



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA PURA E APLICADA

Juan Estuardo Navarro Caballero

Algoritmos proximais e de decomposição proximal inexatos e fortemente convergentes com efeitos de inércia

Florianópolis
2025

Juan Estuardo Navarro Caballero

Algoritmos proximais e de decomposição proximal inexatos e fortemente convergentes com efeitos de inércia

Tese submetida ao Programa de Pós-Graduação em Matemática Pura e Aplicada da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Matemática Pura e Aplicada, com área de concentração em Matemática Aplicada.

Orientador: Prof. Maicon Marques Alves, Dr.

Florianópolis
2025

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Caballero, Juan Estuardo Navarro

Algoritmos proximais e de decomposição proximal inexatos e fortemente convergentes com efeitos de inércia / Juan Estuardo Navarro Caballero ; orientador, Maicon Marques Alves, 2025.

89 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas, Programa de Pós-Graduação em Matemática Pura e Aplicada, Florianópolis, 2025.

Inclui referências.

1. Matemática Pura e Aplicada. 2. Inclusão monótona. 3. Método ponto proximal. 4. Métodos de decomposição projetivo. 5. Inércia . I. Alves, Maicon Marques . II. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Matemática Pura e Aplicada. III. Título.

Juan Estuardo Navarro Caballero

Algoritmos proximais e de decomposição proximal inexatos e fortemente convergentes com efeitos de inércia

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Douglas S. Gonçalves, Dr.
Universidade Federal de Santa Catarina - UFSC

Prof. Felipe Lara, Dr.
Universidad de Tarapacá - UTA

Prof. Max L. N. Gonçalves, Dr.
Universidade Federal de Goiás - UFG

Prof. Roger Behling, Dr.
Universidade Federal de Santa Catarina - UFSC

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Matemática Pura e Aplicada, com área de concentração em Matemática Aplicada.

Prof. Douglas S. Gonçalves, Dr.
Coordenador do Programa de Pós-Graduação

Prof. Maicon Marques Alves, Dr.
Orientador

Florianópolis, 2025.

Para a minha família.

Agradecimentos

A conclusão desta tese representa a concretização de um longo percurso acadêmico e pessoal, o qual não teria sido possível sem o apoio e a contribuição de muitas pessoas, às quais expresso minha mais profunda gratidão.

Em primeiro lugar, agradeço ao meu orientador, Prof. Dr. Maicon Marques Alves, pela orientação precisa, paciência e incentivo contínuo ao longo de todo este trabalho. Sua expertise e dedicação foram fundamentais para o desenvolvimento desta pesquisa.

Agradeço também aos professores e colegas do programa de pós-graduação em matemática, cujos comentários, sugestões e discussões enriqueceram significativamente minha formação acadêmica.

Aos meus amigos, pela companhia nos momentos difíceis e pelas palavras de apoio e encorajamento. Vocês tornaram esta caminhada mais leve e significativa.

À minha família, meus pais, Milly Caballero e Estuardo Navarro, por todo o amor, apoio incondicional e por sempre acreditarem em mim. A minhas queridas irmãs Mónica e Marillyse, por todos os momentos lindos que tivemos durante nossa infância. Este trabalho é também fruto dos valores que me ensinaram.

Por último agradeço à Universidade Federal de Santa Catarina, por minha formação acadêmica, e agradeço o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-CAPES, que foi essencial para a conclusão do curso de doutorado.

A todos, o meu sincero muito obrigado.

Resumo

Nesta tese, propomos e analisamos uma versão inercial do algoritmo de Solodov e Svaiter (2000) para resolver problemas de inclusão monótonos. Provamos a convergência forte global não assintótica, bem como sua complexidade iterativa (taxa de convergência). Determinamos condições suficientes sobre os parâmetros inerciais para garantir a convergência e a complexidade iterativa, mostrando que essas condições são mais flexíveis em comparação com outros métodos inerciais existentes. Como aplicação, utilizamos essa abordagem para obter métodos inerciais com passos do tipo forward-backward e Korpelevich para resolver problemas de inclusão monótonos estruturados.

Para resolver problemas de inclusão monótonos envolvendo a soma de n operadores monótonos maximais compostos, apresentamos e estudamos uma versão inercial baseada no método de decomposição projetivo de Eckstein e Svaiter (2009). Estabelecemos também as condições sobre os parâmetros inerciais que asseguram a convergência forte e a complexidade iterativa. Por fim, o método proposto é aplicado para obter versões inerciais de métodos do tipo forward-backward para resolver problemas de inclusão monótonos estruturados.

Palavras-chave: Operadores monótonos. Inclusão monótona. Algoritmo ponto proximal. Critérios de erro relativo. Métodos inerciais. Convergência forte. Complexidade iterativa. Métodos de decomposição projetivos. Métodos forward-backward. Método de Korpelevich.

Abstract

In this thesis, we propose and analyze an inertial version of the algorithm of Solodov and Svaiter (2000) for solving monotone inclusion problems. We demonstrate global non-asymptotic strong convergence, as well as its iteration complexity (rate of convergence). We determine sufficient conditions on the inertial parameters to guarantee both convergence and iteration complexity, showing that these conditions are more flexible compared to other existing inertial methods. As an application, we use this approach to derive inertial methods with forward-backward and Korpelevich-type steps for solving structured monotone inclusion problems.

To solve monotone inclusion problems involving the sum of n composed maximal monotone operators, we present and study an inertial version based on the projective splitting method of Eckstein and Svaiter (2009). We also establish the conditions on the inertial parameters that ensure strong convergence and iteration complexity. Finally, the proposed method is applied to derive inertial versions of forward-backward-type methods for solving structured monotone inclusion problems.

Keywords: Monotone operators. Monotone inclusion. Proximal point algorithm. Relative error criteria. Inertial methods. Strong convergence. Iteration complexity. Projective splitting method. Forward-backward methods. Korpelevich's method.

Sumário

Introdução	1
1 Preliminares	6
1.1 Notação geral, definições e resultados básicos	6
1.2 Análise convexa e operadores monótonos	7
1.3 Problemas de inclusão monótonos	11
1.4 Método ponto proximal e suas versões inexatas	11
2 Um algoritmo ponto proximal inexato com efeitos inerciais fortemente convergente para inclusões monótonas	14
2.1 Declaração do problema e soluções aproximadas para subproblemas proximais	15
2.2 Um algoritmo ponto proximal inexato inercial fortemente convergente	17
2.2.1 Análise de convergência do Algoritmo 1	20
2.2.2 Análise da complexidade iterativa	24
2.3 Uma variante inercial fortemente convergente do método extragradiente de Korpelevich	26
2.4 Versões inerciais fortemente convergentes dos métodos forward-backward modificado de Tseng e forward-backward clássico	30
2.4.1 Versão inercial com convergência forte do método forward-backward de Tseng	30
2.4.2 Uma versão inercial do método forward-backward com convergência forte	33
3 Um algoritmo de decomposição projetivo inercial inexato com convergência forte	36
3.1 O problema principal e o algoritmo	36
3.1.1 Análise de convergência do Algoritmo 5	43
3.1.2 Análise da complexidade iterativa do Algoritmo 5	50
3.2 Variantes do algoritmo inercial fortemente convergente de decomposição projetivo com passos do tipo forward-backward e forward-backward-forward . .	53
3.2.1 Uma variante do algoritmo inercial de decomposição projetivo com convergência forte e passos do tipo forward-backward	54
3.2.2 Uma variante do algoritmo inercial de decomposição projetivo fortemente convergente com passos do tipo forward-backward-forward de Tseng	57
4 Conclusões	60

A	Boa definição do Algoritmo 1 e alguns resultados auxiliares	62
A.1	Boa definição do Algoritmo 1	62
A.2	Resultados Auxiliares	63
B	Boa definição do Algoritmo 5 e resultados auxiliares adicionais	64
B.1	Boa definição do Algoritmo 5	64
B.2	Resultados auxiliares adicionais	65
	Referências Bibliográficas	68

Introdução

Operadores monótonos surgem em diversas áreas da matemática aplicada, tais como: otimização convexa, equações diferenciais parciais, economia matemática, análise variacional e teoria do controle. A noção de operador monótono foi introduzida pela primeira vez em [51, 60], e intensamente estudada entre as décadas de 1960 a 1980 por Kachurovskii, Minty, Brézis, Browder e Rockafellar, os quais estabeleceram os conceitos fundamentais sobre esse tema. Inicialmente, muitos dos trabalhos sobre operadores monótonos estavam relacionados à análise funcional e às equações diferenciais (ver [30, 31, 52]), mas logo foi notada a sua importância na análise convexa e na otimização convexa [61, 84, 88]. Um problema fundamental nas aplicações dos operadores monótonos é resolver problemas de inclusão monótonos, ou seja, resolver inclusões do tipo

$$0 \in T(x), \quad (1)$$

em que $T: \mathcal{H} \rightrightarrows \mathcal{H}$ é um operador monótono maximal sobre um espaço de Hilbert real. Essas inclusões surgem porque condições de otimalidade para problemas de otimização convexa podem ser expressas por meio delas. Além disso, devido à sua aplicabilidade na teoria das equações diferenciais parciais, desigualdades variacionais e, sobretudo, na otimização, o estudo e a resolução de inclusões monótonas têm sido um tema que tem atraído a atenção da comunidade matemática.

Um dos métodos mais clássicos para resolver inclusões monótonas é o *método ponto proximal* (PP), introduzido por Martinet em [58] para resolver desigualdades variacionais monótonas e generalizado para inclusões monótonas por Rockafellar em [86]. Esse método consiste em construir uma sequência iterativa baseada em cálculos de resolventes do operador T , garantindo convergência fraca sob hipóteses flexíveis. No entanto, embora esse método possua propriedades de convergência global, sua principal desvantagem é que requer avaliações dos resolventes $(\lambda T + I)^{-1}$. O problema é que calcular a inversa de $(\lambda T + I)$ pode ser tão difícil quanto resolver o problema original (1), o que torna o método ponto proximal tradicional insatisfatório na prática. Uma forma de superar esse problema é introduzir cálculos inexatos de resolventes (ou seja, soluções inexatas de subproblemas proximais), dando origem a versões inexatas do método ponto proximal. Em [86], Rockafellar propôs uma versão inexata desse método baseada em critérios de erros somáveis que utilizam sequências de erros cuja soma é finita. No entanto, como foi observado em [34, 36, 83], esse critério pode apresentar algumas desvantagens práticas, pois não indica como essa sequência de erros somável deve ser escolhida. Assim, é importante determinar condições de erro para aproximar soluções de subproblemas proximais que possam ser calculadas a cada passo durante o processo de iteração. Nos últimos anos, como alternativa ao critério de erro somável de Rockafellar, os métodos com critério de erro relativo têm merecido a atenção de muitos pesquisadores (ver [6, 28, 35, 67, 91, 93]). Esse novo critério de aproximação, introduzido pela primeira vez por

Solodov e Svaiter [89, 90], é amplamente utilizado tanto na teoria quanto na prática em otimização numérica (ver [35, 46, 48, 66, 67]). O primeiro desses métodos foi apresentado por Solodov e Svaiter em [90], a saber, o método híbrido ponto proximal projetivo (hybrid projection proximal point). A principal característica desse método é que ele combina um critério de erro relativo com passos projetivos. Especificamente, a cada passo do processo iterativo, constrói-se um hiperplano que separa estritamente a iteração atual do conjunto solução (os zeros do operador), que se assume não-vazio, e em seguida define-se a próxima iteração como a projeção da iteração atual sobre esse hiperplano. Uma das características mais importantes dos métodos com critério de erro relativo é que eles permitem uma considerável relaxação dos requisitos de tolerância impostos à solução dos subproblemas proximais o que resulta em uma estrutura mais prática baseada nos algoritmos proximais.

Como métodos baseados em iterações do tipo ponto proximal geralmente convergem fracamente (ver (1.2)), é relevante em aplicações propor e estudar métodos com garantias de convergência forte. Com o objetivo de obter métodos fortemente convergentes, Solodov e Svaiter, em [92], propuseram um novo método híbrido baseado em iterações proximal projetivas. Em linhas gerais, nesse método projeta-se a iterada inicial sobre a interseção de dois semi-espacos que contêm o conjunto solução e que são construídos utilizando soluções aproximadas de subproblemas proximais. Essa técnica tornou-se um mecanismo padrão para forçar a convergência forte em métodos do tipo ponto proximal.

Métodos inerciais para a resolução de inclusões monótonas, com origem na discretização implícita (no tempo) de equações diferenciais de segunda ordem, têm atraído a atenção de muitos pesquisadores atualmente (ver [3, 14, 18, 16]). Ao discretizar um sistema dinâmico de segunda ordem, obtém-se um método iterativo em que o próximo iterado depende dos dois anteriores. Um dos exemplos mais clássicos desses métodos foi apresentado por Polyak [82], que introduziu o chamado *método da bola pesada* (*heavy ball method*) para minimizar uma função quadrática fortemente convexa, sendo capaz de acelerar significativamente a convergência do método do gradiente (ver também [81], p.65). Posteriormente, Alvarez e Attouch, em [5], generalizaram essa ideia para operadores monótonos maximais utilizando o esquema do método do ponto proximal, obtendo o método do ponto proximal inercial. Desde então, observa-se um grande interesse da comunidade de otimização em algoritmos proximais de primeira ordem envolvendo efeitos inerciais, incluindo, por exemplo, versões inerciais dos métodos dos multiplicadores de direções alternadas (*ADMM method*), Douglas-Rachford, forward-backward e sua modificação de Tseng (ver [19, 37, 38, 56, 72]), bem como as versões inexatas desses métodos (ver, por exemplo, [4, 8, 9, 28]). O desenvolvimento recente em métodos inerciais está, em parte, relacionado à sua conexão com métodos rápidos de primeira ordem para problemas de otimização convexos (e também não convexos) (ver, por exemplo, [21, 17, 25, 76]). Destacam-se ainda o método do gradiente acelerado de Nesterov [74, 75] que é um esquema iterativo com passos inerciais e cuja taxa de convergência melhora de $O(\frac{1}{k})$ (sem inércia) para $O(\frac{1}{k^2})$ em termos dos valores da função.

Por todas as vantagens potencialmente proporcionadas pelos efeitos inerciais, na primeira parte desta tese propomos e estudamos uma versão inercial do algoritmo de Solodov e Svaiter [92, Algorithm 1], demonstrando sua convergência forte e a taxa de convergência em termos de complexidade iterativa.

Paralelamente ao problema de inclusões simples (envolvendo apenas um operador) como em (1), muitos problemas relevantes na prática envolvem somas de operadores compostos com

operadores lineares, assumindo a forma

$$0 \in \sum_{i=1}^n G_i^* T_i(G_i z), \quad (2)$$

em que $T_i: \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ são operadores monótonos maximais e $G_i: \mathcal{H}_n \rightarrow \mathcal{H}_i$ são operadores lineares contínuos. Tal formulação permite modelar uma ampla gama de problemas como problemas inversos, aprendizagem de máquina e otimização. A estrutura composta e desacoplada desses problemas motiva o uso de algoritmos que permitam decomposição e processamento paralelo, ao mesmo tempo em que preservem a convergência sob a estrutura monótona global.

Dado que, em geral, não existe uma fórmula explícita para o resolvente da soma de operadores em termos dos resolventes de cada um deles, não é possível aplicar diretamente o método ponto proximal a esse problema. Os métodos de decomposição de operadores constituem uma alternativa para superar essa dificuldade, onde o termo “decomposição” indica que, no esquema iterativo, os operadores envolvidos são avaliados de forma individual. No caso particular em que apenas dois operadores estão presentes no problema (2) (com $G_i = I$), mencionamos os métodos de Douglas-Rachford [49, 55, 42], Peaceman-Rachford [79], forward-backward [55, 78] e o forward-backward modificado de Tseng [97] como os mais populares para a resolução desse tipo de problema. Para uma soma finita arbitrária de operadores, Spingarn em [94] propôs um método de decomposição de operadores no qual, a cada passo do processo iterativo, a solução dos subproblemas é realizada em paralelo. Recentemente, Eckstein e Svaiter desenvolveram em [43, 44] uma nova família de métodos de decomposição de operadores conhecida como *métodos de decomposição projetivos* (*projective splitting methods*). De modo geral, este método consiste em “incrustar” (2) em um espaço produto ampliado, reformulando o problema de inclusão monótono como um problema de viabilidade convexa que busca um ponto no conjunto de soluções estendidas de (2) (ver (3.4)). Dado o problema de viabilidade, planos cortantes são construídos utilizando funcionais afins derivados de elementos no gráfico dos operadores monótonos maximais associados ao problema original (2). Essa nova estrutura de decomposição é mais flexível do que outros métodos de decomposição e apresenta um desempenho numérico promissor em algumas aplicações (ver, por exemplo, [50]). Os métodos do tipo decomposição projectiva geralmente tem convergência fraca, sendo necessárias modificações no mecanismo de iteração para se obter a convergência forte das iteradas. Na segunda parte desta tese propomos e estudamos um algoritmo inercial de decomposição projetiva com convergência forte para resolver (2), com o objetivo de explorar as vantagens dos efeitos inerciais combinado à decomposição projetiva. Até onde sabemos, esta é a primeira vez na literatura que algoritmos inerciais do tipo decomposição projetiva com convergência forte são propostos.

Motivação e objetivos

A crescente complexidade dos problemas da matemática aplicada como otimização convexa, aprendizado de máquina, problemas inversos entre outros motivou o desenvolvimento de métodos cada vez mais eficientes para a resolução de inclusões monótonas. Em particular, dois tipos de problemas se destacam por sua relevância prática e teórica:

1. Inclusões clássicas da forma

$$0 \in T(x)$$

em que T é um operador monótono maximal. Tais problemas aparecem em formulações canônicas de otimização convexa, desigualdades variacionais, entre outros.

2. Inclusões de soma composta da forma $0 \in \sum_{i=1}^n G_i^* T_i(G_i z)$, que modelam problemas estruturados, onde os operadores atuam em diferentes espaços e são acoplados por operadores lineares.

Embora existam métodos clássicos bem estabelecidos para tratar essas duas classes de problemas, como o método ponto proximal de Rockafellar e o método de decomposição projetivo de Eckstein e Svaiter, ainda há limitações em termos de desempenho computacional, principalmente quando aplicados a problemas de alta dimensão.

A motivação principal deste trabalho reside na busca por algoritmos inerciais que combinem a robustez dos métodos clássicos com o potencial de aceleração obtido por dinâmicas inspiradas em sistemas de segunda ordem. Métodos inerciais têm sido estudados extensivamente na otimização convexa, mostrando vantagens teóricas e empíricas em diversos contextos. No entanto, sua aplicação a problemas de inclusão monótonos, especialmente nas versões projetivas e compostas, ainda está em fase de desenvolvimento e carece de uma análise mais profunda.

Neste contexto, os objetivos específicos deste trabalho são:

- Propor e analisar uma versão inercial do método híbrido com projeções de Solodov e Svaiter, com o objetivo de resolver inclusões da forma $0 \in T(x)$. O foco estará em garantir convergência forte da sequência gerada, mantendo as características de simplicidade e robustez do algoritmo original. Posteriormente, aplicamos o algoritmo proposto para obter versões inerciais de algoritmos que utilizam passos do tipo forward-backward e de Korpelevich para resolver problemas estruturados de inclusão monótona.
- Propor uma versão inercial do método de decomposição projetivo de Eckstein e Svaiter para resolver inclusões do tipo $0 \in \sum_{i=1}^n G_i^* T_i(G_i z)$. De forma análoga ao caso anterior, também utilizamos o algoritmo proposto para resolver problemas estruturados com múltiplos termos.
- Estabelecer uma análise teórica rigorosa dos métodos propostos, incluindo a prova de convergência forte (convergência assintótica) e a complexidade iterativa (taxa de convergência). Determinar as condições suficientes sobre os parâmetros inerciais para garantir esses resultados.

Este trabalho se insere no esforço de modernizar os algoritmos baseados em monotonicidade, integrando técnicas de inércia e decomposição projectiva, com vistas a aplicações em larga escala e alta performance. Espera-se, com isso, contribuir tanto para o avanço teórico da área quanto para o desenvolvimento de ferramentas computacionais mais eficazes em contextos aplicados.

Principais contribuições. A seguir, apresentamos um resumo das principais contribuições desta tese.

- (a) Contribuições do Capítulo 2. Propomos e analisamos uma versão inercial do algoritmo fortemente convergente de Solodov e Svaiter [92, Algorithm 1] para a resolução de problemas de inclusão monótonos como em (1). Provamos a convergência forte assintótica

global e a complexidade iterativa (taxa de convergência $\mathcal{O}(1/\sqrt{k})$) do algoritmo proposto (Algoritmo 1). Além disso, aplicamos o algoritmo principal para resolver problemas estruturados do tipo $0 \in F(x) + B(x)$, obtendo versões inerciais de algoritmos que utilizam passos do tipo Korpelevich, forward-backward modificado de Tseng, forward-backward clássico (Algoritmos 2, 3, 4, respectivamente).

- (b) Contribuições do Capítulo 3. Propomos e analisamos uma versão inercial com convergência forte do método de decomposição projetivo introduzido por Eckstein e Svaiter em [43, 44] para resolver (2). Provamos a convergência assintótica forte e a complexidade iterativa do algoritmo proposto (Algoritmo 5).

De forma análoga ao caso anterior, o algoritmo principal também é utilizado para resolver problemas estruturados (quando $T_i = F_i + B_i$ em (2)), obtendo-se assim algoritmos inerciais com convergência forte que utilizam passos do tipo forward-backward.

Estrutura da tese. Esta tese está organizada da seguinte forma.

Capítulo 1: contém quatro seções. Na Seção 1.1, apresentamos a notação geral, definições e alguns resultados básicos que serão utilizados ao longo desta tese. Na Seção 1.2, revisamos algumas definições e damos uma breve visão geral da teoria dos operadores monótonos e da análise convexa. Na Seção 1.3, revisamos brevemente as inclusões monótonas e fornecemos alguns exemplos dessas. Na Seção 1.4, apresentamos o método ponto proximal e algumas de suas versões inexatas.

Capítulo 2: propõe-se e estuda-se uma versão inercial do algoritmo de Solodov e Svaiter [92, Algorithm 1] para resolver problemas de inclusão monótonos. Este capítulo contém quatro seções. Na Seção 2.1, apresentamos a formulação geral de um problema de inclusão monótono, algumas formas estruturadas desse tipo de problema, e definimos soluções aproximadas para os subproblemas proximais, além de algumas propriedades que essas soluções satisfazem. Na Seção 2.2, é apresentado o algoritmo principal deste capítulo (Algoritmo 1), sendo discutida sua convergência assintótica forte, bem como sua complexidade iterativa. Na Seção 2.3, apresenta-se uma versão inercial com convergência forte do método de Korpelevich. Na Seção 2.4, são apresentadas versões inerciais com convergência forte dos métodos forward-backward e forward-backward modificado de Tseng.

Capítulo 3: apresenta-se uma versão inercial inexata e fortemente convergente do método de decomposição projetivo para resolver problemas de inclusão monótonos estruturados envolvendo a soma finita de operadores monótonos maximais. Este capítulo contém duas seções. Na Seção 3.1, é apresentado o algoritmo principal deste capítulo (Algoritmo 5), sendo demonstradas sua convergência forte e sua complexidade iterativa. Na Seção 3.2, estabelece-se uma aplicação do algoritmo proposto na seção anterior, obtendo-se versões inexatas e fortemente convergentes de métodos do tipo forward-backward para resolver problemas de inclusão monótonos com múltiplos termos.

Capítulo 4: se apresentam as conclusões da tese e alguns possíveis trabalhos futuros.

Capítulo 1

Preliminares

Neste capítulo apresentamos algumas definições e resultados básicos que serão usados ao longo desta tese. Começamos revisando alguns conceitos de análise convexa e a teoria de operadores monótonos. Trataremos também os problemas de inclusão monótonos e apresentamos o método ponto proximal e suas versões inexatas para resolvê-los. Finalmente, damos uma visão geral dos efeitos inerciais em métodos iterativos.

1.1 Notação geral, definições e resultados básicos

Seja $\mathcal{H}$ um espaço de Hilbert real com produto interno $\langle \cdot, \cdot \rangle$ e norma induzida $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. Para $C \subseteq \mathcal{H}$ um subconjunto não-vazio, fechado e convexo, definimos a *projeção ortogonal* de x sobre C , $P_C(x)$, como o único ponto em C que satisfaz

$$\|P_C(x) - x\| \leq \|x - y\| \quad \forall y \in C.$$

Uma sequência $\{x_k\} \subset \mathcal{H}$ *converge fortemente* para $\bar{x} \in \mathcal{H}$ se

$$\|x_k - \bar{x}\| \rightarrow 0, \quad k \rightarrow \infty \quad (1.1)$$

e *converge fracamente* se

$$\langle x_k, x \rangle \rightarrow \langle \bar{x}, x \rangle, \quad k \rightarrow \infty, \quad \forall x \in \mathcal{H}. \quad (1.2)$$

A convergência forte e fraca de uma sequência será denotada por $\rightarrow$ e $\rightharpoonup$, respectivamente. Além disso, $\mathbb{R}_+$ e $\mathbb{R}_{++}$ denotarão o conjunto dos números reais não negativos e (estritamente) positivos, respectivamente.

Seja $G: \mathcal{H} \rightarrow \mathcal{V}$ um operador linear contínuo (limitado), onde $\mathcal{H}$ e $\mathcal{V}$ são espaços de Hilbert reais, o operador $G^*: \mathcal{V} \rightarrow \mathcal{H}$ denota o adjunto de G . A norma de G é dada por

$$\|G\| = \sup_{x \neq 0} \frac{\|Gx\|}{\|x\|}.$$

O seguinte resultado estabelece algumas propriedades conhecidas do operador projeção; para sua demonstração o leitor pode consultar [98].

Lema 1.1.1. *Sejam $C \subseteq \mathcal{H}$ não-vazio, fechado e convexo e $P_C(\cdot)$ a projeção ortogonal sobre C . Então, para todo $x, y \in \mathcal{H}$ e $z \in C$,*

$$\|P_C(x) - P_C(y)\|^2 \leq \|x - y\|^2 - \|[x - P_C(x)] - [y - P_C(y)]\|^2, \quad (1.3)$$

$$\langle x - P_C(x), z - P_C(x) \rangle \leq 0. \quad (1.4)$$

Lema 1.1.2. [50] *Para quaisquer vetores $x_1, \dots, x_m$ em $\mathcal{H}$, tem-se que*

$$\left\| \sum_{i=1}^m x_i \right\|^2 \leq m \sum_{i=1}^m \|x_i\|^2. \quad (1.5)$$

1.2 Análise convexa e operadores monótonos

Um operador $T: \mathcal{H} \rightrightarrows \mathcal{H}$ é chamado de *operador ponto-conjunto* (*set-valued operator*), se T leva cada ponto $x \in \mathcal{H}$ em um subconjunto de $\mathcal{H}$, isto é, $T: \mathcal{H} \rightrightarrows \mathcal{H}$ com $x \in \mathcal{H} \mapsto T(x) \subseteq \mathcal{H}$. O *domínio* (domínio efetivo) e *imagem* de um operador ponto-conjunto $T: \mathcal{H} \rightrightarrows \mathcal{H}$ são definidos, respectivamente, por

$$\text{Dom } T = \{x \in \mathcal{H} \mid T(x) \neq \emptyset\} \quad \text{e} \quad \text{Ran } T = \{u \in \mathcal{H} \mid \exists x \in \mathcal{H} : u \in T(x)\}.$$

O gráfico de T é dado por

$$\text{Gra } T = \{(x, u) \in \mathcal{H} \times \mathcal{H} \mid u \in T(x)\}.$$

A *inversa* de T é o operador ponto-conjunto $T^{-1}: \mathcal{H} \rightrightarrows \mathcal{H}$ definido como $x \in T^{-1}(u)$ se, e somente se, $u \in T(x)$, isto é,

$$T^{-1}(u) := \{x \in \mathcal{H} \mid u \in T(x)\}.$$

Sejam $T: \mathcal{H} \rightrightarrows \mathcal{H}$ e $S: \mathcal{H} \rightrightarrows \mathcal{H}$ operadores ponto-conjunto e $\alpha > 0$, definimos os operadores αT e $T + S$ como

$$(\alpha T)(x) =: \alpha T(x) = \{\alpha u \in \mathcal{H} \mid u \in T(x)\}$$

e

$$(T + S)(x) = \{u + v \in \mathcal{H} \mid u \in T(x), v \in S(x)\}.$$

Se necessário, identificaremos uma função ponto-ponto (single-valued map) $F: \text{Dom } F \subseteq \mathcal{H} \rightarrow \mathcal{H}$ com a sua representação ponto-conjunto $F: \mathcal{H} \rightrightarrows \mathcal{H}$ por

$$F(x) = \begin{cases} \{F(x)\}, & \text{se } x \in \text{Dom } F, \\ \emptyset, & \text{se } x \notin \text{Dom } F. \end{cases}$$

O *resolvente* (com parâmetro $\lambda > 0$) de $T: \mathcal{H} \rightrightarrows \mathcal{H}$, denotado por $J_{\lambda T}$, é definido como

$$J_{\lambda T}(x) = (\lambda T + I)^{-1}(x),$$

onde $I: \mathcal{H} \rightarrow \mathcal{H}$ é o operador identidade sobre $\mathcal{H}$.

Definição 1.2.1. Um operador ponto-conjunto $T: \mathcal{H} \rightrightarrows \mathcal{H}$ é chamado de monótono se

$$\langle x - y, u - v \rangle \geq 0, \quad \forall (x, u), (y, v) \in \text{Gra } T$$

e monótono maximal se é monótono e o seu gráfico não está propriamente contido no gráfico de qualquer outro operador monótono sobre $\mathcal{H}$, isto é, se $S: \mathcal{H} \rightrightarrows \mathcal{H}$ é um operador monótono tal que $\text{Gra } T \subseteq \text{Gra } S$, então $T = S$.

