Monografia de MAC5701 – Tópicos em Ciência da Computação

Universidade de São Paulo

Instituto de Matemática e Estatística

Departamento de Ciência da Computação

Otimização e Análise Convexa

Período: Primeiro Semestre de 2005

Aluna: Ellen Hidemi Fukuda — ellen at ime.usp.br

Orientador: Paulo J. Silva e Silva — pjssilva at ime.usp.br

Resumo

Esta é a monografia da disciplina MAC5701 – Tópicos em Ciência da Computação – da aluna de mestrado Ellen Hidemi Fukuda. O estudo foi realizado no primeiro semestre de 2005 sob a supervisão do professor Dr. Paulo José da Silva e Silva. O objetivo deste trabalho é estudar os principais conceitos de Análise Convexa, uma subárea de Otimização Contínua que utiliza ferramentas como conjuntos convexos, funções convexas e subdiferenciais.

Palavras-chaves: Convexidade, sublinearidade, funções suporte, subdiferenciais.

Sumário

1 Introdução

Consideremos o seguinte problema de Programação Linear (PL) de grande escala:

$$\begin{cases} \min & \langle c, x \rangle \\ \text{s.a} & Ax = b \\ & Dx = e \\ & x \geq 0, \end{cases} \tag{1}$$

em que $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $D \in \mathbb{R}^{d \times n}$, $e \in \mathbb{R}^d$ e D tem posto d. Consideramos que as restrições definidas pela matriz A são "difíceis", ou seja, que o problema (1) sem tais restrições é de fácil solução.

O algoritmo do volume (VA) foi proposto recentemente como alternativa para solução de (1), explorando a relaxação lagrangeana das restrições "difíceis" [BA00]. O método pode ser visto como um método de subgradientes com algumas características de métodos de feixe. Infelizmente a sua convergência não foi demonstrada, mas resultados numéricos encorajadores foram apresentados.

O algoritmo do volume foi recentemente tema de um outro artigo em que os autores propõem modificações que permitem demonstrar resultados de convergência [BMS02]. Isto é obtido com pequenas alterações no VA de modo a ser possível interpretá-lo como um método de feixe.

O projeto do mestrado propõe o estudo dos métodos descritos acima, bem como sua implementação e comparação numérica. Como tópico avançado deseja-se explorar melhor a relação entre esses algoritmos, métodos extragradiente e métodos de ponto proximal. No entanto, tudo isso não será possível sem estudar inicialmente os principais conceitos de Análise Convexa. Este será justamente o assunto a ser abordado na disciplina MAC5701.

O estudo de Análise Convexa se baseia essencialmente no livro *Fundamentals of Convex Analysis* [HUL01] de Jean-Baptiste Hiriart-Urruty e Claude Lemaréchal. Vale ainda ressaltar que nosso objetivo com a monografia é indicar os principais conceitos desse livro, sem longas explicações e demontrações de teoremas e corolários. Além disso, daremos maior importância aos conceitos essencialmente novos.

Para uma boa compreensão deste texto, supõe-se o conhecimento de alguns conceitos elementares, entre eles: limitantes superiores e inferiores, conjunto dos números reais estendidos, álgebra linear e bilinear, diferenciabilidade no espaço euclidiano, análise multivalorada e convexidade de funções com uma variável.

2 Conjuntos convexos

O espaço que trabalharemos aqui é o $\mathbb{R}^n$ equipado com um produto escalar $\langle \cdot, \cdot \rangle$, ou seja, um espaço euclidiano. Utilizando-se de uma base ortonormal, podemos também supor que $\langle x, y \rangle$ é o produto interno usual $x^T y$.

2.1 Definições básicas

Definição 2.1. *O conjunto* $C \subset \mathbb{R}^n$ *é dito* convexo *se* $\alpha x + (1-\alpha)x'$ *pertence a* C *para quaisquer* x *e* x' *em* C *e* $\alpha \in [0, 1]$.

A interpretação geométrica seria a seguinte: o segmento

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

está contido inteiramente em C para quaisquer x e x' em C. A seguir listaremos alguns dos principais conjuntos convexos, fundamentais nas próximas secções do nosso texto.

1. *Hiperplano afim* (ou, equivalentemente, *hiperplano*):

   $$H_{s,r} \doteq \{x \in \mathbb{R}^n : \langle s, x \rangle = r\},$$

   onde $(s, r) \in \mathbb{R}^n \times \mathbb{R}$ e $s \neq 0$.

2. *Variedade afim*: Um conjunto V tal que a linha $\{\alpha x + (1-\alpha)x' : \alpha \in \mathbb{R}\}$ está inteiramente contida em V para quaisquer x e x' em V.

3. *Semi-espaço afim* (ou, equivalentemente, *semi-espaço*):

   $$\begin{aligned} &\{x \in \mathbb{R}^n : \langle s, x \rangle < r\} \quad \text{(aberto)} \\ H^- \doteq &\{x \in \mathbb{R}^n : \langle s, x \rangle \leq r\} \quad \text{(fechado)} \end{aligned}$$

   onde $(s, r) \in \mathbb{R}^n \times \mathbb{R}$ e $s \neq 0$.

4. *Simplex*: Dado $\alpha = (\alpha_1, ..., \alpha_k) \in \mathbb{R}^k$ qualquer, um simplex unitário no $\mathbb{R}^k$ é dado por:

   $$\Delta_k \doteq \left\{\alpha \in \mathbb{R}^k : \sum_{i=1}^{k} \alpha_i = 1, \alpha_i \geq 0 \text{ para } i = 1, ..., k\right\}.$$

5. *Cones convexos*: Um cone K é um conjunto que possui a seguinte propriedade: $\{\alpha x : \alpha > 0\}$ está inteiramente contido em K para qualquer $x \in K$. Um exemplo de cone convexo definido no $\mathbb{R}^n$ é dado a seguir, com p e m escalares quaisquer:

$$\begin{aligned} \langle s_j, x\rangle &= 0 \quad \text{para } j = 1, ..., m \\ \langle s_{m+j}, x\rangle &\leq 0 \quad \text{para } j = 1, ..., p \end{aligned}$$

Existem algumas operações que preservam convexidade: intersecção de conjuntos convexos, produto cartesiano de conjuntos convexos, imagem de um conjunto convexo sobre um mapeamento afim, interior e fecho de um conjunto convexo. Outra operação com essa propriedade é a soma de Minkowski de conjuntos convexos (C_1 e C_2), ou seja,

$$C_1 + C_2 \doteq \{x = x_1 + x_2 : x_1 \in C_1, x_2 \in C_2\}.$$

Podemos generalizar essa afirmação e concluir que o conjunto

$$\alpha_1 C_1 + \alpha_2 C_2 \doteq \{x = \alpha_1 x_1 + \alpha_2 x_2 : x_1 \in C_1, x_2 \in C_2\}$$

também é convexo, onde α_1 e α_2 são números reais quaisquer.

Outra classe de operações interessantes toma conjuntos não convexos e constróem, a partir deles, conjuntos convexos. Para isso, os seguintes conceitos básicos de Álgebra Linear devem ser lembrados: combinação linear de elementos de $\mathbb{R}^n$, subspaço linear de $\mathbb{R}^n$, subspaço gerado por um conjunto S contido em $\mathbb{R}^n$ (denotado por $\operatorname{lin} S$ ou $\operatorname{span} S$; e que também é o menor subspaço que contém S) e independência linear de vetores. Eis as novas definições:

1. Uma *combinação afim* de elementos $x_1, ..., x_k$ de $\mathbb{R}^n$ é um elemento $\sum_{i=1}^k \alpha_i x_i$ onde $\sum_{i=1}^k \alpha_i = 1$.

2. Uma *variendade afim* em $\mathbb{R}^n$ é um conjunto contendo todos as suas combinações afins. Observe que a intersecção de variedades afins continuam sendo variedades afins.

3. A variedade afim gerada por um conjunto $S \subset \mathbb{R}^n$ (intersecção de todas as variedades afins que contêm S) é denotada por $\operatorname{aff} S$ e denominada *casco afim* de S. Temos também que $\operatorname{aff} S$ é a menor variedade afim que contém S.

4. Os pontos $x_0, x_1, ..., x_k$ são *afins de modo independente* se o conjunto $x_0 + \text{lin}\{x_0 - x_0, x_1 - x_0, ..., x_k - x_0\}$ tem dimensão completa k. Se $x \in \text{aff}\{x_0, x_1, ..., x_k\}$, então ele pode se escrito unicamente como sendo $x = \sum_{i=0}^{k} \alpha_i x_i$ com $\sum_{i=0}^{k} \alpha_i = 1$.

Duas outras definições importantes são dadas a seguir.

Definição 2.2. *Uma combinação convexa dos elementos $x_1, ..., x_k \in \mathbb{R}^n$ é um elemento da forma $\sum_{i=1}^{k} \alpha_i x_i$ onde $\sum_{i=1}^{k} \alpha_i = 1$ e $\alpha_i \geq 0$ para $i = 1, ..., k$.*

Note que uma combinação convexa é um tipo particular de combinação afim, que, por sua vez, é um tipo particular de combinação linear. Um fato importante que envolve esse conceito é que um conjunto $C \subset \mathbb{R}^n$ é convexo se, e somente se, contêm todas as combinações convexas de seus elementos.

Definição 2.3. *O* casco convexo *de $S \subset \mathbb{R}^n$ é dado por:*

$$\begin{aligned} \text{co}\, S \ &\doteq \cap\{C : C \text{ é convexo e contém } S\} \\ &= \{\textstyle\sum_{i=1}^{m} \alpha_i x_i : \alpha \in \Delta_m, x_i \in C_i \text{ para } i = 1, ..., m\}, \end{aligned}$$

onde $S = C_1 \cup C_2 \cup ... \cup C_m$, com cada C_i convexo.

As definições acima são úteis, em especial para o seguinte teorema:

Teorema 2.1. *(Carathéodory) Qualquer $x \in \text{co}\, S \subset \mathbb{R}^n$ pode ser representado como uma combinação convexa de $n + 1$ elementos de S.*

Muitos dos conceitos vistos acima podem ser aplicados a conjuntos fechados. O *casco convexo fechado* $\overline{\text{co}}\, S$ de um conjunto não vazio $S \subset \mathbb{R}^n$ é a intersecção de todos os conjuntos convexos fechados que contêm S. Esse casco convexo fechado é o fecho $\text{cl}(\text{co}\, S)$ do casco convexo de S. Ademais, pode-se provar que se S é limitado (resp. compacto), então $\text{co}\, S$ é limitado (resp. compacto).

