



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

MARCUS VINÍCIUS DE MORAIS CHAGAS

**Complexidade por Iteração do Método
HPE e sua Versão Acelerada para
Otimização Convexa**

Goiânia
2023



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

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

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

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

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

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

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

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

2. Nome completo do autor

Marcus Vinícius de Moraes Chagas

3. Título do trabalho

Complexidade por iteração do método HPE e sua versão acelerada para otimização convexa

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

Concorda com a liberação total do documento SIM NÃO¹

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

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

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

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

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

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



Documento assinado eletronicamente por **Jefferson Divino Gonçalves De Melo, Professor do Magistério Superior**, em 11/04/2023, às 12:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marcus Vinícius De Moraes Chagas, Discente**, em 11/04/2023, às 13:32, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3658377** e o código CRC **0082A194**.

MARCUS VINÍCIUS DE MORAIS CHAGAS

Complexidade por Iteração do Método HPE e sua Versão Acelerada para Otimização Convexa

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

Área de concentração: Otimização.

Orientador: Prof. Dr. Jefferson Divino Gonçalves de Melo

Goiânia
2023

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

Chagas, Marcus Vinícius de Moraes

Complexidade por iteração do método HPE e sua versão acelerada
para otimização convexa [manuscrito] / Marcus Vinícius de Moraes
Chagas. - 2023.

LXIII, 63 f.

Orientador: Prof. Dr. Jefferson Divino Gonçalves de Melo .

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto
de Matemática e Estatística (IME), Programa de Pós-Graduação em
Matemática, Goiânia, 2023.

Bibliografia.

Inclui lista de figuras.

1. Otimização convexa. 2. Zeros de operadores monótonos
maximais. 3. Método do ponto proximal. 4. Método HPE. 5. Método
A-HPE. I. , Jefferson Divino Gonçalves de Melo, orient. II. Título.

CDU 51



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 02 da sessão de Defesa de Dissertação de **Marcus Vinícius de Moraes Chagas**, que confere o título de Mestre em Matemática, **na área de Otimização**.

Ao décimo dia do mês de abril do ano de dois mil e vinte e três, a partir das nove horas, via Web Videoconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada "**Complexidade por iteração do método HPE e sua versão acelerada para otimização convexa**." Os trabalhos foram instalados pelo presidente da banca e orientador, Professor Doutor **Jefferson Divino Gonçalves de Melo - IME/UFG** com a participação dos demais membros da Banca Examinadora: Professor Doutor **Max Leandro Nobre Gonçalves - IME/UFG**, membro titular interno, **Glaydston de Carvalho Bento - IME/UFG**, membro titular interno e o Professor Doutor **Maicon Marques Alves - DMAT/UFSC**, membro titular externo. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Jefferson Divino Gonçalves de Melo - IME/UFG**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao décimo dia do mês de abril do ano de dois mil e vinte e três.

TÍTULO SUGERIDO PELA BANCA

Complexidade por iteração do método HPE e sua versão acelerada para otimização convexa



Documento assinado eletronicamente por **Jefferson Divino Gonçalves De Melo, Professor do Magistério Superior**, em 11/04/2023, às 08:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Max Leandro Nobre Gonçalves, Professor do Magistério Superior**, em 11/04/2023, às 17:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maicon Marques Alves, Usuário Externo**, em 12/04/2023, às 09:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Glaydston De Carvalho Bento, Professor do Magistério Superior**, em 12/04/2023, às 22:39, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3565960** e o código CRC **4B4B826F**.

Referência: Processo nº 23070.011682/2023-52

SEI nº 3565960

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Marcus Vinícius de Morais Chagas

Graduou-se em Licenciatura em Matemática na UFG - Universidade Federal de Goiás em 2018.

Dedico este trabalho a Deus, aos familiares e amigos.

Agradecimentos

A Deus que me deu sabedoria para elaborar essa dissertação e sempre esteve comigo.

Aos meus pais Lázaro e Maria Lina que acreditaram em mim e sempre me dando apoio.

A minha esposa Rossineide, minha sobrinha Bárbara e meu cunhado Clemilson que estiveram ao meu lado me dando apoio.

Ao bispo Paulo Henrique e os membros da minha igreja que intercederam por mim.

A minha mãe na fé e intercessora Pra. Divina Vitor.

Ao meu orientador Prof. Dr. Jefferson Melo pelos ensinamentos, paciência e pelo tempo dedicado.

Ao Prof. Dr. Glaydston de Carvalho Bento, Prof. Dr. Max Leandro Nobre Gonçalves, o Prof. Dr. Maicon Marques Alves e o Prof. Dr. Leandro da Fonseca Prudente (mesmo sendo como suplente) por aceitarem o convite para participar da Banca Examinadora dessa dissertação.

A todos os professores, funcionários e amigos do Instituto de Matemática e Estatística da UFG, especialmente a Profa. Dra. Maria Bethânia que a considero como minha mãe no IME sempre me incetivando a nunca desistir e sou grato novamente a Prof.Dr. Glaydston com seus conhecimentos e sugestões enviando *paper* pra mim.

A meu amigo Prof. Ms. Marcos Gomes que contribuiu para confecção das figuras e correções de português desta dissertação e sempre me animando a nunca desistir, e agradeço a Edivaldo e Mirelli Uana que me ajudaram durante o curso com as disciplinas.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”.

Marthin Luther King,
líder que lutou contra racismo no EUA..

Resumo

Chagas, Marcus V.M. **Complexidade por Iteração do Método HPE e sua Versão Acelerada para Otimização Convexa**. Goiânia, 2023. 68p. Dissertação de Mestrado. Programa de Pós-Graduação de Matemática, Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Neste trabalho analisamos o método *Hybrid Proximal Extragradient* (HPE) para encontrar zeros de operador monótonos maximais e sua versão acelerada *Accelerated Hybrid Proximal Extragradient* (A-HPE) para resolver problemas de otimização convexa cuja função objetivo é dada pela soma de duas outras funções convexas, uma diferenciável com gradiente Lipschitz e outra não necessariamente diferenciável.

O método HPE foi proposto por Solodov e Svaiter e consiste em uma versão inexata do método do ponto proximal tendo seus subproblemas proximais resolvidos de forma aproximada utilizando um critério de erro relativo seguido por uma atualização dos iterados por meio de um passo do tipo extragradiente. O HPE pode ser visto também como um framework, no sentido que vários outros métodos para minimizar funções convexas e mais geralmente para encontrar zeros de operadores monótonos maximais podem ser vistos como casos especiais do método HPE, tais como os métodos extragradiente, Newton regularizado, ADMM, etc.

Neste trabalho, analisaremos tanto a convergência assintótica do método HPE quanto sua complexidade por iteração. Também analisaremos a complexidade por iteração do método A-HPE proposto por Monteiro e Svaiter. O A-HPE é um método acelerado de primeira ordem, isto é, um método que utiliza somente informações dos valores funcionais e da primeira derivada ou subgradientes das componentes da função objetivo e que possui complexidade por iteração ótima.

Palavras-chave

Otimização convexa, zeros de operadores monótonos maximais, método do ponto proximal, método HPE, método A-HPE, complexidade por iteração.

Abstract

Chagas, Marcus V.M. **Iteration Complexity of HPE Method and its Accelerated Version for Convex Optimization**. Goiânia, 2023. 68p. MSc. Dissertation. Programa de Pós-Graduação de Matemática, Insitudo de Matemática e Estatística, Universidade Federal de Goiás.

In this work, we analyze the *Hybrid Proximal Extragradiente* (HPE) method to find zeroes of maximal monotone operators and its accelerated version *Accelerated Hybrid Proximal Extragradient* (A-HPE) to solve convex optimization problems whose objective function is given by the sum of two other convex functions, one differentiable with Lipschitz gradient and another one not necessarily differentiable.

The HPE method was introduced by Solodov and Svaiter, it consists of an inexact version of the proximal point method having its proximal subproblems inexactly solved using a relative error criterion followed by an extragradient step. The HPE can also be seen as a framework, in the sense that many other methods for minimizing convex functions and more generally to find zeroes of maximal monotone operators can be seen as instances of the HPE method, such as the extragradient method, regularized Newton type method, ADMM, etc.

In this work, we will analyze both the asymptotic convergence of the HPE method and its iteration-complexity. We will also analyze the iteration-complexity of the A-HPE method proposed by Monteiro and Svaiter. The A-HPE is a first-order accelerated method, i.e., it uses only information of the functional values and the first derivative or subgradients of the objective function and has optimal iteration-complexity.

Keywords

Convex optimization, Zeroes of maximal monotone operators, Proximal point method, HPE method, A-HPE method, iteration-complexity.

Sumário

Lista de Figuras	14
1 Introdução e Preliminares	16
1.1 Introdução	16
1.2 Estrutura da Dissertação	18
1.3 Preliminares	19
1.3.1 Análise Convexa	19
1.3.2 Operador Ponto-Conjunto	23
2 Método do Ponto Proximal	30
2.1 Análise de Convergência	31
2.2 Método do Ponto Proximal para Operadores Monótonos Maximais	33
2.2.1 Convergência do Método do Ponto Proximal	34
3 Método HPE e sua complexidade	36
3.1 Um Conceito de Solução Aproximada	36
3.1.1 Método HPE	39
Exemplo:	40
3.2 Análise de Convergência	43
3.3 Resultados de Complexidade para o Método HPE	46
3.4 Método do Korpelevich	49
4 Método A-HPE para Otimização Convexa	52
4.1 Método A-HPE e sua Complexidade por Iteração	52
4.2 Método A-HPE para uma Classe Especial de Problemas Convexos	61
5 Conclusão	65
Referências Bibliográficas	67

Lista de Figuras

1.1	D_1 é um Conjunto convexo, D_2 é um conjunto não-convexo	19
1.2	Função Convexa	20
1.3	y^1 e y^2 são exemplos de subgradiente de f no ponto x	21
1.4	Gráfico da aplicação S_3 .	29

Notações

Segue a lista de notações usadas ao longo dessa dissertação:

1. \mathbb{N} é o conjunto dos números naturais;
2. \mathbb{R} é o conjunto dos números reais;
3. \mathbb{R}_+ é o conjunto dos números reais não-negativos;
4. \mathbb{R}^n é o espaço Euclidiano cujos elementos são da forma $x = (x_1, x_2, \dots, x_n)$, $x_i \in \mathbb{R}$, $1 \leq i \leq n$;
5. $\mathbb{R}_+^n = \{(x_1, x_2, \dots, x_n) | x_i \geq 0, 1 \leq i \leq n\}$;
6. $\mathbb{R}_{++}^n = \{(x_1, x_2, \dots, x_n) | x_i > 0, 1 \leq i \leq n\}$;
7. $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ é o conjunto dos números reais estendidos;
8. $\|x\|$ é a norma Euclidiana do vetor $x \in \mathbb{R}^n$, i.e., $\|x\| = \sqrt{x_1^2 + \dots + x_n^2}$;
9. $\|x\|_1$ é a norma da soma do vetor $x \in \mathbb{R}^n$, i.e., $\|x\|_1 = \sum_{i=1}^n |x_i|$;
10. $\langle x, y \rangle$ é o produto interno entre $x, y \in \mathbb{R}^n$, i.e. $\langle x, y \rangle = x_1 y_1 + \dots + x_n y_n$;
11. $\partial_\varepsilon f(x)$ é o ε -subdiferencial da função convexa f em $x \in \mathbb{R}^n$;
12. $\text{Dom } f$ é o domínio da função f .

Introdução e Preliminares

1.1 Introdução

Considere o seguinte problema de otimização convexa

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

onde $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é uma função convexa, semicontínua inferiormente e própria. Segue da condição de otimalidade de primeira ordem e da convexidade de f que $\bar{x} \in \text{Dom} f$ é uma solução do problema acima se, e somente se, $0 \in \partial f(\bar{x})$. Logo, o problema (1-1) pode ser visto como o problema de obter um zero do operador monótono maximal $T := \partial f$. Sendo assim, podemos generalizar o problema (1-1) para o seguinte problema de inclusão monótona: encontrar $\bar{x} \in \mathbb{R}^n$ solução de

$$0 \in T(x), \quad x \in \mathbb{R}^n, \quad (1-2)$$

onde $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é um operador ponto-conjunto monótono maximal.

Um método clássico para resolver o problema (1-1), ou mais geralmente (1-2), é o método do ponto proximal. Este método aplicado para resolver o problema (1-1) gera recursivamente, a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, uma sequência de pontos $\{x_k\} \subset \mathbb{R}^n$ tal que x_k é a única solução do seguinte subproblema proximal

$$x_k := \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2\lambda_k} \|x - x_{k-1}\|^2 \right\},$$

onde $\lambda_k > 0$ é um parâmetro de regularização. Utilizando a condição de otimalidade de primeira ordem para o problema acima, podemos ver que ele é equivalente a x_k ser a única solução do seguinte problema de inclusão

$$0 \in \lambda_k \partial f(x) + x - x_{k-1}. \quad (1-3)$$

Logo a extensão natural do método do ponto proximal para resolver o problema de inclusão monótona (1-2) consiste em a partir de um ponto $x_0 \in \mathbb{R}^n$ gerar uma sequência $\{x_k\}$ tal que x_k seja solução do seguinte subproblema proximal

$$0 \in \lambda_k T(x) + x - x_{k-1}. \quad (1-4)$$

O método do ponto proximal foi inicialmente proposto por Martinet em [5] e ficou bastante popular a partir dos trabalhos de Rockafellar na década de 1970, o qual também propôs e analisou uma versão inexata do método para resolver o problema (1-2), veja [13].

Neste trabalho estamos interessados em analisar o método *Hybrid Proximal Extragradient* (HPE) proposto por Solodov e Svaiter em [8] para obter uma solução do problema (1-2). Este método é uma versão inexata do método do ponto proximal, o qual resolve aproximadamente o subproblema proximal (1-4) utilizando um critério de erro relativo e atualiza o iterado x_k por meio de um passo extragradiente. Especificamente, em sua k -ésima iteração, o método HPE consiste em obter um par $(y_k, v_k) \in \mathbb{R}^n \times \mathbb{R}^n$ juntamente com um escalar $\varepsilon_k > 0$ satisfazendo as seguintes relações:

$$v_k \in T^{\varepsilon_k}(y_k), \quad \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|y_k - x_{k-1}\|^2, \quad x_k = x_{k-1} - \lambda_k v_k, \quad (1-5)$$

onde $\{\lambda_k\} \subset \mathbb{R}_{++}$, $0 \leq \sigma < 1$ e T^ε é o operador alargamento (*enlargement*) de T , definido em [12].

Note que considerando $\sigma = 0$ na desigualdade em (1-5), obtemos que $\varepsilon_k = 0$ e $\lambda_k v_k + y_k - x_{k-1} = 0$. Segue dessa última relação combinada com a definição de x_k em (1-5) que $x_k = y_k$, e portanto como $\varepsilon_k = 0$, segue da inclusão em (1-5) que $v_k \in T(x_k)$.¹ Logo, esse fato combinado novamente com a terceira relação em (1-5) implica que $0 \in \lambda_k T(x_k) + x_k - x_{k-1}$, e portanto x_k é a solução exata do subproblema proximal (1-4).

Analisaremos tanto a convergência assintótica quanto a complexidade por iteração do método HPE. Para uma melhor compreensão dessa análise, uma parte inicial deste trabalho dedica-se a analisar a convergência do método do ponto proximal exato tanto para resolver o problema de otimização convexa (2-1) quanto o problema de inclusão monótona (1-2). Para a análise da complexidade por iteração do HPE, estamos interessados em obter soluções aproximadas do problema (1-2) no seguinte sentido: dado $\eta > 0$, dizemos que $(y, v, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_{++}$ é uma solução η -aproximada de (1-2) se as seguintes relações são satisfeitas

$$v \in T^\varepsilon(y), \quad \|v\| \leq \eta, \quad \varepsilon \leq \eta. \quad (1-6)$$

¹A definição do operador *enlargement* combinada com a maximalidade de T implica que $T^0 = T$, veja [12].

Como veremos, o número de iterações do método HPE para obter uma solução η -aproximada do problema (1-2) é, no pior caso,² $k = \mathcal{O}(1/\eta^2)$.

Na segunda parte dessa dissertação, analisaremos a complexidade por iteração da versão acelerada do método HPE, a saber, o método *Accelerated Hybrid Proximal Extragradient* (A-HPE) proposto por Monteiro e Svaiter em [9] para obter uma solução aproximada de um problema de otimização convexa. De fato, nosso principal interesse é especializar esse método para resolver a seguinte classe de problemas de otimização convexa

$$\min_{x \in \mathbb{R}^n} \{g(x) + h(x)\}, \quad (1-7)$$

onde $g : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa diferenciável e $h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é uma função convexa, semicontínua inferiormente e própria. Como veremos, dada uma tolerância $\eta > 0$, o método A-HPE computa uma solução η -aproximada do problema acima, i.e., uma tripla (y, v, ε) satisfazendo $v \in \partial g_\varepsilon(x) + \nabla h(x)$ com $\|v\| \leq \eta$ e $\varepsilon \leq \eta$, em no máximo $k = \mathcal{O}(1/\eta^{2/3})$ iterações.

1.2 Estrutura da Dissertação

No primeiro capítulo apresentamos algumas definições básicas da análise convexa, bem como definições e resultados sobre operadores monótonos maximais necessários para o desenvolvimento deste trabalho.

O segundo capítulo é dedicado a versão exata do método do ponto proximal tanto para resolver problemas de otimização convexa quanto para obter zeros de operadores monótonos maximais.

No terceiro capítulo apresentamos com mais detalhes o método HPE e analisamos tanto sua convergência assintótica quanto sua complexidade por iteração. Neste capítulo também apresentamos o método de Korpelevich e mostramos que este método pode ser visto como um caso especial do método HPE.

No quarto capítulo apresentamos a versão acelerada do método HPE e discutiremos sua complexidade por iteração. Neste capítulo, consideramos também um método acelerado de primeira ordem que pode ser visto como um caso particular dessa versão acelerada do HPE.

²Cabe-se notar que a terceira condição em (1-6) é satisfeita em no máximo $k = \mathcal{O}(1/\eta)$ iterações.

1.3 Preliminares

Nesta seção apresentamos algumas definições e resultados básicos que serão úteis neste trabalho. Iniciamos com a definição de uma aplicação Lipschitz-contínua. Este conceito será utilizado principalmente quando a aplicação for o gradiente de uma função diferenciável.

