Proposição 3.5. *Seja (x^k) a sequência gerada no método **MPG4**. Suponha que $S^* \neq \emptyset$ e que ∇g é Localmente Lipschitz contínuo em torno de qualquer ponto em S^*. Então*

$$\lim_{k\to\infty} k[F(x^k) - \min_{x\in\mathbb{R}^n} F(x)] = 0.$$

Demonstração. Temos pela Proposição 3.4, que α_k é limitado inferiormente por algum $\alpha > 0$. Sendo assim, tomando $x^* \in S^*$ e fazendo a substituição de $x = x^*$, obtemos do Corolário 3.1 (i) que

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \geqslant 2\alpha[F(x^{k+1}) - F(x^*)] \geqslant 0. \tag{3.69}$$

Como $\|x^k - x^*\|^2$ é decrescente, temos que $(\|x^k - x^*\|^2)$ converge, já que a norma é decrescente e limitada inferiormente por 0. (Teorema da convergência monótona). Dado $\varepsilon > 0$, existe $K \in \mathbb{N}$ suficientemente grande, tal que $\|x^{K+k} - x^*\|^2 - \|x^K - x^*\|^2 \leqslant \varepsilon$, para qualquer $k \in \mathbb{N}$. Sendo assim, de (3.69), temos

$$\varepsilon \geqslant \|x^K - x^*\|^2 - \|x^{K+k} - x^*\|^2 \geqslant 2\alpha[F(x^{k+1}) - F(x^*)] \geqslant 0,$$

aplicando o somatório de $l = K$ a $(k + K - 1)$, temos a seguinte soma telescópica.

$$\begin{aligned}\varepsilon \geqslant \|x^k - x^*\|^2 - \|x^{k+K} - x^*\|^2 &= \sum_{l=K}^{k+K-1} (\|x^l - x^*\|^2 - \|x^{l+1} - x^*\|^2) \\ &\geqslant 2\alpha \sum_{l=K}^{k+K-1} [F(x^{l+1}) - F(x^*)].\end{aligned}$$

Como $(F(x^k))$ é monótona não-crescente, ou seja, $F(x^k) \geqslant F(x^{k+K})$, temos que

$$\varepsilon \geqslant 2\alpha k[F(x^{k+K}) - F(x^*)] = 2\alpha\frac{k}{(K+k)}(K+k)(F(x^{K+k}) - F(x^*)),$$

o que por sua vez, implica que

$$\varepsilon \geqslant 2k\alpha(F(x^{K+k}) - F(x^*)),$$

sendo assim,

$$k(F(x^{K+k}) - F(x^*)) \leqslant \frac{\varepsilon}{2\alpha},$$

aplicando lim sup com $k \to \infty$

$$\lim_{k\to\infty} \sup(K+k)(F(x^{K+k}) - F(x^*)) \leqslant \frac{\varepsilon}{2\alpha}.$$

Como K não afeta o limite, temos

$$\limsup_{k\to\infty} k(F(x^k) - F(x^*)) \leqslant \frac{\varepsilon}{2\alpha}.$$

E como $F(x^k) \geqslant F(x^*)$, o limite inferior é não negativo, logo

$$\begin{aligned}0 \leqslant \liminf_{k\to\infty} k(F(x^k) - F(x^*)) &\leqslant \limsup_{k\to\infty} k(F(x^k) - F(x^*)) \\ = \lim_{k\to\infty} \sup(K+k)(F(x^{K+k}) - F(x^*)) &\leqslant \frac{\varepsilon}{2\alpha}.\end{aligned}$$

Fazendo $\varepsilon > 0$ arbitrário tender a zero, pelo Teorema do Confronto, temos

$$\lim_{k\to\infty} k(F(x^k) - F(x^*)) = 0,$$

como $x^* \in S^*$, temos

$$\lim_{k\to\infty} k(F(x^k) - \min_{x\in\mathbb{R}^n} F(x)) = 0.$$

Como queríamos demonstrar. $\square$

Capítulo 4

Análise de Complexidade

O presente capítulo destina-se a abordar os resultados de análise de complexidade "pior-caso" dos métodos **MPG1, MPG2** e **MPG3** no que diz respeito aos valores funcionais. O estudo de complexidade pior-caso é importante na literatura relacionada pois ele oferece uma garantia teórica robusta sobre o desempenho máximo de um algoritmo, mesmo sob as condições mais adversas ou desafiadoras.

4.1 Análise de Complexidade do MPG1

Apresentamos a análise de complexidade do **MPG1**. Sob a suposição de que os comprimentos de passos gerados pelo procedimento de Busca Linear 1 são limitados inferiormente por um número positivo, a seguinte análise irá mostrar que o erro esperado do valor de custo na k-ésima iteração em relação ao valor ótimo é da ordem $o(k^{-1})$. Para maiores detalhes, o leitor também pode consultar [9].

Teorema 4.1. *Sejam* (x^k) *e* (α_k) *as sequências geradas pelo* ***MPG1****. Suponha que* $S^* \neq \emptyset$ *e que existe* $\alpha > 0$ *tal que* $\alpha_k \geqslant \alpha > 0$ *para todo* $k \in \mathbb{N}$*. Então, temos*

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x) \leqslant \frac{1}{2\alpha}\frac{[\mathtt{dist}(x^0, S^*)]^2}{k}, \quad \forall k \in \mathbb{N}. \tag{4.1}$$

Em particular,

$$\lim_{k \to \infty} k[(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x)] = 0. \tag{4.2}$$

Demonstração. Para qualquer $x^* \in S^*$, pela Proposição 3.1 (i) e do fato que $\delta \in (0, 1/2)$, temos:

$$
\begin{aligned}
0 \geqslant (g+h)(x^*) - (g+h)(x^{l+1}) &\geqslant \frac{1}{2\alpha_l}(\|x^{l+1}-x^*\|^2 - \|x^l - x^*\|^2 + (1-2\delta)\|x^l - x^{l+1}\|^2) \\
&\geqslant \frac{1}{2\alpha_l}(\|x^{l+1}-x^*\|^2 - \|x^l - x^*\|^2).
\end{aligned}
$$

Sendo assim,

$$
0 \geqslant (g+h)(x^*) - (g+h)(x^{l+1}) \geqslant \frac{1}{2\alpha_l}(\|x^{l+1}-x^*\|^2 - \|x^l - x^*\|)^2, \quad \forall l \in \mathbb{N}. \tag{4.3}
$$

Agora, por hipótese, supomos que $\alpha_l \geqslant \alpha$, então $\frac{1}{\alpha_l} \leqslant \frac{1}{\alpha}$. Por outro lado, foi provado anteriormente (veja Teorema 3.1 (i)) que a sequência gerada pelo método **MPG1** é Fejér convergente em relação ao conjunto solução S^*, em particular, podemos concluir que $\|x^{l+1}-x^*\|^2 - \|x^l - x^*\| \leqslant 0$, para todo $x^* \in S^*$ e $l \in \mathbb{N}$. Então,

$$
\frac{1}{\alpha_l}\left(\|x^{l+1}-x^*\|^2 - \|x^l - x^*\|\right) \geqslant \frac{1}{\alpha}\left(\|x^{l+1}-x^*\|^2 - \|x^l - x^*\|\right).
$$

Daí, pela desigualdade anterior e (4.3), temos que

$$
0 \geqslant (g+h)(x^*) - (g+h)(x^{l+1}) \geqslant \frac{1}{2\alpha}(\|x^{l+1}-x^*\|^2 - \|x^l - x^*\|)^2.
$$

Dessa forma, aplicando somatório na desigualdade acima de $l = 0, 1, 2, \cdots, k-1$, obtemos uma série telescópica no lado direito que nos fornece

$$
\sum_{l=0}^{k-1}[(g+h)(x^*) - (g+h)(x^{l+1})] \geqslant \frac{1}{2\alpha}(\|x^k - x^*\|^2 - \|x^0 - x^*\|^2),
$$

reorganizando, temos

$$
\sum_{l=0}^{k-1}[(g+h)(x^{l+1}) - (g+h)(x^*)] \leqslant \frac{1}{2\alpha}(\|x^0 - x^*\|^2 - \|x^k - x^*\|^2). \tag{4.4}
$$

Agora, observe que pela Proposição 3.1 (ii), vale $(g+h)(x^{l+1}) \leqslant (g+h)(x^l)$ para todo $l \leqslant k-1$. Em particular,

$$
\sum_{l=0}^{k-1}(g+h)(x^{l+1}) \geqslant k(g+h)(x^k).
$$

Da equação (4.4) e da desigualdade anterior, temos

$$
k[(g+h)(x^k) - (g+h)(x^*)] \leqslant \frac{1}{2\alpha}(\|x^* - x^0\|^2 - \|x^k - x^*\|^2),
$$

como $\|x^k - x^*\|^2 \geqslant 0$, então a desigualdade anterior, se torna

$$
k[(g+h)(x^k) - (g+h)(x^*)] \leqslant \frac{1}{2\alpha}(\|x^* - x^0\|^2). \tag{4.5}
$$

Podemos perceber que não importa como escolhemos $x^* \in S^*$, o valor ótimo $(g+h)(x^*) = \min_{x \in \mathbb{R}^n}(f+g)(x)$ é fixo. Sendo assim, utilizando (4.5) e a definição de distancia de um ponto a um conjunto, temos

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x) \leqslant \frac{1}{2\alpha} \inf_{y \in S^*} \frac{\|y - x^0\|^2}{k} = \frac{1}{2\alpha} \frac{[\text{dist}(x^0, S^*)]^2}{k},$$

o que por sua vez, implica que

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x) \leqslant \frac{1}{2\alpha} \frac{[\text{dist}(x^0, S^*)]^2}{k},$$

provando assim, (4.1). Agora, vamos provar a segunda parte do Teorema. Vimos anteriormente, pelo Teorema 3.1 que a sequência (x^k) converge para algum $x^* \in S^*$, isto é, $\|x^k - x^*\| \to 0$ quando $k \to \infty$. Etão, dado $\varepsilon > 0$, existe $K \in \mathbb{N}$ tal que $\|x^k - x^*\| < \varepsilon$ para todo $k \geqslant K$. Para qualquer $l \geqslant K$ obtemos da Proposição 3.1 (i) que $\|x^l - x^*\| \geqslant \|x^{l+1} - x^*\|$ juntamente com (4.3) que

$$\begin{aligned}
0 \geqslant (g+h)(x^*) - (g+h)(x^{l+1}) &\geqslant \frac{1}{2\alpha_l}\left(\|x^{l+1} - x^*\| - \|x^l - x^*\|\right)^2 \\
&= \frac{1}{2\alpha_l}\left(\|x^{l+1} - x^*\| + \|x^l - x^*\|\right)\left(\|x^{l+1} - x^*\| - \|x^l - x^*\|\right) \\
&\geqslant \frac{1}{2\alpha_l} 2\|x^{l+1} - x^*\|\left(\|x^{l+1} - x^*\| - \|x^l - x^*\|\right) \\
&= \frac{1}{\alpha_l}\|x^{l+1} - x^*\|(\|x^{l+1} - x^*\| - \|x^l - x^*\|) \\
&\geqslant \frac{\varepsilon}{\alpha}(\|x^{l+1} - x^*\| - \|x^l - x^*\|).
\end{aligned}$$

Dessa forma,

$$0 \geqslant (g+h)(x^*) - (g+h)(x^{l+1}) \geqslant \frac{\varepsilon}{\alpha}(\|x^{l+1} - x^*\| - \|x^l - x^*\|).$$

Aplicando somatório na desigualdade acima de $l = K, K+1, \ldots, K+k-1.$, obtemos uma série telescópica no lado direito que nos fornece

$$\sum_{l=K}^{K+k-1} [(g+h)(x^*) - (g+h)(x^{l+1})] \geqslant \frac{\varepsilon}{\alpha}(\|x^{K+k} - x^*\| - \|x^K - x^*\|), \tag{4.6}$$

como $(g+h)(x^l) \geqslant (g+h)(x^{l+1})$ para $l \leqslant K + k - 1$, temos

$$\sum_{l=K}^{K+k-1} (g+h)(x^{l+1}) \geqslant k(g+h)(x^{K+k}).$$

Sendo assim, utilizando essa informação em (4.6), encontramos

$$\begin{aligned}
k[(g+h)(x^*) - (g+h)(x^{K+k})] &\geqslant \frac{\varepsilon}{\alpha}(\|x^{K+k} - x^*\| - \|x^K - x^*\|) \\
&\geqslant -\frac{\varepsilon}{\alpha}\|x^K - x^*\|.
\end{aligned}$$

