Universidade Federal de Minas Gerais
Escola de Engenharia
Departamento de Engenharia Elétrica

Notas de Aula de Otimização

Jaime A. Ramírez
Felipe Campelo
Frederico G. Guimarães
Lucas S. Batista
Ricardo H. C. Takahashi

Sumário

Lista de Figuras

Capítulo 2

Condições de Otimalidade

2.1 Introdução

No capítulo anterior, vimos, de maneira intuitiva a definição de função objetivo e funções de restrição de igualdade e desigualdade, e as diferentes estratégias e princípios que podem ser utilizados para resolver problemas de otimização. Abordou-se também as diferentes características que as funções objetivo e de restrição podem assumir e a implicação dessas características na possível estratégia a ser utilizada na solução do problema de otimização.

Neste capítulo discutiremos a caracterização da função objetivo, das funções de restrição, e as condições de otimalidade que nos auxiliarão encontrar a solução para o problema de otimização definido matematicamente como:

$$
\begin{aligned}
\boldsymbol{x}^* = & \arg \min_{\boldsymbol{x}} \ f(\boldsymbol{x}) \\
& \text{sujeito a:} \begin{cases} g_i(\boldsymbol{x}) \leq 0, i = 1, \ldots, p \\ h_j(\boldsymbol{x}) = 0, j = 1, \ldots, q \end{cases}
\end{aligned}
\tag{2.1}
$$

sendo que $\boldsymbol{x} \in \mathbb{R}^n$, $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^1$, $g_i(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^p$ e $h_j(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^q$.

A escolha de técnicas adequadas para tratar o problema definido em (2.1) depende da natureza das funções $f(\boldsymbol{x})$, $g_i(\boldsymbol{x})$, $h_j(\boldsymbol{x})$. Não há uma técnica de otimização que seja universal, no sentido de ser a melhor técnica para otimizar quaisquer funções, e a escolha das técnicas, frequentemente, basea-se em informações sobre o problema em questão.

Para nos orientar nessa caracterização, apresentaremos os seguintes pontos, relacionadas com a questão de *o quê são* as soluções do problema (2.1):

1. Dado o funcional $f(\cdot)$ o que são os pontos de mínimo desse funcional, ou seja, o que são as soluções do problema de otimização?

2. O que são os pontos de mínimo local desse funcional, se são dadas também as restrições $g_i(\boldsymbol{x}) \leq 0$ e $h_j(\boldsymbol{x}) = 0$?

3. Dado um ponto $\boldsymbol{x} \in \mathbb{R}^n$, que tipo de testes podem ser realizados para determinar se esse ponto é ou não um ponto de mínimo de $f(\cdot)$, nos dois casos

anteriores?

Respostas a essas questões serão fornecidas tanto no sentido local (mínimos locais) quanto global (mínimos globais).

Antes porém de deduzir as condições de otimalidade do problema (2.1), é necessário agregar alguma informação que seja útil para se decidir *como proceder* para encontrar tais soluções. Algumas caracterizações úteis, definidas neste capítulo, são:

1. Função, funcional, continuidade e diferenciabilidade;
2. Curvas de nível, superfície de nível, região subnível;
3. Convexidade, quasi-convexidade, e não convexidade;
4. Unimodalidade e multimodalidade.

Cada uma dessas informações a respeito da função, se estiver disponível, permite a agregação de um certo tipo de informação de caráter global que auxilia o processo de otimização.

A dedução das condições de otimimalidade serão utilizadas para a compreensão e concepção dos algoritmos que utilizam a estratégia de direção de busca para a solução de problemas de otimização. Leitura complementar pode ser encontrada em [1]- [2].

2.2 Caracterização de Funções

2.2.1 Função e Funcional

Função

Uma função é uma relação que associa de maneira única membros de um conjunto A com membros de um conjunto B. Em termos mais formais, uma função do conjunto A para o conjunto B é um objeto f tal que todo elemento "a" que pertence ao conjunto A é associado de maneira única com o objeto $f(a)$ que pertence ao conjunto B, $f(a) \in B$. Em termos matemáticos:

Definição 2.1 (Função) *Sejam A e B dois conjuntos com membros $a_i, \ldots, a_m$ e $b_i, \ldots, b_n$, respectivamente. Uma função f que associa de maneira única membros de A em B é definida como:*

$$f : A \mapsto B \tag{2.2}$$

□

Funcional

Um funcional é uma função que retorna um único valor, i.e. um número escalar. Em termos matemáticos:

Definição 2.2 (Funcional) *Se $f(\cdot)$ é um funcional então:*

$$f : \mathbb{R}^n \mapsto \mathbb{R}^1 \tag{2.3}$$

□

2.2.2 Superfície de Nível e Região Subnível

A caracterização de funções adotada neste capítulo se fundamenta nos conceitos de *superfície de nível* e de *região subnível*.

Superfície de Nível

Definição 2.3 (Superfície de Nível) *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$. A superfície de nível $S(f, \alpha)$, associada ao nível α, é definida como:*

$$S(f, \alpha) = \{\boldsymbol{x} \in C \mid f(\boldsymbol{x}) = \alpha\} \tag{2.4}$$

$\square$

O conceito e definição de superfície de nível pode ser ilustrado usando uma função quadrática. Para esse fim, vamos utilizar a seguinte função:

$$\begin{aligned} & f(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}_0)' Q (\boldsymbol{x} - \boldsymbol{x}_0) \\ & Q = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 1 \end{bmatrix} \qquad \boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned} \tag{2.5}$$

que está ilustrada na Fig. 2.1.

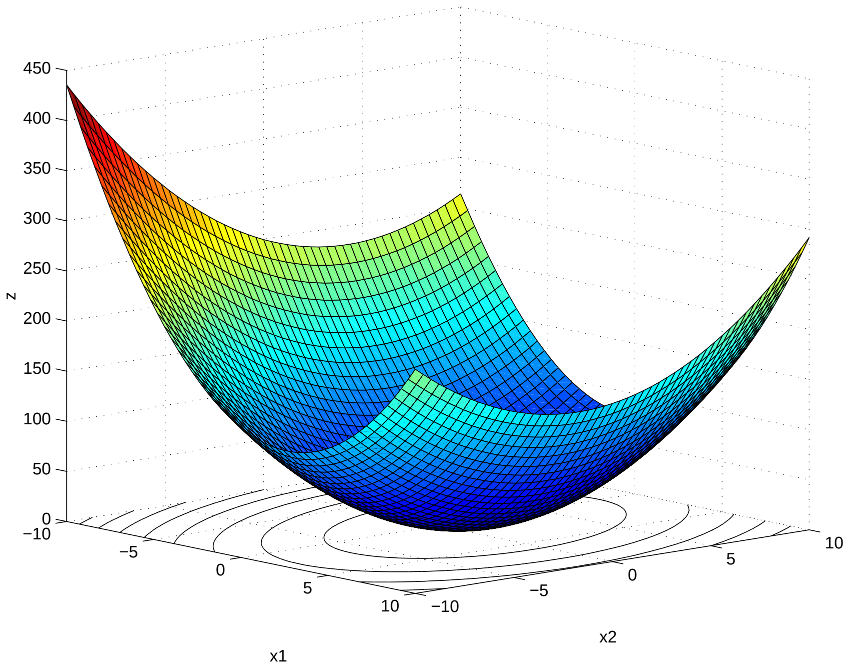

Figura 2.1: Superfície que representa o gráfico da função quadrática (2.5).

As curvas de nível estão representadas no plano $x_1 \times x_2$ na Fig. 2.1. Cada curva contém os pontos que possuem o mesmo valor de função.

Região Sub-Nível

Definição 2.4 (Região Sub-Nível) *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$. A região de sub-nível $R(f, \alpha)$, associada ao nível α, é definida como:*

$$R(f, \alpha) = \{\boldsymbol{x} \in C \mid f(\boldsymbol{x}) \leq \alpha\} \tag{2.6}$$

$\square$

O conceito de região de sub-nível para a função quadrática (2.5) está ilustrado na Fig. 2.2 a seguir.

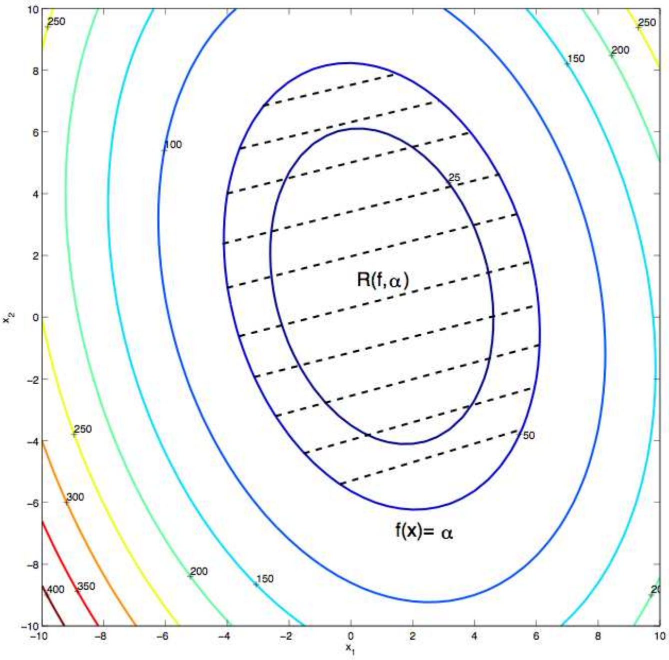

Figura 2.2: Gráfico ilustrando uma região sub-nível $R(f, \alpha)$ - região hachurada - da função quadrática. Indica-se, também, várias curvas de nível, i.e. curvas para as quais a função quadrática possui um mesmo valor α. Destaca-se a curva de nível $\alpha = 50$.

Normalmente $S(f, \alpha)$ corresponde a uma fronteira de $R(f, \alpha)$, embora seja possível escolher α de forma que isso não ocorra. Claramente é válida uma relação de ordenação das regiões de sub-nível de uma função.

Proposição 2.1 *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$. As regiões de sub-nível dessa função obedecem a:*

$$R(f, \alpha_1) \supset R(f, \alpha_2) \Leftrightarrow \alpha_1 > \alpha_2 \tag{2.7}$$

$\square$

Pode-se pensar os problemas de otimização como sendo equivalentes a um problema de determinar pontos que estejam sucessivamente no interior de regiões de sub-nível cada vez menores (de menor valor de α). Em linhas gerais, se constroem

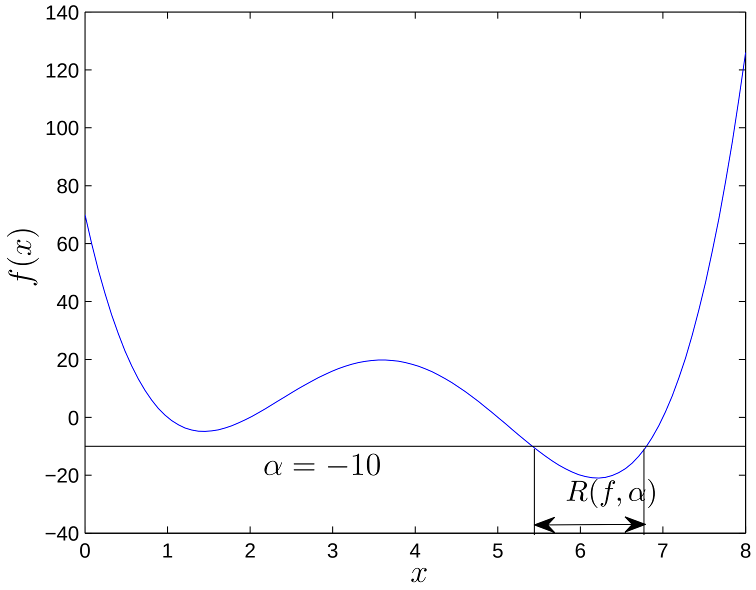

Figura 2.3: Valor de α para o qual a região de subnível é conexa.

algoritmos que produzem tais sequências de pontos. Consequentemente, produz-se uma "contração" do conjunto definido pelas regiões de sub-nível, sendo a solução atingida quando a região de sub-nível se degenerar no ponto de ótimo.

