UMA NOVA CLASSE DE MÉTODOS DE PONTO PROXIMAL COM MÉTRICA VARIÁVEL PARA PROBLEMAS EM OTIMIZAÇÃO COM RESTRIÇÕES DE POSITIVIDADE

Gilvan Lima de Oliveira

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:

Prof. Paulo Roberto Oliveira, Dr. Ing.

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

Prof. Márcia Helena Costa Fampa, D. Sc.

Prof. Maria Helena Cautiero Horta Jardim, D. Sc.

Prof. Susana Scheimberg de Makler, D. Sc.

Prof. Orizon Pereira Ferreira, D. Sc.

Prof. Levi Lopes de Lima, D. Sc.

RIO DE JANEIRO, RJ - BRASIL

ABRIL DE 2002

UMA NOVA CLASSE DE MÉTODOS DE PONTO PROXIMAL COM MÉTRICA VARIÁVEL PARA PROBLEMAS EM OTIMIZAÇÃO COM RESTRIÇÕES DE POSITIVIDADE

Gilvan Lima de Oliveira

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:

____

Prof. Paulo Roberto Oliveira, Dr. Ing.

____

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

____

Prof. Márcia Helena Costa Fampa, D. Sc.

____

Prof. Maria Helena Cautiero Horta Jardim, D. Sc.

____

Prof. Susana Scheimberg de Makler, D. Sc.

____

Prof. Orizon Pereira Ferreira, D. Sc.

____

Prof. Levi Lopes de Lima, D. Sc.

RIO DE JANEIRO, RJ - BRASIL

ABRIL DE 2002

OLIVEIRA, GILVAN LIMA DE

Uma Nova Classe de Métodos de Ponto Proximal com Métrica Variável Para Problemas em Otimização com Restrições de Positividade [Rio de Janeiro] 2002

IX, 46 p. 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computacão, 2002)

Tese – Universidade Federal do Rio de Janeiro, COPPE

1 - Programação Convexa

2 - Método de Ponto Proximal

3 - Métrica Variável

I. COPPE/UFRJ   II. Título (série)

- À memória de meu pai João Cícero e de minhas irmãs:
Iranildes, Fátima e Regina Lúcia.

Agradecimentos

- Aos meus orientadores Paulo Roberto Oliveira e João Xavier da Cruz Neto pela orientação, incentivo, pela confiança creditada a mim e por terem sido, acima de tudo, amigos nos momentos mais difíceis no transcorrer deste trabalho.

- Aos demais componentes da banca examinadora pela leitura desta tese e pelas sugestões dadas que contribuiram bastante para melhoria dela.

- Um agradecimento muito especial à minha esposa Alcynéa e aos meus filhos Camilla e Carlos Eduardo que se constituiram nos maiores responsáveis pelo meu empenho e desempenho na realização deste curso.

- Aos professores do Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ que também contribuíram diretamente ou indiretamente para que este trabalho se realizasse. Em especial aos professores Regina Burachik, Nelson Maculan, Luis Maurício e Susana Scheimberg que enriqueceram-me de muitos conteúdos em Programação Linear e Análise Convexa.

- Aos colegas do Departamento de Matemática da UFPi, que assumiram minha carga horária durante o tempo que estive realizando o doutorado.

- A CAPES, pelo apoio financeiro.

- À minha mãe Francisca Lima e aos meus irmãos: Ivan, Jesus, Evanildes, Ivoneide, Iraneide (Dedé) e Irisvan que sempre me apoiaram.

- A todos os amigos que direta ou indiretamente me ajudaram em mais esta conquista.

– Aos amigos de cursos da UFRJ, representados por: Vera e Adriano Caminha("Krueger"), Rodrigo Leone, Rosely e Geci.

– Um agradecimento especial aos amigos Jurandir de Oliveira Lopes e Alfredo Wagner, pelas discussões valiosas que tivemos a respeito do tema e pelos bons tempos vividos no Rio.

– Ao amigo Paulo Sérgio Marques pelas preciosas ajudas quando da digitação desta tese.

– Ao meu grande amigo Marcílio Flávio Rangel de Farias, pela ajuda permanente nos momentos mais difíceis dessa caminhada e por fazer do ensino de 1° e 2° graus, em especial o de Matemática, no estado do Piauí, uma referência nacional.

– Finalmente, a todos os funcionários do PESC-COPPE pela ajuda em todas as ocasiões.

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

UMA NOVA CLASSE DE MÉTODOS DE PONTO PROXIMAL COM MÉTRICA VARIÁVEL PARA PROBLEMAS EM OTIMIZAÇÃO COM RESTRIÇÕES DE POSITIVIDADE

Gilvan Lima de Oliveira

Abril/2002

Orientadores: Paulo Roberto Oliveira e João Xavier da Cruz Neto

Programa: Engenharia de Sistemas e Computação

Neste trabalho introduzimos uma família de métodos interior-proximal com métrica variável para resolver o problema $\min f(x)$ $s.t.$ $x \geq 0$, onde o passo iterativo é dado por

$$x_j^{k+1} = x_j^k \left(1 - \beta_k^{-1}(x_j^k)^{r-1} \left[\nabla f(x^{k+1})\right]_j\right)$$

com β_k convenientemente escolhido, e $r \geq 1$. Estebelecemos propriedades de convergência similares aquelas conhecidas dos métodos multiplicativos, a saber, convergência fraca a um ponto KKT para funções convexas com gradiente Lipschitziano e convergência forte a uma solução para funções estritamente convexas. No caso quadrático, o algoritmo tem uma expressão explícita. Estudamos a taxa de convergência deste método. Mostramos que a taxa de convergência é linear quando a função é fortemente convexa no ótimo, e sublinear caso contrário. Uma versão inexata é introduzida e analisada.

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

A NEW CLASS OF PROXIMAL POINT METHODS WITH VARIABLE METRIC FOR OPTIMIZATION UNDER POSITIVITY CONSTRAINTS

Gilvan Lima de Oliveira

April/2002

Advisors: Paulo Roberto Oliveira and João Xavier da Cruz Neto

Department: Systems and Computation Engineering

In this work we introduce a family of variable metric interior-proximal methods for the problem $\min f(x)$ $s.t.$ $x \geq 0$, where the iterative step is given by

$$x_j^{k+1} = x_j^k \left(1 - \beta_k^{-1}(x_j^k)^{r-1}\left[\nabla f(x^{k+1})\right]_j\right)$$

with β_k conveniently chosen, and $r \geq 1$. We establish convergence properties similar to those known for multiplicative methods, namely weak convergence to a KKT point for convex and gradient Lipschitzian functions and full convergence to the solution for strictly convex functions. In the quadratic case the algorithm has explicit expression. We study the convergence rate of that method. We show that the rate of convergence is linear when the objective is strongly convex at the optimum, but can be sublinear otherwise. An inexact version is introduced and analyzed.

Índice

Introdução

Um método muito utilizado para minimizar funções convexas $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é o Algoritmo de Ponto Proximal. Mais especificamente, tal método gera uma sequência $\{x^k\}$ obtida pela resolução do subproblema:

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + (2\beta_k)^{-1} \|x - x^k\|^2 \right\} \tag{0.1}$$

onde $\{\beta_k\}$ é uma sequência de números reais e $\|\cdot\|$ denota a norma Euclideana em $\mathbb{R}^n$. Este método foi introduzido por Martinet([M/78]) o qual é baseado na noção de aproximação proximal de f dada por

$$f_\lambda(x) = \inf_{z \in \mathbb{R}^n} \left\{ f(z) + (2\lambda)^{-1} \|x - z\|^2 \right\}, \quad \lambda > 0, \tag{0.2}$$

introduzida anteriormente por Moreau([MO/62] e [MO/65]). No ano de 1976, o algoritmo proximal foi fortemente desenvolvido e estudado por Rockafellar([R/70] e [R/76a]).

Nos anos seguintes, muitos pesquisadores propuseram e estudaram propriedades de convergência de novos métodos de minimização que trazem em si a sistemática utilizada no algoritmo de ponto proximal, substituindo o termo regularizador quadrático por um núcleo diferente. Nessa direção, em 1990, Eggermont([EG/90]) introduziu um método gradiente multiplicativo para resolver o seguinte problema com restrições de positividade:

$$\min\{f(x); \quad x \geq 0\} \tag{0.3}$$

Tais métodos são bastante utilizados em várias aplicações interessantes, como por exemplo, reconstrução de imagens. Para maiores informações, ver ([EG/90]) e referências ali relacionadas.

O método proposto por Eggermont, tem o formato (0.1), onde o termo quadrático foi substituído por uma função tipo-distância, a saber,

$$d(x,y) = \sum_{j=1}^{n} x_j \log\left(\frac{x_j}{y_j} + y_j - x_j\right), \tag{0.4}$$

a chamada φ-divergência de Kullback-Leibler.

Em 1992, Censor e Zenios ([CZ/92]) propuseram um novo método proximal substituindo, desta feita, o núcleo quadrático por uma outra função tipo-distância, denominada distância de Bregman. Este algoritmo foi mais explorado sistematicamente por Eckstein([E/93]), Chen e Teboulle([CT/93]) e A. Iusem, B. F. Svaiter e M. Teboulle ([IST/94]).

Motivado pelo trabalho de Eggermont ([EG/90]), em 1995, A. Iusem ([I/95a]) analisou um método gradiente multiplicativo de pontos interiores para o problema (0.3), tendo sobretudo analisado propriedades de convergência da sequência $\{x^k\}$, obtida através das iterações definidas por:

$$x_j^{k+1} = x_j^k \left(1 - \lambda_k \left[\nabla f(x^k)\right]_j\right) \qquad (1 \leq j \leq n), \tag{0.5}$$

onde λ_k é um número real positivo, obtido através de uma busca linear exata em um intervalo $[0, c_k]$, com c_k escolhido de tal forma a se obter a viabilidade do iterado x^{k+1}, ou seja, de forma que x^{k+1} permaneça no interior do ortante positivo. Os resultados de convergência obtidos para este algoritmo, diante da escolha dos parâmetros λ_k, foram essencialmente os mesmos obtidos para o caso do método de descida máxima no caso da minimização irrestrita. Basicamente, supondo nível limitação da função objetivo f, A. Iusem, mostrou convergência fraca (ver Cap. Preliminares), da sequência gerada pelo método, a uma solução do problema (0.3) e convergência forte diante da hipótese de convexidade estrita de f.