Proposição 1.2.2. Seja $T: \mathcal{H} \rightrightarrows \mathcal{H}$ um operador monótono maximal, então $J_{\lambda T}(x): \mathcal{H} \rightarrow \mathcal{H}$ é uma função ponto-ponto monótono maximal.

Demonstração. [24, Proposition 23.8, Corollary 23.11]. □

Definição 1.2.3. Seja $f: \mathcal{H} \rightarrow (-\infty, \infty]$ uma função real estendida.

(i) f é convexa, se para todo $x, y \in \mathcal{H}$ e $\lambda \in [0, 1]$,

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

(ii) O domínio e epigrafo de f são definidos por

$$\text{dom } f = \{x \in \mathcal{H} \mid f(x) < \infty\} \quad \text{e} \quad \text{epi } f = \{(x, \mu) \in \mathcal{H} \times \mathbb{R} \mid \mu \geq f(x)\},$$

respectivamente.

(iii) f é chamada de própria se $\text{dom } f \neq \emptyset$.

(iv) f é semi-contínua inferiormente (ou fechada) se $\text{epi } f$ é fechado em $\mathcal{H} \times \mathbb{R}$.

Definição 1.2.4. O subdiferencial de uma função $f: \mathcal{H} \rightarrow (-\infty, \infty]$ própria é o operador ponto-conjunto $\partial f: \mathcal{H} \rightrightarrows \mathcal{H}$ dado por

$$\partial f(x) = \begin{cases} \{v \in \mathcal{H} \mid f(y) \geq f(x) + \langle v, y - x \rangle \quad \forall y \in \mathcal{H}\}, & \text{se } x \in \text{dom } f \\ \emptyset, & \text{se } x \notin \text{dom } f. \end{cases}$$

Se existe $x \in \mathcal{H}$ tal que $\partial f(x) \neq \emptyset$ diz-se que f é subdiferenciável em x , e nesse caso um vetor $v \in \partial f(x)$ é chamado de *subgradiente* de f em x . Note que se f é própria e convexa, então ∂f é trivialmente monótono. Outras propriedades da subdiferencial são resumidas na seguinte proposição.

Proposição 1.2.5. Seja $f: \mathcal{H} \rightarrow (-\infty, \infty]$ própria e convexa, e $x \in \text{dom } f$. Então

(a) $\text{Dom } \partial f \subseteq \text{dom } f$.

(b) Se f é Gâteaux diferenciável em x , então $\partial f(x) = \{\nabla f(x)\}$.

(c) Se f é semi-contínua inferiormente, então ∂f é monótono maximal.

Demonstração. Ver [24, Proposition 16.4, Proposition 17.31 e Theorem 20.25]. □

Definição 1.2.6. Sejam $f: \mathcal{H} \rightarrow (-\infty, \infty]$ própria e $x \in \mathcal{H}$. Então x é um minimizador (global) de f se $f(x) = \inf_{x \in \mathcal{H}} f(x) \in \mathbb{R}$. O conjunto de minimizadores de f é denotado por $\text{Argmin } f$. Se $\text{Argmin } f$ é unitário, o seu único elemento é denotado por $\argmin_{x \in \mathcal{H}} f(x)$.

Teorema 1.2.7 (Regra de Fermat). *Seja $f: \mathcal{H} \rightarrow (-\infty, \infty]$ própria. Então*

$$\operatorname{Argmin} f = \{x \in \mathcal{H} \mid 0 \in \partial f(x)\}.$$

Demonstração. [24, Theorem 16.3]. □

Sejam $f: \mathcal{H} \rightarrow (-\infty, \infty]$ uma função própria, convexa e semi-contínua inferiormente e $\lambda > 0$, o operador proximal $\operatorname{prox}_{\lambda f}: \mathcal{H} \rightarrow \mathcal{H}$ é dado por

$$\operatorname{prox}_{\lambda f}(x) := \operatorname{argmin}_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\lambda} \|x - y\|^2 \right\}. \quad (1.6)$$

Um fato importante é que o operador proximal é o resolvente do subdiferencial da função associada [24, Proposition 16.44], isto é,

$$\operatorname{prox}_{\lambda f}(x) = (\lambda \partial f + I)^{-1}(x). \quad (1.7)$$

Dado $C \subseteq \mathcal{H}$ um subconjunto não-vazio, fechado e convexo. O operador cone normal de C , $N_C: \mathcal{H} \rightrightarrows \mathcal{H}$ é definido como

$$N_C(x) = \begin{cases} \{v \in \mathcal{H} \mid \langle v, x - y \rangle \geq 0 \quad \forall y \in C\}, & \text{se } x \in C \\ \emptyset, & \text{se } x \notin C. \end{cases} \quad (1.8)$$

É fácil verificar que [24, Example 16.13]

$$N_C(x) = \partial \delta_C(x),$$

onde, $\delta_C: \mathcal{H} \rightarrow (0, \infty]$ é a função indicadora de C definida como

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

O operador N_C é monótono maximal [24, Example 20.26].

Dado $\varepsilon \geq 0$, o ε -subdiferencial de uma função $f: \mathcal{H} \rightarrow (-\infty, \infty]$ própria é definido em cada $x \in \mathcal{H}$ como

$$\partial_\varepsilon f(x) := \{v \in \mathcal{H} \mid f(y) \geq f(x) + \langle v, y - x \rangle - \varepsilon \quad \forall y \in \mathcal{H}\}.$$

Definição 1.2.8. Uma função $F: \operatorname{Dom} F \subseteq \mathcal{H} \rightarrow \mathcal{H}$ é cocoerciva com constante $\beta > 0$ (ou β -cocoerciva) se

$$\langle x - y, F(x) - F(y) \rangle \geq \beta \|F(x) - F(y)\|^2 \quad \text{para todo } x, y \in \operatorname{Dom} F.$$

e L -Lipschitz contínua (para algum $L > 0$) se

$$\|F(x) - F(y)\| \leq L \|x - y\| \quad \text{para todo } x, y \in \operatorname{Dom} F.$$

Note que se F é β -cocoerciva, então F é $1/\beta$ -Lipschitz contínua.

Definição 1.2.9. *Sejam $T: \mathcal{H} \rightrightarrows \mathcal{H}$ um operador monótono maximal e $\varepsilon \geq 0$, o ε -enlargement de T , $T^{[\varepsilon]}: \mathcal{H} \rightrightarrows \mathcal{H}$ é definido como*

$$T^{[\varepsilon]}(x) := \{u \in \mathcal{H} \mid \langle x - y, u - v \rangle \geq -\varepsilon \text{ para todo } (y, v) \in \text{Gra } T\}. \quad (1.9)$$

Os ε -enlargements de T foram apresentados por Burachik, Iusem e Svaiter em [32], estes representam aproximações externas de T , no sentido que, se T é monótono, então a inclusão $T(x) \subseteq T^{[\varepsilon]}(x)$ é satisfeita trivialmente, justificando assim o nome “enlargement”.

Algumas propriedades importantes dos ε -enlargements e ε -subdiferenciais são resumidas na proposição a seguir (para uma demonstração, ver, por exemplo, [32, 33]).

Proposição 1.2.10. *Sejam $T, S: \mathcal{H} \rightrightarrows \mathcal{H}$ operadores ponto-conjunto. Então*

- (a) *Se $\varepsilon_1 \leq \varepsilon_2$, então $T^{[\varepsilon_1]}(x) \subseteq T^{[\varepsilon_2]}(x)$ para todo $x \in \mathcal{H}$.*
- (b) *$T^{[\varepsilon_1]}(x) + S^{[\varepsilon_2]}(x) \subseteq (T + S)^{[\varepsilon_1 + \varepsilon_2]}(x)$ para todo $x \in \mathcal{H}$ e $\varepsilon_1, \varepsilon_2 \geq 0$.*
- (c) *T é monótono se, e somente se, $\text{Gra } T \subseteq \text{Gra } T^{[0]}$.*
- (d) *T é mononótono maximal se, e somente se, $T = T^{[0]}$.*
- (e) *Se T é monótono maximal, a sequência $(y_k, v_k, \varepsilon_k)$ é tal que $v_k \in T^{[\varepsilon_k]}(y_k)$, $y_k \rightarrow y$, $v_k \rightarrow v$ e $\varepsilon_k \rightarrow \varepsilon$, então $v \in T^{[\varepsilon]}(y)$.*
- (f) *Se $f: \mathcal{H} \rightarrow (-\infty, \infty]$ é própria, convexa e semi-contínua inferiormente, então $\partial_\varepsilon f(x) \subseteq (\partial f)^\varepsilon(x)$ para todo $x \in \mathcal{H}$ e $\varepsilon \geq 0$.*
- (g) *Se $u \in \partial f(x)$ e $y \in \text{dom } f$, então $u \in \partial_\varepsilon f(y)$, onde $\varepsilon = f(y) - (f(x) + \langle y - x, u \rangle)$.*

Lema 1.2.11. *Seja $F: \text{Dom } F \subseteq \mathcal{H} \rightarrow \mathcal{H}$ uma aplicação L^{-1} -cocoerciva. Então, para todo $z, z' \in \text{Dom } F$, temos*

$$F(z) \in F^{[\varepsilon]}(z'), \quad \text{com } \varepsilon = \frac{L\|z' - z\|^2}{4}.$$

Demonstração. [95, Lemma 2.2]. □

A demonstração do próximo lema decorre diretamente da definição de ε -enlargement de operadores ponto-conjunto, combinada com manipulações algébricas simples.

Lema 1.2.12. *Sejam $A: \mathcal{H} \rightrightarrows \mathcal{H}$ e $B: \mathcal{G} \rightrightarrows \mathcal{G}$ operadores monótono maximais, e sejam $\varepsilon^a, \varepsilon^b \geq 0$. Então, para todo $x \in \mathcal{H}$ e $y \in \mathcal{G}$, vale*

$$A^{[\varepsilon^a]}(x) \times B^{[\varepsilon^b]}(y) \subseteq (A \times B)^{[\varepsilon^a + \varepsilon^b]}(x, y),$$

onde $A \times B: \mathcal{H} \times \mathcal{G} \rightrightarrows \mathcal{H} \times \mathcal{G}$ é definido por $(A \times B)(z, v) = A(z) \times B(v)$. ($\mathcal{H} \times \mathcal{G}$ munido do produto interno usual.)

1.3 Problemas de inclusão monótonos

Dado um operador monótono maximal $T: \mathcal{H} \rightrightarrows \mathcal{H}$, um problema de inclusão monótono (PIM) consiste em

$$\text{encontrar } x \in \mathcal{H} \quad \text{tal que} \quad 0 \in T(x). \quad (1.10)$$

Denotaremos por $\text{Zer } T$ ou $T^{-1}(0)$ o conjunto de pontos que satisfaz (1.10), isto é,

$$\text{Zer } T = T^{-1}(0) = \{x \in \mathcal{H} \mid 0 \in T(x)\},$$

que são chamados os *zeros do operador* T . Um grande número de problemas da matemática aplicada tais como otimização convexa, desigualdades variacionais, problemas de pontos sela e de equilíbrio podem ser resolvidos no contexto de (1.10). Vejamos alguns exemplos.

Exemplo 1.3.1 (Problemas de minimização). Seja $f: \mathcal{H} \rightarrow (-\infty, \infty]$ própria, convexa e semi-contínua inferiormente. Considere o problema de minimização (convexa)

$$\min_{x \in \mathcal{H}} f(x). \quad (1.11)$$

Pelo Teorema 1.2.7, resolver (1.11) é equivalente a encontrar $x \in \mathcal{H}$ tal que $0 \in \partial f(x)$ (ou $0 = \nabla f(x)$ se f é diferenciável, ver Proposição 1.2.5 (a)). Portanto, o problema de minimização (1.11) pode ser visto como um PIM com $T = \partial f$.

Exemplo 1.3.2 (Problema de desigualdade variacional (PDV)). Sejam $F: \mathcal{H} \rightarrow \mathcal{H}$ uma função monótona contínua e $C \subseteq \mathcal{H}$ um subconjunto não-vazio, fechado e convexo. O problema de desigualdade variacional para F e C é encontrar $x^* \in C$ tal que

$$\langle F(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C. \quad (1.12)$$

Note que da definição $N_C(\cdot)$ em (1.8), (1.12) é equivalente a $0 \in F(x^*) + N_C(x^*) =: T(x^*)$ (monótono maximal, ver [45, Proposition 12.3.6]), e assim, um PDV também pode ser visto como um PIM.

Outros exemplos de PIM são os problemas de ponto sela, ver [87] (ver também [85] e as referências ali contidas para outros exemplos especiais).

1.4 Método ponto proximal e suas versões inexatas

Um dos métodos iterativos mais conhecidos para resolver o problema (1.10) é o método ponto proximal (PP) (ou algoritmo ponto proximal). O método PP foi inicialmente proposto por Martinet em [58] inspirado pelo trabalho prévio de Moreau [71] e posteriormente generalizado para operadores monótonos maximais por Rockafellar em [86]. Este último formulou o método PP como o seguinte esquema iterativo

$$x_{k+1} = (\lambda_k T + I)^{-1}(x_k), \quad \forall k \geq 0 \quad (1.13)$$

onde $\{\lambda_k\} \subset \mathbb{R}_{++}$ é uma sequência de parâmetros de regularização (ou parâmetros de tamanho de passo). Em [86], Rockafellar mostrou que se a sequência $\{\lambda_k\}$ é limitada inferiormente

longe de zero (existe $\lambda > 0$ tal que $\lambda_k \geq \lambda$) e o conjunto de zeros $T^{-1}(0)$ é não-vazio, então a sequência $\{x_k\}$ gerada em (1.13) converge fracamente para algum ponto solução $x^* \in T^{-1}(0)$. O método PP apresentado em (1.13) pode ser motivado observando que

$$0 \in T(x) \iff J_{\lambda T}(x) = x$$

sendo $J_{\lambda T} = (\lambda T + I)^{-1}$ um operador não expansivo (ver [29, Proposition 2.2]).

Para determinar x_{k+1} em (1.13) é necessário calcular exatamente a inversa de $(\lambda T + I)$. Em muitos casos, isso é tão difícil quanto resolver o problema principal (1.10). Uma forma de superar essa dificuldade é calcular resolventes de forma inexata. O primeiro esquema ponto proximal inexato foi apresentado em [86], utilizando um critério de somabilidade na tolerância do erro. Especificamente, em [86], a equação (1.13) é relaxada e a regra de aproximação com seu critério de tolerância de erro é

$$\|x_{k+1} - (\lambda_k T + I)^{-1}(x_k)\| < e_k, \quad \sum_{k=0}^{\infty} e_k < \infty. \quad (1.14)$$

A condição (1.14), juntamente com a suposição de que $\{\lambda_k\}$ é limitada longe de zero, garante a convergência fraca das iteradas para algum $x^* \in T^{-1}(0)$, desde que esse conjunto seja não-vazio [86, Theorem 1].

Note que, na regra de tolerância de erro em (1.14), $(\lambda_k T + I)^{-1}(x_k)$ é o ponto a ser aproximado, de modo que ele não está disponível. Portanto, o critério anterior não pode ser utilizado diretamente nas implementações do algoritmo. No entanto, em [86, Proposition 3], foi estabelecido que (1.14) é implicado por

$$\text{dist}(0, \lambda_k T(x_{k+1}) + x_{k+1} - x_k) \leq e_k, \quad \sum_{k=0}^{\infty} e_k < \infty. \quad (1.15)$$

A condição (1.15) é mais fácil de verificar na prática, pois requer apenas uma avaliação de T em x_{k+1} . Assim, dado x_k uma aproximação atual, definimos o ponto x_{k+1} como aquele que satisfaz (1.15). Nesse caso, tem-se

$$x_{k+1} \approx (\lambda_k T + I)^{-1}(x_k), \quad \forall k \geq 0. \quad (1.16)$$

Em (1.16), “ $\approx$ ” significa que x_{k+1} é uma aproximação da iterada (exata) $(\lambda_k T + I)^{-1}(x_k)$, definindo assim uma versão inexata do método PP. Em [86], Rockafellar provou a convergência fraca de (1.16) sob o critério de erro de somabilidade dada anteriormente e aplicou os resultados obtidos à justificação de algoritmos de Lagrangianos aumentados para programação convexa [85].

Em geral, as sequências geradas como em (1.16) não convergem fortemente (ver [47]). O problema de forçar a convergência forte do algoritmo PP (1.13) (introduzindo modificações convenientes no processo iterativo) foi abordado por Solodov e Svaiter em [92], no contexto de métodos PP inexatos com critérios de erro relativo. Os métodos do tipo PP inexatos utilizando critérios de erro relativo surgiram como uma alternativa aos algoritmos proximais que empregam critérios de erro somáveis para resolver subproblemas proximais; os primeiros métodos desse tipo foram propostos em [89, 90] e posteriormente estudados, por

exemplo, em [13, 67, 68, 69]. A ideia central consiste em observar que a iteração PP exata $x_+ = (\lambda T + I)^{-1}(x)$ é equivalente ao sistema inclusão-equação

$$\begin{cases} v \in T(x_+), \\ \lambda v + x_+ - x = 0 \end{cases} \quad (1.17)$$

e, em seguida, relaxar (1.17) dentro dos critérios de tolerância de erro relativo - ver Definição 2.1.1 mais adiante.

Capítulo 2

Um algoritmo ponto proximal inexato com efeitos inerciais fortemente convergente para inclusões monótonas

Neste capítulo, propõe-se e estuda-se uma variante inercial do algoritmo ponto proximal (PP) inexato fortemente convergente de Solodov e Svaiter ([92, Algorithm 1]) para inclusões monótonas. Provamos a convergência assintótica do algoritmo principal proposto (a saber, Algoritmo 1) sob hipóteses menos restritivas sobre os parâmetros inerciais em relação a outros algoritmos do tipo proximal inercial (como as apresentadas em [9, Theorem 2.7] e [5, Theorem 2.1]). Também determinamos a sua taxa de convergência não assintótica em termos de complexidade iterativa. Como exemplos especiais, aplicamos nosso algoritmo para resolver problemas estruturados de operadores monótonos, obtendo variantes (inerciais) fortemente convergente dos conhecidos métodos de Korpelevich, forward-backward modificado de Tseng (FBT) e forward-backward (clássico) (FB) (que são os Algoritmos 2, 3 e 4, respectivamente).

A base para o desenvolvimento deste capítulo é o artigo [92]: introduziremos efeitos inerciais sobre a iteração de [92, Algorithm 1], obtendo assim um algoritmo PP inexato (com erro relativo) inercial fortemente convergente. Algoritmos proximais com efeitos inerciais para resolver problemas de inclusão monótonos e problemas relacionados foram propostos pela primeira vez no artigo [5] e posteriormente estudados e desenvolvidos em outras linhas de pesquisa por diversos autores (ver, por exemplo, [8, 22, 18, 20, 16, 27, 39] e referências ali contidas). A ideia principal consiste em, dada uma iteração atual, digamos x_k , introduzir um “efeito inercial” extrapolando x_k na direção dada por $x_k - x_{k-1}$, ou seja,

$$w_k = x_k + \alpha_k(x_k - x_{k-1}),$$

onde $\alpha_k \geq 0$, e então utilizar a informação w_k em vez de x_k para atualizar e obter a próxima iteração x_{k+1} .

Os resultados apresentados neste capítulo estão publicados em [11].

2.1 Declaração do problema e soluções aproximadas para subproblemas proximais

Queremos resolver o problema de inclusão monótono (PIM)

$$0 \in T(x), \quad (2.1)$$

onde $T: \mathcal{H} \rightrightarrows \mathcal{H}$ é um operador monótono maximal, assim como as inclusões estruturadas

$$0 \in F(x) + B(x) \quad (2.2)$$

e

$$0 \in F(x) + N_C(x) \quad (2.3)$$

onde $F: \text{Dom } F \subseteq \mathcal{H} \rightarrow \mathcal{H}$ é um operador monótono (ponto-ponto), $B: \mathcal{H} \rightrightarrows \mathcal{H}$ é um operador monótono maximal (ponto-conjunto) e $N_C: \mathcal{H} \rightrightarrows \mathcal{H}$ é o cone normal do conjunto não-vazio, fechado e convexo $C \subseteq \mathcal{H}$ (ver (1.8)).

O problema (2.1) e suas versões estruturadas (2.2) e (2.3) aparecem em uma ampla classe de aplicações, incluindo otimização convexa, aprendizagem de máquina, economia, entre outros (ver, por exemplo, [24] e referências nele). Como mencionado na Seção 1.3 do Capítulo 1 (ver (1.12)), o problema (2.3) é equivalente ao problema de desigualdade variacional de encontrar $x \in C$ tal que

$$\langle F(x), y - x \rangle \geq 0 \quad \forall y \in C. \quad (2.4)$$

Concentraremos agora nossos esforços na resolução numérica dos problemas (2.1), (2.2) e (2.3). Ou seja, desejamos implementar algoritmos iterativos capazes de calcular soluções aproximadas para esses problemas.

Como foi estabelecido na Seção 1.4 do capítulo anterior (ver (1.17)), o conceito de critério de erro relativo para subproblemas proximais de uma iteração ponto proximal exata $y = (\lambda T + I)^{-1}(x)$ é equivalente a resolver o sistema inclusão-equação

$$\begin{cases} v \in T(y), \\ \lambda v + y - x = 0. \end{cases} \quad (2.5)$$

O sistema (2.5) motiva a seguinte definição de soluções aproximadas para subproblemas proximais, que foi introduzida pela primeira vez por Solodov e Svaiter em [93].

Definição 2.1.1. *Sejam $(x, \lambda) \in \mathcal{H} \times \mathbb{R}_{++}$ e $\sigma \in [0, 1)$. Uma tripla $(y, v, \varepsilon) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}_+$ é uma σ -solução aproximada do sistema proximal (2.5) em (x, λ) se*

$$\begin{cases} v \in T^{[\varepsilon]}(y), \\ \|\lambda v + y - x\|^2 + 2\lambda\varepsilon \leq \sigma^2 (\|\lambda v\|^2 + \|y - x\|^2). \end{cases} \quad (2.6)$$

Observação 2.1.2. Se tomarmos $\sigma = 0$ em (2.6), então temos

$$\varepsilon = 0 \quad \text{e} \quad \lambda v + y - x = 0,$$

o que, à luz da Proposição 1.2.10(d), implica que

$$v \in T(y) \quad \text{e} \quad \lambda v + y - x = 0,$$

e assim, $y = (\lambda T + I)^{-1}(x)$. Em outras palavras, a definição anterior de solução aproximada (inexata) para subproblemas proximais fornece a solução exata quando $\sigma = 0$. Além disso, o erro tanto na inclusão (por meio do ε -enlargement de T) quanto na equação em (2.6) é “relativo” ao deslocamento λv e $y - x$.

A seguir, estabelecem-se algumas propriedades das soluções aproximadas definidas anteriormente. Começamos com o seguinte lema, que foi provado em [93, Lemma 2].

Lema 2.1.3. *Sejam $(x, \lambda) \in \mathcal{H} \times \mathbb{R}_{++}$ e $\sigma \in [0, 1)$. Uma tripla $(y, v, \varepsilon) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}_+$ é uma σ -solução aproximada do sistema proximal (2.5) em (x, λ) (no sentido da Definição 2.1.1) se, e somente se,*

$$\langle x - y, v \rangle - \varepsilon \geq \frac{1 - \sigma^2}{2\lambda} (\|\lambda v\|^2 + \|x - y\|^2).$$

Além disso, as seguintes afirmações são equivalentes

$$\begin{cases} v = 0, \\ x \in T^{-1}(0), \\ y = x, \end{cases}$$

e qualquer uma delas implica que $\varepsilon = 0$.

O lema a seguir será utilizado na análise (da convergência forte e da complexidade iterativa) do Algoritmo 1.

Lema 2.1.4. *Sejam $(x, \lambda) \in \mathcal{H} \times \mathbb{R}_{++}$ e $\sigma \in [0, 1)$. Se (y, v, ε) é uma σ -solução aproximada do sistema proximal (2.5) em (x, λ) , então*

$$\|P_H(x) - x\| \geq \frac{1 - \sigma^2}{2} \max\{\|\lambda v\|, \|x - y\|\}, \quad (2.7)$$

onde

$$H = \{z \in \mathcal{H} \mid \langle z - y, v \rangle \leq \varepsilon\}.$$

Demonstração. Consideraremos os seguintes casos:

Primeiro caso: Se $v = 0$. Nesse caso $H = \mathcal{H}$, e assim $P_H = I$ (operador identidade) o que implica que o lado esquerdo de (2.7) é zero. Por outro lado, do Lema 2.1.3 temos que o lado direito de (2.7) também é zero e assim a desigualdade em (2.7) é satisfeita.

Segundo caso: Se $v \neq 0$. Nesse caso, a projeção sobre o semi-espço H é dada por (ver, por exemplo, [24, Example 28.16])

$$P_H(x) = x - \frac{\langle x - y, v \rangle - \varepsilon}{\|v\|^2} v$$

e como $\langle x - y, v \rangle - \varepsilon \geq 0$ (do Lema 2.1.3), temos

$$\|P_H(x) - x\| = \frac{\langle x - y, v \rangle - \varepsilon}{\|v\|}. \quad (2.8)$$

Usando (novamente) o Lema 2.1.3, obtemos

$$\begin{aligned} \frac{\langle x - y, v \rangle - \varepsilon}{\|v\|} &\geq \frac{1 - \sigma^2}{2} \left(\lambda \|v\| + \frac{\|x - y\|^2}{\lambda \|v\|} \right) \\ &\geq \frac{1 - \sigma^2}{2} \lambda \|v\|. \end{aligned} \quad (2.9)$$

Por outro lado, usando as desigualdades $a + b \geq \sqrt{ab}$ com $a = \lambda \|v\|$ e $b = \|x - y\|^2 / (\lambda \|v\|)$, temos

$$\lambda \|v\| + \frac{\|x - y\|^2}{\lambda \|v\|} \geq \|x - y\|. \quad (2.10)$$

O resultado desejado segue da combinação de (2.8)–(2.10). $\square$

2.2 Um algoritmo ponto proximal inexato inercial fortemente convergente

Considere o problema de inclusão monótono (2.1), isto é,

$$0 \in T(x) \quad (2.11)$$

sob a suposição de que as seguintes hipóteses são satisfeitas:

- A1. $T: \mathcal{H} \rightrightarrows \mathcal{H}$ é um operador monótono maximal.
- A2. O conjunto solução $\mathcal{S} := T^{-1}(0)$ de (2.11) é não-vazio.

Para resolver numericamente o problema (2.11), propõe-se um algoritmo proximal inercial inexato com convergência forte, a saber, o Algoritmo 1. Demonstramos sua convergência forte (Teorema 2.2.6) e sua complexidade iterativa pontual (Teorema 2.2.7).

A seguir, apresentamos algumas características principais do Algoritmo 1.

Efeitos inerciais: Os efeitos inerciais sobre as iteradas geradas são induzidos por duas sequências de “controle”, a saber, $\{\alpha_k\}$ e $\{\beta_k\}$ — ver (2.12).

Crítério de erro relativo para os subproblemas: Adota-se um critério de erro relativo flexível para a resolução dos subproblemas proximais, o qual permite a presença de erros tanto na inclusão quanto na equação do sistema proximal — ver a Definição 2.1.1 e o sistema (2.13).

Projeção ortogonal sobre semi-espços: A próxima iterada x_{k+1} é definida como a projeção ortogonal sobre a interseção dos semi-espços H_k e W_k - ver (2.14) e (2.15).

Algoritmo 1: Um algoritmo PP inexato inercial para resolver (2.11).

Entrada: $x_0 = x_{-1} \in \mathcal{H}$ e $\sigma \in [0, 1)$.

para $k = 0, 1, 2, \dots$ **faça**

Escolha $\alpha_k, \beta_k \geq 0$ e defina

$$\begin{aligned} w_k &= x_k + \alpha_k(x_k - x_{k-1}), \\ \tilde{w}_k &= w_k + \beta_k(w_k - x_0). \end{aligned} \tag{2.12}$$

Escolha $\lambda_k > 0$ e determine uma σ -solução aproximada (no sentido da Definição 2.1.1) em $(\tilde{w}_k, \lambda_k)$, ou seja, calcule $(y_k, v_k, \varepsilon_k)$ tal que

$$\begin{cases} v_k \in T^{[\varepsilon_k]}(y_k), \\ \|\lambda_k v_k + y_k - \tilde{w}_k\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 (\|\lambda_k v_k\|^2 + \|y_k - \tilde{w}_k\|^2). \end{cases} \tag{2.13}$$

Defina

$$\begin{aligned} H_k &= \{z \in \mathcal{H} \mid \langle z - y_k, v_k \rangle \leq \varepsilon_k\}, \\ W_k &= \{z \in \mathcal{H} \mid \langle z - x_k, x_0 - x_k \rangle \leq 0\}. \end{aligned} \tag{2.14}$$

Atualize

$$x_{k+1} = P_{H_k \cap W_k}(x_0). \tag{2.15}$$

fim

A seguir, alguns comentários com respeito ao Algoritmo 1.

- (i) Os passos de extrapolação definidos em (2.12) introduzem efeitos inerciais nas iterações geradas pelo Algoritmo 1. Inicialmente, destacamos que w_k , definido pelo parâmetro inercial α_k , corresponde ao efeito de extrapolação típico, frequentemente utilizado em diversos algoritmos do tipo inercial (ver, por exemplo, [5, 8, 9, 18, 20, 16, 28]). Por outro lado, a extrapolação adicional representada por $\tilde{w}_k$ é especialmente projetada para o mecanismo de iteração específico do Algoritmo 1. As condições sobre as sequências $\{\alpha_k\}$ e $\{\beta_k\}$ que garantem a convergência (forte) e a análise de complexidade iterativa do Algoritmo 1 serão apresentadas na Hipótese 2.2.2 (ver também a Observação 2.2.3).