Definição 1. Uma aplicação $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é Lipschitz-contínua se existir uma constante $L > 0$ tal que:

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

O lema a seguir apresenta uma importante propriedade de uma função diferenciável com gradiente Lipschitz-contínuo.

Lema 2. Seja $g : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável com gradiente Lipschitz-contínuo com constante L . Então, temos que

$$g(x) \leq g(y) + \langle x - y, \nabla g(y) \rangle + \frac{L}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n. \quad (1-8)$$

1.3.1 Análise Convexa

Nessa seção apresentamos alguns conceitos básicos da análise convexa. Para mais detalhes, veja por exemplo [1] ou [3, Capítulo 3].

Definição 3. (Conjunto Convexo) Seja $D \subset \mathbb{R}^n$. O conjunto D é dito convexo quando para quaisquer $x, y \in D$ e $\alpha \in [0, 1]$, tem-se que $(1 - \alpha)x + \alpha y \in D$.

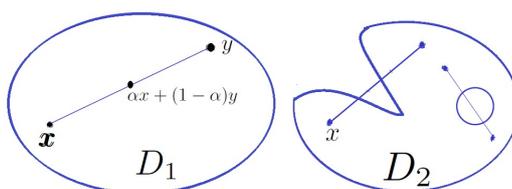


Figura 1.1: D_1 é um Conjunto convexo, D_2 é um conjunto não-convexo

Note que \mathbb{R}^n é um exemplo trivial de conjunto convexo. Note também que a bola fechada em \mathbb{R}^n definida por

$$B(x, r) = \{a \in \mathbb{R}^n \mid \|x - a\| \leq r\}, \quad r > 0,$$

também é um conjunto convexo. De fato, considere $x, y \in \mathbb{R}^n$ tais que $\|x - a\| \leq r$ e $\|y - a\| \leq r$, e assim para $\alpha \in [0, 1]$ temos que

$$\begin{aligned} \|(1 - \alpha)x + \alpha y - a\| &= \|(1 - \alpha)x + \alpha y - (1 - \alpha)a - \alpha a\| \\ &= \|(1 - \alpha)(x - a) + \alpha(y - a)\| \\ &\leq (1 - \alpha)\|x - a\| + \alpha\|y - a\| \\ &\leq (1 - \alpha)r + \alpha r = r. \end{aligned}$$

Em seguida apresentamos a definição do cone normal a um conjunto convexo fechado.

Definição 4. *Seja $D \subset \mathbb{R}^n$ convexo, fechado e não vazio. O cone normal a D em um ponto $x \in D$ é definido por*

$$N_D(x) = \{v \in \mathbb{R}^n, \langle v, y - x \rangle \leq 0, \forall y \in D\}, \text{ e } N_D(x) = \emptyset \text{ se } x \notin D. \quad (1-9)$$

Considere a projeção de um ponto y no conjunto $D \subset \mathbb{R}^n$ convexo, fechado e não-vazio definida por $P_D(y) = \operatorname{argmin}_{x \in D} \|y - x\|$. Então, segue do teorema da projeção (veja por exemplo [3, página 105]) que

$$z - P_D(z) \in N_D(P_D(z)), \quad \forall z \in \mathbb{R}^n. \quad (1-10)$$

Definição 5. *Considere um conjunto convexo não vazio $D \subset \mathbb{R}^n$. Uma função $f : D \rightarrow \overline{\mathbb{R}}$ é dita convexa se para quaisquer $x, y \in D$ e $t \in [0, 1]$, tem-se*

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

A função f é dita *estritamente convexa* quando a desigualdade acima é estrita para todo $x \neq y$ e $t \in (0, 1)$.

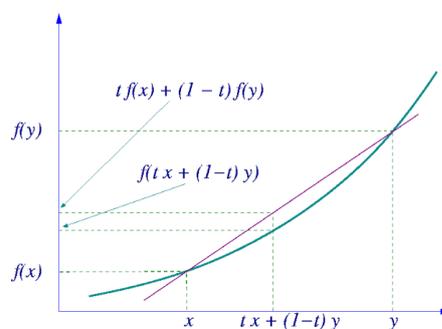


Figura 1.2: Função Convexa

Um exemplo simples de função convexa é a função norma, i.e., $f(x) = \|x\|$.

A função f é dita *fortemente convexa com módulo* $\gamma \geq 0$ se para quaisquer $x, y \in D$ e $\alpha \in [0, 1]$, tem-se

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \gamma \alpha(1 - \alpha)\|x - y\|^2.$$

O resultado a seguir será útil em nossa análise e pode ser facilmente verificado.

Teorema 6. *Sejam $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função convexa e $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função fortemente convexa com módulo $\gamma > 0$. Então $f + g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é fortemente convexa com módulo $\gamma > 0$.*

A seguir apresentamos um resultado que garante que uma função fortemente convexa tem um único minimizador.

Teorema 7. *Seja $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função fortemente convexa, semicontínua inferiormente e própria. Então f possui um único minimizador.*

A seguinte definição introduz o conceito de subgradiente de uma função, o qual pode ser visto como uma generalização do vetor gradiente quando uma função é convexa e não necessariamente diferenciável.

Definição 8. *Seja $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função convexa, semicontínua inferiormente e própria. Um elemento $y \in \mathbb{R}^n$ é dito subgradiente de f no ponto $x \in \mathbb{R}^n$ se*

$$f(z) \geq f(x) + \langle y, z - x \rangle, \forall z \in \mathbb{R}^n. \quad (1-11)$$

O conjunto de todos os subgradientes de f em x , denotado por $\partial f(x)$, é o subdiferencial de f em x .

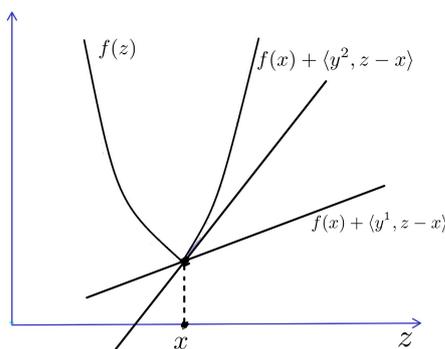


Figura 1.3: y^1 e y^2 são exemplos de subgradiente de f no ponto x

Dado $\varepsilon \geq 0$, um elemento $y \in \mathbb{R}^n$ é um ε -subgradiente de f no ponto $x \in \mathbb{R}^n$ se

$$f(z) \geq f(x) + \langle y, z - x \rangle - \varepsilon, \forall z \in \mathbb{R}^n. \quad (1-12)$$

O conjunto de todos os ε - subgradientes de f em x chama-se ε - subdiferencial de f em x , e é denotado por $\partial_\varepsilon f(x)$.

Para ilustrar o subdiferencial de uma função convexa, considere a função $f: \mathbb{R} \rightarrow \mathbb{R}$ definida por $f(x) = |x|$. Note que o ponto $x = 0$ é único minimizador global de f , mas não é diferenciável em 0. Além disso, segue facilmente da desigualdade triangular que f é convexa. É fácil ver que

$$\partial f(x) = \begin{cases} 1, & x > 0 \\ [-1, 1], & x = 0 \\ -1, & x < 0. \end{cases}$$

Seja D um conjunto convexo, fechado e não vazio. A função indicadora $\delta_D: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é definida como

$$\delta_D(x) = \begin{cases} 0, & x \in D, \\ \infty, & x \notin D. \end{cases} \quad (1-13)$$

Note que se $x \notin D$, então $\partial\delta_D(x) = \emptyset$. De fato, caso contrário, se existisse um $v \in \partial\delta_D(x)$, segue de (1-11) que $\delta_D(y) \geq \delta_D(x) + \langle v, y - x \rangle$ para todo $y \in \mathbb{R}^n$. Tomando $y \in D$ e do fato que $x \notin D$, segue dessa desigualdade e (1-13) que $0 = \delta_D(y) \geq \delta_D(x) + \langle v, y - x \rangle = \infty$, obtendo uma contradição. Agora, se $x \in D$, segue de (1-11) e (1-13) que $v \in \partial\delta_D(x)$ se e somente se $\delta_D(y) \geq \langle v, y - x \rangle$ para todo $y \in \mathbb{R}^n$. Tomando $y \in D$, então temos que $0 = \delta_D(y) \geq \langle v, y - x \rangle$, ou seja $v \in N_D(x)$ e portanto $\partial\delta_D(x) \subset N_D(x)$. Por outro lado se $v \in N_D(x)$, então $0 \geq \langle v, y - x \rangle$ para todo $y \in D$. Logo segue de (1-13) que $\delta_D(y) = 0 \geq \langle v, y - x \rangle$ para todo $y \in D$, além disso se $y \notin D$, então $\delta_D(y) = \infty$ e portanto temos que $\delta_D(y) \geq \langle v, y - x \rangle$ para todo $y \in \mathbb{R}^n$, ou seja $v \in \partial\delta_D(x)$, concluindo portanto que $\partial\delta_D(x) = N_D(x)$.

Em seguida, apresentaremos um resultado que se uma função convexa for diferenciável num ponto $x \in \mathbb{R}^n$, então o subdiferencial dessa função no ponto x contém um único elemento, o gradiente da função no ponto x .

Teorema 9. *Seja $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função convexa própria, e seja $x \in \text{Dom } f$. Então f é diferenciável no ponto x se e somente se o conjunto $\partial f(x)$ possui um único elemento. Neste caso, temos que $\partial f(x) = \{\nabla f(x)\}$.*

Prova. Veja por exemplo [1, Teorema 3.33]. □

O próximo resultado utiliza o subdiferencial de uma função convexa para caracterizar o conjunto dos minimizadores dessa função. Note que essa condição estende para o contexto de função convexa não diferenciável a clássica condição de otimalidade de primeira ordem.

Teorema 10. *Seja $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função convexa, semicontínua inferiormente e própria. Um ponto $\bar{x} \in \mathbb{R}^n$ é um minimizador de f em \mathbb{R}^n se, e somente se,*

$$0 \in \partial f(\bar{x}).$$

Prova. Veja por exemplo [6, Teorema 16.2].

Introduziremos agora a definição de uma sequência Fejér convergente a um conjunto.

Definição 11 ([1], página 205). *Seja $U \subseteq \mathbb{R}^n$ não-vazio. Uma sequência $\{x_k\} \subset \mathbb{R}^n$ é dita Fejér convergente a U se, para todo $u \in U$, temos que*

$$\|u - x_{k+1}\| \leq \|u - x_k\|, \quad \forall k \geq 0. \quad (1-14)$$

Com a definição da sequência Fejér convergente demonstraremos o seguinte lema.

Lema 12. *Seja uma sequência $\{x_k\}$ Fejér convergente a $U \subseteq \mathbb{R}^n$, então $\{x_k\}$ é limitada. Além disso, se existir um ponto de acumulação \bar{x} de $\{x_k\}$ em $U \subseteq \mathbb{R}^n$, então $\{x_k\}$ converge para \bar{x} .*

Prova. Para todo $u \in U$, a desigualdade em (1-14) implica em $\|u - x_k\| \leq \|u - x_0\|$. Logo a sequência $\{x_k\}$ é limitada. Agora assumamos que exista um ponto de acumulação \bar{x} em U e seja $\{x_{k_i}\}$ uma subsequência tal que $\lim_{i \rightarrow \infty} x_{k_i} = \bar{x}$. Como $\bar{x} \in U$ segue de (1-14) que a sequência $\{\|x_{k_i} - \bar{x}\|\}$ é monótona não-crescente e claramente limitada inferiormente por zero. Logo, essa sequência é convergente. Como a subsequência $\{\|x_{k_i} - \bar{x}\|\}$ converge a zero, concluímos que toda a sequência $\{\|x_k - \bar{x}\|\}$ converge a zero. \square

Na seção a seguir, apresentaremos alguns conceitos e resultados sobre operadores ponto-conjunto.

1.3.2 Operador Ponto-Conjunto

Iniciamos com a definição de um operador ponto-conjunto.

Definição 13. *Um operador ponto-conjunto $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é uma relação $T \subseteq \mathbb{R}^n \times \mathbb{R}^n$. Alternativamente, pode-se considerar T como uma aplicação multivalorada de \mathbb{R}^n na família $\wp(\mathbb{R}^n)$ de subconjuntos de \mathbb{R}^n , especificamente*

$$T(z) = \{v \in \mathbb{R}^n \mid (z, v) \in T\} \quad \forall z \in \mathbb{R}^n.$$

A inversa de T é a aplicação $T^{-1} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfazendo a relação $z \in T^{-1}(x) \Leftrightarrow x \in T(z)$. O gráfico de um operador ponto-conjunto $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é definido como

$$Gr(T) = \{(z, v) \in \mathbb{R}^n \times \mathbb{R}^n \mid v \in T(z)\},$$

o seu domínio é o conjunto

$$DomT := \{z \in \mathbb{R}^n \mid T(z) \neq \emptyset\},$$

e sua imagem é definido como

$$ImT = T(\mathbb{R}^n).$$

O operador T é dito monótono se

$$\langle v - \tilde{v}, z - \tilde{z} \rangle \geq 0 \quad \forall (z, v), (\tilde{z}, \tilde{v}) \in Gr(T), \quad (1-15)$$

e T é monótono maximal se é monótono e maximal em relação a ordem parcial de inclusão, ou seja, $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ monótono e $Gr(S) \supset Gr(T)$ implica que $S = T$.

O operador subdiferencial de uma função convexa, própria e semicontínua inferiormente $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é um importante exemplo de operador monótono maximal. A monotonia desse operador pode ser facilmente verificada como segue.

Seja $(x, v), (\tilde{x}, \tilde{v}) \in Gr(\partial f)$, temos por (1-11) que

$$f(\tilde{x}) \geq f(x) + \langle \tilde{x} - x, v \rangle,$$

$$f(x) \geq f(\tilde{x}) + \langle x - \tilde{x}, \tilde{v} \rangle.$$

Somando as desigualdades acima obtemos que

$$\langle \tilde{x} - x, v - \tilde{v} \rangle \geq 0,$$

logo o subdiferencial ∂f é um operador monótono. A maximalidade de ∂f requer uma prova mais elaborada e pode ser encontrada, por exemplo, em [15].

A seguir mostramos que um operador linear monótono é maximal.

Teorema 14. *Seja $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ um operador monótono e linear. Então T é maximal.*

Prova. Seja G um operador monótono tal que $Gr(T) \subset Gr(G)$. Considere $(x, \tilde{x}) \in Gr(G)$. Logo, para todo $z \in \mathbb{R}^n$ e $\rho > 0$, temos que

$$\begin{aligned} 0 &\leq \langle T(x + \rho z) - \tilde{x}, (x + \rho z) - x \rangle = \rho \langle T(x) + \rho T(z) - \tilde{x}, z \rangle \\ &= \rho \langle T(x) - \tilde{x}, z \rangle + \rho^2 \langle T(z) - z \rangle, \end{aligned} \quad (1-16)$$

dividindo (1-16) por ρ e aplicando limite quando $\rho \rightarrow 0$, obtemos

$$\langle T(x) - \tilde{x}, z \rangle = 0, \quad \forall z \in \mathbb{R}^n.$$

Logo, tomando $z = T(x) - \tilde{x}$, temos que $\|T(x) - \tilde{x}\|^2 = 0$, ou seja, $\tilde{x} = T(x)$, concluindo que $(x, \tilde{x}) \in G(T)$ e portanto que T é monótono maximal. \square

A seguir apresentamos um resultado básico sobre continuidade de operador monótono maximal que será útil na análise da convergência do método do ponto proximal no Capítulo 2.

Lema 15. *Seja T um operador monótono maximal e considere sequências $\{y_k\}$ e $\{z_k\}$ tais que $y_k \in T(z_k)$ para todo $k \geq 0$. Se $\lim_{k \rightarrow +\infty} y_k = \bar{y}$ e $\lim_{k \rightarrow +\infty} z_k = \bar{z}$, então $\bar{y} \in T(\bar{z})$.*

Prova. Defina um operador S como

$$S(\bar{z}) = T(\bar{z}) \cup \bar{y}, \quad \text{e} \quad S(z) = T(z), \quad \text{se } z \neq \bar{z}.$$

Como o operador T é monótono e $y_k \in T(z_k)$ para todo k , temos

$$\langle y - y_k, z - z_k \rangle \geq 0, \quad \forall y \in T(z), \quad k \geq 0.$$

Passando o limite nessa desigualdade e utilizando que $\lim_{k \rightarrow +\infty} y_k = \bar{y}$ e $\lim_{k \rightarrow +\infty} z_k = \bar{z}$, temos

$$\langle y - \bar{y}, z - \bar{z} \rangle \geq 0, \quad \forall y \in T(z).$$

Como $S(z) = T(z)$ para $z \neq \bar{z}$ e $\bar{y} \in S(\bar{z})$, segue da desigualdade acima que S é monótono. Como $T(x) \subseteq S(x)$ para todo $x \in \mathbb{R}^n$ e T é monótono maximal, concluimos que $S = T$, e assim $T(\bar{z}) = S(\bar{z}) = T(\bar{z}) \cup \{\bar{y}\}$, ou seja, $\bar{y} \in T(\bar{z})$. \square

Burachik, Iusem e Svaiter introduziram em [12] o ε – *alargamento* de operadores monótonos maximais ponto-conjunto como segue. Dado $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ e um escalar $\varepsilon \geq 0$, ε – *alargamento* de T é o operador $T^\varepsilon : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ definido por

$$T^\varepsilon(z) = \{v \in \mathbb{R}^n \mid \langle v - \tilde{v}, z - \tilde{z} \rangle \geq -\varepsilon, \quad \forall (\tilde{z}, \tilde{v}) \in Gr(T)\}. \quad (1-17)$$

Segue as propriedades do operador T^ε que serão necessárias no decorrer dessa dissertação.

Proposição 16. [12] *Sejam $T, T' : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Então,*

- se $\varepsilon_1 \leq \varepsilon_2$, então $T^{\varepsilon_1}(z) \subseteq T^{\varepsilon_2}(z)$, para todo $z \in \mathbb{R}^n$;*
- $T^\varepsilon(z) + (T')^{\varepsilon'}(z) \subseteq (T + T')^{\varepsilon + \varepsilon'}(z)$, para todo $z \in \mathbb{R}^n$ e $\varepsilon, \varepsilon' \in \mathbb{R}$;*
- T é monótono maximal se, e somente se, $T = T^0$.*

Prova. Para provar o item a) segue de (1-17) que se $v \in T^{\varepsilon_1}(z)$, então

$$\langle v - \tilde{v}, z - \tilde{z} \rangle \geq -\varepsilon_1.$$

Como $\varepsilon_1 \leq \varepsilon_2$, então $-\varepsilon_1 \geq -\varepsilon_2$ e temos

$$\langle v - \tilde{v}, z - \tilde{z} \rangle \geq -\varepsilon_1 \geq -\varepsilon_2 \Rightarrow \langle v - \tilde{v}, z - \tilde{z} \rangle \geq -\varepsilon_2,$$

logo $T^{\varepsilon_1}(z) \subseteq T^{\varepsilon_2}(z)$, e vale igualde se $\varepsilon_1 = \varepsilon_2$.