O uso de métrica variável para o método de ponto proximal, foi discutido por J. B. Hirirt-Urruty e C. Lemaréchal em ([UL/93]), para o caso irrestrito:

$$\min\{f(x); x \in \mathbb{R}^n\}, \tag{0.6}$$

onde a função f é suposta convexa, própria e fechada (f toma valores em $\mathbb{R}^n \bigcup\{+\infty\}$ mas não é identicamente $+\infty$, enquanto que a propriedade de ser fechada, diz respeito a f ser semicontínua inferiormente). Neste modelo, o iterado x^{k+1} é obtido como solução do problema:

$$\min_{y \in \mathbb{R}^n} \left[f(y) + \frac{1}{2}\langle M_k(y - x^k), y - x^k \rangle \right], \tag{0.7}$$

onde M_k é uma matriz simétrica definida positiva, para cada $k \in \mathbb{N}$. Os autores mostraram que se $M_k = \mu_k M$, com $\mu_k > 0$ para todo k, M simétrica definida positiva e se a condição de convergência

$$\sum_{k=1}^{+\infty} \frac{1}{\mu_k} = +\infty, \tag{0.8}$$

ocorrer, então a sequência $\{x^k\}$ gerada pelo algoritmo é minimizante. Além disso, se $\mu_k \geq \sigma > 0$, então a sequência inteira converge, desde que o problema tenha solução. Tal método foi mais aprofundado por J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal e C. A. Sagastizábal em ([BGLS/95]), porém ainda no caso irrestrito. Neste trabalho, os autores estabeleceram, além da convergência global, taxa de convergência superlinear para a sequência gerada pela algoritmo.

Uma ferramenta muito utilizada na programação matemática contínua é a associação entre métricas Riemannianas e direções de descida. Tal associação data desde Luemberger ([L/72]), seguido por Gabay ([G/82]), Bayer e Lagarias ([BL/89]) e Rapcsák ([RT/96]).

Em 1995, J. X. C. Neto e P. R. Olivcira ([XO/95]) cxploraram dc forma sistemática tal ligação. Partindo da propriedade de que o gradiente depende da métrica e a solução do subproblema que define as direções de descida são equivalentes, os autores mostraram que a maioria dos métodos primais clássicos são métodos gradiente em alguma métrica Riemanniana apropriadamente escolhida. Por exemplo, para o problema (0.6), os autores verificaram que se f é continuamente diferenciável, então a direção de Cauchy é a direção oposta do gradiente onde $\mathbb{R}^n$ é munido da métrica euclideana, e, se f é de classe C^3, então o método de Newton aplicado a tal problema é um método gradiente, onde o conjunto

$$M = \{x \in \mathbb{R}^n; \nabla^2 f(x) \ \text{ é definida positiva}\}, \tag{0.9}$$

suposto não-vazio, está munido da métrica definida pela hessiana $\nabla^2 f(\cdot)$, mais especificamente temos,

$$df(x)v = \langle [\nabla^2 f(x)]^{-1} \nabla f(x), v \rangle \tag{0.10}$$

Com relação aos Métodos Multiplicativos, de Eggermont ([EG/90]), para resolver o problema (0.3), os autores mostraram que a direção proposta representava o gradiente da função f relativo à métrica definida no interior do ortante positivo, dada pelo hessiano da barreira entropia $h(x) = \sum_{i=1}^{n} x_i log x_i$, ou seja,

$$[\text{grad} f(x)]_j = x_j \left[\nabla f(x)\right]_j. \tag{0.11}$$

Em geral, considerando $\mathbb{R}^n_{++}$ como uma variedade Riemanniana, com a métrica $G(x) = diag(h_1''(x_1), h_2''(x_2), ..., h_n''(x_n))$, dada pelo hessiano de uma função convexa particular $h(x) = \sum_{i=1}^{n} h_i(x_i)$, os autores mostraram que, para resolver o problema

$$\min\{f(x); \quad \text{sujeito a:} \quad Ax = b, \quad x \in \mathbb{R}^n_{++}\}, \tag{0.12}$$

a direção afim-escala é o gradiente associado à métrica correspondente à barreira logarítmica. Ou seja, quando $h_i(x) = -logx_i$, com a métrica Riemanniana

$$G(x) = X^{-2} = \text{diag}(x_1^{-2}, x_2^{-2}, ..., x_n^{-2}), \tag{0.13}$$

X representando a matriz diagonal cujos componentes não-nulas são os elementos de x, verificou-se que:

$$\text{grad}^G f(x) = X(I - XA^t(AX^2A^t)^{-1}AX)X\nabla f(x), \tag{0.14}$$

a qual é a direção oposta à direção afim-escala de Dikin. Na expressão acima, A^t representa a transposta da matriz A. Os autores, ainda, desenvolveram uma teoria de convergência em que os resultados assemelharam-se aos clássicos da otimização irrestrita (veja ([XLO/98])).

Em seguida, O. P. Ferreira e P. R. Oliveira ([FO/97]), generalizaram os algoritmos de ponto proximal e subgradiente, para minimizar uma função convexa, para o contexto de variedades Riemannianas. Os autores obtiveram alguns resultados de convergência de tais algoritmos mediante algumas hipóteses sobre a curvatura da variedade. Por exemplo, para o algoritmo subgradiente clássico, os autores provaram convergência, mediante a hipótese de não-negatividade da curvatura da variedade, enquanto que para o algoritmo de ponto proximal eles provaram a convergência supondo que a variedade seja de Hadamard. Em tais trabalhos ([XO/95]) e ([FO/97]), é feita busca ao longo de geodésicas.

A. W. M. Pinto e P. R. Oliveira ([WO/00]) generalizaram e analisaram o algoritmo proposto por Eggermont, sob o enfoque de geometria Riemanniana, no qual não é realizada busca geodésica, e sim busca linear na direção do simétrico do gradiente da função objetivo restrita a variedade. Nesse trabalho, o conjunto $\mathbb{R}^n_{++}$ é encarado com uma variedade Riemanniana com a métrica $G(x) = X^{-r}$, onde $r \geq 1$. Observe que para cada valor do parâmetro r, o núcleo a ser usado, representará uma métrica dada pelo hessiano de uma determinada função convexa do tipo discutida no parágrafo anterior, por exemplo, para $r = 1$, o núcleo representará a métrica dada pelo hessiano da barreira entropia $h(x) = \sum_{i=1}^{n} x_i log x_i$; para $r = 2$, a métrica é dada pelo hessiano da barreira logaritmica $h(x) = -\sum_{i=1}^{n} log x_i$ e para $r = 3$, a métrica é dada pelo hessiano da barreira de Fiacco e McCormick $h(x) = \sum_{i=1}^{n} \frac{1}{x_i}$. Nesse trabalho, o conjunto

$$M = \{x \in \mathbb{R}^n_{++}; \quad Ax = b\} \tag{0.15}$$

é visto como uma subvariedade de $\mathbb{R}^n_{++}$ com a métrica induzida. Os autores observaram que a direção de busca proposta por Eggermont ([EG/90]) e por Gonzaga ([GC/90]) são casos particulares da direção que propõem, para determinados valores do parâmetro r (respectivamente, $r = 1$ e $r = 2$). Os autores obtiveram resultados de convergência semelhantes aos obtidos por A. Iusem ([I/95a]), para o algoritmo proposto por Eggermont, e aos obtidos por Gonzaga, para o caso afim-escala, sendo que nesse útimo caso a convergência fraca (ver Cap. Preliminares) foi obtida. Nesse trabalho não é feita análise da taxa de convergência e nem proposta versão inexata para o método em questão.

Motivado pelos estudos desenvolvidos por ([XO/95]), ([FO/97]) e, recentemente, por ([WO/00]), propomo-nos a introduzir e analisar as propriedades de convergência da sequência gerada pela família, a um parâmetro $r \geq 1$, de métodos de ponto proximal com métrica variável, dada pelo hessiano de funções convexas, para resolver o problema (0.3). Trata-se de método de pontos interiores no qual o núcleo quadrático, do algoritmo de ponto proximal tradicional, será substituído pela métrica definida por $G(x) = X^{-r}$ com $r \geq 1$. Uma das vantagens da utilização de tal métrica, além de ser definida a partir de funções convexas, é que a mesma, como veremos, produz demonstrações bastante simples para uma boa parte dos resultados analisados na convergência da sequência a ser obtida, como por exemplo, a prova de que a sequência $\{x^{k+1} - x^k\}$ tende a zero e, por conseguinte, a prova da complementariedade, isto é, $x_j^* \left[\nabla f(x^*)\right]_j = 0$, onde x^* é um ponto de acumulação de $\{x^k\}$.

A nova classe de métodos é diferente das estudadas por J. X. C. Neto e P. R. Oliveira ([XO/95]), O. P. Ferreira e P. R. Oliveira ([FO/97]) e A. W. M. Pinto e P. R. Oliveira ([WO/00]), uma vez que não faremos hipóteses sobre a curvatura, não utilizaremos busca geodésica e, nem tampouco, busca linear. Usaremos a metodologia do ponto proximal com métrica variável analisada em ([UL/93]) e ([BGLS/95]) para o caso irrestrito, restrita ao subconjunto convexo $\mathbb{R}^n_+ \subset \mathbb{R}^n$, obtendo viabilidade mediante escolha conveniente dos parâmetros de regularização. A taxa de convergência, da sequência gerada pelo método, será estabelecida, supondo limitação de tais parâmetros. Além disso, introduziremos e analisaremos uma versão inexata para essa nova classe de métodos.

As iterações analisadas por A. Iusem ([I/95a]), descritas em (0.5), serão vistas como uma realização explícita das iterações dessa nova classe de métodos, para o caso particular $r = 1$.

A tese encontra-se dividida em quatro capítulos:

O Capítulo 1 apresenta todos os preliminares indispensáveis ao desenvolvimento das análises que se seguem. As provas são todas referenciadas.

No Capítulo 2, fazemos uma relação das hipóteses necessárias sobre a função objetivo, definimos a nova classe de métodos e fazemos a análise das propriedades de convergência de tal classe. Mostramos, supondo a nivel-limitação da função objetivo e que a mesma possua gradiente Lipschitziano, que a sequência $\{x^k\}$ gerada pelo algoritmo, para uma escolha conveniente dos parâmetros de regularização, está contida no interior do ortante positivo, e "converge fracamente" (ver Cap. Preliminares). Para o caso quadrático, o algoritmo apresenta-se de forma explícita, inclusive os parâmetros de regularização, o que é bastante vantajoso do ponto de vista de implementação.

No Capítulo 3, fazemos uma análise da taxa de convergência de tal método. Mostramos que a sequência $\{x^k\}$ converge linearmente diante de convexidade forte da função objetivo no interior do ortante positivo. Na fronteira do ortante positivo, a sequência converge sublinearmente, para todo $r > 1$. No caso $r = 1$, mostramos que a taxa de convergência linear, nas componentes nulas do ponto limite da sequência gerada por tal método, pode ser obtida, desde que seja suposta complementariedade estrita.

No Capítulo 4, apresentamos uma versão inexata para tal algoritmo. Com uma escolha apropriada para os erros, mostramos essencialmente os mesmos resultados obtidos na versão exata. Como veremos, não pedimos que a sequência formada pelos módulos dos erros seja somável, apenas que a mesma tenda a zero.

Capítulo 1

Preliminares

1.1 Introdução

Neste capítulo apresentamos resultados indispensáveis às análises que faremos posteriormente. Os dois primeiros resultados, que dizem respeito a matrizes definidas positivas, serão úteis quando estudarmos a taxa de convergência do algoritmo proposto. Suas demonstrações serão referenciadas.

