Método de Newton para encontrar zeros de uma classe especial de funções semi-suaves

Dissertação de Mestrado por

Maurício Silva Louzeiro

Orientado por

Prof. Dr. **Orizon Pereira Ferreira**

Financiado por
CAPES

IME - Instituto de Matemática e Estatística
Universidade Federal de Goiás
Goiânia, Goiás, Brasil
Março 2016

Maurício Silva Louzeiro

Método de Newton para encontrar zeros de uma classe especial de funções semi-suaves

Dissertação apresentada ao Programa de Pós-Graduação do
Instituto de Matemática e Estatística da Universidade
Federal de Goiás, como requisito parcial para obtenção
do título de Mestre em Matemática.
Área de concentração: Otimização.
Orientador: Prof.Dr. Orizon Pereira Ferreira.

Goiânia
2016

Dedicado a:

Meu pai Nelson, minha mãe Ivonete e todos que torcem por mim.

Agradecimentos

A Deus, por tudo.
À minha mãe Ivonete e meu pai Nelson por toda dedicação, confiança e apoio de todas as maneiras.
Aos meus irmãos Jutay e Samara, em especial o primeiro, pelos ensinamentos, conselhos e pelo convívio em Goiânia.
À minha família em geral, primos, tios e avós, pelo apoio.
Ao Marcelo Coelho, meu professor de matemática do ensino médio, pelo incentivo e confiança, fundamentais para que eu seguisse essa área.
Ao meu orientador Dr.Orizon pela oportunidade de trabalhar ao seu lado e pela conduta das orientações.
À banca examinadora em geral, por ler esta dissertação e pelas sugestões dadas que contribuiram bastante para melhoria dela.
Aos amigos e colegas, tanto da minha cidade natal, Corrente-PI, como de Goiânia-GO, responsáveis pelas descontrações.
À CAPES pelo auxílio financeiro.

Abstract

In this work, we will study a new strategy to minimize a convex function on a simplicial cone. This method consists in to obtain the solution of a minimization problem through the root of a semi-smooth equation associated to its optimality conditions. To find this root, we use the semi-smooth version of the Newton's method, where the derivative of the function that defines the semi-smooth equation is replaced by a convenient Clarke subgradient. For the case that the function is quadratic, we will see that it allows us to have weaker conditions for the convergence of the sequence generated by the semi-smooth Newton's method. Motivated by this new minimization strategy we will also use the semi-smooth Newton's method to find roots of two special semi-smooth equations, one associated to x^+ and the another one associated to $|x|$.

Keywords: Newton's method, semi-smooth equation, simplicial cone, quadratic function.

Resumo

Neste trabalho, estudaremos uma nova estratégia para minimizar uma função convexa sobre um cone simplicial. Este método consiste em obter a solução do problema de minimização através da raiz de uma equação semi-suave associada às suas condições de otimalidade. Para encontrar essa raiz, usaremos uma versão semi-suave do método de Newton, onde a derivada da função que define a equação semi-suave é substituída por um subgradiente de Clarke conveniente. Para o caso em que a função é quadrática, veremos que é possível obter condições mais fracas para a convergência da sequência gerada pelo método de Newton semi-suave. Motivados por esta nova estratégia de minimização também usaremos o método de Newton semi-suave para encontrar raízes de dois tipos específicos de equações semi-suaves, uma associada à x^+ e a outra associada à $|x|$.

Palavras-chave: método de Newton, equação semi-suave, cone simplicial, função quadrática.

Notações Básicas

Ω^c: complementar do conjunto Ω,

$\mathbb{R}^+$: conjunto dos números reais não negativos,

$P_D(x)$: projeção do vetor x no conjunto D,

I: matriz identidade,

$\mathbb{R}^n_+$: octante não negativo do espaço Euclideano $\mathbb{R}^n$,

x^+: vetor cuja i-ésima coordenada é $\max\{x_i, 0\}$,

x^-: vetor cuja i-ésima coordenada é $\max\{-x_i, 0\}$,

$|x|$: vetor cuja i-ésima coordenada é $|x_i|$,

$\mathbb{R}^{n\times m}$: espaço das matrizes $n \times m$ com coeficientes reais,

$\langle \cdot, \cdot \rangle$: produto interno canônico em $\mathbb{R}^n$,

$(e_i)_{i=1}^n$: base canônica do $\mathbb{R}^n$,

PG: progressão geométrica,

$\|\cdot\|$: norma euclidiana em $\mathbb{R}^n$,

$\partial\Phi(x)$: subdiferencial de Clarke da função Φ em x,

$co\{\}$: fecho convexo de um conjunto,

$\mathbb{C}^n$: espaço da funções reais n vezes continuamente diferenciáveis,

$F'(x)$: derivada da função F no ponto x.

Sumário

Capítulo 1

Introdução

Neste trabalho, estudamos uma versão semi-suave do método de Newton para resolver o problema de minimizar uma função convexa sobre um cone simplicial. O método de Newton semi-suave para resolver este problema foi recentemente usado em [2]. A estratégia consiste em, a partir das condições de otimalidade do problema de minimização, obter uma equação semi-suave cuja a raiz gera uma solução para problema original. Alguns dos primeiros a estudar esses tipos de equações semi-suaves usando métodos do tipo Newton para encontrar sua raiz, foram Brugnano e Casulli em [6] e Mangasarian em [13]. Testes numéricos apresentados nos artigos [6] e [13], indicam que esta abordagem tem vantagens potenciais sobre os métodos existentes para resolver o problema em certas instâncias. A principal vantagem é a capacidade de encontrar soluções para problemas de grande dimensão em relativamente poucas iterações. As propriedades de convergência do método explicam parcialmente este bom comportamento. Embora neste trabalho estamos interessados apenas na parte teórica desse problema, vale a pena ressaltar que existem muitos trabalhos nessa mesma direção que também focam nos experimentos numéricos, veja [1], [2], [5], [6] e [13].

Como aplicação da metodologia empregada para resolver o problema de minimizar uma função convexa sobre um cone simplicial, vamos estudar o problema de projeção sobre um cone simplicial. É sabido, que este problema é equivalente a minimizar uma função quadrática estritamente convexa sobre um cone simplicial. Com base nisso, vale lembrar que calcular a projeção sobre um cone simplicial em geral é um problema difícil e bem caro computacionalmente. Outras maneiras de resolver o problema em questão incluem algoritmos para resolver LCP's (problemas de complementaridade linear) e métodos especiais com base em sua geometria, veja [4], [14] e [15]. Ainda podem ser incluidos os elegantes métodos baseados no algoritmo clássico de Von Neumann, veja [8]. No entanto, estes métodos são igualmente caros.

A dissertação está organizada da seguinte forma. No Capítulo 2 são feitas definições, exemplos e resultados auxiliares, que tem como principais referências bibliográficas [3], [7], [10] e [16]. No Capítulo 3 introduzimos um método para minimizar uma função convexa sobre um cone simplicial e tratamos as particularidades obtidas no caso em que a função convexa é a quadrática, ver [2]. No Capítulo 4 usamos o método de Newton semi-suave para estudar uma equação semi-suave do tipo modular, cujas principais referências são [12] e [13]. Além disso, estudaremos uma equação semi-suave com o operador x_+, que pode ser vista em [1] e [6], ambos trabalhando maneiras diferentes de resolvê-la.

Capítulo 2

Notações e resultados auxiliares

Neste capítulo serão enunciados e, em sua maioria, provados alguns resultados que darão suporte em demonstrações nos demais capítulos. Além disso, veremos algumas definições e exemplos. Este capítulo será dividido em duas seções: a primeira exibe conceitos relacionados a otimização e a segunda conceitos voltados para Análise no $\mathbb{R}^n$. Começamos com algumas notações. O espaço que trabalharemos será o euclidiano n-dimensional, ou seja, o $\mathbb{R}^n$. O i-ésimo componente do vetor $x \in \mathbb{R}^n$ será denotado por x_i para cada $i = 1, \ldots, n$. O octante não negativo é o conjunto definido por

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \ : \ x_i \geq 0\,,\, i = 1, \ldots, n\}. \tag{2.1}$$

Para cada $x \in \mathbb{R}^n$ denotaremos por $\operatorname{sgn}(x)$ o vetor n-dimensional com coordenadas iguais a 1, 0 ou -1, caso a entrada correspondente do vetor x seja positiva, zero ou negativa, respectivamente. Dado um $a \in \mathbb{R}$, definiremos $a^+ := \max\{a, 0\}$ e $a^- := \max\{-a, 0\}$ e x^+, x^- e $|x|$ como sendo os vetores cuja a i-ésima componente é igual a x_i^+, x_i^- e $|x_i|$, respectivamente. Daí, partindo dessas definições e usando (2.1), é fácil ver que

$$x^+ \in \mathbb{R}^n_+, \qquad\qquad x^- \in \mathbb{R}^n_+, \qquad \forall \ \ x \in \mathbb{R}^n. \tag{2.2}$$

A partir disso, podemos afirmar que vale a igualdade $x = x^+ - x^-$, ou seja, $x_i = x_i^+ - x_i^-$ para $i = 1, ..., n$. De fato, basta observarmos que

$$x_i \geq 0 \Rightarrow x_i^+ - x_i^- = \max\{x_i, 0\} - \max\{-x_i, 0\} = x_i - 0 = x_i,$$

$$x_i < 0 \Rightarrow x_i^+ - x_i^- = \max\{x_i, 0\} - \max\{-x_i, 0\} = 0 - (-x_i) = x_i.$$

Outra importante relação que pode ser obtida desses vetores é que $\langle x^+, x^- \rangle = 0$. De fato, pois

$$x_i^+ x_i^- = \max\{x_i, 0\} \max\{-x_i, 0\} = \begin{cases} 0 = x^i \cdot 0, & \text{se} \quad x^i \geq 0, \\ 0 = 0 \cdot (-x^i), & \text{se} \quad x^i < 0, \end{cases}$$

e como $\langle x^+, x^- \rangle = \sum_{i=1}^n x_i^+ x_i^-$, segue que o produto interno $\langle x^+, x^- \rangle$ de fato será 0. Se $x \in \mathbb{R}^n$ então $\text{diag}(x)$ denotará a matriz diagonal $n \times n$ cuja (i,i)-ésima entrada é igual a x_i, $i = 1, \ldots, n$. Para uma matriz $M \in \mathbb{R}^{n \times n}$ (espaço das matrizes de dimensão $n \times n$ com entradas reais) consideraremos a norma definida por $\|M\| := \max\{\|Mx\| \; : \; x \in \mathbb{R}^n, \; \|x\| = 1\}$. As seguintes desigualdade para a norma de matrizes são válidas:

$$\|Mx\| \leq \|M\|\|x\|, \qquad \|LM\| \leq \|L\|\|M\|, \qquad \forall \; L, M \in \mathbb{R}^{n \times n}, \qquad \forall \; x \in \mathbb{R}^n. \tag{2.3}$$

2.1 Alguns conceitos de otimização

Definição 2.1.1 *Uma projeção $P_D(x)$ de um ponto $x \in \mathbb{R}^n$ sobre um conjunto $D \subseteq \mathbb{R}^n$ é uma solução (global) do seguinte problema de otimização restrito*

$$\begin{aligned} &\text{Minimize} \;\; \|x - y\| \\ &\text{sujeito a} \quad y \in D. \end{aligned} \tag{2.4}$$

Definição 2.1.2 *Seja A um subconjunto de um espaço normado $\mathbb{X}$. Uma aplicação $T : A \longrightarrow \mathbb{X}$ é dita não expansiva se satisfaz a desigualdade*

$$\|T(x) - T(y)\| \leq \|x - y\|, \qquad \forall \; x, y \in A. \tag{2.5}$$

Proposição 2.1.3 *Seja $D \subset \mathbb{R}^n$ um subconjunto convexo e fechado. Então a projeção $P_D(x)$ de um ponto $x \in \mathbb{R}^n$ sobre o conjunto D existe e é unica. Além disso, temos que $\overline{x} = P_D(x)$ se, e somente se, $\overline{x} \in D$ e vale*

$$\langle x - \overline{x}, y - \overline{x} \rangle \leq 0, \qquad \forall \; y \in D. \tag{2.6}$$

Prova. Como a norma é uma aplicação positiva e contínua podemos afirmar pelo Teorema de Weirstrass que se for definida num fechado o seu mínimo será assumido em um ponto do conjunto. Desse argumento e da definição de projeção segue que $P_D(x)$ existe. Para provarmos a sua unicidade suponhamos que $\overline{x}$ e $\overline{x}'$ são projeções de x em D. Daí, pela convexidade do conjunto D temos

$$z(\lambda) = \lambda\overline{x} + (1 - \lambda)\overline{x}' \in D, \qquad \forall \; \lambda \in [0, 1].$$

Tomando $\lambda = 1/2$ e fazendo simples manipulações algébricas podemos afirmar que

$$\left\| x - z\left(\frac{1}{2}\right) \right\| = \left\| x - \left(\frac{1}{2}\overline{x} + \left(1 - \frac{1}{2}\right)\overline{x}'\right) \right\| = \left\| \frac{1}{2}(x - \overline{x}) + \frac{1}{2}(x - \overline{x}') \right\| \leq \frac{1}{2}\left(\|x - \overline{x}\| + \left\|x - \overline{x}'\right\|\right)$$

Tomando $\|x - \overline{x}\| = \beta$ segue que $\|x - \overline{x}'\| = \beta$, pois ambos são soluções de (2.4). Assim, da última desigualdade temos

$$\left\| x - z\left(\frac{1}{2}\right) \right\| \leq \frac{1}{2}(\beta + \beta) = \beta.$$

Se a desigualdade for estrita temos uma contradição com o fato de $\overline{x}$ e $\overline{x}'$ serem projeção de x em D. E caso acontecesse a igualdade, três pontos diferentes ($\overline{x}$, $\overline{x}'$ e $z(1/2)$) do segmento de reta $z(\lambda)$ estariam a uma mesma distância de x, o que é impossível. Portanto, a projeção é única.
Fixando um ponto y no conjunto D e usando a convexidade deste podemos afirmar que

$$x(\alpha) = (1-\alpha)\overline{x} + \alpha y \in D, \qquad \forall\ \alpha \in (0,1]. \tag{2.7}$$

