UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ALGORITMO DE PONTO PROXIMAL PARA OTIMIZAÇÃO EM $\mathbb{R}^n$

por

Sandra Regina Peres da Silva

Dissertação de Mestrado em Matemática

Agradecimentos

Ao professor Doutor Orizon Pereira Ferreira pela escolha do tema, orientação e paciência, que muito contribuíram na elaboração desta dissertação.

À banca examinadora pelas sugestões que contribuíram para a melhoria deste trabalho.

Ao meu esposo Mauricio e ao meu filho Tiago pelo carinho, incentivo e compreensão durante a realização do curso de Mestrado.

Aos meus familiares, em especial à minha mãe Izabel, pelo incentivo irrestrito.

Aos colegas de curso, pela amizade e ajuda mútua.

Aos professores e funcionários do Instituto de Matemática e Estatística da UFG, pela oportunidade e apoio.

Ao meu esposo Mauricio e
ao meu filho Tiago

Resumo

Nesta dissertação estudamos um algoritmo conceitual para minimização de uma função convexa, não necessariamente diferenciável, conhecido por *algoritmo de ponto proximal*. Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. O algoritmo de ponto proximal é um processo iterativo que, a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, gera recursivamente uma seqüência de pontos $\{x_k\} \subset \mathbb{R}^n$ pela iteração: $x_{k+1} := \arg\min_{x\in\mathbb{R}^n}\{f(x) + \lambda_k\|x_k - x\|^2\}$, onde $\{\lambda_k\}$ é uma seqüência de números positivos.

Mostramos, primeiramente, que a função estritamente convexa $f_k(x) := f(x)+\lambda_k\|x_k-x\|^2$ possui um único minimizador. Logo, o algoritmo está bem definido. Em seguida estabelecemos a convergência do algoritmo baseados na *Fejér* convergência da seqüência $\{x_k\}$ ao conjunto dos minimizadores da função convexa f, caso ele seja não vazio. Caso a função convexa f não possua minimizador mostramos que a seqüência gerada pelo algoritmo divergirá.

Também estudamos o algoritmo de ponto proximal para operadores monótonos maximais arbitrários, $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$. Mostramos que este algoritmo está bem definido e que, se existir $\overline{x} \in \mathbb{R}^n$ tal que $0 \in T(\overline{x})$, a seqüência gerada pelo algoritmo convergirá para uma singularidade do operador. Além disso, mostramos que o subdiferencial de uma função convexa é um operador monótono maximal e assim, determinar suas singularidades generaliza, em um certo sentido, o problema de minimização convexa.

Abstract

In this dissertation, we study a conceptual algorithm for minimization of the convex function, not necessarily differential, known by *proximal point algorithm*. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. The proximal point algorithm is an iterative procedure, which starts at a point $x_0 \in \mathbb{R}^n$, generate recursively a sequence of points $\{x_k\} \subset \mathbb{R}^n$ by iteration: $x_{k+1} := \arg\min_{x\in\mathbb{R}^n}\{f(x) + \lambda_k\|x_k - x\|^2\}$, where $\{\lambda_k\}$ is a sequence of positive numbers.

We show, firstly, that the strictly convex function $f_k(x) := f(x) + \lambda_k\|x_k - x\|^2$ has a unique minimizer. Thus, the algorithm is well defined. Next we establish the convergence of algorithm based upon the of *Fejér* convergence of sequence $\{x_k\}$ to set of minimizers of the convex function f, case it is not empty. If the convex function f not has minimizer we show that the sequence generate for the algorithm diverge.

We also study the proximal point algorithm for arbitrary maximal monotone operators, $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$. We show that this algorthm is well defined and that, if exists $\overline{x} \in \mathbb{R}^n$ such that $0 \in T(\overline{x})$, the sequence generated for algorithm converge to a singularity of the operator. In addition, we show that the subdifferential of the a convex function is a maximal monotone operator and so, to determine its singularities generalizes, in a certain sense, the convex minimization problem.

Índice

CAPÍTULO I

Introdução

Nesta dissertação estudaremos um algoritmo conceitual para minimizacão de uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ convexa, não necessariamente diferenciável. Esse algoritmo, conhecido por *algoritmo de ponto proximal* de acordo com a terminologia de Rockafellar [10], foi aplicado pela primeira vez em otimização convexa por Martinet [6], na década de 70.

O algoritmo de ponto proximal é um processo iterativo para minimizar uma função convexa que, a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, gera recursivamente uma seqüência de pontos $\{x_k\} \subset \mathbb{R}^n$ pela iteração:

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

onde $\{\lambda_k\}$ é uma seqüência de números positivos. A iteração acima foi introduzida por Moreau [8] no ano de 1965.

Para estabelecer a convergência do algoritmo de ponto proximal, uma suposição usual é a existência de um mínimo da função objetivo; isto é, se existe $x_* \in \mathbb{R}^n$ tal que $f(x_*) \leq f(x)$, para todo $x \in \mathbb{R}^n$, então a seqüência $\{x_k\}$ converge a este minimizador. Esta suposição foi retirada por Osman Güller [2] em 1991, ou seja, se a função objetivo não possui minimizador, então a seqüência $\{x_k\}$ diverge. Em 1993, Rafael Correa e Claude Lemaréchal [1] estabeleceram de maneira simples e unificada a convergência de vários métodos de minimização convexa, dentre eles, o método de ponto proximal. A análise de convergência que faremos baseia-se na noção de Fejér convergência da seqüência $\{x_k\}$ ao conjunto dos minimizadores da função convexa f (veja Iusem [4]).

O algoritmo de ponto proximal também é fundamental para resolver o problema de encontrar singularidades de operadores monótonos maximais. Estudaremos este importante resultado devido à Rockafellar [11], pois o problema de encontrar pontos singulares de operador monótono maximal generaliza o problema de minimização de funções convexas.

Com o objetivo de tornar este trabalho mais completo, destinamos o Capítulo II à uma breve revisão de análise no $\mathbb{R}^n$ e o Capítulo III à um resumo dos conceitos básicos, relativos aos conjuntos convexos e às funções convexas, que serão necessários ao desenvolvimento dos capítulos seguintes. Apesar dos conceitos de análise convexa já serem conhecidos, nós os discutiremos com detalhes, principalmente para uniformizar as notações.

A parte central deste trabalho encontra-se nos Capítulos IV e VI. No Capítulo IV mostraremos que o algoritmo de ponto proximal para minimização convexa está bem definido e que a seqüência gerada por ele converge para um minimizador da função, caso exista algum. No Capítulo V, introduziremos o conceito de operadores monótonos em $\mathbb{R}^n$, operadores monótonos ponto-conjunto e operadores monótonos maximais e encerraremos o capítulo mostrando que o subdiferencial de uma função convexa é um operador monótono maximal. Finalmente, no Capítulo VI estudaremos o algoritmo de ponto proximal para operadores monótonos maximais arbitrários, $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$. Mostraremos que este algoritmo está bem definido e que, se existir $\overline{x} \in \mathbb{R}^n$ tal que $0 \in T(\overline{x})$, a seqüência gerada pelo algoritmo convergirá para uma singularidade do operador.

CAPÍTULO II

Topologia do $\mathbb{R}^n$: Um resumo dos conceitos básicos

II.1 Introdução

Neste capítulo apenas enunciaremos alguns conceitos relacionados à topologia do espaço Euclideano, os quais serão utilizados nos Capítulos III e IV. Por tratarem-se de resultados básicos de análise no $\mathbb{R}^n$ as demonstrações serão omitidas e poderão ser encontradas na referência [5].

II.2 Noções topológicas no espaço Euclideano

A *bola aberta* de centro $a \in \mathbb{R}^n$ e raio $r > 0$ é o conjunto dos pontos $x \in \mathbb{R}^n$ cuja distância ao ponto a é menor do que r. Usaremos a notação $B(a; r)$ para indicar esse conjunto. Assim

$$B(a; r) = \{x \in \mathbb{R}^n : \|x - a\| < r\}.$$

Definição II.2.1. Um conjunto $X \subset \mathbb{R}^n$ diz-se *limitado* quando existe um número real $c > 0$ tal que $\|x\| \leq c$ para todo $x \in X$.

Seja $X \subset \mathbb{R}^n$. Um ponto $a \in X$ chama-se um *ponto interior* a X quando é centro de alguma bola aberta contida em X, ou seja, quando existe $\delta > 0$ tal que $\|x - a\| < \delta$ implica $x \in X$. O *interior* de X é o conjunto $\operatorname{int} X$, formado pelos pontos interiores a X. Quando $x \in \operatorname{int} V$, dizemos que o conjunto V é uma *vizinhança* do ponto x.

Definição II.2.2. Um conjunto $X \subset \mathbb{R}^n$ chama-se *aberto* quando todos os seus pontos são interiores, isto é, quando para cada $x \in X$ existe $\delta > 0$ tal que $B(x; \delta) \subset X$.

Definição II.2.3. Seja $X \subset \mathbb{R}^n$. Um ponto $a \in \mathbb{R}^n$ chama-se *ponto de acumulação* do conjunto X quando toda bola aberta de centro a contém algum ponto de X, diferente do ponto a. Noutros termos, para todo $\epsilon > 0$, deve existir $x \in X$ tal que $0 < \|x - a\| < \epsilon$.

Diz-se que um ponto x é um *ponto de aderência* de um conjunto $X \subset \mathbb{R}^n$, se qualquer vizinhança de x contém algum elemento de X. Isso significa que x pode ser um elemento de X ou não, mas se não for certamente será ponto de acumulação de X. O conjunto dos pontos aderentes a X chama-se o *fecho* de X.

Definição II.2.4. Um conjunto $X \subset \mathbb{R}^n$ chama-se *fechado* quando contém todos os seus pontos de aderência.

O teorema a seguir caracteriza os conjuntos fechados.

Teorema II.2.5. *Um conjunto é fechado se, e somente se, seu complementar é aberto.*

Definição II.2.6. Diremos que um conjunto $K \subset \mathbb{R}^n$ é *compacto* quando ele for fechado e limitado.

II.3 Seqüências no espaço Euclideano

Uma *seqüência* em $\mathbb{R}^n$ é uma função $x\colon \mathbb{N} \to \mathbb{R}^n$ definida no conjunto $\mathbb{N}$ dos números naturais. O valor que essa função assume no número k é indicado com x_k e chama-se o *k-ésimo termo* da seqüência. Usaremos a notação $\{x_k\}$ para indicar a seqüência cujo k-ésimo termo é $x_k \in \mathbb{R}^n$. Uma *subseqüência* de $\{x_k\}$ é a restrição da seqüência a um subconjunto infinito $\mathbb{N}' = \{k_1 < k_2 < ... < k_i < ...\} \subset \mathbb{N}$. Indicaremos uma subseqüência pela notação $\{x_{k_j}\}$.

Diz-se que a seqüência $\{x_k\}$ é *limitada* quando o conjunto dos seus termos é limitado em $\mathbb{R}^n$, ou seja, quando existe um número real $c > 0$ tal que $\|x_k\| \leq c$ para todo $k \in \mathbb{N}$.

Uma seqüência $\{x_k\}$ de números reais chama-se *monótona não-decrescente* quando se tem $x_k \leq x_{k+1}$ para todo $k \in \mathbb{N}$ e chama-se *monótona não-crescente* quando se tem $x_{k+1} \leq x_k$ para todo $k \in \mathbb{N}$.

Quando existe o limite $a = \lim\limits_{k\to+\infty} x_k$, diz-se que a seqüência $\{x_k\}$ é convergente.

Os teoremas enunciados a seguir têm papéis importantes na demonstração da convergência do algoritmo de ponto proximal para funções convexas. Tratam-se de resultados clássicos de análise no $\mathbb{R}^n$, por isso as demonstrações serão omitidas.

Teorema II.3.1. *Dados $X \subset \mathbb{R}^n$ e $a \in \mathbb{R}^n$, as seguintes afirmações são equivalentes:*
(i) a é ponto de acumulação de X;
(ii) Existe uma seqüência de pontos $x_k \in X$, com $\lim\limits_{k\to+\infty} x_k = a$ e $x_k \neq a$ para todo $k \in \mathbb{N}$;
(iii) Toda bola aberta de centro a contém uma infinidade de pontos de X.

Teorema II.3.2. *(Bolzano-Weierstrass) Toda seqüência limitada em $\mathbb{R}^n$ possui uma subseqüência convergente.*

Teorema II.3.3. *Toda seqüência monótona limitada é convergente.*

Teorema II.3.4. *Toda subseqüência de uma seqüência convergente converge para o mesmo limite da seqüência.*

II.4 Funções contínuas

A propriedade de continuidade das funções convexas permitirá, dentre outros resultados, mostrar a boa definição do algoritmo de ponto proximal. Enunciaremos nesta seção os principais teoremas que serão utilizados no Capítulo IV.