1.2 Símbolos e Notações

Ao longo deste trabalho, adotaremos a seguinte simbologia:

$\mathbb{R}^n = \{x = (x_1, x_2, ..., x_n);\ x_i \in \mathbb{R}\}$: espaço vetorial euclideano;

$\mathbb{R}^n_+ = \{x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n;\ x_i \geq 0\}$;

$\mathbb{R}^n_{++} = \{x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n;\ x_i > 0\}$;

$\langle x, y \rangle = \sum_{i=1}^n x_i y_i$: produto interno canônico em $\mathbb{R}^n$;

$||x|| = \sqrt{\langle x, x \rangle}$: norma euclideana;

$||x||_G = \sqrt{\langle Gx, x \rangle}$: norma dada pela matriz definida positiva G;

∇f e $\nabla^2 f$: gradiente e hessiano, respectivamente, de uma função real f;

$\max(x)$: maior componente do vetor $x \in \mathbb{R}^n$;

$\min(x)$: menor componente do vetor $x \in \mathbb{R}^n$;

$M_{m\times n}(\mathbb{R})$: conjunto das matrizes com elementos reais, de ordem $m \times n$;

$\|A\|_2 := \sup\{\|Ax\|;\ \|x\| \leq 1\}$, onde $A \in M_{m\times n}(\mathbb{R})$;

$(X)^r$: dado $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n_{++}$, a notação $(X)^r$ representará a matriz diagonal, de ordem $n \times n$, cujos elementos da diagonal principal são as r-ésimas potências x_1^r, x_2^r, ..., x_n^r das componentes do vetor x;

I: representará a matriz identidade em $M_{n\times n}(\mathbb{R})$.

Os dois resultados seguintes, a respeito de matrizes definidas positivas, tornar-se-ão as ferramentas cruciais para o estabelecimento da taxa de convergência linear da nova classe de métodos que propomos analisar.

Proposição 1.2.1 *Para cada par de matrizes definidas positivas $A, B \in M_{n\times n}(\mathbb{R})$, existem matrizes $P, D \in M_{n\times n}(\mathbb{R})$ (P não singular, D diagonal positiva) tais que $AB = PDP^{-1}$.*

Prova. Ver R. A. Horn e C. R. Johnson [HJ/85]. ■

Proposição 1.2.2 *Seja $M \in M_{n\times n}(\mathbb{R})$ uma matriz não singular e considere a norma $||| x ||| = \|M^{-1}x\|$. Se A, B são simétricas definidas positivas, $AB = PDP^{-1}$, com D diagonal positiva e β um número real positivo, então*

a) $||| (I + \frac{1}{\beta}AB)^{-1} ||| \leq \|M^{-1}P\|_2 \|P^{-1}M\|_2 (1 + \frac{\lambda\xi}{\beta})^{-1}$

b) $||| (\beta A^{-1} + B)^{-1} ||| \leq \|M^{-1}\|_2 \|M\|_2 \xi^{-1}$

onde λ, ξ são os menores autovalores de A, B respectivamente.

Prova. Ver A. N. Iusem e M. Teboulle [IT/93]. ■

1.3 Taxa de Convergência

Definição 1.3.1 *Seja $\{x^k\}$ uma sequência em $\mathbb{R}^n$ tal que $x^k \to x^*$. Então diremos que:*

1. $\{x^k\}$ *converge q-linearmente a x^*, se*

$$\limsup_{k\to+\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} < 1; \tag{1.1}$$

2. $\{x^k\}$ *converge q-superlinearmente a x^*, se*

$$\limsup_{k\to+\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0; \tag{1.2}$$

3. $\{x^k\}$ *converge q-sublinearmente a x^*, se*

$$\limsup_{k\to+\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 1. \tag{1.3}$$

Observação 1.3.1 *A convergência linear depende da métrica.*

1.4 Convergência Fraca

No próximo capítulo, mostaremos a convergência fraca para a sequência gerada pela nova classe de métodos a ser definida. Tal convergência é distinta daquela convergência fraca clássica estudada em análise funcional, conforme vemos na seguinte definição:

Definição 1.4.1 *Uma sequência $\{x^k\} \subset \mathbb{R}^n$ diz-se fracamente convergente a um conjunto $C \subseteq \mathbb{R}^n$ se:*

1. *a sequência $\{x^k\}$ é limitada,*

2. $\lim_{k\to+\infty}(x^{k+1} - x^k) = 0$,

3. *todo ponto limite de $\{x^k\}$ pertence a C.*

1.5 Operadores Monótonos Maximais

No Capítulo 4, para a garantia da boa definição do par de sequências geradas pelo método inexato a ser proposto, faremos uso de um resultado de existência num contexto mais amplo, o de operadores monótonos maximais. Estes são generalizações de transformações lineares não-negativas para o caso não-linear. Um caso particular desse tipo de operador é o gradiente de funções convexas diferenciáveis, como veremos a seguir. Iniciaremos com uma definição que pode ser dada, num contexto mais geral, para espaços de Hilbert. Mas por questão de simplicidade trabalharemos apenas com operadores ponto-conjunto definidos em $\mathbb{R}^n$.

Definição 1.5.1 *Seja* $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ *uma aplicação ponto-conjunto. Diremos que* T *é monótono se, e somente se,*

$$\langle x - y, u - v\rangle \geq 0, \tag{1.4}$$

para quaisquer $x,\ y \in \mathbb{R}^n,\ \ u \in T(x)$ *e* $v \in T(y)$.

Para incluirmos, também, o caso não diferenciável, precisamos fazer uso da definição de subgradientes.

Definição 1.5.2 ξ *é um subgradiente de* f *em* x *se, e somente se,*

$$\langle \xi, y - x\rangle \leq f(y) - f(x), \tag{1.5}$$

para todo $y \in \mathbb{R}^n$. *O conjunto,*

$$\partial f(x) = \{\xi;\ \xi \text{ é um subgradiente de } f \text{ em } x\} \tag{1.6}$$

denomina-se o subdiferencial de f *em* x.

Os dois resultados a seguir são bastante conhecidos e, portanto, vamos apenas citá-los:

1. Se f é diferenciável em x, e convexa, então $\partial f(x) = \{\nabla f(x)\}$.

2. Se f é convexa, então $\partial f(x) \neq \emptyset$, para todo x pertencente ao interior relativo do domínio efetivo de f ($domf = \{x \in \mathbb{R}^n;\ f(x) < +\infty\}$).

Definição 1.5.3 *Seja $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ um operador monótono. T é dito* maximal *se, e só se, para cada $\tilde{T} : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ monótono tal que $T(x) \subset \tilde{T}(x)$ para todo x, ocorre $T = \tilde{T}$.*

Dito de uma outra maneira, um operador monótono T é dito *maximal* se o gráfico de T, definido por

$$G(T) = \{(x,y);\ x \in dom(T),\ y \in T(x)\}, \tag{1.7}$$

onde $dom(T) = \{x;\ T(x) \neq \emptyset\}$, não está propriamente contido no gráfico de um outro operador monótono.

Exemplo 1.5.0.1 *$T = \partial f$, no caso em que f é convexa. (ver R. T. Rockafellar ([R/70]) pag.340)*

Definição 1.5.4 *Seja $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ um operador monótono maximal. Então,*

1. T é dito sobrejetivo se para todo $y \in \mathbb{R}^n$ existe $x \in \mathbb{R}^n$ tal que $y \in T(x)$ e

2. T é dito injetivo se para $x \neq y$ ocorre que $T(x) \bigcap T(y) = \emptyset$.

Maiores detalhes sobre operadores monótonos maximais podem ser obtidos em ([R/76b]) e ([I/95b]).

Capítulo 2

Definição do Algoritmo e Análise de Convergência

2.1 Introdução

Neste capítulo definiremos a família, a um parâmetro, de métodos do tipo-proximal com métrica variável para resolver o problema:

$$(P) : \begin{cases} \min : & f(x) \\ s.\, a: & x \geq 0 \end{cases} \tag{2.1}$$

Trata-se de um método de pontos interiores, onde a viabilidade é garantida por uma escolha apropriada dos parâmetros de regularização. Em seguida, faremos a análise das propriedades de convergência de tal classe de métodos e verificaremos que, no caso quadrático, temos uma expressão explícita para o iterado, fazendo com que, do ponto de vista computacional, o método seja bastante vantajoso.

2.2 Hipóteses Necessárias

Nesta secção estabeleceremos as hipóteses necessárias sobre a função objetivo, que servirão para a análise de convergência da nova classe de métodos. Hipóteses adicionais poderão ser feitas durante o desenvolvimento do texto, mas o serão em situações específicas.

Hipótese 2.2.1. A função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ é continuamente diferenciável e convexa sobre $\mathbb{R}^n$, de forma que para todo $x, y \in \mathbb{R}^n$,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0. \tag{2.2}$$

Hipótese 2.2.2. Os conjuntos de níveis de f são compactos, isto é, para todo $y \in \mathbb{R}^n$, o conjunto

$$N_f(y) = \{x \in \mathbb{R}^n; f(x) \leq f(y)\} \tag{2.3}$$

é compacto.

Hipótese 2.2.3. A função objetivo f possui gradiente Lipschtiziano, isto é, existe uma constante $0 < L < +\infty$, tal que para todo $x, y \in \mathbb{R}^n$

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|. \tag{2.4}$$

As Hipóteses 2.2.1. e 2.2.2. são essenciais para a proposta a ser desenvolvida. A Hipótese 2.2.3. será muito importante, como veremos, para a definição dos parâmetros de regularização do método, fazendo com que o método proximal proposto, seja de pontos interiores.

2.3 Definição do Algoritmo

No que se segue, $\mathcal{S}$ denotará o conjunto formado pelas soluções do problema (P), e suponhamos que $\mathcal{S} \neq \emptyset$.

Inicialmente, escolha um número real $r \geq 1$ e $\{\sigma_k\}$ uma sequência de números reais tais que $0 < \sigma_k \leq \lambda$, $\forall k \in \mathbb{N}$, para algum $\lambda > 0$.

O algoritmo será definido como segue:

Algoritmo 2.3.1 .

Inicialização:

$$x^0 > 0 \tag{2.5}$$

Passo Iterativo: *Dado* $x^k > 0$, *defina*

$$\beta_k = (\max(x^k))^{r-1} \left\| \nabla f\left(x^k\right) \right\| + L(\max(x^k))^r + \sigma_k \tag{2.6}$$

Compute:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ f\left(x\right) + \frac{\beta_k}{2} \left\| x - x^k \right\|^2_{(X^k)^{-r}} \right\} \tag{2.7}$$

Se $\nabla f(x^k) = 0$ *ou se* $x^k = x^{k+1}$ *pare, senão*

Faça: $k \leftarrow k + 1$,

Repita.

Como nos referimos na introdução desse trabalho, aqui não realizaremos busca geodésica, busca linear e nem tampouco faremos penalizações. Os parâmetros de regularização foram definidos de forma a garantir a permanência da sequência gerada pelo algoritmo no interior do conjunto $\mathbb{R}^n_+$. A constante L, na definição de tais parâmetros β_k (igualdade (2.6)), é a mesma da desigualdade (2.4).

Doravante, $\{x^k\}$ representará a sequência gerada por (2.5)-(2.7).

2.4 Análise de Convergência

Nesta secção estudaremos as propriedades de convergência da classe de algoritmos proposto na secção anterior. Mostraremos a boa definição da sequência $\{x^k\}$, a convergência fraca do método ao conjunto S e convergência forte diante da convexidade estrita da função objetivo.