Como sabemos que $\|x^K - x^*\| < \varepsilon$, então

$$k[(g+h)(x^*) - (g+h)(x^{K+k})] \geqslant -\frac{\varepsilon^2}{\alpha},$$

o que por sua vez, implica que

$$k[(g+h)(x^{K+k}) - (g+h)(x^*)] \leqslant \frac{\varepsilon^2}{\alpha}.$$

Agora, multiplicando por $\frac{K+k}{k} > 0$, em ambos os lados da desigualdade anterior, temos

$$(K+k)[(g+h)(x^{K+k}) - (g+h)(x^*)] \leqslant \frac{K+k}{k} \cdot \frac{\varepsilon^2}{\alpha},$$

tomando $\limsup$ com $k \to +\infty$, temos

$$\limsup_{k\to+\infty} (K+k)[(g+h)(x^{K+k}) - (g+h)(x^*)] \leqslant \limsup_{k\to+\infty} \frac{K+k}{k} \cdot \frac{\varepsilon^2}{\alpha},$$

como temos K fixo, o $\limsup$ ao longo de $K+k$ é igual ao $\limsup$ ao longo de k, sendo assim

$$\begin{aligned}
\limsup_{k\to+\infty} k[(g+h)(x^k) - (g+h)(x^*)] &= \limsup_{k\to+\infty} (K+k)[(g+h)(x^{K+k}) - (g+h)(x^*)] \\
&\leqslant \limsup_{k\to+\infty} \frac{K+k}{k} \cdot \frac{\varepsilon^2}{\alpha} = \frac{\varepsilon^2}{\alpha},
\end{aligned}$$

como $\varepsilon > 0$ é arbitrário, vale

$$\limsup_{k\to+\infty} k[(g+h)(x^k) - (g+h)(x^*)] \leqslant 0.$$

Temos também que $(g+h)(x^k) \geqslant (g+h)(x^*)$, pois $x^* \in S^*$. Sendo assim, $\liminf_{k\to\infty} k[(g+h)(x^k) - (g+h)(x^*)] \geqslant 0$, dessa forma

$$0 \leqslant \liminf_{k\to\infty} k[(g+h)(x^k) - (g+h)(x^*)] \leqslant \limsup_{k\to+\infty} k[(g+h)(x^k) - (g+h)(x^*)] \leqslant 0.$$

Dessa forma,

$$\lim_{k\to\infty} k[(g+h)(x^k) - (g+h)(x^*)] = 0,$$

o que por sua vez, implica que

$$\lim_{k\to\infty} k[(g+h)(x^k) - \min_{x\in\mathbb{R}^n} (g+h)(x)] = 0.$$

O que verifica (4.2) e completa a demonstração do Teorema. $\square$

Vale mencionar que a taxa $o(k^{-1})$ foi obtida em [21, 22, 23]. Anteriormente ao utilizar o método do ponto proximal para resolver o Problema 2.1 com $g \equiv 0$. Nosso resultado acima poderia ser considerado uma extensão de alguns resultados desses artigos, em particular, [[22], Corolário 3.1], para o quadro mais geral do (2.1) com Busca Linear. Quando os tamanhos dos passos não são limitados inferiormente por uma constante positiva, discutimos a possível validade da mesma complexidade da seguinte forma.

Observação 4.1. *Surge a seguinte questão do Teorema acima: podemos ter a complexidade* $o(k^{-1})$ *da diferença* $(g+h)(x^k) - \min_{x\in\mathbb{R}^n}(g+h)(x)$ *quando* $\liminf_{k\to\infty} = 0$? *Suponha que* (x^k) *converge para algum* $x^* \in S^*$. *Veja o Teorema 3.1. Analisando cuidadosamente a demostração de* (4.1) *presente no Teorema* (4.1), *observamos que a complexidade de* $o(k^{-1})$ *permanece quando a seguinte condição é válida: Existe* $\lambda \in [-1, 1)$ *tal que*

$$\limsup_{k\to\infty} \frac{\|x^k - x^*\|^{1+\lambda}}{\alpha_k} < +\infty, \tag{4.7}$$

o que pode permitir que α_k *se aproxime de 0.*

Temos que (4.7) vale quando α_k é limitado inferiormente por um número positivo. O exemplo simples a seguir mostra a possível validade de (4.7) mesmo quando $\alpha_k \to 0$ à medida que $k \to \infty$. Assim, a complexidade $o(k^{-1})$ dos valores da função permanece verdadeira no exemplo abaixo. No entanto, em geral, verificar (4.7) pode não ser trivial, uma vez que x^* é desconhecido.

Exemplo 4.1. *Seja:*

$$g(x) = \frac{1}{1+p}|x|^{1+p} \quad \text{com} \quad 0 < p < 1 \quad e \quad h(x) = \delta_{[0,\infty)}(x).$$

Observe que nesse caso o problema (2.1) *tem solução única, sendo está* $x^* = 0$. *Podemos perceber também que para* $x > 0$, *temos que esse problema se resume a um problema restrito no qual recorremos a um problema irrestrito do tipo*

$$\min_{x\in\mathbb{R}^n} h(x) + \delta_{[0,\infty)}(x).$$

Sendo assim, temos

$$\begin{aligned} J(x,\alpha) = \operatorname{prox}_{\alpha h}(x - \alpha\nabla g(x)) &= \arg\min_{y\in\mathbb{R}^n} h(y) + \frac{1}{2\alpha}\|y - (x - \alpha\nabla g(x))\|^2 \\ &= \arg\min_{y\in\mathbb{R}^n} \delta_{[0,\infty)}(y) + \frac{1}{2\alpha}\|y - (x - \alpha\nabla g(x))\|^2, \end{aligned}$$

como $x > 0$*, então*

$$\nabla g(x) = g'(x) = \frac{1+p}{1+p}|x|^{(1+p)-1} = |x|^p = x^p, \tag{4.8}$$

dessa forma, a igualdade anterior se torna

$$\begin{aligned} J(x,\alpha) &= \arg\min_{y\in\mathbb{R}^n} \delta_{[0,\infty)}(y) + \frac{1}{2\alpha}\|y-(x-\alpha x^p)\|^2 \\ &= \arg\min_{y\in[0,+\infty)} \frac{1}{2\alpha}\|y-(x-\alpha x^p)\|^2 \\ &= P_{[0,\infty)}(x-\alpha x^p) \\ &= \max\{x-\alpha x^p, 0\}. \end{aligned}$$

ou seja,

$$x_{k+1} = J(x,\alpha) = \max\{x-\alpha x^p, 0\}. \tag{4.9}$$

Para que não haja confusão entre a sequência do **MPG1** *e a da iteração, vamos escrever* (x_k) *em vez de* (x^k)*. Com o intuito de evitar o caso trivial, vamos supor que* $x_k > 0$ *para todo* $k \in \mathbb{N}$*. Pois, caso contrário se* $x_k = 0$*, temos*

$$\begin{aligned} x_{k+1} = J(x_k,\alpha_k) &= max\{x_k - \alpha_k {x_k}^p, 0\} \\ &= \max\{0-\alpha_0 0^p, 0\} \\ &= \max\{-\alpha_0, 0\} \\ &= 0. \end{aligned}$$

Dessa forma, a sequência estacionaria em zero, para toda iteração seguinte, o que acaba sendo trivial, pois a solução ótima $x^* = 0$*, já foi alcançada e o algoritmo não precisaria continuar. Sendo assim, supondo que* $x_k > 0$*, temos*

$$0 < x_{k+1} = x_k - \alpha_k x_k^p < x_k,$$

pois, $\alpha_k > 0$ *e* $x_k > 0$*. Sendo assim, utilizando o procedimento de Busca Linear 1* (2.11) *com* (4.8)*, obtemos*

$$\alpha_k\|(J(x,\alpha))^p - {x_k}^p\| \leqslant \delta\|J(x,\alpha) - x_k\|,$$

o que por sua vez, implica que

$$\alpha_k\|x_{k+1}^p - x_k^p\| \leqslant \delta\|x_{k+1} - x_k\|. \tag{4.10}$$

Aplicando o Teorema do Valor Médio na função auxiliar $\varphi(t) = t^p$ *no intervalo entre* x_{k+1} *e* x_k*. Utilizando uma combinação convexa entre os pontos* x_k *e* x_{k+1}*, podemos concluir que existe* $\eta \in [0,1]$ *e* $c = \eta x_{k+1} + (1-\eta)x_k$*, tal que*

$$\varphi(x_{k+1}) - \varphi(x_k) = \varphi'(c)(x_{k+1} - x_k),$$

dessa forma, aplicando a norma euclidiana em ambos os lados da igualdade anterior, temos

$$|x_{k+1}^p - x_k^p| = |x_{k+1} - x_k| \cdot p|\eta(x_{k+1}) + (1-\eta)x_k|^{p-1},$$

reorganizando, temos

$$|x_{k+1}^p - x_k^p| = |x_{k+1} - x_k| \cdot p|\eta(x_{k+1} - x_k) + x_k|^{p-1}, \tag{4.11}$$

como $|\eta(x_{k+1} - x_k) + x_k| \geqslant |x_k|$ *temos que,*

$$|x_{k+1} - x_k| \cdot p \cdot |\eta(x_{k+1} - x_k) + x_k|^{p-1} \geqslant |x_{k+1} - x_k| \cdot p \cdot |x_k|^{p-1}.$$

Utilizando (4.11) *e a desigualdade anterior, temos*

$$|x_{k+1}^p - x_k^p| \geqslant |x_{k+1} - x_k| \cdot p \cdot |x_k|^{p-1}.$$

Daí, e de (4.10)*, temos que*

$$\delta|x_{k+1} - x_k| \geqslant \alpha_k|x_{k+1}^p - x_k^p| \geqslant \alpha_k|x_{k+1} - x_k| \cdot p \cdot |x_k|^{p-1},$$

o que por sua vez, implica que

$$\delta|x_{k+1} - x_k| \geqslant \alpha_k|x_{k+1} - x_k| \cdot p \cdot |x_k|^{p-1}.$$

Como vimos anteriormente $x_{k+1} < x_k$*. Sendo assim, podemos dividir a desigualdade anterior por* $|x_{k+1} - x_k| \neq 0$*, obtendo assim,* $\delta \geqslant \alpha_k \cdot p \cdot |x_k|^{p-1}$ *o que por sua vez, pode ser reescrita como,* $\alpha_k \leqslant \delta \cdot p^{-1}|x_k|^{1-p}$*. Supondo que* $x_k \to 0$*, quando* $k \to \infty$ *podemos concluir que* $\alpha_k \leqslant \delta \cdot p^{-1}|x_k|^{1-p} \to 0$*, quando* $k \to \infty$*, logo* $\alpha_k \to 0$*. Dessa forma, podemos supor sem perca de generalidade que* $\alpha_k < \sigma$*.*

Vamos definir $\hat{\alpha}_k := \frac{\alpha_k}{\theta}$ *e* $\hat{x}_{k+1} = J(x_k, \hat{\alpha}_k)$*. Sendo assim, utilizando a Busca Linear 1* (2.11)*, obtemos*

$$\hat{\alpha_k}|\nabla g(J(x_k, \hat{\alpha}_k) - \nabla g(x)| > \delta|J(x_k, \hat{\alpha}_k) - x_k|.$$

Lembrando que $\nabla g(x) = x^p$, *temos*

$$\hat{\alpha_k}|(J(x_k, \hat{\alpha}_k))^p - x_k^p| > \delta|J(x_k, \hat{\alpha}_k) - x_k|,$$

o que por sua vez, implica que

$$\hat{\alpha_k}|\hat{x}_{k+1}^p - x_k^p| > \delta|\hat{x}_{k+1} - x_k|. \tag{4.12}$$

Utilizando (4.9), *temos que* $0 \leqslant \hat{x}_{k+1} < x_k$, *pois*

$$J(x_k, \hat{\alpha}_k) = J\left(x_k, \frac{\alpha_k}{\theta}\right) = max\{x_k - \hat{\alpha_k} x_k^p, 0\}.$$

Sendo assim, temos dois casos possíveis para o valor de $J(x_k, \hat{\alpha}_k)$.

Caso 1: $J(x_k, \hat{\alpha}_k) = 0$,

$$\begin{aligned} 0 &\leqslant \hat{x}_{k+1} = J(x_k, \hat{\alpha}_k) = max\{x_k - \hat{\alpha}_k(x_k)^p, 0\} = 0 \\ 0 &\leqslant \hat{x}_{k+1} = 0 < x_k, \end{aligned}$$

dessa forma, $0 \leqslant \hat{x}_{k+1} < x_k$.