Da definição de projeção segue a primeira desigualdade abaixo. Portanto, vale a implicação

$$\|x-\overline{x}\| \leq \|x - x(\alpha)\| \Longrightarrow \|x-\overline{x}\|^2 - \|x-x(\alpha)\|^2 \leq 0.$$

Usando simples manipulações algébricas, segue que

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

Das duas últimas relações segue a primeira desigualdade abaixo. Portanto, vale a implicação

$$2\langle x-\overline{x}, \overline{x}-x(\alpha)\rangle + \|\overline{x}-x(\alpha)\|^2 \geq 0 \Longrightarrow 2\langle x-\overline{x}, x(\alpha)-\overline{x}\rangle - \|\overline{x}-x(\alpha)\|^2 \leq 0.$$

Sendo assim, se substituirmos $x(\alpha)$ pela expressão dada em (2.7) e usarmos propriedades da norma e do produto interno, para reduzirmos os termos semelhantes, obteremos

$$2\alpha\langle x-\overline{x}, y-\overline{x}\rangle - \alpha^2\|y-\overline{x}\|^2 \leq 0, \qquad \forall\ y \in D.$$

Portanto, para concluir que vale (2.6) basta dividir os dois membros por α (estamos supondo $\alpha \neq 0$) e fazer α tender a 0.
Agora suponhamos que $\overline{x} \in D$ e que vale a seguinte desigualdade

$$\langle x-\overline{x}, y-\overline{x}\rangle \leq 0, \qquad \forall\ y \in D.$$

Para provarmos que $\overline{x} = P_D(x)$ primeiramente observemos que vale a seguinte igualdade

$$\|x-y\|^2 = \|x-\overline{x}+\overline{x}-y\|^2 = \|x-\overline{x}\|^2 - 2\langle x-\overline{x}, y-\overline{x}\rangle + \|y-\overline{x}\|^2.$$

Daí, reorganizando os termos desta última igualdade e usando (2.6) temos

$$\|x-y\|^2 - \|x-\overline{x}\|^2 = -2\langle x-\overline{x}, y-\overline{x}\rangle + \|y-\overline{x}\|^2 \geq -2\langle x-\overline{x}, y-\overline{x}\rangle \geq 0.$$

Portanto, segue que $\|x-\overline{x}\| \leq \|x-y\|$ para todo $y \in D$. Como estamos supondo que $\overline{x} \in D$ então segue que $\overline{x} = P_D(x)$, o que conclui a prova. ■

Teorema 2.1.4 *Seja $D \subset \mathbb{R}^n$ um conjunto convexo e fechado. O operador projeção é não expansivo em D.*

Prova. Sejam $x, y \in \mathbb{R}^n$ $(x \neq y)$. Pela proposição anterior temos, a existência (e unicidade) de $P_D(x)$ e $P_D(y)$. Como estes estão em D podemos usar a relação (2.6) para ambos e obter

$$\langle x - P_D(x), P_D(y) - P_D(x)\rangle \leq 0 \quad e \quad \langle y - P_D(y), P_D(x) - P_D(y)\rangle \leq 0. \tag{2.8}$$

Portanto, da primeira inequação em (2.8) segue que

$$\langle P_D(x) - x, P_D(x) - P_D(y)\rangle \leq 0.$$

Somando esta última desigualdade com a segunda desigualdade de (2.8) obtemos

$$\langle P_D(x) - P_D(y), y - P_D(y) + P_D(x) - x\rangle \leq 0.$$

Fazendo simples manipulações algébricas no produto interno acima podemos afirmar que

$$\|P_D(x) - P_D(y)\|^2 + \langle P_D(x) - P_D(y), y - x\rangle \leq 0.$$

Daí, vale a primeira desigualdade abaixo e, consequentemente, temos

$$\|P_D(x) - P_D(y)\|^2 \leq \langle x - y, P_D(x) - P_D(y)\rangle \leq \|P_D(x) - P_D(y)\| \, \|x - y\| \, .$$

Para concluirmos que o operador projeção é não expansivo basta dividirmos os dois membros por $\|P_D(x) - P_D(y)\|$ (observe que basta analisarmos o caso em que esse termo é diferente de 0, pois caso seja 0 a desigualdade (2.5) e trivialmente satisfeita). ■

Definição 2.1.5 *Um conjunto não vazio $K \subset \mathbb{R}^n$ é um cone quando para todo $t \geq 0$ e $x \in K$ tem-se $tx \in K$.*

Definição 2.1.6 *O* cone simplicial *associado a uma matriz não singular $A \in \mathbb{R}^{n\times n}$ é definido por*

$$A\mathbb{R}^n_+ := \{Ax : \ x \in \mathbb{R}^n_+\}. \tag{2.9}$$

Exemplo 2.1.7 Um exemplo simples de cone simplicial é o octante não negativo $\mathbb{R}^n_+$, definido em (2.1). Para ver isto, basta tomar A como a matriz identidade na definição anterior.

Definição 2.1.8 *Um conjunto $K \subset \mathbb{R}^n$ é um* cone pontudo *se possue as seguintes propriedades:*

(i) $\lambda x + \mu y \in K$ *para todo* $\lambda, \mu \geq 0$ *e* $x, y \in K$*,*

(ii) $x, -x \in K$ *implica* $x = 0$.

Proposição 2.1.9 *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz não singular. Então $A\mathbb{R}^n_+$ é um cone pontudo (consequentemente, convexo), fechado e com interior não vazio.*

Prova. Fixando os elementos Ax^+ e Ay^+ em $A\mathbb{R}^n_+$ podemos afirmar que

$$\lambda Ax^+ + \mu Ay^+ = A(\lambda x^+ + \mu y^+) \in A\mathbb{R}^n_+, \quad \forall \ \lambda, \mu \geq 0.$$

De fato, pois $(\lambda x^+ + \mu y^+) \in \mathbb{R}^n_+$. Além disso, se w e $-w$ pertencem a $A\mathbb{R}^n_+$ então existe um $z \in \mathbb{R}^n$ tal que $w = Az^+$, o que implica que $-w = -Az^+ = A(-z^+)$. Daí, como também estamos supondo que $-w \in A\mathbb{R}^n_+$ segue que $-z^+ \in \mathbb{R}^n_+$. Portanto z, e consequentemente w, só pode ser 0. Logo, $A\mathbb{R}^n_+$ é um cone pontudo. Para provarmos a convexidade de um cone pontudo basta considerarmos o item $i)$ da Definição 2.1.8 com

$$\mu = t, \quad \lambda = 1 - t, \qquad \forall t \in (0,1).$$

Note que, como A é uma matriz não singular temos que esta define um homeomorfismo. Portanto, $A\mathbb{R}^n_+$ é fechado, pois é imagem do conjunto fechado $\mathbb{R}^n_+$ pelo homeomorfismo $A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$. Para provarmos que $A\mathbb{R}^n_+$ tem interior não vazio consideremos o ponto

$$x_0 = (1, 1, ..., 1) \in \mathbb{R}^n.$$

Observe que a bola aberta $B_{1/2}(x_0)$ centrada em x_0 e com raio $1/2$ está totalmente contida em $\mathbb{R}^n_+$. Logo, usando o fato que A é um homeomorfismo podemos afirmar que $A(B_{1/2}(x_0))$ (conjunto obtido com a multiplicação da matriz A pelos elementos de $B_{1/2}(x_0)$) é um conjunto aberto em $A\mathbb{R}^n_+$. Daí, temos que Ax_0 é um ponto interior de $A\mathbb{R}^n_+$ e, portanto, $A\mathbb{R}^n_+$ tem interior não vazio.

■

Teorema 2.1.10 *A projeção de um ponto $x \in \mathbb{R}^n$ sobre o octante não negativo $\mathbb{R}^n_+$ existe e é x^+. Além disso, vale*

$$\|x^+ - y^+\| \leq \|x - y\|, \qquad \forall\, x, y \in \mathbb{R}^n.$$

Prova. Fazendo $A = I$ na Proposição 2.1.9 segue que $\mathbb{R}^n_+$ é um conjunto fechado e convexo. Além disso, usando as relações

$$x = x^+ - x^-, \qquad \left\langle x^+, x^- \right\rangle = 0 \qquad \forall\, x \in \mathbb{R}^n,$$

podemos afirmar que, para qualquer $z \in \mathbb{R}^n_+$, vale a seguinte igualdade

$$\left\langle x - x^+, z - x^+ \right\rangle = \left\langle -x^-, z - x^+ \right\rangle = \left\langle -x^-, z \right\rangle + \left\langle x^-, x^+ \right\rangle = -\left\langle x^-, z \right\rangle.$$

Da definição de x^- segue que $x^- \in \mathbb{R}^n_+$. Assim, como estamos tomando z em $\mathbb{R}^n_+$ temos que $\langle x^-, z\rangle$ é um somatório de n números não negativos, o que implica $\langle x^-, z\rangle \geq 0$. Portanto, vale

$$\left\langle x - x^+, z - x^+ \right\rangle = -\left\langle x^-, z \right\rangle \leq 0, \quad \forall\, z \in \mathbb{R}^n_+.$$

Sendo assim, segue da Proposição 2.1.3 que

$$x^+ = P_D(x), \qquad\qquad y^+ = P_D(y), \qquad \forall\, x, y \in \mathbb{R}^n.$$

Como a projeção é um operador não expansivo, a prova está completa. ■

Lema 2.1.11 *Sejam x e y pontos arbitrários em $\mathbb{R}^n$. Podemos afirmar que vale*

$$\left\| |x| - |y| \right\| \leq \|x - y\|.$$

Prova. Observe que

$$|z| = z^+ + z^-, \qquad \forall z \in \mathbb{R}^n.$$

Dessa igualdade juntamente com simples manipulações algébricas obtemos

$$\||x| - |y|\| = \left\|(x^+ + x^-) - (y^+ + y^-)\right\| = \left\|(x^+ - y^+) + (x^- - y^-)\right\|.$$

Usando o fato que a norma usada provém do produto interno segue que

$$\left\|(x^+ - y^+) + (x^- - y^-)\right\|^2 = \left\|x^+ - y^+\right\|^2 + 2\left\langle x^+ - y^+, x^- - y^-\right\rangle + \left\|x^- - y^-\right\|^2. \tag{2.10}$$

Note que

$$\|x - y\| = \left\|(x^+ - x^-) - (y^+ - y^-)\right\| = \left\|(x^+ - y^+) - (x^- - y^-)\right\|.$$

Portanto, partindo de novo do fato de que a norma usada provém do produto interno vale

$$\left\|(x^+ - y^+) - (x^- - y^-)\right\|^2 = \left\|x^+ - y^+\right\|^2 - 2\left\langle x^+ - y^+, x^- - y^-\right\rangle + \left\|x^- - y^-\right\|^2.$$

Subtraindo (2.10) desta última igualdade obtemos

$$\|x - y\|^2 - \||x| - |y|\|^2 = -4\left\langle x^+ - y^+, x^- - y^-\right\rangle.$$

Para concluirmos a prova basta mostrarmos que o produto interno $\langle x^+ - y^+, x^- - y^-\rangle$ é menor ou igual a zero. Temos

$$\left\langle x^+ - y^+, x^- - y^-\right\rangle = \left\langle x^+, x^-\right\rangle - \left\langle x^+, y^-\right\rangle - \left\langle y^+, x^-\right\rangle + \left\langle y^+, y^-\right\rangle. \tag{2.11}$$

Entretanto,

$$\left\langle x^+, x^-\right\rangle = 0 = \left\langle y^+, y^-\right\rangle, \quad \left\langle x^+, y^-\right\rangle \geq 0, \qquad \left\langle y^+, x^-\right\rangle \geq 0.$$

Sendo assim, temos de (2.11) que $\langle x^+ - y^+, x^- - y^-\rangle \leq 0$, o que conclui a prova. ■

Definição 2.1.12 *Seja $K \subset \mathbb{R}^n$ um cone. O* cone dual *de K é o seguinte conjunto*

$$K^* := \{x \in \mathbb{R}^n \mid \langle x, y\rangle \geq 0, \forall\ y \in K\}.$$

Observação 2.1.13 Considerando o cone $\mathbb{R}^n_+$ podemos afirmar que o seu dual é igual ao próprio conjunto $\mathbb{R}^n_+$. Objetivando provar isso fixemos um $x \in (\mathbb{R}^n_+)^*$. Logo

$$\langle x, y\rangle \geq 0, \qquad \forall\ y \in \mathbb{R}^n_+.$$

Observemos que $(e_i)_{i=1}^n$ está contido em $\mathbb{R}^n_+$, pois as coordenadas de cada e_i são não negativas (0 e 1). Assim, fazendo o produto interno acima para y igual a e_i verifica-se que

$$x_i = \langle x, e_i\rangle \geq 0, \qquad i = 1, ..., n.$$

Isto implica que $x \in \mathbb{R}^n_+$ e, consequentemente, temos $(\mathbb{R}^n_+)^* \subset \mathbb{R}^n_+$. Agora basta provarmos que $\mathbb{R}^n_+ \subset (\mathbb{R}^n_+)^*$ para chegarmos a igualdade desejada. Se fixarmos um $x \in \mathbb{R}^n_+$ teremos que $\langle x, y\rangle$ será um somatório de números não negativos para todo $y \in \mathbb{R}^n_+$. Donde concluímos que $(x \in \mathbb{R}^n_+)^*$. Assim, podemos afirmar que $\mathbb{R}^n_+ \subset (\mathbb{R}^n_+)^*$, e consequentemente, obter $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$.