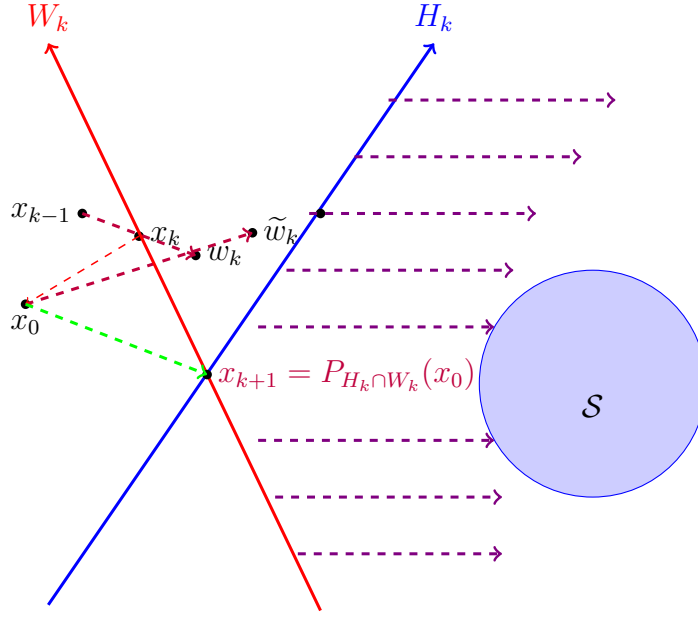


Figura 2.1: Interpretação geométrica dos dois passos de extrapolação (2.12) do Algoritmo 1, bem como da atualização por meio do passo projetivo em (2.15).

- (ii) O Algoritmo 1 constitui uma generalização, mais especificamente, uma versão inercial do [92, Algorithm 1], sendo que diversos resultados apresentados neste trabalho são inspirados naquela referência. A condição de erro relativo estabelecida em (2.13) é mais geral do que a correspondente em [92, Definition 1]; na verdade, a formulação dada em (2.13) foi introduzida pela primeira vez em [93, Definition 1], no contexto de uma versão distinta do algoritmo ponto proximal. Critérios de erro relativo análogos ao de (2.13), aplicados aos subproblemas do algoritmo PP e de suas variantes, têm sido utilizados em diferentes contextos por diversos autores (ver, por exemplo, [13, 67, 68, 69, 93] e as referências ali citadas).
- (iii) Note-se que, ao fixar $\sigma = 0$ em (2.13), obtemos $\varepsilon_k = 0$, $\lambda_k v_k + y_k - \tilde{w}_k = 0$ e $v_k \in T^{[0]}(y_k)$, o que equivale a dizer que $y_k = (\lambda_k T + I)^{-1}(\tilde{w}_k)$ e $v_k = \frac{1}{\lambda_k}(\tilde{w}_k - y_k)$. Recordamos que, como T é monótono maximal, tem-se $T^{[0]} = T$ (ver Proposição 1.2.10(d)). Assim, sob a suposição de que o operador resolvente $(\lambda T + I)^{-1}$ é de fácil avaliação, conclui-se que a tripla $(y_k, v_k, \varepsilon_k) := ((\lambda_k T + I)^{-1}(\tilde{w}_k), \frac{1}{\lambda_k}(\tilde{w}_k - y_k), 0)$ satisfaz, automaticamente, a condição de erro relativo dada em (2.13). No entanto, neste trabalho não assumimos tal hipótese. Pelo contrário, estamos interessados nos casos em que a avaliação do resolvente $(\lambda T + I)^{-1}$ apresenta dificuldade computacional. As estratégias para obter uma tripla $(y_k, v_k, \varepsilon_k)$ que satisfaça a condição (2.13), no caso geral com $\sigma \in [0, 1)$, dependerão da estrutura específica do operador $T(\cdot)$ definido em (2.11) — ver, por exemplo, as Seções 2.3 e 2.4 do presente capítulo.
- (iv) Os semi-espacos H_k e W_k , definidos conforme a (2.14), coincidem com aqueles utilizados em [92, Algorithm 1]. Esses conjuntos desempenham um papel fundamental na garantia da convergência das sequências geradas pelo Algoritmo 1. Enquanto a fronteira de H_k separa o conjunto solução S de $\tilde{w}_k$ (ver Lema 2.1.3, (2.13) e (2.14)), o conjunto W_k é introduzido para forçar a convergência forte. Conforme discutido em [92], a definição da iterada x_{k+1} dada por (2.15) depende do cálculo explícito da projeção ortogonal

sobre a interseção $H_k \cap W_k$, o que, essencialmente, se reduz (no máximo) à resolução de um sistema linear de dimensão 2×2 (ver [92, pp. 195-196] e [24, Proposition 29.22 e Proposition 29.23]). A boa definição do Algoritmo 1 no que se refere à equação (2.15) será discutida no Apêndice A.

- (v) O Algoritmo 1 pode ser empregado tanto como um método prático quando se dispõe de um algoritmo interno capaz de calcular uma tripla $(y_k, v_k, \varepsilon_k)$ que satisfaça a condição (2.13) quanto como uma estrutura geral para o desenvolvimento e análise de esquemas mais específicos. Esta última abordagem será explorada com maior profundidade nas Seções 2.3 e 2.4, com o objetivo de obter versões inerciais fortemente convergentes dos métodos de Korpelevich, forward-backward modificado de Tseng e forward-backward clássico, aplicados a desigualdades variacionais e inclusões monótonas estruturadas.

2.2.1 Análise de convergência do Algoritmo 1

Nesta subseção, estabelecemos os principais resultados de convergência forte do Algoritmo 1 sob as hipóteses apropriadas sobre os parâmetros inerciais. Iniciamos enunciando e demonstrando alguns resultados auxiliares (Proposição 2.2.1, Proposição 2.2.4 e Corolário 2.2.5), os quais serão necessários para a prova do teorema principal (Teorema 2.2.6). Esses resultados mostram as propriedades das sequências geradas pelo Algoritmo 1.

Proposição 2.2.1. *Considere as sequências geradas pelo Algoritmo 1 e seja d_0 a distância entre o ponto inicial x_0 e o conjunto solução $S := T^{-1}(0) \neq \emptyset$ do problema (2.11). Então valem as seguintes afirmações*

- (a) Para todo $k \geq 0$, tem-se

$$\|x_{k+1} - x_0\|^2 \geq \|x_k - x_0\|^2 + \|x_{k+1} - x_k\|^2.$$

- (b) Para todo $k \geq 0$, tem-se

$$\|x_k - x_0\| \leq d_0.$$

- (c) Para todo $k \geq 0$, vale a estimativa

$$\|x_{k+1} - \tilde{w}_k\| \geq \frac{1 - \sigma^2}{2} \max \{ \|\lambda_k v_k\|, \|\tilde{w}_k - y_k\| \}. \quad (2.16)$$

Demonstração. (a) Pela definição de W_k , tem-se que $x_k = P_{W_k}(x_0)$. Aplicando a desigualdade (1.3) com $C = W_k$, $x = x_{k+1}$ e $y = x_0$, e observando que $x_{k+1} = P_{W_k}(x_{k+1})$ (pois $x_{k+1} \in W_k$), obtemos

$$\|P_{W_k}(x_{k+1}) - P_{W_k}(x_0)\|^2 \leq \|x_{k+1} - x_0\|^2 - \|[x_{k+1} - P_{W_k}(x_{k+1})] - [x_0 - P_{W_k}(x_0)]\|^2.$$

Então

$$\|x_{k+1} - x_k\|^2 \leq \|x_{k+1} - x_0\|^2 - \|x_k - x_0\|^2,$$

o que é exatamente a desigualdade afirmada no item (a).

- (b) De (2.15) e (A.2), para todo $k \geq 0$, temos

$$\|x_{k+1} - x_0\| = \|P_{H_k \cap W_k}(x_0) - x_0\| \leq \|P_S(x_0) - x_0\| = d_0,$$

onde $P_{\mathcal{S}}$ denota a projeção ortogonal sobre o conjunto solução $\mathcal{S} = T^{-1}(0)$. Se $k = 0$, a desigualdade é trivialmente satisfeita.

(c) Como $x_{k+1} \in H_k$, pela definição de $P_{H_k}(\tilde{w}_k)$ obtemos:

$$\|x_{k+1} - \tilde{w}_k\| \geq \|P_{H_k}(\tilde{w}_k) - \tilde{w}_k\|.$$

Por outro lado, usando o Lema 2.1.4 e o fato de que $(y_k, v_k, \varepsilon_k)$ é uma σ -solução aproximada em $(\tilde{w}_k, \lambda_k)$ (ver (2.13)) segue que:

$$\|P_{H_k}(\tilde{w}_k) - \tilde{w}_k\| \geq \frac{1 - \sigma^2}{2} \max \{ \|\lambda_k v_k\|, \|\tilde{w}_k - y_k\| \},$$

o que conclui a demonstração da proposição. $\square$

A partir deste ponto, nesta tese, assumiremos que as seguintes condições são satisfeitas pelas sequências $\{\alpha_k\}$ e $\{\beta_k\}$ definidas no Algoritmo 1.

Hipótese 2.2.2. *As sequências de parâmetros $\{\alpha_k\}$ e $\{\beta_k\}$ satisfazem as seguintes condições:*

(a) *A sequência $\{\alpha_k\}$ é limitada, isto é,*

$$\bar{\alpha} := \sup_{k \geq 0} \alpha_k < \infty.$$

(b) *A sequência $\{\beta_k\}$ pertence a ℓ^2 , isto é,*

$$\bar{s} := \sum_{k=0}^{\infty} \beta_k^2 < \infty.$$

Também usaremos a notação

$$\bar{\beta} := \sup_{k \geq 0} \beta_k < \infty.$$

Observação 2.2.3. Ressaltamos que, mesmo no caso em que $\beta_k = 0$ para todo $k \geq 0$, isto é, quando $\bar{s} = \bar{\beta} = 0$, o efeito inercial induzido pela sequência $\{\alpha_k\}$ no Algoritmo 1 permanece compatível com o tipo de extrapolação usualmente introduzido na literatura recente sobre algoritmos com inércia. Além disso, a Hipótese 2.2.2(a), que requer apenas a limitação da sequência $\{\alpha_k\}$, constitui uma suposição consideravelmente mais flexível em comparação com outros trabalhos envolvendo efeitos inerciais. Por exemplo, no artigo seminal [5], é imposto que $\bar{\alpha} < 1/3$. Veja também [9, Section 2] para uma discussão adicional a esse respeito. Essa maior generalidade torna os resultados desenvolvidos neste trabalho relevantes mesmo no contexto de espaços de dimensão finita, onde, naturalmente, as noções de convergência fraca e forte coincidem.

Proposição 2.2.4. *Considere as sequências geradas pelo Algoritmo 1, sejam d_0 a distância de x_0 ao conjunto solução $\mathcal{S} := T^{-1}(0) \neq \emptyset$ de (2.11) e $\bar{\alpha}, \bar{\beta}, \bar{s}$ como na Hipótese 2.2.2. Então*

(a) *Tem-se*

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 \leq d_0^2.$$

(b) *Tem-se*

$$\sum_{k=0}^{\infty} \|x_{k+1} - \tilde{w}_k\|^2 \leq (1 + \bar{\alpha}) \left[(1 + \bar{\beta}) [1 + \bar{\alpha}(1 + \bar{\beta})] + \bar{s} \right] d_0^2.$$

(c) *Tem-se*

$$x_k - y_k \rightarrow 0.$$

Demonstração. (a) O resultado decorre diretamente dos itens (a) e (b) da Proposição 2.2.1, combinados com um argumento baseado em soma telescópica.

(b) Afiramos que

$$\|w_k - z\|^2 = (1 + \alpha_k) \|x_k - z\|^2 + \alpha_k (1 + \alpha_k) \|x_k - x_{k-1}\|^2 - \alpha_k \|x_{k-1} - z\|^2 \quad \forall z \in \mathcal{H}. \quad (2.17)$$

De fato, a partir da definição de w_k em (2.12), temos

$$x_k - z = \frac{1}{1 + \alpha_k} (w_k - z) + \frac{\alpha_k}{1 + \alpha_k} (x_{k-1} - z)$$

e $w_k - x_{k-1} = (1 + \alpha_k)(x_k - x_{k-1})$, o que, combinado com o Lema A.2.1(a) do Apêndice A e alguns simples passos algébricos, resulta em (2.17).

Observe agora que, utilizando o Lema A.2.1(b) e (2.12), obtemos

$$\begin{aligned} \|x_{k+1} - \tilde{w}_k\|^2 &= \|x_{k+1} - w_k - \beta_k(w_k - x_0)\|^2 \\ &= \|x_{k+1} - w_k\|^2 + \beta_k^2 \|w_k - x_0\|^2 + 2\beta_k \langle w_k - x_{k+1}, w_k - x_0 \rangle \\ &= (1 + \beta_k) \|w_k - x_{k+1}\|^2 + \beta_k(1 + \beta_k) \|w_k - x_0\|^2 - \beta_k \|x_{k+1} - x_0\|^2. \end{aligned}$$

Aplicando (2.17) (com $z = x_{k+1}$ e $z = x_0$) na identidade acima, obtemos

$$\begin{aligned} \|x_{k+1} - \tilde{w}_k\|^2 &= (1 + \beta_k) \left[(1 + \alpha_k) \|x_k - x_{k+1}\|^2 + \alpha_k (1 + \alpha_k) \|x_k - x_{k-1}\|^2 - \alpha_k \|x_{k-1} - x_{k+1}\|^2 \right] \\ &\quad + \beta_k(1 + \beta_k) \left[(1 + \alpha_k) \|x_k - x_0\|^2 + \alpha_k (1 + \alpha_k) \|x_k - x_{k-1}\|^2 - \alpha_k \|x_{k-1} - x_0\|^2 \right] \\ &\quad - \beta_k \|x_{k+1} - x_0\|^2 \\ &= (1 + \beta_k)(1 + \alpha_k) \left[\|x_k - x_{k+1}\|^2 + \alpha_k (1 + \beta_k) \|x_k - x_{k-1}\|^2 + \beta_k \|x_k - x_0\|^2 \right] \\ &\quad - \left[\alpha_k(1 + \beta_k) (\|x_{k-1} - x_{k+1}\|^2 + \beta_k \|x_{k-1} - x_0\|^2) + \beta_k \|x_{k+1} - x_0\|^2 \right]. \quad (2.18) \end{aligned}$$

Por outro lado, como $\|x_{k+1} - x_{k-1}\| + \|x_{k-1} - x_0\| \geq \|x_{k+1} - x_0\|$, usando o Lema A.2.2 do Apêndice A, obtemos

$$\|x_{k+1} - x_{k-1}\|^2 + \beta_k \|x_{k-1} - x_0\|^2 \geq \frac{\beta_k}{1 + \beta_k} \|x_{k+1} - x_0\|^2,$$

o que implica

$$\alpha_k(1 + \beta_k) (\|x_{k+1} - x_{k-1}\|^2 + \beta_k \|x_{k-1} - x_0\|^2) + \beta_k \|x_{k+1} - x_0\|^2 \geq \beta_k(1 + \alpha_k) \|x_{k+1} - x_0\|^2. \quad (2.19)$$

Combinando (2.18) e (2.19), e usando o fato de que $\|x_k - x_0\| \leq \|x_{k+1} - x_0\|$ (ver Proposição 2.2.1(a)), temos

$$\|x_{k+1} - \tilde{w}_k\|^2 \leq (1 + \alpha_k) \left[(1 + \beta_k) [\|x_{k+1} - x_k\|^2 + \alpha_k(1 + \beta_k)\|x_k - x_{k-1}\|^2] + \beta_k^2 \|x_{k+1} - x_0\|^2 \right],$$

o que, ao ser combinado com o item (a), Proposição 2.2.1(b) e a Hipótese 2.2.2, nos fornece o resultado desejado.

(c) Observe primeiramente que

$$\begin{aligned} \|x_k - y_k\| &\leq \|x_k - \tilde{w}_k\| + \|\tilde{w}_k - y_k\| \\ &\leq \|x_k - x_{k+1}\| + \|x_{k+1} - \tilde{w}_k\| + \|\tilde{w}_k - y_k\|. \end{aligned}$$

O resultado desejado decorre agora dos itens (a) e (b) e da Proposição 2.2.1(c). $\square$

Corolário 2.2.5. *Seja $\{x_k\}$ a sequência gerada pelo Algoritmo 1 e suponha que existe um $\underline{\lambda} > 0$ tal que*

$$\lambda_k \geq \underline{\lambda} \quad \text{para todo } k \geq 0. \quad (2.20)$$

Então todo ponto de acumulação fraco da sequência $\{x_k\}$ pertence ao conjunto $\mathcal{S} := T^{-1}(0) \neq \emptyset$, ou seja, é solução do problema (2.11).

Demonstração. Inicialmente, observamos que a Proposição 2.2.1(b) garante que a sequência $\{x_k\}$ é limitada. Por sua vez, da Proposição 2.2.4(b), temos que

$$x_{k+1} - \tilde{w}_k \rightarrow 0, \quad (2.21)$$

o que combinado com (2.16), implica que

$$\lambda_k v_k \rightarrow 0 \quad \text{e} \quad \tilde{w}_k - y_k \rightarrow 0.$$

Além disso, pela desigualdade em (2.13) e a hipótese de que $\lambda_k \geq \underline{\lambda} > 0$ para todo $k \geq 0$, concluímos que

$$v_k \rightarrow 0 \quad \text{e} \quad \varepsilon_k \rightarrow 0. \quad (2.22)$$

Seja agora $\{x_{k_j}\}$ uma subsequência de $\{x_k\}$ tal que $x_{k_j} \rightharpoonup \bar{x}$. Aplicando a Proposição 2.2.4(c), obtemos

$$y_{k_j} \rightharpoonup \bar{x}.$$

Por outro lado, de (2.22), temos que $v_{k_j} \rightarrow 0$ e $\varepsilon_{k_j} \rightarrow 0$. Como $v_{k_j} \in T^{[\varepsilon_{k_j}]}(y_{k_j})$, $v_{k_j} \rightarrow 0$, $y_{k_j} \rightharpoonup \bar{x}$ e $\varepsilon_{k_j} \rightarrow 0$, segue da Proposição 1.2.10(e) que

$$0 \in T(\bar{x}), \quad \text{isto é, } \bar{x} \in \mathcal{S}.$$

Portanto, todos os pontos de acumulação fraco da sequência $\{x_k\}$ pertencem ao conjunto solução $\mathcal{S}$, o que conclui a demonstração do corolário. $\square$

Com todos os resultados anteriores, estamos prontos para enunciar o principal resultado deste capítulo: a convergência forte de qualquer sequência $\{x_k\}$ gerada pelo Algoritmo 1.

Teorema 2.2.6 (Convergência forte do Algoritmo 1). *Sejam $\{x_k\}$, $\{y_k\}$ e $\{\lambda_k\}$ as sequências geradas pelo Algoritmo 1, e suponha que $\{\lambda_k\}$ é limitada inferiormente por uma constante positiva, ou seja, que a condição (2.20) é satisfeita. Assuma também que a Hipótese 2.2.2 é válida para as sequências $\{\alpha_k\}$ e $\{\beta_k\}$. Então, $\{x_k\}$ e $\{y_k\}$ convergem fortemente para $x^* := P_S(x_0)$, onde $S := T^{-1}(0) \neq \emptyset$ é o conjunto solução de (2.11).*

Demonstração. Da Proposição 2.2.1(b) e do fato de que $d_0 = \|x^* - x_0\|$, temos que

$$\|x_k - x_0\| \leq \|x^* - x_0\| \quad \text{para todo } k \geq 0. \quad (2.23)$$

Logo, a sequência $\{x_k\}$ é limitada e, pelo Corolário 2.2.5, seus pontos de acumulação fraco pertencem a S . Seja $\{x_{k_j}\}$ uma subsequência de $\{x_k\}$ e $\bar{x} \in S$ tal que

$$x_{k_j} \rightharpoonup \bar{x}.$$

Como a norma $\|\cdot\|$ é semi-contínua inferiormente na topologia fraca, obtemos

$$\|\bar{x} - x_0\| \leq \liminf \|x_{k_j} - x_0\| \leq \limsup \|x_{k_j} - x_0\| \leq \|x^* - x_0\|, \quad (2.24)$$

onde a última desigualdade decorre de (2.23). Como $x^* = P_S(x_0)$ e $\bar{x} \in S$, segue diretamente de (2.24) que $\bar{x} = x^*$. Assim, $\{x_k\}$ é uma sequência limitada com um único ponto de acumulação fraco x^* , logo do Lema A.2.3 temos que $x_k \rightharpoonup x^*$ e, por (2.24),

$$\|x^* - x_0\| \leq \liminf \|x_k - x_0\| \leq \limsup \|x_k - x_0\| \leq \|x^* - x_0\|,$$

o que implica que $\|x_k - x_0\| \rightarrow \|x^* - x_0\|$. Como $x_k - x_0 \rightharpoonup x^* - x_0$, pelo Lema A.2.1(c), obtemos $x_k - x_0 \rightarrow x^* - x_0$, o que é equivalente a $x_k \rightarrow x^*$. O fato que $y_k \rightarrow x^*$ segue diretamente do fato de que $x_k \rightarrow x^*$ e da Proposição 2.2.4(c). $\square$

2.2.2 Análise da complexidade iterativa

Nesta subseção, realizamos uma análise da complexidade iterativa, no sentido pontual, do Algoritmo 1. O resultado central encontra-se no Teorema 2.2.7, apresentado a seguir. Inspirados pelas análises do método HPE desenvolvidas em [67], propomos estudar o desempenho do Algoritmo 1 a partir da seguinte perspectiva: dado um parâmetro de tolerância $\rho > 0$, nosso objetivo é determinar um número de iterações necessárias para obter uma tripla $(y, v, \varepsilon) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}_+$ tal que

$$v \in T^{[\varepsilon]}(y) \quad \text{e} \quad \max\{\|v\|, \varepsilon\} \leq \rho. \quad (2.25)$$

Nesse contexto, dizemos que y constitui uma solução ρ -aproximada do problema (2.11). Vale observar que, fazendo $\rho = 0$ em (2.25), obtemos $v = 0$ e $\varepsilon = 0$, o que implica que $0 \in T^{[0]}(y)$, e portanto $0 \in T(y)$, ou seja, y é uma solução exata de (2.11).

Na sequência, apresentamos o principal resultado teórico sobre a complexidade iterativa do Algoritmo 1.

Teorema 2.2.7 (Complexidade iterativa pontual do Algoritmo 1). *Considere as sequências geradas pelo Algoritmo 1 e suponha que $\{\lambda_k\}$ seja limitada inferiormente por uma constante positiva, ou seja, que a condição (2.20) é satisfeita. Seja d_0 a distância de x_0 ao conjunto*

solução $\mathcal{S} := T^{-1}(0) \neq \emptyset$ de (2.11), e suponha que a Hipótese 2.2.2 seja válida. Então, para todo $k \geq 0$, existe um índice $j \in \{0, \dots, k\}$ tal que

$$v_j \in T^{[\varepsilon_j]}(y_j), \quad (2.26)$$

$$\|v_j\| \leq \frac{2d_0}{\sqrt{k+1}} \left(\frac{\sqrt{(1+\bar{\alpha})[(1+\bar{\beta})[1+\bar{\alpha}(1+\bar{\beta})] + \bar{s}]}}{(1-\sigma^2)\underline{\lambda}} \right), \quad (2.27)$$

$$\varepsilon_j \leq \frac{d_0^2}{k+1} \left(\frac{4\sigma^2(1+\bar{\alpha})[(1+\bar{\beta})[1+\bar{\alpha}(1+\bar{\beta})] + \bar{s}]}{(1-\sigma^2)^2\underline{\lambda}} \right). \quad (2.28)$$

Demonstração. A partir da Proposição 2.2.1(c) e da desigualdade apresentada em (2.13), obtemos, para todo $i \in \{0, \dots, k\}$,

$$\max \left\{ \|\lambda_i v_i\|^2, \frac{\lambda_i \varepsilon_i}{\sigma^2} \right\} \leq \frac{4}{(1-\sigma^2)^2} \|x_{i+1} - \tilde{w}_i\|^2.$$

Com isso, utilizando a Proposição 2.2.4(b), segue que

$$\sum_{i=0}^k \max \left\{ \|\lambda_i v_i\|^2, \frac{\lambda_i \varepsilon_i}{\sigma^2} \right\} \leq \frac{4(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}{(1-\sigma^2)^2} d_0^2.$$

Portanto, existe algum índice $j \in \{0, \dots, k\}$ tal que

$$(k+1) \max \left\{ \|\lambda_j v_j\|^2, \frac{\lambda_j \varepsilon_j}{\sigma^2} \right\} \leq \frac{4(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}{(1-\sigma^2)^2} d_0^2.$$

Combinando esta estimativa com a condição (2.20), obtemos diretamente as desigualdades (2.27) e (2.28). Por fim, vale observar que a inclusão (2.26) decorre diretamente da inclusão em (2.13), o que conclui a demonstração do teorema. $\square$

Apresentamos a seguir alguns comentários a respeito do Teorema 2.2.7.

- (i) Suponha que escolhemos $\beta_k \equiv 0$ no Algoritmo 1, conforme mencionado na Observação 2.2.3. Nesse caso, temos que $\bar{\beta} = \bar{s} = 0$ (ver Hipótese 2.2.2), e as expressões em (2.26)–(2.28) se simplificam para:

$$v_j \in T^{[\varepsilon_j]}(y_j), \quad \|v_j\| \leq \frac{2(1+\bar{\alpha})d_0}{\sqrt{k+1}(1-\sigma^2)\underline{\lambda}}, \quad \varepsilon_j \leq \frac{4\sigma^2(1+\bar{\alpha})^2 d_0^2}{(k+1)(1-\sigma^2)^2 \underline{\lambda}}. \quad (2.29)$$

Gostaríamos de destacar que as estimativas obtidas para v_j e ε_j em (2.29) são consideravelmente mais simples do que aquelas apresentadas em [9, Theorem 2.7], onde se investigou a complexidade iterativa de métodos do tipo ponto proximal inercial com erro relativo para subproblemas proximais. Essa simplificação foi possível, essencialmente, à custa de se computar uma projeção extra sobre a interseção de dois semi-espacos (ver definição de x_{k+1} em (2.15)), a qual, como mencionado na quarta observação após o Algoritmo 1, equivale a resolver no máximo um sistema linear de dimensão 2×2 . Além disso, como observado também na Observação 2.2.3, a hipótese de que a sequência $\{\alpha_k\}$ é limitada (ver Hipótese 2.2.2) é consideravelmente mais fraca do que as normalmente adotadas em análises de algoritmos ponto proximal com inércia (em particular, no trabalho [9]).

- (ii) Para facilitar a exposição, consideremos o caso simplificado dado por (2.29). Nesse cenário, o Teorema 2.2.7 assegura que, para uma tolerância prescrita $\rho > 0$, o Algoritmo 1 encontra uma tripla (y, v, ε) satisfazendo a condição (2.25) em, no máximo,

$$O \left(\left\lceil \left(\frac{1 + \bar{\alpha}}{1 - \sigma^2} \right)^2 \max \left\{ \left(\frac{d_0}{\underline{\lambda}\rho} \right)^2, \frac{\sigma^2 d_0^2}{\underline{\lambda}\rho} \right\} \right\rceil \right) \quad (2.30)$$

iterações, onde $\lceil m \rceil$ representa o menor inteiro maior ou igual a m .

2.3 Uma variante inercial fortemente convergente do método extragradiente de Korpelevich

Nesta seção, considere o problema de desigualdade variacional (PDV) dado em (2.4), isto é, o problema de encontrar $x \in C$ tal que

$$\langle F(x), y - x \rangle \geq 0 \quad \text{para todo } y \in C, \quad (2.31)$$

sob as seguintes hipóteses:

- B1. $C \subseteq \text{Dom } F$ é um subconjunto não-vazio, fechado e convexo de $\mathcal{H}$.
- B2. $F: \text{Dom } F \subseteq \mathcal{H} \rightarrow \mathcal{H}$ é monótono e L -Lipschitz contínuo, isto é, F é monótono em $\text{Dom } F$ e existe $L > 0$ tal que

$$\|F(x) - F(y)\| \leq L\|x - y\| \quad \text{para todo } x, y \in \text{Dom } F. \quad (2.32)$$

- B3. O conjunto solução $\mathcal{S} := (F + N_C)^{-1}(0)$ de (2.31) é não-vazio.

Vale ressaltar que, sob as hipóteses acima sobre C e F , o operador $F + N_C$ é monótono maximal [45, Proposition 12.3.6]. Além disso, conforme discutido no Exemplo 1.3.2 da Seção 1.3, o problema (2.31) é equivalente à inclusão monótona $0 \in T(x)$, onde $T = F + N_C$; essa equivalência permite aplicar os resultados da Seção 2.2 ao problema (2.31).

Para resolver numericamente (2.31), propomos e estudamos uma variante inercial fortemente convergente do consagrado método extragradiente de Korpelevich [53] (ver Algoritmo 2). Os principais resultados sobre convergência forte e complexidade iterativa estão resumidos no Teorema 2.3.2 mais adiante.

Algoritmo 2: Uma variante inercial fortemente convergente do método do extragradiente de Korpelevich para resolver (2.31).