Seja $f\colon X \to \mathbb{R}$ uma função definida no conjunto $X \subset \mathbb{R}^n$. Diz-se que f é *contínua* no ponto $a \in X$ quando, para qualquer $\epsilon > 0$ dado, pode-se obter $\delta > 0$ tal que $\|x - a\| < \delta$ implica $|f(x) - f(a)| < \epsilon$. Se $f\colon X \to \mathbb{R}$ é contínua em todos os pontos do conjunto X, diz-se simplesmente que f é uma função *contínua*.

Dado $X \subset \mathbb{R}^n$, uma função $f\colon X \to \mathbb{R}$ diz-se *Lipschitziana* quando existe $L > 0$ (constante de Lipschitz de f) tal que, para quaisquer $x, y \in X$, tem-se

$$|f(x) - f(y)| \leq L\|x - y\|.$$

Toda função Lipschitziana é contínua: dado $\epsilon > 0$, basta tomar $\delta = \dfrac{\epsilon}{L}$. Então $\|x - a\| < \delta$ implica $|f(x) - f(a)| \leq L\|x - a\| < L.\dfrac{\epsilon}{L} = \epsilon$.

Definição II.4.1. Diz-se que uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ é *localmente lipschitziana* quando, para todo $x_0 \in \mathbb{R}^n$, existem $L_{x_0} > 0$ e $\delta > 0$, tais que $|f(x) - f(y)| \leq L_{x_0}\|x - y\|$, para todo x, $y \in B(x_0; \delta)$.

Teorema II.4.2. *Seja $f\colon X \to \mathbb{R}^n$ uma função definida no subconjunto $X \subset \mathbb{R}^n$. f é contínua no ponto $a \in X$ se, e somente se, para toda seqüência de pontos $x_k \in X$ com $\lim_{k\to+\infty} x_k = a$, tem-se $\lim_{k\to+\infty} f(x_k) = f(a)$.*

Teorema II.4.3. *(Weierstrass) Seja* $f\colon K \to \mathbb{R}$ *uma função contínua, definida num conjunto compacto* $K \subset \mathbb{R}^n$*. Então* f *atinge seu máximo e seu mínimo em* K*; isto é, existem pontos* $x_0, x_1 \in K$ *tais que* $f(x_0) \leq f(x) \leq f(x_1)$ *para qualquer* $x \in K$.

Teorema II.4.4. *(Teorema do Valor Médio) Seja* $f\colon \mathbb{R}^n \to \mathbb{R}$ *diferenciável. Dados* x, $y \in \mathbb{R}^n$*, existe um ponto* ξ *no segmento ligando* x *a* y *tal que*

$$f(y) - f(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\xi).(y - x)_i, \tag{II.4.1}$$

onde $(y - x)_i$ *é a i-ésima componente do vetor* $(y - x) \in \mathbb{R}^n$.

Podemos escrever a equação (II.4.1) na forma

$$f(y) - f(x) = \langle \nabla f(\xi), y - x \rangle.$$

CAPÍTULO III

Análise convexa: Um resumo dos conceitos básicos

III.1 Introdução

Destinamos este capítulo a um estudo dos conceitos relacionados aos conjuntos convexos e às funções convexas em $\mathbb{R}^n$, os quais utilizaremos no restante desta dissertação. Para uniformizar as notações e tornar este trabalho "autocontido", demostraremos a grande maioria dos resultados. As definições e os teoremas enunciados no capítulo anterior, juntamente com os resultados apresentados aqui, são o instrumental necessário para o desenvolvimento do algoritmo de ponto proximal que será introduzido no Capítulo IV. Os resultados de análise convexa em $\mathbb{R}^n$ utilizados nesta seção encontram-se em [3], [10] e [12].

III.2 Conjuntos convexos

Definição III.2.1. Um conjunto $C \subset \mathbb{R}^n$ é dito *convexo* se $\alpha x + (1-\alpha)x'$ pertence a C, sempre que x e x' pertencem a C, e $\alpha \in [0,1]$.

Geometricamente, esta definição nos diz que o segmento de reta

$$[x, x'] := \{\alpha x + (1-\alpha)x' : 0 \leq \alpha \leq 1\}$$

está inteiramente contido em C, sempre que os pontos extremos x e x' estão em C.

Definição III.2.2. Dada $f\colon \mathbb{R}^n \to \mathbb{R}$, o *epigráfico* de f é o conjunto não vazio

$$\operatorname{epi} f := \{(x,r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}.$$

O epigráfico de uma função linear é caracterizado por um vetor $s \in \mathbb{R}^n$, e tem a forma $\{(x,r) \in \mathbb{R}^n \times \mathbb{R} : \langle s, x\rangle - r \leq 0\}$. Assim, o epigráfico de funções afins, são convenientemente escritos em termos de algum $x_0 \in \mathbb{R}^n$:

$$\{(x,r) : r \geq f(x_0) + \langle s, x - x_0\rangle\} = \{(x,r) : \langle s, x\rangle - r \leq \langle s, x_0\rangle - f(x_0)\}.$$

As funções afins possuem papel importante no contexto da análise convexa estudada aqui. Mostraremos no Teorema III.3.4 que se f é uma função convexa então f é minorizada por alguma função afim. A definição de hiperplano suporte e um breve estudo sobre projeção auxiliarão na demonstração do teorema citado acima.

Definição III.2.3. (Hiperplano Suporte) Dado um conjunto $C \subset \mathbb{R}^n$ com fronteira $\partial C \neq \emptyset$ e um ponto $x \in \partial C$, considere o hiperplano afim $H_{s,r} = \{y \in \mathbb{R}^n : \langle s, y\rangle = r\}$ para $s \in \mathbb{R}^n$ e $r \in \mathbb{R}$. Dizemos que $H_{s,r}$ é *suporte* ao conjunto C em x se vale a desigualdade:

$$\langle s, x\rangle \leq r,$$

para todo $x \in C$ e, além disso, $x \in H_{s,r}$.

Seja $C \neq \emptyset$ um conjunto convexo fechado em $\mathbb{R}^n$. Para $x \in \mathbb{R}^n$ fixado, consideremos o seguinte problema:

$$\inf\{\|y - x\|^2 : y \in C\}. \tag{III.2.1}$$

Dado $c \in C$, tome um conjunto de sub-nível $S := \{y \in \mathbb{R}^n : \|y - x\| \leq \|c - x\|\}$. Então (III.2.1) é equivalente a

$$\inf\{\|y - x\| : y \in C \cap S\},$$

o qual possui uma solução, pois $y \mapsto \|y - x\|$ é contínua, S é compacto e $C \cap S$ é compacto. Portanto, deduzimos a *existência* de um ponto em C que minimiza a distância a x e desta forma (III.2.1) é de fato um mínimo. Veremos que existe um único ponto que minimiza a distância de x ao convexo C.

Teorema III.2.4. *Seja $x \in \mathbb{R}^n$. Se, para algum $y_x \in C$ tem-se $\|y_x - x\| \leq \|y - x\|$, para todo $y \in C$, então*

$$\langle x - y_x, y - y_x \rangle \leq 0, \tag{III.2.2}$$

para todo $y \in C$.

Demonstração. Tome y arbitrário em C, de modo que $y_x + \alpha(y - y_x) \in C$, para todo $\alpha \in]0, 1[$. Então podemos escrever

$$\begin{aligned}
\|y_x - x\|^2 &\leq \|y_x + \alpha(y - y_x) - x\|^2 \\
&= \|(y_x - x) + \alpha(y - y_x)\|^2 \\
&= \|y_x - x\|^2 + 2\alpha\langle y_x - x, y - y_x\rangle + \alpha^2\|y - y_x\|^2.
\end{aligned}$$

Esta desigualdade implica que

$$0 \leq \alpha\langle y_x - x, y - y_x\rangle + \frac{1}{2}\alpha^2\|y - y_x\|^2.$$

Dividindo por $\alpha > 0$ e fazendo $\alpha \downarrow 0$ obtemos

$$\langle x - y_x, y - y_x \rangle \leq 0. \quad \square$$

Teorema III.2.5. *Existe um único ponto $y_x \in C$ tal que $\|y_x - x\| \leq \|y - x\|$, para todo $y \in C$.*

Demonstração. Suponha que $y_x \in C$ satisfaça $\langle x - y_x, y - y_x \rangle \leq 0$, para todo $y \in C$. Se $x \in C$, então o resultado vale. Se não, escreva para um arbitrário $y \in C$

$$\begin{aligned}
0 &\geq \langle x - y_x, y - y_x \rangle \\
&= \langle x - y_x, y - x + x - y_x \rangle \\
&= \|x - y_x\|^2 + \langle x - y_x, y - x \rangle \\
&\geq \|y_x - x\|^2 - \|y_x - x\| \|y - x\|,
\end{aligned}$$

onde a desigualdade de *Cauchy-Schwarz* foi utilizada. Dividindo por $\|y_x - x\| > 0$, obtemos $\|y_x - x\| \leq \|y - x\|$. Agora, suponhamos que existam $y_x \in C$ e $y'_x \in C$ tais que

$$\|y_x - x\| \leq \|y - x\| \quad \text{e} \quad \|y'_x - x\| \leq \|y - x\|,$$

para todo $y \in C$. Em particular, valem as desigualdades

$$\langle x - y_x, y'_x - y_x \rangle \leq 0 \quad \text{e} \quad \langle x - y'_x, y_x - y'_x \rangle \leq 0,$$

e somando-as obtemos

$$\langle y'_x - y_x, y'_x - y_x \rangle \leq 0.$$

Logo, $y'_x = y_x$. Portanto tal ponto y_x é único. $\square$

Denotemos por $\mathrm{p}_C(x)$ o único ponto dado pelo teorema anterior, que é chamado de *projeção* de x sobre o convexo C.

Teorema III.2.6. *Seja $C \subset \mathbb{R}^n$ um conjunto não vazio convexo fechado, e seja $x \notin C$. Então existe $s \in \mathbb{R}^n$ tal que*

$$\langle s, x \rangle > \sup\{\langle s, y \rangle \colon y \in C\}. \tag{III.2.3}$$

Demonstração. Seja $s := x - \mathrm{p}_C(x) \neq 0$. Escrevemos (III.2.2) como

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

Assim,

$$\langle s, x \rangle - \|s\|^2 \geq \langle s, y \rangle,$$

para todo $y \in C$. Segue da última desigualdade que $s = x - \mathrm{p}_C(x)$ satisfaz (III.2.3). $\square$

Lema III.2.7. *Seja $x \in \partial C$, onde $C \neq \emptyset$ é convexo em $\mathbb{R}^n$ (naturalmente $C \neq \mathbb{R}^n$). Então existe um hiperplano suporte a C em x.*

Demonstração. Como C, o *fecho* de C e seus complementos possuem a mesma fronteira, uma seqüência $\{x_k\}$ pode ser encontrada tal que x_k não pertença ao *fecho* de C, para $k = 1, 2, ...$ e $\lim_{k \to +\infty} x_k = x$. Para cada k temos, pelo Teorema III.2.6, algum s_k com $\|s_k\| = 1$ tal que $\langle s_k, x_k - y \rangle > 0$, para todo y pertencente a C. Extraindo uma subseqüência, se necessário, temos que $s_k \to s$ (observe que $s \neq 0$) e passando ao limite obtemos

$$\langle s, x - y \rangle \geq 0,$$

para todo $y \in C$. Tomando $r = \langle s, x \rangle$ o hiperplano $H_{s,r} = \{y \in \mathbb{R}^n; \langle s, y \rangle = r\}$ é suporte a C em x. $\square$

Na próxima seção, faremos um estudo detalhado dos resultados de funções convexas que utilizaremos adiante.

III.3 Funções convexas

Uma importante classe de funções, no contexto da análise real, é a classe das funções convexas. A noção de convexidade ocupa posição de destaque no estudo da teoria de otimização. A atratividade da convexidade para resolver problemas de minimização reside no fato de que a determinação de um mínimo local de uma função convexa implica na determinação de um mínimo global dessa função. Além disso, a estrita convexidade de uma função definida num conjunto compacto permitirá escrevermos de maneira única a iteração de ponto proximal que será estudada no capítulo seguinte.

Definição III.3.1. Seja $C \subset \mathbb{R}^n$ um conjunto convexo não vazio. Uma função $f\colon C \to \mathbb{R}$ é dita *convexa* sobre C quando, para todo par $(x, x') \in C \times C$ e todo $\alpha \in]0, 1[$, vale

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x'). \tag{III.3.1}$$

Dizemos que f é *estritamente convexa* sobre C quando (III.3.1) vale como uma desigualdade estrita se $x \neq x'$.

