OTIMIZAÇÃO E ANÁLISE CONVEXA: ASPECTOS TEÓRICOS E APLICAÇÕES *

Matheus Souza† Maria A. Diniz-Ehrhardt‡

Resumo

Otimização Convexa é uma classe especial de problemas de otimização matemática em que a função objetivo e o conjunto factível são ambos convexos. Existem diversos métodos robustos e eficientes, principalmente os métodos de pontos interiores, para resolver problemas desse tipo. Isso se dá devido à teoria bem desenvolvida e à grande frequência com que ocorrem na prática. Problemas clássicos de otimização convexa e suas aplicações serão analisados.

Abstract

Convex Optimization is a special class of problems in mathematical optimization where the cost function and the feasible set are both convex. There are several efficient and trustable methods, mainly interior–point methods, to solve problems like these. This is due to the well developed theory they are based on and to their frequent occurrence in practical applications. Classical problems of convex optimization and their applications are to be analysed.

*Este projeto contou com o apoio da FAPESP.
†FEEC-Unicamp. e-mail: ra071855@fee.unicamp.br
‡DMA-IMECC-Unicamp. e-mail: cheti@ime.unicamp.br

1 Introdução

Otimização convexa é um ramo especial da área de otimização matemática em que a função objetivo e o conjunto factível são ambos convexos. Problemas de otimização convexa surgem em diversas aplicações e já possuem uma teoria rica e desenvolvida, sendo esses os principais fatores que possibilitaram a criação de métodos robustos e eficientes para resolvê-los.

Embora a teoria matemática em que se baseia esta área seja relativamente antiga, foi a partir do final do século XX que houve um grande desenvolvimento no estudo de aplicações e na construção de métodos computacionais específicos para essa classe de problemas. A concepção dos métodos de pontos interiores para problemas de programação linear criou uma nova classe de algoritmos que foram generalizados e que hoje resolvem de forma rápida e confiável os problemas de otimização convexa.

As aplicações surgem em diversos ramos da engenharia, como sistemas de controle, análise e processamento de sinais, análise de circuitos, projeto de sistemas de energia elétrica, entre outros [3]. Os problemas clássicos de otimização convexa com respectivas aplicações serão discutidos e analisados, a fim de explorar esse importante ramo da programação não–linear.

Este relatório está organizado como se segue. Nas seções 2 e 3, descrevemos os principais resultados da análise convexa, sendo a primeira destinada ao estudo dos conjuntos convexos e a segunda ao estudo das funções convexas. Posteriormente, na seção 4 apresentamos o problema de otimização convexa e discutimos vários problemas clássicos que podem ser formulados dessa forma. Na seção 5 estudamos o conceito de dualidade e sua relação com os multiplicadores de Lagrange, finalizando com a análise das condições de qualificação e das condições de otimalidade. Na seção 6, comentamos todos os resultados explorados durante o relatório.

2 Convexidade: Conjuntos

Nesta seção discutiremos fatos básicos sobre a convexidade de conjuntos, enunciando os principais resultados teóricos e descrevendo alguns exemplos notáveis de conjuntos convexos.

2.1 Conjuntos Convexos e Afins

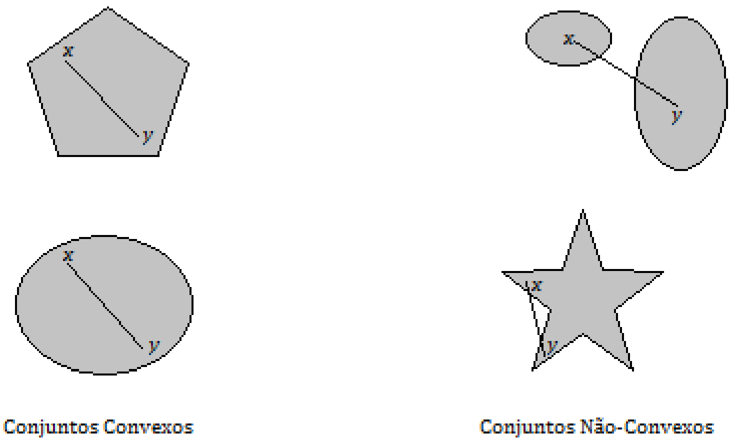
2.1.1 Conjuntos Convexos

Para iniciar nosso estudo sobre os conjuntos convexos, precisamos primeiramente defini-los precisamente.

Definição 2.1 *Um conjunto $S \subset \mathbb{R}^n$ é dito convexo se o segmento de reta entre quaisquer dois pontos em S está contido em S, i.e., se para quaisquer $x, y \in S$ e*

$\theta \in [0, 1]$, *verifica-se que* $\theta x + (1 - \theta)y \in S$.

Figura 1: Ilustração da definição de conjuntos convexos. À esquerda, temos dois exemplos de conjuntos convexos. À direita, dois conjuntos não–convexos



Baseando-se nessa definição, chamamos um ponto da forma $\theta_1 x_1 + \cdots + \theta_k x_k$, onde $\theta_1 + \cdots + \theta_k = 1$ e $\theta_i \geq 0$, $i = 1, \ldots, k$, uma *combinação convexa* dos pontos $x_1, \ldots, x_k$. Podemos considerar uma combinação convexa de pontos como a média aritmética ponderada desses pontos. Pode-se mostrar que um conjunto S é convexo se, e somente se, todas as combinações convexas de seus pontos está em S.

Exemplo 2.1 ***Caixas***. *Uma caixa, i.e., um conjunto da forma*

$$S = \{x \in I\!R^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \ldots, n\},$$

é um conjunto convexo. Para mostrar isso, tomemos $x, y \in S$, *de onde temos que* $\alpha_i \leq x_i$, $y_i \leq \beta_i$, $i = 1, \ldots, n$. *Então, para todo* $\theta \in [0, 1]$, *obtemos*

$$\theta\alpha_i \leq \theta x_i \leq \theta\beta_i$$

e

$$(1 - \theta)\alpha_i \leq (1 - \theta)y_i \leq (1 - \theta)\beta_i.$$

Somando-se ambas as desigualdades e simplificando chegamos a

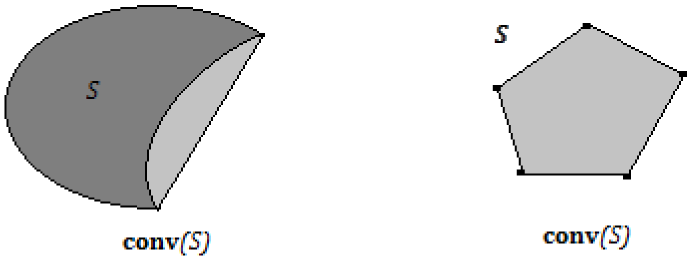
$$\alpha_i \leq \theta x_i + (1 - \theta)y_i \leq \beta_i.$$

Assim, $\theta x_i + (1 - \theta)y_i \in S$, *de onde concluímos que toda caixa é um conjunto convexo.*

Definição 2.2 *O envoltório (ou fecho) convexo de um conjunto $S \subset I\!R^n$, denotado por* $\mathbf{conv}(S)$*, é o conjunto de todas as combinações convexas dos pontos de S:*

$$\mathbf{conv}(S) = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in S, \theta_i \geq 0, i = 1, \ldots, k, \theta_1 + \cdots + \theta_k = 1\}. \quad (1)$$

Figura 2: Representação do fecho convexo de dois conjuntos. À esquerda, o conjunto inicial está representado pela cor mais escura, enquanto que seu fecho é formado por ambas as cores. À direita, o conjunto inicial era formado pelos cinco pontos destacados e o fecho corresponde à área preenchida.



Podemos ver facilmente que o envoltório convexo $\mathbf{conv}(S)$ é sempre um conjunto convexo, para qualquer conjunto S. Além disso, o envoltório convexo $\mathbf{conv}(S)$ é o menor conjunto convexo que contém S: Se C é um conjunto convexo qualquer que contém S, então $\mathbf{conv}(S) \subset C$.

A idéia de combinação convexa pode ser generalizada para incluir somas infinitas e integrais [3].

2.1.2 Conjuntos Afins

Uma classe muito importante de conjuntos e que está diretamente relacionada aos conjuntos convexos são os conjuntos (ou variedades) afins.

Definição 2.3 *Um conjunto $S \subset I\!R^n$ é afim se a reta que passa por quaisquer dois pontos de S está contida em S, i.e., se para quaisquer $x, y \in S$ e $\theta \in I\!R$ temos $\theta x + (1 - \theta y) \in S$.*

Esta idéia pode ser generalizada como fizemos com a convexidade de conjuntos: um ponto da forma $\theta_1 x_1 + \cdots + \theta_k x_k$, onde $\theta_1 + \cdots + \theta_k = 1$, é chamado de *combinação afim* dos pontos $x_1, \ldots, x_k$. Assim como no caso da convexidade, pode-se mostrar que um conjunto afim contém todas as combinações afins de seus pontos.

É importante perceber que qualquer variedade afim é um conjunto convexo, mas que o contrário nem sempre ocorre (vide Exemplo 2.1), de onde podemos concluir que a condição de ser afim é mais fraca que a de ser convexo.

Se S é um conjunto afim e $x_0 \in S$, então o conjunto

$$V = S - x_0 = \{x - x_0 \mid x \in S\} \tag{2}$$

é um subespaço (ou variedade linear) de $\mathbb{R}^n$. De fato, sejam $v_1, v_2 \in V$ e $\alpha, \beta \in \mathbb{R}$. Então, temos $v_1 + x_0 \in S$ e $v_2 + x_0 \in S$ e

$$\alpha v_1 + \beta v_2 + x_0 = \alpha(v_1 + x_0) + \beta(v_2 + x_0) + (1 - \alpha - \beta)x_0 \in S,$$

pois S é afim e $\alpha+\beta+(1-\alpha-\beta) = 1$. Assim, $\alpha v_1+\beta v_2 \in V$, já que $\alpha v_1+\beta v_2+x_0 \in S$.

Portanto, podemos escrever o conjunto afim S como

$$S = V + x_0 = \{v + x_0 \mid v \in V\}, \tag{3}$$

ou seja, como um subespaço somado a um deslocamento. O subespaço V é o subespaço associado ao conjunto S e é independente da escolha do elemento x_0. Com efeito, sejam $x_0, x_0' \in S$. Se $S = V + x_0$, então $x_0' = x_0 + v'$, com $v' \in V$, de onde $x_0 = x_0' - v'$. Assim, temos que

$$S = V + x_0 = V + x_0' - v' = (V - v') + x_0' = V + x_0',$$

já que $V - v' = V$ pois $v' \in V$ e V é variedade linear.

Exemplo 2.2 ***Solução de um Sistema de Equações Lineares****. A solução de um sistema de equações lineares, $S = \{x \mid Ax = b\}$, onde $A \in \mathbb{R}^{m\times n}$ e $b \in \mathbb{R}^m$ é um conjunto afim. De fato, sejam $x, y \in S$. Então, para todo $\theta \in \mathbb{R}$, temos:*

$$\begin{aligned} A(\theta x + (1-\theta)y) &= \theta Ax + (1-\theta)Ay = \\ &= \theta b + (1 - \theta b) = \\ &= b, \end{aligned}$$

o que mostra que a reta $\theta x + (1 - \theta y)$ está em S. O subespaço associado ao conjunto S é o núcleo da matriz A, $\mathcal{N}(A)$.

Nós definimos a *dimensão* de um conjunto afim qualquer como a dimensão de seu subespaço associado.

O conjunto de todas as combinações afins dos pontos de um conjunto $C \subset I\!R^n$ é chamado de *envoltório (ou fecho) afim* e denotado por $\mathbf{aff}(S)$:

$$\mathbf{aff}(S) = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_1, \ldots, x_k \in S, \theta_1 + \cdots + \theta_k = 1\}. \tag{4}$$

Assim como no caso convexo, facilmente percebe-se que o envoltório afim de qualquer conjunto S é sempre afim (e, portanto, convexo) e é o menor conjunto afim que contém S.

Chamamos de *dimensão afim* de um conjunto qualquer S como a dimensão de seu fecho afim. Se a dimensão afim de um conjunto $S \subset I\!R^n$ é menor que n, então o conjunto está contido em $\mathbf{aff}(S) \neq I\!R^n$. Definimos o *interior relativo* do conjunto S, denotado por $\mathbf{relint}(S)$, como o seu interior em relação a seu envoltório afim, ou seja,

$$\mathbf{relint}(S) = \{x \in S \mid B(x,r) \cap \mathbf{aff}(S) \subset S \textit{ para algum } r > 0\}, \tag{5}$$

onde $B(x,r) = \{y \mid ||y - x|| \leq r\}$ é a bola de centro x e raio r com respeito à norma $|| \cdot ||$. Podemos definir a *fronteira relativa* de um conjunto S como $\mathbf{cl}(S) \backslash \mathbf{relint}(S)$, onde $\mathbf{cl}(S)$ é o fecho do conjunto S.