Caso 2: $J(x_k, \hat{\alpha}_k) = x_k - \hat{\alpha_k} x_k^p$,

temos que $\hat{\alpha}_k = \frac{\alpha_k}{\theta} > 0$ *e* $x_k > 0$ *então*

$$0 \leqslant \hat{x}_{k+1} < x_k.$$

Supondo que $\hat{x}_{k+1} > 0$, *temos* $\hat{x}_{k+1} < x_k$ *elevando a potencia p, temos* $\hat{x}_{k+1}^p < x_k^p$ *dessa forma,*

$$x_k^p - \hat{x}_{k+1}^p > 0. \tag{4.13}$$

como $\hat{x}_{k+1} < x_k$, *podemos concluir que,* $\frac{1}{\hat{x}_{k+1}} > \frac{1}{x_k}$, *dessa forma* $\frac{1}{\hat{x}_{k+1}^{1-p}} > \frac{1}{x_k^{1-p}}$, *o que por sua vez, implica que* $\hat{x}_{k+1}^{p-1} > x_k^{p-1}$, *dividindo a última desigualdade por* $\hat{x}_{k+1}$, *temos*

$$\hat{x}_{k+1}^p > \hat{x}_k^{p-1}\hat{x}_{k+1}.$$

Daí e de (4.13), *obtemos*

$$0 < x_k^p - \hat{x}_{k+1}^p < x_k^p - \hat{x}_k^{p-1}\hat{x}_{k+1} = x_k^{p-1}(x_k - \hat{x}_{k+1}),$$

multiplicando por $\hat{\alpha}_k > 0$ *e usando* (4.12),

$$\hat{\alpha}_k|x_k|^{p-1}|x_k - \hat{x}_{k+1}| > \hat{\alpha}_k|\hat{x}_k^p - \hat{x}_{k+1}^p| > \delta|\hat{x}_{k+1} - x_k|,$$

dessa forma,

$$\hat{\alpha}_k|x_k|^{p-1}|x_k - \hat{x}_{k+1}| > \delta|\hat{x}_{k+1} - x_k|,$$

como $|\hat{x}_{k+1} - x_k| \neq 0$, *pois* $\hat{x}_{k+1} < x_k$, *podemos dividir a desigualdade anterior por* $|\hat{x}_{k+1} - x_k|$, *sendo assim, obtemos que* $\delta < \hat{\alpha}_k|x_k|^{p-1}$ *como* $\hat{\alpha}_k = \frac{\alpha_k}{\theta}$, *temos* $\frac{\alpha_k}{\theta} > \frac{\delta}{|x_k|^{p-1}}$, *reorganizando, obtemos* $\frac{1}{\theta} > \frac{\delta}{|x_k|^{p-1}\alpha_k}$, *o que por sua vez, implica que* $\frac{1}{\theta} > \frac{\delta|x_k|^{1-p}}{\alpha_k}$. *Temos por hipótese que* $x^* = 0$ *e pela Busca Linear 1 que* $\delta \in (0, \frac{1}{2})$ *dessa forma,*

$$\frac{1}{\theta} > \frac{|x_k - x^*|^{1-p}}{\alpha_k}.$$

Fazendo $\lambda = -p \in [-1.1)$ *e passando* $\limsup$ *quando* $k \to +\infty$ *na desigualdade anterior, obtemos* (4.7)

$$\limsup_{k\to+\infty} \frac{|x_k - x^*|^{(1+\lambda)}}{\alpha_k} < +\infty.$$

Como queríamos demonstrar.

Proposição 4.1. *Seja* (α_k) *a sequência gerada pelo* ***MPG1***. *As seguintes afirmações são verdadeiras:*

(i) Se o gradiente de g é globalmente Lipschitz continuo (1.3) *em dom(h) com constante* $L > 0$, *então* $\alpha_k \geqslant \min\{\sigma, \frac{\delta\theta}{L}\}$ *para todo* $k \in \mathbb{N}$.

(ii) Suponha que $S^* \neq 0$. *Se* $\nabla g(x)$ *é localmente Lipschitz contínuo em qualquer* $x \in S^*$, *então existe* $x^* \in S^*$ *tal que*

$$\liminf_{k\to+\infty} \alpha_k \geqslant \min\left\{\sigma, \frac{\delta\theta}{L}\right\},$$

onde $L > 0$, *é uma constante de Lipschitz de* ∇g *ao redor de* x^*. *Consequentemente, existe* $\alpha > 0$ *tal que* $\alpha_k \geqslant \alpha$, $\forall k \in \mathbb{N}$.

Demonstração. Vamos supor que ∇g seja globalmente Lipschitz continuo com constante $L > 0$. Se $\alpha_k < \sigma$, vamos definir $\hat{\alpha}_k := \frac{\alpha_k}{\theta} > 0$ e $\hat{x}^k := J(x^k, \hat{\alpha}_k)$. Combinando essas informações com a Busca Linear 1 (2.11), obtemos

$$\hat{\alpha}_k\|\nabla g(\hat{x}^k) - \nabla g(x^k)\| > \delta\|\hat{x}^k - x^k\|. \tag{4.14}$$

Observe que $\|\hat{x}^k - x^k\| \neq 0$. Pois, caso contrário teríamos $\hat{x}^k = x^k$, no qual fazendo estas substituições em (2.7), implicaria que $0 \in \nabla g(x^k) + \partial h(x^k)$, ou seja, $x^k \in S^*$, o que contradiz a falha da condição da Busca Linear para o passo $\hat{\alpha}_k$. Além disso, como ∇g é globalmente Lipschitz continuo com $L > 0$, temos

$$\|\nabla g(x^k) - \nabla g(\hat{x}^k)\| \leqslant L\|x^k - \hat{x}^k\|, \quad \forall k \in \mathbb{N}.$$

Daí, e de (4.14), temos

$$\delta\|\hat{x}^k - x^k\| < \hat{\alpha}_k\|\nabla g(\hat{x}^k) - \nabla g(x^k)\| \leqslant \hat{\alpha}_k L\|x^k - \hat{x}^k\|,$$

sendo assim,

$$\delta\|\hat{x}^k - x^k\| < \hat{\alpha}_k L\|x^k - \hat{x}^k\|,$$

como $\|\hat{x}^k - x^k\| \neq 0$, então podemos concluir que $\delta < \hat{\alpha}_k L$. Temos que $\hat{\alpha}_k := \frac{\alpha_k}{\theta} > 0$, sendo assim $\delta < \frac{\alpha_k L}{\theta}$, o que por sua vez, implica que $\alpha_k \geqslant \frac{\delta\theta}{L}$, quando $\alpha_k < \sigma$. Logo,

$$\alpha_k \geqslant \min\left\{\sigma, \frac{\delta\theta}{L}\right\}.$$

Como queríamos verificar no item (i).

Para provar o item (ii) vamos supor que $S^* \neq \emptyset$ e que g é localmente Lipschitz contínua em qualquer ponto de S^*. Pelo Teorema 3.1, (x^k) converge para algum $x^* \in S^*$. Como temos que ∇g é localmente Lipschitz continuo em x^*, então existem $\varepsilon, L > 0$ tal que

$$\|\nabla g(x) - \nabla g(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in B_\varepsilon(x^*). \tag{4.15}$$

Onde $B_\varepsilon(x^*)$ é a esfera fechada em $\mathbb{R}^n$ com centro x^* e raio ε. Dessa forma, como (x^k) converge para x^*, então existe $K \in \mathbb{N}$ tal que

$$\|x^k - x^*\| \leqslant \frac{\theta\varepsilon}{2+\theta} < \varepsilon, \quad \forall k > K. \tag{4.16}$$

Com $\theta \in (0,1)$ definido anteriormente na Busca linear 1 (2.11). Seja $k > K$ qualquer, se $\alpha_k < \sigma$, analogamente à primeira parte, definimos $\hat{\alpha}_k := \frac{\alpha_k}{\theta} > 0$ e $\hat{x}_k := J(x^k, \hat{\alpha}_k)$. Sendo assim, fazendo estas substituições na Busca linear 1, obtemos novamente (4.14). Utilizando o Lema 2.2, temos que

$$\|x^k - J(x^k, \hat{\alpha}_k)\| \leqslant \frac{\hat{\alpha}_k}{\alpha_k}\|x^k - J(x^k, \alpha_k)\|,$$

dessa forma,

$$\|x^k - \hat{x}^k\| \leqslant \frac{1}{\theta}\|x^k - J(x^k, \alpha_k)\| = \frac{1}{\theta}\|x^k - x^{k+1}\|,$$

o que por sua vez, implica que

$$\|x^k - \hat{x}^k\| \leqslant \frac{1}{\theta}\|x^k - x^{k+1}\|.$$

Somando $\|x^k - x^*\|$ em ambos os lados da desigualdade anterior, obtemos

$$\|x^k - \hat{x}^k\| + \|x^k - x^*\| \leqslant \frac{1}{\theta}\|x^k - x^{k+1}\| + \|x^k - x^*\|.$$

Daí, utilizando desigualdade triangular, temos

$$\|\hat{x}^k - x^*\| \leqslant \|\hat{x}^k - x^k\| + \|x^k - x^*\| \leqslant \frac{1}{\theta}\|x^k - x^{k+1}\| + \|x^k - x^*\|, \tag{4.17}$$

Aplicando a desigualdade triangular em $\|x^k - x^{k+1}\|$ e utilizando (4.16), temos

$$\begin{aligned}\|x^k - x^{k+1}\| &\leqslant \|x^k - x^*\| + \|x^* - x^{k+1}\| \\ &\leqslant \frac{\theta\varepsilon}{2+\theta} + \frac{\theta\varepsilon}{2+\theta}.\end{aligned}$$

Pois, (4.16) vale para todo $k \in K$. Sendo assim, vale para $k = k+1$. Dessa forma, utilizando a desigualdade anterior e (4.17), obtemos

$$\begin{aligned}\|\hat{x}^k - x^*\| \leqslant \|\hat{x}^k - x^k\| + \|x^k - x^*\| &\leqslant \frac{1}{\theta}\cdot\frac{2\theta\varepsilon}{2+\theta} + \frac{\theta\varepsilon}{2+\theta} \\ &= \frac{2\varepsilon}{2+\theta} + \frac{\theta\varepsilon}{2+\theta} \\ &= \frac{\varepsilon(2+\theta)}{2+\theta} = \varepsilon.\end{aligned}$$

Logo, $\|\hat{x}^k - x^*\| \leqslant \varepsilon$. Ou seja, $\hat{x}^k \in B_\varepsilon(x^*)$. Temos de (4.16) que $\|x^k - x^*\| < \varepsilon$, sendo assim, $x^k \in B_\varepsilon(x^*)$. Dessa forma, de (4.15), como $\hat{x}^k$ e $x^k \in B_\varepsilon(x^*)$, vale

$$\|\nabla g(x^k) - \nabla g(\hat{x}^k)\| \leqslant L\,\|x^k - \hat{x}^k\|.$$

Daí, e de (4.14), obtemos

$$\delta\|\hat{x}^k - x^k\| < \hat{\alpha}_k\|\nabla g(x^k) - \nabla g(\hat{x}^k)\| \leqslant \hat{\alpha}_k L\|x^k - \hat{x}^k\|,$$

o que por sua vez, implica que

$$\delta\|\hat{x}^k - x^k\| < \hat{\alpha}_k L\|x^k - \hat{x}^k\|.$$

Como vimos anteriormente $\|\hat{x}^k - x^k\| \neq 0$, dessa forma, utilizando a desigualdade anterior temos $\delta < \hat{\alpha}_k L$. Como $\hat{\alpha}_k = \frac{\alpha_k}{\theta}$, concluímos que $\alpha_k \geqslant \frac{\theta\delta}{L}$. Consequentemente, como $\alpha_k < \sigma$, obtemos $\alpha_k \geqslant \min\left\{\sigma, \frac{\theta\delta}{L}\right\}, \quad \forall k > K$. Dessa forma,

$$\liminf_{k\to\infty} \alpha_k \geqslant \min\left\{\sigma, \frac{\theta\delta}{L}\right\}.$$

Como queríamos, agora como $\alpha_k > 0$ para $k \in \mathbb{N}$, temos que

$$\alpha_k \geqslant \alpha := \min\left\{\alpha_1, \ldots, \alpha_K, \frac{\theta\delta}{L}, \sigma\right\} > 0.$$

Dessa forma, $\alpha_k \geqslant \alpha$, para todo $k \in \mathbb{N}$. Concluímos assim, a última parte da proposição.

$\square$

Vale lembrar que a suposição da Proposição 4.1 (i) de que ∇g é globalmente Lipschitz contínua em $\mathrm{dom}(h)$ também é suficiente para a Hipótese A2. As suposições da Proposição 4.1 (ii) certamente não são suficientes para garantir a Hipótese A2. No entanto, existem muitas classes amplas de funções que satisfazem todas elas. Por exemplo, quando dom(h) é fechado, uma função g, que é diferenciável com gradiente localmente Lipschitz contínuo em dom(h), satisfaz todos os requisitos; veja também a Proposição 2.1.

O Teorema 4.1, junto com a Proposição 4.1 e os Teorema 4.2, nos levam ao seguinte resultado.