Exemplo III.3.2. A função $h(x) = \lambda_k \|x_k - x\|^2$, $\lambda_k > 0$, é estritamente convexa, para todo $k > 0$.

Proposição III.3.3. *Se $f\colon \mathbb{R}^n \to \mathbb{R}$ é uma função convexa e $h\colon \mathbb{R}^n \to \mathbb{R}$ é uma função estritamente convexa, então $f + h\colon \mathbb{R}^n \to \mathbb{R}$ é uma função estritamente convexa.*

Demonstração. Se x' e x'' pertencem a $\mathbb{R}^n$ e $0 \leq \alpha \leq 1$, vale:

$$f(\alpha x' + [1-\alpha]x'') \leq \alpha f(x') + [1-\alpha] f(x'') \quad \text{e}$$
$$h(\alpha x' + [1-\alpha]x'') < \alpha h(x') + [1-\alpha] h(x'').$$

Logo,

$$\begin{aligned}(f+h)(\alpha x'+[1-\alpha]x'') &= f(\alpha x'+[1-\alpha]x'')+h(\alpha x'+[1-\alpha]x'')\\ &\leq \alpha f(x')+[1-\alpha]f(x'')+h(\alpha x'+[1-\alpha]x'')\\ &< \alpha f(x')+[1-\alpha]f(x'')+\alpha h(x')+[1-\alpha]h(x'')\\ &= \alpha(f+h)(x')+[1-\alpha](f+h)(x'').\end{aligned}$$

Assim, $(f+h)$ é estritamente convexa em $\mathbb{R}^n$. □

É fácil verificar que o epigráfico de uma função convexa $f\colon \mathbb{R}^n \to \mathbb{R}$ é um conjunto convexo. Este fato, juntamente com os resultados obtidos na seção anterior, será importante na demonstração do próximo teorema: a existência de uma função afim minorando f.

Teorema III.3.4. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Para todo $x \in \mathbb{R}^n$, existe $s = s(x) \in \mathbb{R}^n$ tal que*

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

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

Demonstração. Sabemos que $\operatorname{epi} f = \{(x,r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}$ é um subconjunto convexo fechado de $\mathbb{R}^n \times \mathbb{R}$. Dado $x \in \mathbb{R}^n$ e $(x, f(x)) \in \partial(\operatorname{epi} f)$, podemos tomar, pelo Lema III.2.7, um hiperplano suporte ao $\operatorname{epi} f$ em $(x, f(x))$. Usando (III.2.3), existem $s \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$, ambos não nulos, tais que

$$\langle s, y\rangle + \alpha r \leq \langle s, x\rangle + \alpha f(x), \tag{III.3.2}$$

para todo $(y,r) \in \mathbb{R}^n \times \mathbb{R}$. Escolha $r \in \mathbb{R}$ e $\delta > 0$, tais que $f(x+\delta s) < r$. Assim, de (III.3.2), temos

$$\begin{aligned}&0 \geq \langle s, \delta s\rangle + \alpha(f(x+\delta s) - f(x)),\\ &0 < \delta\|s\|^2 \leq \alpha(f(x) - f(x+\delta s)).\end{aligned}$$

Isto mostra que $\alpha \neq 0$ (senão s também seria igual a zero). Sem perda de generalidade, podemos assumir $\alpha = -1$ e, retornando à inequação (III.3.2), obtemos

$$\langle s, y\rangle - r \leq \langle s, x\rangle - f(x) \Leftrightarrow f(y) \geq f(x) + \langle s, y - x\rangle,$$

que dá a função afim procurada. $\square$

Como nosso objetivo é resolver um problema de minimização convexa, formalizaremos a definição de mínimo de uma função e demonstraremos os resultados que serão necessários na prova de convergência do algoritmo.

Definição III.3.5. Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função qualquer. Um ponto $x_* \in \mathbb{R}^n$ é:
(i) um *mínimo local* de f se existe $r > 0$ tal que: $\|x - x_*\| \leq r$ implica $f(x_*) \leq f(x)$, para todo $x \in B(x_*; r)$.
(ii) um *mínimo global* de f se $f(x_*) \leq f(x)$, para todo $x \in \mathbb{R}^n$.

Proposição III.3.6. *Sejam $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa e x_* um mínimo local de f em $\mathbb{R}^n$. Então x_* é mínimo global de f em $\mathbb{R}^n$.*

Demonstração. Suponhamos que f admita um mínimo local x_* que não é mínimo global. Sendo x_* um mínimo local, existe $r > 0$ tal que $f(x_*) \leq f(x)$, para todo $x \in B(x_*; r)$. Como x_* não é mínimo global, existe $z \in \mathbb{R}^n$ tal que $f(z) < f(x_*)$. Sendo $\mathbb{R}^n$ convexo, $(\lambda x_* + (1-\lambda)z) \in \mathbb{R}^n$, para todo $\lambda \in]0, 1[$. Escolha λ suficientemente próximo da unidade tal que $(\lambda x_* + (1-\lambda)z) \in B(x_*; r)$. Pela convexidade de f, temos

$$f(\lambda x_* + (1-\lambda)z) \leq \lambda f(x_*) + (1-\lambda) f(z) < f(x_*),$$

pois $f(z) < f(x_*)$. Mas, pela construção, $(\lambda x_* + (1-\lambda)z)$ está na bola $B(x_*; r)$. Logo $f(x_*) \leq f(\lambda x_* + (1-\lambda)z)$, uma contradição. Portanto, x_* é um mínimo global de f em $\mathbb{R}^n$. $\square$

A seguir definiremos coercividade de uma função contínua $f\colon \mathbb{R}^n \to \mathbb{R}$. Uma implicação particular desta definição é o Teorema III.3.8: se uma função contínua f possui a propriedade de coercividade então f possui pelo menos um mínimo global. Este teorema será importante para a boa definição do algoritmo de ponto proximal para funções convexas.

Definição III.3.7. Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função contínua. Diz-se que f é *1-coerciva* se

$$\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty\,.$$

A definição acima diz que f tende ao infinito mais rápido do que qualquer função afim.

Teorema III.3.8. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função contínua. Se f é 1-coerciva, então f possui pelo menos um mínimo global.*

Demonstração. Por hipótese $\displaystyle\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty$. Isto significa que se $\|x\|$ é grande, então $\dfrac{f(x)}{\|x\|}$ também o é. Conseqüentemente, dado $M \geq 0$ existe um número $r > 0$ tal que $\|x\| \geq r$ implica que

$$\frac{f(x)}{\|x\|} \geq M \;\Leftrightarrow\; f(x) \geq M\|x\| \geq Mr\,.$$

Seja $\overline{B}(0;r)$ o conjunto $\{x : \|x\| \leq r\}$. A função f é contínua em cada ponto do conjunto $\overline{B}(0;r)$, e este conjunto é fechado e limitado. Pelo Teorema de *Weierstrass*, segue que f atinge um valor mínimo sobre $\overline{B}(0;r)$ num ponto $x_* \in \overline{B}(0;r)$; ou seja, $x \in \overline{B}(0;r)$ implica $f(x_*) \leq f(x)$. Em particular, como $0 \in \overline{B}(0;r)$, vemos que $f(x_*) \leq f(0)$. Por outro lado, se $x \notin \overline{B}(0;r)$, então $f(x) > f(0) \geq f(x_*)$. Resumindo, vemos que $x \in \overline{B}(0;r)$ implica $f(x) \geq f(x_*)$ e $x \notin \overline{B}(0;r)$ implica $f(x) > f(x_*)$. Isto mostra que x_* é um mínimo global de f. $\square$

Teorema III.3.9. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Então, para toda coleção de pontos $\{x_1, x_2, ..., x_k\}$ em $\mathbb{R}^n$ e toda coleção $\{\alpha_1, \alpha_2, ..., \alpha_k\}$ em $\mathbb{R}^n$ satisfazendo*

$$\alpha_i \geq 0 \quad \text{para} \;\; i = 1, 2, ..., k \quad \text{e} \;\; \sum_{i=1}^{k} \alpha_i = 1,$$

vale a desigualdade de Jensen

$$f(\sum_{i=1}^{k} \alpha_i x_i) \leq \sum_{i=1}^{k} \alpha_i f(x_i).$$

Demonstração. Considere primeiro $k = 2$. A relação é trivial se α_1 ou α_2 são zero. Caso contrário, a convexidade da função f resulta em $f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$. Ou seja, $f(\sum_{i=1}^{2} \alpha_i x_i) \leq \sum_{i=1}^{2} \alpha_i f(x_i)$. Agora, suponha indutivamente que a relação é verdadeira para $k-1$ e considere as coleções $\{x_i\}$ e $\{\alpha_i\}$, com $\alpha_i \geq 0$ para $i = 1, 2, ..., k$ e $\sum_{i=1}^{k} \alpha_i = 1$. Se α_k é 0 ou 1, então nada há para provar. Se não, fixe $\overline{\alpha} := \sum_{i=1}^{k-1} \alpha_i$ e $\overline{\alpha}_i := \dfrac{\alpha_i}{\overline{\alpha}}$ para $i = 1, ..., k-1$. Observe que $\alpha_k = 1 - \overline{\alpha} \in]0, 1[$, $\overline{\alpha}_i \geq 0$ e $\sum_{i=1}^{k-1} \overline{\alpha}_i = 1$. Assim,

$$\sum_{i=1}^{k} \alpha_i x_i = \overline{\alpha} \sum_{i=1}^{k-1} \overline{\alpha}_i x_i + (1 - \overline{\alpha}) x_k.$$

O ponto $\overline{x} := \sum_{i=1}^{k-1} \overline{\alpha}_i x_i$ da útima relação acima pertence a $\mathbb{R}^n$. Usando a convexidade de f e aplicando a hipótese de indução para $\overline{x}$, temos $f(\overline{\alpha x} + (1 - \overline{\alpha}) x_k) \leq \overline{\alpha} f(\overline{x}) + (1 - \overline{\alpha}) f(x_k)$. Ou seja,

$$\begin{aligned} f(\sum_{i=1}^{k} \alpha_i x_i) &\leq \overline{\alpha} f(\sum_{i=1}^{k-1} \overline{\alpha}_i x_i) + \alpha_k f(x_k) = f(\sum_{i=1}^{k-1} \alpha_i x_i) + \alpha_k f(x_k) \\ &\leq \sum_{i=1}^{k-1} \alpha_i f(x_i) + \alpha_k f(x_k) = \sum_{i=1}^{k} \alpha_i f(x_i). \qquad \square \end{aligned}$$

Definição III.3.10. Diz-se que uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ é *localmente limitada* quando para todo $x_0 \in \mathbb{R}^n$ existem $c > 0$ e $\delta > 0$, tais que $|f(x)| \leq c$, para todo $x \in B(x_0;\delta)$.

Teorema III.3.11. *Se $f\colon \mathbb{R}^n \to \mathbb{R}$ é convexa, então f é localmente limitada.*

Demonstração. Dado $x_0 \in \mathbb{R}^n$, tome um "cubo" C em $\mathbb{R}^n$, de vértices $x_1, x_2, ..., x_m$ $(m = 2^n)$ e com centro em x_0. Como o "cubo" C é um conjunto convexo em $\mathbb{R}^n$, para todo ponto $x \in C$, podemos encontrar escalares $\lambda_1, \lambda_2, ..., \lambda_m$, tais que

$$x = \sum_{i=1}^{m} \lambda_i x_i\,, \quad \text{com} \;\; \lambda_i \geq 0 \quad \text{e} \quad \sum_{i=1}^{m} \lambda_i = 1.$$

Sendo f uma função convexa vale a desigualdade de *Jensen*:

$$f(x) \leq \sum_{i=1}^{m} \lambda_i f(x_i) \leq \max_{1 \leq i \leq n} f(x_i) := \tilde{c}.$$

Assim, f é limitada superiormente. Por outro lado, para qualquer $x \in C$, podemos escolher $y \in C$, tal que $x_0 = \frac{1}{2}x + \frac{1}{2}y$. Então $f(x_0) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y)$. Mas isso implica

$$f(x) \geq 2\left[f(x_0) - \frac{1}{2}f(y)\right] \geq 2f(x_0) - \tilde{c}.$$

Logo f é também limitada inferiormente em C. Portanto, f é localmente limitada. $\square$

Teorema III.3.12. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa e suponha existirem $x_0, \delta > 0, m$ e M, tais que $m \leq f(x) \leq M$, para todo $x \in B(x_0;2\delta)$. Então f é lipschitziana em $B(x_0;\delta)$.*