Exemplo 2.3 *Considere o círculo em $I\!R^3$ definido por*

$$S = \{x \in I\!R^3 \mid x_2^2 + x_3^2 \leq 1, x_1 = 2\}.$$

Seu envoltório afim é dado por $\mathbf{aff}(S) = \{x \in I\!R^3 \mid x_1 = 2\}$. O interior do conjunto S é vazio, mas seu interior relativo é

$$\mathbf{relint}(S) = \{x \in I\!R^3 \mid x_2^2 + x_3^2 < 1, x_1 = 2\}.$$

A fronteira do conjunto em $I\!R^3$ é o próprio conjunto S, porém sua fronteira relativa é dada por

$$\mathbf{cl}(S) \backslash \mathbf{relint}(S) = \{x \in I\!R^3 \mid x_2^2 + x_3^2 = 1, x_1 = 2\}.$$

2.1.3 Cones

A terceira classe de conjuntos que iremos analisar e que é muito próxima às duas anteriores são os *cones*.

Definição 2.4 *Um conjunto S é dito um cone se para todo $x \in S$ e $\theta \geq 0$ temos que $\theta x \in S$. Um conjunto S é um cone convexo se ele for um conjunto convexo e um cone, ou seja, se $\forall x_1, x_2 \in S$ e $\theta_1, \theta_2 \geq 0$ tivermos*

$$\theta_1 x_1 + \theta_2 x_2 \in S.$$

Um ponto da forma $\theta_1 x_1 + \cdots + \theta_k x_k$ com $\theta_1, \ldots, \theta_k \geq 0$ é chamado de *combinação cônica* dos pontos $x_1, \ldots, x_k$. Um conjunto S é um cone convexo se, e somente se, ele contém todas as combinações cônicas de seus elementos. O *fecho cônico* de um conjunto S, denotado por $\mathbf{cone}(S)$ é o conjunto de todas as combinações cônicas dos elementos de S, *i.e.*,

$$\mathbf{cone}(S) = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in S, \theta_i \geq 0, i = 1, \ldots, k\}. \tag{6}$$

O fecho cônico de um conjunto S é o menor cone convexo que contém o conjunto S.

Agora, definiremos o cone dual para um cone qualquer. Seja S um cone. O conjunto

$$S^* = \{y \mid x^t y \geq 0, \ \forall x \in S\} \tag{7}$$

é chamado de *cone dual* de S. O cone dual S^* é um cone e é sempre convexo, mesmo que o cone original S não o seja. Essa e diversas outras propriedades a respeito dos cones duais que podem ser buscadas em [3].

2.1.4 Propriedades

Agora que apresentamos as classes de conjuntos com que vamos trabalhar, enunciaremos alguns fatos que são importantes para o ramo da Análise Convexa.

Primeiramente, podemos facilmente perceber que, para todo conjunto S, temos que $\mathbf{conv}(S) \subset \mathbf{aff}(S)$. A demonstração dessa propriedade é trivial já que o conjunto $\mathbf{conv}(S)$ contém todo segmento de reta entre dois pontos de S, enquanto que $\mathbf{aff}(S)$ contém toda reta que passa por dois pontos de S. Além disso, pode-se mostrar também que $\mathbf{aff}(S) = \mathbf{aff}(\mathbf{cone}(S)) = \mathbf{aff}(\mathbf{cl}(S))$.

O teorema a seguir formaliza os resultados já discutidos quando apresentamos os fechos convexo e cônico.

Teorema 2.1 ***(Teorema de Caratheodory)***
Seja S um subconjunto não–vazio de $I\!R^n$.

(a) Todo $x \in \mathbf{cone}(S)$ pode ser representado como uma combinação cônica de vetores $x_1, \ldots, x_m \in S$ que são linearmente independentes.

(b) Todo $x \in \mathbf{conv}(S)$ pode ser representado como uma combinação convexa de vetores $x_1, \ldots, x_m \in S$ tais que $x_2 - x_1, \ldots, x_m - x_1$ são linearmente independentes.

O teorema 2.1 é muito importante para demonstrar diversos resultados da Análise Convexa. Para a demonstração desse resultado e para uma extensão interessante, ver [2].

A proposição a seguir nos dá alguns fatos básicos a respeito do interior relativo de um conjunto.

Proposição 2.1 *Seja S um conjunto convexo não–vazio.*

(a) (Princípio do Segmento de Reta) Se $x \in$ **relint**(S) *e $\overline{x} \in$* **cl**(S)*, então todos os pontos no segmento de reta que conecta x a $\overline{x}$, exceto possivelmente $\overline{x}$, pertencem a* **relint**(S)*.*

(b) (O Interior Relativo é Não–Vazio) **relint**(S) *é um conjunto convexo não–vazio e tem o mesmo fecho afim que o conjunto S.*

(c) $x \in$ **relint**(S) *se, e somente se, todo segmento de reta em S que possui x como uma extremidade pode ser prolongado além de x sem sair de S (i.e., para todo $y \in S$, existe um $\gamma > 1$ tal que $x + (\gamma - 1)(x - y) \in S$).*

A demonstração dessa proposição também pode ser encontrada em [2]. Os fatos enunciados acima nos dão informações notáveis sobre a topologia do interior relativo de um conjunto. Essas noções são muito usadas no estudo de outros resultados da Análise Convexa.

2.2 Exemplos Notáveis

A seguir discutiremos diversos casos clássicos de conjuntos convexos. Alguns exemplos simples:

- O conjunto vazio, um conjunto unitário e o próprio espaço $\mathbb{R}^n$ são subconjuntos afins de $\mathbb{R}^n$.
- Uma reta é um conjunto afim e um segmento de reta qualquer é um conjunto convexo.
- Qualquer subespaço de $\mathbb{R}^n$ é afim e um cone convexo.
- Um *raio*, um conjunto da forma $\{x_0 + \theta v \mid \theta \geq 0\}$, onde $v \neq 0$, é convexo. Ele é um cone convexo se sua base for $x_0 = 0$.

2.2.1 Hiperplanos e Semi-espaços

Um *hiperplano* é um conjunto da forma

$$\mathcal{H} = \{x \mid a^t x = b\}, \tag{8}$$

onde $a \in \mathbb{R}^n$, $a \neq 0$ e $b \in \mathbb{R}$. Trata-se de um conjunto afim, já que o hiperplano é a solução de uma equação linear na variável x. Geometricamente, o vetor a é o *vetor normal* do hiperplano e a constante b determina seu deslocamento em relação à origem, pois podemos reescrever o conjunto como

$$\mathcal{H} = \{x \mid a^t(x - x_0) = 0\},$$

onde $x_0 \in \mathcal{H}$, ou seja, é tal que $a^t x_0 = b$.

Um hiperplano divide o espaço $I\!R^n$ em dois *semi-espaços*. Um semi-espaço fechado é um conjunto da forma

$$\mathcal{S} = \{x \mid a^t x \leq b\}, \tag{9}$$

onde $a \neq 0$. Semi-espaços são conjuntos convexos, mas não são afins. A fronteira do semi-espaço (9) é o hiperplano $\{x \mid a^t x = b\}$. O interior do semi-espaço (9), o conjunto $\{x \mid a^t x < b\}$, é chamado de semi-espaço aberto.

2.2.2 Bolas Euclidianas e Elipsóides

Uma *bola Euclidiana* em $I\!R^n$ tem a forma

$$B(x_0, r) = \{x \mid ||x - x_0||_2 \leq r\} \tag{10}$$

onde $r > 0$ e $||\cdot||_2$ denota a norma Euclidiana. O ponto x_0 é o *centro* da bola e o escalar r é seu *raio*. Uma bola Euclidiana é um conjunto convexo. De fato, sejam $x_1, x_2 \in B(x_0, r)$, ou seja, $||x_1 - x_0||_2 \leq r$ e $||x_2 - x_0||_2 \leq r$. Assim

$$\begin{aligned} ||\theta x_1 + (1-\theta)x_2 - x_0||_2 &= ||\theta(x_1 - x_0) + (1-\theta)(x_2 - x_0)||_2 \leq \\ &\leq \theta||x_1 - x_0||_2 + (1-\theta)||x_2 - x_0||_2 \leq \\ &\leq \theta r + (1-\theta) r = r. \end{aligned}$$

Um *elipsóide* é um conjunto da forma

$$\mathcal{E} = \{x \mid (x - x_0)^t P^{-1} (x - x_0) \leq 1\}, \tag{11}$$

onde P é uma matriz simétrica e definida positiva. Geometricamente, o vetor x_0 é o *centro* do elipsóide e os autovalores da matriz P determinam o tamanho de seus semi–eixos. Uma bola Euclidiana é um elipsóide com $P = r^2 I$. Uma forma mais geral de representar um elipsóide é dada por $\mathcal{E} = \{x_0 + Au \mid ||u||_2 \leq 1\}$, com A semidefinida positiva. Se $A = P^{1/2}$, essa forma é equivalente a (11). Se a matriz A for semidefinida positiva e singular, $\mathcal{E}$ é chamado de *elipsóide degenerado*. Todos os elipsóides, incluindo os casos degenerados, são conjuntos convexos.

2.2.3 Bolas e Cones Normais

A partir da idéia desenvolvida para a norma Euclidiana, podemos tomar uma norma qualquer $||\cdot||$ em $I\!R^n$ e definir uma *bola* de *centro* x_0 e *raio* r como o conjunto

$$\mathcal{B} = \{x \mid ||x - x_0|| \leq r\}. \tag{12}$$

Pode-se mostrar usando raciocínio análogo e propriedades gerais de normas que toda bola é um conjunto convexo.

Um conjunto relacionado é o *cone normal*. O cone normal associado à norma $||\cdot||$ é o conjunto

$$\mathcal{C} = \{(x,t) \mid ||x|| \leq t\} \subset \mathbb{R}^{n+1}. \tag{13}$$

Todo cone normal é um cone convexo.

Exemplo 2.4 ***Cone de Segunda Ordem***. *Um cone de segunda ordem é o cone normal para a norma Euclidiana, ou seja,*

$$\mathcal{C} = \{(x,t) \mid ||x||_2 \leq t\}.$$

O cone de segunda ordem também é chamado de cone quadrático e tem papel importante em problemas clássicos de Otimização Convexa.

2.2.4 Poliedros

Um *poliedro* é o conjunto–solução de um número finito de igualdades e desigualdades lineares:

$$\mathcal{P} = \{x \mid Ax \leq b,\ Cx = d\}, \tag{14}$$

onde $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^m$ e $d \in \mathbb{R}^p$. Conjuntos afins, semi-espaços e segmentos de reta são exemplos de poliedros. Um poliedro limitado é chamado de *politopo*.

Um *simplex* é um importante exemplo de poliedro. Suponhamos que os $k+1$ vetores $x_0, x_1, \ldots, x_k$ são tais que $x_1 - x_0, \ldots, x_k - x_0$ são linearmente independentes. O simplex determinado por esses pontos é dado por:

$$\mathcal{S} = \mathbf{conv}\{x_0, x_1, \ldots, x_k\} = \{\theta_0 x_0 + \cdots + \theta_k x_k,\ \theta_0 + \cdots + \theta_k = 1\}. \tag{15}$$

A dimensão afim desse simplex é k e por isso podemos chamá-lo de simplex k-dimensional em $\mathbb{R}^n$. Para descrever um simplex na forma (14) são necessárias algumas transformações [3].

2.3 Operações que Preservam a Convexidade

Como a convexidade é uma propriedade importante e que nos fornece vários resultados sobre um conjunto, devemos saber que operações a conservam para que esses resultados ainda possam ser usados.

2.3.1 Operações Básicas

Primeiramente, estudamos três operações fundamentais que preservam a convexidade, conforme está enunciado na proposição a seguir.

Proposição 2.2
(i) (Interseção) A interseção $\cap_{i \in I} S_i$ *de uma coleção de conjuntos convexos* $\{S_i \mid i \in I\}$, $I \subset I\!N$, *é um conjunto convexo.*
(ii) (Soma) O conjunto $S_1 + S_2 = \{v_1 + v_2 \mid v_1 \in S_1, v_2 \in S_2\}$, *onde* S_1 *e* S_2 *são conjuntos convexos, é convexo.*
(iii) (Produto por Escalar) O conjunto $\lambda S = \{\lambda x \mid x \in S\}$, *onde* S *é um conjunto convexo, é convexo.*

A demonstração dessa proposição envolve basicamente a definição de convexidade de conjuntos e pode ser encontrada em [2].