Para o item b), dado $v \in T^{\varepsilon}(z) + T^{\varepsilon'}(z)$ temos novamente de (1-17) que

$$\begin{aligned} \langle v - \tilde{v}, z - \tilde{z} \rangle &\geq -\varepsilon \\ \langle v - \tilde{v}, z - \tilde{z} \rangle &\geq -\varepsilon', \end{aligned}$$

e somando essas duas desigualdades obtemos

$$2\langle v - \tilde{v}, z - \tilde{z} \rangle \geq -(\varepsilon + \varepsilon') \Rightarrow \langle v - \tilde{v}, z - \tilde{z} \rangle \geq -\frac{\varepsilon + \varepsilon'}{2} \geq -(\varepsilon + \varepsilon').$$

Logo $T^{\varepsilon}(z) + (T')^{\varepsilon'}(z) \subseteq (T + T')^{\varepsilon + \varepsilon'}(z)$, para todo $z \in \mathbb{R}^n$ e $\varepsilon, \varepsilon' \in \mathbb{R}_+$.

Para o item c), temos que pela equação (1-17), basta tomar $\varepsilon = 0$, e como $T \subset T^{\varepsilon}$, $\forall \varepsilon \geq 0$, e pela equação (1-15), que trata da monotonia do operador T , temos que $T = T^0$. A recíproca é trivial. \square

Em seguida apresentamos algumas propriedades envolvendo o ε -alargamento do subdiferencial de uma função convexa $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ e o seu ε -subdiferencial $\partial_{\varepsilon} f$.

Proposição 17. *Seja $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função convexa semicontínua inferiormente e própria. Então*

- a) $\partial_{\varepsilon} f(x) \subset (\partial f)^{\varepsilon}(x)$ para todo $\varepsilon \geq 0$ e $x \in \mathbb{R}^n$, onde $(\partial f)^{\varepsilon}$ representa o ε -alargamento de ∂f ;
- b) Se $v \in \partial f(x)$ e $f(y) < \infty$, então $v \in \partial_{\varepsilon} f(y)$, onde $\varepsilon := f(y) - [f(x) + \langle y - x, v \rangle]$.

Prova. No item a), tome $u \in \partial_{\varepsilon} f(x)$. Então para todo $y \in \mathbb{R}^n$, $v \in \partial f(y)$. Por definição de $\partial_{\varepsilon} f$ e ∂f , temos

$$\varepsilon + \langle u, x - y \rangle \geq f(x) - f(y) \geq \langle v, x - y \rangle.$$

Logo segue de (1-17) com $T = \partial f$ que $u \in (\partial f)^{\varepsilon}$. No item b) como f é convexa e $v \in \partial f(x)$ temos

$$f(y) \geq f(x) + \langle y - x, v \rangle,$$

Multiplicando por 2 em ambos os membros da desigualdade acima e desenvolvendo, obtemos

$$f(y) \geq f(x) + \langle y - x, v \rangle - f(y) + f(x) + \langle y - x, v \rangle$$

o que resulta em

$$f(y) \geq f(x) + \langle y - x, v \rangle - \varepsilon,$$

logo $v \in \partial_\varepsilon f(y)$. □

Note que em geral a inclusão do item a) da Proposição 17 pode ser estrita. De fato, considerando a função $f(x) = -\alpha \sum_{j=1}^n \log(x_j)$ com $x \in \mathbb{R}_{++}^n$, $\alpha > 0$ e $T = \partial f$, temos que $T^\varepsilon(x) \neq \partial_\varepsilon f(x)$ para todo $\varepsilon \geq \alpha n$, veja [12] para mais detalhes. Observe também que o cone normal a um conjunto convexo, fechado e não vazio D pode ser visto como um operador ponto-conjunto $N_D : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Além disso como visto, temos que $N_D = \partial \delta_D$. Em seguida, mostraremos que para este operador vale a inclusão contrária do item a) da Proposição 17 e portanto temos que $N_D^\varepsilon = (\partial \delta_D)^\varepsilon = \partial_\varepsilon \delta_D$ para todo $\varepsilon \geq 0$.

Proposição 18. *Seja D um conjunto convexo, fechado e não vazio. Então, $(\partial \delta_D)^\varepsilon = N_D^\varepsilon = \partial_\varepsilon \delta_D$, para todo $\varepsilon \geq 0$.*

Prova. A primeira inclusão segue imediatamente do fato que $N_D = \partial \delta_D$. Logo, temos pelo item a) da Proposição 17 que $\partial_\varepsilon \delta_D(x) \subset N_D^\varepsilon(x)$ para todo $x \in \mathbb{R}^n$, e portanto basta mostrar que $N_D^\varepsilon(x) \subset \partial_\varepsilon \delta_D(x)$, para todo $x \in \mathbb{R}^n$. Primeiro notamos que se $x \notin D$, então $N_D^\varepsilon(x) = \emptyset$ para todo $\varepsilon \geq 0$. De fato, pela definição do ε -alargamento em (1-17), temos

$$N_D^\varepsilon(x) = \{w \in \mathbb{R}^n : \langle w - v, x - y \rangle \geq -\varepsilon; \quad \forall (y, v) \in Gr(N_D)\}. \quad (1-18)$$

Em particular, tomando $y = P_D(x)$ e $v = v_t := t(x - y)$, temos que $v_t \neq 0, \forall t > 0$ e $v_t \in N_D(y)$ (veja (1-10) e note que $N_D(y)$ é um cone). Logo de (1-18), temos

$$\langle w, x - y \rangle - \langle v_t, x - y \rangle \geq -\varepsilon,$$

ou seja,

$$\langle w, x - y \rangle - t \|x - y\|^2 \geq -\varepsilon.$$

Tomando $t \uparrow +\infty$, obtemos uma contradição. Logo se $x \notin D$, temos que $N_D^\varepsilon(x) = \emptyset, \forall \varepsilon \geq 0$. Portanto basta mostrar que $N_D^\varepsilon(x) \subset \partial_\varepsilon \delta_D(x)$ vale para todo $x \in D$. Note que da definição do ε -subdiferencial em (1-12), temos

$$\partial_\varepsilon \delta_D(x) = \{z \in \mathbb{R}^n : \delta_D(y) \geq \delta_D(x) + \langle z, y - x \rangle - \varepsilon \quad \forall y \in \mathbb{R}^n\}. \quad (1-19)$$

Considere $x \in D$ e $w \in N_D^\varepsilon(x)$ arbitrários. Temos de (1-18) que

$$\langle w - v, x - y \rangle \geq -\varepsilon, \quad \forall (y, v) \in Gr(N_D). \quad (1-20)$$

Como $0 \in N_D(y)$ para todo $y \in D$, temos da desigualdade acima com $v = 0$ que

$$0 \geq \langle w, y - x \rangle - \varepsilon, \quad \forall y \in D.$$

Observe que para $x, y \in D$, temos $\delta_D(y) = \delta_D(x) = 0$ e portanto segue da desigualdade acima que a desigualdade em (1-19) com $z = w$ vale para todo $y \in D$. Como $\delta_D(y) = \infty$ para todo $y \notin D$, concluímos que a desigualdade (1-19) com $z = w$ vale para todo $y \in \mathbb{R}^n$ e portanto $w \in \partial_\varepsilon \delta_D(x)$, ou seja $N_D^\varepsilon(x) \subset \partial_\varepsilon \delta_D(x)$. \square

Apresentamos em seguida a definição de Lipschitz continuidade local de uma aplicação ponto-conjunto. Essa definição será utilizada para a aplicação $S = T^{-1}$.

Definição 19. Uma aplicação $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é dita *L-Lipschitz contínua em zero* se existe um único $z^* \in S(0)$ e para quaisquer sequências $\{u_k\} \subset \mathbb{R}^n$ e $\{z_k\} \subset \mathbb{R}^n$ satisfazendo $(u_k, z_k) \in Gr(S)$ para todo k e $\lim_{k \rightarrow +\infty} u_k = 0$, existe algum $k_0 \geq 0$ tal que

$$\|z_k - z^*\| \leq L\|u_k\|, \quad \forall k \geq k_0. \quad (1-21)$$

Toda aplicação $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ que seja L-Lipschitz contínua é também L-Lipschitz contínua em zero. Como estamos interessados em aplicações monótonas, note que se A é uma matriz simétrica e semidefinida positiva, então a aplicação S_1 definida por $S_1(x) = Ax$ é monótona e Lipschitz contínua com constante $L = \|A\|$, em particular, essa aplicação é $\|A\|$ -Lipschitz contínua em zero. Note que a aplicação $S_2 : \mathbb{R} \rightarrow \mathbb{R}$ definida por $S_2(x) = x^2$ é L-Lipschitz contínua em zero com constante $L = 1$ (note que como $\{u_k\}$ converge para zero, temos que $|z_k| = u_k^2 \leq |u_k|$ para todo k suficientemente grande satisfazendo $|u_k| \leq 1$), mas S_2 não é Lipschitz contínua. Considere a aplicação $S_3 : \mathbb{R} \rightrightarrows \mathbb{R}$ definida por

$$S_3(x) = \begin{cases} -2, & x < -1 \\ [-2, -1], & x = -1 \\ x, & x \in (-1, 0) \\ x^2, & x \geq 0. \end{cases}$$

O gráfico de S_3 é apresentado logo abaixo. Note que S_3 é uma aplicação ponto-conjunto, monótona, Lipschitz contínua em zero com constante $L = 1$, mas não é Lipschitz contínua em \mathbb{R} .

Em seguida, apresentamos a definição de um operador firmemente não-expansivo.

Definição 20. Dizemos que um operador $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ é um operador firmemente não-expansivo quando

$$\|S(x) - S(y)\|^2 \leq \|x - y\|^2 - \|x - y - (S(x) - S(y))\|^2; \quad \forall x, y \in \mathbb{R}^n.$$

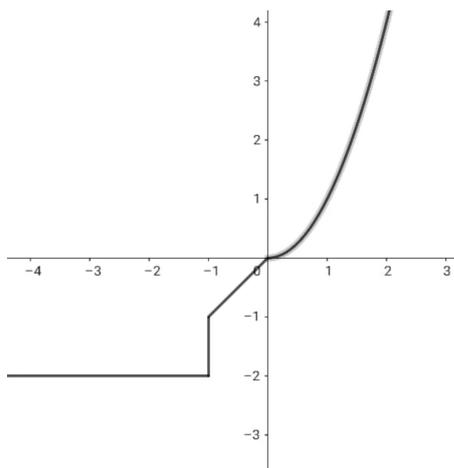


Figura 1.4: Gráfico da aplicação S_3 .

Finalizamos essa seção apresentando o teorema de Minty, o qual é fundamental para mostrar a boa definição do método do ponto proximal no contexto de operadores ponto-conjunto.

Teorema 21 (Minty). *Seja $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ um operador monótono. Então o operador T é monótono maximal se, e somente se, $\text{Im}(I + T) = \mathbb{R}^n$.*

Prova. Veja por exemplo [4, Teorema 4.1] ou [6, Teorema 21.1].

No capítulo a seguir apresentamos e analisamos o método do ponto proximal para resolver problemas de otimização convexa e também, mais geralmente, para obter zeros de operadores monótonos maximais.

Método do Ponto Proximal

O método do ponto proximal originou-se nos trabalhos de Martinet em 1970 [5] e foi bastante analisado e difundido por Rockafellar [14]. Desde então, surgiram vários trabalhos analisando o método ponto proximal, suas variantes e aplicações.

Em seguida apresentaremos esse método para resolver problemas de otimização convexa. Isso nos motivará para introduzir o método no contexto de zero de operadores maximais. Considere o problema de otimização

$$\min_{x \in \mathbb{R}^n} f(x), \quad (2-1)$$

onde $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é uma função convexa, semicontínua inferior, própria e não necessariamente diferenciável. O método do ponto proximal para resolver o problema acima gera, a partir de um ponto inicial $x_0 \in \mathbb{R}^n$ dado, uma sequência de pontos $\{x_k\}$ obtida como solução exata ou aproximada de uma sequência de subproblemas proximais. Em seguida, formalizaremos esse método.

Método do ponto proximal:

Passo 0: Sejam $x_0 \in \mathbb{R}^n$, $\{\lambda_k\} \subset \mathbb{R}_{++}$, e $k = 1$.

Passo 1: Se $0 \in \partial f(x_{k-1})$, então pare;

Passo 2: Calcule

$$x_k = \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2\lambda_k} \|x - x_{k-1}\|^2 \right\}; \quad (2-2)$$

faça $k \leftarrow k + 1$ e vá para o passo 1.

Fim.

Por conveniência, defina, para todo $k \geq 1$, a seguinte função

$$f_k(x) := f(x) + \frac{1}{2\lambda_k} \|x - x_{k-1}\|^2. \quad (2-3)$$

Note que utilizando a condição de otimalidade de primeira ordem para o problema (2-2) (veja Teorema 10) temos que $0 \in \partial f_k(x_k)$, e portanto que

$$0 \in \lambda_k \partial f(x_k) + x_k - x_{k-1}. \quad (2-4)$$

Para justificar a boa definição da sequência $\{x_k\}$ gerada pelo método do ponto proximal, basta notar que a função f_k possui um único minimizador. De fato, como a função f é convexa, e $\|\cdot\|^2$ é fortemente convexa, temos pelo Teorema 6 que a função f_k também é fortemente convexa, e portanto, segue do Teorema 7, que f_k possui um único minimizador.

Na seção a seguir apresentamos a análise de convergência da sequência $\{x_k\}$ gerada pelo método do ponto proximal.

2.1 Análise de Convergência

Nesta seção analisaremos a convergência do método do ponto proximal apresentado na seção anterior. Iniciamos nossa análise por considerar o seguinte resultado básico.

Lema 22. *A sequência $\{x_k\}$ gerada pelo método do ponto proximal (veja (2-2)) satisfaz a seguinte desigualdade*

$$\|x - x_{k+1}\|^2 \leq \|x - x_k\|^2 - \|x_{k+1} - x_k\|^2 + 2\lambda_{k+1}(f(x) - f(x_{k+1})), \quad \forall x \in \mathbb{R}^n, \quad \forall k \geq 0. \quad (2-5)$$

Prova. Primeiramente, note que a seguinte relação vale trivialmente

$$\begin{aligned} \|x - x_k\|^2 &= \|x - x_{k+1} + x_{k+1} - x_k\|^2 \\ &= \|x - x_{k+1}\|^2 - 2\langle x - x_{k+1}, x_k - x_{k+1} \rangle + \|x_{k+1} - x_k\|^2. \end{aligned} \quad (2-6)$$

Por (2-4) temos que

$$\frac{1}{\lambda_{k+1}}(x_k - x_{k+1}) \in \partial f(x_{k+1}).$$

Logo, temos da Definição 8 que

$$f(x) \geq f(x_{k+1}) + \frac{1}{\lambda_{k+1}} \langle x_k - x_{k+1}, x - x_{k+1} \rangle,$$

ou equivalentemente

$$-\langle x_k - x_{k+1}, x - x_{k+1} \rangle \geq \lambda_{k+1}(f(x_{k+1}) - f(x)). \quad (2-7)$$

Substituindo (2-7) em (2-6), e reescrevendo a desigualdade resultante obtemos (2-5). \square

Em seguida analisaremos a convergência da sequência $\{x_k\}$ gerada pelo método do ponto proximal. Assumiremos que o método de fato gere uma sequência infinita, o que é equivalente a dizer, pelo Passo 1, que nenhum ponto da sequência $\{x_k\}$ é solução do problema (2-1).

Teorema 23. *Considere $\{x_k\}$ gerada pelo método do ponto proximal e assumamos que a sequência de passos $\{\lambda_k\}$ satisfaz $\sum_{k=1}^{+\infty} \lambda_k = +\infty$. Então temos que*

$$\lim_{k \rightarrow +\infty} f(x_k) = f^*.$$

Além disso, se o conjunto $X^ \subseteq \mathbb{R}^n$ dos minimizadores de f é não-vazio, então a sequência $\{x_k\}$ converge para algum ponto $x^* \in X^*$.*

Prova. Fazendo $x = x_k$ em (2-5) obtemos como consequência que $\{f(x_k)\}$ é decrescente. Assim afirmamos que $\lim_{k \rightarrow +\infty} f(x_k) = f^*$. Com efeito, se $\lim_{k \rightarrow +\infty} f(x_k) > f^*$, então existem $x \in \mathbb{R}^n$ e $\varepsilon > 0$ tais que

$$f(x) < f(x_k) - \varepsilon, \quad \forall k > 0. \quad (2-8)$$

Segue de (2-5) e (2-8) que

$$\|x - x_{k+1}\|^2 < \|x - x_k\|^2 - 2\lambda_{k+1}\varepsilon, \quad \forall k \geq 0.$$

Reescrevendo essa desigualdade e somando ambos os lados de $k = 1$ até $k = i$ com $i \in \mathbb{N}$ arbitrário, obtemos

$$2 \sum_{k=1}^i \lambda_k \leq \frac{\|x - x_0\|^2}{\varepsilon},$$

o que contradiz a hipótese que $\sum_{k=1}^{+\infty} \lambda_k = +\infty$. Portanto, concluímos que $\lim_{k \rightarrow +\infty} f(x_k) = f^*$.