Demonstração. Para os pontos y e y' pertencentes à $B(x_0;\delta)$, com $y \neq y'$, tome

$$y'' := y' + \delta \frac{y' - y}{\|y' - y\|} \quad \in \quad B(x_0;2\delta).$$

Observe que y' pertence ao segmento $[y, y'']$, isto é, $y' = \dfrac{\|y' - y\|}{\delta + \|y' - y\|}\, y'' \;+\; \dfrac{\delta}{\delta + \|y' - y\|}\, y.$ Aplicando a convexidade de f, obtemos

$$f(y') \leq \frac{\|y' - y\|}{\delta + \|y' - y\|}\, f(y'') \;+\; \frac{\delta}{\delta + \|y' - y\|}\, f(y).$$

Subtraindo $f(y)$ de ambos os lados da desigualdade, resulta

$$f(y') - f(y) \leq \frac{\|y' - y\|}{\delta + \|y' - y\|} f(y'') + (\frac{\delta}{\delta + \|y' - y\|} - 1) f(y) = \frac{\|y' - y\|}{\delta + \|y' - y\|}[f(y'') - f(y)].$$

Como $m \leq f(y'') \leq M$ e $m \leq f(y') \leq M$, temos que $f(y'') \leq M$ e $-f(y') \leq -m$. Daí

$$\frac{\|y' - y\|}{\delta + \|y' - y\|}[f(y'') - f(y)] \leq \frac{\|y' - y\|}{\delta}(M - m).$$

Assim,

$$f(y') - f(y) \leq \frac{\|y' - y\|}{\delta}(M - m).$$

Como y e y' são quaisquer obtemos a desigualdade análoga trocando y por y'. Portanto,

$$|f(y') - f(y)| \leq \frac{M - m}{\delta}\|y' - y\|. \qquad \square$$

Corolário III.3.13. *Se $f\colon \mathbb{R}^n \to \mathbb{R}$ é uma função convexa, então f é localmente lipschitziana.*

Demonstração. Aplicação imediata dos teoremas III.3.11 e III.3.12. $\square$

Teorema III.3.14. *Se $f\colon \mathbb{R}^n \to \mathbb{R}$ é convexa, então f é contínua.*

Demonstração. Dado $x_0 \in \mathbb{R}^n$, existem, pelo Corolário III.3.13, $\delta_{x_0} > 0$ e $L_{x_0} > 0$, tais que para todo $x, y \in B(x_0; \delta_{x_0})$ implica $|f(x) - f(y)| \leq L_{x_0}\|x - y\|$. Dado $\epsilon > 0$, tome $\delta = \min\{\frac{\epsilon}{L_{x_0}}, \delta_{x_0}\}$. Logo, para todo $x \in B(x_0; \delta) \subset B(x_0; \delta_{x_0})$, temos

$$|f(x) - f(x_0)| \leq L_{x_0}\|x - x_0\| \leq L_{x_0}.\frac{\epsilon}{L_{x_0}} = \epsilon.$$

Portanto, f é contínua. $\square$

Neste trabalho a função objetivo $f\colon \mathbb{R}^n \to \mathbb{R}$ não é necessariamente diferenciável. Mas, sendo uma função convexa, f é *"subdiferenciável"*. Dado $x \in \mathbb{R}^n$, o conjunto de todos os vetores $s \in \mathbb{R}^n$ satisfazendo (III.3.4) ou (III.3.5) é chamado o *subdiferencial* de f em x e denotado por $\partial f(x)$. As definições a seguir conduzem a duas interpretações do subdiferencial.

Definição III.3.15. Seja S um conjunto não vazio em $\mathbb{R}^n$. A função

$$\begin{aligned} \sigma_S \colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \sigma_S(x) := \sup\{\langle s, x \rangle : s \in S\} \end{aligned}$$

é chamada *função suporte* de S.

Definição III.3.16. A *derivada direcional* de f em x na direção d é

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

Quando f é uma função convexa a igualdade acima pode ser substituída por

$$f'(x,d) = \inf_{t>0} \frac{f(x+td) - f(x)}{t}. \tag{III.3.3}$$

Definição III.3.17. (Subdiferencial I) O *subdiferencial* $\partial f(x)$ de f em x é o conjunto cuja função suporte é $f'(x,.)$, isto é,

$$\partial f(x) := \{s \in \mathbb{R}^n : \langle s, d \rangle \leq f'(x,d) \text{ para todo } d \in \mathbb{R}^n\}. \tag{III.3.4}$$

Nesta definição de subdiferencial está envolvido o cálculo da derivada direcional e determinação de uma função suporte. É possível dar uma definição mais direta:

Definição III.3.18. (Subdiferencial II) O *subdiferencial* de f em x é o conjunto dos vetores $s \in \mathbb{R}^n$ satisfazendo

$$f(y) \geq f(x) + \langle s, y - x \rangle, \tag{III.3.5}$$

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

Um vetor $s \in \mathbb{R}^n$ satisfazendo (III.3.4) ou (III.3.5) é dito um *subgradiente* de f no ponto x. A multifunção $\partial f \colon x \mapsto \partial f(x)$ é chamada o *subdiferencial* de f.

Teorema III.3.19. *As definições III.3.17 e III.3.18 são equivalentes.*

Demonstração. Seja s satisfazendo (III.3.4), isto é

$$\langle s, d \rangle \leq f'(x, d), \tag{III.3.6}$$

para todo $d \in \mathbb{R}^n$. A equação (III.3.3) é equivalente a

$$\langle s, d \rangle \leq \frac{f(x + td) - f(x)}{t}, \tag{III.3.7}$$

para todo $d \in \mathbb{R}^n$ e $t > 0$. Quando d percorre $\mathbb{R}^n$ e t percorre $\mathbb{R}^*_+$, $y := x + td$ percorre $\mathbb{R}^n$ e então vemos que (III.3.7) é exatamente (III.3.5). $\square$

Proposição III.3.20. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. O conjunto $\partial f(x)$ é não vazio, convexo e compacto, para todo $x \in \mathbb{R}^n$.*

Demonstração. Seja $x \in \mathbb{R}^n$. Juntando os resultados do Teorema III.3.4 e da Definição III.3.18 segue imediatamente que $\partial f(x) \neq \emptyset$. Agora, sejam $s, w \in \partial f(x)$ e $t \in [0, 1]$. Mostraremos que $(1 - t)s + tw \in \partial f(x)$. Novamente, pela Definição III.3.18, valem as desigualdades

$$f(x) + \langle s, y - x \rangle \leq f(y) \quad \text{e} \quad f(x) + \langle w, y - x \rangle \leq f(y),$$

para todo $y \in \mathbb{R}^n$. Logo,

$$\begin{aligned}
f(x) + \langle (1 - t)s + tw, y - x \rangle &= f(x) + (1 - t)\langle s, y - x \rangle + t\langle w, y - x \rangle \\
&= f(x) - tf(x) + tf(x) + (1 - t)\langle s, y - x \rangle + t\langle w, y - x \rangle \\
&= (1 - t)[f(x) + \langle s, y - x \rangle] + t[f(x) + \langle w, y - x \rangle] \\
&\leq (1 - t)f(y) + tf(y) \\
&= f(y).
\end{aligned}$$

Portanto, $\partial f(x)$ é convexo. Para a última afirmação do teorema mostraremos que $\partial f(x)$ é fechado e limitado. Seja $\{s_k\}$ uma seqüência em $\partial f(x)$ convergindo para $s \in \mathbb{R}^n$. Para todo $k \in \mathbb{N}$, temos

$$f(y) \geq f(x_k) + \langle s_k, y - x_k \rangle,$$

para todo $y \in \mathbb{R}^n$. Passando ao limite quando k tende a $+\infty$ e usando a continuidade de f e do produto escalar, obtemos

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

para todo $y \in \mathbb{R}^n$. Logo $s \in \partial f(x)$, o que significa que $\partial f(x)$ é fechado. Para finalizar, dado $s \in \partial f(x)$, $s \neq 0$, tomemos $\delta > 0$ arbitrariamente pequeno de modo que $y = x + \delta \dfrac{s}{\|s\|}$ pertença ao conjunto compacto $\overline{B}(x;\delta)$. Sendo f convexa, para todo $y \in \overline{B}(x;\delta)$, existe uma constante de *Lipschitz* $L > 0$ tal que $f(y) - f(x) \leq L\|y - x\|$. Ou seja,

$$f(y) - f(x) \leq L.\delta, \tag{III.3.8}$$

para todo $y \in \overline{B}(x;\delta)$. Por outro lado, pela convexidade de f, $f(y) \geq f(x) + \langle s, y - x \rangle$, para todo $y \in \mathbb{R}^n$. Substituindo pelo y dado acima obtemos $f(y) \geq f(x) + \langle s, \delta \dfrac{s}{\|s\|} \rangle$, o que implica

$$f(y) \geq f(x) + \delta\|s\|. \tag{III.3.9}$$

Juntando (III.3.8) e (III.3.9) obtemos $\delta\|s\| \leq f(y) - f(x) \leq L.\delta$ o que resulta $\|s\| \leq L$. Assim, $\partial f(x)$ é limitado. Portanto $\partial f(x)$ é compacto. □

Se uma função f é diferenciável em x, então o conjunto $\partial f(x)$ reduz-se ao gradiente $\nabla f(x)$. O próximo teorema será utilizado na demonstração da Proposição V.2.2 no Capítulo V.

Teorema III.3.21. *Seja* $f\colon \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável. Então* f *é convexa se, e somente se,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \tag{III.3.10}$$

para todo x *e todo* y *em* $\mathbb{R}^n$*.*

Demonstração. Seja f uma função convexa; para $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ e $\alpha \in]0, 1[$ arbitrário temos, da definição de convexidade:

$$f(\alpha y + (1 - \alpha)x) - f(x) \leq \alpha[f(y) - f(x)].$$

Observando que $\alpha y + (1 - \alpha)x = x + \alpha(y - x)$, dividindo por α e fazendo $\alpha \downarrow 0$, o lado direito da igualdade acima tende a $\langle \nabla f(x), y - x \rangle$. Logo (III.3.10) está verificada. Reciprocamente, tome x_1 e x_2 em $\mathbb{R}^n$, $\alpha \in]0, 1[$ e defina $x := \alpha x_1 + (1 - \alpha)x_2 \in \mathbb{R}^n$. Por hipótese, $f(x_i) \geq f(x) + \langle \nabla f(x), x_i - x \rangle$, para $i = 1, 2$. Usando combinação convexa, obtemos

$$\alpha f(x_1) + (1 - \alpha)f(x_2) \geq f(x) + \langle \nabla f(x), \alpha x_1 + (1 - \alpha)x_2 - x \rangle,$$

o qual, após simplificação, resulta em $\alpha f(x_1) + (1 - \alpha)f(x_2) \geq f(\alpha x_1 + (1 - \alpha)x_2)$. □

O próximo resultado é elementar. Combinando-o com o Teorema III.3.26 poderemos, no capítulo seguinte, caracterizar a seqüência gerada pelo algoritmo de ponto proximal.

Teorema III.3.22. *Seja* $f\colon \mathbb{R}^n \to \mathbb{R}$ *uma função convexa. O ponto* $x_* \in \mathbb{R}^n$ *é um mínimo de* f *, isto é,* $f(x_*) \leq f(x)$ *para todo* $x \in \mathbb{R}^n$ *se, e somente se* $0 \in \partial f(x_*)$*.*

Demonstração. A equivalência segue imediatamente da inequação (III.3.5), pois:

$$f(x) \geq f(x_*) \Leftrightarrow f(x) \geq f(x_*) + \langle 0, x - x_* \rangle$$

para todo $x \in \mathbb{R}^n$. □

Proposição III.3.23. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. O conjunto dos minimizadores de f é convexo.*

Demonstração. Seja U^* o conjunto dos minimizadores da função convexa f. Dados $x_1, x_2 \in U^*$, temos $f(x_1) = f(x_2)$. Para $\alpha \in [0,1]$, queremos mostrar que $x = \alpha x_1 + (1-\alpha)x_2$ também é um minimizador de f. Pela convexidade de f temos

$$f(x) = f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) = f(x_1).$$

Mas isto deve valer como igualdade, ou x_1 e x_2 não seriam minimizadores de f. Logo, $x \in U^*$. Portanto, U^* é um conjunto convexo. □

Proposição III.3.24. *Se $f\colon K \to \mathbb{R}$ é estritamente convexa sobre o conjunto compacto $K \subset \mathbb{R}^n$, então f possui um único minimizador.*

Demonstração. Como f está definida sobre um conjunto compacto $K \subset \mathbb{R}^n$, pelo Teorema de *Weierstrass*, f atinge seu mínimo. Suponhamos, por absurdo, que $x_* \in K$, $x_{**} \in K$ e $x_* \neq x_{**}$ sejam minimizadores da função estritamente convexa f. Sejam $s \in \partial f(x_{**})$ e $s' \in \partial f(x_*)$, então:

$$f(x_*) > f(x_{**}) + \langle s, x_* - x_{**} \rangle \ \text{ e}$$
$$f(x_{**}) > f(x_*) + \langle s', x_{**} - x_* \rangle.$$

Mas, pelo Teorema III.3.22, para $s = s' = 0$, obtemos

$$f(x_*) > f(x_{**}) \ \text{ e } \ f(x_{**}) > f(x_*).$$

Portanto, f possui um único minimizador. □

Tomando $f\colon \mathbb{R}^n \to \mathbb{R}$ estritamente convexa, se o conjunto U^* dos minimizadores de f é não vazio, então f possui um único minimizador.