Lema 2.1.14 *Seja A uma matriz em $\mathbb{R}^{n\times n}$. A é não singular se, e somente se, $A^\top$ for não singular. Além disso, vale*

$$(A^\top)^{-1} = (A^{-1})^\top. \tag{2.12}$$

Prova. Se A é uma matriz não singular então $AA^{-1} = A^{-1}A = I$. Logo, através de simples propriedades da transposta de matrizes, podemos garantir as igualdades

$$I = I^\top = \left(AA^{-1}\right)^\top = \left(A^{-1}\right)^\top A^\top, \qquad I = I^\top = \left(A^{-1}A\right)^\top = A^\top \left(A^{-1}\right)^\top.$$

Portanto, pela própia definição da inversa de matrizes segue que $A^\top$ é invertível e vale (2.12). A recíproca segue do fato $A = (A^\top)^\top$. ■

Lema 2.1.15 *Seja A uma matriz não singular em $\mathbb{R}^{n\times n}$. Então, vale a igualdade*

$$(A\mathbb{R}^n_+)^* = (A^\top)^{-1}\mathbb{R}^n_+.$$

Prova. Usando a definição de dual de um cone podemos afirmar que

$$x \in (A\mathbb{R}^n_+)^* \iff \langle x, y\rangle \geq 0, \qquad \forall\, y \in A\mathbb{R}^n_+.$$

Observe que podemos reescrever a equivalência acima da seguinte maneira

$$x \in (A\mathbb{R}^n_+)^* \quad \iff \quad \langle x, Ak\rangle \geq 0, \qquad \forall\, k \in \mathbb{R}^n_+.$$

Daí, como $\langle x, Ak\rangle = \left\langle A^\top x, k\right\rangle$, da última equivalência temos

$$x \in (A\mathbb{R}^n_+)^* \quad \iff \quad \left\langle A^\top x, k\right\rangle \geq 0, \qquad \forall\, k \in \mathbb{R}^n_+.$$

Pela Observação 2.1.13 segue que $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$. Logo

$$x \in (A\mathbb{R}^n_+)^* \quad \iff \quad A^\top x \in \mathbb{R}^n_+.$$

Como A é não singular pelo Lema 2.1.14 podemos afirmar que $A^\top$ também será não singular, ou seja, $(A^\top)^{-1}$ está bem definida. Assim, fica fácil ver que $A^\top x \in \mathbb{R}^n_+$ se, e somente se, $x \in (A^\top)^{-1}\mathbb{R}^n_+$. Logo,

$$x \in (\mathbb{R}^n_+)^* \iff x \in \mathbb{R}^n_+,$$

e a demonstração está completa. ■

Teorema 2.1.16 *Sejam $\varphi : \mathbb{R}^n \to \mathbb{R}$ uma função convexa diferenciável e K um cone fechado e convexo em $\mathbb{R}^n$. Então $\bar{x}$ é uma solução do problema*

$$\begin{array}{rl} \text{Minimizar} & \varphi(x) \\ \text{sujeito a} & x \in K, \end{array}$$

se, e somente se, $\bar{x}$ satisfaz as seguintes condições

$$\overline{x} \in K, \qquad \nabla\varphi(\overline{x}) \in K^*, \qquad \langle \nabla\varphi(\overline{x}), \overline{x}\rangle = 0.$$

Prova. Primeiramente supomos que $\overline{x}$ é uma solução do problema de minimização acima. Daí, segue diretamente que $\overline{x} \in K$. Agora, fixemos um $x \in K$ e definamos a função $g : [0,1] \longrightarrow \mathbb{R}$ da seguinte forma

$$g(t) = \varphi(\overline{x} + tw), \quad \text{com} \quad w = x - \overline{x}. \tag{2.13}$$

Como K é um conjunto convexo podemos afirmar que $\overline{x}+tw \in K$ para todo $t \in [0,1]$. De fato, pois $\overline{x} + tw$ com t variando em no intervalo $[0,1]$ é a combinação convexa de x e $\overline{x}$, como segue

$$\overline{x} + tw = \overline{x} + t(x - \overline{x}) = (1-t)\overline{x} + tx.$$

Daí, o fato de $\overline{x}$ ser um minimizador da função φ em K implica que zero é um minimizador da função g, pois $g(0) = \varphi(\overline{x})$. Portanto, temos $g(t) \geq g(0)$ para todo $t \in [0,1]$. Isso implica

$$\frac{g(t) - g(0)}{t} \geq 0, \qquad \forall\, t \in (0,1].$$

Como a função φ é diferenciável temos que a função g é. Daí, segue que

$$g(t) = g(0) + tg'(0) + r(t), \quad \text{com} \qquad \lim_{t\to 0} \frac{r(t)}{t} = 0.$$

Fazendo simples manipulações algébricas obtemos

$$0 \leq \frac{g(t) - g(0)}{t} = g'(0) + \frac{r(t)}{t}, \quad \text{com} \qquad \lim_{t\to 0} \frac{r(t)}{t} = 0.$$

Aplicando limite com t indo para 0 na desigualdade anterior segue que $g'(0) \geq 0$. Usando a definição da função g dada em (2.13) e a Regra da Cadeia é fácil ver que $g'(0) = \langle \nabla\varphi(\overline{x}), x - \overline{x}\rangle \geq 0$. Portanto, do fato de que x foi tomado arbitrariamente, segue que

$$\langle \nabla\varphi(\overline{x}), x - \overline{x}\rangle \geq 0 \Longrightarrow \langle \nabla\varphi(\overline{x}), x\rangle \geq \langle \nabla\varphi(\overline{x}), \overline{x}\rangle\,, \qquad \forall\, x \in K. \tag{2.14}$$

Assim, o mínimo de $\langle \nabla\varphi(\overline{x}), x\rangle$ é atingido em $\overline{x}$. Como K é um cone podemos afirmar que 0 e $2\overline{x}$ pertencem a K. Portanto, usando a desigualdade acima para $x = 0$ e $x = 2\overline{x}$, temos

$$\begin{aligned} \langle \nabla\varphi(\overline{x}), 0 - \overline{x}\rangle \geq 0 \quad &\Longrightarrow \quad \langle \nabla\varphi(\overline{x}), \overline{x}\rangle \leq 0. \\ \langle \nabla\varphi(\overline{x}), 2\overline{x} - \overline{x}\rangle \geq 0 \quad &\Longrightarrow \quad \langle \nabla\varphi(\overline{x}), \overline{x}\rangle \geq 0. \end{aligned}$$

Logo $\langle \nabla\varphi(\overline{x}), \overline{x}\rangle = 0$ que, juntamente com a segunda desigualdade de (2.14) implica que

$$\langle \nabla\varphi(\overline{x}), x\rangle \geq 0, \qquad \forall\, x \in K,$$

donde podemos concluir que $\nabla\varphi(\overline{x}) \in K^*$.
Para provar a recíproca suponhamos que

$$\overline{x} \in K, \qquad \nabla\varphi(\overline{x}) \in K^*, \qquad \langle \nabla\varphi(\overline{x}), \overline{x}\rangle = 0. \tag{2.15}$$

Como a função φ é convexa podemos afirmar que

$$\varphi(x) - \varphi(\overline{x}) \geq \langle \nabla\varphi(\overline{x}), x - \overline{x}\rangle\,, \qquad \forall\, x \in K.$$

O fato de $\nabla\varphi(\overline{x})$ pertencer a K^* nos possibilita afirmar que $\langle\nabla\varphi(\overline{x}), x\rangle \geq 0$ para todo $x \in K$. Portanto, juntamente com (2.15), temos

$$\langle\nabla\varphi(\overline{x}), x - \overline{x}\rangle = \langle\nabla\varphi(\overline{x}), x\rangle - \langle\nabla\varphi(\overline{x}), \overline{x}\rangle \geq 0, \qquad \forall\ x \in K.$$

Assim, usando as últimas desigualdades obtidas torna-se possível afirmar que

$$\varphi(x) \geq \varphi(\overline{x}), \qquad \forall\ x \in K.$$

Desde que $\overline{x} \in K$, segue que $\overline{x}$ é solução do problema dado, o que conclui a prova. ■

2.2 Alguns conceitos de análise

Lema 2.2.1 (Lema de Banach) *Sejam $E \in \mathbb{R}^{n\times n}$ e I a matriz identidade $n \times n$. Se $\|E\| < 1$ então $E - I$ é invertível e vale*

$$\|(E - I)^{-1}\| \leq \frac{1}{(1 - \|E\|)}.$$

Prova. Consideremos as sequências $\{S_k\}$ e $\{t_k\}$ definidas por

$$S_k = I + E + E^2 + \cdots + E^k, \qquad t_k = 1 + \|E\| + \|E\|^2 + \cdots + \|E\|^k.$$

Aplicando a desigualdade de Cauchy-Schwarz podemos afirmar que

$$\|S_{k+1} - S_k\| = \|(I + E + \cdots + E^{k+1}) - (I + E + \cdots + E^k) = \|E^{k+1}\| \leq \|E\|^{k+1} = t_{k+1} - t_k. \quad (2.16)$$

Outra relação que podemos obter a partir dessas definições é a seguinte

$$(I - E)S_k = S_k(I - E) = (I + E + \cdots + E^k)(I - E) = I - E^{k+1}. \quad (2.17)$$

Logo podemos afirmar que $\lim\limits_{k\to\infty} I - E^k = I$. De fato, pois

$$\|I - (I - E^k)\| = \|E^k\| \leq \|E\|^k \quad \text{e} \quad \lim_{k\to\infty} \|E\|^k = 0,$$

uma vez que $\|E\| < 1$. Observemos que $\{t_k\}$ é a soma de uma PG de razão $\|E\|$. Portanto, podemos afirmar que $\{t_k\}$ é uma sequência monótona crescente. Usando a hipótese $\|E\| < 1$ segue que esta sequência é convergente e vale

$$\lim_{k\to\infty} t_k = \frac{1}{1 - \|E\|}.$$

Daí, por (2.16) segue que $\{S_k\}$ é uma sequência de Cauchy em $\mathbb{R}^{n\times n}$, o que garante a existência do limite de $\{S_k\}$. Assim, aplicando limite com k tendendo ao infinito na equação (2.17), obtemos

$$(I - E)(\lim_{k\to\infty} S_k) = (\lim_{k\to\infty} S_k)(I - E) = I.$$

Isso implica que $\lim_{k\to\infty} S_k = (I-E)^{-1}$. Além disso, temos que

$$\|(E-I)^{-1}\| = \|(I-E)^{-1}\| = \|\lim_{k\to\infty} S_k\| \leq \lim_{k\to\infty}\left(\|I\| + \|E\| + \cdots + \|E^k\|\right) \leq \lim_{k\to\infty} t_k = \frac{1}{1-\|E\|}.$$

■

Definição 2.2.2 *Dizemos que $x^* \in \mathbb{X}$ é um ponto fixo de uma função $F : \mathbb{X} \longrightarrow \mathbb{X}$ se $F(x^*) = x^*$.*

Teorema 2.2.3 (Princípio básico da aplicação contração) *Sejam $\mathbb{X}$ um espaço métrico completo, com métrica ρ, e $\phi : \mathbb{X} \to \mathbb{X}$ uma função contínua. Se existe $\alpha \in [0,1)$ tal que*

$$\rho(\phi(x), \phi(y)) \leq \alpha\rho(x,y), \qquad \forall\ x, y \in \mathbb{X},$$

podemos afirmar que ϕ possui um único ponto fixo em $\mathbb{X}$.

Prova. Primeiramente provemos a existência. Para isso definamos, partindo de um ponto inicial arbitrário $x_0 \in \mathbb{X}$, a sequência $x_n = \phi(x_{n-1})$ para $n = 1, 2, 3, ...$ Dessa definição e do fato que ρ é uma contração podemos afirmar que

$$\rho(x_{n+1}, x_n) = \rho(\phi(x_n), \phi(x_{n-1})) \leq \alpha\rho(x_n, x_{n-1}).$$

Fazendo um processo recursivo na relação acima vê-se facilmente que vale

$$\rho(x_{n+1}, x_n) \leq \alpha\rho(x_n, x_{n-1}) \leq ... \leq \alpha^n\rho(x_1, x_0).$$

Daí, usando a desigualdade triangular da métrica ρ e seguindo o raciocínio acima, conclui-se

$$\begin{aligned} \rho(x_n, x_m) &\leq \rho(x_n, x_{n-1}) + ... + \rho(x_{m+1}, x_m) \\ &\leq (\alpha^{n-1} + ... + \alpha^m)\rho(x_1, x_0), \end{aligned}$$

com $n, m \in \mathbb{N}$ $(n > m)$. Fazendo uso da fórmula da soma de uma PG temos

$$\alpha^m + ... + \alpha^{n-1} \leq \alpha^m + ... + \alpha^{n-1} + ... = \alpha^m(1 + \alpha + \alpha^2 + ...) = \frac{\alpha^m}{1-\alpha}.$$

Portanto, com base nas duas últimas desigualdades obtidas acima podemos afirmar que

$$\rho(x_n, x_m) \leq \frac{\alpha^m}{1-\alpha}\rho(x_1, x_0), \qquad \forall\ n, m \in \mathbb{N}\ \ (n > m).$$

Assim, temos que $\{x_n\}$ é uma sequência de Cauchy. De fato, pois como $\alpha \in [0,1)$, da última desigualdade vê-se facilmente que podemos escolher um m de tal forma que $\rho(x_m, x_n)$ seja tão pequeno quanto se queira. Portanto pela completude do espaço métrico $\mathbb{X}$ temos que $\{x_n\}$ é convergente, e converge para um ponto x pertencente a $\mathbb{X}$. Assim, como ϕ é contínua temos

$$\phi(x_n) = x_{n+1} \Longrightarrow \phi(x) = \lim_{n\to\infty}\phi(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

Agora provemos a unicidade. Para isso, suponhamos, por contradição, que existem x e x^* distintos tais que $\phi(x) = x$ e $\phi(x^*) = x^*$. Daí, como ϕ é uma contração, temos

$$\rho(x, x^*) = \rho(\phi(x), \phi(x^*)) \leq \alpha\rho(x, x^*)$$

Como ρ é uma métrica e estamos supondo $x \neq x^*$ segue que $\rho(x, x^*)$ é diferente de 0. Portanto, podemos dividir os dois membros da desigualdade anterior por $\rho(x, x^*)$, obtendo que $\alpha \geq 1$, o que contraria a hipótese. ■

Proposição 2.2.4 *Sejam $\{x_k\}$ uma sequência em $\mathbb{R}^n$ e u um vetor em $\mathbb{R}^n$. Se existir $\alpha \in [0, 1)$ tal que*

$$\|x_{k+1} - u\| \leq \alpha \|x_k - u\|, \qquad \forall\ k \in \mathbb{N}, \tag{2.18}$$

então a sequência $\{x_k\}$ converge para u.