2.3.2 Transformações Afins

Uma *transformação* (ou *função*) $f : I\!R^n \to I\!R^m$ é dita *afim* se for da forma $f(x) = Ax + b$, onde $A \in I\!R^{m \times n}$ e $b \in I\!R^m$. Suponha que $S \subset I\!R^n$ é convexo e $f : I\!R^n \to I\!R^m$ é uma função afim. Então a imagem de S pela função f,

$$f(S) = \{f(x) \mid x \in S\},$$

é convexa. O mesmo ocorre com a imagem inversa por uma função afim.

Exemplos clássicos de transformações que podem ser feitas em um conjunto usando funções afins são: *escalamento*, *deslocamento*, *projeção*, *soma*, entre outros. Para uma análise detalhada de cada caso, ver [3].

As transformações afins e lineares são casos especiais das *funções lineares–fracionárias*. Uma função linear fracionária é uma função do tipo

$$f(x) = \frac{Ax + b}{c^t x + d}, \ \mathbf{dom}(f) = \{x \mid c^t x + d > 0\} \tag{16}$$

onde $A \in I\!R^{m \times n}$, $b \in I\!R^m$, $c \in I\!R^n$ e $d \in I\!R$. Se $c = 0$ e $d > 0$, o domínio de f é o espaço $I\!R^n$ e a função é afim.

Assim como no caso de funções afins, se um conjunto S é convexo, então sua imagem por uma função linear–fracionária também o é.

2.4 Hiperplanos de Suporte e de Separação

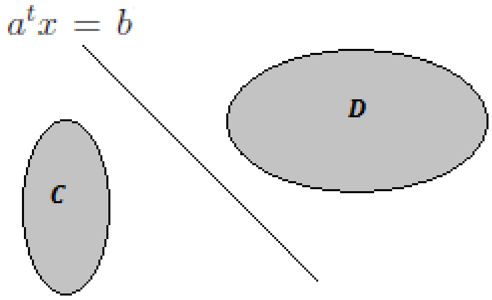
Para finalizar a seção de conjuntos convexos, vamos enunciar dois teoremas que são muito importantes na Análise Convexa e que usam as características fornecidas pela convexidade.

O primeiro teorema tem uma idéia simples: dois conjuntos convexos que são disjuntos são separáveis por hiperplanos.

Teorema 2.2 ***(Teorema do Hiperplano de Separação)*** *Suponha que os conjuntos C e D sejam convexos e tais que $C \cap D = \emptyset$. Então existem $a \neq 0$ e b tais que $a^t x \leq b$ para todo $x \in C$ e $a^t x \geq b$ para todo $x \in D$.*

O Teorema 2.2 nos diz que exite um hiperplano $\{x \mid a^t x = b\}$, chamado de *hiperplano de separação* dos conjuntos C e D, de forma que cada um esteja contido em um semi-espaço distinto.

Figura 3: Representação do hiperplano de separação dos conjuntos C e D.



Quando o hiperplano de separação satisfaz uma condição mais forte como $a^t x < b$ para todo $x \in C$ e $a^t x > b$ para todo $x \in D$, dizemos que esta é uma *separação estrita* dos conjuntos C e D.

A recíproca do Teorema 2.2, ou seja, a existência do hiperplano de separação implica que os conjuntos convexos C e D são disjuntos não é válida, a menos que novas restrições sobre os conjuntos C e D sejam feitas. Um simples contra–exemplo: sejam $C = D = \{0\} \subset I\!R$. Nesse caso, o hiperplano de separação de C e D é $x = 0$, mas evidentemente ambos se intersectam.

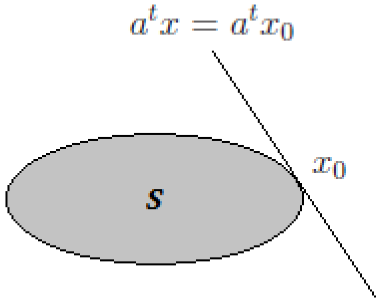
Outro teorema importante é o Teorema do Hiperplano de Suporte. Primeiramente, vamos definir o hiperplano de suporte.

Definição 2.5 *Suponha que $S \subset I\!R^n$ e x_0 é um ponto na fronteira de S, $\mathbf{bd}(S) = \mathbf{cl}(S)\backslash\mathbf{int}(S)$. Se $a \neq 0$ satisfaz $a^t x \leq a^t x_0$ para todo $x \in S$, então o hiperplano $\{x \mid a^t x = a^t x_0\}$ é chamado de hiperplano de suporte para S no ponto x_0.*

Note que a definição de hiperplano de suporte é equivalente à do hiperplano de separação de S e x_0. Geometricamente, o hiperplano de suporte é o hiperplano tangente ao conjunto S no ponto x_0. O teorema a seguir nos garante, com algumas condições, a existência do hiperplano de suporte.

Teorema 2.3 ***(Teorema do Hiperplano de Suporte)*** *Seja S um conjunto convexo não–vazio e x_0 um ponto qualquer de* $\mathbf{bd}(S)$. *Então existe um hiperplano de suporte para S no ponto x_0.*

Figura 4: Ilustração do hiperplano de suporte para o conjunto S em x_0.



Ao contrário do Teorema 2.2, o Teorema 2.3 possui uma recíproca parcial verdadeira: Se um conjunto é fechado, tem interior não–vazio e possui um hiperplano de suporte para todo ponto de sua fronteira, então ele é convexo [3].

Ambos os teoremas desta subseção estão demonstrados em [2,3].

3 Convexidade: Funções

Nesta seção apresentamos a idéia de convexidade para funções e, com isso, estudamos diversas propriedades que essas possuem.

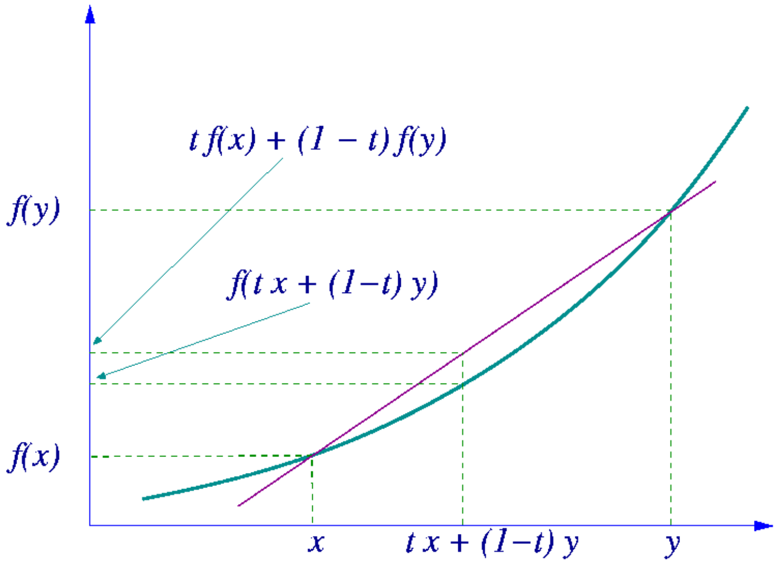
3.1 Funções Convexas

A noção de função convexa é definida abaixo e ilustrada na Figura 5.

Definição 3.1 *Seja S um subconjunto convexo de R^n. Uma função $f : S \to I\!R$ é dita convexa se*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \ \forall x, y \in S, \forall \alpha \in [0,1]. \qquad (17)$$

Figura 5: Representação de uma função $f : \mathbb{R} \to \mathbb{R}$ que é convexa.



Geometricamente, essa desigualdade significa que a *corda* de x para y fica sempre sobre o gráfico de f. Uma função é dita *estritamente convexa* se a desigualdade (17) é satisfeita estritamente, exceto nos pontos extremos x e y. Dizemos que uma função f é *côncava* se $-f$ é convexa e *estritamente côncava* se $-f$ é estritamente convexa.

Uma função afim sempre satisfaz a igualdade em (17). Portanto, toda função afim é convexa e côncava simultaneamente. A recíproca é válida: se f é convexa e côncava, então f é afim.

Se $f : S \to \mathbb{R}$ é uma função e γ é um escalar, então os conjuntos da forma $L_\gamma = \{x \in S \mid f(x) \leq \gamma\}$ são chamados de *conjuntos de nível*. Se f é convexa, então todos os seus conjuntos de nível são convexos. De fato, sejam $x, y \in S$ tais que $f(x) \leq \gamma$ e $f(y) \leq \gamma$, então para todo $\alpha \in [0, 1]$, temos $\alpha x + (1 - \alpha)y \in S$, já que S é convexo e

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

devido à convexidade de f. A convexidade de todos os conjuntos de nível não implica, no entanto, a convexidade da função: a função escalar $f(x) = \sqrt{|x|}$ tem conjuntos de nível convexos e não é convexa.

Muitas vezes é conveniente estender o domínio de uma função convexa a todo o espaço $\mathbb{R}^n$. Se f é convexa, definimos sua *extensão de valor estendido* $\tilde{f} : \mathbb{R}^n \to$

$\mathbb{R} \cup \{\infty\}$ como

$$\tilde{f}(x) = \begin{cases} f(x) & \text{se } x \in \mathbf{dom}(f) \\ \infty & \text{se } x \notin \mathbf{dom}(f). \end{cases}$$

Podemos sempre recuperar o domínio da função original f a partir de sua extensão $\tilde{f}$: $\mathbf{dom}(f) = \{x \mid \tilde{f}(x) < \infty\}$.

Usando a extensão de uma função convexa f, podemos simplificar a notação, não precisando mais declarar seu domínio. A convexidade da função original se preserva para sua função estendida [3].

Usaremos o mesmo símbolo para representar a função e sua extensão de valor estendido sempre que não houver problemas devido à ambiguidade. Assim, todas as funções convexas já estarão implicitamente estendidas.

3.2 Condições de Convexidade

Na subseção anterior, definimos funções convexas usando apenas a desigualdade (17). Nesta subseção vamos apresentar outras condições de convexidade em que são exigidas mais propriedades das funções.

3.2.1 Condições de Primeira–Ordem

As condições enunciadas no teorema a seguir garantem a convexidade de uma função quando se supõe que ela seja diferenciável.

Teorema 3.1 ***(Condições de Primeira–Ordem)*** *Suponha $f \in \mathcal{C}^1$. Então f é convexa se, e somente se, seu domínio* $\mathbf{dom}(f)$ *é convexo e*

$$f(y) \geq f(x) + \nabla f(x)^t (y - x) \tag{18}$$

é satisfeita $\forall x, y \in \mathbf{dom}(f)$.

Demonstração Primeiramente, suponhamos f convexa. Então para todo $\alpha \in [0, 1]$

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

Assim, para $\alpha \neq 0$,

$$\frac{f(x + \alpha(y - x)) - f(x)}{\alpha} \leq f(y) - f(x)$$

Fazendo $\alpha \to 0$, obtemos

$$\frac{f(x + \alpha(y - x)) - f(x)}{\alpha} = \partial_{(y-x)} f(x) = \nabla f(x)^t (y - x),$$

que é a derivada direcional de f na direção do vetor $y - x$. Portanto,

$$\nabla f(x)^t(y - x) \leq f(y) - f(x).$$

Para demonstrar a segunda parte, assumimos

$$f(y) \geq f(x) + \nabla f(x)^t(y - x),\ \forall x, y \in \mathbf{dom}(f).$$

Sejam $x_1, x_2 \in \mathbf{dom}(f)$ e seja $x = \alpha x_1 + (1 - \alpha)x_2$, $\alpha \in [0, 1]$. Quando $y = x_1$ ou $y = x_2$, temos

$$f(x_1) \geq f(x) + \nabla f(x)^t(x_1 - x), \tag{19}$$

$$f(x_2) \geq f(x) + \nabla f(x)^t(x_2 - x). \tag{20}$$

Multiplicando (19) por α e (20) por $(1-\alpha)$ e somando as duas inequações, obtemos

$$\alpha f(x_1) + (1 - \alpha)f(x_2) \geq f(x) + \nabla f(x)^t[\alpha x_1 + (1 - \alpha)x_2 - x].$$

Substituindo $x = \alpha x_1 + (1 - \alpha)x_2$, temos

$$\alpha f(x_1) + (1 - \alpha)f(x_2) \geq f(\alpha x_1 + (1 - \alpha)x_2). \quad \square$$

O Teorema 3.1 é um resultado muito importante pois nos garante que a aproximação de Taylor de primeira–ordem em qualquer ponto é uma cota inferior para a função em todo o seu domínio. Assim, por exemplo, a desigualdade (18) nos mostra que se $\nabla f(x) = 0$, então para todo $y \in \mathbf{dom}(f)$, $f(y) \geq f(x)$, ou seja, x é minimizador global da função f.

3.2.2 Condições de Segunda–Ordem

As condições expostas pelo resultado a seguir nos garantem a convexidade, supondo que f tem derivada de segunda ordem contínua. Em alguns casos, as condições a seguir são mais fáceis de serem verificadas na prática, porém são mais restritivas que as demais.