Proposição III.3.25. *Sejam f e g funções convexas de $\mathbb{R}^n$ em $\mathbb{R}$ e $\sigma_{\partial f}$ e $\sigma_{\partial g}$ as funções suportes dos conjuntos convexos fechados ∂f e ∂g. Então $\sigma_{\partial f} + \sigma_{\partial g}$ é a função suporte de $\partial f + \partial g$.*

Demonstração. Chamemos ∂h o conjunto fechado $\partial f + \partial g$. Pela Definição III.3.15 de função suporte:

$$\sigma_{\partial h}(d) = \sup\{\langle s_1 + s_2, d\rangle : s_1 \in \partial f \quad \text{e} \quad s_2 \in \partial g\}.$$

Na expressão acima, s_1 e s_2 percorrem de forma independente os conjuntos ∂f e ∂g, respectivamente. Assim,

$$\sigma_{\partial h}(d) = \sup_{s \in \partial f} \langle s, d\rangle + \sup_{s \in \partial g} \langle s, d\rangle. \qquad \square$$

Teorema III.3.26. *Sejam f e g funções convexas de $\mathbb{R}^n$ em $\mathbb{R}$. Então $\partial(f+g)(x) = \partial f(x) + \partial g(x)$, para todo $x \in \mathbb{R}^n$.*

Demonstração. Pela Proposição III.3.25 temos que $\partial f(x) + \partial g(x)$ é um conjunto convexo compacto cuja função suporte é

$$f'(x,.) + g'(x,.).$$

Por outro lado, a função suporte de $\partial(f+g)(x)$ é, pela definição de derivada direcional,

$$(f+g)'(x,.)$$

que, de um cálculo elementar, é justamente $f'(x,.) + g'(x,.)$. Portanto, os dois conjuntos convexos e compactos $\partial(f+g)(x)$ e $\partial f(x) + \partial g(x)$ coincidem, pois têm a mesma função suporte. $\square$

No Capítulo VI desenvolveremos o algoritmo de ponto proximal para operadores monótonos maximais arbitrários, $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$. Um problema fundamental é determinar um elemento $z \in \mathbb{R}^n$ tal que $0 \in T(z)$. Agora, se T é o subdiferencial ∂f de uma função convexa $f\colon \mathbb{R}^n \to \mathbb{R}$, então T é monótono maximal, e a relação $0 \in T(z)$ quer dizer que z é um minimizador de f, como já vimos no Teorema III.3.22. Neste sentido, o problema de encontrar uma singularidade de um operador monótono maximal generaliza o problema de minimizar uma função convexa. Para demonstramos que o operador ∂f é monótono maximal, faz-se necessário algumas definições, e começaremos por definir convexidade *própria*.

Definição III.3.27. A função $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, não identicamente $+\infty$, é dita *convexa* quando, para todo $(x, x') \in \mathbb{R}^n \times \mathbb{R}^n$ e todo $\alpha \in]0,1[$, vale

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x'),$$

considerada como uma desigualdade em $\mathbb{R} \cup \{+\infty\}$.

Para compreender a definição acima lembremo-nos da Definição III.3.1: uma função $f\colon C \to \mathbb{R}$, é dita *convexa* sobre o conjunto convexo $\emptyset \neq C \subset \mathbb{R}^n$ quando, para todo par $(x, x') \in C \times C$ e todo $\alpha \in]0,1[$, vale

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x').$$

Então estendemos a função f, fazendo

$$f(x) := +\infty,$$

para $x \notin C$. Assim, obtemos uma nova função f, que chamaremos de função convexa *própria*.

Definição III.3.28. Seja $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa, $f \not\equiv +\infty$. A função $f^*\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ definida por:

$$f^*(s) = \sup_{x \in \mathbb{R}^n} \{\langle x, s\rangle - f(x)\},$$

para todo $s \in \mathbb{R}^n$, é dita a *conjugada* da função f.

Exemplo III.3.29. Dada a função convexa $f(x) = \|x\|$, encontremos sua função conjugada. Calculando o subdiferencial de f obtemos: $\partial f(x) = \{1\}$, se $x > 0$, $\partial f(x) = [-1, 1]$, se $x = 0$ e $\partial f(x) = \{-1\}$, se $x < 0$. Pela definição de conjugada, para todo $s \in \partial f(x)$, $f^*(s)$ é o $\sup\{\langle s, x\rangle - f(x)\}$, $x \in \mathbb{R}^n$. Logo, $f^*(s) = \sup\limits_{x \in \mathbb{R}^n}\{(s-1)x\}$, se $x \geq 0$ e $f^*(s) = \sup\limits_{x\in\mathbb{R}^n}\{(s+1)x\}$, se $x < 0$. Portanto,

$$f^*(s) = \begin{cases} +\infty, & \text{se } \|s\| > 1 \\ 0, & \text{se } \|s\| \leq 1 \end{cases}.$$

Observação III.3.30. Seja $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Para cada $x \in \mathbb{R}^n$ e cada $s \in \mathbb{R}^n$, vale, pela Definição III.3.28, que $f^*(s) \geq \langle x, s\rangle - f(x)$; isto é,

$$f(x) + f^*(s) \geq \langle x, s\rangle,$$

sempre que o lado esquerdo da desigualdade estiver definido. Esta é chamada a *desigualdade de Fenchel*.

Teorema III.3.31. *Sejam $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa, $f \not\equiv +\infty$, e $x \in \mathbb{R}^n$. Então $s \in \partial f(x)$ se, e somente se, $f^*(s) = \langle x, s\rangle - f(x)$.*

Demonstração. Seja s um subgradiente de f em x. Então, $f(y) \geq f(x) + \langle s, y - x\rangle$, para todo $y \in \mathbb{R}^n$. Mas isso equivale a $\langle y, s\rangle - f(y) \leq \langle x, s\rangle - f(x)$, para todo $y \in \mathbb{R}^n$. Como $f^*(s) = \sup\limits_{y\in\mathbb{R}^n}\{\langle y, s\rangle - f(y)\} \leq \langle x, s\rangle - f(x)$, obtemos que $f^*(s) = \langle x, s\rangle - f(x)$. Reciprocamente, se $f^*(s) = \sup\limits_{y\in\mathbb{R}^n}\{\langle y, s\rangle - f(y)\} = \langle x, s\rangle - f(x)$, então $f^*(s) < +\infty$. Ou seja $\langle x, s\rangle - f(x) \geq \langle y, s\rangle - f(y)$, para todo $y \in \mathbb{R}^n$. Logo, $f(y) \geq f(x) + \langle y - x, s\rangle$, para todo $y \in \mathbb{R}^n$. Portanto, $s \in \partial f(x)$. $\square$

Observe que este teorema revela que se $f\colon \mathbb{R}^n \to \mathbb{R}$ é uma função convexa e $s \in \partial f(x)$, para algum $x \in \mathbb{R}^n$, então $f^*(s) < +\infty$.

Teorema III.3.32. *Seja* $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa,* $f \not\equiv +\infty$*. A conjugada* $f^*\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é uma função convexa.*

Demonstração. Dados s e $s' \in \mathbb{R}^n$, consideremos dois casos. Primeiro, s, $s' \in \mathbb{R}^n$ tais que $f^*(s) < +\infty$ e $f^*(s') < +\infty$. Para $t \in [0,1]$,

$$\begin{aligned} f^*[(1-t)s + ts'] &= \sup_{x\in\mathbb{R}^n}\{(1-t)\langle s, x\rangle + t\langle s', x\rangle - f(x)\} \\ &= \sup\{(1-t)\langle s, x\rangle - (1-t)f(x)\} + \sup\{t\langle s', x\rangle - tf(x)\} \\ &\leq (1-t)f^*(s) + tf^*(s'). \end{aligned}$$

Logo, neste caso, f^* é convexa. Agora consideremos s, $s' \in \mathbb{R}^n$ tais que $f^*(s) = +\infty$ ou $f^*(s') = +\infty$. Trivialmente temos: $f^*[(1-t)s + ts'] \leq (1-t)f^*(s) + tf^*(s')$. Portanto, f^* é uma função convexa. □

Definição III.3.33. Seja $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa. Dizemos que a função $f^{**}\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é a função conjugada $(f^*)^*$ da conjugada de f.

Lema III.3.34. *Seja* $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa. Então* f^{**} *é o supremo do conjunto de todas as funções afins que minoram* f*.*

Demonstração. Seja A o conjunto de todas as funções afins que minoram f e considere $\mathcal{F} = \sup\{g : g \in A\}$. Para cada $s \in \mathbb{R}^n$, a função

$$g(x) = \langle x, s\rangle - f^*(s),$$

é uma função afim. Da Observação III.3.30 segue que g é uma minorante de f. Logo $g \in A$. Daí

$$f^{**}(x) = \sup_{s\in\mathbb{R}^n}\{\langle x, s\rangle - f^*(s)\} \leq \mathcal{F}(x), \tag{III.3.11}$$

para todo $x \in \mathbb{R}^n$. Se $h \in A$, então h é da forma $h(x) = \langle x, s \rangle - \alpha$, onde $s \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$, o que implica que, para todo $x \in \mathbb{R}^n$, $\langle x, s \rangle - \alpha \leq f(x)$. Conseqüentemente

$$\alpha \geq \sup_{x \in \mathbb{R}^n} \{\langle x, s \rangle - f(x)\} = f^*(s),$$

e assim

$$h(x) = \langle x, s \rangle - \alpha \leq \langle x, s \rangle - f^*(s).$$

Disto segue que, para cada $x \in \mathbb{R}^n$,

$$\mathcal{F}(x) \leq \sup_{s \in \mathbb{R}^n} \{\langle x, s \rangle - f^*(s)\} = f^{**}(x), \tag{III.3.12}$$

donde, de (III.3.11) e (III.3.12), concluímos que $f^{**} = \mathcal{F}$. □

Proposição III.3.35. *Se $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é uma função convexa, então $f = f^{**}$.*

Demonstração. Primeiramente, observemos que $f^{**} \leq f$. De fato, por definição

$$f^{**}(w) := \sup\{\langle w, y \rangle - f^*(y) : y \in \mathbb{R}^n\}.$$

Mas $f^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle y, x \rangle - f(x)\} \geq \langle y, w \rangle - f(w)$, para todo $w \in \mathbb{R}^n$, o que implica

$$\langle w, y \rangle - f^*(y) \leq \langle w, y \rangle - (\langle y, w \rangle - f(w)) = f(w).$$

Portanto, $f^{**}(w) \leq f(w)$, para todo $w \in \mathbb{R}^n$. Agora, suponhamos por absurdo, que exista $x_0 \in \mathbb{R}^n$ tal que $f^{**}(x_0) < f(x_0)$. Isto implica a existência de $\alpha \in \mathbb{R}$ tal que $f^{**}(x_0) < \alpha < f(x_0)$. Sendo f convexa, existe uma função afim h minorando f definida por $h(x) = \alpha - \langle s, x - x_0 \rangle$; isto é, $h < f$. Pelo Lema III.3.34 temos $f^{**} > h$ e do fato de $h(x_0) = \alpha$ vem que $f^{**}(x_0) \geq h(x_0) = \alpha$, o que é uma contradição. Portanto $f^{**} = f$. □

CAPÍTULO IV

Algoritmo de ponto proximal para otimização em $\mathbb{R}^n$

IV.1 Introdução

O *algoritmo de ponto proximal* é um método para minimizar uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ convexa, não necessariamente diferenciável, que foi aplicado pela primeira vez em otimização convexa por Martinet [6], entre os anos de 1970 e 1972. A terminologia *ponto proximal* é devida à Rockafellar, conforme a referência [10].

O algoritmo de ponto proximal trata-se de um processo iterativo para minimizar uma função convexa que, a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, gera recursivamente uma seqüência de pontos $\{x_k\} \subset \mathbb{R}^n$ que converge para o minimizador da função objetivo. Uma suposição usual para estabelecer a convergência do algoritmo é a existência de um mínimo da função objetivo; isto é, se existe $x_* \in \mathbb{R}^n$ tal que $f(x_*) \leq f(x)$, para todo $x \in \mathbb{R}^n$, então a seqüência $\{x_k\}$ converge a este minimizador.