Terminamos essa secção com a seguinte definição:

Definição 2.4. *A* combinação cônica *de elementos $x_1, ..., x_k$ é um elemento da forma $\sum_{i=1}^{k} \alpha_i x_i$, onde os coeficientes α_i são não negativos. Da mesma forma, podemos definir a combinação cônica de um conjunto $S \subset \mathbb{R}^n$ como o* casco cônico *de S (denotado por $\text{cone}\, S$).*

Analogamente ao casco convexo, temos que o *casco cônico fechado* $\overline{\text{cone}} S$ é o fecho de $\text{cone}\, S$, ou seja, $\text{cl}\,\text{cone}\, S$.

2.2 Conjuntos convexos associados a conjuntos convexos

Nesta secção mostraremos quatro conceitos um pouco mais sofisticados: interior relativo, cone assintótico, pontos extremos e faces expostas. Iniciemos pelo primeiro.

Definição 2.5. *O* interior relativo $\operatorname{ri} C$ *de um conjunto convexo* $C \subset \mathbb{R}^n$ *é o interior de* C *para a topologia relativa ao casco afim de* C*. Em outras palavras,* $x \in \operatorname{ri} C$ *se, e somente se,* $x \in \operatorname{aff} C$ *e existe* $\delta > 0$ *tal que* $(\operatorname{aff} C) \cap B(x, \delta) \subset C$*. A dimensão de um conjunto convexo* C *é a dimensão de seu casco afim, que é a dimensão do subspaço paralelo a* $\operatorname{aff} C$*.*

Da mesma forma, podemos definir a *borda relativa* de C como sendo $\operatorname{rbd} C \doteq \operatorname{cl} C \setminus \operatorname{ri} C$. O *fecho relativo* de C é apenas $\operatorname{cl} C$, porque os pontos de acumulação de um conjunto C já estão em $\operatorname{aff} C$. Podemos também afirmar que se $C \neq \emptyset$ então $\operatorname{ri} C \neq \emptyset$ e $\dim(\operatorname{ri} C) = \dim C$.

Os três conjuntos convexos $\operatorname{ri} C$, C e $\operatorname{cl} C$ têm o nesmo casco afim (e, conseqüentemente a mesma dimensão), o mesmo interior relativo e o mesmo fecho (e, por sua vez, a mesma borda relativa). Sob esses resultados, temos também que C e $\operatorname{cl} C$ têm o mesmo interior (e borda). Além disso, se dois conjuntos convexos C_1 e C_2 têm o mesmo interior relativo, então eles geram a mesma variedade afim e possuem o mesmo interior relativo. Isto ocorre exatamente quando temos a relação $\operatorname{ri} C_1 \subset C_2 \subset \operatorname{cl} C_1$.

O segundo conceito importante é definido abaixo:

Definição 2.6. *O* cone assintótico *(ou* de recessão*) associado a um conjunto convexo* C *fechado é o conjunto de direções que podemos percorrer de um ponto* x *até o infinito, permanecendo em* C*.*

Equivalentemente, temos as seguintes possíveis definições:

$$C_\infty(x) \doteq \{d \in \mathbb{R}^n : x + td \in C \text{ para todo } t > 0\} = \bigcap_{t>0} \frac{C - x}{t}.$$

Podemos provar que o cone convexo fechado $C_\infty(x)$ não depende de $x \in C$, ou seja, podemos denotar tal cone apenas por C_∞. Além disso, um conjunto convexo fechado C é compacto se, e somente se, $C_\infty = \{0\}$. Outras propriedades de cones assintóticos envolvem intersecções, produtos cartesianos e operadores lineares e serão omitidas nesse texto. Veremos agora o próximo conceito.

Definição 2.7. *Dizemos que* $x \in C$ *é um* ponto extremo *de* C *se não existem dois pontos distintos* x_1 *e* x_2 *em* C *tais que* $x = 1/2(x_1 + x_2)$*.*

Naturalmente, podemos substituir o escalar $1/2$ por qualquer $\alpha \in]0,1[$. Considere o conjunto de pontos extremos de C como sendo $\operatorname{ext} C$. Então, se C é compacto, temos $\operatorname{ext} C \neq \emptyset$.

Teorema 2.2. *(Minkowski) Seja C um conjunto compacto em $\mathbb{R}^n$. Então C é o casco convexo de seus pontos extremos, isto é, $C = \operatorname{co}(\operatorname{ext} S)$.*

Combinando o teorema acima com o de Carathéodory, conclui-se que se $\dim C = k$, então todo elemento de C é uma combinação convexa de no máximo $k+1$ pontos extremos de C.

Definição 2.8. *Um subconjunto convexo não vazio $F \subset C$ é uma* face *de C se ele satisfaz a seguinte propriedade: todo segmento de C que possui em seu interior relativo um elemento de F, está inteiramente contido em F.*

Se F é uma face de C, então podemos provar que todo ponto extremo de F é um ponto extremo de C.

Enunciaremos agora o último conceito dessa secção: *face exposta*, um tipo particular de face. Antes disso, dizemos que um hiperplano afim $H_{s,r}$ é um *hiperplano suporte* de um conjunto C quando C é contido inteiramente em um dos semi-espaços associados a $H_{s,r}$. Eis, agora, a definição:

Definição 2.9. *O conjunto $F \subset C$ é uma* face exposta *de C se existe um hiperplano suporte $H_{s,r}$ de C tal que $F = C \cap H_{s,r}$. Além disso, dizemos que um* ponto exposto *(ou* vértice*) é uma face exposta de dimensão zero.*

2.3 Projeção em conjuntos convexos fechados

Seja p_V a projeção ortogonal em um subespaço $V \subset \mathbb{R}^n$. Esse operador é linear, simétrico, positivo semi-definido e idempotente. Queremos generalizar tal operador para o caso em que V é inteiramente convexo. Pode-se demonstrar que um ponto $y_C \in C$ é a projeção $p_C(x)$ de um ponto x em um conjunto convexo C se, e somente se, $\langle x - y_x, y - y_x \rangle \leq 0$ para todo $y \in C$.

Vejamos um exemplo de seu uso tomando-se cones convexos fechados. Considere, inicialmente, o seguinte:

Definição 2.10. *(*Cone polar*) Seja K um cone convexo. Seu polar é*

$$K^\circ \doteq \{s \in \mathbb{R}^n : \langle s, x \rangle \leq 0 \,\text{para todo}\, x \in K\}$$

Note que a mudança de $\langle \cdot, \cdot \rangle$ acarreta mudança no cone polar. É fácil ver também que K° é um cone convexo fechado e que $K \cap K^\circ = \{0\}$. Além disso, se $K' \subset K$ então $(K')^\circ \supset K^\circ$. O contrário vale se $K^{\circ\circ} = K$ para K fechado. Agora que vimos a caracterização de cone polar, podemos mostrar o teorema de Jean-Jacques Moreau:

Teorema 2.3. *(Moreau) Seja K um cone convexo fechado e x, x_1, x_2 em $\mathbb{R}^n$. As seguintes afirmações são equivalentes:*

(a) $x = x_1 + x_2$ com $x_1 \in K$, $x_2 \in K^\circ$ e $\langle x_1, x_2 \rangle = 0$.

(b) $x_1 = p_K(x)$ e $x_2 = p_{K^\circ}(x)$.

2.4 Separação e aplicações

Considere a seguinte propriedade: dados dois conjuntos convexos S_1 e S_2 disjuntos, existe um conjunto convexo (no caso, um hiperplano afim) que pode ser colocado entre S_1 e S_2. Se tais conjuntos são fechados e se um deles é também limitado, podemos ter uma separação estrita.

Algumas das conseqüências dessa propriedade são as seguintes: existência de hiperplanos que suportam conjuntos convexos em um ponto de sua borda, descrições de conjuntos convexos por uma série de aproximações externas (por exemplo, um conjunto convexo fechado qualquer é uma intersecção de infinitos semi-espaços), e o teorema de Minkowski. Uma outra conseqüência é o teorema de Farkas, cuja versão generalizada é dada a seguir:

Teorema 2.4. *(Farkas) Sejam (b, r) e (s_j, ρ_j) em $\mathbb{R}^n \times \mathbb{R}$, com j variando em um conjunto de índices J arbitrário. Considere que o sistema de inequações $\langle s_j, x \rangle \leq \rho_j$ para todo $j \in J$ tem uma solução $x \in \mathbb{R}^n$. As seguintes afirmações são, portanto, equivalentes:*

(a) $\langle b, x \rangle \leq r$ para todo x que satisfaz o sistema;

(b) (b, r) se situa no casco convexo fechado cônico de S, defindo por $\{(0,1)\} \cup \{(s_j, \rho_j)\}_{j \in J}$.

2.5 Aproximações cônicas de conjuntos convexos

Dado um conjunto S e um elemento x de S, gostaríamos de aproximar S perto de x por um conjunto mais "simples". Na geometria diferencial clássica, uma superfície S "suave" pode ser aproximada por uma variedade afim (tangente) no ponto x. O intuito dessa secção é mostrar uma generalização dessa aproximação para conjuntos convexos. Em Análise Convexa, substitutos naturais para subspaços são os cones convexos fechados. Nossa nova definição de tangente é, portanto, a seguinte:

Definição 2.11. *(Direções tangentes) Seja $S \subset \mathbb{R}^n$ não vazio. Dizemos que $d \in \mathbb{R}^n$ é uma* direção tangente *para S em $x \in S$ se existe uma seqüência*

$(x_k) \subset S$ e uma seqüência (t_k) tal que quando $k \to +\infty$, temos:

$$x_k \to x, \qquad t_k \downarrow 0, \qquad \frac{x_k - x}{t_k} \to d.$$

O conjunto de todas essas direções é denominado cone tangente *de S em $x \in S$ e denotado por $T_S(x)$.*

Pode-se mostrar que tal cone tangente é fechado. Agora considere apenas os casos em que S é convexo. Então, podemos ter a seguinte caracterização (global) do cone tangente:

Proposição 2.1. *O* cone tangente *para um conjunto convexo fechado C em $x \in C$ é o fecho do cone gerado por $C - \{x\}$. Esse conjunto é, portanto, convexo.*

Um outro tipo de cone (relacionado ao cone tangente, conforme veremos posteriormente) é definido abaixo:

Definição 2.12. *A direção $s \in \mathbb{R}^n$ é dito* normal *para C em $x \in C$ se $\langle s, y - x\rangle \leq 0$ para todo $y \in C$. O conjunto de todas essas direções é denominado* cone normal *para C em x e denotado por $N_C(x)$.*

A relação entre o cone tangente e o cone normal é simples: um é o polar do outro. Sabendo-se disso, podemos estabelecer várias propriedades que envolvem ambos os cones. Apenas como ilustração, citemos uma delas: Para um x fixo, $T_C(x)$ cresce se, e somente se, $N_C(x)$ decresce e tal propriedade é válida particularmente quando C cresce. Outra propriedade interessante envolve cones normais e faces expostas:

Proposição 2.2. *Sejam $x \in C$ com C convexo e $s \in \mathbb{R}^n$. As seguintes afirmações são equivalentes:*

(a) $s \in N_C(x)$;

(b) x está na face exposta $F_C(x) : \langle s, x\rangle = \max_{y \in C}\langle s, y\rangle$;

(c) $x = p_C(x + s)$.