Entrada: $x_0 = x_{-1} \in \mathcal{H}$, $\sigma \in (0, 1)$ e $\lambda = \sigma/L$.

para $k = 0, 1, 2, \dots$ **faça**

Escolha $\alpha_k, \beta_k \geq 0$ e defina

$$\begin{aligned} w_k &= x_k + \alpha_k(x_k - x_{k-1}), \\ \tilde{w}_k &= w_k + \beta_k(w_k - x_0). \end{aligned} \tag{2.33}$$

Seja $w'_k = P_C(\tilde{w}_k)$ e em seguida compute

$$\begin{aligned} y_k &= P_C(\tilde{w}_k - \lambda F(w'_k)), \\ \tilde{y}_k &= P_C(\tilde{w}_k - \lambda F(y_k)). \end{aligned} \tag{2.34}$$

Defina

$$\begin{aligned} H_k &= \{z \in \mathcal{H} \mid \langle z - y_k, v_k \rangle \leq \varepsilon_k\}, \\ W_k &= \{z \in \mathcal{H} \mid \langle z - x_k, x_0 - x_k \rangle \leq 0\}, \end{aligned} \tag{2.35}$$

onde

$$\begin{aligned} q_k &= \frac{\tilde{w}_k - \tilde{y}_k}{\lambda} - F(y_k), \\ v_k &= F(y_k) + q_k, \\ \varepsilon_k &= \langle q_k, \tilde{y}_k - y_k \rangle. \end{aligned} \tag{2.36}$$

Atualize a próxima iteração por

$$x_{k+1} = P_{H_k \cap W_k}(x_0). \tag{2.37}$$

fim

Algumas observações sobre o Algoritmo 2.

- (i) Os parâmetros inerciais α_k e β_k dados em (2.33) desempenham o mesmo papel que no Algoritmo 1 — veja o primeiro comentário após o Algoritmo 1. De modo análogo ao que foi feito anteriormente, para analisar a convergência forte e a complexidade iterativa do Algoritmo 2, assumiremos as mesmas condições sobre as sequências $\{\alpha_k\}$ e $\{\beta_k\}$ descritas na Hipótese 2.2.2.
- (ii) Um passo fundamental do Algoritmo 2 é a realização das projeções descritas em (2.34). Esse procedimento segue a mesma ideia do clássico método extragradiente de Korpelevich: o ponto y_k representa uma etapa de “gradiente projetado” partindo de $\tilde{w}_k$ na direção de $-F(w'_k)$, enquanto $\tilde{y}_k$ constitui uma etapa de correção (extragradiente). A projeção extra usada para calcular $w'_k = P_C(\tilde{w}_k)$ tem como objetivo garantir a viabilidade de $\tilde{w}_k$ em relação ao conjunto C , o que é essencial, já que, caso $\tilde{w}_k \notin \text{Dom } F$, a avaliação de

$F(\cdot)$ não seria possível. Vale mencionar que, se o operador F estiver definido em todo o espaço $\mathcal{H}$, essa projeção adicional torna-se desnecessária, bastando tomar $w'_k = \tilde{w}_k$.

- (iii) Os conjuntos afins H_k e W_k , definidos em (2.35), seguem a mesma estrutura apresentada no Algoritmo 1 — veja (2.14). Além disso, o cálculo das quantidades q_k , v_k e ε_k dadas em (2.36) não requer avaliações adicionais do operador F . Já a obtenção de x_{k+1} conforme (2.37) foi discutida no quarto comentário após o Algoritmo 1.
- (iv) Como o Algoritmo 2 é um caso particular do Algoritmo 1 (ver a Proposição 2.3.1 mais adiante), segue do Corolário A.1.2 no Apêndice A que o Algoritmo 2 está bem definido. Ver também o quarto comentário após o Algoritmo 1 e a Proposição 2.3.1.

A demonstração da próxima proposição segue essencialmente a mesma linha de raciocínio de [67, Theorem 5.1]. Para fins de completude, incluímos a demonstração a seguir.

Proposição 2.3.1. *Para as sequências geradas pelo Algoritmo 2, vale, para todo $k \geq 0$:*

$$\begin{cases} v_k \in (F + N_C^{[\varepsilon_k]})(y_k) \subseteq (F + N_C)^{[\varepsilon_k]}(y_k), \\ \|\lambda v_k + y_k - \tilde{w}_k\|^2 + 2\lambda\varepsilon_k \leq \sigma^2\|y_k - \tilde{w}_k\|^2. \end{cases} \quad (2.38)$$

Como consequência, concluímos que o Algoritmo 2 é um caso particular do Algoritmo 1 (com $\lambda_k \equiv \lambda$) para resolver o problema (2.11) com $T = F + N_C$.

Demonstração. O uso direto da segunda identidade em (2.34), juntamente com o fato de que $P_C = (\lambda N_C + I)^{-1}$, implica que $q_k \in N_C(\tilde{y}_k) = \partial\delta_C(\tilde{y}_k)$, onde q_k é definido como em (2.36). Assim, pela Proposição 1.2.10(g) e pelo fato de que $\partial_{\varepsilon_k}\delta_C = N_C^{[\varepsilon_k]}$, obtemos

$$q_k \in \partial_{\varepsilon_k}\delta_C(y_k) = N_C^{[\varepsilon_k]}(y_k)$$

onde ε_k é como em (2.36). A partir dessa inclusão e da definição de v_k dada em (2.36), obtemos a primeira inclusão em (2.38). A segunda inclusão decorre da primeira e da Proposição 1.2.10(b). Vamos agora demonstrar a desigualdade em (2.38). Observe primeiro que, pela primeira identidade em (2.34),

$$p_k := \frac{\tilde{w}_k - y_k}{\lambda} - F(w'_k) \in N_C(y_k), \quad (2.39)$$

o que, combinado com o fato de que $\tilde{y}_k \in C$ e com a definição de $N_C(y_k)$, implica que $\langle p_k, \tilde{y}_k - y_k \rangle \leq 0$. Assim,

$$\varepsilon_k = \langle q_k, \tilde{y}_k - y_k \rangle = \langle q_k - p_k, \tilde{y}_k - y_k \rangle + \langle p_k, \tilde{y}_k - y_k \rangle \leq \langle q_k - p_k, \tilde{y}_k - y_k \rangle \quad (2.40)$$

e, portanto,

$$\begin{aligned} \|\lambda v_k + y_k - \tilde{w}_k\|^2 + 2\lambda\varepsilon_k &= \|y_k - \tilde{y}_k\|^2 + 2\lambda\varepsilon_k && [\text{por (2.36)}] \\ &\leq \|y_k - \tilde{y}_k\|^2 + 2\lambda\langle q_k - p_k, \tilde{y}_k - y_k \rangle && [\text{por (2.40)}] \\ &= \|y_k - \tilde{y}_k - \lambda(q_k - p_k)\|^2 - \|\lambda(q_k - p_k)\|^2 \\ &\leq \|y_k - \tilde{y}_k - \lambda(q_k - p_k)\|^2 \\ &= \|\lambda(F(y_k) - F(w'_k))\|^2 && [(2.36) \text{ e } (2.39)] \\ &\leq (\lambda L)^2\|y_k - w'_k\|^2 && [\text{por (2.32)}] \\ &\leq (\lambda L)^2\|y_k - \tilde{w}_k\|^2 \\ &= \sigma^2\|y_k - \tilde{w}_k\|^2 && [\lambda = \sigma/L]. \end{aligned}$$

Por fim, a última afirmação da proposição decorre diretamente de (2.38) e das definições dos Algoritmos 1 e 2. $\square$

A seguir, resumimos os principais resultados obtidos para o Algoritmo 2.

Teorema 2.3.2 (Convergência forte e complexidade iterativa do Algoritmo 2). *Suponha que as hipóteses B1, B2 e B3 mencionadas anteriormente sejam válidas sobre o conjunto C e o operador $F(\cdot)$, e considere as sequências geradas pelo Algoritmo 2. Seja d_0 a distância de x_0 ao conjunto solução $S \neq \emptyset$ de (2.3) (ou equivalentemente (2.31)), e suponha que a Hipótese 2.2.2 seja satisfeita pelas sequências $\{\alpha_k\}$ e $\{\beta_k\}$. Então as seguintes afirmações são válidas:*

- (a) *As sequências $\{x_k\}$ e $\{y_k\}$ convergem fortemente para $x^* := P_S(x_0)$.*
- (b) *Para todo $k \geq 0$, existe $j \in \{0, \dots, k\}$ tal que*

$$\begin{cases} v_j \in (F + N_C^{[\varepsilon_j]})(y_j), & (2.41) \\ \|v_j\| \leq \frac{2d_0L}{\sqrt{k+1}} \left(\frac{\sqrt{(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}}{(1-\sigma^2)\sigma} \right), & (2.42) \\ \varepsilon_j \leq \frac{d_0^2L}{k+1} \left(\frac{4\sigma(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}{(1-\sigma^2)^2} \right). & (2.43) \end{cases}$$

Demonstração. Pela Proposição 2.3.1, sabemos que o Algoritmo 2 é um caso particular do Algoritmo 1 para resolver o problema (2.11) com $T = F + N_C$. Assim, basta aplicar os Teoremas 2.2.6 e 2.2.7, utilizar o fato de que $\lambda_k \equiv \lambda = \sigma/L$ (ver a entrada do Algoritmo 2) e lembrar que o problema de desigualdade variacional (2.31) é equivalente à inclusão $0 \in (F + N_C)(x)$. $\square$

Observação 2.3.3. As estimativas apresentadas no item (b) acima para $\|v_j\|$ e ε_j são análogas às obtidas por Monteiro e Svaiter em [67, Theorem 5.2(a)] no contexto do método extragradiente de Korpelevich (ver também [73]). De forma similar à primeira observação feita após o Teorema 2.2.7, se considerarmos $\beta_k \equiv 0$ no Algoritmo 2, então as expressões (2.41) – (2.43) se reduzem a

$$v_j \in (F + N_C^{[\varepsilon_j]})(y_j), \quad \|v_j\| \leq \frac{2(1+\bar{\alpha})d_0L}{\sqrt{k+1}(1-\sigma^2)\sigma}, \quad \varepsilon_j \leq \frac{4\sigma(1+\bar{\alpha})^2d_0^2L}{(k+1)(1-\sigma^2)^2}.$$

Essas estimativas implicam que, dada uma tolerância $\rho > 0$, o Algoritmo 2 encontra uma tripla (y, v, ε) tal que $v \in (F + N_C^{[\varepsilon]})(y)$ e $\max\{\|v\|, \varepsilon\} \leq \rho$ após no máximo

$$O \left(\left\lceil \left(\frac{1+\bar{\alpha}}{1-\sigma^2} \right)^2 \max \left\{ \left(\frac{d_0L}{\sigma\rho} \right)^2, \frac{d_0^2L\sigma}{\rho} \right\} \right\rceil \right) \quad (2.44)$$

iterações. Além disso, vale observar que a inclusão (2.41) está fortemente relacionada às definições clássicas de *soluções fracas e fortes* para problemas de desigualdade variacional (ver, por exemplo, [67] para uma discussão sobre o tema).

2.4 Versões inerciais fortemente convergentes dos métodos forward-backward modificado de Tseng e forward-backward clássico

Neste seção, considere o problema de inclusão monótono estruturado descrito em (2.2), ou seja, buscamos $x \in \mathcal{H}$ tal que

$$0 \in F(x) + B(x) \quad (2.45)$$

sob as seguintes hipóteses:

- C1. $F: \text{Dom } F \subseteq \mathcal{H} \rightarrow \mathcal{H}$ é um operador (ponto-ponto) monótono e contínuo.
- C2. $B: \mathcal{H} \rightrightarrows \mathcal{H}$ é um operador monótono maximal (ponto-conjunto), com $\text{Dom } B \subseteq C \subseteq \text{Dom } F$, onde C é um subconjunto não-vazio, fechado e convexo de $\mathcal{H}$.
- C3. O conjunto solução $\mathcal{S} := (F + B)^{-1}(0)$ do problema (2.45) é não-vazio.

As hipóteses C1 e C2 garantem que o operador $T = F + B$ é monótono maximal (ver, por exemplo, [66, Proposition A.1]), e portanto os resultados da Seção 2.2 podem ser aplicados à resolução de (2.45).

Assumindo que o resolvente $(\lambda B + I)^{-1}$ do operador $B(\cdot)$ pode ser avaliado de forma eficiente, propomos dois métodos do tipo forward-backward para resolver numericamente (2.45), a saber:

1. Uma variante inercial com convergência forte do método forward-backward de Tseng [97], válida para o caso em que $F(\cdot)$ é L -Lipschitz contínuo, isto é, quando vale (2.32). Este método é apresentado no Algoritmo 3, e os principais resultados de convergência forte e complexidade estão reunidos no Teorema 2.4.2.
2. Uma modificação do método forward-backward para o caso em que $F(\cdot)$ é $(1/L)$ -cocoercivo sobre $\text{Dom } F$, isto é

$$\langle x - y, F(x) - F(y) \rangle \geq \frac{1}{L} \|F(x) - F(y)\|^2 \quad \text{para todo } x, y \in \text{Dom } F, \quad (2.46)$$

para algum $L > 0$. O algoritmo correspondente aparece como Algoritmo 4, e seus principais resultados de convergência e complexidade estão apresentados no Teorema 2.4.5.

2.4.1 Versão inercial com convergência forte do método forward-backward de Tseng

Como mencionado no início desta seção, nesta subseção analisamos o problema (2.45) sob a suposição de que as funções $F(\cdot)$, $B(\cdot)$ e/ou o conjunto C satisfazem as condições C1, C2 e C3, além da condição de L -Lipschitz continuidade dada por (2.32).

Apresentamos a seguir o algoritmo correspondente:

Algoritmo 3: Variante inercial fortemente convergente do método forward-backward de Tseng para resolver (2.45).

Entrada: $x_0 = x_{-1} \in \mathcal{H}$, $\sigma \in (0, 1)$ e $\lambda = \sigma/L$.

para $k = 0, 1, 2, \dots$ **faça**

Escolha $\alpha_k, \beta_k \geq 0$ e defina

$$\begin{aligned} w_k &= x_k + \alpha_k(x_k - x_{k-1}), \\ \tilde{w}_k &= w_k + \beta_k(w_k - x_0). \end{aligned} \tag{2.47}$$

Seja $w'_k = P_C(\tilde{w}_k)$ e calcule

$$\begin{aligned} y_k &= (\lambda B + I)^{-1}(\tilde{w}_k - \lambda F(w'_k)), \\ v_k &= F(y_k) - F(w'_k) + \frac{1}{\lambda}(\tilde{w}_k - y_k). \end{aligned} \tag{2.48}$$

Defina

$$\begin{aligned} H_k &= \{z \in \mathcal{H} \mid \langle z - y_k, v_k \rangle \leq 0\}, \\ W_k &= \{z \in \mathcal{H} \mid \langle z - x_k, x_0 - x_k \rangle \leq 0\}. \end{aligned} \tag{2.49}$$

Atualize

$$x_{k+1} = P_{H_k \cap W_k}(x_0). \tag{2.50}$$

fim

A seguir, algumas observações a respeito do Algoritmo 3.

- (i) Recomendamos ao leitor consultar os comentários apresentados após os Algoritmos 1 e 2 sobre o papel das sequências inerciais $\{\alpha_k\}$ e $\{\beta_k\}$, bem como sobre os semi-espacos H_k e W_k , e também sobre a boa definição do Algoritmo 3 em relação à projeção em (2.50).
- (ii) Do ponto de vista computacional, a operação mais custosa do Algoritmo 3 é o cálculo do resolvente $(\lambda B + I)^{-1}$ de B na equação (2.48); o cálculo de y_k e v_k como em (2.48) se assemelha à iteração do método forward-backward modificado de Tseng [97]. Também destacamos que o Algoritmo 3 pode ser aplicado para resolver o PDV (2.31), tomando $B = N_C$ em (2.45), caso em que $(\lambda B + I)^{-1} = P_C$. Nesse caso, ao comparar os Algoritmos 2 e 3, observa-se que uma vantagem principal deste último está no fato de que (2.48) exige apenas uma projeção, a saber, $y_k = P_C(\tilde{w}_k - \lambda F(w'_k))$, enquanto (2.34) requer duas projeções. Vale destacar ainda que a projeção adicional $w'_k = P_C(\tilde{w}_k)$ no passo 2 do Algoritmo 3 é necessária para recuperar a viabilidade de $\tilde{w}_k$ em relação ao domínio de $F(\cdot)$; se $\text{Dom } F = \mathcal{H}$, então basta tomar $w'_k = \tilde{w}_k$.
- (iii) Uma outra possível vantagem do Algoritmo 3 sobre o Algoritmo 2 (quando ambos são aplicados para resolver (2.31)) é a qualidade da aproximação: no primeiro, obtém-se (y_k, v_k) exatamente no gráfico de $F + N_C$, enquanto no segundo, (y_k, v_k) pertence a um “enlargement” do gráfico de $F + N_C$ — veja as inclusões em (2.38) e (2.51).

Proposição 2.4.1. *Sejam as sequências geradas pelo Algoritmo 3. Então, para todo $k \geq 0$, vale:*

$$\begin{cases} v_k \in (F + B)(y_k), \\ \|\lambda v_k + y_k - \tilde{w}_k\| \leq \sigma \|y_k - \tilde{w}_k\|. \end{cases} \quad (2.51)$$

Como consequência, assumindo que $\lambda_k \equiv \lambda$ e $\varepsilon_k \equiv 0$, conclui-se que o Algoritmo 3 é um caso particular do Algoritmo 1 aplicado para resolver o problema (2.11) com $T = F + B$.

Demonstração. A demonstração de (2.51) segue uma estrutura semelhante à apresentada em [67, Proposition 6.1]. A afirmação de que o Algoritmo 3 é um caso particular do Algoritmo 1 decorre diretamente de (2.51), das definições dos Algoritmos 3 e 1, bem como do item (d) da Proposição 1.2.10. $\square$

A seguir, apresentamos o principal resultado relacionado ao Algoritmo 3.

Teorema 2.4.2 (Convergência forte e complexidade iterativa do Algoritmo 3). *Suponha que as hipóteses C1, C2, C3 e a condição de Lipschitz em (2.32) sejam satisfeitas. Considere as sequências geradas pelo Algoritmo 3 e seja d_0 a distância de x_0 ao conjunto solução $\mathcal{S} \neq \emptyset$ do problema (2.45). Além disso, suponha que a Hipótese 2.2.2 seja válida para as sequências $\{\alpha_k\}$ e $\{\beta_k\}$. Então valem as seguintes afirmações:*

- (a) *As sequências $\{x_k\}$ e $\{y_k\}$ convergem fortemente para $x^* := P_{\mathcal{S}}(x_0)$.*
- (b) *Para todo $k \geq 0$, existe $j \in \{0, \dots, k\}$ tal que*

$$\begin{cases} v_j \in (F + B)(y_j), \\ \|v_j\| \leq \frac{2d_0L}{\sqrt{k+1}} \left(\frac{\sqrt{(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}}{(1-\sigma^2)\sigma} \right). \end{cases} \quad (2.52)$$

$$\quad (2.53)$$

Demonstração. A demonstração segue essencialmente o mesmo esquema da prova do Teorema 2.3.2, substituindo-se agora a Proposição 2.3.1 pela Proposição 2.4.1. $\square$

Observação 2.4.3. De maneira análoga ao que foi feita na Observação 2.3.3, as expressões (2.52) e (2.53) podem ser utilizadas para estimar o número de iterações necessárias do Algoritmo 3 para obter um par (y, v) tal que

$$v \in (F + B)(y) \text{ e } \|v\| \leq \rho, \quad (2.54)$$

onde $\rho > 0$ é uma tolerância previamente fixada. Como comentado na terceira observação após o Algoritmo 3, uma vantagem aparente da relação (2.54), em comparação com o resultado análogo na Observação 2.3.3, é que aqui o par (y, v) pertence exatamente ao gráfico de $F + B$, enquanto no outro caso a inclusão depende de um “enlargement” do operador.

2.4.2 Uma versão inercial do método forward-backward com convergência forte

Nesta subsecção, estudamos o problema (2.45) sob a suposição de que os operadores $F(\cdot)$, $B(\cdot)$ e/ou o conjunto C satisfazem as hipóteses C1, C2 e C3, além da condição de cocoercividade dada por (2.46).

Algoritmo 4: Uma versão inercial com convergência forte do método forward-backward para resolver (2.45).

Entrada: $x_0 = x_{-1} \in \mathcal{H}$, $\sigma \in (0, 1)$ e $\lambda = 2\sigma^2/L$.

para $k = 0, 1, 2, \dots$ **faça**

Escolha $\alpha_k, \beta_k \geq 0$ e defina

$$\begin{aligned} w_k &= x_k + \alpha_k(x_k - x_{k-1}), \\ \tilde{w}_k &= w_k + \beta_k(w_k - x_0). \end{aligned} \tag{2.55}$$

Seja $w'_k = P_C(\tilde{w}_k)$ e calcule

$$y_k = (\lambda B + I)^{-1}(\tilde{w}_k - \lambda F(w'_k)). \tag{2.56}$$

Defina os conjuntos

$$\begin{aligned} H_k &= \{z \in \mathcal{H} \mid \langle z - y_k, v_k \rangle \leq \varepsilon_k\}, \\ W_k &= \{z \in \mathcal{H} \mid \langle z - x_k, x_0 - x_k \rangle \leq 0\}, \end{aligned} \tag{2.57}$$

onde

$$\begin{aligned} v_k &= \frac{\tilde{w}_k - y_k}{\lambda}, \\ \varepsilon_k &= \frac{\|y_k - w'_k\|^2}{4L^{-1}}. \end{aligned} \tag{2.58}$$

Atualize

$$x_{k+1} = P_{H_k \cap W_k}(x_0). \tag{2.59}$$

fim

A seguir, algumas observações a respeito do Algoritmo 4.

- (i) O papel e a interpretação das extrapolações w_k e $\tilde{w}_k$, definidas em (2.55), são os mesmos discutidos anteriormente na primeira observação após o Algoritmo 1. O passo principal do Algoritmo 4 consiste no cálculo do ponto y_k por meio do passo forward-backward descrito em (2.56), o qual remete ao método clássico de passo forward-backward (ver [55, 78]). A projeção w'_k de $\tilde{w}_k$ sobre o conjunto C é necessária para garantir a viabilidade de $\tilde{w}_k$ com respeito ao conjunto de restrição C . Note que, se $C = \mathcal{H}$, isto é, se $F(\cdot)$ estiver definido em todo o espaço, então podemos simplesmente tomar $w'_k = \tilde{w}_k$. Além disso,

os conjuntos afins H_k e W_k e a atualização x_{k+1} , definidos respectivamente em (2.57) e (2.59), seguem a mesma estrutura das versões anteriores apresentadas neste capítulo.

- (ii) A principal vantagem do Algoritmo 4, quando comparado ao Algoritmo 3, é que o primeiro não exige uma avaliação extra de $F(\cdot)$ no ponto y_k , enquanto o segundo necessita calcular $F(y_k)$ para definir v_k (ver (2.48)). Por outro lado, o Algoritmo 3 oferece uma solução aproximada potencialmente mais precisa para o problema (2.45), pois a inclusão em (2.52) garante que o par está exatamente no gráfico de $F(\cdot) + B(\cdot)$, enquanto (2.61), que será apresentado a seguir, fornece um ponto pertencente a um ε -enlargement desse gráfico.
- (iii) Embora o Algoritmo 4 seja semelhante ao algoritmo principal proposto e estudado em [40, Eq. (3.1)], destacamos que, ao contrário do presente trabalho, o artigo [40] não apresenta nenhuma análise de complexidade para o método considerado.

Mostraremos agora que o Algoritmo 4 também pode ser interpretado como um caso particular do Algoritmo 1.

Proposição 2.4.4. *Sejam as sequências geradas pelo Algoritmo 4. Então, para todo $k \geq 0$, valem as seguintes relações:*

$$\begin{aligned} v_k &\in (F^{[\varepsilon_k]} + B)(y_k) \subseteq (F + B)^{[\varepsilon_k]}(y_k), \\ \lambda v_k + y_k - \tilde{w}_k &= 0, \quad 2\lambda\varepsilon_k \leq \sigma^2 \|y_k - \tilde{w}_k\|^2. \end{aligned} \quad (2.60)$$

Como consequência, o Algoritmo 4 se enquadra como um caso especial do Algoritmo 1 (com $\lambda_k \equiv \lambda$) para a resolução do problema (2.45) com $T = F + B$.

Demonstração. A demonstração segue a mesma linha da prova de [95, Proposition 5.3]. $\square$

Apresentamos a seguir o principal resultado relacionado à convergência forte e à complexidade iterativa do Algoritmo 4.

Teorema 2.4.5 (Convergência forte e complexidade iterativa do Algoritmo 4). *Suponha que as suposições C1, C2, C3 e a condição de $(1/L)$ -cocoercividade (2.46) sejam satisfeitas. Considere as sequências geradas pelo Algoritmo 4 e seja d_0 a distância de x_0 ao conjunto solução $\mathcal{S} \neq \emptyset$ de (2.45). Suponha ainda que a Hipótese 2.2.2 seja satisfeita pelas sequências $\{\alpha_k\}$ e $\{\beta_k\}$. Então valem as seguintes afirmações:*

(a) *As sequências $\{x_k\}$ e $\{y_k\}$ convergem fortemente para $x^* := P_{\mathcal{S}}(x_0)$.*

(b) *Para todo $k \geq 0$, existe $j \in \{0, \dots, k\}$ tal que*

$$\begin{cases} v_j \in (F^{[\varepsilon_j]} + B)(y_j), \\ \|v_j\| \leq \frac{2d_0L}{\sqrt{k+1}} \left(\frac{\sqrt{(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}}{2(1-\sigma^2)\sigma^2} \right), \end{cases} \quad (2.61)$$

$$\|v_j\| \leq \frac{2d_0L}{\sqrt{k+1}} \left(\frac{\sqrt{(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}}{2(1-\sigma^2)\sigma^2} \right), \quad (2.62)$$

$$\varepsilon_j \leq \frac{d_0^2L}{k+1} \left(\frac{4\sigma^2(1+\bar{\alpha})[(1+\bar{\beta})(1+\bar{\alpha}(1+\bar{\beta})) + \bar{s}]}{2(1-\sigma^2)^2\sigma^2} \right). \quad (2.63)$$

Demonstração. A demonstração segue a mesma estrutura da prova do Teorema [2.3.2](#), utilizando agora a Proposição [2.4.4](#) no lugar da Proposição [2.3.1](#). $\square$

Capítulo 3

Um algoritmo de decomposição projetivo inercial inexato com convergência forte

Neste capítulo, apresenta-se um algoritmo inercial inexato de decomposição projetivo (DP) (projective splitting (PS)), com convergência forte, para a busca de zeros de inclusões monótonas compostas envolvendo a soma de um número finito de operadores monótonos maximais. O algoritmo apresentado incorpora dois efeitos inerciais, cujos parâmetros são escolhidos de acordo com condições específicas, a fim de garantir a convergência forte dos iterados e possibilitar uma análise da complexidade iterativa. Além disso, para problemas de inclusão monótonos estruturados, desenvolvemos duas variantes do algoritmo principal, que empregam passos do tipo forward-backward e forward-backward-forward.

Os resultados apresentados neste capítulo estão publicados em [12].

3.1 O problema principal e o algoritmo

Nesta seção, propomos e analisamos um algoritmo de decomposição projetivo fortemente convergente, incorporando efeitos inerciais e tolerância a erro relativo, para a resolução do problema de inclusão monótono composto (3.2). Antes de apresentar o algoritmo, introduziremos algumas notações necessárias.

Sejam $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_{n-1}$ espaços de Hilbert reais, e defina $\mathcal{H}_n := \mathcal{H}_0$. Denotaremos por $\langle \cdot, \cdot \rangle$ o produto interno e por $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ a norma associada em cada espaço $\mathcal{H}_i$ ($1 \leq i \leq n$). Além disso, consideremos o produto cartesiano $\mathcal{H} := \mathcal{H}_0 \times \mathcal{H}_1 \times \dots \times \mathcal{H}_{n-1}$, munido do produto interno e da norma definidos, respectivamente, da seguinte maneira (para algum $\gamma > 0$):

$$\langle p, p' \rangle_\gamma = \gamma \langle z, z' \rangle + \sum_{i=1}^{n-1} \langle w_i, w'_i \rangle, \quad \|p\|_\gamma = \sqrt{\langle p, p \rangle_\gamma}, \quad (3.1)$$

onde $p = (z, w)$ e $p' = (z', w')$, com $z, z' \in \mathcal{H}_0$ e $w := (w_1, \dots, w_{n-1})$, $w' := (w'_1, \dots, w'_{n-1}) \in \mathcal{H}_1 \times \dots \times \mathcal{H}_{n-1}$. Os elementos de $\mathcal{H}$ serão denotados tanto por $p = (z, w_1, \dots, w_{n-1})$ quanto, de forma abreviada, por $p = (z, w)$, onde $w = (w_1, \dots, w_{n-1})$.

Consideremos o problema de inclusão monótono de múltiplos termos

$$0 \in \sum_{i=1}^n G_i^* T_i(G_i z) \quad (3.2)$$

onde $n \geq 2$, sob as seguintes hipóteses:

- A1. Para cada $i = 1, \dots, n$, o operador $T_i: \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ é monótono maximal.
- A2. Para cada $i = 1, \dots, n$, o operador $G_i: \mathcal{H}_0 \rightarrow \mathcal{H}_i$ é linear e limitado. Além disso, assume-se que G_n é o operador identidade em $\mathcal{H}_0 = \mathcal{H}_n$, ou seja, $G_n(z) = z$ para todo $z \in \mathcal{H}_0$.
- A3. O conjunto de soluções de (3.2) é não-vazio, isto é, existe ao menos um $z \in \mathcal{H}_0$ que satisfaz a inclusão em (3.2).

O algoritmo proposto nesta seção, a saber Algoritmo 5, é inspirado na estrutura de decomposição projetivo de [2, 43], incorporando efeitos de inércia nas iterações e critérios de erro relativo na resolução inexata dos subproblemas proximais. Além disso, o Algoritmo 5 é especialmente estruturado para garantir a convergência forte das iteradas, definindo o próximo iterado como a projeção da aproximação inicial p^0 sobre a interseção de dois semi-espacos (ver (3.18) e (3.19)). Comentários adicionais e observações relevantes são apresentados imediatamente depois da descrição do Algoritmo 5.