Corolário 4.1. *Seja (x^k) a sequência gerada pelo **MPG1**. Adicionalmente, suponha que $S^* \neq \emptyset$.*

(i) Se o gradiente de g for globalmente Lipschitz contínuo em dom(h), então temos

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x) = O(k^{-1}).$$

(ii) Se o gradiente de g é localmente Lipschitz contínuo em S^*, então temos

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x) = o(k^{-1}).$$

Obtemos convergência linear quando os tamanhos de passo são limitados inferiormente por um número positivo e ou g ou h é fortemente convexa (ver, Observação A.1).

Teorema 4.2. *Sejam (x^k) e (α_k) as sequências geradas pelo **MPG1**. Suponha que $S^* \neq \emptyset$, que existe $\alpha > 0$ satisfazendo $\alpha_k \geqslant \alpha > 0$ para todo $k \in \mathbb{N}$, e que g ou h é fortemente convexa com constante $\mu > 0$. Então $S^* = \{x^*\}$ é um conjunto unitário e*

$$\|x^{k+1} - x^*\| \leqslant \frac{1}{\sqrt{1+\alpha\mu}}\|x^k - x^*\| \leqslant \left(\frac{1}{\sqrt{1+\alpha\mu}}\right)^{k+1}\|x^0 - x^*\|, \quad \forall k \in \mathbb{N}. \tag{4.18}$$

Ou seja, a sequência (x^k) converge para x^* com a taxa linear $\dfrac{1}{\sqrt{1+\alpha\mu}} < 1$. Consequentemente, se g ou h é fortemente convexa e ∇g é Localmente Lipschitz contínua em S^*, então (x^k) converge linearmente para a única solução ótima.

Demonstração. Temos que g ou h são funções fortemente convexas com constante $\mu > 0$, então a soma $g + h$ também é fortemente convexa com constante $\mu > 0$, em particular $g + h$ é estritamente convexa. Logo, temos também que S^* é um conjunto unitário (isto é, $S^* = \{x^*\}$). Além disso, usando a Proposição 3.1 (i) com $x = x^*$, temos que

$$\|x^k - x^*\|^2 \geqslant \|x^{k+1} - x^*\|^2 + 2\alpha_k[(g + h)(x^{k+1}) - (g + h)(x^*)].$$

Agora, utilizando a convexidade forte de $g + h$, o qual nos permite concluir que temos $(g + h)(x^{k+1}) - (g + h)(x^*) \geqslant \frac{\mu}{2}\|x^{k+1} - x^*\|^2$. Dessa forma, subtituindo na desigualdade anterior, obtemos

$$\|x^k - x^*\|^2 \geqslant \|x^{k+1} - x^*\|^2 + \alpha_k\mu\|x^{k+1} - x^*\|^2,$$

o que por sua vez, implica que

$$\|x^k - x^*\|^2 \geqslant (1 + \alpha\mu)\|x^{k+1} - x^*\|^2.$$

Reorganizando e utilizanto o fato que $\alpha_k \geqslant \alpha$, temos

$$\|x^{k+1} - x^*\|^2 \leqslant \frac{\|x^k - x^*\|^2}{1 + \alpha\mu}, \quad \alpha, \mu > 0.$$

Aplicando raiz quadrada de ambos os lados, concluímos que

$$\|x^{k+1} - x^*\| \leqslant \frac{1}{\sqrt{1 + \alpha\mu}}\|x^k - x^*\|. \tag{4.19}$$

Temos que para $k = 0$ em (4.19), obtemos

$$\|x^1 - x^*\| \leqslant \frac{1}{\sqrt{1 + \alpha\mu}}\|x^0 - x^*\|.$$

Para $k = 1$,

$$\begin{aligned}
\|x^2 - x^*\| &\leqslant \frac{1}{\sqrt{1 + \alpha\mu}}\|x^1 - x^*\| \\
&\leqslant \frac{1}{\sqrt{1 + \alpha\mu}}\left(\frac{1}{\sqrt{1 + \alpha\mu}}\|x^0 - x^*\|\right) \\
&\leqslant \left(\frac{1}{\sqrt{1 + \alpha\mu}}\right)^2\|x^0 - x^*\|.
\end{aligned}$$

Para $k = 2$,

$$\begin{aligned}
\|x^3 - x^*\| &\leqslant \frac{1}{\sqrt{1 + \alpha\mu}}\|x^2 - x^*\| \\
&\leqslant \frac{1}{\sqrt{1 + \alpha\mu}}\left(\left(\frac{1}{\sqrt{1 + \alpha\mu}}\right)^2\|x^0 - x^*\|\right) \\
&\leqslant \left(\frac{1}{\sqrt{1 + \alpha\mu}}\right)^3\|x^0 - x^*\|, \\
&\vdots
\end{aligned}$$

dessa forma, em $k = k + 1$, obtemos a partir de (4.19) que

$$\|x^{k+1} - x^*\| \leqslant \frac{1}{\sqrt{1+\alpha\mu}} \cdot \left\|x^k - x^*\right\| \leqslant \left(\frac{1}{\sqrt{1+\alpha\mu}}\right)^{k+1} \left\|x^0 - x^*\right\|, \quad \forall\, k \in \mathbb{N}.$$

como queríamos demonstrar. □

Como a condição $x = J(x, \alpha)$ para $\alpha > 0$ é necessária e suficiente para que x seja uma solução ótima do Problema 2.1 (veja Observação 2.2), é interessante estudar a complexidade de $\|x^k - J(x^k, \alpha_k)\|$ para o método **MPG1**. A velocidade da convergência obtida abaixo não é afetada pelo comportamento dos tamanhos de passo α_k.

Teorema 4.3. *Sejam* (x^k) *e* (α_k) *as sequências geradas pelo **MPG1**. Então temos*

$$\liminf_{k\to\infty} \sqrt{k} \cdot \|x^k - J(x^k, \alpha_k)\| = 0. \tag{4.20}$$

Demonstração. Vamos supor que não ocorre (4.20). Sendo assim, o limite inferior não é zero, ou seja, podemos encontrar um número $\varepsilon > 0$ tal que para algum $K \in \mathbb{N}$ suficientemente grande, temos $\sqrt{k}\|x^k - J(x^k, \alpha_k)\| \geqslant \varepsilon$, o que por sua vez, implica que

$$\|x^k - J(x^k, \alpha_k)\| \geqslant \frac{\varepsilon}{\sqrt{k}}, \quad \forall k \geqslant K.$$

Sendo assim, elevando ambos os lados da desigualdade anterior ao quadrado, obtemos

$$\|x^k - J(x^k, \alpha_k)\|^2 \geqslant \frac{\varepsilon^2}{k},$$

aplicando o somatório de $k = K$ até $+\infty$, temos

$$\sum_{k=K}^{\infty} \|x^k - J(x^k, \alpha_k)\|^2 \geqslant \varepsilon^2 \sum_{k=K}^{\infty} \frac{1}{k} = +\infty, \quad \to \text{série divergente}.$$

Sendo assim,

$$\sum_{k=K}^{\infty} \|x^k - J(x^k, \alpha_k)\|^2 = +\infty. \tag{4.21}$$

Por outro lado, utilizando a Proposição 3.1 (ii), obtemos para todo $k \geqslant K$, que

$$[(g+h)(x^{k+1}) - (g+h)(x^k)] \leqslant -\left(\frac{1-\delta}{\alpha_k}\right)\|x^{k+1} - x^k\|^2,$$

o qual por sua vez, reorganizando os termos e usando a identidade em (3.1) implica que

$$\|x^k - J(x^k, \alpha_k)\|^2 = \|x^k - x^{k+1}\|^2 \leqslant \frac{\alpha_k}{1-\delta}[(g+h)(x^k) - (g+h)(x^{k+1})].$$

Agora, como pelo **MPG1** $\alpha_k < \sigma$, então para todo $k \in K$, temos que a desigualdade anterior se torna,

$$\|x^k - J(x^k - \alpha_k)\|^2 \leqslant \frac{\sigma}{1-\delta}\left[(g+h)(x^k) - (g+h)(x^{k+1})\right].$$

Somando a desigualdade anterior de ambos os lados para $k = K, \ldots, N$, temos

$$\sum_{k=K}^{N} \|x^k - J(x^k - \alpha_k)\|^2 \leqslant \frac{\sigma}{1-\delta}\sum_{k=K}^{N}\left[(g+h)(x^k) - (g+h)(x^{k+1})\right]. \tag{4.22}$$

Agora, escrevendo o somatório do lado direito, obtemos a seguinte soma telescópica,

$$\begin{aligned}\sum_{k=K}^{N}\left[(g+h)(x^k) - (g+h)(x^{k+1})\right] &= (g+h)(x^K) - (g+h)(x^{N+1})\\ &\leqslant (g+h)(x^K) - (g+h)(x^*).\end{aligned}$$

Sendo assim,

$$\sum_{k=K}^{N}\left[(g+h)(x^k) - (g+h)(x^{k+1})\right] \leqslant (g+h)(x^K) - (g+h)(x^N) \leqslant (g+h)(x^K) - (g+h)(x^*).$$

Daí e de (4.22), tomando o limite com $N \to +\infty$ temos,

$$\sum_{k=K}^{\infty} \|x^k - J(x^k - \alpha_k)\|^2 \leqslant \frac{\sigma}{1-\delta}[(g+h)(x^K) - (g+h)(x^*)] < +\infty,$$

absurdo, pois contrária (4.21). Logo,

$$\liminf_{k\to\infty} \sqrt{k}\cdot\|x^k - J(x^k, \alpha_k)\| = 0.$$

Como queríamos demostrar.

□

4.2 Análise de Complexidade do MPG2

Nesta seção, de forma análoga à seção anterior iremos tratar dos resultados de complexidade pior-caso relativos ao método **MPG2**, comprovando, no que diz respeito a complexidade que as alterações estudadas melhoram a complexidade do método.

Teorema 4.4. *Sejam (x^k) e (α^k) as sequências geradas pelo **MPG2**. Suponha que $S^* \neq \emptyset$ e exista $\alpha > 0$ tal que $\alpha_k \geqslant \alpha > 0$ para todo $k \in \mathbb{N}$. Então, temos*

$$(g+h)(x^k) - \min_{x\in\mathbb{R}^n}(g+h)(x) \leqslant \frac{\frac{2}{\alpha}(\|x^0 - x^*\|^2 + 2\alpha[(g+h)(x^0) - \min_{x\in\mathbb{R}^n}(g+h)(x)])}{(k+1)^2},$$

para todo $k \in \mathbb{N}$.

Demonstração. Seja $x^* \in S^*$. Pelo Lema (A.1) (i) temos que, $\frac{1}{t_k} \leqslant \frac{2}{k+1}$, reorganizando, temos $t_k \geqslant \frac{k+1}{2}$ fazendo $k = k+1$, obtemos $t_{k+1} \geqslant \frac{k+1+1}{2}$ o que por sua vez, implica que $t_{k+1} \geqslant 1$.

Como $t_{k+1} \geqslant 1$, podemos concluir que $t_{k+1}^{-1} \in [0,1]$. Utilizando a convexidade de g, e o fato de x^* e x^k pertencerem ao $\mathrm{dom}(g)$, encontramos a seguinte conbinação convexa,

$$x = t_{k+1}^{-1}x^* + (1 - t_{k+1}^{-1})x^k \in \mathrm{dom(h)}.$$

Aplicando a Proposição 3.2 para esse novo x, obtemos

$$\begin{aligned}
&\frac{1}{2\alpha_k}\left(\|x^{k+1} - (t_{k+1}^{-1}x^* + (1 - t_{k+1}^{-1}))x^k\|^2 - \|y^k - (t_{k+1}^{-1}x^* + (1 - t_{k+1}^{-1})x^k)\|^2\right) \\
&\qquad \leqslant (g+h)(t_{k+1}^{-1}x^* + (1 - t_{k+1}^{-1})x^k) - (g+h)(x^{k+1}) \\
&\qquad = t_{k+1}^{-1}(g+h)(x^*) + (1 - t_{k+1}^{-1})(g+h)(x^k) - (g+h)(x^{k+1}),
\end{aligned}$$

reorganizando, temos

$$\begin{aligned}
&(g+h)(x^*) - (g+h)(x^*) + t_{k+1}^{-1}(g+h)(x^*) + (1 - t_{k+1}^{-1})(g+h)(x^k) - (g+h)(x^{k+1}) \\
&\geqslant \frac{1}{2\alpha_k}\left(\|x^{k+1} - (t_{k+1}^{-1}x^* + (1 - t_{k+1}^{-1})x^k)\|^2 - \|y^k - (t_{k+1}^{-1}x^* + (1 - t_{k+1}^{-1})x^k)\|^2\right),
\end{aligned}$$

multiplicando o lado direito da desigualdade anterior por t_{k+1}^2 no numerador e denominador, obtemos

$$\begin{aligned}
&(1 - t_{k+1}^{-1})[(g+h)(x^k) - (g+h)(x^*)] - [(g+h)(x^{k+1}) - (g+h)(x^*)] \\
&\geqslant \frac{1}{2\alpha_k t_{k+1}^2}\left(\|t_{k+1}x^{k+1} - (x^* + (t_{k+1} - 1)x^k)\|^2 - \|t_{k+1}y^k - (x^* + (t_{k+1} - 1)x^k)\|^2\right).
\end{aligned}$$