3 Funções convexas

Similarmente aos conjuntos convexos, introduziremos aqui os principais conceitos relacionados a funções convexas.

3.1 Definições básicas e exemplos

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

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

Além disso, f é dita *estritamente convexa* em C quando a desigualdade acima é estrita para $x \neq x'$. Dizemos ainda que f é *fortemente convexa* em C se existe $c > 0$ tal que, para todo $(x, x') \in C \times C$ e todo $\alpha \in]0, 1[$,

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha)f(x') - \frac{1}{2}c\alpha(1-\alpha)\|x - x'\|^2.$$

O conjunto C dessas definições aparece como domínio da definição de f. Uma outra definição considera a função definida inteiramente em $\mathbb{R}^n$, tomando-se possíveis valores iguais a infinito.

Definição 3.2. *Uma função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, não identicamente $+\infty$, é dita convexa quando, para todo $(x, x') \in \mathbb{R}^n \times \mathbb{R}^n$ e todo $\alpha \in]0, 1[$, temos:*

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

A classe de tais funções é denotada $\operatorname{Conv}\mathbb{R}^n$.

Para mostrar que as duas definições acima são equivalentes, basta estendermos f da definição 3.1, ou seja, $f(x) = +\infty$ se $x \notin C$. Dizemos, ainda, que o domínio efetivo de uma função $f \in \operatorname{Conv}\mathbb{R}^n$ é um conjunto não vazio $\operatorname{dom} f \doteq \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

Lembremos agora de um conceito que vale para qualquer função não identicamente $+\infty$:

Definição 3.3. *(Epígrafe) Dado $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, não identicamente $+\infty$, a* epígrafe *de f é um conjunto não vazio dado por:*

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

A *epígrafe estrita* de f é definida de maneira idêntica, só que com desigualdade estrita. Outra definição muito utilizada é mostrada a seguir:

Definição 3.4. *Um* conjunto de nível *em r de uma função f é dado por*

$$S_r(f) \doteq \{x \in \mathbb{R}^n : f(x) \leq r\}.$$

Relacionando tais conjuntos de nível com epígrafes, temos que $(x,r) \in \operatorname{epi} f$ se, e somente se, $x \in S_r(f)$. A seguinte proposição nos permite passar de uma idéia algébrica para uma idéia geométrica (e vice-versa) facilmente:

Proposição 3.1. *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ não identicamente igual a $+\infty$. As seguintes afirmações são, portanto, equivalentes:*

(a) f é convexa (usando-se a definição 3.2);

(b) A epígrafe de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$;

(c) A epígrafe estrita de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.

Só para relembrar, dizemos que f é *côncava* quando $-f$ é convexa, o que equivale dizer que a *hipógrafe* de f (i.e., definição 3.3 com a desigualdade invertida) é um conjunto convexo. Ademais, os conjuntos de nível de qualquer $f \in \operatorname{Conv} \mathbb{R}^n$ são também convexos.

A desigualde básica da definição 3.1 pode ser generalizada conforme pode ser observado abaixo:

Teorema 3.1. *(Desigualdade de Jensen) Seja $f \in \operatorname{Conv} \mathbb{R}^n$. Então, para toda coleção $\{x_1, ..., x_k\}$ de pontos no $\operatorname{dom} f$ e para todo $\alpha = (\alpha_1, ..., \alpha_k)$ em Δ_k, vale a desigualdade*

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

A epígrafe de uma função linear é caracterizada por $s \in \mathbb{R}^n$ e é dado por $(x,r) \in \mathbb{R}^n \times \mathbb{R}$ tal que $r \geq \langle s, x \rangle$. A epígrafe de uma função afim é, por sua vez, um semi-espaço fechado. É interessante notar que qualquer função convexa é minorada por uma função afim. Mais precisamente, se tomarmos $x_0 \in \operatorname{ri} \operatorname{dom} f$, então existe s no subspaço paralelo ao $\operatorname{aff} \operatorname{dom} f$ tal que $f(x) \geq f(x_0) + \langle s, x - x_0 \rangle$ para todo $x \in \mathbb{R}^n$. Em outras palavras, podemos forçar a função afim a coincidir com f em x_0.

Outra caracterização útil é dada a seguir:

Proposição 3.2. *Para $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, as seguintes três propriedades são equivalentes:*

(a) f é semi-contínua inferior em $\mathbb{R}^n$;

(b) $\operatorname{epi} f$ é um conjunto fechado em $\mathbb{R}^n \times \mathbb{R}$;

(c) Os conjuntos de nível $S_r(f)$ são fechados (possivelmente vazios) para todo $r \in \mathbb{R}$.

Uma função é dita *fechada* se satisfaz (a), (b) ou (c).

Definição 3.5. *O* fecho *de uma função* f *é uma função* $\mathrm{cl}\, f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ *definido por:* $\mathrm{cl}\, f(x) \doteq \liminf_{y \to x} f(y)$ *para todo* $x \in \mathbb{R}^n$*, ou equivalentemente,* $\mathrm{epi}(\mathrm{cl}\, f) \doteq \mathrm{cl}(\mathrm{epi}\, f)$.

O fecho de uma função convexa é também convexo. Denotamos uma função convexa fechada como $\overline{\mathrm{Conv}}\mathbb{R}^n$. Uma maneira de construir tais funções é por aproximações externas, pegando-se o supremo de todas as funções afins que minoram a função.

A partir desses resultados e definições, podemos citar os seguintes exemplos de funções convexas:

1. *Indicador*: Dado um subconjunto $S \subset \mathbb{R}^n$, a função $i_S : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ definida por
$$i_S(x) \doteq \begin{cases} 0 & \text{se } x \in S, \\ +\infty & \text{caso contrário} \end{cases}$$
é denominada função *indicador* de S. Naturalmente, i_S é fechada e convexa se, e somente se, S é fechado e convexo. Ademais, temos que $\mathrm{epi}\, i_S = S \times \mathbb{R}^+$.

2. *Função suporte*: Seja S um subconjunto de $\mathbb{R}^n$. A função suporte de S em um ponto x é dada por $\sigma_S(x) \doteq \sup\{\langle s, x\rangle : s \in S\}$. Tal função é convexa e fechada. Além disso, tanto sua epígrafe quanto seu domínio são cones convexos em $\mathbb{R}^n \times \mathbb{R}$ e em $\mathbb{R}^n$, respectivamente.

3. *Função afim por partes*: Considere $(s_1, b_1), ..., (s_m, b_m)$ elementos em $\mathbb{R}^n \times \mathbb{R}$. Uma função do tipo $\mathbb{R}^n \ni x \mapsto \breve{f}(x) \doteq \max\{\langle s_j, x\rangle - b_j : j = 1, ..., m\}$ é denominada afim por partes, ou seja, $\mathbb{R}^n$ é dividido em no máximo m regiões $\breve{f}$ afins.

4. *Função poliedral*: É aquela cuja epígrafe é um poliedro convexo fechado. Podemos dizer que é uma função de $\overline{\mathrm{Conv}}\mathbb{R}^n$ da forma $\breve{f} + i_P$, onde $\breve{f}$ é afim por partes e i_P é o indicador de um poliedro convexo fechado.

5. *Distâncias*: Seja C um conjunto convexo não vazio de $\mathbb{R}^n$ e $\|\cdot\|$ uma norma arbitrária. A função distância, definida por $d_C(x) \doteq \inf\{\|y - x\| : y \in C\}$, é também convexa.

6. *Formas quadráticas*: Seja $A : \mathbb{R}^n \to \mathbb{R}^n$ um operador linear simétrico. A função quadrática da forma $f(x) \doteq 1/2\langle Ax, x\rangle$ é convexa com domínio no $\mathbb{R}^n$ se, e somente se, A é positiva semi-definida. No caso de A positiva definida, f é estritamente convexa com módulo igual ao menor auto-valor de A.

7. *Soma dos maiores auto-valores de uma matriz*: Considere o espaço das matrizes simétricas $n \times n$ (denotada por $S_n(\mathbb{R})$) e seja $A \in S_n(\mathbb{R})$ com auto-valores $\lambda_1 \geq ... \geq \lambda_n$. A soma f_m dos m maiores auto-valores de A é dada por $f_m(A) \doteq \sup\{\mathrm{tr}(QQ^TA) : Q \in \Omega\}$, onde $\Omega \doteq \{Q : Q^TQ = I_m\}$ é o conjunto das matrizes que possuem m colunas ortonormais de dimensão n. Tal f_m é também convexo.

8. *Volume dos elipsóides*: Defina $S_n^{++}(\mathbb{R})$ [resp. $S_n^{+}(\mathbb{R})$] o espaço das matrizes positivas [resp. semi-] definidas. Considerando o espaço $S_n(\mathbb{R})$, a função definida por

$$A \mapsto f(A) \doteq \begin{cases} \log(\det A^{-1}) & \text{se } A \in S_n^{++}(\mathbb{R}) \\ +\infty & \text{caso contrário} \end{cases}$$

é convexa.

3.2 Operações que preservam convexidade

Vimos na secção 2.1 que podemos construir conjuntos convexos a partir de outros conjuntos convexos. Podemos, então, criar epígrafes convexas a partir de outras. Tal fato mostra que é natural contruirmos funções convexas a partir de outras (mais simples). Mostraremos a seguir quatro operações que tomam funções convexas fechadas e elaboram outras funções também convexas e fechadas:

1. *Combinação positiva de funções*: Sejam $f_1, ..., f_m$ em $\mathrm{Conv}\,\mathbb{R}^n$ [resp. $\overline{\mathrm{Conv}}\mathbb{R}^n$] e $t_1, ..., t_m$ números positivos. Suponha que exista um ponto onde todas as funções são finitas. Então, a função $f \doteq \sum_{i=1}^m t_i f_i$ também pertence a $\mathrm{Conv}\,\mathbb{R}^n$ [resp. $\overline{\mathrm{Conv}}\mathbb{R}^n$].