As regiões de sub-nível, analisadas sob o ponto de vista topológico, definem uma categorização importante para as funções.

2.2.3 Unimodalidade e Multimodalidade

Definição 2.5 (Função Unimodal) *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$. Diz-se que $f(\cdot)$ é unimodal se $R(f,\alpha)$ é conexo para todo $\alpha \in \mathbb{R}$. Diz-se ainda que $f(\cdot)$ é estritamente unimodal se, além disso, $R(f,\alpha)$ é um conjunto compacto para todo $\alpha \in \mathbb{R}$.* $\square$

Por simetria, define-se ainda:

Definição 2.6 (Função Multimodal) *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$. Diz-se que $f(\cdot)$ é multimodal se existe $\alpha \in \mathbb{R}$ tal que $R(f,\alpha)$ não é conexo.* $\square$

As Figs. 2.3 e 2.4 mostram respectivamente uma região de subnível conexa e uma região de subnível desconexa. Para esta função existe um valor de α para o qual $R(f,\alpha)$ não é um conjunto conexo, caracterizando uma função multimodal.

Nota 2.1 *Note-se que uma função unimodal pode possuir múltiplos mínimos, desde que o conjunto deste seja conexo, e uma função estritamente unimodal também pode possuir múltiplos mínimos, desde que o conjunto destes seja conexo compacto. O primeiro caso ocorre, por exemplo, para a função*

$$f(\boldsymbol{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2.8}$$

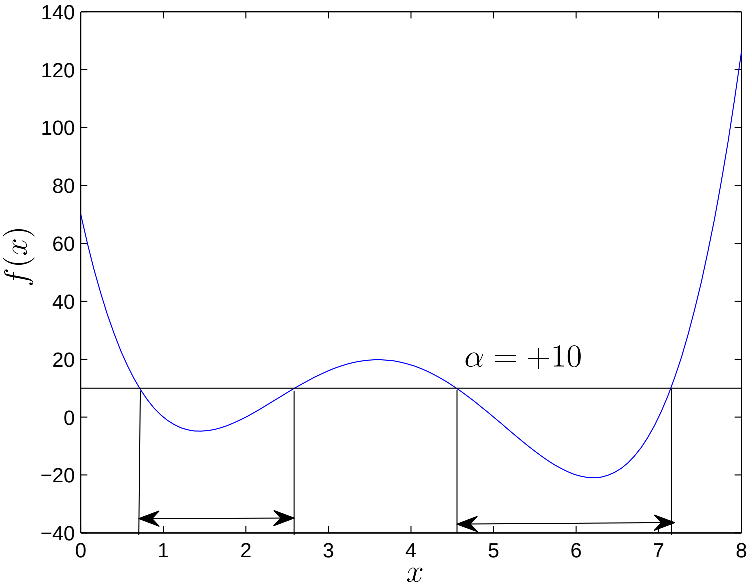

Figura 2.4: Valor de α para o qual a região de subnível é desconexa.

para a qual todos os pontos que pertencem ao eixo $x_1 = 0$ (esses pontos formam um conjunto conexo mas não compacto) constituem mínimos. Essa observação revela uma diferença fundamental das noções de função unimodal e função multimodal aqui definidas em relação às usualmente encontradas na literatura. Os autores acreditam que no formato apresentado neste texto essas definições ganham maior funcionalidade para articularem a teoria de otimização.

$\diamondsuit$

2.2.4 Bacias de Atração

Ao redor de mínimos locais, sempre haverá regiões nas quais a função se comportará de maneira unimodal. Tais regiões são definidas como *bacias de atração* associadas a tais mínimos. Para estabelecer essa definição, é necessário definir preliminarmente a *região conexa de sub-nível*.

Definição 2.7 (Região Conexa de Sub-Nível) *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$, seja a região de sub-nível $R(f, \alpha)$, associada ao nível α, e seja um ponto $\boldsymbol{x}_0 \in R(f, \alpha)$. A região conexa de sub-nível $R(f, \alpha, \boldsymbol{x}_0)$ é definida como o maior subconjunto conexo de $R(f, \alpha)$ que contém $\boldsymbol{x}_0$.* $\square$

Agora é possível definir *bacia de atração*.

Definição 2.8 (Bacia de Atração) *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$, e seja $\boldsymbol{x}^* \in C$ um mínimo local de $f(\cdot)$. A bacia de atração de $\boldsymbol{x}^*$ é definida como a maior região conexa de sub-nível associada a $\boldsymbol{x}^*$, sendo α^* o nível correspondente, tal que a função restrita a essa região*

$$f(\cdot) : \; R_c\,(f, \alpha^*, \boldsymbol{x}^*) \;\mapsto \mathbb{R} \tag{2.9}$$

é unimodal. A bacia de atração é dita estrita se nessa região a função é estritamente unimodal. □

2.2.5 Continuidade e Diferenciabilidade

Suposições de continuidade e de diferenciabilidade das funções são importantes na definição de alguns métodos de otimização. De maneira intuitiva, uma função contínua é aquela para a qual uma pequena variação na entrada gera uma pequena variação no resultado da função, isto é, a função não possui "saltos". Uma definição formal é dada a seguir.

Definição 2.9 (Função contínua) *Uma função* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ *é contínua se* $\forall\ \boldsymbol{x}_0 \in C$*:*

1. $f(\boldsymbol{x}_0)$ *é definido;*
2. $\lim\limits_{\boldsymbol{x} \to \boldsymbol{x}_0} f(\boldsymbol{x}) = f(\boldsymbol{x}_0)$.

□

Definição 2.10 (Função diferenciável) *Uma função* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ *é diferenciável se* $\forall\ \boldsymbol{x}_0 \in C$ *existe o vetor gradiente:*

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{2.10}$$

□

Essas suposições nos permitem extrair propriedades interessantes a respeito de suas superfícies de nível e bacias de atração.

Proposição 2.2 *Seja* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$*. Se* $f(\cdot)$ *é contínua no domínio* C*, então*

$$dist(S(f, \alpha_1), S(f, \alpha_2)) > 0\ \forall\ (\alpha_1, \alpha_2)\ |\ \ |\alpha_1 - \alpha_2|\ > 0 \tag{2.11}$$

sendo $dist(\cdot, \cdot)$ *a função distância.* □

Corolário 2.1 *Superfícies de nível de funções contínuas não se tocam nem se cruzam.* □

Proposição 2.3 *Seja* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$*. Se* $f(\cdot)$ *é diferenciável no domínio* C*, então toda superfície de nível* $S(f, \alpha)$ *é suave, sendo o hiperplano tangente à superfície em cada ponto perpendicular ao gradiente da função no ponto.* □

A hipótese de diferenciabilidade de uma função permite elaborar estratégias de otimização baseadas no fato de que o gradiente de um função (que, no caso de funções diferenciáveis, é sempre bem definido) indica quais são as direções do espaço para as quais, partindo-se de um ponto, ocorre localmente a diminuição da função. Isso equivale à determinação das direções para as quais se caminha para regiões de sub-nível inferiores. A proposição a seguir formaliza esse fato, que deriva da proposição anterior.

Proposição 2.4 *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ uma função diferenciável no domínio C, seja $\boldsymbol{x}_0$ um ponto pertencente à superfície de nível $S(f, \alpha)$, e seja $\nabla f(\boldsymbol{x}_0)$ o gradiente de $f(\cdot)$ no ponto $\boldsymbol{x}_0$. Seja ainda um vetor $\boldsymbol{d} \in \mathbb{R}^n$. Então, se*

$$\boldsymbol{d} \cdot \nabla f(\boldsymbol{x}_0) < 0 \tag{2.12}$$

então existe $\epsilon > 0$ tal que:

$$f(\boldsymbol{x}_0 + \epsilon \boldsymbol{d}) < f(\boldsymbol{x}_0) \tag{2.13}$$

□

Dizemos que $\boldsymbol{d}$ é uma direção minimizante de $f(\cdot)$ no ponto $\boldsymbol{x}_0$.

Por fim, o subgradiente é uma generalização do vetor gradiente para o caso de funções não diferenciáveis.

Definição 2.11 (Subgradiente) *Seja $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}$. Um funcional linear f^{sb} é um subgradiente de $f(\cdot)$ no ponto $\boldsymbol{x}_0$ se:*

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + f^{sb}(\boldsymbol{x} - \boldsymbol{x}_0) \quad , \quad \forall\, \boldsymbol{x} \tag{2.14}$$

□

Por exemplo, seja a função $f(x) = |x|$. A derivada desta função é:

$$f'(x) = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases} \tag{2.15}$$

No ponto $x = 0$ a derivada não é definida, entretanto pode-se definir o subgradiente como qualquer número real no intervalo $[-1, 1]$.

A Fig. ? ilustra o conceito de subgradiente para uma função de duas variáveis. Qualquer vetor no cone formado pelos vetores $\boldsymbol{v}_1$ e $\boldsymbol{v}_2$ é um subgradiente de $f(\cdot)$ no ponto $\boldsymbol{x}_0$, em que o vetor gradiente não é definido.

2.2.6 Convexidade, Quasi-Convexidade e Não Convexidade

Conjunto Convexo

Definição 2.12 (Conjunto Convexo) *Diz-se que um conjunto $C \in \mathbb{R}^n$ é convexo se para quaisquer vetores $\boldsymbol{x},\ \boldsymbol{y} \in C$,*

$$\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{y} \ \in \ C \tag{2.16}$$

para todo $\alpha \in [0, 1]$. □

Em outras palavras, um conjunto C é dito convexo se todos os pontos do segmento de reta que une dois pontos quaisquer de C também pertencem a C. Isso está ilustrado na Fig. 2.5.

Outro tipo de informação que pode ser útil em processos de otimização diz respeito à convexidade das funções.

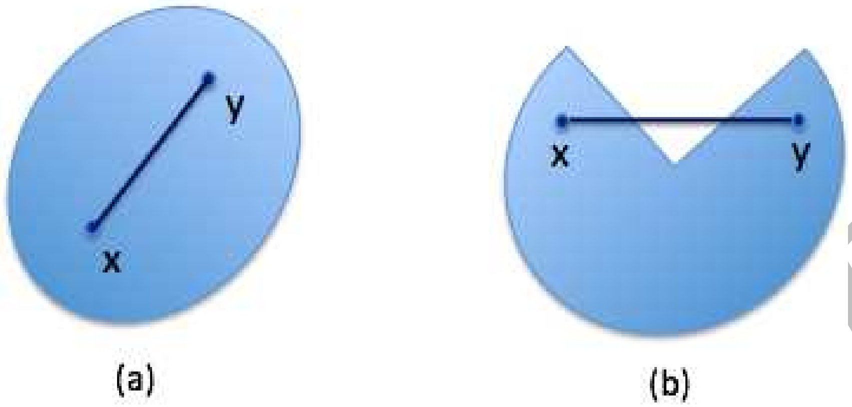

Figura 2.5: Representação: (a) Conjunto convexo, (b) Conjunto não convexo

Função Convexa

Definição 2.13 (Função Convexa) *Diz-se que uma função $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ definida sobre um conjunto convexo C é convexa se para quaisquer $\boldsymbol{x},\ \boldsymbol{y} \in C$,*

$$f\left(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}\right) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \tag{2.17}$$

para todo $\alpha \in [0,1]$. Se para quaisquer $\boldsymbol{x},\ \boldsymbol{y} \in C$, sendo $\boldsymbol{x} \neq \boldsymbol{y}$ e $0 < \alpha < 1$, a desigualdade é estrita, então $f(\cdot)$ é estritamente convexa. □

Analogamente, $f(\cdot)$ é (estritamente) côncava se $-f(\cdot)$ for (estritamente) convexa.