Quanto à análise do Algoritmo 5, os principais resultados desta seção são os Teoremas 3.1.12 e 3.1.13, que tratam, respectivamente, da convergência forte e da análise de complexidade iterativa. Na Subseção 3.1.1, provamos a convergência assintótica forte do Algoritmo 5, enquanto na Subseção 3.1.2 discutimos sua complexidade iterativa.

Antes de introduzir o algoritmo, precisamos definir o conjunto solução estendido (relacionado ao Problema (3.2)) e uma família especial de funcionais afins, conforme (3.4) e (3.7), respectivamente. A principal motivação para considerar o conjunto solução estendido $\mathcal{S}_e$ e a família $\{\varphi_k(\cdot)\}$ é, conforme [43], transformar a inclusão monotona (3.2) em um problema de viabilidade convexa (no espaço ampliado $\mathcal{H} := \mathcal{H}_0 \times \mathcal{H}_1 \times \dots \times \mathcal{H}_{n-1}$), de encontrar um ponto em $\mathcal{S}_e$, utilizando posteriormente os funcionais $\varphi_k(\cdot)$ como planos de corte.

O conjunto de soluções estendido de (3.2). Primeiramente, observamos que, como $G_n = I$ (ver Hipótese A2), o problema de inclusão monótono (3.2) pode ser reescrito como

$$0 \in \sum_{i=1}^{n-1} G_i^* T_i G_i(z) + T_n(z). \quad (3.3)$$

Definimos o conjunto de soluções estendido (também chamado de conjunto generalizado de Kuhn-Tucker) do problema (3.2) como

$$\mathcal{S}_e := \left\{ (z, w_1, \dots, w_{n-1}) \in \mathcal{H} \left| w_i \in T_i(G_i z), i = 1, \dots, n-1, - \sum_{i=1}^{n-1} G_i^* w_i \in T_n(z) \right. \right\}. \quad (3.4)$$

O conjunto $\mathcal{S}_e$ é um subconjunto fechado e convexo de $\mathcal{H}$ (ver, por exemplo, [2]) e, em virtude da Hipótese A3, também é não-vazio. De fato, é simples verificar que um ponto $z \in \mathcal{H}_0$

satisfaz (3.3) se, e somente se, existem elementos $(w_1, \dots, w_{n-1}) \in \mathcal{H}_1 \times \dots \times \mathcal{H}_{n-1}$ tais que $p := (z, w_1, \dots, w_{n-1})$ pertence a $\mathcal{S}_e$. Assim, o problema (3.2) (ou, equivalentemente, (3.3)) pode ser reformulado como o problema de viabilidade convexa de encontrar um ponto no conjunto fechado e convexo $\mathcal{S}_e$. Motivados por essa observação, introduzimos a seguir a família $\{\varphi_k(\cdot)\}$ de funcionais afins que serão utilizados como planos de separação em nosso esquema de decomposição projetivo (ver Eq. (3.18) no Algoritmo 5).

Construção da família de separadores afins. Seja (x_i^k, y_i^k) um par pertencente ao gráfico do enlargement $T_i^{[\varepsilon_i^k]}$ para cada $i = 1, \dots, n$, isto é, $y_i^k \in T_i^{[\varepsilon_i^k]}(x_i^k)$. Em virtude da definição do enlargement $T_i^{[\varepsilon_i^k]}$ em (1.9) e das inclusões em (3.4), segue que

$$\langle G_i z - x_i^k, w_i - y_i^k \rangle \geq -\varepsilon_i^k \quad \text{para todo } (z, w_1, \dots, w_{n-1}) \in \mathcal{S}_e,$$

onde definimos

$$w_n := - \sum_{i=1}^{n-1} G_i^* w_i. \quad (3.5)$$

(Observe que, de acordo com (3.4) e (3.5), temos que $w_n \in T_n(z)$). Ao somar as desigualdades acima para todos os índices $i = 1, \dots, n$, obtemos

$$\varphi_k \left(\underbrace{z, w_1, \dots, w_{n-1}}_p \right) \leq 0 \quad \text{para todo } p \in \mathcal{S}_e, \quad (3.6)$$

sendo $\varphi_k: \mathcal{H} \rightarrow \mathbb{R}$ definida como

$$\varphi_k(p) = \sum_{i=1}^n (\langle G_i z - x_i^k, y_i^k - w_i \rangle - \varepsilon_i^k) \quad \text{para todo } p = (z, w_1, \dots, w_{n-1}) \in \mathcal{H}. \quad (3.7)$$

A relação (3.6) assegura que o conjunto $\mathcal{S}_e$, onde buscamos uma solução, está contido no semi-espço fechado $H_k := \{p \in \mathcal{H} \mid \varphi_k(p) \leq 0\}$. Adicionalmente, outra construção de um semi-espço, também contendo $\mathcal{S}_e$, será apresentada mais adiante (ver Eq. (3.18)), sendo que projeções sucessivas sobre a interseção de ambos semi-espços (ver Eq. (3.19)) constituem a base do esquema iterativo do Algoritmo 5.

O próximo lema, inspirado em [50, Lemma 4], reúne algumas propriedades fundamentais de $\varphi_k(\cdot)$ que serão essenciais ao longo deste trabalho.

Lema 3.1.1. *Sejam $\varphi_k(\cdot)$ e $\mathcal{S}_e$ definidos como em (3.7) e (3.4), respectivamente. Então valem as seguintes afirmações:*

(a) *Temos que*

$$\varphi_k(p) = \left\langle z, \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right\rangle + \sum_{i=1}^{n-1} \langle x_i^k - G_i x_n^k, w_i \rangle - \sum_{i=1}^n \langle x_i^k, y_i^k \rangle - \sum_{i=1}^n \varepsilon_i^k, \quad (3.8)$$

para todo $p = (z, w_1, \dots, w_{n-1}) \in \mathcal{H}$. Em particular, $\varphi_k(\cdot)$ é um funcional afim sobre $\mathcal{H}$.

(b) *Vale que $\varphi_k(p) \leq 0$ para todo $p \in \mathcal{S}_e$.*

(c) O gradiente de $\varphi_k(\cdot)$, considerando o produto interno $\langle \cdot, \cdot \rangle_\gamma$ definido em (3.1), é dado por

$$\nabla \varphi_k = \left(\gamma^{-1} \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right), x_1^k - G_1 x_n^k, \dots, x_{n-1}^k - G_{n-1} x_n^k \right), \quad (3.9)$$

e sua norma é

$$\|\nabla \varphi_k\|_\gamma^2 = \gamma^{-1} \left\| \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right\|^2 + \sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2. \quad (3.10)$$

(d) O gradiente $\nabla \varphi_k$ é nulo se, e somente se, $\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$ e $x_i^k = G_i x_n^k$ para todo $i = 1, \dots, n-1$.

(e) Se $\nabla \varphi_k = 0$, então $(x_n^k, y_1^k, \dots, y_{n-1}^k) \in \mathcal{S}_e$ sempre que uma das seguintes condições for satisfeita:

- (i) $\varepsilon_i^k = 0$, para todo $i = 1, \dots, n$;
- (ii) existe $\tilde{p} = (\tilde{z}, \tilde{w}_1, \dots, \tilde{w}_{n-1})$ tal que $\varphi_k(\tilde{p}) \geq 0$.

Demonstração. A demonstração segue, em linhas gerais, o raciocínio apresentado em [50, Lemma 4].

A partir de (3.7), podemos reescrever φ_k da seguinte forma:

$$\begin{aligned} \varphi_k(z, w_1, \dots, w_{n-1}) &= \sum_{i=1}^n \langle G_i z, y_i^k - w_i \rangle - \sum_{i=1}^n \langle x_i^k, y_i^k - w_i \rangle - \sum_{i=1}^n \varepsilon_i^k \\ &= \sum_{i=1}^n \langle z, G_i^* y_i^k - G_i^* w_i \rangle - \sum_{i=1}^n \langle x_i^k, y_i^k \rangle + \sum_{i=1}^{n-1} \langle x_i^k, w_i \rangle + \langle x_n^k, w_n \rangle - \sum_{i=1}^n \varepsilon_i^k \\ &= \gamma \langle z, \frac{1}{\gamma} \sum_{i=1}^n G_i^* y_i^k \rangle + \sum_{i=1}^{n-1} \langle x_i^k - G_i x_n^k, w_i \rangle - \sum_{i=1}^n \langle x_i^k, y_i^k \rangle - \sum_{i=1}^n \varepsilon_i^k. \end{aligned} \quad (3.11)$$

A última igualdade é obtida utilizando (3.5) e o fato de que $G_n = I$ (ver hipótese A2), o que implica que $\sum_{i=1}^n G_i^* w_i = 0$ e $\langle x_n^k, w_n \rangle = \sum_{i=1}^{n-1} \langle -G_i x_n^k, w_i \rangle$.

As afirmações dos itens (a) e (c) seguem diretamente da expressão (3.11) juntamente com a definição do produto interno em (3.1). Já o item (b) decorre naturalmente da construção de φ_k , conforme apresentado em (3.6). O item (d) decorre de maneira imediata a partir da fórmula do gradiente (3.9).

Finalmente, para provar o item (e), suponha inicialmente que $\nabla \varphi_k = 0$. Como $y_i^k \in T_i^{[\varepsilon_i^k]}(x_i^k)$ para todo $i = 1, \dots, n$, usando o item (d), deduzimos que

$$y_i^k \in T_i^{[\varepsilon_i^k]}(G_i x_n^k) \quad (i = 1, \dots, n-1) \quad \text{e} \quad - \sum_{i=1}^{n-1} G_i^* y_i^k = y_n^k \in T_n^{[\varepsilon_n^k]}(x_n^k).$$

Se $\varepsilon_i^k = 0$ para todo $i = 1, \dots, n$, então, pela definição de $\mathcal{S}_e$ e pela Proposição 1.2.10(d), temos que $(x_n^k, y_1^k, \dots, y_{n-1}^k) \in \mathcal{S}_e$, o que conclui a prova de (i). Para provar (ii), suponha que $\varphi_k(\tilde{p}) \geq 0$ para algum $\tilde{p} \in \mathcal{H}$. Como $\nabla \varphi_k = 0$, pelas expressões (3.8) e (3.9) e usando o item (d), obtemos que $\sum_{i=1}^n \varepsilon_i^k = -\varphi_k(\tilde{p}) \leq 0$. Assim, $\varepsilon_i^k = 0$ para todo $i = 1, \dots, n$, e o resultado segue do item (i). $\square$

Agora, introduzimos o principal algoritmo deste capítulo para a resolução de (3.3). O algoritmo combina três elementos fundamentais: um passo inercial, um procedimento de decomposição proximal com critério de erro relativo para decompor o problema em subproblemas tratáveis, e uma projeção sobre a interseção de semi-espacos para garantir a convergência forte.

Algoritmo 5: Algoritmo de DP inexato inercial fortemente convergente para resolver (3.3).

Entrada: $(z^{-1}, w_1^{-1}, \dots, w_{n-1}^{-1}) = (z^0, w_1^0, \dots, w_{n-1}^0) \in \mathcal{H}$, escalares $0 \leq \sigma < 1$ e $\gamma > 0$.

para $k = 0, 1, \dots$ **faça**

Escolha $\alpha_k \geq 0, \beta_k \geq 0$ e defina

$$\hat{z}^k = z^k + \alpha_k(z^k - z^{k-1}), \quad \tilde{z}^k = \hat{z}^k + \beta_k(\hat{z}^k - z^0), \quad (3.12)$$

$$\hat{w}_i^k = w_i^k + \alpha_k(w_i^k - w_i^{k-1}), \quad \tilde{w}_i^k = \hat{w}_i^k + \beta_k(\hat{w}_i^k - w_i^0) \quad \forall i = \overline{1, n-1}, \quad (3.13)$$

$$\tilde{w}_n^k = - \sum_{i=1}^{n-1} G_i^* \tilde{w}_i^k. \quad (3.14)$$

Para cada $i = 1, \dots, n$, escolha escalares $\lambda_i^k > 0$ e compute

$(x_i^k, y_i^k, \varepsilon_i^k) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}_+$ tais que

$$y_i^k \in T_i^{[\varepsilon_i^k]}(x_i^k),$$

$$\|\lambda_i^k y_i^k + x_i^k - (G_i \tilde{z}^k + \lambda_i^k \tilde{w}_i^k)\|^2 + 2\lambda_i^k \varepsilon_i^k \leq \sigma^2 \left(\|G_i \tilde{z}^k - x_i^k\|^2 + \|\lambda_i^k (\tilde{w}_i^k - y_i^k)\|^2 \right). \quad (3.15)$$

se $x_i^k = G_i x_n^k$ ($i = 1, \dots, n-1$) e $\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$ **então**

retorna

$$(\bar{z}, \bar{w}_1, \dots, \bar{w}_{n-1}) := (x_n^k, y_1^k, \dots, y_{n-1}^k). \quad (3.16)$$

fim

Defina o funcional afim $\varphi_k(\cdot)$ como em (3.7) (ver também (3.8)), seja

$$p^k := (z^k, w_1^k, \dots, w_{n-1}^k), \quad (3.17)$$

defina também

$$H_k = \{p \in \mathcal{H} \mid \varphi_k(p) \leq 0\},$$

$$W_k = \{p \in \mathcal{H} \mid \langle p^0 - p^k, p - p^k \rangle_\gamma \leq 0\}, \quad (3.18)$$

e atualize

$$(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1}) = P_{H_k \cap W_k}(p^0). \quad (3.19)$$

fim

Notação adicional. Para facilitar a análise do Algoritmo 5, introduzimos a seguir algumas notações auxiliares. De maneira análoga à definição de $\tilde{w}_n^k$ em (3.14), definem-se, para as

sequências $\{w_i^k\}$ e $\{\widehat{w}_i^k\}$, com $1 \leq i \leq n-1$,

$$w_n^k := - \sum_{i=1}^{n-1} G_i^* w_i^k \quad \text{e} \quad \widehat{w}_n^k := - \sum_{i=1}^{n-1} G_i^* \widehat{w}_i^k. \quad (3.20)$$

Analogamente à definição de p^k apresentada em (3.17), consideram-se os vetores

$$\widehat{p}^k := (\widehat{z}^k, \widehat{w}_1^k, \dots, \widehat{w}_{n-1}^k) \quad \text{e} \quad \widetilde{p}^k := (\widetilde{z}^k, \widetilde{w}_1^k, \dots, \widetilde{w}_{n-1}^k). \quad (3.21)$$

A partir das definições em (3.12), (3.13), (3.17) e (3.21), obtém-se

$$\widehat{p}^k = p^k + \alpha_k(p^k - p^{k-1}) \quad \text{e} \quad \widetilde{p}^k = \widehat{p}^k + \beta_k(\widehat{p}^k - p^0).$$

Substituindo diretamente a primeira identidade de (3.13) na segunda identidade de (3.20), e a segunda identidade de (3.13) em (3.14), deduz-se que

$$\widehat{w}_n^k = w_n^k + \alpha_k(w_n^k - w_n^{k-1}) \quad \text{e} \quad \widetilde{w}_n^k = \widehat{w}_n^k + \beta_k(\widehat{w}_n^k - w_n^0). \quad (3.22)$$

Sempre que necessário, utilizaremos também a seguinte notação:

$$e_i^k := \lambda_i^k y_i^k + x_i^k - (G_i \widetilde{z}^k + \lambda_i^k \widetilde{w}_i^k), \quad i = 1, \dots, n. \quad (3.23)$$

Com esta notação, a desigualdade apresentada em (3.15) pode ser reescrita como

$$\|e_i^k\|^2 + 2\lambda_i^k \varepsilon_i^k \leq \sigma^2 \left(\|G_i \widetilde{z}^k - x_i^k\|^2 + \|\lambda_i^k (\widetilde{w}_i^k - y_i^k)\|^2 \right). \quad (3.24)$$

Além disso, observa-se que, como G_n é assumido igual ao operador identidade (ver Hipótese A2), a definição (3.14) implica que

$$\sum_{i=1}^n G_i^* \widetilde{w}_i^k = 0. \quad (3.25)$$

A seguir, algumas observações sobre o Algoritmo 5:

- (i) O Algoritmo 5 atua no espaço aumentado $\mathcal{H} = \mathcal{H}_0 \times \mathcal{H}_1 \times \dots \times \mathcal{H}_{n-1}$. Recorde que o conjunto de soluções estendido $\mathcal{S}_e$ é um subconjunto de $\mathcal{H}$ e que nosso problema principal (3.2) é equivalente ao problema de viabilidade convexa de encontrar um ponto $p = (z, w_1, \dots, w_{n-1})$ em $\mathcal{S}_e$. Além disso, pelo Lema 3.1.1(b), o conjunto $\mathcal{S}_e$ é contido no semiespaço “negativo” definido por cada funcional afim $\varphi_k(\cdot)$ (ver definição de H_k em (3.18)). Isso explica a natureza “projetiva” do Algoritmo 5, no qual cada iteração é construída, conforme (3.19), por projeções sucessivas sobre a interseção dos semiespaços H_k e W_k . O principal papel do semiespaço W_k é garantir a convergência forte das sequências geradas pelo Algoritmo 5 (ver Teorema 3.1.12 mais adiante).
- (ii) As etapas de extrapolação definidas em (3.12)–(3.14) introduzem *efeitos inerciais* nas iterações, controlados pelos parâmetros α_k e β_k . Observamos que a primeira extrapolação, envolvendo $\widehat{z}^k$ e $\widehat{w}_i^k$ ($i = 1, \dots, n-1$), corresponde ao tipo mais usual de extrapolação presente em diversos algoritmos com inércia (ver, por exemplo, [5, 8, 9, 18, 16]). Já a segunda extrapolação, envolvendo $\widetilde{z}^k$ e $\widetilde{w}_i^k$ ($i = 1, \dots, n-1$), foi especificamente introduzida para considerar o mecanismo de iteração do Algoritmo 5. As condições sobre as sequências $\{\alpha_k\}$ e $\{\beta_k\}$ que garantem a convergência forte e a complexidade iterativa são as mesmas que na Hipótese 2.2.2 (ver também Hipótese 3.1.6).

- (iii) Em cada iteração, o Algoritmo 5 exige a resolução aproximada de n subproblemas proximais, conforme (3.15). Esses subproblemas podem ser resolvidos de forma independente, em paralelo e aproximadamente, obedecendo a um critério de erro relativo controlado pelo parâmetro $\sigma \in [0, 1)$. Note que, no caso $\sigma = 0$, a equação (3.15) implica que $\varepsilon_i^k = 0$ e

$$\lambda_i^k y_i^k + x_i^k = G_i \tilde{z}^k + \lambda_i^k \tilde{w}_i^k, \quad y_i^k \in T_i(x_i^k),$$

para todo $i = 1, \dots, n$ (recorde que $T^{[0]} = T$). Assim, dado $\lambda_i^k > 0$, a tripla $(x_i^k, y_i^k, \varepsilon_i^k)$ definido como

$$x_i^k := G_i \tilde{z}^k + \lambda_i^k \tilde{w}_i^k - \lambda_i^k y_i^k, \quad y_i^k := (\lambda_i^k T_i + I)^{-1}(G_i \tilde{z}^k + \lambda_i^k \tilde{w}_i^k), \quad \varepsilon_i^k := 0 \quad (3.26)$$

satisfaz automaticamente o critério de erro (3.15). A principal limitação de (3.26) reside no custo computacional para calcular o resolvente exato $(\lambda_i^k T_i + I)^{-1}$ de cada operador monótono maximal T_i . O critério de erro relativo (3.15) contorna essa dificuldade, permitindo o cálculo aproximado das tríplas $(x_i^k, y_i^k, \varepsilon_i^k)$ por meio de métodos internos adequados à estrutura de T_i , utilizando (3.15) como critério de parada. Ressaltamos ainda que o critério (3.15) generaliza um critério de erro relativo anteriormente proposto para algoritmos DP com inércia em [10]. Na Seção 3.2, mostraremos como o Algoritmo 5 pode ser usado como base para projetar novos métodos de decomposição de operadores, onde as tríplas $(x_i^k, y_i^k, \varepsilon_i^k)$ podem ser computadas via passos de tipo forward-backward ou forward-backward-forward (de Tseng), sob hipóteses de cocoercividade ou Lipschitz continuidade dos operadores T_i .

- (iv) No Corolário 3.1.3 mais adiante, mostraremos que, caso o Algoritmo 5 pare (ver (3.16)), então o ponto $(\bar{z}, \bar{w}_1, \dots, \bar{w}_{n-1})$ pertence ao conjunto de soluções estendido $\mathcal{S}_e$. Para a análise da convergência, assumiremos então que o algoritmo gera sequências infinitas (ver também a Observação 3.1.4).
- (v) A definição do próximo iterado $(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1})$ como a projeção ortogonal de p^0 sobre a interseção $H_k \cap W_k$ dos semi-espacos (conforme (3.19)) é um dos principais mecanismos que asseguram a convergência forte do Algoritmo 5 (note que H_k é de fato um semi-espaco em $\mathcal{H}$, como garantido pelo Lema 3.1.1 (a)). No contexto de métodos proximais inexatos com critério de erro relativo, essa ideia remonta a [92] e foi recentemente explorada em [11] (em algoritmos com inércia). A computação dessa projeção ortogonal reduz-se (no máximo) à resolução de um sistema linear 2×2 (ver [92, pp. 195–196] e [24, Propositions 29.22 e 29.23]). Destacamos que a definição do próximo iterado depende do fato que a interseção $H_k \cap W_k$ seja não-vazia, o que será discutido com mais detalhes no Apêndice B (ver Proposição B.1.1).
- (vi) Por fim, é importante ressaltar que tanto a implementação quanto a análise de convergência do Algoritmo 5 são independentes do cálculo explícito de normas ou inversas dos operadores lineares limitados G_i ($1 \leq i \leq n$).

3.1.1 Análise de convergência do Algoritmo 5

Nesta subseção, estudamos as propriedades de *convergência forte* do Algoritmo 5. O principal resultado será apresentado no Teorema 3.1.12, ao passo que outros resultados auxiliares também serão enunciados e demonstrados ao longo da subseção.

Para o estabelecimento da convergência forte do Algoritmo 5, serão necessárias as Hipóteses 3.1.6 e 2.2.2, que impõem condições sobre as sequências de parâmetros proximais $\{\lambda_i^k\}$ ($1 \leq i \leq n$) e os parâmetros inerciais $\{\alpha_k\}$ e $\{\beta_k\}$, respectivamente.

Iniciamos com o seguinte lema técnico.

Lema 3.1.2. *Suponha que a iteração $k \geq 0$ é atingida no Algoritmo 5 e seja $\{\tilde{p}^k\}$ e $\varphi_k(\cdot)$ definidos conforme (3.21) e (3.7), respectivamente. Então valem as seguintes afirmações:*

(a) *Para todo $i = 1, \dots, n$, tem-se*

$$\langle G_i \tilde{z}^k - x_i^k, y_i^k - \tilde{w}_i^k \rangle - \varepsilon_i^k \geq \frac{1 - \sigma^2}{2\lambda_i^k} (\|G_i \tilde{z}^k - x_i^k\|^2 + \|\lambda_i^k(\tilde{w}_i^k - y_i^k)\|^2). \quad (3.27)$$

(b) *Temos a estimativa inferior*

$$\varphi_k(\tilde{p}^k) \geq \frac{1 - \sigma^2}{2} \sum_{i=1}^n \left(\frac{1}{\lambda_i^k} \|G_i \tilde{z}^k - x_i^k\|^2 + \lambda_i^k \|\tilde{w}_i^k - y_i^k\|^2 \right) \geq 0.$$

Demonstração. (a) Aplicando a identidade $\langle a, b \rangle = (1/2)(\|a\|^2 + \|b\|^2 - \|a - b\|^2)$ com $a = G_i \tilde{z}^k - x_i^k$ e $b = \lambda_i^k(y_i^k - \tilde{w}_i^k)$, obtemos:

$$\langle G_i \tilde{z}^k - x_i^k, \lambda_i^k(y_i^k - \tilde{w}_i^k) \rangle = \frac{1}{2} \left(\|G_i \tilde{z}^k - x_i^k\|^2 + \|\lambda_i^k(y_i^k - \tilde{w}_i^k)\|^2 - \underbrace{\|G_i \tilde{z}^k - x_i^k - \lambda_i^k(y_i^k - \tilde{w}_i^k)\|^2}_{-e_i^k} \right),$$

em que e_i^k está definido como em (3.23). Subtraindo ε_i^k de ambos os lados da igualdade acima e utilizando o critério de erro relativo em (3.15), concluímos (ver também (3.24)):

$$\begin{aligned} \langle G_i \tilde{z}^k - x_i^k, y_i^k - \tilde{w}_i^k \rangle - \varepsilon_i^k &= \frac{1}{2\lambda_i^k} (\|G_i \tilde{z}^k - x_i^k\|^2 + \|\lambda_i^k(y_i^k - \tilde{w}_i^k)\|^2 - [\|e_i^k\|^2 + 2\lambda_i^k \varepsilon_i^k]) \\ &\geq \frac{1 - \sigma^2}{2\lambda_i^k} (\|G_i \tilde{z}^k - x_i^k\|^2 + \|\lambda_i^k(\tilde{w}_i^k - y_i^k)\|^2), \end{aligned}$$

o que conclui a demonstração do item (a).

(b) O resultado segue imediatamente da soma das desigualdades (3.27) para $i = 1, \dots, n$ e das definições de $\varphi_k(\cdot)$ e $\tilde{p}^k$ (ver (3.7) e (3.21)). $\square$

A seguir, mostramos que se o Algoritmo 5 parar, então ele para exatamente em uma solução do problema (3.2).

Corolário 3.1.3. *Se o Algoritmo 5 retornar $(\bar{z}, \bar{w}_1, \dots, \bar{w}_{n-1})$ conforme definido em (3.16), então $(\bar{z}, \bar{w}_1, \dots, \bar{w}_{n-1})$ pertence ao conjunto de soluções estendido $\mathcal{S}_e$. Em particular, $\bar{z}$ é uma solução do problema (3.2).*

Demonstração. A demonstração é uma consequência direta do Lema 3.1.2(b) e dos itens (d) e (e)(ii) do Lema 3.1.1. $\square$

Observação 3.1.4. À luz do Corolário 3.1.3 e da Proposição B.1.1 apresentada no Apêndice B, assumiremos, a partir deste ponto da seção, que o Algoritmo 5 não para, ou seja, gera sequências infinitas. Em particular, conforme o item (d) do Lema 3.1.1, podemos garantir que

$$\nabla \varphi_k \neq 0 \quad \text{para todo } k \geq 0. \quad (3.28)$$

Proposição 3.1.5. *Seja $\{p^k\}$ a sequência gerada pelo Algoritmo 5 e seja $d_{0,\gamma}$ a distância de p^0 ao conjunto de soluções estendido S_e do problema (3.2). Os seguintes resultados são válidos:*

(a) Para todo $k \geq 0$,

$$\|p^{k+1} - p^0\|_\gamma^2 \geq \|p^k - p^0\|_\gamma^2 + \|p^{k+1} - p^k\|_\gamma^2.$$

(b) Para todo $k \geq 0$,

$$\|p^k - p^0\|_\gamma \leq d_{0,\gamma}.$$

(c) Temos que

$$\sum_{k=0}^{\infty} \|p^{k+1} - p^k\|_\gamma^2 \leq d_{0,\gamma}^2.$$

Demonstração. (a) Aplicando a desigualdade (1.3) do Lema 1.1.1 (no espaço $\mathcal{H}$ com o produto interno definido em (3.1)), com $C = W_k$, $x = p^{k+1}$ e $y = p^0$, obtemos:

$$\|P_{W_k}(p^{k+1}) - P_{W_k}(p^0)\|_\gamma^2 \leq \|p^{k+1} - p^0\|_\gamma^2 - \|[p^{k+1} - P_{W_k}(p^{k+1})] - [p^0 - P_{W_k}(p^0)]\|_\gamma^2,$$

onde W_k é definido como em (3.18). Como $p^{k+1} = P_{W_k}(p^{k+1})$, pois $p^{k+1} \in W_k$ (ver (3.17) e (3.19)), e $p^k = P_{W_k}(p^0)$ (por definição de W_k), a desigualdade acima se reduz a:

$$\|p^{k+1} - p^k\|_\gamma^2 \leq \|p^{k+1} - p^0\|_\gamma^2 - \|p^k - p^0\|_\gamma^2,$$

o que conclui a demonstração do item (a).

(b) Suponha $k \geq 1$ (para $k = 0$ o resultado é imediato). De (3.17) e (3.19), temos que $p^k = P_{H_{k-1} \cap W_{k-1}}(p^0)$. Usando a Proposição B.1.1 e a definição da projeção $P_{H_{k-1} \cap W_{k-1}}(p^0)$, segue que

$$\|p^k - p^0\|_\gamma = \|P_{H_{k-1} \cap W_{k-1}}(p^0) - p^0\|_\gamma \leq \|P_{S_e}(p^0) - p^0\|_\gamma = d_{0,\gamma}.$$

(c) Este item decorre diretamente dos itens (a) e (b), utilizando o argumento de soma telescópica. $\square$

A seguir, impomos uma suposição de limitação para as sequências de parâmetros proximais $\{\lambda_i^k\}$ geradas pelo Algoritmo 5:

Hipótese 3.1.6. *Para todo $i = 1, \dots, n$, a sequência $\{\lambda_i^k\}$ gerada pelo Algoritmo 5 satisfaz*

$$\lambda_i^k \in [\underline{\lambda}, \bar{\lambda}] \quad \text{para todo } k \geq 0,$$

em que $\bar{\lambda} \geq \underline{\lambda} > 0$.