Comecemos com o estabelecimento da boa definição da sequência $\{x^k\}$, o que será consequência do seguinte lema:

Lema 2.4.1 *A sequência* $\{x^k\}$ *está bem definida.*

Prova. Usaremos indução sobre k; uma vez que (2.1) tem solução, seja α uma cota inferior para f em $\mathbb{R}^n_+$ e

$$f_k(x) = f(x) + \frac{\beta_k}{2}\|x - x^k\|^2_{(X^k)^{-r}}. \tag{2.8}$$

Então

$$f_k(x) \geq \alpha + \frac{\beta_k}{2}\left\|x - x^k\right\|^2_{\left(X^k\right)^{-r}} \tag{2.9}$$

e como $\left\|x - x^k\right\|^2_{\left(X^k\right)^{-r}}$ é nível-limitada, segue-se que f_k também o é. Dessa forma a minimização em (2.7) reduz-se a compactos e o mínimo é atingido. A unicidade desse minizador, x^{k+1}, decorre da convexidade estrita da função perturbada f_k. ■

Proposição 2.4.1 $x^k > 0$, *para todo* $k \in \mathbb{N}$.

Prova. Indução sobre k. Para $k = 0$ por (2.5). Desde que x^{k+1} resolve (2.7), temos

$$0 = \nabla f_k\left(x^{k+1}\right) = \nabla f\left(x^{k+1}\right) + \beta_k\left(X^k\right)^{-r}\left(x^{k+1} - x^k\right) \tag{2.10}$$

ou equivalentemente,

$$x^{k+1} = x^k - \frac{1}{\beta_k}\left(X^k\right)^r \nabla f\left(x^{k+1}\right). \tag{2.11}$$

Portanto, é verdade que

$$\left(X^k\right)^{-1}\left(x^{k+1} - x^k\right) = -\frac{1}{\beta_k}\left(X^k\right)^{r-1}\nabla f\left(x^{k+1}\right). \tag{2.12}$$

Agora é suficiente mostrar que $\left\|\left(X^k\right)^{-1}\left(x^{k+1} - x^k\right)\right\| < 1$; de fato, da igualdade (2.11) acima obtemos

$$\begin{aligned}
\left\|\left(X^k\right)^{-1}\left(x^{k+1} - x^k\right)\right\| &= \tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\nabla f\left(x^{k+1}\right)\right\| \\
&\leq \tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\| + \\
&\quad + \tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\left(\nabla f\left(x^{k+1}\right) - \nabla f\left(x^k\right)\right)\right\|,
\end{aligned}$$

então,

$$\begin{aligned}\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| &\leq \tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\|+\\ &+\tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\right\|\left\|\nabla f\left(x^{k+1}\right)-\nabla f\left(x^k\right)\right\|\end{aligned}$$

Como o gradiente de f é L-lipschtziano, temos que

$$\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| \leq \frac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\|+\frac{L}{\beta_k}(\max(x^k))^{r-1}\left\|x^{k+1}-x^k\right\|.$$

Esta última desigualdade pode ser reescrita como:

$$\begin{aligned}\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| &\leq \tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\|+\\ &+\tfrac{L}{\beta_k}(\max(x^k))^{r}(\max(x^k))^{-1}\left\|x^{k+1}-x^k\right\|.\end{aligned}$$

Como $(\max(x^k))^{-1}=\min[(x^k)^{-1}]:=\min((x_1^k)^{-1},(x_2^k)^{-1},...,(x_n^k)^{-1})$,

$$\begin{aligned}\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| &\leq \tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\|+\\ &+\tfrac{L}{\beta_k}(\max(x^k))^{r}\min\left[\left(x^k\right)^{-1}\right]\left\|x^{k+1}-x^k\right\|.\end{aligned}$$

Por outro lado, temos que

$$\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| \geq \min\left[\left(x^k\right)^{-1}\right]\left\|x^{k+1}-x^k\right\|,$$

dessa forma,

$$\begin{aligned}\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| &\leq \tfrac{1}{\beta_k}\left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\|+\\ &+\tfrac{L}{\beta_k}(\max(x^k))^{r}\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\|\end{aligned}$$

Esta última desigualdade torna-se

$$\left(\beta_k-L(\max(x^k)^r\right)\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| \leq \left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\|.$$

Da definição de β_k em (2.6), segue-se que

$$\left\|\left(X^k\right)^{-1}\left(x^{k+1}-x^k\right)\right\| \leq \frac{\left\|\left(X^k\right)^{r-1}\nabla f\left(x^k\right)\right\|}{\left\|(X^k)^{r-1}\nabla f\left(x^k\right)\right\|+\sigma_k} \tag{2.13}$$

Como $\sigma_k > 0$, temos que

$$\left\| \left(X^k\right)^{-1} \left(x^{k+1} - x^k\right) \right\| < 1.$$

Portanto, componente a componente, esta última desigualdade assegura que

$$\left| \frac{x_i^{k+1}}{x_i^k} - 1 \right| < 1,$$

o qual é equivalente a

$$0 < x_i^{k+1} < 2x_i^k \quad (1 \leq i \leq n). \tag{2.14}$$

■

Seja x^* uma solução do problema (P), ou seja $x^* \in \mathcal{S}$, e defina $f^* = f(x^*)$.

Proposição 2.4.2 *Para todo $k \geq 0$, temos $f^* \leq f\left(x^{k+1}\right) \leq f\left(x^k\right)$.*

Prova. Pela definição de x^{k+1} em (2.7),

$$f\left(x^{k+1}\right) + \frac{\beta_k}{2} \left\| x^{k+1} - x^k \right\|^2_{\left(X^k\right)^{-r}} \leq f(x) + \frac{\beta_k}{2} \left\| x - x^k \right\|^2_{\left(X^k\right)^{-r}}, \tag{2.15}$$

para todo $x \in \mathbb{R}^n$.

Em particular, colocando $x = x^k$ na desigualdade (2.15) acima, obtemos

$$f\left(x^{k+1}\right) + \frac{\beta_k}{2} \left\| x^{k+1} - x^k \right\|^2_{\left(X^k\right)^{-r}} \leq f\left(x^k\right), \tag{2.16}$$

isto é,

$$0 \leq \frac{\beta_k}{2} \left\| x^{k+1} - x^k \right\|^2_{\left(X^k\right)^{-r}} \leq f\left(x^k\right) - f\left(x^{k+1}\right) \tag{2.17}$$

■

Corolário 2.4.1 *A sequência $\left\{f\left(x^k\right)\right\}$ converge.*

Prova. Imediatamente da Proposição (2.4.2). ∎

Corolário 2.4.2 *A sequência* $\left\{x^k\right\}$ *é limitada.*

Prova. Pela Proposição (2.4.2) temos que $x^k \in N_f(x^0)$, para todo $k \geq 0$, os quais são limitados pela Hipótese 2.2.2. da segunda seção deste capítulo. ∎

Corolário 2.4.3 *A sequência dos parâmetros de regularização* $\{\beta_k\}$ *é limitada.*

Prova. A prova segue imediatamente da limitação da sequência $\{\sigma_k\}$ e do Corolário 2.4.2 acima. ∎

Corolário 2.4.4 $\sum_{k=0}^{+\infty} \beta_k \left\|x^{k+1} - x^k\right\|^2_{\left(X^k\right)^{-r}} < +\infty$.

Prova. Da desigualdade (2.17) temos

$$\sum_{k=0}^{j} \frac{\beta_k}{2} \left\|x^{k+1} - x^k\right\|^2_{\left(X^k\right)^{-r}} \leq f\left(x^0\right) - f\left(x^{j+1}\right) \leq f\left(x^0\right) - f\left(x^*\right),$$

para toda solução x^* do problema (P). Portanto,

$$\sum_{k=0}^{+\infty} \beta_k \left\|x^{k+1} - x^k\right\|^2_{\left(X^k\right)^{-r}} \leq 2(f\left(x^0\right) - f\left(x^*\right)) < +\infty$$

∎

Corolário 2.4.5 *Se, na definição dos parâmetros* β_k*,* $\sigma_k \geq \sigma > 0,\ \forall k$*, então*

$$\sum_{k=0}^{+\infty} \left\|x^{k+1} - x^k\right\|^2_{\left(X^k\right)^{-r}} < +\infty. \tag{2.18}$$

Prova. De fato, por hipótese temos que

$$\sigma \left\|x^{k+1} - x^k\right\|^2_{\left(X^k\right)^{-r}} \leq \beta_k \left\|x^{k+1} - x^k\right\|^2_{\left(X^k\right)^{-r}}. \tag{2.19}$$

A afirmação decorre do Corolário (2.4.4). ∎

Corolário 2.4.6 $\lim_{k\to+\infty} \left\| x^{k+1} - x^k \right\|^2_{\left(X^k\right)^{-r}} = 0.$

Prova. Consequência imediata do corolário anterior. ■

Corolário 2.4.7 $\sum_{k=0}^{+\infty} \left\| x^{k+1} - x^k \right\|^2 < +\infty.$

Prova. De fato, usando a definição de β_k, temos

$$\beta_k \| x^{k+1} - x^k \|^2_{(X^k)^{-r}} \geq \frac{\beta_k}{(\max(x^k))^r} \| x^{k+1} - x^k \|^2 \geq L \| x^{k+1} - x^k \|^2. \tag{2.20}$$

O resultado segue do Corolário 2.4.4 acima. ■

Corolário 2.4.8 $\lim_{k\to+\infty} (x^{k+1} - x^k) = 0.$

Prova. Imediatamente do Corolário 2.4.7. ■

Proposição 2.4.3 *Todo ponto de acumulação x^* de $\left\{x^k\right\}$ satisfaz:*

$$x_j^* \geq 0, \qquad x_j^* \left[\nabla f\left(x^*\right)\right]_j = 0, \qquad j = 1, 2, 3, ..., n. \tag{2.21}$$

Prova. Em vista da Proposição 2.4.1 e do Corolário 2.4.8, basta provarmos que $x_j^* \left[\nabla f\left(x^*\right)\right]_j = 0$ para todo j e todo ponto limite x^* de $\left\{x^k\right\}$. Seja x^* um ponto limite de $\left\{x^k\right\}$ e seja $\left\{x^{k_j}\right\}$ uma subsequência de $\left\{x^k\right\}$ convergindo para x^*. Usando a expressão do algoritmo e o Corolário 2.4.8, temos

$$\lim_{k\to+\infty} \left\| \frac{1}{\beta_k} \left(X^k\right)^r \nabla f\left(x^{k+1}\right) \right\| = 0. \tag{2.22}$$

Para cada $j = 1, 2, ..., n$ temos que

$$\left| \frac{1}{\beta_k} (x_j^k)^r [\nabla f(x^{k+1})]_j \right| \leq \left\| \frac{1}{\beta_k} \left(X^k\right)^r \nabla f\left(x^{k+1}\right) \right\|. \tag{2.23}$$

Pelo Corolário 2.4.3, temos que existe $\rho > 0$ tal que $\rho \leq \frac{1}{\beta_k}$, portanto, usando (2.23) segue-se que

$$\rho\left|(x_j^k)^T[\nabla f(x^{k+1})]_j\right| \leq \left\|\frac{1}{\beta_k}\left(X^k\right)^T \nabla f\left(x^{k+1}\right)\right\|. \tag{2.24}$$

Pelo limite (2.22) segue-se que

$$\lim_{k\to+\infty}\left(x^k\right)_j^T\left[\nabla f\left(x^{k+1}\right)\right]_j = 0. \tag{2.25}$$

Seja $\left\{x^{k_l}\right\}$ uma subsequência de $\left\{x^k\right\}$ tal que $x^{k_l} \to x^*$. Então temos

$$\lim_{l\to+\infty}\left(x^{k_l}\right)_j^T\left[\nabla f\left(x^{k_l+1}\right)\right]_j = 0. \tag{2.26}$$

Por outro lado, pelo Corolário 2.4.8, segue-se que $\lim_{l\to+\infty} x^{k_l+1} = \lim_{l\to+\infty} x^{k_l}$. Desse fato e do limite em (2.26) segue-se que $\left(x_j^*\right)^T\left[\nabla f\left(x^*\right)\right]_j = 0$. Portanto, $x_j^*\left[\nabla f\left(x^*\right)\right]_j = 0$ para cada $j = 1, 2, ..., n$. ∎

Proposição 2.4.4 *Todo ponto limite de $\left\{x^k\right\}$ é uma solução do problema (2.1).*

Prova. A prova é a mesma apresentada por A. Iusem em ([I/95a]), exceto por algumas modificações inerentes ao novo método. No entanto, para um maior entendimento do leitor, transcreveremos tais adaptações em detalhes.