Proposição 2.5 (Caracterizações de Funções Convexas) *Seja $f(\cdot)$ uma função duas vezes diferenciável, sobre um conjunto convexo $C \subset \mathbb{R}^n$. Então são equivalentes as afirmativas a seguir:*

i. $f\left(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}\right) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \quad \forall\ \alpha \in [0,1]$

ii. $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})'(\boldsymbol{y} - \boldsymbol{x}) \quad \forall\ \boldsymbol{x},\ \boldsymbol{y} \in C$

iii. $H(\boldsymbol{x}) \geq 0 \quad \forall\ \boldsymbol{x} \in C$

sendo $\nabla f(\boldsymbol{x})$ o vetor gradiente no ponto $\boldsymbol{x}$ e $H(\boldsymbol{x})$ a matriz Hessiana [1] no ponto $\boldsymbol{x}$. □

Como no caso de conjuntos convexos, é possível obter funções convexas a partir de combinações convexas.

Proposição 2.6 (Combinações Convexas) *Sejam $f_i(\cdot) : C_i \subset \mathbb{R}^n \mapsto \mathbb{R}$ funções convexas definidas sobre conjuntos convexos C_i , $i = 1, \ldots, m$. Então:*

[1] A matriz Hessiana é obtida a partir da derivada segunda de $f(\cdot)$ em relação a $\boldsymbol{x}$.

i. $\alpha f_i(\cdot)$ *é convexa sobre* C_i, $\forall\ \alpha \geq 0$

ii. $\displaystyle\sum_{i=1}^{m} \alpha_i f_i(\cdot)$ *é convexa sobre* $\displaystyle\bigcap_{i=1}^{m} C_i$ *para* $\alpha_i \geq 0\ ,\ i = 1, \ldots, m$

$\square$

Proposição 2.7 *Seja* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ *uma função convexa sobre* C *convexo. Então a região de sub-nível* $R(f, \alpha)$ *é convexa para todo* $\alpha \in \mathbb{R}$. $\square$

A recíproca não é verdadeira.

A convexidade de $R(f, \alpha)$ define um novo tipo de função, as funções *quasi-convexas.*

Definição 2.14 (Função Quasi-Convexa) *Seja* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ *uma função tal que suas regiões de sub-nível* $R(f, \alpha)$ *são convexas para todo* $\alpha \in \mathbb{R}$. *Neste caso, diz-se que* $f(\cdot)$ *é quasi-convexa no domínio* C. $\square$

Proposição 2.8 *Se* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ *é uma função quasi-convexa, então:*

$$f\left(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}\right) \leq \max\left\{f(\boldsymbol{x}), f(\boldsymbol{y})\right\} \quad \forall\ \ \boldsymbol{x},\ \boldsymbol{y} \in C\ ,\ \forall\ \alpha \in [0,1] \tag{2.18}$$

$\square$

Outro resultado envolvendo conjuntos e funções convexas pode ser obtido a partir da definição de Epígrafo:

Definição 2.15 (Epígrafo) *O* epígrafo *de uma função* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ *é definido como:*

$$[f, C] = \{(\boldsymbol{x}, \theta) \in \mathbb{R}^n \times \mathbb{R}\ :\ \boldsymbol{x} \in C\ ,\ f(\boldsymbol{x}) \leq \theta\} \tag{2.19}$$

$\square$

Proposição 2.9 *Uma função* $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ *definida sobre* C *convexo é convexa se, e somente se,* $[f, C]$ *é um conjunto convexo.* $\square$

Como todo conjunto convexo, o epígrafo de uma função convexa admite hiperplanos suporte em qualquer ponto de sua fronteira.

A convexidade de funções pode ser relacionada com as regiões de sub-nível, superfícies de nível e bacias de atração.

Proposição 2.10 *Todas as regiões de sub-nível de uma função convexa num domínio convexo são conjuntos convexos.* $\square$

Proposição 2.11 *Uma função convexa em um domínio convexo possui uma única bacia de atração, a qual é um conjunto convexo.* $\square$

Proposição 2.12 *Seja uma função convexa $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$, seja um ponto qualquer $\boldsymbol{x}_0 \in \mathbb{R}^n$, e seja $s(\boldsymbol{x}_0) \in \mathbb{R}^n$ um vetor subgradiente da função no ponto. Então a região de sub-nível que possui o ponto $\boldsymbol{x}_0$ em sua fronteira está contida no semi-espaço fechado negativo definido pelo vetor subgradiente no ponto $\boldsymbol{x}_0$, ou seja:*

$$\begin{aligned} &E_s = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{x} - \boldsymbol{x}_0) \cdot s(\boldsymbol{x}_0) \leq 0\} \\ &R(f, f(\boldsymbol{x}_0)) \subset E_s \end{aligned} \tag{2.20}$$

□

2.2.7 Mínimo Local e Mínimo Global

Introduzimos o conceito de mínimo local como o ponto $\boldsymbol{x}^*$, para o qual qualquer vetor $\boldsymbol{x}$ na vizinhança ϵ de $\boldsymbol{x}^*$ implica em $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$. Matematicamente:

Definição 2.16 (Mínimo Local) *Seja $f(\cdot) : C \subset \mathbb{R}^n \mapsto \mathbb{R}$. Um ponto $\boldsymbol{x}^*$ é um mínimo local de $f(\cdot)$ sobre C se existe $\epsilon > 0$ tal que*

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \quad , \quad \forall\ \boldsymbol{x} \in V(\boldsymbol{x}^*, \epsilon) \cap C \tag{2.21}$$

onde $V(\boldsymbol{x}^*, \epsilon) \triangleq \{\boldsymbol{x} : \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \epsilon\}$. O ponto $\boldsymbol{x}^* \in C$ é um mínimo local estrito se vale a desigualdade estrita. □

Naturalmente, o conjunto C é o subconjunto do espaço $\mathbb{R}^n$ definido pelas restrições:

$$C \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) \leq 0 \ ; \ i = 1, \ldots, p \ ; \ h_j(\boldsymbol{x}) = 0 \ ; \ j = 1, \ldots, q\} \tag{2.22}$$

É possível, a partir desta definição, construir a definição de mínimo global do funcional. Se for possível escolher $\epsilon > 0$ tal que $V(\boldsymbol{x}^*, \epsilon) \cap C = C$, então $\boldsymbol{x}^*$ é um mínimo global de $f(\cdot)$ sobre C. O mínimo global é ainda estrito se a desigualdade for satisfeita de modo estrito.

2.3 Problema Exemplo

Consideremos o problema:

$$\boldsymbol{x}^* = \ \arg\ \min_{\boldsymbol{x}}\ f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$

$$\text{sujeito a:} \begin{cases} g_1(\boldsymbol{x}) : 3x_1 + 2x_2 \leq 12 \\ h_1(\boldsymbol{x}) : x_1 + x_2 = 5 \\ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \end{cases} \tag{2.23}$$

que representa a minimização de uma função de duas variáveis $f(x_1, x_2)$, sujeita a uma restrição de desigualdade $g_1(\boldsymbol{x})$ e outra de igualdade $h_1(\boldsymbol{x})$. Em geral, como viu-se no capítulo anterior, problemas práticos podem ter mais do que uma restrição de igualdade e desigualdade. Entretanto, para facilitar a análise, consideraremos apenas $g_1(\cdot)$ e $h_1(\cdot)$. Por ser uma função de apenas duas variáveis, $f(x_1, x_2)$ pode ser representada no plano $(x_1 \times x_2)$ através de curvas de nível, conforme indicado na Figura 2.6. As duas restrições, $g_1(\cdot)$ e $h_1(\cdot)$, estão também representadas na Figura 2.6. O objetivo nesta seção é explorar conceitos intuitivos e geométricos para caracterizar o mínimo de (2.23). Este problema exemplo será utilizado nas seções seguintes para deduzir as condições necessárias e suficientes de otimização.

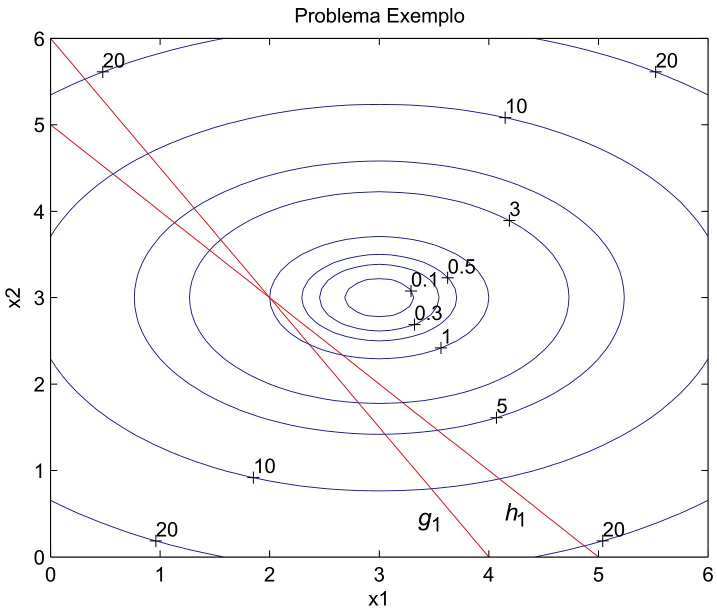

Figura 2.6: Ilustração gráfica do problema exemplo.

2.3.1 Soluções Gráficas

As soluções gráficas serão apresentadas, separadamente, para três tipos de problemas: irrestritos, com restrição de igualdade, e com restrição de desigualdade, seguindo o exemplo definido em (2.23). Em geral, esses três tipos de problemas, analisados separadamente ou em conjunto, representam os tipos possíveis de problemas de otimização.

Problemas Irrestritos

A partir da equação (2.23) pode-se definir o problema irrestrito da seguinte maneira:

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$
$$\text{sujeito a:} \left\{ \; 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \right. \tag{2.24}$$

que como o nome sugere, não possui nenhuma função de restrição imposta a $f(\cdot)$. Neste caso, os limites inferiores e superiores de $\boldsymbol{x}$ definem a região factível. Analisando as curvas de nível da função objetivo $f(\cdot)$, observa-se por inspeção que o valor mínimo ocorre para ($x_1^* = 3$ e $x_2^* = 3$), já que as curvas de nível de $f(\cdot)$ diminuem de valor a medida que se aproxima desse ponto, conforme indicado na Figura 2.7. No ponto solução encontrado por inspeção, constata-se que $f(\boldsymbol{x}^*) = 0$, para tanto basta substituir os valores de $x_1^* = 3$ e $x_2^* = 3$ na equação de $f(\cdot)$.