Lema 3.1.7. *Considere as sequências envolvidas no Algoritmo 5 e seja $\{\tilde{p}^k\}$ definida conforme (3.21). Além disso, seja $\varphi_k(\cdot)$ a função definida em (3.7). Valem as seguintes afirmações:*

(a) Para todo $k \geq 0$,

$$\varphi_k(\tilde{p}^k) \geq \frac{(1 - \sigma^2) \min\{\underline{\lambda}, 1/\bar{\lambda}\}}{2} \sum_{i=1}^n (\|G_i \tilde{z}^k - x_i^k\|^2 + \|\tilde{w}_i^k - y_i^k\|^2). \quad (3.29)$$

(b) Para todo $k \geq 0$,

$$\varphi_k(\tilde{p}^k) \geq \frac{1}{c} \|\nabla \varphi_k\|_\gamma^2 \quad (3.30)$$

em que

$$c := n \max_{i=1, \dots, n} \|G_i\|^2 \left(\frac{4 \max\{1, 1/\gamma\}}{(1 - \sigma^2) \min\{\underline{\lambda}, 1/\bar{\lambda}\}} \right). \quad (3.31)$$

Demonstração. (a) Esta afirmação decorre diretamente do Lema 3.1.2(b) e a Hipótese 3.1.6.

(b) Primeiramente, observe que, a partir das identidades (3.25), (1.5) e do fato de que $\|G_i\| = \|G_i^*\|$ para todo $1 \leq i \leq n$, obtemos:

$$\left\| \sum_{i=1}^n G_i^* y_i^k \right\|^2 = \left\| \sum_{i=1}^n G_i^* (\tilde{w}_i^k - y_i^k) \right\|^2 \leq n \max_{i=1, \dots, n} \|G_i\|^2 \sum_{i=1}^n \|\tilde{w}_i^k - y_i^k\|^2. \quad (3.32)$$

Analogamente (lembrando que $G_n = I$), temos:

$$\begin{aligned} \sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2 &= \sum_{i=1}^{n-1} \|x_i^k - G_i \tilde{z}^k + G_i (\tilde{z}^k - x_n^k)\|^2 \\ &\leq 2 \sum_{i=1}^{n-1} (\|G_i \tilde{z}^k - x_i^k\|^2 + \|G_i (\tilde{z}^k - x_n^k)\|^2) \\ &\leq 2 \left(\sum_{i=1}^{n-1} \|G_i \tilde{z}^k - x_i^k\|^2 + n \max_{i=1, \dots, n} \|G_i\|^2 \|\tilde{z}^k - x_n^k\|^2 \right) \\ &\leq 2n \max_{i=1, \dots, n} \|G_i\|^2 \sum_{i=1}^n \|G_i \tilde{z}^k - x_i^k\|^2. \end{aligned} \quad (3.33)$$

Assim, a partir de (3.10), (3.29), (3.32) e (3.33), concluímos que:

$$\begin{aligned} \|\nabla \varphi_k\|_\gamma^2 &= \gamma^{-1} \left\| \sum_{i=1}^n G_i^* y_i^k \right\|^2 + \sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2 \\ &\leq 2n \max_{i=1, \dots, n} \|G_i\|^2 \max\{1, 1/\gamma\} \sum_{i=1}^n (\|G_i \tilde{z}^k - x_i^k\|^2 + \|\tilde{w}_i^k - y_i^k\|^2) \\ &\leq n \max_{i=1, \dots, n} \|G_i\|^2 \left(\frac{4 \max\{1, 1/\gamma\}}{(1 - \sigma^2) \min\{\underline{\lambda}, 1/\bar{\lambda}\}} \right) \varphi_k(\tilde{p}^k) \\ &= c \varphi_k(\tilde{p}^k), \end{aligned}$$

o que mostra a validade da desigualdade (3.30). $\square$

Lema 3.1.8. *Considere as sequências geradas pelo Algoritmo 5, seja $\varphi_k(\cdot)$ definida como em (3.7) e seja $c > 0$ a constante dada em (3.31). Então, para todo $k \geq 0$, valem as seguintes desigualdades:*

- (a) $\|\nabla\varphi_k\|_\gamma \leq c\|p^{k+1} - \tilde{p}^k\|_\gamma \quad \text{e} \quad \varphi_k(\tilde{p}^k) \leq c\|p^{k+1} - \tilde{p}^k\|_\gamma^2.$
- (b) $\sum_{i=1}^n \varepsilon_i^k \leq \left(\frac{\sigma^2}{1 - \sigma^2}\right) c\|p^{k+1} - \tilde{p}^k\|_\gamma^2.$

Demonstração. (a) Para demonstrar a primeira desigualdade, recorde que estamos assumindo $\nabla\varphi_k \neq 0$ (ver (3.28)). Assim, a partir da definição do semi-espço H_k em (3.18) e do Lema 3.1.1(a), obtemos:

$$P_{H_k}(p) = p - \frac{\max\{0, \varphi_k(p)\}}{\|\nabla\varphi_k\|_\gamma^2} \nabla\varphi_k \quad \text{para todo } p \in \mathcal{H}.$$

Como $p^{k+1} \in H_k$ (ver (3.17) e (3.19)) e utilizando o Lema 3.1.7(b), obtemos:

$$\begin{aligned} \|p^{k+1} - \tilde{p}^k\|_\gamma &\geq \|P_{H_k}(\tilde{p}^k) - \tilde{p}^k\|_\gamma \\ &= \frac{\varphi_k(\tilde{p}^k)}{\|\nabla\varphi_k\|_\gamma} \\ &\geq c^{-1}\|\nabla\varphi_k\|_\gamma, \end{aligned} \tag{3.34}$$

o que conclui a demonstração da primeira desigualdade. A segunda segue diretamente ao combinar (3.34) com a primeira.

(b) A partir do critério de erro dado em (3.15), obtemos, para cada $i = 1, \dots, n$,

$$\varepsilon_i^k \leq \frac{\sigma^2}{2} \left(\frac{1}{\lambda_i^k} \|G_i \tilde{z}^k - x_i^k\|^2 + \lambda_i^k \|\tilde{w}_i^k - y_i^k\|^2 \right).$$

Somando as desigualdades acima para $i = 1, \dots, n$ e usando o Lema 3.1.2(b), obtemos:

$$\sum_{i=1}^n \varepsilon_i^k \leq \frac{\sigma^2}{1 - \sigma^2} \varphi_k(\tilde{p}^k).$$

Por fim, para concluir a demonstração do item (b), basta combinar a desigualdade acima com a segunda do item (a). $\square$

A partir de agora, assumiremos neste capítulo que as sequências $\{\alpha_k\}$ e $\{\beta_k\}$ dadas no Algoritmo 5 satisfazem a Hipótese 2.2.2 do capítulo anterior. Observe que essa hipótese também foi imposta às sequências de parâmetros inerciais consideradas no Algoritmo 1, nas quais também se utilizou a notação α_k e β_k .

O resultado a seguir é fundamental tanto para a análise de convergência quanto para o estudo da complexidade iterativa do Algoritmo 5. A demonstração será omitida pois segue a mesma linha de raciocínio da prova da Proposição 2.2.4(b).

Proposição 3.1.9. *Seja $\{p^k\}$ a sequência gerada pelo Algoritmo 5, e seja $\{\tilde{p}^k\}$ definida como em (3.21). Denote por $d_{0,\gamma}$ a distância de p^0 ao conjunto de soluções estendido S_e de (3.2). Além disso, sejam $\bar{\alpha}$, $\bar{\beta}$ e $\bar{s}$ conforme definidos na Hipótese 2.2.2. Então, vale que*

$$\sum_{k=0}^{\infty} \|p^{k+1} - \tilde{p}^k\|_{\gamma}^2 \leq \Omega_{(\alpha,\beta)} d_{0,\gamma}^2 \quad (3.35)$$

onde

$$\Omega_{(\alpha,\beta)} := (1 + \bar{\alpha}) \left[(1 + \bar{\beta}) [1 + \bar{\alpha}(1 + \bar{\beta})] + \bar{s} \right]. \quad (3.36)$$

Proposição 3.1.10. *Considere as sequências geradas pelo Algoritmo 5. Então, temos:*

- (a) $y_n^k + \sum_{i=1}^{n-1} G_i^* y_i^k \rightarrow 0$, e para todo $i = 1, \dots, n$, tem-se $x_i^k - G_i x_n^k \rightarrow 0$.
- (b) Para todo $i = 1, \dots, n$, vale que $\varepsilon_i^k \rightarrow 0$.
- (c) Para todo $i = 1, \dots, n$, tem-se $G_i \tilde{z}^k - x_i^k \rightarrow 0$ e $\tilde{w}_i^k - y_i^k \rightarrow 0$.
- (d) Para todo $i = 1, \dots, n$, tem-se $G_i \hat{z}^k - x_i^k \rightarrow 0$ e $\hat{w}_i^k - y_i^k \rightarrow 0$.
- (e) Para todo $i = 1, \dots, n$, tem-se $G_i z^k - x_i^k \rightarrow 0$ e $w_i^k - y_i^k \rightarrow 0$.

Demonstração. (a) Este item decorre diretamente da primeira desigualdade do Lema 3.1.8(a), em conjunto com (3.10) e (3.35) (lembre que $G_n = I$).

(b) Resulta como consequência imediata do Lema 3.1.8(b) combinado com (3.35).

(c) Este item segue da segunda desigualdade do Lema 3.1.8(a), da Proposição 3.1.9 e do Lema 3.1.7(a).

(d) Aplicando a desigualdade triangular e a segunda identidade de (3.12), obtemos:

$$\begin{aligned} \|G_i \hat{z}^k - x_i^k\| &\leq \|\tilde{z}^k - \hat{z}^k\| \|G_i\| + \|G_i \tilde{z}^k - x_i^k\| \\ &= \beta_k \|\tilde{z}^k - z^0\| \|G_i\| + \|G_i \tilde{z}^k - x_i^k\|, \quad i = 1, \dots, n. \end{aligned} \quad (3.37)$$

De modo análogo, temos também:

$$\begin{aligned} \|\hat{w}_i^k - y_i^k\| &\leq \|\tilde{w}_i^k - \hat{w}_i^k\| + \|\tilde{w}_i^k - y_i^k\| \\ &= \beta_k \|\tilde{w}_i^k - w_i^0\| + \|\tilde{w}_i^k - y_i^k\|, \quad i = 1, \dots, n-1. \end{aligned} \quad (3.38)$$

Além disso, usando a desigualdade (1.5), (3.14), e a segunda identidade em (3.20) e (3.13), temos:

$$\begin{aligned} \frac{1}{2} \|y_n^k - \hat{w}_n^k\|^2 &\leq \|\tilde{w}_n^k - \hat{w}_n^k\|^2 + \|y_n^k - \tilde{w}_n^k\|^2 \\ &= \left\| \sum_{i=1}^{n-1} G_i^* (\tilde{w}_i^k - \hat{w}_i^k) \right\|^2 + \|y_n^k - \tilde{w}_n^k\|^2 \\ &\leq n \max_{i=1, \dots, n-1} \|G_i\|^2 \beta_k^2 \sum_{i=1}^{n-1} \|\hat{w}_i^k - w_i^0\|^2 + \|y_n^k - \tilde{w}_n^k\|^2. \end{aligned} \quad (3.39)$$

Como a sequência $\{p^k\}$ é limitada (ver Proposição 3.1.5(b)), segue de (3.17) e (3.20) que as sequências $\{z^k\}$ e $\{w_i^k\}$ também são limitadas, para todo $i = 1, \dots, n$. Portanto, as sequências $\{\widehat{z}^k\}$, $\{\widehat{w}_i^k\}$ e $\{\widehat{w}_i^k\}$ ($i = 1, \dots, n$) também são limitadas (basta usar (3.12)–(3.14) e o fato de que $\{\alpha_k\}$ e $\{\beta_k\}$ são limitadas). Além disso, pela Hipótese 2.2.2(b), temos que $\beta_k \rightarrow 0$. Assim, o resultado desejado segue de (3.37)–(3.39) e do item (c).

(e) Argumentando de forma análoga à do item (d), obtemos:

$$\begin{aligned} \|G_i z^k - x_i^k\| &\leq \|z^k - \widehat{z}^k\| \|G_i\| + \|G_i \widehat{z}^k - x_i^k\| \\ &= \alpha_k \|z^k - z^{k-1}\| \|G_i\| + \|G_i \widehat{z}^k - x_i^k\|, \quad i = 1, \dots, n. \end{aligned} \quad (3.40)$$

e

$$\begin{aligned} \|w_i^k - y_i^k\| &\leq \|w_i^k - \widehat{w}_i^k\| + \|\widehat{w}_i^k - y_i^k\| \\ &= \alpha_k \|w_i^k - w_i^{k-1}\| + \|\widehat{w}_i^k - y_i^k\|, \quad i = 1, \dots, n, \quad [\text{por (3.13) e (3.22)}]. \end{aligned} \quad (3.41)$$

Dado que $z^k - z^{k-1} \rightarrow 0$ e $w_i^k - w_i^{k-1} \rightarrow 0$ para todo $i = 1, \dots, n$ (como consequência da Proposição 3.1.5(c) em conjunto com as Eqs. (3.1), (3.17) e (3.20)) e do fato de que $\{\alpha_k\}$ é limitada, o resultado desejado segue diretamente de (3.40), (3.41) e do item (d). $\square$

Lema 3.1.11. *Todo ponto de acumulação fraco da sequência $\{p^k\}$ definida em (3.17) pertence ao conjunto $\mathcal{S}_e$.*

Demonstração. Seja $\bar{p} = (\bar{z}, \bar{w}_1, \dots, \bar{w}_{n-1}) \in \mathcal{H}$ um ponto de acumulação fraco da sequência $\{p^k\}$ (lembre que, pela Proposição 3.1.5(b), essa sequência é limitada). Assim, existe uma subsequência $\{p^{k_j}\} \subseteq \{p^k\}$ tal que $p^{k_j} \rightharpoonup \bar{p}$, isto é,

$$z^{k_j} \rightharpoonup \bar{z} \quad \text{e} \quad w_i^{k_j} \rightharpoonup \bar{w}_i, \quad i = 1, \dots, n-1. \quad (3.42)$$

Utilizando a Proposição 3.1.10(e), (3.42) e o fato de que $G_n = I$ (ver Hipótese A2), concluímos também que

$$x_n^{k_j} \rightharpoonup \bar{z} \quad \text{e} \quad y_i^{k_j} \rightharpoonup \bar{w}_i, \quad i = 1, \dots, n-1.$$

Pela definição do Algoritmo 5, sabemos que $y_i^{k_j} \in T_i^{[\varepsilon_i^{k_j}]}(x_i^{k_j})$ para todo $1 \leq i \leq n$. Além disso, segundo a Proposição 3.1.10 (itens (a) e (b)), temos

$$x_i^{k_j} - G_i x_n^{k_j} \rightarrow 0 \quad (1 \leq i \leq n-1), \quad y_n^{k_j} + \sum_{i=1}^{n-1} G_i^* y_i^{k_j} \rightarrow 0, \quad \varepsilon_i^{k_j} \rightarrow 0.$$

Assim, pelo Corolário B.2.3, obtemos

$$\bar{w}_i \in T_i(G_i \bar{z}), \quad i = 1, \dots, n-1, \quad - \sum_{i=1}^{n-1} G_i^* \bar{w}_i \in T_n(\bar{z}),$$

o que, pela definição do conjunto $\mathcal{S}_e$ dada em (3.4), implica que $\bar{p} \in \mathcal{S}_e$. Portanto, todo ponto de acumulação fraco da sequência $\{p^k\}$ pertence a $\mathcal{S}_e$. $\square$

Teorema 3.1.12 (Convergência forte do Algoritmo 5). *Seja a sequência $\{p^k\}$ gerada pelo Algoritmo 5, conforme definido em (3.17), e seja o conjunto de soluções estendido $\mathcal{S}_e$ dado em (3.4). Suponha que as Hipóteses A1–A3, 3.1.6 e 2.2.2 sejam satisfeitas. Além disso, considere a sequência $\{w_n^k\}$ conforme definida em (3.20).*

Então $\{p^k\}$ converge fortemente para $p^ := (z^*, w_1^*, \dots, w_{n-1}^*)$, onde p^* é a projeção ortogonal de p^0 sobre $\mathcal{S}_e$. A sequência $\{z^k\}$ converge fortemente para z^* , que é um ponto solução do Problema (3.2), e $\{w_i^k\}$ converge fortemente para w_i^* , para todo $1 \leq i \leq n$, com $w_n^* := -\sum_{i=1}^{n-1} G_i^* w_i^*$. Além disso, para $i = 1, \dots, n$, as sequências $\{x_i^k\}$ e $\{y_i^k\}$ convergem fortemente para $G_i z^*$ e w_i^* , respectivamente.*

Demonstração. Passo 1. Primeiramente, provamos que $p^k \rightarrow p^*$. De fato, pela Proposição 3.1.5(b) e pelo fato que $d_{0,\gamma} = \|p^* - p^0\|_\gamma$, tem-se

$$\|p^k - p^0\|_\gamma \leq \|p^* - p^0\|_\gamma \quad \text{para todo } k \geq 0. \quad (3.43)$$

Assim, $\{p^k\}$ é limitada, e, pelo Lema 3.1.11, seus pontos de acumulação fraco pertencem a $\mathcal{S}_e$. Seja $\{p^{k_j}\}$ uma subsequência de $\{p^k\}$ que converge fracamente para algum $\bar{p} \in \mathcal{S}_e$:

$$\bar{p} \in \mathcal{S}_e, \quad p^{k_j} \rightharpoonup \bar{p}.$$

Como a norma $\|\cdot\|_\gamma$ é semi-contínua inferiormente com respeito à topologia fraca em $\mathcal{H}$, obtemos

$$\|\bar{p} - p^0\|_\gamma \leq \liminf \|p^{k_j} - p^0\|_\gamma \leq \limsup \|p^{k_j} - p^0\|_\gamma \leq \|p^* - p^0\|_\gamma, \quad (3.44)$$

onde a última desigualdade segue de (3.43). Como p^* é a projeção ortogonal de p^0 sobre $\mathcal{S}_e$ e $\bar{p} \in \mathcal{S}_e$, a desigualdade (3.44) implica que $\bar{p} = p^*$. Logo, $p^k \rightharpoonup p^*$ e de (3.44), temos

$$\|p^* - p^0\|_\gamma \leq \liminf \|p^k - p^0\|_\gamma \leq \limsup \|p^k - p^0\|_\gamma \leq \|p^* - p^0\|_\gamma.$$

Portanto, $\|p^k - p^0\|_\gamma \rightarrow \|p^* - p^0\|_\gamma$, e como $p^k - p^0 \rightharpoonup p^* - p^0$, segue que $p^k - p^0 \rightarrow p^* - p^0$, ou seja, $p^k \rightarrow p^*$.

Passo 2. As convergências $z^k \rightarrow z^*$ e $w_i^k \rightarrow w_i^*$ para $1 \leq i \leq n$ decorrem diretamente do fato de que $p^k \rightarrow p^*$ e das definições $p^k = (z^k, w_1^k, \dots, w_{n-1}^k)$, $p^* = (z^*, w_1^*, \dots, w_{n-1}^*)$, $w_n^k = -\sum_{i=1}^{n-1} G_i^* w_i^k$ e $w_n^* = -\sum_{i=1}^{n-1} G_i^* w_i^*$. Resta provar que $\{x_i^k\}$ e $\{y_i^k\}$ convergem fortemente para $G_i z^*$ e w_i^* , respectivamente, para cada $i = 1, \dots, n$. Essa afirmação segue diretamente da Proposição 3.1.10(e) e das convergências $z^k \rightarrow z^*$ e $w_i^k \rightarrow w_i^*$ para todo $i = 1, \dots, n$. $\square$

3.1.2 Análise da complexidade iterativa do Algoritmo 5

Nesta subseção, investigamos a complexidade iterativa do Algoritmo 5. Para isso, introduzimos a seguinte noção de solução aproximada para o problema (3.2): $\{x_i\}_{i=1}^n$, $\{y_i\}_{i=1}^n$ e $\{\varepsilon_i\}_{i=1}^n$, onde $x_i, y_i \in \mathcal{H}_i$ e $\varepsilon_i \in \mathbb{R}_+$ para $i = 1, \dots, n$, constituem uma solução aproximada de (3.2) com tolerância $\rho > 0$ se satisfazem

$$y_i \in T_i^{[\varepsilon_i]}(x_i), \quad i = 1, \dots, n, \quad \sum_{i=1}^n \varepsilon_i \leq \rho, \quad \left\| \sum_{i=1}^{n-1} G_i^* y_i + y_n \right\| \leq \rho, \quad (3.45)$$

$$\|x_i - G_i x_n\| \leq \rho, \quad i = 1, \dots, n-1.$$

É fácil verificar que, se fizermos $\rho = 0$ em (3.45), obtemos que x_n é uma solução exata do problema (3.2).

A seguir, apresentamos o resultado de complexidade iterativa, o qual fornece uma estimativa do número de iterações que o Algoritmo 5 necessita para obter vetores $\{x_i\}_{i=1}^n$, $\{y_i\}_{i=1}^n$ e $\{\varepsilon_i\}_{i=1}^n$ que satisfaçam as condições em (3.45), para uma tolerância pré-fixada $\rho > 0$.

Teorema 3.1.13 (Complexidade iterativa pontual do Algoritmo 5). *Considere as sequências geradas pelo Algoritmo 5. Seja $d_{0,\gamma}$ a distância entre p^0 e o conjunto de soluções estendido S_e , conforme definido em (3.4), e seja $\Omega_{(\alpha,\beta)} > 0$ a constante dada em (3.36). Suponha ainda que as Hipóteses A1–A3, 3.1.6 e 2.2.2 sejam satisfeitas. Então, para todo $k \geq 0$, existe um índice $\ell \in \{0, 1, \dots, k\}$ tal que*

$$y_i^\ell \in T_i^{[\varepsilon_i^\ell]}(x_i^\ell), \quad i = 1, \dots, n, \quad (3.46)$$

$$\left\| \sum_{i=1}^{n-1} G_i^* y_i^\ell + y_n^\ell \right\| \leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) d_{0,\gamma}}{\sqrt{k+1}} \left(\frac{4\sqrt{\gamma} \max\{1, 1/\gamma\} \sqrt{\Omega_{(\alpha,\beta)}}}{(1-\sigma^2) \min\{\underline{\lambda}, 1/\bar{\lambda}\}} \right), \quad (3.47)$$

$$\|x_i^\ell - G_i x_n^\ell\| \leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) d_{0,\gamma}}{\sqrt{k+1}} \left(\frac{4 \max\{1, 1/\gamma\} \sqrt{\Omega_{(\alpha,\beta)}}}{(1-\sigma^2) \min\{\underline{\lambda}, 1/\bar{\lambda}\}} \right), \quad i = 1, \dots, n-1, \quad (3.48)$$

$$\sum_{i=1}^n \varepsilon_i^\ell \leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) d_{0,\gamma}^2}{k+1} \left(\frac{4\sigma^2 \max\{1, 1/\gamma\} \Omega_{(\alpha,\beta)}}{(1-\sigma^2)^2 \min\{\underline{\lambda}, 1/\bar{\lambda}\}} \right). \quad (3.49)$$

Demonstração. A aplicação direta da primeira desigualdade do item (a) e do item (b) do Lema 3.1.8 implica, para todo $j \geq 0$,

$$\Delta_j := \max \left\{ \frac{\|\nabla \varphi_j\|_\gamma^2}{c^2}, \frac{1-\sigma^2}{c\sigma^2} \sum_{i=1}^n \varepsilon_i^j \right\} \leq \|p^{j+1} - \tilde{p}^j\|_\gamma^2 \quad (3.50)$$

que, por sua vez, combinada com a Proposição 3.1.9, fornece a estimativa

$$\sum_{j=0}^k \Delta_j \leq \Omega_{(\alpha,\beta)} d_{0,\gamma}^2,$$

onde $\Omega_{(\alpha,\beta)}$ é a constante definida em (3.36). Escolhendo então um índice $\ell \in \{0, \dots, k\}$ tal que $\Delta_\ell = \min_{j=0, \dots, k} \Delta_j$, obtemos

$$(k+1)\Delta_\ell \leq \Omega_{(\alpha,\beta)} d_{0,\gamma}^2. \quad (3.51)$$

A partir de (3.50) e (3.51), concluímos que

$$\|\nabla \varphi_\ell\|_\gamma^2 \leq \frac{c^2 \Omega_{(\alpha,\beta)} d_{0,\gamma}^2}{k+1} \quad \text{e} \quad \sum_{i=1}^n \varepsilon_i^\ell \leq \frac{c\sigma^2 \Omega_{(\alpha,\beta)} d_{0,\gamma}^2}{(1-\sigma^2)(k+1)}. \quad (3.52)$$

(Lembramos que c está definido em (3.31)). Agora, combinando a primeira desigualdade de (3.52) com (3.10), obtemos também

$$\gamma^{-1} \left\| \sum_{i=1}^{n-1} G_i^* y_i^\ell + y_n^\ell \right\|^2 + \sum_{i=1}^{n-1} \|x_i^\ell - G_i x_n^\ell\|^2 \leq \frac{c^2 \Omega_{(\alpha, \beta)} d_{0, \gamma}^2}{k+1},$$

o que, após manipulações algébricas simples, nos leva às estimativas em (3.47) e (3.48). Note também que (3.49) segue diretamente da segunda desigualdade em (3.52). Para concluir a demonstração do teorema, observamos que a inclusão (3.46) decorre diretamente da definição do Algoritmo 5 (ver as inclusões em (3.15)). $\square$

Observação 3.1.14. Para concluir esta seção, vamos mostrar, seguindo a metodologia apresentada por Pentón Machado e Sicre em [80], que o Algoritmo 5 pode ser visto como um caso particular do Algoritmo 1. De fato, considere as aplicações

$$F(x, y_1, \dots, y_{n-1}) := \left(\frac{1}{\gamma} \sum_{i=1}^{n-1} G_i^* y_i, -G_1 x, \dots, -G_{n-1} x \right)$$

e

$$G(x, y_1, \dots, y_{n-1}) := \frac{1}{\gamma} T_n(x) \times T_1^{-1}(y_1) \times \dots \times T_{n-1}^{-1}(y_{n-1}).$$

Com base na prova de [80, Proposition 3.3] (veja também [80, Lemma 2.2]), é fácil mostrar que $(q^k, \nabla \varphi_k, \varepsilon^k)$ é uma $\hat{\sigma}$ -solução aproximada (no sentido da Definition 2.1.1) de $T = F + G$ em $(\tilde{p}^k, \lambda_k \equiv 1)$, onde

$$q^k = (x_n^k, y_1^k, \dots, y_{n-1}^k), \quad \varepsilon^k = \sum_{i=1}^n \varepsilon_i^k, \quad \hat{\sigma} = \sqrt{1 - \frac{(1 - \sigma^2) \min\{\bar{\lambda}^{-1}, \underline{\lambda}\}}{2\delta}}$$

com

$$\delta := \max \left\{ \frac{nD}{\gamma} + 1, 2nD + \gamma \right\} \quad \text{e} \quad D := \max_{1 \leq i \leq n} \|G_i\|^2.$$

Portanto, o Algoritmo 5 é de fato um caso particular do Algoritmo 1 para $T = F + G$, com $\sigma = \hat{\sigma}$ e $\lambda_k \equiv 1$. Assim, pelo Teorema 2.2.7, obtemos as seguintes estimativas: para cada $k \in \mathbb{N}$, existe $\ell \in \{0, \dots, k\}$ tal que

$$\left\| \sum_{i=1}^{n-1} G_i^* y_i^\ell + y_n^\ell \right\| \leq 2\sqrt{\gamma} \frac{d_{0, \gamma}}{\sqrt{k+1}} \frac{\sqrt{\Omega_{(\alpha, \beta)}}}{(1 - \hat{\sigma}^2)}, \quad (3.53)$$

$$\|x_i^\ell - G_i x_n^\ell\| \leq 2\sqrt{n-1} \frac{d_{0, \gamma}}{\sqrt{k+1}} \frac{\sqrt{\Omega_{(\alpha, \beta)}}}{(1 - \hat{\sigma}^2)}, \quad i = 1, \dots, n-1, \quad (3.54)$$

$$\sum_{i=1}^n \varepsilon_i^\ell \leq 4 \hat{\sigma}^2 \frac{d_{0, \gamma}^2}{k+1} \frac{\Omega_{(\alpha, \beta)}}{(1 - \hat{\sigma}^2)^2}. \quad (3.55)$$

Observe que, enquanto a cota em (3.48) é de ordem $O(n)$, seu correspondente em (3.54) é de ordem $O(n\sqrt{n})$ (já que $\hat{\sigma} = O(1/\sqrt{n})$, todas as outras são de ordem $O(n)$), o que mostra que usar a estrutura de decomposição desenvolvida nesta seção nos permite obter cotas mais refinadas para a complexidade iterativa.

3.2 Variantes do algoritmo inercial fortemente convergente de decomposição projetivo com passos do tipo forward-backward e forward-backward-forward

Nesta seção, estudamos uma versão estruturada do problema de inclusão monótono (3.2), na qual cada operador monótono maximal T_i admite uma decomposição do tipo: $T_i = F_i + B_i$ para todo $i = 1, \dots, n$. Mais especificamente, consideramos o problema de inclusão

$$0 \in \sum_{i=1}^n G_i^* \underbrace{(F_i + B_i)}_{T_i} G_i(z), \quad (3.56)$$

com $n \geq 2$ sob as seguintes condições:

- B1. $F_i: \text{Dom } F_i \subseteq \mathcal{H}_i \rightarrow \mathcal{H}_i$ é um operador ponto-ponto, monótono e contínuo.
- B2. $B_i: \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ é um operador monótono maximal tal que $\text{Dom } B_i \subseteq C_i \subseteq \text{Dom } F_i$, onde C_i é um subconjunto não-vazio, fechado e convexo de $\mathcal{H}_i$.
- B3. A suposição A2 da Seção 3.1 é válida para os operadores lineares e limitados G_i ($1 \leq i \leq n$).
- B4. O conjunto de soluções de (3.56) é não-vazio.