Multiplicando por t_{k+1}^2, temos

$$\begin{aligned}
&\frac{1}{2\alpha_k}\left(\|t_{k+1}x^{k+1} - (x^* + (t_{k+1} - 1)x^k)\|^2 - \|t_{k+1}y^k - (x^* + (t_{k+1} - 1)x^k)\|^2\right) \\
&\qquad \leqslant (t_{k+1}^2 - t_{k+1})[(g+h)(x^k) - (g+h)(x^*)] - t_{k+1}^2[(g+h)(x^{k+1}) - (g+h)(x^*)],
\end{aligned}$$

usando (3.23) e o Lema (A.1)(ii), temos

$$\begin{aligned}
&\frac{1}{2\alpha_k}\left(\|t_{k+1}x^{k+1} - (t_{k+1} - 1)x^k - x^*\|^2 - \|t_{k+1}x^k - (t_k - 1)x^{k-1} - x^*\|^2\right) \\
&\leqslant (t_{k+1}^2 - t_{k+1})[(g+h)(x^k) - (g+h)(x^*)] - t_{k+1}^2[(g+h)(x^{k+1}) - (g+h)(x^*)] \\
&= (t_k^2 + t_{k+1} - t_{k+1})[(g+h)(x^k) - (g+h)(x^*)] - t_{k+1}^2[(g+h)(x^{k+1}) - (g+h)(x^*)] \\
&= (t_k^2[(g+h)(x^k) - (g+h)(x^*)] - t_{k+1}^2[(g+h)(x^{k+1}) - (g+h)(x^*)]).
\end{aligned}$$

Sendo assim, multiplicando por (-1), temos

$$\begin{aligned}
&\|t_{k+1}x^k-(t_k-1)x^{k-1}-x^*\|^2-\|t_{k+1}x^{k+1}-(t_{k+1}-1)x^k-x^*\|^2\\
&\geqslant 2\alpha_k(t_{k+1}^2[(g+h)(x^{k+1})-(g+h)(x^*)]-t_k^2[(g+h)(x^k)-(g+h)(x^*))\\
&=2\alpha_k t_{k+1}^2[(g+h)(x^{k+1})-(g+h)(x^*)]-2\alpha_k t_k^2[(g+h)(x^k)-(g+h)(x^*)],
\end{aligned}$$

como $\alpha_k \geqslant \alpha_{k+1} =$ LS1 $(\tilde{y}^k,\alpha_k,\theta,\delta)$ e $(g+h)(x^{k+1})-(g+h)(x^*)\geqslant 0$, temos

$$\begin{aligned}
&\|t_{k+1}x^k-(t_k-1)x^{k-1}-x^*\|^2-\|t_{k+1}x^{k+1}-(t_{k+1}-1)x^k-x^*\|^2\\
&\geqslant 2\alpha_{k+1}t_{k+1}^2[(g+h)(x^{k+1})-(g+h)(x^*)]-2\alpha_k t_k^2[(g+h)(x^k)-(g+h)(x^*)].
\end{aligned}$$

Reorganizando a desigualdade acima e aplicando indução em k decrescendo, obtemos

$$\begin{aligned}
2\alpha_{k+1}t_{k+1}^2&[(g+h)(x^{k+1})-(g+h)(x^*)]\\
&\leqslant \|t_{k+1}x^{k+1}-(t_{k+1}-1)x^k-x^*\|^2+2\alpha_{k+1}t_{k+1}^2[(g+h)(x^{k+1})\\
&-(g+h)(x^*)]\\
&\leqslant \|t_kx^k-(t_k-1)x^{k-1}-x^*\|^2+2\alpha_k t_k^2[(g+h)(x^k)-(g+h)(x^*)]\\
&\leqslant\cdots\leqslant \|t_0x^0-(t_0-1)x^{-1}-x^*\|^2+2\alpha_0t_0^2[(g+h)(x^0)-(g+h)(x^*)],
\end{aligned}$$

como $t_0=1$ pelo **MPG2**, temos

$$\begin{aligned}
2\alpha_{k+1}t_{k+1}^2&[(g+h)(x^{k+1})-(g+h)(x^*)]\\
&\leqslant \|x^0-(1-1)x^{-1}-x^*\|^2+2\alpha_0 1^2[(g+h)(x^0)-(g+h)(x_\star)]\\
&=\|x^0-x^*\|^2+2\alpha_0[(g+h)(x^0)-(g+h)(x^*)],
\end{aligned}$$

lembrando que $\alpha_k\leqslant\sigma$, $\forall k\in\mathbb{N}$, temos

$$2\alpha_k t_k^2[(g+h)(x^k)-(g+h)(x^*)]\leqslant \|x^0-x^*\|^2+2\sigma[(g+h)(x^0)-(g+h)(x^*)],$$

usando o fato de $x^*\in S^*$, a desigualdade anterior e o Lema A.1 (i), obtemos

$$\begin{aligned}
(g+h)(x^k)-\min_{x\in\mathbb{R}^n}(g+h)(x)&\leqslant\frac{1}{2\alpha_k t_k^2}(\|x^0-x^*\|^2+2\sigma[(g+h)(x^0)-\min_{x\in\mathbb{R}^n}(g+h)(x)])\\
&\leqslant\frac{4}{2\alpha_k(k+1)^2}(\|x^0-x^*\|^2\\
&\quad+2\sigma[(g+h)(x^0)-\min_{x\in\mathbb{R}^n}(g+h)(x)])\\
&\leqslant\frac{\frac{2}{\alpha}\cdot(\|x^0-x^*\|^2+2\sigma[(g+h)(x^0)-\min_{x\in\mathbb{R}^n}(g+h)(x)])}{(k+1)^2}.
\end{aligned}$$

para todo $x^*\in S^*$, como queríamos demostrar. □

Podemos perceber com este Teorema que o erro esperado dos iterados gerados pelo **MPG2** após k iterações é $o(k^{-2})$ quando os tamanhos dos passos são limitados inferiormente por uma constante positiva. De maneira semelhante a Proposição 4.1, provaremos no resultado seguinte, que tal requisito é satisfeito sob a suposição de Lipschitz global sobre o gradiente de g. A complexidade $o(k^{-2})$ para o método acelerado, de maneira similar (3.23)-(4.15), foi obtida recentemente em [2, 18] sob a suposição de Lipschitz global. Seria interessante combinar sua tecnica com as nossas para uma complexidade semelhante sob a suposição mais fraca de continuidade de Lipschitz local, como na Proposição 4.1 (ii).

Proposição 4.2. *Seja* (α_k) *a sequência gerada pela Busca Linear 1 em* **MPG2**. *Se o gradiente de* g *for globalmente Lipschitz continuo em dom(h), então existe algum* $\alpha > 0$ *tal que* $\alpha_k \geqslant \alpha$ *para todo* $k \in \mathbb{N}$.

Demonstração. Vamos supor que ∇g é globalmente Lipschitz contínuo no dom(h) com constante $L > 0$. Como α_k é a sequência gerada pela Busca Linear 1, então sabemos que é não negativa, decrescente e limitada, logo existe α tal que $\alpha_k \to \alpha$ quando $k \to +\infty$. Sendo assim, se $\alpha < \frac{\delta\theta}{L}$ podemos encontrar $K \in \mathbb{N}$ tal que $\alpha_k < \frac{\delta\theta}{L}, \quad \forall k > K$. Vamos definir $\hat{\alpha}_k = \frac{\alpha_k}{\theta} > 0$ e $\hat{y}^k = J(\tilde{y}^k, \hat{\alpha}_k) = \text{prox}_{\hat{\alpha}_k h}(\tilde{y}^k - \hat{\alpha}_k \nabla g(\tilde{y}^k)) \in \text{dom(h)}$. Se $\alpha_k < \alpha_{k-1}$ para $k < K$ temos pela Busca Linear 1 (2.11), que

$$\hat{\alpha}_k \|\nabla g(\hat{y}^k) - \nabla g(\tilde{y}^k)\| > \delta \|\hat{y}^k - \tilde{y}^k\|. \tag{4.23}$$

Como ∇g é Lipschitz contínuo em dom(h) com constante L, então para $\hat{y}^k = x$ e $\tilde{y}^k = y$, obtemos

$$\hat{\alpha}_k \|\nabla g(\hat{y}^k) - \nabla g(\tilde{y}^k)\| \leqslant \hat{\alpha}_k L \|\hat{y}^k - \tilde{y}^k\|.$$

Daí, e de (4.23), temos

$$\hat{\alpha}_k L \|\hat{y}^k - \tilde{y}^k\| > \delta \|\hat{y}^k - \tilde{y}^k\|.$$

Dessa forma, dividindo por $\|\hat{y} - \tilde{y}\| > 0$, pois caso contrário a desigualdade não faria sentido. Obtemos,

$$\hat{\alpha}_k L > \delta.$$

Como $\hat{\alpha}_k = \frac{\alpha_k}{\theta}$, temos $\frac{\alpha_k}{\theta} L > \delta$, ou seja, $\alpha_k \geqslant \frac{\delta\theta}{L}$.

Chegamos em um absurdo, pois supomos que $\alpha_k < \frac{\delta\theta}{L}$. Logo, $\alpha_k \geqslant \frac{\delta\theta}{L}$, ou seja, $\alpha > \frac{\delta\theta}{L}$. Logo, existe $\alpha > 0$, tal que $\alpha_k \geqslant \alpha$, para todo $k \in \mathbb{N}$. □

Vamos completar a seção com uma consequência direta da Proposição acima e do Teorema (4.4).

Corolário 4.2. *Seja* (x^k) *a sequência gerada pelo* ***MPG2***. *Suponha que* $S^* \neq \emptyset$ *e que o gradiente de* g *seja continuamente Lipschitz em dom(h). Então, temos*

$$(g+h)(x^k) - \min_{x \in R^n}(g+h)(x) = O\left(\frac{1}{(k+1)^2}\right).$$

4.3 Análise de Complexidade do MPG3.

Nesta subseção, mostraremos a complexidade do **MPG3** com uma taxa semelhante ao Teorema 4.1.

Teorema 4.5. *Sejam* (x^k) *e* (β_k) *as sequências geradas pelo* ***MPG3***. *Suponha que* $S^* \neq \emptyset$ *e exista algum* $\beta > 0$ *satisfazendo* $\beta_k \geqslant \beta > 0$ *para todo* $k \in \mathbb{N}$. *Então,*

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x) \leqslant \frac{1}{2\beta}\frac{[\text{dist}(x^0, S^*)]^2 + 2[(g+h)(x^0) - \min_{x \in \mathbb{R}^n}(g+h)(x)]}{k}. \tag{4.24}$$

Além disso,

$$\lim_{k \to \infty} k[(g+h)(x^k) - \min_{x \in \mathbb{R}^n}(g+h)(x)] = 0. \tag{4.25}$$

Demonstração. Usando a Proposição 3.3, com $l \in \mathbb{N}$ e $x^* \in S^*$, obtemos

$$\begin{aligned} 0 &\geqslant (g+h)(x^*) - (g+h)(x^{l+1}) \\ &\geqslant \frac{1}{2\beta_l}(\|x^{l+1} - x^*\|^2 - \|x^l - x^*\|^2) + 2[(g+h)(x^{l+1}) - (g+h)(x^l)], \end{aligned}$$

como $\beta_k \geqslant \beta$ e a sequência (x^k) gerada pelo **MPG3** é Quasi-Fejér, temos

$$(g+h)(x^*) - (g+h)(x^{l+1}) \geqslant \frac{1}{2\beta}(\|x^{l+1} - x^*\|^2 - \|x^l - x^*\|^2) + 2[(g+h)(x^{l+1}) - (g+h)(x^l)], \tag{4.26}$$

Para todo $l \in \mathbb{N}$. Somando a desigualdade acima para $l = 0, 1, \ldots, k-1$, tem-se a partir da soma telescópica que

$$\sum_{l=0}^{k-1}[(g+h)(x^*) - (g+h)(x^{l+1})] \geqslant \frac{1}{2\beta}(\|x^k - x^*\|^2 - \|x^0 - x^*\|^2) + 2[(g+h)(x^k) - (g+h)(x^0)],$$

como $(g+h)(x^k) \geqslant (g+h)(x^*)$, pois $x^* \in S^*$, então

$$\sum_{l=0}^{k-1}[(g+h)(x^*) - (g+h)(x^{l+1})] \geqslant \frac{1}{2\beta}(\|x^k - x^*\|^2 - \|x^0 - x^*\|^2) + 2[(g+h)(x^*) - (g+h)(x^0)]. \tag{4.27}$$