Uma vez que f é convexa, é suficiente mostrarmos que x^* satisfaz as condições de KKT do problema (P), a saber,

$$x_j^* \geq 0, \quad x_j^*\left[\nabla f\left(x^*\right)\right]_j = 0, \quad \left[\nabla f\left(x^*\right)\right]_j \geq 0 \;\; (1 \leq j \leq n) \tag{2.27}$$

Agora, devido à Proposição 2.4.3, basta mostrarmos que $\left[\nabla f\left(x^*\right)\right]_j \geq 0$. Seja x^* um ponto de acumulação $\left\{x^k\right\}$ e suponha que existe i tal que $\left[\nabla f\left(x^*\right)\right]_i < 0$. Consideremos os seguintes conjuntos:

$$I = \left\{i;\ \left[\nabla f\left(x^*\right)\right]_i < 0\right\}, \quad J = \{1, 2, ..., n\} \backslash I \;\text{ e }\; V = \left\{x \in \mathbb{R}_+^n;\ x_i = 0,\ i \in I\right\}$$

De acordo com nossas hipóteses, $I \neq \emptyset$. Uma vez que $x_i^* \left[\nabla f\left(x^*\right)\right]_i = 0$, concluimos que $x_i^* = 0$, para todo $i \in I$, isto é, $x^* \in V$. Também, como consequência da proposição anterior e definição do conjunto J, temos que

$$x_j^* \geq 0, \qquad x_j^* \left[\nabla f\left(x^*\right)\right]_j = 0, \qquad \left[\nabla f\left(x^*\right)\right]_j \geq 0, \qquad \forall j \in J, \tag{2.28}$$

as quais são precisamente as condições KKT do problema

$$(P_V) \begin{cases} \min: & f(x) \\ s.a: & x \in V. \end{cases}$$

Note que o conjunto $V = \left\{x \in \mathbb{R}_+^n; x_i = 0, i \in I\right\}$ é convexo. Agora, a convexidade de f, implica que x^* é solução de (P_V).

Consideremos o seguinte subconjunto aberto de $\mathbb{R}_+^n$

$$B = \{x \in \mathbb{R}_+^n; \ \ [\nabla f(x)]_i < 0, \ \ \text{para } i \in I\}. \tag{2.29}$$

Por definição, $x^* \in B$, isto é, $B \neq \phi$. Tome a subsequência $\left\{x^{k_j}\right\}$ de $\left\{x^k\right\}$ tal que $\lim_{j \to +\infty} x^{k_j} = x^*$. Uma vez que B é um conjunto aberto de $\mathbb{R}_+^n$, podemos supor que $\left\{x^{k_j}\right\} \subset B$.

Para dar continuidade à demonstração dessa proposição, precisaremos dos resultados dos lemas a seguir.

Lema 2.4.2 *A sequência $\left\{x^k\right\}$, possui uma infinidade de termos $\left\{x^{k_l}\right\}_{l \in \mathbb{N}}$, tais que $x^{k_l} \notin B$, onde B é o conjunto definido em (2.29).*

Prova. Se assim não o fosse, existiria algum $k_0 \in \mathbb{N}$, tal que $x^k \in B$, para todo $k \geq k_0$. Pela definição das iterações, temos que

$$x^{k+1} = x^k - \frac{1}{\beta_k}\left(X^k\right)^r \nabla f\left(x^{k+1}\right). \tag{2.30}$$

Portanto, da equação (2.30) e do fato que $\{x^k\} \subset B$, temos que $x_i^{k+1} \geq x_i^k$, para todo $i \in I$ e $k \geq k_0$. Segue-se daí que, $x_i^{k+1} \geq x_i^{k_0} > 0$, para todo $i \in I$ e $k \geq k_0$. Para os termos da subsequência $\left\{x^{k_j}\right\}$, os quais convergem a x^*, temos que $x_i^{k_j+1} \geq x_i^{k_0} > 0$. Tomando $\lim x^{k_j+1}$ com $j \to +\infty$, obteremos

$$0 = x_i^* = \lim_{j\to+\infty} x_i^{k_j+1} = \lim_{j\to+\infty} x_i^{k_j} \geq x_i^{k_0} > 0,$$

o que é uma contradição, pois x^* é solução de (P_V) e $x^* \in B$, isto é,

$$x_i^*[\nabla f(x^*]_i = 0 \quad \text{e} \quad [\nabla f(x^*]_i < 0, \tag{2.31}$$

ou seja, $x_i^* = 0$. ■

Consideremos a seguinte sequência

$$p_j = \max\{q < k_j + 1;\ x^q \notin B\}.$$

(Convencione: $p_j = 0$, se $x^q \in B$ para todo $q < k_j + 1$). Então, da definição de p_j, para todo q tal que $p_j + 1 \leq q \leq k_j + 1$ temos $x^q \in B$. Portanto, segue-se que para todo q satisfazendo $p_j + 1 \leq q \leq q + 1 \leq k_j + 1$

$$x_i^{k_j+1} \geq x_i^{q+1} \geq x_i^q \geq x_i^{p_j+1}, \quad \forall i \in I. \tag{2.32}$$

Lema 2.4.3 $\lim_{j\to+\infty} p_j = +\infty$.

Prova. Se existe algum $j_0 \in \mathbb{N}$ suficientemente grande, tal que $p_j = p_{j_0}$, para todo $j \geq j_0$, então (2.32) implica

$$x_i^{k_j} \geq x_i^{q+1} \geq x_i^q \geq x_i^{p_j+1} \geq x_i^{p_{j_0}+1}, \quad \forall i \in I. \tag{2.33}$$

Agora, tomando $\lim x_i^{k_j}$ com $j \to +\infty$, segue-se que:

$$0 = x_i^* = \lim_{j\to+\infty} x_i^{k_j} \geq x_i^{p_{j_0}+1} > 0,$$

a qual é uma contradição, pela mesma razão apresentada na conclusão do lema anterior. ■

Continuando a demonstração da proposição em questão segue-se que de (2.32), para cada $i \in I$,

$$\lim_{j \to +\infty} x_i^{p_j+1} = 0 \tag{2.34}$$

e, usando o Corolário (2.4.8), temos que

$$\lim_{j \to +\infty} x_i^{p_j} = 0, \quad \forall i \in I. \tag{2.35}$$

Pela definição de p_j, $x^{p_j} \notin B$, para todo $j \in \mathbb{N}$. Para simplificar a notação usada, considere $\{x^{p_l}\}$ subsequência de $\{x^{p_j}\}$ e suponha que $x^{p_l} \to \overline{x}$. Uma vez que B é aberto, temos $\overline{x} \notin B$, pois $\{x^{p_l}\} \subset \{x^{p_j}\}$. Como $f\left(x^k\right) \to f^*$, quando $k \to +\infty$, temos

$$f\left(x^*\right) = f\left(\overline{x}\right) = \lim_{k \to +\infty} f\left(x^k\right) = f^*. \tag{2.36}$$

Do limite (2.35), segue-se que

$$\overline{x}_i = \lim_{l \to +\infty} x_i^{p_l} = \lim_{j \to +\infty} x_i^{p_j} = 0, \quad \forall i \in I. \tag{2.37}$$

Concluimos que $\overline{x} \in V$ e $f\left(\overline{x}\right) = f\left(x^*\right)$. Isto significa que $\overline{x}$ resolve (P_V). Observemos que se f fosse estritamente convexa já teríamos a contradição, uma vez que $\overline{x} \neq x^*$ $(\overline{x} \notin B \;\text{ e }\; x^* \in B)$ e ambos solucionam o problema (P_V). Mas devemos trabalhar somente com a hipótese de convexidade da função f. Uma vez que, $\overline{x} \notin B$, temos $\left[\nabla f\left(\overline{x}\right)\right]_s \geq 0$, para algum $s \in I$.

Consideremos o seguinte problema

$$(P_W) \begin{cases} \min: & f(x) \\ s.a: & x \in W \end{cases}$$

onde $W = \left\{x \in \mathbb{R}^n_+;\ x_i = 0,\ i \in I \backslash \{s\}\right\}$. As condições de otimalidades (KKT) de (P_W) são:

$$x_j \geq 0, \qquad [\nabla f(x)]_j \geq 0, \qquad x_j [\nabla f(x)]_j = 0, \qquad j \in J \cup \{s\}. \tag{2.38}$$

Claramente, vemos que $\overline{x}$ satisfaz às condições de otimalidade (KKT) do problema $((P_W)$. Devido à convexidade de f (e como $\overline{x} \in W$) temos que $\overline{x}$ é uma solução de $((P_W)$. Como $V \subset W$, $x^* \in V$ e

$$f(\overline{x}) = f(x^*) = \lim_{k \to +\infty} f\left(x^k\right) = f^*,$$

segue-se que x^*, também, satisfaz às condições gerais de otimalidades do problema (P_W), isto é, x^* é solução do problema (P_W). Em particular, $[\nabla f(x^*)]_s \geq 0$. Mas isto é falso, uma vez que $s \in I$. A demonstração está completa. ∎

Proposição 2.4.5 *Se f é estritamente convexa, então $\{x^k\}$ converge para uma única solução do problema (2.1).*

Prova. Note que a convexidade estrita de f e a existência de soluções do problema (2.1) implica que todos os conjuntos de nível da função objetivo f são limitados. Pela Proposição 2.4.4, os pontos de acumulação de $\{x^k\}$, são soluções do problema (2.1) e devido à convexidade estrita de f, a solução x^* do problema (2.1) é única. Portanto, $\lim_{k \to +\infty} x^k = x^*$. ∎

2.5 Aplicação: Caso Quadrático

Nesta secção, consideremos o caso particular em que a função objetivo é dada por

$$f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle c, x\rangle + \gamma \tag{2.39}$$

onde Q is simétrica não-negativa e γ é um número real. Mostraremos que a sequência $\{x^k\}$ tem uma expressão explícita, bem como os precondicionadores β_k, o que é bastante vantajoso do ponto de vista computacional.