Agora assumamos que $X^* \neq \emptyset$ e seja $\bar{x} \in X^*$. Então temos que $f(\bar{x}) \leq f(x)$, para todo $x \in \mathbb{R}^n$. Assim, substituindo $x = \bar{x}$ em (2-5) temos que

$$\|\bar{x} - x_{k+1}\|^2 \leq \|\bar{x} - x_k\|^2, \quad \forall k \geq 0,$$

concluindo que a sequência $\{x_k\}$ é Féjer convergente a X^* e, em particular, limitada. Logo $\{x_k\}$ possui uma subsequência convergente. Seja x^* um ponto de acumulação de $\{x_k\}$ e seja $\{x_{k_i}\}$ convergente a x^* . Segue da primeira parte do teorema e pela semicontinuidade inferior de f que

$$f(x^*) \leq \liminf_{i \rightarrow +\infty} f(x_{k_i}) = f^*$$

e assim $x^* \in X^*$. Portanto, como $\{x_k\}$ é Fejér convergente a X^* , segue do Lema 12 que toda a sequência $\{x_k\}$ converge a um elemento de X^* , concluindo a prova do teorema. \square

2.2 Método do Ponto Proximal para Operadores Monótonos Maximais

Nesta seção analisaremos o método do ponto proximal para obter zero de operadores monótonos maximais.

Primeiramente, note que o problema de minimização convexa (2-1) pode, em função do Teorema 10, ser considerado como o problema de inclusão: obter $\bar{x} \in \mathbb{R}^n$ satisfazendo $0 \in \partial f(\bar{x})$. Mais ainda, como vimos anteriormente o operador ∂f é monótono maximal. Logo, podemos considerar uma extensão deste problema de otimização para o problema mais geral de encontrar zero de um operador monótono maximal $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, i.e., obter $\bar{x} \in \mathbb{R}^n$ tal que $0 \in T(\bar{x})$.

Como vimos na seção anterior, a sequência $\{x_k\}$ gerada pelo método do ponto proximal para minimização convexa satisfaz a seguinte relação (veja (2-4))

$$0 \in \lambda_k \partial f(x) + x - x_{k-1}, \quad x \in \mathbb{R}^n.$$

Logo, o método do ponto proximal para obter zero do operador monótono maximal $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ pode ser considerado como sendo um processo iterativo que, a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, gera uma sequência $\{x_k\}$ onde gerado x_{k-1} para algum $k \geq 1$, o próximo elemento x_k é obtido como sendo a única solução do seguinte subproblema proximal

$$0 \in \lambda_k T(x) + x - x_{k-1}, \quad x \in \mathbb{R}^n,$$

ou equivalentemente

$$x_k = (I + \lambda_k T)^{-1}(x_{k-1}),$$

onde $\{\lambda_k\} \subset \mathbb{R}_{++}$ é uma sequência de parâmetros de regularização dada. Note que a boa definição da sequência $\{x_k\}$ acima segue do Teorema de Minty (veja Teorema 21).

Método do ponto proximal para operadores monótonos maximais:

Passo 0: Sejam $x_0 \in \mathbb{R}^n$, $\{\lambda_k\} \subset \mathbb{R}_{++}$, e $k \geq 1$;

Passo 1: Se $0 \in T(x_{k-1})$, então pare;

Passo 2: Encontre x_k solução do seguinte problema

$$0 \in \lambda_k T(x) + x - x_{k-1}, \quad x \in \mathbb{R}^n, \quad (2-9)$$

faça $k \leftarrow k + 1$ e vá para o passo 1.

Fim.

2.2.1 Convergência do Método do Ponto Proximal

Nesta seção analisaremos a convergência da sequência $\{x_k\} \subset \mathbb{R}^n$ gerada pelo método do ponto proximal.

Teorema 24. *Seja $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ um operador monótono maximal e assumamos que exista $\bar{x} \in \mathbb{R}^n$ tal que $0 \in T(\bar{x})$. Então a sequência $\{x_k\}$ definida em (2-9) com $\lambda_k \geq \bar{\lambda} > 0$ converge para algum ponto $x^* \in \mathbb{R}^n$ tal que $0 \in T(x^*)$.*

Prova. Seja $\bar{x} \in \mathbb{R}^n$ tal que $0 \in T(\bar{x})$ e note que a seguinte igualdade vale trivialmente

$$\|x_k - \bar{x}\|^2 = \|x_{k+1} - \bar{x}\|^2 + \|x_k - x_{k+1}\|^2 + 2\langle x_k - x_{k+1}, x_{k+1} - \bar{x} \rangle.$$

Logo, definindo $v_{k+1} := (x_k - x_{k+1})/\lambda_{k+1}$ temos que

$$\|x_{k+1} - \bar{x}\|^2 = \|x_k - \bar{x}\|^2 - \|x_k - x_{k+1}\|^2 - 2\lambda_{k+1} \langle v_{k+1}, x_{k+1} - \bar{x} \rangle. \quad (2-10)$$

Por outro lado, como $0 \in \lambda_{k+1} T(x_{k+1}) + x_{k+1} - x_k$ (veja (2-9)) temos que $v_{k+1} \in T(x_{k+1})$. Como o operador T é monótono e $0 \in T(\bar{x})$, temos

$$\langle v_{k+1}, x_{k+1} - \bar{x} \rangle \geq 0.$$

Logo temos de (2-10) que

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 - \|x_k - x_{k+1}\|^2, \quad \forall k \geq 0. \quad (2-11)$$

E, portanto concluímos que a sequência $\{x_k\} \subset \mathbb{R}^n$ é Fejér convergente ao conjunto $T^{-1}(0) = \{\bar{x} \in \mathbb{R}^n | 0 \in T(\bar{x})\}$ e, em particular, limitada, veja Lema 12. Segue também

da desigualdade acima que

$$\sum_{k=0}^i \|x_k - x_{k+1}\|^2 \leq \sum_{k=0}^i (\|x_k - \bar{x}\|^2 - \|x_{k+1} - \bar{x}\|^2) \leq \|x_0 - \bar{x}\|^2, \quad \forall i \geq 0.$$

Como consequência, temos que

$$\lim_{k \rightarrow +\infty} \|x_k - x_{k+1}\|^2 = 0. \quad (2-12)$$

Como a sequência $\{x_k\}$ é limitada, ela possui ponto de acumulação. Seja $x^* \in \mathbb{R}^n$ um ponto de acumulação de $\{x_k\}$ e considere uma subsequência $\{x_{k_j}\}$ convergente a x^* . Portanto temos de (2-12) que $\lim_k (x_{k_j} - x_{k_j+1}) = 0$ e $\lim_k x_{k_j+1} = x^*$. Logo, como por hipótese $\lambda_k \geq \bar{\lambda} > 0$ temos que $\{v_{k_j+1}\}$ converge a zero, onde $v_{k_j+1} = \frac{1}{\lambda_{k_j+1}}(x_{k_j} - x_{k_j+1})$. Por outro lado, segue de (2-9) que

$$v_{k_j+1} \in T(x_{k_j+1}). \quad (2-13)$$

Como o operador T é monótono maximal, segue de (2-13) e do Lema 15 com $y_j = v_{k_j+1}$, $\bar{y} = 0$, $z_j = x_{k_j+1}$ e $\bar{z} = x^*$, que $0 \in T(x^*)$, ou equivalentemente $x^* \in T^{-1}(0)$. Como $\{x_k\}$ é Fejér convergente ao conjunto $T^{-1}(0)$, segue do Lema 12 que toda a sequência $\{x_k\}$ converge a um ponto de $T^{-1}(0)$. \square

No próximo capítulo apresentamos e analisamos o método HPE, o qual consiste em uma versão inexata do método do ponto proximal para obter zero de operadores monótonos maximais.

Método HPE e sua complexidade

Neste capítulo analisamos o método HPE para obter zeros de um operador monótono maximal. Analisamos tanto as propriedades assintóticas do método quanto sua complexidade por iteração.

3.1 Um Conceito de Solução Aproximada

Nesta seção consideramos o problema de encontrar zeros de um operador monótono maximal $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. No capítulo anterior vimos que a k -ésima iteração do método do ponto proximal requer a solução do subproblema

$$0 \in \lambda_k T(x) + x - x_{k-1}. \quad (3-1)$$

Em geral, resolver o problema acima é difícil, fazendo-se necessário computar soluções aproximadas. Rockafellar em [14] analisou o método do ponto proximal onde a solução x_k do subproblema proximal acima pode ser computada de forma aproximada satisfazendo uma das seguintes condições

$$\text{dist}(0, \lambda_k T(x_k) + x_k - x_{k-1}) \leq e_k, \quad \sum_{k=1}^{\infty} e_k < \infty, \quad (3-2)$$

ou

$$\text{dist}(0, \lambda_k T(x_k) + x_k - x_{k-1}) \leq d_k \|x_k - x_{k-1}\|, \quad \sum_{k=1}^{\infty} d_k < \infty. \quad (3-3)$$

Considerando o operador T^ε como definido em (1-17), temos a seguinte definição de solução aproximada para o subproblema proximal (3-1).

Definição 25. *Seja $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++}$ e $\sigma \in [0, 1)$. Dizemos que um par $(y, v) \in \text{dom} T \times \mathbb{R}^n$ é uma solução aproximada de (3-1) com erro de tolerância σ se existe $\varepsilon \geq 0$ tal que*

$$v \in T^\varepsilon(y), \quad (3-4)$$

$$\|r\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|y - x\|^2, \quad (3-5)$$

onde $r := \lambda v + y - x$.

Seja $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++}$ e considere uma solução exata y do subproblema proximal $0 \in \lambda T(y) + y - x$, isto é, existe $v \in T(y)$ tal que $0 = \lambda v + y - x$ e portanto tomando $\varepsilon = 0$, vemos que o par (y, v) é uma solução aproximada no sentido da Definição 25 com erro de tolerância σ para qualquer $\sigma \in [0, 1)$. Note que neste caso $r = \lambda v + y - x = 0$ e que $T^0 = T$ pois T é monótono maximal, veja o item c) da Proposição 16.

Note também que considerando $\varepsilon = 0$, da Definição 25 temos que as seguintes relações são satisfeitas para uma solução aproximada (y, v) :

$$\begin{aligned} v &\in T(y), & \lambda v + y - x &= r \\ \|r\| &\leq \sigma \|y - x\|. \end{aligned}$$

O seguinte resultado mostra o quão próximo está uma solução aproximada de uma solução exata do subproblema proximal.

Lema 26. *Seja $x \in \mathbb{R}^n$ e $\lambda > 0$, considere uma solução aproximada (y, v) com tolerância $\sigma \in [0, 1)$ do subproblema proximal $0 \in \lambda T(\cdot) + (\cdot - x)$, e seja z a solução exata deste subproblema. Então temos que*

$$\|z - y\| \leq \sigma \|y - x\|. \quad (3-6)$$

Prova. Como (y, v) é uma solução aproximada de $0 \in \lambda T(\cdot) + (\cdot - x)$ com tolerância $\sigma \in [0, 1)$, temos que existe $\varepsilon \geq 0$ tal que

$$v \in T^\varepsilon(y), \quad \|r\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|y - x\|^2, \quad (3-7)$$

$$\lambda v + y - x = r \quad (3-8)$$

Como z é uma solução exata de $0 \in \lambda T(\cdot) + (\cdot - x)$, existe $u \in T(z)$ tal que

$$\lambda u + (z - x) = 0 \quad (3-9)$$

Segue portanto da definição de T^ε em (1-17) que

$$\langle u - v, z - y \rangle \geq -\varepsilon. \quad (3-10)$$

Observe que de (3-8) e (3-9) temos que

$$\begin{aligned} u - v &= -\frac{1}{\lambda}(z - x) - \frac{1}{\lambda}(r - (y - x)) \\ &= -\frac{1}{\lambda}(r + z - y). \end{aligned} \quad (3-11)$$

Combinando (3-10) e (3-11) obtemos $-(1/\lambda)\langle r+z-y, z-y \rangle \geq -\varepsilon$ ou equivalentemente

$$\|z-y\|^2 + \langle r, z-y \rangle \leq \lambda\varepsilon.$$

Segue da desigualdade de Cauchy-Schwarz e a desigualdade acima que

$$\|z-y\|^2 - \|r\|\|z-y\| \leq \lambda\varepsilon.$$

Multiplicando a desigualdade acima por 4 e completando quadrados temos que

$$(2\|z-y\| - \|r\|)^2 \leq \|r\|^2 + 4\lambda\varepsilon,$$

de onde obtemos

$$\|z-y\| \leq \frac{\|r\| + \sqrt{\|r\|^2 + 4\lambda\varepsilon}}{2} \leq \sqrt{\|r\|^2 + 2\lambda\varepsilon}, \quad (3-12)$$

onde a última desigualdade segue da concavidade da função $f(x) = \sqrt{x}$ ¹. Portanto, a conclusão do lema segue diretamente de (3-12) e (3-5). \square

A seguir apresentamos algumas propriedades básicas de soluções aproximadas do subproblema proximal (3-1) no sentido da Definição 25. Este resultado é importante tanto para a elaboração do método HPE quanto para sua análise de convergência.

Lema 27. *Seja $x \in \mathbb{R}^n$, $\lambda > 0$ e (y, v) uma solução aproximada de $0 \in \lambda T(\cdot) + (\cdot - x)$ com tolerância $\sigma \in [0, 1)$, e defina*

$$x^+ = x - \lambda v. \quad (3-13)$$

Então, as seguintes afirmações valem:

a) *Para todo $\tilde{x} \in \mathbb{R}^n$ temos*

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

b) *Para toda solução x^* do problema de inclusão $0 \in T(x)$, temos*

$$\|x^* - x\|^2 - \|x^* - x^+\|^2 \geq (1 - \sigma^2)\|y - x\|^2.$$

¹De fato como f é côncava, então, $\forall x, y \in \mathbb{R}_+$, temos $\sqrt{x/2 + y/2} \geq \sqrt{x}/2 + \sqrt{y}/2$. Para provar a desigualdade desejada, tome $x = \|r\|^2$ e $y = \|r\|^2 + 4\lambda\varepsilon$.

Prova. a) A prova deste item segue diretamente das seguintes relações

$$\begin{aligned}
\|\bar{x} - x\|^2 - \|\bar{x} - x^+\|^2 &= \|\bar{x} - y + y - x\|^2 - \|\bar{x} - y + y - x^+\|^2 \\
&= \|\bar{x} - y\|^2 + 2\langle \bar{x} - y, y - x \rangle + \|y - x\|^2 \\
&\quad - [\|\bar{x} - y\|^2 + 2\langle \bar{x} - y, y - x^+ \rangle + \|y - x^+\|^2] \\
&= 2\langle \bar{x} - y, x^+ - x \rangle + \|y - x\|^2 - \|y - x^+\|^2.
\end{aligned} \tag{3-14}$$

b) Como (y, v) é uma solução aproximada de $0 \in \lambda T(\cdot) + (\cdot - x)$ no sentido da Definição 25, temos que existe $\varepsilon \geq 0$ tal que

$$\begin{aligned}
v &\in T^\varepsilon(y), \\
\|r\|^2 + 2\lambda\varepsilon &\leq \sigma^2\|y - x\|^2,
\end{aligned} \tag{3-15}$$

onde $r = \lambda v + (y - x)$. Como $0 \in T(x^*)$ e $v \in T^\varepsilon(y)$, segue das definições de T^ε e x^+ em (1-17) e (3-13), respectivamente, que

$$\begin{aligned}
\langle x^* - y, x^+ - x \rangle &= \langle x^* - y, -\lambda v \rangle \\
&= \lambda \langle x^* - y, 0 - v \rangle \geq -\lambda\varepsilon.
\end{aligned}$$

Como $r = \lambda v + y - x$ e $x^+ = x - \lambda v$, temos que $y - x^+ = r$ e portanto segue das relações acima do item a) com $\bar{x} = x^*$ e (3-15) que

$$\begin{aligned}
\|x^* - x\|^2 - \|x^* - x^+\|^2 &\geq -2\lambda\varepsilon + \|y - x\|^2 - \|r\|^2 \\
&= \|y - x\|^2 - (\|r\|^2 + 2\lambda\varepsilon) \\
&\geq (1 - \sigma^2)\|y - x\|^2,
\end{aligned}$$

provando o resultado desejado. □

3.1.1 Método HPE

Nesta seção apresentamos e analisamos o método HPE, o qual é formalmente descrito abaixo.

Método HPE:

Passo 0: Seja $x_0 \in \mathbb{R}^n$, $0 \leq \sigma < 1$ e $k = 1$;

Passo 1: Escolha $\lambda_k > 0$ e encontre $y_k, v_k \in \mathbb{R}^n$ e $\varepsilon_k \geq 0$ tais que

$$\begin{aligned} v_k &\in T^{\varepsilon_k}(y_k), \\ \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k &\leq \sigma^2 \|y_k - x_{k-1}\|^2; \end{aligned}$$

Passo 2: Defina

$$x_k = x_{k-1} - \lambda_k v_k,$$

faça $k \leftarrow k + 1$ e retorne ao Passo 1.

Fim.

Observe que se o k -ésimo subproblema do método do ponto proximal (3-1) for resolvido de forma exata, i.e., $0 \in \lambda_k T(x_k) + x_k - x_{k-1}$, então a tripla $(y_k, v_k, \varepsilon_k) := (x_k, (x_{k-1} - x_k)/\lambda_k, 0)$ é uma solução do subproblema do método HPE (Passo 1). Além disso, se na k -ésima iteração do método HPE consideramos $\sigma = 0$, então $\varepsilon_k = 0$, $\lambda_k v_k + y_k - x_{k-1} = 0$, o qual combinado com a atualização de x_k no Passo 2 implica que $y_k = x_k$, e assim segue da inclusão no Passo 1 que $v_k \in T^0(x_k) = T(x_k)$, ou seja, $0 \in \lambda_k T(x_k) + x_k - x_{k-1}$. Portanto, concluímos que se $\sigma = 0$, a iteração do método HPE coincide com a iteração do método do ponto proximal exato.

Em seguida apresentamos um exemplo que mostra que o passo *extragradiente* definido no Passo 2 do método HPE é essencial para a convergência do método.

Exemplo:

Seja $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ o operador definido por

$$T(x) := Mx, \quad M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (3-16)$$

Então $\forall x, \tilde{x} \in \mathbb{R}^2$, $v = T(x)$ e $\tilde{v} = T(\tilde{x})$ segue que

$$\langle x - \tilde{x}, v - \tilde{v} \rangle = \langle x - \tilde{x}, T(x) - T(\tilde{x}) \rangle = \langle x - \tilde{x}, T(x - \tilde{x}) \rangle = 0,$$

onde a última igualdade segue diretamente da definição de T e da matriz M em (3-16). Concluímos portanto que T é um operador monótono, o qual é maximal em virtude do Teorema 14. Afirmamos que para todo $x \in \mathbb{R}^n$ vale $T^\varepsilon(x) = T(x)$. De fato, segue diretamente da definição de T^ε em (1-17) que $T \subset T^\varepsilon$. Logo, basta mostrar que $T^\varepsilon \subset T$.