Um resultado importante obtido por Osman Güller [2] e apresentado de maneira simples e unificada por Rafael Correa e Claude Lemaréchal [1], foi a retirada da suposição de existência de um mínimo da função objetivo; ou seja, se a função não possui minimizador, então a seqüência $\{x_k\}$ diverge.

Denotando por U^* o conjunto dos minimizadores da função convexa f, nossa análise de convergência do algoritmo baseia-se na *Fejér convergência* da seqüência $\{x_k\}$ ao conjunto não vazio U^*. No entanto, se U^* é vazio, mostraremos que a seqüência $\{x_k\}$ diverge.

IV.2 Algoritmo de ponto proximal

Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. O *algoritmo de ponto proximal* gera, para um ponto de partida $x_0 \in \mathbb{R}^n$, a seqüência $\{x_k\} \subset \mathbb{R}^n$ pela iteração

$$x_{k+1} := \arg\min_{x\in\mathbb{R}^n}\{f(x) + \lambda_k\|x_k - x\|^2\}, \tag{IV.2.1}$$

onde $\{\lambda_k\}$ é uma seqüência de números positivos.

Definindo

$$f_k(x) := f(x) + \lambda_k\|x_k - x\|^2,$$

nossa primeira preocupação é mostrar que o conjunto $\arg\min_{x\in\mathbb{R}^n}\{f_k(x)\}$ dos minimizadores da função f_k possui um único elemento, pois só assim a fórmula (IV.2.1) faz sentido. Em seguida mostraremos que a seqüência gerada pelo algoritmo converge para o minimizador da função objetivo, caso exista algum.

IV.2.1 Boa definição

Para a boa definição da seqüência $\{x_k\}$, mostraremos que, para todo $k > 0$, a função

$$f_k(x) = f(x) + \lambda_k\|x_k - x\|^2$$

possui um único minimizador. Note que a função f_k é uma regularização aproximada de f, isto é, se f_k possuir minimizador, temos a expectativa de que f_k forneça algum tipo de aproximação ao problema de minimizar a função objetivo f.

Pela Proposição III.3.20, para todo $k > 0$, dado $x_k \in \mathbb{R}^n$, existe $s_k \in \mathbb{R}^n$ tal que $f(x) \geq f(x_k) + \langle s_k, x - x_k \rangle$, e isto implica que

$$f_k(x) = f(x) + \lambda_k \|x_k - x\|^2 \geq f(x_k) + \langle s_k, x - x_k \rangle + \lambda_k \|x_k - x\|^2 \ .$$

Dividindo por $\|x - x_k\|$, obtemos

$$\frac{f_k(x)}{\|x - x_k\|} \geq \frac{f(x_k)}{\|x - x_k\|} + \langle s_k, \frac{x - x_k}{\|x - x_k\|} \rangle + \lambda_k \|x_k - x\|$$

e passando ao limite, com $\|x - x_k\|$ tendendo a $+\infty$, vem que

$$\lim_{\|x - x_k\| \to +\infty} \frac{f_k(x)}{\|x - x_k\|} = +\infty \ .$$

Isto significa que f_k é *1-coerciva* e, pelo Teorema III.3.8, f_k possui pelo menos um minimizador.

É fácil ver que a função $h(x) = \lambda_k \|x_k - x\|^2$ é estritamente convexa. Juntando a isto o resultado da Proposição III.3.3, temos que f_k é uma função estritamente convexa e, portanto, f_k possui um único minimizador. Assim, a seqüência $\{x_k\}$ gerada pela iteração (IV.2.1) está bem definida.

Seja x_{k+1} o único minimizador de f_k. Vale, pelos Teorema III.3.22 e Teorema III.3.26, que

$$0 \in \partial f_k(x_{k+1}) = \partial f(x_{k+1}) + \{2\lambda_k(x_{k+1} - x_k)\},$$

o que equivale dizer

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

Desta forma acabamos de demonstrar o seguinte resultado:

Teorema IV.2.1. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. A seqüência $\{x_k\} \subset \mathbb{R}^n$ gerada por* (IV.2.1) *está bem definida e caracterizada pela relação*

$$-2\lambda_k\,(x_{k+1} - x_k) \in \partial f(x_{k+1}). \tag{IV.2.2}$$

IV.2.2 Convergência

Na literatura, até há pouco tempo, os resultados de convergência para o algoritmo de ponto proximal eram obtidos somente para o caso da função convexa possuir minimizador. A demonstração da convergência do algoritmo apresentada aqui, depende somente de alguns conceitos de análise no $\mathbb{R}^n$ e dos lemas a seguir.

Seja U^* o conjunto dos minimizadores da função convexa f e suponhamos que a seqüência de números positivos $\{\lambda_k\}$ é tal que $\sum\limits_{k=0}^{\infty} \dfrac{1}{\lambda_k} = +\infty$. Provaremos nesta seção que a seqüência $\{x_k\}$ dada pela iteração (IV.2.1) converge, se o conjunto U^* é não vazio, e diverge se U^* é vazio.

A Definição IV.2.2 e o próximo lema, são resultados gerais sobre a seqüência, que independem das características da função.

Definição IV.2.2. Uma seqüência $\{x_k\} \subset \mathbb{R}^n$ é dita *Fejér convergente* ao conjunto $U \neq \emptyset$, $U \subset \mathbb{R}^n$, se

$$\|u - x_{k+1}\| \leq \|u - x_k\|, \tag{IV.2.3}$$

para todo $k \geq 0$ e todo $u \in U$.

Lema IV.2.3. *Se $\{x_k\}$ é Fejér convergente ao conjunto $\emptyset \neq U \subset \mathbb{R}^n$, então $\{x_k\}$ é uma seqüência limitada. Se, além disso, um ponto de acumulação $\overline{x}$ da seqüência $\{x_k\}$ pertence a U, então $\overline{x} = \lim\limits_{k\to+\infty} x_k$.*

Demonstração. Dado $u \in U$, a desigualdade (IV.2.3) implica que $\|u - x_k\| \leq \|u - x_0\|$, para todo k. Deste modo a seqüência $\{x_k\}$ é limitada. Agora sejam $\overline{x} \in U$ um ponto de acumulação de $\{x_k\}$ e $\{x_{k_j}\}$ uma subseqüência de $\{x_k\}$ tal que $\lim\limits_{j \to +\infty} x_{k_j} = \overline{x}$. Como $\overline{x} \in U$, segue de (IV.2.3) que a seqüência de números reais positivos $\{\|x_k - \overline{x}\|\}$ é monótona não-crescente e possui uma subseqüência, a saber $\{\|x_{k_j} - \overline{x}\|\}$, convergindo para zero. Então a seqüência converge para zero, isto é, $0 = \lim\limits_{k \to +\infty} \|x_k - \overline{x}\|$, o que significa $\lim\limits_{k \to +\infty} x_k = \overline{x}$. $\square$

A hipótese feita sobre a seqüência $\{\lambda_k\}$ é motivada pelo lema a seguir.

Lema IV.2.4. *Seja $f \colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Se a seqüência $\{x_k\}$ é gerada pelo algoritmo* (IV.2.1), *então vale a desigualdade*

$$\|x - x_{k+1}\|^2 \leq \|x - x_k\|^2 - \|x_{k+1} - x_k\|^2 + \frac{1}{\lambda_k}(f(x) - f(x_{k+1})),$$

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

Demonstração. Seja $x \in \mathbb{R}^n$. Considere as igualdades

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

Isto implica que

$$\|x - x_{k+1}\|^2 = \|x - x_k\|^2 - \|x_{k+1} - x_k\|^2 + 2\langle x_k - x_{k+1}, x - x_{k+1} \rangle \ . \tag{IV.2.4}$$

Pelo Teorema IV.2.1,

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

Da definição de $\partial f(x_{k+1})$ temos

$$\begin{aligned}
f(x) &\geq f(x_{k+1}) + \langle 2\lambda_k(x_k - x_{k+1}), x - x_{k+1} \rangle \\
&= f(x_{k+1}) + 2\lambda_k \langle x_k - x_{k+1}, x - x_{k+1} \rangle,
\end{aligned}$$

e isto implica que

$$\frac{1}{\lambda_k}(f(x) - f(x_{k+1})) \geq 2\langle x_k - x_{k+1}, x - x_{k+1}\rangle. \tag{IV.2.5}$$

Substituindo (IV.2.5) em (IV.2.4) temos

$$\|x - x_{k+1}\|^2 \leq \|x - x_k\|^2 - \|x_{k+1} - x_k\|^2 + \frac{1}{\lambda_k}(f(x) - f(x_{k+1})), \tag{IV.2.6}$$

que é o resultado desejado. □

Corolário IV.2.5. *Se* $\{x_k\}$ *é a seqüência gerada por* (IV.2.1), *então vale a desigualdade*

$$\|x - x_{k+1}\|^2 \leq \|x - x_k\|^2 + \frac{1}{\lambda_k}(f(x) - f(x_{k+1})), \tag{IV.2.7}$$

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

Demonstração. Segue do Lema IV.2.4. □

Demonstraremos a seguir o principal resultado deste capítulo: a convergência do algoritmo de ponto proximal.

Teorema IV.2.6. *Sejam* $f\colon \mathbb{R}^n \to \mathbb{R}$ *uma função convexa e* $\{x_k\}$ *a seqüência gerada por* (IV.2.1). *Se a seqüência* $\{\lambda_k\}$ *é tal que* $\sum_{k=0}^{\infty} \frac{1}{\lambda_k} = +\infty$, *então* $\lim_{k\to+\infty} f(x_k) = f_*$, *onde* $f_* = \inf_{x\in\mathbb{R}^n} f(x)$. *Se, além disso, o conjunto* U^* *dos minimizadores de* f *é não vazio, então* $\lim_{k\to+\infty} x_k = x_*$, *com* $x_* \in U^*$.

Demonstração. Seja U^* o conjunto dos minimizadores da função convexa f. Se $x_k \in U^*$, para algum $k > 0$, pela iteração do algoritmo obtemos $x_k = x_{k+1} = x_{k+2} = ...$ e daí $f(x_k) = f_*$ e não há mais nada a fazer. Agora suponhamos que $x_k \notin U^*$, para todo k. Substituindo x por x_k em (IV.2.6) concluímos que a seqüência $\{f(x_k)\}$ é estritamente decrescente. Devemos provar que $\lim_{k\to+\infty} f(x_k) = f_*$.Suponhamos então, por absurdo, que $\lim_{k\to+\infty} f(x_k) > f_*$. Então existe $x \in \mathbb{R}^n$ e $\delta > 0$ tal que

$$f(x) < f(x_k) - \delta, \tag{IV.2.8}$$

para todo $k > 0$. Substituindo a desigualdade (**??**) na inequação (**??**), temos

$$\|x - x_{k+1}\|^2 \leq \|x - x_k\|^2 - \frac{\delta}{\lambda_k},$$

para todo k. Mas isto implica que

$$\frac{\delta}{\lambda_k} \leq \|x - x_k\|^2 - \|x - x_{k+1}\|^2,$$

para todo $k > 0$. Somando termo a termo

$$\begin{aligned} \sum_{k=0}^{j} \frac{1}{\lambda_k} &\leq \frac{1}{\delta}(\|x - x_0\|^2 - \|x - x_{j+1}\|^2) \\ &\leq \frac{1}{\delta}\|x - x_0\|^2, \end{aligned}$$

para todo j, o que contraria $\sum_{k=0}^{\infty} \frac{1}{\lambda_k} = +\infty$. Logo $\lim_{k\to+\infty} f(x_k) = f_*$. Se, além disso, $U^* \neq \emptyset$, tome $\overline{x} \in U^*$. Deste modo $f(\overline{x}) \leq f(x_k)$, para todo k. Substituindo x por $\overline{x}$ em (IV.2.7) obtemos $\|\overline{x} - x_{k+1}\|^2 \leq \|\overline{x} - x_k\|^2$, para todo k, e assim a seqüência $\{x_k\}$ é Fejér convergente a U^*. Logo, pelo Lema IV.2.3, $\{x_k\}$ é limitada. Seja $\{x_{k_j}\}$ uma subseqüência convergente de $\{x_k\}$, digamos $\lim_{j\to+\infty} x_{k_j} = x_*$. Da primeira parte $\lim_{j\to+\infty} f(x_{k_j}) = f_*$ e, sendo f contínua, $f(x_*) = f_*$, o que implica $x_* \in U^*$. Portanto, novamente do Lema IV.2.3, concluímos que a seqüência $\{x_k\}$ converge a x_*, isto é, $\lim_{k\to+\infty} x_k = x_*$. $\square$

IV.3 Observações finais

Vimos neste capítulo que a seqüência gerada pela iteração do método de ponto proximal é unicamente determinada. Na análise de convergência do método mostramos que o limite dos valores funcionais de $f(x_k)$ tende ao ínfimo da função f quando $k \to +\infty$. Caso este ínfimo seja $-\infty$, ou não seja atingido, a seqüência $\{x_k\}$ diverge, senão a seqüência é Fejér convergente ao conjunto não vazio U^* dos minimizadores da função convexa f.