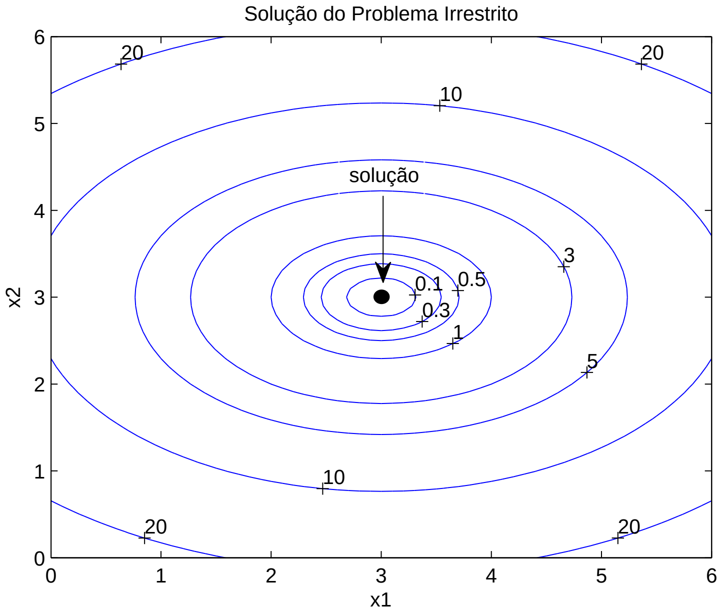

Figura 2.7: Solução gráfica do problema exemplo – irrestrito.

Restrição de Desigualdade

A partir de (2.25) pode-se definir o problema com restrição de desigualdade:

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} \; f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$

$$\text{sujeito a:} \begin{cases} g_1(\boldsymbol{x}) : 3x_1 + 2x_2 \leq 12 \\ \\ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \end{cases} \tag{2.25}$$

Neste caso, ao incluir a restrição de desigualdade $g_1(\cdot) \leq 0$, força-se que a solução do problema (2.25) esteja na região factível, conforme indicado na Figura 2.8. A região factível passa a ser o conjunto dos pontos que satisfaz $g_1(\cdot) \leq 0$ e os limites superiores e inferiores de x_1 e x_2. A solução, em princípio, poderia ser qualquer ponto (x_1, x_2) pertencente a região factível. Entretanto, por inspeção, pode-se identificar que o mínimo é o ponto (x_1^*, x_2^*) definido na curva de nível de $f(\cdot)$ que tangencia $g_1(\cdot)$. Observe que no ponto solução (x_1^*, x_2^*), $\nabla f(\cdot)$ está, exatamente, no sentido oposto de $\nabla g_1(\cdot)$. Essa relação entre os gradientes é a base para estabelecer as condições de otimalidade de primeira ordem para problemas com restrições de desigualdade.

Restrição de Igualdade

A partir da equação (2.23), pode-se definir o problema com a restrição de igualdade:

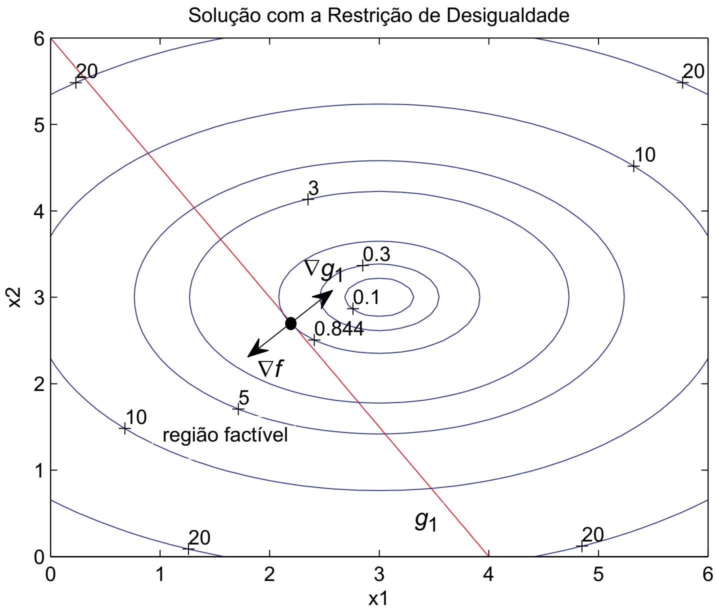

Figura 2.8: Solução gráfica do problema exemplo – restrição de desigualdade.

$$
\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} \; f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2
$$

$$
\text{sujeito a:} \begin{cases} h_1(\boldsymbol{x}) : x_1 + x_2 = 5 \\ \\ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \end{cases} \tag{2.26}
$$

Neste caso, ao incluir a restrição de igualdade $h_1(\cdot)$, força-se que a solução do problema (2.26) esteja sobre a reta $h_1(\cdot)$, uma vez que a solução do problema (2.26) tem que satisfazer a equação de $h_1(\cdot)$. A primeira observação a ser feita é que ao reduzir os pontos solução possíveis aos pontos contidos sobre a reta $h_1(\cdot)$, está-se, de fato, reduzindo a região factível. A solução, em princípio, poderia ser qualquer ponto (x_1, x_2) sobre a reta $h_1(\cdot)$. Entretanto, por inspeção, pode-se identificar que o mínimo é o ponto (x_1^*, x_2^*) definido na curva de nível de $f(\cdot)$ que tangencia $h_1(\cdot)$, conforme indicado na Figura 2.9. Observe que nesse ponto $\nabla f(\cdot)$ está, exatamente, no sentido oposto de $\nabla h_1(\cdot)$ (o caso geral requer que os gradientes estejam alinhados). Essa relação entre os gradientes é a base para estabelecer as condições de otimalidade de primeira ordem para problemas com restrições de igualdade.

Restrições de Desigualdade e Igualdade

O problema com ambas as restrições é apresentado em (2.27).

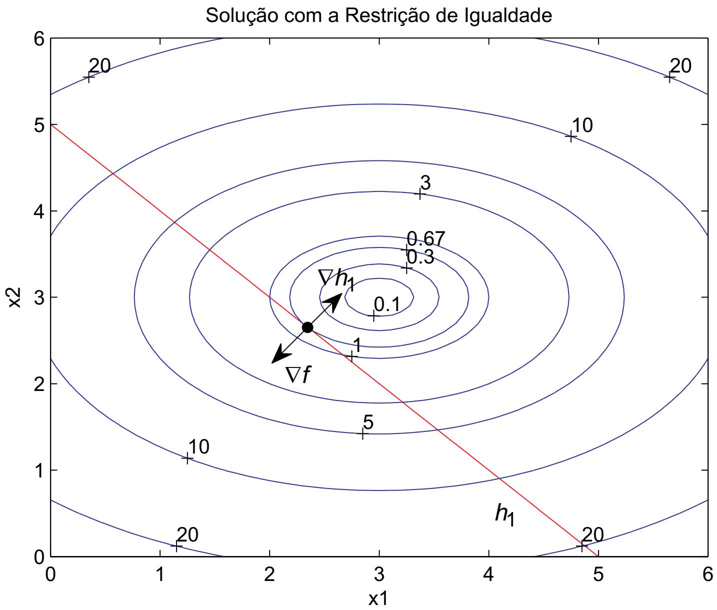

Figura 2.9: Solução gráfica do problema exemplo – restrição de igualdade.

$$\boldsymbol{x}^* = \arg\ \min_{\boldsymbol{x}}\ f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$
$$\text{sujeito a:} \begin{cases} g_1(\boldsymbol{x}) : 3x_1 + 2x_2 \leq 12 \\ h_1(\boldsymbol{x}) : x_1 + x_2 = 5 \\ 0 \leq x_1 \leq 6 \quad 0 \leq x_2 \leq 6 \end{cases} \tag{2.27}$$

Este caso é, em princípio, o caso geral que envolve a função objetivo e restrições de igualdade e desigualdade. A região factível passa a ser a região que satisfaz simultaneamente $h_1(\cdot) = 0$ e $g_1(\cdot) \leq 0$, respeitando-se os limites superiores e inferiores de x_1 e x_2. Por inspeção, pode-se identificar que o mínimo (x_1^*, x_2^*) é o ponto de interseção entre as curvas $h_1(\cdot)$ e $g_1(\cdot)$, uma vez que ambas restrições tem que ser satisfeitas. Resolvendo-se o sistema de equações formado por $h_1(\cdot)$ e $g_1(\cdot)$, obtém-se ($x_1^* = 2$ e $x_2^* = 3$). Nesse ponto $f(\boldsymbol{x}^*) = 1$. Observe que neste caso, no ponto solução, o somatório dos gradientes de $f(\cdot)$, $h_1(\cdot)$ e $g_1(\cdot)$ não se anula automaticamente. Para que isso aconteça é necessário que o $\nabla h_1(\cdot)$ ou o $\nabla g_1(\cdot)$ tenha o seu sentido invertido, i.e. seja multiplicado por uma constante com sinal negativo. Veremos nas seções seguintes que isso só se verifica com $\nabla h_1(\cdot)$. Essa relação entre os gradientes é a base para estabelecer as condições de otimalidade de primeira ordem para o caso geral envolvendo problemas com restrições de igualdade e desigualdade.

2.4 Condições Analíticas: Problemas Irrestritos

Nesta seção, apresentaremos as condições analíticas necessárias e suficientes que permitem afirmar se a solução de um determinado problema de otimização é de

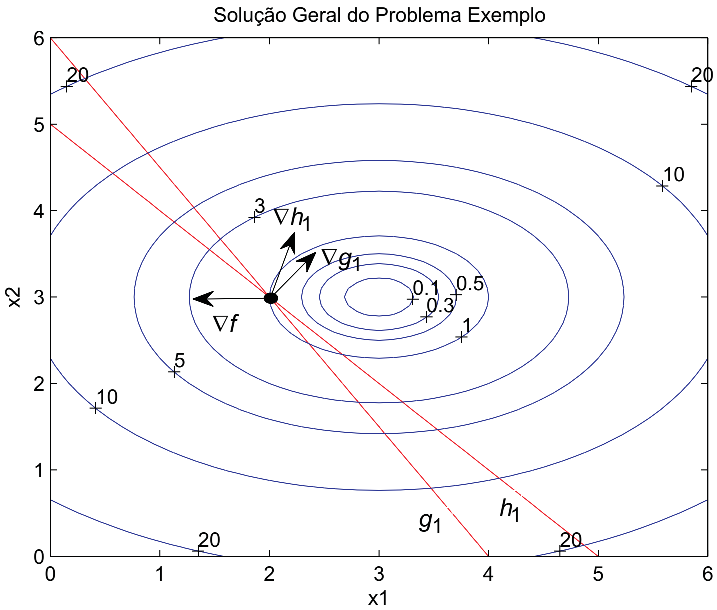

Figura 2.10: Solução gráfica do problema exemplo – restrição de igualdade e desigualdade.

fato a solução ótima. Essas condições serão utilizadas nos capítulos seguintes, como critérios de parada e convergência, quando serão estudados os métodos numéricos determinísticos para problemas irrestritos e restritos. Similarmente à seção anterior, utilizaremos conceitos intuitivos com o auxílio de interpretação geométrica para apresentação das condições necessárias e suficientes. Por conveniência, suporemos que a função objetivo $f(\cdot)$ possui apenas um mínimo e que a solução encontra-se no interior da região factível.

Para facilitar a dedução analítica, utilizaremos como exemplo o mesmo problema (2.23) da seção anterior, envolvendo uma função objetivo $f(\cdot)$ de duas variáveis e apenas uma restrição de desigualdade $g(\cdot)$ e outra de igualdade $h(\cdot)$. Embora o problema de otimização inclua apenas uma restrição de desigualdade e outra de igualdade, a análise que apresentaremos a seguir pode ser generalizada.

O problema irrestrito pode ser definido como:

$$\begin{aligned} \boldsymbol{x}^* = \ & \arg \min_{\boldsymbol{x}} \ f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2 \\ & \text{sujeito a:} \left\{ \ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \right. \end{aligned} \tag{2.28}$$

Conforme já observado na seção anterior, a Figura 2.7 ilustra as curvas de nível da equação (2.28) no plano $x_1 \times x_2$, ao passo que a Figura 2.11 ilustra o gráfico 3D da mesma função. Um plano tangente ao ponto $(x_1 = 3, \ x_2 = 3)$ foi desenhado para realçar o ponto de mínimo da função. A Figura 2.11 será utilizada para identificar as propriedades de $f(\cdot)$ no ponto de mínimo.