2. *Supremo de funções convexas*: Seja $\{f_i\}_{i\in I}$ uma família de funções convexas [resp. fechadas convexas]. Se existe x_0 tal que $\sup_I f_i(x_0) < +\infty$, então seu supremo (ponto a ponto) $f \doteq \sup\{f_i : i \in I\}$ pertence a $\mathrm{Conv}\,\mathbb{R}^n$ [resp. $\overline{\mathrm{Conv}}\mathbb{R}^n$].

   Em particular, seja uma função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ não identicamente $+\infty$, minorada por uma função afim. Então, a função f^* definida por

$\mathbb{R}^n \ni s \mapsto f^* \doteq \sup\{\langle s, x\rangle - f(x) : x \in \operatorname{dom} f\}$ – denominada *função conjugada* de f – pertence a $\overline{\operatorname{Conv}}\mathbb{R}^n$.

3. *Pré-composição com um mapeamento afim*: Seja $f \in \operatorname{Conv}\mathbb{R}^n$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}^n$] e seja A um mapeamento afim de $\mathbb{R}^m$ para $\mathbb{R}^n$ tal que $\operatorname{Im} A \cap \operatorname{dom} f \neq \emptyset$. Então, a função $f \circ A : \mathbb{R}^m \ni x \mapsto (f \circ A)(x) = f(A(x))$ está em $\operatorname{Conv}\mathbb{R}^m$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}^m$].

4. *Pós-composição com uma função convexa crescente*: Considere $f \in \operatorname{Conv}\mathbb{R}^n$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}^n$] e $g \in \operatorname{Conv}\mathbb{R}$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}$] crescente. Suponha que existe um $x_0 \in \mathbb{R}^n$ tal que $f(x_0) \in \operatorname{dom} g$, e considere $g(+\infty) \doteq +\infty$. Então, a função composta $g \circ f : x \mapsto g(f(x))$ pertence a $\operatorname{Conv}\mathbb{R}^n$ [resp. $\overline{\operatorname{Conv}}\mathbb{R}^n$].

Algumas operações, no entanto, não preservam a propriedade das funções continuarem fechadas. Outras devem e/ou podem ser forçadas a preservá-la. Eis a primeira dessas operações:

Proposição 3.3. *Se f está em $\operatorname{Conv}\mathbb{R}^n$, então sua perspectiva $\tilde{f}$ pertence a $\operatorname{Conv}\mathbb{R}^{n+1}$, onde $\tilde{f}$ é definida por:*

$$\tilde{f}(u,x) \doteq \begin{cases} uf(x/u) & \text{se } u > 0, \\ +\infty & \text{caso contrário.} \end{cases}$$

Trocando-se $\operatorname{Conv}\mathbb{R}^n$ acima por $\overline{\operatorname{Conv}}\mathbb{R}^n$, e tendo-se $x' \in \operatorname{ri}\operatorname{dom} f$, temos o fecho $\operatorname{cl}\tilde{f}$ de sua perspectiva dado por:

$$(\operatorname{cl}\tilde{f})(u,x) = \begin{cases} uf(x,u) & \text{se } u > 0, \\ \lim_{\alpha\downarrow 0} \alpha f(x' - x + x/\alpha) & \text{se } u = 0, \\ +\infty & \text{se } u < 0. \end{cases}$$

Outra operação interessante e bem usada é a convolução infimal:

Definição 3.6. *Sejam f_1 e f_2 duas funções de $\mathbb{R}^n$ em $\mathbb{R} \cup \{+\infty\}$. Sua* convolução infimal *é a função de $\mathbb{R}^n$ em $\mathbb{R} \cup \{\pm\infty\}$ definida por:*

$$\begin{aligned}(f_1 *_{\inf} f_2) &\doteq \inf\{f_1(x_1) + f_2(x_2) : x_1 + x_2 = x\} \\ &= \inf_{y\in\mathbb{R}^n}[f_1(y) + f_2(x-y)].\end{aligned}$$

A operação de convolução infimal é dita *exata* em $x = \overline{x}_1 + \overline{x}_2$ quando o mínimo (não necessariamente único) é alcançado em $(\overline{x}_1, \overline{x}_2)$.

Definição 3.7. *(Função imagem por transformação linear) Sejam $A : \mathbb{R}^m \to \mathbb{R}^n$ um operador linear e $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$. A imagem de g sobre A é a função $Ag : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ definida como $(Ag)(x) \doteq \inf\{g(y) : Ay = x\}$.*

Note que essa definição vem da função indicador: quando $g = i_C$, com $C \in \mathbb{R}^m$ não vazio, escrevemos:

$$(Ag)(x) = \begin{cases} 0 & \text{se } x = Ay \text{ para algum } y \in C, \\ +\infty & \text{caso contrário.} \end{cases}$$

Em outras palavras, $Ag = i_{A(C)}$ é a função indicador da imagem de C sobre A. Um exemplo de função imagem muito utilizada é a chamada *função marginal* de g. Esta é a imagem de g sobre o operador linear A projetando cada $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ em $x \in \mathbb{R}^n$, o que mostra que ela pertence a $\operatorname{Conv}\mathbb{R}^n$. Finalizaremos essa secção com a importante definição a seguir:

Definição 3.8. *Seja $g : \mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$, não identicamente $+\infty$, minorada por uma função afim. Dizemos que o casco convexo* $\operatorname{co} g$ *de g é:*

$$\begin{aligned} \operatorname{co} g(x) &\doteq \inf\{r : \langle x, r\rangle \in \operatorname{co}\operatorname{epi} g\} \\ &= \sup\{h(x) : h \in \operatorname{Conv}\mathbb{R}^n, h \leq g\} \\ &= \inf\{\textstyle\sum_{i=1}^k \alpha_i g(x_i) : k = 1, 2, ..., \alpha \in \Delta_k, \\ &\quad x_i \in \operatorname{dom} g, \textstyle\sum_{i=1}^k \alpha_i x_i = x\}. \end{aligned}$$

Equivalentemente, podemos definir o casco convexo fechado $\overline{\operatorname{co}} g$ (que coincide com o fecho $\operatorname{cl}\operatorname{co} g$*) de g como sendo*

$$\begin{aligned} \overline{\operatorname{co}} g(x) &\doteq \inf\{r : \langle x, r\rangle \in \overline{\operatorname{co}}\operatorname{epi} g\} \\ &= \sup\{h(x) : h \in \overline{\operatorname{Conv}}\mathbb{R}^n, h \leq g\} \\ &= \sup\{\langle s, x\rangle - b : \langle s, y\rangle - b \leq g(y) \text{ para todo } y \in \mathbb{R}^n\}. \end{aligned}$$

3.3 Comportamento local e global de funções convexas

Funções convexas são localmente Lipschitz no interior relativo de seu domínio. Isso não ocorre necessariamente na sua borda relativa.

Lema 3.1. *Seja $f \in \operatorname{Conv}\mathbb{R}^n$ e suponha que existam x_0, δ, m e M tais que $m \leq f(x) \leq M$ para todo $x \in B(x_0, 2\delta)$. Então, f é Lipschitz em $B(x_0, \delta)$, mais precisamente, para todo y e y' em $B(x_0, \delta)$, temos:*

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

Usando-se esse lema, podemos provar também o teorema abaixo:

Teorema 3.2. *Sejam $f \in \operatorname{Conv}\mathbb{R}^n$, S um conjunto convexo compacto de* $\operatorname{ri}\operatorname{dom} f$*. Então, existe $L = L(S) \geq 0$ tal que*

$$|f(x) - f(x')| \leq L\|x - x'\| \text{ para todo } x \text{ e } x' \text{em } S.$$

Segue em particular que f é contínua com restrição ao interior relativo de seu domínio, ou seja, para $x_0 \in \operatorname{ri}\operatorname{dom} f$ e $x \in \operatorname{ri}\operatorname{dom} f$ convergindo a x_0, temos que $f(x) \to f(x_0)$.

Considere agora as funções convexas $f_k : \mathbb{R}^n \to \mathbb{R}$ convergindo pontualmente (com $k \to +\infty$) para $f : \mathbb{R}^n \to \mathbb{R}$. Então, podemos provar que f é convexa e, para cada conjunto compacto S, a convergência de f_k para f é uniforme em S. Tendo estudado o comportamento de $f(x)$ quando x aproxima $\operatorname{rbd}\operatorname{dom} f$, falta-nos considerar o caso em que x é " ilimitado", ou seja, para seqüências com normas que convergem para o infinito.

Proposição 3.4. *Para $f \in \overline{\mathrm{Conv}}\mathbb{R}^n$, o cone assintótico de* $\operatorname{epi} f$ *é a epígrafe da função $f'_\infty \in \overline{\mathrm{Conv}}\mathbb{R}^n$, definido por*

$$d \mapsto f'_\infty(d) \doteq \sup_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} = \lim_{t \to +\infty} \frac{f(x_0 + td) - f(x_0)}{t},$$

onde x_0 é arbitrário no $\operatorname{dom} f$. *Tal f'_∞ é denominada* função assinótica.

A importância de funções assintóticas pode ser observada abaixo:

Proposição 3.5. *Seja $f \in \overline{\mathrm{Conv}}\mathbb{R}^n$. Todos os conjuntos de nível não vazios de f possuem o mesmo cone assintótico, que é o conjunto de nível de f'_∞ no nível zero, ou seja:*

$$\forall\, r \in \mathbb{R} \text{ com } S_r(f) \neq \emptyset, \qquad [S_r(f)]_\infty = \{d \in \mathbb{R}^n : f'_\infty(d) \leq 0\}.$$

Em particular, as seguintes afirmações são equivalentes:

(a) Existe r para o qual $S_r(f)$ é não vazio e compacto;

(b) Todos os conjuntos de nível de f são compactos;

(c) $f'_\infty(d) > 0$ para todo $d \in \mathbb{R}^n$ não nulo.

Uma função $f \in \overline{\mathrm{Conv}}\mathbb{R}^n$ que satisfaz (a), (b) ou (c) é dita $0-coersiva$. Em outras palavras, uma função $0-coersiva$ é aquela que cresce no infinito. Se f cresce no infinito mais rápido que qualquer função afim, f é dito $1-coersiva$.

Por fim, as funções assintóticas estão bem relacionadas com problemas de minimização. Seja $f \in \overline{\mathrm{Conv}}\mathbb{R}^n$. Então,

- Se $f'_\infty(d) < 0$ para algum d, então f é ilimitado por baixo.
- A condição $f'_\infty(d) > 0$ para todo $d \neq 0$ é necessária e suficiente para que f tenha um conjunto de pontos mínimos não vazio e limitado.
- Se $f'_\infty \geq 0$, com $f'_\infty(d) = 0$ para algum $d \neq 0$, não se garante existência de mínimos.

3.4 Diferenciabilidade de primeira e segunda ordem

Nesta secção, mostraremos a relação entre diferenciabilidade e convexidade em termos do gradiente da função.