Teorema 3.2 ***(Condições de Segunda–Ordem)*** *Seja $f \in \mathcal{C}^2$ e tal que $\mathbf{dom}(f)$ é não–vazio. Então f é convexa se, e somente se, sua matriz Hessiana é semidefinida positiva em todo o seu domínio, ou seja, $\nabla^2 f(x) \geq 0$, $\forall x \in \mathbf{dom}(f)$.*

Demonstração Pelo Teorema de Taylor, temos

$$f(y) = f(x) + \nabla f(x)^t(y - x) + \frac{1}{2}(y - x)^t\nabla^2 f(x + \alpha(y - x))(y - x), \tag{21}$$

para algum $\alpha \in [0, 1]$. Se a Hessiana é semidefinida positiva, então

$$f(y) \geq f(x) + \nabla f(x)^t(y - x), \tag{22}$$

de onde, observando as condições de primeira–ordem, podemos concluir que f é convexa.

Para demonstrar a segunda parte, partimos para a contrapositiva. Suponhamos que a Hessiana não seja semidefinida positiva em algum $x \in \mathbf{dom}(f)$. Devido à continuidade da Hessiana, podemos assumir sem perda de generalidade que o ponto x pertence ao interior de $\mathbf{dom}(f)$. Então existe y tal que $(y-x)^t\nabla^2 f(x)(y-x) < 0$. Novamente, pela continuidade da Hessiana, podemos escolher y tal que para todo $\alpha \in [0,1]$,

$$(y-x)^t\nabla^2 f(x+\alpha(y-x))(y-x) < 0.$$

Com isso e usando (21), temos que (22) não é satisfeita, concluindo que f não é convexa. □

O Teorema 3.2 é importante por nos fornecer a idéia geométrica de que uma função apenas é convexa se sua concavidade é voltada para cima em todo ponto. Se a matriz for definida positiva é simples verificar as condições acima usando a fatoração de Cholesky, por exemplo. Se a matriz for semidefinida positiva singular ou definida positiva mal–condicionada, a fatoração de Cholesky irá falhar e algum outro método deve ser buscado.

3.3 Exemplos

A seguir, citamos alguns exemplos notáveis de funções convexas.

Exemplo 3.1 ***Função Afim***. *Como já foi mencionado anteriormente, é fácil ver que toda função afim é côncava e convexa ao mesmo tempo.*

Exemplo 3.2 ***Normas***. *Toda norma em $I\!R^n$ é convexa. De fato, sejam $||\cdot|| : I\!R^n \to I\!R$ uma norma qualquer e $\theta \in [0,1]$. Então,*

$$||\theta x + (1-\theta)y|| \leq ||\theta x|| + ||(1-\theta)y|| = \theta||x|| + (1-\theta)||y||,$$

onde usamos propriedades simples de normas, como a desigualdade triangular.

Exemplo 3.3 ***Função Máximo***. *A função* $f(x) = \max_i\{x_i\}$ *é convexa em* $\mathbb{R}^n$*. Com efeito, se* $\theta \in [0,1]$*, temos*

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max_i\{\theta x_i + (1-\theta)y_i\} \leq \\ &\leq \theta \max_i\{x_i\} + (1-\theta)\max_i\{y_i\} = \\ &= \theta f(x) + (1-\theta)f(y). \end{aligned}$$

Exemplo 3.4 ***Função log-sum-exp***. *A função* $f(x) = \ln(e^{x_1} + \cdots + e^{x_n})$ *é convexa em* $\mathbb{R}^n$*. De fato, se* $f(x) = \ln(\sum_i e^{x_i})$*, temos o vetor gradiente de* f

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}\ ,\ \cdots\ ,\ \frac{\partial f}{\partial x_n}\right)^t = \frac{1}{\sum_i e^{x_i}}\,(e^{x_1}\ ,\ \cdots\ ,\ e^{x_n})^t\,.$$

Simplificando a notação,

$$\nabla f(x) = \frac{1}{\mathbf{1}^t z} z,$$

onde $z = (e^{x_1}\ ,\ \cdots\ ,\ e^{x_n})^t$ *e* $\mathbf{1}$ *é o vetor de* $\mathbb{R}^n$ *cujas componentes são todas iguais a 1.*

Para determinar a Hessiana, calculamos seus elementos

$$\frac{\partial^2 f(x)}{\partial x_i^2} = \frac{e^{x_i}\sum_k e^{x_k} - e^{2x_i}}{(\sum_k e^{x_k})^2}$$

e

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = -\frac{e^{x_i}e^{x_j}}{(\sum_k e^{x_k})^2},$$

de onde obtemos

$$\nabla^2 f(x) = \frac{1}{(\mathbf{1}^t z)^2}\left[(\mathbf{1}^t z)\mathbf{diag}(z) - zz^t\right].$$

A Hessiana é semidefinida positiva para todo x*, já que, sendo* $v \in \mathbb{R}^n$*, temos*

$$v^t \nabla^2 f(x) v = \frac{1}{(\mathbf{1}^t z)^2}\left[(\sum_k z_k)(\sum_k v_k^2 z_k) - (\sum_k v_k z_k)^2\right] \geq 0,$$

onde usamos a desigualdade de Cauchy-Schwarz ($||a||_2^2\,||b||_2^2 \geq (a^t b)^2$), com $a_i = v_i\sqrt{z_i}$, $b_i = \sqrt{z_i}$.

Essa função é muito importante por ser uma aproximação diferenciável para a função máximo, mas apresenta problemas computacionais se as componentes do vetor x se tornarem muito grandes.

Existem muitos outros exemplos clássicos de funções convexas e que podem ser encontrados em qualquer referência da área. Em [3], Boyd *et. al.* enumeram um bom número de exemplos além dos discutidos aqui.

3.4 Epigrafo

O gráfico de uma função $f : I\!R^n \to I\!R$ é definido como o conjunto

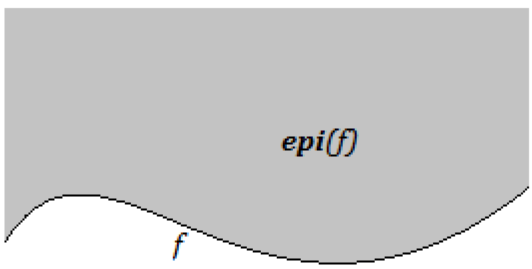
$$\{(x, f(x)) \mid x \in \mathbf{dom}(f)\} \subset I\!R^{n+1}. \tag{23}$$

Assim, a partir dessa idéia, definimos o *epigrafo* da função $f : I\!R^n \to I\!R$ como o conjunto

$$\mathbf{epi}(f) = \{(x, t) \mid x \in \mathbf{dom}(f)\ ,\ f(x) \leq t\} \subset I\!R^{n+1}, \tag{24}$$

ou seja, o epigrafo de uma função é o conjunto de pontos em $I\!R^{n+1}$ que está acima da função.

Figura 6: Ilustração da definição do epigrafo de uma função $f : I\!R \to I\!R$.



A primeira ligação entre os conceitos de convexidade para funções e para conjuntos é feita através do epigrafo: uma função é convexa se, e somente se, seu epigrafo é convexo.

Vários dos resultados discutidos anteriormente para funções convexas podem ter outra interpretação geométrica com a noção de epigrafo. Por exemplo, a condição de primeira–ordem para funções convexas (18) pode ser vista como o Teorema do hiperplano de suporte (Teorema 2.3) aplicado para o conjunto convexo $\mathbf{epi}(f)$.

3.5 Operações que Preservam a Convexidade

De forma análoga à feita na seção anterior, vamos enunciar uma série de operações que preservam a convexidade das funções.

3.5.1 Combinações Não–negativas

É evidente que, se f é uma função convexa e α é um escalar não–negativo, então a função αf é convexa. Da mesma forma, se f_1 e f_2 são funções convexas, então a função $f_1 + f_2$ também o é. Combinando ambas as idéias, pode-se provar que uma combinação não–negativa de funções convexas é convexa, ou seja, se $f_1, \ldots, f_n$ são funções convexas e $w_1, \ldots, w_n \geq 0$, então a função

$$w_1 f_1 + \cdots + w_n f_n$$

é convexa. Analogamente, a combinação não–negativa de funções côncavas é uma função côncava.

A partir desse resultado, percebemos que o conjunto de funções convexas é um cone convexo, por possuir todas as combinações cônicas de seus elementos.

Esse raciocínio pode ser expandido para somas infinitas e integrais.

3.5.2 Composição por uma Transformação Afim

Seja $f : I\!\!R^n \to R$, $A \in I\!\!R^{n \times m}$ e $b \in I\!\!R^n$. Definindo $g : I\!\!R^m \to I\!\!R$ como

$$g(x) = f(Ax + b), \tag{25}$$

temos que, se f é convexa, então g é convexa; se f é côncava, então g é côncava.

3.5.3 Máximo e Supremo

Se f_1 e f_2 são funções convexas e limitadas superiormente, então a função $f(x) = \max\{f_1(x)\ ,\ f_2(x)\}$ também é convexa. Esta propriedade é facilmente verificável e pode ser expandida. Se $f_1, \ldots, f_n$ são funções convexas, então

$$f(x) = \max_i \{f_i(x)\} \tag{26}$$

também é convexa.

Essa idéia pode ser mantida para o supremo: se $f_1, \ldots, f_n(x)$, $f_i : I\!R^n \to I\!R \cup \{\infty\}$, são convexas (estendidas), então a função $g : I\!R^n \to I\!R \cup \{\infty\}$, definida por

$$g(x) = \sup_i \{f_i(x)\} \tag{27}$$

também é convexa. Podemos verificar facilmente esse resultado argumentando que $\mathbf{epi}(g) = \cap_i \mathbf{epi}(f_i)$. Assim, como as funções f_i são convexas, seus epigrafos são convexos e, portanto, $\mathbf{epi}(g)$ é convexo, de onde concluímos que g é convexa.

3.5.4 Composição

Agora analisaremos as condições que devem ser impostas a $h : I\!R \to I\!R$ e $g : I\!R^n \to I\!R$ para que a função $f = h \circ g : I\!R^n \to I\!R$ definida como

$$f(x) = h(g(x)), \tag{28}$$

seja convexa.

Sendo $\tilde{h}$ a extensão de valor estendido da função h, temos no caso geral as condições:

- f é convexa se h é convexa, $\tilde{h}$ é não–decrescente e g é convexa;
- f é convexa se h é convexa, $\tilde{h}$ é não–crescente e g é côncava;
- f é côncava se h é côncava, $\tilde{h}$ é não–decrescente e g é côncava;
- f é côncava se h é côncava, $\tilde{h}$ é não–crescente e g é convexa.

As condições acima podem ser deduzidas a partir do caso particular em que $n = 1$, $h, g \in \mathcal{C}^2$ e $\mathbf{dom}(h) = \mathbf{dom}(g) = I\!R$. Os resultados podem também ser provados a partir da definição de funções convexas [3].

3.6 Funções Quase–Convexas

Uma função $f : S \to I\!R$, $\mathbf{dom}(f) = S \subset I\!R^n$, é dita *quase–convexa* se seu domínio e todos os seus conjuntos de nível

$$L_\alpha = \{x \in \mathbf{dom}(f) \mid f(x) \leq \alpha\},$$

para todo $\alpha \in I\!R$ são convexos. Uma função f é *quase–côncava* se $-f$ é quase–convexa. Uma função que é ao mesmo tempo quase–côncava e quase–convexa é dita *quase–linear*.

Exemplo 3.5 ***(Função Linear–Fracionária)*** *A função*

$$f(x) = \frac{a^t x + b}{c^t x + d},$$

com $\mathbf{dom}(f) = \{x \mid c^t x + d > 0\}$ *é quase–linear. Os conjuntos de nível*

$$L_\alpha = \{x \mid c^t x + d > 0, (a^t x + b)/(c^t x + d) \leq \alpha\}$$

podem ser reescritos da forma

$$L_\alpha = \{x \mid c^t x + d > 0, (a^t x + b) \leq \alpha(c^t x + d)\},$$

sendo todos convexos já que são interseções de semi–espaços. Assim, f é quase–convexa. De forma análoga, pode-se mostrar que $-f$ é quase–convexa.

Outra forma de definir uma função quase convexa é dada a seguir: uma função f é quase convexa se, e somente se, $\mathbf{dom}(f)$ é convexo e $\forall x, y \in \mathbf{dom}(f)$ e $\theta \in [0,1]$

$$f(\theta x + (1-\theta)y) \leq \max\{f(x)\ ,\ f(y)\}. \tag{29}$$

Funções quase–convexas têm papel importante em algumas classes de problemas e apresentam características muito parecidas com as funções convexas. Em [3], pode-se encontrar mais exemplos de funções quase–convexas, além de condições de quase–convexidade que exigem diferenciabilidade, sendo essas análogas às condições de convexidade exploradas na Subseção 3.2.