Analisando a Figura 2.11, observa-se que o mínimo ocorre em $(x_1^* = 3, \ x_2^* = 3)$, e nesse ponto $f(x_1, x_2) = 0$. Se os valores de x_1 e ou x_2 variassem de um pequeno valor, em qualquer direção, o valor da função $f(\cdot)$ certamente aumentaria, uma vez que $\boldsymbol{x}^*$ é o ponto mínimo da superfície convexa que representa a função $f(\cdot)$.

Voltando ao exemplo, representaremos a variação na vizinhança do ponto ótimo como $\Delta\boldsymbol{x}$, e a variação do valor ótimo da função como $\Delta f(\cdot)$. Por observação direta fica evidente que o *mínimo* deve ser um ponto que satisfaça:

$$\Delta f > 0 \quad , \quad \forall \quad \Delta\boldsymbol{x} \tag{2.29}$$

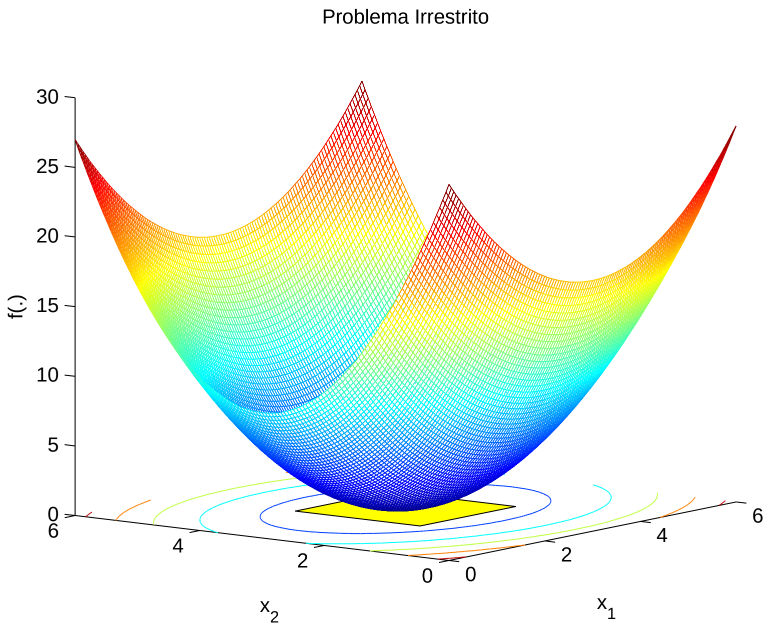

Figura 2.11: Solução gráfica 3D do problema irrestrito.

2.4.1 Condições de Primeira Ordem

O conceito desenvolvido em (2.29) pode ser aplicado no limite, isto é, para incrementos infinitesimais dx_1 e dx_2 sobre $\boldsymbol{x}^*$. A função $f(\cdot)$ pode ser aproximada por um plano tangente no ponto solução, por exemplo utilizando os primeiros termos de uma série de Taylor. A partir do ponto de mínimo, qualquer variação no plano tangente não mudará o valor da função $f(\cdot)$, uma vez que o valor da função $f(\cdot)$ é constante no plano; consequentemente $df = 0$. Por outro lado, observa-se também que qualquer variação no plano, a partir do ponto mínimo, implica que dx_1 e dx_2 *não são zero*. Matematicamente, essa variação pode ser expressa por:

$$df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 = 0 \tag{2.30}$$

ou

$$df = \left[\frac{\partial f}{\partial x_1}\frac{\partial f}{\partial x_2}\right]\left[\begin{array}{c} dx_1 \\ dx_2 \end{array}\right] = 0 \tag{2.31}$$

A equação (2.31) deve ser satisfeita para todos os pontos do plano. Sabendo-se que $dx_1 \neq 0$ e $dx_2 \neq 0$, obtém-se consequentemente:

$$\frac{\partial f}{\partial x_1} = 0 \;\; ; \;\; \frac{\partial f}{\partial x_2} = 0 \tag{2.32}$$

ou, em outras palavras, *o gradiente de $f(\cdot)$ no ponto mínimo deve ser zero*. Matematicamente:

$$\nabla f(x_1^*, x_2^*) = 0 \tag{2.33}$$

A equação (2.33) representa a condição necessária, ou *condição de primeira ordem* para *problemas irrestritos*. Essa equação é utilizada para identificar as soluções possíveis de um problema de otimização. Considerações adicionais devem ser impostas para assegurar se a solução encontrada pela condição de primeira ordem é de fato ótima, neste caso um mínimo, as quais serão tratadas posteriormente. Portanto, para um problema geral de otimização, as condições necessárias de primeira ordem podem ser expressas por:

$$\nabla f(\boldsymbol{x}^*) = 0 \tag{2.34}$$

A equação (2.34) é utilizada para determinar os valores de $\boldsymbol{x}^*$ tanto analítica quanto numericamente.

Proposição 2.13 (Condições Necessárias de 1a Ordem) *Seja $\Omega \subset \mathbb{R}^n$ e f uma função diferenciável sobre Ω. Se $\boldsymbol{x}^*$ é um mínimo local de f sobre Ω, então tem-se que:*

$$\nabla f(\boldsymbol{x}^*) = 0 \tag{2.35}$$

□

2.4.2 Condições de Segunda Ordem

As condições de segunda ordem são normalmente conhecidas como *condições suficientes*. Como o nome sugere, essas condições envolvem a *derivada segunda* da função. As condições de segunda ordem são obtidas através da expansão de Taylor da função. Se $\boldsymbol{x}^*$ é a solução ótima e $\Delta\boldsymbol{x}$ representa uma variação no ponto solução, a qual resulta em uma variação em Δf, então:

$$\Delta f = f(\boldsymbol{x}^* + \Delta\boldsymbol{x}) - f(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*)^T \Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^T H(\boldsymbol{x}^*)\Delta\boldsymbol{x} \tag{2.36}$$

Para avaliarmos (2.36), Δf deve ser maior do que zero, conforme já observado em (2.29). Aplicando as condições necessárias de primeira ordem (2.35), o primeiro termo do lado direito de (2.36) é zero. Isso resulta na seguinte inequação:

$$\frac{1}{2}\Delta\boldsymbol{x}^T H(\boldsymbol{x}^*)\Delta\boldsymbol{x} > 0 \tag{2.37}$$

onde $H(\boldsymbol{x}^*)$ é a matriz Hessiana da função f no ponto mínimo $\boldsymbol{x}^*$. Para que (2.37) seja verdadeira, a matrix $H(\boldsymbol{x}^*)$ deve ser positiva definida. Há três maneiras para determinar se H é positiva definida:

1. Para todos os valores possíveis de $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{x}^T H(\boldsymbol{x}^*)\Delta\boldsymbol{x} > 0$.

2. Todos os autovalores de $H(\boldsymbol{x}^*)$ devem ser positivos.

3. Os determinantes de todas as submatrizes que envolvem a diagonal principal de $H(\boldsymbol{x}^*)$ devem ser positivos.

Das três condições, apenas as duas últimas podem ser testadas. Isso é discutido a seguir. Porém, a última condição não é tão trivial para ordens elevadas.

Exemplo

Seja o problema de minimização definido por:

$$\begin{aligned} \boldsymbol{x}^* = \ & \arg \min_{\boldsymbol{x}} \ f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2 \\ & \text{sujeito a:} \left\{ \ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \right. \end{aligned} \tag{2.38}$$

As condições necessárias de primeira ordem, equação (2.35), requerem:

$$\frac{\partial f(\cdot)}{\partial x_1} = 2(x_1 - 3) = 0 \tag{2.39}$$

$$\frac{\partial f(\cdot)}{\partial x_2} = 4(x_2 - 3) = 0 \tag{2.40}$$

As equações (2.39) e (2.40) podem ser facilmente resolvidas, resultando na solução $x_1^* = 3$ e $x_2^* = 3$. No ponto solução a função $f(\cdot)$ assume o valor $f(\boldsymbol{x}^*) = 0$ e não há outro ponto na região factível em que $f(\boldsymbol{x}) < 0$. Com isso, conclui-se que as condições necessárias de primeira ordem foram satisfeitas. Entretanto, se a função objetivo $f(\cdot)$ fosse mais complexa, com três ou mais variáveis, e não fosse possível representá-la através de curvas de nível, não poderiamos *a priori*, apenas com base nas condições de primeira ordem, afirmar que o ponto encontrado trata-se do mínimo da função. Afinal, o ponto em questão poderia representar o máximo de $f(\cdot)$ ou um ponto de inflexão, por exemplo um ponto de sela. Portanto, é necessário avaliar as condições de segunda ordem.

As condições de segunda ordem requerem que a matriz Hessiana seja positiva definida, que neste caso pode ser obtida facilmente:

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \tag{2.41}$$

Examinando-se as três maneiras para determinar se H é positiva definida, obtém-se:

1. Não é possível testar todos os $\Delta \boldsymbol{x}$

2. Cálculo dos autovalores de H:

$$\begin{vmatrix} 2 - \lambda_1 & 0 \\ 0 & 4 - \lambda_2 \end{vmatrix} = (2 - \lambda_1)(4 - \lambda_2) = 0$$

Os autovalores são $\lambda_1 = 2$, $\lambda_2 = 4$ e a matriz é positiva definida.

3. Cálculo dos determinantes de todas as submatrizes que envolvam a diagonal principal de H:

   $|2| > 0$

   $\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} = 8 > 0$

   A matriz é positiva definida.

Com isso, conclui-se que as condições de segunda ordem são satisfeitas e que $x_1^* = 3$ e $x_2^* = 3$ é de fato o ponto de mínimo da função. Apresenta-se a seguir as condições necessárias de segunda ordem para um caso geral.

Proposição 2.14 (Condições Necessárias de 2a Ordem) *Seja $\Omega \subset \mathbb{R}^n$ e $f(\cdot)$ uma função duas vezes diferenciável sobre Ω. Se $\boldsymbol{x}^*$ é um mínimo local de $f(\cdot)$ sobre Ω, então tem-se que:*

i. $\nabla f(\boldsymbol{x}^*) = 0$

ii. $H(\boldsymbol{x}^*) > 0$

□

Com esse exemplo, conclui-se a dedução analítica das condições necessárias e suficientes que um ponto deve satisfazer para ser considerado um mínimo de uma função $f(\cdot)$ sem restrições. Apresenta-se, a seguir, a proposição que resume esta seção.

Proposição 2.15 (Otimização Irrestrita) *Seja $f(\cdot) \in C^2$ e $\boldsymbol{x}^* \in \mathbb{R}^n$. Se forem simultaneamente satisfeitas:*

i. $\nabla f(\boldsymbol{x}^*) = 0$

ii. $H(\boldsymbol{x}^*) > 0$

então $\boldsymbol{x}^*$ é um mínimo local estrito de $f(\cdot)$ sobre $\mathbb{R}^n$.