Seja um par $(x, v) \in T^\varepsilon$, e assim temos que para todo $\tilde{x} \in \mathbb{R}^n$ vale a seguinte desigualdade

$$\langle x - \tilde{x}, v - T(\tilde{x}) \rangle \geq -\varepsilon.$$

Seja $u \in \mathbb{R}^n$ e $\alpha > 0$ arbitrários e tome $\tilde{x} = x - \alpha u$. Logo temos da desigualdade acima que

$$\langle \alpha u, v - T(x - \alpha u) \rangle \geq -\varepsilon.$$

Usando que T é linear, concluímos que

$$\alpha \langle u, v - T(x) \rangle + \alpha^2 \langle u, T(u) \rangle \geq -\varepsilon.$$

Note que segue da definição de T em (3-16) que $\langle u, T(u) \rangle = 0$. Portanto, dividindo a última desigualdade acima por α e tomando o limite quando α tende a $+\infty$, obtemos

$$\langle u, v - T(x) \rangle \geq -\varepsilon \lim_{\alpha \rightarrow +\infty} \frac{1}{\alpha} = 0.$$

Como u é um elemento arbitrário, podemos tomar $u = T(x) - v$ e concluímos da desigualdade acima que $\langle u, -u \rangle \geq 0$. Logo $-\|u\|^2 \geq 0$ de onde tem-se $\|u\| = 0$, o qual implica que $u = 0$. Logo, $v = T(x)$. Portanto $T^\varepsilon \subset T$.

Como u é um elemento arbitrário em \mathbb{R}^n , tomando $u = T(x) - v$, temos da desigualdade acima que $-\|v - T(x)\|^2 \geq 0$, o qual implica que $v = T(x)$. Portanto, concluímos que $T^\varepsilon \subset T$. Ou seja mostramos que para o operador T definido em (3-16) que $T^\varepsilon = T$.

A seguir apresentaremos um algoritmo que pode ser visto como um caso particular do HPE em um contexto especial onde não consideramos o passo extragradiente do Passo 2 do método HPE, e mostraremos que esse algoritmo gera uma sequência divergente. Isso ilustra a importância de considerar o passo extragradiente no método HPE. Mais precisamente apresentaremos uma sequência $\{(x_k, w_k)\}$ satisfazendo as seguintes relações

$$\begin{aligned} w_k &\in T(x_k), \\ \|\lambda_k w_k + x_k - x_{k-1}\| &\leq \sigma \|x_k - x_{k-1}\|. \end{aligned} \tag{3-17}$$

Em particular, definindo $(v_k, y_k) = (w_k, x_k)$ para todo $k \geq 0$, vemos que a sequência $\{(y_k, v_k)\}$ satisfaz o critério de erro do método HPE, e com isso podemos ver esse procedimento como um caso particular do HPE.

Para definir a sequência $\{(x_k, w_k)\}$ mencionada acima, considere o operador T definido em (3-16), e tome $\sigma \in (1/\sqrt{2}, 1)$, $\lambda_k = 1/2, \forall k$ e $x_0 = (a_0, b_0) \neq (0, 0)$. Além disso

considere a seguinte matriz

$$Q := \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$

e define as duas sequências $\{x_k\}$ e $\{w_k\}$ da seguinte forma

$$x_k := Q^k x_0, \quad w_k := Mx_k \in T(x_k),$$

onde Q^k denota a k -ésima potência de Q . Considerando $x_k = (a_k, b_k)$ para todo $k \geq 0$, temos desse fato e das definições acima que

$$x_k = Q^k x_0 = Qx_{k-1} = (a_{k-1} + b_{k-1}, -a_{k-1} + b_{k-1}),$$

e assim temos que

$$\|x_k\| = \sqrt{2} \sqrt{a_{k-1}^2 + b_{k-1}^2} = \sqrt{2} \|x_{k-1}\|.$$

Logo concluímos que $\|x_k\| = \sqrt{2}^k \|x_0\|$ para todo $k \geq 0$, e portanto a sequência $\{x_k\}$ diverge. Mostraremos que mesmo que $\{x_k\}$ diverja, essa sequência juntamente com $\{w_k\}$ satisfazem a relação (3-17). Como $x_k = Qx_{k-1}$, obtemos

$$\begin{aligned} x_k - x_{k-1} &= (Q - I)x_{k-1}, \\ \lambda_k w_k + x_k - x_{k-1} &= ((1/2)MQ + Q - I)x_{k-1}, \end{aligned}$$

onde I é uma matriz identidade. Observe que

$$(1/2)MQ + Q - I = \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix},$$

e assim,

$$\|\lambda_k w_k + x_k - x_{k-1}\| = \|(1/2)MQx_{k-1} + (Q - I)x_{k-1}\| = \|((1/2)MQ + Q - I)x_{k-1}\|.$$

De onde temos que

$$\|\lambda_k w_k + x_k - x_{k-1}\| = (1/\sqrt{2}) \|x_k - x_{k-1}\| \leq \sigma \|x_k - x_{k-1}\|.$$

Logo a sequência $\{(y_k, v_k)\}$ com $(y_k, v_k) = (x_k, w_k)$ para todo k satisfaz a relação (3-17). Portanto, concluímos que $\{(x_k, w_k)\}$ pode ser visto como um caso particular do método HPE e que a sequência $\{x_k\}$ diverge. Isso mostra que o passo extragradiente do método HPE é fundamental para a convergência do método.

3.2 Análise de Convergência

Iniciamos essa seção apresentando alguns resultados básicos sobre as sequências geradas pelo método HPE.

Lema 28. *Se x^* é solução de (1-2) então*

a) *a seguinte desigualdade vale para todo $k \geq 1$:*

$$\|x^* - x_{k-1}\|^2 - \|x^* - x_k\|^2 \geq (1 - \sigma^2) \|y_k - x_{k-1}\|^2;$$

b) *a sequência $\{\|x^* - x_k\|\}$ é não-crescente e*

$$\|x^* - x_0\|^2 \geq (1 - \sigma^2) \sum_{k=1}^{\infty} \|y_k - x_{k-1}\|^2.$$

Prova. O item a) segue diretamente do item b) do Lema 27 com $x = x_{k-1}$, $y = y_k$ e $x^+ = x_k$. Para demonstrar o item b), pelo Passo 0 do método HPE temos que $0 \leq \sigma < 1$, e aplicando a somatória em ambos os lados da desigualdade do item a) de $k = 1$ a $k = i$, temos que

$$\|x^* - x_0\| \geq \|x^* - x_0\| - \|x^* - x_i\| \geq (1 - \sigma^2) \sum_{k=1}^i \|y_k - x_{k-1}\|^2, \quad (3-18)$$

tomando limite em ambos os lados da desigualdade (3-18) quando i tende a infinito, obtemos a desigualdade do item b). Agora note que $0 \leq \sigma < 1$ e portanto segue do item a) que

$$\|x^* - x_{k-1}\|^2 - \|x^* - x_k\|^2 \geq (1 - \sigma^2) \|y_k - x_{k-1}\|^2 \geq 0,$$

o que implica que a sequência $\{\|x^* - x_k\|\}$ é não-crescente. \square

Lema 29. *Se o problema (1-2) tem solução, então*

a) *a sequência $\{x_k\}$ é limitada;*

b) *temos que $\sum_{k=1}^{\infty} \|y_k - x_{k-1}\|^2 < \infty$;*

c) *a sequência $\{y_k\}$ é limitada;*

d) *se existir $\bar{\lambda} > 0$ tal que $\lambda_k \geq \bar{\lambda} > 0$ para todo $k \geq 1$, temos*

$$\sum_{k=1}^{\infty} \|v_k\|^2 \leq (1 + \sigma)^2 \sum_{k=1}^{\infty} \lambda_k^{-2} \|y_k - x_{k-1}\|^2 < \infty, \quad \sum_{k=1}^{\infty} \varepsilon_k < \infty. \quad (3-19)$$

Prova. Seja x^* uma solução de (1-2). Pelo item b) do Lema 28, a sequência $\{\|x_k - x^*\|\}$ é não-crescente. Logo, $\{\|x_k\|\}$ é limitado. O item b) segue diretamente do item b) do

Lema 28 e do fato que $\sigma \in (0, 1)$. O item c) segue diretamente dos itens a) e b). Para o item d) note que o Passo 1 do HPE garante que para todo k vale

$$\|\lambda_k v_k + y_k - x_{k-1}\| \leq \sigma \|y_k - x_{k-1}\|,$$

e portanto pela desigualdade triangular obtemos

$$\begin{aligned} \lambda_k \|v_k\| &\leq \|\lambda_k v_k + y_k - x_{k-1}\| + \|y_k - x_{k-1}\| \\ &\leq (1 + \sigma) \|y_k - x_{k-1}\|. \end{aligned} \quad (3-20)$$

Segue portanto do item b) e da hipótese $\inf_k \lambda_k \geq \bar{\lambda} > 0$, que

$$\sum_{k=1}^{\infty} \|v_k\|^2 \leq (1 + \sigma)^2 \sum_{k=1}^{\infty} \lambda_k^{-2} \|y_k - x_{k-1}\|^2 \leq (1 + \sigma)^2 \sum_{k=1}^{\infty} \bar{\lambda}^{-2} \|y_k - x_{k-1}\|^2 < \infty.$$

Logo vale a primeira desigualdade do item d). De forma análoga, para provar a segunda desigualdade do item d), note que o Passo 1 do HPE implica que

$$2\lambda_k \varepsilon_k \leq \sigma^2 \|y_k - x_{k-1}\|^2, \quad \forall k \geq 0.$$

Portanto, o resultado desejado segue devido ao item b) e do fato que $\inf_k \lambda_k \geq \bar{\lambda} > 0$. \square

Lema 30. *Se o problema (1-2) tem solução e existe $\bar{\lambda} > 0$ tal que $\lambda_k \geq \bar{\lambda} > 0$ para todo $k \geq 1$, então todo ponto de acumulação da sequência $\{x_k\}$ é solução do problema (1-2).*

Prova. Pelo item a) do Lema 29 temos que a sequência $\{x_k\}$ é limitada. Portanto, pelo Teorema de Bolzano-Weierstrass possui uma subsequência $\{x_{k_j}\}$ convergente para algum ponto \bar{x} . Pelo item b) do Lema 29 segue que $\{\|y_k - x_{k-1}\|\}$ converge para zero, logo a subsequência $\{y_{k_j+1}\}$ também converge para \bar{x} . Pelo item d) do Lema 29 segue também que $\{(v_k, \varepsilon_k)\}$ converge para zero, assim obtemos que a subsequência $\{(v_{k_j+1}, \varepsilon_{k_j+1})\}$ também converge para zero. Além disso, observe que sendo (y_k, v_k) solução aproximada de (3-1) como na Definição 25, temos que

$$v_k \in T^{\varepsilon_k}(y_k), \quad \forall k \geq 1.$$

Tome $x \in \mathbb{R}^n$ e $u \in T(x)$ arbitrários. Então temos que

$$\langle u - v_{k_j+1}, x - y_{k_j+1} \rangle \geq -\varepsilon_{k_j+1}, \quad \forall j \geq 1.$$

Como $y_{k_j+1} \rightarrow \bar{x}$, $v_{k_j+1} \rightarrow 0$ e $\varepsilon_{k_j+1} \rightarrow 0$, passando o limite com $j \rightarrow \infty$ na desigualdade acima, obtemos que

$$\langle u - 0, x - \bar{x} \rangle \geq 0.$$

Observe que (x, u) foi tomado arbitrário no gráfico de T . Como T é monótono maximal, segue que $0 \in T(\bar{x})$, ou seja, \bar{x} é uma solução de (1-2). \square

Teorema 31. *Se o problema (1-2) tem solução e existe $\bar{\lambda} > 0$ tal que $\inf_k \lambda_k \geq \bar{\lambda}$, então a sequência $\{x_k\}$ gerada pelo método HPE converge para uma solução de (1-2).*

Prova. Segue do Lema 29 (a) que a sequência $\{x_k\}$ é limitada e portanto tem pelo menos um ponto de acumulação \bar{x} . Segue do Lema 30 que \bar{x} é uma solução do problema (1-2). Além disso, temos pelo item a) do Lema 28 com $x^* = \bar{x}$ e do fato que $\sigma \in [0, 1)$ que a sequência $\{x_k\}$ é Fejér convergente ao conjunto solução do problema (1-2) e portanto a conclusão do teorema segue do Lema 12. \square

Teorema 32. *Se, além das hipóteses do Teorema 31, temos que T^{-1} é L -Lipschitz-contínua em zero (veja a definição em (1-21)), então a sequência $\{x_k\}$ converge linearmente para uma solução de (1-2).*

Prova. Pelos itens b) e d) do Lema 29, temos que

$$\lim_{k \rightarrow \infty} \|v_k\| = 0, \quad \lim_{k \rightarrow \infty} \|y_k - x_{k-1}\| = 0.$$

Para todo $k \geq 1$, seja $z_k \in \mathbb{R}^n$ a solução exata do subproblema proximal

$$0 \in \lambda_k T(\cdot) + (\cdot - x_{k-1}).$$

Então

$$u_k := -(1/\lambda_k)(z_k - x_{k-1}) \in T(z_k). \quad (3-21)$$

Aplicando o Lema 26, temos que

$$\|z_k - y_k\| \leq \sigma \|y_k - x_{k-1}\|,$$

e portanto pela desigualdade triangular, obtemos

$$\|z_k - x_{k-1}\| \leq \|z_k - y_k\| + \|y_k - x_{k-1}\| \leq (1 + \sigma) \|y_k - x_{k-1}\|. \quad (3-22)$$

Portanto,

$$\|u_k\| \leq (1/\lambda_k)(1 + \sigma) \|y_k - x_{k-1}\| \leq (1/\bar{\lambda})(1 + \sigma) \|y_k - x_{k-1}\|. \quad (3-23)$$

Como a sequência $\{\|y_k - x_{k-1}\|\}$ converge para zero e temos por hipóteses que $\inf_k \lambda_k \geq \bar{\lambda} > 0$, então $\{\|u_k\|\}$ também converge a zero. Seja x^* tal que $0 \in T(x^*)$. Note que (3-21) implica que $z_k \in T^{-1}(u_k)$ para todo $k \geq 1$. Além disso, como T^{-1} é L -Lipschitz contínua em zero, segue da Definição 19 com $S = T^{-1}$ e de (3-21) que existe $k_0 > 0$ tal que

$$\|z_k - x^*\| \leq L \|u_k\| = (L/\lambda_k) \|z_k - x_{k-1}\|, \quad \forall k \geq k_0. \quad (3-24)$$

Usando a desigualdade triangular, obtemos que

$$\begin{aligned}\|x^* - x_{k-1}\| &\leq \|x^* - z_k\| + \|z_k - x_{k-1}\| \\ &\leq (L/\lambda_k)\|z_k - x_{k-1}\| + \|z_k - x_{k-1}\| \\ &= (1 + L/\lambda_k)\|z_k - x_{k-1}\|.\end{aligned}$$

Usando (3-22), temos que

$$\|x^* - x_{k-1}\| \leq (1 + L/\lambda_k)(1 + \sigma)\|y_k - x_{k-1}\|.$$

Segue do Lema 28a), da desigualdade acima e do fato que $\inf_k \lambda_k \geq \bar{\lambda} > 0$ que

$$\begin{aligned}\|x^* - x_k\|^2 &\leq \|x^* - x_{k-1}\|^2 - (1 - \sigma^2)\|y^k - x_{k-1}\|^2 \\ &\leq \left(1 - \frac{1 - \sigma^2}{[(1 + L/\bar{\lambda})(1 + \sigma)]^2}\right) \|x^* - x_{k-1}\|^2 \\ &= \left(1 - \frac{1 - \sigma}{(1 + L/\bar{\lambda})^2(1 + \sigma)}\right) \|x^* - x_{k-1}\|^2, \quad \forall k \geq k_0,\end{aligned}$$

concluindo que $\{x_k\}$ converge para x^* linearmente.

3.3 Resultados de Complexidade para o Método HPE

Antes de estabelecer os resultados de complexidade de iteração para o método HPE, precisamos de alguns resultados técnicos.