Prova. Usando um processo recursivo na desigualdade (2.18) podemos afirmar que

$$\|x_{k+1} - u\| \leq \alpha \|x_k - u\| \leq ... \leq \alpha^{k+1}\|x_0 - u\|, \qquad \forall\ k \in \mathbb{N}.$$

Como $\alpha \in [0, 1)$ vê-se facilmente que vale a seguinte igualdade

$$\lim_{k\to\infty} \alpha^{k+1}\|x_0 - u\| = \|x_0 - u\| \lim_{k\to\infty} \alpha^{k+1} = 0.$$

Logo, juntamente com a última desigualdade, conclui-se o desejado. ■

Lema 2.2.5 *Sejam $I \in \mathbb{R}$ um intervalo aberto e $F : I \longrightarrow \mathbb{R}$ uma função diferenciável num ponto $t \in I$. Podemos afirmar que dado $\epsilon > 0$ existe um $\delta_\epsilon(t) > 0$ tal que se $u, v \in I$ satisfazendo*

$$t - \delta_\epsilon(t) \leq u \leq t \leq v \leq t + \delta_\epsilon(t) \tag{2.19}$$

então vale $|F(v) - F(u) - F'(t)(v - u)| \leq \epsilon(v - u)$.

Prova. Pela definição da derivada $F'(t)$ em um ponto $t \in I$, temos que dado um $\epsilon > 0$ existe um $\delta_\epsilon(t) > 0$ tal que se $0 < |z - t| \leq \delta_\epsilon(t)$, $z \in I$, vale

$$\left| \frac{F(z) - F(t)}{z - t} - F'(t) \right| \leq \epsilon.$$

Multiplicando os dois membros por $|z - t|$ obtemos

$$\big| F(z) - F(t) - F'(t)(z - t) \big| \leq \epsilon\, |z - t|\,, \tag{2.20}$$

para todo $z \in I$ tal que $|z - t| \leq \delta_\epsilon(t)$. Escolhendo u e v como em (2.19), e fazendo algumas manipulações algébricas temos

$$\begin{aligned}
|F(v) - F(u) - F'(t)(v - u)| &= |F(v) - F(u) - F'(t)(v - u) + [F(t) + F'(t)t] - [F(t) + F'(t)t]| \\
&= |F(v) - F(t) - F'(t)(v - t)] - [F(u) - F(t) - F'(t)(u - t)]|\,.
\end{aligned}$$

Usando a desigualdade triangular na última expressão da igualdade acima segue que a mesma é menor que ou igual a

$$\left|F(v) - F(t) - F'(t)(v-t)\right| + \left|F(u) - F(t) - F'(t)(u-t)\right|.$$

Portanto, como $v - t \geq 0$ e $t - u \geq 0$, de (2.20) temos que

$$\left|F(v) - F(u) - F'(t)(v-u)\right| \leq \epsilon(v-t) + \epsilon(t-u) = \epsilon(v-u),$$

o que conclui a prova. ■

Definição 2.2.6 *Sejam* $I := [a,b] \subset \mathbb{R}$ *e* $f : I \longrightarrow \mathbb{R}$. *Dizemos que a função* $F : I \longrightarrow \mathbb{R}$ *é uma c-primitiva de f se F é contínua sobre I e* $F'(x) = f(x)$ *a menos de um conjunto contável de pontos* $x \in I$.

No conjunto no qual não se verifica a igualdade $F'(x) = f(x)$ temos que, ou $F'(x)$ não existe ou não é igual a $f(x)$. Os exemplos ilustrarão esses dois casos.

Exemplo 2.2.7 Defina a função $f : [-1,1] \longrightarrow \mathbb{R}$ por

$$f(t) = \begin{cases} -1, & t \in [-1,0], \\ 1, & t \in (0,1]. \end{cases}$$

Daí, podemos afirmar que a função $F(x) = |x|$ é uma c-primitiva de f, pois esta é contínua e $F'(x)$ coincide com $f(x)$ em todos os pontos onde F é derivável($[-1,1]\backslash\{0\}$).

Exemplo 2.2.8 Definindo $f : [-1,1] \longrightarrow \mathbb{R}$ da seguinte maneira

$$f(t) = \begin{cases} 1, & t \in [-1,1]\backslash\{0\}, \\ 2, & t = 0 \end{cases}$$

podemos afirmar que a função $F(x) = x$ é uma c-primitiva de f, pois apenas no $x = 0$ a igualdade $F'(x) = f(x)$ não ocorre.

Teorema 2.2.9 (Teorema Fundamental do Cálculo Generalizado) *Se* $f : [a,b] \longrightarrow \mathbb{R}$ *tem uma c-primitiva* F *sobre* $[a,b]$, *então* f *é integrável em* $[a,b]$ *e*

$$\int_a^b f(t)dt = F(b) - F(a)$$

Prova. Seja $E = \{c_k\}_{k=1}^{\infty}$ o conjunto em que $F'(x)$ não existe ou $F'(x) \neq f(x)$. Definindo

$$h(t) = \begin{cases} f(t), & t \in [a,b]\backslash E, \\ 0, & t \in E, \end{cases}$$

podemos afirmar que F também é uma c-primitiva de h, pois E tem medida nula.
Dado $\epsilon > 0$, para todo $t \in I - E$, tomemos $\delta_\epsilon(t)$ como no Lema 2.2.5, com $I := [a,b]$. Se $t \in E$, então $t = c_k$ para algum $k \in \mathbb{N}$; da continuidade da F em c_k, temos que existe $\delta_\epsilon(c_k) > 0$ tal que

$$|F(z) - F(c_k)| \leq \epsilon/2^{k+2}, \qquad \forall\ z \in I \cap [c_k - \delta_\epsilon(c_k), c_k + \delta_\epsilon(c_k)].$$

Portanto, considerando os pontos c_k, $c_k + \delta_\epsilon(c_k)$ e $c_k - \delta_\epsilon(c_k)$ para todo k, conseguimos uma partição para o intervalo I. Tomemos uma partição $P := ([x_{i-1}, x_i], t_i)_{i=1}^n$ sobre I que seja mais fina que essa. Assim, se $c_k \in [x_{i-1}, x_i]$ temos

$$\begin{aligned} |F(x_i) - F(x_{i-1}) - h(c_k)(x_i - x_{i-1})|| &\leq |F(x_i) - F(c_k)| + |F(c_k) - F(x_{i-1})| + |h(c_k)(x_i - x_{i-1})| \\ &\leq \epsilon/2^{k+2} + \epsilon/2^{k+2} + 0 = \epsilon/2^{k+1}. \end{aligned}$$

Agora como cada ponto pertecente a E pode ser o representante de dois intervalos em P temos que a soma dos termos com $t_i \in E$ satisfaz

$$\sum_{t_i \in E} |F(x_i) - F(x_{i-1}) - h(t_i)(x_i - x_{i-1})| \leq 2 \sum_{k=1}^{\infty} \frac{\epsilon}{2^{k+1}} = \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \epsilon.$$

Também, pelo Lema 2.2.5, a soma dos termos com $t_i \notin E$ satisfaz

$$\sum_{t_i \notin E} |F(x_i) - F(x_{i-1}) - h(t_i)(x_i - x_{i-1})| \leq \epsilon \sum_{t_i \notin E} (x_i - x_{i-1}) \leq \epsilon(b - a).$$

Fazendo a soma das últimas duas desigualdades obtemos

$$\left| F(b) - F(a) - \sum_{t_i \in I} h(t_i)(x_i - x_{i-1}) \right| = \sum_{t_i \in I} |F(x_i) - F(x_{i-1}) - h(t_i)(x_i - x_{i-1})| \leq \epsilon(1 + b - a).$$

Daí, como o ϵ é arbitrário podemos afirmar que h é integrável, com integral igual a $F(b) - F(a)$. Como E tem conteúdo nulo, pela definição da função h conclui-se que f é integravel (a Lebesgue) e vale

$$\int_a^b f(t)dt = \int_a^b h(t)dt = F(b) - F(a).$$

■

Teorema 2.2.10 (Teorema de Rademacher) *Sejam $\Omega \subset \mathbb{R}^n$ um aberto e $f : \Omega \longrightarrow \mathbb{R}^m$ uma função Lipschitz. Então f é diferenciável em quase todo ponto.*

Prova. Ver [9]. ■

Definição 2.2.11 *Seja $\Omega \subset \mathbb{R}^n$ é o conjunto dos pontos de $\mathbb{R}^n$ onde F é diferenciável. O subdiferencial de Clarke da função $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ em x, denotado por $\partial F(x)$, é o fecho convexo de todos os limites de $F'(x_i)$ quando $\{x_i\} \subset \Omega$ converge para x. Simbolicamente, temos*

$$\partial F(x) = co\{\lim F'(x_i)\ :\ x_i \longrightarrow x, \quad x_i \in \Omega\}. \tag{2.21}$$

Capítulo 3

Programação convexa sob a restrição de um cone simplicial

Neste capítulo estudaremos o método de Newton para resolver o problema de programação convexa restrito a um cone simplicial, o qual tem o seguinte enunciado:

Problema 1 *Sejam $\varphi : \mathbb{R}^n \to \mathbb{R}$ uma função convexa $\mathbb{C}^2$ e $A \in \mathbb{R}^{n\times n}$ uma matriz não singular. Encontre uma solução u do seguinte problema de programação convexa*

$$\begin{aligned} &\text{Minimizar } \ \varphi(x) \\ &\text{sujeito a } \ x \in A\mathbb{R}^n_+. \end{aligned} \tag{3.1}$$

O método de Newton será usado para encontrar a raiz de uma equação semi-suave associada ao Problema 1 e, a partir desta raiz, obtemos a solução do problema (3.1). A seção a seguir discute propriedades dessa equação.

3.1 Equação semi-suave associada ao problema de minimização convexa

A seguir apresentamos uma equação cuja raiz gera uma solução para o Problema 1.

Problema 2 *Sejam $\varphi : \mathbb{R}^n \to \mathbb{R}$ uma função convexa $\mathbb{C}^2$ e $A \in \mathbb{R}^{n\times n}$ uma matriz não singular. Encontre uma solução u da equação semi-suave*

$$A^\top \nabla\varphi(Ax^+) - x^+ + x = 0. \tag{3.2}$$

Na proposição a seguir usaremos o Teorema 2.1.16 para relacionar a raiz da equação (3.2) com uma solução do Problema 1.

Proposição 3.1.1 *Seja $u \in \mathbb{R}^n$. Se u é solução do Problema 2 então Au^+ é solução do Problema 1.*

Prova. Se $u \in \mathbb{R}^n$ é uma solução do Problema 2 então u é uma raiz da equação (3.2). Daí, temos

$$A^\top \nabla\varphi(Au^+) - u^+ + u = 0.$$

Como vale a igualdade $u = u^+ - u^-$ (ou seja, $u^+ - u = u^-$) para todo $u \in \mathbb{R}^n$, segue que

$$A^\top \nabla\varphi(Au^+) = u^-. \tag{3.3}$$

Dado que A (e, consequentemente $A^\top$, pelo Lema 2.1.14) é uma matriz não singular e $u^- \in \mathbb{R}^n_+$, podemos concluir da última igualdade que

$$\nabla\varphi(Au^+) = (A^\top)^{-1}u^- \in (A^\top)^{-1}\mathbb{R}^n_+.$$

Usando o Lema 2.1.15 segue que $\nabla\varphi(Au^+) \in (A\mathbb{R}^n_+)^*$. De (3.3) temos

$$\langle A^\top \nabla\varphi(Au^+), u^+\rangle = \left\langle u^-, u^+\right\rangle = 0.$$

Como vale a seguinte igualdade $\langle A^\top \nabla\varphi(Au^+), u^+\rangle = \langle \nabla\varphi(Au^+), Au^+\rangle$, conclui-se que

$$\langle \nabla\varphi(Au^+), Au^+\rangle = 0.$$

Resumindo os argumentos e observando que $u^+ \in \mathbb{R}^n_+$ podemos afirmar que

$$Au^+ \in A\mathbb{R}^n_+, \qquad \nabla\varphi(Au^+) \in (A\mathbb{R}^n_+)^*, \qquad \left\langle \nabla\varphi(Au^+), Au^+\right\rangle = 0.$$

Uma vez que $A\mathbb{R}^n_+$ é um cone convexo e fechado (ver Proposição 2.1.9), o resultado segue imediatamente pelo Teorema 2.1.16 com $K = A\mathbb{R}^n_+$. ■

Lema 3.1.2 *Seja $\alpha \in \mathbb{R}^+$. Se vale $\left\|A^T\nabla^2\varphi(Ax^+)A - I\right\| \leq \alpha < 1$ para todo $x \in \mathbb{R}^n$, temos que a função $H : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ definida por*

$$H(x) = x^+ - A^T\nabla\varphi(Ax^+), \tag{3.4}$$

é Lipschitz com constante α. Além do mais, o Problema 2 tem uma única solução u.