CAPÍTULO V

Operadores Monótonos Maximais

V.1 Introdução

Encontrar singularidades de operadores é um problema clássico em equações diferenciais. Um importante resultado devido à Rockafellar [11] que estudaremos no Captulo VI, é utilizar o algoritmo de ponto proximal para determinar zeros de um operador monótono maximal arbitrário. No Capítulo III verificamos que um ponto $x_* \in \mathbb{R}^n$ é um minimizador de uma função convexa $f\colon \mathbb{R}^n \to \mathbb{R}$ se, e somente se, $0 \in \partial f(x_*)$. Sendo ∂f um operador monótono maximal, como será visto a seguir, o problema de descobrir suas singularidades generaliza, em um certo sentido, o problema de minimizar funções convexas.

Neste capítulo introduziremos o conceito de operadores monótonos em $\mathbb{R}^n$, operadores monótonos ponto-conjunto e operadores monótonos maximais (veja Iusem [4] e Ortega-Rheinboldt [9]). Encerraremos o capítulo provando a maximalidade do subdiferencial de uma função convexa. No próximo capítulo estudaremos o algoritmo de ponto proximal para operadores monótonos maximais arbitrários.

V.2 Operadores Monótonos contínuos

Nesta seção, veremos que o conceito de convexidade está relacionado com o conceito de monotonicidade, isto é, uma função diferenciável é convexa se, e somente se, o seu gradiente é monótono. Este fato nos fornece uma classe de exemplos de operadores monótonos, a saber, aqueles que são gradientes de funções convexas. Naturalmente surge a pergunta: todo operador monótono é gradiente de uma função convexa? Daremos um exemplo de operadores monótonos que não são gradientes de funções convexas utilizando um "critério" conhecido para determinar convexidade de uma função: verificando se a *hessiana* é uma matriz simétrica, para todo $x \in \mathbb{R}^n$.

Definição V.2.1. Dizemos que o operador $T\colon \mathbb{R}^n \to \mathbb{R}^n$ é *monótono* se

$$\langle T(x) - T(y), x - y \rangle \geq 0, \tag{V.2.1}$$

para todos x e $y \in \mathbb{R}^n$. Dizemos ainda que T é *estritamente monótono* se a desigualdade (V.2.1) é estrita.

Proposição V.2.2. *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável. f é convexa se, e somente se, $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$ é monótono.*

Demonstração. Suponha f convexa. Dados x e $y \in \mathbb{R}^n$ vale, pelo Teorema III.3.21,

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \quad \text{e} \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle),$$

que é equivalente, respectivamente, a

$$\langle -\nabla f(y), x - y \rangle \geq f(y) - f(x) \quad \text{e} \quad \langle \nabla f(x), x - y \rangle \geq f(x) - f(y).$$

Somando-se os lados correspondentes das desigualdades acima obtemos

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$$

Portanto, ∇f é monótono. Reciprocamente, suponha ∇f monótono. Pela inequação (V.2.1), vale: $\langle \nabla f(x), x-y \rangle \geq \langle \nabla f(y), x-y \rangle$. Recorrendo ao *Teorema do Valor Médio*:

$$f(y) - f(x) = \langle \nabla f(\xi), y - x \rangle,$$

para todo $x, y \in \mathbb{R}^n$ e algum $\xi \in (x, y)$. Assim, existe $\alpha \in (0,1)$, tal que:

$$\begin{aligned} f(y) - f(x) &= \langle \nabla f(\xi), \frac{\xi - x}{\alpha} \rangle \ = \ \frac{1}{\alpha} \langle \nabla f(\xi), \xi - x \rangle \\ &\geq \frac{1}{\alpha} \langle \nabla f(x), \xi - x \rangle \ = \langle \nabla f(x), y - x \rangle. \end{aligned}$$

Novamente, pelo Teorema III.3.21, f é convexa se, e somente se, $f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$, para todo $x, y \in \mathbb{R}^n$. Assim, concluímos o que desejávamos. $\square$

Exemplo V.2.3. Seja $A \in M_{n\times n}$ uma matriz anti-simétrica e considere o operador

$$\begin{aligned} T \colon \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto T(x) = Ax. \end{aligned}$$

Dados $x, y \in \mathbb{R}^n$,

$$\begin{aligned} \langle T(x) - T(y), x - y \rangle &= \langle A(x-y), x-y \rangle = \langle x - y, A^t(x-y) \rangle \\ &= \langle x - y, -A(x-y) \rangle = -\langle A(x-y), x-y \rangle \\ &= -\langle T(x) - T(y), x - y \rangle, \end{aligned}$$

o que implica $\langle T(x) - T(y), x - y \rangle = 0$. Portanto, T é monótono. Observemos que $T(x) = Ax$ não pode ser gradiente de nenhuma função, pois senão deveria existir uma função f, duas vezes diferenciável, tal que $\nabla f(x) = Ax = T(x)$ e $\nabla^2 f(x) = A$, o que é um absurdo, pois a *hessiana* $\nabla^2 f(x)$ é simétrica, ao passo que A é anti-simétrica.

V.3 Operadores monótonos ponto-conjunto

Neste momento, consideremos um operador T que associa, para cada $x \in \mathbb{R}^n$, um subconjunto de $\mathbb{R}^n$. Chamaremos este operador com contradomínio no conjunto das partes de $\mathbb{R}^n$, $\mathcal{P}(\mathbb{R}^n)$, de operador ponto-conjunto. Definiremos nesta seção os operadores monótono maximais, que são multifunções estudadas extensivamente por causa de seu importante papel em análise convexa e em equações diferenciais. Mostraremos no final da seção que o subdiferencial de uma função convexa é um operador monótono maximal.

Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ convexa, não necessariamente diferenciável. Dados $x, y \in \mathbb{R}^n$, tomemos $\xi, \eta \in \mathbb{R}^n$, tais que $\xi \in \partial f(x)$ e $\eta \in \partial f(y)$ e

$$\begin{aligned} f(x) \geq f(y) + \langle \eta, x - y\rangle &\Rightarrow \langle\, - \eta, x - y\rangle \geq f(y) - f(x), \\ f(y) \geq f(x) + \langle \xi, y - x\rangle &\Rightarrow \langle \xi\ , x - y\rangle \geq f(x) - f(y). \end{aligned}$$

Somando termo a termo as desigualdades à direita obtemos

$$\langle \xi - \eta, x - y\rangle \geq 0. \tag{V.3.1}$$

Como $\partial f\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ associa para cada x não exatamente um vetor, mas um subconjunto de $\mathbb{R}^n$, estenderemos a noção de operador monótono ponto-ponto para operador monótono ponto-conjunto.

Definição V.3.1. $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é *monótono* se

$$0 \leq \langle x - y, u - v\rangle,$$

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

Observação V.3.2. Segue de (V.3.1) que, se f é convexa, então ∂f é monótono.

Assim para operadores não gradientes, monotonicidade pode ser considerada como uma generalização natural de convexidade.

Definição V.3.3. $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é *monótono maximal* se,
(i) T é monótono;
(ii) Para todo T' monótono tal que $T(x) \subset T'(x)$ para todo x, vale que $T = T'$.

Geometricamente, a definição acima diz que, se um operador T é monótono, e se o seu gráfico $G(T) = \{(z, w) \in \mathbb{R}^n \times \mathbb{R}^n : w \in T(z)\}$ não está contido propriamente no gráfico de nenhum outro operador monótono $T'\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$, então T é monótono maximal.

No próximo capítulo exibiremos o algoritmo de ponto proximal para operadores monótonos maximais arbitrários. Basicamente, este algoritmo resolve o problema de encontrar singularidades deste tipo de operador. Sendo ∂f monótono para toda função convexa f, desejamos verificar a sua maximalidade, pois desta forma estaremos generalizando o problema de minimizar funções convexas. No intuito de demonstrar a maximalidade do operador ∂f provaremos o Lema V.3.4.

Sejam $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $x_0 \in \mathbb{R}^n$. Defina a função

$$\begin{aligned} F\colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto F(x) = f(x) + \frac{1}{2}\|x - x_0\|^2. \end{aligned} \tag{V.3.2}$$

Já mostramos, no Capítulo IV, que F possui um único minimizador. Introduzindo uma notação mais adequada, definimos:

$$\operatorname{prox}_f(x_0) := \operatorname*{argmin}_{x \in \mathbb{R}^n}\{f(x) + \frac{1}{2}\|x - x_0\|^2\}.$$

Lema V.3.4. *Sejam $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa e x, y, $z \in \mathbb{R}^n$. As seguintes condições são equivalentes:*

(i) $z = x + y$ e $f(x) + f^*(y) = \langle x, y\rangle$;

(ii) $x = \operatorname{prox}_f(z)$ e $y = \operatorname{prox}_{f^*}(z)$.

Demonstração. Faremos algumas considerações preliminares antes de provarmos as equivalências. Defina $h\colon \mathbb{R}^n \to \mathbb{R}$ por:

$$h(u) = \frac{1}{2}\|u - z\|^2.$$

Logo, $F = f + h$. Fazendo $x = \operatorname{prox}_f(z)$, temos que x é um minimizador da função F se, e somente se, $0 \in \partial F(x)$, que é equivalente a $0 \in (\partial f(x) + \partial h(x))$. Mas $\nabla h(u) = u - z$, e isto implica

$$x = \operatorname{prox}_f(z) \Leftrightarrow 0 \in (\partial f(x) + \{x - z\}), \tag{V.3.3}$$

ou seja, existe $y \in \partial f(x)$ tal que $z = x + y$. Analogamente, podemos mostrar que

$$y = \operatorname{prox}_{f^*}(z). \tag{V.3.4}$$

Isto é, existe $x \in \partial f^*(y)$ tal que $z = x + y$. O Teorema III.3.31 e a Proposição III.3.35 implicam

$$y \in \partial f(x) \Leftrightarrow f(x) + f^*(y) = \langle x, y\rangle \Leftrightarrow x \in \partial f^*(y). \tag{V.3.5}$$

$(i) \Rightarrow (ii)$ Por hipótese $z = x+y$ e $f(x)+f^*(y) = \langle x, y\rangle$. Então, da equação (V.3.5) obtemos

$$y \in \partial f(x) \ \text{ e } \ x \in \partial f^*(y).$$

Combinando estes resultados com (i), (V.3.3) e (V.3.4) obtemos

$$x = \operatorname{prox}_f(z) \ \text{ e } \ y = \operatorname{prox}_{f^*}(z).$$

$(i) \Leftarrow (ii)$ Seja $x = \operatorname{prox}_f(z)$. Da equação (V.3.3) obtemos a existência de um $y_0 \in \partial f(x)$ tal que $z = x + y_0$. Observemos que a equação (V.3.5) implica

$$x \in \partial f^*(y_0) \text{ e } f(x) + f^*(y_0) = \langle x, y_0 \rangle.$$

Combinando esses resultados com (V.3.4) obtemos $y_0 = \operatorname{prox}_{f^*}(z) = y$. Daí segue que

$$z = x + y \text{ e } f(x) + f^*(y) = \langle x, y \rangle. \qquad \square$$

Teorema V.3.5. *Se $f\colon \mathbb{R}^n \to \mathbb{R}$ é convexa então $\partial f\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é monótono maximal.*

Demonstração. Pela Observação V.3.2, ∂f é monótono. Provemos que ∂f é maximal, para todo $x \in \mathbb{R}^n$. Suponha que exista um operador $T(x)$ tal que $\partial f(x) \subset T(x)$, para todo $x \in \mathbb{R}^n$. Mostremos que se $y \in T(x)$, então $y \in \partial f(x)$. Sejam $x_0, y_0 \in \mathbb{R}^n$, com $y_0 \in T(x_0)$. Se $y \in T(x)$, então

$$\langle x - x_0, y - y_0 \rangle \geq 0, \tag{V.3.6}$$

para todo $x, y \in \mathbb{R}^n$ com $y \in \partial f(x)$, pois T é monótono. Tome $x_1 = \operatorname{prox}_f(x_0 + y_0)$ e $y_1 = \operatorname{prox}_{f^*}(x_0 + y_0)$. O Lema V.3.4 e o Teorema III.3.31 implicam

$$x_0 + y_0 = x_1 + y_1 \text{ e } y_1 \in \partial f(x_1) \subset T(x_1). \tag{V.3.7}$$

Fazendo $x = x_1$ e $y = y_1$ em (V.3.6) obtemos, de (V.3.7), que

$$0 \leq \langle y_1 - y_0, x_1 - x_0 \rangle = \langle x_0 - x_1, x_1 - x_0 \rangle = -\|x_0 - x_1\|^2.$$

Conseqüentemente $x_0 = x_1$ e $y_0 = y_1 \in \partial f(x_1) = \partial f(x_0)$. Assim, concluímos que ∂f é monótono maximal. $\square$

CAPÍTULO VI

O algoritmo para operadores monótonos maximais

VI.1 Introdução

O algoritmo de ponto proximal é fundamental para resolver o problema de encontrar $x \in \mathbb{R}^n$ tal que $0 \in T(x)$, onde T é um operador monótono maximal. O algoritmo é baseado no fato (veja Minty [7]) de que para cada $z \in \mathbb{R}^n$ e $c > 0$, existe um único $u \in \mathbb{R}^n$ tal que $z - u \in cT(u)$, ou, em outras palavras,

$$z \in (I + cT)(u). \tag{VI.1.1}$$

Neste capítulo generalizaremos o algoritmo de ponto proximal apresentado no Capítulo IV. Mostraremos que a seqüência $\{x_k\} \subset \mathbb{R}^n$ gerada pela iteração obtida para o caso de operador monótono maximal está bem definida e que, se existir $\overline{x} \in \mathbb{R}^n$ tal que $0 \in T(\overline{x})$, então a seqüência $\{x_k\}$ convergirá para uma singularidade do operador T.