2.5 Condições Analíticas: Problemas com Restrição de Desigualdade

O problema sujeito à restrição de desigualdade pode ser definido como:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$

$$\text{sujeito a:} \begin{cases} g_1(\boldsymbol{x}) : 3x_1 + 2x_2 \leq 12 \\ \\ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \end{cases} \tag{2.42}$$

A Figura 2.8, da seção anterior, ilustra a solução gráfica para este problema com restrição de desigualdade. Pode-se observar que no ponto solução o vetor gradiente da função objetivo e o da função restrição estão na mesma direção e em sentidos opostos. Examinando outros pontos factíveis na Figura 2.8, pode-se afirmar que essa relação entre $\nabla f(\cdot)$ e $\nabla g(\cdot)$ só é verificada no ponto solução. Existe, portanto, no ponto solução uma relação proporcional entre $\nabla f(\cdot)$ e $\nabla g(\cdot)$. Representanto a constante de proporcionalidade por β_1, pode-se expressar a relação entre os gradientes por:

$$\nabla f(\cdot) = -\beta_1 \nabla g_1(\cdot) \quad \text{ou} \quad \nabla f(\cdot) + \beta_1 \nabla g_1(\cdot) = 0 \tag{2.43}$$

A expressão (2.43) pode ser obtida com critérios mais rigorosos usando o método de multiplicadores de Lagrange.

2.5.1 Condições de Primeira Ordem

Método de Lagrange

No método de Lagrange, o problema original (2.42) é transformado com a introdução de uma função Lagrangeana $\overline{f}(\cdot)$, que consite na função objetivo original $f(\cdot)$ somada a restrição de desigualdade $g_1(\cdot)$ ponderada pelo multiplicador β_1. Todavia, a restrição de desigualdade precisa ser tratada como uma restrição de igualdade; para que isso possa ser considerado basta somar uma variável de folga z_1^2 a $g_1(\cdot)$.

Então, um problema composto por uma função objetivo $f(\boldsymbol{x})$, de duas variáveis x_1 e x_2, e apenas uma função de restrição de desigualdade $g_1(\boldsymbol{x})$, similar ao problema original (2.42), poderia ser expresso de maneira genérica por:

$$\begin{aligned} \boldsymbol{x}^* = \ & \arg\min_{\boldsymbol{x}} \ \overline{f}(\boldsymbol{x}, \beta_1, z_1^2) = f(\boldsymbol{x}) + \beta_1[g_1(\boldsymbol{x}) + z_1^2] \\ & \text{sujeito a:} \begin{cases} g_1(\boldsymbol{x}) + z_1^2 = 0 \\ x_1^{min} \leq x_1 \leq x_1^{max} \quad \text{e} \quad x_2^{min} \leq x_2 \leq x_2^{max} \end{cases} \end{aligned} \tag{2.44}$$

Considerando o Lagrangeano em (2.44) como uma função irrestrita, as condições de primeira ordem tem que ser satisfeitas, ou seja:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = \frac{\partial f(\cdot)}{\partial x_1} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_1} = 0 \tag{2.45}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = \frac{\partial f(\cdot)}{\partial x_2} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_2} = 0 \tag{2.46}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial z_1} = 2\beta_1 z_1 = 0 \tag{2.47}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial \beta_1} = g_1(\cdot) + z_1^2 = 0 \tag{2.48}$$

que representa um sistema de quatro equações e quatro incógnitas, x_1, x_2, β_1 e z_1. As equações (2.47) e (2.48) podem ser combinadas, basta multiplicar (2.47) por z_1 e obtém-se $2\beta_1 z_1^2 = 0$. Da expressão (2.48) tem-se que $z_1^2 = -g_1(\cdot)$. Assim, elimina-se a variável de folga z_1 e obtém-se $\beta_1 g_1(\cdot) = 0$. Então o sistema de equações (2.45)–(2.48) pode ser reduzido e reescrito como:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = \frac{\partial f(\cdot)}{\partial x_1} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_1} = 0 \tag{2.49}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = \frac{\partial f(\cdot)}{\partial x_2} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_2} = 0 \tag{2.50}$$

$$\beta_1 g_1(\cdot) = 0 \tag{2.51}$$

Em linhas gerais, as condições de primeira ordem para um problema com restrição de desigualdade podem ser resumidas da seguinte maneira:

$$\nabla f(\cdot) + \beta_1 \nabla g_1(\cdot) = 0 \tag{2.52}$$

$$\beta_1 g_1(\cdot) = 0 \tag{2.53}$$

A generalização para p restrições de desigualdade será apresentada no final do capíulo.

Voltando ao problema exemplo e substituindo os valores no sistema de equações (2.49)–(2.51), obtém-se:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = 2x_1 - 6 + 3\beta_1 = 0 \tag{2.54}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = 4x_2 - 12 + 2\beta_1 = 0 \tag{2.55}$$

$$\beta_1 g_1(\cdot) = \beta_1(3x_1 + 2x_2 - 12) = 0 \tag{2.56}$$

Neste caso, as equações (2.54)–(2.56) representam um sistema de três equações simultâneas e três incógnitas, e soluções não triviais. Para que a equação (2.56) seja satisfeita é necessário que $\beta_1 = 0$ ou que $g_1(\cdot) = 0$. O sistema de equações simultâneas (2.54)–(2.56) requer que as condições no multiplicador β_1 e na restrição $g_1(\cdot)$ sejam satisfeitas simultaneamente. Isso resulta nos seguintes casos:

1. Caso a: $\beta_1 = 0$ $[g_1 < 0]$
2. Caso b: $\beta_1 \neq 0$ $[g_1 = 0]$

Observa-se claramente, no caso b, que se $\beta_1 \neq 0$, correspondendo a $g_1 = 0$, então a restrição $g_1(\cdot)$ transforma-se em uma igualdade.

O sistema de equações (2.54)–(2.56) deve ser resolvido considerando-se os casos a e b. Fica por conta do leitor mostrar que, no exemplo em questão, o caso a leva a uma solução inviável. Para o caso b, obtém-se $x_1^* = 2.18$, $x_2^* = 2.73$ e $\beta_1^* = 0.55$. Observando a Figura 2.8, solução geométrica obtida na seção anterior, nota-se que a solução analítica encontrada usando o multiplicador de Lagrange indica que é necessário multiplicar o $\nabla g_1(\cdot)$ por $\beta_1 = 0.55$ para que o somatório de $\nabla f(\cdot)$ e $\nabla g_1(\cdot)$ seja zero no ponto solução.

Como no caso irrestrito da seção anterior, é preciso investigar as condições de segunda ordem, ou condições suficientes, para ter segurança em relação à solução encontrada pelas condições necessárias.

2.5.2 Condições de Segunda Ordem

No ponto determinado pelas condições de primeira ordem, qualquer variação $\boldsymbol{\Delta x}$ implicará um aumento na função $f(\cdot)$. Variações $\boldsymbol{\Delta x}$ não são arbitrárias. Elas tem que satisfazer as restrições de desigualdade no ponto solução. As condições de segunda ordem requerem que as equações seguintes sejam satisfeitas:

$$\Delta\overline{f}(\cdot) = \overline{f}(\boldsymbol{x}^* + \Delta\boldsymbol{x}) - \overline{f}(\boldsymbol{x}^*) = \nabla\overline{f}(\boldsymbol{x}^*)^T\Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^T[\overline{H}(\boldsymbol{x}^*)]\Delta\boldsymbol{x} > 0 \tag{2.57}$$

$$\nabla g_1(\cdot)^T\Delta\boldsymbol{x} = 0 \tag{2.58}$$

Na equação (2.57) o termo $[H(\boldsymbol{x}^*)]$ representa a matriz Hessiana da função Lagrangeana $\overline{f}$ calculada no ponto solução. As condições de primeira ordem requerem que $\nabla\overline{f}(\boldsymbol{x}^*) = 0$. Considerando um problema com duas variáveis e uma restrição de desigualdade, tem-se:

$$\Delta\overline{f}(\cdot) = \frac{1}{2}\left[\frac{\partial^2\overline{f}(\cdot)}{\partial x_1^2}(\Delta x_1)^2 + 2\frac{\partial^2\overline{f}(\cdot)}{\partial x_1\partial x_2}(\Delta x_1)(\Delta x_2) + \frac{\partial^2\overline{f}(\cdot)}{\partial x_2^2}(\Delta x_2)^2\right] > 0 \tag{2.59}$$

rearranjando os termos de (2.59) resulta em:

$$\Delta\overline{f}(\cdot) = \frac{1}{2}\left[\frac{\partial^2\overline{f}(\cdot)}{\partial x_1^2}(\frac{\Delta x_1}{\Delta x_2})^2 + 2\frac{\partial^2\overline{f}(\cdot)}{\partial x_1\partial x_2}(\frac{\Delta x_1}{\Delta x_2}) + \frac{\partial^2\overline{f}(\cdot)}{\partial x_2^2}\right](\Delta x_2)^2 > 0 \tag{2.60}$$

A equação (2.58) pode ser expressa por:

$$\frac{\Delta x_1}{\Delta x_2} = -\frac{\partial g_1(\cdot)/\partial x_2}{\partial g_1(\cdot)/\partial x_1} \tag{2.61}$$

Substituindo a equação (2.61) na equação (2.60) obtém-se:

$$\Delta\overline{f}(\cdot) = \frac{1}{2}\left[\frac{\partial^2\overline{f}(\cdot)}{\partial x_1^2}(\frac{\partial g_1(\cdot)/\partial x_2}{\partial g_1(\cdot)/\partial x_1})^2 - 2\frac{\partial^2\overline{f}(\cdot)}{\partial x_1\partial x_2}(\frac{\partial g_1(\cdot)/\partial x_2}{\partial g_1(\cdot)/\partial x_1}) + \frac{\partial^2\overline{f}(\cdot)}{\partial x_2^2}\right](\Delta x_2)^2 > 0 \tag{2.62}$$

que representa a condição de segunda ordem, ou condição suficiente, que deve ser satisfeita.

A equação (2.62) deve ser avaliada no ponto de mínimo encontrado pela condição de primeira ordem. Para assegurar que $\Delta\overline{f}(\cdot) > 0$, deve-se avaliar a expressão entre colchetes, já que o termo $(\Delta x_2)^2$ será sempre positivo. Para o problema definido em

(2.42), avaliando (2.62) obtém-se $[\,\cdot\,] = 44/9 > 0$. Com isso, conclui-se que o ponto $x_1^* = 2.18$, $x_2^* = 2.73$ é de fato o mínimo do problema.

Cabe destacar que a equação (2.62) não é fácil de ser avaliada, particularmente se o problema de otimização for mais complexo, por exemplo envolvendo mais do que duas variáveis e duas restrições, ou contendo restrições quadráticas ou de ordem superior. Do ponto de vista prático, como veremos nos capítulos seguintes, as condições de segunda ordem não serão utilizadas nos algoritmos.

Para concluir esta seção, apresenta-se a seguir a proposição que resume as condições de primeira e segunda ordem que devem ser satisfeitas para assegurar a solução de um problema de otimização com restrição de desigualdade.

Proposição 2.16 (Otimização Restrita – Desigualdade) *Sejam* $f(\cdot) \in C^2$ *e* $g_i(\cdot) \in C^2$, $i = 1, \ldots, p$, *e* $\boldsymbol{x}^*$ *tal que* $g_i(\boldsymbol{x}^*) \leq 0$, $i = 1, \ldots, p$. *Se existem multiplicadores* $\beta_1, \beta_2, \ldots, \beta_p$ *tais que*

i. $\beta_i \geq 0\ ,\ i = 1, \ldots, p$

ii. $\beta_i g_i(\boldsymbol{x}^*) = 0\ ,\ i = 1, \ldots, p$

iii. $\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{p} \beta_i \nabla g_i(\boldsymbol{x}^*) = 0$

iv. $H(\boldsymbol{x}^*) + \sum_{i=1}^{p} \beta_i G_i(\boldsymbol{x}^*) > 0$ *sobre* $M = \{\boldsymbol{y} \in \mathbb{R}^n : \nabla g_i(\boldsymbol{x}^*)'\boldsymbol{y} = 0\ ,\ i \in I(\boldsymbol{x}^*)\}$,
$I(\boldsymbol{x}^*) = \{i\ :\ g_i(\boldsymbol{x}^*) = 0\ ,\ \beta_i > 0\}$

são simultaneamente satisfeitos, então $\boldsymbol{x}^*$ *é um mínimo local estrito de* $f(\cdot)$ *sobre* $g_i(\boldsymbol{x}) \leq 0\ ,\ i = 1, \ldots, p$.