Prova. Como $\mathrm{diag}(\mathrm{sgn}(x^+))x = x^+$ podemos reescrever (3.4) da seguinte maneira

$$H(x) = diag(sgn(x^+))x - A^T\nabla\varphi(A(diag(sgn(x^+))x)).$$

Note que, se todas as coordenadas de x forem diferentes de zero, existirá um $\epsilon > 0$ tal que

$$sgn(y) = sgn(x), \qquad \forall\ y \in B_\epsilon(x),$$

que implica, evidentemente,

$$sgn(y^+) = sgn(x^+), \qquad \forall\ y \in B_\epsilon(x).$$

Daí, temos que $\text{sgn}(y^+)$ é constante para todo y numa vizinhança de x, implicando que H será diferenciável nessa vizinhança com diferencial igual a

$$H'(x) = [I - A^T \nabla^2 \varphi(Ax^+)A]D(x^+), \tag{3.5}$$

em que $D(x^+) = diag(sgn(x^+))$. Definindo por Ω o conjunto dos pontos $x \in \mathbb{R}^n$ cujas coordenadas são todas diferentes de zero e usando a propriedade da norma dada em (2.3), segue que

$$\left\|H'(x)\right\| \leq \left\|A^T \nabla^2 \varphi(Ax^+)A - I\right\| \left\|D(x^+)\right\|, \qquad \forall\, x \in \Omega.$$

Como $D(x^+)$ é uma matriz diagonal com entradas iguais a 0 ou 1 vê-se facilmente que $\|D(x^+)\| \leq 1$. Logo, pela hipótese do Lema temos

$$\left\|H'(x)\right\| \leq \alpha, \qquad \forall\, x \in \Omega. \tag{3.6}$$

Da maneira como Ω foi definido segue que Ω^c tem medida nula. Além disso, fixando $x, y \in \Omega$ e considerando o segmento de reta $[x, y]$, podemos afirmar que o conjunto $[x, y] \cap \Omega^c$ é finito (no máximo n elementos). Consideremos as funções $H_i(t) : \mathbb{R} \longrightarrow \mathbb{R}$ definidas por

$$H_i(t) = H(x + (1-t)(y-x)), \qquad i = 1, ..., n.$$

Daí, tomando $z(t) = x + (1-t)(y-x)$ definamos as funções $G_i : \mathbb{R} \longrightarrow \mathbb{R}$ assim:

$$G_i(t) = \begin{cases} H_i'(z(t)), & z(t) \in \Omega, \\ \\ 0, & z(t) \in \Omega^c. \end{cases}$$

Note que a função H (consequentemente, H_i para $i = 1, ..., n$) é contínua. De fato, pois depende de $\nabla\varphi(x)$ que é $\mathbb{C}^1$ (devido $\varphi(x)$ ser $\mathbb{C}^2$) e de x^+ que é não expansiva (e consequentemente contínua) pelo Teorema 2.1.10. Sendo assim, podemos afirmar que H_i é uma c-primitiva de G_i para $i = 1, ..., n$. Então, pelo Teorema 2.2.9 (Teorema Fundamental do Cálculo Generalizado) temos

$$H_i(x) - H_i(y) = \int_0^1 G_i(t)dt = \int_0^1 H_i'(t)dt = \int_0^1 \langle \nabla H_i(z(t)), x - y \rangle \, dt.$$

Sendo assim, podemos afirmar que

$$H(x) - H(y) = \int_0^1 H'(z(t))(x-y)dt, \qquad \forall\, x, y \in \Omega. \tag{3.7}$$

Assim, aplicando norma no segundo membro da última igualdade e usando propriedades destas, juntamente com a relação (3.6), temos

$$\left\|\int_0^1 H'(z(t))(x-y)dt\right\| \leq \int_0^1 \left\|H'(z(t))\,\|\|\,x - y\right\| dt \leq \alpha \left\|x - y\right\|, \quad \forall\, x, y \in \Omega. \tag{3.8}$$

Caso x ou y não pertença a Ω (suponhamos os dois, sem perda de generalidade) pela densidade de Ω no $\mathbb{R}^n$ podemos afirmar que existem sequências $\{x_k\}$ e $\{y_k\}$ em Ω tais que

$$x_k \longrightarrow x, \qquad\qquad y_k \longrightarrow y.$$

Sendo assim,

$$\|H(x_k) - H(y_k)\| \leq \alpha \|x_k - y_k\|, \qquad \forall\, k \in \mathbb{N}.$$

Fazendo $k \longrightarrow \infty$ na desigualdade acima e usando a continuidade da função H obtemos

$$\|H(x) - H(y)\| \leq \alpha \|x - y\|, \qquad \forall\, x, y \in \Omega^c.$$

Portanto, concluimos a prova da primeira parte do Lema. Para provarmos a segunda parte note que, da própria definição da função H em (3.4), basta mostrarmos a existência e unicidade de um ponto fixo de H. Porém, como H é uma contração, isto segue diretamente do Teorema 2.2.3. ■

3.1.1 Método de Newton semi-suave

Com base na equação (3.2) definamos a função $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ da seguinte maneira

$$F(x) = A^\top \nabla\varphi(Ax^+) - x^+ + x. \tag{3.9}$$

Usando a definição da função H, dada em (3.4), podemos reescrever F como: $F(x) = x - H(x)$. Partindo dessa definição e aplicando a desigualdade triangular vê-se facilmente que

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

Daí, usando o Lema 3.1.2, segue que F é Lipschitz. De fato, para todo $x, y \in \mathbb{R}^n$ vale

$$\|F(x) - F(y)\| \leq \|x - y\| + \alpha\|x - y\| = (1 + \alpha)\|x - y\|.$$

Aplicando o Teorema 2.2.10 (Teorema de Rademacher) temos que F é diferenciável em quase todos os pontos do $\mathbb{R}^n$, ou seja, se definirmos por Ω o conjunto no qual ela é diferenciável obtemos que Ω^c terá medida nula, o que implica que Ω é denso em $\mathbb{R}^n$. Daí, fixando um $x \in \mathbb{R}^n$ existe uma sequência $\{x_i\} \subset \Omega$ que converge para x. Portanto o subdiferencial de Clarke de F em x é não vazio, consequentemente, podemos definir formalmente a iteração do Método de Newton semi-suave para resolver a equação (3.2) por

$$F(x_k) + V_k\,(x_{k+1} - x_k) = 0, \qquad k = 0, 1, \ldots, \tag{3.10}$$

em que V_k é algum subgradiente do subdiferencial de Clarke de F em x_k. Agora, afirmamos que

$$S(x) := \Big[A^\top \nabla^2\varphi(Ax^+)A - I\Big]\,\mathrm{diag}(\mathrm{sgn}(x^+)) + I \in \partial F(x), \qquad x \in \mathbb{R}^n. \tag{3.11}$$

De fato, da demonstração do Lema 3.1.2, em especial a igualdade (3.5), essa inclusão é válida para todo $x \in \Omega$. Sendo assim, basta justificarmos quando $x \in \Omega^c$. Com esse intuito, tomemos uma sequência $\{y_n\}$ convergindo para x no interior de um octante cujo vetor sinal deste octante coincide com $\mathrm{sgn}(x^+)$. Para isso basta consideramos um octante com entrada negativa onde a respectiva entrada de $\mathrm{sgn}(x^+)$ for 0. Assim,

$$\mathrm{diag}(\mathrm{sgn}(\{y_n\}^+)) = \mathrm{diag}(\mathrm{sgn}(x^+)) \qquad \forall\, n \in \mathbb{N}.$$

Consequentemente, $S(y_n)$ converge para $S(x)$ pois $\nabla^2\varphi(Ax^+)$ é contínua. Portanto, $S(x)$ é um dos limites do conjunto

$$\partial F(x) = co\{\lim F'(x_k) \ : \ x_k \longrightarrow x, \quad x_k \in \Omega\}.$$

Partindo disso podemos definir a iteração do método de Newton para a função F da forma

$$F(x_k) + S(x_k)\,(x_{k+1} - x_k) = 0, \qquad k = 0, 1, \ldots.$$

Se $S(x_k)$ for invertível para todo k, podemos reescrever a igualdade anterior da seguinte maneira

$$x_{k+1} = x_k - S(x_k)^{-1}F(x_k), \qquad k = 0, 1, \ldots. \tag{3.12}$$

Esta será a iteração de Newton que usaremos, pois o resultado a seguir prova que $S(x)$ é não singular para todo $x \in \mathbb{R}^n$.

Lema 3.1.3 *Se $\nabla^2\varphi(x)$ é positiva definida então a matriz $S(x)$ definida em* (3.11) *é não singular para todo $x \in \mathbb{R}^n$. Como consequência, o método de Newton semi-suave* (3.12) *está bem definido.*

Prova. Para simplificar as notações tomemos $D = \text{diag}(\text{sgn}(x_+))$. Daí, a matriz $S(x)$ torna-se

$$\Big[A^\top \nabla^2\varphi(Ax^+)A - I\Big]\,D + I.$$

Suponhamos, por contradição, que esta matriz é singular, ou seja, que existe um $u \in \mathbb{R}^n$ tal que

$$\Big[\Big[A^\top \nabla^2\varphi(Ax^+)A - I\Big]\,D + I\Big]\,u = 0, \qquad u \neq 0.$$

Fazendo simples manipulações algébricas é fácil ver que a última equação equivale a

$$A^\top \nabla^2\varphi(Ax^+)ADu = (D - I)u, \qquad u \neq 0. \tag{3.13}$$

Como a matriz hessiana $\nabla^2\varphi(x)$ é simétrica e positiva definida para todo $x \in \mathbb{R}^n$ temos, pela decomposição de Cholesky, que para cada $x \in \mathbb{R}^n$ existe uma matriz não singular $L(x) \in \mathbb{R}^{n\times n}$ tal que $\nabla^2\varphi(Ax^+) = L(x)L(x)^\top$. Assim, desenvolvendo $\left\|L(x)^\top ADu\right\|^2$ obtemos

$$\Big\langle L(x)^\top ADu, L(x)^\top ADu\Big\rangle = \Big\langle D^\top A^\top L(x)L(x)^\top ADu, u\Big\rangle = \Big\langle D^\top A^\top \nabla^2\varphi(Ax^+)ADu, u\Big\rangle.$$

Como D é uma matriz diagonal, cujos os elementos da diagonal são 0 e 1, podemos afirmar que $D = D^\top$ e $D^2 = D$. Portanto, da equação (3.13) e da última igualdade, temos

$$\Big\|L(x)^\top ADu\Big\|^2 = \Big\langle DA^\top \nabla^2\varphi(Ax^+)ADu, u\Big\rangle = \big\langle (D^2 - D)u, u\big\rangle = 0.$$

Como $\nabla^2\varphi(Ax^+) = L(x)L(x)^\top$ e $L(x)^\top ADu = 0$ podemos afirmar que a equação (3.13) implica que $(D - I)u = 0$ ou, equivalentemente, $Du = u$. Portanto,

$$L(x)^\top Au = L(x)^\top ADu = 0, \qquad u \neq 0.$$

Mas isto contradiz o fato da matriz A ser não singular, dado que $L(x)$ é não singular. Portanto, a matriz $S(x)$ é não singular para todo $x \in \mathbb{R}^n$ e o Lema está provada. ■

Lema 3.1.4 *Se* $\left\|A^\top\nabla^2\varphi(Ax^+)A - I\right\| < 1$ *e* $S(x)$ *é definida em* (3.11) *então*

$$\|S(x)^{-1}\| \leq \frac{1}{1 - \|A^\top\nabla^2\varphi(Ax^+)A - I\|}, \qquad \forall \ \ x \in \mathbb{R}^n.$$

Prova. Para simplificar a notação tomemos $S(x) = -(E - I)$, em que a matriz E é definida por

$$E = \left[I - A^\top\nabla^2\varphi(Ax^+)A\right] \operatorname{diag}(\operatorname{sgn}(x^+)).$$

Como a matriz diagonal $\operatorname{diag}(\operatorname{sgn}(x^+))$ tem componentes iguais a 1 ou 0, a hipótese $\|A^\top\nabla^2\varphi(Ax^+)A - I\| < 1$ implica que

$$\|E\| \leq \|A^\top\nabla^2\varphi(Ax^+)A - I\| < 1.$$

Portanto, como $S(x) = -(E - I)$ aplicando o Lema 2.2.1 e a desigualdade anterior, segue a desigualdade desejada. ■

Lema 3.1.5 *Sejam* $\varphi : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável estritamente convexa e* $A \in \mathbb{R}^{n\times n}$ *uma matriz não singular. Suponha* $\|[A^\top\nabla^2\varphi(Ax^+)A - I]\| \leq \alpha < 1$. *Se* $S(x)$ *é definida em* (3.11) *então*

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

Como consequência,

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

Prova. Tomando $x, y \in \mathbb{R}^n$ e usando a definição (3.11) podemos afirmar que vale

$$\|S(x) - S(y)\| = \|[A^\top\nabla^2\varphi(Ax^+)A - I]D(x^+) - [A^\top\nabla^2\varphi(Ay^+)A - I]D(y^+)\|.$$

Aplicando a desigualdade triangular e as propriedades da norma, dadas em (2.3), temos

$$\|S(x) - S(y)\| \leq \|A^\top\nabla^2\varphi(Ax^+)A - I\|\|D(x^+)\| + \|A^\top\nabla^2\varphi(Ay^+)A - I\|\|D(y^+)\|.$$

Como $\|D(x)\| \leq 1$ para todo $x \in \mathbb{R}^n$ segue que

$$\|S(x) - S(y)\| \leq \|A^\top\nabla^2\varphi(Ax^+)A - I\| + \|A^\top\nabla^2\varphi(Ay^+)A - I\| \leq 2\alpha.$$

O que prova a primeira desigualdade do Lema. Para provar a segunda desigualdade note que, como foi observado no início da presente seção, vale

$$F(x) - F(y) = x - H(x) - y + H(y) = (x - y) - (H(x) - H(y)).$$

Daí, usando essa relação juntamente com a equação (3.7) obtemos

$$F(x) - F(y) = (x - y) - \int_0^1 H'(z(t))(x - y)dt = \int_0^1 (x - y) - H'(z(t))(x - y)dt,$$

com $z(t) = y + t(x - y)$. Como, a partir das definições (3.5) e (3.11), é possível afirmar que

$$S(z(t))(x - y) = [I - H'(z(t))](x - y) = (x - y) - H'(z(t)))(x - y), \quad \forall t \in [0, 1]$$

segue, fazendo simples manipulações algébricas nesta duas últimas relações, que vale

$$F(x) - F(y) = \int_0^1 [S(y + t(x - y))] (x - y)dt.$$

Portanto, subtraindo o termo $S(y)(x - y)$ em ambos os membros, obtemos a seguinte igualdade

$$F(x) - F(y) - S(y)(x - y) = \int_0^1 [S(y + t(x - y)) - S(y)] (x - y)\mathrm{d}t. \tag{3.14}$$

Aplicando norma nos dois lados da igualdade acima e usando propriedades da mesma temos

$$\|F(x) - F(y) - S(y)(x - y)\| \leq \int_0^1 \|S(y + t(x - y)) - S(y)\| \, \|x - y\| \, \mathrm{d}t.$$

Daí, a conclusão segue diretamente da primeira desigualdade do Lema. ■

Teorema 3.1.6 *Seja $A \in \mathbb{R}^{n \times n}$ uma matriz não singular. Se $\|A^\top \nabla^2 \varphi(Ax^+)A - I\| \leq \alpha < 1/3$ então, para qualquer ponto inicial $x_0 \in \mathbb{R}^n$, a sequência $\{x_k\}$ dado por* (3.12) *converge linearmente para uma solução u do Problema 2, como segue*

$$\|u - x_{k+1}\| \leq \frac{2\alpha}{1 - \alpha} \|u - x_k\|, \qquad k = 0, 1, \ldots.$$

Consequentemente, Au^+ é uma solução do Problema 1.