De (3.37) temos que $(g+h)(x^l) \geqslant (g+h)(x^{l+1})$ para todo $l = 0, 1, \cdots, k-1$. Combinando este resultado com a desigualdade anterior, temos

$$k[(g+h)(x^*) - (g+h)(x^k)] \geqslant \frac{1}{2\beta}(\|x^k - x^*\|^2 - \|x^0 - x^*\|^2) + 2[(g+h)(x^*) - (g+h)(x^0)].$$

Multiplicando por (-1), temos

$$(g+h)(x^k) - (g+h)(x^*) \leqslant \frac{1}{2\beta}\frac{(\|x^0 - x^*\|^2 - \|x^k - x^*\|^2) + 2[(g+h)(x^0) - (g+h)(x^*)]}{k},$$

o que por sua vez, implica que

$$(g+h)(x^k) - (g+h)(x^*) \leqslant \frac{1}{2\beta}\frac{\|x^0 - x^*\|^2 + 2[(g+h)(x^0) - (g+h)(x^*)]}{k}.$$

Utilizando a definição de ponto conjunto, temos

$$(g+h)(x^k) - \min_{x\in\mathbb{R}^n}(g+h)(x) \leqslant \frac{1}{2\beta}\frac{[\text{dist}(x^0, S^*)]^2 + 2[(g+h)(x^0) - \min_{x\in\mathbb{R}^n}(g+h)(x)]}{k}. \tag{4.28}$$

Para todo $x^* \in S^*$, como queríamos em (4.24). Para justificar (4.25), vamos supor pelo Teorema (3.2) que (x^k) converge para algum $x^* \in S^*$. Sendo assim, para qualquer $\varepsilon > 0$, existe algum $k > 0$ tal que

$$\|x^k - x^*\| \leqslant \varepsilon \quad \text{e} \quad (g+h)(x^k) - (g+h)(x^*) \leqslant \varepsilon, \quad \forall\, k \in K. \tag{4.29}$$

Onde a última desigualdade vem de (4.28). Aplicando somatório em (4.26) para $l = K, K+1, \cdots, K+k-1$ e utilizando soma telescópica, temos

$$\begin{aligned}\sum_{l=k}^{K+k-1}[(g+h)(x^*) - (g+h)(x^{l+1})] &\geqslant \frac{1}{2\beta}(\|x^{K+k} - x^*\|^2 - \|x^K - x^*\|^2)\\ &\quad + 2[(g+h)(x^{K+k}) - (g+h)(x^K)],\end{aligned}$$

como por (4.26) vale $(g+h)(x^{l+1}) \geqslant (g+h)(x^{K+k})$ para todo $l = K, K+1, \cdots, K+k-1$, temos que

$$\begin{aligned}k[(g+h)(x^*) - (g+h)(x^{K+k})] &\geqslant \frac{1}{2\beta}(\|x^{K+k} - x^*\|^2 - \|x^K - x^*\|^2)\\ &\quad + 2[(g+h)(x^{K+k}) - (g+h)(x^K)],\end{aligned}$$

utilizando (4.29), obtemos

$$\begin{aligned}k[(g+h)(x^*) - (g+h)(x^{K+k})] &\geqslant \frac{1}{2\beta}(-\|x^K - x^*\|^2 + 2[(g+h)(x^*)) - (g+h)(x^K)])\\ &\geqslant \frac{1}{2\beta}\left(-\varepsilon^2 - 2\varepsilon\right).\end{aligned}$$

Aplicando lim sup com $k \to \infty$ e utilizando a desigualdade anterior, obtemos

$$\begin{aligned}
\limsup_{k\to\infty} k[(g+h)(x^k) - (g+h)(x^*)] &= \limsup_{k\to\infty} = (K+k)[(g+h)(x^{K+k}) - (g+h)(x^*)] \\
&\leqslant \limsup_{k\to\infty} \frac{K+k}{k} \cdot \frac{\varepsilon^2 + 2\varepsilon}{2\beta} \\
&= \frac{\varepsilon^2 + 2\varepsilon}{2\beta}.
\end{aligned}$$

Como está desigualdade é valida para todo $\varepsilon > 0$. Podemos concluir a seguinte desigualdade $\limsup\limits_{k\to\infty} k[(g+h)(x^k) - (g+h)(x^*)] \leqslant 0$. Por outro lado, como $x^* \in S^*$, podemos perceber que $(g+h)(x^k) - (g+h)(x^*) \geqslant 0, \quad \forall k \in \mathbb{N}$. Dessa forma,

$$(g+h)(x^k) - (g+h)(x^*) = 0.$$

Ou seja,

$$(g+h)(x^k) - \min_{x\in\mathbb{R}^n}(g+h)(x) = 0,$$

completando assim, a prova de (4.25). □

Similarmente a Proposição 4.1, apresentamos algumas condições suficientes para que o passo gerado pela Busca Linear 2 seja limitado inferiormente por uma constante positiva.

Proposição 4.3. *Seja* (β_k) *a sequência gerada pela Busca Linear 2 no* ***MPG3****. As seguintes informações são válidas:*

(i) Se o gradiente de g *é globalmente Lipschitz contínuo em dom(h) com constante* $L > 0$, *então* $\beta_k \geqslant \min\{1, \frac{\theta}{2L}\}$ *para todo* $k \in \mathbb{N}$.

(ii) Suponha que $S^* \neq \emptyset$. *Se* ∇g *é localmente Lipschitz contínua em qualquer* $x \in S^*$, *então existe* $x^* \in S^*$ *tal que*

$$\liminf_{k\to\infty} \beta_k \geqslant \min\left\{1, \frac{\theta}{2L}\right\}, \tag{4.30}$$

onde, $L > 0$ *é a constante de Lipschitz continua do* ∇g *em torno de* x^*. *Consequentemente, existe* $\beta > 0$ *tal que* $\beta_k \geqslant \beta$, *para todo* $k \in \mathbb{N}$.

Demonstração. Vamos verificar primeiro o item (i), supondo que o gradiente de g seja globalmente Lipschitz contínuo em dom(h) com constante $L > 0$. Defina $\hat{\beta}_k := \frac{\beta}{\theta} > 0$ e

$$\hat{y}^k := \hat{\beta}_k J_k + (1 - \hat{\beta}_k)x^k = x^k - \hat{\beta}_k(x^k - J_k). \tag{4.31}$$

Se $\beta_k < 1$, obtemos da Busca Linear 2 (2.14) que

$$(g+h)(\hat{y}^k) > (g+h)(x^k) - \hat{\beta}_k[h(x^k) - h(J_k)] - \hat{\beta}_k\langle \nabla g(x^k), x^k - J_k\rangle + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2.$$

Podemos observar que $x^k \neq J_k$, pois caso contrário utilizando (4.31) e $x^k = J_k$, em (2.7), implicaria que $0 \in \nabla g(x^k) + \partial h(x^k)$, ou seja, $x^k \in S^*$, o que contradiz a falha da condição da Busca Linear para o passo $\hat{\beta}_k$. Outra consequência deste fato é que $\hat{y}^k \neq x^k$, pois por (4.31), temos $\hat{y}^k - x^k = \hat{\beta}_k(x^k - J_k)$, aplicando norma, $\|\hat{y}^k - x^k\| = \hat{\beta}_k\|(x^k - J_k)\|$, como $\hat{\beta}_k > 0$ e $\|x^k - J_k\| \neq 0$, então $\|\hat{y}^k - x^k\| > 0$, logo $\hat{y}^k \neq x^k$. Além disso, é similar a (3.46) na prova do Teorema 3.2 que $\frac{1}{2}\|x^k - J_k\| \leqslant \|\nabla g(\hat{y}^k) - \nabla g(x^k)\|$. Daí, juntamente à continuidade de Lipschitz com constante L de ∇g e (4.31), temos

$$\frac{1}{2}\|x^k - J_k\| \leqslant L\|\hat{y}^k - x^k\| = L\hat{\beta}_k\|x^k - J_k\|.$$

Como $\|x^k - J_k\| \neq 0$, a desigualdade acima resulta em $\hat{\beta}_k \geqslant \frac{1}{2L}$ e portanto $\hat{\beta}_k \geqslant \frac{\theta}{2L}$ quando $\beta_k < 1$. Dessa forma, concluímos que $\beta_k \geqslant \min\left\{1, \frac{\theta}{2L}\right\}$, como queríamos demonstrar.

Para verificar a segunda parte, vamos supor que ∇g seja localmente Lipschitz contínuo em qualquer $x \in S^*$. Pelo Teorema 3.2, suponha que (x^k) converge para $x^* \in S^*$. Portanto, existem $\varepsilon, L > 0$ tais que

$$\|\nabla g(x) - \nabla g(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in \mathbb{B}_\varepsilon(x^*).$$

Como $x^k \to x^*$ quando $k \to \infty$, encontramos $k > 0$ tal que $\|x^k - x^*\| < \varepsilon$ para todo $k > K$. Para qualquer $k > K$, se $\beta_k < 1$, Vamos definir $\hat{\beta}_k = \frac{\beta_k}{\theta}$ e $\hat{y}^k = \hat{\beta}_k J_k + (1 - \hat{\beta}_k)x^k$. De forma semelhante ao argumento acima da primeira parte, temos $\|x^k - J_k\| \neq 0$ e

$$\frac{1}{2}\|x^k - J_k\| \leqslant \|\nabla g(\hat{y}^k) - \nabla g(x^k)\|. \tag{4.32}$$

Agora consideramos dois casos semelhantemente ao Teorema 3.2.

Caso 1: A sequência (β_k) é limitada inferiormente por um número positivo $\beta > 0$. Dessa forma, $\beta_k \geqslant \beta > 0$. Utilizando (3.33), temos que $x^k - x^{k+1} = \beta_k(J_k - x^k)$, aplicando norma, obtemos

$$\|J_k - x^k\| = \frac{\|x^k - x^{k+1}\|}{\beta_k} \leqslant \frac{\|x^k - x^{k+1}\|}{\beta}.$$

Temos que β_k é uma sequência limitada e $x^k \to x^*$, dessa forma $\|x^k - x^{k+1}\| \to 0$, o que por sua vez, implica que $\|J_k - x^k\| \to 0$ quando $k \to \infty$. Utilizando (4.31) e aplicando norma, temos

$$\|x^k - \hat{y}^k\| = \frac{\beta_k}{\theta}\|x^k - J_k\|,$$

como $\|x^k - J_k\| \to 0$, obtemos que $\|x^k - \hat{y}^k\| \to 0$, quando $k \to \infty$. O que nos diz que $(\hat{y}^k)$ está convergindo para x^*. Portanto, existe $K_1 > K$ tal que $\hat{y}^k \in \mathbb{B}_\varepsilon(x^*)$ para todo $k > K_1$. Daí e de (4.32) obtemos

$$\frac{1}{2}\|x^k - J_k\| \leqslant L\|\hat{y}^k - x^k\| = L\hat{\beta}_k\|x^k - J_k\|, \quad \forall k > K_1.$$

Como $\|x^k - J_k\| \neq 0$, podemos concluir que $\frac{1}{2} \leqslant L\hat{\beta}_k$, isto é, $\beta_k \geqslant \frac{\theta}{2L}$ para todo $k > K_1$.