Sob hipóteses razoáveis sobre as funções $f_i(\cdot)$ e $g_i(\cdot)$ ($1 \leq i \leq n$), o problema (3.56) envolve o problema de minimização convexa em soma finita

$$\min_{x \in \mathcal{H}_0} \sum_{i=1}^n f_i(G_i x) + g_i(G_i x),$$

em que, para cada $i = 1, \dots, n$, $f_i: \mathcal{H}_i \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma função própria, convexa e semi-contínua inferiormente, e $g_i: \mathcal{H}_i \rightarrow \mathbb{R}$ é uma função convexa suficientemente suave (ver [26, 41]).

Observação 3.2.1. As suposições B1 e B2 garantem que $T_i = F_i + B_i$ ($1 \leq i \leq n$) é monótono maximal (ver, por exemplo, [66]). Combinando isso com as hipóteses B3 e B4, temos que as condições das suposições A1–A3 da Seção 3.1 são satisfeitas no contexto desta seção. Portanto, os resultados da Seção 3.1 podem ser aplicados ao problema (3.56).

Supondo que os resolventes $(B_i + I)^{-1}$ dos operadores $B_i(\cdot)$ ($1 \leq i \leq n$) sejam de fácil computação, propomos nesta seção duas variantes do Algoritmo 5 para resolver numericamente o problema (3.56), a saber:

1. Uma primeira variante com passos do tipo forward-backward para os subproblemas, assumindo que $F_i(\cdot)$ é L_i^{-1} -cocoercivo no domínio de F_i , ou seja, para todo $i = 1, \dots, n$,

$$\langle F_i(x) - F_i(y), x - y \rangle \geq \frac{1}{L_i} \|F_i(x) - F_i(y)\|^2 \quad \text{para todo } x, y \in \text{Dom } F_i, \quad (3.57)$$

com $L_i > 0$. O algoritmo correspondente é apresentado como Algoritmo 6 e os principais resultados sobre sua convergência e complexidade iterativa estão resumidos no Teorema 3.2.5.

2. Uma segunda variante com passos do tipo forward-backward-forward (no estilo de Tseng), para o caso em que $F_i(\cdot)$ é L_i -Lipschitz contínua, ou seja, para todo $i = 1, \dots, n$,

$$\|F_i(x) - F_i(y)\| \leq L_i\|x - y\| \quad \text{para todo } x, y \in \text{Dom } F_i, \quad (3.58)$$

com $L_i > 0$. O método é descrito no Algoritmo 7, e os principais resultados sobre sua convergência forte e complexidade iterativa são apresentados no Teorema 3.2.9.

3.2.1 Uma variante do algoritmo inercial de decomposição projetivo com convergência forte e passos do tipo forward-backward

Nesta subseção, assumindo que as condições B1–B4 e a desigualdade (3.57) são satisfeitas, propomos o Algoritmo 6 para resolver numericamente o problema (3.56).

A seguir, alguns comentários sobre o Algoritmo 6:

- (i) O Algoritmo 6 pode ser visto como um caso particular do Algoritmo 5 aplicado à inclusão estrutural (3.56), onde a tripla $(x_i^k, y_i^k, \varepsilon_i^k)$, para $1 \leq i \leq n$, que satisfaz a condição de erro relativo (3.15) é explicitamente computada conforme descrito em (3.62) e (3.64); ver a Proposição 3.2.2 mais adiante. Ressaltamos que a estratégia de aproveitar propriedades estruturais em métodos de decomposição de operadores por meio de passos forward-backward remonta, ao menos, a [55, 78], aparecendo também em trabalhos mais recentes (ver, por exemplo, [67]) no contexto de algoritmos do ponto proximal inexatos com erro relativo moderno.
- (ii) Por se tratar de um caso particular do Algoritmo 5, quase todos os comentários e observações feitos antes ou logo após aquele algoritmo também são válidos para o Algoritmo 6 (ver os itens (i)–(vi) após o Algoritmo 5).
- (iii) Dado que se assume que os resolventes $(B_i + I)^{-1}$ de $B_i(\cdot)$, para $1 \leq i \leq n$, podem ser computados de forma eficiente, a obtenção de x_i^k conforme (3.62) depende, basicamente, da avaliação do operador linear G_i em $\tilde{z}^k$ e do operador monótono $F_i(\cdot)$ em $\tilde{z}_i^k$. A principal motivação para introduzir a projeção ortogonal $\tilde{z}_i^k$ é recuperar a viabilidade de $G_i\tilde{z}^k$ em relação ao domínio de $F_i(\cdot)$ conforme exigido na Hipótese B2. No caso em que $\text{Dom } F_i = \mathcal{H}_i$, ou seja, quando o domínio de F_i é o espaço inteiro, poderíamos simplesmente definir $\tilde{z}_i^k = G_i\tilde{z}^k$, eliminando assim a necessidade de projetar sobre o conjunto C_i na implementação do Algoritmo 6.

A seguir, mostramos que o Algoritmo 6 é de fato um caso particular do Algoritmo 5.

Proposição 3.2.2. *Sejam (x_i^k, y_i^k) e ε_i^k ($1 \leq i \leq n$) definidos como em (3.62) e (3.64), respectivamente, e sejam $\tilde{z}^k$ e $\tilde{w}_i^k$ ($1 \leq i \leq n$) conforme em (3.59)–(3.61). Considere também $\lambda_i = 2\sigma^2/L_i$ como na entrada do Algoritmo 6. Então, para todo $i = 1, \dots, n$ e $k \geq 0$, temos:*

$$\begin{aligned} y_i^k &\in (F_i^{[\varepsilon_i^k]} + B_i)(x_i^k) \subseteq (F_i + B_i)^{[\varepsilon_i^k]}(x_i^k), \\ \lambda_i y_i^k + x_i^k - (G_i \tilde{z}^k + \lambda_i \tilde{w}_i^k) &= 0, \quad 2\lambda_i \varepsilon_i^k \leq \sigma^2 \|G_i \tilde{z}^k - x_i^k\|^2. \end{aligned} \quad (3.68)$$

Em particular, o Algoritmo 6 é um caso particular do Algoritmo 5 (com $\lambda_i^k \equiv \lambda_i = 2\sigma^2/L_i$ para todo $1 \leq i \leq n$) para resolver (3.56).

Algoritmo 6: Uma variante do algoritmo inercial de decomposição projetivo com convergência forte e passos forward-backward para resolução de (3.56).

Entrada: $(z^{-1}, w_1^{-1}, \dots, w_{n-1}^{-1}) = (z^0, w_1^0, \dots, w_{n-1}^0) \in \mathcal{H}$, escalares $0 < \sigma < 1$ e $\gamma > 0$. Para $i = 1, \dots, n$, defina $\lambda_i = 2\sigma^2/L_i$.

para $k = 0, 1, \dots$ **faça**

Escolha $\alpha_k \geq 0, \beta_k \geq 0$ e defina

$$\hat{z}^k = z^k + \alpha_k(z^k - z^{k-1}), \quad \tilde{z}^k = \hat{z}^k + \beta_k(\hat{z}^k - z^0), \quad (3.59)$$

$$\hat{w}_i^k = w_i^k + \alpha_k(w_i^k - w_i^{k-1}), \quad \tilde{w}_i^k = \hat{w}_i^k + \beta_k(\hat{w}_i^k - w_i^0) \quad \forall i = \overline{1, n-1}, \quad (3.60)$$

$$\tilde{w}_n^k = - \sum_{i=1}^{n-1} G_i^* \tilde{w}_i^k. \quad (3.61)$$

Para $i = 1, \dots, n$, seja $\bar{z}_i^k := P_{C_i}(G_i \tilde{z}^k)$ e compute

$$\begin{cases} x_i^k := (\lambda_i B_i + I)^{-1} (G_i \tilde{z}^k + \lambda_i \tilde{w}_i^k - \lambda_i F_i(\bar{z}_i^k)), \\ y_i^k := \frac{1}{\lambda_i} (G_i \tilde{z}^k + \lambda_i \tilde{w}_i^k - x_i^k). \end{cases} \quad (3.62)$$

se $x_i^k = G_i x_n^k$ ($i = 1, \dots, n-1$) e $\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$ **então**

retorna

$$(\bar{z}, \bar{w}_1, \dots, \bar{w}_{n-1}) := (x_n^k, y_1^k, \dots, y_{n-1}^k). \quad (3.63)$$

fim

Para $i = 1, \dots, n$, defina

$$\varepsilon_i^k := \frac{L_i \|x_i^k - \bar{z}_i^k\|^2}{4}. \quad (3.64)$$

Seja o funcional afim $\varphi_k(\cdot)$ definido conforme (3.7) (ver também (3.8)), defina

$$p^k := (z^k, w_1^k, \dots, w_{n-1}^k), \quad (3.65)$$

$$\begin{aligned} H_k &= \{p \in \mathcal{H} \mid \varphi_k(p) \leq 0\}, \\ W_k &= \{p \in \mathcal{H} \mid \langle p^0 - p^k, p - p^k \rangle_\gamma \leq 0\}, \end{aligned} \quad (3.66)$$

e atualize

$$(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1}) = P_{H_k \cap W_k}(p^0). \quad (3.67)$$

fim

Demonstração. A partir de (3.62), a definição do resolvente $(\lambda_i B_i + I)^{-1}$ e simples manipulações algébricas, obtemos que $y_i^k \in F_i(\bar{z}_i^k) + B_i(x_i^k)$. Combinando essa inclusão com o Lema 1.2.11 e a definição de ε_i^k dada em (3.64), resulta a primeira inclusão em (3.68). A segunda inclusão decorre diretamente da primeira e da Proposição 1.2.10((b) e (d)). A identidade presente em (3.68) segue diretamente da definição de y_i^k dada em (3.62). Resta agora provar a desigualdade. Para isso, usamos (3.64), a definição de $\lambda_i = 2\sigma^2/L_i$ e simples manipulações, obtendo

$$2\lambda_i \varepsilon_i^k \leq \sigma^2 \|x_i^k - \bar{z}_i^k\|^2 \leq \sigma^2 \|G_i \tilde{z}^k - x_i^k\|^2,$$

sendo que a segunda desigualdade resulta da definição $\bar{z}_i^k := P_{C_i}(G_i \tilde{z}^k)$ e da inclusão $x_i^k \in C_i$, pois $x_i^k \in \text{Dom } B_i \subseteq C_i$. A última afirmação da proposição decorre de (3.68) e das definições dos Algoritmos 5 e 6. $\square$

Observação 3.2.3. À luz da Proposição 3.2.2 e do Corolário 3.1.3, assumiremos a partir deste ponto da subseção que o Algoritmo 6 não para.

Observação 3.2.4. Considerando a definição de $\lambda_i^k \equiv \lambda_i = 2\sigma^2/L_i$ utilizada no Algoritmo 6, verifica-se que a condição da Hipótese 3.1.6 é automaticamente satisfeita com

$$\underline{\lambda} := \frac{2\sigma^2}{\bar{L}} \quad \text{e} \quad \bar{\lambda} := \frac{2\sigma^2}{\underline{L}}, \quad (3.69)$$

onde definimos

$$\bar{L} := \max_{i=1,\dots,n} L_i \quad \text{e} \quad \underline{L} := \min_{i=1,\dots,n} L_i. \quad (3.70)$$

A seguir, resumimos os principais resultados sobre a convergência forte e a complexidade iterativa do Algoritmo 6.

Teorema 3.2.5 (Convergência forte e complexidade iterativa do Algoritmo 6). *Considere as sequências geradas pelo Algoritmo 6, e seja $d_{0,\gamma}$ a distância do ponto inicial p^0 ao conjunto solução estendido $\mathcal{S}_e \neq \emptyset$ de (3.56). Sejam ainda $\bar{L}, \underline{L} > 0$ definidos conforme (3.70) e $\Omega_{(\alpha,\beta)} > 0$ como em (3.36). Suponha que as hipóteses B1–B4 e 2.2.2 são válidas (ver também as Observações 3.2.1 e 3.2.4), e que a condição de cocoercividade (3.57) é satisfeita para cada operador $F_i(\cdot)$, com $1 \leq i \leq n$. Então valem as seguintes afirmações*

- (a) (Convergência forte) *Todos os resultados do Teorema 3.1.12 permanecem válidos para o Algoritmo 6.*

(b) (Complexidade iterativa) Para todo $k \geq 0$, existe um índice $\ell \in \{0, 1, \dots, k\}$ tal que

$$\left\{ \begin{array}{l} y_i^\ell \in (F_i^{[\varepsilon_i^\ell]} + B_i)(x_i^\ell), \quad i = 1, \dots, n, \\ \left\| \sum_{i=1}^{n-1} G_i^* y_i^\ell + y_n^\ell \right\| \leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) \max \left\{ \frac{\bar{L}}{2\sigma^2}, \frac{2\sigma^2}{\underline{L}} \right\} d_{0,\gamma}}{\sqrt{k+1}} \sqrt{\gamma} MR, \\ \left\| x_i^\ell - G_i x_n^\ell \right\| \leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) \max \left\{ \frac{\bar{L}}{2\sigma^2}, \frac{2\sigma^2}{\underline{L}} \right\} d_{0,\gamma}}{\sqrt{k+1}} MR, \quad i = 1, \dots, n, \\ \sum_{i=1}^n \varepsilon_i^\ell \leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) \max \left\{ \frac{\bar{L}}{2\sigma^2}, \frac{2\sigma^2}{\underline{L}} \right\} d_{0,\gamma}^2}{k+1} \sigma^2 MR^2, \end{array} \right.$$

onde,

$$M = 4 \max\{1, 1/\gamma\} \quad \text{e} \quad R = \frac{\sqrt{\Omega_{(\alpha, \beta)}}}{1 - \sigma^2}. \quad (3.71)$$

Demonstração. O resultado decorre diretamente da Proposição 3.2.2, das Observações 3.2.1, 3.2.3 e 3.2.4, e dos Teoremas 3.1.12 e 3.1.13. $\square$

3.2.2 Uma variante do algoritmo inercial de decomposição projetivo fortemente convergente com passos do tipo forward-backward-forward de Tseng

Nesta subseção, assumindo que as condições B1–B4 e (3.58) são satisfeitas, propomos uma variante do Algoritmo 5 (veja o Algoritmo 7 mais adiante), especialmente adaptada para a inclusão estruturada (3.56).

A seguir, algumas observações a respeito do Algoritmo 7:

- (i) Assim como aconteceu com o Algoritmo 6, o Algoritmo 7 também é um caso particular do Algoritmo 5 (ver a Proposição 3.2.6 mais adiante). Mais precisamente, o Algoritmo 7 corresponde essencialmente ao Algoritmo 5 aplicado à inclusão monótona estruturada (3.56), no qual a tripla $(x_i^k, y_i^k, \varepsilon_i^k)$ ($1 \leq i \leq n$), que satisfaz a condição (3.15), é explicitamente computada segundo as fórmulas (3.75) e (3.77). A ideia de utilizar passos do tipo forward-backward-forward como em (3.75) foi introduzida por P. Tseng no artigo seminal [97] (ver também [67]).
- (ii) Assim como foi observado no caso do Algoritmo 6, uma vez que o Algoritmo 7 é um caso particular do Algoritmo 5, quase todos os comentários e observações (i)–(vi) feitos antes ou depois do Algoritmo 5 também se aplicam diretamente ao Algoritmo 7.
- (iii) O Algoritmo 7 é aplicável em um cenário mais geral para o problema de inclusão monótono (3.56), uma vez que a hipótese de Lipschitz continuidade (ver (3.58)) é menos restritiva do que a de cocoercividade (ver (3.57)). No entanto, isso exige, em comparação com (3.62), uma avaliação adicional do operador $F_i(\cdot)$ no ponto x_i^k .

Algoritmo 7: Uma variante do algoritmo inercial de decomposição projetivo fortemente convergente com passos do tipo forward-backward-forward para resolver (3.56).

Entrada: $(z^{-1}, w_1^{-1}, \dots, w_{n-1}^{-1}) = (z^0, w_1^0, \dots, w_{n-1}^0) \in \mathcal{H}$, escalares $0 < \sigma < 1$ e $\gamma > 0$. Para $i = 1, \dots, n$, defina $\lambda_i = \sigma/L_i$.

para $k = 0, 1, \dots$ **faça**

Escolha $\alpha_k \geq 0, \beta_k \geq 0$ e defina

$$\hat{z}^k = z^k + \alpha_k(z^k - z^{k-1}), \quad \tilde{z}^k = \hat{z}^k + \beta_k(\hat{z}^k - z^0), \quad (3.72)$$

$$\hat{w}_i^k = w_i^k + \alpha_k(w_i^k - w_i^{k-1}), \quad \tilde{w}_i^k = \hat{w}_i^k + \beta_k(\hat{w}_i^k - w_i^0) \quad \forall i = \overline{1, n-1}, \quad (3.73)$$

$$\tilde{w}_n^k = - \sum_{i=1}^{n-1} G_i^* \tilde{w}_i^k. \quad (3.74)$$

Para $i = 1, \dots, n$, defina $\bar{z}_i^k := P_{C_i}(G_i \tilde{z}^k)$ e compute

$$\begin{cases} x_i^k = (\lambda_i B_i + I)^{-1} (G_i \tilde{z}^k + \lambda_i \tilde{w}_i^k - \lambda_i F_i(\bar{z}_i^k)), \\ y_i^k = \frac{G_i \tilde{z}^k + \lambda_i \tilde{w}_i^k - x_i^k}{\lambda_i} + F_i(x_i^k) - F_i(\bar{z}_i^k). \end{cases} \quad (3.75)$$

se $x_i^k = G_i x_n^k$ ($i = 1, \dots, n-1$) e $\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$ **então**

retorna

$$(\bar{z}, \bar{w}_1, \dots, \bar{w}_{n-1}) := (x_n^k, y_1^k, \dots, y_{n-1}^k). \quad (3.76)$$

fim

Para $i = 1, \dots, n$, defina

$$\varepsilon_i^k := 0. \quad (3.77)$$

Considere a funcional afim $\varphi_k(\cdot)$ como em (3.7) (ver também (3.8)), defina

$$p^k := (z^k, w_1^k, \dots, w_{n-1}^k), \quad (3.78)$$

$$\begin{aligned} H_k &= \{p \in \mathcal{H} \mid \varphi_k(p) \leq 0\}, \\ W_k &= \{p \in \mathcal{H} \mid \langle p^0 - p^k, p - p^k \rangle_\gamma \leq 0\}, \end{aligned} \quad (3.79)$$

e atualize

$$(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1}) = P_{H_k \cap W_k}(p^0). \quad (3.80)$$

fim

O seguinte resultado mostra que o Algoritmo 7 também pode ser interpretado como um caso particular do Algoritmo 5.

Proposição 3.2.6. *Sejam (x_i^k, y_i^k) ($1 \leq i \leq n$) definidos conforme (3.75), e sejam $\tilde{z}^k$ e $\tilde{w}_i^k$ ($1 \leq i \leq n$) definidos conforme (3.72)–(3.74). Seja ainda $\lambda_i = \sigma/L_i$ como especificado na entrada do Algoritmo 7. Então, para todo $i = 1, \dots, n$ e $k \geq 0$, valem as seguintes relações:*

$$\begin{aligned} y_i^k &\in (F_i + B_i)(x_i^k), \\ \|\lambda_i y_i^k + x_i^k - (G_i \tilde{z}^k + \lambda_i \tilde{w}_i^k)\| &\leq \sigma \|G_i \tilde{z}_i^k - x_i^k\|. \end{aligned} \quad (3.81)$$

Em particular, o Algoritmo 7 é um caso especial do Algoritmo 5 (com $\lambda_i^k \equiv \lambda_i = \sigma/L_i$ para todo $i = 1, \dots, n$) aplicado à resolução de (3.56).

Demonstração. A demonstração segue a mesma linha de raciocínio da [11, Proposition 5.1]. $\square$

Observação 3.2.7. À luz da Proposição 3.2.6 e do Corolário 3.1.3, de forma análoga à Observação 3.2.3, assumiremos a partir deste ponto da subseção que o Algoritmo 7 não para.

Observação 3.2.8. De modo semelhante às definições em (3.69) e (3.70), introduzimos agora

$$\underline{\lambda} := \frac{\sigma}{\bar{L}} \quad \text{e} \quad \bar{\lambda} := \frac{\sigma}{\underline{L}}, \quad (3.82)$$

em que

$$\bar{L} := \max_{i=1, \dots, n} L_i \quad \text{e} \quad \underline{L} := \min_{i=1, \dots, n} L_i. \quad (3.83)$$

Recordamos que, nesta subseção, L_i ($1 \leq i \leq n$) representa a constante de Lipschitz continuidade da aplicação $F_i(\cdot)$ (veja (3.58)).

Teorema 3.2.9 (Convergência forte e complexidade iterativa do Algoritmo 7). *Considere as sequências geradas pelo Algoritmo 7, seja $d_{0,\gamma}$ a distância de p^0 ao conjunto de soluções estendido $\mathcal{S}_e \neq \emptyset$, e sejam $\bar{L}, \underline{L} > 0$ definidos como em (3.83). Considere também $\Omega_{(\alpha,\beta)} > 0$ conforme definido em (3.36). Suponha que as hipóteses B1–B4 e 2.2.2 sejam satisfeitas (ver também as Observações 3.2.1 e 3.2.4), e que a condição de Lipschitz continuidade (3.58) seja válida para cada $F_i(\cdot)$, $1 \leq i \leq n$. Então valem as seguintes afirmações:*

(a) (Convergência forte) *Todas as conclusões do Teorema 3.1.12 continuam válidas para o Algoritmo 7.*

(b) (Complexidade iterativa) *Para todo $k \geq 0$, existe um índice $\ell \in \{0, 1, \dots, k\}$ tal que*

$$\left\{ \begin{aligned} y_i^\ell &\in (F_i + B_i)(x_i^\ell), \quad i = 1, \dots, n, \\ \left\| \sum_{i=1}^{n-1} G_i^* y_i^\ell + y_n^\ell \right\| &\leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) \max \left\{ \frac{\bar{L}}{\sigma}, \frac{\sigma}{\underline{L}} \right\} d_{0,\gamma}}{\sqrt{k+1}} \sqrt{\gamma} M R, \\ \|x_i^\ell - G_i x_n^\ell\| &\leq \frac{n \left(\max_{i=1, \dots, n} \|G_i\|^2 \right) \max \left\{ \frac{\bar{L}}{\sigma}, \frac{\sigma}{\underline{L}} \right\} d_{0,\gamma}}{\sqrt{k+1}} M R, \quad \forall i = \overline{1, n-1}, \end{aligned} \right.$$

onde M e R são dados como em (3.71).

Demonstração. O resultado decorre diretamente da Proposição 3.2.6, das Observações 3.2.1, 3.2.7 e 3.2.8, e dos Teoremas 3.1.12 e 3.1.13. $\square$

Capítulo 4

Conclusões

Nesta tese, foi realizado o estudo e desenvolvimento de variantes inerciais com convergência forte de algoritmos proximais projetivos inexatos para a resolução de problemas de inclusão monótonos. Em particular, foram consideradas duas classes fundamentais de problemas: a primeira da forma $0 \in T(x)$, onde T é um operador monótono maximal, e a segunda, mais geral, da forma $0 \in \sum_{i=1}^n G_i^* T_i(G_i z)$, que modelam inclusões monótonas compostas por meio de operadores lineares contínuos. Em ambos casos, foram propostos novos algoritmos incorporando efeitos inerciais.

Na primeira parte desta tese, foi proposta uma versão inercial com convergência forte do algoritmo de Solodov e Svaiter [92, Algorithm 1], na qual o mecanismo iterativo foi modificado com a introdução de efeitos de inércia inspirados em trabalhos anteriores de Nesterov, Alvarez-Attouch, entre outros. Demonstrou-se que, sob condições razoáveis sobre os parâmetros do algoritmo, a sequência gerada converge fortemente para um ponto solução do problema. Além disso, foi obtida uma estimativa explícita da complexidade iterativa, isto é, do número de iterações necessárias para atingir uma solução aproximada dentro de uma tolerância previamente estabelecida. Esses resultados não apenas reforçam a validade do método proposto, como também mostram que o uso de critérios de erro relativo facilita sua implementação computacional.

Na segunda parte desta tese, foi estudada uma classe mais geral de inclusões monótonas, nas quais intervêm uma soma finita de operadores monótonos maximais compostos com operadores lineares contínuos. Para esse caso, foi proposta uma versão com convergência forte do método de decomposição projetivo introduzido por Eckstein e Svaiter. Esse método representa um avanço teórico, pois combina a capacidade do método de decomposição projetivo de explorar a estrutura do problema com a introdução de efeitos inerciais. Assim como no caso anterior, foi provada rigorosamente a convergência forte das iterações geradas e foram obtidas estimativas da complexidade iterativa, as quais dependem da tolerância desejada, da dimensão do problema (número de operadores monótonos maximais) e dos operadores lineares envolvidos.

As contribuições desta tese são tanto teóricas quanto metodológicas. Do ponto de vista teórico, reforçam-se os fundamentos da teoria de algoritmos inerciais no contexto de inclusões monótonas, ao se estabelecerem condições que garantem a convergência forte mesmo na presença de erros relativos. Metodologicamente, os algoritmos propostos ampliam o repertório de ferramentas disponíveis para a resolução de problemas de otimização, desigualdades variacionais, entre outros, especialmente em contextos nos quais a convergência forte é essencial,

como em aplicações que requerem soluções precisas. Finalmente, uma linha promissora de pesquisa futura consiste em estender os algoritmos aqui propostos a contextos mais gerais, tais como inclusões monótonas em espaços de Banach, ou inclusões não monótonas, nas quais as propriedades do operador exigem uma análise mais cuidadosa.

Apêndice A

Boa definição do Algoritmo 1 e alguns resultados auxiliares

A.1 Boa definição do Algoritmo 1

Neste seção do Apêndice A apresentamos a demonstração da boa definição do Algoritmo 1 — ou seja, mostramos que x_{k+1} , dado por (2.15), está bem definido graças a que o conjunto $H_k \cap W_k$ é não-vazio. Lembramos que estamos assumindo que o conjunto solução $\mathcal{S} := T^{-1}(0)$ do problema (2.11) é não-vazio. A prova da proposição a seguir segue a mesma linha de raciocínio de [92, Proposition 3].

Proposição A.1.1. *Suponha que o Algoritmo 1 atinge a iteração k e que $x_k \in M_{\mathcal{S}}(x_0)$, onde*

$$M_{\mathcal{S}}(x_0) := \{x \in \mathcal{H} \mid \langle z - x, x_0 - x \rangle \leq 0 \text{ para todo } z \in \mathcal{S}\}. \quad (\text{A.1})$$

Então valem as seguintes afirmações:

- (a) $\mathcal{S} \subseteq H_k \cap W_k$.
- (b) x_{k+1} está bem definido e $x_{k+1} \in M_{\mathcal{S}}(x_0)$.

Demonstração. (a) Como o Algoritmo 1 atinge a iteração k , os conjuntos H_k e W_k , como em (2.14), estão bem definidos. Além disso, utilizando a inclusão em (2.13) e a definição do ε -enlargement T^ε dada em (1.9), concluímos que $\mathcal{S} \subseteq H_k$. Por outro lado, assumimos que $x_k \in M_{\mathcal{S}}(x_0)$, o que, pela definição em (A.1), significa que

$$\langle z - x_k, x_0 - x_k \rangle \leq 0 \text{ para todo } z \in \mathcal{S},$$

isto é, $\mathcal{S} \subseteq W_k$ (ver (2.14)). Portanto, temos $\mathcal{S} \subseteq H_k \cap W_k$.

(b) Como $\mathcal{S} \neq \emptyset$, segue do item (a) que $H_k \cap W_k \neq \emptyset$, de modo que a próxima iteração x_{k+1} está bem definida (já que $H_k \cap W_k$ é trivialmente fechado e convexo). Como x_{k+1} é a projeção de x_0 sobre $H_k \cap W_k$, pela caracterização da projeção (ver (1.4)), temos

$$\langle z - x_{k+1}, x_0 - x_{k+1} \rangle \leq 0 \text{ para todo } z \in H_k \cap W_k.$$

Dado que $\mathcal{S} \subseteq H_k \cap W_k$, a desigualdade acima vale, em particular, para todo $z \in \mathcal{S}$, o que implica que $x_{k+1} \in M_{\mathcal{S}}(x_0)$. Assim, concluímos a demonstração da proposição. $\square$

Como consequência da proposição anterior, concluímos que todo o algoritmo está bem definido.

Corolário A.1.2 (Boa definição do Algoritmo 1). *O Algoritmo 1 está bem definido e gera sequências infinitas. Além disso,*

$$\mathcal{S} \subseteq H_k \cap W_k \quad \text{para todo } k \geq 0, \quad (\text{A.2})$$

onde $\mathcal{S} := T^{-1}(0) \neq \emptyset$ é o conjunto solução de (2.11).