Prova. Seja u a solução do Problema 2, cuja a existência é garantida pelo Lema 3.1.2. Da definição de $\{x_k\}$ em (3.12) temos

$$u - x_{k+1} = u - x_k + S(x_k)^{-1} F(x_k), \qquad k = 0, 1, \ldots.$$

Daí, como $F(u) = 0$ podemos afirmar, após simples manipulações algébricas, que

$$u - x_{k+1} = -S(x_k)^{-1} [F(u) - F(x_k) - S(x_k)(u - x_k)], \qquad k = 0, 1, \ldots.$$

Aplicando norma nessa igualdade e usando a desigualdade de Cauchy-Schwars obtemos

$$\|u - x_{k+1}\| \leq \|S(x_k)^{-1}\| \| [F(u) - F(x_k) - S(x_k)(u - x_k)] \|, \qquad k = 0, 1, \ldots.$$

Combinando o Lema (3.1.4) e a segunda parte do Lemma (3.1.5) podemos concluir que

$$\|u - x_{k+1}\| \leq \frac{2\alpha}{1 - \alpha} \|u - x_k\|, \qquad k = 0, 1, \ldots.$$

Dado que $\alpha < 1/3$ pode-se concluir que $2\alpha/(1 - \alpha) < 1$. Portanto, da Proposição 2.2.4 segue que $\{x_k\}$ converge linearmente para a única solução u do Problema 2 para qualquer ponto inicial. Daí, a primeira parte do Lema está provada. A segunda parte segue da Proposição 3.1.1. ■

3.2 Programação quadrática sob a restrição de um cone simplicial

A seguir está enunciado um problema que é um caso particular do Problema 1, pois a função objetivo convexa considerada é quadrática.

Problema 3 *Seja $Q \in \mathbb{R}^{n\times n}$ uma matriz simétrica definida positiva, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ e $A \in \mathbb{R}^{n\times n}$ uma matriz não singular. Encontre uma solução u do problema de programação quadrática*

$$\begin{aligned} &\text{Minimizar } \ \frac{1}{2}x^\top Qx + x^\top b + c \\ &\text{sujeito a } \ x \in A\mathbb{R}^n_+. \end{aligned} \tag{3.15}$$

No exemplo a seguir é apresentado um importante caso particular do Problema 3.

Exemplo 3.2.1 Seja $A \in \mathbb{R}^{n\times n}$ uma matriz não singular e $z \in \mathbb{R}^n$. Uma *projeção $P_{A\mathbb{R}^n_+}(z)$ do ponto z sobre o cone $A\mathbb{R}^n_+$* pode ser definida por

$$P_{A\mathbb{R}^n_+}(z) := \operatorname{argmin}\left\{\frac{1}{2}\|z-y\|^2 \ : \ y \in A\mathbb{R}^n_+\right\}.$$

Equivalentemente, encontrar a projeção de um ponto $z \in \mathbb{R}^n$ sobre o cone simplicial $A\mathbb{R}^n_+$ pode ser visto como o seguinte problema de programação quadrática

$$\begin{aligned} &\text{Minimizar } \ \frac{1}{2}\|z-y\|^2 \\ &\text{sujeito a } \ \ y \in A\mathbb{R}^n_+. \end{aligned}$$

Daí, se $v \in \mathbb{R}^n$ é uma solução deste problema então segue que $P_{A\mathbb{R}^n_+}(z) = Av^+$.

3.2.1 Equação semi-suave associada ao problema de minimização quadrática

Observemos que o Problema 3 é um caso particular do Problema 1, basta tomarmos

$$\varphi(x) = \frac{1}{2}x^\top Qx + x^\top b + c. \tag{3.16}$$

De fato, $\varphi(x)$ é convexa, pois sua hessiana é igual a Q que é definida positiva por hipótese. Dessa última definição, podemos ver facilmente que vale a igualdade e, consequentemente, a implicação a seguir

$$\nabla\varphi(x) = Qx + b \Longrightarrow \nabla\varphi(Ax^+) = QAx^+ + b.$$

Portanto, fazendo simples manipulações algébricas podemos afirmar que

$$A^\top\nabla\varphi(Ax^+) - x^+ + x = A^\top[QAx^+ + b] - x^+ + x = \left[A^\top QA - I\right]x^+ + x + A^\top b.$$

Com base no Problema 2 e dessa última igualdade enuciamos o seguinte problema.

Problema 4 *Sejam $Q \in \mathbb{R}^{n\times n}$ uma matriz simétrica definida positiva, $b \in \mathbb{R}^n$ e $A \in \mathbb{R}^{n\times n}$ uma matriz não singular. Encontre uma solução u da equação semi-suave*

$$\left[A^\top QA - I\right] x^+ + x + A^\top b = 0. \tag{3.17}$$

A seguir, enuciaremos os resultados que correspondem, respectivamente, à Proposicão 3.1.1 e ao Lema 3.1.2, com $\varphi(x)$ sendo a função definida em (3.16).

Proposição 3.2.2 *Se u é solução do Problema 4, então Au^+ é uma solução do Problema 3.*

Lema 3.2.3 *Tomemos $\lambda \in \mathbb{R}$. Se $\left\|A^\top QA - I\right\| < 1$ então o Problema 4 tem uma única solução.*

3.2.2 Método de Newton semi-suave

Com base nas discussões feitas na subseção 3.1.1 e na subseção anterior, consideremos para a função

$$F(x) := \left[A^\top QA - I\right] x^+ + x + A^\top b, \tag{3.18}$$

a sequência semi-suave de Newton é dada por

$$F(x_k) + S(x_k)\,(x_{k+1} - x_k) = 0, \qquad k = 0, 1, \ldots, \tag{3.19}$$

com um ponto inicial x_0 escolhido arbitrariamente e com o seguinte subgradiente

$$S(x_k) := \left[A^\top QA - I\right] \operatorname{diag}(\operatorname{sgn}(x_k^+)) + I, \qquad k = 0, 1, \ldots. \tag{3.20}$$

Como $\operatorname{diag}(\operatorname{sgn}(x^+))x = x^+$ para todo $x \in \mathbb{R}^n$, podemos afirmar por (3.18) e (3.20) que vale

$$S(x_k)x_k = \left[A^\top QA - I\right] x_k^+ + x_k = F(x_k) - A^\top b.$$

Fazendo simples manipulações na equação (3.19) e usando essa última igualdade obtemos

$$S(x_k)x_{k+1} = S(x_k)x_k - F(x_k) = -A^\top b, \qquad k = 0, 1, \ldots. \tag{3.21}$$

Se supormos que $S(x_k)$ é invertível para todo k podemos reescrever essa relação assim

$$x_{k+1} = -S(x_k)^{-1}A^\top b, \qquad k = 0, 1, \ldots. \tag{3.22}$$

Esta é uma definição equivalente da sequência semi-suave de Newton para resolver o Problema 4.

Os dois Lemas a seguir são casos particulares do Lema 3.1.3 e do Lema 3.1.4, respectivamente, em que a função convexa considerada é a função quadrática dada em (3.16).

Lema 3.2.4 *A matriz $S(x)$ definida em (3.20) é não singular para todo $x \in \mathbb{R}^n$. Como consequência, a sequência semi-suave de Newton $\{x_k\}$ está bem definida, para algum ponto inicial $x_0 \in \mathbb{R}^n$.*

Lema 3.2.5 *Tomando $S(x)$ sendo como definido em (3.20). Se $\left\|A^\top QA - I\right\| < 1$ para todo $x \in \mathbb{R}^n$ então*

$$\|S(x)^{-1}\| \leq \frac{1}{1 - \|A^\top QA - I\|}, \qquad \forall \;\; x \in \mathbb{R}^n.$$

O Lema a seguir toma por base o Lema 3.1.5. No entanto, devido a particularidade da função envolvida, ele é mais preciso; possui limitantes melhores que os obtidos pelo Lema 3.1.5 quando φ é a função quadrática.

Lema 3.2.6 *Sejam F uma função definida em (3.18) e $S(x)$ uma matriz definida em (3.20). Podemos afirmar que:*

$$\|S(x) - S(y)\| \leq \|A^\top QA - I\|, \qquad \forall \;\; x, y \in \mathbb{R}^n.$$

Como consequência,

$$\|F(x) - F(y) - S(y)(x - y)\| \leq \|A^\top QA - I\|\|x - y\|, \qquad \forall \;\; x, y \in \mathbb{R}^n.$$

Prova. Tomando $x, y \in \mathbb{R}^n$ e usando a definição (3.20) podemos afirmar que vale

$$\|S(x) - S(y)\| \leq \|A^\top QA - I\| \left\|\mathrm{diag}(\mathrm{sgn}(x^+)) - \mathrm{diag}(\mathrm{sgn}(y^+))\right\| \leq \|A^\top QA - I\|,$$

provando a primeira desigualdade do Lema. Já que estamos trabalhando em um caso particular do problema do capítulo passado, podemos afirmar, com base na equação (3.14), que vale

$$F(x) - F(y) - S(y)(x - y) = \int_0^1 \left[S(y + t(x - y)) - S(y)\right](x - y)\mathrm{d}t.$$

Aplicando norma nos dois lados da igualdade acima e usando propriedades da mesma, temos

$$\|F(x) - F(y) - S(y)(x - y)\| \leq \int_0^1 \|S(y + t(x - y)) - S(y)\| \, \|x - y\| \, \mathrm{d}t.$$

Usando a primeira desigualdade do Lema conclui-se a prova. ■

Teorema 3.2.7 *Seja $Q \in \mathbb{R}^{n\times n}$ sendo uma matriz simétrica positiva definida, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ e $A \in \mathbb{R}^{n\times n}$ uma matriz não singular. Se $\left\|A^\top QA - I\right\| < 1/2$ podemos afirmar que a sequência $\{x_k\}$ gerada pelo método de Newton semi-suave para resolver o Problema 4 converge linearmente para a única solução $u \in \mathbb{R}^n$ do Problema 4, como segue*

$$\|u - x_{k+1}\| \leq \frac{\|A^\top QA - I\|}{1 - \|A^\top QA - I\|}\|u - x_k\|, \qquad k = 0, 1, \ldots, \tag{3.23}$$

Como consequência, Au^+ é a solução do Problema 3.

Prova. Tomando F sendo a função definida em (3.18) e $S(x)$ a matriz definida em (3.20) podemos afirmar que se $u \in \mathbb{R}^n$ é uma solução do Problema 4, temos $F(u) = 0$. Partindo disso e usando a definição de $\{x_{k+1}\}$, dada em (3.19), temos

$$u - x_{k+1} = -S(x_k)^{-1} \left[F(u) - F(x_k) - S(x_k)(u - x_k) \right], \qquad k = 0, 1, \ldots.$$

Aplicando a norma nos dois membros e, posteriormente, usando propriedades da norma dadas em (2.3), segue que

$$\|u - x_{k+1}\| \leq \|S(x_k)^{-1}\| \, \|[F(u) - F(x_k) - S(x_k)(u - x_k)]\|, \qquad k = 0, 1, \ldots.$$

Combinando o Lema 3.2.5 com a segunda parte do Lema 3.2.6 concluimos que vale

$$\|u - x_{k+1}\| \leq \frac{\|A^\top QA - I\|}{1 - \|A^\top QA - I\|} \|u - x_k\|, \qquad k = 0, 1, \ldots. \tag{3.24}$$

Como $\|A^\top QA - I\| < 1/2$, vê-se facilmente que $\|A^\top QA - I\|/(1 - \|A^\top QA - I\|) < 1$. Portanto, a desigualdade em (3.24) implica que $\{x_k\}$ converge linearmente, a partir de qualquer ponto inicial, para a solução u do Problema 4. Daí, conclui-se a primeira parte do teorema. Como $u \in \mathbb{R}^n$ é uma solução do Problema 4, a segunda parte do teorema segue da Proposição 3.2.2. ■

A próxima proposição prova que, sob determinadas condições, a sequência converge para a solução do nosso problema em um número finito de passos.

Proposição 3.2.8 *Se $sgn(x_{k+1}^+) = sgn(x_k^+)$ para algum k então x_{k+1} é uma solução do Problema 4 e, consequentemente, Ax_{k+1}^+ é uma solução do Problema 3.*

Prova. Supondo que $\operatorname{sgn}(x_{k+1}^+) = \operatorname{sgn}(x_k^+)$ temos de (3.20) e de (3.21) que vale a igualdade

$$\left\{ \left[A^\top QA - I \right] \operatorname{diag}(\operatorname{sgn}(x_{k+1}^+)) + I \right\} x_{k+1} = -A^\top b. \tag{3.25}$$

Como $\operatorname{diag}(\operatorname{sgn}(x_{k+1}^+))x_{k+1} = x_{k+1}^+$ segue dessa igualdade que

$$\left[A^\top QA - I \right] x_{k+1}^+ + x_{k+1} = -A^\top b,$$

Logo x_{k+1} satisfaz a equação (3.17), ou seja, é uma solução do Problema 4. Portanto, seque da Proposição 3.2.2 que Ax_{k+1}^+ é solução do Problema 3. ■

A próxima proposição irá mostrar que a sequência gerada pelo método semi-suave de Newton $\{x_k\}$ é limitada. Além disso, tira conclusão sobre os seus pontos de acumulação.

Proposição 3.2.9 *A sequência $\{x_k\}$ gerada pelo método de Newton semi-suave* (3.22) *é limitada para qualquer ponto inicial. Além disso, para cada ponto de acumulação $\bar{x}$ de $\{x_k\}$ existe um $\hat{x} \in \mathbb{R}^n$ tal que*

$$\left(\left(A^\top QA - I \right) diag(sgn(\hat{x}^+)) + I \right) \bar{x} = -A^\top b. \tag{3.26}$$

Em particular, se $sgn(\bar{x}^+) = sgn(\hat{x}^+)$, então $\bar{x}$ é uma solução do Problema 4 e $A\bar{x}^+$ é solução do Problema 3.