4 Otimização Convexa

Durante esta seção falaremos de problemas de Otimização Convexa, dessa forma usaremos vários termos e fatos básicos de Otimização e estes podem ser buscados em qualquer livro da área, como a referência [5].

4.1 Problemas de Otimização Convexa

Um *problema de Otimização Convexa* tem a forma

$$\begin{aligned} Minimizar \quad & f(x) \\ Sujeita\,a \quad & g(x) \leq 0, \\ & Ax = b, \end{aligned} \tag{30}$$

onde $f : I\!R^n \to I\!R$ e cada componente de $g : I\!R^n \to I\!R^m$, *i.e.*, $g_i : I\!R^n \to I\!R$, $i = 1, ..., m$, são funções convexas, $A \in I\!R^{p\times n}$ e $b \in I\!R^p$.

Podemos ver que a região factível é um conjunto convexo, já que, se $\mathcal{S}$ é o conjunto factível, temos que $\mathcal{S} = \{x \mid g(x) \leq 0\} \cap \{x \mid Ax = b\}$, que é a interseção de conjuntos convexos. Assim, em um problema de Otimização Convexa, minimizamos uma função convexa sobre um conjunto convexo.

4.1.1 Otimizadores Locais e Globais

A grande propriedade que problemas de Otimização Convexa possuem é que todo minimizador local também é global. De fato, seja x^* um minimizador local do problema (30), ou seja, x^* é factível e

$$f(x^*) = \inf\{f(z) \mid z \in \mathcal{S},\ ||z - x^*||_2 \leq \varepsilon\}, \tag{31}$$

para algum $\varepsilon > 0$. Suponhamos que x^* não é minimizador global, *i.e.*, existe y factível tal que $f(y) < f(x^*)$. Evidentemente que $||y - x^*|| > \varepsilon$, já que em caso contrário $f(x^*) \leq f(y)$. Considere o ponto z dado por

$$z = (1-\theta)x^* + \theta y, \quad \theta = \frac{\varepsilon}{2||y - x^*||_2}.$$

Então, temos $||z - x^*|| = \varepsilon/2 < \varepsilon$, e, por convexidade de $\mathcal{S}$, z é factível. Por convexidade de f,

$$f(z) \leq (1-\theta)f(x^*) + \theta f(y) < f(x^*),$$

contradizendo (31). Então, não existe y factível com $f(y) < f(x^*)$, de onde temos que x^* é otimizador global de f em $\mathcal{S}$.

A seguir, enunciamos um critério de otimalidade para função objetivo diferenciável.

Teorema 4.1 ***(Critério de Otimalidade para Função Objetivo Diferenciável)*** *Suponha que a função objetivo f do problema (30) é diferenciável e seja $\mathcal{S}$ seu conjunto factível. Então x^* é minimizador global de (30) se, e somente se, $x \in \mathcal{S}$ e*

$$\nabla f(x^*)^t(y - x^*) \geq 0,\ \forall y \in \mathcal{S}. \tag{32}$$

Demonstração Primeiramente, suponhamos $x^* \in \mathcal{S}$ tal que x^* satisfaz (32). Então, se $y \in \mathcal{S}$ temos, por (18), $f(y) \geq f(x^*)$. Isso mostra que x^* é minimizador global de (30).

Para a segunda parte, suponhamos que x^* é minimizador de f em $\mathcal{S}$ tal que a condição (32) não seja válida, *i.e.*, para algum $y \in \mathcal{S}$ temos

$$\nabla f(x^*)^t(y - x^*) < 0.$$

Considere o ponto $z(t) = ty + (1-t)x^*$, onde $t \in [0,1]$. Como $z(t)$ está no segmento entre os pontos x^* e y e $\mathcal{S}$ é convexo, $z(t)$ é factível. Agora note que

$$\frac{d}{dt} f(z(t))\bigg|_{t=0} = \nabla f(x^*)^t (y - x^*) < 0,$$

o que mostra que para algum t suficientemente pequeno temos que $f(z(t)) < f(x^*)$, o que contradiz a hipótese. Logo, um minimizador x^* deve satisfazer (32). $\square$

O teorema acima apresenta a seguinte justificativa geométrica: um ponto x^* é minimizador local de f em $\mathcal{S}$ se, e somente se, não existem direções de descida factíveis em $\mathcal{S}$ a partir de x^*, ou seja, não existe $v \in \mathbb{R}^n$ tal que $f(x^* + v) < f(x^*)$ e $(x^* + v) \in \mathcal{S}$.

4.1.2 Problemas de Otimização Convexa Equivalentes

Neste texto diremos que dois problemas de otimização são equivalentes quando, ao ser encontrada a solução para um deles, a solução para o outro é obtida trivialmente. Assim, veremos algumas transformações que mantêm a convexidade e podem nos ajudar a obter problemas equivalentes e que são mais fáceis de serem resolvidos.

Eliminando as Restrições de Igualdade

Para um problema convexo, as restrições de igualdade devem ser lineares, ou seja, da forma $Ax = b$. Primeiramente, sabemos que uma solução qualquer do sistema $\tilde{x}$ pode ser escrita como $\tilde{x} = x_0 + x_N$, onde x_0 é uma solução particular e $x_N \in \mathcal{N}(A)$. Assim, encontrando uma matriz Z cujas colunas formam uma base para o núcleo da matriz A, podemos eliminar as restrições de igualdade, resultando no problema

$$\begin{array}{rl} Minimizar & f(Zy + x_0) \\ Sujeita\, a & g(Zy + x_0) \leq 0, \end{array}$$

onde y se torna a variável de otimização. Ao encontrar a solução y^* do problema acima, podemos facilmente encontrar a solução x^* do problema original, pois $x^* = Zy^* + x_0$.

Como a composição de uma função convexa com uma função afim é convexa, a eliminação das restrições de igualdade preserva a convexidade do problema. Porém, essa técnica não deve ser usada em todos os problemas, pois obter a matriz Z é computacionalmente caro e isso pode também arruinar alguma estrutura especial.

Variáveis de Folga

Quando temos restrições de desigualdade $g_i(x) \leq 0$ tais que g_i são funções afins, podemos substituir essas restrições por $g_i(x) + s_i = 0$, onde $s_i \geq 0$ são chamadas de variáveis de folga. Assim, introduzir variáveis de folga para *desigualdades afins* preservam a convexidade do problema.

Forma de Epigrafo
O problema de otimização convexa (30) pode ser reescrito da seguinte forma

$$\begin{array}{rl} Minimizar & t \\ Sujeita\, a & f(x) - t \leq 0, \\ & g(x) \leq 0, \\ & Ax = b. \end{array}$$

Como $f(x) \leq t$ representa o epigrafo da função, essa forma é chamada de *forma de epigrafo* do problema (30). Assim, buscamos o menor valor t tal que (x^*, t) pertença ao epigrafo, sendo x^* então o minimizador.

Assim, a função objetivo se torna linear e a nova restrição adicionada $f(x) - t$ é convexa e, portanto, a convexidade do problema é mantida. Portanto, pode-se considerar que todo problema de otimização convexa tem a função objetivo linear e, devido a esse fato, a forma de epigrafo é também chamada de *forma universal*.

O uso da forma universal é muito útil em casos práticos. Por exemplo, a análise teórica é muito mais simples para uma função objetivo linear. Além disso, algoritmos desenvolvidos para problemas convexos podem supor que a função objetivo é linear, já que todo problema convexo poder ser reescrito dessa forma.

4.1.3 Problema de Maximização Côncava

Podemos também nos referir ao problema (abusando da notação)

$$\begin{array}{rl} Maximizar & f(x) \\ Sujeita\, a & g(x) \leq 0, \\ & Ax = b, \end{array} \tag{33}$$

onde $f : I\!R^n \to I\!R$ é uma função côncava e cada componente de $g : I\!R^n \to I\!R^m$, *i.e.*, $g_i : I\!R^n \to I\!R$, $i = 1, ..., m$, é uma função convexa, $A \in I\!R^{p \times n}$ e $b \in I\!R^p$, como um problema de Otimização Convexa. Podemos construir um problema equivalente em que minimizamos a função $-f$, sendo que esse é um problema convexo de fato. Sendo assim, todos os resultados utilizados podem ser utilizados para *problemas de maximização côncava*.

4.1.4 Otimização Quase–Convexa

Um *problema de otimização quase–convexa* tem a forma padrão

$$\begin{array}{rl} Minimizar & f(x) \\ Sujeita\, a & g(x) \leq 0, \\ & Ax = b, \end{array} \tag{34}$$

onde a função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ é quase–convexa e cada componente de $g : \mathbb{R}^n \to \mathbb{R}^m$, *i.e.*, $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., m$, é uma função convexa, $A \in \mathbb{R}^{p\times n}$ e $b \in \mathbb{R}^p$. Assim, essa classe de problemas é muito próxima à de otimização convexa e, portanto, algumas propriedades podem ser estendidas enquanto que outras dever ser adaptadas.

Um método computacional para resolver esse tipo de problema pode ser encontrado em [3].

Minimizadores Locais e Condições de Otimalidade

A primeira grande diferença entre otimização convexa e quase–convexa é que a segunda pode apresentar minimizadores locais que não são necessariamente globais. Podemos adaptar as condições de primeira ordem para o problema quase–convexo: sendo $\mathcal{S}$ a região factível do problema de otimização quase–convexa (34), temos que x^* é minimizador do problema se

$$x \in \mathcal{S}, \quad \nabla f(x^*)^t(y - x^*) > 0,\ \forall y \in \mathcal{S}\backslash\{x^*\}. \tag{35}$$

Essa condição é semelhante à dada em (32), mas note que, enquanto a condição para problemas convexos é necessária e suficiente, essa é uma condição apenas suficiente e, portanto, existem minimizadores que não satisfazem (35).

4.2 Problemas de Otimização Linear

Problemas de *programação* (ou *otimização*) *linear* (PL) são muito utilizados para modelar aplicações práticas nas áreas de administração, logística e engenharia [1]. Um problema geral de programação linear tem a seguinte forma

$$\begin{array}{rl} Minimizar & c^t x + d \\ Sujeita\, a & Gx \leq h, \\ & Ax = b, \end{array} \tag{36}$$

onde $G \in \mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{p\times n}$, $c \in \mathbb{R}^n$, $h \in \mathbb{R}^m$, $b \in \mathbb{R}^p$ e $d \in \mathbb{R}$, ou seja, tem a função objetivo e todas as restrições afins, sendo obviamente um problema convexo.

Normalmente omitimos a constante d pois isso não altera a solução do problema. Além disso, existem problemas de maximização linear que podem facilmente ser transformados em problemas de minimização devido à linearidade da função objetivo.

Um dos modos mais recorrentes de representar (37) é a chamada *forma padrão* de um programa linear que é dada por

$$\begin{array}{rl} Minimizar & c^t x + d \\ Sujeita\, a & Ax = b, \\ & x \geq 0, \end{array} \tag{37}$$

i.e., as únicas restrições de desigualdade são de não–negatividade das componentes da variável de otimização.

Para representar um programa linear geral (37) na forma padrão (38), primeiramente introduzimos variáveis de folga nas desigualdades, resultando em

$$\begin{array}{rl} Minimizar & c^t x + d \\ Sujeita\, a & Gx + s = h, \\ & Ax = b, \\ & s \geq 0. \end{array}$$

A seguir, tomamos a variável x, que é suposta livre, e a expressamos como a diferença de duas variáveis não–negativas u e v, ou seja, $x = u - v$, $u, v \geq 0$. Assim, chegamos ao problema

$$\begin{array}{rl} Minimizar & c^t u - c^t v + d \\ Sujeita\, a & Gu - Gv + s = h, \\ & Au - Av = b, \\ & s \geq 0, \quad u \geq 0, \quad v \geq 0, \end{array}$$

que é um programa linear na forma padrão com variáveis u, v e s.

Exemplo 4.1 ***(Problema da Dieta)*** *Um problema muito simples que pode ser modelado como um programa linear é o problema da dieta: suponhamos que, em uma dieta saudável, devemos ingerir m nutrientes em quantidades mínimas iguais a $b_1, \ldots, b_m$.*

Vamos compor uma dieta com n tipos de alimento, cada um com quantidades $x_1, \ldots, x_n$. Uma unidade do alimento j tem uma quantidade a_{ij} do nutriente i e um custo c_j. Queremos determinar a dieta mais barata que pode *nos fornecer todos esses nutrientes. Esse problema pode ser formulado como um programa linear da seguinte forma*

$$\begin{aligned} Minimizar \quad & c^t x \\ Sujeita\, a \quad & Ax \geq b, \\ & x \geq 0. \end{aligned}$$

Várias alterações podem ser feitas sobre esse problema, como a inserção de cotas superiores para alguns nutrientes que são nocivos em excesso (sódio, por exemplo).