Teorema 3.3. *Sejam f uma função diferenciável em um conjunto aberto $\Omega \subset \mathbb{R}^n$ e C um subconjunto convexo de Ω. Então,*

(a) f é convexa em C se, e somente se,

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle \text{ para todo } (x_0, x) \in C \times C;$$

(b) f é estritamente convexa em C se, e somente se, a desigualdade acima vale de maneira estrita para $x \neq x_0$;

(c) f é fortemente convexa com módulo c em C se, e só se, para todo $(x_0, x) \in C \times C$,

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2} c \|x - x_0\|^2.$$

A seguinte definição é essencial para o próximo teorema que mostraremos.

Definição 3.9. *Seja $C \subset \mathbb{R}^n$ convexo. O mapeamento $F : C \to \mathbb{R}^n$ é dito* monótono *[resp.* estritamente monótono*, resp.* fortemente monótono *com módulo $c > 0$] em C quando, para todo x e x' em C, temos:*

$$\begin{gathered} \langle F(x) - F(x'), x - x' \rangle \geq 0 \\ [\text{resp. } \langle F(x) - F(x'), x - x' \rangle > 0 \text{ quando } x \neq x' \\ \text{resp. } \langle F(x) - F(x'), x - x' \rangle \geq c\|x - x'\|^2] \end{gathered}$$

Note que com apenas uma variável, tal F é dito crescente. Além disso, se considerarmos $F = \nabla f$, temos a seguinte caracterização de função convexa: f é convexa [resp. estritamente convexa, resp. fortemente convexa com módulo c] em C se, e somente se, seu gradiente ∇f é monótono [resp. estritamente monótono, resp. fortemente monótono com módulo c] em C.

Uma função convexa não precisa ser diferenciável em todo o interior de seu domínio. Veremos tal propriedade com maiores detalhes na secção 5. Por ora, citaremos apenas que ela precisa ser diferenciável em "muitos pontos". Utilizando um resultado de Hademacher (1919), que diz que uma função localmente Lipschitz em um conjunto aberto é diferenciável em quase todo lugar de tal conjunto, e usando o lema 3.1, concluímos que o subconjunto do interior da função formado pelos pontos não diferenciáveis tem medida (de Lebesgue) nula.

Outro critério para reconhecimento de funções convexas usa derivadas segundas. Considere o seguinte teorema:

Teorema 3.4. *Seja f duas vezes diferenciável em um conjunto convexo aberto $\Omega \subset \mathbb{R}^n$. Então,*

(a) f é convexa em Ω se, e só se, $\nabla^2 f(x_0)$ é positiva semi-definida para todo $x_0 \in \Omega$.

(b) Se $\nabla^2 f(x_0)$ é positiva definida para todo $x_0 \in \Omega$, então f é estritamente convexa em Ω. Note que o contrário não é necessariamente verdade.

(c) f é fortemente convexa com módulo c em Ω se, e somente se, o menor auto-valor de $\nabla^2 f(\cdot)$ é minorado por c em Ω, ou seja, para todo $x_0 \in \Omega$ e todo $d \in \mathbb{R}^n$, temos $\langle \nabla^2 f(x_0)d, d\rangle \geq c\|d\|^2$.

4 Sublinearidade e Funções Suporte

Mostraremos agora os principais resultados associados a funções sublineares, uma generalização do conceito de linearidade da Análise Real clássica para a Análise Convexa. Tal estudo é motivado pela aproximação tangencial de quaisquer funções convexas por funções sublineares e pela simples correspondência que existe entre elas e conjuntos convexos fechados.

4.1 Funções sublineares

Definição 4.1. *Uma função $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é dita* sublinear *se é convexa e positivamente homogênea (de grau 1), ou seja, se $\sigma \in \operatorname{Conv} \mathbb{R}^n$ e $\sigma(tx) = t\sigma(x)$ para todo $x \in \mathbb{R}^n$ e $t > 0$.*

O seguinte resultado caracteriza geometricamente as funções sublineares:

Proposição 4.1. *Uma função $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é sublinear se, e somente se, sua epígrafe $\operatorname{epi} \sigma$ é um cone convexo não vazio em $\mathbb{R}^n \times \mathbb{R}$.*

Uma outra maneira de definir tais funções usa o conceito de subaditividade:

Definição 4.2. *Uma função σ é subaditiva se $\sigma(x_1 + x_2) \leq \sigma(x_1) + \sigma(x_2)$ para todo $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$.*

Eis, portanto, a nova proposição:

Proposição 4.2. *Uma função $\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, não identicamente $+\infty$, é sublinear se, e somente se, vale uma das seguintes propriedades: (i) $\sigma(t_1x_1 + t_2x_2) \leq t_1\sigma(x_1) + t_2\sigma(x_2)$ para todo $x_1, x_2 \in \mathbb{R}^n$ e $t_1, t_2 > 0$ ou (ii) σ é positivamente homogênea e subaditiva.*

Como corolário da propriedade acima, observamos que uma função sublinear σ satisfaz $\sigma(x) + \sigma(-x) \geq 0$ para todo $x \in \mathbb{R}^n$. Se vale a igualdade, então tal função torna-se linear, conforme pode ser comprovado na proposição abaixo.

Proposição 4.3. *Seja σ sublinear e suponha que existam $x_1, ..., x_m \in \operatorname{dom} \sigma$ tais que $\sigma(x_i) + \sigma(-x_i) = 0$ para $i = 1, ..., m$. Então, σ é linear no subspaço gerado por $x_1, ..., x_m$.*

É comum utilizarmos $U \doteq \{x \in \mathbb{R}^n : \sigma(x) + \sigma(-x) = 0\}$, o subspaço de $\mathbb{R}^n$ no qual σ é linear. Esse U é denominado *espaço de linearidade* de σ.

Citaremos agora alguns exemplos de funções sublineares:

1. O *indicador* e a *função distância* de um cone convexo são sublineares.

2. Uma *norma* de $\mathbb{R}^n$ é uma função sublinear.

3. *Semi-norma quadrática*: Seja Q um operador simétrico positivo semi-definido de $\mathbb{R}^n$ em $\mathbb{R}^n$. Definindo $f(x) \doteq \sqrt{\langle Qx, x\rangle}$ para todo $x \in \mathbb{R}^n$, temos que f é sublinear.

Um outro exemplo mais interessante pode ser dado pelo calibre de um conjunto convexo fechado, o qual é definido a seguir.

Definição 4.3. *(Calibre) Seja C um conjunto convexo fechado que contém a origem. A função γ_C definida por:*

$$\gamma_C(x) \doteq \inf\{\lambda > 0 : x \in \lambda C\}$$

é denominada calibre *de C. Consideramos que $\gamma_C(x) \doteq +\infty$ se não existe $\lambda > 0$ para $x \in \lambda C$.*

Geometricamente, γ_C pode ser obtido com um deslocamento de C no hiperplano $\mathbb{R}^n \times \{1\}$. Assim, a epígrafe de γ_C é o cone que passa por esse C deslocado. O próximo resultado oferece uma das principais propriedades de um calibre.

Teorema 4.1. *Considere C um conjunto convexo fechado que contém a origem. Então,*

(a) *seu calibre γ_C é uma função sublinear fechada não negativa;*

(b) *γ_C é finito em todo lugar se, e só se, zero está no interior de C;*

(c) $\{x \in \mathbb{R}^n : \gamma_C(x) \leq r\} = rC$ para todo $r > 0$ e $\{x \in \mathbb{R}^n : \gamma_C(x) = 0\} = C_\infty$, com C_∞ sendo o cone assintótico de C.

Um corolário do teorema acima é o seguinte: C é compacto (i.e., $C_\infty = \{0\}$) se, e somente se, $\gamma_C(x) > 0$ para todo $x \neq 0$.

Da mesma forma que as funções convexas, as funções sublineares podem ser combinadas de maneira a formarem novas funções sublineares. Por exemplo, o supremo de funções sublineares [fechados] também é sublinear [fechado], se não for identicamente $+\infty$. Além disso, se $\{\sigma\}_{i\in I}$ é uma família de funções sublineares minorados por uma função linear, então $\text{co}\{\inf_{i\in I} \sigma_i\}$ é sublinear. Se $I = \{1, ..., m\}$, podemos ainda obter a convolução infimal dessas funções: $\text{co}\min\{\sigma_1, ..., \sigma_m\} = \sigma_1 *_{\text{inf}} ... *_{\text{inf}} \sigma_m$.

4.2 Funções suporte

Definição 4.4. *Seja S um conjunto não vazio em $\mathbb{R}^n$. Uma função $\sigma_S : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ definida por*

$$\mathbb{R}^n \ni x \mapsto \sigma_S(x) \doteq \sup\{\langle s, x\rangle : s \in S\}$$

é denominada função suporte *de S.*

O conjunto S pode ser interpretado como um indexador: $\sigma_S(\cdot)$ é o supremo de lineares $\langle s, \cdot\rangle$ sobre S. Portanto, uma função suporte é, necessariamente, fechada e sublinear. O domínio de σ_S é um cone convexo, fechado ou não. De fato, se $x \in \text{dom}\,\sigma_S$, então S está contido em um semi-espaço oposto ao x. Ademais, dizemos que uma função suporte de S é finito em todo lugar se, e somente se, S é limitado. Considere agora um conceito muito utilizado relacionado a essas funções:

Definição 4.5. *A* largura *de um conjunto não vazio S ao longo de $x \neq 0$ é*

$$\sigma_S(x) + \sigma_S(-x) = \sup_{s\in S}\langle s, x\rangle - \inf_{s\in S}\langle s, x\rangle,$$

um número em $[0, +\infty]$. Ela vale zero se, e somente se, S está inteiramente em algum hiperplano afim ortogonal a x. Tal hiperplano é escrito como $\{y \in \mathbb{R}^n : \langle y, x\rangle = \sigma_S(x)\}$. A intersecção de todos esses hiperplanos é o casco afim de S.