Prova. Suponhamos, por contradição, que a sequência $\{x_k\}$ não é limitada. Partindo disso e sabendo que a esfera unitária em $\mathbb{R}^n$ é compacta podemos afirmar, respectivamente, que existe uma subsequência $\{x_{k_i}\} \subset \{x_k\}$ e um vetor $v \in \mathbb{R}^n$ tal que

$$\lim_{i\to\infty} \|x_{k_i+1}\| = \infty, \qquad \lim_{i\to\infty} \frac{x_{k_i+1}}{\|x_{k_i+1}\|} = v \neq 0. \tag{3.27}$$

Observe que existe uma quantidade finita de possibilidades para os vetores $\mathrm{sgn}(x_{k_i}^+)$. Logo, existem $\tilde{x} \in \mathbb{R}^n$ e uma subsequência $\{x_{k_j}\}$ de $\{x_{k_i}\}$ tal que

$$\mathrm{sgn}(x_{k_j}^+) = \mathrm{sgn}(\tilde{x}^+), \qquad \forall\, j \in \mathbb{N}.$$

Logo, usando a definição da sequência $\{x_k\}$ dada em (3.21) vê-se facilmente que

$$\left(\left(A^\top QA - I\right)\mathrm{diag}(\mathrm{sgn}(\tilde{x}^+)) + I\right)\frac{x_{k_i+1}}{\|x_{k_i+1}\|} = -\frac{A^\top b}{\|x_{k_i+1}\|}, \qquad i = 0, 1, 2, \ldots.$$

Fazendo i tender para o infinito nesta última igualdade e assumindo (3.27) temos

$$\left(\left(A^\top QA - I\right)\mathrm{diag}(\mathrm{sgn}(\tilde{x}^+)) + I\right) v = 0,$$

Isto contradiz a primeira parte do Lema 3.1.3. Portanto, a sequência $\{x_k\}$ é limitada, o que prova a primeira parte da proposição. Para provar a segunda parte, consideremos um ponto de acumulação $\bar{x}$ da sequência $\{x_k\}$. Dos argumentos usados no início da demonstração e da definição de ponto de acumulação segue que existe um vetor $\hat{x} \in \mathbb{R}^m$ e uma subsequência $\{x_{k_j}\} \subset \{x_k\}$ tal que

$$\lim_{j\to\infty} x_{k_j+1} = \bar{x}, \qquad \mathrm{sgn}(x_{k_j}^+) = \mathrm{sgn}(\hat{x}^+) \qquad \forall\, j \in \mathbb{N}. \tag{3.28}$$

Daí, usando novamente a definição da sequência $\{x_k\}$ dada em (3.21) obtemos que vale

$$\left(\left(A^\top QA - I\right)\mathrm{diag}(\mathrm{sgn}(\hat{x}^+)) + I\right) x_{k_j+1} = -A^\top b, \qquad j = 0, 1, 2, \ldots.$$

Tomando o limite com j indo para o ∞ e usando (3.28) obtemos

$$\left(\left(A^\top QA - I\right)\mathrm{diag}(\mathrm{sgn}(\hat{x}^+)) + I\right) \bar{x} = -A^\top b, \qquad j = 0, 1, 2, \ldots.$$

Assim, concluimos a prova da segunda parte da proposição. Partindo desta e assumindo $\mathrm{sgn}(\bar{x}^+) = \mathrm{sgn}(\hat{x}^+)$ temos

$$\left(\left(A^\top QA - I\right)\mathrm{diag}(\mathrm{sgn}(\bar{x}^+)) + I\right) \bar{x} = -A^\top b.$$

Portanto, como $\mathrm{diag}(\mathrm{sgn}(\bar{x}^+))\bar{x} = \bar{x}^+$, é fácil concluir que $\bar{x}^+$ é uma solução do Problema 4 e, usando a Proposição 3.2.2, obtemos que $A\bar{x}^+$ é solução do Problema 3, o que conclui a prova. ■

Capítulo 4

Método de Newton para equações semi-suaves

Na seção a seguir estudaremos uma equação semi-suave estudada inicialmente em [12] e [13]. Esta é análoga à estudada no capítulo anterior. A principal diferença é que a sua semi-suavidade é do tipo modular.

4.1 Equação com valor absoluto

Faremos o estudo do método de Newton semi-suave para resolver o seguinte problema.

Problema 5 *Encontre a solução da seguinte equação com valor absoluto*

$$Ax - |x| = b, \tag{4.1}$$

em que $A \in \mathbb{R}^{n\times n}$ é uma matriz não singular, $b \in \mathbb{R}^n$ é um vetor e $|x|$ denota o vetor cuja i-ésima coordenada é o módulo da i-ésima coordenada do vetor x.

Na proposição a seguir, provaremos que, sob uma determinada condição, podemos garantir a existência (e unicidade) de solução para o Problema 5.

Proposição 4.1.1 *Seja $\lambda \in (0,1)$. Se $\left\|A^{-1}\right\| < 1$, então o Problema 5 tem uma única solução.*

Prova. Multiplicando na esquerda os dois membros de (4.1) por A^{-1} obtemos

$$x - A^{-1}|x| = A^{-1}b.$$

Fazendo simples manipulações algébricas, podemos afirmar que esta igualdade equivale a

$$x = A^{-1}(b + |x|).$$

Agora, observemos que o Problema 5 tem uma solução se, e somente se, a função $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ definida por

$$\phi(x) = A^{-1}(b + |x|).$$

tem um ponto fixo.
Da definição da função ϕ temos que para todos $x, y \in \mathbb{R}^n$ vale a seguinte igualdade

$$\phi(x) - \phi(y) = A^{-1}(b + |x|) - A^{-1}(b + |y|).$$

Pondo A^{-1} em evidência e fazendo simples manipulações algébricas vê-se facilmente que vale

$$\phi(x) - \phi(y) = A^{-1}(|x| - |y|).$$

Aplicando a norma nos dois membros e usando propriedades da norma, temos

$$\|\phi(x) - \phi(y)\| \leq \left\|A^{-1}\right\| \| \, |x| - |y| \, \|, \qquad \forall \;\; x, y \in \mathbb{R}^n.$$

Usando a hipótese $\left\|A^{-1}\right\| < 1$ e a desigualdade provada no Lema 2.1.11, segue que

$$\|\phi(x) - \phi(y)\| \leq \|A^{-1}\| \|x - y\|, \qquad \forall \;\; x, y \in \mathbb{R}^n.$$

Assim ϕ é uma contração. Aplicando o Teorema 2.2.3 com $\mathbb{X} = \mathbb{R}^n$ e $\rho = \|.\|$, podemos concluir que ϕ tem precisamente um único ponto fixo. Como consequência, podemos afirmar que o Problema 5 tem solução, e esta é única.

■

4.1.1 O método de Newton semi-suave

Com base na equação (4.1) definamos a função semi-suave $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ da seguinte forma

$$g(x) = Ax - |x| - b. \tag{4.2}$$

Considerando essa definição e discussões análogas, feitas anteriormente, afirmamos que

$$S(x) = A - \mathrm{diag}(\mathrm{sgn}(x)) \in \partial g(x), \qquad \forall \; x \in \mathbb{R}^n. \tag{4.3}$$

Apesar dessa afirmação ser aceitável, sua justificativa é um pouco técnica. Portanto vale a pena ser feita. Note que, denotando $\mathrm{diag}(\mathrm{sgn}(x))$ por $D(x)$, podemos reescrever (4.2) da forma

$$g(x) = Ax - D(x)x - b,$$

pois pela definição da matriz $D(x)$, podemos afirmar que $D(x)x = |x|$. Logo, podemos concluir que (4.3) se verifica quando x pertence ao interior de qualquer octante (ou seja, quando todas as suas coordenadas forem diferentes de 0), pois dessa maneira teremos uma vizinhança na qual $sgn(x)$ e, consequentemente, $D(x)$ é constante e, portanto, $g(x)$ torna-se diferenciável nessa vizinhança. Logo,

$$g'(x) = A - D(x), \qquad \forall \; x \in \Omega,$$

em $\Omega \subset \mathbb{R}^n$ é o conjunto no qual g é diferenciável. Sendo assim, basta verificarmos o caso em que $x \in \Omega^c$. Fixe um $x = (x_1, ..., x_n)$ nesse conjunto e tome duas sequências, $\{z_m\} = (z_{1(m)}, ..., z_{n(m)})$ e $\{y_m\} = (y_{1(m)}, ..., y_{n(m)})$, definidas da seguinte maneira

$$\begin{cases} \text{se} \quad x_i \neq 0, \quad z_{i(m)} = x_i, \qquad y_{i(m)} = x_i, \quad i = 1, ..., n, \\ \text{se} \quad x_i = 0, \quad z_{i(m)} = -\frac{1}{m}, \qquad y_{i(m)} = \frac{1}{m}, \quad i = 1, ..., n. \end{cases} \tag{4.4}$$

Note que ambas as sequências convergem para x e pertencem ao interior de algum octante (ou seja, $\text{sgn}(\{x_m\})$ e $\text{sgn}(\{y_m\})$ são constantes). Assim, das discussões acima, podemos afirmar que

$$S(z_m) = A - D(z_m) = A - D_1, \qquad S(y_m) = A - D(y_m) = A - D_2$$

em que $D_1 = \text{diag}(\text{sgn}(\{z_m\}))$ e $D_2 = \text{diag}(\text{sgn}(\{y_m\}))$ para todo $m \in \mathbb{N}$. Daí,

$$S(z_m) \longrightarrow A - D_1, \qquad S(y_m) \longrightarrow A - D_2.$$

Portanto, pela Definição 2.2.11, $A - D_1$ e $A - D_2$ pertencem a $\partial g(x)$. Além disso, vale

$$A - D(x) = A - \frac{D_1 + D_2}{2} = \frac{1}{2}(A - D_1) + \frac{1}{2}(A - D_2) \in \partial g(x).$$

Daí, concluimos que (4.3), de fato, se verifica para todo $x \in \mathbb{R}^n$. Portanto, usando o método de Newton semi-suave para encontrar a solução da equação $g(x) = 0$ recaímos em uma igualdade análoga a (3.19). Como segue,

$$g(x_k) + S(x_k)(x_{k+1} - x_k) = 0.$$

Da definição de g, dada em (4.2), juntamente com o seu subgradiente de Clarke, dado em (4.3), temos

$$Ax_k - |x_k| - b + (A - D(x_k))(x_{k+1} - x_k) = 0.$$

Aplicando a distributiva e reduzindo os termos semelhantes na igualdade acima, obtemos

$$Ax_{k+1} - |x_k| - b - D(x_k)x_{k+1} + D(x_k)x_k = 0.$$

Observemos que $D(x_k)x_k = |x_k|$. Portanto, podemos reescrever a última equação da seguinte maneira

$$Ax_{k+1} - b - D(x_k)x_{k+1} = 0.$$

Após simples manipulações algébricas, podemos afirmar que a igualdade acima equivale a

$$[A - D(x_k)]x_{k+1} = b. \tag{4.5}$$

Assim, se supormos que $A - D(x_k)$ é invertível, obtemos x_{k+1} em função de x_k. Como segue:

$$x_{k+1} = (A - D(x_k))^{-1}b. \tag{4.6}$$

Esta é a iteração do método semi-suave de Newton para resolver o Problema 5.
Com base na discussões feitas no início desta subseção, vê-se a necessidade de alguns resultados para provar a convergência e boa definição do método. Então, comecemos com o seguinte Lema.

Lema 4.1.2 *Tomando $S(x)$ sendo como definido em* (3.20)*. Se* $\left\|A^{-1}\right\| < 1$ *temos que* $A - D(x)$ *é não singular para todo* $x \in \mathbb{R}^n$*, e vale*

$$\|(A - D(x))^{-1}\| \leq \frac{\left\|A^{-1}\right\|}{1 - \|A^{-1}\|}, \qquad \forall \ \ x \in \mathbb{R}^n.$$

Como consequência, a iteração do Método de Newton semi-suave, dada em (4.6)*, está bem definida.*

Prova. Para simplificar as contas, trataremos o $D(x)$ apenas por D. Usando o fato que a matriz A é invertível, temos

$$A - D = A(I - A^{-1}D). \tag{4.7}$$

Como a matriz D é diagonal e possui apenas -1, $+1$ ou 0 como entradas podemos afirmar que $\|D\| \leq 1$. Assim, da hipótese do Lema, obtemos

$$\left\|A^{-1}D\right\| \leq \left\|A^{-1}\right\| \|D\| \leq \left\|A^{-1}\right\| < 1. \tag{4.8}$$

Usando o Lema 2.2.1 com $A^{-1}D$ no lugar de E temos que $I - A^{-1}D$ é invertível e vale

$$\left\|(I - A^{-1}D)^{-1}\right\| \leq \frac{1}{1 - \|A^{-1}D\|}.$$

Fazendo uso das desigualdades em (4.8) e da desigualdade anterior vê-se que

$$\left\|(I - A^{-1}D)^{-1}\right\| \leq \frac{1}{1 - \|A^{-1}D\|} \leq \frac{1}{1 - \|A^{-1}\|}. \tag{4.9}$$

Como A e $I - A^{-1}D$ são invertíveis, se usarmos propriedades da inversa do produto de matrizes em (4.7), podemos concluir que $A - D$ é invertível e vale

$$(A - D)^{-1} = (I - A^{-1}D)^{-1}A^{-1}.$$

Aplicando norma nos dois membros, obtemos

$$\left\|(A - D)^{-1}\right\| = \left\|(I - A^{-1}D)^{-1}A^{-1}\right\| \leq \left\|(I - A^{-1}D)^{-1}\right\| \left\|A^{-1}\right\|.$$

Dessa última desigualdade e de (4.9), obtemos a desigualdade do Lema. ■

No teorema a seguir será provado que, mediante hipótese adequada, a iteração (4.6) converge para a solução, cuja a existência foi garantida pela Proposição 4.1.1.