De fato, da definição do Algoritmo 2.3.1, temos que o iterado x^{k+1} é obtido como solução da equação

$$\nabla f(x^{k+1}) + \beta_k (X^k)^{-r}(x^{k+1} - x^k) = 0, \tag{2.40}$$

onde $\nabla f(x^{k+1}) = Qx^{k+1} + c$. Substituindo esta última iqualdade em 2.40 obtemos

$$Qx^{k+1} + c + \beta_k (X^k)^{-r}(x^{k+1} - x^k) = 0. \tag{2.41}$$

Por outro lado, como $\nabla f(x^k) = Qx^k + c$, resulta que $c = \nabla f(x^k) - Qx^k$. Dessa forma a equação 2.41 torna-se

$$(Q + \beta_k (X^k)^{-r})(x^{k+1} - x^k) = -\nabla f(x^k). \tag{2.42}$$

Uma vez que Q é não-negativa e $\beta_k (X^k)^{-r}$ é positiva, temos que $Q + \beta_k (X^k)^{-r}$ é positiva para todo $k \in \mathbb{N}$, consequentemente, inversível.
Portanto a equação (2.42) torna-se

$$x^{k+1} = x^k - (Q + \beta_k (X^k)^{-r})^{-1} \nabla f(x^k). \tag{2.43}$$

Por outro lado, como

$$\left(Q + \beta_k (X^k)^{-r}\right)^{-1} = \frac{1}{\beta_k}\left(I + \frac{1}{\beta_k}(X^k)^r Q\right)^{-1} (X^k)^r, \tag{2.44}$$

segue-se que

$$x^{k+1} = x^k - \frac{1}{\beta_k}\left(I + \frac{1}{\beta_k}(X^k)^r Q\right)^{-1} (X^k)^r \nabla f(x^k) \tag{2.45}$$

Na definição dos parâmetros β_k, a constante L pode ser trocada por $||Q||_2$.

Capítulo 3

Taxa de Convergência

3.1 Introdução

Nesse capítulo analisaremos a taxa de convergência do algoritmo proposto. Mostraremos que o algoritmo converge linearmente no interior do $\mathbb{R}^n_+$, mediante a hipótese de convexidade forte da função objetivo na solução do problema (2.1). Na fronteira do conjunto $\mathbb{R}^n_+$, para todo $r > 1$, mostraremos convergência sublinear da sequência gerada pelo algoritmo, enquanto que no caso específico $r = 1$, nas componentes nulas da solução, a convergência linear permanece.

3.2 Taxa de Convergência

Proposição 3.2.1 *Suponha que a sequência $\{x^k\}$ converge e seja x^* seu limite. Se*

1. *f é duas vezes continuamente diferenciável e fortemente convexa em x^*,*

2. *$x^* > 0$,*

então existe uma norma $|||\cdot|||$ tal que

$$\limsup_{k \to +\infty} \frac{|||\, x^{k+1} - x^* \,|||}{|||\, x^k - x^* \,|||} \leq \left(1 + \frac{\lambda \xi}{C}\right)^{-1} < 1, \tag{3.1}$$

onde $\lambda = (\min(x^*))^r > 0$, ξ é o menor autovalor de $\nabla^2 f(x^*)$ e C é tal que $\beta_k \leq C$.

Prova. Pela proposição 2.4.3, desde que $x^* > 0$, temos $\nabla f(x^*) = 0$. Expandindo $\nabla f(x)$ em torno de x^* obtemos

$$\nabla f(x) = \nabla^2 f(x^*)(x - x^*) + v, \tag{3.2}$$

onde $\nabla^2 f(x^*)$ é simétrica positiva definida por (*1.*) e $\|v\| \leq \mu \|x - x^*\|^2$ para algum μ e todo x em uma vizinhança de x^*. Pode-se mostrar facilmente, usando fórmula de Taylor com resto integral, que μ depende apenas da constante de Lipschitz L de $\nabla f(\cdot)$.

Da definição do algoritmo, temos que

$$\nabla f(x^{k+1}) = -\beta_k (X^k)^{-r}(x^{k+1} - x^k). \tag{3.3}$$

Substituindo essa equação na igualdade (3.2), obtemos

$$\nabla^2 f(x^*)(x^{k+1} - x^*) = -\beta_k (X^k)^{-r}(x^{k+1} - x^k) - v^k, \tag{3.4}$$

onde v^k satisfaz

$$\|v^k\| \leq \mu \|x^{k+1} - x^*\|^2. \tag{3.5}$$

A equação (3.4) pode ser reescrita na seguinte forma

$$\left(\nabla^2 f(x^*) + \beta_k (X^k)^{-r}\right)(x^{k+1} - x^*) = \beta_k (X^k)^{-r}(x^k - x^*) - v^k. \tag{3.6}$$

Para todo k, temos que a matriz

$\nabla^2 f(x^*) + \beta_k (X^k)^{-r}$ é inversível (na verdade, tal matriz é simétrica definida positiva). Portanto a equação (3.6), é equivalente a

$$x^{k+1} - x^* = \left(\nabla^2 f(x^*) + \beta_k (X^k)^{-r}\right)^{-1}\left(\beta_k (X^k)^{-r}(x^k - x^*) - v^k\right). \tag{3.7}$$

Dessa última igualdade e da fórmula em (2.44), temos

$$x^{k+1} - x^* = \left(I + \frac{1}{\beta_k}(X^k)^r \nabla^2 f(x^*)\right)^{-1}(x^k - x^*) - \left(\nabla^2 f(x^*) + \beta_k (X^k)^{-r}\right)^{-1} v^k. \tag{3.8}$$

Agora aplicando norma em ambos os lados da igualdade anterior, obtemos

$$\begin{aligned} \|x^{k+1} - x^*\| &\leq \left\|(I + \tfrac{1}{\beta_k}(X^k)^r \nabla^2 f(x^*))^{-1}\right\| \|x^k - x^*\| + \\ &+ \left\|(\nabla^2 f(x^*) + \beta(X^k)^{-r})^{-1}\right\| \|v^k\|. \end{aligned} \tag{3.9}$$

(Observe que esta última desigualdade é válida para qualquer norma.) Pela Proposição (1.2.1) existem matrizes P, D, P_k e D_k tais que

$$(X^*)^{-r}\nabla^2 f(x^*) = PDP^{-1} \quad \text{e} \quad (X^k)^{-r}\nabla^2 f(x^*) = P_k D_k P_k^{-1}, \tag{3.10}$$

onde D, D_k são diagonais positivas. Uma vez que

$$\lim_{k\to+\infty} (X^k)^{-r} = (X^*)^{-r}, \tag{3.11}$$

as matrizes P_k e D_k podem ser escolhidas de forma que

$$\lim_{k\to+\infty} P_k = P \quad e \quad \lim_{k\to+\infty} D_k = D. \tag{3.12}$$

Seja $|||\cdot|||$ definida por $|||\, x \,||| = \|P^{-1}x\|$. Supondo que μ foi escolhido de forma que (3.5) ocorra para esta nova norma, temos que

$$\begin{aligned} |||\, x^{k+1} - x^* \,||| &\leq |||\, \left(I + \tfrac{1}{\beta_k}(X^k)^r \nabla^2 f(x^*)\right)^{-1} |||\,|||\, x^k - x^* \,||| + \\ &+ \mu |||\, \nabla^2 f(x^*) + \beta(X^k)^{-r})^{-1} \,|||\,|||\, x^{k+1} - x^* \,|||^2. \end{aligned} \tag{3.13}$$

Usando a Proposição 1.2.2 com

$$A = (X^k)^r, \quad B = \nabla^2 f(x^*) \quad e \quad M = P \tag{3.14}$$

na desigualdade acima, obtemos

$$\begin{aligned} |||\, x^{k+1} - x^* \,||| &\leq \|P^{-1}P_k\|_2 \|P_k^{-1}P\|_2 (1 + \tfrac{(\min(x^k))^r \xi}{\beta_k})^{-1} |||\, x^k - x^* \,||| + \\ &+ \mu \|P^{-1}\|_2 \|P\|_2 \xi^{-1} |||\, x^{k+1} - x^* \,|||^2. \end{aligned} \tag{3.15}$$

Desde que $lim_{k\to+\infty} x^k = x^*$, para k suficientmente grande, segue-se que

$$1 - \frac{\mu}{\xi}\|P^{-1}\|_2 \|P\|_2 |||\, x^{k+1} - x^* \,||| > 0. \tag{3.16}$$

Daí, usando a limitação dos parâmetros de regularização $\beta_k \leq C$, (3.15) pode ser reescrita como

$$\frac{\| x^{k+1} - x^* \|}{\| x^k - x^* \|} \leq \frac{||P^{-1}P_k||_2||P_k^{-1}P||_2}{1 - \frac{\mu}{\xi}||P^{-1}||_2||P||_2\| x^{k+1} - x^* \|}\left(1 + \frac{(\min(x^k))^r\xi}{C}\right)^{-1} \tag{3.17}$$

Observando que

$$lim_{\rightarrow+\infty}P_k = P, \quad lim_{\rightarrow+\infty}||P^{-1}P_k||_2 = lim_{\rightarrow+\infty}||P_k^{-1}P||_2 = 1 \tag{3.18}$$

e

$$lim_{\rightarrow+\infty}(\min(x^k))^r = (\min(x^*))^r \tag{3.19}$$

obtemos ao limite superior em (3.17)

$$\limsup_{k\rightarrow+\infty}\frac{\| x^{k+1} - x^* \|}{\| x^k - x^* \|} \leq \left(1 + \frac{\xi(\min(x^*))^r}{C}\right)^{-1} < 1. \tag{3.20}$$

■

O resultado a seguir caracteriza o comportamento assintótico da sequência no bordo do conjunto $\mathbb{R}^n_+$, mostrando assim que as condições da proposição anterior não são muito restritivas, e que não podem ser facilmente retiradas.

Proposição 3.2.2 *Suponha que a sequência $\{x^k\}$ converge a x^* e que exista l tal que $x_l^* = 0$. Então:*

1. *para qualquer número real $r > 1$, $\{x^k\}$ converge sublinearmente;*

2. *para $r = 1$, a convergência linear, nas componentes nulas da solução x^*, pode ser obtida desde que suponhamos a complementariedade estrita (isto é, se $x_l^* = 0$, então $[\nabla f(x^*)]_l > 0$, $\quad$ se $x_l^* > 0$, então $[\nabla f(x^*)]_l = 0$).*

Prova. Pela definição de x^{k+1} em (2.7) temos que

$$x^{k+1} = x^k - \frac{1}{\beta_k}(X^k)^r \nabla f(x^{k+1}) \tag{3.21}$$

Portanto, para todo $l \in \mathbb{N}$,

$$x_l^{k+1} = x_l^k(1 - \frac{1}{\beta_k}(x^k)_l^{r-1}\left[\nabla f(x^{k+1})\right]_l). \tag{3.22}$$

No primeiro caso, usando $x_l^* = 0$ e $r > 1$, obtemos

$$\lim_{k\to+\infty} \frac{|x_l^{k+1} - x_l^*|}{|x_l^k - x_l^*|} = lim_{k\to+\infty}(1 - \frac{1}{\beta_k}(x^k)_l^{r-1}\left[\nabla f(x^{k+1})\right]_l) = 1, \tag{3.23}$$

uma vez que f é continuamente diferenciável e $\{\beta_k\}$ é limitada.