Exemplo 4.2 ***(Fluxo em Rede)*** *Considere uma rede (i.e., um grafo orientado) com n nós e com arcos (direcionados) ligando cada par de nós. As variáveis do problema são os fluxos em cada arco: x_{ij} denota o fluxo do nó i para o nó j. Dadas as constantes c_{ij}, o custo do fluxo ao longo do arco de i para j é dado por $c_{ij}x_{ij}$. Assim, queremos minimizar o custo total do fluxo na rede, dado por*

$$C = \sum_{i,j=1}^{n} c_{ij}x_{ij}.$$

Cada fluxo deve obedecer às cotas inferiores (l_{ij}) e superiores (u_{ij}), ou seja, $l_{ij} \leq x_{ij} \leq u_{ij}$. Em cada nó, existem trocas com o meio externo à rede dadas por b_i, onde $b_i > 0$ se uma quantidade b_i entra na rede no nó i e $b_i < 0$ se uma quantidade $|b_i|$ sai da rede no nó i. Assumimos que $\sum_i b_i = 0$, ou seja, o fornecimento externo é igual à demanda externa. Além disso, em cada nó há a conservação do fluxo, ou seja, o total de fluxo que entra no nó i é igual ao fluxo que sai do nó i.

Se formularmos esse problema como um programa linear chegamos a

$$\begin{aligned} Minimizar \quad & \sum_{i,j=1}^{n} c_{ij}x_{ij} \\ Sujeita\, a \quad & l_{ij} \leq x_{ij} \leq u_{ij}, \\ & \sum_{k=1}^{n} x_{ki} + b_i - \sum_{k=1}^{n} x_{ik} = 0. \end{aligned}$$

Problemas de fluxos em redes têm várias aplicações em engenharia e logística e exemplos aplicados podem ser buscados em livros clássicos de programação linear.

Existem diversos métodos eficazes para programação linear. O *método Simplex* foi o primeiro a ser desenvolvido e, apesar de ter complexidade exponencial, ele funciona muito bem na prática e é muito rápido para problemas de pequeno e médio porte. Apesar de terem surgido mais tarde, os *métodos de Pontos Interiores* são atualmente muito utilizados para resolver problemas de grande porte com estrutura especial [4].

4.3 Problemas de Otimização Quadrática

Um *problema de programação* (ou *otimização*) *quadrática* (PQ) tem a forma

$$\begin{array}{rl} Minimizar & \frac{1}{2}x^t P x + q^t x + r \\ Sujeita\, a & Gx \leq h, \\ & Ax = b, \end{array} \tag{38}$$

onde $P \in \mathbb{R}^{n\times n}$ é semidefinida positiva, $G \in \mathbb{R}^{m\times n}$ e $A \in \mathbb{R}^{p\times n}$. Assim, em um PQ, minimizamos uma função quadrática convexa sobre um poliedro.

O programa quadrático pode ser generalizado se substituirmos as restrições afins de desigualdade por restrições quadráticas convexas:

$$\begin{array}{rll} Minimizar & \frac{1}{2}x^t P x + q^t x + r & \\ Sujeita\, a & \frac{1}{2}x^t P_i x + q_i^t x + r_i \leq 0, & i = 1, ..., m \\ & Ax = b, & \end{array} \tag{39}$$

onde $P_i \in \mathbb{R}^{n\times n}$ são semidefinidas positivas, $i = 1, ..., m$. Esse problema é chamado de *problema de programação quadrática restrito quadraticamente* (QCQP). Qualquer PQ e qualquer QCQP são problemas convexos. Note que um PL é um caso especial de um PQ quando $P = 0$ e que um PQ é um caso especial de um QCQP quando $P_i = 0$ para $i = 1, ..., m$.

Exemplo 4.3 ***(Quadrados Mínimos Lineares)*** *O problema de encontrar $x \in \mathbb{R}^n$ que "melhor resolve" o sistema sobredeterminado $Ax = b$, ou seja, que minimiza $||r||_2 = ||Ax - b||_2$, pode ser resolvido pelo seguinte PQ irrestrito*

$$Minimizar \qquad ||Ax - b||_2^2 = x^t A^t A x - 2b^t A x + b^t b.$$

Esse problema pode ser resolvido de diversas formas e tem solução analítica conhecida, dada por $x^* = A^\dagger b$, onde $A^\dagger$ é a matriz pseudo-inversa de A.

Quando restrições lineares são adicionadas, o problema passa a não ter uma solução analítica simples. Por exemplo, podemos ter cotas inferiores e superiores para as variáveis, ou seja,

$$\begin{array}{rll} Minimizar & ||Ax - b||_2^2 & \\ Sujeita\,a & l_i \leq x_i \leq u_i, & i = 1, ..., n. \end{array}$$

Problemas de quadrados mínimos lineares são muito utilizados em ajustes de dados ou aparecem como subproblemas de otimização.

Exemplo 4.4 ***(Distância entre Poliedros)*** *Definimos a distância entre dois poliedros* $\mathcal{P}_1 = \{x \mid A_1 x \leq b_1\}$ *e* $\mathcal{P}_2 = \{x \mid A_2 x \leq b_2\}$ *em* $\mathbb{R}^n$ *como*

$$\mathbf{dist}(\mathcal{P}_1, \mathcal{P}_2) = \inf\{||x_1 - x_2||_2 \mid x_1 \in \mathcal{P}_1,\ x_2 \in \mathcal{P}_2\}.$$

Para encontrar a distânca entre $\mathcal{P}_1$ *e* $\mathcal{P}_2$ *podemos resolver o seguinte PQ*

$$\begin{array}{rll} Minimizar & ||x_1 - x_2||_2^2 & \\ Sujeita\,a & A_1 x_1 \leq b_1, & A_2 x_2 \leq b_2, \end{array}$$

com variáveis $x_1, x_2 \in \mathbb{R}^n$. *O problema é infactível se, e somente se, algum dos poliedros é vazio. Caso contrário, o problema nos fornece os pontos mais próximos um do outro e a distância entre eles.*

Esse problema serve como base para outros que formam a classe de problemas geométricos. Problemas geométricos envolvem classificação e separação de pontos, localização de estruturas e determinação de distância entre conjuntos convexos quaisquer.

Exemplo 4.5 ***(Solução de Norma-2 Mínima)*** *Tomemos um sistema linear indeterminado* $Ax = b$, *com* $A \in \mathbb{R}^{m \times n}$, $m < n$, *e* $\mathbf{posto}(A) = m$, *ou seja, com menos equações do que incógnitas. Portanto, esse sistema tem infinitas soluções* *e uma solução especial que podemos determinar é a solução de norma-2 mínima, que pode ser calculada resolvendo-se o PQ*

$$\begin{array}{rl} Minimizar & ||x||_2^2 \\ Sujeita\, a & Ax = b. \end{array}$$

Esse problema surge muito na área de controle ótimo. As variáveis $x_1, ..., x_n$ são chamadas de variáveis de projeto e o vetor Ax nos dá m resultados do projeto dado por x. As equações $Ax = b$ são os requisitos de projeto. Como o sistema tem graus de liberdade, os requisitos podem ser atingidos com vários valores para x. Assim, a solução de norma-2 mínima representa o menor projeto, com respeito à norma $|| \cdot ||_2$.

4.4 Second–Order Cone Programming

Um problema que é muito próximo aos de programação quadrática é o *problema cônico de segunda ordem* (ou *second–order cone program*) (SOCP):

$$\begin{array}{rll} Minimizar & f^t x & \\ Sujeita\, a & ||A_i x + b_i||_2 \leq c_i^t x + d_i, & i = 1, ..., m \\ & Fx = g, & \end{array} \tag{40}$$

onde $A_i \in I\!R^{n_i \times n}$, $i = 1, ..., m$ e $F \in I\!R^{p \times n}$. Chamamos uma restrição da forma

$$||Ax + b||_2 \leq c^t x + d,$$

com $A \in I\!R^{k \times n}$, de *restrição cônica de segunda ordem*, já que é o mesmo que exigir que os pontos da forma $(Ax + b, c^t x + d)$ estejam em um cone de segunda ordem em $I\!R^{k+1}$.

Note que, quando $c_i = 0$, $i = 1, ..., m$, o SOCP (41) se reduz a um QCQP, obtido elevando-se ao quadrado cada restrição de desigualdade. Analogamente, quando $A_i = 0$, $i = 1, ..., m$, então o SOCP (41) se reduz a um PL geral.

5 Dualidade

Nesta seção estudaremos alguns tópicos que envolvem a Teoria de Dualidade, assunto central na área de Otimização Convexa, dentre eles os multiplicadores de Lagrange e sua relação com as condições de qualificação (ou *constraint qualifications*), a função Lagrangiano, condições de otimalidade e os teoremas de dualidade.

Nesta seção, em muitos casos, teremos resultados gerais de Otimização Não–Linear e, portanto, especificaremos bem quais resultados são válidos apenas para problemas convexos.

5.1 Introdução aos Multiplicadores de Lagrange

Nesta subseção veremos a idéia básica a respeito dos multiplicadores de Lagrange, que terão papel fundamental durante toda a seção.

Primeiramente, vamos considerar o problema que apresenta apenas restrições de igualdade

$$\begin{array}{rll} Minimizar & f(x) & \\ Sujeita\,a & h_i(x) = 0, & i = 1, ..., m, \end{array}$$

onde $f : I\!R^n \to I\!R$ e $h_i : I\!R^n \to I\!R$ são funções suaves. Então, um resultado básico, devido a Lagrange, nos diz que, sob algumas hipóteses, para um minimizador local x^* existem escalares $\lambda_1^*, \ldots, \lambda_m^*$, chamados de *multiplicadores de Lagrange*, tais que

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0. \tag{41}$$

Essa condição pode ser vista como n equações, que juntamente com as restrições $h(x^*) = 0$, formam um sistema de $m + n$ equações e $m + n$ incógnitas, o vetor x^* e os multiplicadores λ^*. Portanto, problemas de otimização não–linear e sistemas de equações não–lineares são fortemente relacionados, sendo essa a principal idéia de alguns métodos, como o Primal–Dual.

As hipóteses adicionais que deveriam ser tomadas são chamadas de *condições de qualificação* (CQ). Sabemos que se x^* é minimizador e x^* satisfaz CQ, então existem os multiplicadores de Lagrange λ^*. Uma dessas qualificações afirma que os gradientes $\nabla h_i(x^*)$ devem ser linearmente independentes. Essa CQ é chamada de *Regularidade* ou *LICQ* (de *Linear Independence Constraint Qualification*). Infelizmente, em alguns casos os multiplicadores de Lagrange podem não existir, como veremos no exemplo a seguir.

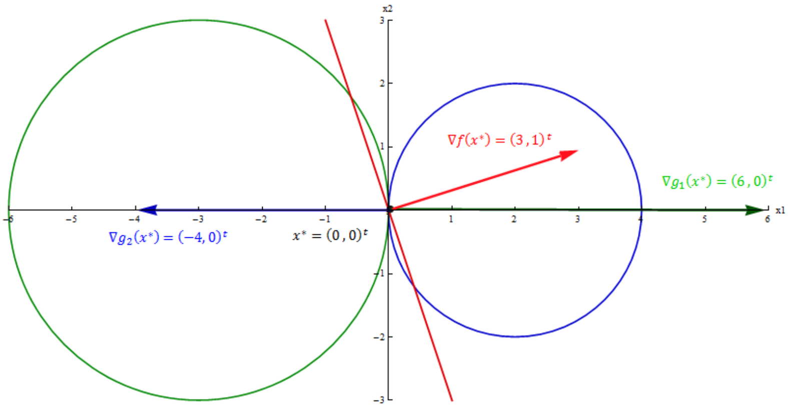
Exemplo 5.1 *Consideremos o seguinte problema*

$$\begin{array}{rl} Minimizar & f(x) = 3x_1 + x_2 \\ Sujeita\,a & h_1(x) = (x_1+3)^2 + x_2^2 - 9 = 0, \\ & h_2(x) = (x_1-2)^2 + x_2^2 - 4 = 0. \end{array}$$

É fácil ver que a solução global para o problema é o ponto $x^* = (0,0)^t$, *sendo esse o único ponto factível, como pode ser observado na figura 7.*

Agora, os multiplicadores de Lagrange, quando existem, são tais que $\nabla f(x^*) = \lambda_1^* \nabla h_1(x^*) + \lambda_2^* \nabla h_2(x^*)$. *Como os gradientes das restrições são linearmente dependentes e o gradiente da função objetivo não é paralelo a esses, não existem os multiplicadores de Lagrange para o ponto* x^*.

Figura 7: Gradientes das restrições e da função objetivo em $x^* = (0,0)^t$, onde não existem multiplicadores de Lagrange.