2.6 Condições Analíticas: Problemas com Restrição de Igualdade

O problema sujeito à restrição de igualdade pode ser definido como:

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}}\ f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$

$$\text{sujeito a:} \begin{cases} h_1(\boldsymbol{x}) : x_1 + x_2 = 5 \\ \\ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \end{cases} \tag{2.63}$$

A Figura 2.26, da seção anterior, ilustra a solução gráfica para este problema. De maneira similar ao caso anterior, observa-se que no ponto solução, o vetor gradiente da função objetivo está na mesma direção e no sentido oposto do vetor gradiente da função restrição. Examinando outros pontos factíveis na Figura 2.26, i.e., outros pontos sobre a reta $h(\cdot)$, uma vez que a restrição de igualdade tem que ser satisfeita, pode-se afirmar que essa relação entre $\nabla f(\cdot)$ e $\nabla h(\cdot)$ só é possível no ponto solução. Existe, portanto, no ponto solução uma relação proporcional entre $\nabla f(\cdot)$ e $\nabla h(\cdot)$.

Representando a constante de proporcionalidade por λ_1, pode-se expressar a relação entre os gradientes por:

$$\nabla f(\cdot) = -\lambda_1 \nabla h_1(\cdot) \quad \text{ou} \quad \nabla f(\cdot) + \lambda_1 \nabla h_1(\cdot) = 0 \tag{2.64}$$

Como discutido na seção anterior, a equação (2.64) pode ser obtida, com critérios matemáticos mais rigorosos, usando o método de multiplicadores de Lagrange.

2.6.1 Condições de Primeira Ordem

Método de Lagrange

Neste caso a função Lagrangeana consite na função objetivo original $f(\cdot)$ somada a restrição de igualdade $h_1(\cdot)$ ponderada pelo multiplicador λ_1 (para diferenciar do multiplicador da função de desigualdade). Então, um problema de composto por uma função objetivo $f(\boldsymbol{x})$, de duas variáveis x_1 e x_2, e apenas uma função de restrição de igualdade $h_1(\boldsymbol{x})$, similar ao problema original (2.63), poderia ser expresso de maneira genérica por:

$$\begin{aligned} &\boldsymbol{x}^* = \text{ arg } \min_{\boldsymbol{x}} \ \overline{f}(\boldsymbol{x}, \lambda_1) = f(\boldsymbol{x}) + \lambda_1 h_1(\boldsymbol{x}) \\ &\text{sujeito a:} \begin{cases} h_1(\boldsymbol{x}) = 0 \\ \\ x_1^{min} \leq x_1 \leq x_1^{max} \text{ e } x_2^{min} \leq x_2 \leq x_2^{max} \end{cases} \end{aligned} \tag{2.65}$$

Considerando o Lagrangeano em (2.65) como uma função irrestrita de duas variáveis, as condições de primeira ordem tem que ser satisfeitas, ou seja:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = \frac{\partial f(\cdot)}{\partial x_1} + \lambda_1 \frac{\partial h_1(\cdot)}{\partial x_1} = 0 \tag{2.66}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = \frac{\partial f(\cdot)}{\partial x_2} + \lambda_1 \frac{\partial h_1(\cdot)}{\partial x_2} = 0 \tag{2.67}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial \lambda_1} = h_1(\cdot) = 0 \tag{2.68}$$

As equações (2.66)–(2.68) expressam as condições de primeira ordem, ou condições necessárias, para o problema definido em (2.65). Essas expressões constituem um sistema de três equações e três incógnitas, x_1, x_2 e λ_1. A equação (2.68) representa a restrição de igualdade.

Em linhas gerais, as condições de primeira ordem para um problema com restrição de igualdade podem ser resumidas da seguinte maneira:

$$\nabla f(\cdot) + \lambda_1 \nabla h_1(\cdot) = 0 \tag{2.69}$$

$$h_1(\cdot) = 0 \tag{2.70}$$

A generalização para q restrições de igualdade será apresentada no final do capítulo.

Substituindo os valores do problema original, obtém-se:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = 2x_1 - 6 + \lambda_1 = 0 \tag{2.71}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = 4x_2 - 12 + \lambda_1 = 0 \tag{2.72}$$

$$h_1(\cdot) = x_1 + x_2 = 5 \tag{2.73}$$

As equações (2.71)–(2.73) representam um sistema de três equações e três incógnitas, cuja solução resulta em $x_1^* = 2,33$, $x_2^* = 2,67$ e $\lambda_1^* = 1,34$. Observando a Figura 2.9, nota-se que a solução analítica encontrada usando o multiplicador de Lagrange coincide com a solução geométrica obtida na seção anterior. Como no caso irrestrito da seção anterior, é preciso investigar as condições de segunda ordem, ou condições suficientes, para ter segurança em relação à solução encontrada pelas condições necessárias. Essa análise é similar a desenvolvida na seção anterior (deixada como exercício para o leitor) e leva a conclusão que a solução encontrada é de fato o mínimo do problema.

Antes de concluir a seção, analisaremos uma interpretação geométrica associada à introdução dos multiplicadores de Lagrange.

2.6.2 Interpretação Geométrica dos Multiplicadores de Lagrange

A interpretação geométrica dos multiplicadores de Lagrange indica que, no ponto solução, λ expressa a relação da variação da função objetivo em relação à variação da restrição. Consideremos:

$$\overline{f}(\cdot) = f(\cdot) + \lambda_1 h_1(\cdot) \tag{2.74}$$

Diferenciando obtém-se

$$d\overline{f}(\cdot) = df(\cdot) + \lambda_1 dh_1(\cdot) \tag{2.75}$$

que resulta em

$$d\overline{f}(\cdot) = \frac{\partial \overline{f}(\cdot)}{\partial x_1}dx_1 + \frac{\partial \overline{f}(\cdot)}{\partial x_2}dx_2 + \frac{\partial \overline{f}(\cdot)}{\partial \lambda_1}d\lambda_1 \tag{2.76}$$

No ponto solução, as condições de primeira ordem requerem que $d\overline{f}(\cdot) = 0$. Consequentemente,

$$\lambda_1 = -\frac{df(\cdot)}{dh(\cdot)} = -\frac{\Delta f(\cdot)}{\Delta h(\cdot)} \tag{2.77}$$

A equação (2.77) indica claramente, como queria-se demonstrar, que o multiplicador de Lagrange λ_1 é a relação entre a variação da função objetivo Δf e a variação da função restrição de igualdade Δh. Essa relação não afeta a determinação da solução ótima, tendo, entretanto, um papel importante na discussão de *análise de sensibilidade*.

Para concluir esta seção, apresenta-se a seguir a proposição que resume as condições de primeira e segunda ordem que devem ser satisfeitas para assegurar a solução de um problema de otimização com restrições de igualdade.

Proposição 2.17 (Otimização Restrita – Igualdade) *Sejam* $f(\cdot) \in C^2$ *e* $h_j(\cdot) \in C^2$, $j = 1, \ldots, q$ *e* $\boldsymbol{x}^*$ *tal que* $h_j(\boldsymbol{x}^*) = 0$, $j = 1, \ldots, q$. *Se existem multiplicadores* $\lambda_1, \lambda_2, \ldots, \lambda_q$ *tais que*

i. $\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\boldsymbol{x}^*) = 0$

ii. $H(\boldsymbol{x}^*) + \sum_{j=1}^{q} \lambda_j H_j(\boldsymbol{x}^*) > 0$ *sobre* $M = \{\boldsymbol{y} \in \mathbb{R}^n \ : \ \nabla h_j(\boldsymbol{x}^*)'\boldsymbol{y} = 0 \ , \ j = 1, \ldots, q\}$

são simultaneamente satisfeitos, então $\boldsymbol{x}^*$ *é um mínimo local estrito de* $f(\cdot)$ *sujeito a* $h_j(x) = 0$, $j = 1, \ldots, q$.

2.7 O Problema Geral de Otimização

O problema geral de otimização é definido incluindo simultaneamente as restrições de desigualdade e igualdade:

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} \ f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$

$$\text{sujeito a:} \begin{cases} g_1(\boldsymbol{x}) : 3x_1 + 2x_2 \leq 12 \\ h_1(\boldsymbol{x}) : x_1 + x_2 = 5 \\ 0 \leq x_1 \leq 6; \quad 0 \leq x_2 \leq 6 \end{cases} \tag{2.78}$$

Usando o método de multiplicadores de Lagrange, um problema de otimização similar ao apresentado em (2.78), composto por uma função objetivo $f(\boldsymbol{x})$, de duas variáveis x_1 e x_2, e apenas uma função de restrição de desigualdade $g_1(\boldsymbol{x})$ e uma restrição de igualdade $h_1(\boldsymbol{x})$, pode ser expresso de maneira genérica por:

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} \ \overline{f}(\boldsymbol{x}, \beta_1, \lambda_1, z_1^2) = f(\boldsymbol{x}) + \beta_1[g_1(\boldsymbol{x}) + z_1^2] + \lambda_1 h_1(\boldsymbol{x})$$

$$\text{sujeito a:} \begin{cases} g_1(\boldsymbol{x}) + z_1^2 = 0 \\ h_1(\boldsymbol{x}) = 0 \\ x_1^{min} \leq x_1 \leq x_1^{max} \text{ e } x_2^{min} \leq x_2 \leq x_2^{max} \end{cases} \tag{2.79}$$

Considerando a função Lagrangeana $\overline{f}(\cdot)$ como uma função irrestrita, as condições de primeira ordem para este caso podem ser obtidas por:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = \frac{\partial f(\cdot)}{\partial x_1} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_1} + \lambda_1 \frac{\partial h_1(\cdot)}{\partial x_1} = 0 \tag{2.80}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = \frac{\partial f(\cdot)}{\partial x_2} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_2} + \lambda_1 \frac{\partial h_1(\cdot)}{\partial x_2} = 0 \tag{2.81}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial z_1} = 2\beta_1 z_1 = 0 \tag{2.82}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial \beta_1} = g_1(\cdot) + z_1^2 = 0 \tag{2.83}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial \lambda_1} = h_1(\cdot) = 0 \tag{2.84}$$

que são uma associação das subseções anteriores referentes a discussão das condições de primeira ordem para problemas com restrições de desigualdade e igualdade. Como discutido no caso do problema com restrições de desigualdade, as equações (2.82) e (2.83) podem ser combinadas eliminando-se a variável z_1. Assim, as equações (2.80) a (2.84) podem ser reduzidas a quatro expressões:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = \frac{\partial f(\cdot)}{\partial x_1} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_1} + \lambda_1 \frac{\partial h_1(\cdot)}{\partial x_1} = 0 \tag{2.85}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = \frac{\partial f(\cdot)}{\partial x_2} + \beta_1 \frac{\partial g_1(\cdot)}{\partial x_2} + \lambda_1 \frac{\partial h_1(\cdot)}{\partial x_2} = 0 \tag{2.86}$$

$$\beta_1 g_1(\cdot) = 0 \tag{2.87}$$

$$h_1(\cdot) = 0 \tag{2.88}$$

A equação (2.87) requer que duas possibilidades sejam testadas:

1. Caso a: $\beta_1 = 0$ $[g_1 < 0]$
2. Caso b: $\beta_1 \neq 0$ $[g_1 = 0]$

Assim, as equações (2.85)–(2.88) correspondem a um sistema de quatro equações e quatro incógnitas, x_1, x_2, β_1 e λ_1, e representam as condições de primeira ordem que devem ser satisfeitas para um problema com uma restrição de desigualdade e uma restrição de igualdade.