No segundo caso, a igualdade (3.22) torna-se

$$x_l^{k+1} = x_l^k(1 - \frac{1}{\beta_k}\left[\nabla f(x^{k+1})\right]_l). \tag{3.24}$$

Seja C tal que $\beta_k \leq C$, para todo $k \in \mathbb{N}$. Uma vez que, $[\nabla f(x^*)]_l > 0$, temos que $\left[\nabla f(x^{k+1})\right]_l > 0$, para k suficientemente grande. Portanto, da equação (3.24) temos,

$$\frac{|x_l^{k+1} - x_l^*|}{|x_l^k - x_l^*|} \leq 1 - \frac{\left[\nabla f(x^{k+1})\right]_l}{C}, \tag{3.25}$$

para k suficientemente grande. Ao limite com $k \to +\infty$, segue-se que

$$\lim_{k\to+\infty} \frac{|x_l^{k+1} - x_l^*|}{|x_l^k - x_l^*|} \leq 1 - \frac{[\nabla f(x^*)]_l}{C} < 1. \tag{3.26}$$

■

Como observado por A. Iusem em ([IT/93]), a convergência linear pode falhar, mesmo no caso em que $x_j^* > 0$, $\forall j \in 1, 2, .., n$, se não supusermos a convexidade forte de f em x^*. A título de verificação em nosso caso, consideremos a família de funções definidas por

$$f(x) = \frac{1}{2m}\sum_{j=1}^{n}(x_j - 1)^{2m}, \quad \text{para } m \in \mathbb{N}, \quad m \geq 2 \tag{3.27}$$

(ver [IT/93]). Da definição de nosso algoritmo, temos que

$$x_j^{k+1} = x_j^k - \frac{1}{\beta_k}(x_j^k)^r \left[\nabla f(x^{k+1})\right]_j. \tag{3.28}$$

Substituindo $\left[\nabla f(x^{k+1})\right]_j = (x_j^{k+1} - 1)^{2m-1}$ na equação acima, obtemos

$$\frac{x_j^{k+1} - 1}{x_j^k - 1} = \frac{1}{1 + \frac{1}{\beta_k}(x_j^k)^r (x_j^{k+1} - 1)^{2(m-1)}}. \tag{3.29}$$

Portanto, segue-se que

$$\lim_{k\to\infty} \frac{|x_j^{k+1} - 1|}{|x_j^k - 1|} = 1, \tag{3.30}$$

ou seja, a sequência $\{x^k\}$ converge sublinearmente, para qualquer valor do parâmetro $r \geq 1$.

Observe que a função dada é estritamente convexa para cada valor $m \geq 2$, garantindo assim que convexidade estrita não é suficiente para a garantia da taxa de convergência linear.

Capítulo 4

Versão Inexata

4.1 Introdução

Neste capítulo apresentaremos e analisaremos uma versão inexata para o algoritmo proposto no Capítulo 1. Mostraremos, como na versão exata, que a sequência gerada pelo método converge fracamente ao conjunto solução do problema (2.1). A sequência de erros proposta tende a zero, e, apesar de estar definida em termos do iterado x^{k+1}, o qual é desconhecido a priori, é de fácil obtenção na prática, como discutido em ([E/93]).

4.2 Definição da Versão Inexata do Algoritmo (2.3.1)

A versão inexata para o algoritmo (2.3.1) proposto acima será definido como segue:

Algoritmo 4.2.1 .

Inicialização:

$$x^0 > 0 \tag{4.1}$$

Passo Iterativo: *Dado* $x^k > 0$, *defina*

$$\beta_k = 2(\max(x^k))^{r-1}\left\|\nabla f\left(x^k\right)\right\| + (L+1)(\max(x^k))^r \tag{4.2}$$

Obtenha a sequência $(x^{k+1}, e^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^n_+$ *satisfazendo*

$$\nabla f(x^{k+1}) + \beta_k (X^k)^{-r}(x^{k+1} - x^k) = e^{k+1}, \tag{4.3}$$

onde $\{e^{k+1}\}$ *satisfaz*

$$\|e^{k+1}\| \leq \min\{\|\nabla f(x^k)\|, \|x^{k+1} - x^k\|\}. \tag{4.4}$$

Se $\nabla f(x^k) = 0$ *ou se* $x^k = x^{k+1}$ *pare, senão*

Faça: $k \leftarrow k+1$,

Repita.

4.3 Análise de Convergência do Algoritmo Inexato (4.2.1)

Inicialmente, precisamos garantir que a equação (4.3) possui sempre solução. Isso é uma consequência do resultado abaixo, enunciado num contexto mais geral:

Teorema 4.3.1 *(***Resultado de Existência***):*

Sejam $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ *um operador monótono maximal e* ∂g *o subdiferencial de uma função convexa própria fechada definida em* $\mathbb{R}^n$, *tais que* $domT \cap int\ domg \neq \phi$. *Se* ∂g *é sobrejetivo, então* $T + \partial g$ *é sobrejetivo.*

Prova. Ver ([BI/95]). ∎

Observe que no nosso caso,

$$T(x) = \nabla f(x) \quad e \quad g(x) = \frac{\beta_k}{2}\|x - x^k\|^2_{(X^k)^{-r}} \tag{4.5}$$

Uma vez que f é convexa, temos que $\nabla f(\cdot)$ é monótono maximal e, claramente, o subdiferencial de g,

$$\partial g(x) = \beta_k (X^k)^{-r}(x - x^k), \tag{4.6}$$

é sobrejetivo, para cada $k \in \mathbb{N}$.

Agora, queremos que o iterado e^{k+1} possa ser obtido, satisfazendo a inequação (4.4), numa quantidade finita de escolhas. Isso garantirá a boa definição da sequência (x^{k+1}, e^{k+1}), o que será demonstrado a seguir:

Lema 4.3.1 *A sequência $(x^{k+1}, e^{k+1}, \beta_k) \in \mathbb{R}^n \times \mathbb{R}^n_+ \times \mathbb{R}_+$, gerada pelo Algoritmo (4.2.1), obtida em (4.3) - (4.4), está bem definida.*

Prova. Com base no Teorema (4.3.1), basta provarmos que em cada iteração o iterado e^{k+1} pode ser obtido após uma quantidade finita de escolhas, até que a inequação (4.4) seja satisfeita. De fato, suponhamos que, na iteração k, isso não ocorre, ou seja, suponhamos que existam infinitos índices $j \in \mathbb{N}$ tais que

$$\|e^{k+j}\| > \min\{\|\nabla f(x^k)\|, \|x^{k+j} - x^k\|\}. \tag{4.7}$$

Podemos, sem perda de generalidade, escolher a sequência $\{e^{k+1}\} \in \mathbb{R}^n_+$, de forma que $e^k \to 0$. Assim temos que

$$0 = \lim_{j \to +\infty} \|e^{k+j}\| \geq \lim_{j \to +\infty} \min\{\|\nabla f(x^k)\|, \|x^{k+j} - x^k\|\}. \tag{4.8}$$

Portanto, segue-se que

$$\|\nabla f(x^k)\| = 0 \quad \text{ou} \quad \lim_{j \to +\infty} \|x^{k+j} - x^k\| = 0. \tag{4.9}$$

Se $\|\nabla f(x^k)\| = 0$ o algoritmo pára, pois $x^k > 0$ será solução do problema (2.1). Por outro lado, se $\lim_{j \to +\infty} \|x^{k+j} - x^k\| = 0$, temos que $x^{k+j} \to x^k$ e, usando a equação

(4.3) segue-se, mais uma vez, que $\|\nabla f(x^k)\| = 0$. Portanto, ou o iterado e^{k+1} pode ser obtido, satisfazendo (4.4), numa quantidade finita de vezes ou o termo $x^k > 0$ é uma solução do problema (2.1). ■

Proposição 4.3.1 *Considerando a terna* $(x^{k+1}, e^{k+1}, \beta_k) \in \mathbb{R}^n \times \mathbb{R}^n_+ \times \mathbb{R}_+$ *obtida em (4.2)-(4.4), temos que:*

1. $x^{k+1} > 0$;
2. $\left(\frac{2\|\nabla f(x^k)\|}{\max(x^k)} + L\right) \|x^k - x^{k+1}\|^2 \leq f(x^k) - f(x^{k+1})$;
3. $\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty$;
4. $\lim_{k\to+\infty}(x^{k+1} - x^k) = 0$;
5. $\lim_{k\to+\infty} e^{k+1} = 0$;

Além disso, obtemos que

6. $\sum_{k=0}^{+\infty} |\langle e^{k+1}, x^k - x^{k+1}\rangle| < +\infty$;
7. $\sum_{k=0}^{+\infty} \beta_k \|x^{k+1} - x^k\|^2_{(X^k)^{-r}} < +\infty$.

Prova.

1. De (4.3) temos que

$$x^{k+1} = x^k - \frac{1}{\beta_k}(X^k)^r(\nabla f(x^{k+1}) - e^{k+1}). \tag{4.10}$$

Procedendo da mesma forma descrita na versão exata na Proposição (2.4.1), temos que

$$\|(X^k)^{-1}(x^{k+1} - x^k)\| \leq \frac{(\max(x^k))^{r-1}(\|\nabla f(x^k)\| + \|e^{k+1})\|)}{\beta_k - L(\max(x^k))^r}. \tag{4.11}$$

Então, usando a definição dos parâmetros β_k e a condição (4.4) sobre a sequência $\{e^k\}$, obtemos

$$\|(X^k)^{-1}(x^{k+1} - x^k)\| < 1, \tag{4.12}$$

ou seja,

$$0 < x_i^{k+1} < 2x_i^k \quad (1 \leq i \leq n). \tag{4.13}$$

2. Uma vez que f é uma função convexa, pela desigualdade do gradiente, temos que

$$\langle \nabla f(x^{k+1}), x^k - x^{k+1} \rangle \leq f(x^k) - f(x^{k+1}) \tag{4.14}$$

Usando a equação (4.3) na equação acima, obtemos

$$\beta_k \|x^k - x^{k+1}\|^2_{(X^k)^{-r}} + \langle e^{k+1}, x^k - x^{k+1} \rangle \leq f(x^k) - f(x^{k+1}). \tag{4.15}$$

Mas sabemos que

$$\beta_k \|x^k - x^{k+1}\|^2_{(X^k)^{-r}} \geq \frac{\beta_k}{(\max(x^k))^r} \|x^k - x^{k+1}\|^2 \tag{4.16}$$

e, pela desigualdade de Cauchy-Schwarz,

$$\langle e^{k+1}, x^k - x^{k+1} \rangle \geq -\|e^{k+1}\| \|x^k - x^{k+1}\|. \tag{4.17}$$

Portanto, substituindo estas duas últimas desigualdades em (4.15), obtemos

$$\frac{\beta_k - (\max(x^k))^r}{(\max(x^k))^r} \|x^k - x^{k+1}\|^2 \leq f(x^k) - f(x^{k+1}). \tag{4.18}$$

Pela definição de β_k, segue-se que

$$\left(\frac{2\|\nabla f(x^k)\|}{\max(x^k)} + L \right) \|x^k - x^{k+1}\|^2 \leq f(x^k) - f(x^{k+1}). \tag{4.19}$$

3. Segue-se da útima desigualdade acima, pois estamos supondo que o conjunto solução do Problema (2.1) é não vazio, mais especificamente,

$$\sum_{k=0}^{+\infty} \|x^k - x^{k+1}\|^2 \leq \frac{f(x^0) - f^*}{L + 1}. \tag{4.20}$$

4. Imediato do item *3.* acima.

5. Imediato da definição dos erros, pois $\|e^{k+1}\| \leq \|x^k - x^{k+1}\|$ e do item anterior.

6. Usando a definição dos erros em (4.4) e usando a desigualdade de Cauchy-Schwarz, temos que

$$|\langle e^{k+1}, x^k - x^{k+1}\rangle| \leq \|x^k - x^{k+1}\|^2. \tag{4.21}$$

A prova segue do item *3.* demonstrado anteriormente.