Lema 33. Para todo $k \in \mathbb{N}$,

$$(1 - \sigma)\|y_k - x_{k-1}\| \leq \|\lambda_k v_k\| \leq (1 + \sigma)\|y_k - x_{k-1}\|. \quad (3-25)$$

Prova. Tendo em vista a desigualdade do Passo 1 do método HPE, e pela desigualdade triangular, obtemos

$$\left| \|\lambda_k v_k\| - \|y_k - x_{k-1}\| \right| \leq \|\lambda_k v_k + y_k - x_{k-1}\| \leq \sigma\|y_k - x_{k-1}\|, \quad \forall k \in \mathbb{N},$$

o que implica claramente (3-25). □

Lema 34. Para todo $\tilde{x} \in \mathbb{R}^n$ e $i \in \mathbb{N}$ temos que

$$\|\tilde{x} - x_{i-1}\|^2 = \|\tilde{x} - x_i\|^2 + 2\lambda_i \langle y_i - \tilde{x}, v_i \rangle + \|y_i - x_{i-1}\|^2 - \|x_i - y_i\|^2. \quad (3-26)$$

Prova. Seja $\tilde{x} \in \mathbb{R}^n$ e $i \in \mathbb{N}$. Então, pelo Lema 27 item a), com $x = x_{i-1}$, $x^+ = x_i$ e $y = y_i$ temos

$$\|\tilde{x} - x_{i-1}\|^2 = \|\tilde{x} - x_i\|^2 + 2\langle \tilde{x} - y_i, x_i - x_{i-1} \rangle + \|y_i - x_{i-1}\|^2 - \|y_i - x_i\|^2.$$

O resultado segue do fato que $x_i - x_{i-1} = -\lambda_i v_i$. □

Lema 35. *Seja d_0 a distância de x_0 ao conjunto $T^{-1}(0)$. Para quaisquer $\alpha \in \mathbb{R}$ e $k \in \mathbb{N}$, existe um índice $i \leq k$ tal que*

$$\|v_i\| \leq d_0 \sqrt{\frac{(1+\sigma)}{(1-\sigma)} \left(\frac{\lambda_i^{\alpha-2}}{\sum_{j=1}^k \lambda_j^\alpha} \right)}, \quad \varepsilon_i \leq \frac{d_0^2 \sigma^2}{2(1-\sigma^2)} \left(\frac{\lambda_i^{\alpha-1}}{\sum_{j=1}^k \lambda_j^\alpha} \right). \quad (3-27)$$

Prova. Defina, para qualquer $k \in \mathbb{N}$,

$$\tau_k := \max \left\{ \frac{2\varepsilon_k \lambda_k^{1-\alpha}}{\sigma^2}, \frac{\|v_k\|^2 \lambda_k^{2-\alpha}}{(1+\sigma)^2} \right\}. \quad (3-28)$$

Então, pela desigualdade do Passo 1 do HPE e pelo (3-25), obtemos

$$\tau_k \lambda_k^\alpha := \max \left\{ \frac{2\varepsilon_k \lambda_k}{\sigma^2}, \frac{\|v_k\|^2 \lambda_k^2}{(1+\sigma)^2} \right\} \leq \|y_k - x_{k-1}\|^2.$$

Sendo $x^* \in T^{-1}(0)$ tal que $d_0 = \|x_0 - x^*\|$, segue da última desigualdade e do Lema 28 que

$$\sum_{j=1}^k \tau_j \lambda_j^\alpha \leq \sum_{j=1}^k \|y_j - x_{j-1}\|^2 \leq \frac{\|x_0 - x^*\|^2}{1-\sigma^2} = \frac{d_0^2}{1-\sigma^2}.$$

Logo, temos que

$$\left(\min_{j=1, \dots, k} \tau_j \right) \sum_{j=1}^k \lambda_j^\alpha \leq \sum_{j=1}^k \tau_j \lambda_j^\alpha \leq \frac{d_0^2}{1-\sigma^2}. \quad (3-29)$$

Portanto as desigualdades em (3-27) seguem facilmente de (3-29) e (3-28). □

Teorema 36. *Seja d_0 a distância de x_0 para $T^{-1}(0)$. Segue que:*

a) *Se $\underline{\lambda} := \inf \lambda_k > 0$, então para qualquer $k \in \mathbb{N}$ existe um índice $i \leq k$ tal que*

$$\|v_i\| \leq d_0 \sqrt{\frac{(1+\sigma)}{(1-\sigma)} \left(\frac{\underline{\lambda}^{-1}}{\sum_{j=1}^k \lambda_j} \right)} \leq \frac{d_0}{\underline{\lambda} \sqrt{k}} \sqrt{\frac{1+\sigma}{1-\sigma}},$$

$$\varepsilon_i \leq \frac{\sigma^2 d_0^2}{2(1-\sigma^2)} \frac{1}{\sum_{i=1}^k \lambda_i} \leq \frac{\sigma^2 d_0^2}{2(1-\sigma^2) \underline{\lambda} k}.$$

b) Para qualquer $k \in \mathbb{N}$, existe um índice $i \leq k$ tal que

$$\|v_i\| \leq d_0 \sqrt{\frac{(1+\sigma)}{(1-\sigma)} \left(\frac{1}{\sum_{j=1}^k \lambda_j^2} \right)}, \quad \varepsilon_i \leq \frac{\sigma^2 d_0^2 \lambda_i}{2(1-\sigma^2) \sum_{j=1}^k \lambda_j^2}. \quad (3-30)$$

c) Se $\sum_{k=1}^{\infty} \lambda_k^2 = \infty$, então as sequências $\{y_k\}$ e $\{x_k\}$ convergem para algum ponto x^* que é solução de (1-2).

Prova. Os itens a) e b) seguem imediatamente do Lema 35 com α igual a 1 e 2, respectivamente. Para o item c), primeiro note que se $v_i = 0$ e $\varepsilon_i = 0$ para algum i , então $x_i = y_i \in T^{-1}(0)$ e $x_k = x_i$ para todo $k \geq i$. Então suponha que

$$a_i := \max\{\|v_i\|, \varepsilon_i\} > 0, \quad \forall i.$$

Como por hipótese temos $\sum_{i=1}^{\infty} \lambda_i^2 = +\infty$, é fácil ver que

$$\lim_{k \rightarrow \infty} \max_{j=1, \dots, k} \frac{\lambda_j}{\sum_{i=1}^k \lambda_i^2} = 0.$$

Portanto, usando (3-30) concluímos que

$$\lim_{k \rightarrow \infty} \min_{i=1, \dots, k} a_i = 0.$$

Portanto existe uma subsequência $\{a_{i_k}\}$ que converge a 0. Pelo itens a) e b) do Lema 29 temos que $\{x_k\}$ é limitada e

$$\lim_{k \rightarrow \infty} \|y_k - x_{k-1}\| = 0, \quad (3-31)$$

e, portanto, $\{y_k\}$ é também limitado, em particular $\{y_{i_k}\}$ é limitada. Logo, existe uma subsequência que sem perda de generalização assumimos ser $\{y_{i_k}\}$ que converge para algum x^* . Afirmamos que x^* é solução de (1-2). De fato, como $\lim_{k \rightarrow \infty} a_{i_k} = 0$, temos que $\lim_{k \rightarrow \infty} v_{i_k} = 0$ e $\lim_{k \rightarrow \infty} \varepsilon_{i_k} = 0$. Por outro lado, temos que

$$\begin{aligned} v_{i_k} &\in T^{\varepsilon_{i_k}}(y_{i_k}), \\ \langle v_{i_k} - u, y_{i_k} - y \rangle &\geq -\varepsilon_{i_k}, \quad \forall (y, u) \in GrT, \end{aligned}$$

passando limite quando k tende a infinito temos que $\langle 0 - u, x^* - y \rangle \geq 0, \forall (y, u) \in GrT$. Como T é monótono maximal temos que $(x^*, 0) \in GrT$, e portanto $0 \in T(x^*)$.

Agora note que (3-31) e o fato que $\{y_{i_k}\}$ converge para x^* implica que x^* é também um ponto de acumulação de $\{x_k\}$. Como x^* é solução de (1-2), temos que a sequência $\{\|x_k - x^*\|\}$ é não-crescente pelo item b) do Lema 28, e portanto concluímos que $\lim_{k \rightarrow \infty} x_k = x^*$. Pelo (3-31), temos que a sequência $\{y_k\}$ também converge a x^* . \square

3.4 Método do Korpelevich

Nesta seção apresentaremos o método *Extragradiente de Korpelevich* (EK) para resolver problemas de desigualdades variacionais.

O principal objetivo dessa seção é mostrar que o método EK pode ser interpretado como um caso particular do método HPE, para mais detalhes, veja [10]. Considere um operador monótono e L -Lipschitz contínuo $F : X \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ e um conjunto $X \subset \mathbb{R}^n$ convexo, fechado e não-vazio. O problema de desigualdade variacional $VIP(F, X)$ consiste em encontrar $x^* \in X$ tal que

$$\langle x - x^*, F(x^*) \rangle \geq 0, \quad \forall x \in X. \quad (3-32)$$

Neste trabalho, assumimos que o conjunto de soluções de $VIP(F, X)$ seja não vazio.

Enunciaremos a seguir o método extragradiente de Korpelevich. Denotamos por P_X o operador projeção no conjunto X .

Método EK:

Passo 0: Sejam $x_0 \in \mathbb{R}^n$, $0 \leq \sigma < 1$ e $\lambda = \sigma/L$ e $k = 1$.

Passo 1: Calcule

$$y_k = P_X(x_{k-1} - \lambda F(x_{k-1})), \quad x_k = P_X(x_{k-1} - \lambda F(y_k)). \quad (3-33)$$

Passo 2: Faça $k \leftarrow k + 1$ e vá para o passo 1.

Fim.

Observe que resolver $VIP(F, X)$ é equivalente a resolver o problema de inclusão monótona (1-2) com $T = F + N_X$; onde $N_X : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é o operador ponto-conjunto (cone normal) definido em (1-9).

Da definição acima, segue que (3-32) é equivalente a $-F(x^*) \in N_X(x^*)$ e, portanto, ao problema de inclusão monótona que corresponde a encontrar $x^* \in X$ tal que

$$0 \in (F + N_X)(x^*). \quad (3-34)$$

Mostraremos a seguir que o método EK para resolver o $VIP(F, X)$ pode ser visto como um caso particular do método HPE para resolver o problema de inclusão monótona (3-34).

Teorema 37. *Sejam $\{x_k\}$ e $\{y_k\}$ sequências geradas pelo método EK e, para $k \in \mathbb{N}$,*

defina

$$\mathbf{q}_k = \frac{1}{\lambda}(\mathbf{x}_{k-1} - \lambda F(\mathbf{y}_k) - \mathbf{x}_k), \quad \varepsilon_k = \langle \mathbf{q}_k, \mathbf{x}_k - \mathbf{y}_k \rangle, \quad \mathbf{v}_k = F(\mathbf{y}_k) + \mathbf{q}_k. \quad (3-35)$$

Então temos que as seguintes relações valem:

- a) $\mathbf{x}_k = \mathbf{x}_{k-1} - \lambda \mathbf{v}_k$;
- b) $\mathbf{q}_k \in \partial_{\varepsilon_k} \delta_X(\mathbf{y}_k)$ e $\mathbf{v}_k \in (F + N_X^{\varepsilon_k})(\mathbf{y}_k) \subset (F + N_X)^{\varepsilon_k}(\mathbf{y}_k)$;
- c) $\|\lambda \mathbf{v}_k + \mathbf{y}_k - \mathbf{x}_{k-1}\|^2 + 2\lambda \varepsilon_k \leq \sigma^2 \|\mathbf{y}_k - \mathbf{x}_{k-1}\|^2$.

Em particular, o método EK é um caso especial do método HPE para resolver o problema (3-34).

Prova. a) A prova deste item segue facilmente das definições de \mathbf{q}_k e \mathbf{v}_k em (3-35).

b) Note que utilizando (3-33), (3-35), e (1-10) segue que

$$\mathbf{q}_k \in N_X(\mathbf{x}_k) = \partial \delta_X(\mathbf{x}_k), \quad (3-36)$$

onde δ_X é a função indicadora (veja definição 1-13).

A primeira inclusão no item b) segue de (3-36) e item b) da Proposição 17. Usando a inclusão $(F + N_X^{\varepsilon_k})(\mathbf{y}_k) \subset (F + N_X)^{\varepsilon_k}(\mathbf{y}_k)$, em virtude do item b) da Proposição 16, a definição de \mathbf{v}_k e pelo fato que $(N_X)^{\varepsilon} = \partial_{\varepsilon} \delta_X$, que foi demonstrada na Proposição 18, temos que

$$\mathbf{v}_k = F(\mathbf{y}_k) + \mathbf{q}_k \in F(\mathbf{y}_k) + \partial_{\varepsilon_k} \delta_X(\mathbf{y}_k) = F(\mathbf{y}_k) + (N_X)^{\varepsilon_k}(\mathbf{y}_k) \subset (F + N_X)^{\varepsilon_k}(\mathbf{y}_k),$$

onde a última inclusão segue da monotonicidade de F e do item b) da Proposição 16.

c) Defina

$$\mathbf{p}_k = \frac{1}{\lambda}(\mathbf{x}_{k-1} - \lambda F(\mathbf{x}_{k-1}) - \mathbf{y}_k), \quad \forall k > 0. \quad (3-37)$$

Logo, segue de (1-10) e da definição de \mathbf{y}_k em (3-33), que $\mathbf{p}_k \in N_X(\mathbf{y}_k)$. Este fato juntamente com (3-35) implicam que

$$\varepsilon_k = \langle \mathbf{q}_k - \mathbf{p}_k, \mathbf{x}_k - \mathbf{y}_k \rangle + \langle \mathbf{p}_k, \mathbf{x}_k - \mathbf{y}_k \rangle \leq \langle \mathbf{q}_k - \mathbf{p}_k, \mathbf{x}_k - \mathbf{y}_k \rangle. \quad (3-38)$$

Combinando a relação acima com o fato que $x_k = x_{k-1} - \lambda v_k$ (veja o item a)), temos

$$\begin{aligned}
 \|\lambda v_k + y_k - x_{k-1}\|^2 + 2\lambda \varepsilon_k &= \|y_k - x_k\|^2 + 2\lambda \varepsilon_k \\
 &\leq \|y_k - x_k\|^2 + 2\lambda \langle q_k - p_k, x_k - y_k \rangle \\
 &= \|\lambda(q_k - p_k) + x_k - y_k\|^2 - \lambda^2 \|q_k - p_k\|^2 \\
 &\leq \|\lambda(q_k - p_k) + x_k - y_k\|^2 = \|\lambda(F(x_{k-1}) - F(y_k))\|^2, \quad (3-39)
 \end{aligned}$$

onde a última igualdade segue da definição de q_k e p_k em (3-35) e (3-37). Portanto, a prova do item c) segue da desigualdade (3-39), do fato que F é L -Lipschitz contínuo em X e da hipótese que $\lambda = \sigma/L$. \square

Método A-HPE para Otimização Convexa

4.1 Método A-HPE e sua Complexidade por Iteração

Nesta seção analisaremos a complexidade por iteração de uma variante acelerada do método HPE para resolver problemas de otimização. Estamos interessados no seguinte problema

$$f_* := \inf\{f(x), x \in \mathbb{R}^n\}, \quad (4-1)$$

onde $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é uma função convexa, semicontínua e própria. Assumimos ao longo dessa seção que o conjunto de soluções ótimas X_* de (4-1) é não-vazio, e portanto $f_* \in \mathbb{R}$.

Método A-HPE:

Passo 0: Sejam $x_0, y_0 \in \mathbb{R}^n$, $0 \leq \sigma \leq 1$ e defina $A_0 = 0$ e $k = 0$.

Passo 1: Calcule $\lambda_{k+1} > 0$ e $(\tilde{y}_{k+1}, v_{k+1}, \varepsilon_{k+1}) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ tal que

$$v_{k+1} \in \partial_{\varepsilon_{k+1}} f(\tilde{y}_{k+1}); \quad (4-2)$$

$$\|\lambda_{k+1} v_{k+1} + \tilde{y}_{k+1} - \tilde{x}_k\|^2 + 2\lambda_{k+1} v_{k+1} \varepsilon_{k+1} \leq \sigma^2 \|\tilde{y}_{k+1} - \tilde{x}_k\|^2, \quad (4-3)$$

onde

$$\tilde{x}_k = \frac{A_k}{A_k + a_{k+1}} y_k + \frac{a_{k+1}}{A_k + a_{k+1}} x_k, \quad (4-4)$$

$$a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1} A_k}}{2}. \quad (4-5)$$

Passo 2: Encontre y_{k+1} tal que $f(y_{k+1}) \leq f(\tilde{y}_{k+1})$ e seja

$$A_{k+1} = A_k + a_{k+1}, \quad (4-6)$$

$$x_{k+1} = x_k - a_{k+1} v_{k+1}. \quad (4-7)$$

Passo 3: Defina $k \leftarrow k + 1$ e vá para o passo 1. **Fim.**

Em seguida apresentamos algumas observações sobre o Método A-HPE.

Note que se ao invés da atualização (4-6) consideramos $A_k = 0$, então o método A-HPE se torna um caso particular do método HPE de [8], estudado no Capítulo 3. De fato, veja que neste caso, segue de (4-4) e (4-5) que $\tilde{x}_k = x_k$, $a_k = \lambda_k$, e portanto (4-7) se torna $x_{k+1} = x_k - \lambda_{k+1} v_{k+1}$, para todo $k \in \mathbb{N}$. Logo o resultado segue deste fato combinado com as relações (4-2) e (4-3).

A método A-HPE não especifica como calcular λ_{k+1} e $(\tilde{y}_{k+1}, v_{k+1}, \varepsilon_{k+1})$ como no passo 1. Para computar essas quantidades apropriadamente, necessitamos de uma implementação específica do algoritmo utilizando propriedades que a função objetivo possa ter. Nas próximas seções, descreveremos alguns procedimentos para encontrar essas quantidades em uma variante do algoritmo de Nesterov que pode ser visto como um caso particular do método A-HPE.

Considerando λ_{k+1} e \tilde{x}_k como no Passo 1, a iteração exata do ponto proximal \tilde{y} e o vetor v definidos como

$$\begin{aligned} \tilde{y} &:= \arg \min_{x \in \mathbb{R}^n} \left\{ \lambda_{k+1} f(x) + \frac{1}{2} \|x - \tilde{x}_k\|^2 \right\}, \\ v &:= \frac{1}{\lambda_{k+1}} (\tilde{x}_k - \tilde{y}), \end{aligned}$$

são caracterizados por

$$v \in \partial f(\tilde{y}), \quad \lambda_{k+1} v + \tilde{y} - \tilde{x}_k = 0, \quad (4-8)$$

e, portanto, $\varepsilon_{k+1} = 0$, $\tilde{y}_{k+1} = \tilde{y}$ e $v_{k+1} = v$ satisfazem o erro de tolerância em (4-2)-(4-3) com $\sigma = 0$. Logo, o critério de erro (4-2)-(4-3) é um relaxamento da caracterização (4-8) da iteração do método do ponto proximal em que a inclusão em (4-8) é relaxada para $v_{k+1} \in \partial_{\varepsilon_{k+1}} f(\tilde{y}_{k+1})$, e na última relação em (4-8) é considerado um resíduo $r_{k+1} = \lambda_{k+1} v_{k+1} + \tilde{y}_{k+1} - \tilde{x}_k$, tal que o par residual $(r_{k+1}, \varepsilon_{k+1})$ é relativamente pequeno em relação a quantidade $\|\tilde{y}_{k+1} - \tilde{x}_k\|$, no sentido que a seguinte desigualdade equivalente a (4-3) é satisfeita: $\|r_{k+1}\|^2 + 2\lambda_{k+1}\varepsilon_{k+1} \leq \sigma^2 \|\tilde{y}_{k+1} - \tilde{x}_k\|^2$.

O erro de tolerância (4-2)-(4-3) é uma versão do critério de erro considerado pelo método HPE o problema de otimização, em que utiliza-se o operador ε -subdiferencial da função f , ao invés do ε -alargamento do operador ∂f , isto é, $(\partial f)^\varepsilon$, veja (1-17). Note que a seguinte inclusão vale $\partial_\varepsilon f(x) \subset (\partial f)^\varepsilon(x)$ para todo $x \in \mathbb{R}^n$ e que, em geral, essa inclusão é estrita, veja Proposição 17a) e os comentários logo após essa proposição.

Quinta observação: Duas maneiras simples de escolher o ponto y_{k+1} como no passo 2 do método A-HPE são definir $y_{k+1} = \tilde{y}_{k+1}$ ou definir $y_{k+1} := \arg \min \{f(y) : y \in \{y_k, \tilde{y}_{k+1}\}\}$. Note que a segunda regra garante que a sequência $\{f(y_k)\}$ seja não-crescente.