Se x tem norma 1 e é interpretado como uma direção, a largura de S mede o quão "largo" S é ao longo de x. Por conveniência, troquemos a notação $\sigma(x)$ por $\sigma(d)$. Daremos agora duas interpretações geométricas que auxiliam a entender melhor o conceito de função suporte. Eis a primeira delas:

Interpretação 4.1. *(Construção no $\mathbb{R}^n$) Dados $S \subset \mathbb{R}^n$ e $d \neq 0$, considere para cada $r \in \mathbb{R}$ o semi-espaço fechado $H^-_{d,r} \doteq \{z \in \mathbb{R}^n : \langle z, d\rangle \leq r\}$. Se S for limitado na direção d, podemos encontrar r grande o suficiente para que $S \subset H^-_{d,r}$. O valor $\sigma_S(d)$ é, justamente, o menor desses r. Se, pelo contrário, S é ilimitado nessa direção, então $\sigma_S(d) = +\infty$.*

A outra interpretação vem de uma construção no $\mathbb{R}^{n+1}$:

Interpretação 4.2. *(Construção no $\mathbb{R}^{n+1}$) Tome o espaço $\mathbb{R}^n \times \mathbb{R}$ e desloque o conjunto S para $\mathbb{R}^n \times \{-1\}$. Considere o casco cônico convexo K_S gerado por esse S deslocado. Então, o cone polar $(K_S)^\circ$ de K_S é a epígrafe de σ_S.*

Ilustraremos agora algumas propriedades de funções suporte. Para um conjunto $S \subset \mathbb{R}^n$ não vazio, $\sigma_S = \sigma_{\mathrm{cl}\, S} = \sigma_{\mathrm{co}\, S} = \sigma_{\overline{\mathrm{co}} S}$. Isto significa que o conceito de função suporte não distingue um conjunto S de seu casco convexo fechado.

Teorema 4.2. *Para $S \subset \mathbb{R}^n$ não vazio e sua função suporte σ_S, temos que $s \in \overline{\mathrm{co}} S \Leftrightarrow [\langle s, d\rangle \leq \sigma_S(d)$ para todo $d \in X]$, onde X pode ser o espaço $\mathbb{R}^n$ inteiro, a bola unitária $B(0,1)$, sua borda ou $\mathrm{dom}\, \sigma_S$.*

Isso mostra que existe uma correspondência entre conjuntos convexos fechados e funções suporte. Ademais, pode-se também verificar se um ponto pertence a um conjunto convexo fechado. A função suporte filtra o interior, o interior relativo e o casco afim de conjuntos convexos fechados.

Teorema 4.3. *Seja $S \in \overline{\mathrm{Conv}} \mathbb{R}^n$ não vazio. Então,*

(a) $s \in \mathrm{aff}\, S$ se, e só se, $\langle s, d\rangle = \sigma_S(d)$ para todo d com $\sigma_S(d) + \sigma_S(-d) = 0$;

(b) $s \in \mathrm{ri}\, S$ se, e só se, $\langle s, d\rangle < \sigma_S(d)$ para todo d com $\sigma_S(d) + \sigma_S(-d) > 0$;

(c) $s \in \mathrm{int}\, S$ se, e somente se, $\langle s, d\rangle < \sigma_S(d)$ para todo $d \neq 0$.

Uma última propriedade é dada a seguir:

Proposição 4.4. *Seja $S \in \overline{\mathrm{Conv}} \mathbb{R}^n$ não vazio. Então, $\mathrm{cl}\, \mathrm{dom}\, \sigma_S$ é o polar do cone assintótico S_∞ de S e vice-versa.*

Exemplifiquemos agora o uso de funções suporte com cones: Seja K um cone convexo fechado no $\mathbb{R}^n$. Então,

$$\sigma_K(d) = \begin{cases} 0 & \text{se } \langle s, d\rangle \leq 0 \text{ para todo } s \in K, \\ +\infty & \text{caso contrário.} \end{cases}$$

Em outras palavras, σ_K é o indicador do cone polar K°. Como $K^{\circ\circ} = K$, a função suporte de K° é também o indicador de K, ou seja, temos: $i_K = \sigma_{K^\circ}$ e $\sigma_K = i_{K^\circ}$.

4.3 Relação entre conjuntos convexos e funções sublineares

Veremos nesta secção que qualquer função sublinear pode ser vista como função suporte de algum conjunto. Para isso, representaremos uma função convexa fechada por uma série de funções afins que a minoram.

Teorema 4.4. *Seja σ uma função fechada sublinear. Então, existe uma função linear que minora σ. Em outras palavras, σ é a função suporte de conjuntos fechados não vazios $S_\sigma \doteq \{s \in \mathbb{R}^n : \langle s, d\rangle \leq \sigma(d)$ para todo $d \in \mathbb{R}^n\}$.*

Juntando-se o teorema 4.2 com o anterior, temos uma bijeção entre conjuntos convexos fechados e funções sublineares fechadas. Em outras palavras, se σ é uma função sublinear fechada, então ela é uma função suporte de um conjunto convexo fechado S não vazio dado por $S = \{s : \langle s, d\rangle \leq \sigma(d)$ para todo $d \in X\}$, onde X pode ser o espaço $\mathbb{R}^n$ inteiro, uma bola unitária $B(0, 1)$, sua borda ou $\operatorname{dom}\sigma$.

No contexto de funções suporte, temos também a seguinte definição:

Definição 4.6. *Seja C um conjunto convexo fechado não vazio com função suporte σ. Para um $d \neq 0$, o conjunto $F_C(d) \doteq \{x \in C : \langle s, d\rangle = \sigma(d)\}$ é chamado* face exposta *de C associada a d.*

Podemos relacionar tal face com o cone normal de C: $x \in F_C(d)$ se, e somente se, $d \in N_C(x)$. Quando d descreve o conjunto de direções normalizadas, as faces expostas correspondentes descrevem exatamente a borda de C, ou seja, vale $\operatorname{bd} C = \cup\{F_C(d) : d \in X\}$, onde X pode ser o espaço $\mathbb{R}^n \setminus \{0\}$, a esfera unitária ou $\operatorname{dom}_C \setminus\{0\}$.

Veremos agora o conceito de norma dual:

Proposição 4.5. *Seja $B \doteq \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ e $B^* \doteq \{s \in \mathbb{R}^n : \langle s, x\rangle \leq \|x\|$ para todo $x \in \mathbb{R}^n\}$, onde $\|\cdot\|$ é uma norma arbitrária em $\mathbb{R}^n$. A função suporte de B e o calibre de B^* são a mesma função $\|\cdot\|^*$, definida por*

$$\|s\|^* \doteq \max\{\langle s, x\rangle : \|x\| \leq 1\}.$$

Além disso, $\|\cdot\|^*$ é uma norma no $\mathbb{R}^n$. A função suporte de sua bola unitária B^* e o calibre do conjunto suportado B são a mesma função $\|\cdot\|$, ou seja,

$$\|x\| = \max\{\langle s, x\rangle : \|s\|^* \leq 1\}.$$

Portanto, uma norma automaticamente define uma outra norma, seu dual, e a operação é simétrica. Outra importante norma é a norma quadrática, definida por $\|x\|_Q \doteq \sqrt{\langle Qx, x\rangle}$, onde Q é um operador simétrico positivo definido. A importância deles vem do fato de derivarem de produtos escalares no $\mathbb{R}^n$: $\langle x, y\rangle_Q \doteq \langle Qx, y\rangle$. A sua norma dual é dada por $(\|s\|_Q)^* = \sqrt{\langle s, Q^{-1}x\rangle} = \|s\|_{Q^{-1}}$.

Segue um outro resultado que envolve calibre e funções suporte:

Proposição 4.6. *Seja C um conjunto convexo fechado que contém a origem. Seu calibre γ_C é a função suporte de um conjunto convexo fechado que contém a origem, dado por:*

$$C^\circ \doteq \{s \in \mathbb{R}^n : \langle s, d\rangle \leq 1 \, \text{para todo} \, d \in C\},$$

o qual define o polar de C.

A versão dual da proposição acima é dada por:

Corolário 4.1. *Seja C um conjunto convexo fechado que contém a origem. Sua função de suporte σ_C é o calibre de C°.*

Com as definições e os resultados anteriores, falta-nos verificar agora os possíveis cálculos que podem ser realizados com funções suporte.

Teorema 4.5. *Sejam S_1 e S_2 conjuntos convexos fechados não vazios e σ_1, σ_2 suas funções suporte. Então,*

$$S_1 \subset S_2 \Leftrightarrow \sigma_1(d) \leq \sigma_2(d) \, \text{para todo} \, d \in \mathbb{R}^n.$$

Isto é uma generalização do resultado obtido no teorema 4.2. Os seguintes resultados, apesar de simples, são também úteis quando se trabalha com funções suporte:

Proposição 4.7. *(a) Sejam σ_1 e σ_2 funções suporte de conjuntos convexos fechados não vazios S_1 e S_2. Se t_1 e t_2 são positivos, então, $t_1\sigma_1 + t_2\sigma_2$ é a função suporte de $\mathrm{cl}(t_1S_1 + t_2S_2)$.*

(b) Seja $\{\sigma_i\}_{i\in I}$ funções suporte de uma família de conjuntos convexos fechados não vazios $\{S_i\}_{i\in I}$. Então, $\sup_{i\in I} \sigma_i$ é a função suporte de $\overline{\mathrm{co}}\{\cup S_i : i \in I\}$.

(c) Considere a mesma família de conjuntos acima e suas respectivas funções suporte. Se $S \doteq \bigcap_{i\in I} S_i \neq \emptyset$, então $\sigma_S = \overline{\mathrm{co}}\{\inf \sigma_i : i \in I\}$.

Outros resultados que vimos, mas que omitiremos aqui, envolvem operadores lineares, distância de Hausdorff, auto-valores maximais e poliedros convexos fechados.

5 Subdiferenciais de funções convexas finitas

A subdiferenciabilidade é uma espécie de generalização do conceito de diferenciabilidade comum. A importância das regras de cálculo crescem dentro da Análise Convexa. Algumas operações com funções convexas destroem a diferenciabilidade mas preservam a convexidade. Nesta secção trataremos de funções convexas de valores finitos.

5.1 Subdiferenciabilidade: definições e interpretações

Iniciaremos com o conceito de derivada direcional:

Definição 5.1. *A* derivada direcional *de uma função f em um ponto x na direção d é*

$$f'(x,d) \doteq \lim \left\{ \frac{f(x+td)-f(x)}{t} : t \downarrow 0 \right\} = \inf \left\{ \frac{f(x+td)-f(x)}{t} : t > 0 \right\}.$$

Para um x fixo, essa função $f'(x,\cdot)$ é finito e sublinear. Ainda, ela é uma função suporte, conforme pode ser observado na definição abaixo:

Definição 5.2. *O* subdiferencial *$\partial f(x)$ de uma função f em x é o conjunto convexo compacto não vazio de $\mathbb{R}^n$ cuja função suporte é $f'(x,\cdot)$, isto é,*

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

Um vetor $s \in \partial f(x)$ é dito subgradiente *de f em x.*

Naturalmente, a mudança do produto escalar acarreta na mudança de $\partial f(x)$. Ademais, todas as propriedades da correspondência de conjuntos convexos fechados (compactos) e funções sublineares finitas podem ser reformuladas por $\partial f(x)$ e $f'(x,\cdot)$. Por exemplo, a largura de $\partial f(x)$ ao longo de uma direção d normalizada é $f'(x,d) + f'(x,-d)$.