Caso 2: A sequência β_k não é limitada inferiormente por um número positivo β. Portanto, podemos encontrar uma subsequência β_k que converge para 0, isto é, $\beta_k \to 0$. Por (4.31), temos que $\hat{y}^k - x^k = \hat{\beta}_k(x^k - J_k)$ aplicando norma,

$$\|\hat{y}^k - x^k\| = \hat{\beta}_k\|(x^k - J_k)\| > 0. \tag{4.33}$$

A demonstração de que (x^k) e (J_k) são limitadas é semelhante ao do Teorema 3.2. Temos de (4.31) e do fato de $\beta_k \to 0$ (por hipótese no caso 2), que $\|\hat{y}^k - x^k\| \to 0$ quando $k \to \infty$. Sendo assim, como (x^k) e (J_k) são limitadas, temos que $\|x^k - J_k\|$ é limitada, dessa forma, aplicando limite em (4.33) com $k \to \infty$, temos

$$\lim_{k\to\infty} \|\hat{y}^k - x^k\| = \lim_{k\to\infty} \frac{\beta_k}{\theta}\|x^k - J_k\| = 0,$$

pois, $\theta > 0$ e $\beta_k \to 0$. Portanto como $(x^k) \to x^*$, concluímos que $(\hat{y}^k) \to x^*$. Semelhantemente a prova do caso 1, temos que, como $(\hat{y}^k) \to x^*$ existe $K_1 > K$ tal que $\hat{y}^k \in \mathbb{B}_\varepsilon(x^*)$ para todo $k > K_1$. Daí e de (4.32) obtemos

$$\frac{1}{2}\|x^k - J_k\| \leqslant L\|\hat{y}^k - x^k\| = L\hat{\beta}_k\|x^k - J_k\|, \quad \forall k > K_1.$$

Como $\|x^k - J_k\| \neq 0$, podemos concluir que $\frac{1}{2} \leqslant L\hat{\beta}_k$, isto é, $\beta_k \geqslant \frac{\theta}{2L}$ para todo $k > K_1$. Chegamos em um absurdo, pois supomos no início que $\beta_k \to 0$ e $\frac{\theta}{2L} > 0$. Portanto, o Caso 2 não pode ocorrer, ou seja, (β_k) é de fato limitada inferiormente por uma constante positiva. A partir da análise de ambos os casos acima, encontramos $K_1 > 0$ tal que $\beta_k \geqslant \frac{\theta}{2L}$ se $\beta_k < 1$ para qualquer $k > K_1$. Isso significa que $\beta_k \geqslant \min\{1, \frac{\theta}{2L}\}$ para $k > K_1$. Em particular

$$\liminf_{k\to\infty} \beta_k \geqslant \min\left\{1, \frac{\theta}{2L}\right\},$$

e existe $\beta > 0$ tal que $\beta_k \geqslant \beta$ para todo $k \in \mathbb{N}$. □

Vamos concluir a seção apresentando um corolário correspondente ao Corolário 4.1, que é facilmente derivado do Teorema 4.5 e da Proposição 4.3.

Corolário 4.3. *Seja* (x^k) *a sequência gerada pelo* ***MPG3****. Suponha que* $S^* \neq \emptyset$.

(i) Se o gradiente de g *for globalmente Lipschitz contínuo em dom(h), então*

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n} (g+h)(x) = O(k^{-1}).$$

(ii) Se o gradiente de g *é localmente Lipschitz contínuo em* S^**, então temos*

$$(g+h)(x^k) - \min_{x \in \mathbb{R}^n} (g+h)(x) = o(k^{-1}).$$

Capítulo 5

Conclusão

No espaço euclidiano $\mathbb{R}^n$, é bem conhecido que a convexidade das funções envolvidas, juntamente com a continuidade global de Lipschitz do gradiente da função g, são condições suficientes para garantir a convergência da sequência gerada pelos métodos de separação forward–backward na resolução do Problema (1.19). No entanto, a hipótese de Lipschitz para o gradiente, em muitas situações particulares é uma restrição significativa. Neste trabalho, analisamos a convergência do método de divisão forward–backward aplicado a problemas de otimização convexa no espaço euclidiano $\mathbb{R}^n$, explorando estratégias de Busca Linear. Essa abordagem não apenas elimina a principal desvantagem associada à necessidade de estimar previamente a constante de Lipschitz para a escolha do tamanho do passo, como também permite estabelecer diversos resultados de complexidade sem impor tal hipótese. Ademais, os esquemas estudados, baseados em procedimentos de Busca Linear, fornecem mecanismos rigorosos e implementáveis para a atualização das iterações, os quais podem ser facilmente adaptados a diferentes aplicações práticas.

Apêndice A

Apêndice

O presente apêndice destina-se a reunir, de forma organizada e acessível, uma coleção de resultados fundamentais, definições, lemas e teoremas que foram empregados ao longo deste trabalho. A inclusão desses elementos visa proporcionar uma referência rápida e conveniente. Ao compilar este material, objetiva-se facilitar a compreensão e a validação dos argumentos desenvolvidos nos capítulos principais, garantindo que o leitor tenha à disposição todos os subsídios teóricos necessários para acompanhar as demonstrações e discussões apresentadas. Para maiores detalhes técnicos, o leitor pode ver [26, 27].

Análise no $\mathbb{R}^n$

Em seguida veremos o Teorema do valor Intermediário e Teorema do Valor Médio, os mesmos serão de suma importância para a prova de convergência do Método do Gradiente que por sua vez utiliza a Busca de Armijo.

Teorema A.1. *(Teorema do Valor Intermediário - TVI) Seja* $f : D \to \mathbb{R}$ *uma função contínua definida em um conjunto* $X \subset \mathbb{R}^n$*. Se existirem* $a, b \in X$ *e* $d \in \mathbb{R}$ *tais que* $f(a) < d < f(b)$ *então existe* $c \in X$ *tal que* $f(c) = d$.

Teorema A.2. *(Teorema do Valor Médio - TVM) Seja* $f : [a, b] \to \mathbb{R}^n$ *um caminho contínuo, diferenciável no intervalo aberto* (a, b)*. Se* $|f'(t)| \leqslant M$ *para todo* $t \in (a, b)$*, então* $|f(b) - f(a)| \leqslant M(b - a)$.

Teorema A.3. *Dada a função diferenciável* $f : D \to \mathbb{R}$*, definida no aberto* $D \subset \mathbb{R}^n$ *definiremos o gradiente de* f *no ponto* $a \in D$ *como o vetor* $\nabla f(a)$*, que corresponde ao* *vetor*

$$\langle \nabla f(a), v\rangle = \frac{\partial f}{\partial v}(a) = \sum \frac{\partial f}{\partial x_i}(a)\cdot \alpha_i$$

para todo $v = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. *Em particular* $\langle \nabla f(a), e_i\rangle = \sum \frac{\partial f}{\partial x_i}(a)$, *logo*

$$\mathtt{grad}\ g(a) = \left(\frac{\partial g}{\partial x_1}(a), \ldots, \frac{\partial g}{\partial x_n}(a)\right).$$

Definição A.1. *Dizemos que um ponto* $x^* \in D \subset \mathbb{R}^n$ *é*

(i) minimizador global de f *em D, se*

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

(ii) minimizador local de f *em D, se existe uma vizinhança U de* $\bar{x}$ *tal que*

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

De forma equivalente, $\bar{x} \in D$ *é minimizador local se existe* $\varepsilon > 0$ *tal que* $f(\bar{x}) \leqslant f(x)$ *para todo* $x \in \{x \in D \mid \|x - \bar{x}\| \leqslant \varepsilon\}$.

Lema A.1. *A sequência positiva* (t_k) *gerada pelo **MPG2** vista em* (3.22) *satisfaz, para todo* $k \in \mathbb{N}$,

(i) $\frac{1}{t_k} \leqslant \frac{2}{k+1}$.

(ii) $t_{k+1}^2 - t_{k+1} = t_k^2$.

Análise Convexa

A seguir, trazemos uma coleção de resultados de Análise Convexa utilizados ao longo do trabalho.

Definição A.2. *(Conjunto Convexo) Um conjunto* $C \subset \mathbb{R}^n$ *é dito convexo quando dados* $x, y \in C$, *o segmento* $[x, y] = \{(1-t)x + ty | t \in [0, 1]\}$ *estiver inteiramente contido em C.*

Definição A.3. *Seja* D *um subconjunto não-vazio, fechado e convexo. O **operador projeção** em* D, *denotado por* $P_D : \mathcal{H} \to D$, *é definido da seguinte forma:*

Para cada $x \in \mathbb{R}^n$, $P_D(x)$ *é o único ponto em* D *que minimiza a distância euclidiana de* x *a* D. *Em outras palavras,* $P_D(x)$ *é o ponto* $\bar{x} \in D$ *tal que:*

$$\bar{x} = P_D(x) = \arg\min_{y \in D} \|x - y\|.$$

Teorema A.4. *(Teorema da Projeção) Seja* $D \subset \mathbb{R}^n$ *um conjunto convexo e fechado. Então, para todo* $x \in \mathbb{R}^n$*, a projeção de* x *sobre* D*, denotada por* $P_D(x)$*, existe e é única. Além disso,* $\bar{x} = P_D(x)$ *se, e somente se,*

$$\bar{x} \in D, \quad \langle x - \bar{x}, y - \bar{x} \rangle \leqslant 0, \quad \forall y \in D.$$

Corolário A.1. *(Operador de projeção é monótono e não-expansivo) Seja* $D \subseteq \mathbb{R}^n$ *um conjunto convexo e fechado. Então para* $x \in \mathbb{R}^n$ *e* $y \in \mathbb{R}^n$ *quaisquer,*

$$(P_D(x) - P_D(y), x - y) \geqslant \|P_D(x) - P_D(y)\|^2 \geqslant 0,$$

$$\|P_D(x) - P_D(y)\| \leqslant \|x - y\|.$$

Em particular, $P_D(\cdot)$ *é contínuo no* $\mathbb{R}^n$*.*

Dado $a \in \mathbb{R}$ e $\lambda > 0$, usaremos as seguintes convenções aritméticas envolvendo $+\infty$:

$$a + (+\infty) = +\infty, \quad 0 \cdot (+\infty) = 0, \quad \lambda \cdot (+\infty) = +\infty.$$

Definição A.4. *(i) Uma função* $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ *é dita* ***convexa*** *se para quaisquer* $x, y \in \mathbb{R}^n$ *e qualquer* $\lambda \in [0, 1]$*, a seguinte desigualdade é satisfeita:*

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

(ii) Uma função $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ *é dita* ***estritamente convexa*** *se para quaisquer* $x, y \in \mathbb{R}^n$ *com* $x \neq y$*, e qualquer* $\lambda \in (0, 1)$*, a seguinte desigualdade é satisfeita:*

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

(iii) Uma função $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ *é dita* ***fortemente convexa*** *com módulo* $\mu > 0$ *se para quaisquer* $x, y \in \mathbb{R}^n$ *e qualquer* $\lambda \in [0, 1]$*, a seguinte desigualdade é satisfeita:*

$$f(\lambda x + (1 - \lambda)y) \leqslant \lambda f(x) + (1 - \lambda)f(y) - \frac{1}{2}\mu\lambda(1 - \lambda)\|x - y\|^2,$$

onde $\|\cdot\|$ *denota a norma euclidiana em* $\mathbb{R}^n$*.*

Teorema A.5. *(Desigualdade do gradiente) Seja* $f : D \to \overline{\mathbb{R}}$ *uma função convexa e diferenciável. Dado,* $x \in \mathrm{dom}(f)$*, então para todo* $y \in \mathbb{R}^n$ *temos que*

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle. \tag{A.1}$$

Observação A.1. *Note que segue do Fato A.1 que a definição A.4 (iii) se torna equivalente à seguinte condição*

$$f(x) \geqslant f(y) + \langle v, x - y \rangle + \frac{\lambda}{2}\|x - y\|^2, \quad \forall x \in \mathbb{R}^n, (y, v) \in \text{Gph } \partial f.$$

Definição A.5. *O conjunto de nível da função* $f : \mathbb{R}^n \to \mathbb{R}$ *associado a* $c \in \mathbb{R}$*, é o conjunto dado por,*

$$S_{lev}(c) = \{x \in \mathbb{R}^n \mid f(x) \leqslant c\}.$$

Definição A.6. *Dizemos que a função* $f : \mathbb{R}^n \to \mathbb{R}$ *é coerciva em* $\mathbb{R}^n$*, quando para cada sequência* $(x^k) \subset \mathbb{R}^n$ *tal que* $\|x^k\| \to \infty$*, tem-se que*

$$\lim_{k \to \infty} f(x^k) = +\infty.$$

Teorema A.6. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função contínua e coerciva. Então,* f *tem um minimizador global.*

Demonstração. Dado $a \in \mathbb{R}^n$. Seja $b = f(a)$. Como f é coerciva, então vale

$$\lim_{\|x\| \to \infty} f(x) = \infty,$$

ou seja, existe $r > 0$ tal que $f(x) > b$ param todo $x \in \mathbb{R}^n$ sempre que $\|x\| > r$. Observe que o conjunto $B = \{x \in \mathbb{R}^n | \ \|x\| \leqslant r\}$ é compacto e f é contínua, então existe $x^* \in B$ tal que $f(x^*) \leqslant f(x), \quad \forall x \in B$. Além disso, $a \in B$ pois $f(a) = b$. Sendo assim, temos que

$$f(x) > b = f(a) \geqslant f(x^*),$$

dessa forma,

$$f(x^*) < f(x), \quad \forall x \in \mathbb{R}^n.$$

Logo, f tem um mínimo global. □

Definição A.7. *(Direção de descida) Dizemos que* $d \in \mathbb{R}^n$ *é uma direção de descida de* $f : \mathbb{R}^n \to \mathbb{R}$ *no ponto* $x \in \mathbb{R}^n$*, se existe* $\varepsilon > 0$ *tal que,*

$$f(x + td) < f(x), \quad \forall t \in (0, \varepsilon].$$

Denotamos por $D_f(x)$ *o conjunto de todas as direções de descida da função* f *no ponto* x*.*

Lema A.2. *(Direções de descida) Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável no ponto* $x \in \mathbb{R}^n$*. Então:*

(i) Para todo $d \in D_f(x)$*, tem-se* $\langle \nabla f(x), d\rangle \leqslant 0$.