Em seguida apresentamos algumas relações que serão úteis na análise do método A-HPE.

Lema 38. Para todo inteiro $k \geq 0$ obtemos $\lambda_{k+1}A_{k+1} = a_{k+1}^2 > 0$.

Prova. Óbvio que $a_{k+1} > 0$ satisfaz (4-5) se, e somente se, $0 < a := a_{k+1}$ satisfaz

$$a^2 - \lambda_{k+1}a - \lambda_{k+1}A_k = 0,$$

ou de modo equivalente, $a_{k+1}^2 = \lambda_{k+1}(A_k + a_{k+1}) = \lambda_{k+1}A_{k+1}$, onde essa última desigualdade resulta de (4-6). \square

Para analisar as propriedades das sequências $\{x_k\}$ e $\{y_k\}$, defina as aplicações afins $\gamma_k : \mathbb{R}^n \rightarrow \mathbb{R}$ como

$$\gamma_k(x) = f(\tilde{y}_k) + \langle x - \tilde{y}_k, v_k \rangle - \varepsilon_k, \quad \forall x \in \mathbb{R}^n, k \geq 1, \quad (4-9)$$

e as aplicações afins agregadas $\Gamma_k : \mathbb{R}^n \rightarrow \mathbb{R}$ recursivamente como

$$\Gamma_0(x) \equiv 0, \quad \Gamma_{k+1}(x) = \frac{A_k}{A_{k+1}}\Gamma_k(x) + \frac{a_{k+1}}{A_{k+1}}\gamma_{k+1}(x), \quad \forall k \geq 0, \quad \forall x \in \mathbb{R}^n. \quad (4-10)$$

Lema 39. Para todo inteiro $k \geq 0$ e $x \in \mathbb{R}^n$, temos que

- a) $\gamma_{k+1}(x)$ é afim e $\gamma_{k+1}(x) \leq f(x)$;
- b) $\Gamma_k(x)$ é afim e $A_k\Gamma_k(x) \leq A_k f(x)$;
- c) $x_k = \arg \min_{x \in \mathbb{R}^n} \{A_k\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2\}$.

Prova. No item a) segue de (4-9) para todo $x \in \mathbb{R}^n$, e por (1-12) temos que

$$f(x) \geq f(\tilde{y}_{k+1}) + \langle x - \tilde{y}_{k+1}, v_{k+1} \rangle - \varepsilon_{k+1} = \gamma_{k+1}(x), \quad \forall x \in \mathbb{R}^n, k \geq 1,$$

o que implica que $v_{k+1} \in \partial f(\tilde{y}_{k+1})$, logo $\gamma_{k+1}(x) \leq f(x)$. No item b) segue de a) e (4-10) usando um argumento de indução simples à partir de $k = 0$. No item c), primeiro note que a condição de otimalidade de primeira ordem para minimizar $A_k\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2$ nos fornece $A_k\nabla\Gamma_k(z_k) + z_k - x_0 = 0$, o que implica que $z_k = x_0 - A_k\nabla\Gamma_k(z_k)$, onde z_k é o único minimizador da função acima. Note que usando argumento de indução sobre k em (4-10) e do fato que $\nabla\gamma_j = v_j$, obtemos

$$A_k\nabla\Gamma_k(x) = \sum_{j=1}^k a_j v_j, \quad \forall x \in \mathbb{R}^n.$$

Além disso, (4-7) implica que $x_k = x_0 - \sum_{j=1}^k a_j v_j$. Logo concluímos que $x_k = z_k$, o que prova o item c). \square

A seguir apresentamos um lema que será usado para analisar o A-HPE.

Lema 40. Sejam $\bar{x}, \bar{y}, \bar{v} \in \mathbb{R}^n$ e escalares $\lambda > 0, \varepsilon \geq 0$ e $\sigma \geq 0$. Então a desigualdade

$$\|\lambda \bar{v} + \bar{y} - \bar{x}\|^2 + 2\lambda \varepsilon \leq \sigma^2 \|\bar{y} - \bar{x}\|^2$$

é equivalente a desigualdade

$$\min_{x \in \mathbb{R}^n} \left\{ \langle \bar{v}, x - \bar{y} \rangle - \varepsilon + \frac{1}{2\lambda} \|x - \bar{x}\|^2 \right\} \geq \frac{1 - \sigma^2}{2\lambda} \|\bar{y} - \bar{x}\|^2.$$

Lema 41. Para todo inteiro $k \geq 0$, definindo

$$\beta_k = \inf_{x \in \mathbb{R}^n} \left\{ A_k \Gamma_k(x) + \frac{1}{2} \|x - x_0\|^2 \right\} - A_k f(y_k), \quad (4-11)$$

temos que $\beta_0 = 0$ e

$$\beta_{k+1} \geq \beta_k + \frac{(1 - \sigma^2) A_{k+1}}{2\lambda_{k+1}} \|\tilde{y}_{k+1} - \tilde{x}_k\|^2, \quad \forall k \geq 0. \quad (4-12)$$

Prova. Pelo item c) do Lema 39, tomando $k = 0$ e pelo fato de que $A_0 = 0$ (em virtude do Passo 0 do método A-HPE), temos que $\beta_0 = 0$. Para demonstrar (4-12), seja um arbitrário $x \in \mathbb{R}^n$. Defina

$$\tilde{x} = \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x \quad (4-13)$$

e observe que (4-4) e (4-6), e pelo fato que γ_{k+1} é afim, obtemos

$$\tilde{x} - \tilde{x}_k = \frac{a_{k+1}}{A_{k+1}} (x - x_k); \quad (4-14)$$

$$\gamma_{k+1}(\tilde{x}) = \frac{A_k}{A_{k+1}} \gamma_{k+1}(y_k) + \frac{a_{k+1}}{A_{k+1}} \gamma_{k+1}(x). \quad (4-15)$$

Usando a definição de Γ_{k+1} em (4-10) e os itens b) e c) do Lema 39, concluímos que, para todo $x \in \mathbb{R}^n$,

$$\begin{aligned} A_{k+1} \Gamma_{k+1}(x) + \frac{1}{2} \|x - x_0\|^2 &= a_{k+1} \gamma_{k+1}(x) + A_k \Gamma_k(x) + \frac{1}{2} \|x - x_0\|^2 \\ &= a_{k+1} \gamma_{k+1}(x) + A_k \Gamma_k(x_k) + \frac{1}{2} \|x_k - x_0\|^2 + \frac{1}{2} \|x - x_k\|^2 \\ &= a_{k+1} \gamma_{k+1}(x) + A_k f(y_k) + \beta_k + \frac{1}{2} \|x - x_k\|^2, \end{aligned}$$

onde a última igualdade segue da definição de β_k em (4-11). Combinando a relação acima

com o Lema 39, a definição de \tilde{x} em (4-13), e relações (4-14) e (4-15), concluímos que

$$\begin{aligned} A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 &= a_{k+1}\gamma_{k+1}(x) + A_{k+1}\gamma_{k+1}(y_k) + \beta_k + \frac{1}{2}\|x - x_k\|^2 \\ &= \beta_k + A_{k+1}\gamma_{k+1}(\tilde{x}) + \frac{A_{k+1}^2}{2a_{k+1}^2}\|\tilde{x} - \tilde{x}_k\|^2 \\ &= \beta_k + A_{k+1}\gamma_{k+1}(\tilde{x}) + \frac{A_{k+1}}{2\lambda_{k+1}}\|\tilde{x} - \tilde{x}_k\|^2, \end{aligned}$$

onde a última igualdade segue do Lema 38. Usando a definição (4-9) com $k = k + 1$, (4-3), e Lema 40 com $\bar{v} = v_{k+1}$, $\bar{y} = \tilde{y}_{k+1}$, $\bar{x} = \tilde{x}_k$, e $\varepsilon = \varepsilon_{k+1}$, concluímos que

$$\begin{aligned} &\gamma_{k+1}(\tilde{x}) + \frac{1}{2\lambda_{k+1}}\|\tilde{x} - \tilde{x}_k\|^2 \\ &= f(\tilde{y}_{k+1}) + \left(\langle \tilde{x} - \tilde{y}_{k+1}, v_{k+1} \rangle - \varepsilon_{k+1} + \frac{1}{2\lambda_{k+1}}\|\tilde{x} - \tilde{x}_k\|^2 \right) \\ &\geq f(\tilde{y}_{k+1}) + \frac{1 - \sigma^2}{2\lambda_{k+1}}\|\tilde{y}_{k+1} - \tilde{x}_k\|^2. \end{aligned}$$

Usando as duas relações acima e o fato de que A_{k+1} é não-negativo, concluímos que

$$\begin{aligned} &\inf_{x \in \mathbb{R}^n} \left(A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 \right) \\ &\geq \beta_k + A_{k+1}f(\tilde{y}_{k+1}) + \frac{(1 - \sigma^2)A_{k+1}}{2\lambda_{k+1}}\|\tilde{y}_{k+1} - \tilde{x}_k\|^2, \end{aligned}$$

combinado com a definição (4-11) com $k = k + 1$ e a desigualdade no Passo 2 do método A-HPE, assim concluímos (4-12). \square

A próxima proposição segue diretamente do Lema 41.

Proposição 42. Para todo inteiro $k \geq 1$ e $x \in \mathbb{R}^n$,

$$A_k f(y_k) + \frac{1 - \sigma^2}{2} \sum_{j=1}^k \frac{A_j}{\lambda_j} \|\tilde{y}_j - \tilde{x}_{j-1}\|^2 + \frac{1}{2}\|x - x_k\|^2 \leq A_k \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2.$$

Prova. Aplicando somatória em (4-12) de $k = 0$ para $k = k - 1$ temos

$$\sum_{j=0}^{k-1} (\beta_{j+1} - \beta_j) \geq \frac{(1 - \sigma^2)}{2} \sum_{j=0}^{k-1} \frac{A_{j+1}}{\lambda_{j+1}} \|\tilde{y}_{j+1} - \tilde{x}_j\|^2,$$

e usando o fato que $\beta_0 = 0$ e que o primeiro membro da desigualdade acima é uma soma telescópica, obtemos

$$\frac{(1 - \sigma^2)}{2} \sum_{j=1}^k \frac{A_j}{\lambda_j} \|\tilde{y}_j - \tilde{x}_{j-1}\|^2 \leq \beta_k,$$

juntamente com (4-11) implica que

$$A_k f(y_k) + \frac{(1 - \sigma^2)}{2} \sum_{j=1}^k \frac{A_j}{\lambda_j} \|\tilde{y}_j - \tilde{x}_{j-1}\|^2 \leq \inf_{\tilde{x} \in \mathbb{R}^n} \left\{ A_k \Gamma_k(\tilde{x}) + \frac{1}{2} \|\tilde{x} - x_0\|^2 \right\}. \quad (4-16)$$

Usando os itens b) e c) do Lema 39, concluímos que para todo $x \in \mathbb{R}^n$,

$$A_k \Gamma_k(x) + \frac{1}{2} \|x - x_0\|^2 = \inf_{\tilde{x} \in \mathbb{R}^n} \left\{ A_k \Gamma_k(\tilde{x}) + \frac{1}{2} \|\tilde{x} - x_0\|^2 \right\} + \frac{1}{2} \|x - x_k\|^2.$$

Subtraindo $\|x - x_k\|^2/2$ em ambos os lados da igualdade acima e substituindo em (4-16) teremos o resultado desejado. \square

O seguinte resultado garante a limitação da sequência $\{x_k\}$ e nos mostra uma importante relação entre a aproximação da sequência $\{f(y_k) - f_*\}$ a zero e a sequência de escalares $\{A_k\}$ definidas no passo 2 do método A-HPE.

Teorema 43. *Seja x_* uma projeção de x_0 em X_* e d_0 a distância de x_0 para X_* . Então, para todo inteiro $k \geq 1$,*

$$\frac{1}{2} \|x_* - x_k\|^2 + A_k [f(y_k) - f_*] + \frac{1 - \sigma^2}{2} \sum_{j=1}^k \frac{A_j}{\lambda_j} \|\tilde{y}_j - \tilde{x}_{j-1}\|^2 \leq \frac{1}{2} d_0^2. \quad (4-17)$$

Consequentemente, para todo inteiro $k \geq 1$,

$$f(y_k) - f_* \leq \frac{d_0^2}{2A_k}, \quad \|x_k - x_*\| \leq d_0, \quad (4-18)$$

e se $\sigma < 1$, então

$$\sum_{j=1}^k \frac{A_j}{\lambda_j} \|\tilde{y}_j - \tilde{x}_{j-1}\|^2 \leq \frac{d_0^2}{1 - \sigma^2}. \quad (4-19)$$

Prova. O resultado segue imediatamente do item b) do Lema 39 e da Proposição 42 com $x = x_*$. De fato, basta observar que pelo item b) do Lema 39 temos que $-A_k f_* \leq -A_k \Gamma_k$, para todo $k \geq 1$ onde f_* foi definido em (4-1), e com isso substituindo essa desigualdade na Proposição 42, com $x = x_*$, e pelo fato de $\|x_* - x_0\| \leq d_0, \forall k \geq 1$, obtemos a desigualdade (4-17). \square

O resultado a seguir mostra como a sequência $\{A_k\}$ cresce em termos da sequência de tamanhos de passos $\{\lambda_k\}$.

Lema 44. *Para todo inteiro $k \geq 0$,*

$$\sqrt{A_{k+1}} \geq \sqrt{A_k} + \frac{1}{2} \sqrt{\lambda_{k+1}}. \quad (4-20)$$

Em consequência, as seguintes afirmações valem:

a) para todo inteiro $k \geq 1$,

$$A_k \geq \frac{1}{4} \left(\sum_{j=1}^k \sqrt{\lambda_j} \right)^2; \quad (4-21)$$

b) se $\sigma < 1$, então $\sum_{j=1}^{\infty} \|\tilde{y}_j - \tilde{x}_{j-1}\|^2 \leq 4d_0^2/(1 - \sigma^2)$.

Prova. Note que pela definição de a_{k+1} em (4-5) implica que

$$a_{k+1} \geq \frac{\lambda_{k+1}}{2} + \sqrt{\lambda_{k+1} A_k}$$

e usando (4-6), concluímos que

$$A_{k+1} \geq A_k + \left(\frac{\lambda_{k+1}}{2} + \sqrt{\lambda_{k+1} A_k} \right) \geq \left(\sqrt{A_k} + \frac{1}{2} \sqrt{\lambda_{k+1}} \right)^2, \quad \forall k \geq 0.$$

Portanto, (4-20) vale para todo $k \geq 0$. Considere um $l \in \mathbb{N}$ arbitrário e note que somando a desigualdade (4-20) de $k = 0$ a $k = l - 1$ e usando o fato de que $A_0 = 0$, obtemos

$$\sqrt{A_l} = \sum_{k=0}^{l-1} (\sqrt{A_{k+1}} - \sqrt{A_k}) \geq \sum_{k=0}^{l-1} \frac{1}{2} \sqrt{\lambda_{k+1}},$$

e portanto a conclusão do item a) segue. Para o item b), observe que do item a) temos que, $\forall k \geq 1$

$$A_k \geq \frac{\lambda_k}{4},$$

substituindo essa desigualdade em (4-19) e aplicando o limite quando k tende a infinito, obtemos o resultado desejado. \square

O próximo teorema segue como consequência imediata do Teorema 43 e Lema 44.

Teorema 45. Para todo inteiro $k \geq 1$ temos

$$f(y_k) - f_* \leq 2d_0^2 \left(\sum_{j=1}^k \sqrt{\lambda_j} \right)^{-2}.$$

A proposição a seguir mostra a complexidade para cada iteração i , com $1 \leq i \leq k$, $k \geq 1$.

Proposição 46. Suponha que $0 < \sigma < 1$. Para todo inteiro $k \geq 1$, obtemos $v_k \in \partial_{\varepsilon_k} f(\tilde{y}_k)$. Então existe $1 \leq i \leq k$ tal que

$$\sqrt{\lambda_i} \|v_i\| \leq \frac{d_0}{\sqrt{\sum_{j=1}^k A_j}}, \quad \varepsilon_i \leq \frac{\sigma^2}{2(1 - \sigma^2)} \frac{d_0^2}{\sqrt{\sum_{j=1}^k A_j}}.$$

Prova. Para todo inteiro $k \geq 1$, definimos

$$\tau_k = \max \left\{ \frac{2\varepsilon_k}{\sigma^2}, \frac{\lambda_k \|v_k\|^2}{(1+\sigma)^2} \right\}. \quad (4-22)$$

Segue de (4-3) com $k := k - 1$, do fato que $\varepsilon_k \geq 0$ e da desigualdade triangular que

$$\begin{aligned} 2\lambda_k \varepsilon_k &\leq \sigma^2 \|\tilde{y}_k - \tilde{x}_{k-1}\|^2 \\ \|\lambda_k v_k\| &\leq \|\tilde{y}_k - \tilde{x}_{k-1}\| + \|\lambda_k v_k + \tilde{y}_k - \tilde{x}_{k-1}\| \leq (1+\sigma) \|\tilde{y}_k - \tilde{x}_{k-1}\|. \end{aligned}$$

Logo, segue das desigualdades acima e da definição da τ_k em (4-22) que

$$\lambda_k \tau_k \leq \|\tilde{y}_k - \tilde{x}_{k-1}\|^2, \quad \forall k \geq 1.$$

Logo, temos de (4-19) que

$$\frac{d_0^2}{1-\sigma^2} \geq \sum_{j=1}^k A_j \tau_j,$$

o qual implica que

$$\frac{d_0^2}{1-\sigma^2} \geq \left(\min_{j=1, \dots, k} \tau_j \right) \left(\sum_{j=1}^k A_j \right).$$

Seja $i \in \{1, \dots, k\}$ tal que $\tau_i = \min_{j=1, \dots, k} \tau_j$. Logo, temos da desigualdade acima que

$$\tau_i \leq \frac{1}{(1-\sigma^2)} \frac{d_0^2}{\sum_{j=1}^k A_j}. \quad (4-23)$$

Por outro lado, note que (4-22) implica que

$$\frac{2\varepsilon_i}{\sigma^2} \leq \tau_i \quad \text{e} \quad \frac{\lambda_i \|v_i\|^2}{(1+\sigma)^2} \leq \tau_i. \quad (4-24)$$

Portanto a prova da proposição segue de (4-23) e (4-24). \square

Em seguida apresentamos um lema básico que será usado para provar o próximo teorema, o qual será útil para estabelecer a complexidade do algoritmo apresentado na próxima seção.