Teorema 4.1.3 *Sejam* $A \in \mathbb{R}^{n \times n}$ *uma matriz não singular e* b *um vetor de* $\mathbb{R}^n$*. A sequência* $\{x_k\}$ *gerada pelo método de Newton semi-suave* (4.6) *para resolver o Problema 5 está bem definida para qualquer ponto inicial* $x_0 \in \mathbb{R}^n$*. Além disso, se* $\|A^{-1}\| < 1/3$*, então a sequência* $\{x_k\}$ *converge linearmente para* $u \in \mathbb{R}^n$*, a única solução do Problema 5, como segue*

$$\|u - x_{k+1}\| \leq \frac{2\|A^{-1}\|}{1 - \|A^{-1}\|}\|u - x_k\|, \qquad k = 0, 1, \ldots.$$

Prova. Como, por hipótese, $\|A^{-1}\|$ é, em particular, menor que 1, podemos afirmar, pela Proposição 4.1.1, que o Problema 5 tem uma única solução. Denotando essa solução por u, temos

$$Au - |u| = b.$$

Logo, por (4.5), segue que

$$[A - D(x_k)]x_{k+1} = Au - |u|\,.$$

Fazendo simples manipulações algébricas, podemos afirmar que

$$A(x_{k+1} - u) = D(x_k)x_{k+1} - |u|\,.$$

Somando o termo $-D(x_k)(x_{k+1} - u)$ em ambos os membros da última igualdade, obtemos

$$(A - D(x_k))(x_{k+1} - u) = D(x_k)u - |u|\,.$$

Usando o fato que $D(x_k)x_k = |x_k|$, ou seja, $|x_k| - D(x_k)x_k = 0$, podemos afirmar que

$$D(x_k)u - |u| = |x_k| - D(x_k)x_k + D(x_k)u - |u|\,.$$

Das duas últimas igualdades temos

$$(A - D(x_k))(x_{k+1} - u) = |x_k| - |u| - D(x_k)(x_k - u).$$

Pelo Lema 4.1.2, $A - D(x_k)$ é invertível, multiplicando a última igualdade por $(A - D(x_k))^{-1}$, colocando norma nos dois membros e aplicando a desigualdade triangular obtemos

$$\|x_{k+1} - u\| \leq \left\|(A - D(x_k))^{-1}\right\| \left(\||x_k| - |u|\| + \|D(x_k)\| \, \|x_k - u\|\right).$$

Usando o fato que $\|D(x_k)\| \leq 1$ e as desigualdades do Lema 4.1.2 e do Lema 2.1.11, segue que

$$\|x_{k+1} - u\| \leq \frac{2\left\|A^{-1}\right\|}{1 - \|A^{-1}\|} \|x_k - u\|\,.$$

Como $\|A^{-1}\| < 1/3$, vê-se que $2\|A^{-1}\|/(1 - \|A^{-1}\|) < 1$. Portanto, usando a Proposição 2.2.4 com

$$\alpha = \frac{2\|A^{-1}\|}{1 - \|A^{-1}\|} < 1,$$

obtemos que a sequência $\{x_k\}$ converge linearmente, a partir de qualquer ponto inicial, para a solução u do Problema 5. ■

4.2 Equação semi-suave com operador x^+

Agora, abordaremos um problema parecido com o da seção anterior. Porém, na equação a ser resolvida aparece o termo x^+ ao invés de $|x|$.

Problema 6 *Encontre as soluções da equação linear por partes*

$$Tx + x^+ = m, \tag{4.10}$$

em que $T \in \mathbb{R}^{n\times n}$ é uma matriz não singular e $m \in \mathbb{R}^n$ um vetor dado.

A estratégia para resolvermos esse problema, mediante determinadas condições, será tranformá-lo no Problema 5 e tirar conclusões a partir dos resultados da seção anterior. A seguir, apresentamos essa estratégia com detalhes.

4.2.1 Discussão e resolução da equação semi-suave com operador x_+

Do exposto no início do Capítulo 2, podemos afirmar que valem as igualdades

$$x = x^+ - x^-, \qquad |x| = x^+ + x^-, \qquad \forall\ x \in \mathbb{R}^n.$$

Somando essas igualdades, é fácil ver que $x + |x| = 2x^+$. Portanto,

$$x^+ = \frac{x + |x|}{2}, \qquad \forall\ x \in \mathbb{R}^n.$$

Sendo assim, com o objetivo de transformar o Problema 6 no Problema 5, podemos reescrever a equação (4.10) da seguinte maneira

$$Tx + \frac{x + |x|}{2} = m.$$

Fazendo simples manipulações algébricas, vê-se que a equação anterior equivale a

$$(-2T - I)x - |x| = -2m.$$

Fazendo $\tilde{T} = -2T - I$ e $\tilde{m} = -2m$, conclui-se que

$$\tilde{T}x - |x| = \tilde{m}. \tag{4.11}$$

Assim, colocamos a equação do Problema 6 na forma da equação do Problema 5. Portanto, em relação ao problema 4.11 valem os resultados que foram provados na seção anterior. Reescreveremos estes com as alterações necessárias. Primeiramente, a proposição que garante a existência (e unicidade) de uma solução v da equação (4.11), equivalentemente, do Problema 6.

Proposição 4.2.1 *Seja $\lambda \in \mathbb{R}$. Se $\left\|\tilde{T}^{-1}\right\| < 1$, então o Problema 6 tem uma única solução.*

O Lema a seguir garante que a iteração semi-suave de Newton para resolver o Problema 6 está bem definida.

Lema 4.2.2 *Se $\left\|\tilde{T}^{-1}\right\| < 1$, temos que $\tilde{T} - D(x)$ é não singular para todo $x \in \mathbb{R}^n$ e vale*

$$\|(\tilde{T} - D(x))^{-1}\| \leq \frac{\left\|\tilde{T}^{-1}\right\|}{1 - \left\|\tilde{T}^{-1}\right\|}, \qquad \forall\ \ x \in \mathbb{R}^n.$$

O próximo resultado garante que a sequência gerada pelo método semi-suave de Newton converge para a solução do nosso problema.

Teorema 4.2.3 *Tomando $\tilde{T} \in \mathbb{R}^{n\times n}$ sendo uma matriz não singular e $\tilde{m} \in \mathbb{R}^n$. Então, a sequência $\{x_k\}$ gerada pelo método de Newton semi-suave para resolver o Problema 6,*

$$x_{k+1} = (\tilde{T} - D(x_k))^{-1}\tilde{m}, \qquad k = 0, 1, \ldots,$$

está bem definida para algum ponto inicial $x_0 \in \mathbb{R}^n$. Além disso, se $\|\tilde{T}^{-1}\| < 1/3$ então a sequência $\{x_k\}$ converge linearmente para $v \in \mathbb{R}^n$, a única solução do Problema 6, como segue

$$\|v - x_{k+1}\| \leq \frac{2\|\tilde{T}^{-1}\|}{1 - \|\tilde{T}^{-1}\|}\|v - x_k\|, \qquad k = 0, 1, \ldots.$$

No corolário a seguir, veremos uma condição suficiente para garantirmos todos os resultados expostos na seção anterior.

Corolário 4.2.4 *Seja T uma matriz não singular tal que $\left\|T^{-1}\right\| < 1$. A matriz $\tilde{T} = -2T - I$ é não-singular e tem a norma $\left\|\tilde{T}^{-1}\right\|$ menor que 1. E quando $\left\|T^{-1}\right\|$ for menor que $1/2$ teremos*

$$\left\|\tilde{T}^{-1}\right\| < 1/3.$$

Prova. Como a matriz T é invertível, podemos reescrever $\tilde{T}$ da seguinte maneira

$$\tilde{T} = 2T\left(-I - \frac{T^{-1}}{2}\right). \tag{4.12}$$

Usando o fato que a matriz T é invertível e fazendo $E = -T^{-1}/2$ no Lema 2.2.1, para concluirmos que $-I - T^{-1}/2$ também é invertível, podemos afirmar, da última igualdade, que $\tilde{T}$ é invertível. E aplicando propriedades da inversa do produto de matrizes em (4.12) obtemos

$$\tilde{T}^{-1} = \left(-I - \frac{T^{-1}}{2}\right)^{-1}(2T)^{-1} = \frac{1}{2}\left(-I - \frac{T^{-1}}{2}\right)^{-1}T^{-1}.$$

Da última igualdade obtemos que

$$\left\|\tilde{T}^{-1}\right\| \leq \frac{1}{2}\left\|\left(-I - \frac{T^{-1}}{2}\right)^{-1}\right\|\left\|T^{-1}\right\|. \tag{4.13}$$

Como $\left\|T^{-1}\right\| < 1$ segue que $\left\|T^{-1}/2\right\| < 1$. Portanto, aplicando o Lema 2.2.1 com $E = -T^{-1}/2$ obtemos a seguinte desigualdade

$$\left\|\left(I + \frac{T^{-1}}{2}\right)^{-1}\right\| \leq \frac{1}{1 - \left\|T^{-1}\right\|/2} = \frac{2}{2 - \left\|T^{-1}\right\|}.$$

Usando esta última desigualdade juntamente com a desigualdade (4.13), temos que

$$\left\|\tilde{T}^{-1}\right\| \leq \frac{1}{2}\left\|\left(I+\frac{T^{-1}}{2}\right)^{-1}\right\|\left\|T^{-1}\right\| \leq \frac{\left\|T^{-1}\right\|}{2-\left\|T^{-1}\right\|}.$$

A partir dessa desigualdade, a prova segue diretamente. ■

A seguir, apresentaremos um exemplo de uma matriz T, cuja norma da sua inversa é maior que $1/2$ mas $\left\|\tilde{T}^{-1}\right\| = \left\|(2T+I)^{-1}\right\| < 1/3$. Assim, conseguimos explicitar que realmente a condição $\left\|T^{-1}\right\| < 1/2$ não é necessária, apenas suficiente. Ou seja, a condição $\left\|\tilde{T}^{-1}\right\| < 1/3$ é satisfeita por mais matrizes que a hipótese $\left\|T^{-1}\right\| < 1/2$.

Exemplo 4.2.5 Considere

$$T = \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

Daí, podemos afirmar que

$$T^{-1} = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{2}{3} \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Sendo assim, como a norma da matriz identidade é 1, segue que $\left\|T^{-1}\right\| = 2/3$. Portanto, temos que $\left\|T^{-1}\right\|$ é maior que $1/2$. Agora, analisemos $\left\|(2T+I)^{-1}\right\|$. Para isso, calculemos

$$2T+I = 2\begin{pmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix},$$

segue que

$$(2T+I)^{-1} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Logo $\left\|(2T+I)^{-1}\right\| = 1/4$. Como $1/4$ é menor que $1/3$, temos que a matriz T satisfaz as condições desejadas.

Observação 4.2.6 Note que podemos reescrever a equação (4.10) da seguinte forma

$$T^{-1}x^{+} + x - T^{-1}m = 0. \tag{4.14}$$

Assim, podemos afirmar que o Problema 6 é uma espécie de generalização do Problema 4, pois se tomarmos

$$T = (A^{\top}QA - I)^{-1}, \qquad m = -(A^{\top}QA - I)^{-1}A^{\top}b.$$

obteremos exatamente a equação do Problema 4. Para verificarmos, basta substituirmos T e m em (4.14) como segue:

$$[(A^{\top}QA - I)^{-1}]^{-1}x^{+} + x - [(A^{\top}QA - I)^{-1}]^{-1}[-(A^{\top}QA - I)^{-1}A^{\top}b] = 0.$$

Daí, fazendo simples manipulações algébricas podemos afirmar que a equação anterior equivale a

$$(A^\top QA - I)x^+ + x + A^\top b = 0,$$

obtendo a equação (4).

Capítulo 5

Considerações finais

Devido a metodologia para minimizar uma função quadrática sobre um cone simplicial empregada no capítulo 3 ser recente, ainda há muito a ser feito nesta direção. Objetivando contribuir com o tema decidimos trabalhar no caso convexo com o tentativa de generalizar o caso quadrático. Apesar do sucesso em boa parte do estudo, não conseguimos obter os mesmos limitantes do caso quadrático, a saber, tivemos que enfraquecer uma hipótese no Lema 3.1.5. Outra contribuição deste trabalho foi a estratégia para resolver o problema 6 dado na seção 4.2. Esta, por sua vez, permite que tenhamos limitantes melhores que os encontrados em [1] para resolver o mesmo problema. Essa vantagem é explicitada pelo Corolário 4.2.4 e pelo Exemplo 4.2.5.

Outro caso interessante, que está por fazer, é usar esse método para minimizar uma função (linear, por exemplo) sobre o cone das matrizes semi-definidas positivas. Para isso, com base no que foi feito, não é difícil ver que seria necessário encontrar um subgradiente de Clarke específico da projeção de uma matriz arbitrária de $\mathbb{R}^{n\times n}$ sobre este cone, o que parece ser algo não trivial. Há alguns trabalhos que fazem algo parecido. Por exemplo [11], que explicita a projeção das matrizes simétricas e encontra um subgradiente de Clarke para esta. Mas isto é, aparentemente, insuficiente para resolver nosso problema, pois precisamos desse resultado para uma matriz arbitrária de $\mathbb{R}^{n\times n}$.

Referências Bibliográficas

[1] J. Barrios, J. B. Cruz, O. Ferreira, and S. Németh. A semi-smooth newton method for a special piecewise linear system with application to positively constrained convex quadratic programming. *Journal of Computational and Applied Mathematics*, 301:91 – 100, 2016.

[2] J. Barrios, O. Ferreira, and S. Németh. A semi-smooth newton method for solving convex quadratic programming problem under simplicial cone constraint. *arXiv preprint arXiv:1503.02753*, 2015.

[3] R. G. Bartle. *A modern theory of integration*, volume 32. American mathematical society Providence, 2001.

[4] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming: theory and algorithms*. John Wiley & Sons, 2013.

[5] J. Y. Bello Cruz, O. P. Ferreira, and L. F. Prudente. On the global convergence of the inexact semi-smooth newton method for absolute value equation. *Computational Optimization and Applications*, pages 1–16, 2016.

[6] L. Brugnano and V. Casulli. Iterative solution of piecewise linear systems. *SIAM Journal on Scientific Computing*, 30(1):463–472, 2008.

[7] F. H. Clarke. *Optimization and nonsmooth analysis*, volume 5 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990.

[8] A. L. Dontchev and R. T. Rockafellar. Implicit functions and solution mappings. *Springer Monographs in Mathematics. Springer*, 208, 2009.

[9] J. Heinonen. *Lectures on Lipschitz analysis*. Univ., 2005.

[10] A. Izmailov and M. Solodov. Otimização, volume i. *Rio de Janeiro, IMPA*, 2007.

[11] J. Malick and H. S. Sendov. Clarke generalized jacobian of the projection onto the cone of positive semidefinite matrices. *Set-Valued Analysis*, 14(3):273–293, 2006.

[12] O. Mangasarian and R. Meyer. Absolute value equations. *Linear Algebra and its Applications*, 419(2-3):359 – 367, 2006.

[13] O. L. Mangasarian. A generalized Newton method for absolute value equations. *Optim. Lett.*, 3(1):101–108, 2009.

[14] K. G. Murty and Y. Fathi. A critical index algorithm for nearest point problems on simplicial cones. *Mathematical Programming*, 23(1):206–215, 1982.

[15] K. G. Murty and F.-T. Yu. *Linear complementarity, linear and nonlinear programming.* Citeseer, 1988.

[16] L. Qi and J. Sun. A nonsmooth version of newton's method. *Mathematical programming*, 58(1-3):353–367, 1993.