(ii) Se $d \in \mathbb{R}^n$ *satisfaz* $\langle \nabla f(x), d\rangle < 0$*, tem-se que* $d \in D_f(x)$.

Demonstração. Vamos provar primeiro o item (i). Seja $d \in D_f(x)$. Ou seja, existe $\varepsilon > 0$ tal que

$$f(x + td) < f(x), \quad \forall\, t \in (0, \varepsilon],\ k \in \mathbb{N}.$$

Utilizando a diferenciabilidade de f no ponto x , temos

$$f(x + td) = f(x) + t\langle \nabla f(x), d\rangle + o(t).$$

Sendo assim, da desigualdade anterior e do fato de d ser direção de descida, obtemos

$$0 > f(x + td) - f(x) = t\langle \nabla f(x), d\rangle + o(t),$$

o que por sua vez, implica que

$$t\langle \nabla f(x), d\rangle + o(t) < 0,$$

dividindo a desigualdade anterior por $t > 0$, temos

$$\langle \nabla f(x), d\rangle + \frac{o(t)}{t} < 0.$$

Agora, passando limite quando $t \to 0^+$, obtemos

$$\lim_{t\to 0^+}\left(\langle \nabla f(x), d\rangle + \frac{o(t)}{t}\right) < 0.$$

Daí, $\langle \nabla f(x), d\rangle \leqslant 0$.

Agora, vamos provar o item (ii). Vamos supor que $\langle \nabla f(x), d\rangle < 0$. Sendo assim, pela diferenciabilidade de f no ponto x , temos

$$f(x + td) = f(x) + t\langle \nabla f(x), d\rangle + o(t).$$

Dividindo por $t > 0$,

$$f(x + td) - f(x) = t\langle \nabla f(x), d\rangle + \frac{o(t)}{t}.$$

Levando em conta que $\langle \nabla f(x), d\rangle < 0$ e utilizando $t > 0$ suficientemente pequeno, obtemos

$$\langle \nabla f(x), d\rangle + \frac{o(t)}{t} \leqslant \frac{1}{2}\langle \nabla f(x), d\rangle < 0.$$

Dessa forma, para $t > 0$ suficientemente pequeno, temos que

$$f(x + td) - f(x) < 0,$$

ou seja,

$$f(x + td) < f(x).$$

Logo, $d \in D_f(x)$. $\square$

Teorema A.7. *(Condição necessária de primeira ordem) Suponhamos que a função* $f : \mathbb{R}^n \to \mathbb{R}$ *seja diferenciável no ponto* $\bar{x} \in \mathbb{R}^n$. *Se* $\bar{x}$ *é um minimizador local irrestrito de* f, *então*

$$\nabla f(\bar{x}) = 0.$$

Teorema A.8. *(Convexidade de conjuntos de nível de funções convexas) Suponhamos que o conjunto* $D \subset \mathbb{R}^n$ *seja convexo e a função* $f : D \to \mathbb{R}$ *seja convexa em* D. *Então o conjunto de nível,*

$$S_{lev}(c) = \{x \in D \mid f(x) \leqslant c\},$$

é convexo para todo $c \in \mathbb{R}$.

Proposição A.1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa e* $\bar{x}$ *tal que* $\nabla f(\bar{x}) = 0$. *Então* $\bar{x}$ *é um minimizador de f.*

Demonstração. Temos que f é convexa, logo vale a desigualdade do gradiente (A.1)

$$f(y) - f(x) \geqslant \langle \nabla f(x), y - x \rangle,$$

substituindo $x = \bar{x}$ temos que para todo $y \in \mathbb{R}^n$,

$$f(y) - f(\bar{x}) \geqslant \langle \nabla f(\bar{x}), y - \bar{x} \rangle,$$

como $\nabla f(\bar{x}) = 0$, temos

$$f(y) \geqslant f(\bar{x}).$$

Logo, $\bar{x}$ é ponto de mínimo de f. $\square$

Definição A.8. *(Subgradiente) Seja* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *uma função convexa. Dizemos que* $y \in \mathbb{R}^n$ *é um subgradiente de* f *no ponto* $x \in \mathrm{dom}(f)$ *se,*

$$f(z) \geqslant f(x) + \langle y, z - x \rangle, \quad \forall z \in \mathbb{R}^n.$$

O conjunto de todos os subgradientes de f *em x, denotado por* $\partial f(x)$, *se chama o subdiferencial de* f *em x.*

Proposição A.2. *(Monotonicidade do subdiferencial) Seja* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *uma função convexa. Então,*

$$\langle u - v, x - y \rangle \geqslant 0, \quad \forall u \in \partial f(x), \ \forall v \in \partial f(y).$$

Fato A.1. *([17] Teorema 4.7.1 e Proposição 4.2.1(i)) O operador subdiferencial* ∂f *é maximalmente monótono, ou seja, não possui extensão monótona própria no sentido de inclusão de gráfico. Além disso, o gráfico de* ∂f, $\mathrm{Gph}(\partial f) = \{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n | v \in \partial f(x)\}$ *é semi-fechado, ou seja, se a sequência* $(x^k, v^k) \subset \mathrm{Gph}(\partial f)$ *satisfaz que* (x^k) *converge para* x *e* (v^k) *converge para* v *então* $(x, v) \in \mathrm{Gph}(\partial f)$.

Considerando $z \in \mathbb{R}^n$ como $z = x + \alpha d$, onde $d \in \mathbb{R}^n$, $\alpha > 0$, a definição de subgradiente pode ser escrita da maneira seguinte:

$$y \in \partial f(x) \iff \frac{f(x + \alpha d) - f(x)}{\alpha} \geqslant \langle y, d \rangle, \quad \forall d \in \mathbb{R}^n, \quad \forall \alpha > 0.$$

Quando a função é não-decrescente e possui um limite quando $\alpha \to 0+$. Temos que a desigualdade anterior é equivalente à condição seguinte:

$$f'(x; d) = \lim_{\alpha \to 0+} \frac{f(x + \alpha d) - f(x)}{\alpha} \geqslant \langle y, d \rangle, \quad \forall d \in \mathbb{R}^n.$$

Sendo assim, temos as seguintes definições equivalentes do subdiferencial:

$$\begin{aligned} \partial f(x) &= \{y \in \mathbb{R}^n \mid f(z) \geqslant f(x) + \langle y, z - x \rangle \quad \forall z \in \mathbb{R}^n\} \\ &= \{y \in \mathbb{R}^n \mid f'(x; d) \geqslant \langle y, d \rangle \quad \forall d \in \mathbb{R}^n\}. \end{aligned}$$

Teorema A.9. *(O subdiferencial de função convexa) Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa. Então para todo* $x \in \mathbb{R}^n$*, o conjunto* $\partial f(x)$ *é convexo, compacto e não-vazio. Além disso, para todo* $d \in \mathbb{R}^n$*, tem-se*

$$f'(x; d) = \max_{y \in \partial f(x)} \langle y, d \rangle.$$

Teorema A.10. *(Condição de otimalidade para minimização de uma função convexa num conjunto convexo) Sejam* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa e* $D \subset \mathbb{R}^n$ *um conjunto convexo. Então* $\bar{x} \in D$ *é um minimizador de* f *em* D *se, e somente se,*

$$\exists\ y \in \partial f(\bar{x}) \ \textit{tal que} \ \langle y, x - \bar{x} \rangle \geqslant 0, \quad \forall x \in D,$$

ou, equivalentemente,

$$0 \in \partial f(x) + N_D(x).$$

Em particular, $x \in \mathbb{R}^n$ é um minimizador de f no $\mathbb{R}^n$ se, e somente se,

$$0 \in \partial f(x). \tag{A.2}$$

Referências Bibliográficas

[1] Armijo, L.: Minimization of functions having Lipschitz continuous first partial derivatives, Pacific J. Math. **16**, 1–3 (1966)

[2] Attouch, H., and Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method isactually $o(k^{-2})$, preprint (2015)

[3] Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. CMS Books in Mathematics/Ouvrages de mathématiques de la SMC. New York(NY), Springer (2017)

[4] Bauschke, H. H., and Patrick L., Combettes. "Correction to: convex analysis and monotone operator theory in Hilbert spaces". Convex analysis and monotone operator theory in Hilbert spaces. Cham: Springer International Publishing, C1-C4 (2020)

[5] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)

[6] Bauschke, H.H., Bolte, J., Teboulle, M.: A descent lemma beyond Lipschitz gradient continuity: first order methods revisited and applications. Math. Oper. Res. **42**, 330–348 (2017)

[7] Beck, A, Teboulle, M.: Gradient-based algorithms with applications to signal recovery problems, in Convex Optimization in Signal Processing and Communications, D. Palomar and Y. Eldar, eds., University Press, Cambridge, 42–88 (2010)

[8] Bello Cruz, J.Y., de Oliveira, W.: On weak and strong convergence of the projected gradient method for convex optimization in real Hilbert spaces. Numer. Funct. Anal. Optim. **37**(2), 129–144 (2016)

[9] Bello Cruz, J.Y., Nghia, T.T.A.: On the convergence of the forward-backward splitting method with linesearches. Optim. Methods Softw. **31**(6), 1209–1238 (2016)

[10] Bello Cruz, J.Y., LI, G; Nghia, T.T.A.: On the linear convergence of forward–backward splitting method: Part I-convergence analysis. J. Optim. Theory Appl. **188**(2), 378-401 (2021)

[11] Bello Cruz, J.Y., Melo, J. G., Serra, R. V. G.: A proximal gradient splitting method for solving convex vector optimization problems. Optimization, **71**(1), 33-53 (2022)

[12] Bello Cruz, J.Y., Melo, J. G., Prudente, L. F., Serra, R. V. G.: A proximal gradient method with an explicit line search for multiobjective optimization. Computational Optimization and Applications, **92**(2), 437-469 (2025)

[13] Bello Cruz, J.Y., Gonçalves, M. L., Melo, J. G., Mohr, C.: A relative inexact proximal gradient method with an explicit linesearch. Journal of Optimization Theory and Applications, **206**(1), 9 (2025)

[14] Bertsekas, D.: Nonlinear Programming, Athena Scientific, Belmont (1995)

[15] Bonettini, S., Ruggiero, V.: On the convergence of primal-dual hybrid gradient algorithms for total variation image restoration. J. Math. Imaging Vision **44**(3), 236–253 (2012)

[16] Bredies, K.: A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space. Inverse Probl. **25**(1), 015005 (2009)

[17] Burachik, R.S., and Iusem, A.N.: Set-Valued Mappings and Enlargements of Monotone Operators, Springer, Berlin (2008)

[18] Chambolle, A., Dossal, C.: On the convergence of the iterates of FISTA, preprint (2014)

[19] Csiszár, I.: Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. Ann. Stat. **19**, 2032–2066 (1991)

[20] Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. In: Splitting Methods in Communications, Image Science, and Engineering. Scientific Computation, Springer, Cham (2016)

[21] Dong, Y.: The proximal point algorithm revisited, J. Optim. Theory Appl. **161**, 478–489 (2014)

[22] Dong, Y.: Comments on 'the proximal point algorithm revisited', J. Optim. Theory Appl. **166**, 343–349 (2015)

[23] Guler, O.: On the convergence of the proximal point algorithm for convex minimization, SIAM J. Optim. **29**, 403–419 (1991)

[24] Iusem, A.: Métodos de ponto proximal em otimização. IMPA (1995)

[25] Iusem, A.N., Svaiter, B.F., and Teboulle, M.: Entropy-like proximal methods in convex programming, Math.Oper. Res. **19**, 790–814 (1994)

[26] Izmailov, A. Mikhail S.:"Otimizaçao, volume i."Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade. IMPA, Rio de Janeiro, 2 (2007)

[27] Lima, E. L.: Curso de Análise, vol. 2.(6a ediçao). Projeto Euclides, IMPA (2000)

[28] Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)

[29] Ribeiro, A. A., Karas E.W: "Um curso de otimização."Universidade de Curitiba, Curitiba, Brasil (2001)

[30] Schuster, T., Kaltenbacher, B., Hofmann, B., Kazimierski, K.S.: Regularization Methods in Banach Spaces. De Gruyter, Berlin (2012)

[31] Salzo, S., Masecchia, S., Verri, A., Barla, A.: Alternating proximal regularized dictionary learning. Neural Comput. **26**(12), 2855–2895 (2014)

[32] Salzo, S.: The variable metric forward–backward splitting algorithm under mild differentiability assumptions. SIAM J. Optim. **27**(4), 2153–2181 (2017)

[33] Vardi, Y., Shepp, L.A., Kaufman, L.: A statistical model for positron emission tomography. J. Am. Stat. Assoc. **80**(386), 8–37 (1985)