Agora, vamos considerar um problema com apenas restrições de desigualdade:

$$\begin{array}{rll} Minimizar & f(x) & \\ Sujeita\,a & g_i(x) \leq 0, & j = 1, ..., p. \end{array}$$

Um segundo resultado básico nos diz que, se x^* é minimizador local do problema e se x^* satisfaz CQ, então existem $\mu_1^*, ..., \mu_p^*$, também chamados de *multiplicadores de Lagrange*, tais que

$$\nabla f(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0,$$

$$\mu_j^* \geq 0, \quad \forall j = 1, ..., p, \qquad \mu_j^* = 0, \quad \forall j \in \mathcal{I}(x^*),$$

onde $\mathcal{I}(x^*)$ denota o conjunto de índices das restrições ativas em x^*, ou seja, $j \in \mathcal{I}(x^*) \iff g_j(x^*) = 0$. Notamos que, nesse caso, é necessária alguma informação auxiliar para podemos resolver esse problema de otimização como um sistema não–linear. Posteriormente veremos como obter novas condições a fim de formar um sistema não–linear com mesma quantidade de incógnitas e variáveis para esse problema.

O resultado a seguir estabelece regras gerais para os multiplicadores de Lagrange, sem nenhuma condição de qualificação.

Teorema 5.1 ***(Condições de Fritz–John)*** *Se x^* é minimizador de f no conjunto $\mathcal{S} = \{x \mid h(x) = 0, \ g(x) \leq 0\}$, então existem escalares $\mu_0, \lambda_1, ..., \lambda_m, \mu_1, ..., \mu_p$, não todos nulos, tais que $\mu_0, \mu_1, ..., \mu_p \geq 0$ e*

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0. \tag{42}$$

Existem duas possibilidades dadas pelas condições de Fritz-John: se $\mu_0 > 0$, temos o caso em que os multiplicadores de Lagrange existem e são dados por $\lambda_i^* = \lambda_i / \mu_0$ e $\mu_j^* = \mu_j / \mu_0$; se $\mu_0 = 0$, então os gradientes das restrições devem ser linearmente dependentes e não se obtém nenhuma informação a respeito da existência dos multiplicadores de Lagrange.

Proposições como as condições de Fritz–John, que estabelecem propriedades necessárias ou suficientes a respeito de minimizadores, são chamadas *condições de otimalidade*. Apesar de ser uma condição de otimalidade simples e sem requerer condições de qualificação, as condições de Fritz–John apresentam o problema de não afirmar quando o escalar μ_0 pode ser tomado como positivo. Portanto, boas condições de otimalidade devem ser restritivas, a fim de que pontos que não sejam minimizadores não as satisfaçam. Falaremos mais sobre condições de otimalidade posteriormente.

5.2 Condições de Qualificação

Como vimos na subseção anterior, *condições de qualificação* são propriedades sobre os pontos factíveis de um problema de otimização que, quando satisfeitas por um minimizador local x^*, a existência dos multiplicadores de Lagrange é garantida. Resumindo, temos

$$x^* \ minimizador \ e \ x^* \ satisfaz \ CQ \implies Existem \ multiplicadores \ de \ Lagrange,$$

de onde observamos que uma CQ mais fraca nos permite encontrar mais candidatos a minimizador local do que uma CQ mais forte, já que existiriam mais pontos que a tornariam válida.

A seguir, veremos três condições de qualificação clássicas.

CQ 1 ***(Linear Independence Constraint Qualification – LICQ)*** *Os gradientes das restrições de igualdade $\nabla h_i(x^*)$, $i = 1, ..., m$, e os gradientes das restrições de desigualdade ativas $\nabla g_j(x^*)$, $j \in \mathcal{I}(x^*)$, são linearmente independentes.*

CQ 2 ***(Mangasarian–Fromovitz Constraint Qualification – MFCQ)*** *Os gradientes das restrições de igualdade $\nabla h_i(x^*)$, $i = 1, ..., m$, são linearmente independentes e existe $y \in I\!R^n$ tal que*

$$\nabla h_i(x^*)^t y = 0, \quad i = 1, ..., m, \qquad \nabla g_j(x^*)^t y < 0, \quad \forall j \in \mathcal{I}(x^*).$$

CQ 3 ***(Slater's Constraint Qualification – SCQ)*** *As funções f e g_j, $j = 1, ..., p$, são convexas, as funções h_i, $i = 1, ..., m$, são afins e existe um ponto factível $\tilde{x} \in \mathbf{relint}(\mathcal{S})$, onde $\mathcal{S} = \{x \in I\!R^n \mid h(x) = 0,\ g(x) \leq 0\}$, tal que $g_j(\tilde{x}) < 0$ sempre que a função g_j não for afim.*

As duas condições de qualificação são clássicas e podem ser usadas para qualquer problema de programação não–linear, já a última é uma condição específica para problemas convexos e será de grande importância para a Teoria de Dualidade.

Pode-se mostrar que, se um ponto satisfaz LICQ, então ele satisfaz MFCQ, mas que a recíproca é falsa (ver Exemplo 5.2). Portanto, MFCQ é uma condição mais fraca, nos fornecendo um maior conjunto de candidatos a minimizador.

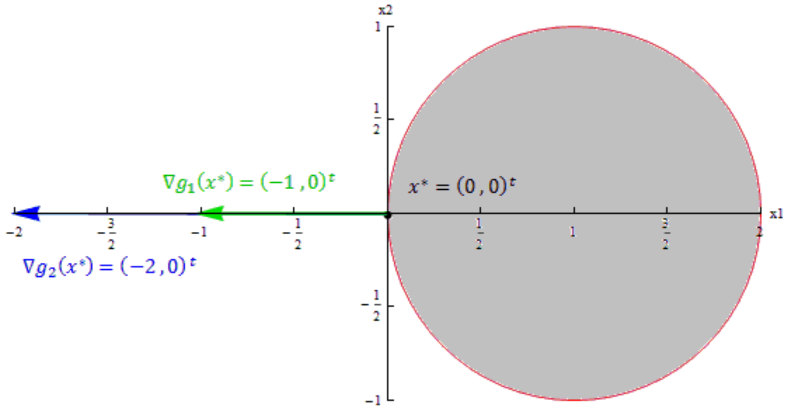
Exemplo 5.2 *Considere o QCQP abaixo*

$$\begin{array}{rl} Minimizar & f(x) = x_1^2 + x_2^2 \\ Sujeita\, a & g_1(x) = -x_1 \leq 0, \\ & g_2(x) = (x_1 - 1)^2 + x_2^2 - 1 \leq 0. \end{array}$$

Observamos que a solução global do problema é o ponto $x^* = (0, 0)^t$. Na figura 8, estão representados o conjunto factível e os gradientes das restrições.

Da figura, percebemos que x^* não satisfaz LICQ, pois os gradientes das restrições ativas em x^*, $\nabla g_1(x^*)$ e $\nabla g_2(x^*)$, são linearmente dependentes. Agora, o ponto satisfaz MFCQ, já que qualquer vetor da forma $y = (\alpha, 0)^t$, $\alpha > 0$, é tal que $\nabla g_1(x^*)^t y < 0$ e $\nabla g_2(x^*)^t y < 0$, e satisfaz também SCQ, já que todas as funções são convexas e o interior relativo do conjunto factível é não–vazio. Conforme as CQ garantem, os multiplicadores μ_1^*, μ_2^* existem e são ambos nulos.

Figura 8: Conjunto factível e gradientes das restrições em $x^* = (0, 0)^t$.



A busca por condições de qualificação mais fracas e que sejam verificadas computacionalmente de forma mais simples é um dos novos objetivos da pesquisa em otimização matemática. Outras condições como a *condição de posto constante* (CRCQ) e CPLD e uma análise detalhada a respeito de condições de qualificação pode ser encontrada em [6].

5.3 Dualidade Lagrangiana

5.3.1 Função Lagrangiano

Considere o problema geral de programação não–linear na forma padrão:

$$\begin{array}{rl} Minimizar & f(x) \\ Sujeita\, a & g(x) \leq 0, \\ & h(x) = 0, \end{array} \tag{43}$$

onde $f : I\!R^n \to I\!R$, $g : I\!R^n \to I\!R^p$ e $h : I\!R^n \to I\!R^m$. A partir de agora, chamaremos esse problema de *problema primal* ou *problema original* e denotaremos seu valor ótimo por p^*. A idéia básica da dualidade Lagrangiana é levar em conta as restrições aumentando a função objetivo com a soma ponderada das restrições. Definimos a *função Lagrangiano* $\mathcal{L} : I\!R^n \times I\!R^m \times I\!R^p \to I\!R$ associada ao problema (43) como

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x). \tag{44}$$

Os escalares λ_i e μ_j, $i = 1, ..., m$, $j = 1, ..., p$, são os multiplicadores de Lagrange associados às respectivas restrições. Os vetores $\lambda = (\lambda_1, ..., \lambda_m)^t$ e $\mu = (\mu_1, ..., \mu_p)^t$ são chamados de *variáveis duais* associadas ao problema (43).

5.3.2 Função Dual

Definimos a função dual associada ao problema primal (43) como a função $\phi : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ que é dada pelo valor mínimo da função Lagrangiano sobre a variável x, ou seja, tomando-se as variáveis duais do problema λ, μ,

$$\phi(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu) = \inf_x \left(f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x) \right). \tag{45}$$

Quando a função Lagrangiano não é limitada inferiormente em x, a função dual assume o valor $-\infty$. Como a função dual é o ínfimo de uma família de funções afins em λ, μ, temos, por um resultado análogo ao demonstrado para o supremo de uma família de funções, que a função dual é côncava mesmo que o problema (43) não seja convexo.

Agora, o resultado a seguir nos garante uma cota inferior para o valor ótimo do problema primal.

Proposição 5.1 *Considere o problema primal (43), cujo valor ótimo é p^*. Então, para quaisquer variáveis duais λ, μ tais que $\mu \geq 0$, temos*

$$\phi(\lambda, \mu) \leq p^*. \tag{46}$$

Demonstração Suponha que $\tilde{x}$ é um ponto factível do problema (43), ou seja, $g(\tilde{x}) \leq 0$ e $h(\tilde{x}) = 0$, e que $\mu \geq 0$. Então,

$$\sum_{i=1}^{m} \lambda_i h_i(\tilde{x}) + \sum_{j=1}^{p} \mu_j g_j(\tilde{x}) \leq 0,$$

já que cada termo da primeira soma é não–positivo e cada termo da segunda é nulo. Portanto,

$$\mathcal{L}(\tilde{x}, \lambda, \mu) = f(\tilde{x}) + \sum_{i=1}^{m} \lambda_i h_i(\tilde{x}) + \sum_{j=1}^{p} \mu_j g_j(\tilde{x}) \leq f(\tilde{x}).$$

Assim,

$$\phi(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu) \leq \mathcal{L}(\tilde{x}, \lambda, \mu) \leq f(\tilde{x}).$$

Como $\phi(\lambda, \mu) \leq f(\tilde{x})$ para qualquer ponto factível $\tilde{x}$, chegamos a (46). $\Box$

5.3.3 Problema Dual

Para cada par (λ, μ), a função dual nos fornece uma cota inferior para o valor ótimo p^* do problema primal. Nosso objetivo agora se torna encontrar a melhor cota inferior para o problema original, construindo o seguinte problema de otimização:

$$\begin{array}{rl} Maximizar & \phi(\lambda, \mu) \\ Sujeita\, a & \mu \geq 0. \end{array} \tag{47}$$

Esse problema é chamado de *problema dual* de (43). Um par de variáveis duais (λ, μ) é dito *dual factível* se $\mu \geq 0$ e se $\phi(\lambda, \mu) > -\infty$, ou seja, se for factível para o problema dual (47). Um par (λ^*, μ^*) é uma *solução dual ótima* se for um otimizador de (47).

Note que o problema dual (47) é convexo, já que se trata de um problema de maximização côncava, e isso ocorre para qualquer problema primal, convexo ou não.

Nos exemplos a seguir, determinamos os duais de problemas clássicos de programação linear e programação quadrática.