VI.2 O algoritmo

Nesta seção obteremos a iteração que define o método de ponto proximal para operadores monótonos maximais. Para tanto definiremos a inversa de um operador ponto-conjunto.

Definição VI.2.1. Seja $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$. Definimos a *inversa* de T como

$$\begin{aligned} T^{-1}\colon \mathbb{R}^n &\to \mathcal{P}(\mathbb{R}^n) \\ x &\mapsto T^{-1}(x), \end{aligned}$$

onde $y \in T^{-1}(x)$ se, e somente se, $x \in T(y)$.

No Capítulo IV vimos que, dada uma função convexa f, o algoritmo de ponto proximal para minimização em $\mathbb{R}^n$ gera, para um ponto inicial $x_0 \in \mathbb{R}^n$, uma seqüência $\{x_k\}$ pela iteração $x_{k+1} := \arg\min\limits_{x\in\mathbb{R}^n}\{f(x) + \lambda_k\|x_k - x\|^2\}$, o que implicava

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

Disto e de (VI.1.1) temos que

$$x_k \in (I + \frac{1}{2\lambda_k}\partial f)(x_{k+1}).$$

A partir desta relação de pertinência definimos o algoritmo de ponto proximal para encontrar os zeros de um operador monótono maximal arbitrário $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ da seguinte forma:

Para um ponto inicial $x_0 \in \mathbb{R}^n$ e $\{\lambda_k\}$ uma seqüência de números reais tais que $\lambda_k > 0$, para todo k, a iteração de ponto proximal

$$x_{k+1} \in (I + \frac{1}{2\lambda_k}T)^{-1}(x_k) \tag{VI.2.1}$$

gera a seqüência $\{x_k\} \subset \mathbb{R}^n$, que define o algoritmo.

VI.2.1 Boa definição

A boa definição do algoritmo está fundamentada no Teorema de Minty, enunciado como Teorema VI.2.4, cuja demonstração será omitida e pode ser encontrada em Minty [7]. As definições a seguir serão introduzidas por fazerem parte das hipóteses do Teorema de Minty.

Definição VI.2.2. Seja $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$.
(i) Dizemos que T é *sobrejetivo* se, para todo $y \in \mathbb{R}^n$, existe $x \in \mathbb{R}^n$ tal que $y \in T(x)$.
(ii) Dizemos que T é *injetivo* se, para todos x, $y \in \mathbb{R}^n$, $x \neq y$, implica $T(x) \cap T(y) = \emptyset$.

Definição VI.2.3. Dizemos que $P\colon \mathbb{R}^n \to \mathbb{R}^n$ é firmemente não-expansivo se

$$\|P(x) - P(y)\|^2 \leq \|x - y\|^2 - \|(x - y) - (P(x) - P(y))\|^2,$$

para todos x e y em $\mathbb{R}^n$.

Teorema VI.2.4. ***(Teorema de Minty)*** *Se $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é um operador monótono maximal e $\mu > 0$, então $I + \mu T$ é injetivo e sobrejetivo e $(I + \mu T)^{-1}$ é firmemente não expansivo.*

Em particular, o Teorema de Minty é um teorema de existência: se T é um operador monótono maximal, então a sentença $y \in T(x)$ tem solução para todo x.

Proposição VI.2.5. *A seqüência $\{x_k\}$ gerada pelo algoritmo* (VI.2.1) *está bem definida.*

Demonstração. Do Teorema de Minty, T é sobrejetivo e, para todo k, existe $x_{k+1} \in \mathbb{R}^n$ tal que

$$x_{k+1} \in (I + \frac{1}{2\lambda_k} T)^{-1}(x_k). \qquad \square$$

VI.2.2 Convergência

A convergência do algoritmo de ponto proximal para operadores monótonos maximais arbitrários, tal como no artigo de Rockafellar [11], deve-se à suposição inicial da existência de um $\overline{x} \in \mathbb{R}^n$ tal que $0 \in T(\overline{x})$. Na demonstração do teorema da convergência usamos o resultado do Lema VI.2.6 que supõe existirem seqüências y_k e z_k tais que $y_k \in T(z_k)$, onde T é monótono maximal.

Lema VI.2.6. *Se* $\lim_{k\to+\infty} y_k = \overline{y}$, $\lim_{k\to+\infty} z_k = \overline{z}$, *$T$ é monótono maximal e* $y_k \in T(z_k)$, *então* $\overline{y} \in T(\overline{z})$.

Demonstração. Defina T' como

$$T'(z) = \begin{cases} T(z), & \text{se } z \neq \overline{z} \\ T(\overline{z}) \cup \{\overline{y}\}, & \text{se } z = \overline{z} \end{cases}.$$

Queremos mostrar que T' é monótono. Para isso devemos checar

$$\langle y - y', z - z' \rangle \geq 0, \tag{VI.2.2}$$

para todo z e $z' \in \mathbb{R}^n$ e para todo $y \in T'(z)$ e $y' \in T'(z')$. Como T é monótono basta verificar (VI.2.2) apenas para $y' = \overline{y}$ e $z' = \overline{z}$. Pela monotonicidade de T, para todo k:

$$\langle y - y_k, z - z_k \rangle \geq 0,$$

para todo z e para todo $y \in T(z)$. Usando o fato que $y_k \to \overline{y}$ e $z_k \to \overline{z}$ implicam $\langle y - \overline{y}, z - \overline{z} \rangle \geq 0$, temos que T' é monótono. Como $T(x) \subset T'(x)$ para todo x e T é monótono maximal, concluímos que $T = T'$ e em particular $T(\overline{z}) = T'(\overline{z}) = T(\overline{z}) \cup \{\overline{y}\}$, isto é, $\overline{y} \in T(\overline{z})$. $\square$

Teorema VI.2.7. *Se $T\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é monótono maximal e existe $\overline{x}$ tal que $0 \in T(\overline{x})$, então a seqüência $\{x_k\}$ definida em* (VI.2.1)*, com $0 < \lambda_k < \tilde{\lambda}$, converge para um vetor x_* tal que $0 \in T(x_*)$.*

Demonstração. Tome $\overline{x} \in \mathbb{R}^n$ tal que $0 \in T(\overline{x})$. Temos

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

o que implica que

$$\|x_{k+1} - \overline{x}\|^2 = \|x_k - \overline{x}\|^2 - \|x_k - x_{k+1}\|^2 - \frac{1}{\lambda_k}\langle 2\lambda_k(x_k - x_{k+1}), x_{k+1} - \overline{x}\rangle. \qquad \text{(VI.2.3)}$$

Por hipótese $0 \in T(\overline{x})$ e, pelo Teorema IV.2.1, $2\lambda_k(x_k - x_{k+1}) \in T(x_{k+1})$. Sendo T monótono temos que

$$\langle 2\lambda_k(x_k - x_{k+1}) - 0, x_{k+1} - \overline{x}\rangle \geq 0,$$

donde obtemos

$$0 < \|x_{k+1} - \overline{x}\|^2 \leq \|x_k - \overline{x}\|^2 - \|x_k - x_{k+1}\|^2, \qquad \text{(VI.2.4)}$$

para todo $k > 0$. Desta desigualdade resulta que a seqüência $\{\|x_k - \overline{x}\|\}$ é monótona decrescente e limitada, portanto convergente. Também vale

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

para todo k. Passando ao limite quando $k \to +\infty$, obtemos $\lim\limits_{k\to+\infty}(x_k - x_{k+1}) = 0$.
Da equação (VI.2.4) também resulta que a seqüência $\{x_k\}$ é Fejér convergente ao conjunto $O^* = \{z \in \mathbb{R}^n : 0 \in T(z)\}$. Assim, pelo Lema IV.2.3, $\{x_k\}$ é limitada. Seja x_* um ponto de acumulação de $\{x_k\}$ e $\{x_{j_k}\}$ uma subseqüência de $\{x_k\}$ tal que $\lim\limits_{k\to+\infty} x_{j_k} = x_*$, e vale que $\lim\limits_{k\to+\infty}(x_{j_k} - x_{j_k+1}) = 0$. Pela iteração do algoritmo definido em (VI.2.1), temos que

$$2\lambda_{j_k}(x_{j_k} - x_{j_k+1}) \in T(x_{j_k+1}).$$

Como, por hipótese, $0 < \lambda_k < \tilde{\lambda}$, segue que $\lim\limits_{k\to+\infty}(2\lambda_{j_k}(x_{j_k} - x_{j_k+1}) = 0$. Sendo T maximal, do Lema VI.2.6, com $y_k = 2\lambda_{j_k}(x_{j_k} - x_{j_k+1})$, $\overline{y} = 0$, $z_k = x_{j_k+1}$ e $\overline{z} = x_*$ vem que $0 \in T(x_*)$, o que significa dizer que $x_* \in O^*$. Pela Fejér convergência da seqüência $\{x_k\}$ ao conjunto O^* segue, do Lema IV.2.3, que

$$\lim_{k\to+\infty} x_k = x_* \,. \qquad \square$$

VI.2.3 Observações finais

Para a convergência do algoritmo de ponto proximal no caso de operadores monótonos maximais, foi essencial a suposição da existência de um $\overline{x} \in \mathbb{R}^n$ tal que $0 \in T(\overline{x})$. Apesar da determinação das singularidades de um operador monótono maximal generalizar o problema de minimização de funções convexas, pois o operador ∂f é monótono maximal, na análise de convergência feita neste capítulo foi essencial a hipótese de que o conjunto dos minimizadores U^* da função objetivo fosse não vazio, o que não ocorreu na análise de convergência feita no Capítulo IV.

Esta dissertação é um estudo introdutório do algoritmo de ponto proximal para otimização convexa. Mas por tratar-se de um tema muito atual ele será de grande valor para estudar outras variantes do algoritmo, como é feito, por exemplo, em [4].

Bibliografia

[1] R. Correa and C. Lemaréchal, Convergence of some algorithms for convex minimization, Mathematical Programming **62** (1993), 267–275.

[2] O. Güler, On the convergence of the proximal point algorithm for convex minimization, SIAM J. Cont. and Opt., **29**, (1991), 403–419.

[3] J.-B. Hiriart-Urruty and C. Lemaréchal, Convex analysis and minimization algorithms I and II, Springer-Verlag (1993).

[4] A. N. Iusem, Proximal point methods in optimization, 20^0 Colóquio Brasileiro de Matemática, IMPA, (1995).

[5] E. L. Lima, Curso de análise, Volume 2, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1981.

[6] B. Martinet, Régularisation, d'inéquations variationelles par approximations successives, Revue Française d'Informatique et de Recherche Operationelle, (1970), 154–159.

[7] G. J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J., **29** (1962), 341–346.

[8] J. J. Moreau, Proximité et dualité dans un espace Hilbertien, Bull. Soc. Math., France, **93** (1965), 273–299.

[9] J. M. Ortega, W.C. Rheinboldt. Iterative solution of nonlinear equations in several variables, Academic Press, Inc, New York (1970).

[10] R. T. Rockafellar, Convex analysis, Princeton University Press, Princeton, 1970.

[11] R. T. Rockafellar. Monotone operators and the proximal point algoritm. SIAM J. cont. optim. **14** (1976), 877–898.

[12] Jan van Tiel. Convex analysis: an introductory text, Royal Netherlands Meteorological Institute, 1984.