7. Segue, imediatamente, da desigualdade (4.15) e da somabilidade das séries $\sum_{k=0}^{+\infty}\langle e^{k+1}, x^k - x^{k+1}\rangle$ e $\sum_{k=0}^{+\infty}(f(x^k) - f(x^{k+1}))$. ■

Proposição 4.3.2 *Todo ponto limite x^* da sequência $\{x^k\}$, gerada pelo algoritmo (4.3), é solução do problema (2.1).*

Prova. Inicialmente, observemos que a sequência $\{x^k\}$, gerada pelo algoritmo (4.3), é limitada, pois estamos supondo nível-limitação da função objetivo f, portanto o conjunto de seus pontos de acumulação é não-vazio. Seja x^* um ponto de acumulação de $\{x^k\}$. Devemos mostrar que as condições de KKT, em x^*, são satisfeitas.

A prova segue os mesmos passos daquela desenvolvida no caso da versão exata, com suas devidas modificações.

Inicialmente, é óbvio que $x_i^* \geq 0$, para todo $i \in \{1, 2, ..., n\}$.

De (4.3) temos que

$$x^{k+1} = x^k - \frac{1}{\beta_k}(X^k)^r(\nabla f(x^{k+1}) - e^{k+1}). \tag{4.22}$$

Como

$$\lim_{k\to+\infty}(x^{k+1} - x^k) = 0 \quad \text{e} \quad \lim_{k\to+\infty} e^{k+1} = 0 \tag{4.23}$$

e, pela limitação da sequência de parâmetros $\{\beta_k\}$, nos termos da subsequência $\{x^{k_j}\}$ que converge a x^*, segue-se de (4.22) que,

$$(x^*)_i^T [\nabla f(x^*)]_i = 0. \tag{4.24}$$

Assim, $x_i^* [\nabla f(x^*)]_i = 0$, para todo $i \in \{1, 2, ..., n\}$.

Para provar que $[\nabla f(x^*)]_i \geq 0$, basta que sigamos os mesmos passos utilizados na prova do caso exato observando que, para a validade dos lemas (2.4.2) e (2.4.3), é necessário tomarmos a sequência $\{e^k\}$ com $e_i^k \geq 0$, $(i = 1, 2, ..., n)$, para k suficientemente grande. ■

Conclusões

Nesse trabalho, desenvolvemos uma teoria de convergência, para essa nova classe de métodos, em que os resultados assemelham-se aos resultados obtidos por A. Iusem em ([I/95a]). Temos a propriedade de que no caso quadrático, o método possui uma expressão explícita para os iterados, o que é bastante vantajoso do ponto de vista computacional.

Com relação ao método estudado por C. Sagastizábal et alli, em ([BGLS/95]) para o caso irrestrito, esta classe de métodos mostra-se como uma extensão daquela para o caso em que a minimizacão é feita no ortante positivo.

Esse procedimento, difere dos demais, por não se fazer penalização. A viabilidade é obtida alterando-se, convenientemente, os parâmetros de regularização β_k. Porém, tais parâmetros ficam dados em termos da constante de Lipschitz, do gradiente, em geral desconhecida.

Como vimos, no Capítulo 4, introduzimos uma versão inexata para esta classe de métodos e obtemos resultados de convergência fraca, ou seja, a sequência $\{x^k\}$ gerada pelo algoritmo satisfaz $\lim_{k\to+\infty}(x^{k+1}-x^k)=0$, e seus pontos de acumulação satisfazem às condições de KKT do problema (2.1), para uma escolha conveniente da sequência de erros $\{e^k\}$.

A seguir relacionamos alguns pontos que, naturalmente, podem ser objetos de estudos futuros:

1. A análise desse algoritmo, sem a hipótese de nivel-limitação da função objetivo f.

2. Obter uma expressão para os parâmetros de regularização β_k, independente da constante de Lipschitz.

3. Análise do caso quadrático, com o objetivo de se estabelecer a convergência da sequência gerada pelo método.

4. Aplicar o algoritmo para o problema com restrições lineares: ou seja, determinar

$$f^* = \arg\min \{f(x); x \in C\}$$

onde $C = \{x \in \mathrm{R}^n_+; Ax \leq b\}$, com A_{mxn} uma matriz de posto máximo, supondo, a priori, que

$$int(C) = \{x \in \mathrm{R}^n; Ax < b\} \neq \phi.$$

5. Obter, se possível, por uma modificação conveniente da sequência $\{\beta_k\}$, convergência superlinear do método.

6. Estender o algoritmo para funções convexas, não necessariamente continuamente diferenciáveis.

7. Estender o algoritmo para o caso de operadores monótonos maximais.

8. Obter uma formulação primal-dual do algoritmo.

9. Propor novas versões inexatas, nas quais os erros dependam apenas do iterado corrente x^k.

Referências Bibliográficas

[BGLS/95] Bonnans, J. F., Gilbert, J. Ch., Lemaréchal C. e Sagastizábal, C. A.: **A Family of Variable Metric Proximal Methods.** Mathematical Programming 68, 15-47, (1995).

[BI/95] Burachik, R. S. and Iusem, A. N. **A generalized proximal point algorithm for the variational inequality problem in a Hilbert space.** IMPA-Instituto de Matemática Pura e Aplicada, Rio de Janeiro, (1995).

[BL/89] Bayer, D. A. and Lagarias, J. C.: **The Nonlinear Geometry of Linear Programming I Affine and Projective Scalling Trajectories.** Transactions of the American Mathematical Society, vol 314, número 2, August 1989.

[CT/93] Chen, G. and Teboulle, M. **Convergence Analysis of a Proximal-like Minimization Algorithm Using Bregman Functions.** SIAM J. Optim. **3**, 538-543, (1993).

[CZ/92] Censor, Y. and Zenios, S.A. **The Proximal Minimization Algorithm with D-Functions.** J. Optim. Theory and Apllic. **73**, 451-464, (1992).

[E/93] Eckestein, J. **Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming.** Math. Oper. Res. **18**, 202-226, (1993).

[EG/90] Eggermont, P. P. B.: **Multiplicative Iterative Algorithms for Convex Programming**. Linear Algebra Appl. 130, 25-42, (1990).

[FO/97] Orizon Pereira Ferreira e P. R. Oliveira **Programação Matemática em Variedades Riemannianas: Algoritmos Subgradiente e Ponto Proximal**, Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro (1997).

[G/82] Gabay, D.: **Minimizing a Differentiable Function over a Diferential Manifolds.** Journal of Optimization Theory and Applications, volume 37, número 2, 1982.

[GC/90] C. C. Gonzaga and L. A. Carlos: **A Primal Affine-Scalling Algorithms for Linearly Constrained Convex Programs**. Technical Report - 230/90, COPPE/UFRJ, Rio de Janeiro, (1990).

[HJ/85] Horn, R. A. and Johnson, C. R.: **Matrix Analysis**, Cambridge Univ. Press, Cambridge (1985).

[I/95a] Iusem, A. N.: **An Interior Multiplicative Method for Optimization Under Positivity Constraints**. Acta Applicadae Mathematicae 38, 163-184, (1995).

[I/95b] Iusem, A.: **Métodos de Ponto Proximal em Otimização**. 20 Colóquio Brasileiro de Matemática, IMPA, 24-28 de Julho de 1995.

[IST/94] Iusem, A. N., Svaiter, B. F. e Teboulle, M.: **Entropy-Like Proximal Methods in Convex Programming**. Mathematics of Operations Research 19, 790-814, (1994).

[IT/93] Iusem, A. N. and Teboulle, M.: **On the Convergence Rate of Entropic Proximal Optimization Methods.** Comp. Appl. Math., v.12, n.2, 153-168, (1993).

[L/72] Luemberger, D. G.: **The Gradient Projection Methods along Geodesics.** Management Science 18, 620-631, 1972.

[M/78] Martinet, B.(1978). **Pertubation des Methodes d'Optimisation.** Application, R.A.I.R.O. Numer. Anal., manuscrit reçu le 20 juin 1975.

[MO/62] Moreau, J.-J. **Fonctions convexes duales et points proximaux dans un espace hilbertien**, Comptes Rendus de l'Académie des Science de Paris, 256, 1069-1071, (1962).

[MO/65] Moreau, J.-J. **Proximité et dualité dans um espace hilbertien**, Bulletin de la Societé Mathématique de France, 93, 273-299, (1965).

[R/70] Rockafellar, R.T. **Convex Analysis.**Princeton University Press, New Jersey (1970).

[R/76a] Rockafellar, R.T. **Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming.**Math. Oper. Res. **1** 97-116, (1976).

[R/76b] Rockafellar, R.T. **Monotone Operators and the Proximal point Algorithm.** SIAM J. Control. Optim. **14**, 877-898, (1976).

[RT/96] Rapcsác, T. and Thang, T. T.: **A Class of Polinomial Variable Metric Algorithms for Linear Optimization**, Mathematical Programming 74, 319-331, 1996.

[XLO/98] J. X. da Cruz Neto, L. L. de Lima and P. R. Oliveira: **Geodesic Algorithms in Riemannian Geometry**, Balkan Journal of Geometry and its Applications (BJGA), vol.03, n.2, 89-100, (1998).

[XO/95] J. X. da Cruz Neto and P. R. Oliveira: **Métodos Geodesicos na Programação Matemática**, Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro (1995).

[UL/93] J. B. Hiriart-Urruty and C. Lemaréchal. **Convex Analysis and Minimization Algorithms**, vols. 1 and 2 (Springer-Verlag, Berlin, Heidelberg, New York, 1993).

[WO/00] A. Wagner Martins Pinto e P. R. Oliveira: **Sobre Abertura e Densidade de Politopos Não-Degenerados e uma Família de Algoritmos Primais de Pontos Interiores para Programação Convexa**, Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro (2000).