Demonstração. Como $x_0 \in M_{\mathcal{S}}(x_0)$, o resultado segue por indução em $k \geq 0$, aplicando a Proposição A.1.1. $\square$

A.2 Resultados Auxiliares

Lema A.2.1. *Seja $\mathcal{H}$ um espaço de Hilbert real. Valem as seguintes afirmações:*

(a) *Para quaisquer $x, y \in \mathcal{H}$ e $t \in \mathbb{R}$, temos:*

$$\|tx + (1-t)y\|^2 = t\|x\|^2 + (1-t)\|y\|^2 - t(1-t)\|x - y\|^2.$$

(b) *Para quaisquer $x, y, z \in \mathcal{H}$, vale:*

$$2\langle x - y, x - z \rangle = \|x - y\|^2 + \|x - z\|^2 - \|y - z\|^2.$$

(c) *Para toda sequência $\{x_n\}$ e ponto x em $\mathcal{H}$, tem-se:*

$$x_n \rightarrow x \quad \text{se, e somente se,} \quad x_n \rightharpoonup x \quad \text{e} \quad \|x_n\| \rightarrow \|x\|.$$

Lema A.2.2. *Sejam $a, b \geq 0$ com $a + b > 0$ e $c \in \mathbb{R}_+$. Então*

$$\min\{as^2 + bt^2 \mid s, t \geq 0 \text{ e } s + t \geq c\} = \frac{ab}{a+b}c^2.$$

Lema A.2.3. *Seja $\{x_k\}$ uma sequência limitada em um espaço de Hilbert $\mathcal{H}$. Se $\{x_k\}$ possui apenas um ponto de acumulação fraco $\bar{x}$, então $x_k \rightharpoonup \bar{x}$.*

Apêndice B

Boa definição do Algoritmo 5 e resultados auxiliares adicionais

B.1 Boa definição do Algoritmo 5

A proposição a seguir assegura a boa definição do Algoritmo 5. Observe que a definição do próximo iterado $(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1})$, conforme (3.19), depende do fato que o conjunto fechado e convexo $H_k \cap W_k$ seja não-vazio. Além disso, lembramos que estamos assumindo que o conjunto de soluções estendidas $\mathcal{S}_e$ em (3.4) é não-vazio.

Proposição B.1.1. *Em toda iteração $k \geq 0$, se o Algoritmo 5 não para, então as seguintes condições são satisfeitas: $\mathcal{S}_e \subseteq H_k \cap W_k$, o próximo iterado $(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1})$, conforme definido em (3.19), está bem definido, e $\mathcal{S}_e \subseteq W_{k+1}$.*

Demonstração. A demonstração é feita por indução em $k \geq 0$. Para $k = 0$, suponha que o Algoritmo 5 não para. Nesse caso, os semi-espacos H_0 e W_0 , como definidos em (3.18), podem ser construídos. Pelo Lema 3.1.1(b), em conjunto com a definição de H_0 , obtemos que $\mathcal{S}_e \subseteq H_0$. Como, trivialmente, $W_0 = \mathcal{H}$, segue que $\mathcal{S}_e \subseteq H_0 \cap W_0$. Dado que o conjunto de soluções estendidas $\mathcal{S}_e$ é não-vazio, concluímos que $H_0 \cap W_0$ também é não-vazio, além de fechado e convexo. Assim, o próximo iterado $p^1 := (z^1, w_1^1, \dots, w_{n-1}^1) = P_{H_0 \cap W_0}(p^0)$ está bem definido e, pela Eq. (1.4) no Lema 1.1.1, temos que $\langle p^0 - p^1, p - p^1 \rangle_\gamma \leq 0$ para todo $p \in H_0 \cap W_0$. Em particular, esta desigualdade vale para todo $p \in \mathcal{S}_e$, pois $\mathcal{S}_e \subseteq H_0 \cap W_0$, o que implica que $\mathcal{S}_e \subseteq W_1$. Concluímos, então, que a proposição é válida para $k = 0$.

Agora, suponha que a proposição vale para algum $k \geq 0$ e mostremos que também vale para $k + 1$, assumindo que o algoritmo prossegue para a iteração $k + 1$, ou seja, que ele não para em $k \geq 0$: $\mathcal{S}_e \subseteq H_k \cap W_k$, o próximo iterado $(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1})$ como em (3.19) está bem definido, e $\mathcal{S}_e \subseteq W_{k+1}$. Então, se o Algoritmo 5 não para na iteração $k + 1$, segue do Lema 3.1.1(b) e da definição de H_{k+1} que $\mathcal{S}_e \subseteq H_{k+1}$. Portanto, $\mathcal{S}_e \subseteq H_{k+1} \cap W_{k+1}$ e, como $\mathcal{S}_e$ é não-vazio e $H_{k+1} \cap W_{k+1}$ é fechado e convexo, o próximo iterado $p^{k+2} := (z^{k+2}, w_1^{k+2}, \dots, w_{n-1}^{k+2}) = P_{H_{k+1} \cap W_{k+1}}(p^0)$ está bem definido. Além disso, temos que $\langle p^0 - p^{k+2}, p - p^{k+2} \rangle_\gamma \leq 0$ para todo $p \in H_{k+1} \cap W_{k+1}$, o que, em particular, vale para todo $p \in \mathcal{S}_e$, implicando que $\mathcal{S}_e \subseteq W_{k+2}$. Isso conclui o processo indutivo e, portanto, a prova da proposição. $\square$

B.2 Resultados auxiliares adicionais

Ao longo desta seção, $\mathcal{H}$ e $\mathcal{G}$ denotam espaços de Hilbert reais. O próximo resultado foi apresentado em [7], como uma generalização de um resultado anterior devido a H. H. Bauschke.

Lema B.2.1. *Seja $T: \mathcal{H} \rightrightarrows \mathcal{H}$ um operador monótono maximal e V um subespaço linear fechado de $\mathcal{H}$. Considere uma sequência $\{(x^k, u^k)\}$ em $\text{Gra } T^{[\varepsilon^k]}$ que converge fracamente para um elemento $(\bar{x}, \bar{u})$ em $\mathcal{H} \times \mathcal{H}$, e suponha que $\varepsilon^k \rightarrow 0$. Se $P_{V^\perp} x^k \rightarrow 0$ e $P_V u^k \rightarrow 0$ (onde P_V e $P_{V^\perp}$ denotam as projeções ortogonais sobre V e $V^\perp$, respectivamente), então $(\bar{x}, \bar{u}) \in (\text{Gra } T) \cap (V \times V^\perp)$ e $\langle x^k, u^k \rangle \rightarrow \langle \bar{x}, \bar{u} \rangle = 0$.*

O lema a seguir generaliza [2, Proposition 2.4], e sua demonstração, que apresentamos aqui, segue as mesmas ideias da prova daquela proposição.

Lema B.2.2. *Sejam $A: \mathcal{H} \rightrightarrows \mathcal{H}$ e $B: \mathcal{G} \rightrightarrows \mathcal{G}$ operadores monótonos maximais, e seja $G: \mathcal{H} \rightarrow \mathcal{G}$ um operador linear e limitado. Suponha que as sequências $\{(r^k, a^k)\}$ e $\{(s^k, b^k)\}$ pertencem a $\text{Gra } A^{[\varepsilon^{a,k}]}$ e $\text{Gra } B^{[\varepsilon^{b,k}]}$, respectivamente, tais que $r^k \rightharpoonup \bar{r} \in \mathcal{H}$ e $b^k \rightharpoonup \bar{b} \in \mathcal{G}$, com $\varepsilon^{a,k} \rightarrow 0$ e $\varepsilon^{b,k} \rightarrow 0$. Se $Gr^k - s^k \rightarrow 0$ e $a^k + G^* b^k \rightarrow 0$, então $(\bar{r}, -G^* \bar{b}) \in \text{Gra } A$ e $(G\bar{r}, \bar{b}) \in \text{Gra } B$.*

Demonstração. Aplicamos o Lema B.2.1 no espaço de Hilbert $\mathcal{H} \times \mathcal{G}$. Para isso, definimos

$$\begin{aligned} T &:= A \times B: \mathcal{H} \times \mathcal{G} \rightrightarrows \mathcal{H} \times \mathcal{G}, \quad T(z, v) = A(z) \times B(v), \\ V &:= \{(z, v) \in \mathcal{H} \times \mathcal{G} \mid v = Gz\}. \end{aligned} \tag{B.1}$$

Note que T é um operador monótono maximal e que V é um subespaço linear fechado de $\mathcal{H} \times \mathcal{G}$. Usando as hipóteses sobre as sequências $\{(r^k, a^k)\}$ e $\{(s^k, b^k)\}$, a definição de $T(\cdot)$ dada em (B.1) e o Lema 1.2.12, obtemos

$$\begin{aligned} (x^k, u^k) &\in \text{Gra } T^{[\varepsilon^k]} \quad \text{para todo } k \geq 0, \quad \text{onde} \\ x^k &:= (r^k, s^k), \quad u^k := (a^k, b^k), \quad \varepsilon^k := \varepsilon^{a,k} + \varepsilon^{b,k}. \end{aligned} \tag{B.2}$$

Suponhamos agora que

$$Gr^k - s^k \rightarrow 0 \quad \text{e} \quad a^k + G^* b^k \rightarrow 0. \tag{B.3}$$

Dado que $r^k \rightharpoonup \bar{r}$ e $b^k \rightharpoonup \bar{b}$, e que os operadores lineares G e G^* são fracamente contínuos, segue que $s^k \rightharpoonup G\bar{r}$ e $a^k \rightharpoonup -G^* \bar{b}$. Assim, usando ainda que $\varepsilon^{a,k} \rightarrow 0$ e $\varepsilon^{b,k} \rightarrow 0$, temos

$$x^k \rightharpoonup (\bar{r}, G\bar{r}) =: \bar{x}, \quad u^k \rightharpoonup (-G^* \bar{b}, \bar{b}) =: \bar{u}, \quad \text{e} \quad \varepsilon^k \rightarrow 0, \tag{B.4}$$

com x^k , u^k e ε^k definidos como em (B.2).

Verifiquemos agora as condições $x^k - P_V x^k \rightarrow 0$ e $P_V u^k \rightarrow 0$. De acordo com [1, Lemma 3.1], para todo $w = (z, v) \in \mathcal{H} \times \mathcal{G}$, temos

$$\begin{aligned} P_V w &= ((I + G^* G)^{-1}(z + G^* v), G(I + G^* G)^{-1}(z + G^* v)), \\ P_{V^\perp} w &= (G^*(I + GG^*)^{-1}(Gz - v), -(I + GG^*)^{-1}(Gz - v)). \end{aligned}$$

Combinando isso com (B.3), concluímos que

$$P_{V^\perp} x^k \rightarrow 0 \quad \text{e} \quad P_V u^k \rightarrow 0, \quad (\text{B.5})$$

onde x^k e u^k são como em (B.2). Assim, a partir de (B.2), (B.4), (B.5) e do Lema B.2.1, concluímos que $(\bar{x}, \bar{u}) \in \text{Gra } T$, ou seja, $\bar{u} \in T(\bar{x})$, o que implica — veja (B.4) — que

$$(-G^* \bar{b}, \bar{b}) \in (A \times B)(\bar{r}, G \bar{r}),$$

logo, $-G^* \bar{b} \in A(\bar{r})$ e $\bar{b} \in B(G \bar{r})$, encerrando a demonstração do lema. $\square$

No próximo corolário, consideramos que $\mathcal{H}_i$ ($1 \leq i \leq n$) são espaços de Hilbert reais, com $\mathcal{H}_n = \mathcal{H}_0$.

Corolário B.2.3. *Sejam $T_i: \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ operadores monótonos maximais para $i = 1, \dots, n$. Suponha que $\{(x_i^k, y_i^k)\}$ é uma sequência em $\text{Gra } T_i^{[\varepsilon_i^k]}$, com $\varepsilon_i^k \rightarrow 0$, e que $x_n^k \rightharpoonup \bar{z} \in \mathcal{H}_0$ e $y_i^k \rightharpoonup \bar{w}_i \in \mathcal{H}_i$ para todo $1 \leq i \leq n-1$. Além disso, para cada $i = 1, \dots, n-1$, seja $G_i: \mathcal{H}_0 \rightarrow \mathcal{H}_i$ um operador linear contínuo. Se*

$$G_i x_n^k - x_i^k \rightarrow 0 \quad \text{para } 1 \leq i \leq n-1 \quad \text{e} \quad y_n^k + \sum_{i=1}^{n-1} G_i^* y_i^k \rightarrow 0,$$

então vale

$$\bar{w}_i \in T_i(G_i \bar{z}), \quad \text{para } i = 1, \dots, n-1, \quad \text{e} \quad -\sum_{i=1}^{n-1} G_i^* \bar{w}_i \in T_n(\bar{z}). \quad (\text{B.6})$$

Demonstração. Seja $\mathcal{H} := \mathcal{H}_0$ e $\mathcal{G} := \mathcal{H}_1 \times \dots \times \mathcal{H}_{n-1}$, equipada com o produto interno cartesiano usual. Definimos os operadores monótonos maximais $A: \mathcal{H} \rightrightarrows \mathcal{H}$ e $B: \mathcal{G} \rightrightarrows \mathcal{G}$ por

$$A := T_n, \quad B := T_1 \times \dots \times T_{n-1}.$$

Definimos também o operador linear contínuo $G: \mathcal{H} \rightarrow \mathcal{G}$ por $G = (G_1, \dots, G_{n-1})$, ou seja,

$$Gz := (G_1 z, \dots, G_{n-1} z) \quad \text{para todo } z \in \mathcal{H}.$$

Agora, consideramos as seguintes sequências:

$$\begin{aligned} a^k &:= y_n^k, \quad \varepsilon^{a,k} := \varepsilon_n^k, \quad r^k := x_n^k, \\ b^k &:= (y_1^k, \dots, y_{n-1}^k), \quad \varepsilon^{b,k} := \sum_{i=1}^{n-1} \varepsilon_i^k, \quad s^k := (x_1^k, \dots, x_{n-1}^k). \end{aligned}$$

Com essas definições e aplicando o Lema 1.2.12, concluímos que $\{(r^k, a^k)\}$ e $\{(s^k, b^k)\}$ são sequências em $\text{Gra } A^{[\varepsilon^{a,k}]}$ e $\text{Gra } B^{[\varepsilon^{b,k}]}$, respectivamente. Pelas hipóteses do corolário, temos ainda que $\varepsilon^{a,k} \rightarrow 0$, $\varepsilon^{b,k} \rightarrow 0$, e

$$r^k \rightharpoonup \bar{z} =: \bar{r}, \quad b^k \rightharpoonup (\bar{w}_1, \dots, \bar{w}_{n-1}) =: \bar{b}.$$

Para aplicar o Lema B.2.2, é necessário verificar que $Gr^k - s^k \rightarrow 0$ e $a^k + G^*b^k \rightarrow 0$. De fato, observamos que

$$\begin{aligned} Gr^k - s^k &= (G_1r^k, \dots, G_{n-1}r^k) - (x_1^k, \dots, x_{n-1}^k) \\ &= (G_1x_n^k - x_1^k, \dots, G_{n-1}x_n^k - x_{n-1}^k), \end{aligned}$$

e, pelas hipóteses do corolário, obtemos $Gr^k - s^k \rightarrow 0$. Além disso, como

$$G^*(w_1, \dots, w_{n-1}) = \sum_{i=1}^{n-1} G_i^*w_i, \quad \text{para todo } (w_1, \dots, w_{n-1}) \in \mathcal{G},$$

segue que

$$a^k + G^*b^k = y_n^k + \sum_{i=1}^{n-1} G_i^*y_i^k \rightarrow 0,$$

também conforme assumido nas hipóteses do corolário. Assim, pelo Lema B.2.2, concluímos que $(\bar{r}, -G^*\bar{b}) \in \text{Gra } A$ e $(G\bar{r}, \bar{b}) \in \text{Gra } B$, o que equivale exatamente à relação expressa em (B.6). $\square$

Referências Bibliográficas

- [1] M. A. Alghamdi, A. Alotaibi, P. L. Combettes, e N. Shahzad. A primal-dual method of partial inverses for composite inclusions. *Optim. Lett.*, 8(8):2271–2284, 2014.
- [2] A. Alotaibi, P. L. Combettes, e N. Shahzad. Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set. *SIAM J. Optim.*, 24(4):2076–2095, 2014.
- [3] F. Alvarez. On minimization property of a second order dissipative system in Hilbert spaces. *SIAM J. Control Optim.*, 38(4):1102–1119, 2000.
- [4] F. Alvarez. Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert spaces. *SIAM J. Optim.*, 14(3):773–782, 2003.
- [5] F. Alvarez e H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Var. Anal.*, 9(1-2):3–11, 2001.
- [6] M. Marques Alves e S. C. Lima. An inexact Spingarn’s partial inverse method with applications to operator splitting and composite optimization. *J. Optim. Theory Appl.*, 175(3):818–847, 2017.
- [7] M. M. Alves. Another proof and a generalization of a theorem of H. H. Bauschke on monotone operators. *Optimization*, 71(1):91–96, 2022.
- [8] M. Marques Alves, J. Eckstein, M. Geremia e Melo, J.G. Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. *Comput. Optim. Appl.*, 75:389–422, 2020.
- [9] M. Marques Alves e R. T. Marcavillaca. On inexact relative-error hybrid proximal extragradient, forward-backward and Tseng’s modified forward-backward methods with inertial effects. *Set-Valued Var. Anal.*, 28:301–325, 2020.
- [10] M. M. Alves, M. Geremia, e R. T. Marcavillaca. A relative-error inertial-relaxed inexact projective splitting algorithm. *J. Convex Anal.*, 31(1):1–24, 2024.
- [11] M. M. Alves, J. E. Navarro Caballero, e R. T. Marcavillaca. A strongly convergent inertial inexact proximal-point algorithm for monotone inclusions with applications to variational inequalities. *arXiv:2407.03485 [math.OC]*, 2024.

- [12] M. Marques. Alves, J. E. Navarro Caballero, e R. T. Marcavillaca. An inertial projective splitting algorithm with strong convergence. *J. Optim. Theory Appl.* 207, 65. <https://doi.org/10.1007/s10957-025-02827-w>, 2025.
- [13] M. Marques Alves, R. D. C. Monteiro, e B. F. Svaiter. Regularized HPE-type methods for solving monotone inclusions with improved pointwise iteration-complexity bounds. *SIAM J. Optim.*, 26(4):2730–2743, 2016.
- [14] A. S. Antipin. Minimization of convex functions on convex sets by means of differential equations. *Set-Valued Var. Anal.*, 28:301–325, 2019.
- [15] H. Attouch, A. Cabot, Z. Chbani, e H. Riahi. Inertial forward-backward algorithms with perturbations: application to Tikhonov regularization. *J. Optim. Theory Appl.*, 179(1):1–36, 2018.
- [16] H. Attouch e J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Math. Program.*, 174(1-2, Ser. B):391–432, 2019.
- [17] H. Attouch, J. Peypouquet, e P. Redont. Fast convex optimization via inertial dynamics with Hessian driven damping. *J. Differential Equations*, 261(10):5734–5783, 2016.
- [18] H. Attouch e A. Cabot. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. *Math. Program.*, 184:243—287, 2020.
- [19] H. Attouch e A. Cabot. Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions. *Appl. Math. Optim.*, 80(3):547–598, 2019.
- [20] H. Attouch, A. Cabot, Z. Chbani, e H. Riahi. Inertial forward-backward algorithms with perturbations: application to Tikhonov regularization. *J. Optim. Theory Appl.*, 179(1):1–36, 2018.
- [21] H. Attouch, Z. Chbani, J. Peypouquet, e P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing. *Math. Program.*, 168(1-2, Ser. B):123–175, 2018.
- [22] H. Attouch. Fast inertial proximal ADMM algorithms for convex structured optimization with linear constraint. *Minimax Theory Appl.*, 6(1):1–24, 2021.
- [23] H. H. Bauschke e P. L. Combettes. A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Math. Oper. Res.*, 26(2):248–264, 2001.
- [24] H. H. Bauschke e P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces, second edition*. Springer, New York, 2017.
- [25] A. Beck e M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
- [26] D. P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Mathematical programming*, 129(2):163–195, 2011.

- [27] R. I. Boţ e E. R. Csetnek. ADMM for monotone operators: convergence analysis and rates. *Adv. Comput. Math.*, 45(1):327–359, 2019.
- [28] R. I. Boţ e E. R. Csetnek. A hybrid proximal-extragradient algorithm with inertial effects. *Numer. Funct. Anal. Optim.*, 36(8):951–963, 2015.
- [29] H. Brezis. Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. *North-Holland Mathematics Studies. 5*. Amsterdam-London: North-Holland, New York, American Elsevier Publishing Comp., 1973.
- [30] F. E. Browder. Nonlinear elliptic boundary problems. *Bull. Amer. Math. Soc.*, 69(6):862–874, 1963.
- [31] F. E. Browder. The solvability of non-linear functionals equations. *Duke Math. J.*, 30(4):557–566, 1963.
- [32] R. S. Burachik, A. N. Iusem, e B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, 5(2):159–180, 1997.
- [33] R. S. Burachik, C. A. Sagastizábal, e B. F. Svaiter. ϵ -enlargements of maximal monotone operators: theory and applications. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, volume 22 of *Appl. Optim.*, pages 25–43. Kluwer Acad. Publ., Dordrecht, 1999.
- [34] J. V. Burke. A variable metric proximal point algorithm for monotone operators. *SIAM J. Control Optim.*, 37(2):353–375, 1999.
- [35] L. C. Ceng, B. S. Mordukhovich, e J. C. Yao. Hybrid approximate proximal method with auxiliary variational inequality for vector optimization. *J. Optim. Theory Appl.*, 146(2):267–303, 2010.
- [36] X. Chen, e M. Fukushima. Proximal quasi-Newton methods for nondifferentiable convex optimization. *Math. Program.*, 85(2):313–334, 1999.
- [37] C. Chen, R. H. Chan, S. Ma, e J. Yang. Inertial proximal ADMM for linearly constrained separable convex optimization. *SIAM J. Imaging Sci.*, 8(4):2239–2267, 2015.
- [38] C. Chen, S. Ma, e J. Yang. A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM J. Optim.*, 25(4):2110–2142, 2015.
- [39] P. L. Combettes e L. E. Glaudin. Quasi-nonexpansive iterations on the affine hull of orbits: from Mann’s mean value algorithm to inertial methods. *SIAM J. Optim.*, 27(4):2356–2380, 2017.
- [40] Q.-L. Dong, D. Jiang, P. Chalamjiak, e Y. Shehu. A strong convergence result involving an inertial forward–backward algorithm for monotone inclusions. *Journal of Fixed Point Theory and Applications*, 19, 12 2017.
- [41] J. Eckstein. A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 173(1):155–182, 2017.

- [42] J. Eckstein e D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(3):293–318, 1992.
- [43] J. Eckstein e B. F. Svaiter. A family of projective splitting methods for the sum of two maximal monotone operators. *Math. Program.*, 111(1-2, Ser. B):173–199, 2008.
- [44] J. Eckstein e B. F. Svaiter. General projective splitting methods for sums of maximal monotone operators. *SIAM J. Control Optim.*, 48(2):787–811, 2009.
- [45] F. Facchinei e J. S. Pang. Finite-dimensional variational inequalities and complementarity problems. *Springer*, 2003.
- [46] M. L. N. Goncalves, J. G. Melo, e R. D. C. Monteiro. Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. *SIAM Journal on Optimization*, 27(1):379–407, 2017.
- [47] O. Güler. New proximal point algorithms for convex minimization. *SIAM J. Optim.*, 2(4):649–664, 1992.
- [48] A. N. Iusem e W. Sosa. On the proximal point method for equilibrium problems in Hilbert spaces. *Optimization*, 59(8):1259–1274, 2010.
- [49] Jr. Jim Douglas e H. H. Rachford Jr. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.
- [50] P. R. Johnstone e J. Eckstein. Projective splitting with forward steps. *Math. Program.*, 191(2, Ser. A):631–670, 2022.
- [51] R. I. Kachurovskii. Monotone operators and convex functionals. *Usp. Mat. Nauk*, 15(4):213–215, 1960.
- [52] R. I. Kachurovskii. Non-linear monotone operators in Banach spaces. *Uspekhi Mat. Nauk*, 23(2):121–168, 1968.
- [53] G. M. Korpelevič. An extragradient method for finding saddle points and for other problems. *Èkonom. i Mat. Metody*, 12(4):747–756, 1976.
- [54] M. Lichman. UCI machine learning repository: Center for Machine Learning and Intelligent Systems, 2013. <http://archive.ics.uci.edu/ml>
- [55] P.-L. Lions e B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.
- [56] D. A. Lorenz e T. Pock. An inertial forward-backward algorithm for monotone inclusions. *J. Math. Imaging Vision*, 51(2):311–325, 2015.
- [57] P.-E. Maingé e N. Merabet. A new inertial-type hybrid projection-proximal algorithm for monotone inclusions. *Appl. Math. Comput.*, 215(9):3149–3162, 2010.
- [58] B. Martinet. Régularisation d’inéquations variationnelles par approximations successives. *Recherche Opérationnelle*, 4(Ser. R-3):154–158, 1970.

- [59] J.-E. Martinez-Legaz e M. Théra. ϵ -subdifferentials in terms of subdifferentials. *Set-Valued Anal.*, 4(4):327–332, 1996.
- [60] G. J. Minty. Monotone networks. *Roc. R. Soc. Land. A: Math. Phys. Eng. Sci.*, 257(1289):194–212, 1960.
- [61] G. J. Minty. On the monotonicity of the gradient of a convex function. *Pac. J. Math.*, 14(1):243–247, 1964.
- [62] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Rev. Française Informat. Recherche Opérationnelle*, 29:154–158, 1970.
- [63] R. D. C. Monteiro, C. Ortiz, e B. F. Svaiter. A first-order block-decomposition method for solving two-easy-block structured semidefinite programs. *Math. Program. Comput.*, 6(2):103–150, 2014.
- [64] R. D. C. Monteiro, C. Ortiz, e B. F. F. Svaiter. Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems. *Comput. Optim. Appl.*, 57(1):45–69, 2014.
- [65] R. D. C. Monteiro, C. Ortiz, e B. F. F. Svaiter. An adaptive accelerated first-order method for convex optimization. *Comput. Optim. Appl.*, 64(1):31–73, 2016.
- [66] R. D. C. Monteiro e B. F. Svaiter. Complexity of variants of Tseng’s modified F-B splitting and Korpelevich’s methods for hemivariational inequalities with applications to saddle point and convex optimization problems. *SIAM Journal on Optimization*, 21:1688–1720, 2010.
- [67] R. D. C. Monteiro e B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20:2755–2787, 2010.
- [68] R. D. C. Monteiro e B. F. Svaiter. Iteration-Complexity of a Newton Proximal Extragradient Method for Monotone Variational Inequalities and Inclusion Problems. *SIAM J. Optim.*, 22(3):914–935, 2012.
- [69] R. D. C. Monteiro e B. F. Svaiter. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. *SIAM J. Optim.*, 23(2):1092–1125, 2013.
- [70] R.D.C Monteiro e S. Chee-Khian. Complexity of the relaxed Peaceman-Rachford splitting method for the sum of two maximal strongly monotone operators. *Comput. Optim. Appl.*, 70:763–790, 2018.
- [71] J.J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.
- [72] A. Moudafi e M. Oliny. Convergence of a splitting inertial proximal method for monotone operators. *J. Comput. Appl. Math.*, 155(2):447–454, 2003.

- [73] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15:229–251, 2005.
- [74] Y. E. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$. *Doklady AN SSSR* (translated as Soviet Math. Doel.), 269:543–547, 1983.
- [75] Y. E. Nesterov. Introductory lectures on convex optimization. A basic course. Kluwer Academic Publishers, Boston, 2004.
- [76] P. Ochs, Y. Chen, T. Brox, e T. Pock. Ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sciences*, 7(2):1388–1419, 2014.
- [77] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.
- [78] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.*, 72(2):383–390, 1979.
- [79] D. W. Peaceman e Jr. H. H. Rachford. The numerical solution of parabolic and elliptic differential equations. *J. Soc. Indust. Appl. Math.*, 3:28–41, 1955.
- [80] M. Pentón Machado e M. R. Sire. A Projective Splitting Method for Monotone Inclusions: Iteration-Complexity and Application to Composite Optimization. *J. Optim. Theory Appl.*, 198, 552–587, 2023.
- [81] B. T. Polyak. Introduction to Optimization. *Translations Series in Mathematics and Engineering. Optimization Software*, 1987.
- [82] B. T. Polyak. Some methods of speeding up the convergence of iteration method. *U.S.S.R. Comput. Math.*, 4(5):1–17, 1964.
- [83] L. Qi e X. Chen. A preconditioning proximal newton method for nondifferentiable convex optimization. *Math. Program.*, 76(3):411–429, 1997.
- [84] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pac. J. Math.*, 17(3):497–510, 1966.
- [85] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of operations research.*, 1.2:97–116, 1976.
- [86] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.
- [87] R. T. Rockafellar. Monotone operators associates with saddle-functions and minimax problems. *American Math.Soc.*, 5:159–180, 1970.
- [88] R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math*, 33:209–216, 1970.

- [89] M. V. Solodov e B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Var. Anal.*, 7(4):323–345, 1999.
- [90] M. V. Solodov e B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.
- [91] M. V. Solodov e B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.*, 25(2):214–230, 2000.
- [92] M. V. Solodov e B. F. Svaiter. Forcing strong convergence of proximal point iterations in a Hilbert space. *Math. Program.*, 87(1, Ser. A):189–202, 2000.
- [93] M. V. Solodov e B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.
- [94] J. E. Spingarn. Partial inverse of a monotone operator. *Appl. Math. Optim.*, 10(3):247–265, 1983.
- [95] B. F. Svaiter. A class of Fejér convergent algorithms, approximate resolvents and the hybrid proximal-extragradient method. *J. Optim. Theory Appl.*, 162(1):133–153, 2014.
- [96] B. F. Svaiter. A weakly convergent fully inexact Douglas-Rachford method with relative error tolerance. *ESAIM: COCV*, 25:57–70, 2018.
- [97] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446 (electronic), 2000.
- [98] E. H. Zarantonello. Projections on convex sets in Hilbert space and spectral theory. In: Zarantonello, E.H., ed., *Contributions to Nonlinear Functional Analysis*, Academic Press, 237–424, 1971.