Por ser uma função convexa $d \mapsto f'(x,d)$ tem, por si só, derivadas direcionais e subdiferenciais. Tem-se particular interesse pelo caso $d = 0$:

Proposição 5.1. *A função sublinear finita $d \mapsto \sigma(d) \doteq f'(x,d)$ satisfaz*

$$\begin{cases} \sigma'(0,\delta) = f'(x,\delta) & \text{para todo } \delta \in \mathbb{R}^n; \\ \sigma(\delta) = \sigma(0) + \sigma'(0,\delta) = \sigma'(0,\delta) & \text{para todo } \delta \in \mathbb{R}^n; \\ \partial\sigma(0) = \partial f(x). \end{cases}$$

Uma segunda definição de subdiferencial mais direta, sem referência a diferenciabilidade, é dada abaixo:

Definição 5.3. *O subdiferencial de uma função f em x é o conjunto de vetores $s \in \mathbb{R}^n$ que satisfaz*

$$f(y) \geq f(x) + \langle s, y - x\rangle \qquad \text{para todo } y \in \mathbb{R}^n.$$

Tendo em vista a definição acima, podemos dizer que os elementos de $\partial f(x)$ são as inclinações dos hiperplanos que suportam a epígrafe de f em $(x, f(x)) \in \mathbb{R}^n \times \mathbb{R}$. Em termos de cones tangentes e normais, expressamos tal resultado da seguinte forma:

Proposição 5.2. *(Subgradiente em termos de cone tangente e cone normal)*

(a) Um vetor $s \in \mathbb{R}^n$ é um subgradiente de f em x se, e somente se, $(s, -1) \in \mathbb{R}^n \times \mathbb{R}$ é normal a $\operatorname{epi} f$ em $(x, f(x))$. Em outras palavras,

$$N_{\operatorname{epi} f}(x, f(x)) = \{(\lambda s, -\lambda) : s \in \partial f(x), \lambda \geq 0\}.$$

(b) O cone tangente do conjunto $\operatorname{epi} f$ em $(x, f(x))$ é a epígrafe da função $d \mapsto f'(x, d)$:

$$T_{\operatorname{epi} f}(x, f(x)) = \{(d, r) : r \geq f'(x, d)\}.$$

Nesse contexo, o conjunto de nível que passa por x, definido por $Sf(x) \doteq S_{f(x)}(f) = \{y \in \mathbb{R}^n : f(y) \leq f(x)\}$, possui uma interessante relação com o subdiferencial $\partial f(x)$:

Lema 5.1. *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $Sf(x)$ um conjunto de nível que passa por x. Então, $T_{Sf(x)}(x) \subset \{d : f'(x, d) \leq 0\}$.*

Podemos, então, caracterizar elementos tangenciais para conjuntos de nível:

Teorema 5.1. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa e suponha que $0 \notin \partial f(x)$. Com $Sf(x)$ sendo um conjunto de nível, temos:*

$$T_{Sf(x)}(x) = \{d \in \mathbb{R}^n : f'(x, d) \leq 0\};$$
$$\operatorname{int}(T_{Sf(x)}(x)) = \{d \in \mathbb{R}^n : f'(x, d) < 0\} \neq \emptyset.$$

E esse resultado pode ainda ser reformulado em termos de cones normais:

Teorema 5.2. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa e suponha que $0 \notin \partial f(x)$. Então, a direção d é normal para $Sf(x)$ em x se, e somente se, existe um $t \geq 0$ e algum $s \in \partial f(x)$ tal que $d = ts$, isto é, $N_{Sf(x)}(x) = \mathbb{R}^+ \partial f(x)$.*

A generalização do gradiente pelo subdiferencial pode ser assim explicada: se $\partial f(x)$ é igual a $\{\nabla f(x)\}$, o conjunto de nível $f(\cdot) = f(x)$ possui um hiperplano tangente a $(x, f(x))$, o cone $T_{Sf(x)}(x)$ equivale ao semi-espaço oposto a $\nabla f(x)$ e o cone $N_{Sf(x)}(x)$ é igual a $\mathbb{R}^+ \nabla f(x)$. Note também que quando $\partial f(x)$ se torna "maior", então $Sf(x)$ fica "menor" ao redor de x.

5.2 Propriedades locais de um subdiferencial

Estudaremos agora algumas propriedades do subdifencial $\partial f(x)$ de uma função convexa f para um dado x. A aproximação de primeira ordem nos diz que f' aproxima f em torno de x, ou seja, $f(x+h) = f(x)+f'(x,h)+o(\|h\|)$. Sabemos também que $f'(x,d) = \sup\{\langle s,d\rangle : s \in \partial f(x)\}$. Assim, para todo $d \in \mathbb{R}^n$, existe um $s_d \in \partial f(x)$ tal que

$$f(x+td) = f(x) + t\langle s_d, d\rangle + t\varepsilon_d(t) \text{ para } t \geq 0. \tag{2}$$

Nesse caso, $\varepsilon_d(t) \to 0$ quando $t \downarrow 0$. Um resultado importante é que a convergência em (2) é uniforme em d em qualquer conjunto limitado, e assim, ε_d pode ser tomada independentemente do d normalizado.

Lema 5.2. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa e $x \in \mathbb{R}^n$. Para todo $\varepsilon > 0$, existe $\delta > 0$ tal que $\|h\| \leq \delta$ implica $|f(x+h) - f(x) - f'(x,h)| \leq \varepsilon\|h\|$.*

Um corolário do lema acima é dado abaixo:

Corolário 5.1. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Para qualquer x, temos $f(x+h) = f(x) + \langle s,h\rangle + o(\|h\|)$ se uma das seguintes propriedades valem:*

$$s \in F_{\partial f(x)}(h) \Leftrightarrow h \in N_{\partial f(x)}(s) \Leftrightarrow s = p_{\partial f(x)}(s+h).$$

Quando definimos o conceito de "derivada" $D : \mathbb{R}^n \to \mathbb{R}$ associada a uma função $f : \mathbb{R}^n \to \mathbb{R}$ em $x \in \mathbb{R}^n$, pode-se considerar a propriedade $(f(x+h) - f(x) - D(h))/\|h\| \to 0$, com h tendendo a zero. Sabemos que no Cálculo clássico, D é linear, enquanto que na Análise Convexa D é sublinear. De fato, mesmo sem especificar a classe de funções a qual D pertence, muitas outras derivadas podem ser definidas, dependendo do tipo de convergência permitida para h. No ponto de vista de Gâteaux têm-se $h = td$, com d fixo no $\mathbb{R}^n$ e $t \to 0$. O ponto de vista de Fréchet, por sua vez, é dada para $h \to 0$.

Corolário 5.2. *Se a função convexa f é diferenciável (do ponto de vista de Gâteaux) em x, seu único subgradiente em x é o gradiente $\nabla f(x)$. Da mesma forma, se $\partial f(x)$ contém somente um elemento s, então f é diferenciável (do ponto de vista de Fréchet) em x, com $\nabla f(x) = s$.*

Uma outra caracterização de faces expostas pode ser dada a partir de subdiferenciais, conforme pode ser observado abaixo:

Proposição 5.3. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Para todo x e d em $\mathbb{R}^n$, temos $F_{\partial f(x)}(d) = \partial(f'(x,\cdot))(d)$.*

Um outro resultado obtido diretamente da definição de subdiferencial é dado a seguir.

Teorema 5.3. *Para uma função $f : \mathbb{R}^n \to \mathbb{R}$ convexa, as três afirmações seguintes são equivalentes:*

(a) f é minorado por x sobre $\mathbb{R}^n$, isto é, $f(x) \leq f(y)$ para todo $y \in \mathbb{R}^n$;

(b) $0 \in \partial f(x)$;

(c) $f'(x, d) \geq 0$ para todo $d \in \mathbb{R}^n$.

Concluiremos nossa secção com o teorema do valor médio para casos não diferenciáveis:

Teorema 5.4. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Dados dois pontos $x \neq y$ em $\mathbb{R}^n$, existe $t \in]0, 1[$ e $s \in \partial f(x_t)$ tal que $f(y) - f(x) = \langle s, y - x\rangle$. Em outras palavras,*

$$f(y) - f(x) \in \bigcup_{t \in]0,1[} \{\langle \partial f(x_t), y - x\rangle\}.$$

O teorema acima também pode ser assim escrito:

Teorema 5.5. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Para $x, y \in \mathbb{R}^n$, temos:*

$$f(y) - f(x) = \int_0^1 \langle \partial f(x_t), y - x\rangle \, dt.$$

5.3 Regras de cálculo com subdiferenciais

Para desenvolver as regras de cálculo com subdiferenciais, é importante lembrarmos dos cálculos com funções suporte. Eis os principais cálculos realizados com subdiferenciais:

1. *Combinação positiva de funções*: Sejam f_1, f_2 funções convexas de $\mathbb{R}^n$ em $\mathbb{R}$ e t_1, t_2 positivas. Então,

$$\partial(t_1 f_1 + t_2 f_2)(x) = t_1 \partial f_1(x) + t_2 \partial f_2(x)$$

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

2. *Pré-composição com um mapeamento afim*: Seja $A : \mathbb{R}^n \to \mathbb{R}^m$ um mapeamento afim ($Ax = A_0 x + b$ com A_0 linear e $b \in \mathbb{R}^m$) e seja $g \in \operatorname{Conv} \mathbb{R}^m$ finita. Então,

$$\partial(g \circ A)(x) = A_0^* \partial g(Ax)$$

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

3. *Pós-composição com uma função convexa crescente de várias variáveis*: Sejam $f_1, ..., f_m$ funções convexas de $\mathbb{R}^n$ em $\mathbb{R}$ e o mapeamento F tal que $\mathbb{R}^n \ni x \mapsto F(x) \doteq (f_1(x), ..., f_m(x)) \in \mathbb{R}^m$. Associe $\mathbb{R}^m$ com um um produto interno e seja $g : \mathbb{R}^m \to \mathbb{R}$ convexa e crescente para cada componente, isto é, $y^i \geq z^i$ para $i = 1, ..., m$ implica $g(y) \geq g(z)$. Então, para todo $x \in \mathbb{R}^n$,

$$\partial(g \circ F)(x) = \Big\{ \begin{array}{l} \sum_{i=1}^m \rho^i s_i : (\rho^1, ..., \rho^m) \in \partial g(F(x)), \\ s_i \in \partial f_i(x) \text{ para } i = 1, ..., m \Big\}. \end{array}$$

4. *Supremo de funções convexas*: Sejam J um conjunto indexador arbitrário e $\{f_j\}_{j \in J}$ um coleção de conjuntos convexos de $\mathbb{R}^n$ em $\mathbb{R}$. Dado $f(x) \doteq \sup\{f_j(x) : j \in J\} < +\infty$ para todo $x \in \mathbb{R}^n$, queremos computar seu subdiferencial. Considere $J(x) \doteq \{j \in J : f_j(x) = f(x)\}$ conjunto dos indexadores ativos em um dado x.