Lema 47. Para todo $C \geq 0$, $\{\alpha_j, s_j\} \in \mathbb{R}$ com $1 \leq j \leq k$ e $\beta_1, \dots, \beta_k > 0$, temos que

$$\max \left\{ \sum_{j=1}^k \alpha_j s_j : \sum_{j=1}^k \beta_j s_j^2 \leq C \right\} = \sqrt{C \sum_{j=1}^k \frac{\alpha_j^2}{\beta_j}}.$$

Prova. Veja por exemplo [9, Lema A.2]. \square

Teorema 48. *Sejam x^* a projeção de x_0 em X_* e denote por d_0 a distância de x_0 a X_* . Considere o método A-HPE com $\sigma < 1$ e y_{k+1} escolhido como $y_{k+1} = \tilde{y}_{k+1}$, para cada $k \geq 0$. Então, para cada $k \geq 1$, temos que*

$$\|y_k - x^*\| \leq \left(\frac{2}{\sqrt{1 - \sigma^2}} + 1 \right) d_0.$$

Prova. Afirmamos primeiramente que para todo $k \geq 1$,

$$\|y_k - x^*\| \leq \frac{1}{A_k} \left(\sum_{j=1}^k A_j \|y_j - \tilde{x}_{j-1}\| \right) + d_0, \quad (4-25)$$

e mostraremos por indução sobre k . Observe que pela desigualdade triangular, pelas relações (4-4)-(4-6), pela convexidade da norma, e pelo Teorema 43 temos que

$$\begin{aligned} \|y_{k+1} - x^*\| &\leq \|y_{k+1} - \tilde{x}_k\| + \|\tilde{x}_k - x^*\| \\ &\leq \|y_{k+1} - \tilde{x}_k\| + \frac{A_k}{A_{k+1}} \|y_k - x^*\| + \frac{a_{k+1}}{A_{k+1}} \|x_k - x^*\| \\ &\leq \|y_{k+1} - \tilde{x}_k\| + \frac{A_k}{A_{k+1}} \|y_k - x^*\| + \frac{a_{k+1}}{A_{k+1}} d_0, \quad \forall k \geq 0. \end{aligned} \quad (4-26)$$

A desigualdade (4-26) com $k = 0$ e o fato de que $A_0 = 0$ implicam claramente (4-25) com $k = 1$. Suponha agora que (4-25) vale para algum k e vamos mostrar que também vale para $k + 1$. De fato, pela desigualdade (4-26), pela relação (4-6), e pela hipótese de indução temos que

$$\begin{aligned} \|y_{k+1} - x^*\| &\leq \|y_{k+1} - \tilde{x}_k\| + \frac{A_k}{A_{k+1}} \left\{ \frac{1}{A_k} \left(\sum_{j=1}^k A_j \|y_j - \tilde{x}_{j-1}\| \right) + d_0 \right\} + \frac{a_{k+1}}{A_{k+1}} d_0 \\ &= \frac{1}{A_{k+1}} \left(\sum_{j=1}^{k+1} A_j \|y_j - \tilde{x}_{j-1}\| \right) + d_0, \end{aligned}$$

e, portanto, (4-25) vale para $k + 1$. E concluímos que por indução vale (4-25) para todo $k \geq 1$. Fazendo $s_k = \|y_k - \tilde{x}_{k-1}\|$, segue do Teorema 43 e da suposição de que $y_k = \tilde{y}_k$ para todo $k \geq 1$ que

$$\sum_{j=1}^k \frac{A_j}{\lambda_j} s_j^2 \leq \frac{d_0^2}{1 - \sigma^2}.$$

Portanto, pelo Lema 47 com $C = d_0^2/(1 - \sigma^2)$, $\alpha_j = A_j$, e $\beta_j = A_j/\lambda_j$, para cada $j = 1, \dots, k$,

concluimos que

$$\sum_{j=1}^k A_j \|y_j - \bar{x}_{j-1}\| \leq \frac{d_0}{\sqrt{1-\sigma^2}} \sqrt{\sum_{j=1}^k A_j \lambda_j} \leq \frac{d_0}{\sqrt{1-\sigma^2}} \sqrt{A_k} \sqrt{\sum_{j=1}^k \lambda_j} \quad (4-27)$$

$$\leq \frac{d_0}{\sqrt{1-\sigma^2}} \sqrt{A_k} \sum_{j=1}^k \sqrt{\lambda_j}, \quad (4-28)$$

onde a segunda desigualdade de (4-27) decorre do fato que a sequência $\{A_k\}$ é crescente e a desigualdade (4-28) segue do fato que a norma Euclidiana é majorizada pela norma da soma. A desigualdade (4-28) juntamente com (4-25) implica que

$$\|y_k - x^*\| \leq \frac{d_0}{\sqrt{1-\sigma^2}} \frac{1}{A_k} \sum_{j=1}^k \sqrt{\lambda_j} + d_0. \quad (4-29)$$

A conclusão da proposição segue agora da desigualdade (4-29) e do Lema 44. \square

Na seção a seguir analisamos o método A-HPE para resolver uma classe especial de problemas de otimização convexa.

4.2 Método A-HPE para uma Classe Especial de Problemas Convexos

O problema de interesse nessa seção é

$$f_* := \min\{f(x) := g(x) + h(x) : x \in \mathbb{R}^n\} \quad (4-30)$$

onde:

1. a função $h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é convexa, semicontinua inferiormente e própria;
2. a função $g : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa e diferenciável com gradiente L_0 -Lipschitz contínuo;
3. o conjunto solução X_* de (4-30) é não-vazio.

Note que o problema (4-30) é equivalente ao seguinte problema de inclusão monótona

$$0 \in (\nabla g + \partial h)(x). \quad (4-31)$$

Em seguida apresentaremos um algoritmo do tipo proximal gradiente (PG) para resolver o problema acima e mostraremos que este algoritmo pode ser visto como uma variante do método A-HPE.

Algoritmo PG

Passo 0: Sejam $x_0, y_0 \in \mathbb{R}^n$, $0 < \sigma \leq 1$ e defina $A_0 = 0$, $\lambda = \sigma^2/L_0$ e $k = 0$.

Passo 1: Defina

$$\tilde{x}_k = \frac{A_k}{A_k + a_{k+1}} y_k + \frac{a_{k+1}}{A_k + a_{k+1}} x_k, \quad (4-32)$$

$$a_{k+1} = \frac{\lambda + \sqrt{\lambda^2 + 4\lambda A_k}}{2}. \quad (4-33)$$

e calcule

$$y_{k+1} = (I + \lambda \partial h)^{-1}(\tilde{x}_k - \lambda \nabla g(\tilde{x}_k)). \quad (4-34)$$

Passo 2: Defina

$$A_{k+1} = A_k + a_{k+1}, \quad (4-35)$$

$$x_{k+1} = x_k - \frac{a_{k+1}}{\lambda} (\tilde{x}_k - y_{k+1}). \quad (4-36)$$

Passo 3: Defina $k \leftarrow k + 1$ e vá para o passo 1.

Fim.

Observe que o cálculo de (4-34) é equivalente a

$$y_{k+1} = \arg \min_{x \in \mathbb{R}^n} \left(\lambda [\langle \nabla g(\tilde{x}_k), x \rangle + h(x)] + \frac{1}{2} \|x - \tilde{x}_k\|^2 \right). \quad (4-37)$$

O resultado a seguir mostra que o algoritmo acima é um caso especial do método A-HPE com $\lambda_k = \lambda$, para todo $k \geq 1$.

Proposição 49. Para todo $k \geq 0$, defina

$$\lambda_{k+1} = \lambda, \quad \tilde{y}_{k+1} = y_{k+1}, \quad v_{k+1} = \frac{1}{\lambda} (\tilde{x}_k - y_{k+1}), \quad (4-38)$$

$$\varepsilon_{k+1} = g(y_{k+1}) - [g(\tilde{x}_k) + \langle y_{k+1} - \tilde{x}_k, \nabla g(\tilde{x}_k) \rangle]. \quad (4-39)$$

Então

$$v_{k+1} \in \partial_{\varepsilon_{k+1}}(g + h)(\tilde{y}_{k+1}), \quad \|\lambda_{k+1} v_{k+1} + \tilde{y}_{k+1} - \tilde{x}_k\|^2 + 2\lambda_{k+1} \varepsilon_{k+1} \leq \sigma^2 \|\tilde{y}_{k+1} - \tilde{x}_k\|^2.$$

Prova. Pelo (4-34) e (4-38), e pelo item b da Proposição 17 juntamente com (4-39) temos que

$$v_{k+1} - \nabla g(\tilde{x}_k) \in \partial h(y_{k+1}), \quad \nabla g(\tilde{x}_k) \in \partial_{\varepsilon_{k+1}} g(y_{k+1}), \quad (4-40)$$

respectivamente. Como $(\partial_\varepsilon g + \partial h)(y) \subseteq \partial_\varepsilon(g + h)(y)$, concluímos que

$$v_{k+1} \in \partial_{\varepsilon_{k+1}}(g + h)(y_{k+1}).$$

Lembrando que $y_{k+1} \in \text{dom } \partial h \subseteq \mathbb{R}^n$, segue de (4-39), do fato de que ∇g é L_0 -Lipschitz contínua em \mathbb{R}^n (em virtude do Lema 2), temos que

$$\varepsilon_{k+1} \leq \frac{L_0}{2} \|y_{k+1} - \tilde{x}_k\|^2 \leq \frac{L_0}{2} \|y_{k+1} - \tilde{x}_k\|^2.$$

E, juntamente com (4-38) e o fato de que $\lambda = \sigma^2/L_0$ implica que

$$\|\lambda v_{k+1} + y_{k+1} - \tilde{x}_k\|^2 + 2\lambda\varepsilon_{k+1} = 2\lambda\varepsilon_{k+1} \leq \lambda L_0 \|y_{k+1} - \tilde{x}_k\|^2 = \sigma^2 \|\tilde{y}_{k+1} - \tilde{x}_k\|^2.$$

Além disso, note que (4-38) implica que a fórmula de atualização para x_{k+1} na etapa 2 do Algoritmo acima é a mesma do do Método A-HPE. Desse modo, segue o resultado desejado. \square

Os resultados a seguir fornecem a complexidade do método de primeira ordem como consequência da Proposição 49 e os resultados de complexidade para o método A-HPE da seção 4.1.

Proposição 50. *Seja uma sequência $\{y_k\}$ gerada como em (4-34), e as sequências $\{v_k\}$ e $\{\varepsilon_k\}$ definidas pela proposição 49 e a sequência $\{w_k\}$ definido como*

$$w_k = v_k + \nabla g(y_k) - \nabla g(\tilde{x}_{k-1}), \quad k \geq 1,$$

então valem as seguintes afirmações:

a) Para todo $k \geq 1$, temos

$$f(y_k) - f_* \leq \frac{2L_0 d_0^2}{k^2 \sigma^2};$$

b) Se $\sigma < 1$, então

$$\|y_k - x_*\| \leq \left(\frac{2}{\sqrt{1 - \sigma^2}} + 1 \right) d_0;$$

c) Se $\sigma < 1$, então para cada $k \geq 1$, $v_k \in \partial_{\varepsilon_k} g(y_k) + \partial h(y_k) \subseteq \partial_{\varepsilon_k}(g + h)(y_k)$, e existe índice $i \leq k$ tal que

$$\|v_i\| = O\left(\frac{L_0 d_0}{k^{3/2}}\right) \quad \text{e} \quad \varepsilon_i = O\left(\frac{L_0 d_0^2}{k^3}\right);$$

d) Se $\sigma < 1$, então para cada $k \geq 1$, $w_k \in (\nabla g + \partial h)(y_k)$, e existe índice $i \leq k$ tal que

$$\|w_i\| = O\left(\frac{L_0 d_0}{k^{3/2}}\right).$$

Prova. O item a) segue diretamente por Proposição 49, Teorema 45 com $\lambda_k = \lambda = \sigma^2/L_0$, e o item b) segue pelo Teorema 48. O item c) segue pela Proposição 49, juntamente com o item a) do Lema 44 e Proposição 46, com $\lambda_k = \lambda := \sigma^2/L_0$. A inclusão no item d), segue da primeira inclusão em (4-40) com $k := k - 1$ e pela definição de w_k . Observe que $y_k \in \text{Dom}\partial h \subseteq \mathbb{R}^n$ e $\tilde{x}_{k-1} \in \mathbb{R}^n$, segue-se da definição de w_k e pela desigualdade (4-38) que

$$\|w_k\| \leq \|v_k\| + \|\nabla g(y_k) - \nabla g(\tilde{x}_{k-1})\| \leq \|v_k\| + L_0\|y_k - \tilde{x}_{k-1}\| = (1 + \lambda L_0)\|v_k\|.$$

A estimativa da complexidade do item d) segue da desigualdade anterior, o item c), e pela definição do λ . Assim, a proposição está provada. \square

Conclusão

Nessa dissertação apresentamos a análise da complexidade por iteração do método *HPE* (Hybrid Proximal Extragradient) e da sua versão acelerada, o método *A – HPE* (Accelerated Hybrid Proximal Extragradient). Uma análise assintótica dos métodos do ponto proximal exato e do método HPE foi também considerada. Neste trabalho, utilizamos como principais referências os artigos [8, 9, 10, 14].

O método *HPE* consiste em uma versão inexata do método do ponto proximal para obter zero de operadores monótonos maximais. Este método além de resolver os subproblemas proximais de forma inexata utilizando um critério de erro relativo, ele possui uma atualização dos iterados usando um passo extra-gradiente. Apresentamos um exemplo mostrando a importância desse passo para garantir a convergência do método. Vimos também que o HPE pode ser visto como um framework no sentido que outros métodos para resolver problemas de otimização e mais geralmente desigualdade variacional podem ser considerados como um caso particular do HPE. Ilustramos este fato mostrando que o método de Korpelevich pode ser visto como um caso especial do método HPE.

Em uma segunda parte da dissertação, estudamos o método *A – HPE*, o qual é uma versão acelerada do método HPE para resolver problemas de otimização convexa. O método *A – HPE* combina as vantagens do *HPE* com a técnica de aceleração de Nesterov para obter uma melhor taxa de complexidade por iteração. Mostramos que (veja o Teorema 45) o A-HPE utilizado para resolver o problema de otimização convexa $f_* := \min_x f(x)$, gera uma sequência $\{x_k\}$ tal que a sequência de valores funcionais $\{f(x_k) - f_*\}$ aproxima de zero a uma taxa $\mathcal{O}(1/k^2)$. Essa é a mesma complexidade ótima obtida pelos métodos de Nesterov bem como o método *FISTA*, veja [1, 2]. Por outro lado, a complexidade do método HPE é analisada sobre os resíduos da seguinte inclusão monótona $0 \in T(x)$, mais especificamente, vimos que o HPE obtém uma solução η -aproximada deste problema, isto é, uma tripla (x, v, ε) tal que $v \in T^\varepsilon(x)$ e $\max\{\|v\|, \varepsilon\} \leq \eta$, em no máximo $\mathcal{O}(1/\eta^2)$. Também analisamos a complexidade do método *A – HPE* em termos dos resíduos da inclusão monótona $0 \in \partial f(x)$. Mostramos que este método obtém uma solução aproximada para este problema, i.e., uma tripla (x, v, ε) tal que $v \in \partial_\varepsilon f(x)$

e $\max\{\|v\|, \varepsilon\} \leq \eta$, em no máximo $\mathcal{O}(1/\eta^{2/3})$, apresentando assim uma melhor taxa de complexidade do que o método HPE quando utilizado para resolver um problema de otimização convexa.

Para trabalhos futuros, estamos interessados em analisar outras variantes do método HPE utilizando termos inerciais e também analisar do ponto de vista computacional alguns algoritmos que podem ser vistos como casos especiais do método HPE.

Referências Bibliográficas

- [1] A. Beck. *First-Order Methods in Optimization*. MOS SIAM Series on Optimization, (2017).
- [2] A. Beck, and M. Teboulle. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. 2, (2009).
- [3] A. Izmailov and M. Solodov. *Otimização Vol.1. Condições de otimalidade, elementos de análise convexa e de dualidade*. 3.ed. Rio de Janeiro:IMPA, (2014).
- [4] A. N. Iusem. *Proximal point methods in optimization*, 20° Colóquio Brasileiro de Matemática, IMPA, (1995).
- [5] B. Martinet. *Régularisation, d'inequations variationnelles par approximations successives*, Revue Française d'Informatique et de Recherche Operationelle, 154-158,(1970).
- [6] Heinz H. Bauschke, Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Second Ed. Springer, (2017).
- [7] J.J.Moreau. *Proximité et dualité dans un espace Hilbertien*, Bull. Soc. Math. France, 93, 278-279,(1965).
- [8] M. V. Solodov and B. F. Svaiter. *A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator*, Set-Valued Anal., 7(4):323–345, (1999).
- [9] R.D.C. Monteiro and B.F. Svaiter. *An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods*, SIAM J. Optim. 23, 1092-1125, (2013).
- [10] R.D.C. Monteiro and B.F. Svaiter. *On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean*, SIAM J. Optim. 20, 2755-2787, (2010).

- [11] R.D.C. Monteiro, C. Ortiz, and B.F. Svaiter. *An adaptive accelerated first-order method for convex optimization*, Comput. Optim. Appl. 64, 31–73, (2016).
- [12] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. *Enlargement of monotone operators with applications to variational inequalities*, Set-Valued Anal., 5, 159–180, (1997).
- [13] R. T. Rockafellar. *Convex analysis*, Princeton University Press, Princeton, (1970).
- [14] R. T. Rockafellar. *Monotone operators and the proximal point algorithm*. SIAM J. cont. optim. 14, 877–898, (1976).
- [15] R. T. Rockafellar. *On the maximal monotonicity of subdifferential mappings*, Pacific Journal of Mathematics, 33, 209-216, (1970).
- [16] Y. E. Nesterov. *A method for solving the convex programming problem with convergence rate $O(1/k^2)$* , Dokl. Akad. Nauk SSSR, 269, 543–547 (1983).
- [17] Y.E. Nesterov. *Gradient methods for minimizing composite functions*, Math. Program. 140, 1–37, (2012).
- [18] Z. Opial. *Weak convergence of the sequence of successive approximations for nonexpansive mappings*. Bulletin of the American Mathematical Society, 73, 591-597, (1967).