Em linhas gerais, as condições de primeira ordem podem ser resumidas da seguinte maneira:

$$\nabla f(\cdot) + \beta_1 \nabla g_1(\cdot) + \lambda_1 \nabla h_1(\cdot) = 0 \tag{2.89}$$

$$\beta_1 g_1(\cdot) = 0 \tag{2.90}$$

$$h_1(\cdot) = 0 \tag{2.91}$$

A generalização para p restrições de desigualdade e q restrições de igualdade será apresentada no final da seção.

Voltando ao problema exemplo, as condições de primeira ordem resultam em:

$$\frac{\partial \overline{f}(\cdot)}{\partial x_1} = 2x_1 - 6 + \lambda_1 + 3\beta_1 = 0 \tag{2.92}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial x_2} = 4x_2 - 12 + \lambda_1 + 2\beta_1 = 0 \tag{2.93}$$

$$\beta_1 g_1(\cdot) = \beta_1(3x_1 + 2x_2 - 12) = 0 \tag{2.94}$$

$$\frac{\partial \overline{f}(\cdot)}{\partial \lambda_1} = x_1 + x_2 - 5 = 0 \tag{2.95}$$

Duas soluções devem ser examinadas. A primeira, Caso a ($\beta_1 = 0$ e $[g_1 < 0]$), requer a solução de um sistema de três equações e três incógnitas, x_1, x_2 e λ_1, já que a equação (2.94) é eliminada. A segunda, Caso b ($\beta_1 \neq 0$ $[g_1 = 0]$), é um sistema de quatro equações e quatro incógnitas, x_1, x_2, β_1 e λ_1, envolvendo as equações (2.92)–(2.95). A solução do caso a fornece $x_1 = 7/3$, $x_2 = 8/3$ e $\lambda_1 = 4/3$, porém a restrição de desigualdade é violada, inviabilizando a solução encontrada. Resolvendo-se a segunda opção obtém-se $x_1 = 2$, $x_2 = 3$, $\beta_1 = 2$ e $\lambda_1 = -4$.

A Figura 2.12, a seguir, ilustra a solução gráfica e analítica do caso geral. Observe que no ponto solução ($x_1 = 2$; $x_2 = 3$) o somatório do gradiente da função objetivo e das funções de restrição, ponderadas pelos respectivos multiplicadores de Lagrange, se anula. Em outras palavras, $\nabla f(\cdot) + \beta_1 g_1(\cdot) + \lambda_1 h_1(\cdot) = 0$.

2.7.1 Condições de Karush-Kuhn-Tucker

Para concluir esta seção, apresenta-se a seguir a proposição conhecida como condições de Karush-Kuhn-Tucker, as quais devem ser satisfeitas para assegurar a solução de um problema geral de otimização.

Proposição 2.18 (Condições Necessárias de Karush-Kuhn-Tucker para Otimalidade) *Seja $\boldsymbol{x}^*$ um ponto regular das restrições do problema de otimização:*

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\textit{sujeito a: } \begin{cases} g_i(\boldsymbol{x}) \leq 0, \ i = 1, \ldots, p \\ h_j(\boldsymbol{x}) = 0, \ j = 1, \ldots, q \end{cases} \tag{2.96}$$

sendo que $f(\cdot)$, $g(\cdot)$, $h(\cdot) \in C^1$. Para $\boldsymbol{x}^*$ ser um ótimo local do problema, deve existir um conjunto de multiplicadores de Karush-Kuhn-Tucker $\boldsymbol{\beta}^* \in \mathbb{R}^p$ e $\boldsymbol{\lambda}^* \in \mathbb{R}^q$, com

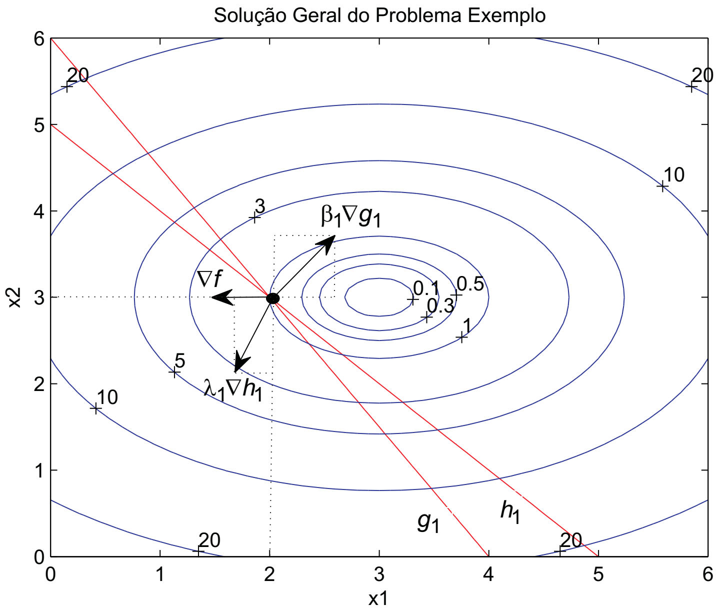

Figura 2.12: Solução analítica do caso geral.

$\beta_i^* \geq 0$ tal que:

$$
\begin{aligned}
&\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{p} \beta_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{q} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = 0 \\
&\beta_i \geq 0 \quad e \quad \beta_i^* g_i(\boldsymbol{x}^*) = 0 \quad \forall\ i = 1, \ldots, p \\
&h_j(\boldsymbol{x}) = 0 \quad \forall\ j = 1, \ldots, q
\end{aligned}
\tag{2.97}
$$

sendo que $\boldsymbol{x} \in \mathbb{R}^n$, $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^1$, $g_i(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^p$ e $h_j(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^q$. □

2.8 Exercícios

1. Conceitue um mínimo local, mínimo global e direções viáveis [3].

2. Considere a função $f(x) = xe^{-2x}$. Calcular todos os pontos de mínimo e máximo local, e também os pontos de inflexão. O que pode ser afirmado a respeito do ponto de mínimo e máximo global dessa função. Justifique analiticamente a resposta [4].

3. Calcular a derivada primeira e segunda da função, definida a seguir, em $x = 0$ [5]:
$$f(\boldsymbol{x}) = x_1^4 + x_1x_2 + (1 + x_2)^2 \tag{2.98}$$
Mostre que $H(\mathbf{0})$ não é positiva definida. Verifique que o mínimo local é $\boldsymbol{x}^* = (0.6959; -1.3479)^T$.

4. Calcular os pontos estacionários da função [5]:
$$f(\boldsymbol{x}) = 2x_1^3 - 3x_1^2 - 6x_1x_2(x_1 - x_2 - 1) \tag{2.99}$$
Quais pontos são mínimos locais e quais são máximos locais?

5. Mostre que a função $f(\boldsymbol{x}) = (x_2 - x_1^2)^2 + x_1^5$ possui apenas um ponto estacionário que não é máximo local ou mínimo local [5].

6. O problema [3]:
$$\text{minimize } f(\boldsymbol{x}) = 2x_1^2 + x_1x_2 + x_2^2 + x_2x_3 + x_3^2 - 6x_1 - 7x_2 - 8x_3 + 9 \tag{2.100}$$
possui um ponto de mínimo local em $\boldsymbol{x}^* = (6/5, 6/5, 17/5)^T$. Verifique se as condições necessárias para um mínimo local são satisfeistas nesse ponto. Esse mínimo local é também um ponto de mínimo global?

7. O problema [5], [3]:
$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\ &\text{sujeito a: } g(\boldsymbol{x}) : (x_1^2 + x_2^2) \leq 2 \end{aligned} \tag{2.101}$$
possui um ponto de mínimo local em $\boldsymbol{x}^* = (1, 1)^T$. Verifique se as condições necessárias para um mínimo local são satisfeistas nesse ponto. Mostre que $H(\boldsymbol{x})$ é singular se, e somente se, $\boldsymbol{x}$ satisfizer a condição: $x_2 - x_1^2 = 0.005$.

8. Considere o problema [4]:
$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}) = x_1^2 + 2x_2^2 \\ &\text{sujeito a: } h(\boldsymbol{x}) : (x_1 + x_2 - 2) = 0 \end{aligned} \tag{2.102}$$

Encontrar um ponto que satisfaça as condições de Karush-Kuhn-Tucker e verificar se esse ponto é a solução ótima. Resolva o problema novamente substituindo a função objetivo por $f(\boldsymbol{x}) = x_1^3 + x_2^3$.

9. Considere o problema [4]:

$$\text{minimize } f(\boldsymbol{x}) = x_1^4 + x_2^4 + 12x_1^2 + 6x_2^2 - x_1x_2 - x_1 - x_2$$

$$\text{sujeito a: } \begin{cases} g_1(\boldsymbol{x}) : x_1 + x_2 \geq 6 \\ g_2(\boldsymbol{x}) : 2x_1 - x_2 \geq 3 \\ x_1 \geq 0; \;\; x_2 \geq 0 \end{cases} \tag{2.103}$$

Escreva as expressões para as condições de Karush-Kuhn-Tucker e mostre que $(x_1, x_2) = (3, 3)$ é o único ponto solução.

10. Suponha o problema de uma variável: max x^2, com $-1 \leq x \leq 2$. Mostre que as condições de Karush-Kuhn-Tucker são satisfeitas nesse problema em $x = 1$, $x = 0$ e $x = 2$, embora o único ponto de ótimo global seja $x = 2$.

11. Resolver graficamente os seguintes problemas:

(i)

$$\text{maximize } f(\boldsymbol{x}) = x_1$$

$$\text{sujeito a: } \begin{cases} g_1(\boldsymbol{x}) : (1 - x_1)^3 - x_2 \geq 0 \\ g_2(\boldsymbol{x}) : x_1 \geq 0 \\ g_3(\boldsymbol{x}) : x_2 \geq 0 \end{cases} \tag{2.104}$$

(ii)

$$\text{minimize } f(\boldsymbol{x}) = x_1^2 + 4x_2^2$$

$$\text{sujeito a: } \begin{cases} g_1(\boldsymbol{x}) : x_1 + x_2 \geq 0 \\ g_2(\boldsymbol{x}) : 2x_1 + x_2 \geq 0 \end{cases} \tag{2.105}$$

(iii)

$$\text{minimize } f(\boldsymbol{x}) = -x_1x_2$$

$$\text{sujeito a: } \begin{cases} h_1(\boldsymbol{x}) : 20x_1 + 15x_2 - 30 = 0 \\ g_1(\boldsymbol{x}) : x_1^2/4 + x_2^2 - 1 \leq 0 \\ 0 \leq x_1 \leq 3; \;\; 0 \leq x_2 \leq 3 \end{cases} \tag{2.106}$$

Para cada um dos gráficos, desenhe as direções dos vetores gradiente da função objetivo e das restrições ativas no ponto ótimo. Verifique se as condições de Karush-Kuhn-Tucker são satisfeitas no ponto solução.

Referências Bibliográficas

[1] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 2 edition, 1989.

[2] P. Venkataraman. *Applied Optimization with Matlab Programming*. John Wiley, 1 edition, 2002.

[3] G.R. Mateus e H.P.L. Luna. *Programação Não-Linear*. V Escola de Computação, 1 edition, 1986.

[4] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley, 3 edition, 2006.

[5] R. Fletcher. *Practical Methods of Optimization*. John Wiley, 2 edition, 1987.