   Temos, então, que $\partial f(x) \supset \overline{\text{co}}\{\cup \partial f_j(x) : j \in J(x)\}$. O contrário vale com algumas suposições a mais, ou seja, se J é também compacto e $j \mapsto f_j(x)$ são semi-contínuas superiores para cada $x \in \mathbb{R}^n$, então, $\partial f(x) = \text{co}\{\cup \partial f_j(x) : j \in J(x)\}$.

   Um caso especial é quando cada f_j é diferenciável. Nesse caso, temos $\partial f(x) = \text{co}\{\nabla f_j(x) : j \in J(x)\}$.

5. *Imagem de uma função sobre um mapeamento linear*: Seja $g : \mathbb{R}^m \to \mathbb{R}$ convexa e $A : \mathbb{R}^m \to \mathbb{R}^n$ um operador linear. Sabemos que $\mathbb{R}^n \ni x \mapsto (Ag)(x) \doteq \inf\{g(y) : Ay = x\}$ é convexa. Considere o conjunto de minimizadores de Ag sendo $Y(x) \doteq \{y \in \mathbb{R}^m : Ay = x, g(y) = (Ag)(x)\}$. Suponha ainda que A seja sobrejetora. Tome x tal que $Y(x)$ não seja vazia. Então, para um $y \in Y(x)$ arbitrário, temos:

$$\partial(Ag)(x) = \{s \in \mathbb{R}^n : A^* \in \partial g(y)\} = (A^*)^{-1}(\partial g(y)).$$

   Além disso, se g é diferenciável para algum $y \in Y(x)$, então Ag é diferenciável em x.

5.4 Exemplo: relaxação lagrangeana

Três tipos de problemas foram desenvolvidos no livro: o maior auto-valor de uma matriz simétrica, a melhor aproximação de uma função contínua em um intervalo compacto, mínimos quadrados lineares parciais e relaxação lagrangeana. Tomaremos este último como nosso exemplo.

Dado um conjunto U e $n+1$ funções $c_0, c_1, ..., c_n$ de U a $\mathbb{R}$, temos o seguinte problema de otimização:

$$\begin{cases} \sup c_0(u), & u \in U \\ c_i(u) = 0, & \text{para } i = 1, ..., n. \end{cases}$$

A *função lagrangeana* associada a essa função, que depende de $u \in U$ e $x = (\xi^1, ..., \xi^n) \in \mathbb{R}^n$, é dada por:

$$L(u, x) \doteq c_0(u) + \sum_{i=1}^{n} \xi^i c_i(u).$$

Por sua vez, a *função dual* é dada por:

$$f(x) \doteq \sup\{L(x, u) : u \in U\}. \tag{3}$$

Temos, dessa forma, um problema de minimização em f. Tal função dual é convexa, sendo supremo de funções afins $L(\cdot, u)$. Além disso, $\nabla_x L(x, u) = c(u)$, com $c(u)$ sendo um vetor com componentes $c_i(u)$, $i = 1, ..., n$. Suponha que c é semi-contínua inferior. Suponha também que $U(x) \doteq \{x \in U : L(x, u) = f(x)\}$, conjunto dos ótimos de (3), é compacto para cada x. Temos, portanto, o seguinte subdiferencial:

$$\partial f(x) = \text{co}\{c(u) : u \in U(x)\}.$$

Um minimizador $\overline{x}$ de f é caracterizado pela seguinte condição: para algum inteiro positivo $p \leq n+1$, existe $u_1, ..., u_p$ em $U(\overline{x})$ e um conjunto de multiplicadores $\alpha = (\alpha_1, ..., \alpha_p) \in \Delta_p$ tais que

$$L(\overline{x}, u_k) = f(\overline{x}) \qquad \text{e} \qquad \sum_{k=1}^{p} \alpha_k c(u_k) = 0 \in \mathbb{R}^n.$$

Em particular, se $L(\overline{x}, \cdot)$ tem apenas um único máximo $\overline{u}$, então, $p = 1$, e assim $c(\overline{u}) = 0$. Para um x não ótimo, temos

$$\text{int}\, T_{Sf(x)}(x) = \{d \in \mathbb{R}^n : d^T c(u) < 0 \text{ para todo } u \in U(x)\}.$$

como sendo o conjunto dos d's que satisfazem $f'(x, d) < 0$, denominadas *direções de descida*.

5.5 O subdiferencial como uma função múltipla

Consideramos agora o caso em que o subdiferencial $\partial f(x)$ varia com o ponto x e com a função f. Agora, a monotonicidade tem uma formulação mesmo na ausência de diferenciabilidade:

Proposição 5.4. *O mapeamento subdiferencial é monótono quando, para todo x_1 e x_2 em $\mathbb{R}^n$,*

$$\langle s_2 - s_1, x_2 - x_1 \rangle \geq 0 \text{ para todo } s_i \in \partial f(x_i), i = 1, 2.$$

Utilizando-se desse conceito, podemos caracterizar funções fortemente convexas da seguinte forma:

Teorema 5.6. *Uma condição necessária e suficiente para que uma função convexa $f : \mathbb{R}^n \to \mathbb{R}$ seja fortemente convexa com módulo $c > 0$ em um conjunto convexo C é: para todo $x_1, x_2 \in C$,*

$$f(x_2) \geq f(x_1) + \langle s, x_2 - x_1 \rangle + \frac{c}{2}\|x_2 - x_1\|^2 \text{ para todo } s \in \partial f(x_1),$$

ou, equivalentemente,

$$\langle s_2 - s_1, x_2 - x_1 \rangle \geq c\|x_2 - x_1\|^2 \text{ para todo } s_i \in \partial f(x_i), i = 1, 2.$$

Da mesma forma, podemos caracterizar funções estritamente convexas:

Proposição 5.5. *Uma condição necessária e suficiente para que uma função convexa $f : \mathbb{R}^n \to \mathbb{R}$ seja estritamente convexa em um conjunto convexo C é: para todo $x_1, x_2 \in C$ com $x_1 \neq x_2$,*

$$f(x_2) > f(x_1) + \langle s, x_2 - x_1 \rangle \text{ para todo } s \in \partial f(x_1),$$

ou, equivalentemente,

$$\langle s_2 - s_1, x_2 - x_1 \rangle > 0 \text{ para todo } s_i \in \partial f(x_i), i = 1, 2.$$

Completaremos esta secção citando as principais propriedades de um subdiferencial:

1. Seja $f \in \mathbb{R}^n \to \mathbb{R}$ convexa. O gráfico de seu mapeamento subdiferencial é fechado em $\mathbb{R}^n \times \mathbb{R}^n$.

2. O mapeamento ∂f é localmente limitado, ou seja, a imagem $\partial f(B)$ de um conjunto limitado $B \subset \mathbb{R}^n$ é um conjunto limitado em $\mathbb{R}^n$.

3. O mapeamento subdiferencial de uma função convexa $f : \mathbb{R}^n \to \mathbb{R}$ é semi-contínua exterior para todo $x \in \mathbb{R}^n$, ou seja,

$$\forall \varepsilon > 0, \exists \delta > 0 : y \in B(x, \delta) \Rightarrow \partial f(y) \subset \partial f(x) + B(0, \varepsilon).$$

4. Para todo $f : \mathbb{R}^n \to \mathbb{R}$ convexa, a função $f'(\cdot, d)$ é semi-contínua superior, ou seja, para todo $x \in \mathbb{R}^n$,

$$f'(x, d) = \limsup_{y \to x} f'(y, d) \text{ para todo } d \in \mathbb{R}^n.$$

5. Seja $\{f_k\}$ uma seqüência de função convexas finitas que convergem ponto a ponto para $f : \mathbb{R}^n \to \mathbb{R}$ e seja $\{x_k\}$ uma seqüência que converge para $x \in \mathbb{R}^n$. Para todo $\varepsilon > 0$,

$$\partial f_k(x_k) \subset \partial f(x) + B(0, \varepsilon)$$

   para um k suficientemente grande.

6. Seja (f_k) uma seqüência de função convexas finitas diferenciáveis que convergem ponto a ponto para $f : \mathbb{R}^n \to \mathbb{R}$ também diferenciável. Então, ∇f_k converge para ∇f uniformemente em um conjunto compacto de $\mathbb{R}^n$.

6 Conclusão

Nesta monografia citamos os principais conceitos e resultados da Análise Convexa, tomando-se como base o livro [HUL01], de Jean-Baptiste Hiriart-Urruty e Claude Lemaréchal. Apesar do estudo envolver todos os capítulos do livro, demos importância aos capítulos A a D para a escrita da monografia, pois são esses que contêm, de fato, os conceitos necessários para a projeto do mestrado. Por esse motivo, e também pela falta de tempo, optamos por não incluir o capítulo E. Como trabalho futuro, estudaremos teoria de dualidade, relaxação lagrangeana, métodos de feixe [HUL93b, Sho85, SS01, Aus87], métodos de subgradientes [HUL93a] e algoritmo do volume [BA00].

Referências

[Aus87] Alfred Auslender. Numerical methods for nondifferentiable convex optimization. *Mathematical Programming Study*, (30):102–126, 1987. Nonlinear analysis and optimization (Louvain-la-Neuve, 1983).

[BA00] F. Barahona and R. Anbil. The volume algorithm: producing primal solutions with a subgradient method. *Mathematica Programming*, 87:385–399, 2000.

[BMS02] L. Bahiense, N. Maculan, and C. Sagastizábal. The volume algorithm revisited: relation with bundle methods. *Mathematical Programming*, 94:41–69, 2002.

[HUL93a] J.-B. Hiriat-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I: Fundamentals*. Springer, 1993.

[HUL93b] J.-B. Hiriat-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*. Springer, 1993.

[HUL01] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer Verlag, Heidelberg, 2001.

[Sho85] N. Z. Shor. *Minimization methods for non-differentiable functions*. Springer, 1985.

[SS01] Claudia A. Sagastizábal and Mikhail V. Solodov. On the relation between bundle methods for maximal monotone inclusions and hybrid proximal point algorithms. In *Inherently parallel algorithms in feasibility and optimization and their applications (Haifa, 2000)*, volume 8 of *Stud. Comput. Math.*, pages 441–455. North-Holland, Amsterdam, 2001.