Exemplo 5.3 ***(Dual de um PL na Forma Padrão)*** *Considere o PL na forma padrão:*

$$\begin{array}{rl} Minimizar & c^t x \\ Sujeita\, a & Ax = b, \\ & x \geq 0. \end{array}$$

Para encontrar seu problema dual, primeiramente devemos encontrar sua função dual associada. A função Lagrangiano para o problema é dada por:

$$\begin{aligned} \mathcal{L}(x, \lambda, \mu) &= c^t x + \lambda^t (Ax - b) + \mu^t(-x) = \\ &= -b^t \lambda + (c + A^t \lambda - \mu)^t x. \end{aligned} \tag{48}$$

Assim, pela definição da função dual, temos:

$$\begin{aligned} \phi(\lambda, \mu) &= \inf_x \{-b^t \lambda + (c + A^t \lambda - \mu)^t x\} = \\ &= -b^t \lambda + \inf_x \{(c + A^t \lambda - \mu)^t x\}. \end{aligned} \tag{49}$$

Agora, uma função afim é limitada quando for identicamente nula. Assim, temos

$$\phi(\lambda, \mu) = \begin{cases} -b^t\lambda & se\ c + A^t\lambda - \mu = 0, \\ -\infty & caso\ contrário. \end{cases}$$

Portanto, podemos dizer que o problema dual é dado por

$$\begin{aligned} Maximizar \quad & \phi(\lambda, \mu) \\ Sujeita\, a \quad & \mu \geq 0, \end{aligned}$$

onde ϕ é a função dual descrita acima. Mas, como a função dual apenas é finita quando $c + A^t\lambda - \mu = 0$, podemos escrever o seguinte problema equivalente

$$\begin{aligned} Maximizar \quad & -b^t\lambda \\ Sujeita\, a \quad & c + A^t\lambda - \mu = 0, \\ & \mu \geq 0, \end{aligned} \tag{50}$$

explicitando as restrições de igualdade. Note que esse problema é um PL na forma padrão e pode ser reescrito apenas com restrições de desigualdade da seguinte forma:

$$\begin{aligned} Maximizar \quad & -b^t\lambda \\ Sujeita\, a \quad & c + A^t\lambda \geq 0. \end{aligned} \tag{51}$$

Assim, os problemas (51) e (52) são equivalentes ao problema dual padrão e são normalmente chamados de problemas duais para um PL na forma padrão.

Exemplo 5.4 ***(Dual de um PQ com Restrições de Desigualdade)*** *Considere o seguinte problema quadrático com restrições de desigualdade (desprezamos o termo constante por simplicidade):*

$$\begin{aligned} Minimizar \quad & \frac{1}{2}x^tPx + q^tx \\ Sujeita\, a \quad & Gx \leq h, \end{aligned}$$

tomando-se, nesse problema, P definida positiva. A função Lagrangiano para o problema acima é dada por

$$\mathcal{L}(x,\mu) = \frac{1}{2}x^tPx + q^tx + \mu^t(Gx-h) =$$
$$= -h^t\mu + \left(\frac{1}{2}x^tPx + q^tx + \mu^tGx\right). \tag{52}$$

Portanto, pela definição de função dual, temos:

$$\phi(\mu) = \inf_x \left\{-h^t\mu + \left(\frac{1}{2}x^tPx + q^tx + \mu^tGx\right)\right\} =$$
$$= -h^t\mu + \inf_x \left\{\frac{1}{2}x^tPx + q^tx + \mu^tGx\right\}. \tag{53}$$

Agora, seja $r(x,\mu) = \frac{1}{2}x^tPx + q^tx + \mu^tGx$. *Como* P *é definida positiva, temos que o valor de* x *que minimiza* r *pode ser obtido resolvendo*

$$\nabla_x r(x,\mu) = Px + q + G^t\mu = 0,$$

de onde temos que o minimizador x^* *é dado por*

$$x^* = -P^{-1}(q + G^t\mu).$$

Assim,

$$\phi(\lambda,\mu) = -h^t\mu + \frac{1}{2}[-P^{-1}(q+G^t\mu)]^tP[-P^{-1}(q+G^t\mu)] +$$
$$+ q^t[-P^{-1}(q+G^t\mu)] + \mu^tG[-P^{-1}(q+G^t\mu)] =$$
$$= -\frac{1}{2}\mu^tGP^{-1}G^t\mu - \frac{1}{2}q^tP^{-1}q - q^tP^{-1}G^t\mu - h^t\mu =$$
$$= -\frac{1}{2}\mu^tQ\mu - v^t\mu - \frac{1}{2}q^tP^{-1}q, \tag{54}$$

onde $Q = GP^{-1}G^t$ *e* $v = h + GP^{-1}q$.

Finalmente, chegamos ao problema dual, onde desprezamos a constante $\frac{1}{2}q^tP^{-1}q$,

$$\begin{aligned} &Minimizar \quad && -\frac{1}{2}\mu^tQ\mu - v^t\mu \\ &Sujeita\,a && \mu \geq 0. \end{aligned}$$

Notamos que o problema dual é mais simples de ser resolvido que o primal, pois as restrições agora são de não–negatividade das variáveis. Além disso, o tamanho do problema diminui se a quantidade de restrições for menor que o número de variáveis primais. Essa redução de tamanho vem com o preço de encontrar a inversa da matriz P, *com a finalidade de encontrar a matriz* Q *e o vetor* v.

5.3.4 Dualidade Fraca e Dualidade Forte

Seja d^* o valor ótimo do problema dual, que é, por definição, a melhor cota inferior para o valor ótimo do problema primal p^* que pode ser obtida a partir da função dual. Assim, podemos concluir que

$$d^* \leq p^*, \tag{55}$$

para qualquer problema de otimização. Essa propriedade é chamada de *dualidade fraca*.

Quando os valores ótimos são infinitos, podemos fazer uma análise interessante. Se o problema primal não for limitado inferiormente, temos $p^* = -\infty$, de onde devemos ter $d^* = -\infty$, ou seja, o problema dual é infactível. Analogamente, se o problema dual não é limitado superiormente, $d^* = \infty$, de onde devemos ter $p^* = \infty$, ou seja, o problema primal é infactível.

Chamamos a diferença $p^* - d^*$ de *gap de dualidade* (*ótimo*), já que ele representa a lacuna existente entre a melhor cota inferior para o valor ótimo com o seu valor verdadeiro.

Usamos o resultado (55) para encontrar um limitante inferior para o valor ótimo do problema original através do cálculo de d^*, já que o problema dual é sempre convexo e pode ser resolvido de forma eficiente.

O caso em que a igualdade

$$d^* = p^* \tag{56}$$

é válida, ou seja, em que o gap de dualidade é nulo, dizemos que o problema satisfaz *dualidade forte*. Neste caso, ao resolvermos o problema dual, encontramos o mesmo valor ótimo do problema primal, o que é geralmente mais simples. Agora, a maioria dos problemas não satisfaz a dualidade forte, mas, se o problema for convexo, temos que a dualidade forte é satisfeita na grande maioria dos casos. O resultado a seguir nos garante a dualidade forte para alguns problemas convexos.

Teorema 5.2 *Seja (P) um problema de otimização convexa com função objetivo f que satisfaz as condições de qualificação de Slater (SCQ). Então, o gap de dualidade para o problema (P) é nulo e existem x^* e (λ^*, μ^*) factíveis tais que $\phi(\lambda^*, \mu^*) = d^* = p^* = f(x^*)$, onde ϕ é a função dual para (P).*

A demonstração do teorema acima pode ser encontrada em [2,3]. Além disso, essas referências ainda fazem uma análise geométrica muito interessante a respeito desses resultados de dualidade.

5.4 Condições de Otimalidade

Nesta subseção, fechamos a parte de Teoria de Dualidade com as condições de otimalidade, que são resultados semelhantes às condições de Fritz–John discutidas no início dessa seção. Primeiramente, introduziremos o conceito de folga complementar.

5.4.1 Folgas Complementares

Suponha que, para o problema (43), seja satisfeita *dualidade forte*. Assim, se x^* é solução primal e (λ^*, μ^*) são soluções duais, temos que

$$\begin{aligned}
f(x^*) = \phi(\lambda^*, \mu^*) = & \\
& = \inf_x \left(f(x) + \sum_{i=1}^{m} \lambda_i^* h_i(x) + \sum_{j=1}^{p} \mu_j^* g_j(x) \right) \leq \\
& \leq f(x^*) + \sum_{i=1}^{m} \lambda_i^* h_i(x^*) + \sum_{j=1}^{p} \mu_j^* g_j(x^*) \leq \\
& \leq f(x^*).
\end{aligned} \tag{57}$$

A partir da terceira linha, podemos concluir que

$$\sum_{j=1}^{p} \mu_j^* g_j(x^*) = 0.$$

Assim, como cada elemento da soma é não–positivo, concluímos que

$$\mu_j^* g_j(x^*) = 0, \qquad j = 1, ..., p. \tag{58}$$

Essas condições são conhecidas como *folgas complementares* e são satisfeitas para quaisquer soluções x^* e $(\lambda^*, \mu*)$, quando temos dualidade forte. Portanto, podemos dizer que o j-ésimo multiplicador de Lagrange é nulo a menos que a j-ésima restrição é ativa em x^*.

5.4.2 Condições de Karush–Kuhn–Tucker

Primeiramente, vamos assumir que o problema (43) é um programa não–linear geral.

Teorema 5.3 ***(Condições Necessárias de Primeira–Ordem)*** *Sejam x^* uma solução primal e (λ^*, μ^*) uma solução dual para o problema (43), tais que alguma CQ é satisfeita. Então:*

$$\begin{aligned}
& g(x^*) \leq 0, && \textit{(Factibilidade Primal)} \\
& h(x^*) = 0, && \textit{(Factibilidade Primal)} \\
& \mu^* \geq 0, && \textit{(Factibilidade Dual)} \\
& \mu_j^* g_j(x^*) = 0, \qquad j = 1, ..., m && \textit{(Folgas Complementares)} \\
& \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0. && \textit{(Estacionaridade)}
\end{aligned} \tag{59}$$

Assim, para qualquer problema geral de otimização, qualquer par de soluções primais e duais em que uma condição de qualificação é válida devem satisfazer as condições de Karush–Kuhn–Tucker, ou KKT (59).

Agora, se considerarmos um problema convexo com dualidade forte, as condições de KKT são necessárias e suficientes.

Teorema 5.4 ***(Condições Necessárias e Suficientes para Problemas Convexos)*** *Considere que o problema (43) é convexo e tal que a dualidade forte é satisfeita. Então, x^* e (λ^*, μ^*) formam um par de soluções primais e duais para o problema se, e somente se,*

$$
\begin{aligned}
&g(x^*) \leq 0, && \textit{(Factibilidade Primal)}\\
&h(x^*) = 0, && \textit{(Factibilidade Primal)}\\
&\mu^* \geq 0, && \textit{(Factibilidade Dual)}\\
&\mu_j^* g_j(x^*) = 0, \qquad j = 1, ..., m && \textit{(Folgas Complementares)}\\
&\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0. && \textit{(Estacionaridade)}
\end{aligned}
\tag{60}
$$

Portanto, em um problema convexo em que as condições de qualificação de Slater são satisfeitas, então as condições de KKT são necessárias e suficientes.

As condições de KKT têm papel muito importante no ramo de otimização. Muitos métodos baseiam suas iterações ou critérios de parada nas condições KKT ou em aproximações dessas condições.

6 Conclusões

Resultados de convexidade são usados em diversos ramos da matemática, principalmente na análise e na otimização. Mesmo para problemas não–lineares, teoremas decorrentes da análise convexa podem ser generalizados ou usados como base para a construção de novas teorias mais amplas.

Sabemos que problemas convexos, quando factíveis, podem ser resolvidos de forma eficiente e rápida. Portanto, o ponto chave na área de otimização convexa está na modelagem dos problemas, já que sua resolução é obtida sem muito esforço. Por isso, a construção de modelos convexos evoluiu tanto que ramos dentro da área possuem nomenclatura própria e são, digamos, totalmente separados dos outros. Talvez seja devido a esse fato que a otimização matemática seja fortemente dividida em linear e não–linear.

Os métodos de pontos interiores são os grandes responsáveis pelo desenvolvimento de aplicações práticas da área de otimização de grande porte. O Simplex tornou a programação linear possível, mas os métodos do tipo primal–dual se tornaram a principal ferramenta para resolver programas convexos de grande porte, onde há a possibilidade de aproveitar a estrutura especial de problemas.

Finalmente, com o grande volume de aplicações descritas neste relatório, foi possível perceber os vários problemas práticos que podem ser resolvidos com otimização convexa. Mesmo com estruturas diferentes, todos compartilham diversas qualidades em comum. Devido a essas características comuns, os métodos se tornam muito eficientes para toda essa classe de problemas, e esses algoritmos podem servir de base para outros mais gerais.

Referências

[1] M. S. Bazaraa & J.J. Jarvis, *Linear Programming and Network Flows.* John Wiley and Sons, New York, NY, 1977.

[2] D. P. Bertsekas, A. Nedic & A. E. Ozdaglar, *Convex Analysis and Optimization.* Athena Scientific, Belmont, MA, 2003.

[3] S. Boyd & L. Vandenberghe, *Convex Optimization.* Cambridge University Press, Cambridge, UK, 2004.

[4] C. C. Gonzaga, *Algoritmos de Pontos Interiores para Programação Linear.* 17º Colóquio Brasileiro de Matemática, IMPA, 1989.

[5] D. G. Luenberger & Y. Ye, *Linear and Nonlinear Programming.* Third Edition, Springer, New York, NY, 2008.

[6] M. L. Schuverdt, *Métodos de Lagrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante.* Tese de Doutorado em Matemática Aplicada, IMECC – UNICAMP, Campinas, SP, 2006.